

AD-A267 453



RL-TR-93-79
Final Technical Report
May 1993

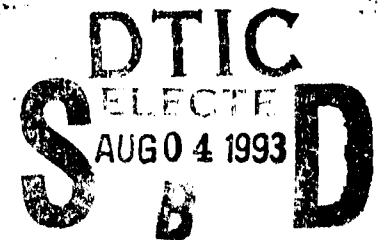


2

SIGNAL DETECTION IN CORRELATED GAUSSIAN AND NON-GAUSSIAN RADAR CLUTTER

Kaman Sciences Corporation

M. Rangaswamy, P. Chakravarthi, Dr. D. Weiner, Dr. L. Cai,
Dr. H. Wang, Dr. A. Ozturk



APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

93-17391



278 Pg

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

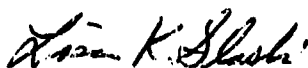
93 8 3

098

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-93-79 has been reviewed and is approved for publication.

APPROVED:



LISA K. SLASKI
Project Engineer

FOR THE COMMANDER



JAMES W. YOUNGBERG, Lt Col, USAF
Deputy Director
Surveillance and Photonics Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (OCTS) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE May 1993		3. REPORT TYPE AND DATES COVERED Final Aug 90 - Dec 91
4. TITLE AND SUBTITLE SIGNAL DETECTION IN CORRELATED GAUSSIAN AND NON-GAUSSIAN RADAR CLUTTER			5. FUNDING NUMBERS C - F30602-89-C-0082, PE - 63741D Task 10 PR - 3640 TA - 06 WU - 05	
6. AUTHOR(S) M. Rangaswamy, P. Chakravarthi, Dr D. Weiner, Dr L. Cai, Dr H. Wang, Dr A. Ozturk				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) KAMAN Sciences Corporation 258 Genesee Street, Suite 103 Utica NY 13502			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Rome Laboratory (OCTS) 26 Electronic Pky Griffiss AFB NY 13441-4514			10. SPONSORING/MONITORING AGENCY REPORT NUMBER RL-TR-93-79	
11. SUPPLEMENTARY NOTES Rome Laboratory Project Engineer: Lisa Slaski/OCTS/(315) 330-4437 Authors were employees of Syracuse University, Syracuse NY, under contract with KAMAN Sciences Corporation. Dr Ozturk was a visiting professor at Syracuse University from EGE University, Izmir, Turkey.				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The subject of this report is the detection of weak targets in a strong clutter environment. Two situations arise depending on whether or not the weak targets can be separated from the clutter. For both cases new receivers are derived which provide significant improvement in performance over other recently proposed techniques. This work includes development of an adaptive joint-domain space-time processor, effective non-gaussian weak signal detectors based on spherically invariant random processes, and a new method for approximating the underlying probability density function of random data which works extremely well with only 100 samples.				
14. SUBJECT TERMS Locally Optimum Detector, Spherically Invariant, Random Processes, Probability Density Function, Weak Signal Detector, Radar, Space-Time Processing, Clutter, Non-Gaussian			15. NUMBER OF PAGES 284	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Executive Summary

The subject of this report is the detection of weak targets in a strong clutter environment. Two situations arise depending upon whether or not the weak targets can be separated from the clutter. For both cases new receivers are derived which provide significant improvement in performance over other recently proposed techniques. This work includes development of an adaptive joint-domain space-time processor, effective non-Gaussian weak signal detectors based on spherically invariant random processes, and a new method for approximating the underlying probability density function of random data which works extremely well with only 100 samples.

When the target and clutter are separable, space-time processing is effective in detecting the target. In effect, this approach maximizes the signal-to-clutter ratio by using two-dimensional filters on the joint spatial and Doppler spectra to isolate the target from the clutter. Furthermore, for Gaussian clutter, space-time processing is the optimum approach for detecting weak targets in a strong clutter background whether or not the targets and clutter are separable.

Unfortunately, when the target and clutter spectra completely overlap, space-time processing is ineffective in detecting weak targets. Nothing can be done to improve performance for the Gaussian clutter case. However, for non-Gaussian clutter, effective weak signal detectors do exist. Nevertheless, this is an area which, in spite of its importance, has received relatively little attention. Much of this report is devoted to

1. the characterization, generation, and approximation of correlated non-Gaussian radar clutter samples and

2. the design and performance of the corresponding weak signal detectors.

Many new and significant results are discussed in this report and are summarized below:

- (1) An adaptive joint-domain space-time processor is derived which not only outperforms currently proposed space-time processors but also converges more rapidly and processes data more efficiently.
- (2) Spherically invariant random processes (SIRPs) are shown to be an attractive approach to the extremely difficult problem of modeling correlated non-Gaussian random variables. Many useful and desirable properties of SIRPs are derived in a straight-forward tutorial manner.
- (3) To make it possible to model many different types of correlated non-Gaussian clutter (e.g.- Weibull, K-distributed, Rician, etc.) an extensive library of SIRPs is developed.
- (4) To enable computer simulation of correlated non-Gaussian radar clutter samples, which are needed for evaluating receiver performance, two different canonical generation schemes are derived.
- (5) Since the probability distribution underlying clutter is not likely to be known in advance, a new method for approximating the univariate probability density function of random data is developed which outperforms existing techniques while using significantly fewer data samples.
- (6) To approximate the probability distribution underlying the N correlated non-Gaussian radar returns received during a coherent processing interval, the technique developed in item 5 is extended in a simple manner to the multivariate probability density function arising from spherically invariant random processes.

- (7) Weak signal receivers, known as locally optimum detectors, are derived for correlated non-Gaussian clutter that can be approximated by SIRPs. These detectors are shown to be canonical in form and combine the conventional Gaussian receiver with the appropriate nonlinearity.
- (8) Because the locally optimum detectors are nonlinear and involve non-Gaussian inputs, their performance must be evaluated by Monte Carlo simulation. A technique is developed for determining the receiver thresholds that reduces by several orders of magnitude the number of Monte Carlo trials required.
- (9) The locally optimum detector for multivariate Student-T clutter is shown to significantly outperform the conventional Gaussian receiver when the target and clutter spectra completely overlap.

DTIC QUALITY INSPECTED 3

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Contents

1	Introduction	1
1.1	Adaptive Implementation of Optimum Space-Time Processing	2
1.2	Weak Signal Detection	3
1.3	Literature Review on Spherically Invariant Random Processes	3
1.4	Radar Clutter Modeling Using SIRPs	3
1.5	Computer Generation of Simulated Radar Clutter Characterized as SIRPs	4
1.6	A New Method for Univariate Distribution Approximation	4
1.7	Distribution Approximation of Radar Clutter by SIRPs	4
1.8	Weak Signal Detection	5
1.9	The Locally Optimum Detector	5
1.10	Determining Thresholds for the Locally Optimum Detector	5
1.11	Performance of the LOD for the Multivariate Student-T Distribution	5
2	Adaptive Implementation of Optimum Space-Time Processing	6
2.1	Introduction	6
2.2	Data Modeling	8
2.3	Difference among The Performance Potentials of The Cascade and Joint-Do main Processors	10
2.4	The Joint-Domain Localized GLR Algorithm	17
2.4.1	The JDL-GLR Principle	18
2.4.2	The JDL-GLR Detection Performance	22
2.4.3	Detection Performance Comparison	23
2.4.4	Other Features of JDL-GLR	25
2.5	Conclusions and Discussion	27

3	Literature Review on Spherically Invariant Random Processes	28
3.1	Introduction	28
3.2	Definitions	29
3.3	Characterization of SIRPs	30
3.4	Determining the PDF of an SIRV	34
3.5	Properties of SIRVs	38
3.5.1	PDF Characterization	38
3.5.2	Closure Under Linear Transformation	38
3.5.3	Minimum Mean Square Error Estimation	38
3.5.4	Distribution of Sums of SIRVs	42
3.5.5	Markov Property for SIRPs	43
3.5.6	Kalman Filter for SIRPs	45
3.5.7	Statistical Independence	47
3.5.8	Ergodicity of SIRPs	47
3.6	Conclusion	47
4	Radar Clutter Modeling Using Spherically Invariant Random Processes	49
4.1	Introduction	49
4.2	Problem Statement	51
4.3	Techniques for Determining the SIRV PDF	54
4.3.1	SIRVs with Known Characteristic PDF	55
4.3.2	SIRVs with Unknown Characteristic PDFs	55
4.3.3	Hankel Transform Approach	56
4.4	Examples of complex SIRVs	59
4.4.1	Examples Based on the Characteristic PDF	59
4.4.2	Examples Based on Marginal Envelope PDF	71
4.4.3	Examples Using the Marginal Characteristic Function	89
4.5	Significance of the Quadratic form of the SIRV PDF	98
4.6	Conclusion	109
5	Computer Generation of Simulated Radar Clutter Characterized as SIRPs	110
5.1	Introduction	110

5.2	Preliminaries	111
5.3	Two Canonical Simulation Procedures for Generating SIRVs	115
5.3.1	Simulation Procedure for SIRVs with Known Characteristic PDF	115
5.3.2	Simulation Scheme for SIRVs with Unknown Characteristic PDF	117
5.4	Performance Assessment of the Simulation Schemes	121
5.5	Conclusions	122
6	A New Method for Univariate Distribution Approximation	131
6.1	Introduction	131
6.2	Definitions	132
6.3	Goodness of Fit Test	133
6.4	Distribution Approximation	137
6.5	Parameter Estimation	138
6.5.1	Estimation of Location and Scale Parameters	139
6.5.2	Shape Parameter Estimation	140
6.6	Conclusions	141
7	Distribution Approximation of Radar Clutter by SIRPs	145
7.1	Introduction	145
7.2	Characterization of Elliptically Symmetric Distributions	146
7.3	Assessing the Distributional Properties	148
7.4	Distribution Identification of SIRVs	150
7.5	Parameter Estimation	154
7.6	Conclusions	156
8	Weak Signal Detection - Literature Review	157
8.1	Weak Signal Problem	157
8.1.1	Literature Review	160
8.2	Non-Gaussian Correlated Data	161
9	The Locally Optimum Detector	162
9.1	The Series Approach	163
9.1.1	The Known Signal Case	163

9.1.2	The Random Signal Case	16
9.2	The Lagrangian approach	16
9.2.1	The Known Signal Case	16
9.2.2	The Random Signal Case	16
9.3	Special Cases	17
9.3.1	The Known Signal Case	17
9.3.2	The Random Signal Case	17
10	Determining Thresholds for the Locally Optimum Detector	18
10.1	Introduction	18
10.2	Methods for Estimating Thresholds	18
10.2.1	Estimates Based on Raw Data	18
10.2.2	Estimates Motivated by the Extreme Value Theory	18
10.3	The Generalized Pareto Distribution	18
10.3.1	Methods for Estimating the Parameters of the GPD	18
10.3.2	Estimation of Thresholds	19
10.4	Numerical Results	19
10.4.1	Characterization of Tail Shape for Known Distributions	19
10.4.2	Empirical Properties of the Estimators for Known Distributions	19
10.4.3	Effect of the Choice of α on the Threshold Estimates	20
10.5	Examples	21
10.5.1	Known Distribution Case	21
10.5.2	An Unknown Distribution Case	21
11	Performance of the Locally Optimum Detector for the Multivariate Student-T Distribution	223
11.1	The Multivariate Student-T Distribution	224
11.2	The Locally Optimum Detector	226
11.3	Computer Simulation of Performance	227
11.4	Conclusions	230
A	Properties of SIRVs	236
A.1	Statistical Independence	236

A.2 Spherically Symmetric Characteristic function	238
A.3 Relationship Between Higher Order and Lower Order SIRV PDFs	238
B Computer Generation of SIRVs Using the Rejection Method	241
B.1 Rejection Method	241
B.2 Rejection Theorem	242
C Issues Related to Extreme Value Theory	245
C.1 Limiting Forms for the Largest Order Statistic	245
C.1.1 Case 1	246
C.1.2 Cases 2 and 3	249
C.2 Tails of Probability Density Functions	251
C.2.1 Case 1	251
C.2.2 Case 2	251
C.2.3 Case 3	251
C.3 PDF of the r^{th} Ordered Statistic	252

List of Figures

2.1	Performance comparison of the three processing configurations: Case 1.	14
2.2	Performance comparison of the three processing configurations: Case 2.	15
2.3	Performance comparison of the three processing configurations: Case 3.	16
2.4	Blockdiagram for illustration of the principle of the Joint-Domain Localized GLR Processor.	20
2.5	An JDL-GLR example.	21
2.6	Two dimensional power spectral density for the clutter/interference used in the example.	24
2.7	Detection performance comparison of the five processors.	26
4.1	K-distribution, $b = 0.31$, $\alpha = 0.05$	61
4.2	K-distribution, $b = 0.77$, $\alpha = 0.3$	62
4.3	K-distribution, $b = 1$, $\alpha = 0.5$	63
4.4	K-distribution, $b = 1.4$, $\alpha = 0.99$	64
4.5	Student-t distribution, $b = 0.14$, $\nu = 0.01$	67
4.6	Student-t distribution, $b = 0.45$, $\nu = 0.1$	68
4.7	Student-t distribution, $b = 1$, $\nu = 0.5$	69
4.8	Student-t distribution, $b = 2.23$, $\nu = 2.5$	70
4.9	Chi Envelope PDF, $b = 0.22$, $\nu = 0.1$	72
4.10	Chi Envelope PDF, $b = 0.5$, $\nu = 0.5$	73
4.11	Chi Envelope PDF, $b = 0.70$, $\nu = 1.0$	74
4.12	Weibull distributed Envelope PDF, $b = 0.5$, $a = 1.86$	77
4.13	Weibull distributed, $b = 1$, $a = 1$	78
4.14	Weibull distributed, $b = 2$, $a = 0.5$	79
4.15	Generalized Rayleigh distributed Envelope PDF, $\alpha = 0.1$, $\beta = 3.45 \times 10^{-15}$	81

4.16	Generalized Rayleigh distributed Envelope PDF, $\alpha = 0.5$, $\beta = 0.048$	82
4.17	Generalized Rayleigh distributed Envelope PDF, $\alpha = 1$, $\beta = 0.577$	83
4.18	Generalized Rayleigh distributed Envelope PDF, $\alpha = 2$, $\beta = 1.414$	84
4.19	Rician Envelope PDF, $\rho = 0.25$	86
4.20	Rician Envelope PDF, $\rho = 0.5$	87
4.21	Rician Envelope PDF, $\rho = 0.9$	88
4.22	Rician Envelope PDF, $a = 0.25$, $\alpha = 1$	93
4.23	Rician Envelope PDF, $a = 0.5$, $\alpha = 1$	94
4.24	Rician Envelope PDF, $a = 0.9$, $\alpha = 1$	95
4.25	Gaussian distribution, zero mean, unit variance	99
4.26	Laplace Distribution, $b = 1$	100
4.27	Cauchy Distribution, $b = 1$	101
4.28	K-distribution, $b = 1$, $\alpha = 0.5$	102
4.29	Student-t Distribution, $b = 1$, $\nu = 1.5$	103
4.30	Chi-distribution, $b = 1$, $\nu = 1$	104
4.31	Generalized Rayleigh PDF, $\alpha = 0.5$, $\beta = 0.05$	105
4.32	Weibull distribution, $a = 1$, $b = 1.0$	106
4.33	Rician PDF, $\rho = 0.5$	107
5.1	Simulation Scheme for SIRVs with Known Characteristic PDF	118
5.2	Simulation Scheme for SIRVs with Unknown Characteristic PDF	120
5.3	Theoretical and Empirical Quadratic form PDFs for Laplace SIRV	123
5.4	Theoretical and Empirical Quadratic form PDFs for Cauchy SIRV	124
5.5	Theoretical and Empirical Quadratic form PDFs for K-distributed SIRV	125
5.6	Theoretical and Empirical Quadratic form PDFs for Student-t SIRV	126
5.7	Theoretical and Empirical Quadratic form PDFs for Chi distributed SIRV	127
5.8	Theoretical and Empirical Quadratic form PDFs for Generalized Rayleigh SIRV	128
5.9	Theoretical and Empirical Quadratic form PDFs for Weibull SIRV	129
5.10	Theoretical and Empirical Quadratic form PDFs for Rician SIRV	130
6.1	Linked Vector Chart:Dashed lines P_0 = Null Linked Vectors, Solid Lines P_1 = Sample Linked Vectors	142

6.2	Empirical Distribution of Q_n for several values of n	143
6.3	Identification Chart for Univariate Distributions Based on 1000 samples ($n=50$)	144
7.1	Goodness of Fit Test using the Q_n Procedure. 90, 95 and 99% contours for the Gaussian distribution. Broken Line = Null distribution Pattern	152
7.2	Identification Chart for SIRVs ($n=2000$, $N=4$) 1 = Gaussian, 2 = Laplace, 3 = Cauchy, 4, 5, 6, 7 = Student-t, 8 = K-distribution, 9 = Chi, 10 = Generalized Rayleigh, 11 = Weibull, 12 = Rician	153
9.1	Canonical form of LOD assuming known signal and independent random variables.	174
9.2	Canonical form of LOD assuming known signal and random variables arising from an SIRP.	176
9.3	Canonical form of LOD assuming random signal and independent random Variables.	177
9.4	Canonical form of LOD assuming random signal and random disturbance arising from an SIRP.	179
10.1	Shaded areas indicating P_F and P_D	181
10.2	Generalized Pareto PDF, $\gamma = -1$	186
10.3	Generalized Pareto PDF, $\gamma = 1$	187
10.4	PDF of test statistic with tail region defined for $t \geq t_0$	188
10.5	Tail of the test statistic shifted to origin.	189
10.6	Normal distribution, $n=10,000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a:True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$	209
10.6	Normal distribution, $n=1000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a:True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$	210
10.6	Lognormal distribution, $n=10,000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a:True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$	211
10.6	Lognormal distribution, $n=1000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a:True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$	212
10.7	Histograms of threshold values. (A) $P_F = 10^{-2}$ (B) $P_F = 10^{-3}$ (C) $P_F = 10^{-4}$ (D) $P_F = 10^{-5}$ (E) $P_F = 10^{-6}$ (F) $P_F = 10^{-7}$	217
10.7	Fig 10.7 Contd.	218
10.7	Fig 10.7 Contd.	219

10.7 Fig 10.7 Contd.	220
10.7 Fig 10.7 Contd.	221
10.7 Fig 10.7 Contd.	222
11.1 Nonlinearity for the student-T distribution.	228

List of Tables

4.1	Marginal PDF	96
4.2	SIRVs obtained from the marginal envelope PDF	97
4.3	SIRVs obtained from the marginal characteristic function	97
5.1	$h_{2N}(p)$ for SIRVs with Known Characteristic PDF	117
5.2	$h_{2N}(p)$ for SIRVs with Unknown Characteristic PDFs	117
5.3	Characteristic PDF for SIRVs listed in Table 4.1 [$E(S^2) = 1$]	118
5.4	Related PDF $f_V(v)$	118
7.1	SIRVs obtained from the marginal envelope PDF	151
7.2	SIRVs obtained from the marginal characteristic function	151
7.3	Shape Parameters of the SIRVs Used for the Identification Chart	154
10.1	Tail parameter γ describing the upper ten percent of various distributions.	198
10.2	Median of the normalized bias values for different percentiles. OLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments	199
10.2	Median of the normalized bias values for different percentiles. (contd.)	200
10.2	Median of the normalized bias values for different percentiles. (contd.)	201
10.2	Median of the normalized bias values for different percentiles.	202
10.3	Median RMS errors for various percentiles. OLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments	203
10.3	Median RMS errors for various percentiles. (contd.)	204
10.3	Median RMS errors for various percentiles. (contd.)	205
10.3	Median RMS errors for various percentiles.	206

11.1	Sample Size=16, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	232
11.2	Sample Size=16, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	232
11.3	Sample Size=16, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	232
11.4	Sample Size=32, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	232
11.5	Sample Size=32, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	233
11.6	Sample Size=32, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	233
11.7	Sample Size=64, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	233
11.8	Sample Size=64, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	233
11.9	Sample Size=64, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	234
11.10	Sample size=64, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver	234
11.11	Sample Size=128, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Opti- mum Detector, GR:Gaussian Receiver	234
11.12	Sample Size=128, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Opti- mum Detector, GR:Gaussian Receiver	234
11.13	Sample Size=128, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Opti- mum Detector, GR:Gaussian Receiver	235
11.14	Sample Size=128, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Opti- mum Detector, GR:Gaussian Receiver	235

Chapter 1

Introduction

The subject of this report is the detection of weak targets in a strong clutter environment. Two situations arise depending upon whether or not the weak targets can be separated from the clutter. For both cases new receivers are derived which provide significant improvement in performance over other recently proposed techniques. This work includes development of an adaptive joint-domain space-time processor, effective non-Gaussian weak signal detectors based on spherically invariant random processes, and a new method for approximating the underlying probability density function of random data which works extremely well with only 100 samples. Many new algorithms were developed for this purpose and resulted in extensive new software. In a companion volume the quality of some of this software is evaluated and discussed.

Two situations arise depending upon whether or not the weak targets can be separated from the clutter. For example, consider the situation illustrated in Figure 1.1 where the joint spatial and Doppler spectra of the received radar samples are shown for targets T_1 and T_2 and a single clutter patch. Obviously, target T_1 can be separated from the clutter by means of filtering whereas the target T_2 cannot.

When the target can be separated from the clutter, performance is limited by the background noise. Assuming a large signal-to-noise ratio, we refer to this as the strong signal case. When the target and clutter overlap and the clutter-to-noise ratio is large, performance is limited by the clutter. Assuming a small signal-to-clutter ratio, we refer to this case as the weak signal case. Finally, when the clutter spectrum partially overlaps the target spectrum, performance is limited by both the clutter and noise. We refer to this situation as the intermediate signal case. The strong and intermediate signal cases are suitable for the adaptive joint-domain space-time

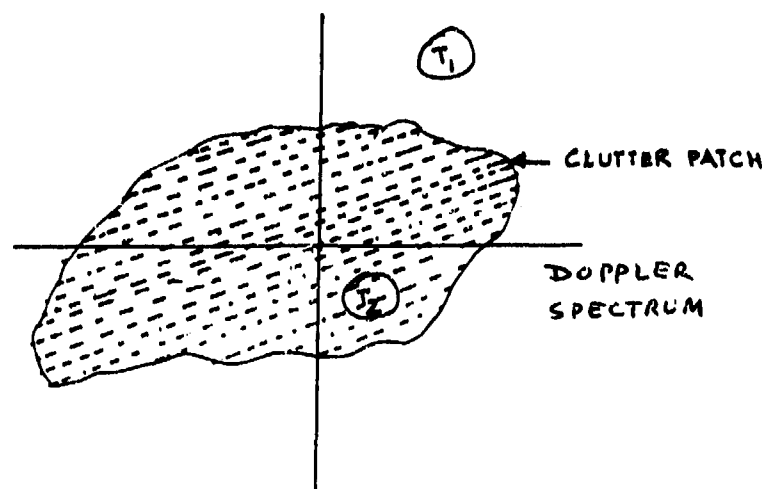


Figure 1.1: Illustration of Target and Clutter Spectra

processor discussed in Chapter 2. The remainder of this report, Chapters 3-11, are devoted to the solution of the weak signal case.

1.1 Adaptive Implementation of Optimum Space-Time Processing

A new adaptive algorithm, called the Joint-Domain Localized Generalized Likelihood Ratio (JDL-GLR) detection algorithm, is presented in Chapter 2. This algorithm takes advantage of the fact that it may be possible to separate the weak target from the strong clutter (interference) by means of space-time processing. Specifically, space-time processing transforms the received samples in space and time to a two-dimensional power spectral density involving both spatial and Doppler frequencies. The spatial frequency is a function of the angle of arrival of the radar pulse return (interference) plane waves with respect to the broadside of the antenna array while the Doppler frequency is linearly proportional to the radial velocity of the object from which the radar pulse is reflected (platforms from which the interference is emitted). When the radar target's angle of arrival and/or radial velocity differs significantly from those of the clutter (interference), it is possible to separate out the target return. System performance is then limited primarily by the background noise. Because the clutter (interference) environment is unknown a priori and is likely to change with time and spatial position, the algorithm must be adaptive with a sufficiently fast convergence rate. The JDL-GLR algorithm presented in Chapter 2 is

both data and computationally efficient and converges quickly for Gaussian random processes. Embedded CFAR and robustness in non-Gaussian clutter (interference) are other properties of this algorithm.

1.2 Weak Signal Detection

The algorithms presented in Chapters 3-11 were developed to handle the case for which it is not possible to separate the target return from the clutter (interference). In other words, these algorithms are intended to be applied only when the target and clutter (interference) spectra overlap significantly. We refer to this situation as the weak signal problem. For this problem, system performance is limited primarily by the clutter (interference). Several new algorithms have been developed for the weak signal detection problem. Although these algorithms can be used to combat both clutter and interference, for ease of discussion, the presentation focuses only on weak signal detection in a strong clutter background. The statistics of clutter have been observed to be both Gaussian and non-Gaussian. Because the weak signal detector for Gaussian processes is identical to that for strong signals, only the non-Gaussian case is considered in Chapters 3-11.

1.3 Literature Review on Spherically Invariant Random Processes

In general, the radar receiver receives N complex (or $2N$ quadrature component) samples from each radar resolution cell. To develop an optimal receiver, it is necessary to have a closed form analytical expression for the joint probability density function (PDF) of the received samples. When the N samples are statistically independent, the joint PDF is simply the product of the marginal PDFs. However, clutter samples are likely to be correlated. Because this correlation is useful for canceling the clutter, it is important that the correlation be modeled. Unfortunately, when the received samples are correlated and non-Gaussian, there are no unique analytical expressions for their joint PDF. A search of the mathematical and signal processing literature reveals that the theory of spherically invariant random processes (SIRP) provides a powerful mechanism for obtaining the joint PDF of N correlated non-Gaussian random variables. The literature search on SIRPs is reviewed in Chapter 3.

1.4 Radar Clutter Modeling Using SIRPs

As mentioned previously, the clutter is unknown apriori and is likely to change with time and spatial position. Consequently, it is necessary to continuously monitor the environment in order

to determine the statistical nature of the clutter. To be able to model as many different types of clutter as possible, a large library of multivariate non-Gaussian PDFs is necessary. Based on the properties of SIRPs, a library of joint PDFs is developed in Chapter 4 for correlated non-Gaussian random variables.

1.5 Computer Generation of Simulated Radar Clutter Characterized as SIRPs

When dealing with non-Gaussian random processes, it is usually difficult, if not impossible to analytically evaluate system performance. Performance must then be determined by means of computer simulation. Two canonical procedures are presented in Chapter 5 for generating correlated non-Gaussian random variables which can be used to simulate samples from SIRPs.

1.6 A New Method for Univariate Distribution Approximation

Because the clutter environment is unknown a priori, the PDF underlying a set of N samples must be approximated using measured samples from the environment. Chapter 6 describes an algorithm for analyzing univariate random data. This algorithm has two modes of operation. In the first mode, the algorithm performs a goodness-of-fit test. Specifically, the test determines, to a desired confidence level, whether random data is statistically consistent with a specified probability distribution. In the second mode of operation, the algorithm approximates the PDF underlying the random data. In particular, by analyzing the random data and without any a priori knowledge, the algorithm identifies from a stored library of PDFs that density function which best approximates the data. Estimates of the scale, location, and shape parameters of the PDF are provided by the algorithm. Of particular note is the observation that the algorithm typically works well with small sample sizes of between 50 and 100 samples.

1.7 Distribution Approximation of Radar Clutter by SIRPs

As noted earlier, the N complex samples received from each radar resolution cell are characterized by a multivariate PDF. For SIRPs, it is shown in Chapter 7 that the multivariate distribution approximation problem can be reduced to an equivalent univariate distribution approximation problem. Consequently, the algorithm of Chapter 6 is also used in Chapter 7 to approximate the joint PDF underlying N correlated non-Gaussian clutter samples provided they are generated from an SIRP.

1.8 Weak Signal Detection

The weak signal detection problem is developed in Chapter 8. Problems encountered in the optimum likelihood ratio test (LRT) are pointed out. The concept of the locally optimum detector (LOD) is introduced as a practical detector structure for the weak signal problem.

1.9 The Locally Optimum Detector

The LOD is derived in Chapter 9 using two different approaches. Both deterministic and random target signals are considered. It is shown that the LOD determines whether a target is present or not by comparing a statistic computed from the data to a set threshold. The receiver structures are specialized to the case for which the clutter plus noise can be approximated as an SIRP.

1.10 Determining Thresholds for the Locally Optimum Detector

Not only is the clutter assumed to be non-Gaussian, the LOD receiver structure is non-linear. As a result, system performance must be determined by means of computer simulation. The threshold is conventionally determined through a Monte Carlo procedure. Unfortunately, the number of trials is inversely proportional to the false alarm probability P_F . For example, when $P_F = 10^{-6}$, a minimum of ten million trials need to be generated. To avoid carrying out so many trials, a new technique, based on extreme value theory is presented in Chapter 10. It is demonstrated that fairly accurate thresholds can be determined for false alarm probabilities as small as 10^{-7} with as few as 5000-10,000 trials.

1.11 Performance of the LOD for the Multivariate Student-T Distribution

Assuming that the clutter plus noise can be approximated by the multivariate Student-T distribution, the LOD is developed in Chapter 11 for the weak signal detection problem. The system performance is evaluated by means of computer simulation. When P_F is less than or equal to 10^{-2} , it is shown that the Gaussian receiver requires a signal to clutter ratio of 10-20 dB larger than that required by the LOD for the same values of P_D and P_F .

Chapter 2

Adaptive Implementation of Optimum Space-Time Processing

2.1 Introduction

It is highly desirable for an airborne surveillance radar system to have the optimum or near optimum performance for detection of weak targets in strong clutter/interference of complicated angle-Doppler spectrum. As the clutter/interference spectrum is unknown to the system and the clutter/interference environment may be varying in both time and space, i.e., nonstationary and nonhomogeneous, the signal processor must be adaptive with a sufficiently fast convergence rate.

Consider a system which employs N_s spatial channels (subarrays of a phased-array) and has N_t pulses in its Coherent Processing Interval (CPI). The optimum processor, or the Neyman-Pearson's likelihood ratio test for such a system, is well developed in [1] under the assumption of Gaussian clutter/interference. This processor, to be referred to as the joint-domain optimum processor in this paper, has the highest performance potential which can be approached by adaptive algorithms such as the Sample-Matrix-Inversion (SMI) [2], the Generalized Likelihood Ratio (GLR) [3, 4], and the Modified SMI [5, 6]. To approach this detection performance potential, however, these algorithms require that the training data set (i.e., the so-called secondary data set) have at least $2N_s N_t \sim 3N_s N_t$ independent and identically distributed (iid) data vectors. Obviously such a training-data size requirement is impractical even for moderate N_s and N_t , as the environment in which an airborne surveillance system operates is usually severely nonstationary and nonhomogeneous. Besides, the computation load can easily become unbearable in practice since it is proportional to $(N_s N_t)^3$. One should also note that lowering N_s and N_t is not

necessarily desirable in practice as the performance potential critically depends on them if the angle-Doppler spectrum of the clutter/interference is complicated.

A much more popular approach to space-time processing can be classified as cascade processing with either the beamformer-Doppler processor configuration or the opposite order configuration. In this paper the former will be called the *Space-time (S-T) configuration* and the latter the *Time-Space (T-S) configuration*. Obviously the optimum detection theory can be applied separately to both spatial and temporal parts of both S-T and T-S configurations, together with various adaptive algorithms available for each part. Of course, the convergence rate and computation load problems associated with adaptive implementation of the joint-domain optimum processor also appear with the cascade configurations, only to a lesser extent. When the convergence does occur, the performance of an adaptive implementation with the S-T (T-S) configuration should approach that of the optimum processor with the same configuration. Cascade processing, especially the S-T configuration, has been so popular in recent years that it seems to replace the joint-domain processor in the airborne surveillance application. Moreover, arguments can often be heard about which cascade configuration has higher detection performance potential.

The first objective of this chapter is to show that

- (1) neither of the two cascade configurations is better than the other, and
- (2) the performance potential of both cascade configurations can fall far below that of the joint-domain optimum processor. In other words, we show that if one wants to approach the highest performance potential offered by the joint-domain optimum processor, both cascade configurations should be avoided.

As pointed out earlier in this section, it is difficult in practice to approach the performance potential of the joint-domain optimum processor with the straightforward application of adaptive algorithms such as the SMI, Modified SMI, GLR, etc., especially in a severely nonstationary and nonhomogeneous environment, even if the heavy real-time computation could become affordable. Therefore, the second objective of this chapter is to develop a new adaptive algorithm for the joint-domain optimum processor, which should be much more data-efficient and computationally efficient than the aforementioned ones. This new algorithm is an extension of our recent work reported in [7, 8] for adaptive Doppler-domain processing.

This chapter is organized as follows. We will first formulate the data model in Section 2.2. In Section 2.3 we will compare the performance potentials of the cascade and joint-domain

processors. The new adaptive algorithm for the joint-domain optimum processor is presented in Section 2.4, together with its performance analysis and comparison. Finally, Section 2.5 summarizes the conclusions with some discussion of related issues.

2.2 Data Modeling

Consider a narrowband antenna array with N_s spatial channels (subarrays). Each channel receives N_t data samples corresponding to the return of a train of N_t coherent pulses for a given range cell. Let the column vector $\mathbf{x}_{tn}$, $N_t \times 1$, represent the N_t baseband complex (I/Q) data samples of the n ,th channel. The data matrix $\mathbf{X}$, $N_t \times N_s$, is defined by

$$\mathbf{X} = [\mathbf{x}_{t1} \quad \mathbf{x}_{t2} \quad \dots \quad \mathbf{x}_{tN_s}] = \begin{bmatrix} \mathbf{x}_{s1}^T \\ \mathbf{x}_{s2}^T \\ \vdots \\ \mathbf{x}_{sN_t}^T \end{bmatrix}, \quad (2.1)$$

where “ T ” denotes the transpose, and the row vectors of $\mathbf{X}$, $\mathbf{x}_{sn}^T$, $n = 1, 2, \dots, N_t$, are the “snapshots” obtained along the spatial channels.

Under the signal-absence hypothesis H_0 , the data matrix $\mathbf{X}$ consists of clutter/interference and noise components only, i.e.,

$$\mathbf{X} = \mathbf{C} + \mathbf{N} \quad (2.2)$$

where $\mathbf{C}$ and $\mathbf{N}$ represent the clutter/interference and noise, respectively, and are assumed to be independent. Under the signal-presence hypothesis H_1 , a target signal component also appears in the data matrix, i.e.,

$$\mathbf{X} = \alpha \mathbf{S} + \mathbf{C} + \mathbf{N} \quad (2.3)$$

where α is an unknown complex constant representing the amplitude of the signal and $\mathbf{S}$ the signal matrix of a known form. We call $\mathbf{X}$ the primary data set as it is from the range cell under the hypothesis test.

For simplicity of discussion only, we assume that the spatial channels are colinear, identical, omni-directional, and equally spaced with spacing d ; and that the pulses of the coherent pulse train are identical with a constant Pulse Repetition Frequency (PRF). Under these assumptions,

the $n_t n_s$ th entry of the signal matrix $\mathbf{S}$ has the following form

$$s(n_t, n_s) = \exp[i2\pi(n_t - 1)\frac{2v}{\lambda\text{PRF}} + i2\pi(n_s - 1)\frac{d \sin \theta}{\lambda}], \quad (2.4)$$

where v is the radial velocity of the target, θ the direction of arrival of the target-return planewave with respect to the broadside of the array, and λ the radar wavelength. Denoting

$$f_{st} = \frac{2v}{\lambda\text{PRF}} \quad (2.5)$$

as the "normalized Doppler frequency" of the target signal, and

$$f_{ss} = \frac{d \sin \theta}{\lambda} \quad (2.6)$$

as the "spatial frequency", $\mathbf{S}$ can be expressed by

$$\mathbf{S} = \mathbf{s}_s^T \otimes \mathbf{s}_t \quad (2.7)$$

where $\otimes$ is the Kronecker product, and

$$\mathbf{s}_t = [1 \quad \exp(i2\pi f_{st}) \quad \dots \quad \exp(i2\pi(N_t - 1)f_{st})]^T \quad (2.8)$$

and

$$\mathbf{s}_s = [1 \quad \exp(i2\pi f_{ss}) \quad \dots \quad \exp(i2\pi(N_s - 1)f_{ss})]^T \quad (2.9)$$

are the signal vectors in time and space domains, respectively. We assume that the parameters PRF, λ , and d have been properly chosen so that f_{st} and f_{ss} are confined within $[-0.5, 0.5]$.

To statistically characterize the clutter/interference and noise components $\mathbf{C}$ and $\mathbf{N}$, we introduce the notation $\text{Vec}(\cdot)$ for a matrix operation that stacks the columns of a matrix under each other to form a new column vector. We assume that the $N_t N_s \times 1$ vector $\text{Vec}(\mathbf{C} + \mathbf{N})$ has a multivariate complex Gaussian distribution with zero mean and a covariance matrix $\mathbf{R}$. Under this assumption, $\mathbf{x}_{tn_s}, n_s = 1, 2, \dots, N_s$ and $\mathbf{x}_{sn_t}, n_t = 1, 2, \dots, N_t$ will also be complex zero-mean Gaussian. Let $\mathbf{R}_t$ and $\mathbf{R}_s$ be the covariance matrices of $\mathbf{x}_{tn_s}$ and $\mathbf{x}_{sn_t}$, respectively. It is easy to see that $\mathbf{R}_t$ and $\mathbf{R}_s$ are the submatrices of $\mathbf{R}$.

In the cases of unknown clutter/interference statistics, the data from the adjacent range cells, conventionally referred to as the secondary data set, are also needed for estimating the covariance

of clutter/interference. Under both H_1 and H_0 , they consist of the clutter/interference and noise components only, and they are denoted by

$$\mathbf{Y}_k = \mathbf{C}_k + \mathbf{N}_k, \quad N_t \times N_s, \quad k = 1, 2, \dots, K \quad (2.10)$$

where K is the number of range cells available. We assume that $\mathbf{Y}_k, k = 1, 2, \dots, K$ and $\mathbf{X}$ are independent of each other and bear the same clutter/interference statistics, i.e., $\text{Vec}(\mathbf{Y}_k)$ should also have a complex-Gaussian distribution with zero mean and a covariance matrix $\mathbf{R}$.

2.3 Difference among The Performance Potentials of The Cascade and Joint-Domain Processors

We will compare the detection performance potentials of the two cascade configurations and the joint-domain processor under the assumption that the clutter/interference-plus-noise covariance matrix is known. With the known covariance, the Space-Time (S-T) configuration is the N_s th-order optimum spatial processor followed by the N_t th-order optimum temporal (Doppler) processor, the Time-Space(T-S) configuration takes the opposite cascade, and the joint-domain processor is the $N_s N_t$ th-order optimum processor. Applying the result in [1] to the above three, we list the optimum weight vectors below for easy reference.

The S-T Configuration: we have

$$\mathbf{w}_{s,s-t} = c_{s,s-t} \mathbf{R}_s^{-1} \mathbf{s}_s \quad (2.11)$$

for the spatial domain weight vector, and

$$\mathbf{w}_{t,s-t} = c_{t,s-t} [(\mathbf{w}_{s,s-t}^H \otimes \mathbf{I}) \mathbf{R} (\mathbf{w}_{s,s-t} \otimes \mathbf{I})]^{-1} \mathbf{s}_t \quad (2.12)$$

for the temporal domain weight vector, where $c_{s,s-t}$ and $c_{t,s-t}$ are constants. We recall that $\mathbf{R}_s$ and $\mathbf{R}_t$ are the covariance matrices for the rows and columns of $\mathbf{X}$, respectively; and $\mathbf{s}_s$ and $\mathbf{s}_t$ are specified by Eq.(2.8) and Eq.(2.9). The test statistic is

$$\eta_{s-t} = \mathbf{w}_{t,s-t}^H \mathbf{X} \mathbf{w}_{s,s-t}^* \quad (2.13)$$

The T-S Configuration: we have

$$\mathbf{w}_{t,t-s} = c_{t,t-s} \mathbf{R}_t^{-1} \mathbf{s}_t \quad (2.14)$$

and

$$\mathbf{w}_{s,t-s} = c_{s,t-s} [(\mathbf{I} \otimes \mathbf{w}_{t,t-s}^H) \mathbf{R} (\mathbf{I} \otimes \mathbf{w}_{t,t-s})]^{-1} \mathbf{s}_s \quad (2.15)$$

for the temporal and spatial weight vectors, respectively. The test statistic is

$$\eta_{t-s} = \mathbf{w}_{t,t-s}^H \mathbf{X} \mathbf{w}_{s,t-s}^* \quad (2.16)$$

The joint-domain optimum processor: the whole set of the data is processed all together by an optimum weight vector as

$$\eta_J = \mathbf{w}_J^H \text{Vec}(\mathbf{X}) \quad (2.17)$$

where $\mathbf{w}_J$ is

$$\mathbf{w}_J = c_J \mathbf{R}^{-1} \quad (2.18)$$

with c_J being a constant scalar.

One should note that the overall weight vectors for the two cascade configurations can have the following equivalent expressions

$$\mathbf{w}_{s-t} = \mathbf{w}_{s,s-t} \otimes \mathbf{w}_{t,s-t} \quad (2.19)$$

and

$$\mathbf{w}_{t-s} = \mathbf{w}_{s,t-s} \otimes \mathbf{w}_{t,t-s} \quad (2.20)$$

The squared magnitude of the test statistic is compared with a chosen threshold η_0 which is determined by the required probability of false alarm P_f as

$$\eta_0 = -\ln P_f; \quad (2.21)$$

and the signal presence is claimed if the test statistic surpasses the threshold.

From the result in [1], the probability of detection of the above three processors has the same

form below with their own weight vectors, i.e., $\mathbf{w}_{s-t}$, $\mathbf{w}_{t-s}$, and $\mathbf{w}_J$ to replace $\mathbf{w}$ therein

$$P_d = 1 - \exp(-\gamma) \int_0^{\eta_0} \exp(-t) I_0(2\sqrt{\gamma t}) dt \quad (2.22)$$

where

$$\gamma = |\alpha|^2 \frac{\mathbf{w}^H \mathbf{s} \mathbf{s}^H \mathbf{w}}{\mathbf{w}^H \mathbf{R} \mathbf{w}} \quad (2.23)$$

and $I_0(\cdot)$ denotes the zero-th order modified Bessel function of the first kind.

The key to achieving the objective of the comparison easily is to identify few typical cases, from the vast number of varieties of clutter/interference conditions, which are also simple enough for numerical evaluation. To do so, the following specifics are necessary.

- (1) The covariance matrix of the receiver noise is given by

$$\mathbf{E}(\text{Vec}(\mathbf{N})\text{Vec}(\mathbf{N})^H) = \sigma_n^2 \mathbf{I} \quad (2.24)$$

with $\mathbf{I}$ being the $N_t N_s \times N_t N_s$ identity matrix.

- (2) The clutter/interference is assumed to have a two-dimension power spectral density of the Gaussian shape centered at $[f_{ct}, f_{cs}]$

$$P_c(f_t, f_s) = \sigma_c^2 \frac{1}{2\pi\sigma_{f_t}\sigma_{f_s}} \exp\left[-\left(\frac{(f_t - f_{ct})^2}{2\sigma_{f_t}^2} + \frac{(f_s - f_{cs})^2}{2\sigma_{f_s}^2}\right)\right] \quad (2.25)$$

where f_t and f_s are the normalized Doppler frequency and spatial frequency, respectively, and σ_{f_t} and σ_{f_s} the parameters controlling the spread of the clutter/interference spectrum. The separation between the signal and the center of the clutter/interference spectrum is denoted by $\Delta f_t = f_{st} - f_{ct}$ and $\Delta f_s = f_{ss} - f_{cs}$.

- (3) The covariance of the clutter/interference corresponding to the above spectrum is then found to be

$$\mathbf{E}(\text{Vec}(\mathbf{C})\text{Vec}(\mathbf{C})^H) = \sigma_c^2 \mathbf{C}_s \otimes \mathbf{C}_t \quad (2.26)$$

where C_t and C_s are Toeplitz matrices specified by

$$C_t = \text{Toeplitz}\{[1 \quad e^{-2(\pi\sigma_{f_t})^2 - i2\pi f_{ct}} \quad \dots \quad e^{-2(\pi\sigma_{f_t}(N_t-1))^2 - i(N_t-1)2\pi f_{ct}}]\}, \quad (2.27)$$

and

$$C_s = \text{Toeplitz}\{[1 \quad e^{-2(\pi\sigma_{f_s})^2 - i2\pi f_{cs}} \quad \dots \quad e^{-2(\pi\sigma_{f_s}(N_s-1))^2 - i(N_s-1)2\pi f_{cs}}]\}, \quad (2.28)$$

respectively. It is easy to verify that (1) and (3) will lead to $R_t = \sigma_c^2 C_t + \sigma_n^2 I$ and $R_s = \sigma_c^2 C_s + \sigma_n^2 I$.

We define the clutter/interference-to-noise-ratio (INR) and signal-to-clutter/interference-plus-noise-ratio (SINR) by

$$\text{INR} = \frac{\sigma_c^2}{\sigma_n^2}, \quad (2.29)$$

and

$$\text{SINR} = \frac{|\alpha|^2}{(\sigma_n^2 + \sigma_c^2)}. \quad (2.30)$$

Three simple cases are identified below in each of which at least one of the cascade configurations suffers severe performance degradation, i.e., significantly departing from the joint-domain optimum.

Case 1. The signal and interference are "well" separated in the angle domain (in the sense that $\Delta f_s > 1/N_s$) but close to each other in the Doppler domain ($\Delta f_t < 1/N_t$). This situation is shown in the subplot in Fig. 2.1. The detection performance vs. SINR for the three processors are plotted in Fig. 2.1 with $\text{INR}=40$ dB and $P_f = 10^{-5}$. The S-T configuration shows almost the same performance potential as the joint-domain optimum in this special case, while the performance loss for the T-S configuration becomes significantly large.

Case 2. The signal and interference are "well" separated in the Doppler domain but close to each other in the angle domain, as indicated by the subplot in Fig. 2.2. The T-S configuration is now close to the joint-domain optimum while the S-T configuration departs significantly.

Case 3. The clutter/interference spectrum has two peaks with one close to the signal in the angle domain while the other in the Doppler domain. In this case *both* cascade configurations fail to approach the joint-domain optimum, as shown in Fig. 2.3.

The above three cases are typical in the sense that we can draw from them the following

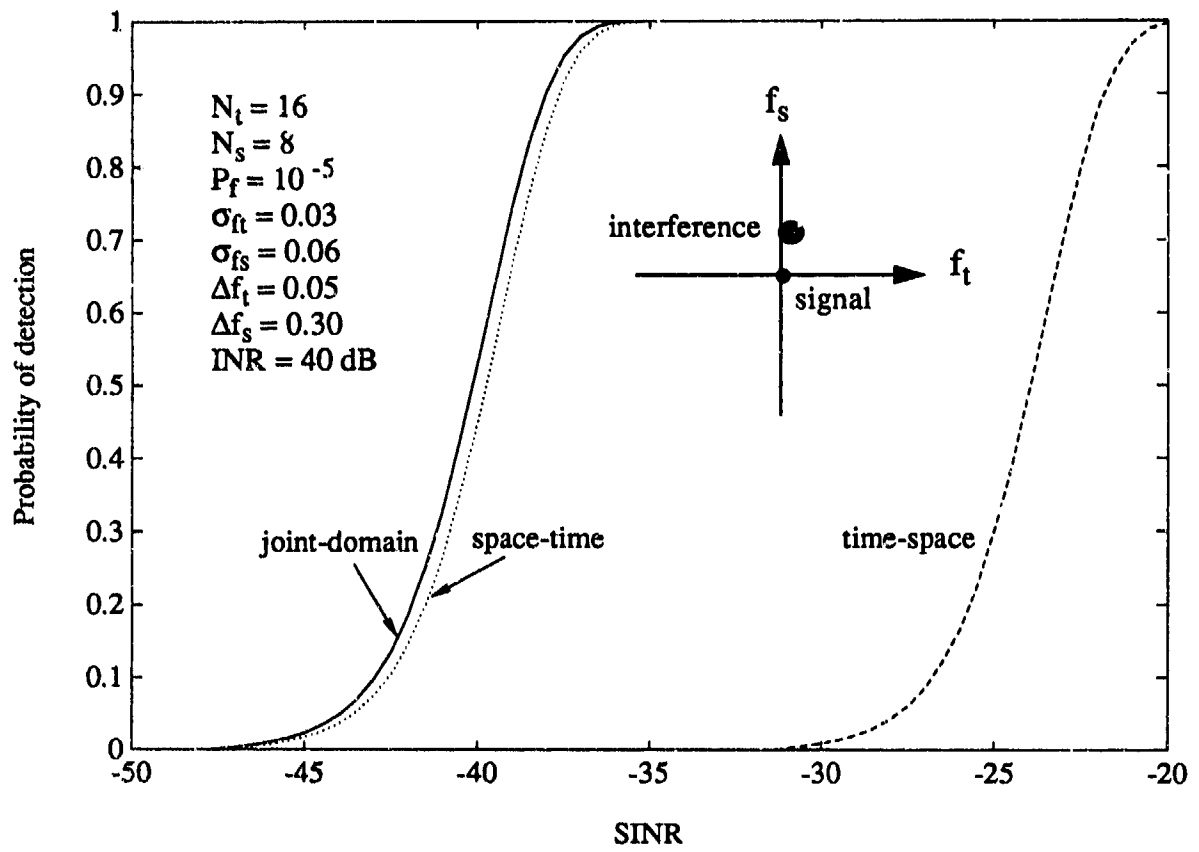


Figure 2.1 Performance comparison of the three processing configurations: Case 1.

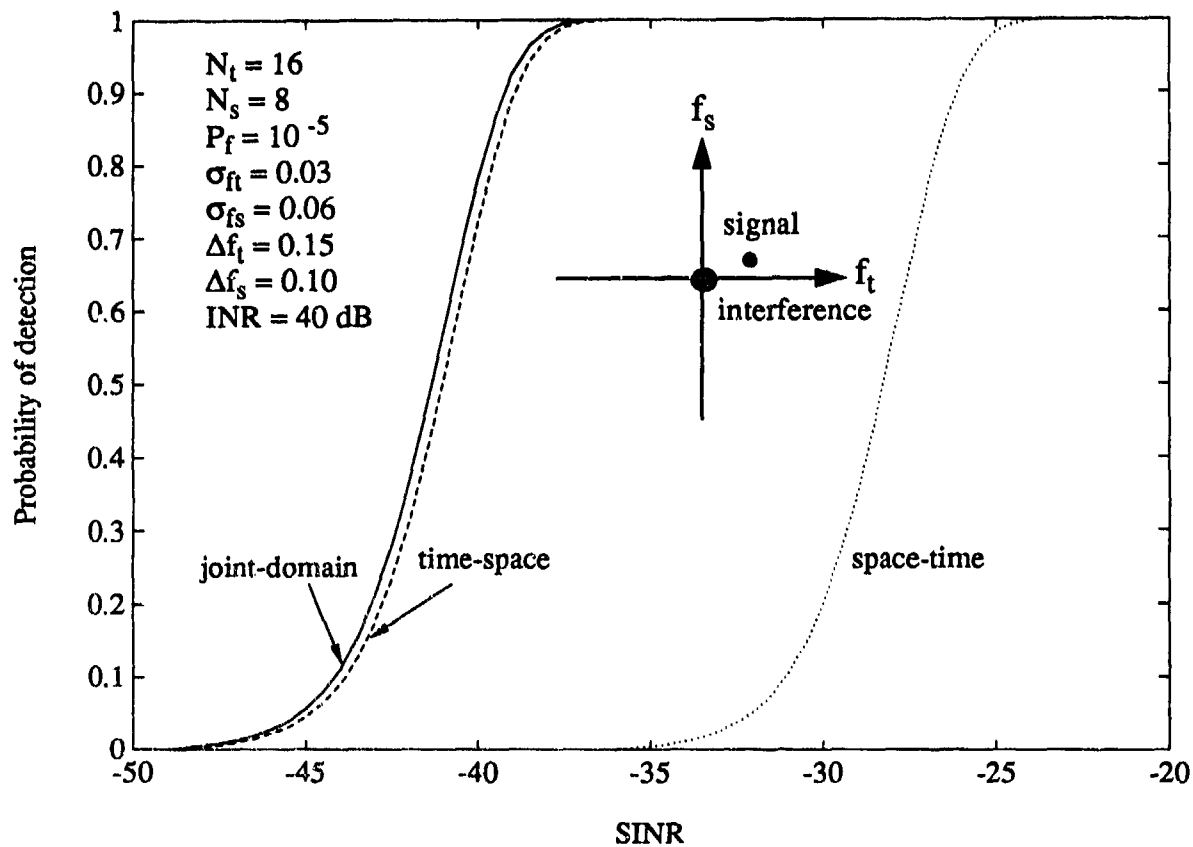


Figure 2.2. Performance comparison of the three processing configurations: Case 2.

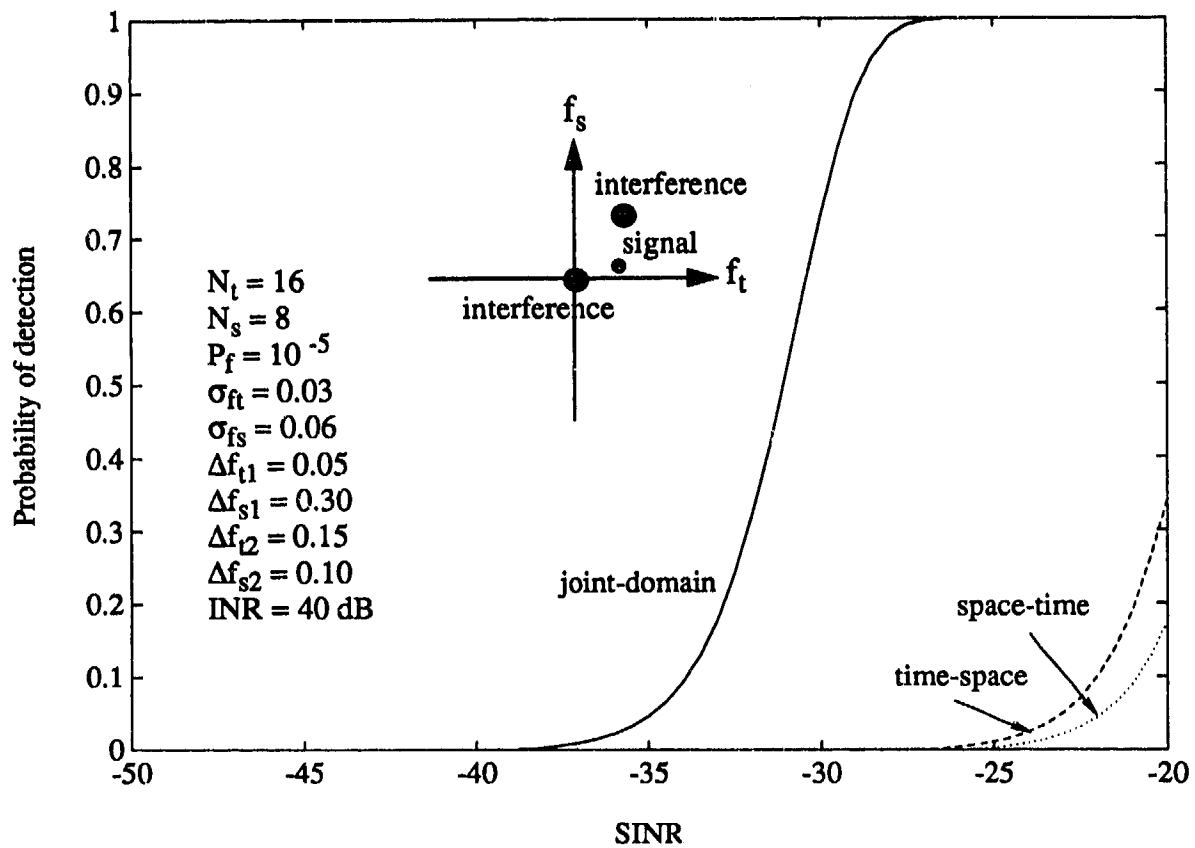


Figure 2.3. Performance comparison of the three processing configurations: Case 3.

conclusions:

- (1) neither of the two cascade configurations is better than the other, and
- (2) the performance potential of both cascade configurations can fall far below that of the joint-domain optimum processor.

Intuitively the above conclusions are also well justified. The T-S configuration in Case 1 suppresses the signal as well as the clutter/interference as they have little separation in the Doppler frequency domain, so does the S-T configuration in Case 2 in the angle domain. As both Case 1 and Case 2 can appear in practical situations without *a priori* knowledge, preselection of either cascade configuration is thus not appropriate. In Case 3 the signal and clutter/interference have little separation in either of the two domains, which results in the failure of both cascade configurations. However, the separation in the joint domain in Case 3 is still sufficiently large to lead to the success of joint-domain optimum processor. As an airborne system has to deal with clutter/interference having both angle and Doppler spectral spread, it is thus important to make full use of the signal-clutter/interference separation, which cannot always be achieved by either of the two cascade configurations.

Although our study so far in this chapter is centered around the detection performance potentials, i.e., under the assumption of known clutter/interference statistics, it is sufficient for us to direct our attention only to the adaptive implementation of the joint-domain optimum processor, since the two cascade configurations have been shown to have limited potentials. This will be the focus of the remaining part of this chapter. Before we proceed, we should point out that, in addition to the problem of limited potentials, the two cascade configurations may have other serious problems of practical importance which are associated with their adaptive implementations, e.g., the difficulty to achieve a high-quality Constant False Alarm Rate (CFAR). This issue will be briefly discussed later in Section 2.5 to preserve the continuity of our main course.

2.4 The Joint-Domain Localized GLR Algorithm

As pointed out in the introduction, the straightforward application of available adaptive algorithms such as the SMI, Modified SMI, and GLR, etc., has considerable difficulty to approach the joint-domain optimum processor in practice, especially in severely nonstationary and non-homogeneous environments. Our goal here is to develop an adaptive implementation which is more data-efficient (in the sense of faster convergence/requiring fewer training data) as well as more computationally efficient. In addition, it is highly desirable in practice to have the adap-

tive algorithm possess an embedded CFAR feature and a low sensitivity to the deviation of the clutter/interference distribution from the assumed Gaussian.

To achieve the above goal we will follow the idea of localized adaptive processing as presented in [7, 8] for adaptive MTD. Although this idea is similar to that of beam-space processing in [9, 10, 11] under the term of partially adaptive array processing, the work in [7, 8] distinguishes itself from the previous study on beam-space processing in the following ways. References [7, 8] are the first to point out that for the cases of the limited training-data size the use of localized adaptive processing is almost mandatory, and they have shown that localized adaptive processing can actually *outperform* fully adaptive processing in nonstationary and nonhomogeneous environments. Furthermore, References [7, 8] are also the first to study localized adaptive processing with the *detection* performance measure, which is of course the primary concern of surveillance systems. In contrast, the previous work on beam-space processing focuses on the *steady state* performance and uses the signal *estimation* performance measure. As the primary concern of this paper is again detection in severely nonstationary and nonhomogeneous environments, it is natural to follow the work in [7, 8]. Of course, the extension represents a nontrivial task as indicated by the complexity of the joint angle- Doppler domain.

As discussed in [7, 8], the localized processing idea can be applied with a variety of adaptive algorithms such as the SMI, Modified SMI, and GLR. We will again pick up the GLR because it offers the desirable embedded CFAR feature as well as possesses the desirable robustness in non-Gaussian clutter/interference [5, 6]. Hence, the new algorithm presented in this section will be called the Joint-Domain Localized GLR (JDL-GLR).

2.4.1 The JDL-GLR Principle

Figure 2.4 illustrates the principle of the JDL-GLR processor we propose. The data in the space-time domain, $\mathbf{X}$, $N_t \times N_s$ is first transformed to the angle-Doppler domain. This multi-dimensional transform should be invertible to avoid any information loss, and it can be done most conveniently via the standard two-dimensional DFT (which is linear and orthogonal) under the assumption made in Section 2.2 for the spatial channels and pulse train. One should note that the gaussianarity assumed for $\mathbf{X}$ will not be affected if the transformation is linear. The angle-Doppler domain data matrix $\mathcal{X}$, $N_t \times N_s$, represents the data at the N_t Doppler-bins and N_s angle-bins of the range cell under the hypothesis test. The same transform is also performed on the secondary data $\mathbf{Y}_k$, $k = 1, 2, \dots, K$, where K is the number of adjacent iid cells, to obtain

the angle-Doppler domain secondary data $\mathcal{Y}_k, N_t \times N_s, k = 1, 2, \dots, K$.

In practice, only the few angle-bins covering the angle section centered at the broadside of the array (i.e., around the look direction where most of the transmitted energy is contained) need to be tested, while at most all Doppler-bins should be tested as the target Doppler frequency shift is unknown to the processor. Let N_{s0} be the number of angle-bins of interest. The $N_t \times N_s$ bins to be tested will be divided into L groups, each of which contains N_{s0} angle-bins and a small number of adjacent Doppler-bins. An example for this grouping is given in Fig. 2.5 where $N_t = 24, N_s = 12$, and $N_{s0} = 3$. We note that the number of Doppler-bins in each group needs not be the same and that some overlap can also be justified. The purpose of dividing along the Doppler axis is to avoid the use of an adaptive processor with large degrees of freedom, which demands a large training-data set as well as a large amount of computation. This opportunity of "divide-and-conquer" is, of course, made available by the multidimensional transformation from the space-time data domain to the angle-Doppler domain, which decouples the degrees of freedom necessary for handling complicated clutter/interference, from the number of data points to be processed. Based on our experience gained from the work in [7, 8], the number of bins in each group is expected to have only minor influence on the detection performance and should be in the range of $2 \times N_{s0} \sim 4 \times N_{s0}$ in general. The angle-Doppler domain secondary data $\mathcal{Y}_k, k = 1, 2, \dots, K$ should be grouped in the same way.

Let N_{tl} be the number of Doppler-bins and $N_l = N_{tl} \times N_{s0}$ the total number of angle-Doppler bins in the l th group. An N_l th-order GLR processor will perform the threshold detection on the N_l bins of the l th group with the test statistic

$$\eta_{nm}^{(l)} = \frac{|\text{Vec}(\mathcal{S}_{nm}^{(l)})^H \hat{\mathbf{R}}_l^{-1} \text{Vec}(\mathcal{X}_l)|^2}{\text{Vec}(\mathcal{S}_{nm}^{(l)})^H \hat{\mathbf{R}}_l^{-1} \text{Vec}(\mathcal{S}_{nm}^{(l)}) [1 + \text{Vec}(\mathcal{X}_l^H) \hat{\mathbf{R}}_l^{-1} \text{Vec}(\mathcal{X}_l)]} \stackrel{H_1}{\underset{H_0}{>}} \eta_0^{(l)} \quad n = 1, 2, \dots, N_{tl} \quad m = 1, 2, \dots, N_{s0} \quad (2.31)$$

where

$$\hat{\mathbf{R}}_l = \sum_{k=1}^K \text{Vec}(\mathcal{Y}_{lk}) \text{Vec}(\mathcal{Y}_{lk})^H, \quad (2.32)$$

and $\mathcal{S}_{nm}^{(l)}, N_{tl} \times N_{s0}$, is the signal-steering matrix in the angle-Doppler domain for the nm th bin of the l th GLR. For a uniform PRF and array spacing, it is easy to see that $\mathcal{S}_{nm}^{(l)}$ has all its entries equal to zero except the nm th one which is $\sqrt{N_t N_s}$. We note that the threshold $\eta_0^{(l)}$ need not be the same across the L groups as evidenced in Subsection 2.4.2 below.

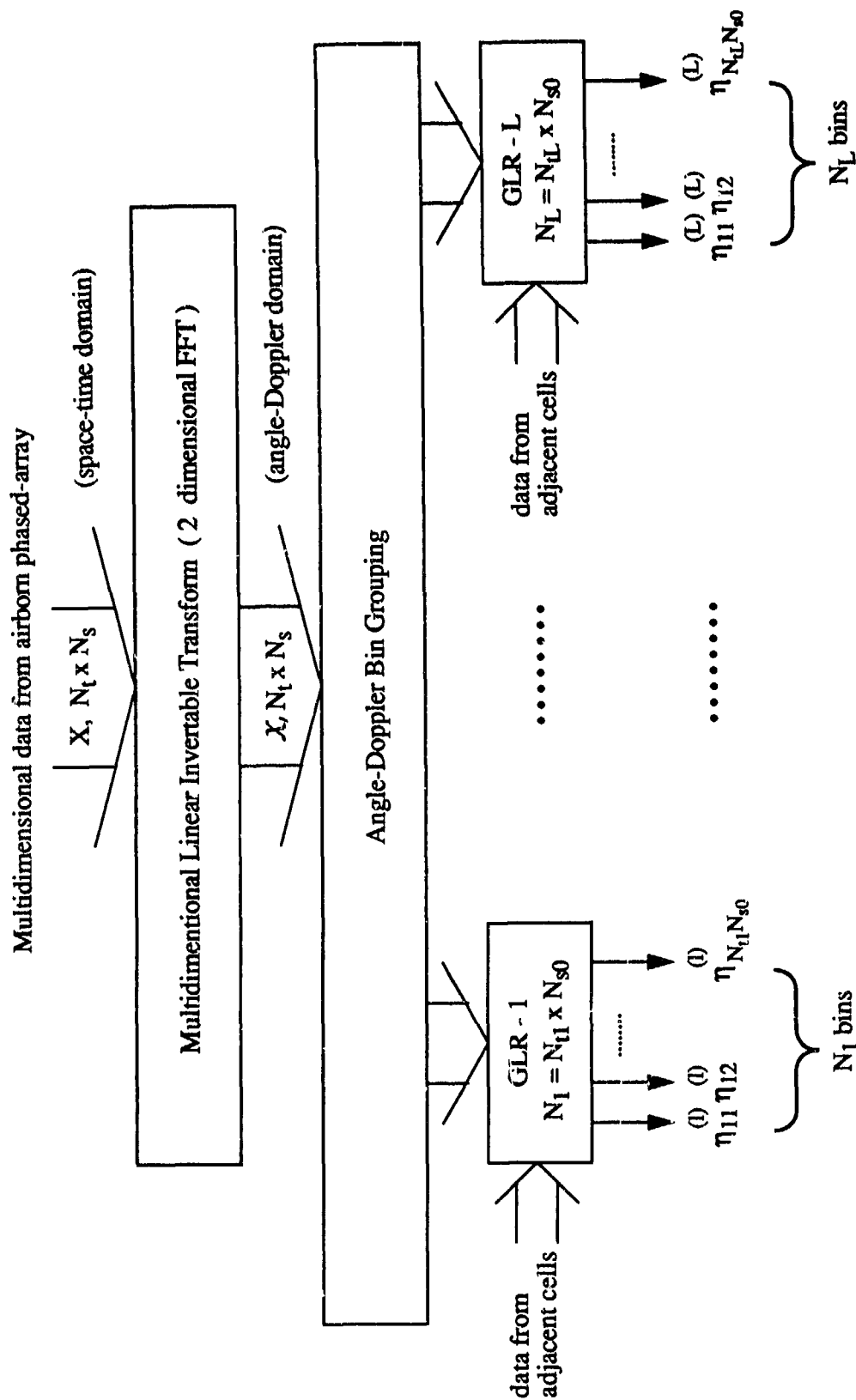


Figure 2.4. Blockdiagram for illustration of the principle of the Joint-Domain Localized GLR Processor.

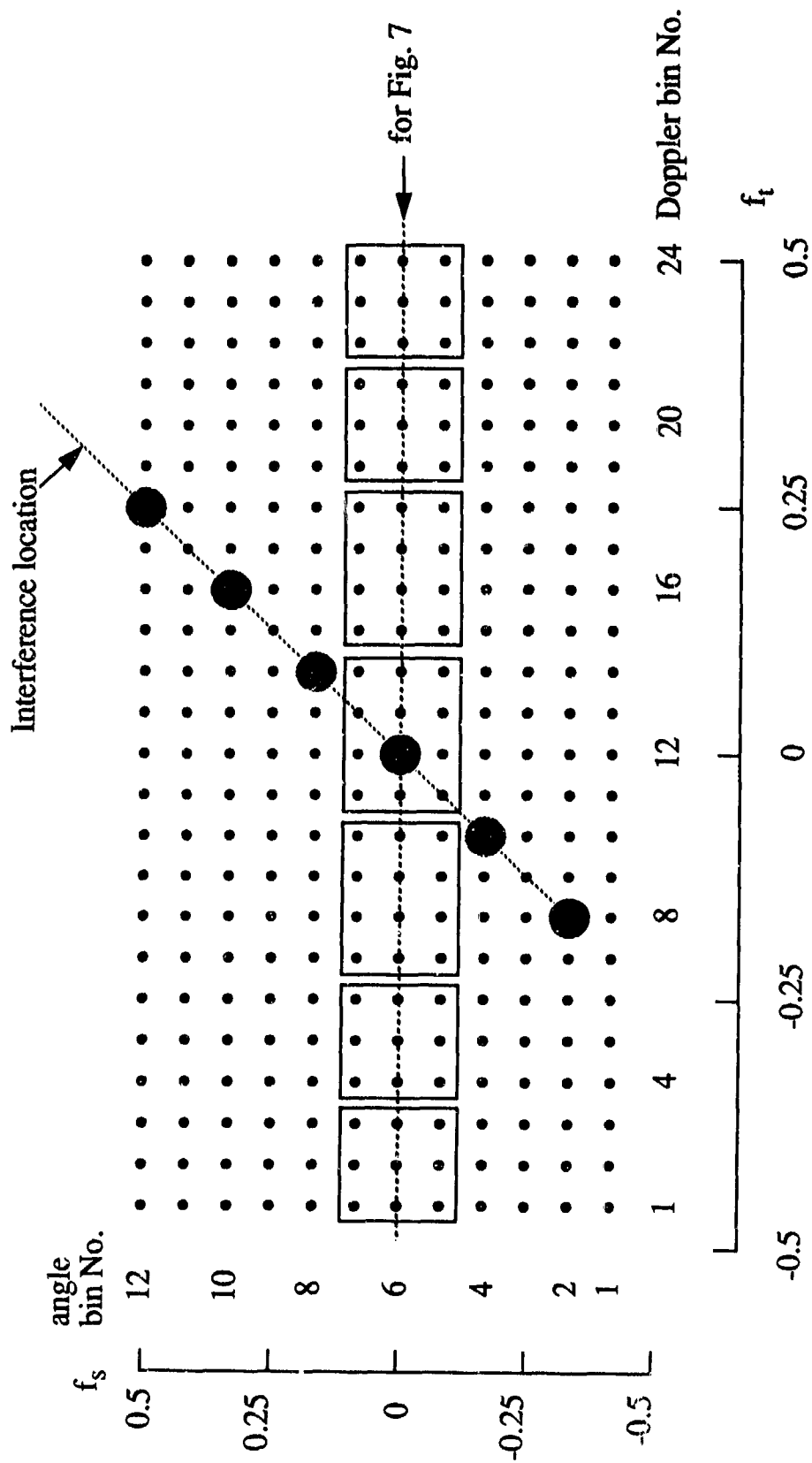


Figure 2.5. An JDL-GLR example.

2.4.2 The JDL-GLR Detection Performance

The detection performance of the original GLR in Gaussian clutter/interference is given in [3, 4] with deterministic modeling and in [12] with stochastic target modeling. As for the Doppler domain localized GLR of [7, 8], it is straightforward to extend the results in [3, 4, 12] to obtain the probabilities of detection and false alarm, P_d and P_f , of the JDL-GLR with both target models. Below we just list the results for the case of non-fluctuating targets with the trivial derivation omitted.

The probability of detection at the nm th bin of the l th GLR is found to be

$$P_d^{(l)}(n, m) = \int_1^0 P_{d|\rho}^{(l)}(n, m) f_{nm}^{(l)}(\rho) d\rho \quad (2.33)$$

where

$$P_{d|\rho}^{(l)}(n, m) = 1 - (1 - \eta_0^{(l)})^{K-N_l+1} \sum_{k=1}^{K-N_l+1} \binom{K-N_l+1}{k} \times \left(\frac{\eta_0^{(l)}}{1 - \eta_0^{(l)}} \right)^k e^{-\rho \beta_{nm}^{(l)} (1 - \eta_0^{(l)})} \sum_{m=0}^{K-1} \frac{[\rho \beta_{nm}^{(l)} (1 - \eta_0^{(l)})]^m}{m!}, \quad (2.34)$$

$$f_{nm}^{(l)}(\rho) = \frac{(K)!}{(K - N_l + 1)!(N_l - 2)!} \rho^{K-N_l+1} (1 - \rho)^{N_l-2}, \quad (2.35)$$

and

$$\beta_{nm}^{(l)} = |\alpha|^2 \text{Vec}(\mathcal{S}_{nm}^{(l)})^H \mathcal{R}_l^{-1} \text{Vec}(\mathcal{S}_{nm}^{(l)}), \quad (2.36)$$

with $\mathcal{R}$ being the covariance matrix of $\text{Vec}(\mathcal{X}_l)$.

The probability of false alarm for all bins in the l th GLR is given by

$$P_f^{(l)} = (1 - \eta_0^{(l)})^{K-N_l+1}. \quad (2.37)$$

Obviously the probability of false alarm can be made equal across the L groups by choosing different $\eta_0^{(l)}, l = 1, 2, \dots, L$. Eq.(2.37) also clearly indicates that, like the original GLR and the Doppler-domain localized GLR, the JDL-GLR has the "integrated/embedded" CFAR feature as $P_f^{(l)}, l = 1, 2, \dots, L$ do not depend on the covariance of the clutter/interference.

2.4.3 Detection Performance Comparison

Although the convergence-rate advantage of the JDL-GLR can be seen intuitively from the fact that the localized GLR's have much lower degrees of freedom than a high-order GLR directly applied to the space-time domain data, the numerical example below should demonstrate this advantage clearly.

Consider a system with $N_s = 12$ and $N_t = 24$. The clutter/interference is assumed to have the two-dimensional multipeak Gaussian-shaped power spectrum density (psd) as shown in Fig. 2.6. For convenience of reference we have also indicated the center locations of this multipeak spectrum in Fig. 2.5. The exact expression of this psd is given by

$$P_c(f_t, f_s) = \sum_{d=1}^6 \sigma_{cd}^2 \frac{1}{2\pi\sigma_{ft}\sigma_{fs}} \exp\left[-\left(\frac{(f_t - f_{ctd})^2}{2\sigma_{ft}^2} + \frac{(f_s - f_{csd})^2}{2\sigma_{fs}^2}\right)\right] \quad (2.38)$$

where σ_{cd}^2 is the power of the d th component. Obviously, the total clutter/interference power σ_c^2 is

$$\sigma_c^2 = \sum_{d=1}^6 \sigma_{cd}^2. \quad (2.39)$$

We set $\sigma_{c1}^2 = \sigma_{c2}^2 = \sigma_{c4}^2 = \sigma_{c5}^2 = \sigma_{c6}^2 = \sigma_{c3}^2/10^{2.5}$, $\text{INR}=50$, and $\text{SNR}=0\text{dB}$ which gives $\text{SINR} \simeq -50\text{dB}$. The thresholds for the processors to be compared are such that every processor has a probability of false alarm $P_f = 10^{-5}$ at each tested bin. We assume that there are $K = 24$ adjacent cells from which the iid secondary data set is obtained.

Consider the following five processors:

- (1) the joint-domain optimal,
- (2) the JDL-GLR with $L = 7$ localized GLR processors with their coverage shown in Fig. 2.5,
- (3) the T-S configuration with the optimal processor for each part,
- (4) the S-T configuration with the optimal processor for each part,
and
- (5) a conventional beamformer followed by the optimal temporal processor (i.e., the optimal MTI).

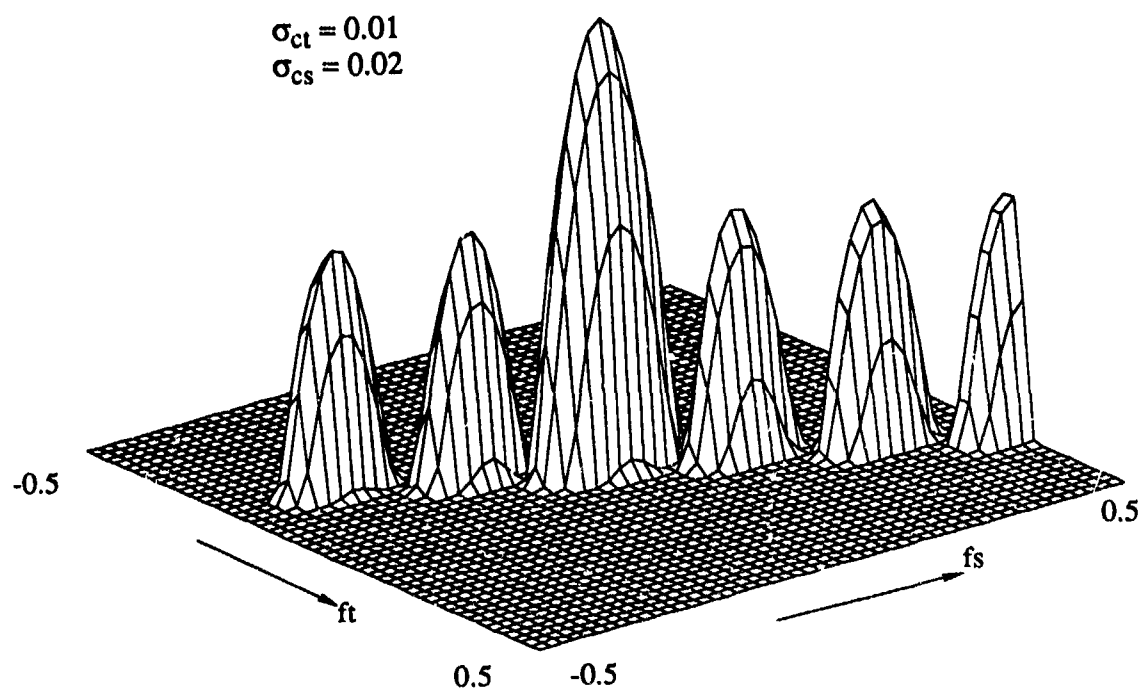


Figure 2.6. Two dimensional power spectral density for the clutter/interference used in the example.

We note that with $N_s = 12$, $N_t = 24$ but $K = 24$ only, any straightforward adaptive implementation of the joint-domain optimal, any adaptive processor with the S-T configuration, and any adaptive processor with the T-S configuration will fail to deliver an acceptable detection performance for this example since $K = 24$ is too small with respect to their degrees of freedom. Therefore, these adaptive processors are excluded from the above list for detailed comparison.

Fig. 2.7 shows the probability of detection of the five processors listed at the 6th angle bin which is the assumed angle of arrival of the target signal. Obviously, the JDL-GLR is the only one that approaches the joint-domain optimal, except at few bins adjacent to the center of the strongest clutter/interference spectrum component. The poor performance of the two optimal cascade configurations should not be a surprise from the discussion in Section 2.3. The fact shown in Fig. 2.7 that the ad. hoc processor of No. 5 can outperform them (especially the optimal S-T configuration) is also a strong evidence that the optimality does not always mean much with a wrong configuration. Of course, the poorest performance of the optimal S-T configuration is due to the fact that the optimal spatial part of processing nulls the clutter/interference as well as the target signal. Finally, we comment that a CFAR loss is inevitably associated with any adaptive implementation of the four optimal/partially optimal processor in Fig. 2.7, while the embedded CFAR feature of the JDL-GLR makes any other additional CFAR processing unnecessary.

2.4.4 Other Features of JDL-GLR

The computation advantage of the JDL-GLR is clear. Recall that the N -th order GLR has a computation load proportional to N^3 . Assume that each localized GLR spans three angle-bins and four Doppler bins and that $N_t/4$ localized GLR are required. This leads to a computation load proportional to $(N_t/4)(3 \times 4)^3 = 432N_t$ for the JDL-GLR. With a load of $N_t^3 N_s^3$ for the straightforward application of the GLR to the space-time domain data, the JDL-GLR will show a computation advantage when $N_t > 4$ and $N_s > 3$. For large N_t and N_s the JDL-GLR offers a computation load reduction by a factor of

$$\gamma = N_t^2 N_s^3 / 432. \quad (2.40)$$

For the example of $N_t = 24$ and $N_s = 12$ in this section, the JDL-GLR's computation load is only 1/2304 of that for the straightforward application of the GLR (or SMI) to the space-time domain data. Like the Doppler-domain localized GLR in [7, 8], the JDL-GLR can further reduce its computation load via deleting the localized GLR processors for the region where the

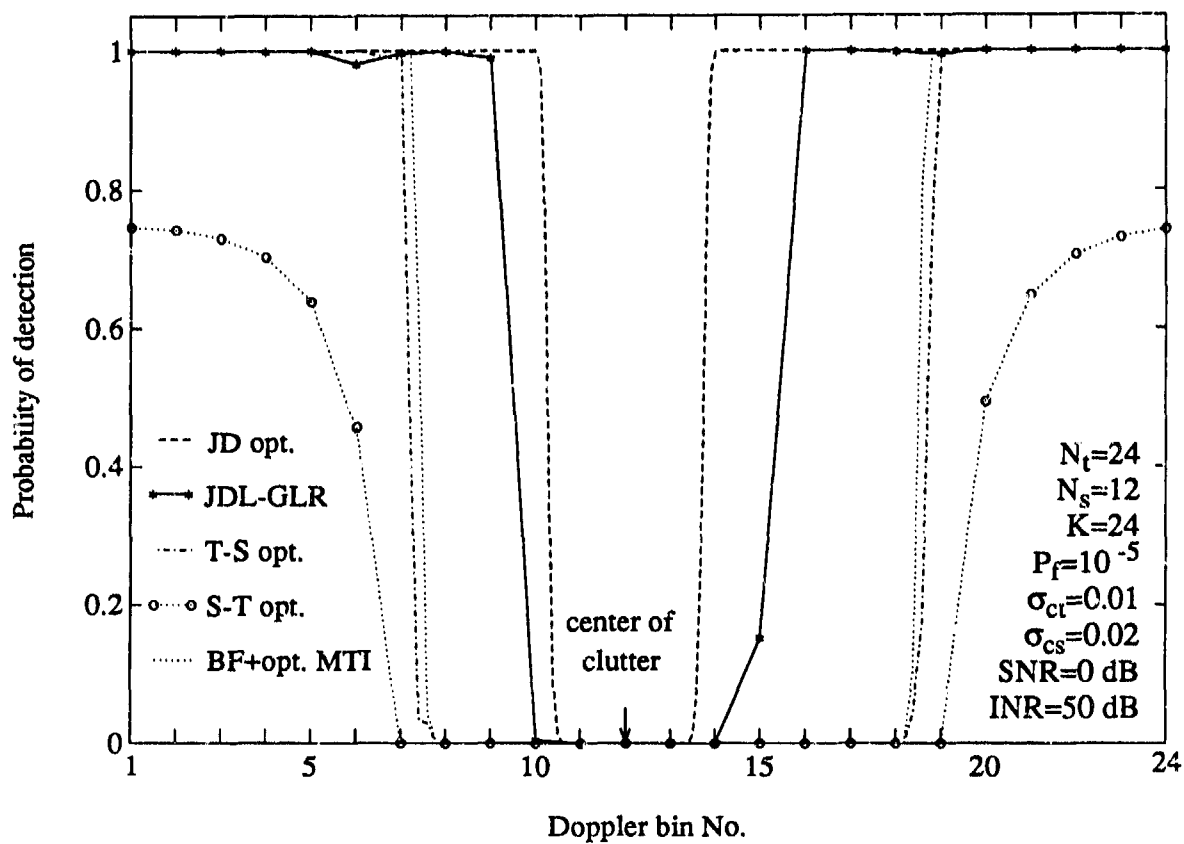


Figure 2.7. Detection performance comparison of the five processors.

detection performance improvement is unnecessary or impossible. This can be done when some *a priori* information is available about the power concentration of the clutter/interference in the angle-Doppler domain. Furthermore, the realization of the JDL-GLR benefits from the available parallel processing techniques as its localized GLRs all operate in parallel.

Since the robustness feature in non-Gaussian clutter/interference resides with the GLR processor which will not be affected by the linear transformation, the JDL-GLR is expected to maintain its robustness. Computationally intensive simulation is being conducted to confirm this feature and the result will be published separately [13].

2.5 Conclusions and Discussion

This chapter shows:

- (1) neither of the two cascade configurations is better than the other;
- (2) the performance potential of both cascade configurations can fall far below that of the joint-domain optimum processor; and
- (3) the Joint-Domain Localized GLR algorithm (JDL-GLR) offers an attractive solution to the problem of approaching the performance potential of the joint-domain optimum processor of a high order ($N_s \times N_t$) with a fast convergence rate and high computation efficiency, together with such highly desirable features as the embedded CFAR and robustness in non-Gaussian clutter/interference.

Finally, we would like to point out that both cascade configurations may have considerable difficulty to achieve a high quality CFAR in practice when both spatial and temporal parts are adaptive. This is because of the random modulation introduced by the adaptive algorithm for the early part of the cascaded two parts. The problem may become more severe in highly non-stationary and nonhomogeneous environments where there is a shortage of a sufficient amount of iid training data to smooth out the extra random modulation. In contrast, the JDL-GLR presented in this chapter is free of such random modulation and can maintain its CFAR performance with a much smaller amount of iid training data. Simulation-based comparison of the CFAR performance of adaptive spatial-temporal processors, can be found in [13].

Chapter 3

Literature Review on Spherically Invariant Random Processes

3.1 Introduction

We present an overview of the literature as it pertains to the modeling of radar clutter by spherically invariant random processes. In addition, relevant mathematical preliminaries are presented in this chapter. When a radar transmits a signal, the received echo may consist of returns from one or more targets, buildings, trees, water, land and weather depending on the environment. The target returns contribute to the desired signal while the other returns contribute to the clutter. Many investigators [14, 15, 16, 17] have reported experimental measurements for which the clutter probability density function has an extended tail. The extended tail gives rise to relatively large probabilities of false alarm. The Gaussian model for the clutter fails to predict this behavior. Two approaches have been used to explain the non-Gaussian behavior. One of them is based on the fact that the assumptions under the central limit theorem (CLT) may fail. The other approach is based on the nonstationary reflectivity properties of the scanned areas. In any event, non-Gaussian models for the univariate (marginal) clutter PDF have been proposed. Commonly reported marginal non-Gaussian PDFs for the clutter are Weibull [14], Log-normal [18, 19] and K-distributions [16, 20, 15]. Second order statistics for these models have been reported in terms of autocorrelation functions or power spectral densities [21, 17].

The Weibull [14] and Log-normal [15] models for radar clutter are primarily based on empirical studies, while the K-distribution has been shown to have physical significance [22, 15] in that the observed statistical properties can be related to the electromagnetic and geometric factors

pertaining to the scattering surface. Computer simulation schemes for Weibull and Log-normal clutter based on the univariate PDFs and correlation functions have been developed in [23] and [24], respectively. Extension of the Weibull and Log-normal and K-distributed clutter models for coherent radar processing have been developed in [25, 18, 26] respectively.

Statistical characterization of the clutter is necessary in order to obtain the optimal radar signal processor. Usually, radars process N pulses at a time. A complete statistical characterization of the clutter requires the specification of the joint probability density function (PDF) of the N samples. When the pulse returns are statistically independent, the joint PDF is simply the product of the marginal PDFs. However, the clutter can be highly correlated. In fact, the correlation between samples is useful in canceling the clutter. Consequently, it is desirable to include the correlation information in the multivariate PDF. For non-Gaussian processes this can be done in more than one way. The theory of spherically invariant random processes (SIRP) provides a powerful mechanism for obtaining the joint PDF of the N correlated non-Gaussian random variables. Applications for the theory of SIRPs can be found in the problem of random flights [27], signal detection and estimation problems in communication theory [28, 29], speech signal processing [30, 31], radar clutter modeling and simulation [32, 26, 33, 34, 35]. The following sections provide a brief overview of literature on the theory of SIRPs.

3.2 Definitions

In this section we present certain definitions and mathematical preliminaries pertaining to the theory of SIRPs. A random vector $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$ is said to be a spherically invariant random vector (SIRV) if its PDF has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = k|\Sigma|^{-\frac{1}{2}}h_N[(\mathbf{y} - \mathbf{b})^T\Sigma^{-1}(\mathbf{y} - \mathbf{b})] \quad (3.1)$$

where k is a normalization chosen so that the volume under the curve of the PDF is unity, $\mathbf{b}$ is a N by 1 vector, Σ is a N by N non-negative definite matrix, and $h_N(\cdot)$ is a one dimensional, positive, real valued monotonically decreasing function. Note that the PDF of an SIRV is elliptically symmetric (i.e., constant contours of $f_{\mathbf{Y}}(\mathbf{y})$ are composed of ellipses). If every random vector obtained by sampling a random process $y(t)$ is a spherically invariant random vector, regardless of the sampling instants or the number of samples, then the process $y(t)$ is defined to be a spherically invariant random process (SIRP).

Kingman [27] introduced the definition of spherically symmetric random vectors (SSRV). In

particular, a random vector $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$ is said to be spherically symmetric provided its PDF has the form

$$f_{\mathbf{X}}(\mathbf{x}) = kh_N[(x_1^2 + x_2^2 + \dots + x_N^2)^{\frac{1}{2}}] = kh_N(\mathbf{x}^T \mathbf{x}) \quad (3.2)$$

where $h_N(\cdot)$ is an arbitrary, non-negative, monotonically decreasing radial function of dimension N and k is a normalization constant chosen so that the volume under the curve of the PDF is unity. The subscript N is used to emphasize that we are dealing with N random variables. Throughout the manuscript, it is assumed that the PDF of a random vector is the joint PDF of its components. Equivalently, if $\omega = [\omega_1, \omega_2, \dots, \omega_N]^T$, the characteristic function of the SSRV $\mathbf{X}$ defined by $\Phi_{\mathbf{X}}(\omega) = E[\exp(j\omega^T \mathbf{x})]$, has the form

$$\Phi_{\mathbf{X}}(\omega) = g_N[(\omega_1^2 + \omega_2^2 + \dots + \omega_N^2)^{\frac{1}{2}}] \quad (3.3)$$

where $g_N(\cdot)$ is a non-negative conjugate symmetric function which is magnitude integrable. An SSRV is a special case of an SIRV, arising from eq (3.1) when $\mathbf{b} = 0$ and $\Sigma = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. In Appendix A, we prove that the characteristic function of an SSRV is also spherically symmetric.

3.3 Characterization of SIRPs

In this section we present some important theorems that help us to characterize the PDF of a SIRV. The work of Yao [28] and Kingman [36] gave rise to a representation theorem for SSRVs. The representation theorem can be stated as follows.

Theorem 1 *If a random vector $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$ is an SSRV for any N , then there exists a non-negative random variable T such that the random variables X_i , ($i = 1, 2, \dots, N$) conditioned on $T = t$ are independent, identically distributed, Gaussian random variables with zero mean and variance equal to $2t$.*

Proof: By definition, the characteristic function of $\mathbf{X}$ is

$$\begin{aligned} \Phi_{\mathbf{X}}(\omega) &= E[\exp(j\omega^T \mathbf{X})] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(j\omega^T \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (3.4)$$

The PDF on the random variable T is introduced by noting that

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_{-\infty}^{\infty} f_{\mathbf{X},T}(\mathbf{x},t)dt \\ &= \int_{-\infty}^{\infty} f_{\mathbf{X}|T}(\mathbf{x}|t)f_T(t)dt. \end{aligned} \quad (3.5)$$

Substituting into the expression for the characteristic function and interchanging the order of integration we obtain

$$\Phi_{\mathbf{X}}(\omega) = \int_{-\infty}^{\infty} \Phi_{\mathbf{X}|T}(\omega, t)f_T(t)dt \quad (3.6)$$

where

$$\Phi_{\mathbf{X}|T}(\omega, t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(j\omega^T \mathbf{x}) f_{\mathbf{X}|T}(\mathbf{x}|t) d\mathbf{x}. \quad (3.7)$$

Since $\mathbf{X}$ is an SSRV for any N , its characteristic function has the form of eq (3.3). This requires that $\Phi_{\mathbf{X}|T}(\omega, t)$ also be a function of $(\omega_1^2 + \omega_2^2 + \dots + \omega_N^2)$ for any choice of N . The only characteristic function having this property [36] is

$$\Phi_{\mathbf{X}|t}(\omega, t) = \exp[-t(\omega_1^2 + \omega_2^2 + \dots + \omega_N^2)] \quad (3.8)$$

where the conditional PDF of $\mathbf{X}$, given $T = t$, is recognized to be multivariate Gaussian, with X_i , ($i = 1, 2, \dots, N$) being statistically independent identically distributed, zero mean Gaussian random variables with variance $2t$. Because the variance equals $2t$, T must be a non-negative random variable. This establishes the theorem. Note that the theorem does not give any physical significance for T . Neither does it reveal how to determine $f_T(t)$.

The representation theorem for SSRVs allows us to write the random vector $\mathbf{X}$ as a product of a Gaussian random vector $\mathbf{Z}$ having zero mean and identity covariance matrix and a non-negative random variable $S = \sqrt{2T}$ with PDF $f_S(s)$. In particular, consider the product $\mathbf{X} = \mathbf{Z}S$. S is assumed to be non-negative for convenience. The PDF of $\mathbf{X}$ conditioned on S is then given by

$$f_{\mathbf{X}|S}(\mathbf{x}|s) = (2\pi)^{-\frac{N}{2}} s^{-N} \exp\left(-\frac{\mathbf{p}'}{2s^2}\right) \quad (3.9)$$

where $\mathbf{p}' = \mathbf{x}^T \mathbf{x}$. From the theorem on total probability, the PDF of $\mathbf{X}$ can be written as

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} \int_0^{\infty} s^{-N} \exp\left(-\frac{\mathbf{p}'}{2s^2}\right) f_S(s) ds. \quad (3.10)$$

Comparing eqs (3.10) and (3.2), we can write $k = (2\pi)^{-\frac{N}{2}}$ and

$$h_N(p') = \int_0^\infty s^{-N} \exp\left(-\frac{p'}{2s^2}\right) f_S(s) ds. \quad (3.11)$$

Thus, it is clear that the PDF of an SSRV is uniquely determined by the specification of a Gaussian random vector having zero mean and identity covariance matrix and a first order PDF $f_S(s)$ called the characteristic PDF.

The following theorem in [37] states that a SIRV is related to an SSRV by a linear transformation.

Theorem 2 *If $\mathbf{X}$ is an SSRV, with characteristic PDF $f_S(s)$, then the deterministic linear transformation*

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} \quad (3.12)$$

results in $\mathbf{Y}$ being an SIRV having mean vector $\mathbf{b}$, covariance matrix $\Sigma = \mathbf{A}\mathbf{A}^T$ and the same characteristic PDF. It is required that $\mathbf{A}\mathbf{A}^T$ be nonsingular.

Proof: Since $\mathbf{X}$ is an SSRV, we can express $\mathbf{X}$ as $\mathbf{X} = \mathbf{Z}S$, where $\mathbf{Z}$ is a Gaussian random vector having zero mean and identity covariance matrix and S is a non-negative random variable. Hence,

$$\mathbf{Y} = \mathbf{A}\mathbf{Z}S + \mathbf{b}. \quad (3.13)$$

Conditioned on S , the PDF of $\mathbf{Y}$ is Gaussian, with mean vector equal to $\mathbf{b}$ and covariance matrix equal to $\mathbf{A}\mathbf{A}^T s^2$. The PDF of $\mathbf{Y}$ conditioned on S is given by

$$f_{\mathbf{Y}|S}(\mathbf{y}|s) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} s^{-N} \exp\left(-\frac{p}{2s^2}\right) \quad (3.14)$$

where $p = (\mathbf{y} - \mathbf{b})^T \Sigma^{-1} (\mathbf{y} - \mathbf{b})$ and $|\Sigma|$ denotes the determinant of the covariance matrix $\Sigma = \mathbf{A}\mathbf{A}^T$. Using the theorem on total probability, the PDF of $\mathbf{Y}$ can be written as

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} h_N(p) \quad (3.15)$$

where

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds. \quad (3.16)$$

The PDF of $\mathbf{Y}$ is of the form of eq (3.1). Therefore, $\mathbf{Y}$ is an SIRV. The PDF of an SIRV is uniquely determined by the specification of a mean vector, a covariance matrix and a first order PDF called the characteristic PDF. Theorem 1 for SSRVs generalizes for SIRVs in a straightforward manner. The only difference is that conditioned on the non-negative random variable T , the $\{Y_k : (k = 1, 2, \dots, N)\}$ are no longer statistically independent. Instead, the PDF of $\mathbf{Y}$ conditioned on T is a multivariate Gaussian PDF. By the same argument used for SSRVs, an SIRV can be written as a product of a Gaussian random vector and a non-negative random variable. The only difference is that the mean of the Gaussian random vector need not be zero and its covariance matrix is not the identity matrix. As a corollary of Theorem 2 [28], it can be readily shown that every linear transformation on an SIRV results in another SIRV having the same characteristic PDF. As a special case, when $f_S(s) = \delta(s - 1)$ where $\delta(\cdot)$ is the unit impulse function, $h_N(p) = \exp(-\frac{p}{2})$ and the corresponding SIRV PDF given by eq (3.15) is the multivariate Gaussian PDF. Therefore, the multivariate Gaussian PDF is a special case of the SIRV PDF.

The following theorem from [29] provides an interesting property of SSRVs when represented in generalized spherical co-ordinates $R \in (0, \infty)$, $\Theta \in (0, 2\pi)$ and $\Phi_k \in (0, \pi)$, $(k = 1, \dots, N - 2)$.

Theorem 3 *When the components of the random vector $\mathbf{X} = [X_1 \dots X_N]^T$ are represented in the generalized spherical coordinates given by*

$$\begin{aligned} X_1 &= R \cos(\Phi_1) \\ X_k &= R \cos(\Phi_k) \prod_{i=1}^{k-1} \sin(\Phi_i) \quad (1 < k \leq N - 2) \\ X_{N-1} &= R \cos(\Theta) \prod_{i=1}^{N-2} \sin(\Phi_i) \\ X_N &= R \sin(\Theta) \prod_{i=1}^{N-2} \sin(\Phi_i), \end{aligned} \tag{3.17}$$

$\mathbf{X}$ is an SSRV if and only if R , Θ and Φ_k are mutually and statistically independent random variables having PDFs of the form

$$\begin{aligned} f_R(r) &= \frac{r^{N-1}}{2^{\frac{N-1}{2}} \Gamma(\frac{N}{2})} h_N(r^2) u(r) \\ f_{\Phi_k}(\phi_k) &= \frac{\Gamma(\frac{N-k+1}{2})}{\sqrt{\pi} \Gamma(\frac{N-k}{2})} \sin^{N-1-k}(\phi_k) [u(\phi_k) - u(\phi_k - \pi)] \\ f_{\Theta}(\theta) &= (2\pi)^{-1} [u(\theta) - u(\theta - 2\pi)] \end{aligned} \tag{3.18}$$

where $\Gamma(\cdot)$ is the Eulero Gamma function and $u(\cdot)$ is the unit step function.

Proof: Since the random vector $\mathbf{X}$ is an SSRV, its PDF is of the form of eq (3.2) with $h_N(p')$ being given by eq (3.11). The Jacobian of the transformation given by eq (3.17) is obtained in [38] as

$$J = (R^{N-1} \prod_{k=1}^{N-2} \sin^{N-1-k}(\phi_k))^{-1}. \quad (3.19)$$

Using eq (3.2) and eq (3.19) and noting that $R^2 = \sum_{k=1}^N X_k^2$, the joint PDF of R , Θ and Φ_k ($k = 1, 2, \dots, N-2$) becomes

$$f_{R,\Theta,\Phi_1,\dots,\Phi_{N-2}}(r, \theta, \phi_1 \dots \phi_{N-2}) = \frac{r^{N-1}}{(2\pi)^{\frac{N}{2}}} h_N(r^2) \prod_{k=1}^{N-2} \sin^{N-1-k}(\phi_k) \quad (3.20)$$

Since the joint PDF in eq (3.20), can be written as a product of the marginal PDFs given in eq (3.18), the variables R , Θ and Φ_k , are mutually and statistically independent with the prescribed PDFs. In order to prove the sufficient part of the property, we start with the marginal PDFs of R , Θ and Φ_k given by eq (3.18) and, under the assumption of statistical independence, obtain the joint PDF of eq (3.20). Using the inverse Jacobian of that given by eq (3.19), results in the PDF of $\mathbf{X}$ being given by eq (3.2).

3.4 Determining the PDF of an SIRV

In this section we shall present schemes for determining the PDF of an SIRV. We recognize that the PDF of an SIRV is uniquely determined by the specification of a mean vector, a covariance matrix and a characteristic first order PDF and that the SIRV PDF has the form of eq (3.15). Several techniques are available in the literature for specifying $h_N(p)$. The simplest technique is to use eq (3.16). However, this procedure requires the knowledge of the characteristic PDF $f_S(s)$. Therefore, when $f_S(s)$ is not known in closed form or it is difficult to evaluate the integral in eq (3.16), alternate methods for specifying $h_N(p)$ must be examined.

To study the behavior of $h_N(p)$, it is convenient to replace p , which is a quadratic form depending on N , by the dummy scalar variable w . We then write

$$h_N(w) = \int_0^\infty s^{-N} \exp(-\frac{w}{2s^2}) f_S(s) ds. \quad (3.21)$$

When both sides of eq (3.21) are differentiated with respect to w , we obtain

$$\frac{dh_N(w)}{dw} = -\frac{1}{2} \int_0^\infty s^{-N-2} \exp\left(-\frac{w}{2s^2}\right) f_S(s) ds. \quad (3.22)$$

The right hand side of eq (3.22) is related to $h_{N+2}(w)$ by the factor of $-\frac{1}{2}$. Thus, we have an interesting result pointed out in [32] that

$$h_{N+2}(w) = (-2) \frac{dh_N(w)}{dw}. \quad (3.23)$$

Because

$$f_Y(y) = (2\pi)^{-\frac{N+2}{2}} |\Sigma|^{-\frac{1}{2}} h_{N+2}(p) \quad (3.24)$$

when Y is of dimension $N+2$, it follows that $h_N(w)$ must be a monotonically decreasing function for all N . Eq (3.23) provides a mechanism for relating higher order PDFs with those of lower order. More precisely, starting with $N=1$ and $N=2$, and using eq (3.23) repeatedly, gives the following pair of recurrence relations.

$$\begin{aligned} h_{2N+1}(w) &= (-2)^N \frac{d^N h_1(w)}{dw^N} \\ h_{2N+2}(w) &= (-2)^N \frac{d^N h_2(w)}{dw^N}. \end{aligned} \quad (3.25)$$

Therefore, starting from $h_1(w)$ and $h_2(w)$ all PDFs of odd and even order, respectively, can be generated by the use of eq (3.25). However, since $h_N(\cdot)$ is defined to be a non-negative monotonically decreasing function for all N , $h_1(\cdot)$ and $h_2(\cdot)$ must belong to a class of functions that are positive and monotonically decreasing. Consequently, their successive derivatives will alternate between negative and positive functions that are monotonically increasing and decreasing, respectively. Given $h_N(w)$, the N^{th} order SIRV PDF is given by

$$f_Y(y) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} h_N(p) \quad (3.26)$$

where $h_N(p)$ is nothing more than $h_N(w)$ with w replaced by p .

Another approach for specifying $h_N(p)$ that begins with the univariate characteristic function has been proposed in [39, 28, 29]. It is required that the univariate characteristic function be a real even function whose magnitude is integrable. Also, it is assumed that the components of

the SIRV are identically distributed. Under these conditions, it has been shown that

$$h_N(p) = (\sqrt{p})^{1-\frac{N}{2}} \int_0^\infty \omega^{\frac{N}{2}} \phi(\omega) J_{\frac{N-1}{2}}(\omega\sqrt{p}) d\omega \quad (3.27)$$

where $\phi(\omega)$ is the univariate characteristic function and $J_\alpha(\eta)$ is the Bessel function of order α . Eq (3.27) has an elegant proof by induction which is presented here. From eq (3.15) it follows that $h_1(p)$ is related to the first order SIRV PDF of the i^{th} component. More explicitly, we can write

$$f_{Y_i}(y_i) = (\sqrt{2\pi}\sigma)^{-1} h_1(p_i) \quad (i = 1, 2, \dots, N) \quad (3.28)$$

where $p_i = \frac{y_i^2}{\sigma^2}$ and σ^2 is the common variance of the random variables Y_i ($i = 1, 2, \dots, N$). For convenience, assume that σ^2 is unity. The univariate characteristic function is then given by

$$\phi_i(\omega) = \int_{-\infty}^\infty f_{Y_i}(y_i) \exp(j\omega y_i) dy_i. \quad (3.29)$$

Using the inverse Fourier transform and noting that $y_i = \sqrt{p_i}$, $h_1(p_i)$ can be expressed in terms of the characteristic function as

$$h_1(p_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \phi_i(\omega) \exp(-j\omega\sqrt{p}) d\omega. \quad (3.30)$$

Since $\phi_i(\omega)$ is the same for all i , the subscript i in eq (3.30) can be dropped. In addition, because $\phi(\omega)$ is an even function, we can rewrite eq (3.30) as

$$h_1(p) = \sqrt{\frac{2}{\pi}} \int_0^\infty \phi(\omega) \cos(\omega\sqrt{p}) d\omega. \quad (3.31)$$

Recognizing that $\cos(x) = \sqrt{\frac{\pi x}{2}} J_{-\frac{1}{2}}(x)$, and replacing p by the dummy variable w , we have

$$h_1(w) = (\sqrt{w})^{\frac{1}{2}} \int_0^\infty \omega^{\frac{1}{2}} \phi(\omega) J_{-\frac{1}{2}}(\omega\sqrt{w}) d\omega. \quad (3.32)$$

Since the derivation makes use of eq (3.23) it is necessary to consider odd and even values of N separately. For odd values of N , eq (3.27) can be written as

$$h_{2N-1}(w) = (\sqrt{w})^{\frac{3}{2}-N} \int_0^\infty \omega^{N-\frac{1}{2}} \phi(\omega) J_{\frac{2N-3}{2}}(\omega\sqrt{w}) d\omega. \quad (3.33)$$

Equation (3.33) is now shown to hold for all N by means of induction. With $N = 1$, eq (3.33)

reduces to eq (3.32). It remains to show that eq (3.33) is valid when N is replaced by $N + 1$. Differentiating both sides of eq (3.33) with respect to w , we obtain

$$\frac{dh_{2N-1}(w)}{dw} = \int_0^\infty \omega^{N-\frac{1}{2}} \phi(\omega) \frac{d}{dw} [(\sqrt{w})^{\frac{3}{2}-N} J_{\frac{2N-3}{2}}(\omega\sqrt{w})] d\omega. \quad (3.34)$$

First, focus on the term $\frac{d}{dw} [(\sqrt{w})^{\frac{3}{2}-N} J_{\frac{2N-3}{2}}(\omega\sqrt{w})]$. Since this involves the derivative of a product, we can write

$$\frac{d}{dw} [(\sqrt{w})^{\frac{3}{2}-N} J_{\frac{2N-3}{2}}(\omega\sqrt{w})] = \frac{1}{2} \left(\frac{3}{2} - N \right) (\sqrt{w})^{\frac{1}{2}-N} J_{\frac{2N-3}{2}}(\omega\sqrt{w}) + (\sqrt{w})^{\frac{3}{2}-N} \frac{d}{dw} [J_{\frac{2N-3}{2}}(\omega\sqrt{w})]. \quad (3.35)$$

Using the identity [40]

$$\frac{dJ_\alpha(\eta)}{d\eta} = \frac{\alpha}{\eta} J_\alpha(\eta) - J_{\alpha+1}(\eta) \quad (3.36)$$

we have

$$\frac{d}{dw} [J_{\frac{2N-3}{2}}(\omega\sqrt{w})] = \frac{\omega}{2} (\sqrt{w})^{-1} \left[\frac{2N-3}{2\omega\sqrt{w}} J_{\frac{2N-3}{2}}(\omega\sqrt{w}) - J_{\frac{2N-1}{2}}(\omega\sqrt{w}) \right]. \quad (3.37)$$

Substituting eq (3.37) in eq (3.35) gives

$$\frac{d}{dw} [(\sqrt{w})^{\frac{3}{2}-N} J_{\frac{2N-3}{2}}(\omega\sqrt{w})] = -\frac{\omega}{2} (\sqrt{w})^{\frac{1}{2}-N} J_{\frac{2N-1}{2}}(\omega\sqrt{w}). \quad (3.38)$$

Consequently, eq (3.34) reduces to

$$\frac{dh_{2N-1}(w)}{dw} = -\frac{1}{2} (\sqrt{w})^{\frac{1}{2}-N} \int_0^\infty \omega^{N+\frac{1}{2}} \phi(\omega) J_{\frac{2N-1}{2}}(\omega\sqrt{w}) d\omega. \quad (3.39)$$

However, from eq (3.23) we know that $h_{2N+1}(w) = (-2) \frac{dh_{2N-1}(w)}{dw}$. Hence, we have from eq (3.39)

$$h_{2N+1}(w) = (\sqrt{w})^{\frac{1}{2}-N} \int_0^\infty \omega^{N+\frac{1}{2}} \phi(\omega) J_{\frac{2N-1}{2}}(\omega\sqrt{w}) d\omega. \quad (3.40)$$

Because eq (3.40) is identical to eq (3.33) with N replaced by $N + 1$, it has been shown by induction that eq (3.33) is valid for all N . It follows that eq (3.27) is valid for all odd values of N .

In a similar manner, starting with $h_2(p)$, it can be shown that

$$h_{2N+2}(p) = \sqrt{p}^{-N} \int_0^\infty \omega^{N+1} \phi(\omega) J_N(\omega\sqrt{p}) d\omega \quad (3.41)$$

for all N . Note that eq (3.41) is identical to eq (3.27) with N replaced by $2N + 2$. The proof of

this result is presented in Chapter 3. Thus, in general, for any N (odd or even), we can write $h_N(p)$ as in eq (3.27).

3.5 Properties of SIRVs

In this section we present certain important properties of SIRVs.

3.5.1 PDF Characterization

The multivariate PDF of an SIRV as given by eqs. (3.15) and (3.16) is uniquely determined by the specification of a mean vector $\mathbf{b}$, a covariance matrix Σ and a characteristic first order PDF $f_S(s)$. The PDF involves a non-negative, real valued monotonically decreasing function $h_N(\cdot)$ of a non-negative quadratic form. The type of SIRV is determined by the form of $h_N(\cdot)$ or, equivalently, the choice of $f_S(s)$. Higher order PDFs can be obtained by the use of eq (3.27) whereas lower order PDFs can be obtained in the usual manner by integrating out the unwanted variables. We discuss this procedure in Appendix A. The PDFs of all orders are of the same type. The marginal PDFs are used to classify the type of SIRV.

3.5.2 Closure Under Linear Transformation

As shown in Theorem 2 of Section 2.3, every linear transformation of the form of eq (3.12) on an SIRV results in another SIRV having the same characteristic PDF. This feature is called the closure property of SIRVs [28, 29].

3.5.3 Minimum Mean Square Error Estimation

In minimum mean square error estimation (MMSE) problems, given a set of data, SIRVs are found to result in linear estimators [39, 28, 41]. An interesting proof of this property is presented here. Let $\mathbf{Y} = [\mathbf{Y}_1^T \mathbf{Y}_2^T]^T$ where $\mathbf{Y}_1 = [Y_1, Y_2, \dots, Y_m]^T$ and $\mathbf{Y}_2 = [Y_{m+1}, Y_{m+2}, \dots, Y_N]^T$ denote the partitions of $\mathbf{Y}$. It has been pointed out in [42] that the minimum mean square error estimate of the random vector $\mathbf{Y}_2$ given the observations from the random vector $\mathbf{Y}_1$, is given by

$$\hat{\mathbf{Y}}_2 = E[\mathbf{Y}_2|\mathbf{Y}_1] \quad (3.42)$$

where $E[\mathbf{Y}_2|\mathbf{Y}_1]$ denotes the conditional mean or the expected value of $\mathbf{Y}_2$ given $\mathbf{Y}_1$. Assume that $\mathbf{Y}$ is an SIRV of dimension N with characteristic PDF $f_S(s)$. Also, for convenience, it is assumed that the mean of $\mathbf{Y}$ is zero. The covariance matrix of $\mathbf{Y}$ denoted by Σ can be partitioned

as

$$\Sigma = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad (3.43)$$

where C_{11} denotes the covariance matrix of Y_1 , C_{12} denotes the cross covariance matrix of the vectors Y_1 and Y_2 , C_{21} is the transpose of C_{12} , and C_{22} denotes the covariance matrix of the vector Y_2 . The PDF of Y_2 given Y_1 is expressed as

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_Y(y)}{f_{Y_1}(y_1)}. \quad (3.44)$$

Recall from eqs. (3.15) and (3.16) that

$$f_Y(y) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} h_N(p) \quad (3.45)$$

where

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds \quad (3.46)$$

and, assuming $b = 0$ $p = y^T \Sigma^{-1} y$. Note that the inverse covariance matrix can be partitioned as [38]

$$\Sigma^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (3.47)$$

where

$$\begin{aligned} A &= (C_{11} - C_{12} C_{22}^{-1} C_{21})^{-1} \\ B &= -A C_{12} C_{22}^{-1} \\ C &= -D C_{21} C_{11}^{-1} \\ D &= (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1}. \end{aligned} \quad (3.48)$$

Expanding the quadratic form, we have

$$p = y_1^T A y_1 + y_1^T B y_2 + y_2^T C y_1 + y_2^T D y_2. \quad (3.49)$$

Adding and subtracting $y_1^T C_{11}^{-1} y_1$ to the right hand side of eq (3.49) gives

$$p = y_1^T (A - C_{11}^{-1}) y_1 + y_1^T C_{11}^{-1} y_1 + y_1^T B y_2 + y_2^T C y_1 + y_2^T D y_2. \quad (3.50)$$

Note that

$$A - C_{11}^{-1} = -BC_{21}C_{11}^{-1}. \quad (3.51)$$

Hence,

$$p = y_1^T C_{11}^{-1} y_1 - y_1^T BC_{21}C_{11}^{-1} y_1 + y_1^T B y_2 + y_2^T C y_1 + y_2^T D y_2. \quad (3.52)$$

However, it can be shown that

$$\begin{aligned} y_2^T C y_1 &= -y_2^T D C_{21} C_{11}^{-1} y_1 \\ y_1^T B y_2 &= -y_1^T C_{11}^{-1} C_{12} D y_2 \\ -y_1^T B C_{21} C_{11}^{-1} y_1 &= y_1^T C_{11}^{-1} C_{12} D C_{21} C_{11}^{-1} y_1 \end{aligned} \quad (3.53)$$

Making these substitutions in the expression for p , it follows that

$$p = y_1^T C_{11}^{-1} y_1 + y_2^T D y_2 - y_2^T D C_{21} C_{11}^{-1} y_1 - y_1^T C_{11}^{-1} C_{12} D y_2 + y_1^T C_{11}^{-1} C_{12} D C_{21} C_{11}^{-1} y_1. \quad (3.54)$$

This can be rewritten as

$$p = y_1^T C_{11}^{-1} y_1 + (y_2 - C_{21} C_{11}^{-1} y_1)^T D (y_2 - C_{21} C_{11}^{-1} y_1) \quad (3.55)$$

For simplicity, we define

$$\begin{aligned} p_1 &= y_1^T C_{11}^{-1} y_1 \\ p_2 &= (y_2 - C_{21} C_{11}^{-1} y_1)^T D (y_2 - C_{21} C_{11}^{-1} y_1). \end{aligned} \quad (3.56)$$

Then

$$p = p_1 + p_2. \quad (3.57)$$

From eqs (3.57) and (3.44)-(3.46), we have

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{k}{f_{Y_1}(y_1)} \int_0^\infty s^{-N} \exp\left(-\frac{p_1 + p_2}{2s^2}\right) f_S(s) ds. \quad (3.58)$$

where $k = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}}$. Next, consider

$$E(Y_2|Y_1) = \frac{k}{f_{Y_1}(y_1)} \int_0^\infty s^{-N} \exp\left(-\frac{p_1}{2s^2}\right) \int_{Y_2} y_2 \exp\left(-\frac{p_2}{2s^2}\right) dy_2 f_S(s) ds. \quad (3.59)$$

Noting that

$$\int_{Y_2} y_2 \exp\left(-\frac{p_2}{2s^2}\right) dy_2 = (2\pi)^{\frac{N-m}{2}} |D|^{-\frac{1}{2}} s^{N-m} [C_{21} C_{11}^{-1} y_1], \quad (3.60)$$

gives

$$E(Y_2|Y_1) = \frac{k_1}{f_{Y_1}(y_1)} \int_0^\infty s^{-m} \exp\left(-\frac{p_1}{2s^2}\right) f_S(s) ds \quad (3.61)$$

where $k_1 = (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} |D|^{-\frac{1}{2}} [C_{21} C_{11}^{-1} y_1]$. When a matrix is partitioned as in eq (3.47), it is known that [43]

$$|\Sigma| = |C_{11}| |C_{22} - C_{21} C_{11}^{-1} C_{12}|. \quad (3.62)$$

Since

$$D = (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1}, \quad (3.63)$$

it follows that

$$|\Sigma| = |C_{11}| |D^{-1}| \quad (3.64)$$

Thus,

$$|\Sigma^{-1}| = |C_{11}|^{-1} |D|. \quad (3.65)$$

Hence, $k_1 = (2\pi)^{-\frac{m}{2}} |C_{11}|^{-\frac{1}{2}} [C_{21} C_{11}^{-1} y_1]$. Finally, since

$$f_{Y_1}(y_1) = (2\pi)^{-\frac{m}{2}} |C_{11}|^{-\frac{1}{2}} \int_0^\infty s^{-m} \exp\left(-\frac{p_1}{2s^2}\right) f_S(s) ds, \quad (3.66)$$

$$\hat{Y}_2 = E(Y_2|Y_1) = [C_{21} C_{11}^{-1} y_1]. \quad (3.67)$$

It is seen that the MMSE estimate of Y_2 given the data Y_1 is a linear function of Y_1 .

If the random vectors Y_1 and Y_2 have non-zero means denoted by b_1 and b_2 respectively,

then eq (3.67) takes the form

$$E(\mathbf{Y}_2|\mathbf{Y}_1) = \mathbf{b}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{y}_1 - \mathbf{b}_1). \quad (3.68)$$

As a consequence of this property, when the random vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are uncorrelated so that $\mathbf{C}_{21} = 0$, then we have

$$E[\mathbf{Y}_2|\mathbf{Y}_1] = \mathbf{b}_2 = E[\mathbf{Y}_2]. \quad (3.69)$$

This property is referred to as semi independence in [39, 44, 28]. However, for all SIRVs except the Gaussian, this result does not imply that

$$f_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2|\mathbf{y}_1) = f_{\mathbf{Y}_2}(\mathbf{y}_2) \quad (3.70)$$

This emphasizes the property that although uncorrelatedness guarantees statistical independence for Gaussian random vectors, it is not a general property of SIRVs.

3.5.4 Distribution of Sums of SIRVs

While it is true that the sum of two jointly Gaussian random vectors is also Gaussian, the same is not true for SIRVs in general. This result holds for two SIRVs when they are statistically independent, have zero mean and when the covariance matrix of the first is within a multiplicative constant of the covariance matrix of the second [28, 29]. More precisely, let $\mathbf{Y}_1 = [Y_{11}, Y_{12}, \dots, Y_{1N}]^T$ and $\mathbf{Y}_2 = [Y_{21}, Y_{22}, \dots, Y_{2N}]^T$ denote two independent zero mean SIRVs. The covariance matrix and characteristic PDF of $\mathbf{Y}_1$ are denoted by Σ_1 and $f_{S_1}(s_1)$. The corresponding quantities for $\mathbf{Y}_2$ are denoted by Σ_2 and $f_{S_2}(s_2)$. We are interested in obtaining the distribution of the sum given by

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2. \quad (3.71)$$

The characteristic function of $\mathbf{Y}$ is given by

$$E[\exp(j\omega\mathbf{Y})] = g_1(\omega^T\Sigma_1\omega)g_2(\omega^T\Sigma_2\omega) \quad (3.72)$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are the characteristic functions of $\mathbf{Y}_1$ and $\mathbf{Y}_2$, respectively. If $\mathbf{Y}$ is a zero mean SIRV, then its characteristic function has the form

$$E[\exp(j\omega\mathbf{Y})] = g(\omega^T\Sigma\omega). \quad (3.73)$$

In order to write eq (3.72) as a function of a single quadratic form, Σ_2 must be within a multiplicative constant of Σ_1 .

3.5.5 Markov Property for SIRPs

An interesting property of SIRPs is that a zero mean wide sense stationary SIRP is Markov if and only if its autocorrelation function has the form

$$R(t_1, t_2) = \exp(-a|(t_1 - t_2|). \quad (3.74)$$

This result is well known for the special case of a zero mean wide sense stationary Gaussian random process. To demonstrate the more general result we consider N samples from a zero mean wide sense stationary SIRP $y(t)$. Let $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$ denote the vector of successive samples obtained from the SIRP.

Given that $y(t)$ is a zero mean wide sense stationary Markov SIRP, we first show that its autocorrelation function must have the form of eq (3.74). Let Y_1, Y_2 and Y_3 denote the random variables obtained by sampling $y(t)$ at time instants t_1, t_2 and t_3 such that $t_1 \leq t_2 \leq t_3$. Since $y(t)$ is a Markov process, the joint PDF of Y_1, Y_2 and Y_3 can be expressed as

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = f_{Y_1}(y_1)f_{Y_2|Y_1}(y_2|y_1)f_{Y_3|Y_2}(y_3|y_2). \quad (3.75)$$

The autocorrelation function $R(t_3, t_1) = E[Y_3 Y_1]$ is given by

$$R(t_3, t_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_3 y_1 f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) dy_1 dy_2 dy_3. \quad (3.76)$$

Also,

$$R(t_2, t_2) = E[Y_2^2] = \int_{-\infty}^{\infty} y_2^2 f_{Y_2}(y_2) dy_2. \quad (3.77)$$

Hence,

$$R(t_3, t_1)R(t_2, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_3 y_1 f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) dy_1 dy_2 dy_3 y_2^2 f_{Y_2}(y_2) dy_2. \quad (3.78)$$

Using eq (3.75) we can rewrite the above equation as

$$R(t_3, t_1)R(t_2, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_3 y_2 f_{Y_3, Y_2}(y_3, y_2) dy_3 dy_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_2 y_1 f_{Y_2, Y_1}(y_2, y_1) dy_2 dy_1. \quad (3.79)$$

Consequently,

$$R(t_3, t_1)R(t_2, t_2) = R(t_3, t_2)R(t_2, t_1). \quad (3.80)$$

The only non-trivial autocorrelation function satisfying this property is given by eq (3.74).

Since $y(t)$ is a zero mean SIRP, it follows that $E[\mathbf{Y}] = \mathbf{0}$. Letting $b = \exp(-a)$, we can write the covariance matrix of $\mathbf{Y}$ as

$$\Sigma = \begin{bmatrix} 1 & b & \dots & b^{N-1} \\ b & 1 & \dots & b^{N-2} \\ b^2 & b & \dots & b^{N-3} \\ \dots & \dots & \dots & \dots \\ b^{N-1} & b^{N-2} & \dots & 1 \end{bmatrix} \quad (3.81)$$

We then make use of eq (3.68) to obtain

$$E[Y_N | Y_{N-1}, Y_{N-2}, \dots, Y_1] = [b^{N-1} \ b^{N-2} \ \dots \ b] \Sigma_{\mathbf{Y}'}^{-1} \mathbf{Y}' \quad (3.82)$$

where $\mathbf{Y}' = [Y_1, Y_2, \dots, Y_{N-1}]^T$ and

$$\Sigma_{\mathbf{Y}'} = \begin{bmatrix} 1 & b & \dots & b^{N-2} \\ b & 1 & \dots & b^{N-3} \\ \dots & \dots & \dots & \dots \\ b^{N-2} & b^{N-3} & \dots & 1 \end{bmatrix} \quad (3.83)$$

Recognizing that

$$\Sigma_{y'}^{-1} = \frac{1}{1-b^2} \begin{bmatrix} 1 & -b & 0 & \dots & \dots & 0 \\ -b & 1+b^2 & -b & 0 & \dots & 0 \\ 0 & -b & 1+b^2 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & -b & 1+b^2 & -b \\ 0 & \dots & \dots & \dots & -b & 1 \end{bmatrix}. \quad (3.84)$$

Therefore, we can rewrite eq (3.82) as

$$E[Y_N | Y_{N-1}, Y_{N-2}, \dots, Y_1] = bY_{N-1}. \quad (3.85)$$

From eq (3.68), we also obtain

$$E[Y_N | Y_{N-1}] = bY_{N-1}. \quad (3.86)$$

Clearly $E[Y_N | Y_{N-1}] = E[Y_N | Y_{N-1}, Y_{N-2}, \dots, Y_1]$. Since this must be true for all choices of $Y_1, Y_2, \dots, Y_{N-1}$, it follows that $f_{Y_N | Y_{N-1}, Y_{N-2}, \dots, Y_1}(y_N | y_{N-1}, y_{N-2}, \dots, y_1) = f_{Y_N | Y_{N-1}}(y_N | y_{N-1})$. Hence, $y(t)$ is Markov.

3.5.6 Kalman Filter for SIRPs

It has been shown by Chu in [41] that the Kalman filter for SIRPs is identical to the corresponding filter for a Gaussian random process. The model considered in [41] is given by

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k \quad (k = 0, 1, \dots, N-1) \\ \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad (k = 0, 1, \dots, N-1) \end{aligned} \quad (3.87)$$

where $\mathbf{x}_k$ denotes the state vector of the underlying process, $\mathbf{w}_k$ is its excitation vector, $\mathbf{y}_k$ denotes the observation vector and $\mathbf{v}_k$ is the measurement noise. It is assumed that $\mathbf{x}_k$, $\mathbf{w}_k$ and

$\mathbf{v}_k$ are jointly SIRP with a common characteristic PDF $f_S(s)$. Also, let

$$\begin{aligned}
E[\mathbf{x}_k] &= \bar{\mathbf{x}}_k \quad (k = 0, 1, \dots, N-1) \\
E[(\mathbf{x}_k - \bar{\mathbf{x}}_k)(\mathbf{x}_k - \bar{\mathbf{x}}_k)^T] &= \mathbf{M}_k \\
E[\mathbf{w}_k] &= E[\mathbf{v}_k] = 0 \\
E[(\mathbf{x}_k - \bar{\mathbf{x}}_k)\mathbf{w}_k^T] &= E[(\mathbf{x}_k - \bar{\mathbf{x}}_k)\mathbf{v}_k^T] = E[\mathbf{w}_k\mathbf{v}_k^T] = 0 \\
E[\mathbf{w}_{kl}\mathbf{w}_{km}] &= \mathbf{Q}_k\delta_{l,m} \\
E[\mathbf{v}_{kl}\mathbf{v}_{km}] &= \mathbf{R}_k\delta_{l,m}
\end{aligned} \tag{3.88}$$

where $\mathbf{w}_{km}$ and $\mathbf{v}_{km}$ are the m^{th} components of $\mathbf{w}_k$ and $\mathbf{v}_k$ respectively, and $\delta_{l,m}$ is the Kronecker delta function. Hence, $\mathbf{x}_k$, $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually uncorrelated while $\mathbf{w}_k$ and $\mathbf{v}_k$ are each white with zero mean.

The innovations vector is defined as

$$\tilde{\mathbf{y}}_{k|k-1} = \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \tag{3.89}$$

where $\hat{\mathbf{x}}_{k|k-1}$ is the MMSE estimate of $\mathbf{x}_k$ given the observation vectors up to $k-1$. The covariance matrix of the innovations can be shown to be

$$\text{Cov}(\tilde{\mathbf{y}}_{k|k-1}) = \mathbf{S}_{k|k-1} = (\mathbf{H}_k\mathbf{M}_k\mathbf{H}_k^T + \mathbf{R}_k). \tag{3.90}$$

It can be readily shown that $\mathbf{x}_k$ and $\mathbf{y}_k$ are jointly SIRP. Therefore, the MMSE estimate of $\mathbf{x}_k$ given the observation vectors up to $k-1$ is a linear function of $\mathbf{y}_m$ $m = 1, 2, \dots, k-1$, as shown by eq (3.68). Hence, the Kalman filter equations for SIRPs are identical to those for the Gaussian case. The Kalman gain denoted by $\mathbf{K}_{k|k}$ is expressed as

$$\mathbf{K}_{k|k} = \mathbf{M}_k\mathbf{H}_k^T\mathbf{S}_{k|k-1}^{-1}. \tag{3.91}$$

The measurement update $\hat{\mathbf{x}}_{k|k}$ is given by

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k|k}\tilde{\mathbf{y}}_{k|k-1} = (\mathbf{I} - \mathbf{K}_{k|k})\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k|k}\mathbf{y}_k. \tag{3.92}$$

The covariance matrix of the error in the update can be written as

$$C_k = M_k - M_k H_k^T (H_k M_k H_k^T + R_k)^{-1} H_k M_k. \quad (3.93)$$

The prediction is then given by

$$\hat{x}_{k+1|k} = F_k \hat{x}_{k|k}. \quad (3.94)$$

Finally, the covariance matrix of the prediction is expressed as

$$M_{k+1} = F_k C_k F_k^T + G_k Q_k G_k^T. \quad (3.95)$$

When systems driven by non-Gaussian noise are encountered in practice, under the assumption of joint SIRPs, these equations provide an efficient computation formula for the Kalman filter.

3.5.7 Statistical Independence

We point out that the only case for which the components of an SSRV are statistically independent occurs when the SSRV is Gaussian. This property is proved in Appendix A.

3.5.8 Ergodicity of SIRPs

It has been pointed out in [39] that an ergodic SIRD is necessarily Gaussian. The proof of the non-ergodicity of SIRPs (except Gaussian) can be easily obtained using the representation theorem [28] for SIRPs which states that an SIRD is a univariate randomization of the Gaussian random process. More precisely, if $y(t)$ is an SIRD, then it can be expressed as $y(t) = Sz(t)$, where S is a non-negative random variable and $z(t)$ is a Gaussian random process. Clearly, if $z(t)$ is stationary, then $y(t)$ will also be stationary. However, different realizations of S result in different scale factors for the sample functions of $y(t)$. Therefore, time averages will differ from one sample function to another and, in general, will not equal the corresponding ensemble average. Consequently, $y(t)$ cannot be ergodic. When S is a non-random constant, $y(t)$ is a Gaussian random process. Then $y(t)$ will be ergodic provided $z(t)$ is also ergodic. It is concluded that only Gaussian SIRPs can be ergodic.

3.6 Conclusion

In this chapter, we have presented an overview of the literature on both the modeling of radar clutter and the theory of SIRPs. It is clear from this chapter that the PDF of an SIRD is uniquely determined by the specification of a mean vector, a covariance matrix and a characteristic first order PDF. It is also seen that many interesting properties of Gaussian random processes extend

readily to SIRPs. A major difference with non-Gaussian SIRPs is their non-ergodic behavior. Consequently, time averages do not result in corresponding ensemble averages. However, if ensemble averages are used instead of time averages, then non-ergodicity is not a serious problem. In the following chapters, we shall present the application of SIRPs for non-Gaussian radar clutter modeling, simulation and distribution identification.

Chapter 4

Radar Clutter Modeling Using Spherically Invariant Random Processes

4.1 Introduction

In this chapter we consider the use of the theory of spherically invariant random processes (SIRP) for modeling correlated non-Gaussian radar clutter. It has been pointed out in chapter 2 that radar clutter can be non-Gaussian and that radars process N pulses at a time. Furthermore, the clutter can be highly correlated. Therefore, by clutter modeling we mean the specification of the joint probability density function (PDF) of the N correlated clutter samples. Since we are dealing with correlated clutter, the joint PDF cannot be constructed by simply taking the product of the marginal PDFs. This chapter presents a mathematically elegant and tractable approach for specifying the joint PDF of N clutter samples. In addition, we discuss the characterization of Gaussian and non-Gaussian correlated random vectors, the need for a library of multivariate PDFs for modeling correlated non-Gaussian clutter, several techniques for establishing this library and, finally, a key result for the distribution identification of multivariate correlated non-Gaussian random vectors.

Specifically, the problem of modeling a random vector obtained by sampling a stochastic process $y(t)$ at N time instants is of interest to us. The stochastic process may be real or complex. In addition, there is no restriction on the number of samples obtained or the sampling time instants. In order to completely characterize the random vector we need to specify the joint probability density function of the N samples (real or complex) or, equivalently, specify the joint

characteristic function. This problem is very well treated when the underlying stochastic process is Gaussian. The joint PDF in this case can be written as $(2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{p}{2})$, where p is a non-negative quadratic form given by $p = [\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{y} - \boldsymbol{\mu}]$. Here $\boldsymbol{\mu}$ and Σ denote the mean vector and covariance matrix of the Gaussian random vector $\mathbf{Y}$ whose components are the N samples of $y(t)$. However, if $y(t)$ is not a Gaussian random process, there is no unique specification for the joint PDF of the N samples except when the samples are statistically independent.

When processing real world data, neither the Gaussianity of the underlying stochastic process nor the statistical independence of the samples is guaranteed. In fact, it is likely that the samples may be correlated. Hence, we need to obtain multivariate non-Gaussian PDFs which can model the correlation between samples. In practice, radar clutter can vary from one application to another. Therefore, we need to have available a library of possible multivariate non-Gaussian PDFs so that an appropriate PDF can be chosen to approximate the data for each clutter scenario.

The theory of **Spherically Invariant Random Processes (SIRP)** provides us with elegant and mathematically tractable techniques to construct multivariate non-Gaussian PDFs. Spherically invariant random processes are generalizations of the familiar Gaussian random process. The PDF of every random vector obtained by sampling a SIRP is uniquely determined by the specification of a mean vector, a covariance matrix and a characteristic first order PDF. In addition, the PDF of a random vector obtained by sampling a SIRP is a function of a non-negative quadratic form. However, the PDF does not necessarily involve an exponential dependence on the quadratic form, as in the Gaussian case. Such a random vector is called a **Spherically Invariant Random Vector (SIRV)**.

There are two kinds of models for non-Gaussian radar clutter. One is called the endogenous model, where the desired non-Gaussian process with prescribed envelope PDF and correlation function is realized by using a zero memory non-linear transformation on a Gaussian process having a prespecified correlation function. In this approach it is not possible to independently control the envelope PDF and the correlation properties of the non-Gaussian process. In addition, not all nonlinearities give rise to a non-negative definite covariance matrix at their outputs. The second model is called an exogenous product model [26]. In this model, the desired non-Gaussian clutter is generated by the product of a Gaussian random process and an independent non-Gaussian process which can be highly correlated. In this scheme, the desired envelope PDF and

the correlation properties can be controlled independently. The exogenous model can be thought of as a slowly time variant non-Gaussian process modulating a Gaussian random process. The SIRP is a special case of the exogenous model, arising when the modulating process does not change rapidly during the observation interval and can be approximated as a random variable. This is due to the fact that the representation theorem for SIRPs allows us to explicitly write the non-Gaussian process as a product of a Gaussian process and a non-negative random variable. By assuming statistical independence between the modulating random variable and the Gaussian process, it is possible to independently control the non-Gaussian envelope PDF and its correlation properties. The SIRP is the only known case of the exogenous multiplicative model which allows the specification of the N^{th} order PDF.

Section 4.2 outlines the problem of interest. In Section 4.3 we present several techniques to obtain SIRVs. Examples based on various techniques described in Section 4.3 are used to obtain a library of SIRV PDFs in Section 4.4. Finally, in Section 4.5, we present a key result which characterizes SIRVs by using the quadratic form appearing in their PDFs.

4.2 Problem Statement

We assume we are dealing with coherent radar clutter. By coherent radar clutter, we mean that the clutter is processed in terms of its in phase and out of phase quadrature components. Pre-detection radar clutter, being a bandpass random process, admits a representation of the form

$$y(t) = \text{Re}\{\tilde{y}(t)\exp(j\omega_0 t)\} \quad (4.1)$$

where $\tilde{y}(t) = y_c(t) + jy_s(t)$ denotes the complex envelope of the clutter process, ω_0 is a known carrier frequency, $y_c(t)$ and $y_s(t)$ denote the in phase and out of phase quadrature components of the complex process $\tilde{y}(t)$. Equation (4.1) can be rewritten as

$$y(t) = y_c(t)\cos(\omega_0 t) - y_s(t)\sin(\omega_0 t). \quad (4.2)$$

We are interested in specifying the joint PDF of N samples obtained by sampling the process $y(t)$. Since it is always more convenient to work with the associated low pass process, we consider the equivalent problem of specifying the PDF of N complex samples obtained from the complex process $\tilde{y}(t)$. The PDF of a complex random variable is defined to be the joint PDF of its in phase and out of phase quadrature components. Therefore, it follows that the joint PDF of N

complex random variables is the joint PDF of the $2N$ in phase and out of phase quadrature components. While dealing with complex random variables, it is sometimes more convenient to work with their envelope and phase. The envelope R and phase Θ of a complex random variable $\tilde{Y}_i = Y_{ci} + jY_{si}$ are defined by

$$\begin{aligned} R_i &= \sqrt{Y_{ci}^2 + Y_{si}^2} \\ \Theta_i &= \arctan\left(\frac{Y_{si}}{Y_{ci}}\right). \end{aligned} \tag{4.3}$$

We consider the problem of specifying the PDF of a random vector $\mathbf{Y}^T = [\mathbf{Y}_c^T; \mathbf{Y}_s^T]$ obtained by sampling the random process $\tilde{y}(t)$, where $\mathbf{Y}_c = [Y_{c1}, Y_{c2}, \dots, Y_{cN}]^T$ and $\mathbf{Y}_s = [Y_{s1}, Y_{s2}, \dots, Y_{sN}]^T$. The subscripts c and s denote the in phase and out of phase quadrature components, respectively. We assume that the process $y(t)$ is a wide sense stationary random process. The necessary and sufficient conditions for $y(t)$ to be temporally wide sense stationary [42] are:

- (A) The quadrature components have zero mean.
- (B) The envelope of the pair wise quadrature components is statistically independent of the phase and the phase is uniformly distributed over the interval $(0, 2\pi)$. This results in the pair wise quadrature components being identically distributed and their joint PDF being circularly symmetric. This also results in the orthogonality of the pair wise quadrature components at each sampling instant.
- (C) The autocovariance function and crosscovariance function of the quadrature processes of the complex process $\tilde{y}(t) = y_c(t) + jy_s(t)$ satisfy the conditions given by

$$\begin{aligned} K_{cc}(\tau) &= K_{ss}(\tau) \\ K_{cs}(\tau) &= -K_{sc}(\tau) \end{aligned} \tag{4.4}$$

where

$$\begin{aligned}
K_{cc}(\tau) &= E\{X_c(t)X_c(t-\tau)\} \\
K_{ss}(\tau) &= E\{X_s(t)X_s(t-\tau)\} \\
K_{cs}(\tau) &= E\{X_c(t)X_s(t-\tau)\} \\
K_{sc}(\tau) &= E\{X_s(t)X_c(t-\tau)\}.
\end{aligned} \tag{4.5}$$

Also, the nonnegative definite property of the covariance matrix of $\mathbf{Y}$ must be satisfied.

- (D) Any choice of autocovariance and crosscovariance functions is allowed as long as requirement (C) is satisfied.

Due to requirement (A), it follows that $E(\mathbf{Y}) = 0$. Hence, $E(\mathbf{Y}_c) = E(\mathbf{Y}_s) = 0$. As a consequence of requirements (B) and (C), the covariance matrix of $\mathbf{Y}$, given by

$$\Sigma = \begin{bmatrix} \Sigma_{cc} & | & \Sigma_{cs} \\ - & | & - \\ \Sigma_{sc} & | & \Sigma_{ss} \end{bmatrix}, \tag{4.6}$$

must satisfy the conditions:

$$\begin{aligned}
\Sigma_{cc} &= \Sigma_{ss} \\
\Sigma_{cs} &= -\Sigma_{sc}
\end{aligned} \tag{4.7}$$

with the elements of the main diagonal of the matrices Σ_{cs} and Σ_{sc} being equal to zero. Note that $\Sigma_{cc} = E\{\mathbf{Y}_c \mathbf{Y}_c^T\}$, $\Sigma_{cs} = E\{\mathbf{Y}_c \mathbf{Y}_s^T\}$, $\Sigma_{sc} = E\{\mathbf{Y}_s \mathbf{Y}_c^T\}$ and $\Sigma_{ss} = E\{\mathbf{Y}_s \mathbf{Y}_s^T\}$. Finally, we point out, regardless of the value of N , we always have an even order PDF when dealing with quadrature components. We are now in a position to proceed with the characterization of $\mathbf{Y}$ as an SIRV.

For an SIRV, it is pointed out that the PDF of a given order automatically implies all lower order PDFs. For example, if N random variables are jointly Gaussian, it is well known that the i^{th} order PDF, $i = 1, 2, \dots, N-1$ is multivariate Gaussian. This property of SIRVs is called internal consistency. The requirements (A)-(D) arising from the wide sense stationarity

requirements of the process $y(t)$ are called external consistency conditions. Requirements (A)-(D) are not inherent to the SIRP and do not hold when the SIRP is not wide sense stationary.

4.3 Techniques for Determining the SIRV PDF

In this section, several techniques are presented for obtaining $h_{2N}(p)$. For convenience, temporal wide sense stationarity of the underlying bandpass process is assumed. However, the functional form of $h_{2N}(\cdot)$ is unaffected whether or not the random process is temporally wide sense stationary. Hence, it is allowable to let $p = (\mathbf{y} - \mathbf{b})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{b})$ in the final result, where $\mathbf{b}$ is any mean non-zero vector and $\boldsymbol{\Sigma}$ is any non-negative definite matrix.

Recall from Chapter 2 that the PDF of an SIRV $\mathbf{Y}^T = [\mathbf{Y}_c^T; \mathbf{Y}_s]$ with $\mathbf{Y}_c$ and $\mathbf{Y}_s$ defined in Section 4.2 is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-N} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h_{2N}(p) \quad (4.8)$$

Assuming temporal wide sense stationarity, $p = \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}$ where $\boldsymbol{\Sigma}$ is given by eq (4.6). The mean vector of $\mathbf{Y}$ is zero due to requirement (A) in Section 4.2. The covariance matrix $\boldsymbol{\Sigma}$ having the form of eq (4.6) and satisfying the requirements of eq (4.7) is readily determined when the autocorrelation function of the process is specified. Given $\boldsymbol{\Sigma}$, several techniques for obtaining $h_{2N}(p)$ are presented in this section.

The representation theorem for SIRVs allows us to express $\mathbf{Y}$ as a product of a Gaussian random vector $\mathbf{Z}$, having the same dimensions as $\mathbf{Y}$ and a non-negative variable S . For the problem of radar clutter modeling, since it is desirable to control the non-Gaussian nature of $\mathbf{Y}$ and its correlation properties independently, we assume that the random variable S is statistically independent of $\mathbf{Z}$. In addition, the covariance matrix of the SIRV can be made equal to the covariance matrix of the Gaussian random vector by requiring $E(S^2)$ to be unity. Finally, it is pointed out that the mean of $\mathbf{Z}$ is necessarily zero.

A physical interpretation can be given to $\mathbf{Z}$ and S . Consider a surveillance volume subdivided into contiguous range-Doppler-azimuth cells. Assuming a large enough cell size such that many scatterers are located in each cell, the N pulse returns from a given cell can be modeled as the Gaussian vector $\mathbf{Z}$ due to the central limit theorem. Also assume that the average clutter power remains constant over the N pulse returns in a coherent processing interval. However, the average clutter power is allowed to vary independently from cell to cell since different sets of scatterers are located in each cell. The variation of the average clutter power from cell to cell is modeled by the square of the non-negative random variable S .

4.3.1 SIRVs with Known Characteristic PDF

We consider specification of the PDF of the SIRV $\mathbf{Y}$ when its characteristic PDF is known in closed form. We have pointed out in the previous section that the mean vector of $\mathbf{Y}$ is zero. Also, we have discussed the specification of the covariance matrix of $\mathbf{Y}$. We now focus on the specification of $h_{2N}(p)$. As a consequence of the representation theorem, we can write

$$h_{2N}(p) = \int_0^\infty s^{-2N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds. \quad (4.9)$$

Equation (4.9) enables us to specify $h_{2N}(p)$ when the characteristic PDF $f_S(s)$ is known in closed form. However, in some cases, even though an analytical expression is known for the characteristic PDF, it may be difficult to evaluate the integral in eq (4.9) in closed form. In such instances, an alternate method for specifying $h_{2N}(p)$ must be examined. The method presented in the next section is useful for these cases.

4.3.2 SIRVs with Unknown Characteristic PDFs

When the characteristic PDF of the SIRV is unknown or when the integral in eq (4.9) is difficult to evaluate, we propose an alternate method to obtain $h_{2N}(p)$. Recall that we are dealing with an even order PDF. Therefore, we can use eq (3.25) starting with $h_2(w)$ to obtain $h_{2N}(w)$. It is worthwhile pointing out that $h_2(\cdot)$ is related to the first order envelope PDF. From requirement (B) of Section 3.2, the joint PDF of the i^{th} in phase and out of phase quadrature components can be expressed as

$$f_{Y_{ci}, Y_{si}}(y_{ci}, y_{si}) = (2\pi)^{-1} \sigma^{-2} h_2(p) \quad (i = 1, 2, \dots, N) \quad (4.10)$$

where $p = \frac{(y_{ci}^2 + y_{si}^2)}{\sigma^2}$ and σ^2 denotes the common variance of the in phase and out of phase quadrature components. The envelope and phase corresponding to the i^{th} quadrature components is given by

$$\begin{aligned} R_i &= \sqrt{Y_{ci}^2 + Y_{si}^2} \\ \Theta_i &= \arctan \frac{Y_{si}}{Y_{ci}}. \end{aligned} \quad (4.11)$$

Due to the assumption of wide sense stationarity, we can drop the subscript i in eq (4.11). The Jacobian of the transformation given by eq (4.11) is $J = R^{-1}$, where J denotes the Jacobian.

Using the Jacobian in eq (4.10) results in the joint PDF of R and Θ being given by

$$f_{R,\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} h_2\left(\frac{r^2}{\sigma^2}\right). \quad (4.12)$$

Clearly, the joint PDF in eq (4.12) can be factored as a product of the marginal PDFs of the random variables R and Θ . Consequently, the random variables R and Θ are statistically independent with PDFs given by

$$\begin{aligned} f_R(r) &= \frac{r}{\sigma^2} h_2\left(\frac{r^2}{\sigma^2}\right) \quad (0 \leq r \leq \infty) \\ f_\Theta(\theta) &= (2\pi)^{-1} \quad (0 \leq \theta \leq 2\pi). \end{aligned} \quad (4.13)$$

Equation (4.13) relates the envelope PDF to $h_2(\cdot)$. Hence, we can write

$$h_2\left(\frac{r^2}{\sigma^2}\right) = \frac{\sigma^2}{r} f_R(r). \quad (4.14)$$

Thus, eq (4.14) provides a mechanism to obtain $h_2(w)$. Starting from $h_2(w)$, we then use eq (3.25) to obtain $h_{2N}(w)$. Since not all non-Gaussian envelope PDFs are admissible for characterization as SJRVs, we must check that $h_2(w)$ and its derivatives satisfy the monotonicity conditions stated in Chapter 2. Finally, $h_{2N}(p)$ is obtained by simply replacing w by $p = (\mathbf{y} - \mathbf{b})^T \Sigma^{-1} (\mathbf{y} - \mathbf{b})$ in $h_{2N}(w)$.

4.3.3 Hankel Transform Approach

In this section we present an approach based on the Hankel transform for specifying $h_{2N}(p)$. Recall that the joint PDF of the i^{th} in phase and out of phase quadrature components of $\mathbf{Y}$ is given by eq (4.10). For convenience, it is assumed that σ^2 is unity. Dropping the subscript i from eq (4.10), the joint characteristic function of Y_{ci} and Y_{si} is expressed as

$$\phi_{Y_c, Y_s}(\omega_1, \omega_2) = (2\pi)^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(j\omega_1 y_c + j\omega_2 y_s) h_2(y_c^2 + y_s^2) dy_c dy_s. \quad (4.15)$$

Introducing the transformations

$$\begin{aligned} R &= \sqrt{Y_c^2 + Y_s^2} \\ \Theta &= \arctan \frac{Y_s}{Y_c} \\ \omega &= \sqrt{\omega_1^2 + \omega_2^2} \\ \alpha &= \arctan \frac{\omega_2}{\omega_1} \end{aligned} \quad (4.16)$$

we can rewrite eq (4.15) as

$$\phi_{Y_c, Y_s}(\omega_1, \omega_2) = (2\pi)^{-1} \int_0^\infty \int_0^{2\pi} \exp[j\omega r \{\cos(\theta)\cos(\alpha) + \sin(\theta)\sin(\alpha)\}] r h_2(r^2) dr d\theta. \quad (4.17)$$

Noting that $\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B)$, we can rewrite eq (4.17) as

$$\phi_{Y_c, Y_s}(\omega_1, \omega_2) = (2\pi)^{-1} \int_0^\infty \int_0^{2\pi} \exp[j\omega r \cos(\theta - \alpha)] r h_2(r^2) dr d\theta. \quad (4.18)$$

Interchanging the order of integration in eq (4.18), and recognizing that [45]

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp[jx \cos(\beta - \gamma)] d\beta, \quad (4.19)$$

where $J_0(x)$ is the Bessel function of order zero, we have

$$\phi_{Y_c, Y_s}(\omega_1, \omega_2) = \int_0^\infty r h_2(r^2) J_0(\omega r) dr. \quad (4.20)$$

From eq (4.20), it is clear that the joint characteristic function of Y_c and Y_s is a function of $\omega = \sqrt{\omega_1^2 + \omega_2^2}$. Hence, it is a circularly symmetric characteristic function. Denoting this function by $\Psi(\omega)$, we can write

$$\Psi(\omega) = \int_0^\infty r h_2(r^2) J_0(\omega r) dr. \quad (4.21)$$

Equation (4.21) is recognized as the Hankel transform of order zero of $h_2(r^2)$. Using the inverse Hankel transform, we obtain

$$h_2(r^2) = \int_0^\infty \omega \Psi(\omega) J_0(\omega r) d\omega. \quad (4.22)$$

Introducing the dummy variable w , we can write

$$h_2(w) = \int_0^\infty \omega \Psi(\omega) J_0(\omega \sqrt{w}) d\omega. \quad (4.23)$$

We then use eq (3.25) to obtain $h_{2N}(w)$. More explicitly, we can write

$$h_{2N}(w) = (-2)^{N-1} \int_0^\infty \omega \Psi(\omega) \frac{d^{N-1}}{dw^{N-1}} [J_0(\omega \sqrt{w})] d\omega. \quad (4.24)$$

Using the identity [45]

$$\frac{dJ_0(\eta)}{d\eta} = -J_1(\eta) \quad (4.25)$$

we have

$$\frac{dJ_0(\omega \sqrt{w})}{dw} = -\frac{\omega}{2} w^{-\frac{1}{2}} J_1(\omega \sqrt{w}). \quad (4.26)$$

Use of the recurrence relation [45]

$$\frac{d}{d\eta} [\eta^{-\alpha} J_\alpha(\eta)] = -\eta^{-\alpha} J_{\alpha+1}(\eta) \quad (4.27)$$

results in

$$\frac{d^2}{dw^2} [J_0(\omega \sqrt{w})] = \frac{\omega^2}{4} (\sqrt{w})^{-2} J_2(\omega \sqrt{w}). \quad (4.28)$$

Repeated use of eq (4.27) gives

$$\frac{d^{N-1}}{dw^{N-1}} [J_0(\omega \sqrt{w})] = (-1)^{N-1} \frac{\omega^{N-1}}{2^{N-1}} (\sqrt{w})^{-N+1} J_{N-1}(\omega \sqrt{w}). \quad (4.29)$$

Substituting eq (4.29) in eq (4.24) gives

$$h_{2N}(w) = (\sqrt{w})^{1-N} \int_0^\infty \omega^N \Psi(\omega) J_{N-1}(\omega \sqrt{w}) d\omega. \quad (4.30)$$

Finally, $h_{2N}(p)$ is obtained from eq (4.30) by replacing w by $p = (\mathbf{y} - \mathbf{b})^T \Sigma^{-1} (\mathbf{y} - \mathbf{b})$. This completes the proof of eq (3.27) for even values of N which had been previously deferred. The integral in eq (4.30) is recognized as the Hankel transform of order $N - 1$ of $\Psi(\omega)$. A number of Hankel transforms have been provided in [46] and these will be made use of in the examples presented in Section 3.4.

4.4 Examples of complex SIRVs

This section presents examples based on the approaches discussed in Section 4.3 and is divided into three parts. In section 4.4.1, we present examples that assumes the knowledge of the characteristic PDF. In Section 4.4.2, the marginal envelope PDF is assumed to be known whereas in Section 4.4.3, knowledge of the marginal characteristic function is assumed. Finally, in 4.4.4 we point out some univariate PDFs that cannot be generalized to SIRV characterization. We consider the problem of determining the PDF of the random vector $\mathbf{Y}^T = [\mathbf{Y}_c^T; \mathbf{Y}_s]$ specified in Section 4.2. It is assumed that the mean vector of $\mathbf{Y}$ and its covariance matrix Σ are known. Consequently, specification of the PDF of $\mathbf{Y}$ of the form of eq (4.8) reduces to determination of $h_{2N}(p)$.

4.4.1 Examples Based on the Characteristic PDF

4.4.1.1 Gaussian Distribution

The Gaussian marginal PDF for the quadrature components having mean b_k and variance σ_k^2 is

$$f_{Y_k}(y_k) = \frac{1}{\sqrt{(2\pi)\sigma_k^2}} \exp\left(-\frac{(y_k - b_k)^2}{2\sigma_k^2}\right) \quad (-\infty \leq y_k \leq \infty). \quad (4.31)$$

The characteristic PDF for this example is given by

$$f_S(s) = \delta(s - 1) \quad (4.32)$$

where $\delta(\cdot)$ is the unit impulse function. Using eq (3.16), it is seen that the resulting $h_N(p)$ is given by

$$h_N(p) = \exp\left(-\frac{p}{2}\right). \quad (4.33)$$

where $p = (\mathbf{y} - \mathbf{b})^T \Sigma^{-1} (\mathbf{y} - \mathbf{b})$. The corresponding PDF for any N is given by eq (3.15). For $N = 1$, this result reduces to eq (4.31). While dealing with quadrature components, we obtain the the corresponding $h_{2N}(p)$ by simply replacing N by $2N$ in eq (4.33). Whenever a characteristic PDF can be made to approach a unit impulse function displaced to the right of the origin by appropriate choice of its parameters, it follows that the corresponding SIRV PDF will approach the Gaussian PDF.

4.4.1.2 K-Distribution

The K-distributed envelope PDF, by definition, is given by

$$f_R(r) = \frac{2b}{\Gamma(\alpha)} \left(\frac{br}{2}\right)^\alpha K_{\alpha-1}(br) u(r) \quad (4.34)$$

where α is the shape parameter of the distribution, b denotes the scale parameter of the distribution, $K_N(t)$ is the N^{th} order modified Bessel function of the second kind and $u(r)$ is the unit step function. The K-distributed envelope PDF is commonly used for modeling radar clutter PDFs that have extended tails [32]- [33] and [15]-[22]. In particular, the PDF becomes heavy tailed as α approaches zero. Plots of eq (4.34) for several values of α are shown in Figures 4.1-4.4.

The K-distributed envelope PDF arises when we consider the product of a Rayleigh distributed random variable R' and an independent random variable V having the generalized-Chi distribution. More precisely, we consider the product $R = R'V$, with R' and V being statistically independent. Their PDFs are given by

$$f_{R'}(r') = r' \exp\left(-\frac{[r']^2}{2}\right) \quad 0 \leq r' \leq \infty \quad (4.35)$$

and

$$f_V(v) = \frac{2b}{\Gamma(\alpha)2^\alpha} (bv)^{2\alpha-1} \exp\left(-\frac{b^2 v^2}{2}\right) u(v), \quad (4.36)$$

respectively. Consequently, the PDF of R is given by

$$\begin{aligned} f_R(r) &= \int_0^\infty f_{R|V}(r|v) f_V(v) dv \\ &= \int_0^\infty \frac{r}{v^2} \exp\left(-\frac{r^2}{2v^2}\right) \frac{2b}{\Gamma(\alpha)2^\alpha} (bv)^{2\alpha-1} \exp\left(-\frac{b^2 v^2}{2}\right) dv. \end{aligned} \quad (4.37)$$

From [45], we have

$$K_\nu(xz) = \frac{z^\nu}{2} \int_0^\infty \exp\left[-\frac{x}{2}\left(t + \frac{z^2}{t}\right)\right] t^{-\nu-1} dt \quad [|\arg z| < \frac{\pi}{4}], \quad z > 0. \quad (4.38)$$

Letting $v^2 = t$ in eq (4.37) and using the result of eq (4.38), the PDF of eq (4.34) follows.

As a matter of interest, we demonstrate the derivation of the PDF for the quadrature components arising from the K-distributed envelope PDF. The quadrature components corresponding to the Rayleigh envelope PDF $f_{R'}(r')$, are independent identically distributed zero mean Gaussian

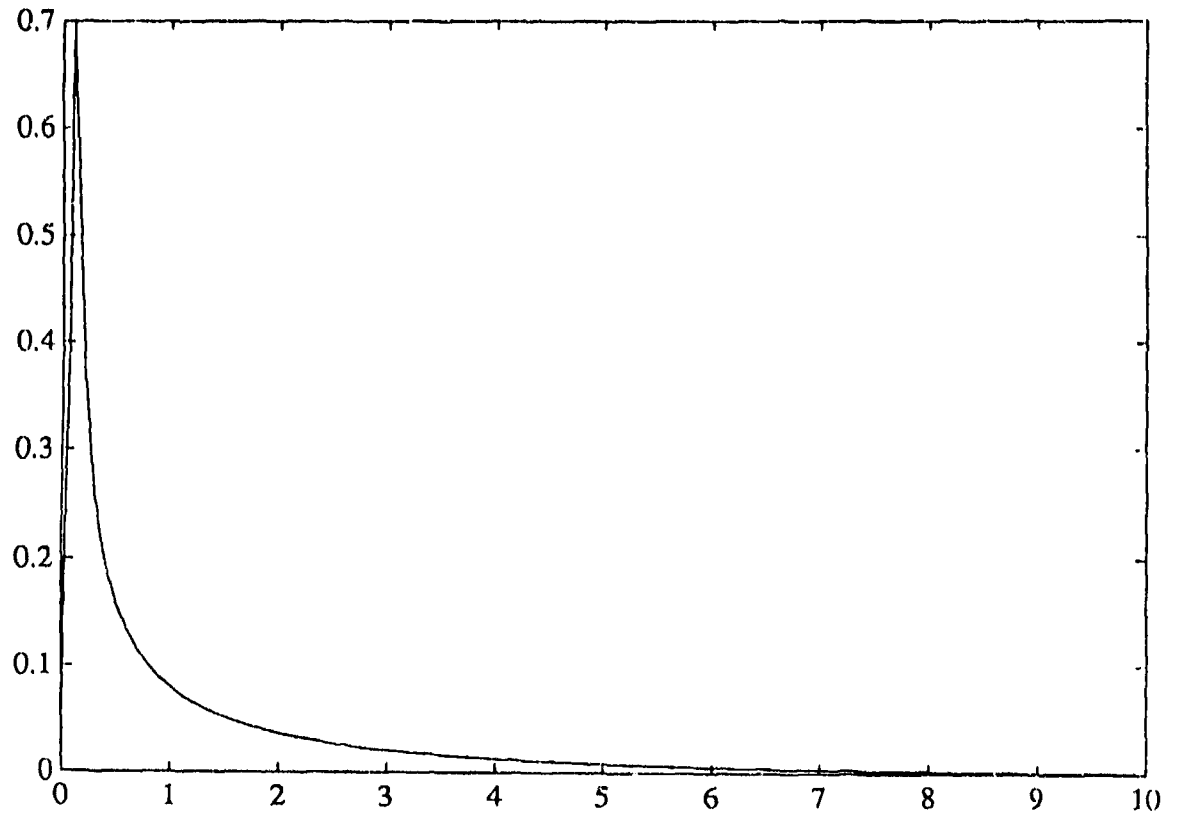


Figure 4.1: K-distribution, $b = 0.31$, $\alpha = 0.05$

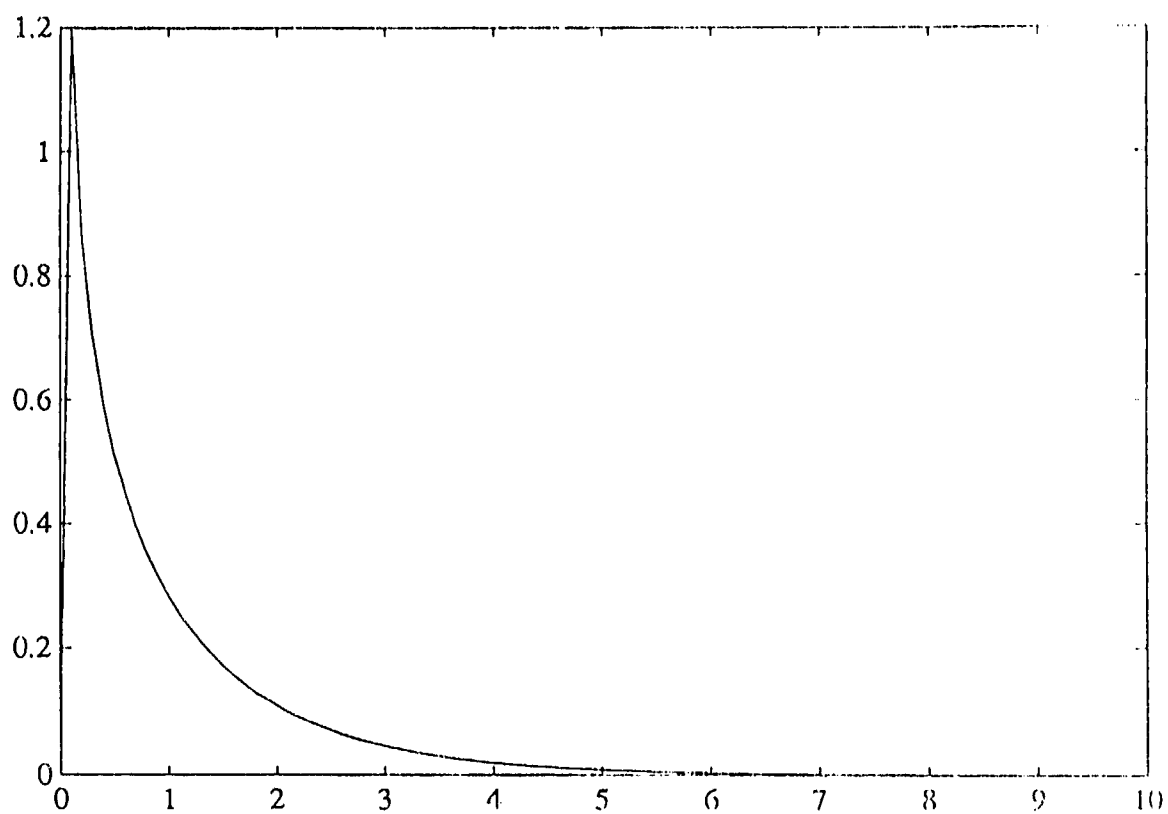


Figure 4.2: K-distribution, $b = 0.77$, $\alpha = 0.3$

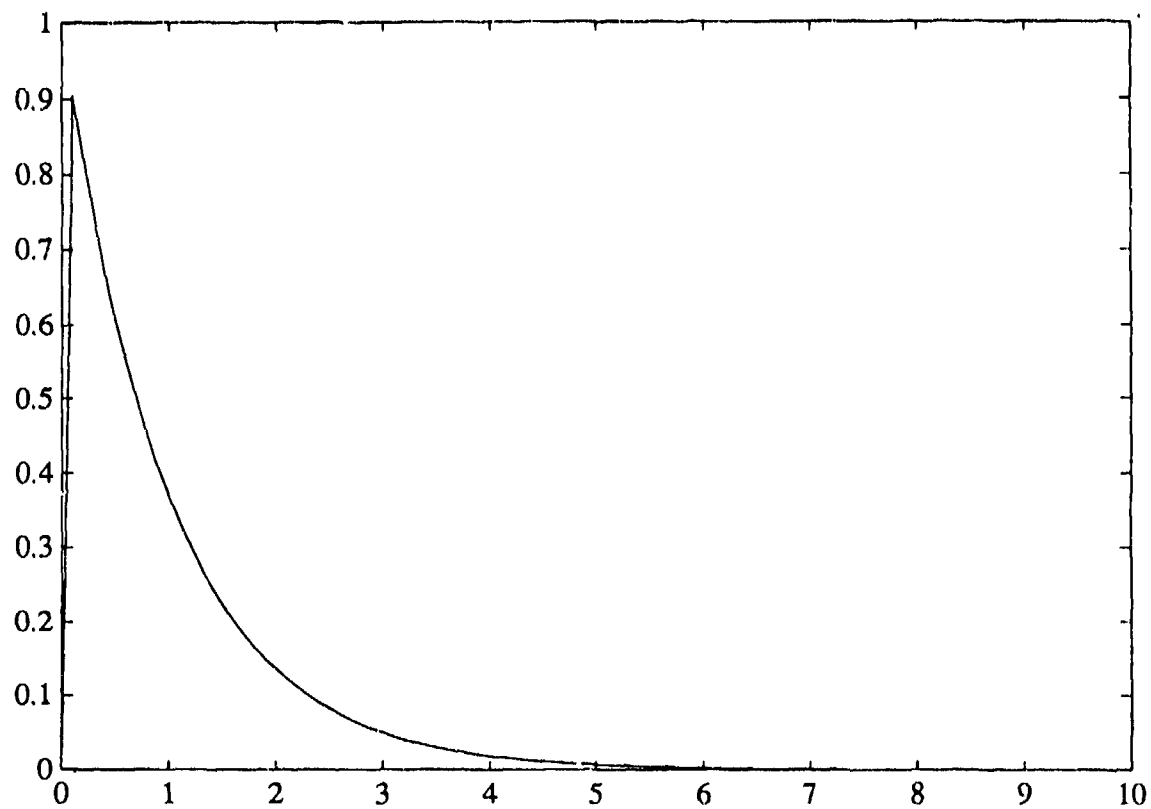


Figure 4.3: K-distribution, $b = 1$, $\alpha = 0.5$

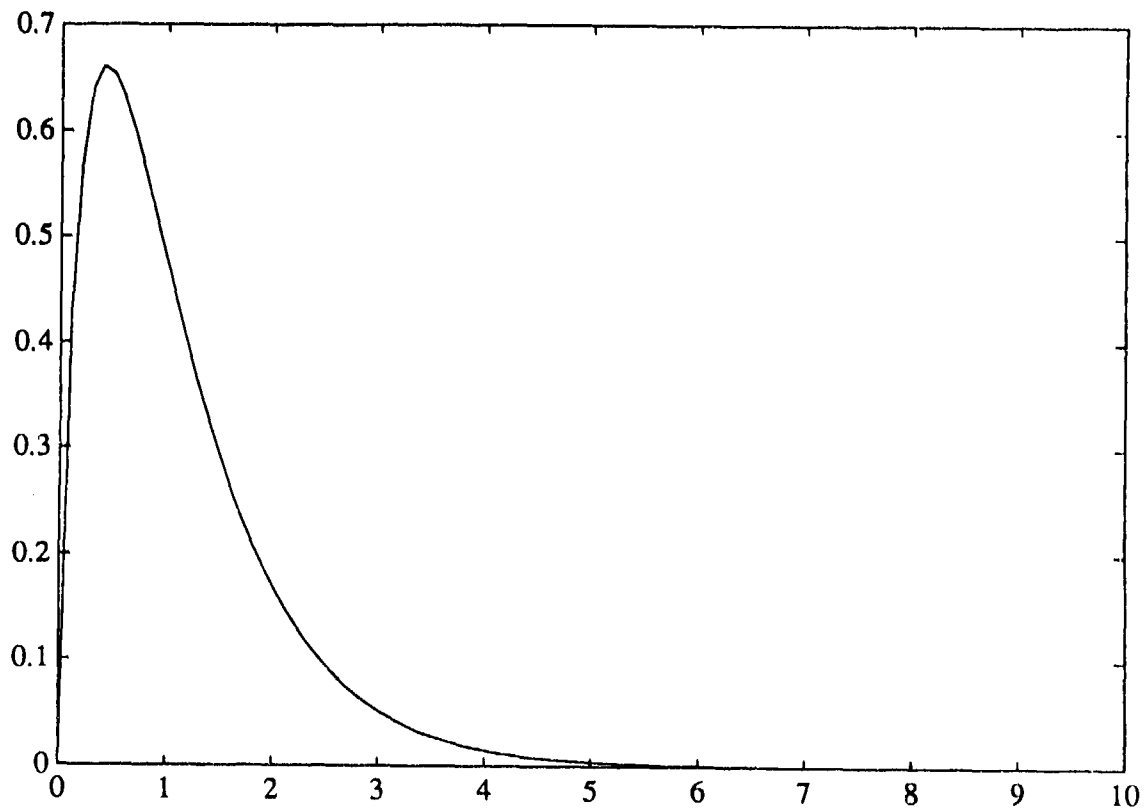


Figure 4.4: K-distribution. $b = 1.4$, $\alpha = 0.99$

random variables having unit variance. The PDF of the quadrature components corresponding to R' is expressed as

$$f_{Z_c}(z) = f_{Z_s}(z) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{z^2}{2}\right) \quad (4.39)$$

where Z_c and Z_s denote the in phase and out of phase quadrature components. The quadrature components arising from the K-distributed envelope PDF, denoted by Y_c and Y_s , respectively, can be expressed as

$$Y_c = Z_c V \quad (4.40)$$

$$Y_s = Z_s V.$$

Note that $|\tilde{Y}| = |\tilde{Z}|$ and $\Theta_{\tilde{Y}} = \Theta_{\tilde{Z}}$. Consequently, the PDF of Y_c is given by

$$f_{Y_c}(y_c) = \frac{b^{2\alpha}}{\sqrt{2\pi}\Gamma(\alpha)2^{\alpha-1}} \int_0^\infty v^{2\alpha-2} \exp\left[-\frac{1}{2}\left(\frac{y_c^2}{v^2} + b^2 v^2\right)\right] dv. \quad (4.41)$$

Making the change of variables $t = b^2 v^2$ and $z^2 = b^2 y_c^2$, and using the result of eq (4.38), the PDF of Y_c is expressed as

$$f_{Y_c}(y_c) = \frac{2b}{\Gamma(\alpha)\sqrt{2\pi}2^\alpha} |by_c|^{\alpha-\frac{1}{2}} K_{\frac{1}{2}-\alpha}(b|y_c|) \quad -\infty < y_c < \infty \quad (4.42)$$

where the absolute value denoted by $|\cdot|$ is used on account of the requirement that $z > 0$. In a similar manner, it can be shown that the PDF of Y_s has the same functional form as eq (4.42). The PDF of eq (4.42) is called the Generalized Laplace PDF [29].

The characteristic PDF for the K-distributed SIRV is

$$f_S(s) = \frac{2}{\Gamma(\alpha)2^\alpha} (bs)^{2\alpha-1} \exp\left(-\frac{b^2 s^2}{2}\right) u(s). \quad (4.43)$$

Using eqs (3.16) and (4.38),

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) \frac{2}{\Gamma(\alpha)2^\alpha} (bs)^{2\alpha-1} \exp\left(-\frac{b^2 s^2}{2}\right) ds. \quad (4.44)$$

Making the change of variables $t = b^2 s^2$ and $z^2 = b^2 p$, the resulting $h_N(p)$ is given by

$$h_N(p) = \frac{b^N}{\Gamma(\alpha)} \frac{(b\sqrt{p})^{\alpha-\frac{N}{2}}}{2^{\alpha-1}} K_{\frac{N}{2}-\alpha}(b\sqrt{p}). \quad (4.45)$$

The corresponding SIRV PDF for any N is given by using eq (3.15). For the case when $N = 1$, this reduces to eq (4.42). When dealing with quadrature components, we use eq (4.45) with N replaced by $2N$

4.4.1.3 Student-t Distribution

The Student-t distribution for the quadrature components is given by

$$f_{Y_k}(y_k) = \frac{\Gamma(\nu + \frac{1}{2})}{b\sqrt{\pi}\Gamma(\nu)} (1 + \frac{y_k^2}{b^2})^{-\nu-\frac{1}{2}} \quad (-\infty < x_k < \infty), \quad \nu > 0 \quad (4.46)$$

where b is the scale parameter, ν is the shape parameter $\Gamma(\nu)$ is the Eulero-Gamma function and $k = c, s$. Plots of the Student-t distribution are shown for several values of ν in Figures 4.5-4.7.

The characteristic PDF for this example is

$$f_S(s) = \frac{2}{\Gamma(\nu)} \left(\frac{1}{2}\right)^\nu b^{2\nu-1} (s^{-1})^{2\nu+1} \exp\left(-\frac{b^2}{2s^2}\right) u(s). \quad (4.47)$$

Use of eq (3.16) results in $h_N(p)$ being given by

$$h_N(p) = \frac{2^{\frac{N}{2}} b^{2\nu} \Gamma(\nu + \frac{N}{2})}{\Gamma(\nu) (b^2 + p)^{\frac{N}{2} + \nu}}. \quad (4.48)$$

The corresponding SIRV PDF for any N is given by eq (3.15). For $N = 1$, this result reduces to eq (4.46). When dealing with quadrature components, we make use of eq (4.48) with N replaced by $2N$.

4.4.1.4 Mixture of Gaussian PDFs

An interesting non-Gaussian marginal PDF that is admissible as an SIRV is the mixture of Gaussian PDFs. We consider the PDF given by

$$f_{Y_k}(y_k) = \sum_i a_i (2\pi k_i^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_k - b_k)^2}{2k_i^2}\right) \quad (4.49)$$

for the quadrature components of $\mathbf{Y}$. The characteristic PDF for this example is given by

$$f_S(s) = \sum_i a_i \delta(s - k_i). \quad (4.50)$$

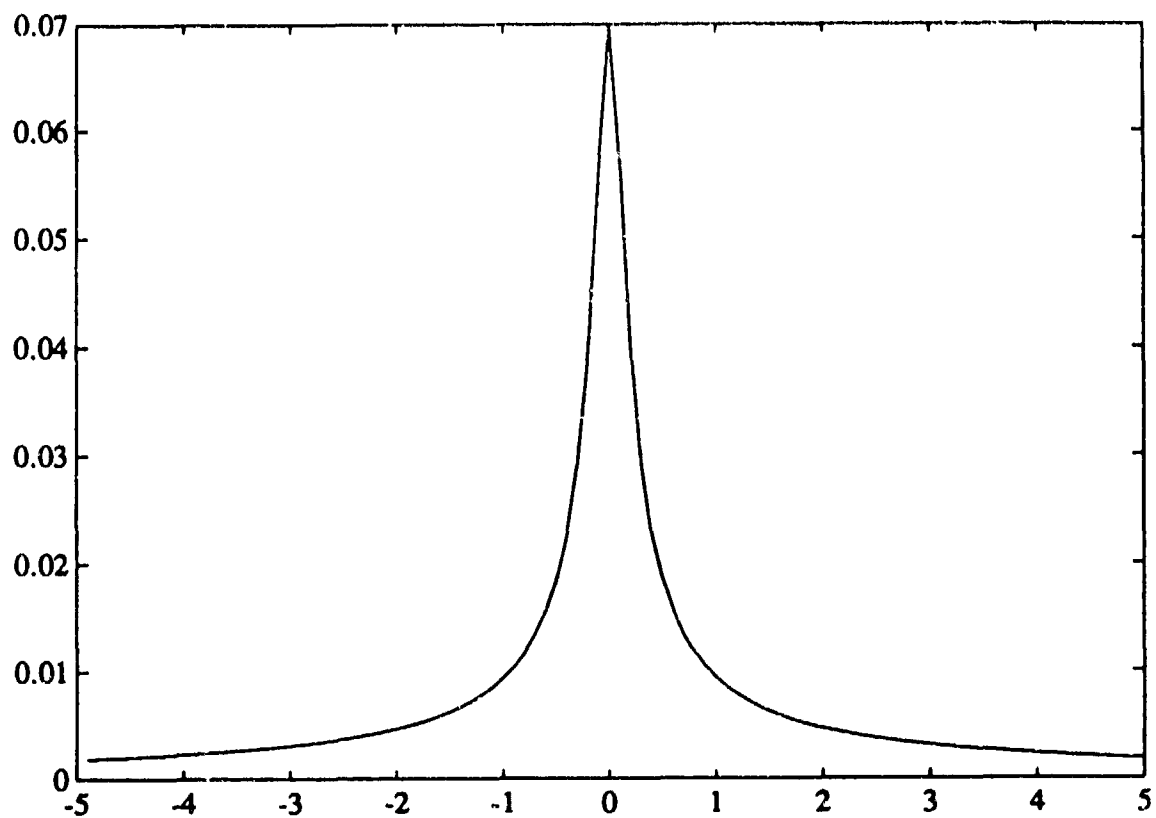


Figure 4.5: Student-t distribution, $b = 0.14$, $\nu = 0.01$

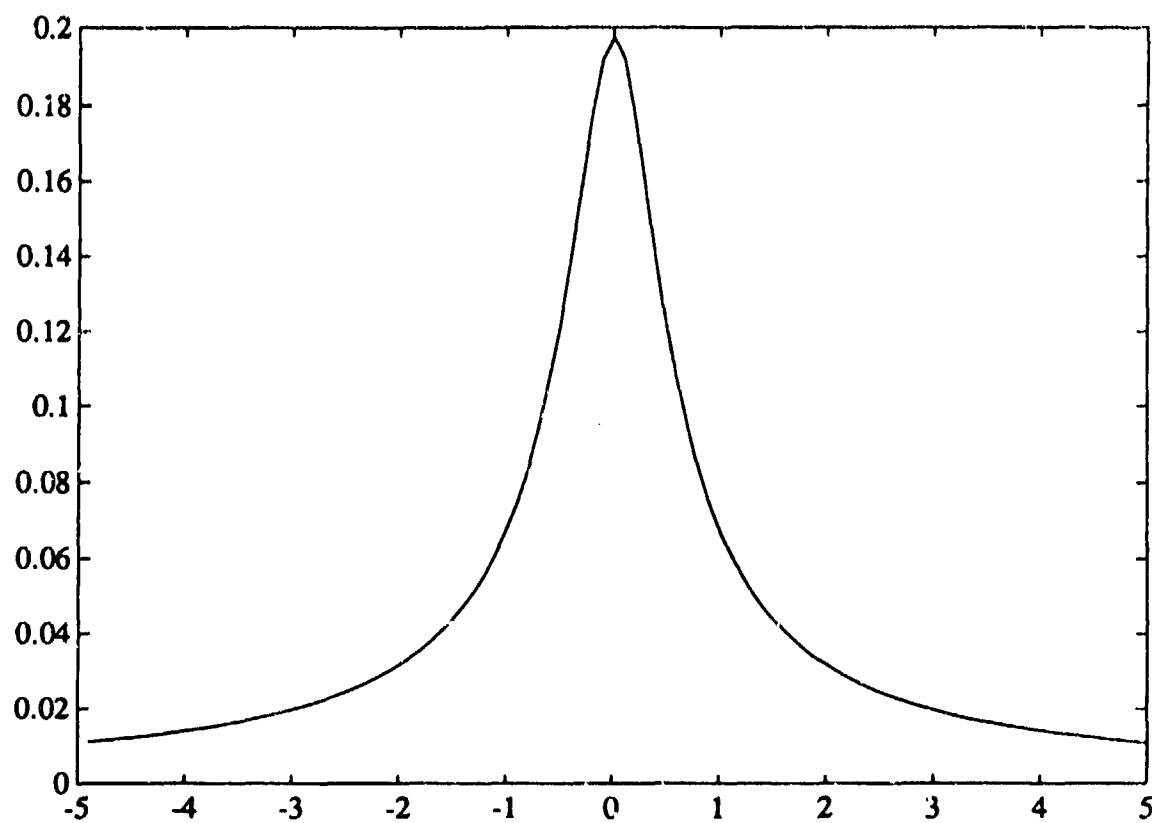


Figure 4.6: Student-t distribution, $b = 0.45$, $\nu = 0.1$

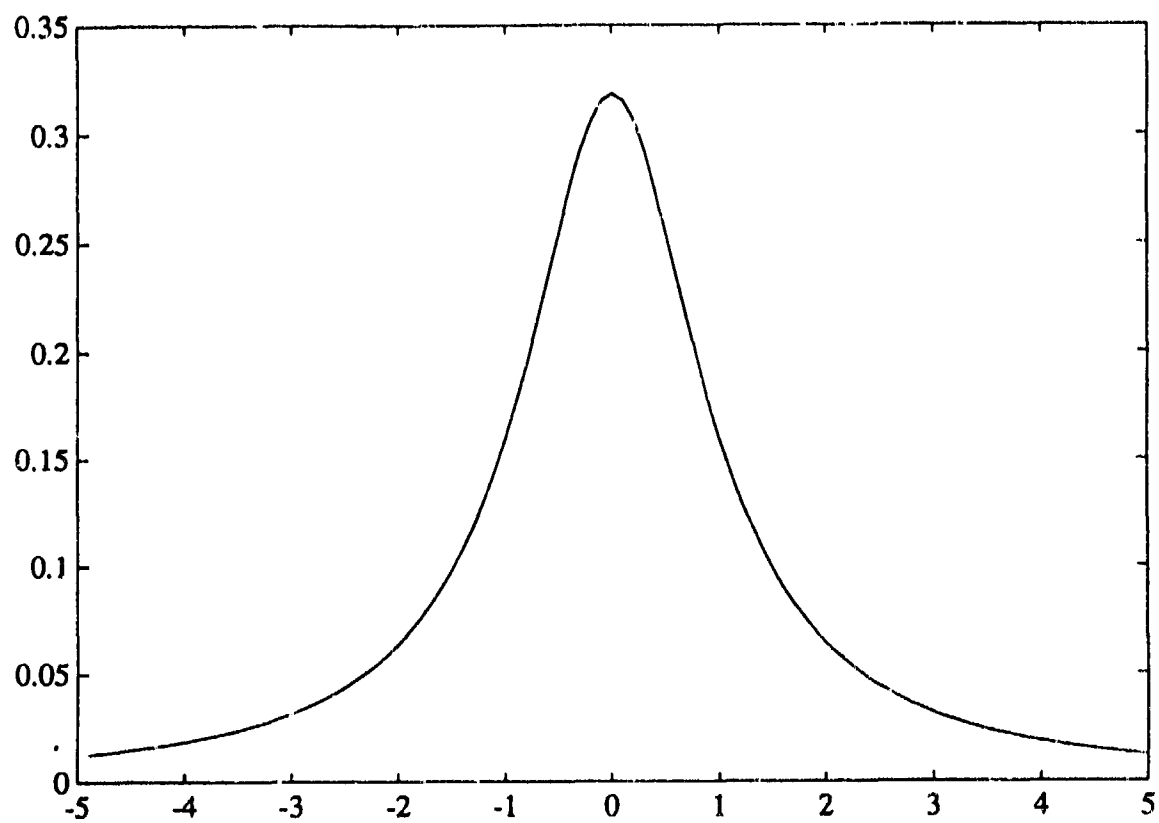


Figure 4.7: Student-t distribution, $b = 1$, $\nu = 0.5$

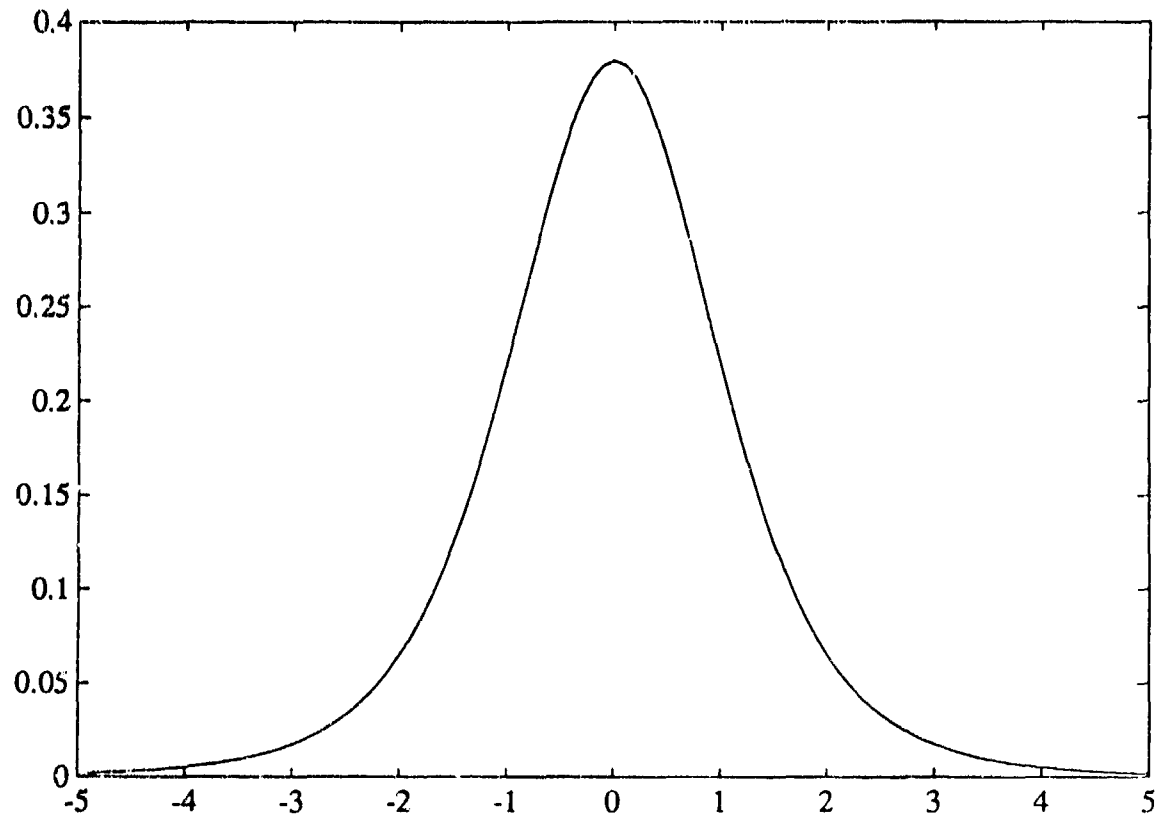


Figure 4.8: Student-t distribution, $b = 2.23$, $\nu = 2.5$

Note that S is a discrete random variable, with a_i denoting the probability $P(S = k_i)$. Also, it is required that

$$\begin{aligned} a_i &\geq 0 \quad i = 1, 2, \dots \\ \sum_i a_i &= 1. \end{aligned} \quad (4.51)$$

Using eq (3.16), it is seen that

$$h_N(p) = \sum_i k_i^{-N} a_i \exp\left(-\frac{p}{2k_i^2}\right). \quad (4.52)$$

The corresponding SIRV PDF for any N is given by eq (3.15). For $N = 1$, this result reduces to eq (4.49). When dealing with quadrature components, we make use of the result of eq (4.52) with N replaced by $2N$. Note that the a_i 's can be assigned any convenient discrete distribution.

4.4.2 Examples Based on Marginal Envelope PDF

We shall report here on some new SIRV PDFs obtained starting from the marginal envelope PDF. In general, note that the characteristic PDF for all the examples considered here are not available in closed form. Since σ^2 is the common variance of the in phase and out of phase quadrature components, σ^2 is equal to $\frac{1}{2}E(R^2)$. In addition, recall that the binomial coefficient is defined by

$$\binom{l}{i} = \frac{l!}{i!(l-i)!}. \quad (4.53)$$

In all the examples in this section, we start with $h_2(w)$ and obtain $h_{2N}(w)$ by the process of successive differentiation. The corresponding $h_{2N}(p)$ for each example is obtained by replacing w by p in $h_{2N}(w)$. In all the examples presented in this section, note that the envelope PDFs reduce to the Rayleigh envelope PDF for appropriately chosen parameters.

4.4.2.1 Chi Envelope PDF

We consider the Chi distributed envelope PDF given by

$$f_R(r) = \frac{2b}{\Gamma(\nu)} (br)^{2\nu-1} \exp(-b^2 r^2) \quad (0 \leq r \leq \infty) \quad (4.54)$$

where b denotes the scale parameter and ν denotes the shape parameter. Plots of the Chi envelope PDF are shown in Figures 4.8-4.10 for several values of ν . Using eq(4.14), we can

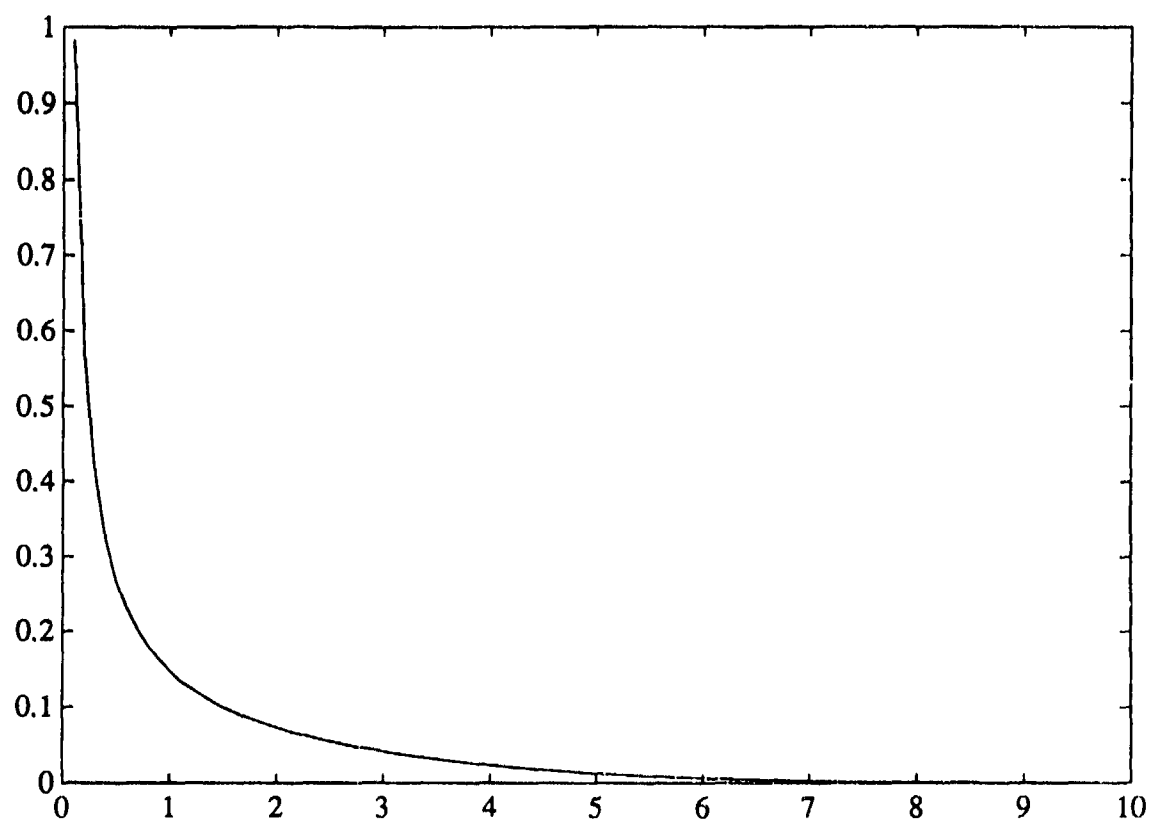


Figure 4.9: Chi Envelope PDF, $b = 0.22$, $\nu = 0.1$

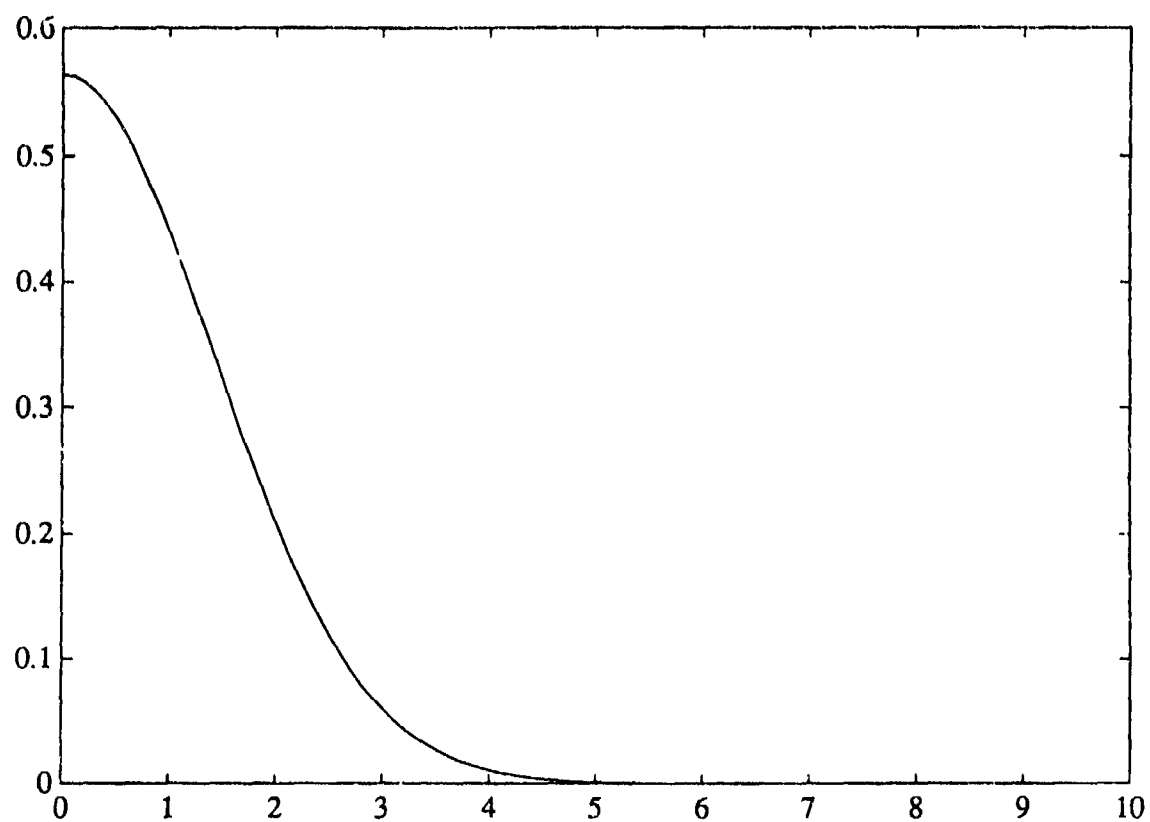


Figure 4.10: Chi Envelope PDF, $b = 0.5$, $\nu = 0.5$

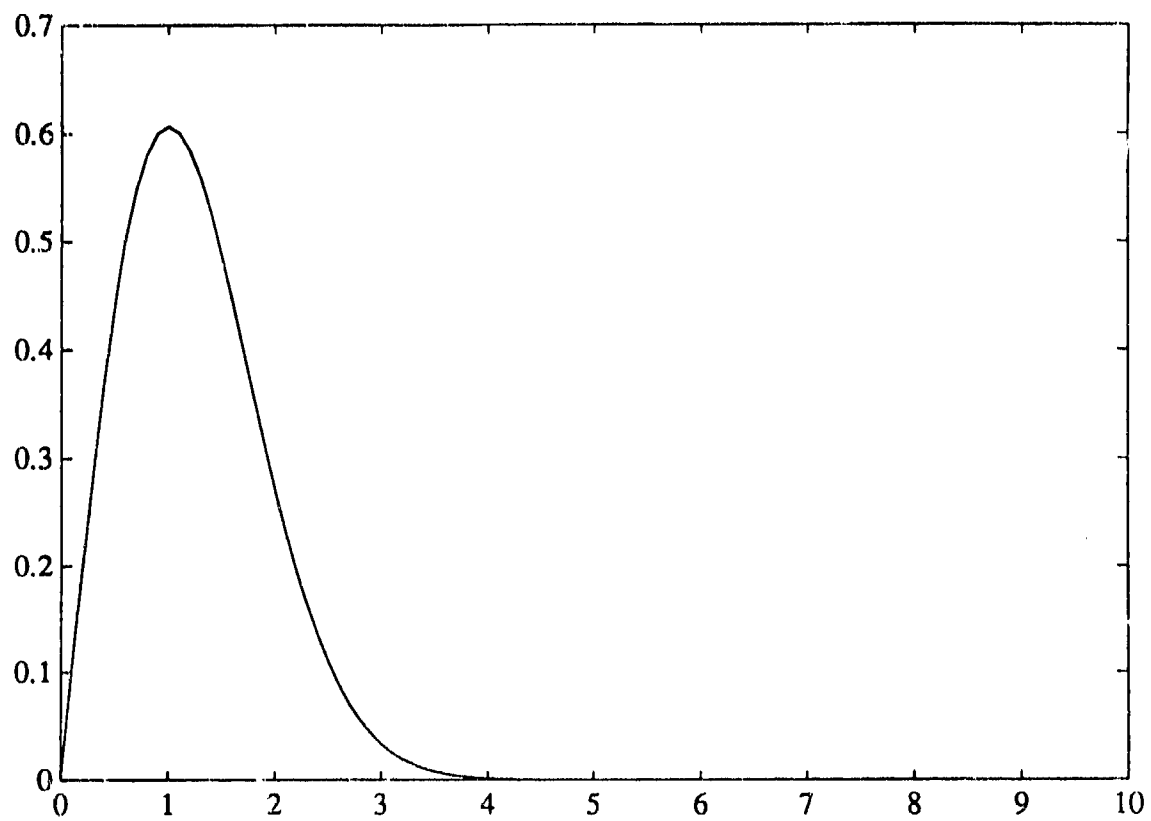


Figure 4.11: Chi Envelope PDF, $b = 0.70$, $\nu = 1.0$

write

$$h_2(w) = \frac{2}{\Gamma(\nu)} (b\sigma)^{2\nu} w^{\nu-1} \exp(-b^2\sigma^2 w). \quad (4.55)$$

Using eq(3.25), we have

$$\begin{aligned} h_{2N}(w) &= (-2)^{N-1} \frac{d^{N-1} h_2(w)}{dw^{N-1}} \\ &= \frac{(-2)^{N-1}}{\Gamma(\nu)} 2(b\sigma)^{2\nu} \frac{d^{N-1}}{dw^{N-1}} [w^{\nu-1} \exp(-b^2\sigma^2 w)]. \end{aligned} \quad (4.56)$$

Recall Leibnitz's theorem for the n^{th} derivative of a product [45], which states that

$$\frac{d^n(uv)}{dx^n} = \sum_{k=0}^n \binom{n}{k} \frac{d^k u}{dx^k} \frac{d^{n-k} v}{dx^{n-k}} \quad (4.57)$$

where u and v are functions of x . Noting that

$$\frac{d^k(w^{\nu-1})}{dw^k} = \frac{\Gamma(\nu)}{\Gamma(\nu-k)} w^{\nu-k-1}, \quad (4.58)$$

it follows that

$$h_{2N}(w) = (-2)^{N-1} A \sum_{k=1}^N G_k w^{\nu-k} \exp(-Bw) \quad (4.59)$$

where

$$\begin{aligned} G_k &= \binom{N-1}{k-1} (-1)^{N-k} B^{N-k} \frac{\Gamma(\nu)}{\Gamma(\nu-k+1)} \\ A &= \frac{2}{\Gamma(\nu)} (b\sigma)^{2\nu} \\ B &= b^2\sigma^2. \end{aligned} \quad (4.60)$$

An important condition that must be pointed out is that the SIRV PDF is valid only for $\nu \leq 1$. This is due to the fact that $h_2(p)$ and its derivatives are monotonically decreasing functions only in the range of values of ν mentioned above. Finally, for $\nu = 1$, note that the Chi envelope PDF reduces to the Rayleigh envelope PDF. The corresponding SIRV PDF then becomes Gaussian.

4.4.2.2 Weibull Envelope PDF

The Weibull distributed envelope PDF is given by

$$f_R(r) = abr^{b-1} \exp(-ar^b) \quad (0 \leq r \leq \infty). \quad (4.61)$$

where a is the scale parameter and b is the shape parameter. Plots of the Weibull distribution for several values of b are shown in Figures 4.12-4.14. Using eq (4.14), we have

$$h_2(w) = ab\sigma^b w^{\frac{b}{2}-1} \exp(-a\sigma^b w^{\frac{b}{2}}) = (-2) \frac{d}{dw} [\exp(-Aw^{\frac{b}{2}})] \quad (4.62)$$

where $A = a\sigma^b$. From eq (3.25), we have

$$h_{2N}(w) = (-2)^N \frac{d^N}{dw^N} [\exp(-Aw^{\frac{b}{2}})]. \quad (4.63)$$

The rule for obtaining the N^{th} derivative of a composite function is [45]: If $f(x) = F(y)$ and $y = \varphi(x)$, then

$$\frac{d^N}{dx^N} \{f(x)\} = \sum_{k=1}^N \frac{U_k}{k!} \frac{d^k}{dy^k} [F(y)] \quad (4.64)$$

where

$$U_k = \sum_{m=1}^k (-1)^{k-m} \binom{k}{m} y^{k-m} \frac{d^N y^m}{dx^N}. \quad (4.65)$$

Making the association $x = w$ and $y = -Aw^{\frac{b}{2}}$, we have

$$h_{2N}(w) = \sum_{k=1}^N C_k w^{\frac{kb}{2}-N} \exp(-Aw^{\frac{b}{2}}) \quad (4.66)$$

where

$$C_k = \sum_{m=1}^k (-1)^{m+N} 2^N \frac{A^k}{k!} \binom{k}{m} \frac{\Gamma(1 + \frac{mb}{2})}{\Gamma(1 + \frac{mb}{2} - N)}. \quad (4.67)$$

The Weibull envelope PDF is admissible for characterization as an SIRV for values of b less than or equal to 2. This is due to the fact that $h_2(w)$ and its derivatives fail to satisfy the monotonicity condition for other values of b . However, this is not a serious restriction from the point of view of radar clutter modeling because the Weibull envelope PDF is of interest in modeling large tailed

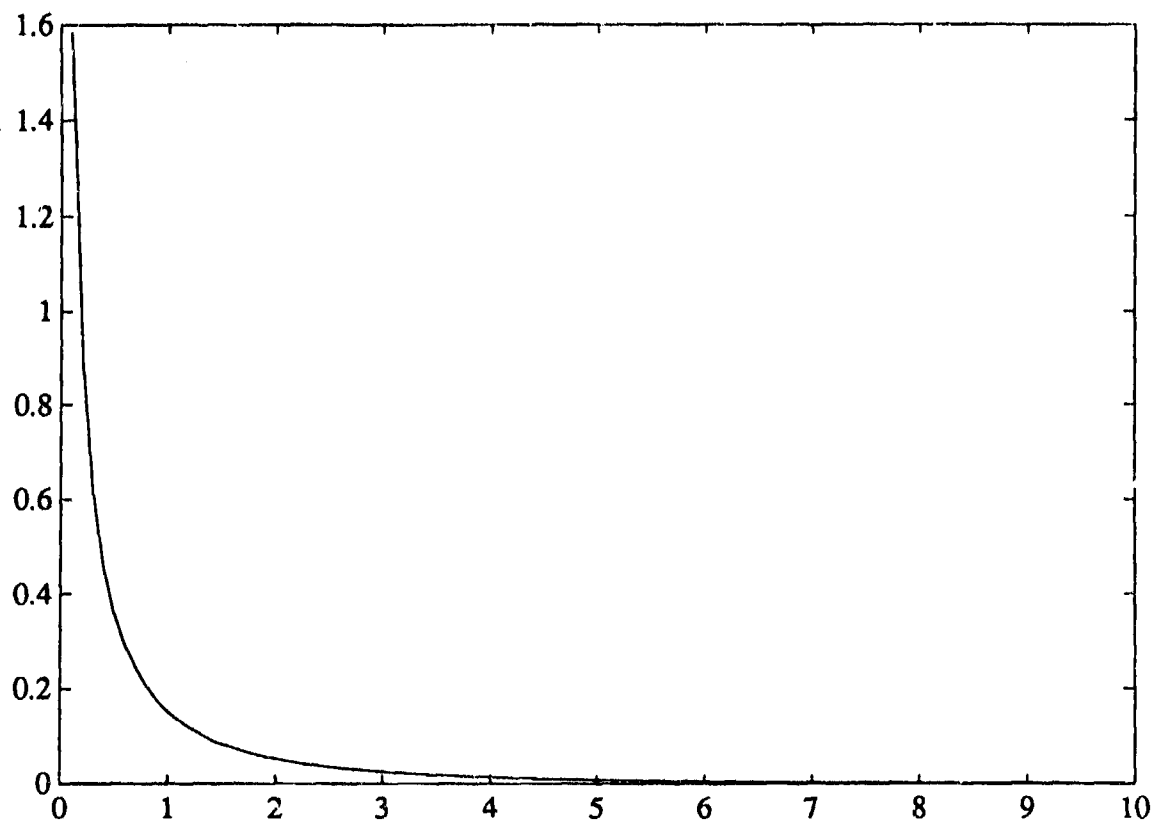


Figure 4.12: Weibull distributed Envelope PDF, $b = 0.5$, $a = 1.86$

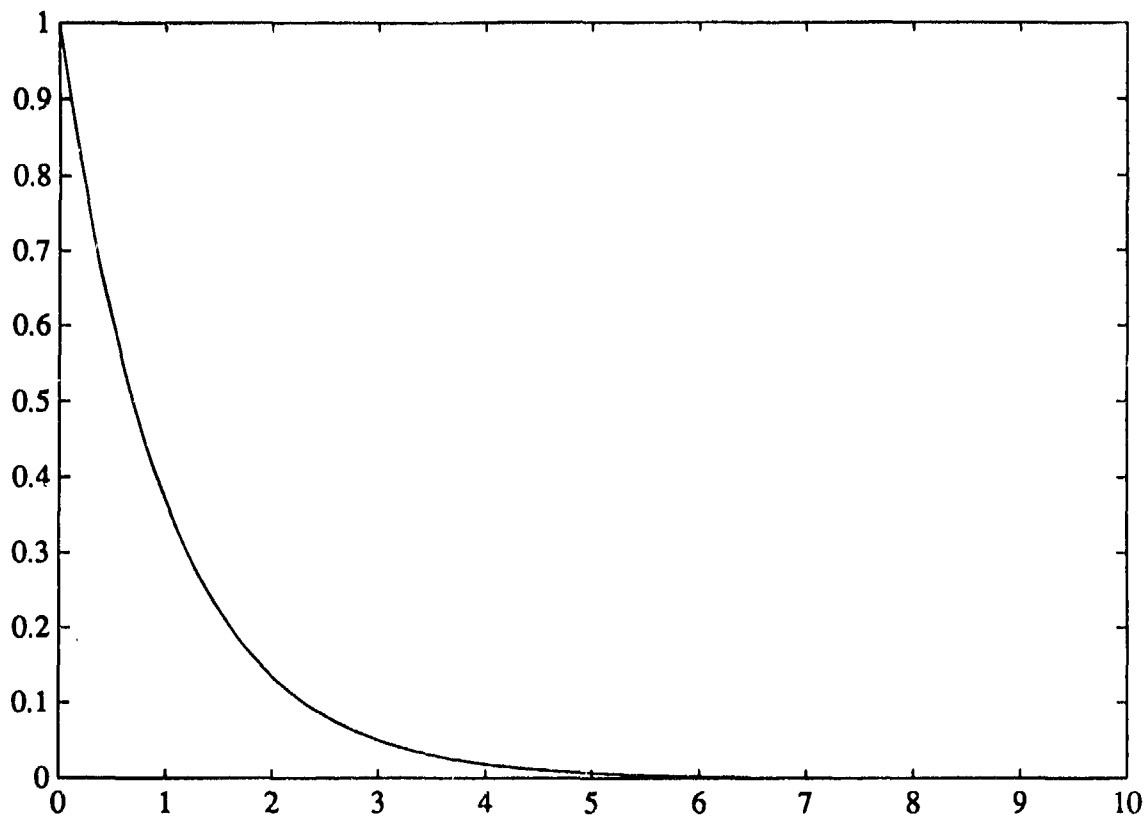


Figure 4.13: Weibull distributed, $b = 1$, $a = 1$

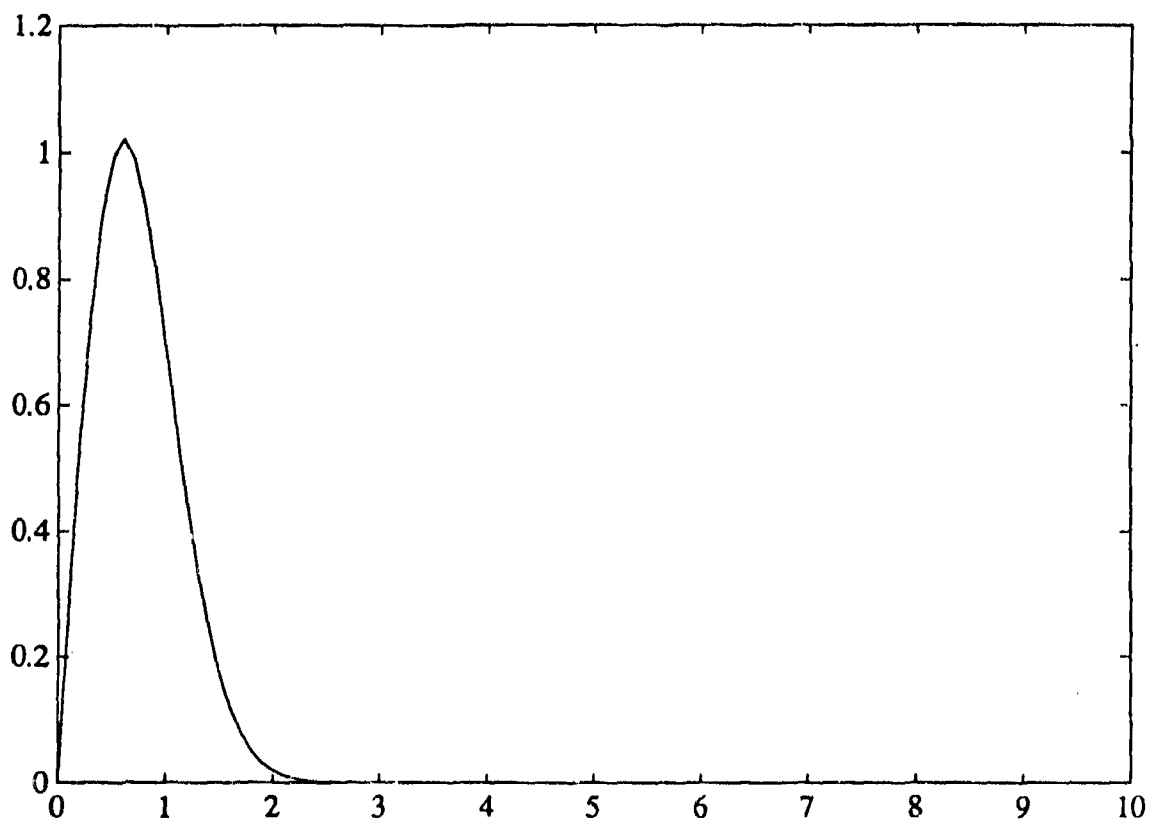


Figure 4.14: Weibull distributed, $b = 2$, $a = 0.5$

clutter. Such a situation arises only when $0 < b \leq 2$. The Weibull envelope PDF reduces to the Rayleigh envelope PDF when $b = 2$. The corresponding SIRV PDF then becomes Gaussian. Another case of interest arises when $b = 1$. In this case the Weibull envelope PDF corresponds to the Exponential envelope PDF.

4.4.2.3 Generalized Rayleigh Envelope PDF

The next PDF considered is for the Generalized Rayleigh envelope which is given by

$$f_R(r) = \frac{\alpha r}{\beta^2 \Gamma(\frac{2}{\alpha})} \exp[-(\frac{r}{\beta})^\alpha] \quad (0 \leq r \leq \infty) \quad (4.68)$$

where α is the shape parameter and β is the scale parameter. Plots of the Generalized Rayleigh distribution are shown for several values of α in Figures 4.15-4.18.

Proceeding as in the previous example, we find that

$$h_2(w) = A \exp(-Bw^{\frac{\alpha}{2}}) \quad (4.69)$$

where

$$\begin{aligned} A &= \frac{\sigma^2 \alpha}{\beta^2 \Gamma(\frac{2}{\alpha})} \\ B &= \beta^{-\alpha} \sigma^\alpha \end{aligned} \quad (4.70)$$

Using eqs (3.25), (3.63) and (3.64), we have

$$h_{2N}(w) = \sum_{k=1}^{N-1} D_k w^{\frac{k\alpha}{2} - N + 1} \exp(-Bw^{\frac{\alpha}{2}}) \quad (4.71)$$

where

$$D_k = \sum_{m=1}^k (-1)^{m+N-1} 2^{N-1} \frac{B^k}{k!} \binom{k}{m} \frac{\Gamma(1 + \frac{m\alpha}{2})}{\Gamma(2 + \frac{m\alpha}{2} - N)} \quad (4.72)$$

Note that the SIRV PDF is valid only in the range $(0 \leq \alpha \leq 2)$. This is because of the fact that the monotonicity conditions for the derivatives of $h_2(p)$ are satisfied only for the specified range of α . The Generalized Rayleigh envelope PDF reduces to the Rayleigh envelope PDF when $\alpha = 2$.

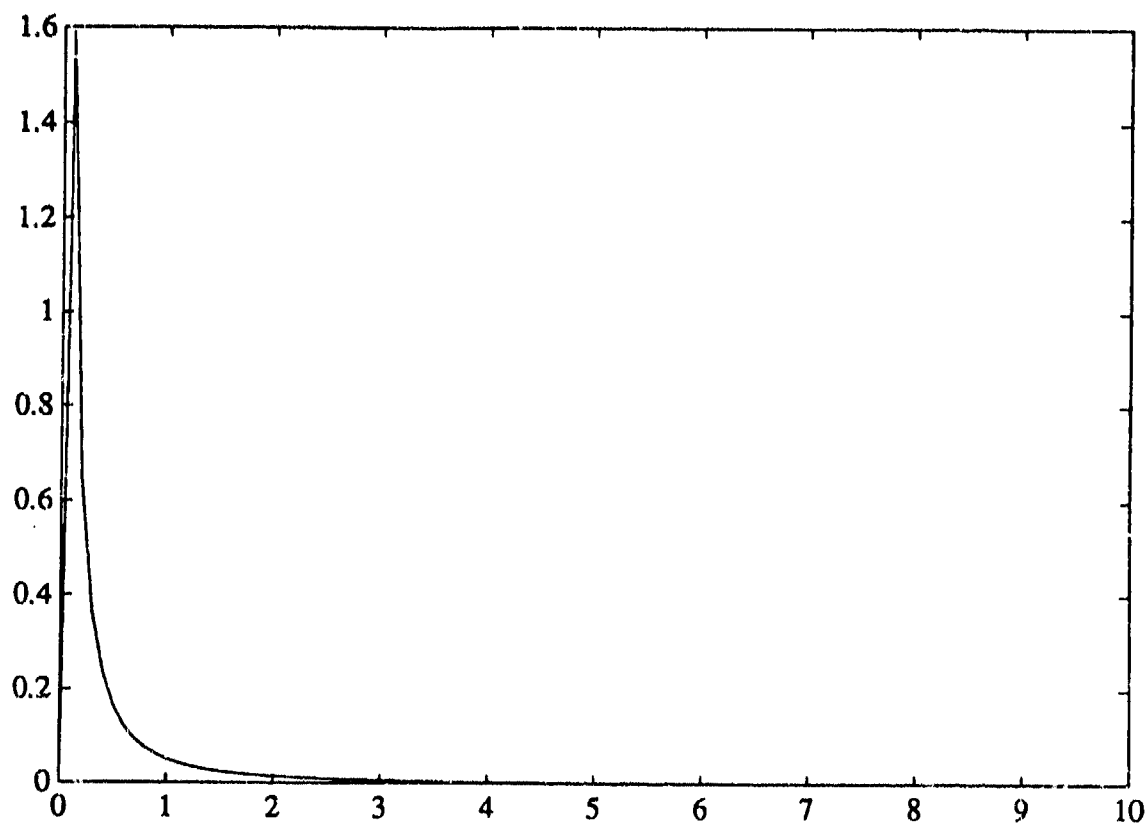


Figure 4.15: Generalized Rayleigh distributed Envelope PDF, $\alpha = 0.1$, $\beta = 3.45 \times 10^{-15}$

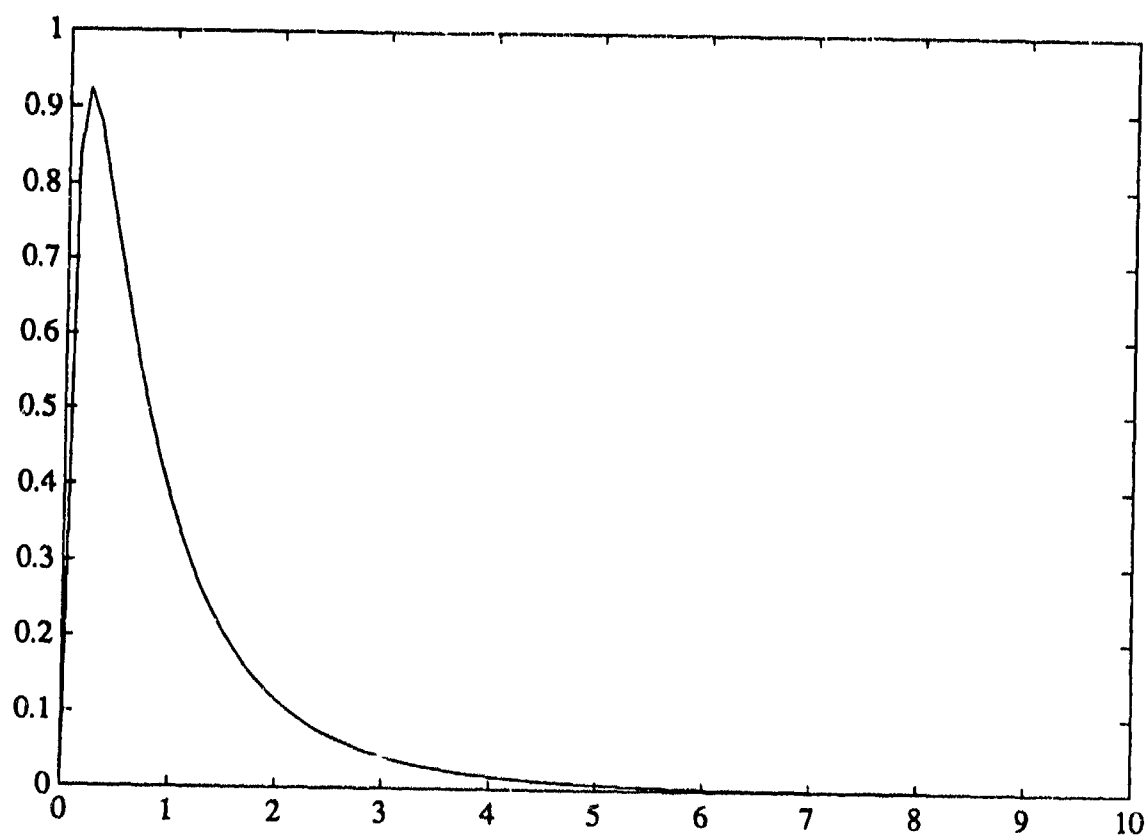


Figure 4.16: Generalized Rayleigh distributed Envelope PDF, $\alpha = 0.5$, $\beta = 0.048$

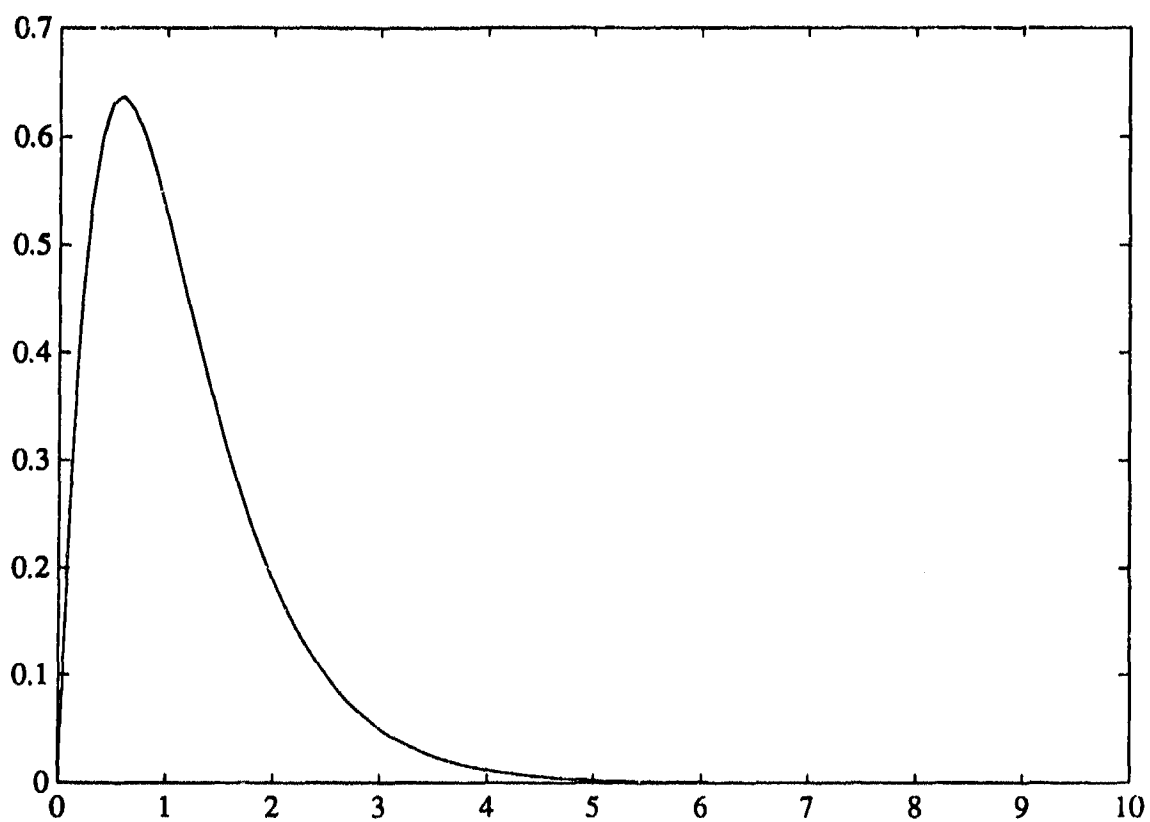


Figure 4.17: Generalized Rayleigh distributed Envelope PDF, $\alpha = 1$, $\beta = 0.577$

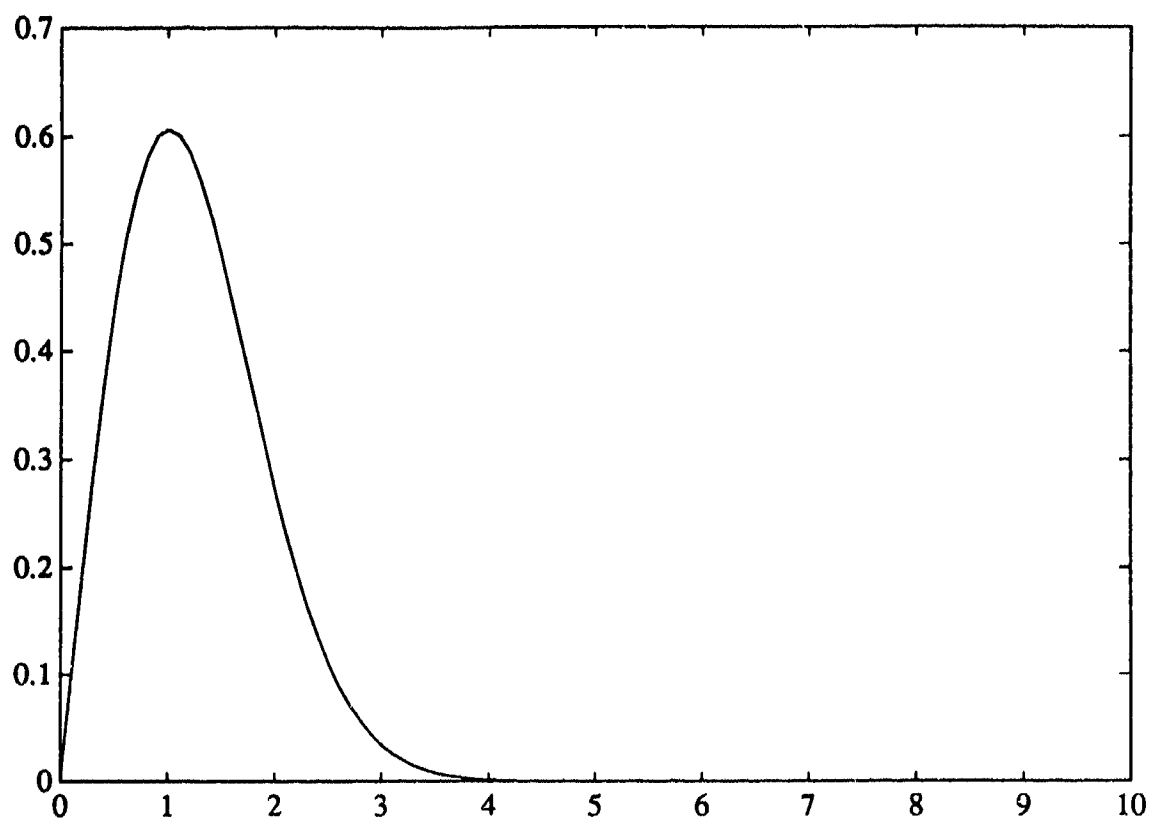


Figure 4.18: Generalized Rayleigh distributed Envelope PDF, $\alpha = 2$, $\beta = 1.414$

4.4.2.4 Rician Envelope PDF Arising from a zeromean complex Gaussian Process with correlated quadrature components

There are two possible ways in which the Rician envelope PDF occurs. One possibility arises through a complex zero mean random process with correlated quadrature components that are Gaussian. The other is through a non-zero mean complex Gaussian process. The former case is considered here, since the SIRV PDF can be obtained by differentiation of $h_2(w)$. For this case, the envelope PDF is given by

$$f_R(r) = \frac{r}{\sqrt{1-\rho^2}} \exp\left[-\frac{r^2}{2(1-\rho^2)}\right] I_0\left[\frac{\rho r^2}{2(1-\rho^2)}\right] \quad (4.73)$$

(0 ≤ r ≤ ∞)
(0 < ρ ≤ 1)

where $I_0(x)$ is the modified Bessel's function of the first kind of order zero. Plots of the Rician envelope PDF for several values of ρ are shown in Figures 4.19-4.21. Let

$$A = \frac{\sigma^2}{2(1-\rho^2)}. \quad (4.74)$$

Using eq (4.14) we have

$$h_2(w) = \frac{\sigma^2}{\sqrt{1-\rho^2}} \exp(-Aw) I_0(\rho Aw). \quad (4.75)$$

From eq (3.25)

$$h_{2N}(w) = (-2)^{N-1} \frac{d^{N-1} h_2(w)}{dw^{N-1}}. \quad (4.76)$$

We then use eq (4.57) and the identities [45]

$$I_n(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos(n\theta) \exp[x \cos(\theta)] d\theta$$

$$\cos^k(\theta) = \frac{1}{2^k} \sum_{m=0}^k \binom{k}{m} \cos[(k-2m)\theta] \quad (4.77)$$

to obtain

$$h_{2N}(w) = \frac{\sigma^{2N}}{(1-\rho^2)^{N-\frac{1}{2}}} \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \left(\frac{\rho}{2}\right)^k \xi_k \exp(-Aw) \quad (4.78)$$

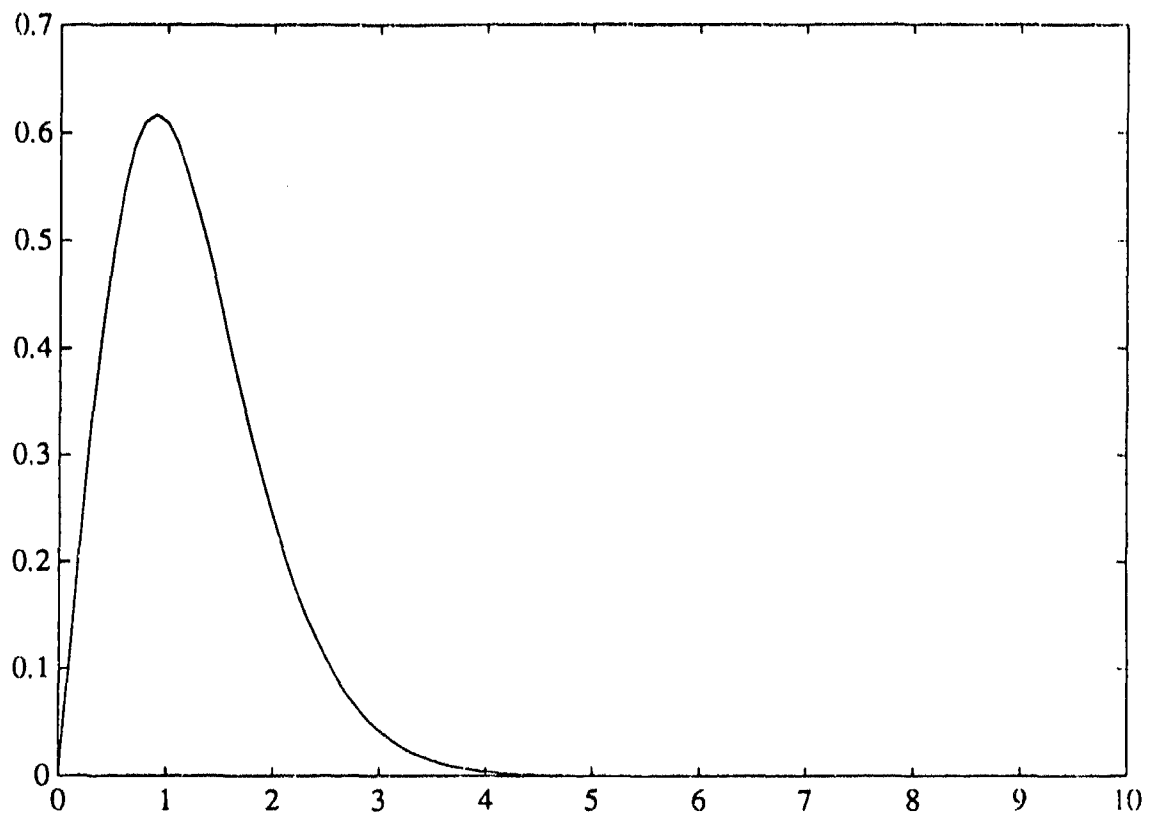


Figure 4.19: Rician Envelope PDF, $\rho = 0.25$

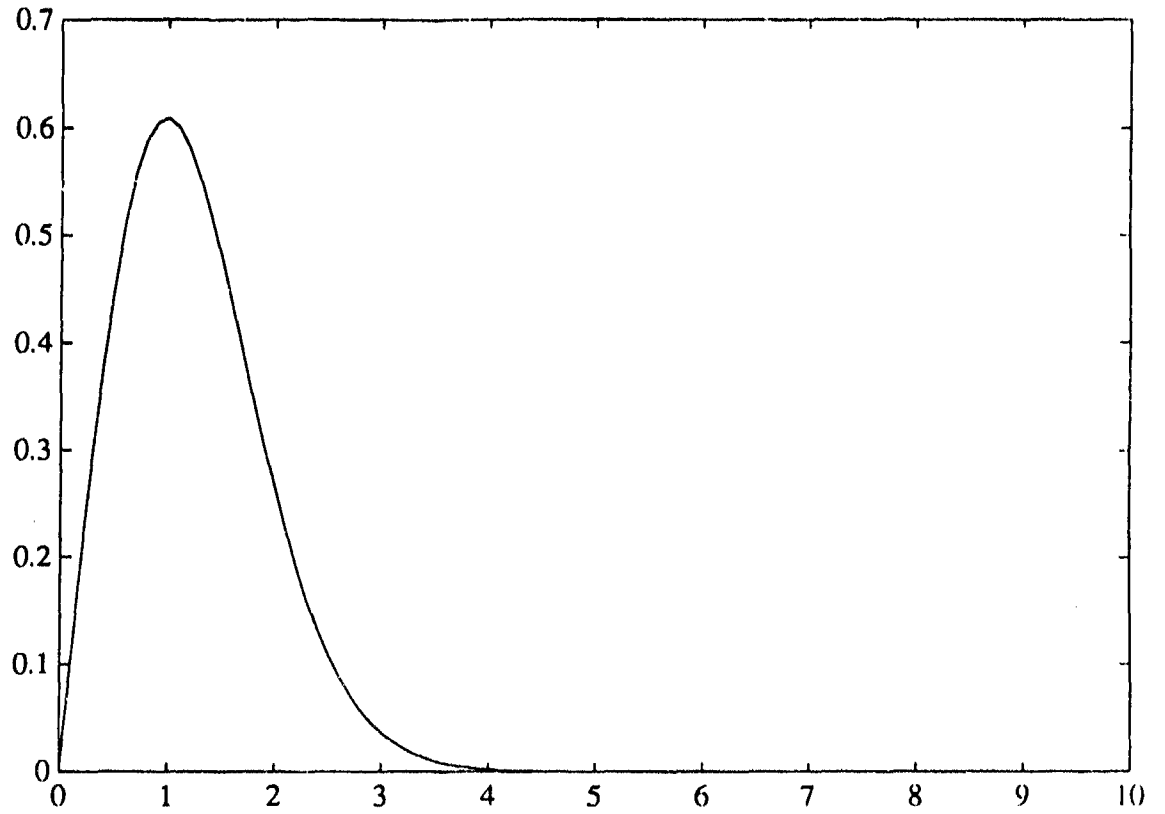


Figure 4.20: Rician Envelope PDF, $\rho = 0.5$

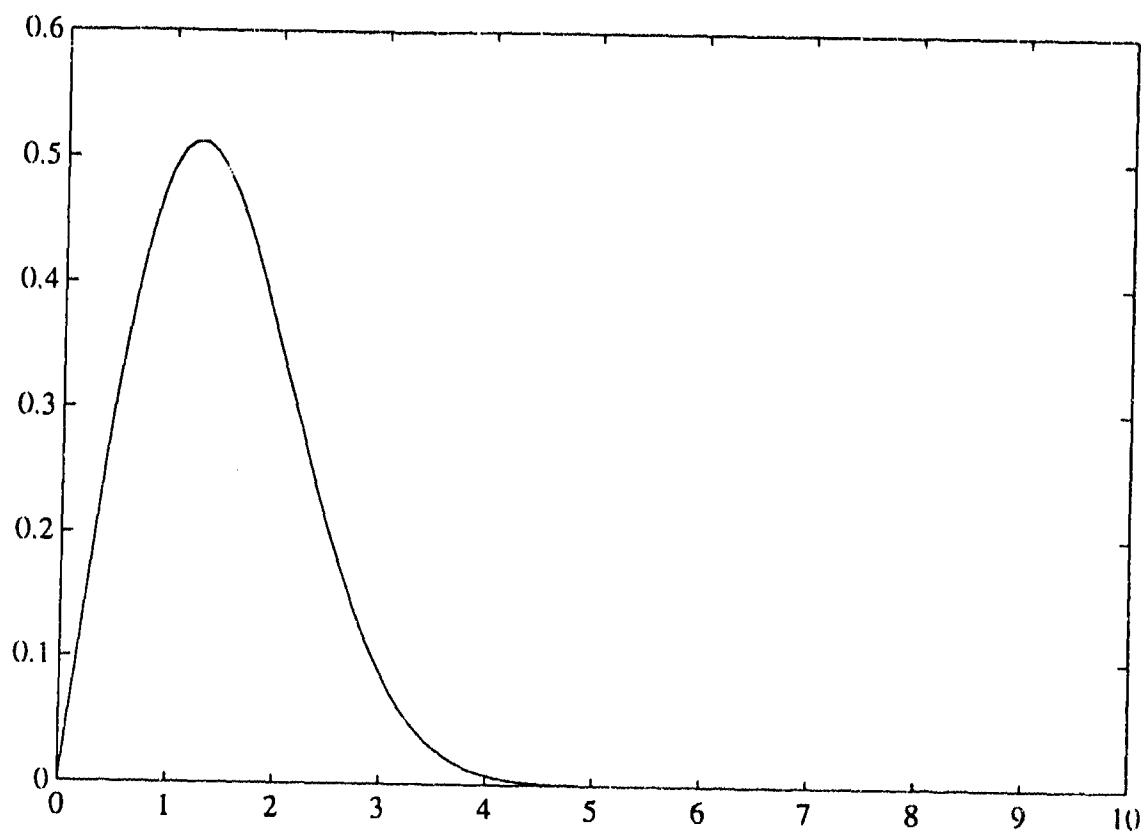


Figure 4.21: Rician Envelope PDF, $\rho = 0.9$

where

$$\xi_k = \sum_{m=0}^k \binom{k}{m} I_{k-2m}(\rho A w). \quad (4.79)$$

For $\rho = 0$, note that the Rician envelope PDF corresponds to the Rayleigh envelope PDF.

4.4.3 Examples Using the Marginal Characteristic Function

Successful use of the marginal characteristic function approach requires the knowledge of various Hankel transforms. For each example, the particular transform used is cited by equation and page number as it appears in [46]. To illustrate the procedure followed, a detailed derivation is presented in the first example. However, in the remaining examples, we simply list the univariate characteristic function of the quadrature components, the corresponding marginal PDF and the resulting $h_{2N}(w)$. Finally, $h_{2N}(p)$ is obtained by replacing w with p in the expressions for $h_{2N}(w)$.

4.4.3.1 Gaussian Distribution

First, we consider the characteristic function given by

$$\Psi(\omega) = \exp\left(-\frac{\omega^2}{2}\right). \quad (4.80)$$

The corresponding marginal PDF of the quadrature components is

$$f_{Y_k}(y_k) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{y_k^2}{2}\right) \quad (-\infty \leq y_k \leq \infty). \quad (4.81)$$

Equation (4.81) is the PDF of a zero mean unit variance Gaussian random variable. Substitution of eq (4.80) in eq (4.30) yields

$$h_{2N}(w) = (\sqrt{w})^{1-N} \int_0^\infty \omega^N \exp\left(-\frac{\omega^2}{2}\right) J_{N-1}(w\sqrt{\omega}) d\omega. \quad (4.82)$$

From [46], eq (10), p29, we have the Hankel transform

$$\int_0^\infty x^{\nu+\frac{1}{2}} \exp(-ax^2) J_\nu(xy) \sqrt{xy} dx = \frac{y^{\nu+\frac{1}{2}}}{(2a)^{\nu+1}} \exp\left(-\frac{y^2}{4a}\right). \quad (4.83)$$

By making the association that $a = 0.5$, $\nu = N-1$, $x = \omega$ and $y = \sqrt{w}$, the above result becomes

$$\int_0^\infty \omega^N \exp\left(-\frac{\omega^2}{2}\right) J_{N-1}(w\sqrt{\omega}) \sqrt{w} d\omega = \sqrt{w}^{N-1+\frac{1}{2}} \exp\left(-\frac{w}{2}\right). \quad (4.84)$$

It follows that

$$h_{2N}(w) = \exp\left(-\frac{w}{2}\right). \quad (4.85)$$

From eq (4.1), it is seen that the resulting SIRV PDF is the familiar multivariate Gaussian PDF, given by

$$f_Y(y) = (2\pi)^{-N} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{p}{2}\right). \quad (4.86)$$

4.4.3.2 K-Distribution

The marginal characteristic function given by

$$\Psi(\omega) = \left(1 + \frac{\omega^2}{b^2}\right)^{-\alpha} \quad (4.87)$$

corresponds to the K-distributed envelope whose PDF is

$$f_R(r) = \frac{2b}{\Gamma(\alpha)} \left(\frac{br}{2}\right)^\alpha K_{\alpha-1}(br) u(r) \quad (4.88)$$

where α is the shape parameter of the distribution, b denotes its scale parameter, $K_N(t)$ is the N^{th} order modified Bessel function of the second kind and $u(r)$ is the unit step function. The pertinent Hankel transform for this example is found as [46] eq (20), p24:

$$\int_0^\infty x^{v+\frac{1}{2}} (x^2 + a^2)^{-u-1} J_v(xy) \sqrt{xy} dx = \frac{a^{v-u} y^{u+\frac{1}{2}} K_{v-u}(ay)}{2^u \Gamma(u+1)} \quad (4.89)$$

The resulting $h_{2N}(w)$ is

$$h_{2N}(w) = \frac{b^{2N}}{\Gamma(\alpha)} \frac{(b\sqrt{w})^{\alpha-N}}{2^{\alpha-1}} K_{N-\alpha}(b\sqrt{w}). \quad (4.90)$$

As a special case, when α is equal to unity, eq (4.87) is the characteristic function of the Laplace distribution for the quadrature components whose PDF is given by

$$f_{Y_k}(y_k) = \frac{b}{2} \exp(-b|y_k|) \quad (-\infty \leq y_k \leq \infty) \quad (4.91)$$

where $|y_k|$ denotes the absolute value of y_k and b denotes the scale parameter. The corresponding $h_{2N}(w)$ is given by

$$h_{2N}(w) = b^{2N} (b\sqrt{w})^{1-N} K_{N-1}(b\sqrt{w}). \quad (4.92)$$

Another interesting case of the K-distribution arises when $\alpha = 0.5$. Since $K_{\frac{1}{2}}(t) = \sqrt{\frac{\pi}{2t}} \exp(-t)$, this corresponds to the exponential distribution for the marginal envelope PDF. Therefore, the K-

distributed envelope PDF with $\alpha = 0.5$ is identical to the Weibull distributed envelope with $b = 1$. Although the characteristic PDF of the Weibull SIRV is unknown in general, the characteristic PDF of the Weibull SIRV for $b = 1$ is obtained when $\alpha = 0.5$ in eq (4.43). Finally, we point out that the K-distributed envelope reduces to the Rayleigh envelope PDF when α tends to ∞ .

4.4.3.3 Student-t Distribution

The characteristic function for the Student-t distribution with scale parameter b and shape parameter ν is given by

$$\Psi(\omega) = \frac{K_\nu(b\omega)(b\omega)^\nu}{2^{\nu-1}\Gamma(\nu)}. \quad (4.93)$$

Note the functional similarity with the envelope PDF given by eq (4.88). The Student-t distribution is referred to as the generalized Cauchy distribution in [47] because the marginal PDF of the quadrature components is given by

$$f_{Y_k}(y_k) = \frac{\Gamma(\nu + \frac{1}{2})}{b\sqrt{\pi}\Gamma(\nu)} \left(1 + \frac{y_k^2}{b^2}\right)^{-\nu-\frac{1}{2}} \quad (-\infty \leq x_k \leq \infty), \quad \nu > 0 \quad (4.94)$$

where $\Gamma(\nu)$ is the Euler-Gamma function. The relevant Hankel transform, [46] eq (3), p63 is

$$\int_0^\infty x^{u+v+\frac{1}{2}} K_u(ax) J_v(xy) \sqrt{xy} dx = \frac{2^{v+u} a^u \Gamma(u+v+1) y^{v+\frac{1}{2}}}{(y^2 + a^2)^{u+v+1}}. \quad (4.95)$$

Using eq (4.30), $h_{2N}(w)$ is expressed as

$$h_{2N}(w) = \frac{2^N b^{2\nu} \Gamma(\nu + N)}{\Gamma(\nu) (b^2 + w)^{N+\nu}}. \quad (4.96)$$

The Cauchy PDF for the quadrature components arises when ν is set equal to $\frac{1}{2}$ in eq (4.94) and is given by

$$f_{Y_k}(y_k) = \frac{b}{\pi(b^2 + y_k^2)} \quad (-\infty \leq x_k \leq \infty) \quad (4.97)$$

where b is the scale parameter. The corresponding $h_{2N}(w)$ is

$$h_{2N}(w) = \frac{2^N b \Gamma(\frac{1}{2} + N)}{\sqrt{\pi} (b^2 + w)^{N+\frac{1}{2}}}. \quad (4.98)$$

Note that the Cauchy PDF does not have finite variance. However, this PDF is useful in modeling impulsive noise [48]. Finally, we point out that when $b = \sqrt{2\nu}$ and ν tends to ∞ in eq (4.94), the Student-t distribution reduces to the Gaussian distribution.

4.4.3.4 Rician Envelope PDF arising from a non-zero mean complex Gaussian Process

We consider the Rician envelope PDF, arising from a non-zero mean complex Gaussian process, given by

$$f_R(r) = \frac{r}{\alpha^2} \exp\left[-\frac{(r^2 + a^2)}{2\alpha^2}\right] I_0\left(\frac{ar}{\alpha^2}\right). \quad (4.99)$$

Plots of the Rician envelope PDF are shown in Figures 4.20-4.22 for $\alpha = 1$ and several values of a . Note that this PDF approaches the Rayleigh PDF as a tends to zero. For convenience, we assume that $\sigma^2 = \frac{1}{2}E(R^2) = 1$. Using eq (4.14), we have

$$h_2(r^2) = A \exp\left(-\frac{r^2}{2\alpha^2}\right) I_0\left(\frac{ar}{\alpha^2}\right) \quad (4.100)$$

where $A = \frac{\exp(-\frac{a^2}{2\alpha^2})}{\alpha^2}$. Noting that [45]

$$\int_0^\infty x \exp(-\alpha x^2) I_\nu(\beta x) J_\nu(\gamma x) dx = \frac{1}{2\alpha} \exp\left(\frac{\beta^2 - \gamma^2}{4\alpha}\right) J_\nu\left(\frac{\beta\gamma}{2\alpha}\right) \quad (4.101)$$

$$\operatorname{Re}\{\alpha\} > 0, \operatorname{Re}\{\nu\} > -1,$$

eq (4.21) results in the characteristic function

$$\Psi(\omega) = \exp\left(-\frac{\omega^2 \alpha^2}{2}\right) J_0\left(\frac{\omega a}{\alpha^2}\right). \quad (4.102)$$

Recognizing that [45]

$$\int_0^\infty x^{\lambda-1} \exp(-\alpha x^2) J_\mu(\beta x) J_\nu(\gamma x) dx = \frac{\beta^\mu \gamma^\nu \alpha^{-\frac{(\mu+\nu+\lambda+2)}{2}}}{2^{\mu+\nu+1} \Gamma(\nu+1)} \sum_{m=0}^\infty \frac{\Gamma(m+\frac{\mu}{2}+\frac{\nu}{2}+\frac{\lambda+1}{2})}{m! \Gamma(m+\mu+1)} \left(-\frac{\beta^2}{4\alpha}\right)^m F(-m, -\mu-m; \nu+1; \frac{\gamma^2}{\beta^2}) \quad (4.103)$$

$$\operatorname{Re}\{\alpha\} > 0, \operatorname{Re}\{\mu + \nu + \lambda\} > -2, \beta > 0, \gamma > 0$$

where $F(., .; .; .)$ is the four parameter hypergeometric function, it follows from eq (4.30) that

$$h_{2N}(w) = \frac{\alpha^{2N+2}}{2^{2N+1} \Gamma(N)} \sum_{m=0}^\infty \frac{\Gamma(m+N+1)}{m! \Gamma(m+1)} \left(\frac{-a^2}{2\alpha^6}\right)^m F(-m, -m; N; \frac{w\alpha^4}{a^2}) \quad (4.104)$$

Since $h_{2N}(w)$ for this example involves an infinite series of hypergeometric functions, form is mathematically intractable. Therefore, the corresponding multivariate SIRV PDF does not lend

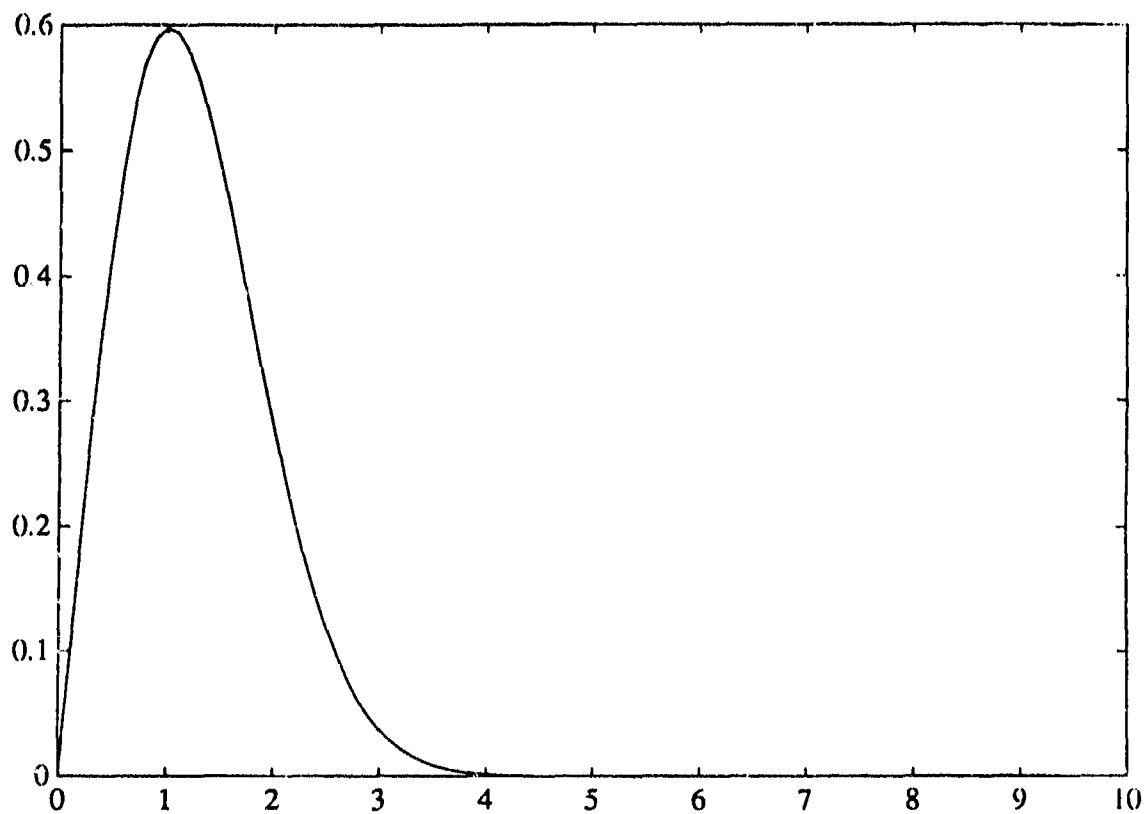


Figure 4.22: Rician Envelope PDF, $a = 0.25$, $\alpha = 1$

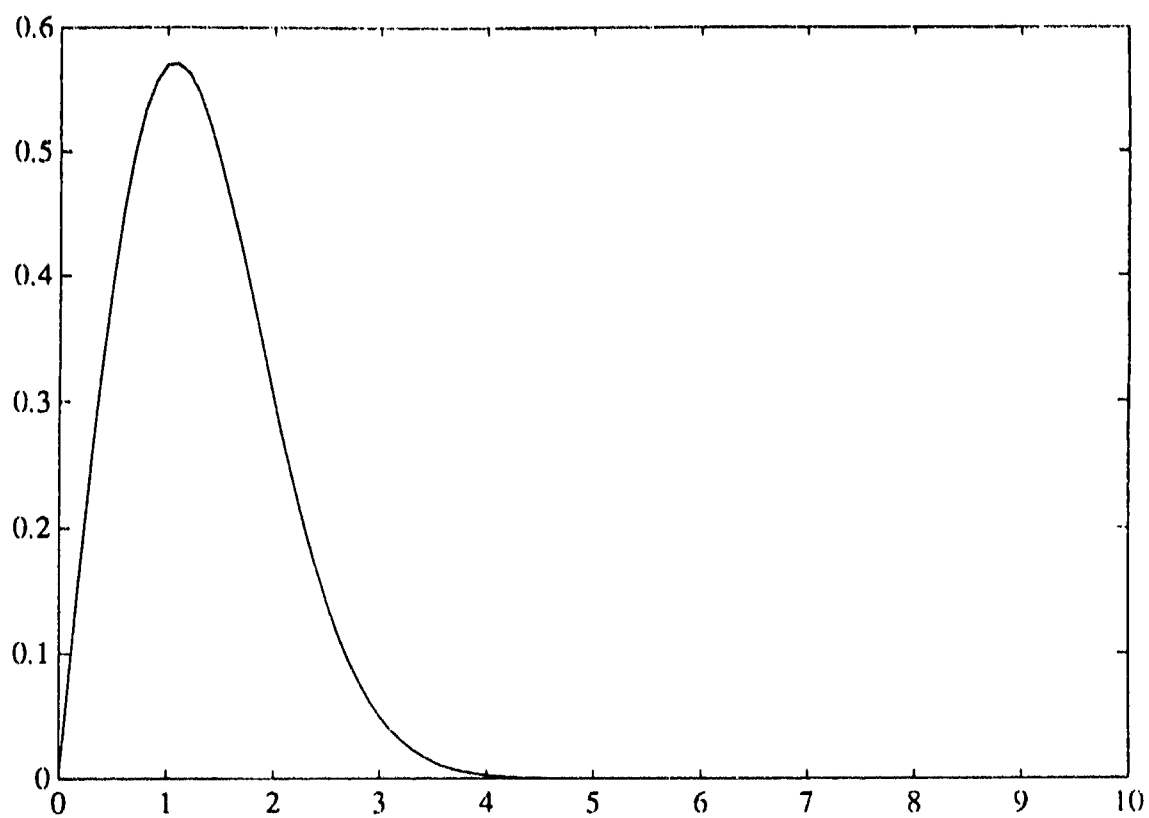


Figure 4.23: Rician Envelope PDF, $a = 0.5$, $\alpha = 1$

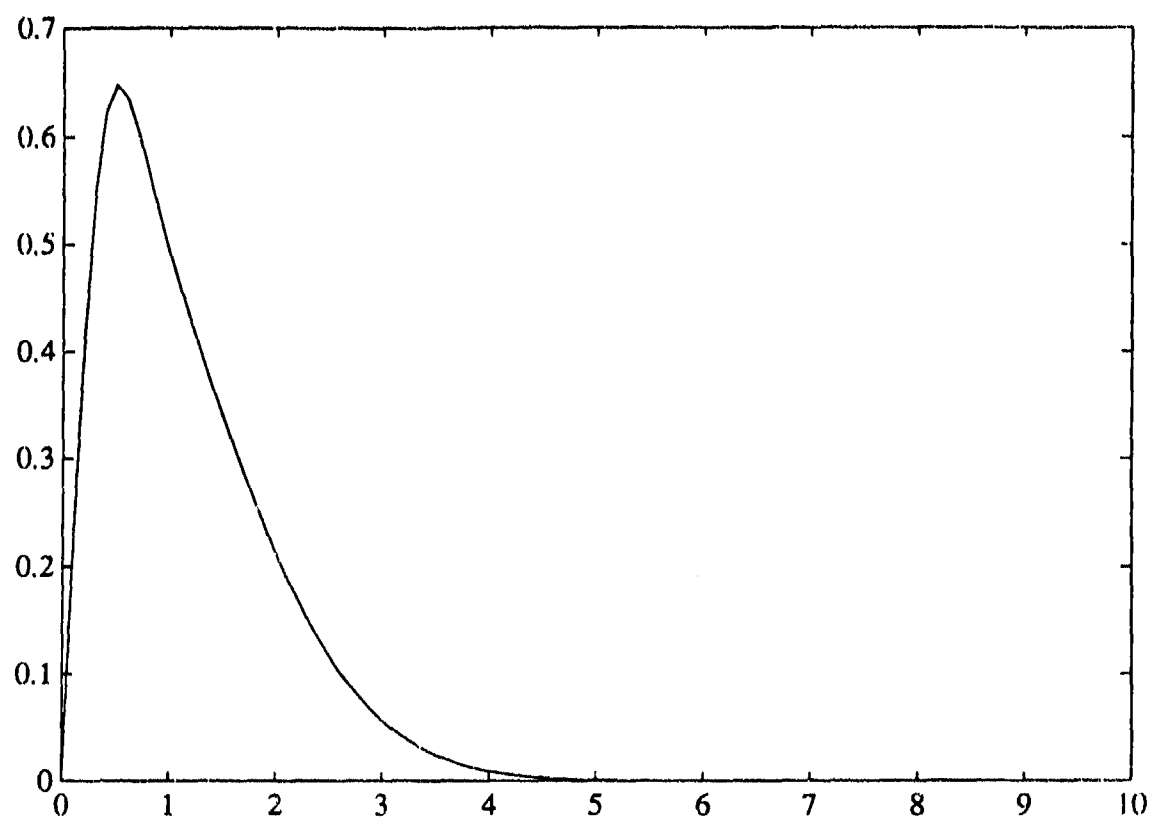


Figure 4.24: Rician Envelope PDF, $a = 0.9$, $\alpha = 1$

Table 4.1: Marginal PDF

Marginal PDF	$f_X(x)$
Chi	$\frac{1}{\Gamma(\nu)}(bx)^{\nu-1}\exp(-b^2x^2)$
Weibull	$abx^{b-1}\exp(-ax^b)$
Generalized Rayleigh	$\frac{a}{b^2}\exp[-(\frac{x}{b})^a]$
Rician	$\frac{x}{\sqrt{1-\rho^2}}\exp[-\frac{x^2}{2(1-\rho^2)}]I_0[\frac{\rho x^2}{2(1-\rho^2)}]$
Gaussian	$\sqrt{2\pi}^{-1}\exp(-\frac{x^2}{2})$
Laplace	$\frac{1}{2}\exp(-b x_k)$
Cauchy	$\frac{1}{\pi(b^2+x^2)}$
K-distribution	$\frac{1}{\Gamma(\frac{\alpha}{2})(\frac{b}{2})^\alpha}K_{\alpha-1}(bx)u(x)$
Student-t	$\frac{\Gamma(\frac{\nu+1}{2})}{b\sqrt{\pi}\Gamma(\frac{\nu}{2})}(1+\frac{x^2}{b^2})^{-\nu-\frac{1}{2}}$

itself for use in practical applications.

4.4.3.5 Summary

The results derived in this section are summarized here. As a point of interest, it is mentioned that the log-normal envelope PDF given by

$$f_R(r) = \frac{1}{\sqrt{2\pi}y} \exp\left[-\frac{\{\log(r)\}^2}{2}\right] \quad (4.105)$$

and the Johnson (unbounded) distribution whose PDF is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\delta\sqrt{1+y^2}} \exp\left[-\frac{\{\sinh^{-1}(y) - \gamma\}^2}{2\delta^2}\right] \quad (4.106)$$

cannot be extended to SIRVs because $h_2(w)$ for each of these distributions fails to satisfy the monotonicity conditions stated in section 4.3.

Table 4.1, presents a list of marginal PDFs suitable for extension to SIRVs. Table 4.2 tabulates $h_{2N}(p)$ for those marginal PDFs treated as envelope PDFs while Table 4.3 gives those $h_{2N}(p)$ obtained from the associated marginal characteristic function.

Plots of eq (4.8) with $N = 1$ for the various SIRV PDFs are shown in Figures 4.25-4.33. In all the plots, the covariance matrix used is given by

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (4.107)$$

Observe that each PDF is unimodal. However, the width and height of the peak along with the

Table 4.2: SIRVs obtained from the marginal envelope PDF

Marginal PDF	$h_{2N}(p)$
Chi	$(-2)^{N-1} A \sum_{k=1}^N G_k p^{\nu-k} \exp(-Bp)$
	$G_k = \binom{N-1}{k-1} (-1)^{k-1} B^{k-1} \frac{\Gamma(\nu)}{\Gamma(\nu-k+1)}$
	$A = \frac{1}{\Gamma(\nu)} (b\sigma)^{2\nu}$
	$B = b^2 \sigma^2$
	$\nu \leq 1$
Weibull	$\sum_{k=1}^N C_k p^{\frac{N}{2}-N} \exp(-Ap^{\frac{1}{2}})$
	$A = a\sigma^b$
	$C_k = \sum_{m=1}^k (-1)^{m+N} 2^N \frac{A^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{2})}{\Gamma(1+\frac{m}{2}-N)}$
	$b \leq 2$
Gen. Rayleigh	$\sum_{k=1}^{N-1} D_k p^{\frac{N}{2}-N+1} \exp(-Bp^{\frac{1}{2}})$
	$A = \frac{\sigma^2 \alpha}{\beta^2 \Gamma(\frac{1}{2})}$
	$B = \beta^{-\alpha} \sigma^{\alpha}$
	$D_k = \sum_{m=1}^k (-1)^{m+N-1} 2^{N-1} \frac{B^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{2})}{\Gamma(2+\frac{m}{2}-N)}$
	$\alpha \leq 2$
Rician	$\frac{\sigma^{2N}}{(1-\rho^2)^{N-\frac{1}{2}}} \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \left(\frac{\rho}{2}\right)^k \xi_k \exp(-A)$
	$\xi_k = \sum_{m=0}^k \binom{k}{m} I_{k-2m}(\rho A), A = \frac{p\sigma^2}{2(1-\rho^2)}$

Table 4.3: SIRVs obtained from the marginal characteristic function

Marginal PDF	$h_{2N}(p)$
Gaussian	$\exp(-\frac{p}{2})$
Laplace	$b^{2N} (b\sqrt{p})^{1-N} K_{N-1}(b\sqrt{p})$
Cauchy	$\frac{2^N b \Gamma(\frac{1}{2}+N)}{\sqrt{\pi} (b^2+p)^{N+\frac{1}{2}}}$
K-distribution	$\frac{b^{2N} (b\sqrt{p})^{\alpha-N}}{\Gamma(\alpha)} \frac{1}{2^{\alpha-1}} K_{N-\alpha}(b\sqrt{p})$
Student-t	$\frac{2^N b^{2\nu} \Gamma(\nu+N)}{\Gamma(\nu) (b^2+p)^{N+\nu}}$

behavior of the extreme values (i.e. the tails) differ significantly.

4.5 Significance of the Quadratic form of the SIRV PDF

Thus far, our discussion has focused on techniques that can be used to obtain the PDF of an SIRV starting from either the first order PDF or the first order characteristic function. Given random data, we are also interested in the problem of approximating the distribution of the underlying data. The problem of multivariate distribution identification is of interest in radar signal detection. Since the background clutter is not known a priori, there is a need to identify the underlying clutter PDF based on measurements obtained from a given environment. Since the radar processes N pulses at a time, knowledge of the joint PDF of the N samples is necessary in order to obtain the optimal radar signal processor for the given clutter background. We present an important theorem here which enables us to address the distribution approximation of an SIRV.

Theorem 4 *The PDF of the quadratic form appearing in eq (3.15) is given by*

$$f_P(p) = \frac{1}{2^{\frac{N}{2}} \Gamma(\frac{N}{2})} p^{\frac{N}{2}-1} h_N(p) \quad (0 \leq p \leq \infty). \quad (4.108)$$

Proof: First, we consider a spherically symmetric random vector (SSRV) $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$. Because an SSRV is a special case of the SIRV, the representation theorem can be used to express $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{Z}S \quad (4.109)$$

where $\mathbf{Z}$ is a Gaussian random vector having zero mean and identity covariance matrix and S is a non-negative random variable with PDF $f_S(s)$. Consider the random variable

$$P' = \mathbf{X}^T \mathbf{X}. \quad (4.110)$$

Using eq (4.109) in eq (4.110) gives

$$P' = \mathbf{Z}^T \mathbf{Z} S^2. \quad (4.111)$$

Since $\mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^N Z_i^2$ is the sum of the squares of independent identically distributed Gaussian random variables having zero mean and unit variance, the PDF of $V = \mathbf{Z}^T \mathbf{Z}$ is a Chi square

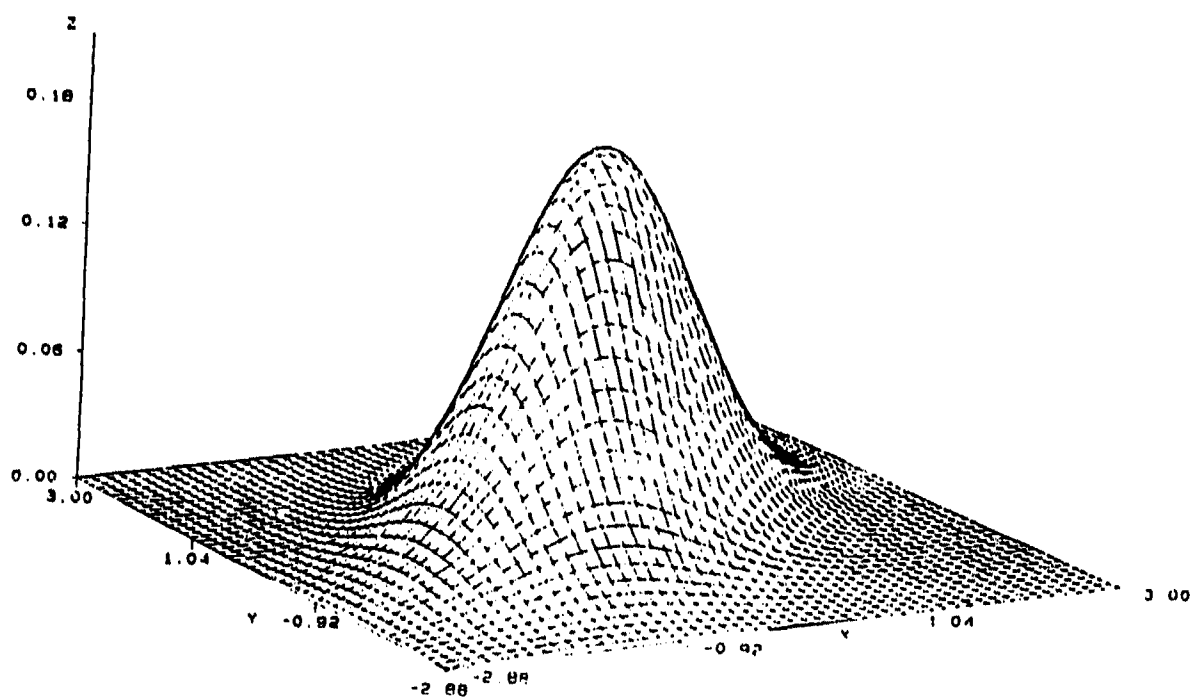


Figure 4.25: Gaussian distribution, zero mean, unit variance

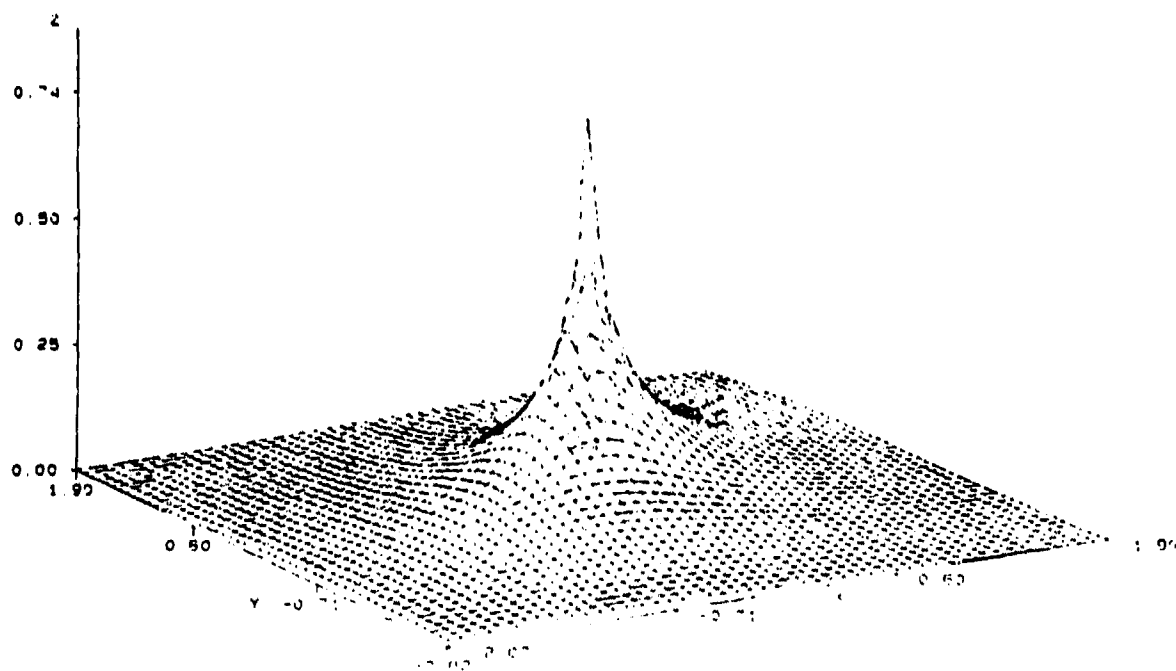


Figure 4.26: Laplace Distribution, $b = 1$

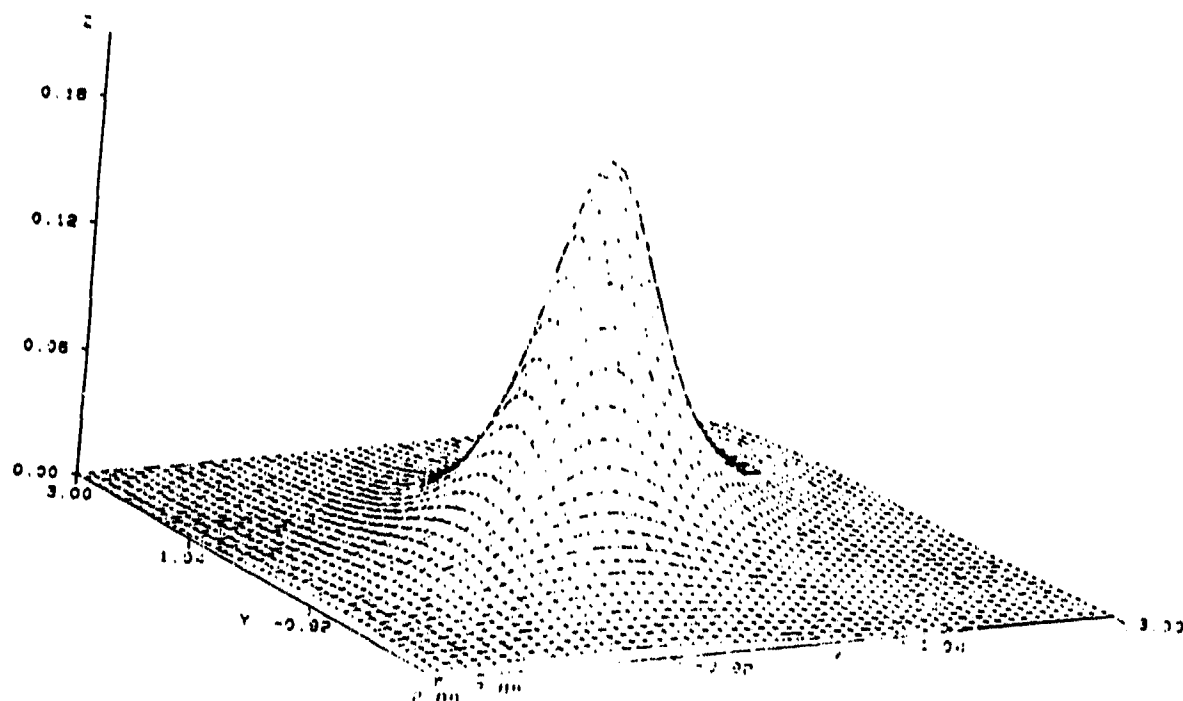


Figure 4.27: Cauchy Distribution, $b = 1$

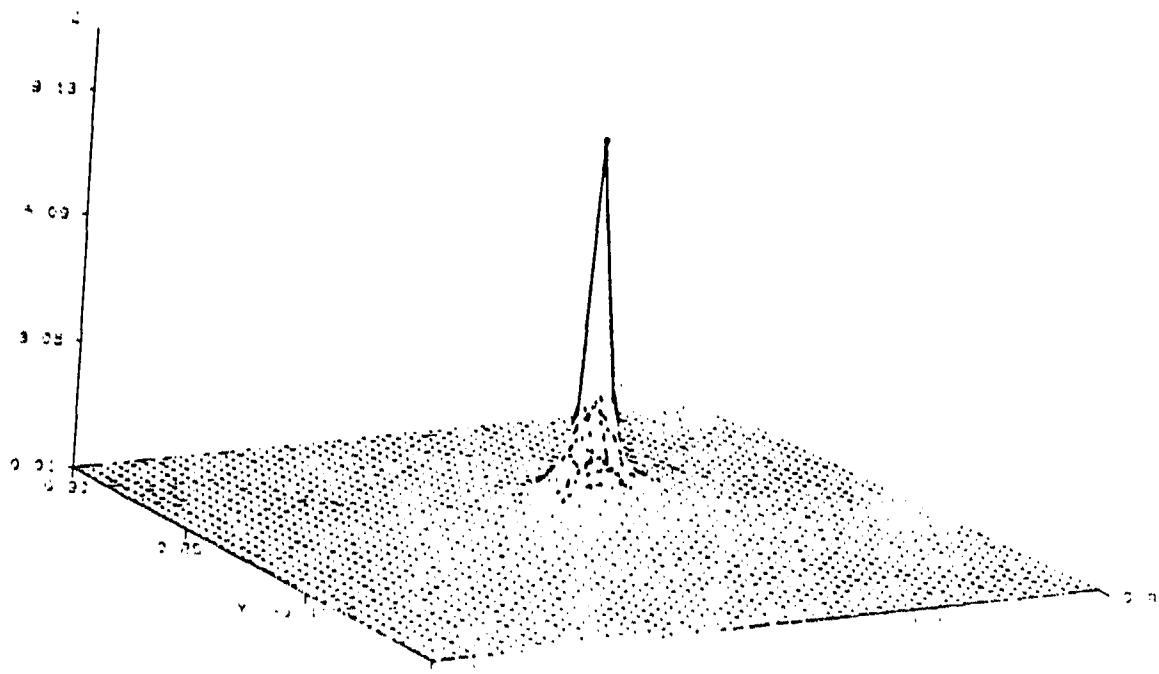


Figure 4.28: K-distribution, $b = 1$, $\alpha = 0.5$

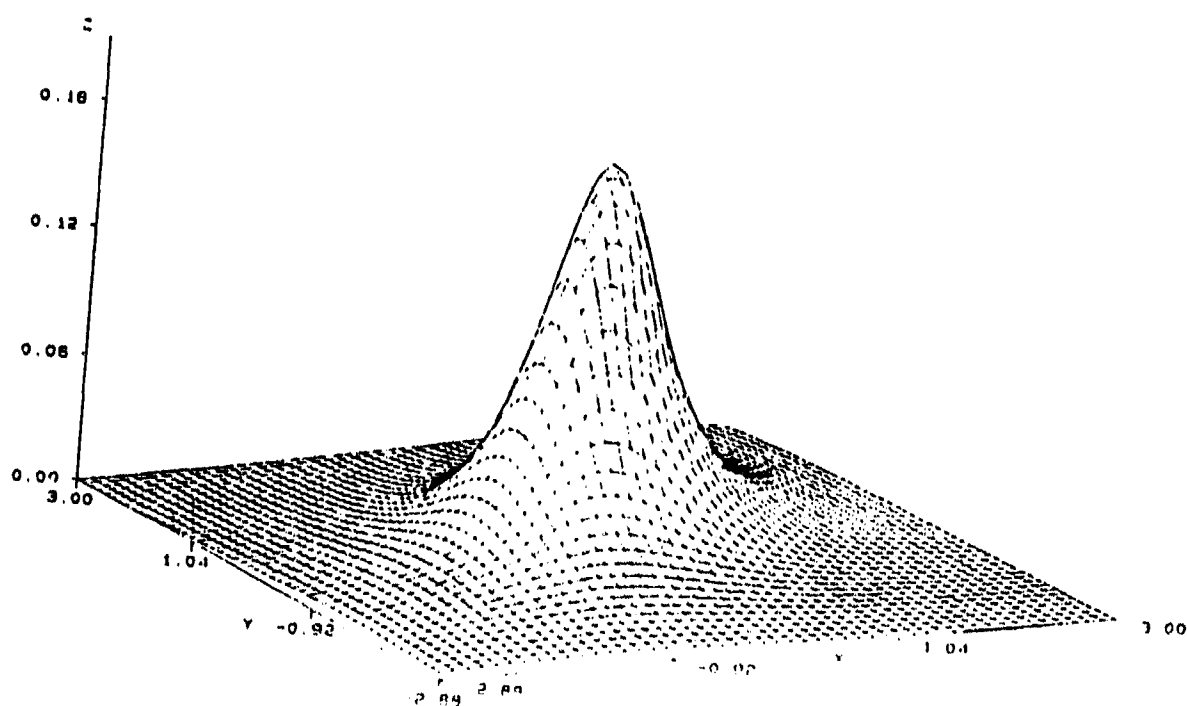


Figure 4.29: Student-t Distribution, $b = 1$, $\nu = 1.5$

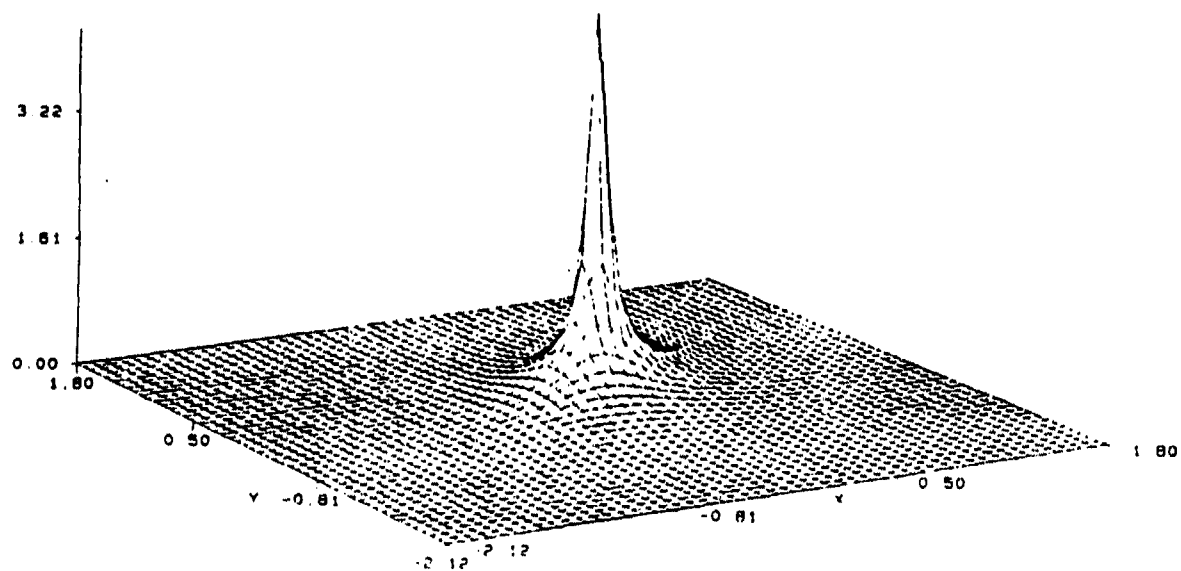


Figure 4.30: Chi-distribution, $b = 1, \nu = 1$

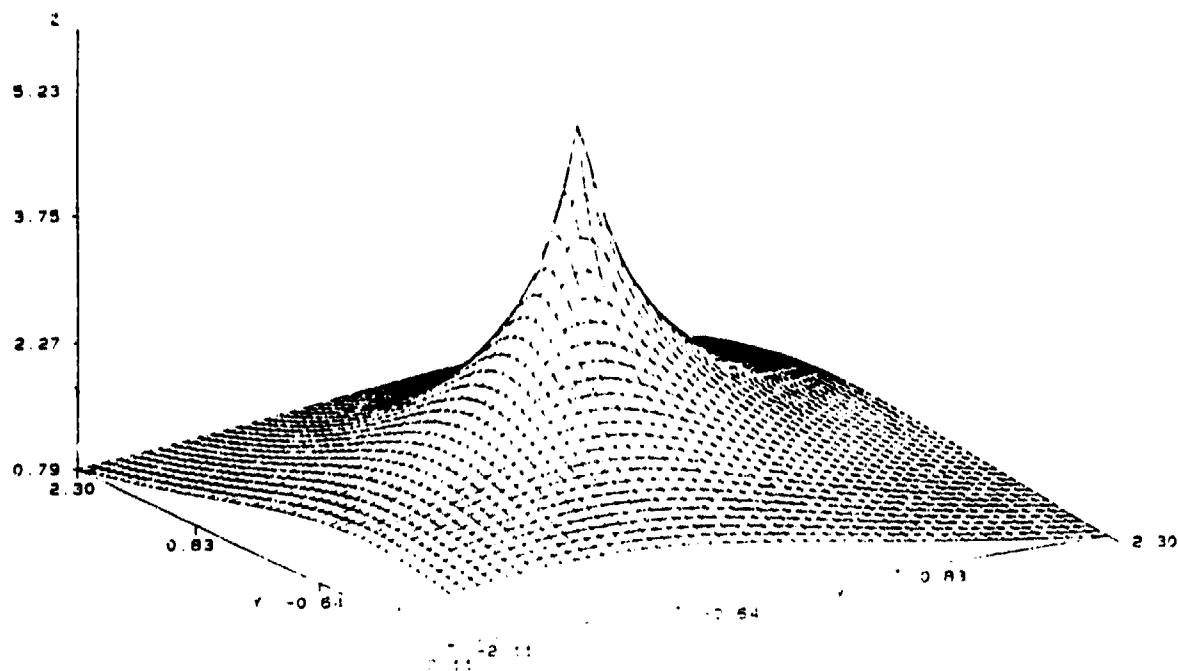


Figure 4.31: Generalized Rayleigh PDF, $\alpha = 0.5$, $\beta = 0.05$

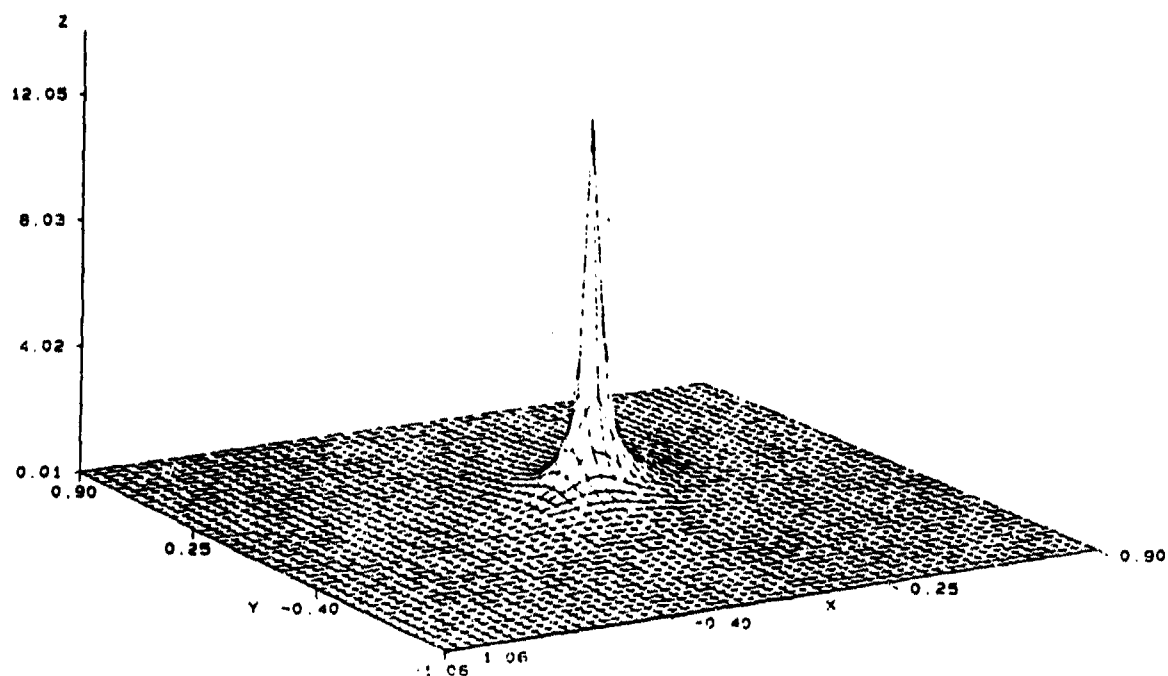


Figure 4.32: Weibull distribution, $a = 1$, $b = 1.0$

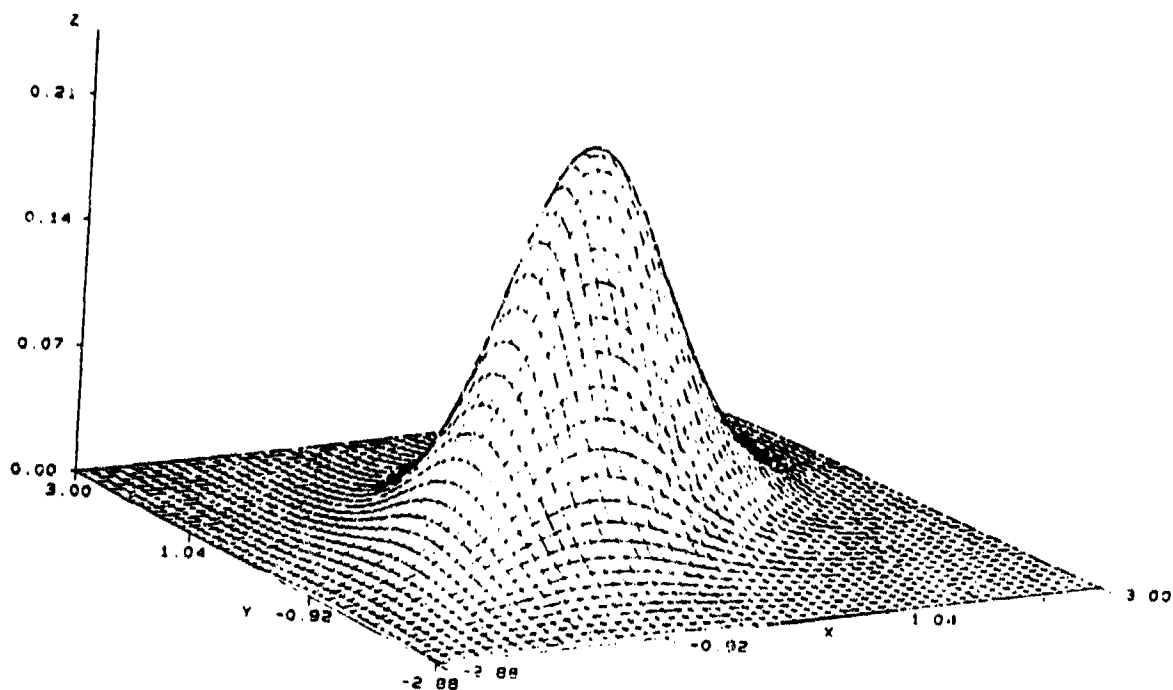


Figure 4.33: Rician PDF, $\rho = 0.5$

distribution with N degrees of freedom. Consequently,

$$f_V(v) = \frac{v^{\frac{N}{2}-1}}{2^{\frac{N}{2}}\Gamma(\frac{N}{2})} \exp(-\frac{v}{2}) \quad ; \quad v \geq 0. \quad (4.112)$$

Noting that $P' = VS^2$, it follows that

$$f_{P'|S}(p'|s) = \frac{(p')^{\frac{N}{2}-1}}{2^{\frac{N}{2}}\Gamma(\frac{N}{2})} s^{-N} \exp(-\frac{p'}{2s^2}). \quad (4.113)$$

From the theorem of total probability, we have

$$f_{P'}(p') = \int_0^\infty \frac{(p')^{\frac{N}{2}-1}}{2^{\frac{N}{2}}\Gamma(\frac{N}{2})} s^{-N} \exp(-\frac{p'}{2s^2}) f_S(s) ds. \quad (4.114)$$

Recall from Theorem 2 that

$$h_N(p') = \int_0^\infty s^{-N} \exp(-\frac{p'}{2s^2}) f_S(s) ds. \quad (4.115)$$

Consequently, the PDF of P' is expressed as

$$f_{P'}(p') = \frac{(p')^{\frac{N}{2}-1}}{2^{\frac{N}{2}}\Gamma(\frac{N}{2})} h_N(p'). \quad (4.116)$$

Recall that an SIRV $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$ having a mean vector $\mathbf{b}$ and covariance matrix Σ is related to the SSRV $\mathbf{X}$ by the linear transformation

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} \quad (4.117)$$

where $\Sigma = \mathbf{A}\mathbf{A}^T$. Observe that

$$\begin{aligned} P &= (\mathbf{Y} - \mathbf{b})^T \Sigma^{-1} (\mathbf{Y} - \mathbf{b}) \\ &= (\mathbf{A}\mathbf{X})(\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{A}\mathbf{X} \\ &= \mathbf{X}^T \mathbf{X}. \end{aligned} \quad (4.118)$$

Since $P = P'$, the PDF of the quadratic form P which is associated with $\mathbf{Y}$ is

$$f_P(p) = \frac{(p)^{\frac{N}{2}-1}}{2^{\frac{N}{2}} \Gamma(\frac{N}{2})} h_N(p). \quad (4.119)$$

This establishes the theorem. Thus, an SIRV is uniquely characterized by the quadratic form appearing in its PDF. Knowledge of the quadratic form PDF is sufficient to identify the SIRV PDF. This is an important result since it allows us to reduce the multivariate distribution identification problem to the equivalent problem of univariate distribution identification of the quadratic form. We point out that the invariance of the distribution of the quadratic form, regardless of whether we are dealing with an SIRV or an SSRV, arises from the fact that the random vector is distributed over an N dimensional hypersphere of radius R . The radius of the hypersphere remains unchanged regardless of whether we consider an SIRV or an SSRV. Only the azimuthal angles and radial angle change depending on whether the random vector is a SSRV or an SIRV. In context of the radar problem, we are dealing with N complex samples or $2N$ quadrature components. The results presented in this section are applicable when N is replaced by $2N$.

4.6 Conclusion

In this chapter we have pointed out a method to obtain the PDF of correlated non-Gaussian random vectors arising in the problem of radar clutter modeling. The theory of SIRPs has been used to develop the multivariate PDFs. Various techniques have been presented to obtain SIRV PDFs. Several examples are provided to illustrate these techniques. Finally, we have obtained the PDF of the quadratic form of a SIRV and we have shown that this PDF remains unchanged regardless of whether we are dealing with an SSRV or an SIRV. We have also established that the quadratic form contains all the information that is required in order to identify the SIRV PDF. As a consequence of this result, the problem of an SIRV (multivariate) distribution identification has been reduced to the equivalent identification of the univariate distribution of the non-negative quadratic form.

Chapter 5

Computer Generation of Simulated Radar Clutter Characterized as SIRPs

5.1 Introduction

This investigation is motivated by a desire to simulate correlated non-Gaussian radar clutter. Various investigators have reported experimental results where non-Gaussian marginal probability density functions (PDF) have been used to model the clutter. Usually, radars process N samples at a time. Statistical characterization of the clutter requires the specification of the joint PDF of the N samples. In addition, the clutter may be highly correlated. Hence, the joint PDF must take into account the correlation between samples. Statistical characterization of the clutter is necessary if an optimal radar signal processor is to be obtained. For use of the well known likelihood ratio test, it is necessary to have closed form expressions for the joint PDF of the N clutter samples in order to obtain the optimal radar signal processor. In most cases, it is difficult to evaluate the performance of the optimal radar signal processor analytically when the clutter samples are correlated and non-Gaussian. Then computer simulation may be necessary. Therefore, there is a need to develop efficient procedures that facilitate computer simulation of the clutter. A library of multivariate non-Gaussian PDFs has been developed in Chapter 4, using the theory of **Spherically Invariant Random Processes (SIRP)** and **Spherically Invariant Random Vectors (SIRV)**. In view of the large number of parameters that are free to be specified, the library of multivariate non-Gaussian PDFs can be used to approximate many different radar clutter scenarios. In this chapter we concern ourselves with the development of computer simulation procedures for the library of non-Gaussian PDFs obtained in Chapter 4

so that the performance of any radar signal processor can be evaluated for a variety of different clutter scenarios. Another issue addressed in this chapter is performance assessment of the simulation procedures. It has been pointed out in Chapter 4 that the quadratic form appearing in the PDF of the SIRV contains all the information necessary to identify the PDF of the underlying SIRV. We make use of this result in order to assess the performance of the simulation procedures. Some interesting simulation techniques have been proposed for SIRVs in [31] and [33]. The technique suggested in [31] makes use of Meijer's-G functions. These functions are generalizations of Hypergeometric functions which do not lend themselves to the development of simple and elegant simulation procedures. The technique suggested in [33] requires transformations from rectangular to spherical co-ordinates and then back again. Secondly, this simulation procedure involves the use of the inverse distribution function approach for a rather complicated distribution function. The approach developed in this chapter is simpler to implement than those proposed in [31] and [33]. In addition, a new approach is proposed for assessing the effectiveness of the simulation procedure.

In Section 5.2, we review some definitions and background information pertaining to the theory of spherically invariant random processes. Section 5.3 presents two canonical simulation procedures for generating SIRVs. Performance assessment of the simulation procedures is discussed in Section 5.4. Finally, conclusions are presented in Section 5.5.

5.2 Preliminaries

We begin by restating the definitions for spherically invariant random vector and spherically invariant random processes. A spherically invariant random vector (SIRV) is a random vector (real or complex) whose PDF is uniquely determined by the specification of a mean vector, a covariance matrix and a characteristic first order PDF. Equivalently, the PDF of an SIRV can also be referred to as an elliptically contoured distribution. A spherically invariant random process (SIRP) is a random process (real or complex) such that every random vector obtained by sampling this process is an SIRV. The work of Yao [28] gave rise to a representation theorem which can be stated as follows (see Theorem 1):

If a random vector is a SIRV, then there exists a non-negative random variable S such that the PDF of the random vector conditioned on S is a multivariate Gaussian PDF.

We consider the product given by $\mathbf{X} = \mathbf{Z}S$, where $\mathbf{X} = [X_1 \dots X_N]^T$ denotes the SIRV, $\mathbf{Z} = [Z_1 \dots Z_N]^T$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{M}$ and

S is a non-negative random variable with PDF $f_S(s)$. Since it is desirable to independently control the correlation properties and the non-Gaussian envelope PDF, Z and S are assumed to be statistically independent. The PDF of $\mathbf{X}$ conditioned on S is (see eq (3.14))

$$f_{\mathbf{X}|S}(\mathbf{x}|s) = (2\pi)^{-\frac{N}{2}} |\mathbf{M}|^{-\frac{1}{2}} s^{-N} \exp\left(-\frac{p}{2s^2}\right) \quad (5.1)$$

where p is a non-negative quadratic form given by $p = \mathbf{x}^T \mathbf{M}^{-1} \mathbf{x}$ and $|\mathbf{M}|$ denotes the determinant of the covariance matrix $\mathbf{M}$. The PDF of $\mathbf{X}$ is given by (see eqs (3.15) and (3.16))

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} |\mathbf{M}|^{-\frac{1}{2}} h_N(p) \quad (5.2)$$

where

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds. \quad (5.3)$$

The PDF of the random variable S is called the characteristic PDF of the SIRV. Therefore, it is apparent that the PDF of a SIRV is completely determined by the specification of a mean vector, a covariance matrix and a characteristic first order PDF. In addition, the PDF of the SIRV is a function of a non-negative quadratic form. However, unlike the Gaussian case, dependence on the quadratic form is more complicated than the simple exponential. Therefore, an SIRP can be regarded as a generalization of the familiar Gaussian random process. We point out that the covariance matrix of the SIRV is given by $\Sigma = \mathbf{M}E(S^2)$ where $E(S^2)$ is the mean square value of the random variable S . It is seen that the covariance matrix of the SIRV normalized by the mean square value of S is the covariance matrix of the Gaussian random vector. Note that it is possible to set the covariance matrix of the SIRV equal to that of the Gaussian random vector by requiring that $E(S^2)$ be equal to unity. The desired non-Gaussian PDF can be obtained by choosing $f_S(s)$ appropriately. Thus, it is seen that the SIRV formulation for radar clutter modeling affords independent control over the non-Gaussian PDF of the clutter and its correlation properties. Several techniques are available in Chapter 4 for obtaining $h_N(p)$. Note that the Gaussian random vector is a special case of an SIRV and is obtained when $f_S(s) = \delta(s - 1)$ where $\delta(t)$ is the unit impulse function. An interesting interpretation of the representation theorem is that every SIRV is the modulation of a Gaussian random vector by a non-negative random variable.

Many of the attractive properties of Gaussian random vectors also apply to SIRVs. The most relevant property of SIRVs for the purpose of computer simulation is the closure property under

linear transformation [28] stated below (see Theorem 2):

If $\mathbf{X}$ is an SIRV with characteristic PDF $f_S(s)$, then

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} \quad (5.4)$$

is also an SIRV with the same characteristic PDF. It is assumed that $\mathbf{A}\mathbf{A}^T$ is a nonsingular matrix and $\mathbf{b}$ is a known vector having the same dimension as $\mathbf{X}$.

Theorem 2 provides us with a powerful technique for simulating SIRVs. A white SIRV is defined as one that has a diagonal covariance matrix. In other words, the components of the white SIRV are uncorrelated but not necessarily independent. We can start with a zero mean white SIRV $\mathbf{X}$ having identity covariance matrix and perform the linear transformation given by eq (5.4) to obtain an SIRV $\mathbf{Y}$ having a non-zero mean and desired covariance matrix Σ . The matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are given by

$$\begin{aligned} \mathbf{A} &= \mathbf{E}\mathbf{D}^{\frac{1}{2}} \\ \mathbf{b} &= \mu_{\mathbf{Y}} \end{aligned} \quad (5.5)$$

where $\mathbf{E}$ is the matrix of normalized eigen vectors of the covariance matrix Σ , $\mathbf{D}$ is the diagonal matrix of eigen values of Σ and $\mu_{\mathbf{Y}}$ is the desired non-zero mean vector.

In many instances it is not possible to obtain $f_S(s)$ for an SIRV in closed form, even though its existence is guaranteed. In such cases, an alternate approach must be used in order to characterize the SIRV. The following theorem can be used to completely characterize a white SIRV having zero mean and identity covariance matrix (see Theorem 3):

A random vector $\mathbf{X} = [X_1 \dots X_N]^T$ is a zero mean white SIRV having identity covariance matrix if and only if there exist random variables $R \in (0, \infty)$, $\Theta \in (0, 2\pi)$ and $\Phi_k \in (0, \pi)$, ($k = 1, \dots, N - 2$) such that when the components of $\mathbf{X}$ are expressed in the generalized spherical

coordinates

$$\begin{aligned}
 X_1 &= R \cos(\Phi_1) \\
 X_k &= R \cos(\Phi_k) \prod_{i=1}^{k-1} \sin(\Phi_i) \quad (1 < k \leq N-2) \\
 X_{N-1} &= R \cos(\Theta) \prod_{i=1}^{N-2} \sin(\Phi_i) \\
 X_N &= R \sin(\Theta) \prod_{i=1}^{N-2} \sin(\Phi_i)
 \end{aligned} \tag{5.6}$$

then the random variables R , Θ and Φ_k are mutually and statistically independent having PDFs of the form

$$\begin{aligned}
 f_R(r) &= \frac{r^{N-1}}{2^{\frac{N}{2}-1} \Gamma(\frac{N}{2})} h_N(r^2) u(r) \\
 f_{\Phi_k}(\phi_k) &= \frac{\Gamma(\frac{N-k+1}{2})}{\sqrt{\pi} \Gamma(\frac{N-k}{2})} \sin^{N-1-k}(\phi_k) [u(\phi_k) - u(\phi_k - \pi)] \\
 f_{\Theta}(\theta) &= (2\pi)^{-1} [u(\theta) - u(\theta - 2\pi)]
 \end{aligned} \tag{5.7}$$

where $\Gamma(\nu)$ is the Euler Gamma function and $u(t)$ is the unit step function.

As a consequence of Theorem 3, any SIRV with zero mean and identity covariance matrix can be represented in generalized spherical coordinates which are mutually and statistically independent regardless of the SIRV considered. Also, note that the PDFs of Θ and Φ_k , ($k = 1, \dots, N-2$) are functionally independent of the white SIRV considered. Only the PDF of R changes from one white SIRV to another. Note that $R^2 = \sum_{k=1}^N X_k^2 = \mathbf{X}^T \mathbf{X}$. Hence R is the norm of the SIRV.

Another important feature of the SIRV is that the quadratic form appearing in its PDF contains all the information necessary to identify the PDF. It follows that knowledge of the PDF of the quadratic form of the SIRV is sufficient to identify the PDF of the corresponding SIRV [34] (see Theorem 4):

The PDF of the quadratic form appearing in eq (5.2) is given by

$$f_P(p) = \frac{1}{2^{\frac{N}{2}-1} \Gamma(\frac{N}{2})} p^{\frac{N}{2}-1} h_N(p) \quad (0 \leq p < \infty) \tag{5.8}$$

and remains unchanged regardless of whether or not the SIRV is white.

The theorems reviewed in this section will be made use of in the proposed simulation approach, discussed in Section 5.3, and in assessing the performance of the simulation procedure, discussed in Section 5.4.

In the context of the problem of radar clutter modeling and simulation, the bandpass process $Y(t) = \text{Re}[\tilde{Y}(t)\exp(j\omega_0 t)]$ can be expressed in terms of the equivalent complex, wide sense stationary random processes $\tilde{Y}(t)$. More precisely, we obtain N complex samples by sampling the complex random process $\tilde{Y}(t) = Y_c(t) + jY_s(t)$, where the subscripts c and s denote the in phase and out of phase quadrature components. This is equivalent to working with a real vector of $2N$ quadrature components which is the approach taken in this chapter. Therefore, the results presented in this section are applied to the problem of radar clutter modeling with N replaced by $2N$. For ease of reference, the library of non-Gaussian SIRV PDFs obtained in Chapter 4 is repeated here. However, $h_{2N}(p)$ for those SIRVs for which the characteristic PDF is known are listed in Table 5.1. The corresponding characteristic PDFs are listed in Table 5.2. Table 5.3 lists $h_{2N}(p)$ for those SIRVs whose characteristic PDF is unknown.

5.3 Two Canonical Simulation Procedures for Generating SIRVs

In this section, we concern ourselves with two simulation procedures for generating the SIRVs listed in Table 5.1 and Table 5.2. The first simulation procedure to be discussed is applicable when the characteristic PDF, $f_S(s)$, is known. For each of the PDFs listed in Table 5.1, the characteristic PDF $f_S(s)$ is tabulated in Table 5.3, where $E(S^2) = 1$. Since the representation theorem results in the covariance matrix of the SIRV being given by $\Sigma = ME(S^2)$, the choice of $E(S^2) = 1$ makes Σ identical to M , the covariance matrix of the Gaussian random vector Z . However, as listed in Table 5.4, the PDFs commonly encountered in statistical tables do not have unit mean square value. In order to obtain the random variable S , having unit mean square value and the corresponding PDF $f_S(s)$, we generate the random variable V having PDF $f_V(v)$ and mean square value $E(V^2) = a^2$, and perform the linear transformation $S = \frac{V}{a}$ to obtain the desired S . In Table 5.1, and Table 5.4, the scale parameter b , as well as the shape parameter ν are identical in both cases and $u(v)$ denotes the unit step function. The simulation procedure for these SIRV PDFs is fairly simple and is stated below:

5.3.1 Simulation Procedure for SIRVs with Known Characteristic PDF

- (1) Generate a white zero mean Gaussian random vector Z , having identity covariance matrix.

- (2) Then generate a random variable V from the PDF $f_V(v)$. Denote the mean square value of V by a^2 .
- (3) Normalize the random variable V by a to obtain the modulating random variable S . In other words generate $S = \frac{V}{a}$.
- (4) Generate the product given by $\mathbf{X} = \mathbf{Z}S$. At this step, we have a white SIRV having zero mean and identity covariance matrix.
- (5) Finally, perform the linear transformation given by eq (5.5) to obtain the SIRV $\mathbf{Y}$ with desired mean and covariance matrix.

Fig 5.1 shows the simulation procedure presented above.

The subroutine RNNOR in IMSL was used for generating the Gaussian random vector $\mathbf{Z}$. Interestingly enough, the PDFs listed in Table 4.4 can be related to the PDF of the Gamma distribution as discussed below. The PDF $f_V(v)$ for the K-distributed SIRV is a Chi PDF. We first address the random variable generation for the Chi PDF and then provide the transformations for obtaining the random variables for the other PDFs listed in Table 4.4.

Consider the standard Gamma distribution given by

$$f_T(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)} \exp(-t) \quad t > 0 \quad (5.9)$$

where α denotes the shape parameter and $\Gamma(\alpha)$ is the Euler-Gamma function. The random variable T is readily generated by using the IMSL subroutine RNGAM. The procedure for generating the Chi distributed random variable V needed for the K-distributed SIRV is summarized below.

1. Generate the random variable T from the standard Gamma distribution of eq (5.9) by using the IMSL subroutine RNGAM.
2. Perform the transformation $V = \sqrt{2T}$.

The PDF $f_V(v)$ for the Laplace SIRV is a Rayleigh PDF and is obtained from $f_V(v)$ of the K-distributed SIRV by letting $\alpha = 1$. The random variable V for the PDF $f_v(v)$ listed in Table 4.4 for the Student-t SIRV is obtained from the standard Gamma PDF of eq (5.9) by the transformation $V = \frac{b}{\sqrt{2T}}$ and letting $\alpha = \nu$. Finally, the PDF $f_V(v)$ for the Cauchy SIRV is

Table 5.1: $h_{2N}(p)$ for SIRVs with Known Characteristic PDF

Marginal PDF	$h_{2N}(p)$
Laplace	$b^{2N} (b\sqrt{p})^{1-N} K_{N-1}(b\sqrt{p})$
Cauchy	$\frac{2^N \Gamma(\frac{1}{2} + N)}{\sqrt{\pi} (b^2 + p)^{N + \frac{1}{2}}}$
K-distribution	$\frac{b^{2N} (b\sqrt{p})^{N-1}}{\Gamma(\alpha) 2^{\alpha-1}} K_{N-\alpha}(b\sqrt{p})$
Student-t	$\frac{2^N b^{2\nu} \Gamma(\nu + N)}{\Gamma(\nu) (b^2 + p)^{N+\nu}}$

Table 5.2: $h_{2N}(p)$ for SIRVs with Unknown Characteristic PDFs

Marginal PDF	$h_{2N}(p)$
Chi	$(-2)^{N-1} A \sum_{k=1}^N G_k p^{\nu-k} \exp(-Bp)$
	$G_k = \binom{N-1}{k-1} (-1)^{k-1} B^{k-1} \frac{\Gamma(\nu)}{\Gamma(\nu-k+1)}$
	$A = \frac{2}{\Gamma(\nu)} (b\sigma)^{2\nu}$
	$B = b^2 \sigma^2$
	$\nu \leq 1$
Weibull	$\sum_{k=1}^N C_k p^{\frac{k}{2}-N} \exp(-Ap^{\frac{1}{2}})$
	$A = a\sigma^b$
	$C_k = \sum_{m=1}^k (-1)^{m+N} 2^N \frac{A^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{2})}{\Gamma(1+\frac{m}{2}-N)}$
	$b \leq 2$
Gen. Rayleigh	$\sum_{k=1}^{N-1} D_k p^{\frac{k}{2}-N+1} \exp(-Bp^{\frac{1}{2}})$
	$A = \frac{\sigma^2 \alpha}{\beta^2 \Gamma(\frac{\alpha}{2})}$
	$B = \beta^{-\alpha} \sigma^\alpha$
	$D_k = \sum_{m=1}^k (-1)^{m+N-1} 2^{N-1} \frac{B^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{2})}{\Gamma(2+\frac{m}{2}-N)}$
	$\alpha \leq 2$
Rician	$\frac{\sigma^{2N}}{(1-\rho^2)^{N-\frac{1}{2}}} \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \left(\frac{\rho}{2}\right)^k \xi_k \exp(-A)$
	$\xi_k = \sum_{m=0}^k \binom{k}{m} I_{k-2m}(\rho A), A = \frac{\rho \sigma^2}{2(1-\rho^2)}$

obtained from $f_V(v)$ of the Student-t SIRV by letting $\nu = 1$. The procedure for generating the random variable V needed for the Student-t SIRV is summarized below.

1. Generate the random variable T from the standard Gamma distribution of eq (5.9) by using the IMSL subroutine RNGAM.
2. Perform the transformation $V = \frac{b}{\sqrt{2T}}$.

5.3.2 Simulation Scheme for SIRVs with Unknown Characteristic PDF

We now concern ourselves with the second simulation procedure which is applicable when the characteristic PDF is unknown, as is the case for SIRVs listed in Table 4.2. The alternate approach makes use of Theorem 3 and the representation theorem. As pointed out previously,

Table 5.3: Characteristic PDF for SIRVs listed in Table 4.1 [$E(S^2) = 1$]

Marginal PDF	$f_S(s)$
Laplace	$ab^2 s \exp(-\frac{a^2 b^2 s^2}{2}) u(s)$
Cauchy	$a^2 b^2 s^{-3} \exp(-\frac{b^2}{2as^2}) u(s)$
K-distribution	$\frac{2ab}{\Gamma(\alpha)2^\alpha} (bas)^{2\alpha-1} \exp(-\frac{b^2 a^2 s^2}{2}) u(s)$
Student-t	$\frac{2ab}{\Gamma(\nu)2^\nu} b^{2\nu-1} (as)^{-(2\nu+1)} \exp(-\frac{b^2}{2as^2}) u(s)$

Table 5.4: Related PDF $f_V(v)$

Marginal PDF	$f_V(v)$	$a^2 = E(V^2)$
Laplace	$b^2 v \exp(-\frac{b^2 v^2}{2}) u(v)$	$\frac{2}{b^2}$
Cauchy	$b^2 v^{-3} \exp(-\frac{b^2}{2v^2}) u(v)$	$\frac{b^2}{2}$
K-distribution	$\frac{2b}{\Gamma(\alpha)2^\alpha} (bv)^{2\alpha-1} \exp(-\frac{b^2 v^2}{2}) u(v)$	$\frac{2\alpha}{b^2}$
Student-t	$\frac{2b}{\Gamma(\nu)2^\nu} b^{2\nu-1} v^{-(2\nu+1)} \exp(-\frac{b^2}{2v^2}) u(v)$	$\frac{b^2}{2(\nu-1)}$

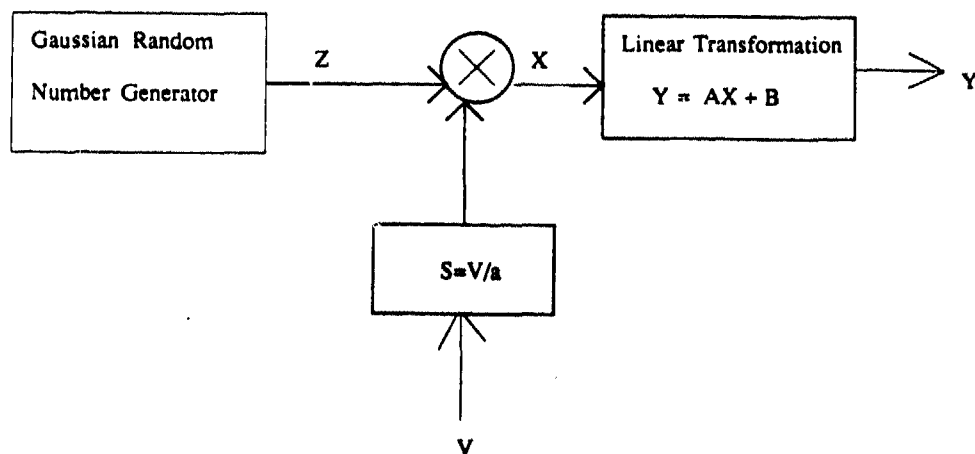


Figure 5.1: Simulation Scheme for SIRVs with Known Characteristic PDF

the PDFs of Θ and Φ_k ($k = 1, 2, \dots, N - 2$) are independent of the white SIRV being considered. Only the PDF of R changes from one white SIRV to another. As a result, the second simulation procedure requires the capability to generate the random variable R whose PDF is given by eq (5.7). Since the Gaussian random vector belongs to the family of SIRVs, a zero mean white Gaussian random vector $\mathbf{Z}$ with identity covariance matrix admits a representation of the form of eq(5.6). Let R_G denote the norm of the white Gaussian random vector. The simulation procedure is stated below:

- (1) Generate a white, zero mean Gaussian random vector $\mathbf{Z}$ having identity covariance matrix.
- (2) Compute the norm $R_G = \|\mathbf{Z}\| = \sqrt{\mathbf{Z}^T \mathbf{Z}}$ of the white Gaussian random vector.
- (3) Generate the norm $R = \|\mathbf{X}\| = \sqrt{\mathbf{X}^T \mathbf{X}}$ of the white SIRV from the PDF of R given by eq (5.7).
- (4) Generate the white SIRV $\mathbf{X}$ by taking the product $\mathbf{X} = \mathbf{Z} \frac{R}{R_G}$.
- (5) Finally, perform the linear transformation given by eq (5.5) to obtain the SIRV $\mathbf{Y}$ with desired mean and covariance matrix.

The simulation procedure is shown schematically in Fig 5.2.

Note that this simulation procedure avoids the explicit generation of the variables Θ and Φ_k ($k = 1, \dots, N - 2$). The generation procedure for a white Gaussian random vector is well known. Therefore, we need to concern ourselves only with the development of a suitable generation scheme for the norm R of the white SIRV $\mathbf{X}$. Generation of the norm R is not trivial. This is due to the fact that the PDF of R is usually not in a simple functional form. Consequently, it may not be possible to conveniently evaluate analytically the distribution function and its inverse. As a result, generation methods based on the inverse distribution function do not offer a practical solution to this problem. Therefore, in this chapter we generate R by making use of the approach called the 'Rejection Method'. The rejection method can be used to generate random variables whose cumulative distribution functions are not known, but whose PDFs are known explicitly [49]. The rejection procedure assumes knowledge of the maximum value of the PDF of R for a given SIRV PDF and a finite estimate to the range of the PDF of R so that the area

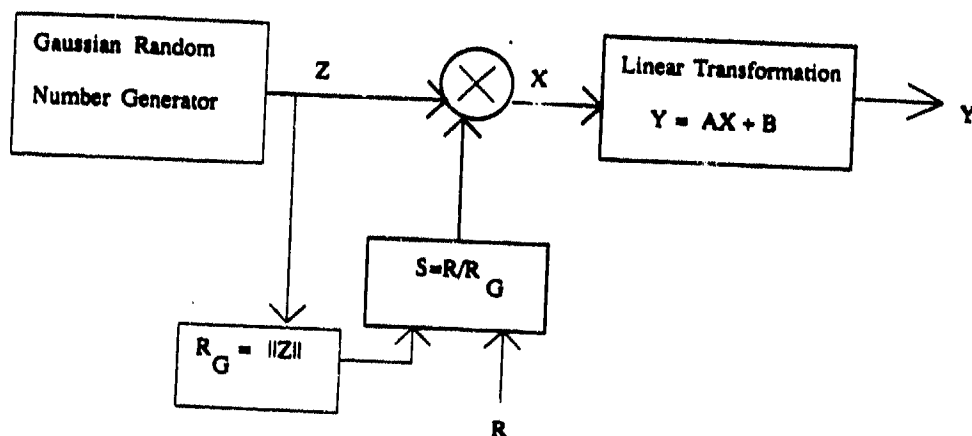


Figure 5.2: Simulation Scheme for SIRVs with Unknown Characteristic PDF

under the PDF curve is close to unity. These quantities are denoted by c and b , respectively. We discuss the rejection procedure in detail in Appendix B. The Rejection method is summarized below:

- (1) Generate a uniform random variate U_1 on the interval $(0, b)$.
- (2) Generate another uniform variate U_2 on the interval $(0, c)$.
- (3) If $U_2 \leq f_R(U_1)$, then $R = U_1$. Otherwise, reject U_1 and return to step 1.

Note that the simulation procedures of Fig 5.1 and Fig 5.2 are canonical in the sense that their forms remain unchanged from the simulation of one SIRV to another. Even though, the scheme of Fig 5.2 can be used even when $f_S(s)$ is known, the scheme of Fig 5.1 is preferred when S can be generated easily. The linear transformation of eq (5.5) is a filtering operation. In both schemes, pre-modulation filtering is equivalent to post-modulation filtering. This results from the fact that the representation theorem is valid whether or not the SIRV X and the Gaussian random vector Z are white.

5.4 Performance Assessment of the Simulation Schemes

In this section we concern ourselves with the performance assessment of the simulation procedures developed in section 5.3. We point out that the simulation procedures developed in section 5.3 are exact in the sense that they are derived without approximation from theory. Hence, departures from the exact SIRVs will depend for the most part on the nonideality of the uniform random number generators used. Empirical assessment of the simulation procedures is necessary for practical applications.

One possible approach for assessing the distributional properties of the simulated data is to perform a hypothesis test on the marginal distributions of the components of the SIRV where the hypothesis are given by

H_0 : The hypothesis that the simulated data is from the desired distribution

H_1 : The hypothesis that the simulated data is not from the desired distribution.

For a fixed Type-1 error probability (i.e., the probability that H_1 is accepted given that H_0 is true) each marginal distribution can be checked by employing one of the commonly used goodness of fit procedures. Since the components of the random vectors are not statistically independent, we are now confronted with the problem of developing a goodness of fit test for the multivariate data. In general, it is very difficult to obtain the overall significance level of the test (i.e., the probability that H_0 is accepted given that H_0 is true) for the multivariate goodness of fit testing procedure.

However, an attractive feature of SIRVs is that the quadratic form p appearing in the SIRV PDF contains all the information necessary for identifying the PDF of the SIRV. In other words, knowledge of the PDF of the quadratic form is sufficient to determine the underlying SIRV PDF. Furthermore, the quadratic form PDF remains unchanged regardless of whether the SIRV is white or colored. The PDF of the quadratic form appearing in the SIRV PDF is given by eq (5.8). For the radar problem where we deal with N complex samples or $2N$ quadrature components, note that we make use of eq (5.8) with N replaced by $2N$. Hence, we base our goodness of fit test procedure for the generated SIRVs on the PDF of the quadratic form p . Note that we have now reduced the multivariate problem to an equivalent univariate problem involving the goodness of fit test for the PDF of the quadratic form.

In the examples presented in this section, we generated $m = 1000$ realizations of the random vector $\mathbf{Y}$ with $N = 2$ complex samples and obtained one thousand samples of the quadratic form

P for each of the non-Gaussian SIRVs whose PDFs are listed in Tables 5.1 and 5.3. In each case, we used the corresponding theoretical PDF of the quadratic form given by eq (5.8) to test for the distribution of the generated quadratic form. The frequency histograms for the generated data and the corresponding theoretical PDFs are shown in figures 5.3-5.10. In addition, a Chi-Square test was performed on the generated data with the Type-1 error fixed at 0.05 and the null hypothesis was not rejected in each case. The histograms provide a good idea about the true distributions for large sample sizes. Observe that the empirical PDFs are very close to the theoretical PDFs. Note that the procedure used in this section to assess the distributional assumptions of the random samples from the SIRV PDFs is a formal goodness of fit test. Similar procedures have been proposed to test for multivariate normality in [50] and [51].

5.5 Conclusions

In this Chapter, we have presented two schemes that can be used in practice to simulate correlated non-gaussian radar clutter when the clutter can be modeled as a spherically invariant random process. We pointed out that the simulation schemes developed are canonical schemes and do not change form from the simulation of one SIRV to another. A new approach, based on the PDF of the quadratic form appearing in the SIRV PDF, was used to perform a goodness of fit test in order to assess performance of the proposed simulation schemes. Performance assessment based on this scheme showed excellent agreement between the theoretical and empirical PDFs of the quadratic form. Finally, it was pointed out that use of this technique reduced the goodness of fit test from a multivariate testing procedure to a univariate testing procedure resulting in tremendous processing simplicity. Therefore, this procedure lends itself very well to practical applications.

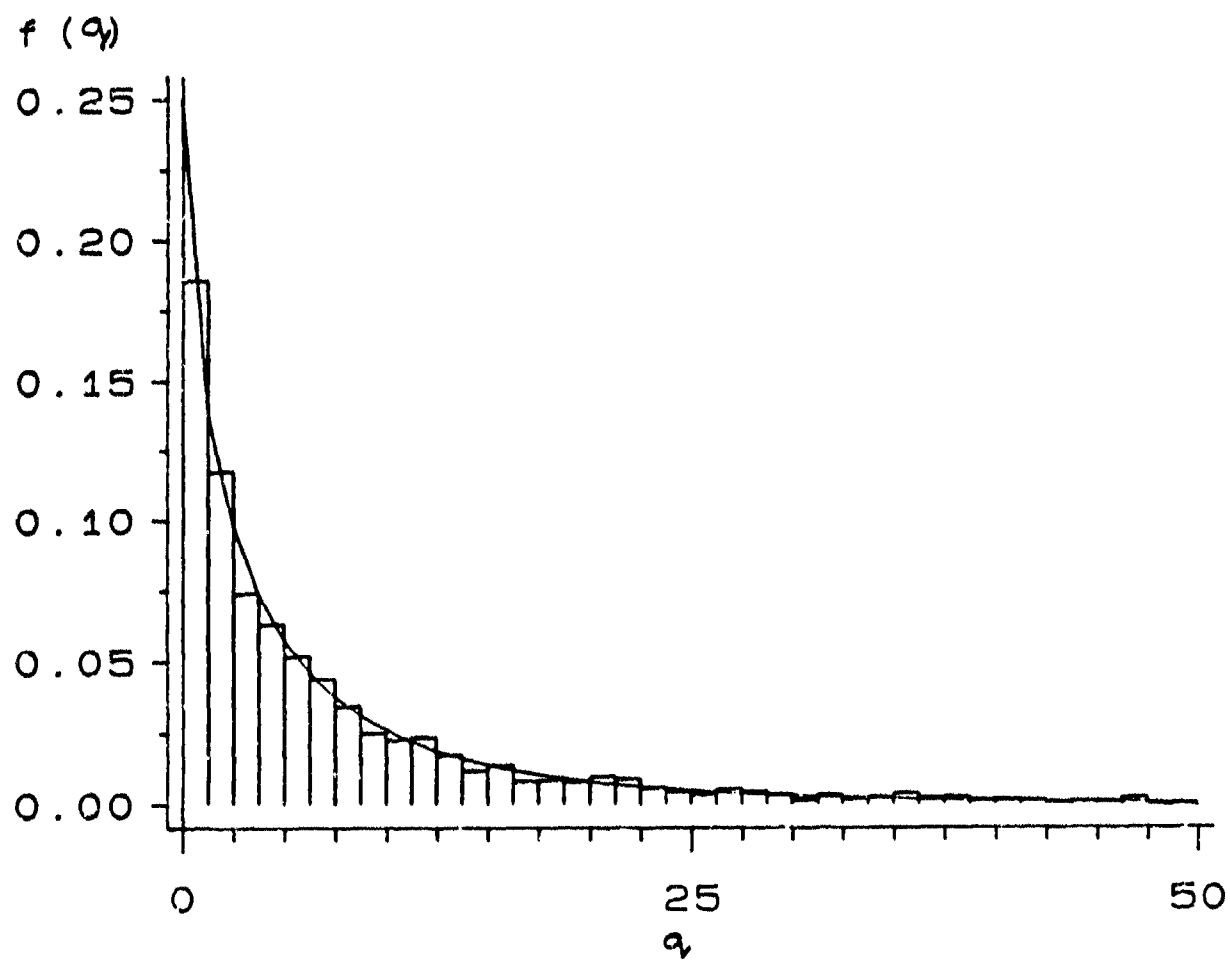


Figure 5.3: Theoretical and Empirical Quadratic form PDFs for Laplace SIRV

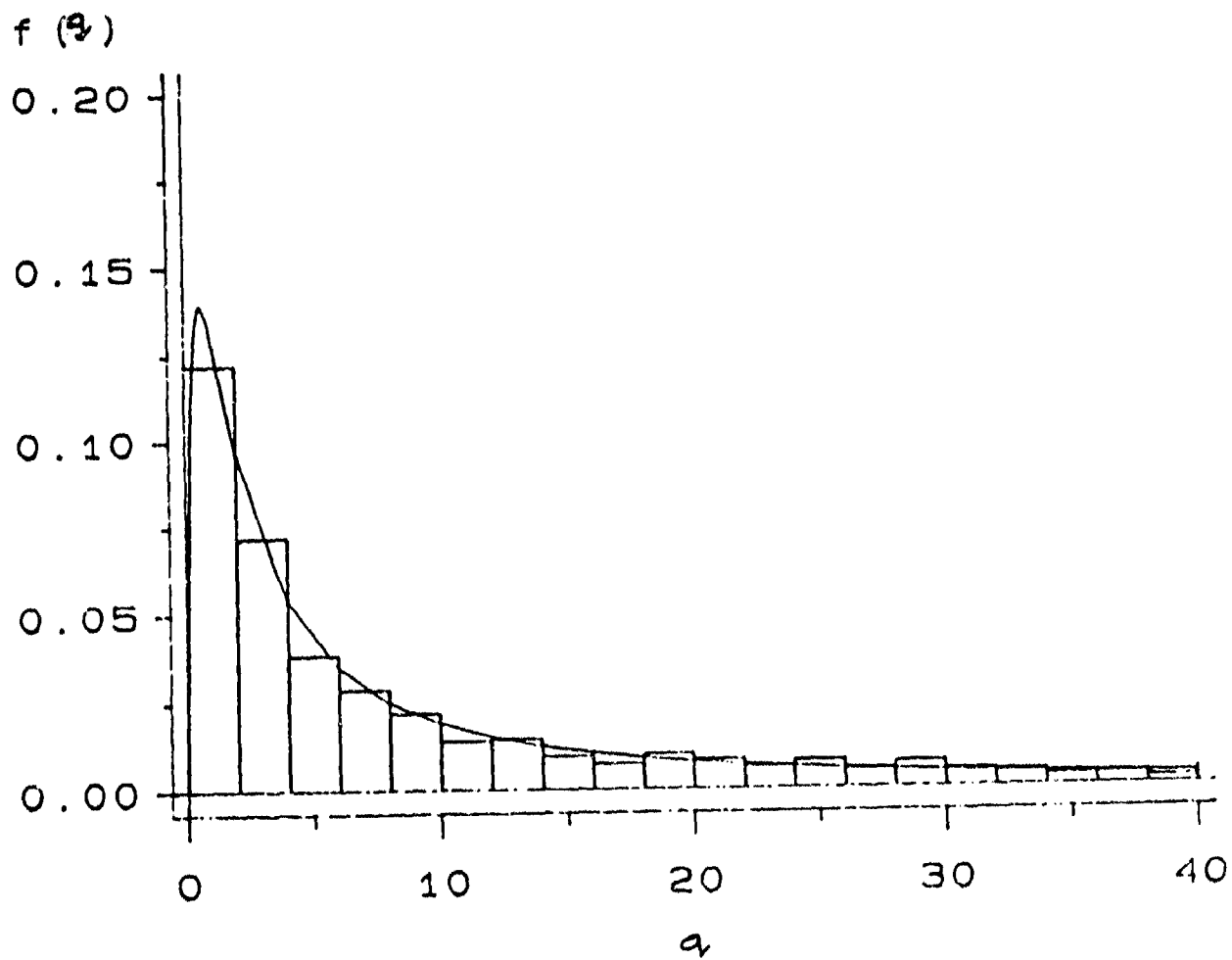


Figure 5.4: Theoretical and Empirical Quadratic form PDFs for Cauchy SIRV

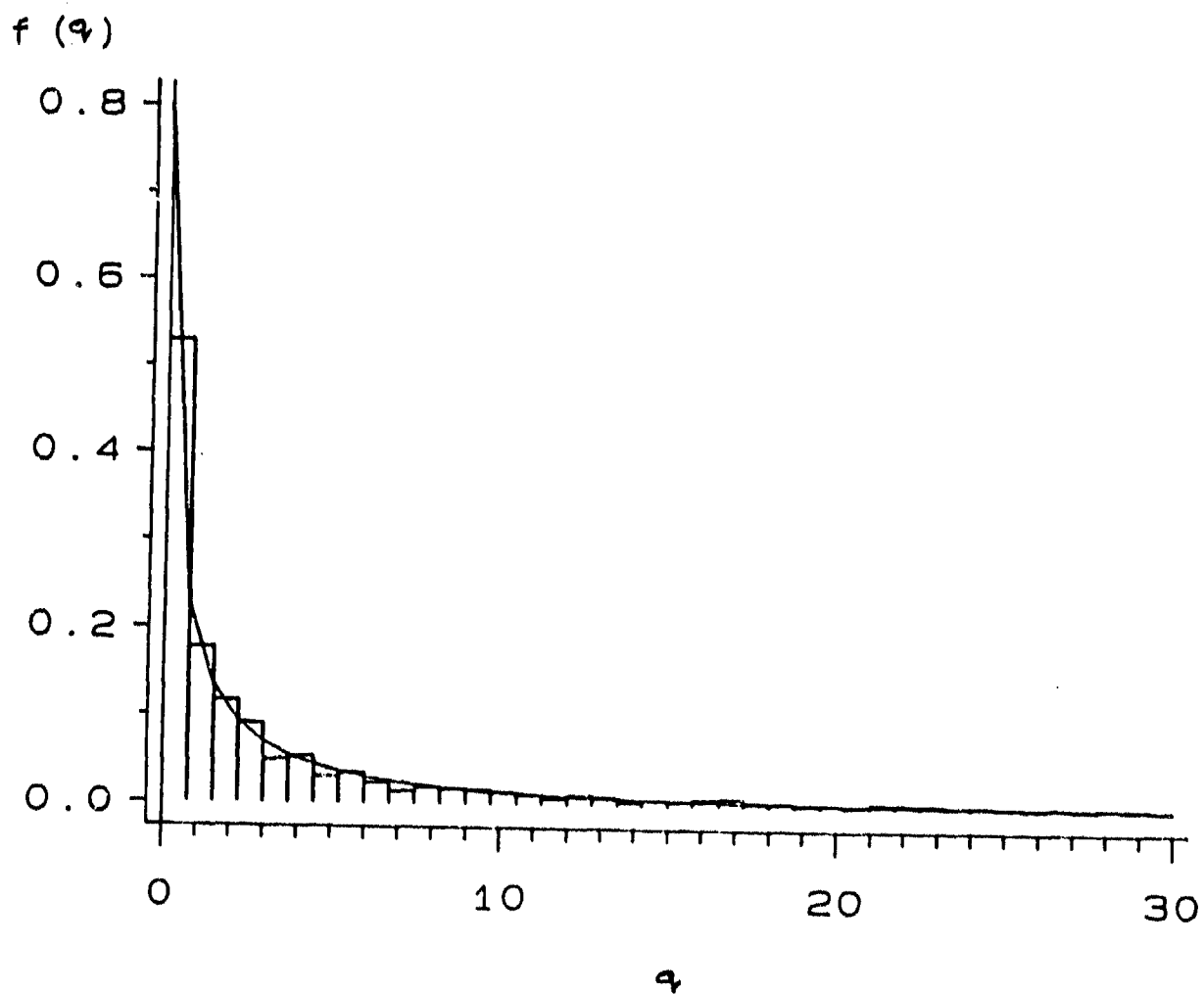


Figure 5.5: Theoretical and Empirical Quadratic form PDFs for K-distributed SIRV

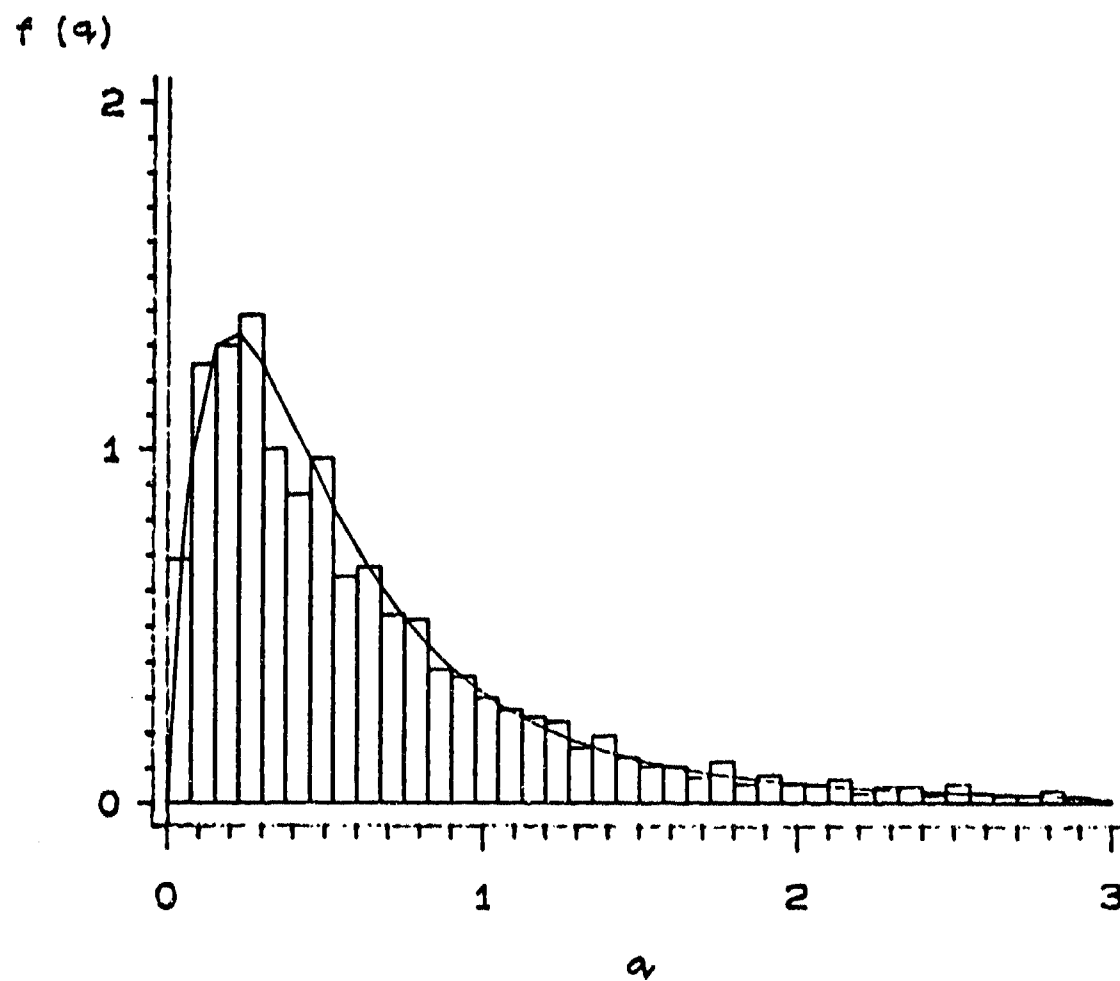


Figure 5.6: Theoretical and Empirical Quadratic form PDFs for Student-t SIRV

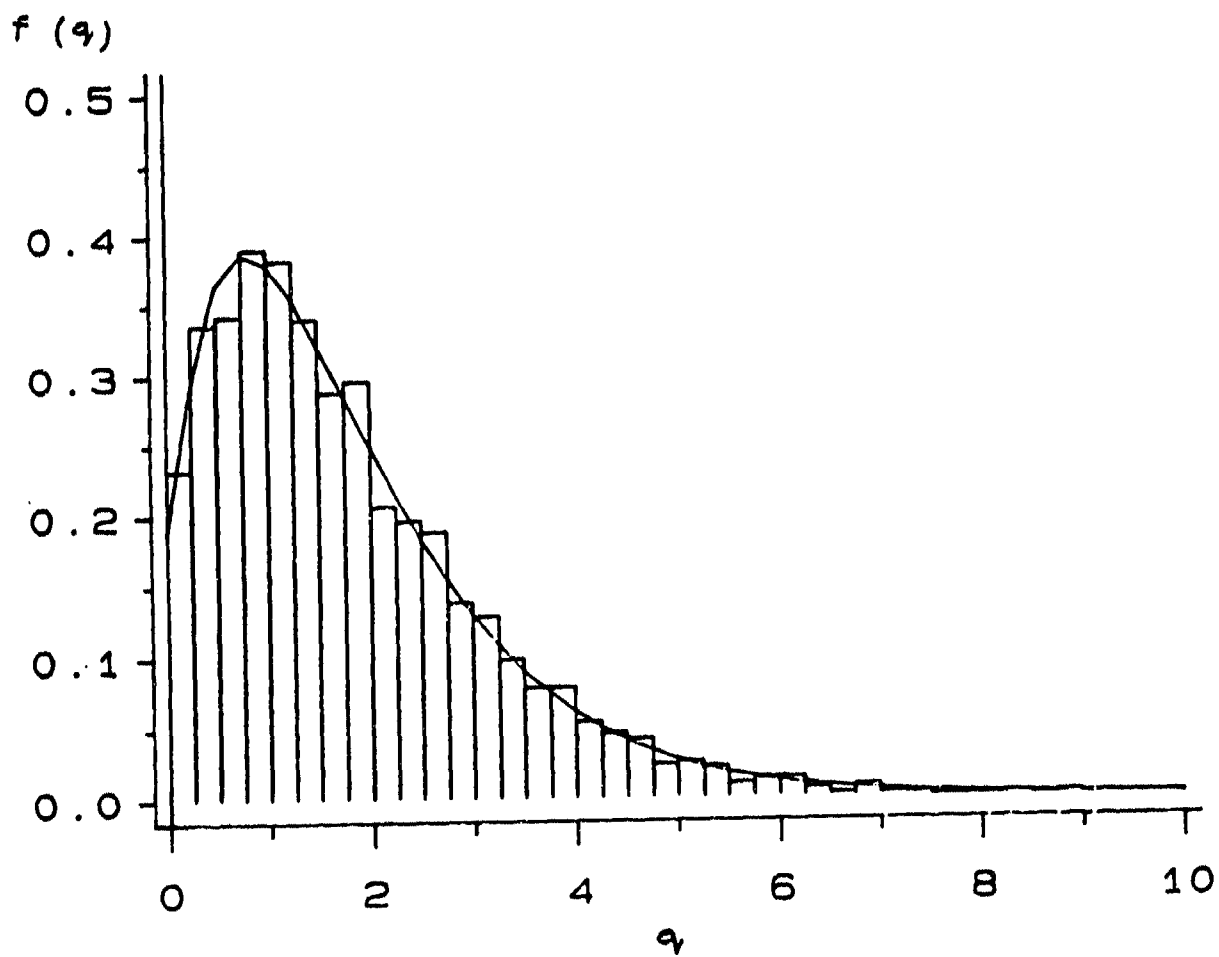


Figure 5.7: Theoretical and Empirical Quadratic form PDFs for Chi distributed SIRV

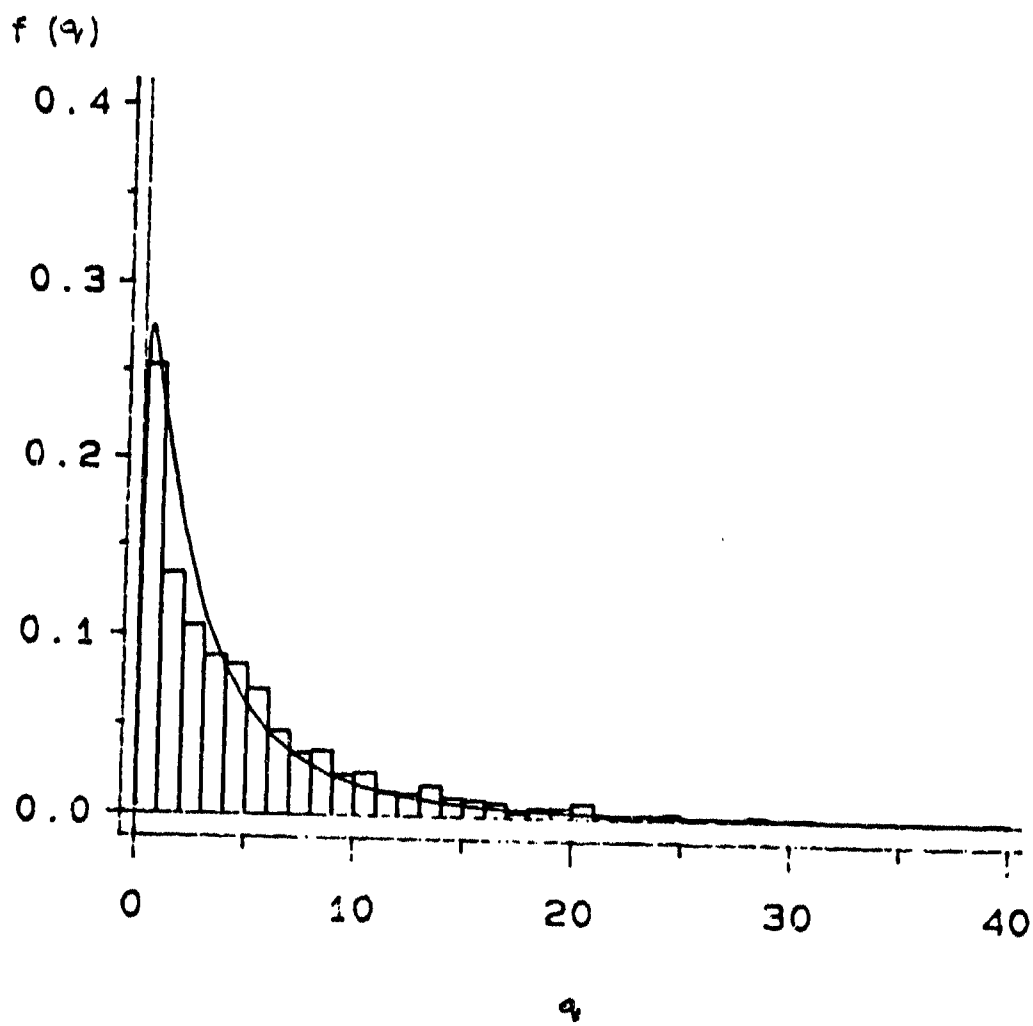


Figure 5.8: Theoretical and Empirical Quadratic form PDFs for Generalized Rayleigh SIRV

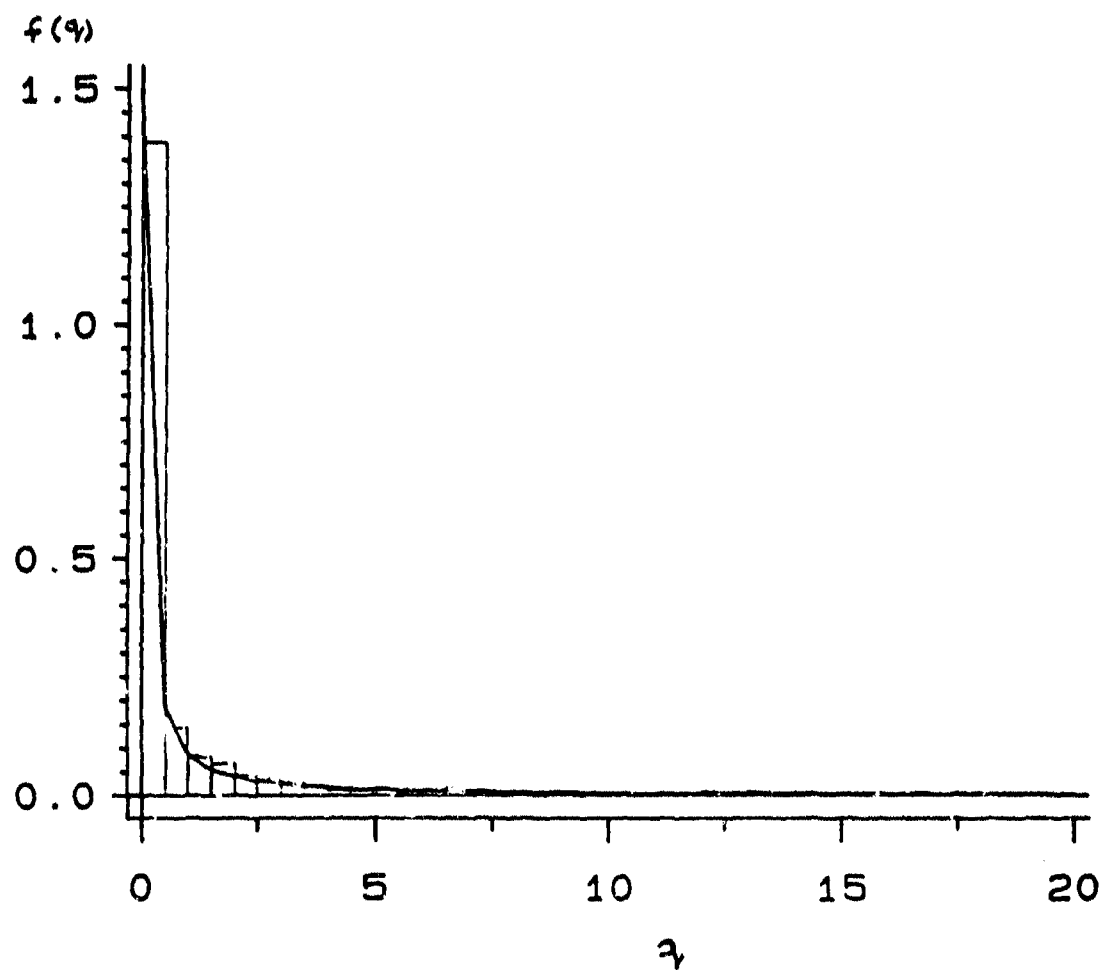


Figure 5.9: Theoretical and Empirical Quadratic form PDFs for Weibull SIRV

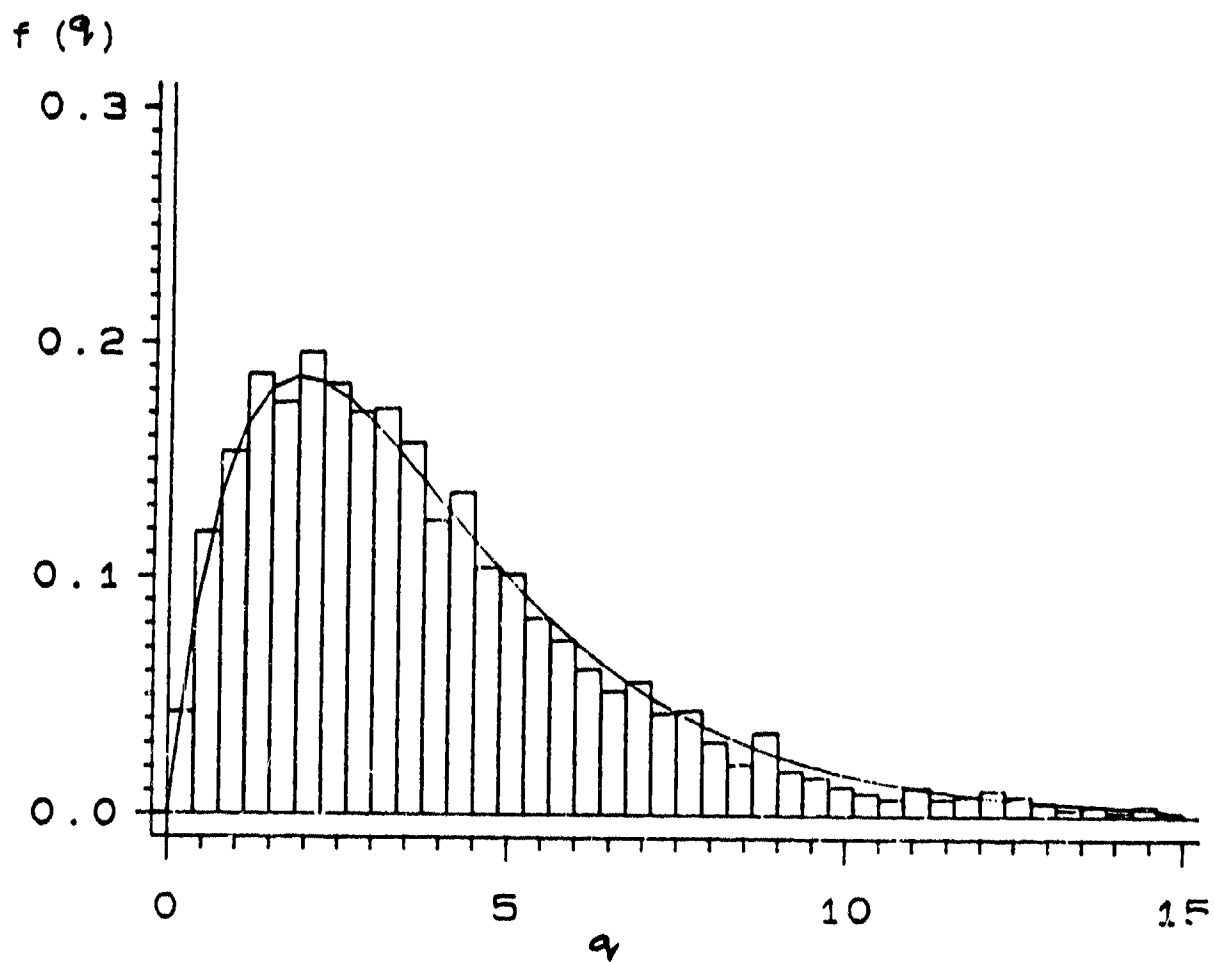


Figure 5.10: Theoretical and Empirical Quadratic form PDFs for Rician SIRV

Chapter 6

A New Method for Univariate Distribution Approximation

6.1 Introduction

In this chapter we address the problem of approximating the PDF of a set of random data. In practice, the clutter PDF encountered in radar signal processing is not known apriori. Consequently, a scheme that approximates the clutter PDF based on a set of measured data is necessary. Currently, available tests such as the Kolmogorov-Smirnov test and the Chi-Square test address the problem of goodness-of-fit for random data. In particular, these tests provide information about whether a set of random data is statistically consistent with a specified distribution, to within a certain confidence level. However, if the specified distribution is rejected, these tests cannot be used for approximating the underlying PDF of the random data. Moreover, these tests require large sample sizes for reliable results.

In practice, only a small number of samples may be available. Therefore, the scheme used should be efficient for small sample sizes. A new algorithm based on sample order statistics has been developed in [50] for univariate distribution identification. This algorithm has two modes of operation. In the first mode the algorithm performs a goodness-of-fit test. Specifically, the test determines, to a desired confidence level, whether random data is statistically consistent with a specified probability distribution. In the second mode of operation the algorithm approximates the PDF underlying the random data. In particular, by analyzing the random data and without any a priori knowledge, the algorithm identifies from a stored library of PDFs that density function which best approximates the data. Estimates of the scale, location, and shape param-

eters of the PDF are provided by the algorithm. The algorithm typically works well with small sample sizes of between 50 and 100 samples. An extension of this algorithm for the multivariate Gaussian PDF has been considered in [50] and [52].

In this chapter we describe a new method for univariate distribution approximation. In section 6.2 we present definitions. Section 6.3 describes the algorithm developed in [50] for univariate distribution identification. The proposed distribution identification algorithm is discussed in Section 6.4. Section 6.5 proposes a method to estimate the shape parameter based on the procedure developed in Section 6.4. Finally, conclusions are presented in Section 6.6.

6.2 Definitions

Let $f_Y(y)$ denote the PDF OF Y which has been standardized in a specified manner. Introduce the linear transformation defined by

$$x = \beta y + \alpha \quad (6.1)$$

The PDF of X is given by

$$f_X(x) = \frac{1}{|\beta|} f_Y\left(\frac{x - \alpha}{\beta}\right) \quad (6.2)$$

where α and β are defined to be the location and scale parameters of X , respectively. The mean μ_x and variance σ_x of the random variable X are given by

$$\begin{aligned} \mu_x &= E(X) \\ \sigma_x^2 &= E[(X - \mu_x)^2]. \end{aligned} \quad (6.3)$$

Although the mean and the variance are related to the location and scale parameters, note that the location parameter is not the mean value and the scale parameter is not the square root of the variance, in general. However, for a standardized Gaussian PDF $f_Y(y)$ for which the mean is zero and the variance is unity, the location parameter is the mean of X and the scale parameter is the standard deviation (square root of the variance) of X .

The coefficient of skewness, α_3 , and the coefficient of kurtosis, α_4 , are defined to be

$$\begin{aligned} \alpha_3 &= \frac{E[(X - \mu_x)^3]}{\sigma_x^3} \\ \alpha_4 &= \frac{E[(X - \mu_x)^4]}{\sigma_x^4}. \end{aligned} \quad (6.4)$$

It is readily shown that α_3 and α_4 are invariant to the values of μ_x and σ_x . For any PDF that is symmetric about the mean, $\alpha_3 = 0$. For the case of the Gaussian distribution, $\alpha_3 = 0$ and $\alpha_4 = 3$.

6.3 Goodness of Fit Test

In this section, we introduce a general graphical method for testing whether a set of random data is statistically consistent with a specified univariate distribution. The proposed method not only yields a formal goodness-of-fit test but also provides a graphical representation that gives insight into how well the random data is representative of the specified distribution (null hypothesis). Using the normal distribution as a *reference distribution*, the standardized sample order statistics are represented by a system of linked vectors. Both the terminal point of these linked vectors and the shape of their trajectories are used in determining whether or not to accept the null hypothesis.

In this section we first give a brief description of the corresponding test statistic and then explain the goodness of fit test procedure. For illustration purposes, we assume that the null distribution is Gaussian. However, the proposed procedure works for any null hypothesis.

Let X_k ; $k = 1, 2, \dots, n$ denote the k^{th} sample from a Gaussian distribution with mean μ and variance σ^2 . We define

$$Y_k = \frac{X_k - \bar{X}}{S} \quad k = 1, 2, \dots, n \quad (6.5)$$

where $\bar{X} = \Sigma X_k/n$ is the sample mean and $S = \{\Sigma(X_i - \bar{X})^2/(n-1)\}^{1/2}$ is the sample standard deviation. The standardized order statistics are denoted by $Y_{i:n}$ $i = 1, 2, \dots, n$ and are obtained by ordering the Y_k ; $k = 1, 2, \dots, n$ such that $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$. The i^{th} linked vector is characterized by its length and orientation with respect to the horizontal axis. Let $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ denote the ordered samples obtained by ordering X_k ; $k = 1, 2, \dots, n$. Let $m_{1:n}, m_{2:n}, \dots, m_{n:n}$ denote the expected values of the standard normal order statistics, where $m_{i:n} = E[(X_{i:n} - \mu)/\sigma]$. The length of the i^{th} vector a_i is obtained from the absolute value of the i^{th} standardized sample order statistic $Y_{i:n}$, while its orientation θ_i is related to $m_{i:n}$. More specifically, by definition,

$$a_i = \frac{|Y_{i:n}|}{n} \quad (6.6)$$

$$\theta_i = \pi \Phi(m_{i:n})$$

where $\Phi(x) = (\sqrt{2\pi})^{-1} \int_{-\infty}^x \exp(-\frac{t^2}{2}) dt$ is the distribution function of the standard Gaussian

distribution. We define the sample points Q_k in a two dimensional plane by

$$Q_k = (U_k, V_k) \quad k = 1, 2, \dots, n \quad (6.7)$$

where $U_0 = V_0 = 0$ and

$$\begin{aligned} U_k &= \frac{1}{k} \sum_{i=1}^k \{ \cos(\theta_i) \} |Y_{i:n}| \\ V_k &= \frac{1}{k} \sum_{i=1}^k \{ \sin(\theta_i) \} |Y_{i:n}| \\ k &= 1, 2, \dots, n. \end{aligned} \quad (6.8)$$

The sample linked vectors are obtained by joining the points Q_k . Note that $Q_0 = (0, 0)$. It should also be noted that the statistic Q_n given in eq (6.7) represents the terminal point of the linked vectors defined above. Figure 6.1 shows the linked vectors obtained for the Gaussian distribution with $n = 6$. The null distribution was obtained by averaging the results for 50,000 Monte Carlo trials. The solid curve in Figure 5.1 shows the linked vectors for the sample distribution while the dashed curve shows the linked vector for the null distribution. The magnitude and angles of the linked vectors are obtained from eq (6.6). Note that the angles are independent of the data and depend only on the sample size n . Only the magnitudes of the linked vectors are dependent on the samples drawn and change from one trial to another.

For a typical set of ordered samples (i.e., ordered samples drawn from the null distribution) it is reasonable to expect that the sample linked vectors would closely follow the null pattern. If the ordered set of samples is not from the null distribution, the sample linked vectors are not expected to closely follow the null pattern. Hence, the procedure provides visual information about how well the ordered set of samples fit the null distribution.

An important property of the Q_n statistic is that it is invariant under linear transformation. In particular, we consider the standardization used in eq (6.5). Let $Z_i = aX_i + b$, where a and b are known constants. Let S' denote the sample standard deviation of the samples Z_i . Then, it is readily shown that $\frac{|X_i - \bar{X}|}{S} = \frac{|Z_i - \bar{Z}|}{S'}$. The invariance property follows as a consequence. The advantage of this property is that the PDF of $Q_n = (U_n, V_n)$ depends only on the sample size n and is unaffected by the location and scale parameters. Since it is difficult to determine the joint PDF of U_n and V_n analytically, it is necessary to obtain empirical results.

Assuming that the conditions under the central limit theorem are satisfied, the marginal PDFs of U_n and V_n can be approximated as Gaussian, in the limit of large n . In addition, it is assumed

that the joint PDF of U_n and V_n is approximately bivariate Gaussian. Consequently, all that is needed to determine the bivariate PDF is the specification of $E(U_n)$, $E(V_n)$, $E(U_n V_n)$, $Var(U_n)$ and $Var(V_n)$. Drawing samples from the Gaussian distribution, it has been shown empirically in [50] that for $3 \leq n \leq 100$

$$\begin{aligned} E(U_n) &= 0 \\ E(V_n) &= \mu_v \approx 0.326601 + \frac{0.412921}{n} \\ E(U_n V_n) &= 0 \\ Var(U_n) &= \sigma_u^2 \approx \frac{0.02123}{n} + \frac{0.01765}{n^2} \\ Var(V_n) &= \sigma_v^2 \approx \frac{0.94427}{n} - \frac{0.0951}{n^2}. \end{aligned} \tag{6.9}$$

Since U_n and V_n are approximately bivariate Gaussian for large or moderate sample sizes, their joint PDF can be written as

$$f_{U_n, V_n}(u_n, v_n) = (2\pi)^{-1}(\sigma_u \sigma_v)^{-1} \exp\left(-\frac{t}{2}\right) \tag{6.10}$$

where

$$t = \frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2}. \tag{6.11}$$

Let $t = t_0$. Then the equation

$$t_0 = \frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2} \tag{6.12}$$

is that of an ellipse in the u_n, v_n plane for which

$$f_{U_n, V_n}(u_n, v_n) = (2\pi)^{-1}(\sigma_u \sigma_v)^{-1} \exp\left(-\frac{t_0}{2}\right). \tag{6.13}$$

Points that fall within the ellipse correspond to those points in the u_n, v_n plane for which

$$f_{U_n, V_n}(u_n, v_n) > (2\pi)^{-1}(\sigma_u \sigma_v)^{-1} \exp\left(-\frac{t_0}{2}\right). \tag{6.14}$$

Let

$$\alpha = P(T > t_0) = P(u_n, v_n \text{ fall outside the ellipse given by eq (6.12)}). \tag{6.15}$$

It is well known that the PDF of the random variable T defined by eq (6.11) has a Chi-Square distribution with two degrees of freedom [53] and is given by

$$f_T(t) = 0.5 \exp(-\frac{t}{2}). \quad (6.16)$$

Hence,

$$\alpha = 1 - \exp(-\frac{t_0}{2}). \quad (6.17)$$

Consequently, $t_0 = -2\ln(1 - \alpha)$. Thus, eq (6.12) becomes

$$\frac{u_n^2}{\sigma_u^2} + \frac{(v_n - \mu_v)^2}{\sigma_v^2} = -2\ln(1 - \alpha). \quad (6.18)$$

α is known as the significance level of the test. It is the probability that Q_n falls outside the ellipse specified by eq (6.18) given that the data is coming from a Gaussian distribution. $1 - \alpha$ is known as the confidence level and the corresponding ellipse is known as the confidence ellipse.

Eq (6.12) can be written in the standardized form

$$1 = \frac{u_n^2}{\sigma_u^2 t_0} + \frac{(v_n - \mu_v)^2}{\sigma_v^2 t_0} \quad (6.19)$$

where the lengths of the major and minor axes are given by $\max [\sigma_u \sqrt{t_0}, \sigma_v \sqrt{t_0}]$ and $\min [\sigma_u \sqrt{t_0}, \sigma_v \sqrt{t_0}]$ respectively. From eq (6.17), observe that smaller values of α correspond to larger values of t_0 . Consequently, the confidence ellipses become larger as the confidence level is increased.

For a given sample size n ($n \leq 100$) approximate values of μ_v , σ_u^2 and σ_v^2 can be obtained from eq (6.9). The confidence ellipse of eq (6.18) can then be used to make a visual test of the null hypothesis. If the terminal sample point falls inside the ellipse, then the data is declared as being consistent with the Gaussian distribution with confidence level $1 - \alpha$. Otherwise the null hypothesis is rejected with a significance level α .

A major difficulty in determining the joint PDF of U_n and V_n is that the coefficients of skewness and kurtosis of U_n and V_n (see Table 5.1) indicate that the Gaussian approximation for the bivariate PDF may not be satisfactory for $n < 10$. The empirical bivariate PDF of U_n and V_n were obtained by using 50,000 Monte-Carlo trials for $n=3, 10, 20, 30, 50$ and 100 . The corresponding probability contours are shown in Figure 6.2. The same procedure is used even when the null distribution is different from the Gaussian distribution. However, note that the

standard Gaussian distribution is always used as the reference distribution for determining the angles θ_i .

6.4 Distribution Approximation

In this section we present a graphical procedure for approximating the underlying PDF of a set of random data based on the goodness-of-fit test procedure discussed in section 6.3.

Following a similar approach to that outlined in section 6.3, random samples are generated from many different univariate probability distributions. For each specified distribution and for a given n , the statistic $Q_n = (U_n, V_n)$ given by eq (6.8) is obtained for various choices of the shape parameter. Thus, each distribution is represented by a trajectory in the two dimensional plane whose coordinates are U_n and V_n . Figure 6.3 shows an example of such a representation. Twelve distributions, namely Gaussian (1), Uniform (2), Exponential (3), Laplace (4), Logistic (5), Cauchy (6), Extreme Value (7), Gumbel type-2 (8), Gamma (9), Pareto (10), Weibull (11) and Lognormal (12), are represented in this chart. The value of Q_n at each point of the trajectories is obtained by Monte-Carlo experiments using the standard Gaussian distribution as the reference distribution for determining the angles θ_i . The results are based on averaging 1000 trials of 50 samples from each distribution. The samples from each distribution are obtained by using the IMSL subroutines for specified values of the shape parameter. Since the procedure is location and scale invariant, the trajectory reduces to a single point for those PDFs which do not have shape parameters but are characterized only in terms of their location and scale parameters. By way of example, the Gaussian, Laplace, Exponential, Uniform and Cauchy PDFs are represented by single points in the $U_n - V_n$ plane. However, those PDFs which have shape parameters are represented by trajectories. For a given value of the shape parameter, a single point is obtained in the $U_n - V_n$ plane. By varying the shape parameter, isolated points are determined along the trajectory. The trajectory for the PDF is obtained by joining these points. In a sense the trajectory represents a family of PDFs having the same distribution but with different shape parameter values. For example, the trajectory corresponding to the Gamma distribution in Figure 6.3 is obtained by joining the points for which the shape parameters are 0.2, 0.3, 0.5, 0.7, 1.0, 2.0, 3.0, 4.0, 6.0, 10.0. As the shape parameter increases, note that the Gamma distribution approaches the Gaussian distribution. The representation of Figure 6.3 is called an identification chart. Some distributions such as the β distribution and the SU-Johnson system of distributions, have two shape parameters. For these cases, the trajectories are obtained by holding one shape

parameter fixed while the other is varied. For these distributions, several different trajectories are generated in order to cover as much of the $U_n - V_n$ plane as possible. For certain choices of the shape parameters, two or more PDFs become identical. When this occurs, their trajectories intersect on the identification chart.

It is apparent that the identification chart of Figure 6.3 provides a one to one graphical representation for each PDF for a given n . Therefore, every point in the identification chart corresponds to a specific distribution. Thus, if the null hypothesis in the goodness-of-fit test discussed in section 6.3 is rejected, then the distribution which approximates the underlying PDF of the set of random data can be obtained by comparing Q_n obtained for the samples with the existing trajectories in the chart. The closest point or trajectory to the sample Q_n is chosen as an approximation to the PDF underlying the random data. The closest point or trajectory to the sample point is determined by projecting the sample point Q_n to neighboring points or trajectories on the chart and choosing that point or trajectory whose perpendicular distance from the sample point is the smallest. The complete approximation algorithm is summarized as follows.

1. Compute Y_k as specified in section 6.3
2. Obtain the standardized order statistic $Y_{i:n}$.
3. Compute U_n and V_n from eq (6.8).
4. Obtain an identification chart based on the sample size n as discussed in this section. Plot the sample point Q_n on this chart.
5. Compare the sample point Q_n with the existing distributions on the chart. The nearest neighboring point (or trajectory) on the chart is used as an approximation to the PDF of the samples.

The accuracy of this procedure can be increased by including as many distributions as possible in the identification chart. However, it is emphasized that this procedure does not identify the underlying PDF. Rather it identifies a suitable approximation to the underlying PDF.

6.5 Parameter Estimation

Once the distribution of the samples is approximated, the next step is to estimate its parameters. The method discussed in section 6.4 lends itself for estimating the parameters of the

approximated distribution. We present the estimation procedure for the location, scale and shape parameters in this section.

6.5.1 Estimation of Location and Scale Parameters

Let $f(x; \alpha, \beta,)$ denote the distribution which approximates the PDF of the set of random data, where α and β are the location parameter and scale parameter, respectively of the approximating PDF. Let $X_{i:n}$ denote the ordered statistics of X from a sample of size n . The standardized ordered statistics are defined by

$$W_{i:n} = \frac{X_{i:n} - \alpha}{\beta}. \quad (6.20)$$

Let

$$\mu_{i:n} = E[W_{i:n}]. \quad (6.21)$$

Then

$$E[X_{i:n}] = \beta \mu_{i:n} + \alpha \quad (6.22)$$

We consider the following statistics

$$T_1 = \sum_i \cos(\theta_i) X_{i:n} \quad (6.23)$$

$$T_2 = \sum_i \sin(\theta_i) X_{i:n}$$

where θ_i is the angle defined in eq (6.6). The expected values of T_1 and T_2 are

$$E[T_1] = \sum_i \cos(\theta_i) [\beta \mu_{i:n} + \alpha] \quad (6.24)$$

$$E[T_2] = \sum_i \sin(\theta_i) [\beta \mu_{i:n} + \alpha].$$

These can be written as

$$E(T_1) = a\alpha + b\beta \quad (6.25)$$

$$E(T_2) = c\alpha + d\beta$$

where

$$\begin{aligned}
 a &= \sum_i \cos(\theta_i) \\
 b &= \sum_i \mu_{i:n} \cos(\theta_i) \\
 c &= \sum_i \sin(\theta_i) \\
 d &= \sum_i \mu_{i:n} \sin(\theta_i).
 \end{aligned} \tag{6.26}$$

Because the standardized Gaussian distribution is used as the reference distribution for θ_i , it can be shown that $a = 0$. It follows that

$$\begin{aligned}
 \hat{\beta} &= \frac{E[T_1]}{b} \\
 \hat{\alpha} &= \frac{E[T_2 - d\hat{\beta}]}{c}
 \end{aligned} \tag{6.27}$$

where the symbol $\hat{\cdot}$ is used to denote an estimate. For n sufficiently large (i.e., $n > 50$), suitable estimates for $E[T_1]$ and $E[T_2]$ are

$$\begin{aligned}
 \hat{E}[T_1] &= T_1 \\
 \hat{E}[T_2] &= T_2.
 \end{aligned} \tag{6.28}$$

Estimates for b and d rely upon an estimate of $\mu_{i:n}$. $\hat{\mu}_{i:n}$ is obtained from a Monte Carlo simulation of $W_{i:n}$ where $W_{i:n}$ is generated from the known approximating distribution $f(x; 0, 1)$ having zero location and unity scale parameters. $\hat{\mu}_{i:n}$ is the sample mean of $W_{i:n}$ based upon 1000 Monte Carlo trials. Having $\hat{\mu}_{i:n}$, the estimates for b and d are given by

$$\begin{aligned}
 \hat{b} &= \sum_i^n \hat{\mu}_{i:n} \cos(\theta_i) \\
 \hat{d} &= \sum_i^n \hat{\mu}_{i:n} \sin(\theta_i).
 \end{aligned} \tag{6.29}$$

The scale and location parameters are then estimated by application of eq (6.27).

6.5.2 Shape Parameter Estimation

In this section we present an approximate method for estimating the shape parameter of the approximating PDF. This procedure can be used only when one of the shape parameters is unknown. Let γ denote the shape parameter of the approximating PDF being estimated. Since U_n and V_n are location and scale invariant, the point Q_n depends only on the sample size n and

the shape parameter γ . The expected value of U_n and V_n can be expressed as

$$E(U_n) = \varphi_1(n, \gamma) \quad (6.30)$$

$$E(V_n) = \varphi_2(n, \gamma)$$

where $\varphi_1(\cdot, \cdot)$, $\varphi_2(\cdot, \cdot)$ are some functions of γ and n . For a given sample size n and shape parameter γ_0 the corresponding expected point $\varphi_1(n, \gamma_0)$, $\varphi_2(n, \gamma_0)$ can be determined approximately in the $U_n - V_n$ plane.

The proposed shape parameter estimation method is based on finding a point such that

$$U_n = \varphi_1(n, \hat{\gamma}) \quad (6.31)$$

$$V_n = \varphi_2(n, \hat{\gamma})$$

where $\hat{\gamma}$ is the sample estimator of γ . However, in many instances the sample point may not correspond exactly to a particular trajectory. In such a case, let $E(Q_{1n}) = (u_1, v_1)$, $E(Q_{2n}) = (u_2, v_2)$ denote the expected points corresponding to two different shape parameter values $\gamma = \gamma_1$ and $\gamma = \gamma_2$. It is assumed that the sample point lies in between the points corresponding to γ_1 and γ_2 . Assuming that linear interpolation provides a satisfactory approximation, the estimate of the shape parameter corresponding to the sample point is given by

$$\hat{\gamma} \approx \gamma_1 + \frac{(\gamma_2 - \gamma_1)(x_0 - u_1)}{(u_2 - u_1)} \quad (6.32)$$

where

$$x_0 = \frac{A(V_n - v_1) + A^2 u_1 + U_n}{(A^2 + 1)} \quad (6.33)$$

$$A = \frac{(v_2 - v_1)}{(u_2 - u_1)}.$$

The accuracy of the procedure can be improved by employing a non-linear interpolation method. It must be emphasized that the shape parameter estimation procedure presented in this section is an approximate procedure.

6.6 Conclusions

This chapter has presented a new algorithm for analyzing univariate random data. The algorithm provides a graphical representation for goodness-of-fit test which determines whether a set

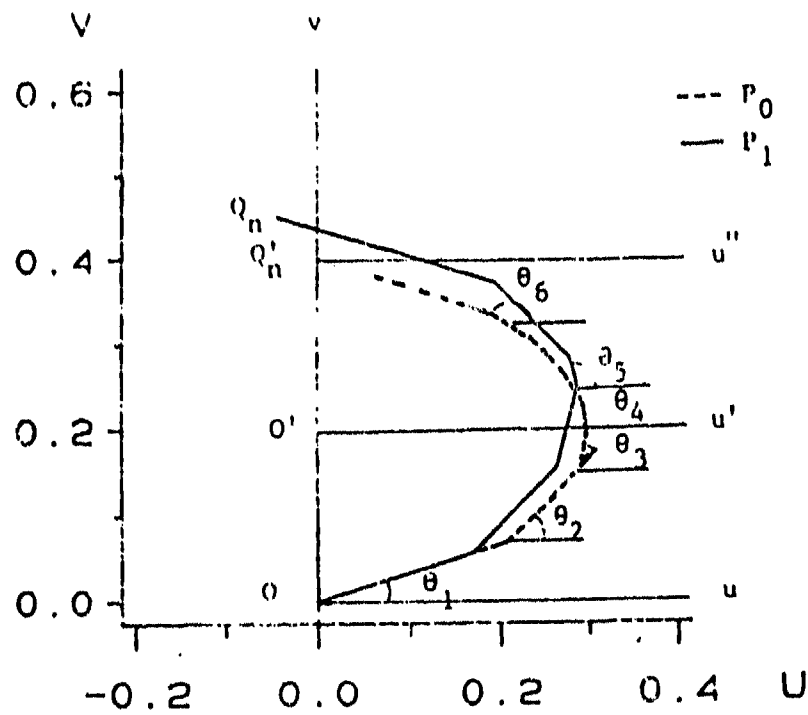


Figure 6.1: Linked Vector Chart: Dashed lines P_0 = Null Linked Vectors, Solid Lines P_1 = Sample Linked Vectors

of random data is statistically consistent with a specified PDF. Also, a graphical procedure is presented for the problem of approximating the underlying PDF of a set of random data. Estimation of location, scale and shape parameters of the approximating PDF have been discussed. Finally, it must be pointed out that the chief advantage of the algorithm presented in this chapter is that it works well for small sample sizes between 50 and 100 samples.

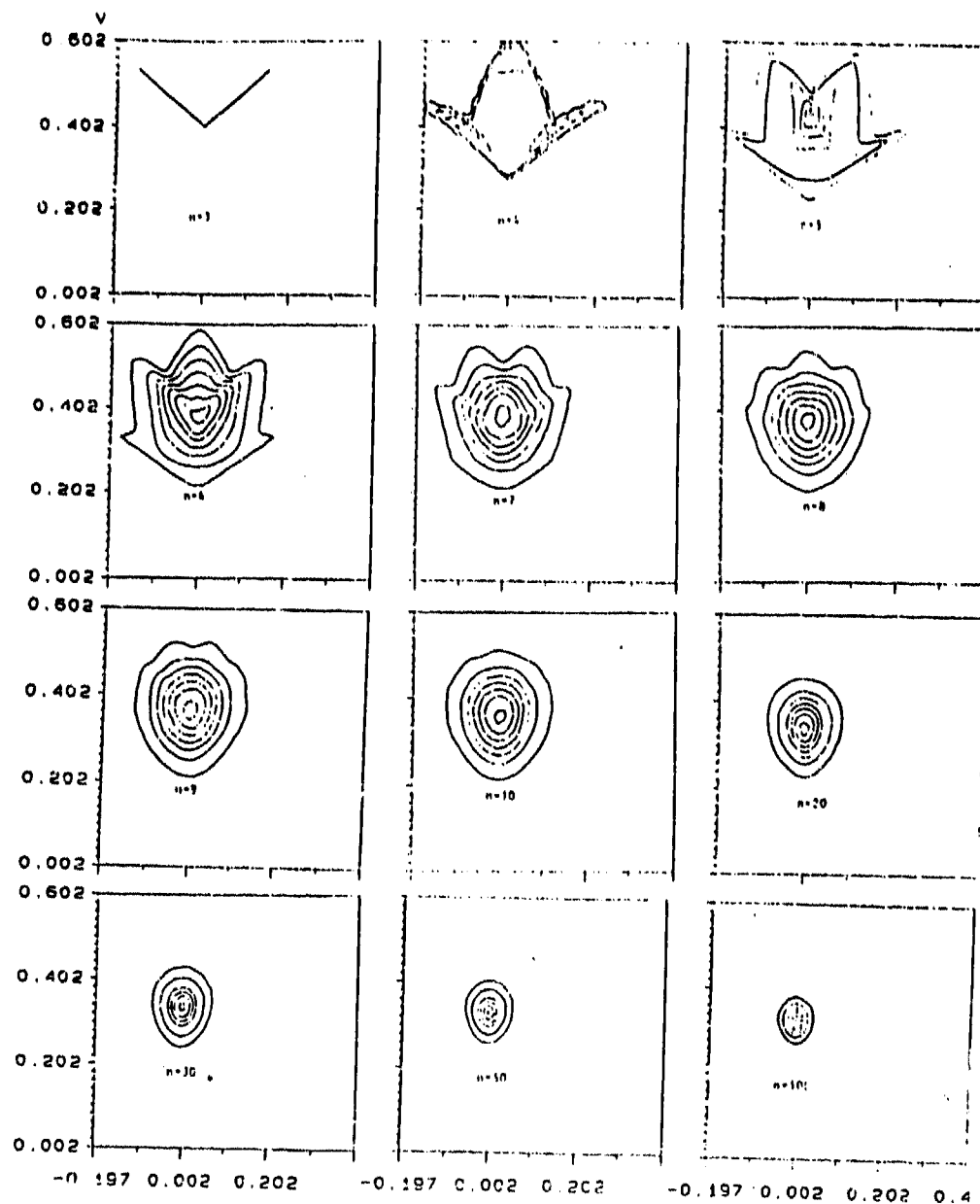


Figure 6.2: Empirical Distribution of Q_n for several values of n

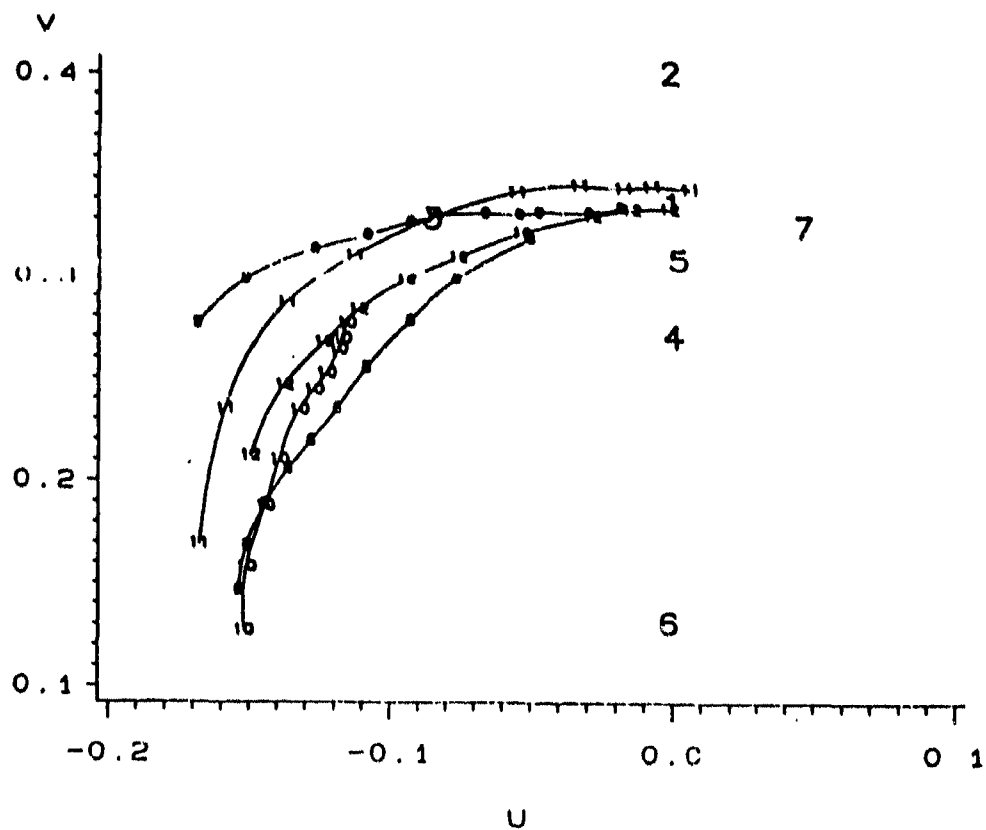


Figure 6.3: Identification Chart for Univariate Distributions Based on 1000 samples ($n=50$)

Table 64 Some Monte Carlo moments of U and V

N	U					V					U					V				
	MEAN	VARIANCE	SKEN.	CURTOS.	MEAN	N	MEAN	VARIANCE	SKEN.	CURTOS.	MEAN	VARIANCE	SKEN.	CURTOS.	MEAN	VARIANCE	SKEN.	CURTOS.		
3	0.00037	0.00046	-0.0036	1.7594	0.47067	52	-0.00011	0.00041	0.00041	2.9636	52	-0.00011	0.00041	0.00041	2.9636	52	-0.00011	0.00041	0.00041	2.9636
4	0.00052	0.00058	-0.0157	3.1854	0.43458	53	-0.00008	0.00040	0.00040	3.0510	53	-0.00008	0.00040	0.00040	3.0510	53	-0.00008	0.00040	0.00040	3.0510
5	0.00070	0.00072	0.0186	2.6987	0.41276	54	-0.00008	0.00040	0.00040	2.9991	54	-0.00008	0.00040	0.00040	2.9991	54	-0.00008	0.00040	0.00040	2.9991
6	0.00097	0.00091	-0.0055	2.8623	0.39793	55	-0.00006	0.00038	0.00038	3.0644	55	-0.00006	0.00038	0.00038	3.0644	55	-0.00006	0.00038	0.00038	3.0644
7	0.00121	0.00113	-0.0073	2.7035	0.38763	56	-0.00003	0.00036	0.00036	3.0245	56	-0.00003	0.00036	0.00036	3.0245	56	-0.00003	0.00036	0.00036	3.0245
8	0.00155	0.00145	-0.0025	2.9141	0.38000	57	-0.00003	0.00036	0.00036	2.9809	57	-0.00003	0.00036	0.00036	2.9809	57	-0.00003	0.00036	0.00036	2.9809
9	0.00198	0.00185	-0.0050	2.9201	0.37362	58	-0.00003	0.00037	0.00037	3.0141	58	-0.00003	0.00037	0.00037	3.0141	58	-0.00003	0.00037	0.00037	3.0141
10	0.00251	0.00229	0.0041	2.9426	0.36865	59	-0.00004	0.00037	0.00037	2.9954	59	-0.00004	0.00037	0.00037	2.9954	59	-0.00004	0.00037	0.00037	2.9954
11	0.00312	0.00276	0.0110	2.9936	0.36516	60	-0.00006	0.00036	0.00036	3.0362	60	-0.00006	0.00036	0.00036	3.0362	60	-0.00006	0.00036	0.00036	3.0362
12	0.00382	0.00319	-0.0213	2.9812	0.36151	61	-0.00004	0.00035	0.00035	2.9801	61	-0.00004	0.00035	0.00035	2.9801	61	-0.00004	0.00035	0.00035	2.9801
13	0.00465	0.00389	-0.0273	2.9912	0.35858	62	-0.00011	0.00035	0.00035	2.9575	62	-0.00011	0.00035	0.00035	2.9575	62	-0.00011	0.00035	0.00035	2.9575
14	0.00562	0.00460	-0.0223	3.0230	0.35633	63	-0.00007	0.00033	0.00033	3.0105	63	-0.00007	0.00033	0.00033	3.0105	63	-0.00007	0.00033	0.00033	3.0105
15	0.00683	0.00558	-0.0133	3.0185	0.35271	64	-0.00007	0.00033	0.00033	3.0385	64	-0.00007	0.00033	0.00033	3.0385	64	-0.00007	0.00033	0.00033	3.0385
16	0.00830	0.00683	-0.0073	3.0413	0.34973	65	-0.00001	0.00033	0.00033	3.0282	65	-0.00001	0.00033	0.00033	3.0282	65	-0.00001	0.00033	0.00033	3.0282
17	0.01000	0.00830	-0.0018	3.0604	0.34728	66	-0.00007	0.00033	0.00033	3.0373	66	-0.00007	0.00033	0.00033	3.0373	66	-0.00007	0.00033	0.00033	3.0373
18	0.01190	0.01000	-0.0070	3.0843	0.34525	67	-0.00007	0.00033	0.00033	3.0265	67	-0.00007	0.00033	0.00033	3.0265	67	-0.00007	0.00033	0.00033	3.0265
19	0.01400	0.01190	-0.0043	3.1136	0.34325	68	-0.00007	0.00033	0.00033	3.0174	68	-0.00007	0.00033	0.00033	3.0174	68	-0.00007	0.00033	0.00033	3.0174
20	0.01630	0.01400	-0.0043	3.1482	0.34167	69	-0.00001	0.00030	0.00030	3.0459	69	-0.00001	0.00030	0.00030	3.0459	69	-0.00001	0.00030	0.00030	3.0459
21	0.01880	0.01630	-0.0043	3.1873	0.34009	70	-0.00001	0.00030	0.00030	3.0329	70	-0.00001	0.00030	0.00030	3.0329	70	-0.00001	0.00030	0.00030	3.0329
22	0.02150	0.01880	-0.0043	3.2306	0.33894	71	-0.00001	0.00030	0.00030	3.0255	71	-0.00001	0.00030	0.00030	3.0255	71	-0.00001	0.00030	0.00030	3.0255
23	0.02450	0.02150	-0.0043	3.2776	0.33801	72	-0.00002	0.00029	0.00029	3.0184	72	-0.00002	0.00029	0.00029	3.0184	72	-0.00002	0.00029	0.00029	3.0184
24	0.02780	0.02450	-0.0043	3.3282	0.33721	73	-0.00002	0.00028	0.00028	3.0114	73	-0.00002	0.00028	0.00028	3.0114	73	-0.00002	0.00028	0.00028	3.0114
25	0.03130	0.02780	-0.0043	3.3826	0.33656	74	-0.00001	0.00028	0.00028	3.0045	74	-0.00001	0.00028	0.00028	3.0045	74	-0.00001	0.00028	0.00028	3.0045
26	0.03500	0.03130	-0.0043	3.4407	0.33602	75	-0.00001	0.00028	0.00028	2.9954	75	-0.00001	0.00028	0.00028	2.9954	75	-0.00001	0.00028	0.00028	2.9954
27	0.03890	0.03500	-0.0043	3.5026	0.33578	76	-0.00001	0.00028	0.00028	2.9909	76	-0.00001	0.00028	0.00028	2.9909	76	-0.00001	0.00028	0.00028	2.9909
28	0.04300	0.03890	-0.0043	3.5683	0.33568	77	-0.00001	0.00028	0.00028	2.9864	77	-0.00001	0.00028	0.00028	2.9864	77	-0.00001	0.00028	0.00028	2.9864
29	0.04730	0.04300	-0.0043	3.6379	0.33548	78	-0.00001	0.00028	0.00028	2.9819	78	-0.00001	0.00028	0.00028	2.9819	78	-0.00001	0.00028	0.00028	2.9819
30	0.05180	0.04730	-0.0043	3.7112	0.33532	79	-0.00001	0.00028	0.00028	2.9775	79	-0.00001	0.00028	0.00028	2.9775	79	-0.00001	0.00028	0.00028	2.9775
31	0.05650	0.05180	-0.0043	3.7882	0.33520	80	-0.00002	0.00028	0.00028	2.9731	80	-0.00002	0.00028	0.00028	2.9731	80	-0.00002	0.00028	0.00028	2.9731
32	0.06140	0.05650	-0.0043	3.8694	0.33509	81	-0.00002	0.00028	0.00028	2.9687	81	-0.00002	0.00028	0.00028	2.9687	81	-0.00002	0.00028	0.00028	2.9687
33	0.06650	0.06140	-0.0043	3.9549	0.33501	82	-0.00002	0.00028	0.00028	2.9644	82	-0.00002	0.00028	0.00028	2.9644	82	-0.00002	0.00028	0.00028	2.9644
34	0.07180	0.06650	-0.0043	4.0449	0.33497	83	-0.00001	0.00028	0.00028	2.9601	83	-0.00001	0.00028	0.00028	2.9601	83	-0.00001	0.00028	0.00028	2.9601
35	0.07730	0.07180	-0.0043	4.1394	0.33485	84	-0.00001	0.00028	0.00028	2.9558	84	-0.00001	0.00028	0.00028	2.9558	84	-0.00001	0.00028	0.00028	2.9558
36	0.08300	0.07730	-0.0043	4.2376	0.33479	85	-0.00001	0.00028	0.00028	2.9515	85	-0.00001	0.00028	0.00028	2.9515	85	-0.00001	0.00028	0.00028	2.9515
37	0.08890	0.08300	-0.0043	4.3397	0.33472	86	-0.00001	0.00028	0.00028	2.9472	86	-0.00001	0.00028	0.00028	2.9472	86	-0.00001	0.00028	0.00028	2.9472
38	0.09500	0.08890	-0.0043	4.4459	0.33456	87	-0.00001	0.00028	0.00028	2.9429	87	-0.00001	0.00028	0.00028	2.9429	87	-0.00001	0.00028	0.00028	2.9429
39	0.09990	0.09500	-0.0043	4.5562	0.33450	88	-0.00001	0.00028	0.00028	2.9386	88	-0.00001	0.00028	0.00028	2.9386	88	-0.00001	0.00028	0.00028	2.9386
40	0.09990	0.09990	-0.0043	4.6707	0.33444	89	-0.00001	0.00028	0.00028	2.9343	89	-0.00001	0.00028	0.00028	2.9343	89	-0.00001	0.00028	0.00028	2.9343
41	0.09990	0.09990	-0.0043	4.7894	0.33438	90	-0.00001	0.00028	0.00028	2.9300	90	-0.00001	0.00028	0.00028	2.9300	90	-0.00001	0.00028	0.00028	2.9300
42	0.09990	0.09990	-0.0043	4.9122	0.33432	91	-0.00001	0.00028	0.00028	2.9257	91	-0.00001	0.00028	0.00028	2.9257	91	-0.00001	0.00028	0.00028	2.9257
43	0.09990	0.09990	-0.0043	5.0399	0.33426	92	-0.00001	0.00028	0.00028	2.9214	92	-0.00001	0.00028	0.00028	2.9214	92	-0.00001	0.00028	0.00028	2.9214
44	0.09990	0.09990	-0.0043	5.1726	0.33420	93	-0.00001	0.00028	0.00028	2.9171	93	-0.00001	0.00028	0.00028	2.9171	93	-0.00001	0.00028	0.00028	2.9171
45	0.09990	0.09990	-0.0043	5.3103	0.33414	94	-0.00001	0.00028	0.00028	2.9128	94	-0.00001	0.00028	0.00028	2.9128	94	-0.00001	0.00028	0.00028	2.9128
46	0.09990	0.09990	-0.0043	5.4530	0.33408	95	-0.00001	0.00028	0.00028	2.9085	95	-0.00001	0.00028	0.00028	2.9085	95	-0.00001	0.00028	0.00028	2.9085
47	0.09990	0.09990	-0.0043	5.6007	0.33402	96	-0.00001	0.00028	0.00028	2.9042	96	-0.00001	0.00028	0.00028	2.9042	96	-0.00001	0.00028	0.00028	2.9042
48	0.09990	0.09990	-0.0043	5.7534	0.33396	97	-0.00001	0.00028	0.00028	2.9000	97	-0.00001	0.00028	0.00028	2.9000	97	-0.00001	0.00028	0.00028	2.9000
49	0.09990	0.09990	-0.0043	5.9111	0.33390	98	-0.00001	0.00028	0.00028	2.8957	98	-0.00001	0.00028	0.00028	2.8957	98	-0.00001	0.00028	0.00028	2.8957
50	0.09990	0.09990	-0.0043	6.0738	0.33384	99	-0.00001	0.00028	0.00028	2.8914	99	-0.00001	0.00028	0.00028	2.8914	99	-0.00001	0.00028	0.00028	2.8914
51	0.09990	0.09990	-0.0043	6.2415	0.33378	100	-0.00001	0.00028	0.00028	2.8871	100	-0.00001	0.00028	0.00028	2.8871	100	-0.00001	0.00028	0.00028	2.8871

Chapter 7

Distribution Approximation of Radar Clutter by SIRPs

7.1 Introduction

This investigation is motivated by a desire to characterize correlated non-Gaussian radar clutter by approximating the underlying probability density function of the clutter. Various investigators have reported experimental results where non-Gaussian marginal probability density functions (PDF) have been used to model the clutter. Usually, radars process N samples at a time. Statistical characterization of the clutter requires the specification of the joint PDF of the N samples. In addition, the clutter may be highly correlated. Hence, the joint PDF must take into account the correlation between samples. Statistical characterization of the clutter is necessary if an optimal radar signal processor is to be obtained. For use of the well known likelihood ratio test, it is desirable to have closed form expressions for the joint PDF of the N clutter samples in order to obtain the optimal radar signal processor. The joint PDF of the N clutter samples can be easily specified when the clutter is Gaussian. However, when the clutter is non-Gaussian and is correlated, many different joint PDFs of the clutter samples can result in the same set of marginal (univariate) distributions having a specified non-Gaussian behavior. The multivariate non-Gaussian PDF can be specified uniquely only when the random variables are statistically independent.

Specification of the multivariate PDF is generally a non-trivial problem with no simple best solution [54]. As explained earlier, the theory of Spherically Invariant Random Processes (SIRP) provides a powerful mechanism to obtain the joint PDF of the N correlated, non-Gaussian clutter

samples. Many of the tractable properties of the Gaussian random process also apply to SIRPs. SIRPs have received considerable attention over the past two decades since most of the elegant and mathematically tractable properties of the multivariate Gaussian distribution generalize to this class of distributions. Applications of SIRPs can be found in the random flight problem [27], signal detection [29], speech signal modeling [30] and radar clutter modeling [32] and [34].

In this Chapter, using certain properties of SIRPs, we adopt an algorithm developed in [50] to identify the underlying distribution of a given set of data. Section 7.2 provides background information about SIRPs. In Section 7.3 we present a procedure for the goodness of fit test for PDFs arising from SIRPs. The proposed distribution identification algorithm is discussed in Section 7.4. Section 7.5 proposes a method to estimate the shape parameter based on the procedure developed in Section 7.4. Finally, conclusions are presented in Section 7.6.

7.2 Characterization of Elliptically Symmetric Distributions

A random vector $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$ is said to have an elliptically contoured distribution if the characteristic function of $\mathbf{X}$ can be expressed as

$$\Phi_{\mathbf{X}}(\omega) = \exp(j\omega^T \mu) \Psi(\omega^T \Sigma \omega) \quad (7.1)$$

where ω and μ is an N by 1 vector, Σ is an N by N positive definite matrix and Ψ is an arbitrary function [37]. In many practical applications involving Monte Carlo experiments, a more restricted class of elliptically contoured distributions are used because of their relative simplicity. This class of distributions called elliptically symmetric distributions (ESD) and has a PDF of the form

$$f_{\mathbf{X}}(\mathbf{x}) = k |\Sigma|^{-\frac{1}{2}} h_N(p) \quad (7.2)$$

where k is a normalization constant chosen so that the volume under the curve of $f_{\mathbf{X}}(\mathbf{x})$ is unity, $p = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)$ is a non-negative quadratic form and $h_N(p)$ is a non-negative, monotonically decreasing, real valued function. The random vector $\mathbf{X}$ having a PDF of the form of eq (7.2) is also called a spherically invariant random vector (SIRV). The constant k is equal to $(2\pi)^{-\frac{N}{2}}$. In this Chapter we shall restrict our attention to SIRVs. A representation theorem for SIRVs [28] states that if a random vector is an SIRV then there exists a non-negative random variable S such that the PDF of the random vector conditioned on S is a multivariate Gaussian PDF. In mathematical terms, we consider the product given by $\mathbf{X} = \mathbf{Z}S$ where $\mathbf{X}$ is an SIRV, S

is a non-negative random variable having PDF $f_S(s)$ and $\mathbf{Z}$ is a Gaussian random vector having the same dimensions as $\mathbf{X}$. Then, we can express $h_N(p)$ as

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds \quad (7.3)$$

where p is the previously defined quadratic form. The PDF of the random variable S (i.e. $f_S(s)$) is called the characteristic PDF of the SIRV. We define a spherically invariant random process as random process (real or complex) such that every random vector obtained by sampling this process is a SIRV having the same characteristic PDF.

In the special case when Σ is the identity matrix, eq (7.2) represents the PDF of a spherically symmetric random vector. This is due to the fact that the PDF in such a case is a function of $\mathbf{x}^T \mathbf{x}$. Elliptically symmetric distributions are related to spherically symmetric distributions in an interesting way. If $\mathbf{Y}$ is a spherically symmetric random vector, then the random vector $\mathbf{X}$ which has an ESD can then be obtained by the linear transformation [28]

$$\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{b} \quad (7.4)$$

where $\mathbf{A}$ is an N by N matrix such that

$$\Sigma = \mathbf{A}\mathbf{A}^T \quad (7.5)$$

and $\mathbf{b}$ is a known $N \times 1$ vector. Thus, in many applications it is sufficient to deal with spherically symmetric distributions and generalize the results to elliptically symmetric distributions.

Finally, the PDF of the quadratic form appearing in eq (7.2) is given by

$$f_P(p) = \frac{p^{\frac{N}{2}-1}}{2^{\frac{N}{2}} \Gamma(\frac{N}{2})} h_N(p) u(p) \quad (7.6)$$

where $\Gamma(\alpha)$ is the Euler-Gamma function and $u(p)$ is the unit step function [34]. It has also been pointed out in Chapter 3 that the PDF of the quadratic form remains unchanged regardless of whether the PDF of the random vector is spherically symmetric or elliptically symmetric. For example, in the multivariate Gaussian case, the PDF of the quadratic form is the well known Chi-square distribution with N degrees of freedom. Therefore, for a given N , the SIRV (or spherically symmetric distribution) is uniquely characterized by the quadratic form. In order to

identify the PDF of the underlying SIRV it is sufficient to identify the PDF of the quadratic form. This attractive property of SIRVs enables us to study various distributional aspects of the corresponding multivariate samples. When a radar uses coherent processing, the joint PDF of the $2N$ quadrature components is of interest. The above results are then applicable with N replaced by $2N$.

7.3 Assessing the Distributional Properties

In modeling real world data, the first step is to determine the most appropriate PDF that approximates the data. In the univariate case, the fit and assessment of the goodness of fit for various distributions has been studied extensively and several methods are available for this purpose. However, limited success has been achieved for the multivariate situation. Although a number of multivariate distributions have been developed, the multivariate Gaussian distribution has been the focus of much of the techniques for multivariate analysis [55].

Assessment of the distributional assumptions for multivariate data is a non trivial problem. Several techniques have been proposed to assess multivariate Gaussianity. In a recent paper Ozturk and Romeu [52] a review of the methods for testing multivariate Gaussianity is given. Many of these methods can be modified or generalized to develop goodness of fit methods for elliptically symmetric distributions. If a random vector $\mathbf{Y}$ is an SIRV, then the corresponding marginal distributions must be identical except for their location and scale parameters. Based on this property, one can use the standard univariate goodness of fit testing procedures to assess the degree of similarity of the marginal distributions of the multivariate data. However, such an approach does not provide a way to assess the joint distribution of the components of the multivariate sample. Recall from Section 4.5 that SIRVs can be characterized in terms of the quadratic form P . Equation (7.6) provides an important property for developing goodness of fit test procedures for SIRVs. Specifically, if the PDF of P can be identified, then the corresponding PDF of the SIRV can also be identified. In fact, many tests for assessment of multivariate Gaussianity are based on the use of this quadratic form [56]. By use of this technique, note that the multivariate distribution identification problem is reduced to a corresponding univariate distribution identification of the quadratic form. Any of the classical goodness of fit testing procedures like the Kolmogorov-Smirnov and Chi-Square tests can be used to address the problem of distribution identification of the quadratic form. However, the requirement of large sample sizes for specifying the parameters of the distribution and low power of the test necessitate use

of alternate procedures that are more efficient.

A general algorithm was developed in [50] to test for univariate and multivariate normality. In this section we propose the use of this algorithm for performing the goodness of fit test for SIRVs. The procedure is summarized here for completeness. Let $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$ denote a vector of observations. For each observation vector of size n , we compute the corresponding quadratic form P_i ($i = 1, 2, \dots, n$). Our goal is to test whether the transformed sample belongs to a certain distribution $F(p; \alpha, \beta, \gamma)$ where α, β are the location and scale parameters, respectively and γ is the shape parameter.

The standardized order statistics are denoted by $Y_{i:n}$ $i = 1, 2, \dots, n$ and are obtained by ordering the Y_k ; $k = 1, 2, \dots, n$ such that $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$.

$$Y_{i:n} = \frac{(P_{i:n} - \bar{P})}{S_P} \quad (7.7)$$

where $\bar{P}$ and S_P are the sample mean and sample standard deviation, respectively of P_k , $k = 1, 2, \dots, n$. The i^{th} standardized ordered quadratic form sample is represented by a point $Q_i = (U_i, V_i)$ in a two dimensional plane where

$$\begin{aligned} U_i &= \frac{1}{n} \sum_{j=1}^i \cos\{\pi \Phi(m_{j:n})\} |Y_{j:n}| \\ V_i &= \frac{1}{n} \sum_{j=1}^i \sin\{\pi \Phi(m_{j:n})\} |Y_{j:n}| \end{aligned} \quad (7.8)$$

In the above equations $\pi = 3.14159$, Φ is the distribution function of the standard normal PDF and $m_{j:n}$ is the expected value of the j^{th} order statistic from the standard normal PDF.

For a given multivariate sample, the points Q_i ($i = 1, 2, \dots, n$) are plotted and joined to obtain a linked vector chart. Similarly, using the expected values of the statistic $Y_{j:n}$, ($j = 1, 2, \dots, n$) under the null hypothesis an expected linked vector chart can also be obtained. The proposed test is based on comparing the sample and expected linked vectors. If the null hypothesis is true, then we expect that the sample linked vectors will follow the expected linked vectors closely.

Finally, a formal goodness of fit test is performed using the terminal point of the expected linked vectors (i.e $Q_n = (U_n, V_n)$). A confidence contour for the true point is obtained to provide a test hypothesis. If the terminal point of the sample does not fall inside the $100(1-\alpha)\%$ confidence ellipse, then the corresponding null hypothesis is rejected at the α level of significance. Note that the Q_n test provides an interesting graphical representation of the data. An example of such

graphical representation is given in Fig 7.1 for testing a multivariate Gaussian distribution with $n = 50$ and $N = 4$.

It should be noted that the Q_n statistic is location and scale invariant. In other words it is independent of the location and scale parameters. However, it depends on the shape parameter of the null distribution. Assessment of the distributional assumptions of distributions that have shape parameters is conceptually different from the corresponding problem for distributions that do not have shape parameters. In the former case, we test whether the sample comes from a particular member of a family of distributions while in the latter case, we test for a single distribution. One possibility for dealing with this problem is to specify the value of the shape parameter and perform the test in the usual way. If the shape parameter cannot be specified, then an adaptive approach which uses the sample estimate of the shape parameter must be employed.

Advantages of using the Q_n procedure are explained in [50]. Usually the classical goodness of fit tests end up with either rejecting or accepting the null hypothesis. An attractive property of the Q_n procedure is that it provides some information about the true distributions if the null hypothesis is rejected. Using this property an algorithm for characterizing and identifying the distributions can be developed. The next section explains these ideas.

7.4 Distribution Identification of SIRVs

Following the same procedure described in Section 6.4, where the reference distribution was Gaussian, an identification chart can be generated for each of quadratic form PDFs of the SIRVs listed in Tables 7.1 and 7.2. Recall from Chapter 4 that the PDF of the quadratic form is invariant to the choice of μ and Σ . Hence, for simplicity, the trajectories for the PDFs of the quadratic forms of the SIRVs listed in Tables 7.1 and 7.2 are obtained by generating the SIRVs having zero mean and identity covariance matrix. Each point on a trajectory is obtained by averaging the results of 2000 Monte Carlo trials of size 100. As before, PDFs which do not have shape parameters are represented by a single point in the U-V plane while those which have shape parameters generate a trajectory in the U-V plane by changing the shape parameter.

An example of the identification chart is given in Fig 7.2 for $N = 4$ and $n = 50$ where the expected values of $Q_n = (U_n, V_n)$ is plotted for various distributions. The Gaussian distribution was used as the reference distribution for determining the angles of the linked vectors. The SIRVs listed in Table 5.1 and Table 5.2 are included in the chart and labeled by number. It is noted

Table 7.1: SIRVs obtained from the marginal envelope PDF

Marginal PDF	$h_{2N}(p)$
Chi	$(-2)^{N-1} A \sum_{k=1}^N G_k p^{\nu-k} \exp(-Bp)$
	$G_k = \binom{N-1}{k-1} (-1)^{k-1} B^{k-1} \frac{\Gamma(\nu)}{\Gamma(\nu-k+1)}$
	$A = \frac{\Gamma(\nu)}{\Gamma(\nu)} (b\sigma)^{2\nu}$
	$B = b^2 \sigma^2$
	$\nu \leq 1$
Weibull	$\sum_{k=1}^N C_k p^{\frac{1}{b}-N} \exp(-Ap^{\frac{1}{b}})$
	$A = a\sigma^b$
	$C_k = \sum_{m=1}^k (-1)^{m+N} 2^N \frac{4^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{b})}{\Gamma(1+\frac{m}{b}-N)}$
	$b \leq 2$
Gen. Rayleigh	$\sum_{k=1}^{N-1} D_k p^{\frac{1}{2}-N+1} \exp(-Bp^{\frac{1}{2}})$
	$A = \frac{\sigma^2 \alpha}{\beta^2 \Gamma(\frac{\alpha}{2})}$
	$B = \beta^{-\alpha} \sigma^\alpha$
	$D_k = \sum_{m=1}^k (-1)^{m+N-1} 2^{N-1} \frac{B^k}{k!} \binom{k}{m} \frac{\Gamma(1+\frac{m}{2})}{\Gamma(1+\frac{m}{2}-N)}$
	$\alpha \leq 2$
Rician	$\frac{e^{2N}}{(1-\rho^2)^{N-\frac{1}{2}}} \sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \left(\frac{\rho}{2}\right)^k \xi_k \exp(-A)$
	$\xi_k = \sum_{m=0}^k \binom{k}{m} I_{k-2m}(\rho A), A = \frac{\rho \sigma^2}{2(1-\rho^2)}$

Table 7.2: SIRVs obtained from the marginal characteristic function

Marginal PDF	$h_{2N}(p)$
Gaussian	$\exp(-\frac{p}{2})$
Laplace	$b^{2N} (b\sqrt{p})^{1-N} K_{N-1}(b\sqrt{p})$
Cauchy	$\frac{2^N b \Gamma(\frac{1}{2}+N)}{\sqrt{\pi} (b^2+p)^{N+\frac{1}{2}}}$
K-distribution	$\frac{b^{2N}}{\Gamma(\alpha)} \frac{(b\sqrt{p})^{\alpha-N}}{2^{\alpha-1}} K_{N-\alpha}(b\sqrt{p})$
Student-t	$\frac{2^N b^{2\nu} \Gamma(\nu+N)}{\Gamma(\nu) (b^2+p)^{N+\nu}}$

that the multivariate Gaussian (1), Laplace (2) and Cauchy (3) distributions are represented by single points on the chart while the multivariate K-distribution (8), Chi (9), Generalized Rayleigh (10) Weibull (11) and Rician (12) are represented by trajectories. The Student-t distribution (4, 5, 6, 7) with degrees of freedom 3, 5, 10 and 15, respectively, is also shown in the chart. The trajectories for each distribution were obtained by joining 10 points resulting from the use of the distributions with parameter values listed in Table 5.3. Each point in the chart is obtained by simulating 2000 samples from the corresponding distributions. The methods developed by Rangaswamy et al. [35, 57] were used to generate the multivariate samples.

The identification chart that provides an interesting display for identifying and characterizing

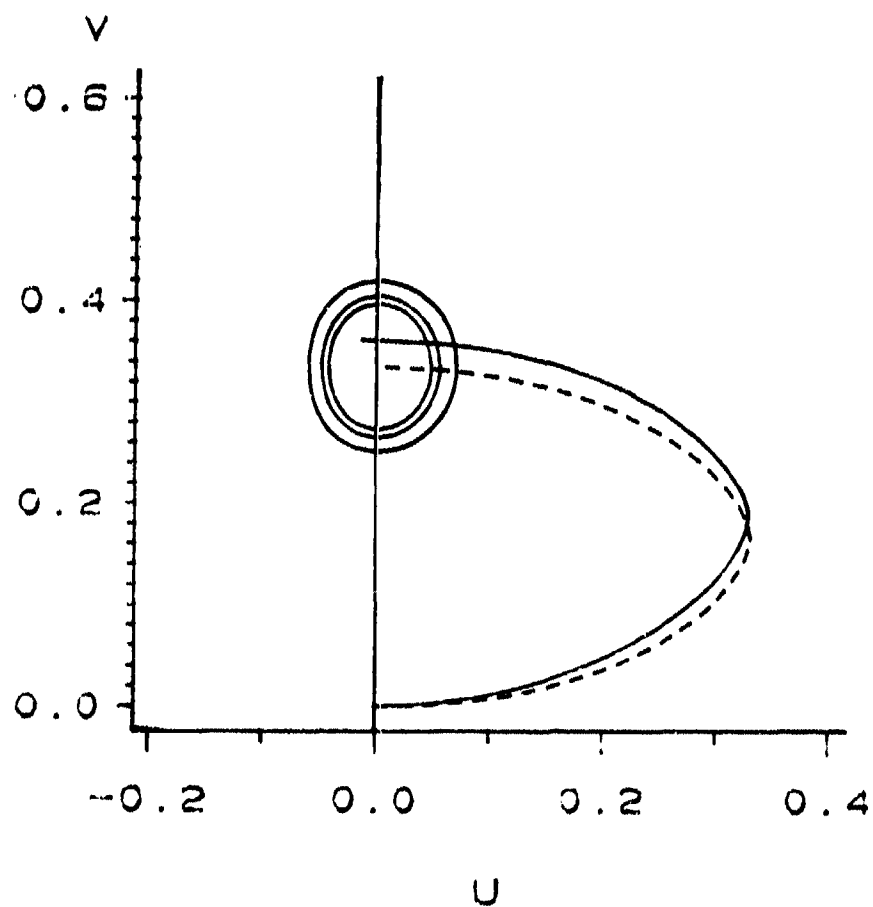


Figure 7.1: Goodness of Fit Test using the Q_n Procedure. 90, 95 and 99% contours for the Gaussian distribution. Broken Line = Null distribution Pattern

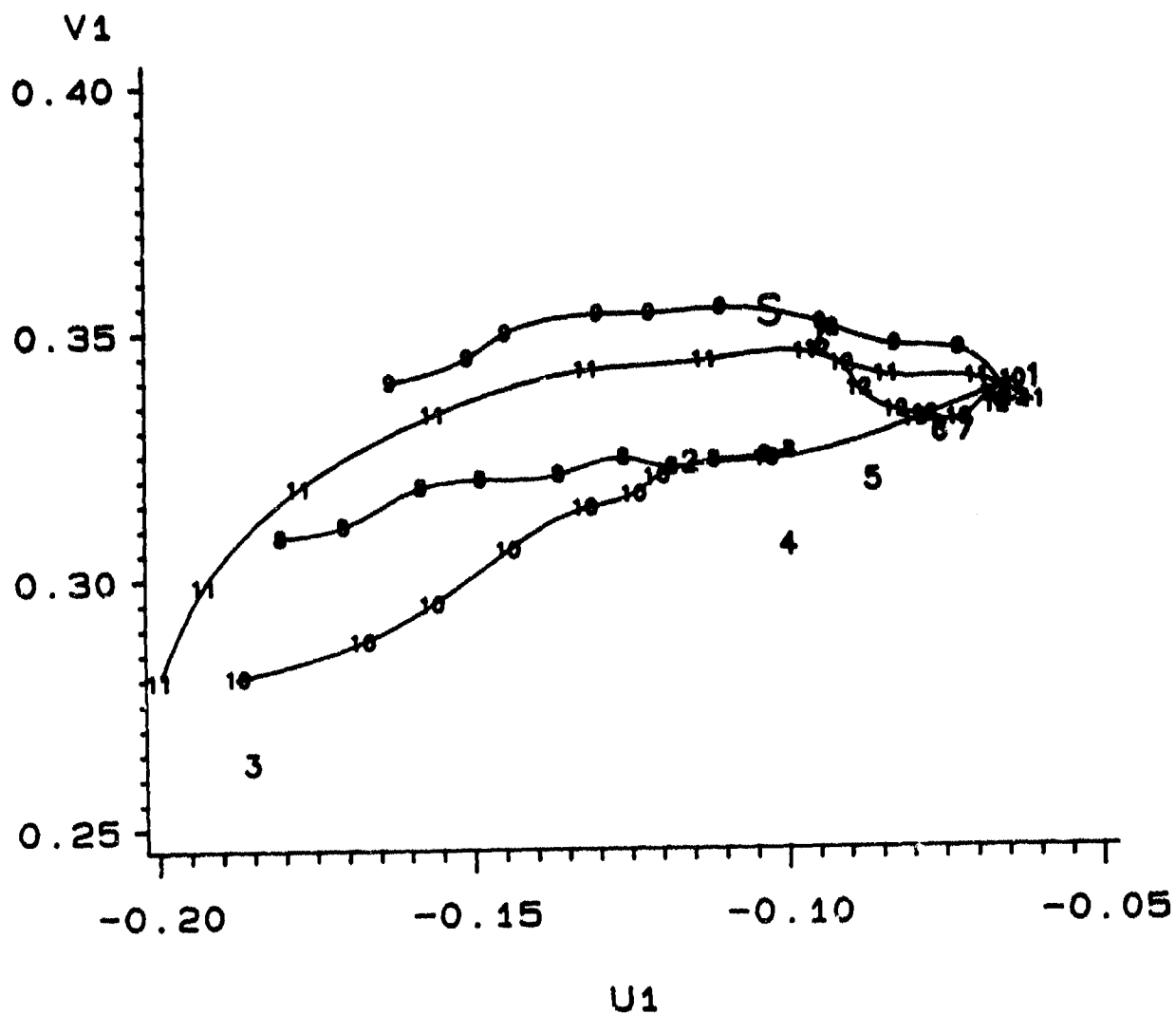


Figure 7.2: Identification Chart for SIRVs ($n=2000$, $N=4$) 1 = Gaussian, 2 = Laplace, 3 = Cauchy, 4, 5, 6, 7 = Student-t, 8 = K-distribution, 9 = Chi, 10 = Generalized Rayleigh, 11 = Weibull, 12 = Rician

Table 7.3: Shape Parameters of the SIRVs Used for the Identification Chart

K-Distribution	0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9, 1.1, 1.5, 1.9
Chi	0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5, 0.6, 0.75, 0.95
Gen. Rayleigh	0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0, 1.5, 2.0
Weibull	0.3, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0
Rician	0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9

the distributions. Also, relationships between the various distributions are clearly seen. For example, as their parameters are varied, certain distributions approach the multivariate Gaussian distribution. Also, for appropriately chosen parameters, the multivariate Weibull distribution and the Generalized Rayleigh distribution coincide. For a given N -variate sample of size n , the statistic Q_n based on the sample quadratic forms can be computed and plotted on the identification chart. Then the nearest distribution to the sample point is identified as the best candidate for the underlying true distribution of the data. An example of such an identification is shown in Figure 7.2 where a well known data set (i.e. Iris Setosa [58]) is used to obtain a value for Q_n and is denoted by the point S. The Iris Setosa data consists of four measurements taken from 50 plants. It is seen from Figure 7.2 that the best candidate for approximating the data is the multivariate Chi (9) distribution.

We point out that there are other methods which can be used for the distribution identification problem. A commonly used technique is the $Q - Q$ plot. To identify the underlying distribution the sample quantiles are plotted against the expected quantiles of a reference distribution. Then the resulting shape of the plotted curve is taken as a basis for identifying the corresponding candidate for the true distributions. However, the identification is made on a subjective basis. Even then the procedure is not very easy. Another well known approach for identifying the distribution is to characterize them via their skewness (α_3) and kurtosis (α_4) coefficients. In this case, all the distributions are represented by points on the α_3 - α_4 plane and the sample data point is compared with the theoretical distributions in the same way as in the Q_n procedure. However, estimates of α_3 and α_4 are known to be highly sensitive to extreme observations and therefore, large sample sizes are necessary to perform the identification for a given degree of accuracy.

7.5 Parameter Estimation

It is well known that the maximum likelihood estimate of the covariance matrix of a Gaussian random vector is the sample covariance matrix. Interestingly enough, it has been shown

in [59] that the maximum likelihood estimate of the covariance matrix Σ is the same sample covariance matrix used in the Gaussian case to within a multiplicative constant. Because Q_n is scale invariant, the identification procedure for SIRVs can proceed without knowledge of the multiplicative constant.

From eq (7.6), it is clear that the expected value of the quadratic form can be expressed as

$$E(P) = \varphi(N, \gamma) \quad (7.9)$$

where γ is the shape parameter of the distribution. For those SIRVs where $\varphi(\cdot)$ can be evaluated in closed form and is invertible, the sample mean of P , denoted by $\bar{P}$ can be used to estimate the shape parameter according to

$$\hat{\gamma} = \varphi^{-1}\{\bar{P}, N\}. \quad (7.10)$$

where $\bar{P} = \frac{1}{n} \sum_{i=1}^n P_i$. For example, in case of the K-distribution, we have $E(P) = 2\nu N$ where ν is the shape parameter of the K-distribution. Clearly, the shape parameter can be approximated as $\hat{\nu} = \frac{\bar{P}}{2N}$. Unfortunately, it is not always possible to obtain an invertible closed form expression for $\varphi(\cdot, \cdot)$. The shape parameter estimation procedure suggested here is not suitable in such a case. An alternate method for the parameter estimation problem is then needed.

In this Chapter we propose to use the Q_n statistic to obtain an approximate estimator for the shape parameter. The underlying procedure is explained in [50] and is summarized here. Let the points (U_1, V_1) and (U_2, V_2) denote expected points corresponding to parameters γ_1 and γ_2 respectively, of a given SIRV. If these points are the nearest points on the curve for the identified distribution to the sample point $Q_n = (U_n, V_n)$, then by using a linear interpolation, an approximate estimator of γ is given by

$$\hat{\gamma} \approx \gamma_1 + \frac{(\gamma_2 - \gamma_1)(x_0 - U_1)}{(U_2 - U_1)} \quad (7.11)$$

where

$$x_0 = \frac{A(V_n - V_1) + A^2 U_1 + U_n}{(A^2 + 1)} \quad (7.12)$$

$$A = \frac{(V_2 - V_1)}{(U_2 - U_1)}$$

The accuracy of the proposed estimator for γ depends on the distance between the sample point Q_n and the corresponding curve. If necessary, the approximation can be improved by using

non-linear interpolation methods.

7.6 Conclusions

In this Chapter we have addressed the problem of distribution approximation of radar clutter under the assumption that the clutter can be characterized as a SIRP. First and foremost, we have shown that the multivariate distribution identification problem for SIRPs can be reduced to an equivalent univariate distribution identification problem of a non-negative quadratic form, resulting in considerable processing simplicity. A new algorithm which provides a graphical representation for the goodness of fit test and the distribution identification has been used. This algorithm, while conceptually simple, is extremely efficient while dealing with small sample sizes. Therefore, it is suitable for use in a variety of practical applications. Finally, based on this algorithm, a new approach has been proposed for estimating the shape parameter of SIRPs.

Chapter 8

Weak Signal Detection - Literature Review

8.1 Weak Signal Problem

In radar applications it is found that the received target signal is contaminated with clutter and thermal noise. The received signal due to undesired reflections from land, sea, atmosphere etc. is called *clutter*. The thermal noise, which is generated by the receiver hardware, is typically modeled as a Gaussian random process. This kind of noise is always present. Depending upon the situation, the clutter may or may not be modeled as a Gaussian random process. Also, the power associated with the background clutter may be orders of magnitude larger than the receiver thermal noise or the desired signal power.

In modern radars, temporal and spatial processing are used to separate the target from the clutter. For example, the received signal from a target having a radial velocity with respect to the radar will experience a Doppler shift. If the target spectrum appears in the tail of the clutter spectrum, then conventional frequency domain techniques can be used to extract the target from the clutter. Similarly, if the spatial spectrum of the target does not overlap that of the clutter, performance will be limited by the background noise rather than the clutter. In this research use is also made of temporal and spatial processing. However, we are interested in the case where the target temporal and spatial spectra cannot be separated from the clutter. By definition, this is referred to as the weak signal detection problem. Given a Range-Doppler-Azimuth cell in which a target is to be detected, it is assumed that the signal is larger than the background noise but much smaller than the clutter. Hence, even after temporal and spatial processing, performance

is limited by the clutter.

Therefore, it becomes very important to identify the clutter plus noise probability density function. This density function is the N^{th} order joint density function of the received radar samples $r_1, r_2, \dots, r_N$ in the absence of a target signal. The received waveform can be modeled as a random process. Since we will be sampling this process at N time instants, we need to have the knowledge of the N^{th} order joint probability density function (PDF) of the N random variables. In this research effort the performance measures of radar receivers are analyzed, given the N^{th} order PDF associated with the random process.

In the hypothesis testing problem, where we have to decide whether the target is present or absent, two kinds of errors can occur: 1) A false alarm which occurs when it is decided that the target is present when it is not, 2) A miss which occurs when it is decided that the target is not present when it is. In many radar problems the chosen criterion is to fix the false alarm at a certain value and then to maximize the probability of detection. In statistical decision theory the *Likelihood Ratio Test (LRT)* is optimum for these kinds of problems. The *LRT* evaluates the likelihood ratio which is the ratio of the N^{th} order joint PDF under the alternative hypothesis H_1 (signal present case) to the N^{th} order joint PDF under the null hypothesis H_0 (signal not present case). This ratio is then compared to a certain threshold to make a decision. Under the constraint of a fixed false alarm, the Neyman-Pearson receiver obtained on the basis of the likelihood ratio test is the optimum receiver.

The components of the received vector $\underline{r}$ can be written mathematically as

$$H_1 : r_i = s_i + d_i \quad (8.1)$$

$$H_0 : r_i = d_i \quad i = 1, 2, \dots, N \quad (8.2)$$

where s_i , and d_i represent the desired signal return and the additive disturbance, respectively. Also, let $f_{\underline{R}}(\underline{r}|H_1)$, $f_{\underline{R}}(\underline{r}|H_0)$, $f_{\underline{D}}(\underline{d})$, denote the N^{th} order PDFs of $\underline{R}$ under H_1 , $\underline{R}$ under H_0 and the disturbance. In general, the disturbance may be composed of clutter plus noise. Since it is not possible to separate the clutter and noise components of the disturbance when the disturbance is measured, we focus on the disturbance itself. As the signal becomes very weak (i.e. as the signal to clutter plus noise ratio (SCNR) approaches zero), the numerator and the

denominator of the *LRT* tend to become identical. This is due to the fact that

$$f_R(r|H_1) \approx f_R(r|H_0) = f_D(d). \quad (8.3)$$

This will result in the likelihood ratio being approximately equal to unity independent of the received signal. Thus, if T_s denotes the likelihood ratio,

$$P_D = \int_{\eta}^{\infty} f_{T_s}(T_s|H_1)dt_s \approx P_F = \int_{\eta}^{\infty} f_{T_s}(T_s|H_0)dt_s, \quad (8.4)$$

where P_D and P_F represent the detection and false alarm probabilities. Therefore, the *LRT* performs poorly in the limit as the signal strength tends to zero.

Even though the problem of weak signal detection in radar applications is of great interest, most of the literature by various researchers has been devoted to strong signals in a clutter plus noise background. Optimal and/or very good sub-optimal schemes have been proposed to achieve the desired level of performance. Only a relatively small fraction of the literature is devoted to the design of practical schemes for the detection of weak signals. In this report we present a general theory for developing practical detector structures for weak signal problems. Also, analysis of performance is carried out for a specific case where the background clutter is assumed to have a multivariate student-T distribution and the signal to clutter plus noise ratio (SCNR) is very small. In such problems the concept of the Locally Optimum Detector (*LOD*) is used to come up with the decision rule which is also a ratio test. For a deterministic signal, a statistic is obtained by taking the ratio of the derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0 . The limit of this ratio as the signal strength tends to zero is evaluated to obtain the test statistic for the decision rule. In the random signal case the test statistic is a ratio, in the limit as the signal strength tends to zero, of the second derivative with respect to the signal strength of the N^{th} order joint PDF under H_1 to the N^{th} order joint PDF under H_0 . This approach is valid when it is known that the SCNR ratio is very small but the actual value of SCNR is unknown. Thus, the *LOD* turns out to be a Uniformly Most Powerful (UMP) test for the class of problems where the SCNR is in the neighborhood of zero. The theory of *LODs* is explained in detail in the next chapter.

8.1.1 Literature Review

The concept of the locally optimum detector was first established by Neyman and Pearson in their paper 'Contributions to the Theory of Statistical Hypothesis Testing' [60, 61]. Subsequently this was applied to statistical communication and signal processing by several researchers.

David Middleton's work [62] on the LOD is based on expanding the LRT in terms of a power series expansion and truncating the series to a first order approximation. In the limit as the signal tends to zero, the canonical structure of the locally optimum detector is established with very weak restrictions on the statistical properties of signal and noise. The analysis applies equally well to non-Gaussian as well as Gaussian, non-stationary as well as stationary processes, for stochastic as well as deterministic signals, continuous as well as discrete time signals and for combinations of signal and noise that need not be additive. In fact, the general character of the results is independent of the particular nature of the signal and noise, although specific noise distributions determine the specific detector structures. Middleton shows that the locally optimum detector is a threshold detector with very strong optimality features in the limit of an infinitely large number of samples. However, in our research, we are interested in applications where the number of samples may not be too large.

For a variety of detection problems, Jack Capon [63] concludes that implementation of the LOD is either less, or no more complicated than the Neyman-Pearson detector. Other researchers in this area such as John Thomas [64], Saleem Kassam [48], Conte and Longo [65], Shishkov and Penev [66] have all obtained performance of the LOD under the asymptotic condition of an infinitely large number of samples. These researchers have modeled the noise samples as independent, identically distributed random variables. This enables them to have a closed form expression for the N^{th} order PDF of multivariate non-Gaussian noise. Applying the LOD test, they have arrived at the decision statistic. Using the central limit theorem, the test statistic is shown to approach Gaussian in the limit of very large sample size. Then the performance measures are evaluated. Shishkov and Penev [66] have considered correlated interference, but have restricted themselves to multivariate Gaussian interference. Modestino and Ningo [47] were amongst the earliest researchers to consider weak signal detection arising from bandpass processes. They have modeled the received signal as statistically independent complex samples and then obtained the joint density function of the inphase and quadrature components. Under the assumption that the clutter density function is circularly symmetric, they transform the joint

density function to an equivalent one involving the envelope and phase. Martinez, Swaszek and Thomas[54], have considered the case where the noise has a multivariate Laplace distribution, where any non-negative definite matrix can be used to model the correlation between the random variables. However, they do not analyze the receiver performance for small sample sizes which is the case of practical interest.

8.2 Non-Gaussian Correlated Data

Previously, general analytic expressions for the various applicable N^{th} order joint non-Gaussian PDFs which allow for correlation between the variables were unavailable. As a result, researchers in the past assumed independence between the samples. By assuming independence between the samples, they were able to get the N^{th} order PDF as a product of the marginals. If we carry out the locally optimum test using the N^{th} order density function based upon independence and evaluate its performance, it is found that an unreasonably large number of samples is needed for acceptable performance. This arises because independent samples imply a white spectrum. Consequently, space-time processing cannot be used to filter the target from the clutter. Based on the concept of Spherically Invariant Random Processes (SIRP), analytical expressions for some N^{th} order joint Non-Gaussian PDFs which allow for correlation between the variables are now available. The SIRP was explained in great detail in Chapters 3-7. Since theoretical evaluation of receiver performance is very difficult for non-Gaussian PDFs, it is done through computer simulation. The computer simulation procedure for receiver performance evaluation is explained in chapter 11. This performance is compared with that of the Gaussian receiver to see the gain obtained due to the added complexity of the locally optimum detector.

Chapter 9

The Locally Optimum Detector

The usual criterion in radar problems is to maximize the probability of detection under a fixed false alarm probability constraint. This receiver is called the Neyman-Pearson receiver. The receiver implements the Likelihood Ratio Test (LRT) and compares it against a threshold whose value is designed to give the desired false alarm probability. In particular, consider the received vector $\underline{R}^T = [R_1, R_2, \dots, R_N]$. Introduce the two hypotheses H_0 and H_1 as described below:

$$H_0 : r_i = c_i + n_i \quad (9.1)$$

$$H_1 : r_i = \theta s_i + c_i + n_i \quad i = 1, 2, \dots, N. \quad (9.2)$$

Thus, H_0 pertains to the hypothesis that the received signal consists solely of clutter plus noise while target signal is assumed to be present under the hypothesis H_1 . Let the joint probability density function of $R_1, R_2, \dots, R_N$ under hypothesis H_k ($k = 0, 1$) be denoted by $f_{\underline{R}}(\underline{r}|H_k)$. The Neyman-Pearson receiver performs the LRT

$$T_s(\underline{r}) = \frac{f_{\underline{R}}(\underline{r}|H_1)}{f_{\underline{R}}(\underline{r}|H_0)} \underset{H_0}{\overset{H_1}{>}} \eta \quad (9.3)$$

where η is specified to satisfy the false alarm constraint

$$P_F = \int_{\eta}^{\infty} f_{T_s}(t_s|H_0) dt_s \quad (9.4)$$

and $f_{T_s}(t_s|H_k)$ is the conditional probability density function of the test statistic T_s given hypothesis H_k .

However, when the signal strength is very small relative to the clutter plus noise, the joint

density function of the received random variables under H_1 approaches that under H_0 . Then the numerator and the denominator of the LRT become approximately equal leading to numerical difficulties in discriminating between the two hypotheses. The Neyman-Pearson test is of course optimum. However, the form of the LRT can be rearranged to yield a test statistic which is more sensitive to perturbations in the received data. This gives rise to the concept of the Locally Optimum Detector (LOD). In this chapter the concept of the LOD is developed in detail using two approaches. The first approach is based on a power series expansion of the LRT and the second approach derives the LOD by an optimization using the principle of Lagrangian multipliers. It is shown that both approaches yield identical detector structures, though starting from different theoretical points of view. As the signal strength becomes weaker, the LOD becomes optimum even though its performance may not be as good as desired for a fixed sample size.

9.1 The Series Approach

9.1.1 The Known Signal Case

Let the additive clutter component $\underline{C} = [C_1, C_2, \dots, C_N]^T$ be stationary and independent of the stationary white Gaussian background noise $\underline{N} = [N_1, N_2, \dots, N_N]^T$. The noise variance σ_n^2 is assumed to be several orders of magnitude below the clutter variance σ_c^2 which is taken to be unity without loss of generality. The signal is assumed to be of the form $\theta \underline{S}$, where $\underline{S}$ is known. The components of $\underline{S}$ are chosen to have $|S_i|^2 = 1$ so that the positive parameter θ is a measure of the signal to clutter ratio (SCR) defined by

$$SCR = \frac{\theta^2 |S_i|^2}{\sigma_c^2} = \theta^2. \quad (9.5)$$

Because the clutter and noise are statistically independent with the noise assumed to have zero mean, the covariance matrix of the disturbance vector $\underline{D} = \underline{C} + \underline{N}$, denoted by M_D , is equal to the covariance matrix of the clutter M_C plus the covariance matrix of the noise M_N . Since the noise is white and stationary, the covariance matrix of the noise is of the form $M_N = \sigma_n^2 I$, where I is the identity matrix. When the clutter is highly correlated, the covariance matrix M_C tends to be ill-conditioned. However, M_D will not be ill-conditioned because, by adding the small value σ_n^2 to the diagonal elements of M_C , the smallest eigenvalue of M_D is guaranteed to be no smaller than σ_n^2 . Also, addition of M_N to M_C ensures that the disturbance spectrum will limit performance even in those frequency intervals where the clutter spectrum is negligible.

With this approximation the LRT takes the form

$$T_s = \frac{f_R(\underline{r}|H_1)}{f_R(\underline{r}|H_0)} = \frac{f_D(\underline{r} - \theta \underline{s})}{f_D(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta'. \quad (9.6)$$

As mentioned previously, when $\theta \ll 1$, the signal $\theta \underline{s}$ represents a small perturbation in the received vector under hypothesis H_1 . Hence, $f_R(\underline{r}|H_1)$ approximately equals $f_R(\underline{r}|H_0)$. As a result, T_s is relatively insensitive to $\theta \underline{s}$. One approach at deriving a weak signal detector is to expand the numerator of the LRT in a Taylor series.

For this purpose, let $\underline{y} = \underline{r} - \theta \underline{s}$. Then

$$f_R(\underline{r}|H_1) = f_D(\underline{y}). \quad (9.7)$$

Expanding $f_D(\underline{y})$ in a Taylor series about the received vector $\underline{r}$, we obtain

$$\begin{aligned} f_D(\underline{y}) &= f_D(\underline{r}) + \sum_{k_1=1}^N (y_{k_1} - r_{k_1}) \frac{\partial f_D(\underline{y})}{\partial y_{k_1}} \Big|_{\underline{y}=\underline{r}} \\ &+ \frac{1}{2!} \sum_{k_1=1}^N \sum_{k_2=1}^N (y_{k_1} - r_{k_1})(y_{k_2} - r_{k_2}) \frac{\partial^2 f_D(\underline{y})}{\partial y_{k_1} \partial y_{k_2}} \Big|_{\underline{y}=\underline{r}} \\ &+ \dots \\ &+ \frac{1}{n!} \sum_{k_1=1}^N \sum_{k_2=1}^N \dots \sum_{k_n=1}^N (y_{k_1} - r_{k_1})(y_{k_2} - r_{k_2}) \dots (y_{k_n} - r_{k_n}) \frac{\partial^n f_D(\underline{y})}{\partial y_{k_1} \partial y_{k_2} \dots \partial y_{k_n}} \Big|_{\underline{y}=\underline{r}} \\ &+ \dots \end{aligned} \quad (9.8)$$

This can be expressed in vector form by introducing the operator

$$(\underline{y} - \underline{r})^T \nabla_{\underline{y}} = \sum_{k=1}^N (y_k - r_k) \frac{\partial}{\partial y_k} \quad (9.9)$$

where the subscript $\underline{y}$ on ∇ indicates partial differentiation with respect to the components of $\underline{y}$.

The expansion of $f_D(\underline{y})$ about the point $\underline{y} = \underline{r}$ then becomes

$$\begin{aligned} f_D(\underline{y}) &= f_D(\underline{r}) + [(\underline{y} - \underline{r})^T \nabla_{\underline{y}}] f_D(\underline{y}) \Big|_{\underline{y}=\underline{r}} \\ &+ \frac{1}{2!} [(\underline{y} - \underline{r})^T \nabla_{\underline{y}}]^2 f_D(\underline{y}) \Big|_{\underline{y}=\underline{r}} \\ &+ \dots \\ &+ \frac{1}{n!} [(\underline{y} - \underline{r})^T \nabla_{\underline{y}}]^n f_D(\underline{y}) \Big|_{\underline{y}=\underline{r}} \end{aligned}$$

$$\begin{aligned}
& + \dots \\
& = f_{\underline{D}}(\underline{r}) + \sum_{n=1}^{\infty} \frac{1}{n!} [(\underline{y} - \underline{r})^T \nabla_{\underline{y}}]^n f_{\underline{D}}(\underline{y})|_{\underline{y}=\underline{r}}.
\end{aligned} \tag{9.10}$$

Recall that $\underline{y} = \underline{r} - \theta \underline{s}$, where θ and $\underline{s}$ are constants. Note that $\underline{y} - \underline{r} = -\theta \underline{s}$ and $\frac{\partial}{\partial y_k} = \frac{\partial}{\partial r_k}$. Then

$$(\underline{y} - \underline{r})^T \nabla_{\underline{y}} = \sum_{k=1}^N (-\theta s_k) \frac{\partial}{\partial r_k} = -\theta \underline{s}^T \nabla_r \tag{9.11}$$

where the subscript r on ∇ indicates partial differentiation with respect to the components of $\underline{r}$. It follows that the expansion may be written as

$$f_{\underline{D}}(\underline{r} - \theta \underline{s}) = f_{\underline{D}}(\underline{r}) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \theta^n [\underline{s}^T \nabla_r]^n f_{\underline{D}}(\underline{r}). \tag{9.12}$$

In order for the above expansion to be meaningful, it is necessary that all the derivatives in the above expansion exist.

Thus, using the above expansion of $f_{\underline{D}}(\underline{r} - \theta \underline{s})$, the Taylor series expansion of the likelihood ratio about the received vector $\underline{r}$ in equation 9.6 can be written as

$$T_s(\underline{r}) = 1 + \left[\sum_{n=1}^{\infty} \frac{(-1)^n \theta^n}{n!} (s^T \nabla_r)^n \right] f_{\underline{D}}(\underline{r}). \tag{9.13}$$

The first term, being a constant, can be combined with the threshold without loss of optimality. The LOD is defined to be the term corresponding to $n = 1$ of the infinite summation. For $\theta \ll 1$, it is assumed that the remaining terms in the summation are negligible. On the other hand, because $\underline{r}$ is random and the partial derivatives of the PDF may be large, the remaining terms may not be negligible. However, it is assumed that this occurs with small probability. The resulting detector structure can be expressed as

$$T_{LOD}(\underline{r}) = - \frac{(s^T \nabla_r) f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta \tag{9.14}$$

where η is chosen so as to achieve the desired false alarm probability.

9.1.2 The Random Signal Case

When the signal is random, $f_{\underline{R}}(\underline{r}|H_1)$ is obtained by integrating the joint density function $f_{\underline{R}, \underline{S}}(\underline{r}, \underline{s}|H_1)$ over all possible values of $\underline{s}$. Hence,

$$f_{\underline{R}}(\underline{r}|H_1) = \int_{-\infty}^{\infty} f_{\underline{R}, \underline{S}}(\underline{r}, \underline{s}|H_1) d\underline{s} = \int_{-\infty}^{\infty} f_{\underline{R}|\underline{S}=\underline{s}}(\underline{r}|\underline{s}, H_1) f_{\underline{S}}(\underline{s}) d\underline{s} = E_s[f_{\underline{R}|\underline{S}=\underline{s}}(\underline{r}|\underline{s}, H_1)] \quad (9.15)$$

where E_s denotes the expectation operation carried out with respect to the random vector $\underline{S}$. Because the denominator of T_s in equation (9.6) is independent of $\underline{s}$, the Taylor series expansion of the likelihood ratio can now be written as

$$T_s(\underline{r}) = 1 + \left[\sum_{n=1}^{\infty} \frac{(-1)^n \theta^n}{n! f_{\underline{D}}(\underline{r})} E_s[(\underline{s}^T \nabla_r)^n] f_{\underline{D}}(\underline{r}) \right]. \quad (9.16)$$

Once again, as in the known signal case, the unity term appearing in the test statistic can be put into the threshold. If we make the assumption that the expected value of the signal vector is $\underline{0}$, then the $n = 1$ term in the infinite series of equation (9.16) goes to zero. Thus, for the random signal case, where the signal vector has zero mean the LOD is defined to be the second term ($n = 2$) in the infinite series. As in the deterministic signal case, θ is assumed to be small enough such that the remaining terms of the series are negligible with high probability. Consequently, the LOD for the random signal case is given by

$$T_{s2}(\underline{r}) = \frac{\theta^2}{2f_{\underline{D}}(\underline{r})} E_s[(\underline{s}^T \nabla_r)^2] f_{\underline{D}}(\underline{r}) \Big|_{\substack{H_1 \\ > \eta''}} \quad (9.17)$$

where T_{s2} represents the second order term in the Taylor series expansion of T_s . The above equation can be rewritten as

$$T_{s2}(\underline{r}) = \frac{\theta^2}{2f_{\underline{D}}(\underline{r})} E_s[\nabla_r^T \underline{s} \underline{s}^T \nabla_r] f_{\underline{D}}(\underline{r}) \Big|_{\substack{H_1 \\ > \eta''}} \quad (9.18)$$

where, as before, η'' is chosen to achieve the specified false alarm probability. Lumping the constant $\frac{\theta^2}{2}$ with the threshold and recognizing that

$$E_s[(\underline{s}^T \nabla_r)^2] = E_s[\nabla_r^T \underline{s} \underline{s}^T \nabla_r] = \nabla_r^T P \nabla_r, \quad (9.19)$$

where P is the covariance matrix of the signal vector, then the detector structure for the locally optimal test becomes

$$T_{LOD}(\underline{r}) = \frac{\nabla_{\underline{r}}^T P \nabla_{\underline{r}} [f_{\underline{D}}(\underline{r})]}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.20)$$

9.2 The Lagrangian approach

Consider again the hypotheses testing problem defined in equation (9.2). Let us define a nonrandomized decision rule $\phi(\underline{r})$ such that

$$\phi(\underline{r}) = \begin{cases} 1; & H_1 \text{ true (target present)} \\ 0; & H_0 \text{ true (target absent).} \end{cases} \quad (9.21)$$

This amounts to partitioning the decision space into two regions, S_1 and S_0 . A target is declared if the vector $\underline{r}$ is present in the space S_1 . If it falls in the space S_0 , then the decision is made that the target is absent. The probability of detection equals the probability that the nonrandomized decision rule equals unity, given that hypothesis H_1 is indeed true. This probability will, in general, be a function of θ , the signal to clutter ratio. Denoting $\beta(\theta)$ as the probability of detection we have

$$P_D = \beta(\theta) = p[\phi(\underline{r}) = 1 | H_1] = \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r} | H_1) d\underline{r}. \quad (9.22)$$

$\beta(\theta)$ is defined to be the power function of the test. The false alarm probability is given by

$$P_F = p[\phi(\underline{r}) = 1 | H_0] = \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{R}}(\underline{r} | H_0) d\underline{r} = \alpha. \quad (9.23)$$

The optimization problem to be discussed in the next section imposes the constraint that the false alarm probability be equal to α . α is also defined to be the significance level of the test.

9.2.1 The Known Signal Case

As discussed earlier, in the limit as the signal strength tends to zero, the probability of detection becomes approximately equal to the probability of false alarm. Therefore, instead of maximizing the probability of detection, one approach is to maximize the slope of the power function ($\beta(\theta)$) curve at the point θ equal to zero. The function to be maximized and the constraint are given

in the following two equations. Maximize

$$\frac{\partial \beta(\theta)}{\partial \theta} \Big|_{\theta=0} = \left[\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \phi(r) f_R(r|H_1) dr \right]_{\theta=0} \quad (9.24)$$

subject to the constraint

$$\int_{-\infty}^{\infty} \phi(r) f_R(r|H_0) dr = \alpha. \quad (9.25)$$

We also require that the test be uniformly most powerful (UMP) in the sense that $\phi(r)$ be independent of θ for small neighborhoods in the vicinity of $\theta = 0$. Notice that there is a derivative with respect to θ outside the integral in equation (9.24). If the function $f_R(r|H_1)$ is a well behaved function such that its derivative exists at all points, the derivative can be moved inside the integral resulting in

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \phi(r) f_R(r|H_1) dr = \int_{-\infty}^{\infty} \frac{\partial \phi(r)}{\partial \theta} f_R(r|H_1) dr + \int_{-\infty}^{\infty} \phi(r) \frac{\partial f_R(r|H_1)}{\partial \theta} dr. \quad (9.26)$$

Because of the UMP requirement, $\frac{\partial \phi(r)}{\partial \theta} = 0$ and the first integral in equation (9.26) integrates to zero. It follows that

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \phi(r) f_R(r|H_1) dr = \int_{-\infty}^{\infty} \phi(r) \frac{\partial f_R(r|H_1)}{\partial \theta} dr. \quad (9.27)$$

Given the function $\frac{\partial \beta(\theta)}{\partial \theta} \Big|_{\theta=0}$ to be maximized along with the false alarm probability constraint, the functional form of the maximization problem using the Lagrange multiplier approach is

$$\max \left[\int_{-\infty}^{\infty} \phi(r) \frac{\partial f_R(r|H_1)}{\partial \theta} dr \Big|_{\theta=0} + \mu \left[\alpha - \int_{-\infty}^{\infty} \phi(r) f_R(r|H_0) dr \right] \right] \quad (9.28)$$

where η is the Lagrange multiplier. Expression (9.28) can be rewritten as

$$\max \left[\int_{-\infty}^{\infty} \phi(r) \left[\frac{\partial f_R(r|H_1)}{\partial \theta} - \eta f_R(r|H_0) \right] dr \Big|_{\theta=0} + \eta \alpha \right] \quad (9.29)$$

To maximize the above integral, the decision regions should be chosen such that the integrand is always positive. In other words, the decision regions are chosen such that

$$\frac{\partial f_R(r|H_1)}{\partial \theta} \Big|_{\theta=0} \geq_{H_0} \eta f_R(r|H_0). \quad (9.30)$$

As was pointed out in the previous section, $f_{\underline{R}}(\underline{r}|H_1)$ is identical to $f_{\underline{D}}(\underline{r} - \theta \underline{s})$. Therefore, the decision rule becomes

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} \Big|_{\theta=0} \underset{H_0}{\overset{H_1}{>}} \eta f_{\underline{D}}(\underline{r}). \quad (9.31)$$

The locally optimum detector is defined to be that detector which implements the ratio test

$$\frac{\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} \Big|_{\theta=0}}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.32)$$

The Lagrange multiplier η is chosen to satisfy the false alarm constraint. Note that

$$f_{\underline{D}}(\underline{r} - \theta \underline{s}) = f_{\underline{D}}(r_1 - \theta s_1, r_2 - \theta s_2, \dots, r_N - \theta s_N). \quad (9.33)$$

As a result,

$$\begin{aligned} \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} &= \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial(r_1 - \theta s_1)} \frac{\partial(r_1 - \theta s_1)}{\partial \theta} + \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial(r_2 - \theta s_2)} \frac{\partial(r_2 - \theta s_2)}{\partial \theta} \\ &+ \dots + \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial(r_N - \theta s_N)} \frac{\partial(r_N - \theta s_N)}{\partial \theta} \\ &= \sum_{k=1}^N \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial(r_k - \theta s_k)} (-s_k). \end{aligned} \quad (9.34)$$

Consequently,

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} \Big|_{\theta=0} = - \sum_{k=1}^N \frac{\partial f_{\underline{D}}(\underline{r})}{\partial r_k} s_k = -(\underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}). \quad (9.35)$$

Thus, the locally optimum detector can also be written as

$$T_{LOD}(\underline{r}) = - \frac{(\underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.36)$$

It can be seen that this detector is identical to the one in equation (9.14) obtained through the series approach.

9.2.2 The Random Signal Case

Consider a random signal $\underline{S}$ and let its joint PDF be denoted by $f_{\underline{S}}(\underline{s})$. Also, without loss of generality, we can make the assumption that the signal vector has zero mean and that each component of the vector has unit variance. Given the signal vector $\underline{S}$ the joint density function

on the received vector under hypothesis H_1 is

$$f_R(\underline{r}|\underline{s}, H_1) = f_D(\underline{r} - \theta \underline{s}). \quad (9.37)$$

The power function for the locally optimum test was given in the previous section in equation (9.22). However, in the random signal case the unconditional density function $f_R(\underline{r}|H_1)$ is obtained by integrating out the random vector $\underline{s}$ from the joint PDF $f_{R,\underline{s}}(\underline{r}, \underline{s}|H_1) = f_R(\underline{r}|\underline{s}, H_1)f_{\underline{s}}(\underline{s})$. Use of equation (9.37) results in

$$\beta(\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}) f_D(\underline{r} - \theta \underline{s}) f_{\underline{s}}(\underline{s}) d\underline{r} d\underline{s}. \quad (9.38)$$

The false alarm constraint is once again given by

$$\int_{-\infty}^{\infty} \phi(\underline{r}) f_R(\underline{r}|H_0) d\underline{r} = \alpha. \quad (9.39)$$

As before, we wish to maximize $\frac{\partial \beta(\theta)}{\partial \theta}|_{\theta=0}$. If the function $f_D(\underline{r} - \theta \underline{s})$ is a well behaved function such that its derivative exists at all points, then

$$\frac{\partial \beta(\theta)}{\partial \theta} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}) \frac{\partial f_D(\underline{r} - \theta \underline{s})}{\partial \theta} f_{\underline{s}}(\underline{s}) d\underline{r} d\underline{s}. \quad (9.40)$$

It follows from equations (9.35) and (9.36) that

$$\frac{\partial \beta(\theta)}{\partial \theta}|_{\theta=0} = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}) \left[\sum_{k=1}^N \frac{\partial f_D(\underline{r})}{\partial r_k} s_k \right] f_{\underline{s}}(\underline{s}) d\underline{r} d\underline{s}. \quad (9.41)$$

Because of the zero mean assumption

$$\int_{-\infty}^{\infty} s_k f_{\underline{s}}(\underline{s}) d\underline{s} = 0. \quad (9.42)$$

We conclude that

$$\frac{\partial \beta(\theta)}{\partial \theta}|_{\theta=0} = 0 \quad (9.43)$$

independent of the choice of $\phi(\underline{r})$. Therefore, to maximize the ability of the power function to increase in the vicinity of the origin, we maximize $\frac{\partial^2 \beta(\theta)}{\partial \theta^2}|_{\theta=0} = 0$. As before, assuming that the

role of integration and differentiation can be interchanged,

$$\frac{\partial^2 \beta(\theta)}{\partial \theta^2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}) \frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta^2} f_{\underline{s}}(\underline{s}) d\underline{r} d\underline{s}. \quad (9.44)$$

However from equation (9.34)

$$\begin{aligned} \frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \sum_{k=1}^N \frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial (r_k - \theta s_k)} (-s_k) \\ &= \sum_{j=1}^N \sum_{k=1}^N \frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial (r_j - \theta s_j) \partial (r_k - \theta s_k)} \frac{\partial (r_j - \theta s_j)}{\partial \theta} (-s_k) \\ &= \sum_{j=1}^N \sum_{k=1}^N \frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial (r_j - \theta s_j) \partial (r_k - \theta s_k)} s_j s_k. \end{aligned} \quad (9.45)$$

Hence,

$$\frac{\partial^2 f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta^2} \Big|_{\theta=0} = \sum_{j=1}^N \sum_{k=1}^N \frac{\partial^2 f_{\underline{D}}(\underline{r})}{\partial r_j \partial r_k} s_j s_k = (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}). \quad (9.46)$$

Then the second derivative of the power function at the origin takes the form

$$\frac{\partial^2 \beta(\theta)}{\partial \theta^2} \Big|_{\theta=0} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}) (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}) f_{\underline{s}}(\underline{s}) d\underline{r} d\underline{s} = \int_{-\infty}^{\infty} \phi(\underline{r}) E_{\underline{s}} (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}) d\underline{r}. \quad (9.47)$$

Using the approach of Lagrange multipliers to maximize the function in equation (9.47) along with the constraint (9.39), the optimization problem can be written as

$$\max \left[\int_{-\infty}^{\infty} \phi(\underline{r}) E_{\underline{s}} (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}) d\underline{r} + \eta [\alpha - \int_{-\infty}^{\infty} \phi(\underline{r}) f_{\underline{D}}(\underline{r}) d\underline{r}] \right]. \quad (9.48)$$

The above expression can be rewritten as

$$\max \left[\int_{-\infty}^{\infty} \phi(\underline{r}) [E_{\underline{s}} (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}) - \eta f_{\underline{D}}(\underline{r})] d\underline{r} \right] + \eta \alpha. \quad (9.49)$$

To maximize the integral the decision regions have to be chosen such that the integrand is always nonnegative. The resulting decision regions yield the inequalities

$$E_{\underline{s}} (\nabla_{\underline{r}}^T \underline{s} \underline{s}^T \nabla_{\underline{r}}) f_{\underline{D}}(\underline{r}) \stackrel{H_1}{\geq} \eta f_{\underline{D}}(\underline{r}). \quad (9.50)$$

If the covariance matrix of the signal vector is denoted by P , then the locally optimum detector can be written as

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P \nabla_r) f_D(\underline{r})}{f_D(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.51)$$

As a general rule for deriving locally optimum tests, note that we maximize at the origin the first non-vanishing derivative of the power function. For the known and the purely random signal cases the first non-vanishing derivative is the first and the second derivative, respectively.

9.3 Special Cases

In this section LOD structures will be derived for two special cases. In the first it is assumed that the N random variables in the disturbance vector $\underline{D}$ are statistically independent. With this assumption, the joint PDF of the N random variables is obtained as a product of the marginal density functions of the individual random variable. In the second the N random variables are modeled as arising from an SIRP. This model enables us to write the joint PDF of the random variables analytically, accounting for the correlation between the random variables. The locally optimum detector structures are derived for both cases. It turns out in both cases that the detector can be expressed in a canonical form. This canonical expression is derived for both the known and the random signal cases.

9.3.1 The Known Signal Case

9.3.1.1 Independent Random Variables

From equation (9.32), the LOD structure in the known signal case is given as

$$\frac{\frac{\partial f_D(\underline{r} - \theta \underline{s})}{\partial \theta} \big|_{\theta=0}}{f_D(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.52)$$

Let the N random variables in the vector $\underline{D}$ be independent such that the PDF of the i^{th} random variable is $f_{D_i}(d_i)$. Therefore, the conditional joint density functions of the N received random variables are given by

$$f_{R_1, R_2, \dots, R_N}(r_1, r_2, \dots, r_N | H_0) = \prod_{i=1}^N f_{D_i}(r_i) \quad (9.53)$$

$$f_{R_1, R_2, \dots, R_N}(r_1, r_2, \dots, r_N | H_1) = \prod_{i=1}^N f_{D_i}(r_i - \theta s_i). \quad (9.54)$$

The numerator in the ratio test of equation (9.52) is evaluated as

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} \Big|_{\theta=0} = \frac{\partial}{\partial \theta} \left[\prod_{i=1}^N f_{D_i}(r_i - \theta s_i) \right] \Big|_{\theta=0} = \sum_{i=1}^N \left\{ (-s_i) \frac{df_{D_i}(r_i)}{dr_i} \prod_{j=1}^N f_{D_j}(r_j) \right\} \quad j \neq i. \quad (9.55)$$

Thus, from equation (9.52) the LOD statistic for independent random variables is given by

$$T_{LOD}(r_1, r_2, \dots, r_N) = - \sum_{i=1}^N s_i \frac{f'_{D_i}(r_i)}{f_{D_i}(r_i)} \quad (9.56)$$

where $f'_{D_i}(r_i)$ denotes the derivative of $f_{D_i}(r_i)$ with respect to r_i . The above equation for the LOD statistic is the canonical form obtained when the random variables are independent. For different density functions, $f_{D_i}(r_i)$, the detector will be different, although its structure remains the same. The canonical form of the detector is shown in Fig. 9.1.

9.3.1.2 Random Variables Arising from an SIRP Distribution

When the random variables of the disturbance are drawn from an SIRP distribution, the joint PDF can be written as

$$f_{\underline{D}}(\underline{d}) = \frac{1}{2\pi^{N/2} |\underline{M}|^{1/2}} h_N(p) \quad (9.57)$$

where $p = \underline{d}^T \underline{M}^{-1} \underline{d}$, $\underline{M}$ is the covariance matrix for the N random variables and $h_N(p)$ is a positive valued, nonlinear function of p . The numerator of the ratio test in equation (9.52) is given by

$$\frac{\partial f_{\underline{D}}(\underline{r} - \theta \underline{s})}{\partial \theta} \Big|_{\theta=0} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{2\pi^{N/2} |\underline{M}|^{1/2}} h_N(p) \right\} \Big|_{\theta=0} = \frac{1}{2\pi^{N/2} |\underline{M}|^{1/2}} \frac{\partial}{\partial \theta} \{h_N(p)\} \Big|_{\theta=0}. \quad (9.58)$$

In terms of θ and $\underline{d} = \underline{r} - \theta \underline{s}$, the quadratic form p equals $(\underline{r} - \theta \underline{s})^T \underline{M}^{-1} (\underline{r} - \theta \underline{s})$. From the chain rule for differentiation we have

$$\frac{\partial}{\partial \theta} (h_N(p)) = \frac{\partial}{\partial p} (h_N(p)) \frac{\partial p}{\partial \theta} \quad (9.59)$$

From the expression for p

$$\frac{\partial p}{\partial \theta} \Big|_{\theta=0} = -2(\underline{s}^T \underline{M}^{-1} \underline{r}). \quad (9.60)$$

Making use of equations (9.58-9.60) the LOD statistic in equation (9.52) becomes

$$T_{LOD}(\underline{r}) = -2(\underline{s}^T \underline{M}^{-1} \underline{r}) \frac{h'_N(p)}{h_N(p)} \quad (9.61)$$

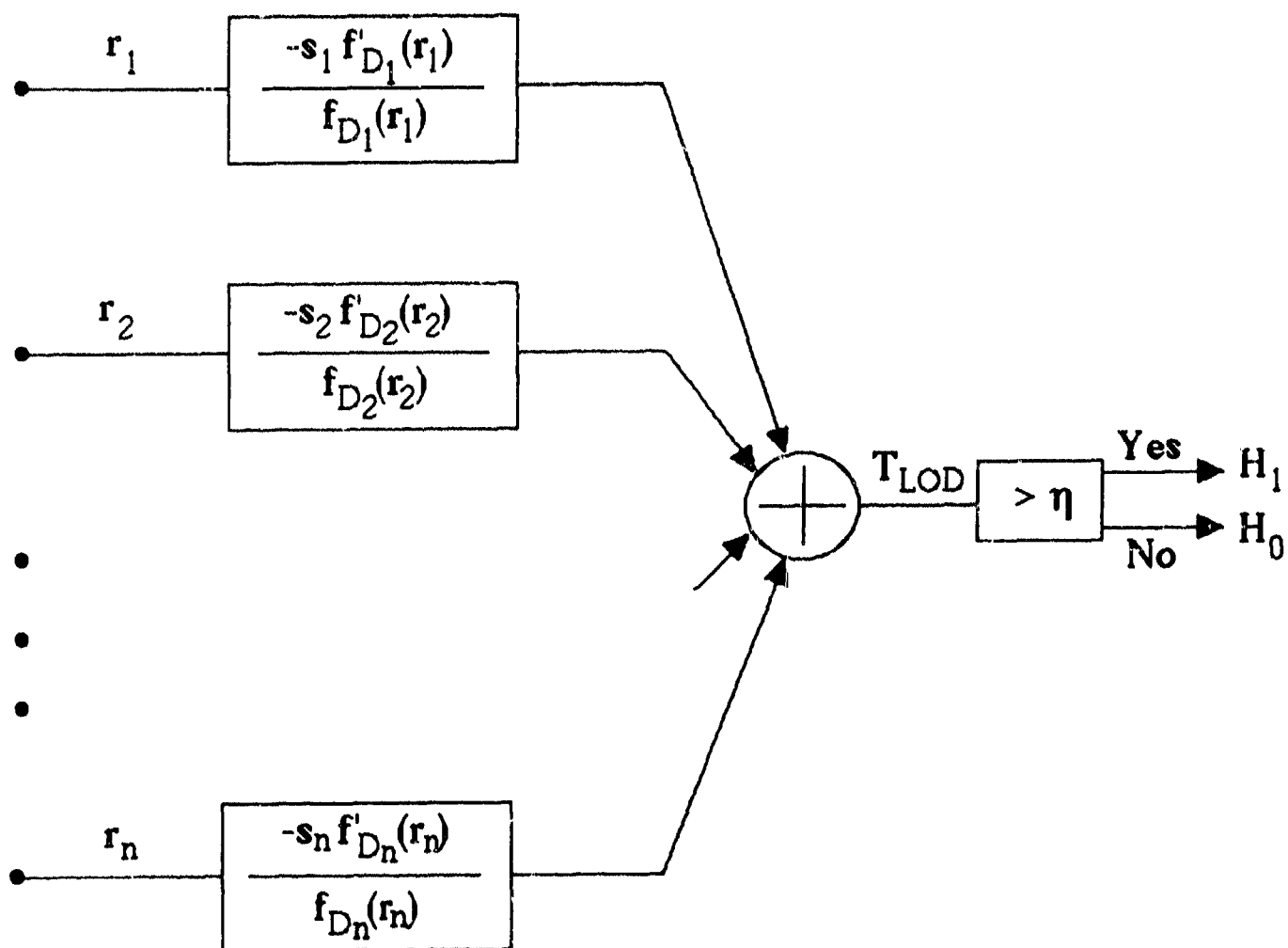


Figure 9.1: Canonical form of LOD assuming known signal and independent random variables.

where $h'_N(p)$ denotes the derivative of the function $h_N(p)$ with respect to the argument p . The LOD statistic in equation (9.61) represents the canonical structure when the disturbance is modeled as an SIRP. The nonlinear function $h_N(q)$ depends on the particular joint density function used to model the disturbance. The canonical structure for the detector is shown in Fig. 9.2.

9.3.2 The Random Signal Case

9.3.2.1 Independent Random Variables

The locally optimum detector is given by equation (9.51) when the signal is random. Rewriting equation (9.51) the LOD structure is

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P \nabla_r) f_{\underline{D}}(\underline{r})}{f_{\underline{D}}(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \eta. \quad (9.62)$$

P is the random signal covariance matrix. For convenience, the signal random variables are assumed to be independent in which case the covariance matrix P is diagonal. Let the diagonal elements of the matrix P be represented by σ_i^2 , $i = 1, 2, \dots, N$. Because the disturbance random variables are also assumed to be independent, the joint density function $f_{\underline{D}}(\underline{r})$ is again given by the product of the marginal density functions of the individual random variables. Specifically,

$$f_{\underline{D}}(\underline{r}) = \prod_{i=1}^N f_{D_i}(r_i) \quad (9.63)$$

Also, when P is diagonal,

$$\nabla_r^T P \nabla_r = \sum_{i=1}^N \sigma_i^2 \frac{\partial^2}{\partial r_i^2}. \quad (9.64)$$

Using equations (9.62-9.64) and following the same steps as in the known signal case, the LOD statistic can be derived as

$$T_{LOD}(\underline{r}) = \sum_{i=1}^N \sigma_i^2 \frac{f''_{D_i}(r_i)}{f_{D_i}(r_i)} \quad (9.65)$$

where the double prime indicates second derivative with respect to the argument. The canonical structure derived above is shown in Fig. 9.3.

9.3.2.2 Random Variables Arising from an SIRP Distribution

When the disturbance vector is modeled as having an SIRP distribution, the joint PDF and the LOD structure are given by equations (9.57) and (9.62), respectively. Since the constant terms in the joint density function cancel out in the numerator and denominator of the ratio test

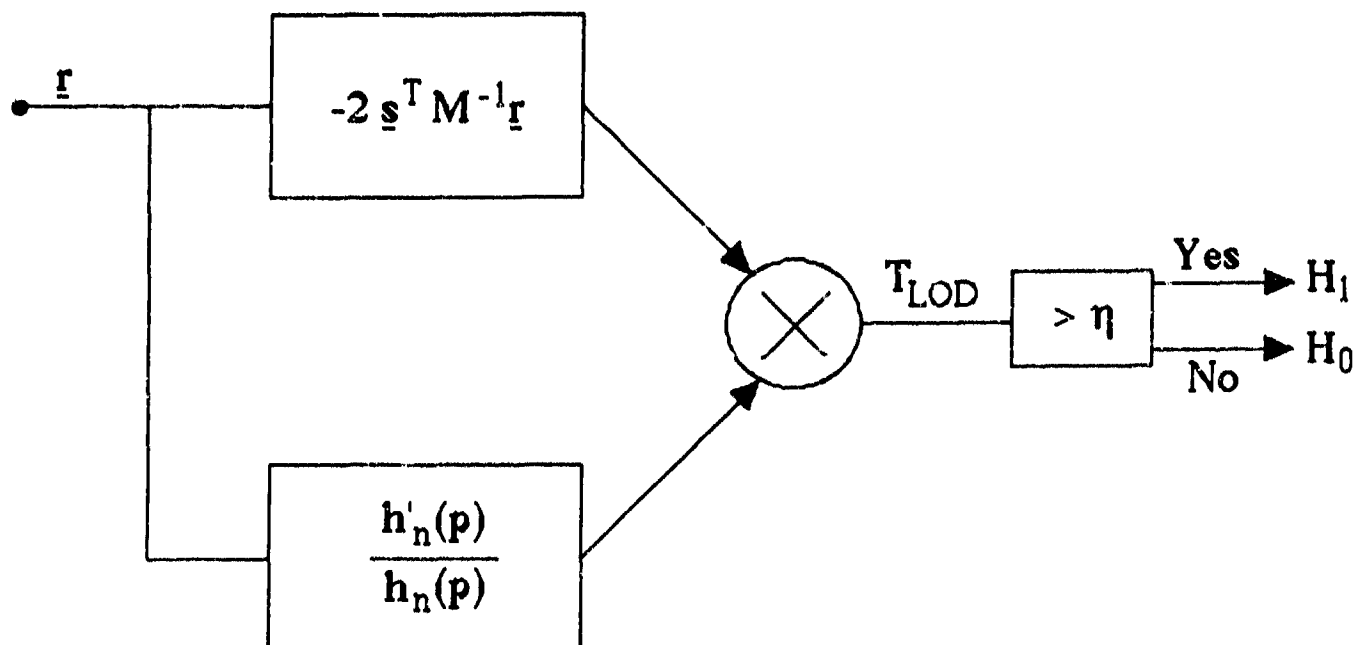


Figure 9.2: Canonical form of LOD assuming known signal and random variables arising from an SIRP.

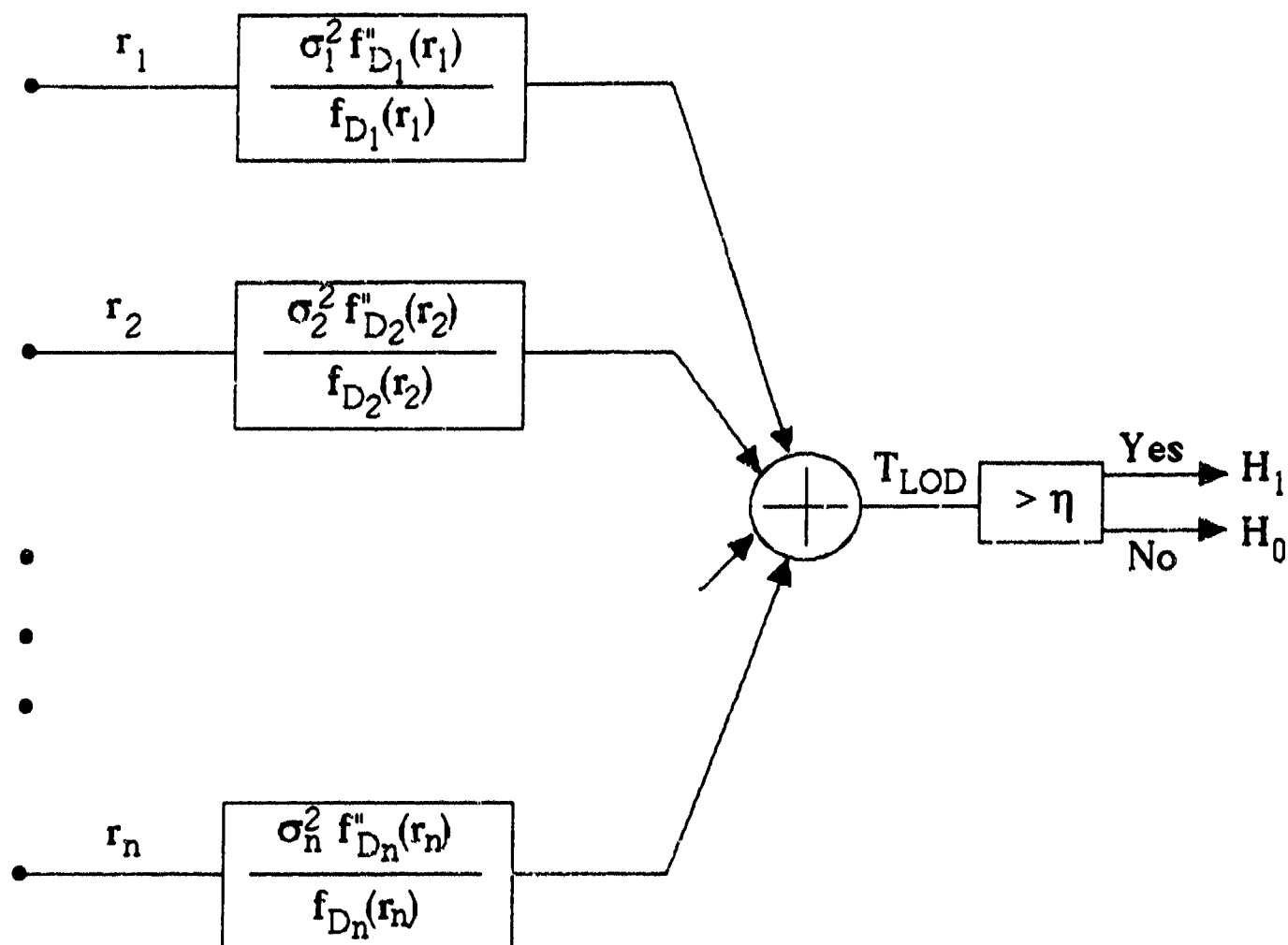


Figure 9.3: Canonical form of LOD assuming random signal and independent random Variables.

in equation (9.62), the LOD statistic is obtained by evaluating

$$T_{LOD}(\underline{r}) = \frac{(\nabla_r^T P \nabla_r) h_N(p)}{h_N(p)}. \quad (9.66)$$

The locally optimum detector statistic that results from the above equation can be written as

$$T_{LOD}(\underline{r}) = \frac{2h'_N(p)S_M}{h_N(p)} + \frac{4h''_N(p)}{h_N(p)} \left(\sum_{i=1}^N \underline{r}^T \underline{M}_i^{-1} \right)^2 \quad (9.67)$$

where S_M represents the sum of all the elements of the matrix M^{-1} and $\underline{M}_i^{-1}$ represents the i^{th} column of M^{-1} . The canonical structure of the detector is shown in Fig. 9.4.

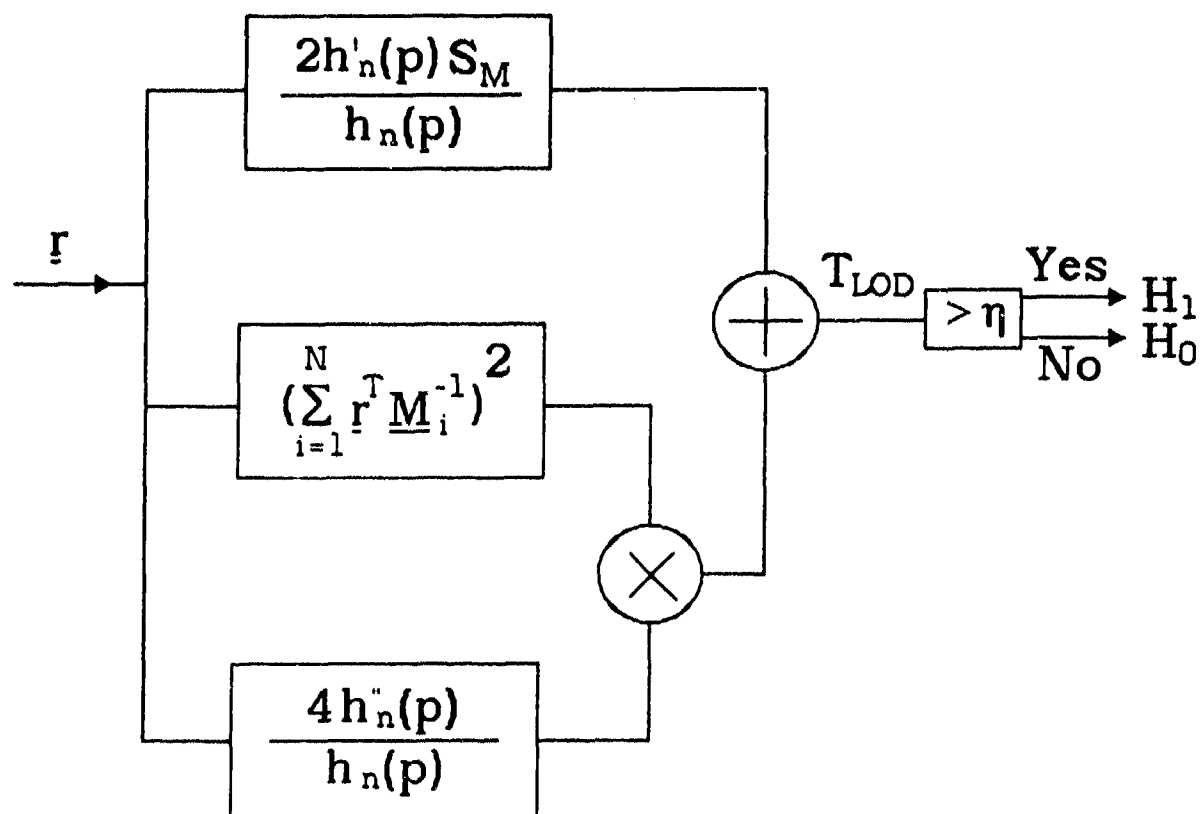


Figure 9.4: Canonical form of LOD assuming random signal and random disturbance arising from an SIRP.

Chapter 10

Determining Thresholds for the Locally Optimum Detector

10.1 Introduction

The hypothesis testing problem for deciding whether or not a target is present is given by equations (9.1-9.2) in Chapter 9. For weak signal applications, it was shown that the Locally Optimum Detector is useful in coming up with a decision rule. For the known signal case, the LOD structure is given by equation (9.32). Since the test statistic is a nonlinear function when $f_{\underline{D}}(\underline{r}|H_0)$ and $f_{\underline{D}}(\underline{r}|H_1)$ are multivariate non-Gaussian density functions, it is not possible, in general, to analytically evaluate in closed form the threshold η for a specified false alarm probability. Given the probability density functions (PDF) of the test statistic denoted by T , under hypotheses H_1 and H_0 , the detection and false alarm probabilities are

$$P_D = \int_{\eta}^{\infty} f_T(t|H_1)dt \quad (10.1)$$

$$P_F = \int_{\eta}^{\infty} f_T(t|H_0)dt. \quad (10.2)$$

P_D and P_F are represented by the shaded areas shown in Fig. 10.1. As indicated in the figure P_F is typically much smaller than P_D .

In practice, the density function of T is not known in advance. For example, depending upon various conditions such as terrain, weather etc., the clutter may be from Gaussian, K-distributed, Weibull or some other probability distribution. It has recently been shown [50] that approximations for the PDFs on T can be determined experimentally using a relatively small

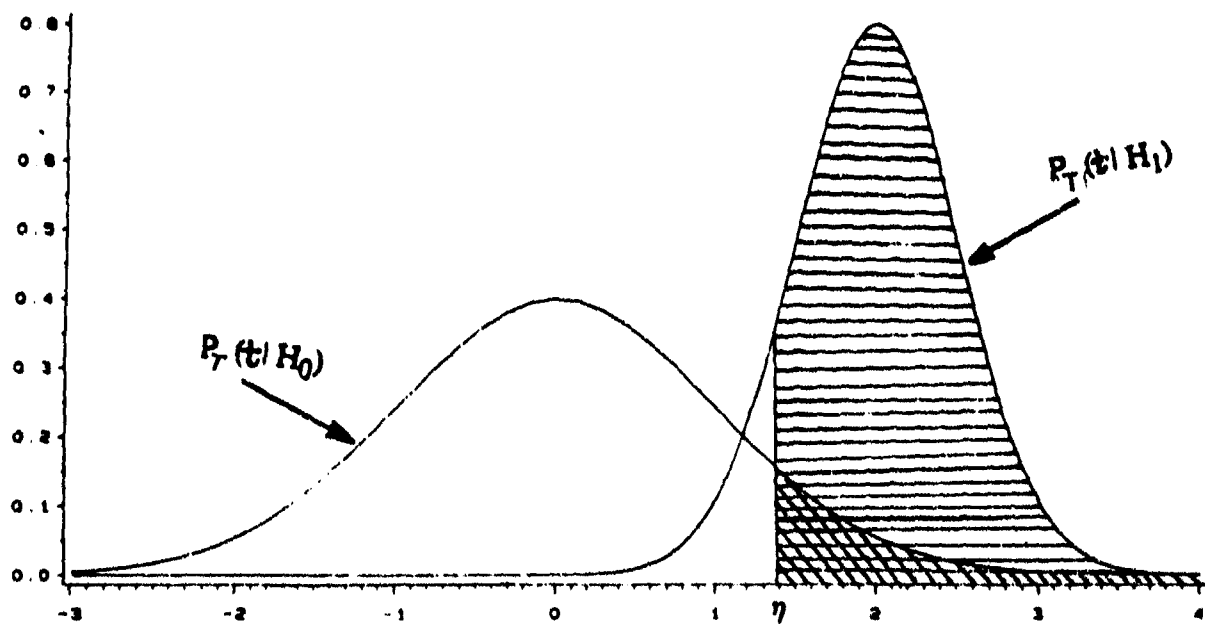


Figure 10.1: Shaded areas indicating P_F and P_D .

number of samples (eg: 50-100 samples give good fits depending on the distribution). Because the number of samples required by Ozturk's technique is small, it is unlikely that actual data samples will be from the extreme tails of the PDFs. Consequently, the good fit mentioned above applies to the main body of the density function.

In order to establish the threshold for a specified P_F , it is necessary to accurately know the behavior of the tail of $f_T(t|H_0)$. The threshold can be determined through Monte Carlo techniques. Unfortunately, the number of trials M required is given by the rule of thumb

$$M \geq \frac{10}{P_F}. \quad (10.3)$$

Hence, if $P_F = 10^{-5}$, at least one million trials should be generated. Clearly, this is not a very desirable situation. In this Chapter a new approach is developed for experimentally determining the extreme tail of $f_T(t|H_0)$, where the number of samples required is several orders of magnitude smaller than that suggested by equation (10.3). Once the tail of $f_T(t|H_0)$ has been estimated, the threshold can be determined by use of equation (10.2).

10.2 Methods for Estimating Thresholds

10.2.1 Estimates Based on Raw Data

In this section we consider some commonly used threshold estimates. These estimates are called raw estimates and are already included in some statistical package programs (eg: the UNIVARIATE procedure in the SAS package).

Let $X_1 \leq X_2 \leq \dots \leq X_n$ denote the sample order statistics from a distribution function $F(x)$. Let p denote the desired false alarm probability. Also, let $n(1-p) = j + g$ where j is the integer part of $n(1-p)$. We denote the threshold estimate based on the k^{th} procedure to be described below by $\eta_p^{(k)}$. Four different threshold estimates are given as follows:

$$\eta_p^{(1)} = (1 - g)X_j + gX_{j+1} \quad (10.4)$$

$$\eta_p^{(2)} = X_k, \text{ where } k \text{ is the integer part of } [n(1 - p) + 1/2] \quad (10.5)$$

$$\eta_p^{(3)} = (1 - \delta)X_j + \delta X_{j+1}, \delta = 0 \text{ if } g = 0; \delta = 1 \text{ if } g > 0 \quad (10.6)$$

$$\eta_p^{(4)} = \delta X_{j+1} + (1 - \delta)(X_j + X_{j+1})/2, \delta = 0 \text{ if } g = 0; \delta = 1 \text{ if } g > 0. \quad (10.7)$$

It is known that all of the above methods are asymptotically equivalent. Thus, if a large sample size is used (where for example M is determined from equation (10.3)), the choice of the

best method is no longer critical. However, in an empirical study [67], it has been shown that $\eta_p^{(4)}$ outperformed the other estimators when $g = 0$. It is noted that the methods based on the above estimators are restricted by the condition that $1 \leq n(1 - p) \leq n - 1$. This implies that the smallest value of the false alarm probability p cannot be lower than $1/n$. Consequently, the threshold corresponding to the smallest false alarm probability which can be estimated by these procedures depends on the sample size. Thus, for a reasonable size of n , estimation of thresholds for small false alarm probabilities cannot be made when these methods are used.

10.2.2 Estimates Motivated by the Extreme Value Theory

Extreme value distributions are obtained as limiting distributions of largest (or smallest) values of sample order statistics. Assuming independent trials, if $X_1 \leq X_2 \leq \dots \leq X_n$ are order statistics from a common distribution function $F(x)$, then the cumulative distribution function of the largest order statistic is given by

$$G_n(x) = P(X_n \leq x) = [F(x)]^n. \quad (10.8)$$

It is clear, as $n \rightarrow \infty$, that the limiting value of $G_n(x)$ approaches zero if $F(x)$ is less than 1 and unity if $F(x)$ is equal to 1 for a specified value of x . A standardized limiting distribution of X_n may be obtained by introducing the linear transformation, $a_n X_n + b_n$, where a_n and b_n are finite constants depending on the sample size n .

In Appendix C, using the theory of limiting distributions [68], it is shown that if there exist sequences a_n and b_n such that

$$\lim_{n \rightarrow \infty} P\left(\frac{X_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_n(a_n x + b_n) \rightarrow \Lambda(x) \quad (10.9)$$

then the solution of the above functional equation yields all the possible limiting forms for the distribution function $G_n(x)$. The solutions to the above equation are derived in Appendix C and are rewritten here:

$$\Lambda(x) = \exp(-e^{-x}) \quad x \geq 0 \quad (10.10)$$

$$\Lambda(x) = \exp(-x^{-k}) \quad x \geq 0, \quad k > 0 \quad (10.11)$$

$$\Lambda(x) = \exp(-(-x)^k) \quad x \leq 0, \quad k > 0. \quad (10.12)$$

In the limit, as n gets large, these are the three types of distribution functions to which the largest

order statistic drawn from almost any smooth and continuous distribution function converge. Therefore, for large x , the tails of almost all smooth and continuous probability density functions for the largest order statistic also converge to three limiting forms. From equations (10.10) and (10.11), respectively, the two limiting forms that pertain to the right tail (the case of interest for the locally optimum detector test statistic) are [68]

$$1. \frac{d\Lambda(x)}{dx} \approx H(x) = e^{-x} \quad (10.13)$$

$$2. \frac{d\Lambda(x)}{dx} \approx H(x) = kx^{-(k+1)} \quad k \geq 0 \quad (10.14)$$

where $H(x)$ approximates the probability density function for large x . The first equation above is the well known exponential distribution and the second equation is related to the Pareto distribution. The details that lead to the limiting distributions of the tails are shown in Appendix C.

It remains to be explained how the distribution of the largest order statistic is related to the tails of the underlying PDF from which the samples are drawn. The relationship is based on the observation that inferences from short sequences are likely to be unreliable. In particular, instead of observing k sets of n samples and taking the largest order statistic from each of the k sets, it is better to observe a single set of nk samples and use the largest k samples from this set [69]. The k largest order statistics from a vector of nk observations constitute the tail of the underlying distribution especially when n is very large. Therefore, the limiting distribution of the largest order statistic closely approximates the tail of the underlying PDF for large n .

10.3 The Generalized Pareto Distribution

The Generalized Pareto Distribution (GPD) is defined for $x > 0$ by the distribution function

$$G(x) = 1 - (1 + \gamma x/\sigma)^{-1/\gamma}, \quad -\infty < \gamma < \infty, \sigma > 0, \gamma x > -\sigma. \quad (10.15)$$

This distribution has a simple closed form and includes a range of distributions depending upon the choice of γ and σ . For example, the exponential distribution results for $\gamma = 0$ and the uniform distribution is obtained when $\gamma = -1$. The GPD defined in equation (10.15) is valid for all $x > 0$ while equations (10.13) and (10.14) are valid only for large x .

The probability density function corresponding to the GPD is given by

$$g(x) = \frac{d}{dx} [1 - (1 + \frac{\gamma x}{\sigma})^{-1/\gamma}] = \frac{1}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1}. \quad (10.16)$$

If we let $\gamma \rightarrow 0$ in the above equation, note that

$$\lim_{\gamma \rightarrow 0} \frac{1}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1} = \frac{1}{\sigma} e^{-\frac{x}{\sigma}}. \quad (10.17)$$

Also, if we let x be large in equation (10.16), note that

$$\frac{1}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}-1} \approx \frac{1}{\sigma} (\frac{\gamma}{\sigma})^{-\frac{1}{\gamma}-1} x^{-\frac{1}{\gamma}-1}. \quad (10.18)$$

Equations (10.17) and (10.18) are of the same form as equations (10.13) and (10.14). Thus, the GPD can be used to approximate both types of tail behavior exhibited by the right tail. Typical plots of the Generalized Pareto PDF are shown for $\gamma < 0$ and $\gamma > 0$ in figures 10.2 & 10.3.

We wish to set thresholds for specified false alarm probabilities when the underlying density functions are unknown. To set very small false alarm probabilities, the tail of the PDF $p_L(l|H_0)$ has to be accurately modeled. Figure 10.4 represents a typical PDF of the test statistic with the tail region of the PDF being defined as that to the right of $t = t_0$. Figure 10.5 shows the tail translated to the origin. The choice for t_0 is somewhat arbitrary. For example, t_0 can be chosen such that the area in the shaded region equals 0.1, 0.05 or 0.01. It is the portion of the PDF to the right of t_0 that we are interested in modeling by the GPD. In particular, the tail region of the PDF is translated to the origin and modeled as a GPD. Once the estimates of σ and γ have been obtained, the GPD is scaled by the area of the shaded region and translated back to the point t_0 . In this way, the area under the PDF of the test statistic is maintained at unity.

10.3.1 Methods for Estimating the Parameters of the GPD

Suppose that the sample ordered statistics $X_1 \leq X_2 \leq \dots \leq X_n$ are drawn from the distribution function $F(x)$. To estimate the right tail of this distribution it is necessary to determine a value (say x_0) and then use those sample observations which are greater than x_0 to obtain the quantity $z = x - x_0$. Once the tail observations have been chosen, the Generalized Pareto Distribution can be fitted to these observations by using standard methods of parameter estimation. Observe that the portion of the observations used from a complete set of samples depends on the choice

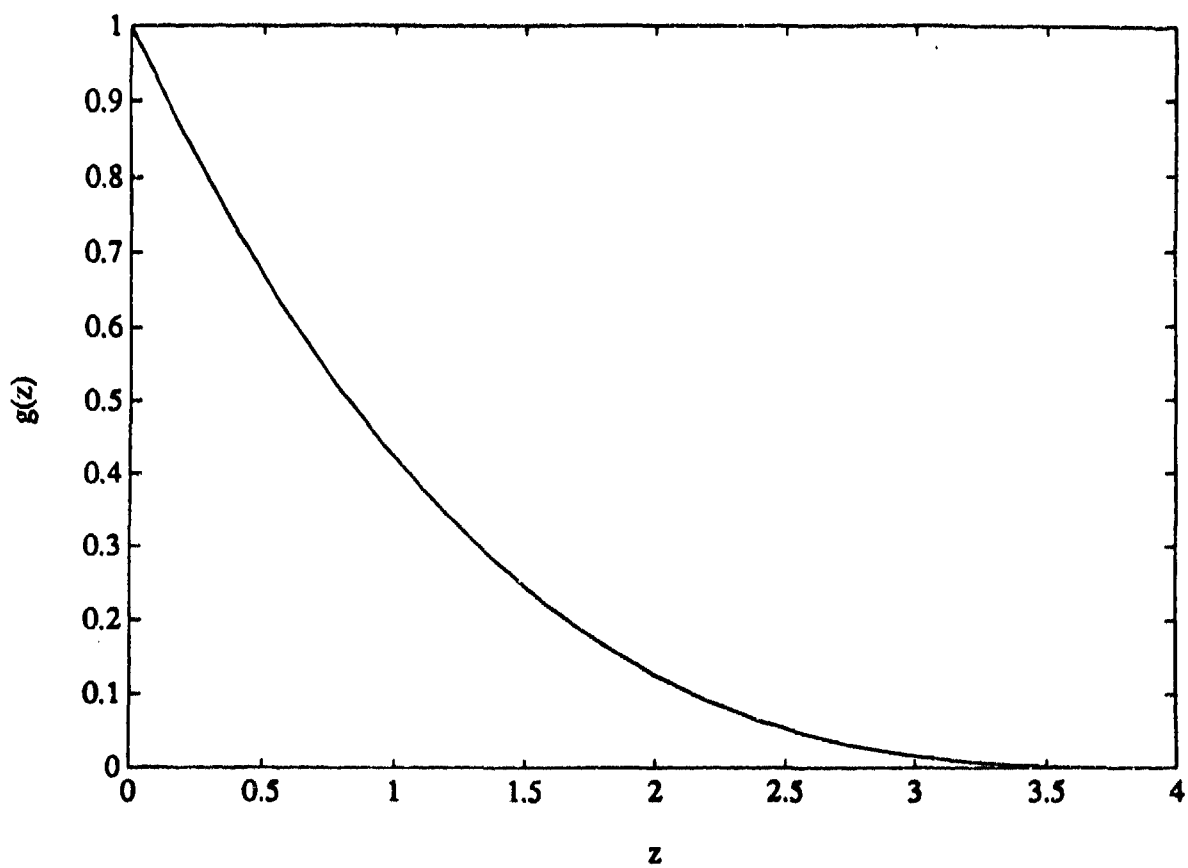


Figure 10.2: Generalized Pareto PDF, $\gamma = -1$.

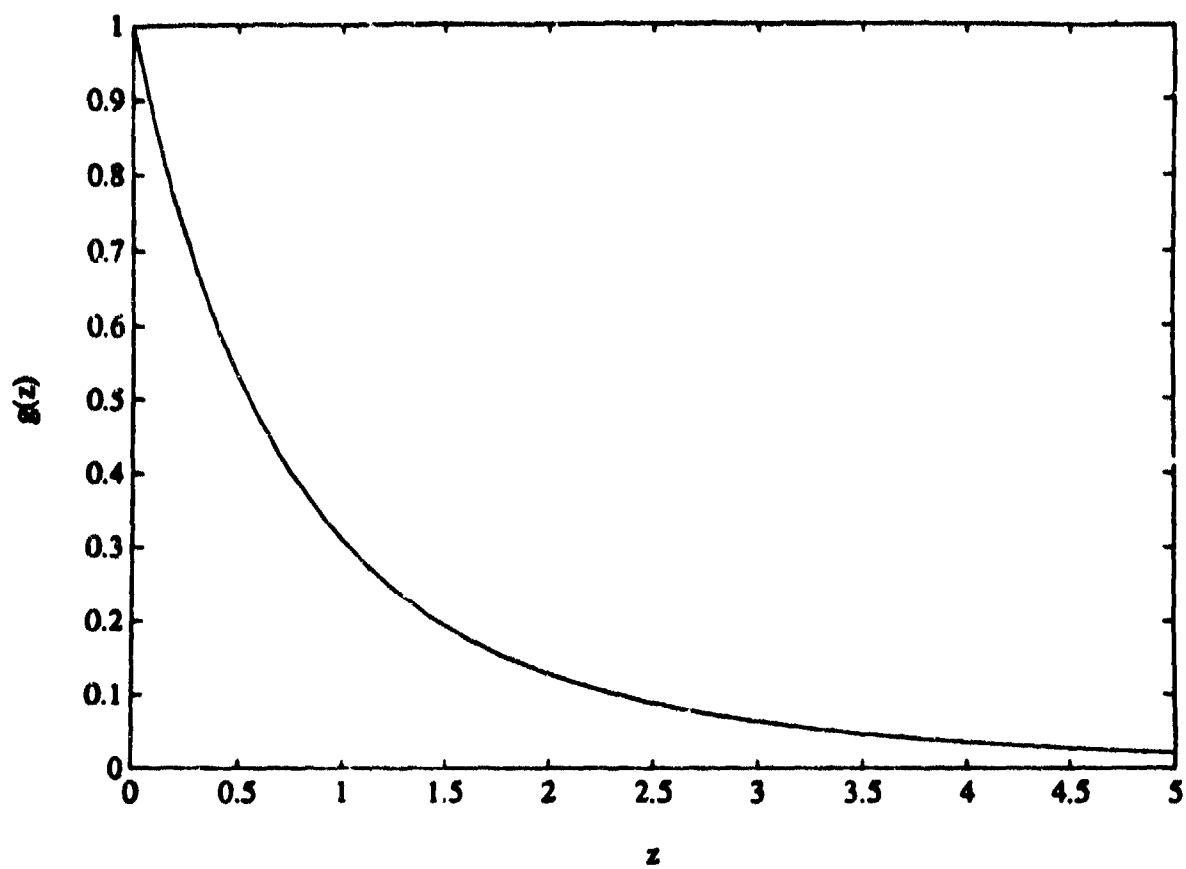


Figure 10.3: Generalized Pareto PDF, $\gamma = 1$.

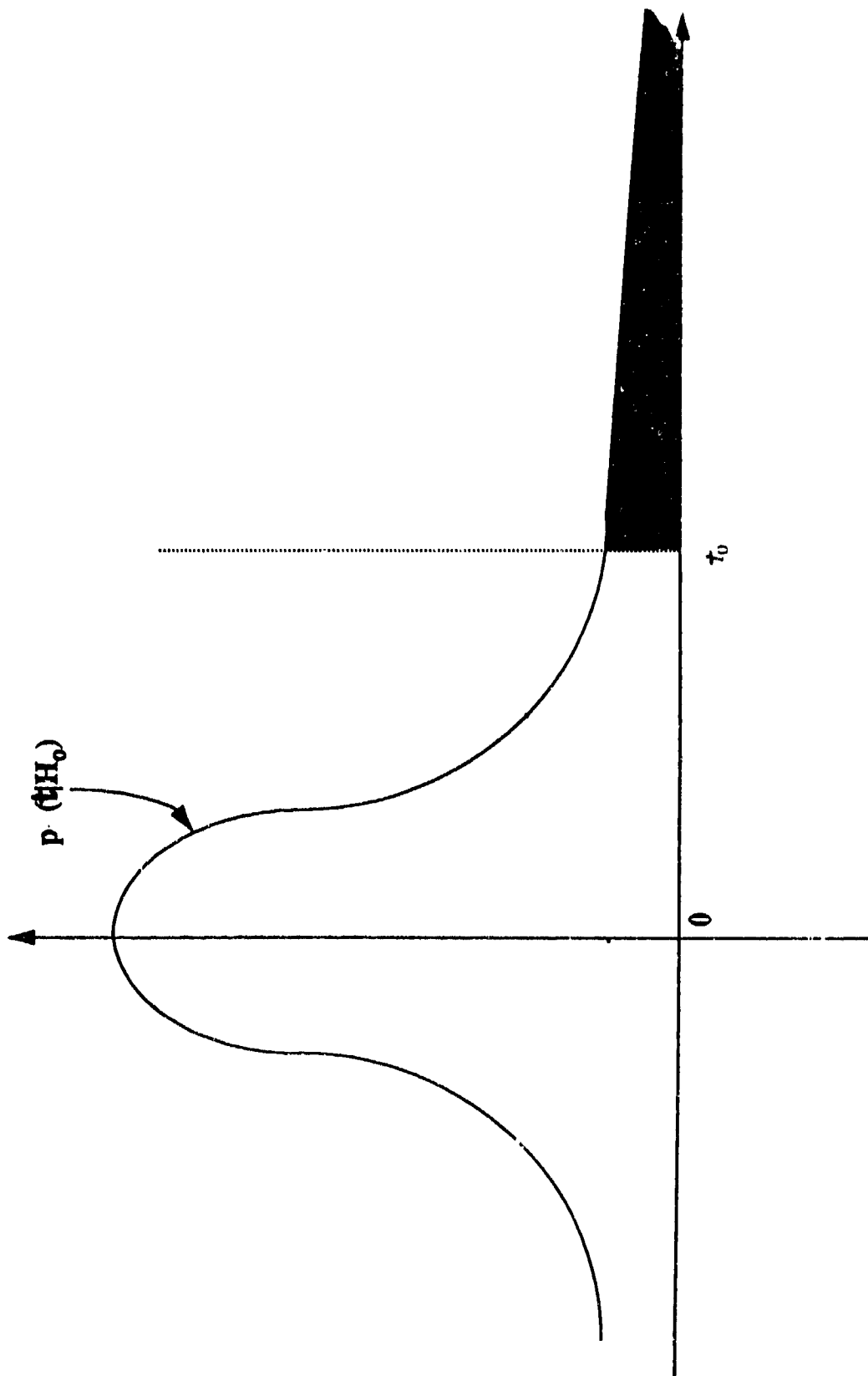


Figure 10.4: PDF of test statistic with tail region defined for $t \geq t_0$.

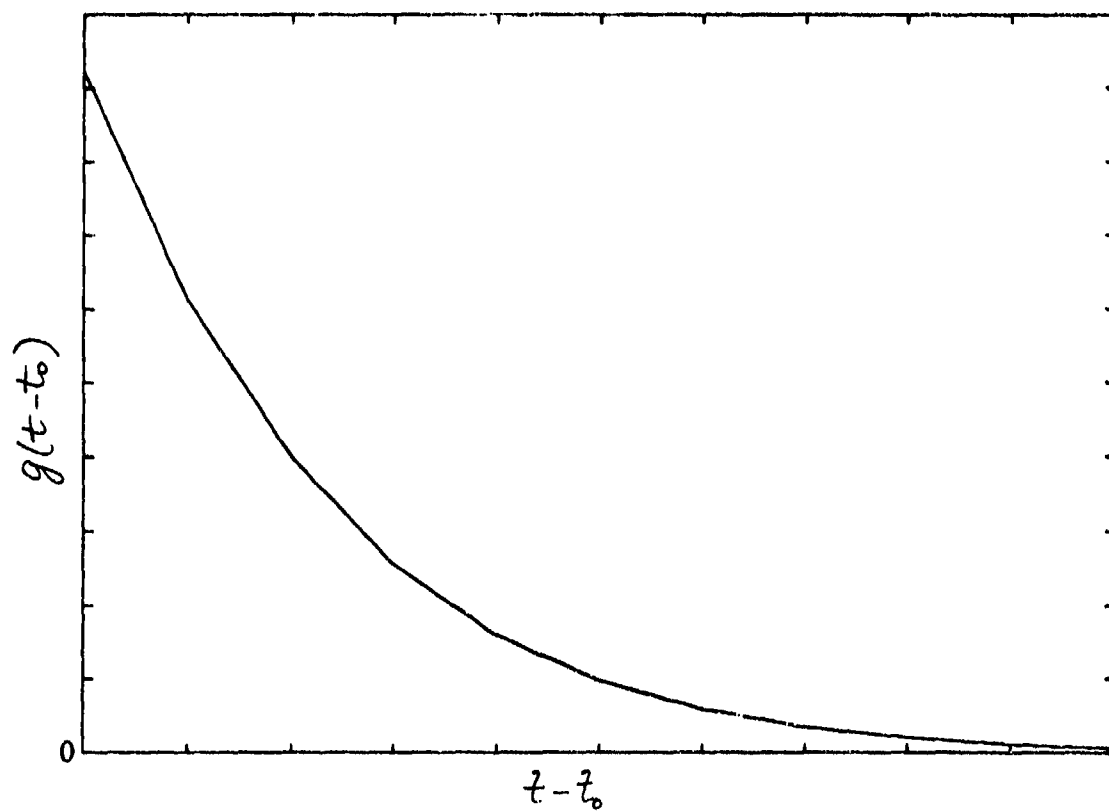


Figure 10.5: Tail of the test statistic shifted to origin.

of x_0 . One approach to selecting x_0 is to make a histogram of the data set and choose x_0 to be near the point of inflection of the histogram. DuMouchel [70] proposed choosing x_0 to be the value such that $\int_{x_0}^{\infty} f_X(x)dx = 0.1$. Such an approach is less subjective and appears to be satisfactory for many applications. However, it is noted by DuMouchel that "using an even smaller fraction of observations would restrict profitable use of the statistic to much larger sizes. On the other hand, to use more than the upper one tenth of a sample would seem to allow too much dependence on the central part of the distribution." In other words, if a smaller fraction is used, we need larger sample sizes to get an adequate number of samples for estimation and if a larger fraction is used, the body of the distribution may influence estimation of the tail.

Let x_0 be chosen as the value such that $1 - F(x_0) = \int_{x_0}^{\infty} f_X(x)dx = \alpha$. The distribution function to be used in approximating the tail can be written as

$$\tilde{F}(x) = (1 - \alpha) + \alpha G(x - x_0) = 1 - \alpha[1 + \frac{\gamma}{\sigma}(x - x_0)]^{-1/\gamma} \quad x > x_0 \quad (10.19)$$

where $G(x)$ is given in equation (10.15). Assuming that the tail of a given distribution can be approximated by equation (10.19), then the estimation problem of the distribution in the tail region is reduced to estimation of the parameters of the Generalized Pareto distribution.

In this chapter we consider three methods for the parameter estimation of the Generalized Pareto distribution. The three methods are maximum likelihood estimation, the method of probability weighted moments, and the ordered sample least squares approach. The first two methods, applied to the GPD, are discussed by Hosking and Wallis [71]. The ordered sample least squares approach is a new technique developed in this work. The performance of the three estimation procedures are compared on the basis of estimation bias and mean square error.

10.3.1.1 Maximum Likelihood Estimation

The probability density function corresponding to the GPD from equation (10.16), with x replaced by z , is

$$g(z) = \frac{1}{\sigma} \left(1 + \frac{\gamma z}{\sigma}\right)^{-\frac{1}{\gamma}-1}. \quad (10.20)$$

Given a sample vector $[z_1, z_2, \dots, z_m]$ from the GPD the joint density function $L_{\underline{z}}(\underline{z})$ of the m samples, assuming independence is given by

$$L_{\underline{z}}(\underline{z}) = \frac{1}{\sigma^m} \prod_{i=1}^m \left(1 + \frac{\gamma z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}. \quad (10.21)$$

To theoretically obtain the maximum likelihood estimates of σ and γ , the logarithm of the joint density function in equation (10.21) is differentiated with respect to σ and γ , respectively, and the derivatives are set to zero. Let the largest m observations from the unknown distribution whose tail is being modeled by GPD be placed in the vector $[x_{n-m+1}, x_{n-m+2}, \dots, x_n]$. Translation of the tail region to the origin results in the vector $[x_{n-m+1} - x_0, x_{n-m+2} - x_0, \dots, x_n - x_0] = [z_1, z_2, \dots, z_m]$. Letting $\tau = \gamma/\sigma$ in equation (10.21) and differentiating the logarithm of the joint density function with respect to σ we get

$$\begin{aligned} -\frac{d}{d\sigma} \log L_{\underline{Z}}(\underline{z}) &= \frac{d}{d\sigma} [m \log(\sigma) + (1 + (\tau\sigma)^{-1}) \sum_{i=1}^m \log(1 + \tau z_i)] \\ &= \frac{m}{\sigma} + (1 - \frac{1}{\sigma^2 \tau}) \sum_{i=1}^m \log(1 + \gamma z_i / \sigma). \end{aligned} \quad (10.22)$$

By setting equation (10.22) to zero, an expression for σ that satisfies the equation is

$$\sigma(\tau) = \sum_{i=1}^m \log(1 + \tau z_i) / (m\tau). \quad (10.23)$$

The expression for σ is now substituted into equation (10.22), so as to obtain a function of τ alone. $\hat{\tau}$ is derived by differentiating the quantity

$$m \log \sigma(\tau) + (1 + 1/(\sigma(\tau)\tau)) \sum_{i=1}^m \log(1 + \tau z_i) \quad (10.24)$$

with respect to τ and setting the derivative equal to zero with the constraint that $\tau z_i > -1$. However, the differentiation leads to a nonlinear equation whose analytical solution is not known. This difficulty is circumvented by minimizing equation (10.24) numerically with respect to τ . The numerical minimization was performed using the Nelder-Mead algorithm [72]. Once the estimate for τ has been obtained, then $\hat{\sigma}$ is obtained from equation (10.23) and γ is estimated by $\hat{\gamma} = \hat{\sigma} \hat{\tau}$.

10.3.1.2 Probability Weighted Moments

The probability weighted moments of a continuous random variable Z with distribution function G are the quantities

$$M_{p,r,s} = E[Z^p G^r(Z) (1 - G(Z))^s] \quad (10.25)$$

where E is the expectation operator and p , r and s are real numbers. For the GPD it is convenient to choose $p = 1$ and $r = 0$, respectively. Then the probability weighted moments are

$$M_{1,0,s} = E[Z(1 - G(Z))^s] \quad (10.26)$$

For the GPD there are two parameters to be estimated, σ and γ . Substituting $s = 0$, in equation (10.26), we get

$$\epsilon_0 = M_{1,0,0} = E[Z] = \int_0^\infty \frac{Z}{\sigma} \left(1 + \frac{\gamma Z}{\sigma}\right)^{-\frac{1}{\gamma}-1} dZ. \quad (10.27)$$

Letting $1 + \frac{\gamma Z}{\sigma} = Y$, equation (10.27) results in

$$\begin{aligned} \epsilon_0 &= \frac{\sigma}{\gamma^2} \int_1^\infty (Y-1) Y^{-\frac{1}{\gamma}-1} dY \\ &= \frac{\sigma}{\gamma^2} \left[\frac{Y^{-\frac{1}{\gamma}-1}}{-\frac{1}{\gamma}+1} - \frac{Y^{-\frac{1}{\gamma}}}{-\frac{1}{\gamma}} \right]_1^\infty \\ &= \frac{\sigma}{1-\gamma}. \end{aligned} \quad (10.28)$$

Letting $s = 1$ in equation (10.26) we obtain

$$\epsilon_1 = M_{1,0,1} = E[Z(1 - G(Z))] = \int_0^\infty \frac{Z}{\sigma} \left(1 + \frac{\gamma Z}{\sigma}\right)^{-\frac{1}{\gamma}} \left(1 + \frac{\gamma Z}{\sigma}\right)^{-\frac{1}{\gamma}-1} dZ. \quad (10.29)$$

Letting $1 + \frac{\gamma Z}{\sigma} = Y$, as before, equation (10.29) results in

$$\begin{aligned} \epsilon_1 &= \frac{\sigma}{\gamma^2} \int_1^\infty (Y-1) Y^{-\frac{2}{\gamma}-1} dY \\ &= \frac{\sigma}{\gamma^2} \left[\frac{Y^{-\frac{2}{\gamma}-1}}{-\frac{2}{\gamma}+1} - \frac{Y^{-\frac{2}{\gamma}}}{-\frac{2}{\gamma}} \right]_1^\infty \\ &= \frac{\sigma}{2(2-\gamma)}. \end{aligned} \quad (10.30)$$

The values of ϵ_0 and ϵ_1 are obtained from equations (10.28) and (10.30), respectively, for given values of σ and γ . Since there are two equations in two unknowns σ and γ can be obtained as functions of ϵ_0 and ϵ_1 . Solving for σ and γ we obtain

$$\hat{\sigma} = 2\epsilon_0\epsilon_1/(\epsilon_0 - 2\epsilon_1) \quad (10.31)$$

and

$$\hat{\gamma} = 2 - \epsilon_0 / (\epsilon_0 - 2\epsilon_1) \quad (10.32)$$

where ϵ_0 and ϵ_1 are estimated from the data by the estimators $\hat{\epsilon}_0 = \sum_{i=1}^m z_i / m$ and $\hat{\epsilon}_1 = \sum_{i=1}^m (m-i)z_i / \{m(m-1)\}$ [71]. Once the values of $\hat{\epsilon}_0$ and $\hat{\epsilon}_1$ are obtained the estimates of σ and γ are obtained by making use of equations (10.31) and (10.32). Note that the method of probability weighted moments involves computationally simple expression for the estimates.

10.3.1.3 The Ordered Sample Least Squares Method - A new approach

The procedure used in maximum likelihood estimation is based on minimizing the quantity in equation (10.24). Similarly, the probability weighted moment estimates are obtained by equating with the sample based values the theoretical values of the quantity $E[Z(1-G(Z))^s]$, $s=0,1$, where $Z = X - x_0$. The ordered sample least squares method is based on the principle of minimizing the squared distance between the ordered sample and the expected value of the ordered sample. Computer simulations reveal that this can be a more suitable approach for estimating the parameters.

In Appendix C the method for evaluating the mean and the variance of the r^{th} ordered statistic from a sample size n is presented. For the Generalized Pareto Distribution the mean and the variance of the r^{th} order statistic can be derived since the probability distribution function is known in closed form. Let x be replaced by z in equation (10.15) and let $G(z) = u$. Solution for z results in

$$z = G^{-1}(u) = \frac{\sigma}{\gamma} [(1-u)^{-\gamma} - 1]. \quad (10.33)$$

Making use of the above equation and equation (C.62) in Appendix A, the expected value of Z_r is

$$E(Z_r) = \frac{\sigma}{\gamma} \frac{n!}{(r-1)!(n-r)!} \left[\int_0^1 ((1-u)^{-\gamma} - 1) u^{r-1} (1-u)^{n-r} du \right]. \quad (10.34)$$

The integral in the above equation can be broken into two parts as follows.

$$E(Z_r) = \frac{\sigma}{\gamma} \frac{n!}{(r-1)!(n-r)!} \left[\int_0^1 (1-u)^{-\gamma} u^{r-1} (1-u)^{n-r} du - \int_0^1 u^{r-1} (1-u)^{n-r} du \right]. \quad (10.35)$$

From results presented in Gradshtyn and Ryzhik [45], the expression for $E(Z_r)$ becomes

$$E(Z_r) = \frac{\sigma}{\gamma} \frac{n!}{(r-1)!(n-r)!} \left[\frac{(r-1)!(n-r-\gamma)!}{(n-\gamma)!} - \frac{(r-1)!(n-r)!}{n!} \right]$$

$$\begin{aligned}
&= \frac{\sigma}{\gamma} \left[\frac{n!(n-r-\gamma)!}{(n-r)!(n-\gamma)!} - 1 \right] \\
&= \frac{\sigma}{\gamma} \left[\frac{\Gamma(n+1)\Gamma(n-r-\gamma+1)}{\Gamma(n-r+1)\Gamma(n-\gamma+1)} - 1 \right].
\end{aligned} \tag{10.36}$$

To calculate the variance of Z_r , we first calculate $E(Z_r^2)$. Making use of equation (10.33) and equation (C.65) in Appendix C, the expected value of Z_r^2 is

$$E(Z_r^2) = \frac{\sigma^2}{\gamma^2} \frac{n!}{(r-1)!(n-r)!} \left[\int_0^1 ((1-u)^{-\gamma} - 1)^2 u^{r-1} (1-u)^{n-r} du \right]. \tag{10.37}$$

The integral in the above equation can be rewritten as follows:

$$E(Z_r^2) = \frac{\sigma^2}{\gamma^2} \frac{n!}{(r-1)!(n-r)!} \left[\int_0^1 ((1-u)^{-2\gamma} - 2(1-u)^{-\gamma} + 1) u^{r-1} (1-u)^{n-r} du \right]. \tag{10.38}$$

Making use of results from [73], the above integral evaluates to

$$\begin{aligned}
E(Z_r^2) &= \frac{\sigma^2}{\gamma^2} \frac{n!}{(n-r)!} \left[\frac{(n-r-2\gamma)!}{(n-2\gamma)!} - \frac{2(n-r-\gamma)!}{(n-\gamma)!} + 1 \right] \\
&= \frac{\sigma^2}{\gamma^2} \frac{\Gamma(n+1)}{\Gamma(n-r+1)} \left[\frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} - \frac{2\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)} + 1 \right].
\end{aligned} \tag{10.39}$$

From equations (10.36) and (10.39) and using the result $Var(Z_r) = E(Z_r^2) - E^2(Z_r)$, we have

$$\begin{aligned}
Var(Z_r) &= \frac{\sigma^2}{\gamma^2} \frac{\Gamma(n+1)}{\Gamma(n-r+1)} \left[\frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} \right. \\
&\quad \left. - \frac{2\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)} + 1 \right] - \left\{ \frac{\sigma}{\gamma} \left[\frac{\Gamma(n+1)\Gamma(n-r-\gamma+1)}{\Gamma(n-r+1)\Gamma(n-\gamma+1)} - 1 \right] \right\}^2.
\end{aligned} \tag{10.40}$$

Simplifying the above equation results in

$$Var(Z_r) = \frac{\sigma^2}{\gamma^2} \left[\frac{\Gamma(n+1)}{\Gamma(n-r+1)} \frac{\Gamma(n-r-2\gamma+1)}{\Gamma(n-2\gamma+1)} - \frac{\Gamma^2(n+1)}{\Gamma^2(n-r+1)} \frac{\Gamma^2(n-r-\gamma+1)}{\Gamma^2(n-\gamma+1)} \right]. \tag{10.41}$$

Letting $Q_r(\gamma) = \frac{\Gamma(n+1)}{\Gamma(n-r+1)} \frac{\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)}$, results in

$$E(Z_r) = \mu_r = \frac{\sigma}{\gamma} \{Q_r(\gamma) - 1\} \tag{10.42}$$

$$Var(Z_r) = \sigma_r^2 = \frac{\sigma^2}{\gamma^2} \{Q_r(2\gamma) - (Q_r(\gamma))^2\}. \tag{10.43}$$

A computationally simpler expression can be found for $Q_r(\gamma)$ by making use of the properties of gamma functions. Dividing $Q_r(\gamma)$ by $Q_{r-1}(\gamma)$ we get

$$\frac{Q_r(\gamma)}{Q_{r-1}(\gamma)} = \frac{\frac{\Gamma(n+1)}{\Gamma(n-r+1)} \frac{\Gamma(n-r-\gamma+1)}{\Gamma(n-\gamma+1)}}{\frac{\Gamma(n+1)}{\Gamma(n-r+2)} \frac{\Gamma(n-r-\gamma+2)}{\Gamma(n-\gamma+1)}} = \frac{n-r+1}{n-r-\gamma+1}. \quad (10.44)$$

Equation (10.44) reduces to

$$Q_r(\gamma) = \prod_{i=1}^r (n-i+1)/(n-i+1-\gamma). \quad (10.45)$$

To find the least squares estimates of the parameters we write the following non-linear model for the r^{th} sample order statistic

$$Z_r = E(Z_r) + e_r, \quad r = 1, 2, \dots, m \quad (10.46)$$

where the error term e_r has a distribution with mean 0 and variance σ_r^2 . Since the order statistics are not independent, the errors are also not independent. Because of the non-linear structure of the model in equation (10.46) and correlated errors, least squares estimation does not offer a straightforward solution to the estimation problem. Even so, in this study we proceed to use the ordered sample least squares (OSLS) procedure to estimate the parameters.

In equation (10.42), we note that the scale parameter σ appears linearly whereas the shape parameter γ does not. The least squares estimates are obtained by minimizing the quantity

$$S = \sum_{r=1}^m e_r^2 = \sum_{r=1}^m (Z_r - \sigma(Q_r(\gamma) - 1)/\gamma)^2. \quad (10.47)$$

Since σ appears linearly in the above expression, minimization can be achieved analytically. Differentiating equation (10.47) with respect to σ and setting the derivative equal to zero results in

$$2 \sum_{r=1}^m (Z_r - \frac{\sigma}{\gamma}(Q_r(\gamma) - 1))(-\frac{1}{\gamma}(Q_r(\gamma) - 1)) = 0. \quad (10.48)$$

The solution for σ from the above equation is

$$\hat{\sigma}(\gamma) = \gamma \frac{\sum_{r=1}^m Z_r(Q_r(\gamma) - 1)}{\sum_{r=1}^m (Q_r(\gamma) - 1)^2}. \quad (10.49)$$

The expression for $\hat{\sigma}$ is substituted in equation (10.47) and the resulting expression is minimized

with respect to γ . The resulting expression after the substitution is nonlinear and minimization cannot be performed analytically. Using the Nelder-Mead algorithm [72], the minimization is done numerically. Once the estimate of $\hat{\gamma}$ is obtained, $\hat{\sigma}$ is obtained from equation (10.49).

Recall that the GPD is being used to approximate the tail of the underlying distribution. Hence, the ordered statistics $Z_r, r = 1, 2, \dots, m$, from the GPD actually correspond to the ordered statistics $X_{n-m+1} - x_0, X_{n-m+2} - x_0, \dots, X_n - x_0$ from the underlying distribution.

The least squares procedure results in a computationally convenient algorithm. It is emphasized that the minimization of S is carried out only with respect to the single parameter γ . Furthermore, the underlying criterion is based on minimizing the distance between the empirical values and the expected values of the ordered samples. Some numerical comparisons are given in section 10.4.

10.3.2 Estimation of Thresholds

The Generalized Pareto Distribution that is estimated from the data is used to approximate the tail of the unknown, underlying distribution. We now show that the threshold is related to the approximating distribution function in a direct manner. With reference to equation (10.19), let $\hat{\eta}_p$ denote the threshold estimate of the threshold corresponding to a false alarm probability p . We then have

$$\tilde{F}(\hat{\eta}_p) = 1 - p = 1 - \alpha[1 + \frac{\gamma}{\sigma}(\hat{\eta}_p - x_0)]^{-1/\gamma}. \quad (10.50)$$

Solution for $\hat{\eta}_p$ results in

$$\hat{\eta}_p = x_0 + \sigma(q^{-\gamma} - 1)/\gamma \quad (10.51)$$

where $\alpha = 1 - F(x_0)$, $q = (1 - p)/\alpha$ and $x_0 = F^{-1}(1 - \alpha)$. For many applications DuMouchel [70] suggests that $\alpha = 0.1$ be used. As will be discussed in the subsequent sections, the optimal value of α depends on the threshold being estimated. Since the distribution function $F(x)$ is not known, x_0 cannot be determined for a given value of α . Therefore, following common practice, the sample order statistic X_{n-m} , where $m = [\alpha n]$ and $[.]$ denotes the integer part operator, is used as an estimate of x_0 .

10.4 Numerical Results

10.4.1 Characterization of Tail Shape for Known Distributions

We first discuss a method for estimating the parameters of the GPD when the underlying distribution is known. Choose x_0 such that $1 - F(x_0) = 0.1$. Then define the points p_i $i=1,2,\dots,1000$ by

$$p_i = 0.90005 + 0.0001(i - 1). \quad (10.52)$$

Analytically evaluate $x_i = F^{-1}(p_i)$ from the known distribution. Using the 1000 values of x_i , the maximum likelihood estimation, the ordered sample least squares and the probability weighted moments procedures were applied to determine the corresponding γ values for various distributions. The results are given in Table 10.1. The number in parentheses for the Weibull and Lognormal distributions is the value of the shape parameter. For the remaining distributions the number denotes the degrees of freedom. Since σ is a scale parameter, the shape parameter γ best describes the tail shape. For the exponential and the uniform distributions the value of γ can be obtained theoretically. $\gamma = 0$ for the exponential distribution and is -1 for the uniform distribution. Since the size of the tail decreases with decreasing γ , the relationship between the tail behavior and the corresponding values of the shape parameter γ can be clearly inferred from this table.

Distribution	OSLS	ML	PWM
Gaussian	-0.144	-0.151	-0.174
Weibull(3)	-0.163	-0.168	-0.194
Weibull(.67)	0.108	0.129	0.137
Weibull(.5)	0.201	0.265	0.263
Student-t(3)	0.290	0.260	0.261
Student-t(5)	0.132	0.099	0.090
Student-t(8)	0.031	0.006	-0.010
Lognormal(1)	0.232	0.259	0.258
Chi-square(1)	0.030	0.034	0.044
Chi-square(4)	-0.024	-0.033	-0.034
Chi-square(8)	-0.047	-0.058	-0.064

Table 10.1: Tail parameter γ describing the upper ten percent of various distributions.

10.4.2 Empirical Properties of the Estimators for Known Distributions

Seven distributions with widely differing tail behaviors were chosen in order to investigate the adequacy of the approximation of extreme tails by the GPD and to compare the three estimation procedures. The gamma distribution and Weibull distribution with shape parameter of value 3 have tails lighter than those of the exponential PDF. The tails of the chi-square distribution with 4 degrees of freedom and the student-T distribution with 8 degrees of freedom are approximately the same as those of the exponential PDF. Finally, the student-T distribution with 4 degrees of freedom and the Lognormal distribution with shape parameter of value 1 have tails heavier than those of the exponential PDF.

Let η and $\hat{\eta}$ denote the true and estimated thresholds, respectively. A Monte Carlo experiment was performed to investigate the normalized bias, $\frac{\hat{\eta} - \eta}{\eta}$ and the normalized mean square error $(\frac{\hat{\eta} - \eta}{\eta})^2$ of the proposed threshold estimates. The four sample sizes given by $m = 25, 50, 100$ and 1000 were considered. Each set of samples was obtained by generating n observations and taking the largest $m = 0.1n$ observations. For example, a set of samples of size 25 was obtained by selecting the largest 25 observations from a collection of size 250 samples. For all four different values of m , $k=200,000/m$ trials were performed for each of the seven distributions. The median of the normalized bias values were computed for each distribution and estimation procedure. The results for $P_F = 10^{-k}$, $k=2,3,\dots,7$ are given in Table 10.2. Similarly the median of the positive square root of the normalized mean square error are presented in Table 10.3. The results in the two tables differ because the sign of $(\hat{\eta} - \eta)/\eta$ is lost in the normalized root mean square values computed in Table 10.3. Extremely poor estimates for η were obtained in some of the trials.

		m=25					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	-0.0112	0.0043	-0.0040	-0.0276	-0.0571	-0.0872
Normal	ML	-0.0034	0.0187	0.0328	0.0358	0.0281	0.0137
Normal	PWM	-0.0084	-0.0208	-0.0560	-0.1015	-0.1464	-0.1924
Weibull(3)	OSLS	-0.0048	0.0013	-0.0041	-0.0202	-0.0418	-0.0619
Weibull(3)	ML	0.0039	0.0481	0.0938	0.1374	0.1776	0.2137
Weibull(3)	PWM	-0.0037	-0.0106	-0.0333	-0.0635	-0.0919	-0.1216
t(4)	OSLS	-0.0424	-0.0792	-0.1658	-0.2727	-0.3872	-0.4922
t(4)	ML	-0.0166	-0.1115	-0.2526	-0.4045	-0.5416	-0.6541
t(4)	PWM	-0.0218	-0.0929	-0.2160	-0.3498	-0.4761	-0.5881
t(8)	OSLS	-0.0221	-0.0186	-0.0572	-0.1164	-0.1975	-0.2879
t(8)	ML	-0.0104	-0.0468	-0.1169	-0.2077	-0.3055	-0.4033
t(8)	PWM	-0.0129	-0.0452	-0.1095	-0.2039	-0.3063	-0.4115
Chi-sq(4)	OSLS	-0.0209	-0.0039	0.0241	0.0333	-0.0088	-0.0104
Chi-sq(4)	ML	-0.0037	0.0943	0.2518	0.4571	0.6185	0.8810
Chi-sq(4)	PWM	-0.0144	-0.0205	-0.0334	-0.0576	-0.1254	-0.1624
Lognormal	OSLS	-0.0835	-0.0982	-0.0634	0.0016	0.1007	0.2567
Lognormal	ML	-0.0058	0.1836	0.5932	1.2736	2.4832	4.4947
Lognormal	PWM	-0.0543	-0.0878	-0.0931	-0.0728	-0.0228	0.0639
Pareto(-0.25)	OSLS	-0.0092	0.0208	0.0423	0.0631	0.0780	0.0874
Pareto(-0.25)	ML	-0.0030	0.0523	0.1190	0.1868	0.2479	0.2969
Pareto(-0.25)	PWM	-0.0077	0.0052	0.0121	0.0199	0.0237	0.0278

Table 10.2: Median of the normalized bias values for different percentiles. OSLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments

		m=50					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0036	0.0073	-0.0068	-0.0354	-0.0676	-0.1022
Normal	ML	0.0042	0.0323	0.0497	0.0578	0.0528	0.0380
Normal	PWM	-0.0012	-0.0118	-0.0459	-0.0861	-0.1318	-0.1742
Weibull(3)	OSLS	-0.0022	-0.0007	-0.0133	-0.0337	-0.0571	-0.0838
Weibull(3)	ML	0.0056	0.0500	0.0991	0.1436	0.1847	0.2199
Weibull(3)	PWM	-0.0014	-0.0105	-0.0342	-0.0629	-0.0937	-0.1256
t(4)	OSLS	-0.0147	-0.0646	-0.1800	-0.3209	-0.4501	-0.5063
t(4)	ML	-0.0068	-0.0867	-0.2264	-0.3736	-0.5120	-0.6291
t(4)	PWM	-0.0078	-0.0622	-0.1662	-0.2973	-0.4233	-0.5391
t(8)	OSLS	-0.0062	-0.0222	-0.0841	-0.1723	-0.2694	-0.3703
t(8)	ML	-0.0031	-0.0502	-0.1352	-0.2385	-0.3460	-0.4517
t(8)	PWM	-0.0032	-0.0336	-0.1064	-0.2041	-0.3051	-0.4046
Chi-sq(4)	OSLS	-0.0092	-0.0004	0.0051	0.0060	-0.0498	-0.0686
Chi-sq(4)	ML	0.0115	0.1134	0.2755	0.4775	0.6368	0.9150
Chi-sq(4)	PWM	-0.0041	-0.0087	-0.0191	-0.0407	-0.1123	-0.1488
Lognormal	OSLS	-0.0544	-0.0594	-0.0272	0.0458	0.1573	0.3274
Lognormal	ML	0.0092	0.2177	0.6336	1.3811	2.6197	4.7101
Lognormal	PWM	-0.0302	-0.0391	-0.0185	0.0413	0.1480	0.2977
Pareto(-0.25)	OSLS	-0.0052	0.0100	0.0214	0.0326	0.0404	0.0448
Pareto(-0.25)	ML	0.0005	0.0463	0.1011	0.1560	0.2003	0.2357
Pareto(-0.25)	PWM	-0.0050	-0.0018	-0.0012	-0.0019	-0.0023	-0.0012

Table 10.2: Median of the normalized bias values for different percentiles. (contd.)

		m=100					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0017	-0.0016	-0.0253	-0.0637	-0.1040	-0.1464
Normal	ML	0.0068	0.0263	0.0306	0.0229	0.0063	-0.0185
Normal	PWM	0.0018	-0.0181	-0.0549	-0.1022	-0.1524	-0.1986
Weibull(3)	OSLS	0.0005	-0.0017	-0.0164	-0.0376	-0.0624	-0.0888
Weibull(3)	ML	0.0037	0.0270	0.0564	0.0840	0.1003	0.1158
Weibull(3)	PWM	0.0004	-0.0095	-0.0320	-0.0607	-0.0918	-0.1220
t(4)	OSLS	-0.0064	-0.0441	-0.1421	-0.2680	-0.3922	-0.5031
t(4)	ML	-0.0004	-0.0564	-0.1650	-0.2907	-0.4174	-0.5354
t(4)	PWM	-0.0003	-0.0478	-0.1403	-0.2636	-0.3809	-0.4949
t(8)	OSLS	-0.0024	-0.0134	-0.0751	-0.1606	-0.2578	-0.3548
t(8)	ML	0.0011	-0.0342	-0.1145	-0.2123	-0.3157	-0.4216
t(8)	PWM	0.0013	-0.0271	-0.0955	-0.1888	-0.2892	-0.3916
Chi-sq(4)	OSLS	-0.0032	-0.0028	-0.0077	-0.0198	-0.0841	-0.1111
Chi-sq(4)	ML	0.0175	0.1189	0.2655	0.4581	0.5917	0.8298
Chi-sq(4)	PWM	-0.0004	-0.0089	-0.0238	-0.0448	-0.1143	-0.1520
Lognormal	OSLS	-0.0159	-0.0542	-0.0876	-0.1089	-0.0940	-0.0617
Lognormal	ML	-0.0111	-0.0251	-0.0068	0.0536	0.1499	0.3104
Lognormal	PWM	-0.0165	-0.0210	0.0141	0.0924	0.2315	0.3965
Pareto(-0.25)	OSLS	-0.0023	0.0109	0.0255	0.0350	0.0419	0.0471
Pareto(-0.25)	ML	0.0033	0.0544	0.1170	0.1739	0.2215	0.2611
Pareto(-0.25)	PWM	-0.0014	0.0004	0.0052	0.0084	0.0112	0.0129

Table 10.2: Median of the normalized bias values for different percentiles. (contd.)

		m=1000					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0035	-0.0013	-0.0244	-0.0613	-0.1010	-0.1432
Normal	ML	0.0059	0.0017	-0.0259	-0.0626	-0.1075	-0.1476
Normal	PWM	0.0028	-0.0192	-0.0586	-0.1064	-0.1560	-0.2016
Weibull(3)	OSLS	0.0013	-0.0023	-0.0175	-0.0381	-0.0627	-0.0885
Weibull(3)	ML	0.0020	-0.0018	-0.0159	-0.0386	-0.0641	-0.0909
Weibull(3)	PWM	0.0010	-0.0092	-0.0297	-0.0578	-0.0880	-0.1192
t(4)	OSLS	0.0058	-0.0044	-0.0605	-0.1574	-0.2690	-0.3715
t(4)	ML	0.0141	-0.0137	-0.1018	-0.2167	-0.3326	-0.4406
t(4)	PWM	0.0141	-0.0176	-0.1104	-0.2277	-0.3479	-0.4598
t(8)	OSLS	0.0033	-0.0021	-0.0452	-0.1167	-0.2001	-0.2896
t(8)	ML	0.0070	-0.0117	-0.0664	-0.1464	-0.2404	-0.3382
t(8)	PWM	0.0045	-0.0219	-0.0896	-0.1825	-0.2857	-0.3862
Chi-sq(4)	OSLS	0.0003	0.0012	-0.0057	-0.0167	-0.0826	-0.1107
Chi-sq(4)	ML	0.0012	-0.0021	-0.0152	-0.0354	-0.1026	-0.1349
Chi-sq(4)	PWM	0.0006	-0.0011	-0.0080	-0.0263	-0.0934	-0.1211
Lognormal	OSLS	-0.0038	-0.0221	-0.0259	0.0055	0.0646	0.1638
Lognormal	ML	-0.0098	0.0063	0.0616	0.1767	0.3495	0.5999
Lognormal	PWM	-0.0128	-0.0004	0.0567	0.1683	0.3400	0.5771
Pareto(-0.25)	OSLS	0.0002	0.0002	0.0012	0.0007	0.0003	0.0000
Pareto(-0.25)	ML	-0.0002	-0.0010	-0.0044	-0.0061	-0.0081	-0.0094
Pareto(-0.25)	PWM	0.0003	-0.0011	-0.0007	-0.0006	-0.0035	-0.0038

Table 10.2: Median of the normalized bias values for different percentiles.

		m=25					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0558	0.1127	0.2022	0.2825	0.3507	0.4044
Normal	ML	0.0558	0.0909	0.1459	0.2057	0.2588	0.3070
Normal	PWM	0.0559	0.1215	0.2121	0.2920	0.3586	0.4117
Weibull(3)	OSLS	0.0257	0.0577	0.1089	0.1580	0.2031	0.2415
Weibull(3)	ML	0.0258	0.0531	0.0950	0.1378	0.1780	0.2139
Weibull(3)	PWM	0.0256	0.0624	0.1149	0.1659	0.2110	0.2495
t(4)	OSLS	0.1069	0.2261	0.4160	0.5989	0.7397	0.8405
t(4)	ML	0.1051	0.2353	0.4157	0.5812	0.7127	0.8097
t(4)	PWM	0.1019	0.2329	0.4213	0.5956	0.7368	0.8344
t(8)	OSLS	0.0781	0.1666	0.3073	0.4455	0.5701	0.6730
t(8)	ML	0.0779	0.1493	0.2554	0.3648	0.4689	0.5649
t(8)	PWM	0.0775	0.1752	0.3180	0.4544	0.5783	0.6787
Chi-sq(4)	OSLS	0.0610	0.1313	0.2441	0.3592	0.4650	0.5455
Chi-sq(4)	ML	0.0721	0.2179	0.4459	0.7901	1.1783	1.7789
Chi-sq(4)	PWM	0.0592	0.1384	0.2500	0.3622	0.4666	0.5446
Lognormal	OSLS	0.1335	0.2452	0.4362	0.6271	0.7785	0.8785
Lognormal	ML	0.1439	0.4007	0.7303	1.4149	2.7312	5.0774
Lognormal	PWM	0.1260	0.2582	0.4463	0.6281	0.7737	0.8705
Pareto(-0.25)	OSLS	0.0409	0.0787	0.1348	0.1752	0.2017	0.219
Pareto(-0.25)	ML	0.0402	0.0763	0.1419	0.2075	0.2640	0.3127
Pareto(-0.25)	PWM	0.0411	0.0866	0.1430	0.1817	0.2084	0.2240

Table 10.3: Median RMS errors for various percentiles. OSLS:Ordered Sample Least Square, ML:Maximum Likelihood, PWM:Probability Weighted Moments

		m=50					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0401	0.0772	0.1391	0.1981	0.2548	0.3042
Normal	ML	0.0394	0.0689	0.1122	0.1559	0.1959	0.2328
Normal	PWM	0.0399	0.0865	0.1530	0.2192	0.2759	0.3273
Weibull(3)	OSLS	0.0180	0.0393	0.0743	0.1135	0.1511	0.1854
Weibull(3)	ML	0.0185	0.0509	0.0997	0.1447	0.1859	0.2214
Weibull(3)	PWM	0.0180	0.0442	0.0852	0.1263	0.1661	0.2017
t(4)	OSLS	0.0779	0.1826	0.3506	0.5179	0.6633	0.7724
t(4)	ML	0.0768	0.1910	0.3602	0.5244	0.6688	0.7762
t(4)	PWM	0.0760	0.1778	0.3332	0.4899	0.6303	0.7386
t(8)	OSLS	0.0561	0.1228	0.2316	0.3503	0.4666	0.5698
t(8)	ML	0.0553	0.1219	0.2226	0.3385	0.4504	0.5529
t(8)	PWM	0.0554	0.1306	0.2405	0.3613	0.4793	0.5807
Chi-sq(4)	OSLS	0.0431	0.0890	0.1678	0.2509	0.3351	0.4109
Chi-sq(4)	ML	0.0489	0.1661	0.3386	0.5487	0.7664	1.1112
Chi-sq(4)	PWM	0.0426	0.0939	0.1747	0.2584	0.3431	0.4185
Lognormal	OSLS	0.0975	0.1834	0.3439	0.5155	0.6660	0.7990
Lognormal	ML	0.0993	0.3381	0.6769	1.3921	2.6297	4.7240
Lognormal	PWM	0.0864	0.1954	0.3510	0.5143	0.6621	0.8012
Pareto(-0.25)	OSLS	0.0289	0.0534	0.0890	0.1162	0.1346	0.1486
Pareto(-0.25)	ML	0.0284	0.0602	0.1149	0.1675	0.2084	0.2417
Pareto(-0.25)	PWM	0.0293	0.0616	0.1032	0.1320	0.1533	0.1666

Table 10.3: Median RMS errors for various percentiles. (contd.)

		m=100					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0284	0.0522	0.0964	0.1414	0.1863	0.2305
Normal	ML	0.0290	0.0517	0.0840	0.1123	0.1433	0.1689
Normal	PWM	0.0281	0.0584	0.1090	0.1635	0.2134	0.2585
Weibull(3)	OSLS	0.0128	0.0273	0.0529	0.0811	0.1101	0.1378
Weibull(3)	ML	0.0131	0.0389	0.0790	0.1202	0.1570	0.1868
Weibull(3)	PWM	0.0126	0.0312	0.0622	0.0942	0.1284	0.1590
t(4)	OSLS	0.0550	0.1400	0.2801	0.4336	0.5739	0.6909
t(4)	ML	0.0525	0.1377	0.2716	0.4165	0.5497	0.6627
t(4)	PWM	0.0527	0.1334	0.2619	0.4046	0.5323	0.6469
t(8)	OSLS	0.0386	0.0914	0.1770	0.2761	0.3758	0.4732
t(8)	ML	0.0388	0.0896	0.1735	0.2710	0.3734	0.4763
t(8)	PWM	0.0384	0.0869	0.1727	0.2750	0.3777	0.4817
Chi-sq(4)	OSLS	0.0287	0.0649	0.1264	0.1932	0.2699	0.3373
Chi-sq(4)	ML	0.0350	0.1437	0.2959	0.4688	0.6092	0.8592
Chi-sq(4)	PWM	0.0283	0.0686	0.1289	0.1948	0.2730	0.3383
Lognormal	OSLS	0.0683	0.1527	0.2794	0.4174	0.5299	0.6290
Lognormal	ML	0.0652	0.1515	0.2690	0.4039	0.5465	0.6769
Lognormal	PWM	0.0647	0.1417	0.2519	0.3805	0.5218	0.6710
Pareto(-0.25)	OSLS	0.0201	0.0372	0.0637	0.0845	0.0997	0.1110
Pareto(-0.25)	ML	0.0197	0.0569	0.1192	0.1746	0.2221	0.2613
Pareto(-0.25)	PWM	0.0201	0.0434	0.0718	0.0952	0.1108	0.1220

Table 10.3: Median RMS errors for various percentiles. (contd.)

		m=1000					
P_F		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Normal	OSLS	0.0077	0.0182	0.0373	0.0643	0.1017	0.1440
Normal	ML	0.0087	0.0160	0.0362	0.0632	0.1075	0.1476
Normal	PWM	0.0081	0.0247	0.0586	0.1064	0.1560	0.2016
Weibull(3)	OSLS	0.0037	0.0086	0.0194	0.0393	0.0630	0.0890
Weibull(3)	ML	0.0040	0.0078	0.0191	0.0397	0.0649	0.0909
Weibull(3)	PWM	0.0036	0.0108	0.0300	0.0578	0.0880	0.1192
t(4)	OSLS	0.0203	0.0534	0.1383	0.2476	0.3717	0.4763
t(4)	ML	0.0213	0.0447	0.1083	0.2168	0.3326	0.4406
t(4)	PWM	0.0213	0.0499	0.1207	0.2306	0.3479	0.4598
t(8)	OSLS	0.0135	0.0298	0.0726	0.1379	0.2121	0.3018
t(8)	ML	0.0129	0.0272	0.0750	0.1518	0.2436	0.3406
t(8)	PWM	0.0129	0.0349	0.0939	0.1830	0.2863	0.3863
Chi-sq(4)	OSLS	0.0104	0.0207	0.0362	0.0588	0.1094	0.1408
Chi-sq(4)	ML	0.0099	0.0192	0.0363	0.0589	0.1095	0.1429
Chi-sq(4)	PWM	0.0100	0.0211	0.0400	0.0602	0.1103	0.1433
Lognormal	OSLS	0.0206	0.0528	0.1222	0.1836	0.2429	0.3276
Lognormal	ML	0.0195	0.0434	0.0984	0.2012	0.3581	0.5999
Lognormal	PWM	0.0201	0.0410	0.0927	0.1919	0.3445	0.5770
Pareto(-0.25)	OSLS	0.0061	0.0101	0.0158	0.0213	0.0247	0.0278
Pareto(-0.25)	ML	0.0063	0.0092	0.0154	0.0198	0.0243	0.0268
Pareto(-0.25)	PWM	0.0065	0.0126	0.0222	0.0306	0.0375	0.0428

Table 10.3: Median RMS errors for various percentiles.

These poor estimates could severely influence an arithmetic mean of the estimates. To avoid this problem, median values were used in place of the arithmetic means.

The empirical results in Table 10.2 indicate that the newly proposed ordered sample least squares estimator generally has a smaller normalized bias than the other estimators for small or moderate sample sizes. Overall the second smallest normalized bias is achieved by the probability weighted moments method. The maximum likelihood estimator has the largest normalized bias when $P_F \geq 10^{-5}$, especially for the long tailed distributions. The normalized bias of all three estimators decrease as the sample size increases. When the parent distribution is GPD, all three estimators perform very well. Even so, the ordered sample least square estimator outperforms the others. The relatively strong performance for the GPD is explained as follows. The extreme value theory is based on the premise that tails of smooth continuous distributions tend towards the GPD. For the GPD, this premise is exactly satisfied. Hence, the corresponding performance is noticeably better than for other distributions.

The results for the median of the normalized root mean square error are surprising. The maximum likelihood estimator is known to be asymptotically efficient. This is always true when the samples are drawn from the underlying distribution (in our case from the generalized Pareto distribution). This property of the maximum likelihood estimator can be observed in Table 10.3 when $m = 1000$ but not for smaller sample sizes. Although the ordered sample least squares method has a smaller normalized root mean square error in many cases, there is no clear winner with respect to this criterion.

From the empirical results which are based on a limited number of distributions and sample sizes, it is not easy to make a strong recommendation as to which method to use in practice. However, in terms of the normalized bias, the ordered samples least squares estimator appears to perform better over the other estimators in estimating the large thresholds when $P_F \leq 10^{-6}$. In any event, it is seen that the extreme value theory can be used successfully to determine threshold values, when the false alarm probability is very small.

Two practical advantages of estimation based on extreme value theory are: 1) When there is a constraint on the number of samples, the thresholds obtained from extreme value theory are theoretically expected to be closer to the true thresholds than those obtained by conventional Monte Carlo techniques. However, in both techniques an increase in sample size offers greater accuracy in estimating thresholds. 2) Because the estimate of the tail of the underlying distribu-

tion is in closed form, estimation can be made for thresholds corresponding to extremely small false alarm probabilities independent of the sample size. In experiments with fixed amounts of data, this is an important advantage.

10.4.3 Effect of the Choice of α on the Threshold Estimates

As was mentioned previously, only samples whose value exceed x_0 are used in estimating the GPD parameters. The value of x_0 is determined by α . The results presented in Tables 10.2-10.3 were obtained by means of Monte Carlo experiments where $\alpha = 0.1$ was used independent of the value of P_F for which the threshold was being estimated. When the false alarm probability was extremely small, the bias and root mean square errors were quite large for some distributions. This is due to the fact that the GPD is intended to model the extreme tail of the underlying distribution. The smaller the value of α , the better will be the GPD approximation over the extreme tail being approximated. When α is chosen too large, a better fit is found for that portion of the distribution closer to the center at the expense of lesser accuracy in the extreme tail. Of course, there is a tradeoff between the choice of α and the number of data samples available for determining the parameters of the GPD.

In our application the major objective is to approximate the extreme tails corresponding to thresholds of 10^{-6} or smaller. Consequently, we explored the implications of selecting values less than 0.1 for α . To accomplish this, we obtained the theoretical values of x_i for the standard Normal and Lognormal distributions corresponding to $F^{-1}(p_i)$ where $p_i = \frac{i-0.5}{n}$: $i=1,2,\dots,n$, and $n = 1,000$ and $10,000$ respectively. These two distributions are chosen because they represent extremes: The Normal distribution is light tailed while the Lognormal is a heavy tailed distribution.

The number of the x_i samples used to determine the parameters of the GPD is given by αn . The parameters were estimated using the OLS procedure for values of α equal to 0.1, 0.05 and 0.01. The resulting GPDs were then used to determine the thresholds for false alarm probabilities given by $P_F = 10^{-k}$ where $k=2,3,\dots,7$. These results are presented in figure 10.6, where both the theoretical and approximated thresholds are plotted as a function of k for (A) Normal distribution ($n=10,000$), (B) Normal distribution ($n=1000$), (C) Lognormal distribution ($n=10,000$), (D) Lognormal distribution ($n=1000$). For $k \geq 5$, it is seen that $\alpha = 0.01$ (curve b) appears to be the best choice for approximating the thresholds. The best results were obtained with $n = 10,000$. However, good results were obtained with $n = 1,000$.

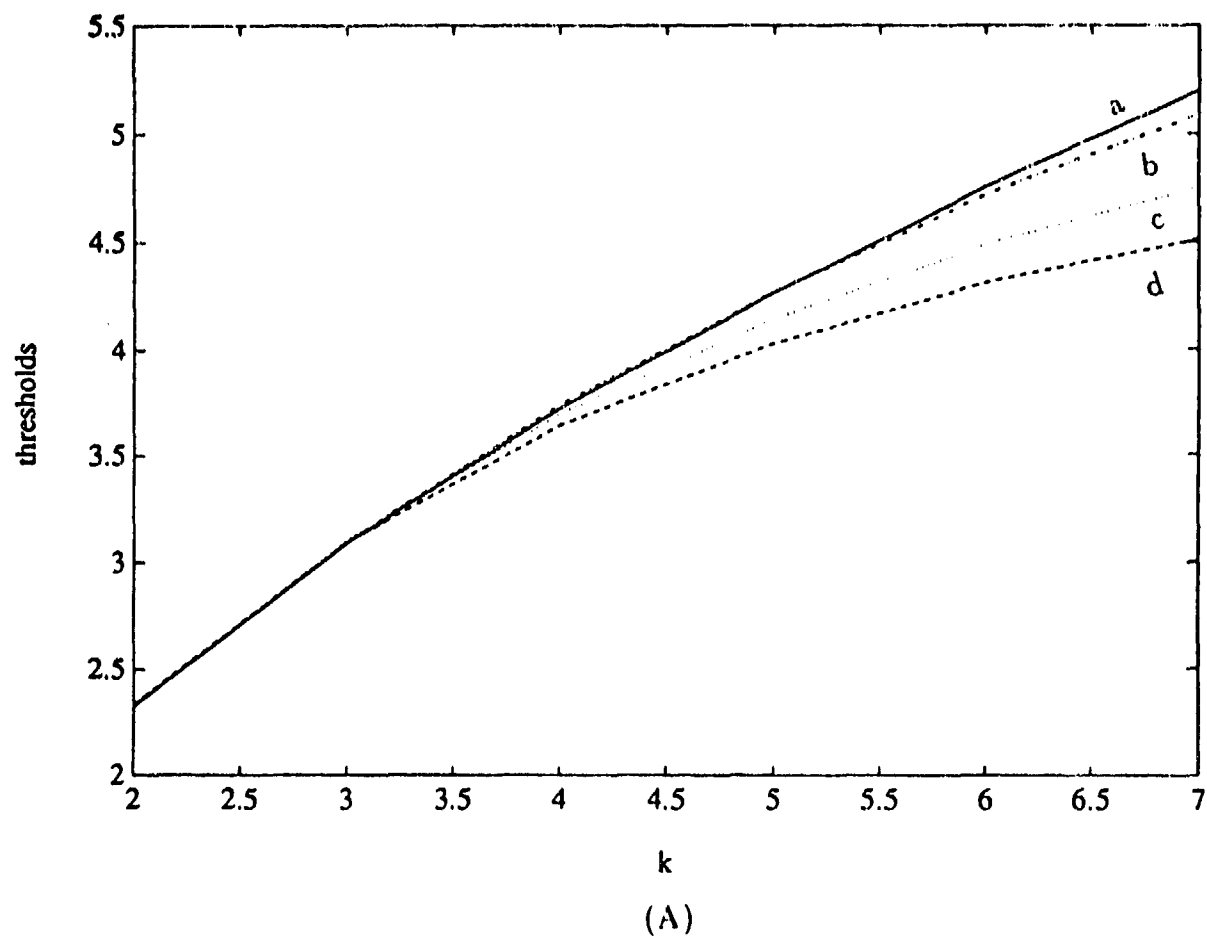


Figure 10.6: Normal distribution, $n=10,000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a: True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$.

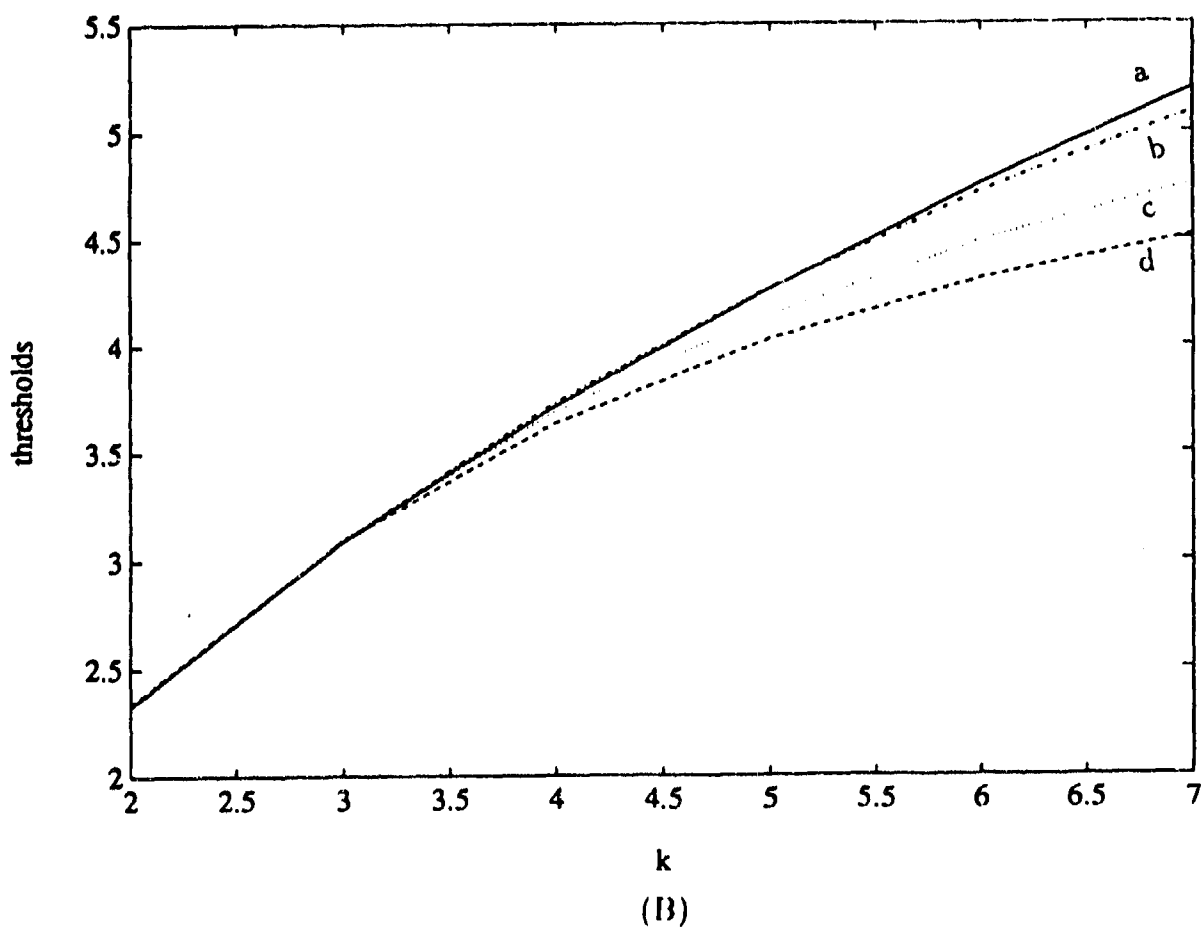


Figure 10.6: Normal distribution, $n=1000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a: True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$.

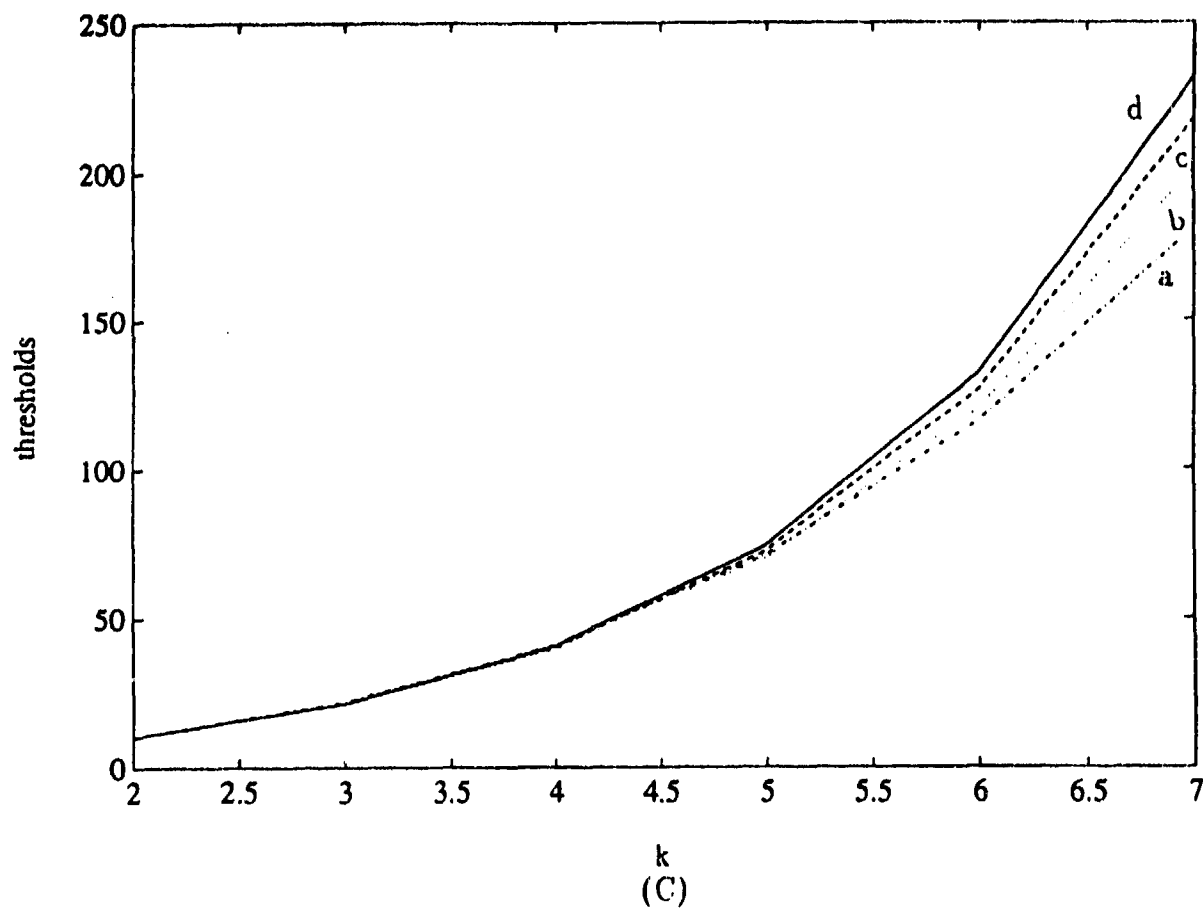


Figure 10.6: Lognormal distribution, $n=10,000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a: True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$.

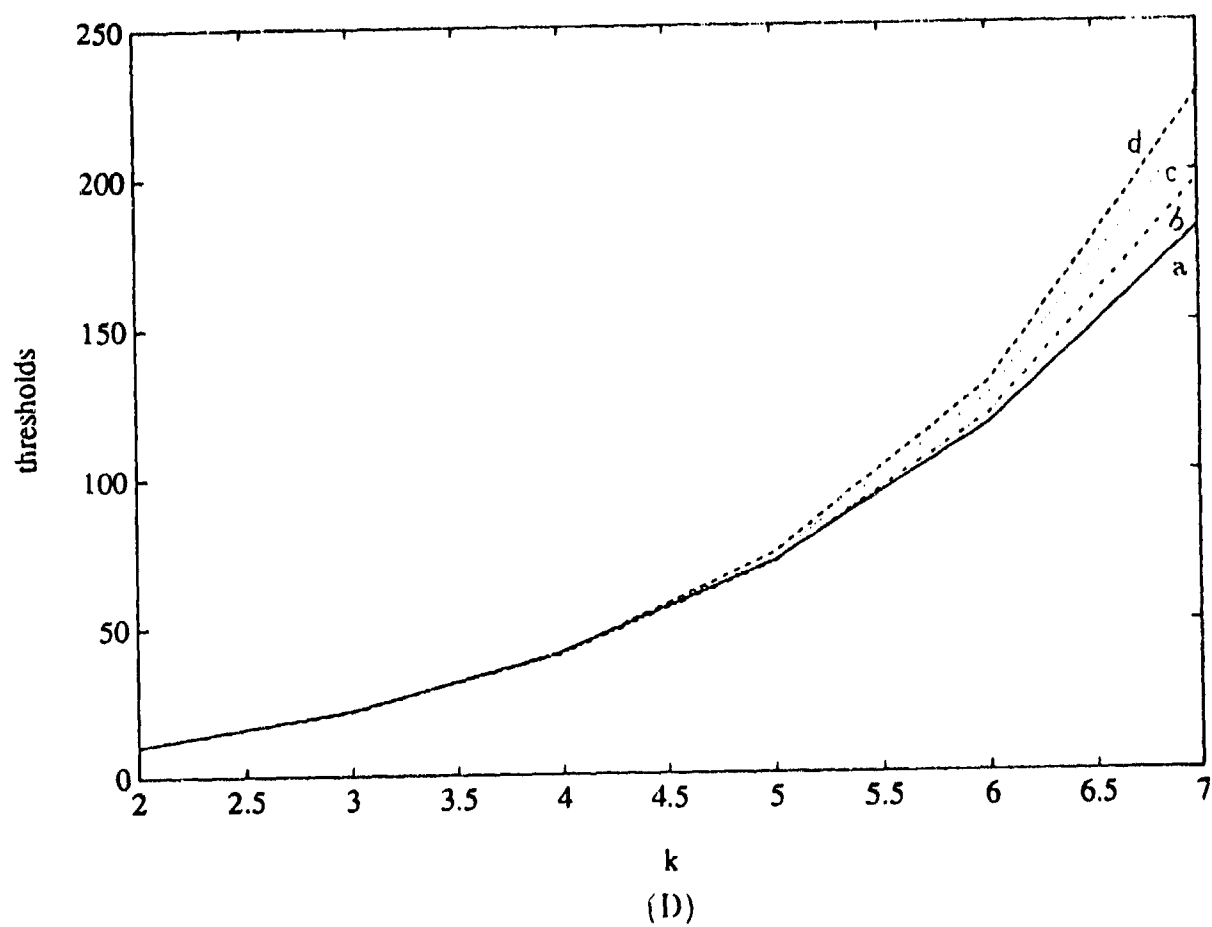


Figure 10.6: Lognormal distribution, $n=1000$ Thresholds for $P_F = 10^{-k}$. Data points correspond to $k = 2, 3, \dots, 7$. a: True, b: $\alpha=0.01$, c: $\alpha=0.05$, d: $\alpha=0.10$.

10.5 Examples

10.5.1 Known Distribution Case

To evaluate the accuracy of the threshold value estimates, 10000 random samples were generated from the Gaussian and Lognormal distributions and the upper tails of these two distributions were modeled as Generalized Pareto. In sections 10.4.1 and 10.4.3, theoretical values given by $x_i = F^{-1}(p_i)$ were used to estimate the tail. In this section randomly generated samples are used in place of the theoretical values. Choosing $\alpha = 0.01$, the theoretical thresholds of the Gaussian distribution for $P_F = 10^{-k}$ $k = 2, 3, \dots, 7$ are 2.326, 3.090, 3.719, 4.265, 4.753 and 5.199, respectively. The thresholds estimated are 2.315, 3.223, 3.847, 4.370, 4.855 and 5.292. For the Lognormal distribution the theoretical thresholds corresponding to $P_F = 10^{-k}$ $k = 2, 3, \dots, 7$ are 10.240, 21.982, 41.224, 71.157, 115.981 and 181.152. Once again, using $\alpha=0.01$, the thresholds estimated are 10.449, 22.862, 42.473, 69.216, 112.229 and 183.495. Note that the estimated results are very close to the true thresholds. We note here that these results were obtained on the basis of one set of observations from the two known distributions, corresponding to a particular seed value. For a different set of samples the estimates will be different depending on the tail behavior of that set of samples. But, unless the samples are really not a true representative of the distribution from which they are drawn, we expect that the estimates based on different samples should give threshold values that yield false alarm probabilities close to the design value.

10.5.2 An Unknown Distribution Case

In the previous section the underlying distributions were known to us and the estimates based on the extreme value theory were encouraging for both light and heavy tail behavior. In this example, we take a non-Gaussian problem where the underlying distribution is unknown.

The two hypotheses characterizing the detection problem are given in equations (9.1-9.2). We consider the weak signal case for which the clutter is much stronger than the background noise. The locally optimum detector (LOD) [74] has been shown to be suitable for the weak signal detection problem. Under hypothesis H_1 , the signal is denoted by θs_i , where θ is a measure of the signal strength. For a deterministic signal and a given set of observations $\underline{r} = [r_1, r_2, \dots, r_N]^T$ the LOD performs the LRT

$$L(\underline{r}) = \frac{\partial P_{R|H_1}(\underline{r}|H_1)}{\partial \theta} \bigg|_{\theta=0} > \eta \quad (10.53)$$

where $P_{R|H_i}(\underline{r}|H_i)$ is the joint PDF of $r_1, r_2, \dots, r_N$ under hypothesis H_i ; $i=0,1$.

Martinez, Swaszek and Thomas [75] studied the locally optimal detection problem for non-

Gaussian distributions and considered the bivariate Laplace distribution as an example. In this section we illustrate the procedure for determining the thresholds of a LOD based on $N=2$ and the received samples having the bivariate Laplace distribution given by

$$f_R(r_1, r_2) = \frac{1}{2\pi|M|^{1/2}} K_0[(2\mathbf{r}^T M^{-1} \mathbf{r})^{1/2}] \quad (10.54)$$

where M is the covariance matrix for the two samples, $|M|$ denotes its determinant, $\mathbf{r}^T M^{-1} \mathbf{r}$ is equal to $(r_1^2 - 2\rho r_1 r_2 + r_2^2)/(1 - \rho^2)$, ρ is the correlation coefficient between R_1 and R_2 and $K_0(\cdot)$ is the modified Bessel function of the second kind of zero order. The resulting locally optimum detector statistic is [75]

$$T_{LOD}(r_1, r_2) = \left(\frac{2}{\mathbf{r}^T M^{-1} \mathbf{r}} \right)^{1/2} \frac{K_1[(2\mathbf{r}^T M^{-1} \mathbf{r})^{1/2}]}{K_0[(2\mathbf{r}^T M^{-1} \mathbf{r})^{1/2}]} \times \mathbf{s}^T M^{-1} \mathbf{r} \quad (10.55)$$

where $\mathbf{s} = (s_1, s_2)^T$, $\mathbf{s}^T M^{-1} \mathbf{r} = (r_1 - \rho r_2)s_1 + (r_2 - \rho r_1)s_2$ and $K_1(\cdot)$ is the modified Bessel function of the second kind of first order. s_1 and s_2 are the known signal levels. In this example we take $s_1 = 1$ and $s_2 = -1$. Because of the complexity of $T_{LOD}(\cdot)$, it is not possible to determine a closed form expression for its probability density function.

In many applications in radar, thresholds have to be set to achieve desired false alarm probabilities based on a sample size which is orders of magnitude less than $10/P_F$. As will be pointed out later, the statistic in equation (10.55) represents a worst case situation in the sense that our simulations indicate that the variance of the test statistic is extremely large. To investigate the reliability of the thresholds estimated based on extreme value theory with smaller sample sizes, 10,000 pairs of observations (r_1, r_2) were generated from the bivariate Laplace distribution given in equation (10.54), with $\rho = 0.90$. The values of $T_{LOD}(r_1, r_2)$ were computed for each pair and sorted in increasing order. Corresponding to $\alpha = 0.01$, the largest 100 values of the underlying statistic (the top one per cent) were selected to fit the Generalized Pareto Distribution. This experiment was repeated 250 times. The threshold corresponding to a certain false alarm probability P_F of the distribution of the statistic $T_{LOD}(r_1, r_2)$ is estimated from equation (10.51) as $\hat{\eta}_{P_F} = x_0 + \hat{\sigma}[(\frac{1-P_F}{0.01})^{-\hat{\gamma}} - 1]/\hat{\gamma}$ where x_0 is the 9900th largest value of the statistic. Thresholds were estimated for false alarm probabilities $P_F = 10^{-k}$, $k = 2, \dots, 7$ for each repetition of the experiment. Histograms of these threshold values are shown in figure 10.7, for the different P_F s. To give a better appreciation for the range of values, the bins are not necessarily of equal

width. The histograms give an indication of the spread in the threshold values depending on the particular samples collected. From the histograms corresponding to false alarm probabilities of 10^{-2} , 10^{-3} and 10^{-4} we can see that the threshold estimates obtained on the basis of even one set of samples is likely to approximately yield the desired P_F . Since the underlying distribution of $T_{LOD}(\cdot)$ is unknown, one measure of the accuracy of the estimate is the extent to which most of the estimates fall in one bin of the histogram. Also, we can see that there is negligible overlap between the estimated threshold values in the histograms for the three different P_F s. This supports the claim that the estimated threshold is likely to yield a false alarm probability which is of the same order as the desired P_F . There is a higher overlap in the thresholds of the histograms for $P_F=10^{-5}$, 10^{-6} and 10^{-7} . Also, there is much higher spread in the threshold values estimated. Based on the excellent results obtained for the same choices of P_F s in the known cases of the previous section, these results are surprising. However, it is explained as follows. The γ values of the GPD estimated for the different repetitions of this experiment lie in the range 0.45 – 0.55. This represents an extremely heavy tailed distribution. From Table 10.1 we see that the Lognormal distribution, which is quite a heavy tailed distribution, has $\gamma=0.232$. The heavy tailed nature of the detector statistic can also be observed by comparing the large threshold values seen in the histograms with the corresponding thresholds of the Gaussian and the Lognormal distributions. The variance of the GPD is given by

$$\begin{aligned} \text{Var}(X) &= \frac{\sigma^2}{(1-\gamma^2)(1-2\gamma)} & \gamma < 0.5 \\ &= \infty & \gamma \geq 0.5 \end{aligned} \quad (10.56)$$

Thus, the bivariate Laplace results in a very highly fluctuating statistic with an extremely large variance. As such, it represents a 'worst case' situation for empirically determining the threshold. A much larger sample size is needed to obtain reliable threshold estimates because of the exceedingly large tail of the underlying distribution.

In general, an indication of how heavy the true tail may be for an unknown distribution is given by the estimate of γ for the GPD. When an extremely heavy tail is indicated, another strategy for estimating the thresholds when P_F is very small is to choose the median value of the thresholds estimated when the experiment is repeated a specified number of times with 10,000 samples in each repetition. The choice of the median as the estimator ensures that very large and very small values do not affect the results. For the present example, we chose to repeat the

250 trials three times. By counting the number of estimates that fell into the bins centered at 20, 28 and 36 for $P_F=10^{-5}$, 40, 50, 70 and 90 for $P_F=10^{-6}$ and 100 and 150 for $P_F=10^{-7}$, it was found that 88 percent of the estimates fell into these bins. Thus, even for this extremely large tailed example, we believe that use of the GPD has allowed us to estimate useful values for the thresholds with sample sizes much smaller than $10/P_F$.

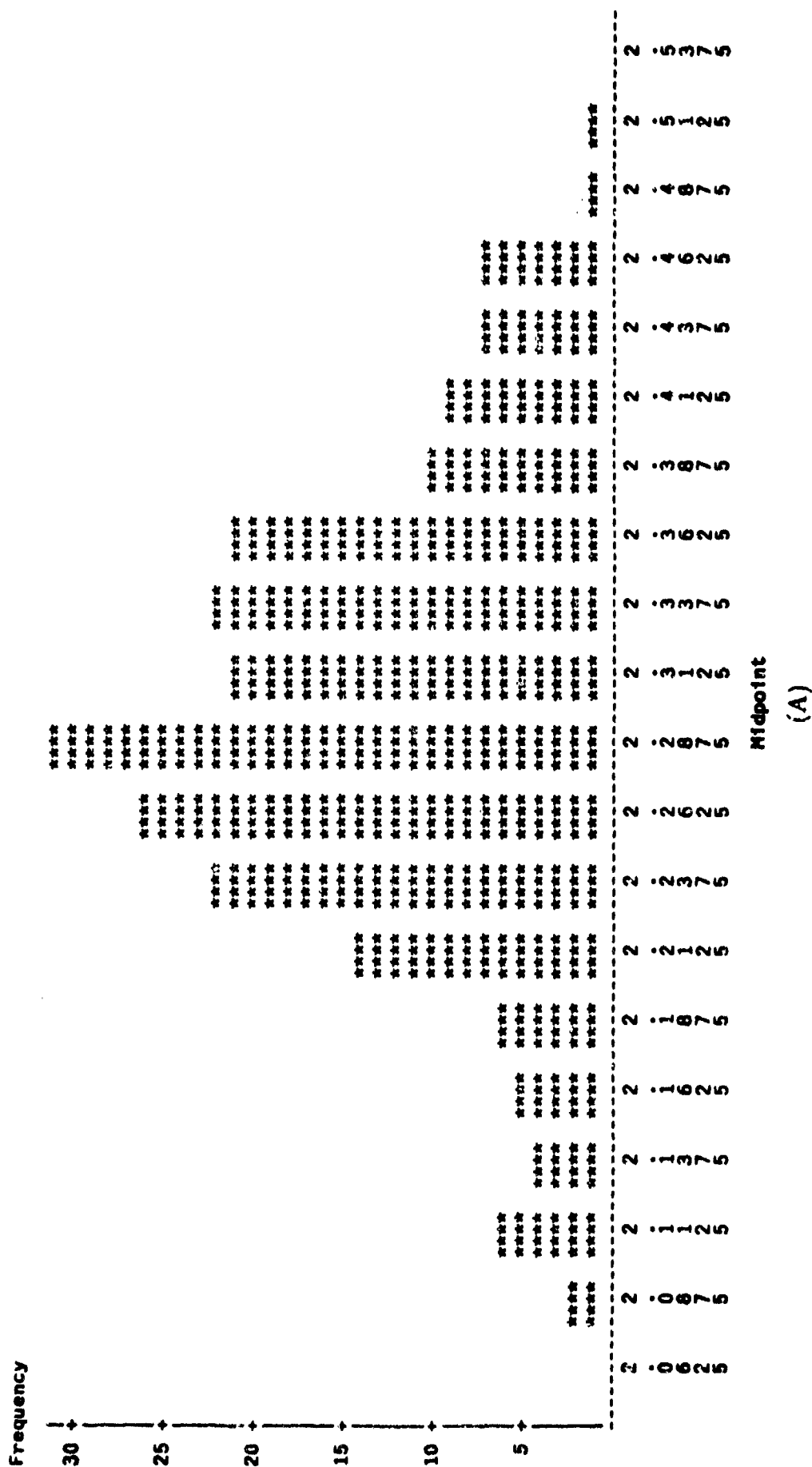


Figure 10.7: Histograms of threshold values. (A) $P_F = 10^{-2}$ (B) $P_F = 10^{-3}$ (C) $P_F = 10^{-4}$ (D) $P_F = 10^{-5}$ (E) $P_F = 10^{-6}$ (F) $P_F = 10^{-7}$

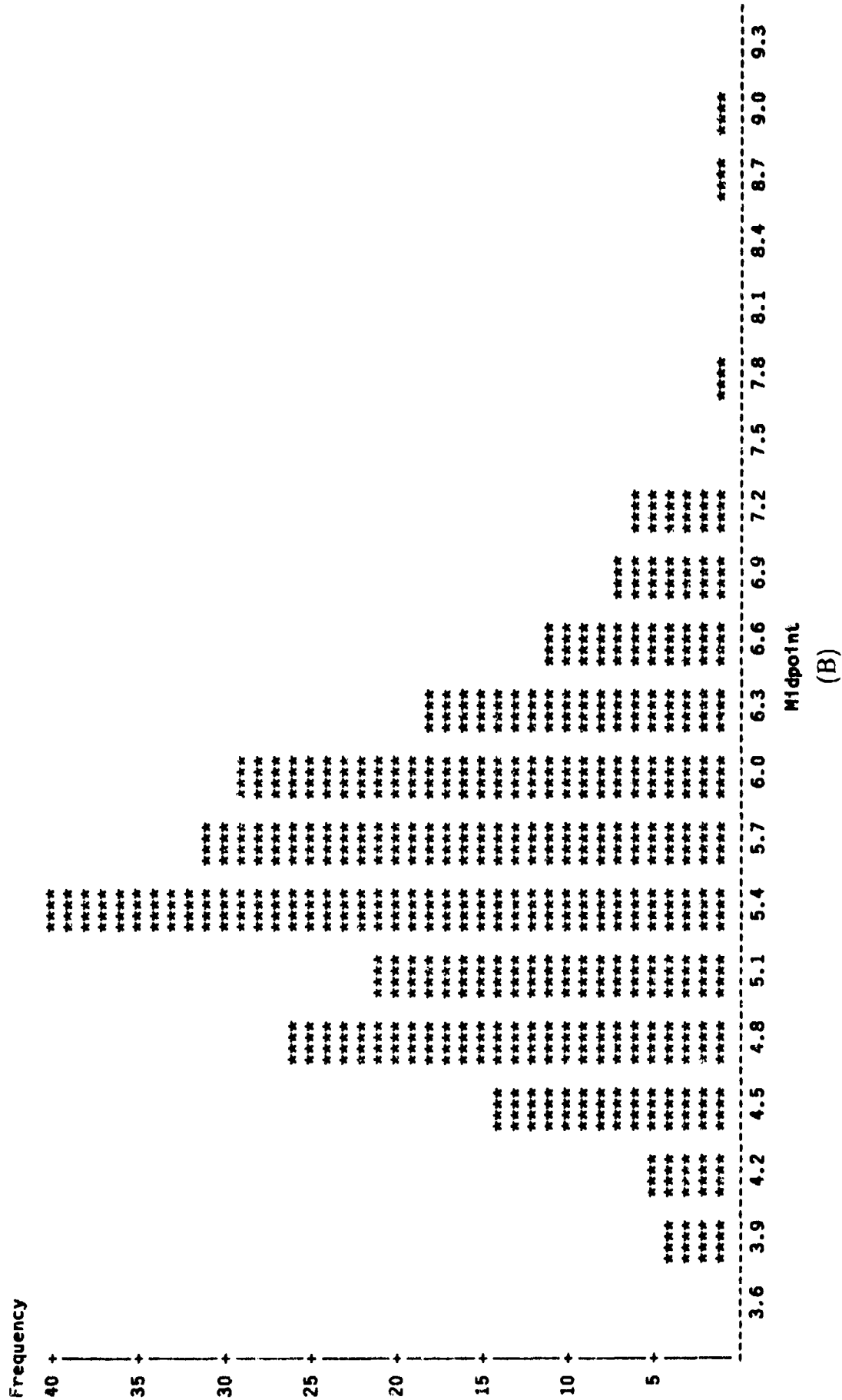


Figure 10.7: Fig 10.7 Contd.

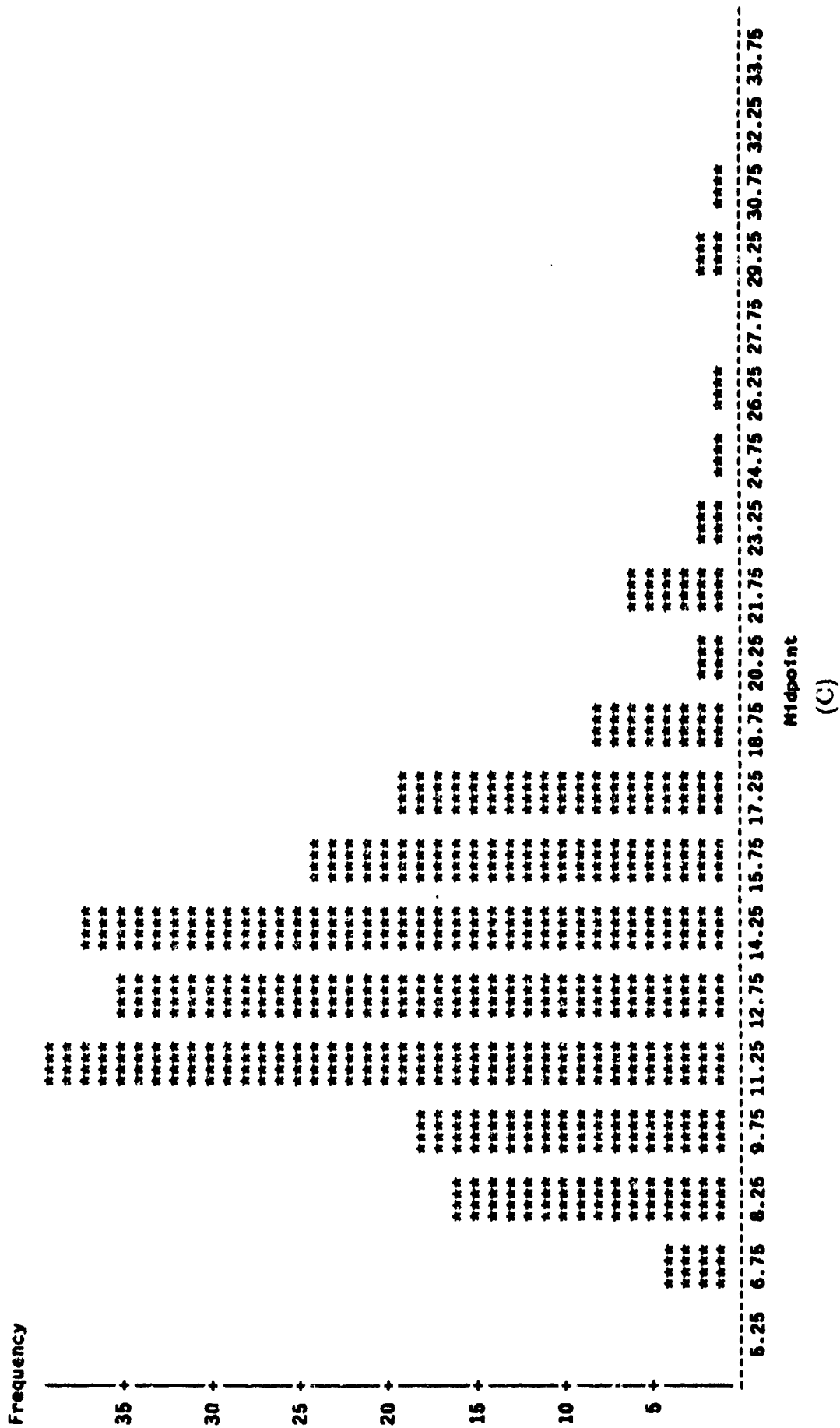


Figure 10.7: Fig 10.7 Contd.

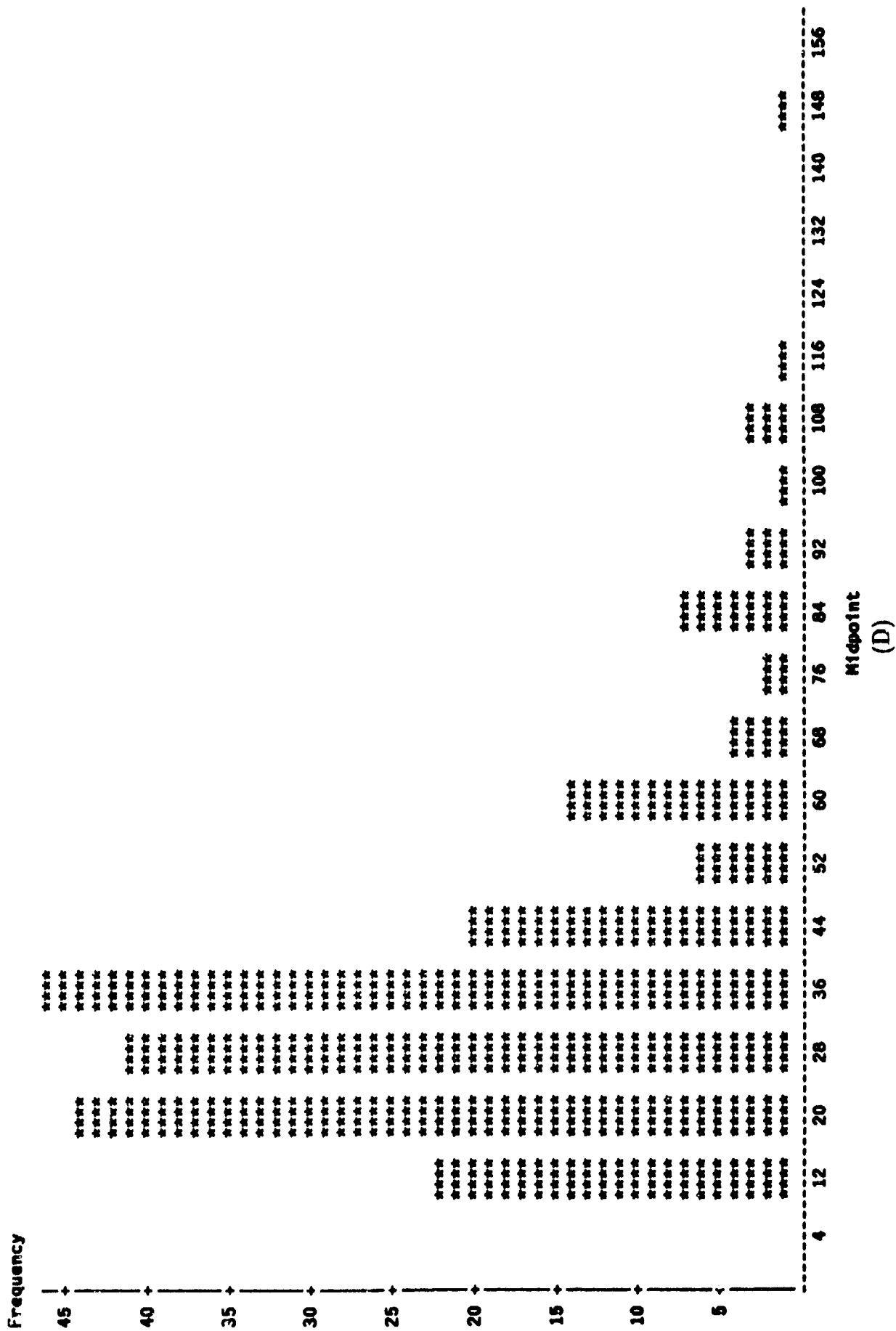


Figure 10.7: Fig 10.7 Contd.

Chapter 11

Performance of the Locally Optimum Detector for the Multivariate Student-T Distribution

In radar problems involving weak signal applications, it is found that the large returns due to clutter can lead to a small signal to disturbance ratio. The large returns from clutter result when the density function of the clutter exhibits an extended tail behavior. Consequently, the probability density function of the disturbance can no longer be modeled as Gaussian. The significance of a non-Gaussian PDF with an extended tail is that many more large returns result than would be the case for a Gaussian PDF having the same variance. Hence, there is a need to be able to model non-Gaussian random processes.

The multivariate student-T distribution is a member of the class of joint PDFs arising from Spherically Invariant Random Processes (SIRP). SIRPs are explained in detail in the earlier chapters. When an SIRP is sampled at N instants in time, the resulting vector is said to be spherically invariant. The theory of SIRPs offers a way to model the joint density function on these N samples where the correlation between the individual random variables in the vector is accounted for. With this approach locally optimum detector structures can be derived for non-Gaussian disturbances without the need to assume that the random variables are statistically independent. In this chapter we analyze the performance of the LOD when the background disturbance consisting of clutter and noise can be modeled as having a multivariate student-T distribution.

11.1 The Multivariate Student-T Distribution

A convenient procedure for generating a multivariate student-T distribution is discussed in this section. Let the random vector $\underline{X}$ have a multivariate Gaussian distribution with zero mean and covariance matrix M . The zero mean assumption will not affect the generality of the results that follow. The joint density function on the elements of $\underline{X}$ is given by

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^N |M|^{1/2}} e^{-\frac{\underline{x}^T M^{-1} \underline{x}}{2}} \quad (11.1)$$

where the vector $\underline{X}$ has $2N$ elements from N inphase and N quadrature samples. Consider the vector $\underline{W} = \underline{X}/\nu$, where ν is a nonnegative random variable statistically independent of $\underline{X}$. Let $\underline{w}^T M^{-1} \underline{w}$ be denoted by the variable p . Then, the conditional density function of the vector $\underline{W}$ given ν can be written as

$$f_{\underline{W}}(\underline{w}|\nu) = \frac{1}{(2\pi)^N |M|^{1/2}} \nu^{2N} e^{-\frac{\nu^2 p}{2}} \quad (11.2)$$

The unconditional density function on $\underline{W}$ is given by

$$f_{\underline{W}}(\underline{w}) = \int_0^\infty f_{\underline{W}}(\underline{w}|\nu) f_\nu(\nu) d\nu \quad (11.3)$$

where $f_\nu(\nu)$ is the probability density function of the random variable ν . Because $\underline{X}$ and ν are statistically independent, it follows that

$$E(\underline{W}) = E\left(\frac{\underline{X}}{\nu}\right) = E(\underline{X})E(\nu^{-1}) = \underline{0} \quad (11.4)$$

$$E(\underline{W} \underline{W}^T) = E(\underline{X} \underline{X}^T)E(\nu^{-2}) = E(\nu^{-2})M. \quad (11.5)$$

It can be seen from the above equation that the level for the variance of the elements of the vector $\underline{W}$ can be adjusted by appropriate choice of $E(\nu^{-2})$.

With respect to equation (11.3), let $f_\nu(\nu)$ be the generalized chi PDF given by

$$f_\nu(\nu) = 2 \frac{\nu^{2\beta-1} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)}. \quad (11.6)$$

From equation (11.6), $E(\nu^{-2})$ can be calculated. Specifically,

$$E(\nu^{-2}) = \int_0^\infty 2\nu^{-2} \frac{\nu^{2\beta-1} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)} d\nu = \int_0^\infty 2 \frac{\nu^{2\beta-3} e^{-\alpha\nu^2} \alpha^\beta}{\Gamma(\beta)} d\nu \quad (11.7)$$

Letting $\alpha\nu^2 = x$ in the above equation we get

$$E(\nu^{-2}) = \alpha \int_0^\infty \frac{x^{\beta-2} e^{-x} dx}{\Gamma(\beta)} = \alpha \frac{\Gamma(\beta-1)}{\Gamma(\beta)} = \frac{\alpha}{\beta-1}. \quad (11.8)$$

If we let $\alpha = \beta - 1$, then the generalized chi PDF in equation (11.6) is such that $E(\nu^{-2}) = 1$ irrespective of the choice for the parameter β . Then the generalized chi PDF takes the form

$$f_\nu(\nu) = \frac{2\nu^{2\beta-1} e^{-(\beta-1)\nu^2} (\beta-1)^\beta}{\Gamma(\beta)} \quad \beta > 1. \quad (11.9)$$

In general, we can set the value of $E(\nu^{-2})$ to a desired constant C by choosing $\alpha = C(\beta - 1)$.

Integrating the conditional density function $f_{\underline{W}}(\underline{w}|\nu)$ as given by equation (11.2), over the PDF of the nonnegative random variable ν , we obtain the multivariate student-T distribution. The details are given below. Choosing $\alpha = \beta - 1$ in equation (11.6) we can write

$$\begin{aligned} f_{\underline{W}}(\underline{w}) &= \int_0^\infty \frac{1}{(2\pi)^N |M|^{1/2}} \nu^{2N} e^{-\frac{\nu^2 p}{2}} \frac{2\nu^{2\beta-1} e^{-(\beta-1)\nu^2} (\beta-1)^\beta}{\Gamma(\beta)} d\nu \\ &= \frac{(\beta-1)^\beta}{(2\pi)^N |M|^{1/2} \Gamma(\beta)} \int_0^\infty 2\nu^{2N+2\beta-1} e^{-\nu^2(\beta-1+p/2)} d\nu. \end{aligned} \quad (11.10)$$

Letting $(\beta - 1 + p/2)\nu^2 = y$ we get

$$\begin{aligned} f_{\underline{W}}(\underline{w}) &= \frac{(\beta-1)^\beta}{(2\pi)^N |M|^{1/2} \Gamma(\beta)} \int_0^\infty \frac{y^{N+\beta-1} e^{-y}}{(\beta-1+p/2)^{N+\beta}} dy \\ &= \frac{(\beta-1)^\beta \Gamma(N+\beta)}{(2\pi)^N |M|^{1/2} \Gamma(\beta) (\beta-1+p/2)^{N+\beta}}. \end{aligned} \quad (11.11)$$

The above expression is defined to be the $2N$ -dimensional multivariate student-T distribution with parameters N and β . N represents the number of complex samples and β determines the tail behavior of the multivariate density function.

For simulation purposes, the density function in equation (11.9) can be simulated as follows. The first step is to generate a standard Gamma variate from the density function $f_Y(y) = \frac{y^{\beta-1} e^{-y}}{\Gamma(\beta)}$. The IMSL package was used to generate standard Gamma variates. The next step is to divide the generated random variable by the parameter $\beta - 1$. Let $X = Y/(\beta - 1)$. The density function of X is

$$f_X(x) = \frac{(\beta-1)^\beta x^{\beta-1} e^{-x(\beta-1)}}{\Gamma(\beta)}. \quad (11.12)$$

The positive square root of $\frac{Y}{\beta-1}$ results in the desired density function. Let $\nu = X^{\frac{1}{2}}$. Therefore $X = \nu^2$. Introducing the Jacobian of the transformation, the density function of ν becomes

$$f_{\nu}(\nu) = \frac{2\nu^{2\beta-1}e^{-(\beta-1)\nu^2}(\beta-1)^{\beta}}{\Gamma(\beta)} \quad (11.13)$$

which is identical to that in equation (11.9).

11.2 The Locally Optimum Detector

The locally optimum detector for the multivariate student-t distribution can now be derived. From equation (9.32) the locally optimum detector is given as

$$\frac{\frac{\partial f_D(\underline{r}-\theta\underline{s})}{\partial \theta}|_{\theta=0}}{f_D(\underline{r})} \underset{H_0}{\overset{H_1}{>}} \mu. \quad (11.14)$$

Assuming the disturbance can be modeled by a multivariate student-T distribution, $f_D(\underline{r})$ is given by equation (11.11), where $p = \underline{r}^T M^{-1} \underline{r}$. Since equation (11.14) is a ratio test and all constants can be placed in the threshold which is determined by specifying a false alarm probability, all multiplicative constants are ignored for convenience. Hence, we will be concerned only with the terms containing the variable $\underline{R}$. Excluding the constant term the numerator in the ratio test is given by

$$\frac{\partial f_D(\underline{r}-\theta\underline{s})}{\partial \theta}|_{\theta=0} = \frac{\partial}{\partial \theta} \left[\frac{1}{(\beta-1+p/2)^{N+\beta}} \right] |_{\theta=0}. \quad (11.15)$$

Applying the chain rule, the derivative with respect to θ can be expressed as the derivative with respect to p times the derivative of p with respect to θ . The derivative of p with respect to θ at $\theta = 0$ can be derived as

$$\frac{\partial p}{\partial \theta}|_{\theta=0} = \left(\frac{\partial}{\partial \theta} (\underline{r}-\theta\underline{s})^T M^{-1} (\underline{r}-\theta\underline{s}) \right) |_{\theta=0} = -2\underline{s}^T M^{-1} \underline{r}. \quad (11.16)$$

Therefore, the numerator in the ratio test, excluding the constant, is given by

$$\frac{\partial f_D(\underline{r}-\theta\underline{s})}{\partial \theta}|_{\theta=0} = (\beta-1+p/2)^{-(N+\beta+1)} \times \underline{s}^T M^{-1} \underline{r}. \quad (11.17)$$

From the above equation, the sufficient statistic for the locally optimum detector for the multivariate student-T distribution can be written as

$$T_{LOD}(\underline{r}) = \frac{\underline{s}^T M^{-1} \underline{r}}{\beta - 1 + p/2}. \quad (11.18)$$

The above result for the LOD statistic is very significant. The numerator in equation (11.18) is recognized as the Gaussian linear detector. This detector is a matched filter which maximizes the signal-to-disturbance ratio whether or not the disturbance is Gaussian. In weak signal applications the signal to disturbance ratio will still be low after matched filtering. The denominator of the LOD statistic is the nonlinear term in the statistic. The behavior of the nonlinearity is such that it scales down large values of p and enhances small values of p . The nonlinearity is plotted as a function of p in Fig. 11.1. This is reasonable because large values of radar returns result in large p while small values of the returns yields small values of p . Because it is known a priori that we are dealing with the weak signal problem, large returns cannot be due to the signal. Consequently, the output of the matched filter is weighted by a small number. On the other hand, the matched filter output is weighted by a large number when the return is small and the contribution due to the signal, if present, can be detected.

11.3 Computer Simulation of Performance

The performance of the locally optimum detector in a multivariate student-T distributed clutter is obtained through computer simulations for weak signal applications. For simulation purposes a multivariate student-T distributed disturbance vector $\underline{D}$ and a transmitted signal vector $\underline{S}$ have to be generated. The first step in generating the correlated multivariate student-T distributed random variables is to generate a $2N$ -dimensional white Gaussian random vector. Subroutine DRNNOA from the IMSL package is used to generate a white Gaussian vector of desired dimension. Each element of the white Gaussian vector is divided by the random variable generated from the density function in equation (11.9). This results in a white student-T distributed vector. The next step is to introduce correlation between the random variables. The covariance matrix of the clutter process is assumed known with unit elements along the diagonal. To get the covariance matrix M of the disturbance we add a small number, determined by the clutter to noise ratio, to the diagonal elements of the clutter covariance matrix. This serves to limit the performance of the receiver even where the clutter power is negligible. In this simulation, the clutter to noise ratio is taken to be 80 dB. Given the covariance matrix,

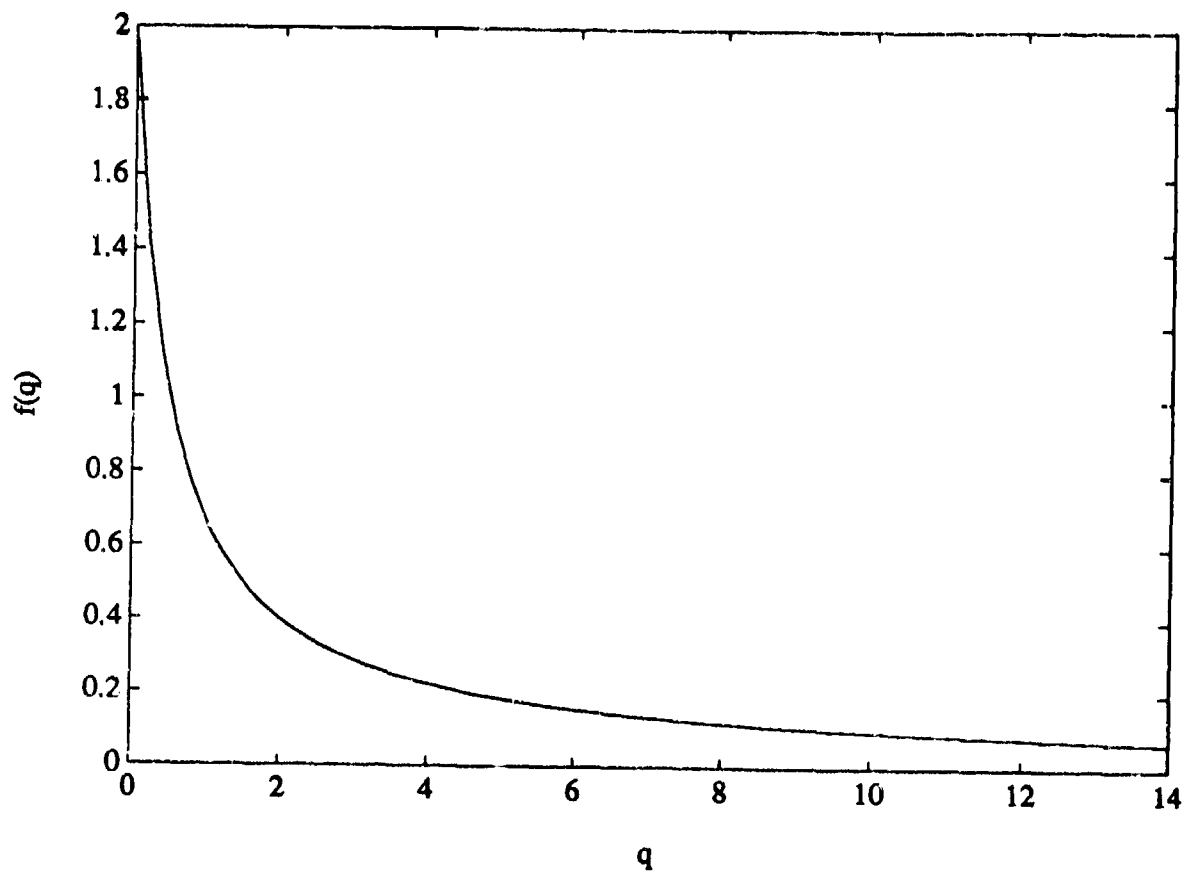


Figure 11.1: Nonlinearity for the student-T distribution.

a cholesky decomposition is carried out such that the matrix $M = KK^T$ where K is a lower triangular matrix. Multiplying the matrix K and the white student-T distributed vector we obtain a student-T distributed vector with the desired correlation between the random variables.

The autocorrelation of the clutter process is taken to be a geometric function in this problem. Assuming radar returns from clutter cells to be highly correlated, as is the case with ground clutter, the sample to sample correlation is taken as 0.95 in this problem. Specifically, the sample autocorrelation function is chosen as

$$R_{CC}(n) = (0.95)^n \quad n = 0, 1, \dots, N - 1 \quad (11.19)$$

where $R_{CC}(n)$ is the discrete time autocorrelation function of the clutter process. Using the above function the elements of the covariance matrix of the disturbance can be filled appropriately. The elements of the signal vector are chosen such that the n^{th} element $S_n = e^{j2\pi f_D(n-1)T}$, $n = 1, 2, \dots, N$. f_D represents the Doppler frequency shift of the received signal and T represents the time separation between sampling instants.

The detector in equation (11.18) is now simulated. A value of $\beta = 1.5$ for the multivariate student-T distribution is chosen because this value results in a relatively long tail for the corresponding marginal PDF of one element of the vector. By evaluating thresholds for specified false alarm probabilities, the student-T distribution was seen to have heavier tails than the Gaussian distribution for false alarm probabilities less than 10^{-4} but smaller tails than the Gaussian otherwise.

The thresholds corresponding to false alarm probabilities 10^{-k} ; $k = 1, 2, 3, 4$ are obtained through the method of extreme value theory explained in Chapter 10. Once the threshold is set the detection probabilities are obtained by simulating the LOD for received vectors consisting of the sum of the signal and disturbance vectors for various signal-to-disturbance ratios. The value of f_D is chosen to be zero in this simulation. The number of trials in the Monte Carlo simulation for each case is equal to 10,000. The performance of the LOD is compared to that of the Gaussian detector for the same multivariate student-T distributed clutter. The test statistic for the Gaussian detector is the same as the numerator of the LOD, which is $\underline{g}^T M^{-1} \underline{z}$. The results are shown in Tables 11.1-11.14.

From the tables it can be seen that, when the false alarm probability is 10^{-1} , the LOD and the Gaussian receiver have comparable performances for the various signal to clutter ratios con-

sidered. For lower false alarm probabilities, the LOD always outperforms the Gaussian receiver except for the zero dB entries in Tables 11.1, 11.4, 11.7, 11.11 and 11.12. The difference is especially significant for false alarm probabilities equal to 10^{-3} and 10^{-4} .

From our computer simulations we expect that the performance improvement of the LOD over the linear Gaussian receiver depends on the shape of the tail disturbance PDF. The heavier the tail of the disturbance PDF, the better is the expected improvement in performance. The student-T distribution, while being heavier tailed than the Gaussian, is not as heavy tailed as the K-distribution and Weibull distribution. In fact, the student-T distribution may not be a likely candidate for modeling the radar disturbance. The student-T distribution was chosen as the first distribution to be studied only because of the mathematical simplicity and well behaved nature of its multivariate PDF. Nevertheless, the analysis done with the student-T distribution confirms that the LOD outperforms the Gaussian receiver for weak signal applications.

11.4 Conclusions

It can be observed from the tables that the Gaussian receiver performance degrades abruptly for false alarm probabilities less than or equal to 10^{-2} whereas the LOD shows a gentler degradation in performance. Both the receivers show an improvement in performance as the number of samples is increased. However, the LOD shows a dramatic improvement in performance when the sample size is greater than 64. From Table 11.14, it can be seen that for $SCR=0$ dB and $P_F = 10^{-4}$, the detection probability for the LOD is 0.3720 while that for the Gaussian receiver is 0.0003. This represents an improvement factor in the vicinity of 3 orders of magnitude for the LOD. Also, from Tables 11.3, 11.6, 11.9 and 11.13, we observe that when the P_F is set to 10^{-3} the LOD shows a performance improvement of two orders of magnitude compared to the Gaussian receiver. For larger sample sizes eg. 64, 128 the detection probability of the LOD is in the tenths for $SCR=-10$ dB and $P_F = 10^{-2}$, while for the Gaussian receiver it is in the hundredths. Overall, when P_F is less than or equal to 10^{-2} , the Gaussian receiver requires a signal-to-clutter ratio 10-20 dB larger than that required by the LOD for the same values of P_F and P_D .

The LOD does not work well if the signal to clutter ratio is too large. The performance degrades rapidly for signal to clutter ratios exceeding zero dB. The LOD is designed for detecting targets when it is known that the signal is weak. The aim of using a LOD is to obtain detection in range-Doppler-azimuth cells where conventional space-time processing does not help in getting acceptable performance. These cells are now ignored because it is felt that they are hopeless for

target detection purposes. The nonlinearity present in the LOD plays the role of suppressing large returns. However, if the SCR is high, the large returns are more likely to be caused due to the signal than due to the clutter. Hence, the detection performance will drop off compared to the Gaussian receiver. In general, when the SCR is relatively high (10 dB) the likelihood ratio test is the optimal test for target detection under a fixed false alarm constraint. When the signal to clutter ratio becomes very close to zero, the LOD receiver will hardly show any detections even though it would still outperform the Gaussian receiver. This is because the PDFs under H_0 and H_1 are so close to each other that it is impossible to separate them without increasing the sample size by orders of magnitude.

The concept of spherically invariant random processes and locally optimum detectors are particularly relevant in the context of modern radar applications. When the radar scans a volume searching for targets there might be certain regions in the volume where the clutter returns are so strong that signal returns get blanked out. It is in these regions that we can obtain detections with LODs. There is a need to monitor the environment so that we are able to separate the clutter regions from volumes that are just limited by background noise. When detections are limited by background noise alone, LODs are inapplicable. In this research effort work is beginning in the area of using artificial intelligence (AI) for monitoring the volume. Using AI, clutter patches can be identified and the underlying multivariate PDF of the clutter returns can be approximated using the library of SIRPs that have been developed. From the library of LODs the LOD corresponding to the approximated SIRP can be used in clutter regions to obtain detections if the target is present, where earlier it would not have been possible.

SCR		LOD	GR
0 dB	P_D	0.7047	0.8600
-10 dB	P_D	0.3220	0.2800
-20 dB	P_D	0.1611	0.1460
-30 dB	P_D	0.1175	0.1190

Table 11.1: Sample Size=16, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.3761	0.1050
-10 dB	P_D	0.0838	0.0190
-20 dB	P_D	0.0246	0.0120
-30 dB	P_D	0.0141	0.0100

Table 11.2: Sample Size=16, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.1604	0.0030
-10 dB	P_D	0.0198	0.0014
-20 dB	P_D	0.0027	0.0011
-30 dB	P_D	0.0012	0.0001

Table 11.3: Sample Size=16, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.7607	0.9090
-10 dB	P_D	0.3608	0.3200
-20 dB	P_D	0.1704	0.1540
-30 dB	P_D	0.1202	0.1190

Table 11.4: Sample Size=32, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.4573	0.1750
-10 dB	P_D	0.1052	0.0220
-20 dB	P_D	0.0255	0.0130
-30 dB	P_D	0.0145	0.0120

Table 11.5: Sample Size=32, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.2621	0.0035
-10 dB	P_D	0.0289	0.0015
-20 dB	P_D	0.0042	0.0012
-30 dB	P_D	0.0013	0.0001

Table 11.6: Sample Size=32, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.8117	0.9510
-10 dB	P_D	0.4302	0.3790
-20 dB	P_D	0.1278	0.1590
-30 dB	P_D	0.1252	0.1195

Table 11.7: Sample Size=64, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.5484	0.3000
-10 dB	P_D	0.1446	0.0230
-20 dB	P_D	0.0301	0.0120
-30 dB	P_D	0.0152	0.0010

Table 11.8: Sample Size=64, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.3643	0.0048
-10 dB	P_D	0.0492	0.0016
-20 dB	P_D	0.0057	0.0012
-30 dB	P_D	0.0019	0.0001

Table 11.9: Sample Size=64, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.2522	0.0002
-10 dB	P_D	0.0202	0.0001
-20 dB	P_D	0.0018	0.0001
-30 dB	P_D	0.0005	0.0000

Table 11.10: Sample size=64, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.8517	0.9790
-10 dB	P_D	0.4987	0.4870
-20 dB	P_D	0.2080	0.1710
-30 dB	P_D	0.1314	0.1186
-40 dB	P_D	0.1106	0.1059

Table 11.11: Sample Size=128, $P_F = 10^{-1}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.6511	0.7050
-10 dB	P_D	0.2190	0.0320
-20 dB	P_D	0.0445	0.0150
-30 dB	P_D	0.0198	0.0116
-40 dB	P_D	0.0147	0.0010

Table 11.12: Sample Size=128, $P_F = 10^{-2}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.4777	0.0090
-10 dB	P_D	0.0869	0.0020
-20 dB	P_D	0.0098	0.0013
-30 dB	P_D	0.0037	0.0011
-40 dB	P_D	0.0021	0.0001

Table 11.13: Sample Size=128, $P_F = 10^{-3}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

SCR		LOD	GR
0 dB	P_D	0.3720	0.0003
-10 dB	P_D	0.0430	0.0002
-20 dB	P_D	0.0039	0.0001
-30 dB	P_D	0.0007	0.0001
-40 dB	P_D	0.0003	0.0000

Table 11.14: Sample Size=128, $P_F = 10^{-4}$, SCR:Signal to Clutter Ratio, LOD:Locally Optimum Detector, GR:Gaussian Receiver

Appendix A

Properties of SIRVs

In this appendix we present some original proofs for properties of SIRPs stated in the literature.

A.1 Statistical Independence

An SSRV $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$ results in statistical independence of the X_i $i = 1, 2, \dots, N$ if and only if the SSRV is Gaussian.

Proof: Recall that the PDF of $\mathbf{X}$ can be expressed as

$$f_{\mathbf{X}}(\mathbf{x}) = kh_N[(x_1^2 + x_2^2 + \dots + x_N^2)^{\frac{1}{2}}] = (2\pi)^{-\frac{N}{2}} h_N(\sqrt{\mathbf{x}^T \mathbf{x}}). \quad (\text{A.1})$$

If the components of $\mathbf{X}$ are statistically independent, then the PDF given by eq (A.1) must factor into the product of the marginal PDFs of the components of $\mathbf{X}$. It then follows that

$$h_N[(x_1^2 + x_2^2 + \dots + x_N^2)^{\frac{1}{2}}] = \prod_{i=1}^N g(x_i). \quad (\text{A.2})$$

Letting $r = (x_1^2 + x_2^2 + \dots + x_N^2)^{\frac{1}{2}}$ and differentiating both sides of eq (A.2) with respect to x_i , results in

$$\frac{x_i}{r} h'_N(r) = \prod_{\substack{j=1 \\ j \neq i}}^N g(x_j) g'(x_i). \quad (\text{A.3})$$

Dividing both sides of eq (A.3) by $x_i h_N(r)$ results in

$$\frac{h'_N(r)}{r h_N(r)} = \frac{g'(x_i)}{x_i g(x_i)}. \quad (\text{A.4})$$

Equality holds in eq (A.4) if and only if the left and right sides of eq (A.4) are equal to the same constant. Denoting this constant by $-\lambda$, we have

$$\frac{h'_N(r)}{r h_N(r)} = -\lambda. \quad (\text{A.5})$$

Integrating both sides of eq (A.5) with respect to r gives

$$h_N(r) = a \exp\left(-\frac{\lambda r^2}{2}\right) \quad (\text{A.6})$$

where a is the constant of integration. Hence,

$$h_N[(x_1^2 + x_2^2 + \dots + x_N^2)^{\frac{1}{2}}] = a \exp\left[-\frac{\lambda}{2}(x_1^2 + x_2^2 + \dots + x_N^2)\right] \quad (\text{A.7})$$

Substitution of eq (A.7) in eq (A.1) clearly results in the Gaussian PDF. The constraint of unity volume under the PDF results in $a = \lambda^{\frac{N}{2}}$.

In order to prove the sufficient part of the property, we start with the marginal PDFs of the components of $\mathbf{X}$ given by

$$f_{X_i}(x_i) = \left(\frac{2\pi}{\lambda}\right)^{-\frac{1}{2}} \exp\left(-\frac{\lambda x_i^2}{2}\right). \quad (\text{A.8})$$

Under the assumption of statistical independence, we obtain the PDF of $\mathbf{X}$ by taking the product of the marginal PDFs of its components as

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{2\pi}{\lambda}\right)^{-\frac{N}{2}} \exp\left(-\frac{\lambda}{2} \sum_{i=1}^N x_i^2\right). \quad (\text{A.9})$$

Clearly the PDF given by eq (A.9) is of the form of eq (A.1). Hence, the sufficient part of the property follows.

An alternate proof of this property can be obtained by using the representation theorem. The representation theorem allows us to express the SSRV $\mathbf{X}$ as a product of a Gaussian random vector $\mathbf{Z}$ having zero mean and identity covariance matrix and a non-negative random variable

S. More precisely, we can write

$$\mathbf{X} = \mathbf{Z}S. \quad (\text{A.10})$$

The components of $\mathbf{X}$ can be statistically independent if and only if S is a constant. When S is a constant, $\mathbf{X}$ is a Gaussian SSRV. As is often the case, the representation theorem provides a simplified approach for determining properties of SIRVs.

A.2 Spherically Symmetric Characteristic function

In this section, we prove that the characteristic function of an SSRV is spherically symmetric.

Proof: We consider the SSRV $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$. From the representation theorem, we can write $\mathbf{X} = \mathbf{Z}S$ where $\mathbf{Z}$ is a Gaussian random vector having zero mean and identity covariance matrix and S is a non-negative random variable with PDF $f_S(s)$. The characteristic function of $\mathbf{X}$ given by

$$\Phi_{\mathbf{X}}(\omega) = E[\exp(j\omega^T \mathbf{X})] \quad (\text{A.11})$$

where $\omega = [\omega_1, \omega_2, \dots, \omega_N]^T$ can be expressed as

$$\Phi_{\mathbf{X}}(\omega) = E_S[\Phi_{\mathbf{X}|S=s}(\omega)] \quad (\text{A.12})$$

where $\Phi_{\mathbf{X}|S=s}(\omega) = E[\exp(j\omega^T \mathbf{Z}s)]$. However,

$$E[\exp(j\omega^T \mathbf{Z}s)] = \exp\left(-\frac{s^2}{2} \sum_{i=1}^N \omega_i^2\right). \quad (\text{A.13})$$

Using eq (A.13) in eq (A.12) results in

$$\Phi_{\mathbf{X}}(\omega) = \int_0^\infty \exp\left(-\frac{s^2}{2} \sum_{i=1}^N \omega_i^2\right) f_S(s) ds. \quad (\text{A.14})$$

The characteristic function given by eq (A.14) can be expressed as a function of $\sqrt{\omega^T \omega}$. Hence it is spherically symmetric.

A.3 Relationship Between Higher Order and Lower Order SIRV PDFs

In this section we examine the relationship between the higher order and lower order SIRV PDFs. More precisely we consider an SIRV $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$ having mean vector μ , covari-

ance matrix Σ and characteristic PDF $f_S(s)$. The PDF of $\mathbf{Y}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} h_N(p) \quad (\text{A.15})$$

where $p = (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu)$ and

$$h_N(p) = \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds. \quad (\text{A.16})$$

The vector $\mathbf{Y}$ can be partitioned as $\mathbf{Y} = [\mathbf{Y}_1^T \ \mathbf{Y}_2^T]^T$ where $\mathbf{Y}_1 = [Y_1, Y_2, \dots, Y_m]^T$ and $\mathbf{Y}_2 = [Y_{m+1}, Y_{m+2}, \dots, Y_N]^T$. Let μ_1 and μ_2 denote the mean vectors of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ respectively, and Σ_1 and Σ_2 denote the corresponding covariance matrices. We need to obtain the PDF of $\mathbf{Y}_1$ from the PDF of $\mathbf{Y}$ by integrating out over the $N - m$ random variables (i.e., the components of $\mathbf{Y}_2$). Let $p_1 = (\mathbf{y}_1 - \mu_1)^T \Sigma_1^{-1} (\mathbf{y}_1 - \mu_1)$ and $p_2 = (\mathbf{y}_2 - \mu_2)^T \Sigma_2^{-1} (\mathbf{y}_2 - \mu_2)$. The PDF of $\mathbf{Y}_1$ is given by

$$f_{\mathbf{Y}_1}(\mathbf{y}_1) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_0^\infty s^{-N} \exp\left(-\frac{p}{2s^2}\right) f_S(s) ds d\mathbf{Y}_2. \quad (\text{A.17})$$

From [38] (p17 eq.8, p18 eq.11) we have

$$(2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{p}{2s^2}\right) d\mathbf{Y}_2 = (2\pi)^{-\frac{m}{2}} |\Sigma_1|^{-\frac{1}{2}} s^{N-m} \exp\left(-\frac{p_1}{2s^2}\right). \quad (\text{A.18})$$

Using eq (A.18) in eq (A.17) gives

$$f_{\mathbf{Y}_1}(\mathbf{y}_1) = (2\pi)^{-\frac{m}{2}} |\Sigma_1|^{-\frac{1}{2}} \int_0^\infty s^{-m} \exp\left(-\frac{p_1}{2s^2}\right) f_S(s) ds. \quad (\text{A.19})$$

The PDF of $\mathbf{Y}_1$ can be expressed as

$$f_{\mathbf{Y}_1}(\mathbf{y}_1) = (2\pi)^{-\frac{m}{2}} |\Sigma_1|^{-\frac{1}{2}} h_m(p_1) \quad (\text{A.20})$$

where

$$h_m(p_1) = \int_0^\infty s^{-m} \exp\left(-\frac{p_1}{2s^2}\right) f_S(s) ds. \quad (\text{A.21})$$

Clearly, $h_m(p_1)$ given by eq (A.21) can be obtained from eq (A.16) by simply replacing N by m and p by p_1 . To determine the PDF of $\mathbf{Y}_1$, all that is needed is the specification of its mean vector and covariance matrix. As a special case, when $m = 1$, eq (A.19) gives us the first order SIRV PDF. Therefore, to obtain the first order SIRV PDF of the i^{th} component of $\mathbf{Y}$ starting

from the N^{th} order SIRV PDF, we simply use eq (A.19) with $m = 1$, $\Sigma_1 = \sigma_i$ and $p_1 = \frac{(y_i - \mu_i)^2}{\sigma_i^2}$.

Appendix B

Computer Generation of SIRVs Using the Rejection Method

B.1 Rejection Method

We present a proof of the rejection procedure [42] used for generating the norm R of the white SIRV $\mathbf{X}$ in Chapter 4. In many instances, it is likely that the PDF of a random variable is known explicitly, but its cumulative distribution function is either unknown or has a complicated functional form. Consequently, the cumulative distribution function cannot be inverted easily. Therefore, the use of the inverse distribution function for generating the random variable does not offer a practical solution for this problem. Hence, it is necessary to use a different scheme for generating the random variable. We consider the problem of generating a sequence of random numbers with PDF $f_R(r)$ of a random variable R , in terms of a random number sequence with PDF $f_{U_1}(u_1)$ of a random variable U_1 . The underlying assumption is that the random number sequence from the PDF of U_1 can be readily generated.

The rejection method used in Chapter 4 is based on the relative frequency interpretation of the conditional PDF

$$f_{U_1}(u_1|\mathcal{M})du_1 = \frac{P\{u_1 < U_1 \leq u_1 + du_1, \mathcal{M}\}}{P(\mathcal{M})} \quad (\text{B.1})$$

of a random variable U_1 given the event $\mathcal{M}$. $\mathcal{M}$ is expressed in terms of the random variable U_1 and another random variable U_2 and is chosen so that the resulting conditional PDF $f_{U_1}(u_1|\mathcal{M})$ equals $f_R(r)$. The desired sequence is generated by setting $R = U_1$ given that the event $\mathcal{M}$ has occurred and rejecting U_1 otherwise. The problem has a solution only if the domains of r and u_1 are such that $f_R(r) = 0$ in every interval for which $f_{U_1}(u_1) = 0$. Therefore, we can assume that

the ratio $\frac{f_{U_1}(u_1)}{f_R(u_1)}$ is bounded from below by some positive constant a :

$$\frac{f_{U_1}(u_1)}{f_R(u_1)} \geq a > 0 \text{ for every } u_1 \quad (\text{B.2})$$

B.2 Rejection Theorem

It is desired to generate a random variable R with PDF $f_R(r)$. Let U_1 be any random variable with PDF $f_{U_1}(u_1)$ such that $f_{U_1}(u_1) = 0$ whenever $f_R(r) = 0$. Let U_2 be a uniformly distributed random variable on the interval $(0, 1)$. If the random variables U_1 and U_2 are statistically independent and

$$\mathcal{M} = \{U_2 \leq g(U_1)\} \quad (\text{B.3})$$

where

$$g(u_1) = a \frac{f_R(u_1)}{f_{U_1}(u_1)} \leq 1, \quad (\text{B.4})$$

then

$$f_{U_1}(u_1|\mathcal{M}) = f_R(u_1). \quad (\text{B.5})$$

Proof: The joint PDF of the random variables U_1 and U_2 can be written as $f_{U_1, U_2}(u_1, u_2) = f_{U_1}(u_1)f_{U_2}(u_2)$, since U_1 and U_2 are statistically independent. Hence, we have

$$P(\mathcal{M}) = \int_{-\infty}^{\infty} \int_0^{g(u_1)} f_{U_1}(u_1)f_{U_2}(u_2)du_1du_2. \quad (\text{B.6})$$

However, since U_2 is uniformly distributed in the interval $(0, 1)$ and $g(u_1) \leq 1$,

$$\int_0^{g(u_1)} f_{U_2}(u_2)du_2 = g(u_1). \quad (\text{B.7})$$

Using eq (B.7) in eq (B.6) gives

$$P(\mathcal{M}) = \int_{-\infty}^{\infty} g(u_1)f_{U_1}(u_1)du_1. \quad (\text{B.8})$$

However, $g(u_1) = a \frac{f_R(u_1)}{f_{U_1}(u_1)}$. Therefore, we have

$$P(\mathcal{M}) = a \int_{-\infty}^{\infty} f_R(u_1)du_1 = a. \quad (\text{B.9})$$

We can express the numerator of eq (B.1) as

$$P\{u_1 < U_1 \leq u_1 + du_1, \mathcal{M}\} = \int_0^{g(u_1)} f_{U_1}(u_1) f_{U_2}(u_2) du_1 du_2 = g(u_1) f_{U_1}(u_1) du_1 = a f_R(u_1) du_1. \quad (\text{B.10})$$

Using eqs (B.9) and (B.10) in eq (B.1) results in eq (B.5).

Thus, we have the following algorithm for generating the sequence of random numbers from the PDF of R .

1. Generate U_1 and U_2 .
2. If $U_2 \leq a \frac{f_R(u_1)}{f_{U_1}(u_1)}$, then $U_1 = R$
3. Otherwise reject U_1 .

With reference to the generation of the norm R in Chapter 4, U_1 and U_2 were uniformly distributed random variables. Let c denote the maximum value of the PDF of R and b denote a finite range for the PDF of R such that the area under the curve of the PDF is close to unity. U_1 is assumed to be uniformly distributed in the interval $(0, b)$. Clearly, $\frac{f_{U_1}(u_1)}{f_R(u_1)} \geq \frac{1}{bc}$. Hence, $\frac{f_R(u_1)}{bc f_{U_1}(u_1)} \leq 1$. Therefore, $a = \frac{1}{bc}$. Step 2 above becomes: If $U_2 \leq \frac{f_R(u_1)}{bc f_{U_1}(u_1)} = \frac{f_R(u_1)}{c}$, then $U_1 = R$. This can be rewritten as: If $cU_2 \leq f_R(u_1)$, then $U_1 = R$. For ease of implementation, this latter form is used in conjunction with a uniform random variable U'_2 that is uniformly distributed over the interval $(0, c)$. This is the procedure followed in Chapter 4.

The method used in Chapter 4 becomes inefficient if U_1 is rejected frequently in step 3, resulting in the necessity to generate the two uniformly distributed random variables of step 1 an inordinate number of times. This problem can be overcome by using for U_1 a PDF which bounds the PDF of R and satisfies the conditions stated in section B.1 and in the rejection theorem. Then a random variable from this PDF is used in step 1 instead of the uniform random variable U_1 .

A second drawback of using a uniformly distributed random variable U_1 is that it may not be possible to efficiently generate SIRVs of length greater than 8. This is due to the fact that the PDF of R depends on N . Consequently, the uniform distribution for U_1 may not satisfactorily bound the PDF of the norm R for all N . This drawback can be overcome by choosing a different PDF for U_1 for each choice of N , such that the conditions stated in section B.1 and in the rejection theorem are satisfied. This method would require the use of an exhaustive table which tabulates the appropriate PDF of U_1 for each desired value of N .

Finally, it is pointed out that by using a composite function for the PDF of U_1 , it is possible to improve the simulation procedure in terms of being able to generate random numbers from the body and the tail of the PDF of R . These issues are suitable topics for future investigation as an extension of this work.

Appendix C

C.1 Limiting Forms for the Largest Order Statistic

Let $X_1 \leq X_2 \leq \dots \leq X_n$ be the ordered statistics of n random variables having a common distribution function $F(x)$. Assuming that the trials of drawing the random variables from the distribution function $F(x)$ are independent, the distribution function of the largest order statistic X_n is given by

$$\begin{aligned} P(X_n \leq x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= F^n(x). \end{aligned} \quad (\text{C.1})$$

When F is continuous but unknown, an asymptotic theory is developed for F in the range 0_+ to 1_- [68]. It is shown that positive sequences $\{a_n\}$ and $\{b_n\}$ exist such that

$$\lim_{n \rightarrow \infty} P\left(\frac{X_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} P(X_n \leq a_n x + b_n) \rightarrow \Lambda(x) \quad (\text{C.2})$$

or equivalently, by means of equation (C.1), that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) \rightarrow \Lambda(x). \quad (\text{C.3})$$

Let $n = md$ in equation (C.3). d is a fixed positive constant so that as $n \rightarrow \infty$, $m \rightarrow \infty$. Using the fact that $n = md$, we can write

$$\lim_{m \rightarrow \infty} F^{md}(a_{md}x + b_{md}) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) \rightarrow \Lambda(x). \quad (\text{C.4})$$

It is also true that

$$\lim_{m \rightarrow \infty} [F^m(a_m x + b_m)]^d = \lim_{m \rightarrow \infty} F^{md}(a_m x + b_m) \rightarrow \Lambda^d(x). \quad (\text{C.5})$$

If equations (C.4) and (C.5) hold, then from a theorem of Hintchin [76], there exist numbers $A_d > 0$ and $B_d > 0$ such that

$$\Lambda^d(A_d x + B_d) = \Lambda(x) \quad (\text{C.6})$$

for all integer values of d .

Solution of the above functional equation yields all the possible limiting forms for the distribution function $F^n(x)$. The constant A_d may or may not be unity. If it is unity, then the functional equation to be solved is given by

$$\Lambda^d(x + B_d) = \Lambda(x). \quad (\text{C.7})$$

On the other hand, if A_d is not unity, the form of equation (C.6) stands and there exists a value $x_{0d} = B_d/(1 - A_d)$ such that

$$\Lambda^d(x_{0d}) = \Lambda(x_{0d}). \quad (\text{C.8})$$

Constraining the solution to the above equation to be real and nonnegative, the solution is either $\Lambda = 0$ or 1. However, because $\Lambda(x)$ is a distribution function the value of Λ can be 0 only if x_{0d} is the lower endpoint at which $\Lambda(x_{0d}) = 0_+$ and Λ can be 1 only if x_{0d} is the upper end at which $\Lambda(x_{0d}) = 1_-$. Since A_d and B_d are assumed to be finite, x_{0d} must also be finite. Consequently, there is no loss in generality by assuming that the endpoint of interest is located at the origin (i.e., $x_{0d} = 0$). When $A_d \neq 1$, note that $x_{0d} = 0$ implies $B_d = 0$. As a result, the solutions for equation (C.6) fall into three cases which are given below.

$$1) \quad \Lambda^d(x + B_d) = \Lambda(x) \quad A_d = 1 \quad (\text{C.9})$$

$$2) \quad \Lambda^d(A_d x) = \Lambda(x) \quad A_d \neq 1 \quad F = 0 \text{ when } x = 0 \quad (\text{C.10})$$

$$3) \quad \Lambda^d(A_d x) = \Lambda(x) \quad A_d \neq 1 \quad F = 1 \text{ when } x = 0 \quad (\text{C.11})$$

C.1.1 Case 1

Case (1) of equation (C.9) is solved as follows. Taking the logarithm, we have

$$\log \Lambda(x) = d \log \Lambda(x + B_d). \quad (\text{C.12})$$

Multiplying through by a minus sign and taking the logarithm of both sides, we obtain

$$\log[-\log \Lambda(x)] = \log d + \log[-\log \Lambda(x + B_d)]. \quad (\text{C.13})$$

For simplicity, let

$$g(x) = \log[-\log \Lambda(x)]. \quad (\text{C.14})$$

Then equation (C.13) becomes

$$g(x) = \log d + g(x + B_d). \quad (\text{C.15})$$

Equivalently,

$$g(x - B_d) = \log d + g(x) \quad (\text{C.16})$$

or

$$g(x) = g(x - B_d) - \log d. \quad (\text{C.17})$$

Adding equations (C.15) and (C.17), we obtain

$$g(x + B_d) + g(x - B_d) = 2g(x). \quad (\text{C.18})$$

The above equation is valid for all x if and only if $g(x)$ is linear in x . Specifically, let

$$g(x) = kx + j \quad (\text{C.19})$$

where j and k are constants. Then

$$g(x + B_d) = k(x + B_d) + j = g(x) - \log d = kx + j - \log d. \quad (\text{C.20})$$

It follows that

$$kB_d = -\log d \quad \text{or} \quad k = -\frac{\log d}{B_d}. \quad (\text{C.21})$$

Substituting equation (C.21) in equation (C.19), we see that

$$g(x) + \frac{x \log d}{B_d} = j. \quad (\text{C.22})$$

Using equation (C.14), this result becomes

$$\log[-\log \Lambda(x)] + \frac{x \log d}{B_d} = j. \quad (\text{C.23})$$

Thus, we have

$$\log[-\log \Lambda(x)] = -\frac{x \log d}{B_d} + j. \quad (\text{C.24})$$

Hence, for case (1) of equation (C.9) to hold, $\log[-\log \Lambda(x)]$ must be linear in x .

We now solve for the sequence $\{B_d\}$. For this purpose, let $d = pq$ where p and q are both integers. Note that

$$\Lambda^{pq}(x + B_{pq}) = \Lambda(x). \quad (\text{C.25})$$

From the above equation we get

$$\begin{aligned} \Lambda(x + B_{pq}) &= \Lambda^{\frac{1}{pq}}(x) \\ &= [\Lambda^{\frac{1}{p}}(x)]^{\frac{1}{q}} = [\Lambda(x + B_p)]^{\frac{1}{q}} \\ &= \Lambda^{\frac{1}{q}}(x + B_p) = \Lambda((x + B_p) + B_q) = \Lambda(x + B_p + B_q). \end{aligned} \quad (\text{C.26})$$

Equation (C.26) implies that

$$B_{pq} = B_p + B_q. \quad (\text{C.27})$$

We now determine the functional dependence of the sequence $\{B_d\}$ on the subscript d . To emphasize this functional dependence, we rewrite equation (C.27) as

$$B(pq) = B(p) + B(q). \quad (\text{C.28})$$

From the above equation, it is clear that the functional dependence is logarithmic. Thus, the solution for B_d is given by

$$B(d) = B_d = \log d \quad (\text{C.29})$$

Substituting equation (C.29) into equation (C.24) yields

$$\log[-\log \Lambda(x)] = -x + j \quad (\text{C.30})$$

where j plays the role of a location parameter. Hence, without loss of generality, j is chosen to be zero. The above equation then simplifies to

$$\log[-\log \Lambda(x)] = -x. \quad (\text{C.31})$$

Solution for $\Lambda(x)$ results in

$$\Lambda(x) = \exp(-e^{-x}). \quad (\text{C.32})$$

Equation (C.32) is the solution of equation (C.9) for case 1.

C.1.2 Cases 2 and 3

The solutions to Cases (2) and (3) of equation (C.10) and (C.11) are now derived. In both cases we have

$$\Lambda^d(A_d x) = \Lambda(x). \quad (\text{C.33})$$

From equation (C.33) we get

$$\log \Lambda(x) = d \log \Lambda(A_d x). \quad (\text{C.34})$$

Multiplying through by a minus sign and taking the logarithm of both sides, we obtain

$$\log[-\log \Lambda(x)] = \log d + \log[-\log \Lambda(A_d x)]. \quad (\text{C.35})$$

As in case 1, let

$$g(x) = \log[-\log \Lambda(x)]. \quad (\text{C.36})$$

Then equation (C.35) becomes

$$g(x) = \log d + g(A_d x). \quad (\text{C.37})$$

Alternatively,

$$g\left(\frac{x}{A_d}\right) = \log d + g(x) \quad (\text{C.38})$$

or equivalently,

$$g(x) = -\log d + g\left(\frac{x}{A_d}\right). \quad (\text{C.39})$$

Adding equations (C.37) and (C.39) results in

$$g(A_d x) + g\left(\frac{x}{A_d}\right) = 2g(x). \quad (\text{C.40})$$

The solution to the above equation is

$$g(x) = \pm k \log x \quad \text{for } x > 0 \quad (\text{C.41})$$

and

$$g(x) = \pm k \log(-x) \quad \text{for } x < 0 \quad (\text{C.42})$$

where k is a positive constant. Use of equation (C.36) in equations (C.41) and (C.42) yields

$$\log[-\log \Lambda(x)] = \pm k \log x \quad \text{for } x > 0 \quad (\text{C.43})$$

$$\log[-\log \Lambda(x)] = \pm k \log(-x) \quad \text{for } x < 0. \quad (\text{C.44})$$

For Case 2, $\Lambda = 0$ when $x = 0$. This implies $x = 0$ is the lower end point of $\Lambda(x)$. Hence, $\Lambda(x)$ is nonzero for $x \geq 0$. Therefore, our solution is given by equation (C.43) where we must choose the sign in front of k to be negative. Then

$$\log[-\log \Lambda(x)] = -k \log x \quad x \geq 0 \quad (\text{C.45})$$

which results in

$$\Lambda(x) = \exp(-x^{-k}) \quad x \geq 0. \quad (\text{C.46})$$

For case 3, $\Lambda = 1$ when $x = 0$. This implies that $x = 0$ is the upper endpoint of $\Lambda(x)$. Hence, $\Lambda(x)$ is nonzero for $x \leq 0$. Consequently, the solution is given by equation (C.44) where we choose the sign in front of k to be positive. Then

$$\log[-\log \Lambda(x)] = k \log(-x) \quad x \leq 0 \quad (\text{C.47})$$

resulting in

$$\Lambda(x) = \exp(-(-x)^k) \quad x \leq 0. \quad (\text{C.48})$$

Thus, the three possible limiting forms for the distribution $\Lambda(x)$ that arise as solutions to equation 1 are given as follows:

$$1) \quad \Lambda(x) = \exp(-e^{-x}) \quad (\text{C.49})$$

$$2) \quad \Lambda(x) = \exp(-x^{-k}) \quad x \geq 0, \quad k > 0 \quad (\text{C.50})$$

$$3) \quad \Lambda(x) = \exp(-(-x)^k) \quad x \leq 0, \quad k > 0. \quad (\text{C.51})$$

C.2 Tails of Probability Density Functions

Equations (C.49-C.51) represent the three possible limiting forms of the distribution function for almost all smooth and continuous probability density functions. By differentiating the three functions, we obtain the three possible limiting forms for the probability density functions themselves.

C.2.1 Case 1

The derivative of $\Lambda(x)$ is given by

$$H(x) = \frac{d}{dx}\Lambda(x) = \exp(-e^{-x}) \cdot (-e^{-x})(-1) = e^{-x} \exp(-e^{-x}) = \exp(-x - e^{-x}). \quad (\text{C.52})$$

In our application we are interested in the right tail of the probability density function. Since we have to set thresholds corresponding to small false alarm probabilities, the thresholds will be in the right tail of the probability density function. When x is very large, $x \gg e^{-x}$. Therefore, equation (C.52) can be simplified to obtain the PDF of the tail as

$$H(x) = e^{-x} \quad x \text{ large}. \quad (\text{C.53})$$

C.2.2 Case 2

The derivative of $\Lambda(x)$ is given by

$$\begin{aligned} H(x) &= \frac{d}{dx}\Lambda(x) = \exp(-x^{-k}) \cdot (kx^{-k-1}) \\ &= k \exp(-x^{-k}) e^{(k-1)\log x} = k \exp(-x^{-k} - (k+1)\log x). \end{aligned} \quad (\text{C.54})$$

When x is very large $\log x \gg x^{-k}$. Therefore, equation (C.54) can be simplified to obtain the PDF of the tail as

$$H(x) = k e^{-(k+1)\log x} = k x^{-(k+1)} \quad x > 0, x \text{ large } k > 0. \quad (\text{C.55})$$

C.2.3 Case 3

The derivative of $\Lambda(x)$ for this case is given by

$$\begin{aligned} H(x) &= \frac{d}{dx}\Lambda(x) = \exp(-(-x)^k) \cdot (k(-x)^{k-1}) \\ &= k \exp(-(-x)^k) e^{(k-1)\log(-x)} = k \exp(-(-x)^k + (k-1)\log x). \end{aligned} \quad (\text{C.56})$$

When $-x$ is very large, $(-x)^k \gg \log x$. Therefore, equation (C.56) can be simplified to obtain the PDF of the tail as

$$H(x) = ke^{-(-x)^k} \quad x < 0, -x \text{ large } k > 0. \quad (\text{C.57})$$

A basic assumption in the above development is that successive trials are independent. This led to equation (C.1). In practice, as n becomes large, it may be difficult to ensure the independence of successive trials. To the extent that the assumption holds, the results in equations (C.49-C.51) are valid.

C.3 PDF of the r^{th} Ordered Statistic

Suppose that the ordered samples $X_1 \leq X_2 \leq \dots \leq X_n$ are drawn from the distribution function $F(x)$. Let us further assume that the trials used to draw the samples from the distribution are independent. Consider the r^{th} ordered statistic X_r . Recall that $P(X_r \leq x)$ is the distribution function of X_r . This, in turn, is the probability that at least r of the X_i 's are less than or equal to x . Treating this as a Binomial problem, the distribution function is

$$F_{X_r}(x) = P(X_r \leq x) = \sum_{i=r}^n \frac{n!}{i!(n-i)!} F^i(x) [1 - F(x)]^{n-i} \quad (\text{C.58})$$

where the i^{th} term in the summation is the binomial probability that exactly i of $X_1, X_2, \dots, X_n$ are less than or equal to x . Using integration by parts, it can be shown that equation (C.58) can be represented in terms of integral

$$F_{X_r}(x) = \frac{n!}{(r-1)!(n-r)!} \int_0^{F(x)} t^{r-1} (1-t)^{n-r} dt. \quad (\text{C.59})$$

The probability density function of the r^{th} ordered statistic is the derivative of $F_{X_r}(x)$ and is given by

$$\begin{aligned} f_{X_r}(x) = \frac{d}{dx} F_{X_r}(x) &= \frac{n!}{(r-1)!(n-r)!} \frac{d}{dx} \int_0^{F(x)} t^{r-1} (1-t)^{n-r} dt \\ &= \frac{n!}{(r-1)!(n-r)!} F^{r-1}(x) [1 - F(x)]^{n-r} f(x) \end{aligned} \quad (\text{C.60})$$

where $f(x) = \frac{d}{dx} F(x)$. Equation (C.60) represents the general form of the PDF of the r^{th} ordered statistic. If $F(x)$ is known, then the mean and the variance of the r^{th} ordered statistic can be

calculated. The expected value of X_r is given by

$$E(X_r) = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} x F^{r-1}(x) [1 - F(x)]^{n-r} f(x) dx. \quad (C.61)$$

An alternate form for the expected value of X_r can be obtained by letting $u = F(x)$. Therefore, $x = F^{-1}(u)$. The infinite limits of the integral in the above equation then becomes finite after the transformation. The transformed integral is

$$E(X_r) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 F^{-1}(u) u^{r-1} (1-u)^{n-r} du. \quad (C.62)$$

The variance of the r^{th} ordered statistic is expressed as

$$\text{Var}(X_r) = E[(X_r - E(X_r))^2] = E(X_r^2) - E^2(X_r). \quad (C.63)$$

Making use of equation (C.60), $E(X_r^2)$ can be written as follows.

$$E(X_r^2) = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} x^2 F^{r-1}(x) [1 - F(x)]^{n-r} f(x) dx. \quad (C.64)$$

An alternate form for the expected value of X_r can be obtained by again letting $u = F(x)$. We then get

$$E(X_r^2) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 [F^{-1}(u)]^2 u^{r-1} (1-u)^{n-r} du. \quad (C.65)$$

The variance of X_r can be calculated from equations (C.62) and (C.65) when $F^{-1}(u)$ is known.

Bibliography

- [1] L. E. Brennan and I. S. Reed. Theory of adaptive radar. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-9**:pp.237-252, 1973.
- [2] I.S. Reed, J.D. Mallett, and L.E. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-10**:pp.853-863, 1974.
- [3] E.J. Kelly. An adaptive detection algorithm. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-22**:pp. 115-127, 1986.
- [4] C. G. Khatri and C. R. Rao. Effects of estimated noise covariance matrix in optimum signal detection. *IEEE Trans. on Acoust., Speech, Signal Processing*, **ASSP-35**:pp.671-679, 1987.
- [5] L. Cai and H. Wang. On adaptive filtering with the cfar feature and its performance sensitivity to non-gaussian interference. *Proceedings of the 24th Annual conference on Information Sciences and Systems*, pages pp.558-563, 1990.
- [6] L. Cai and H. Wang. Performance comparisons of the modified smi and glr algorithms. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-27**, 1991.
- [7] H. Wang and L. Cai. On adaptive implementation of optimum mti in severely nonhomogeneous environments. *Proc. IEEE Int. Radar Conference*, pages pp.351-355, 1990.

- [8] H. Wang and L. Cai. A localized adaptive mtd processor. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-27**, 1991.
- [9] R. N. Adams, L. L. Horowitz, and K. D. Senne. Adaptive main-beam nulling for narrow beam antenna arrays. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-16**:pp.509-516, 1980.
- [10] R. A. Monzingo and T. W. Miller. *Introduction to Adaptive Arrays*. Wiley, New York, 1980.
- [11] J. E. Hudson. *Adaptive Array Principles*. Peter Peregrinus, New York, 1981.
- [12] H. Wang and L. Cai. On adaptive multiband signal detection with the glr algorithm. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-27**, 1991.
- [13] H. Wang et. al. Cfar performance of adaptive spatial-temporal processors. Technical report, Rome Laboratory, 1991.
- [14] Sekine M., Musha T., Tomita Y., Hagsawa T., Irabu T., and Kiuchi E. On Weibull Distributed Weather Clutter. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-15**:pp.824-828, 1979.
- [15] Jakeman E. and Pusey P.N. A model for non-Rayleigh Sea Echo. *IEEE Trans. on Antennas and Propagation*, **AP-24**:pp.806-814, 1976.
- [16] Watts S. and Ward K.D. Spatial Correlation in K-distributed Sea Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **134**, (6):pp.526-532, 1987.
- [17] Hawkes C.W. and Haykin S.S. Modeling of Clutter for Coherent Pulsed Radar. *IEEE Trans. on Information Theory*, **IT-21**:pp.703-707, 1975.
- [18] Farina A., Russo A., and Studer F.A. Coherent Radar Detection in Log-normal Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **133**, (1):pp.39-54, 1986.

- [19] Conte E. and Longo M. On a Coherent Model For Log-normal Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **134**, (2):pp.198-201, 1987.
- [20] Watts S. Radar Detection Prediction in K-distributed Sea Clutter and Thermal Noise. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-23**:pp.40-45, 1987.
- [21] Gang Li and Kai-Bor Yu. Modeling and Simulation of Coherent Weibull Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **136**, (1):pp.2-12, 1989.
- [22] Jakeman E. On the Statistics of K-distributed Noise. *J. Phys. A.*, **13**:pp.31-48, 1980.
- [23] Szajnowski W.J. The Generation of Correlated Weibull Clutter for Signal Detection Problems. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-13**:pp.536-540, 1977.
- [24] Peebles P. The Generation of Correlated Log Normal Clutter for Radar Simulation. *IEEE Trans. on Aerospace and Electronic Systems*, **AES-7**:pp.1215-1217, 1971.
- [25] Farina A., Russo A., Scannapieco F., and Barbarossa S. Theory of Radar Detection in Coherent Weibull Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **134**, (2):pp.174-190, 1987.
- [26] Conte E., Galati G., and Longo M. Exogenous Modeling of Non-Gaussian Clutter. *J. IERE*, **57**, (4):pp.151-155, 1987.
- [27] Kingman J.F.C. Random Walks with Spherical Symmetry. *Acta Math.*, **109**:pp.11-53, 1963.
- [28] Yao K. A Representation Theorem and Its Applications to Spherically Invariant Random Processes. *IEEE Trans. on Information Theory*, **IT-19**:pp.600-608, 1973.

- [29] Goldman J. Detection in the Presence of Spherically Symmetric Random Vectors. *IEEE Trans. on Information Theory*, **IT-22**:pp.52-58, 1976.
- [30] Brehm H. Description of Spherically Invariant Random Processes by Means of G-functions. *Springer Lecture Notes*, **969**:pp.39-73, 1982.
- [31] Brehm H. and Stammmler W. Description and Generation of Spherically Invariant Speech-Model Signals. *Signal Process.*, **12**, (2):pp.119-141, 1987.
- [32] Conte E. and Longo M. Characterization of Radar Clutter as a Spherically Invariant Random Process. *IEE Proc.F, Commun., Radar, & Signal Process.*, **134**, (2):pp.191-197, 1987.
- [33] Conte E., Longo M., and Lops M. Modelling and Simulation of non-Rayleigh Radar Clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, **138**, (2):pp.121-130, 1991.
- [34] Rangaswamy M., Weiner D.D., and Ozturk A. Spherically Invariant Random Processes for Modeling and Distribution Identification of Non-Gaussian Random Vectors. Accepted for publication in **IEEE-AES** trans.
- [35] Rangaswamy M., Weiner D.D., and Ozturk A. Computer Generation of Correlated Non-Gaussian Clutter for Radar Signal Detection. Accepted for publication in **IEEE-AES** trans.
- [36] Kingman J.F.C. On Random sequences with Spherical Symmetry. *Biometrika*, 59:pp.492-494, 1972.
- [37] Johnson M.E. *Multivariate Statistical Simulation*. John Wiley and sons, New York, 1987.
- [38] Miller K.S. *Multidimensional Gaussian Distributions*. Wiley, New York, 1964.

- [39] Blake I.F. and Thomas J.B. On a Class of Processes Arising in Linear Estimation Theory. *IEEE Trans. on Information Theory*, IT-14:pp.12-16, 1968.
- [40] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications Inc., New York, 1972.
- [41] Chu K.C. Estimation and Decision for Linear Systems with Elliptical Random Processes. *IEEE Trans. on Automatic Control*, AC-18:pp.499-505, 1973.
- [42] Papoulis A. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, 1984.
- [43] Frame J.S. Matrix Functions and Applications Part-1-Matrix Operations and Generalized Inverses. *IEEE Spectrum*, 3:pp.212, 1964.
- [44] Picinbono B. Spherically Invariant and Compound Gaussian Stochastic Processes. *IEEE Trans. on Information Theory*, IT-16:pp.77-79, 1970.
- [45] Gradshteyn I.S. and Ryzhik I.M. *Table of Integrals, Series and Products*. Academic Press Inc., New York, 1980.
- [46] Erdelyi A., Magnus W., and Oberhettinger F. *Tables of Integral Transforms*. McGraw-Hill, New York, 1954.
- [47] Modestino J.W. and Ningo A.Y. Detection of Weak Signals in Narrowband Non-Gaussian Noise. *IEEE Trans. on Information Theory*, IT-25:pp.592-600, 1979.
- [48] Kassam S.A. *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, New York, 1988.
- [49] B.L. Bratley, P. and Fox and L.E. Schrage. *A Guide to Simulation*. Springer-Verlag, New York, 1987.

- [50] Ozturk A. A New Method For Univariate and Multivariate Distribution Identification. Submitted for publication to JASA.
- [51] Koziol J.A. A class of Invariant Procedures for Assessing Multivariate Normality. *Biometrika*, 69:pp.423-427, 1982.
- [52] Ozturk A. and Romeu J.L. A New Method For Assessing Multivariate Normality With Graphical Applications. Accepted for Publication in *Commun. in Statistics*.
- [53] Johnson N.L. and Kotz S. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and sons, New York, 1976.
- [54] Martinez A.B., Swaszek P.F., and Thomas J.B. Locally Optimal Detection in Multivariate Non-Gaussian Noise. *IEEE Trans. on Information Theory*, IT-30:pp.815-822, 1984.
- [55] Gnanadesikan R. *Methods of Statistical Data Analysis of Multivariate Observations*. John Wiley and sons, New York, 1977.
- [56] Mardia K.V. Test of Univariate and Multivariate Normality. *Handbook of Statistics*, 1:pp.279-320, 1980.
- [57] Rangaswamy M., Weiner D.D., and Ozturk A. Simulation of Correlated Non-Gaussian Interference for Radar Signal Detection. In *Proceedings of twentyfifth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 1991.
- [58] Fisher R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann.Eugenics*, 7:pp.179-188, 1936.
- [59] Fang K.T. and Anderson T.W. *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press inc., New York, 1990.

- [60] J. Neyman and E. S. Pearson. Contributions to the theory of testing statistical hypothesis. *Annals of Mathematics*, pages pp. 25-138, 1930.
- [61] J. Neyman and E. S. Pearson. Sufficient statistics and uniformly most powerful tests of statistical hypothesis. *Annals of Mathematics*, pages pp. 113-137, 1930.
- [62] Middleton D. Canonically optimum threshold detection. *Memorandum RM-4687-PR*, 1965.
- [63] Capon J. On the asymptotic efficiency of locally optimum detectors. *IRE Transactions on Information Theory*, pages pp. 67-71, 1961.
- [64] J. H. Miller and J. Thomas. Detectors for discrete time signals in non-gaussian noise. *IEEE Trans. on Information Theory*, IT-18, No. 2:pp. 241-250, 1972.
- [65] E. Conte, L. Izzo, M. Longo, and L. Paura. Asymptotically optimum radar detectors in non-rayleigh clutter. *IEE Proc.F, Commun., Radar, & Signal Process.*, 134, No. 7 :pp. 667-672, 1987.
- [66] B. B. Shishkov and S. I. Penev. Asymptotically optimum algorithms for detection of signals in a background of correlated interference and white noise. *Radiotekhnika i elektronika*, 7:pp. 1419-1424, 1988.
- [67] D. D. Boos. Using extreme value theory to estimate large samples. *Technometrics*, 20:pp. 33-39, 1984.
- [68] Editor: Tiago de Oliveira. *Statistical Extremes and Applications*. D. Reidel Publishing Co., Boston, 1984.
- [69] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*, 52:pp. 393-442, 1990.

- [70] W. H. DuMouchel. Estimating the stable index α in order to measure tail thickness: A critique. *The Annals of Statistics*, 11:pp. 1019-1031, 1983.
- [71] J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation of the generalized pareto distribution. *Technometrics*, 29, No. 3:pp. 339-349, 1987.
- [72] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:pp. 308-313, 1965.
- [73] Gradshteyn I. S. and I. M. Ryzhik. *Tables of Integrals, Series and Products*. Academic Press, NY, 1980.
- [74] S. A. Kassam. *Signal Detection in Non-Gaussian Noise*. Springer Verlag, NY, 1989.
- [75] A. B. Martinez, P. F. Swaszek, and J. B. Thomas. Locally optimum detection in multivariate non-gaussian noise. *IEEE Trans. on Information Theory*, IT-30, No. 6:pp. 815-822, 1984.
- [76] J. Galambos. *Asymptotic Theory of Extreme Order Statistics*. Wiley, NY, 1978.

**MISSION
OF
ROME LABORATORY**

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C³I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C³I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.