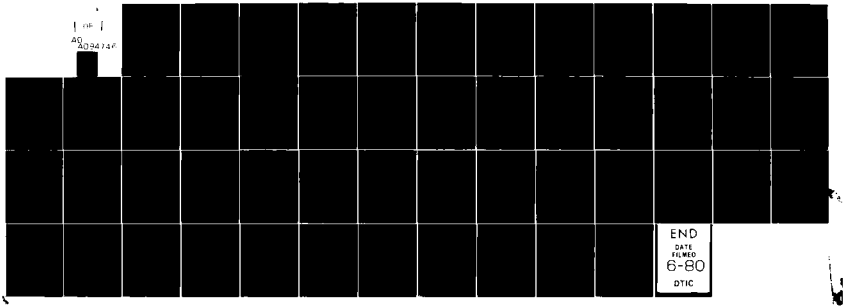
AD-A084 746 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 20/10
DATA MODELING USING QUANTILE AND DENSITY-QUANTILE FUNCTIONS.(U)
APR 80 E PARZEN DAAG29-80-C-0070
UNCLASSIFIED TR-8-2 ARO-16998.2-M NL



TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

DATA MODELING USING QUANTILE AND DENSITY-QUANTILE FUNCTIONS

by Emanuel Parzen
Institute of Statistics, Texas A&M University

Texas A & M Research Foundation
Project No. 4226

"Robust Statistical Data Analysis and Modeling"
Sponsored by the U. S. Army Research Office
Grant (15) DAAG29-80-C-0070

Professor Emanuel Parzen, Principal Investigator

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report No. 3-2 / 2. GOVT ACCESSION NO.: AD-A084746	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Data Modeling Using Quantile and Density-Quantile Functions	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Emanuel Parzen	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0070 /
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P. O. Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: April 1980
	13. NUMBER OF PAGES: 47
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quantile function, density-quantile function, score function, sample quantile function, sample quantile-density function, histogram-quantile function, sample entropy, score deviation, tail exponents, mode percentile, quantile box plot, cumulative weighted spacings plot, quantile bootstrap, minimum residual score deviation estimation, and 19 quantile values for universal data summary.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Secs. 1-3 introduce the role of quantile functions in statistical modeling, the sample quantile function, and location and scale parameter models. Quantile function based descriptors of a probability distribution are defined (Sec. 4). Sec. 5 defines quantile box plots and transformation distribution functions; an example of their application is discussed in Sec. 6. A quantile version of "bootstrap" simulation methods is outlined in Sec. 7. Data summary by a few values of the sample quantile function is discussed (Sec. 9). Sec. 8 discusses quantile function formulations of robust estimators of location and scale.

DTIC

DATA MODELING USING QUANTILE AND DENSITY-QUANTILE FUNCTIONS

by Emanuel Parzen

Texas A&M University
Institute of Statistics

Abstract

Statistical data modeling is a field of statistical reasoning that seeks to fit models to data without using models based on prior theory; rather one seeks to learn the model by a process which could be called statistical model identification. When analyzing a sample $X_1, \ldots, X_n$, statisticians should not confine themselves to either fitting a Gaussian distribution, or transforming the data to be Gaussian. Such an approach ignores the importance of bimodality as a feature of observed data, and also ignores the need to fit to data probability model based distributions which could suggest probability models for the causes generating the data. This paper describes an approach to statistical data modeling which emphasizes estimation of quantile and density-quantile functions; it treats the Gaussian distribution as just one of the available distributions.

---

*Emanuel Parzen is Distinguished Professor at the Institute of Statistics, Texas A & M University, College Station, Texas, 77843. This research was supported in part by the Army Research Office (Grant DAAG29-80-C-0070).

Sections 1-3 introduce the role of quantile functions in statistical modeling, the sample quantile function, and location and scale parameter models. Quantile function based descriptors of a probability distribution are defined (Section 4). Section 5 defines quantile box plots and transformation distribution functions; an example of their application is discussed in Section 6. A quantile version of "bootstrap" simulation methods is outlined in Section 7. Data summary by a few values of the sample quantile function is discussed (Section 9). Section 8 discusses quantile function formulations of robust estimators of location and scale.

The concepts discussed in this paper are best summarized by a list of some of the terminology defined: quantile function, density-quantile function, score function, sample quantile function, sample quantile-density function, histogram-quantile function, sample entropy, score deviation, tail exponents, mode percentile, quantile box plot, cumulative weighted spacings plot, quantile bootstrap, minimum residual score deviation estimation, and 19 quantile values for universal data summary.

1. Some Basic Concepts of Statistical Modeling and Estimation

One of the basic problems of statistical data analysis is the one-sample problem: given a sample $X_1, \ldots, X_n$ which we assume initially to be independent observations of a population characteristic represented by a random variable X, we would like to infer the probability distribution of X.

The probability distribution of X is usually represented by its distribution function

$$F(x) = \Pr [X \leq x]$$

and by its probability density function

$$f(x) = F'(x) .$$

In this paper we assume X is continuous and possesses a probability density function.

The problem of statistical inference is often defined to be parameter estimation; then one assumes that the true probability density function f(x) belongs to a family of functions $f_\theta(x)$ indexed by a vector θ of parameters $\theta_1, \ldots, \theta_r$.

The maximum likelihood estimator of θ is defined to be a function $\hat{\theta}$ of $X_1, \ldots, X_n$ satisfying $L(\hat{\theta}) = \max_\theta L(\theta)$, defining

$$L(\theta) = f_\theta(X_1, \ldots, X_n) = \prod_{j=1}^{n} f_\theta(X_j) ;$$

$L(\theta)$ is the joint probability density of the observed data when θ is the true parameter value.

Maximum likelihood estimation is not a principle to be accepted uncritically; statisticians delight in constructing examples in which it leads to unbelievable conclusions. To understand when and why maximum likelihood estimation works, we have to introduce empirical distribution function (EDF) $\tilde{F}(x)$ defined by

$$\tilde{F}(x) = \text{fraction of } X_1, \ldots, X_n \leq x$$

To graph $F(x)$, one determines the order statistics $X_{(1)} \leq X_{(2)} , < \ldots < X_{(n)}$ which are the sample values (assumed to be distinct) arranged in increasing order; then

$$F(x) = \frac{j}{n} , \quad X_{(j)} \leq x < X_{(j+1)} , \quad j = 0, 1, 2, \ldots, n$$

where $X_{(0)} = -\infty$ and $X_{(n+1)} = \infty$.

The concept of likelihood is now defined as average log likelihood

$$L_n(\theta) = \frac{1}{n} \log \prod_{j=1}^{n} f_\theta(X_j)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \log f_\theta(X_j)$$

$$= \int_{-\infty}^{\infty} \log f_\theta(x) \, d\tilde{F}(x)$$

One can regard $L_n(\theta)$ as a measure of "distance" between the data represented by $\tilde{F}$, and the model represented by $f_\theta(x)$.

Another important interpretation of $L_n(\theta)$ is an estimator of a "distance" between the true probability density f(x) and the model $f_\theta(x)$. An important role in the theory of statistical inference is played by the Kullback-Liebler information number, or directed divergence (see Zacks (1971); it is defined by

$$I(f;f_\theta) = E_f\left[\log \frac{f}{f_\theta}\right]$$

$$= \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{f_\theta(x)}\, dx$$

$$= H(f;f) - H(f;f_\theta)$$

defining

$$H(f;g) = \int_{-\infty}^{\infty} f(x) \log g(x)\, dx \ .$$

It has the properties: $I(f;f_\theta) \geq 0$ and $I(f;f) = 0$.
The average directed divergence between f and f_θ given a sample $X_1, \ldots, X_n$ is

$$I_n(f;f_\theta) = \frac{1}{n} E_f \log \frac{f(X_1, \ldots, X_n)}{f_\theta(X_1, \ldots, X_n)}$$

$$= \frac{1}{n}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1,\ldots,x_n) \log \frac{f(x_1,\ldots,x_n)}{f_\theta(x_1,\ldots,x_n)}\, dx_1 \cdots dx_n$$

$$= I(f;f_\theta) \ .$$

A criterion for model fitting is to choose f_θ to minimize $I(f;f_\theta)$ or an estimator of $I(f;f_\theta)$; an estimator would be

$$I(\tilde{f};f_\theta) = H(\tilde{f};\tilde{f}) - H(\tilde{f};f_\theta)$$

if $\tilde{f}$ were a non-parametric estimator of f. While $\tilde{F}$ is a natural non-parametric estimator of F, there does not exist a natural non-parametric estimator of f. However a natural non-parametric estimator $\tilde{H}(f;f_\theta)$ of $H(f;f_\theta)$ does exist, namely the average log likelihood $L_n(\theta)$; in symbols,

$$\tilde{H}(f;f_\theta) = \int_{-\infty}^{\infty} \log f_\theta(x)\, d\tilde{F}(x)$$

A natural estimator $\tilde{H}(f;f)$ will be given below. Akaike (1973) has pioneered in emphasizing that to find $\hat{\theta}$, the parameter values θ which minimize

$$\tilde{I}(f;f_\theta) = \tilde{H}(f;f) - \tilde{H}(f;f_\theta) ,$$

it is not necessary to know $\tilde{H}(f;f)$; one need only choose $\hat{\theta}$ to maximize $L_n(\theta)$. One approach to measuring how well the maximum likelihood model $f_{\hat{\theta}}$ "matches" the data would be to measure how significantly different from zero is $\tilde{I}(f;f_{\hat{\theta}})$. Other approaches to measuring the mathematical fit of a model to data are introduced in this paper using various representing functions of the data and model which are called the "raw" and "smooth" representing functions respectively. One of our goals is to develop means of judging goodness of fit of a family f_θ of probability densities to a true probability density <u>before</u> forming

estimators $\hat{\theta}$ of the parameters.

This paper discusses the increased insight to be obtained by describing the probability distribution of a random variable X by its quantile function $Q(u)$, $0 \leq u \leq 1$, and density-quantile function $fQ(u)$, $0 \leq u \leq 1$. Define

$$Q(u) = F^{-1}(u) = \inf \{x:F(x) \geq u\} \quad ,$$

$$fQ(u) = f(Q(u)) \ .$$

The quantile-density function $q(u)$, $0 \leq u \leq 1$, is the derivative of the quantile function:

$$q(u) = Q'(u) \ .$$

The score function is (-1) times the derivative of the density-quantile function:

$$J(u) = -(fQ)'(u)$$

An important identity is

$$fQ(u)\ q(u) = 1$$

which follows by differentiating the identity

$$FQ(u) = u \ .$$

We can now give an example of the advantages of thinking "quantile" in the sense of thinking in terms of $fQ(u)$ rather than $f(x)$. Two measures of the smoothness of a function are

the integral of its logarithm and the integral of its derivative squared. Thus

$$\int_0^1 \log fQ(u) = \int_{-\infty}^{\infty} f(x) \log f(x)\, dx = H(f;f)$$

is the Shannon information measure or entropy of f, while the Fisher information measure of f is

$$\int_0^1 |J(u)|^2 du = \int_0^1 |fQ'(u)|^2\, du$$

$$= \int_0^1 \left|\frac{f'Q(u)}{fQ(u)}\right|^2 du = \int_{-\infty}^{\infty} \frac{|f'(x)|^2}{f(x)}\, dx$$

One can give a natural estimator of entropy:

$$\tilde{H}(f;f) = -\int_0^1 \log \tilde{q}(u)\, du$$

where $\tilde{q}(u)$ is the sample quantile-density defined below. We call $\tilde{H}$ the sample entropy.

The density quantile function as a function of interest for itself was introduced by Parzen (1979). Tukey (1965) pointed out the significance of Q(u) and q(u) under the names "representing" function and "sparsity" function. A review of some standard approaches to statistical modelling is given by Ord and Patil (1975).

2. Sample Quantile Function

To a batch of data one can define a sample quantile function $\tilde{Q}(u)$, $0 \leq u \leq 1$, which provides a "universal" description and summary of the data. However, there is no universally accepted definition of $\tilde{Q}(u)$.

Given a sample $X_1, \ldots, X_n$, with order statistics $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ one could define $\tilde{Q}$ by

$$\tilde{Q}(u) = \tilde{F}^{-1}(u) , \quad 0 \leq u \leq 1 ;$$

then $\tilde{Q}$ is piecewise constant,

$$\tilde{Q}(u) = X_{(j)} , \quad \frac{j-1}{n} < u \leq \frac{j}{n} , \quad j = 1, 2, \ldots, n .$$

One often prefers a piecewise-linear definition of $\tilde{Q}(u)$; then one defines

$$\tilde{Q}(u) = X_{(j)} \text{ if } u = \frac{j-0.5}{n} \text{ or } \frac{j}{n+1} .$$

One also defines values for $u = 0$ or 1, say $\tilde{Q}(0) = X_{(1)}$ and $\tilde{Q}(1) = X_{(n)}$. At other values of u in $0 < u < 1$, one defines $\tilde{Q}(u)$ by linear interpolation of its values at the grid points $(j-0.5)/n$ or $j/(n+1)$. Then $\tilde{Q}(u)$ is differentiable, and $\tilde{q}(u) = \tilde{Q}'(u)$ may be expressed in terms of the sample "spacings"

$$n \{X_{(j+1)} - X_{(j)}\} .$$

When $\tilde{Q}(u) = X_{(j)}$ at $u = (j-0.5)/n$, then ($j = 1, 2, \ldots, n-1$),

$$\tilde{q}(u) = n \{X_{(j+1)} - X_{(j)}\} , \quad \frac{j-0.5}{n} < u < \frac{j+0.5}{n} .$$

A favorite tool of statistical data analysis is the histogram which can be defined as a piece-wise constant estimator $\tilde{f}(x)$ of the density function. The sample quantile function $\tilde{Q}(u)$ is then defined as the inverse of the sample distribution function $\tilde{F}(x) = \int_{-\infty}^{x} \tilde{f}(y)\, dy$. The insight in a histogram seems to me to be made more visible by plotting instead the histogram-quantile function $\tilde{f}(\tilde{Q}(u))$, $0 \leq u \leq 1$.

A raw estimator of $fQ(u)$, called a raw density-quantile function and denoted $\tilde{f}Q(u)$, can be formed from the reciprocal of a slightly smoothed estimator of $q(u)$; for example, one might define

$$\tilde{f}Q(u) = \frac{2h}{\tilde{Q}(u+2h) - \tilde{Q}(u-2h)}$$

The sample quantile function $Q(u)$, $0 \leq u \leq 1$, is a stochastic process (or time series) whose asymptotic distribution can be shown to satisfy (under suitable assumptions on fQ; see Csorgo and Revesz (1978)).

$$\{\sqrt{n} fQ(u)\{\tilde{Q}(u) - Q(u)\} \quad , \ 0 \leq u \leq 1\} \xrightarrow{L} \{B(u), \ 0 \leq u \leq 1\}$$

where $\{B(u), \ 0 \leq u \leq 1\}$, denotes a Brownian Bridge stochastic process with covariance function

$$E[B(u_1)B(u_2)] = u_1(1-u_2) \text{ for } 0 \leq u_1 \leq u_2 \leq 1 \ ,$$

$\overset{L}{=}$ denotes "identically distributed as", and the convergence is in the sense of convergence of distribution of stochastic processes.

The asymptotic distribution of the sample spacings, and thus of $\tilde{q}(u)$, also have been extensively investigated but is difficult to summarize briefly. One important fact is that for any fixed $u_1, \ldots, u_k$

$$fQ(u_1)\tilde{q}(u_1), \ldots, fQ(u_k)\tilde{q}(u_k)$$

are asymptotically independent and distributed as an exponential distribution with mean 1.

The difference between the roles of distribtuion functions and quantile functions in statistical inference is made clear by considering the basic goodness of fit problem: test the hypothesis H_0 ,

$$H_0: \quad F(x) = F_0(x) \ , \ -\infty < \infty < \infty \ ,$$

that the true distribution function $F(x)$ equals a specified distribution function $F_0(x)$. One could compare the sample distribution $\tilde{F}(x)$ to $F_0(x)$ (or equivalently test whether the transformed random variables $F_0(X_1), \ldots, F_0(X_n)$ are uniformly distributed) by comparing $\tilde{F}(Q_0(u))$ to u. The applicable asymptotic distribution theorem is

$$\{\sqrt{n}\{\tilde{F}(Q_0(u)) - u\}, \quad 0 \leq u \leq 1\} \xrightarrow{L} = \{B(u) \ , \ 0 \leq u \leq 1\}$$

Alternately one could compare quantile functions. Instead of comparing $\tilde{Q}(u)$ to $Q_0(u) = F_0^{-1}(u)$, one could compare the sample quantile function of $F_0(X_1), \ldots, F_0(X_n)$, which equals

$F_0(Q(u))$, to u. The relevant asymptotic distribution theorem is

$$\{\sqrt{n}\{F_0(\tilde{Q}(u)) - u\} \ , \ 0 \leq u \leq 1\} \xrightarrow{L} \{B(u) \ , \ 0 \leq u \leq 1\}$$

The problem of statistical modeling can be elegantly defined in terms of quantile functions: one seeks to determine distribution functions $F_0(x)$ such that $F_0(\tilde{Q}(u))$ is not significantly different from a uniform quantile function u. Given a parametric family of distribution functions $F_\theta(x)$ an optimal estimator $\hat{\theta}$ of θ could be defined as the value of θ which minimizes the distance $||F_\theta(\tilde{Q}(u)) - u||$ for a suitable measure of distance between functions on the interval 0 to 1.

An example of a distance is the conventional L_2 distance

$$||g_1 - g_2||^2 = \int_0^1 |g_1(u) - g_2(u)|^2 \, du \ .$$

However, one would like to choose the distance so that the estimator $\hat{\theta}$ would be asymptotically efficient. Such a distance is provided by the reproducing kernel Hilbert space (RKHS) norm of the covariance kernel of the Brownian Bridge stochastic process; it can be defined over any sub-interval $0 \leq p < u \leq q < 1$:

$$||g_1-g_2||^2_{p,q} = \int_p^q |g_1'(u) - g_2'(u)^2 du + \frac{1}{p}|g_1(p) - g_2(p)|^2$$

$$+ \frac{1}{1-q}|g_1(q) - g_2(q)|^2 \ ,$$

$$||g_1-g_2||^2_{0,1} = \int_0^1 |g_1'(u) - g_2'(u)|^2 \, du \ .$$

The inner product is

$$(g_1, g_2)_{p,q} = \int_p^q g_1'(u)\ g_2'(u)\ du + \frac{1}{p}\ g_1(p)\ g_2(p)$$

$$+ \frac{1}{1-q} g_1(q)\ g_2(q)$$

A minimum distance criterion for statistical estimation of the parameters θ of a parametric family F_θ of distribution functions is to choose θ to minimize

$$|| F_\theta(\tilde{Q}(u)) - u||^2 = \int_0^1 |f_\theta(\tilde{Q}(u))\ \tilde{q}(u) - 1|^2\ du$$

One may show this criterion to be asymptotically equivalent to maximizing likelihood, or minimizing the estimated directed divergence $I(\tilde{f}, f_\theta)$:

$$I(\tilde{f}; f_\theta) = \int_{-\infty}^{\infty} \tilde{f}(x)\ \log \frac{\tilde{f}(x)}{f\ (x)}\ dx$$

$$= - \int_0^1 \log\ \{f_\theta(\tilde{Q}(u)\ \tilde{q}(u)\}\ du$$

3. Location and Scale Parameter Models.

An important parametric model for a distribution function $F(x)$ is

$$F(x) = F_0\left(\frac{x-\mu}{\sigma}\right)$$

where F_0 is specified, and μ and σ are unknown (location and scale) parameters to be estimated. Then

$$Q(u) = \mu + \sigma\, Q_0(u) ,$$

$$q(u) = \sigma\, q_0(u) ,$$

$$fQ(u) = \frac{1}{\sigma}\, f_0 Q_0(u) .$$

Two important choices for F_0 are:

(1) the normal or Gaussian case:

$$F_0(x) = \Phi(x) = \int_{-\infty}^{x} \phi(y)\, dy ,$$

$$f_0(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2} x^2 ;$$

(2) the exponential case:

$$F_0(x) = 1-e^{-x} , \quad f_0(x) = e^{-x} .$$

The quantile functions, score functions, and density-quantile functions of some standard probability laws are given in the Table. Graphs of density-quantile functions are given in the Figures.

Because of the way that $fQ(u)$ depends on μ and σ, one can introduce functions to test hypothesis H_0: $Q(u) = \mu + \sigma Q_0(u)$ which do not require estimation of μ and σ before testing the hypothesis. Define

$$\sigma_0 = \int_0^1 f_0 Q_0(u)\ q(u)\ du\ ,$$

$$d(u) = \frac{1}{\sigma_0}\ f_0 Q_0(u)\ q(u)\ ,$$

$$D(u) = \int_0^u d(u')\ du'\ .$$

We call $D(u)$ a transformation distribution function, and $d(u)$ a transformation density. The null hypothesis H_0 is equivalent to

$$D(u) = u\ ,\quad d(u) = 1\ ,\ 0 \leq u \leq 1\ .$$

Given an estimator $\tilde{D}(u)$, $0 \leq u \leq 1$, the deviations of $\tilde{D}(u)$ from linearity can be used to test whether a sample consists of random variables satisfying H_0, or consists of random variables satisfying H_o plus outliers. Such techniques would be useful for many diverse applications.

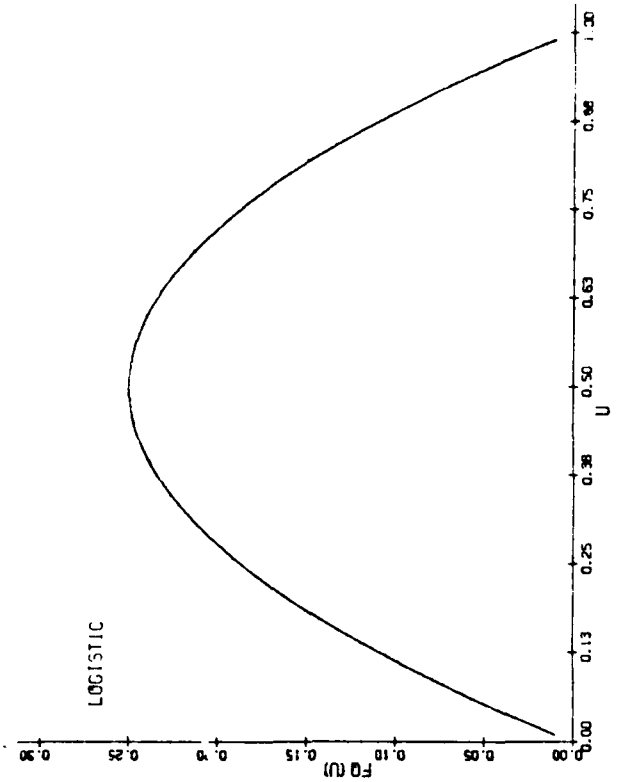
Table

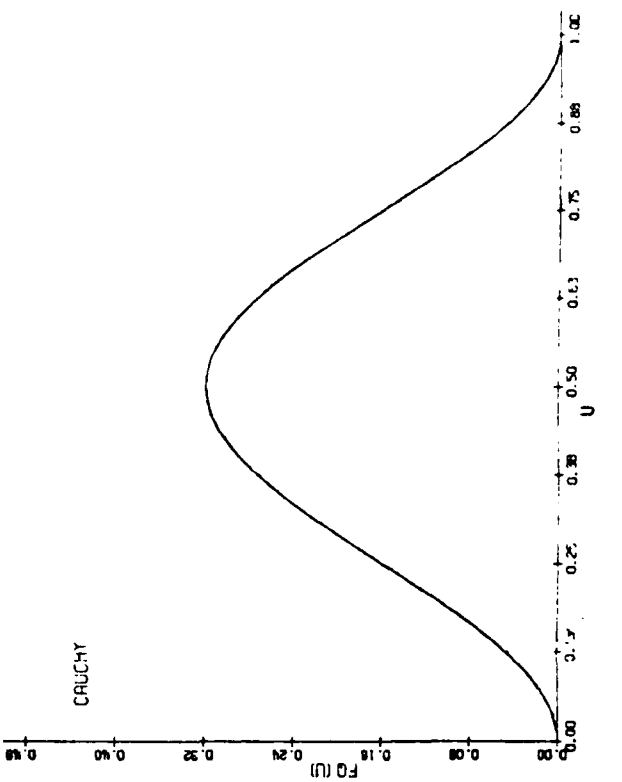
QUANTILE FUNCTIONS, SCORE FUNCTIONS, AND DENSITY QUANTILE FUNCTIONS

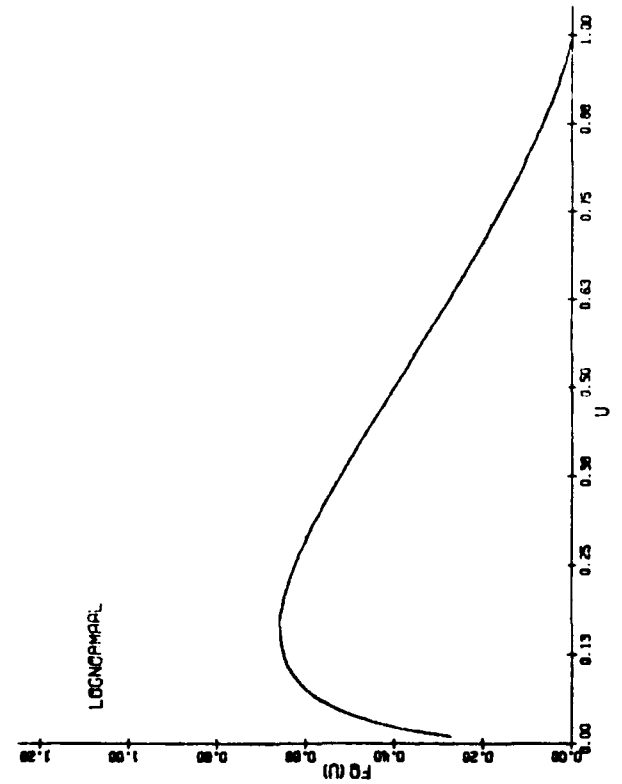
Probability Law	Q(u)	J(u)	fQ(u)
Normal	$\Phi^{-1}(u)$	$\Phi^{-1}(u)$	$\phi\Phi^{-1}(u) = (2\pi)^{-1/2}\exp-\frac{1}{2}\{\Phi^{-1}(u)\}^2$
Log-normal	$e^{\Phi^{-1}(u)}$	$e^{-\Phi^{-1}(u)}\{\Phi^{-1}(u)+1\}$	$\phi\Phi^{-1}(u)\exp - \Phi^{-1}(u)$
Exponential	$\log (1-u)^{-1}$	1	1-u
Extreme Value	$\log\log (1-u)^{-1}$	$1 + \log (1-u)^{-1}$	$(1-u)\log (1-u)^{-1}$
Weibull	$\frac{1}{\beta}\{\log (1-u)^{-1}\}^{\beta}$	$\{\log (1-u)^{-1}\}^{-\beta}\{1-\beta+\log(1-u)^{-1}\}$	$\frac{1}{\beta}(1-u)\{\log (1-u)^{-1}\}^{1-\beta}$
Logistic	$\log \frac{u}{1-u}$	2u - 1	u(1-u)
Double-Exponential	log 2u , u<1/2 -log 2(1-u) , u>1/2	sign (2u-1)	u , u < 0.5 ; 1-u , u > 0.5
Cauchy	$\tan \pi (u- \frac{1}{2})$	$-\sin 2\pi u$	$\frac{1}{\pi}\sin^2 \pi u$
Pareto	$\frac{1}{\beta}(1-u)^{-\beta}$	$(1+\beta)(1-u)^{\beta}$	$\frac{1}{\beta}(1-u)^{1+\beta}$

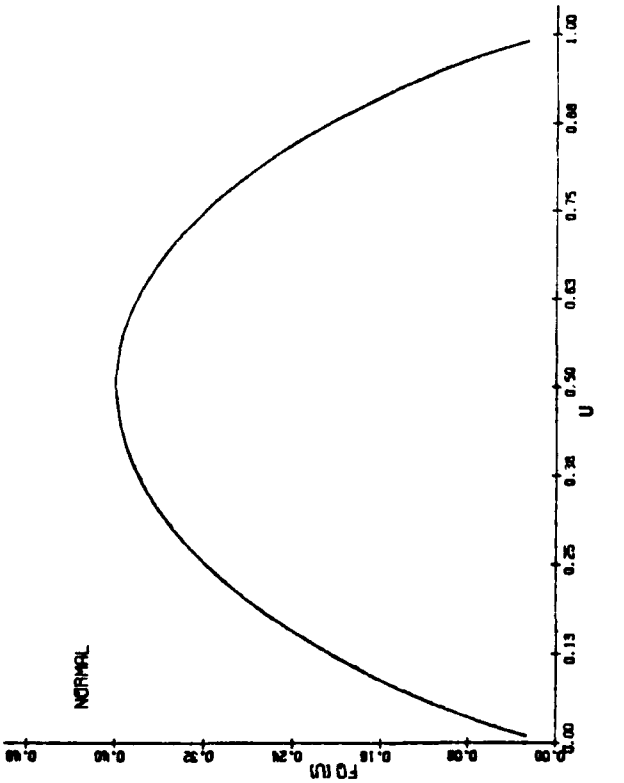
Figure A

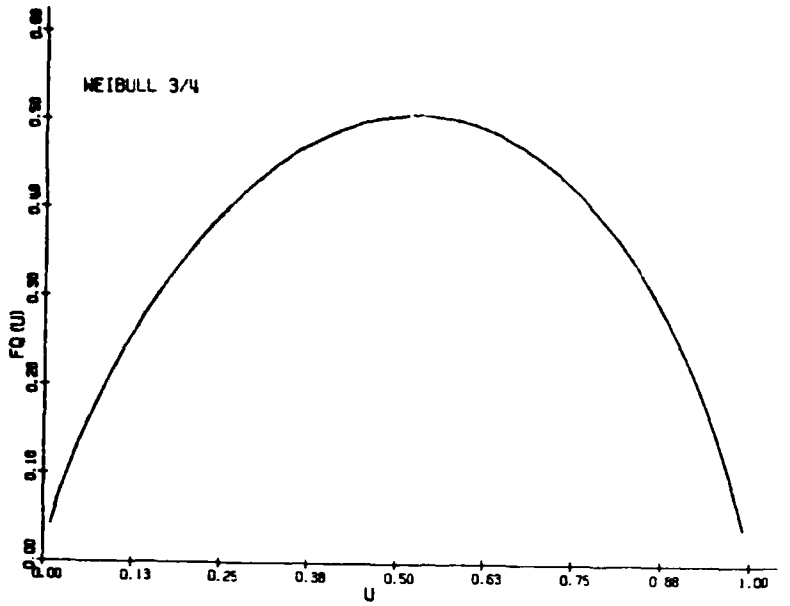
Density Quantile Functions $fQ(u)$, $0 \leq u \leq 1$ of some common probability distributions Lognormal, Logistic, Normal, Cauchy, Weibull with various shape parameters, and Extreme Value.

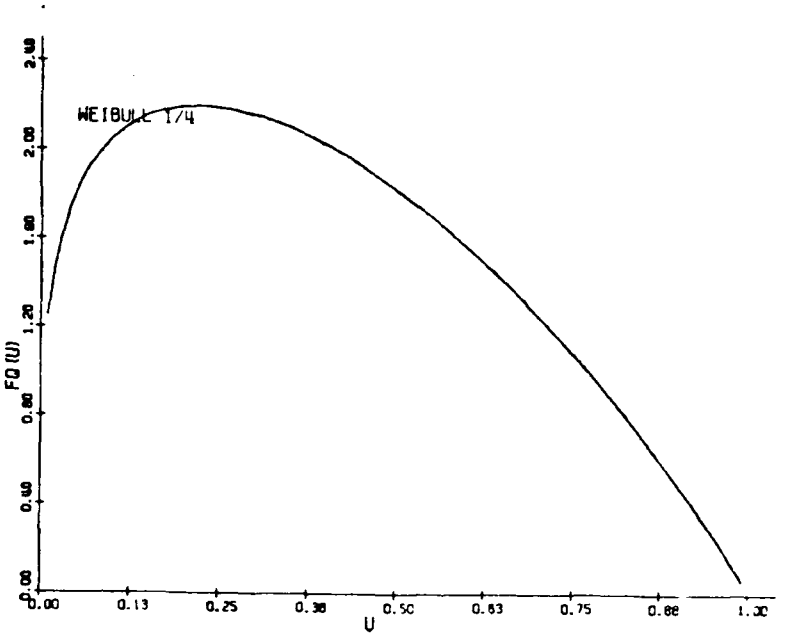

LOGISTIC
FQ (U)
U

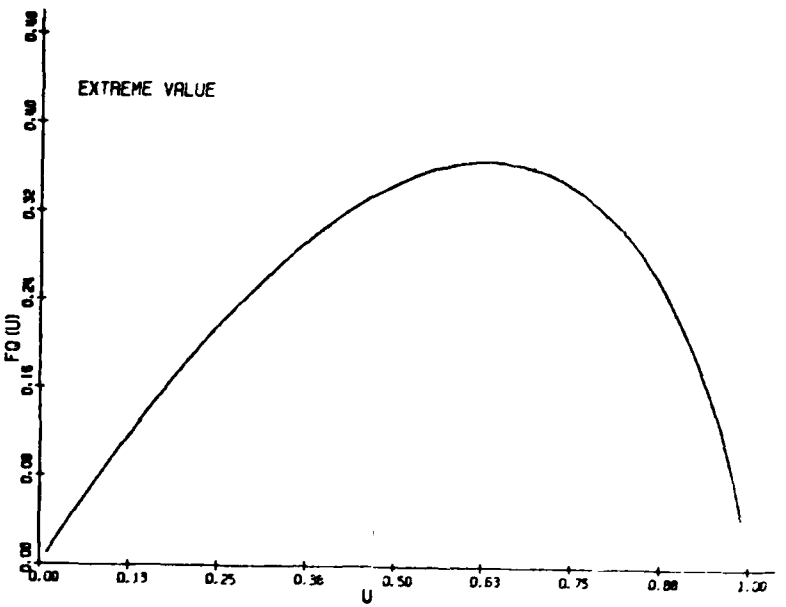

CAUCHY
FQ (U)
U

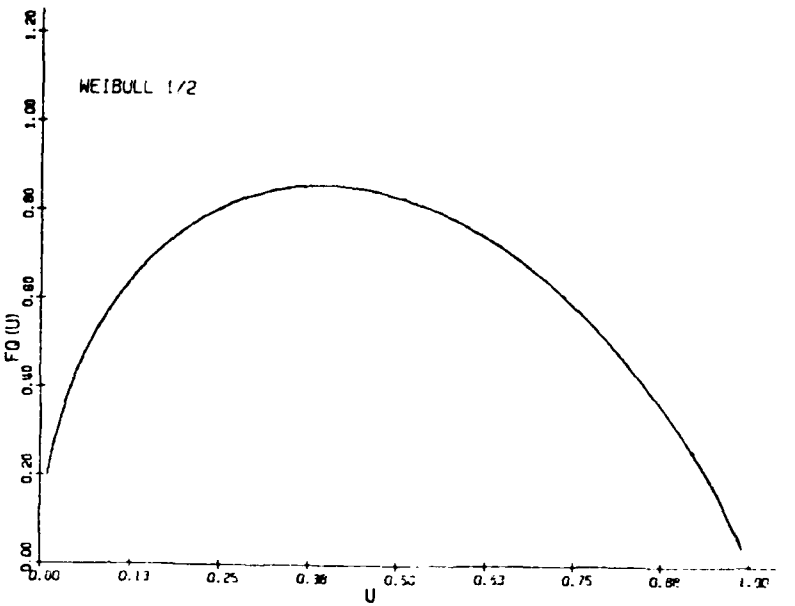

LOGNORMAL
FQ (U)
U


NORMAL
FQ (U)
U


WEIBULL 3/4
FQ(U)
U
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00


WEIBULL 1/4
FQ(U)
U
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00


EXTREME VALUE
FQ(U)
U
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00


WEIBULL 1/2
FQ(U)
U
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00

4. Quantile Based Measures of Average, Deviation, Tail Behavior, and Modes.

We propose that the sample quantile function provides a representing function for the sample in the following senses:

(1) Models for the data should be viewed as being in one to one correspondence with the smooth quantile functions $\hat{Q}(u)$, $0 \leq u \leq 1$ which are their representing functions,

(2) The criteria for testing whether a model fits the data, should be based on measures of fit between the representating functions $\hat{Q}(u)$ and $\tilde{Q}(u)$,

(3) Since the sample is summarized by its representing function $\tilde{Q}(u)$, any descriptor of the sample should be expressible as a function of $\tilde{Q}(u)$. Similarly any descriptor of the distribution of X should be expressible as a function of $Q(u)$.

There are four characteristics of a probability distribution which we would like to infer from the data:

(1) location, represented by a measure of average;

(2) spread, represented by a measure of deviation ,

(3) tail behavior, represented by the behavior of $fQ(u)$ as u tends to 0 and 1 ,

(4) modality, represented by the number of modes (relative maximum) in the probability density or in the density-quantile function.

Location and spread parameters for a distribution seem to be meaningful only when it is unimodal; otherwise, one may want to find an associated variable whose values can be used to divide the original sample of X values into two or more samples, each of which is unimodal.

A parameter representing average or location will be denoted by μ; it could be defined by one of the following concepts:

$$\text{median } \mu = Q(0.5) ,$$

$$\text{mid-quartile } \mu = \tfrac{1}{2}\{Q(0.25) + Q(0.75)\} ,$$

$$\text{mid-range } \mu = \tfrac{1}{2}\{Q(0) + Q(1)\} ,$$

$$\text{mean } \mu = \int_0^1 Q(u)du = \int_{-\infty}^{\infty} x f(x)dx .$$

A parameter representing deviation or spread or scale will be denoted by σ; it could be defined by one of the following concepts:

$$\text{interquartile range } \sigma = Q(0.75) - Q(0.25) ,$$

$$\underline{\text{Score deviation}}\ \sigma = \int_0^1 J_0(u)Q(u)du \text{ with score function } J_0(u)$$

$$\text{standard deviation } \sigma = \{\int_0^1 \{Q(u) - \int_0^1 Q(t)dt\}^2 du\}^{\frac{1}{2}}$$

The properties of $fQ(u)$ describe the tail behavior, modality, and symmetry of the distribution. Indices α_1 and α_2 such that

$$fQ(u) \sim u^{\alpha_1} \text{ as } u \longrightarrow 0 ,$$

$$fQ(u) \sim (1-u)^{\alpha_2} \text{ as } u \longrightarrow 1$$

may be rigorously defined when they exist by

$$\alpha_1 = \lim_{u \longrightarrow 0} \frac{uJ(u)}{fQ(u)} \quad , \qquad \alpha_2 = \lim_{u \longrightarrow 1} \frac{(1-u)J(u)}{fQ(u)}$$

We call α_1 the left tail exponent, and α_2 the right tail exponent.

The tail exponent α indicates whether the tail is short, medium, or long: $\alpha < 1$, short; $\alpha = 1$, medium; $\alpha > 1$, long.

The Gaussian distribution has $\alpha_1 = \alpha_2 = \alpha = 1$; the exponential distribution has $\alpha_1 = 0$ and $\alpha_2 = 1$; the Cauchy distribution has $\alpha_1 = \alpha_2 = \alpha = 2$. The graphs of their fQ functions are given in Figure A. Our ideas of the canonical shapes of distributions seem to me to become unified when they are formulated not in terms of the shape of f(x) but in terms of the shape of fQ(u); for example, J and U-shaped distributions correspond to $\alpha \leq 0$.

When fQ(u) is unimodal, an important descriptor is the mode percentile, denoted p_{mode} . It is defined to be the value of u at which fQ(u) achieves its mode (or maximum value). The value of p_{mode} and its relation to 0.5, is a quick summary of the skewness of the distribution.

When p-mode ≥ 0.5 , the distribution is conventionally described as being skewed to the left; this occurs if we assume that fQ satisfies $fQ(u) \leq fQ(1-u)$, $0 \leq u \leq 0.5$, which implies that $Q(u) + Q(1-u) \leq 2Q(0.5)$, and consequently that

$$\text{mean} \leq \text{median} \leq \text{mode} \ .$$

Similarly p-mode ≤ 0.5 (and the distribution is skewed to the right) if we assume that $fQ(u) \geq fQ(1-u)$, $0 \leq u \leq 0.5$, which implies that $Q(u) + Q(1-u) \geq 2Q(0.5)$, and consequently that

$$\text{mean} \geq \text{median} \geq \text{mode} .$$

The fact that a density-quantile function is always defined on the unit interval, while a density function $f(x)$ is defined on an infinite interval, seems to me to make the former easier to estimate.

5. Quantile Box-Plots and Transform Distribution Functions

The most dramatic new data-analytic tools suggested by the quantile and density-quantile approach are Quantile Box plots of $\tilde{Q}(u)$, $0 \leq u \leq 1$, and plots of sample transformation distribution functions $\tilde{D}(u)$, $0 \leq u \leq 1$.

Quantile box plots are formed of the original data and of the data after transformations such as square root, logarithm, and reciprocal. They provide quick procedures for estimating location, scale, and shape. A quantile box plot consists of a graph of a quantile function on which is superimposed various boxes with vertices $(p,\tilde{Q}(p))$, $(p,\tilde{Q}(1-p))$, $(1-p, \tilde{Q}(p))$, $(1-p, \tilde{Q}(1-p))$ which we call a p-box. One usually chooses $p = 1/4$, $1/8$, $1/16$. Within the quartile box ($p = 0.25$), one draws a median line with vertices $(0.25, \tilde{Q}(0.5))$, $(0.75, \tilde{Q}(0.5))$. An approximate confidence interval for the median $\tilde{Q}(0.5)$ is indicated by a vertical line with vertices $(0.5, \tilde{Q}(0.5) \pm IQ/\sqrt{n})$ where n is the sample size and $IQ = \tilde{Q}(0.75) - \tilde{Q}(0.25)$ is the inter-quartile range. The symmetry of the distribution is judged by the symmetry of $\tilde{Q}(u)$ within the quartile box.

A quantile box plot is an extension of the idea of a box plot introduced by Tukey (1977).

A transformation distribution function, or cumulative weighted spacings function, is defined by

$$\tilde{D}(u) = \int_0^u \tilde{d}(t)\, dt \ , \quad 0 \leq u \leq 1 \ ,$$

where

$$\tilde{d}(u) = \frac{1}{\tilde{\sigma}_0} f_0 Q_0(u)\, \tilde{q}(u) \ , \quad \tilde{\sigma}_0 = \int_0^1 f_0 Q_0(u)\, \tilde{q}(u)\, du \ .$$

Its pseudo-correlations are defined by

$$\tilde{\rho}(v) = \int_0^1 e^{2\pi iuv}\, \tilde{d}(u)\, du\ , \quad v = 0, \pm 1, \ldots\ .$$

The asymptotic distributional properties of $\tilde{d}(u)$ are similar to those of the sample spectral density of a stationary time series. Tests of H_0 could be based on $\int_0^1 \log \tilde{d}(u)\, du$; the deviation from $D(u) = u$ of $\tilde{D}(u)$; the deviation from $\rho(v) = 0$ of $\tilde{\rho}(v)$, $v = 1, 2, \ldots$.

To estimate the density-quantile function $fQ(u)$, one uses $\tilde{d}(u)$ to form smooth estimators $\hat{d}(u)$ of $d(u)$. Two main approaches are:

(1) kernel method -- for a suitable kernel K

$$\hat{d}_K(u) = \int_0^1 \tilde{d}(t)\, K(u-t)\, dt\ ;$$

(2) autoregressive method -- for a suitable order m

$$\hat{d}_m(u) = \hat{\sigma}_m^2 \left| 1 + \hat{\alpha}_m(1)e^{2\pi iu} + \ldots + \hat{\alpha}_m(m)e^{2\pi ium} \right|^{-2}$$

where $\hat{\sigma}_m^2$, $\hat{\alpha}_m(j)$, $j = 1, \ldots, m$ are determined from certain linear equations (Yule-Walker equations) in

$$\hat{\rho}(v) = \int_0^1 e^{2\pi iuv}\, \tilde{d}(u)\, du, \quad v = 0, \pm 1, \ldots, \pm m.$$

The autoregressive estimator, including procedures for selecting the order m, are implemented in a computer program ONESAM whose use is illustrated. It should be noted that choosing order $m = 0$ is equivalent to accepting H_0.

A solution to the important problem of estimating fQ(u) is provided by the "autoregressive" estimator

$$\hat{fQ}_m(u) = f_0 Q_0(u) \quad \{\tilde{\sigma}_0 \, \hat{d}_m(u)\}^{-1}$$

Some diagnostics we use for choosing the order m of the autoregressive estimator $\hat{fQ}_m(u)$ are the square modulus pseudo-correlations $|\hat{\rho}(v)|^2$; the residual variances $\sigma^2(m)$; the Akaike order determining criterion, for sample size T,

$$AIC(m) = \log \hat{\sigma}_m^2 + \frac{2m}{T} ;$$

and Parzen's criterion

$$CAT(m) = \frac{1}{T} \sum_{j=1}^{m} \hat{\sigma}_j^{-2} - \hat{\sigma}_m^{-2}$$

whose shape in practice is similar to the shape of AIC.

Another approach to estimating fQ(u) which deserves investigation is to estimate log fQ(u) by smoothing - log $\tilde{q}(u)$.

6. Examples

To illustrate how to use Q, D, and fQ in statistical data analysis, let us consider data from Tukey (1977), p. 117 which lists seasonal snowfall in Buffalo, New York and Cairo, Illinois, from 1918-19 to 1937-38, and asks "What light do these two batches throw on how they should be expressed." To answer this question one approach might be to examine the quantile box-plots of the batches (Figure B); the quartile box in Buffalo appears symmetric while in Cairo it does not. One might attempt a transformation of the Cairo data; we choose the square root and conclude that its quartile box is symmetric. Does this prove that Buffalo snowfall and square root of Cairo snowfall are Gaussian?

A rigorous approach is to form the cumulative weighted spacings function

$$\tilde{D}(u) = \frac{\int_0^u \phi\Phi^{-1}(t)\tilde{q}(t)dt}{\int_0^1 \phi\Phi^{-1}(t)\tilde{q}(t)dt} \quad , \quad 0 \leq u \leq 1$$

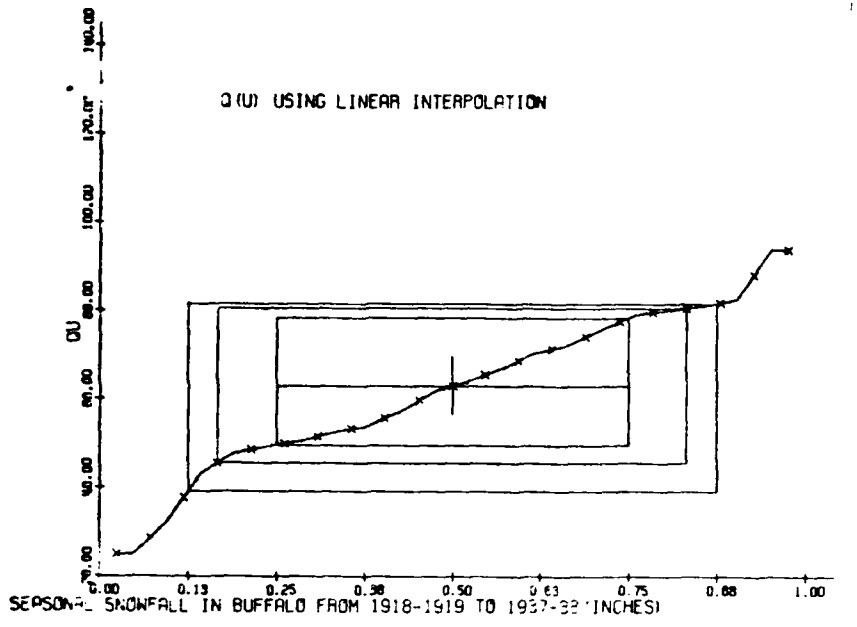
whose deviations from u provide a test of Gaussian-ness which does not first require estimation of μ and σ (mean and standard deviation). The graphs of $\tilde{D}(u)$ in Figure C indicate clearly that Buffalo snowfall and square root of Cairo are Gaussian, while Cairo snowfall is not Gaussian.

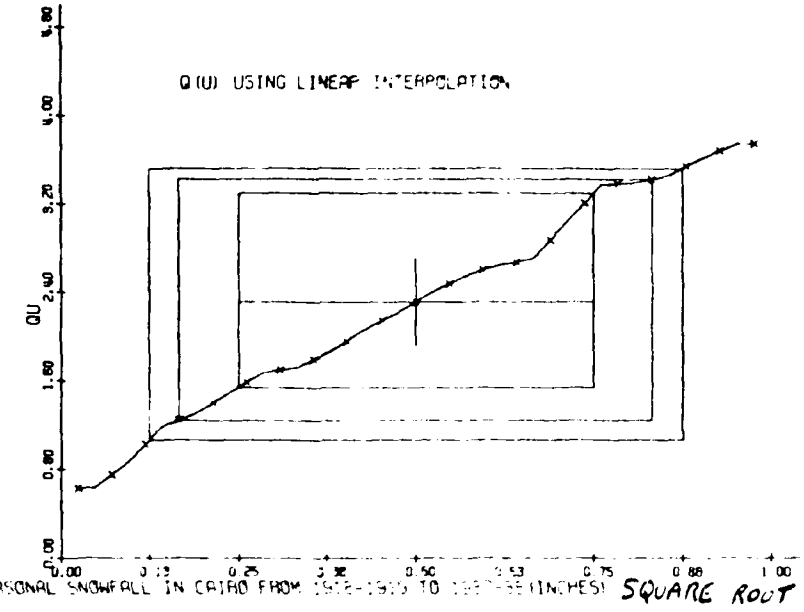
The true character of the Cairo snowfall data emerges when one estimates its fQ function; it turns out to be bimodal, which we interpret to mean that there are two kinds of snowfall years in Cairo, Illinois -- light and heavy.

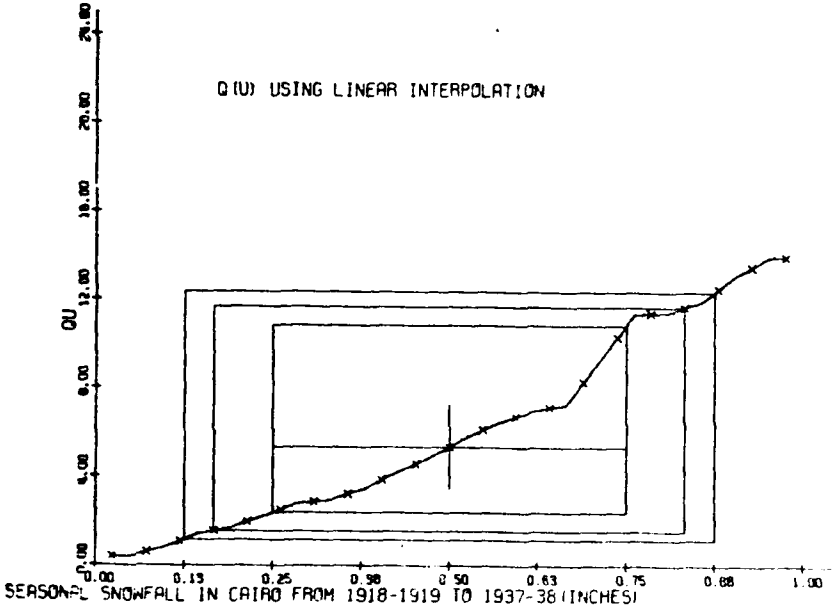
Even though various diagnostic tests of the square roots of Cairo snowfall data indicate that it is Gaussian the order 1 autoregressive estimator of the density quantile indicates that bimodality is a possible alternative hypothesis.

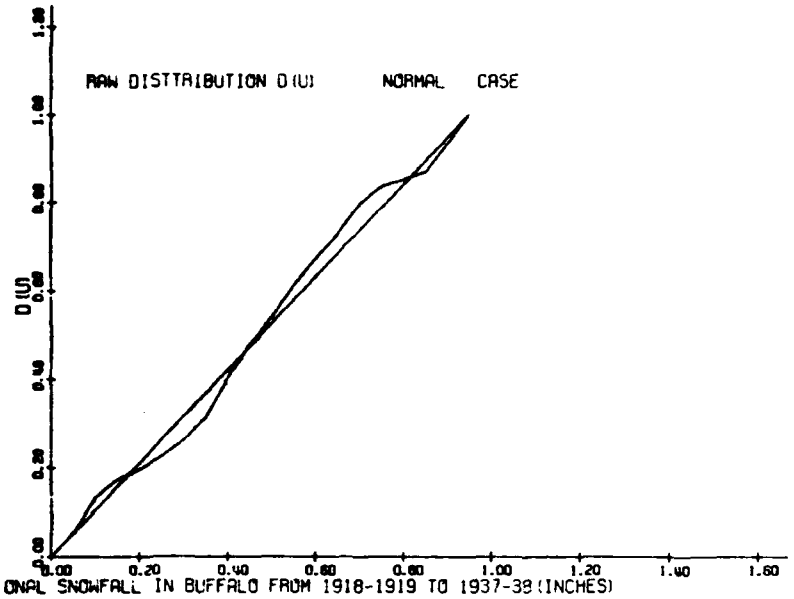
Order Statistics	Buffalo	Cairo	Cairo Square Root
1	25.0	0.4	.6325
2	39.8	1.2	1.0954
3	46.7	1.6	1.2649
4	49.1	1.8	1.3416
5	49.6	2.7	1.6432
6	51.6	2.9	1.7029
7	53.5	3.0	1.7320
8	54.7	4.0	2.0000
9	60.3	4.5	2.1213
10	63.6	5.4	2.3238
11	64.8	6.2	2.4900
12	69.4	6.8	2.6077
13	71.8	7.2	2.6833
14	72.9	7.4	2.7203
15	79.0	11.3	3.3615
16	79.6	11.5	3.3912
17	80.7	11.5	3.3912
18	81.6	12.4	3.5214
19	83.6	13.9	3.7283
20	103.9	14.1	3.7550
Mean $\hat{\mu}$	64.1	6.5	2.375
Q(0.25)	49.6	2.7	1.6432
Q(0.50)	64.2	5.8	2.4069
Q(0.75)	79.6	11.5	3.3912
IQ	30.0	8.8	1.748
S.D. $\hat{\sigma}$	18.4	4.5	0.945

$\lvert\hat{b}(v)\rvert^2$	Buffalo	Cairo	Cairo Square Root
0	1.0000	1.0000	1.0000
1	.0203	.1391	.0705
2	.0187	.0192	.0142
3	.0053	.0215	.0222
4	.0196	.0496	.0289
5	.0169	.0317	.0181
$\sigma^2(m)$			
0	1.0000	1.0000	1.0000
1	.9797	.8609	.9295
2	.9657	.8199	.8963
3	.9559	.8179	.8603
4	.9367	.7883	.8570
5	.9021	.7610	.8404
AIC(m)			
0	.0000	.0000	.0000
1	.0795	-.0497	.0269
2	.1651	.0014	.0905
3	.2549	.0990	.1495
4	.3346	.1622	.2457
5	.3969	.2269	.3261
Minimum At m =	0	1	0
CAT			
0	-1.0000	-1.0000	-1.0000
1	- .9212	-1.0483	- .9709
2	- .8369	- .9877	- .9028
3	- .7497	- .8772	- .8373
4	- .6718	- .8020	- .7361
5	- .6076	- .7235	- .6504

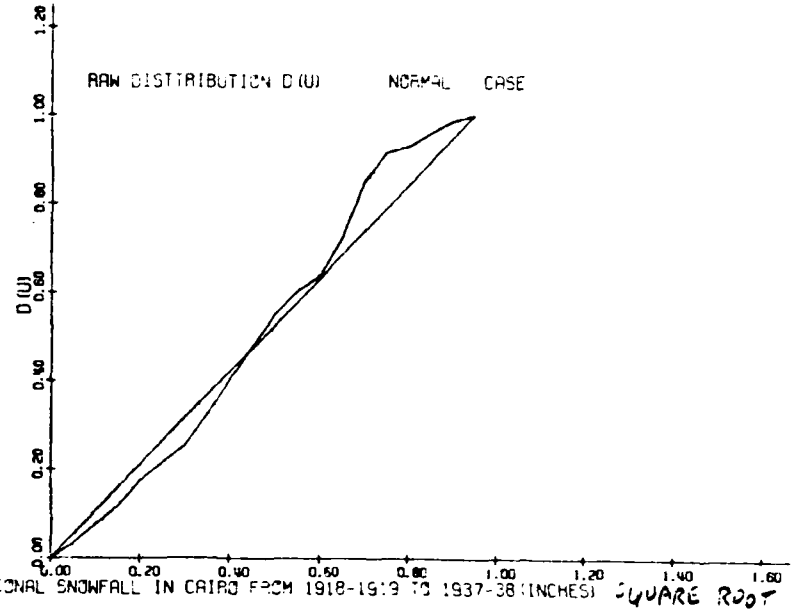

Q(U) USING LINEAR INTERPOLATION
QU
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00
SEASONAL SNOWFALL IN BUFFALO FROM 1918-1919 TO 1937-38 (INCHES)


Q(U) USING LINEAR INTERPOLATION
QU
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00
SEASONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES)
SQUARE ROOT

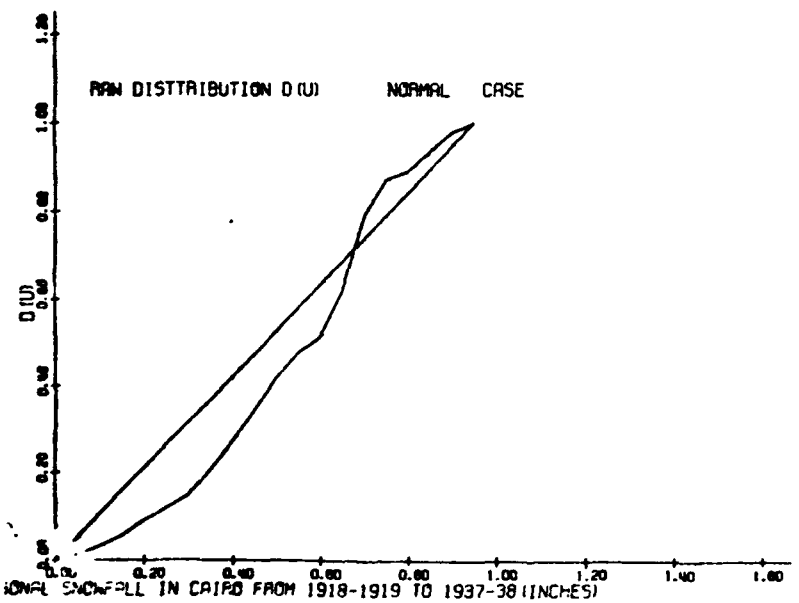

Q(U) USING LINEAR INTERPOLATION
QU
0.00
0.13
0.25
0.38
0.50
0.63
0.75
0.88
1.00
SEASONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES)



ONAL SNOWFALL IN BUFFALO FROM 1918-1919 TO 1937-38 (INCHES)

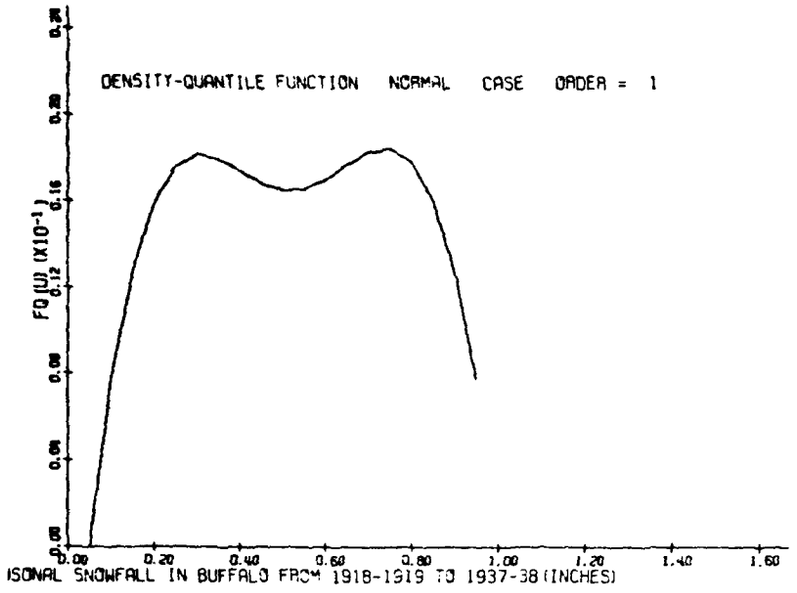


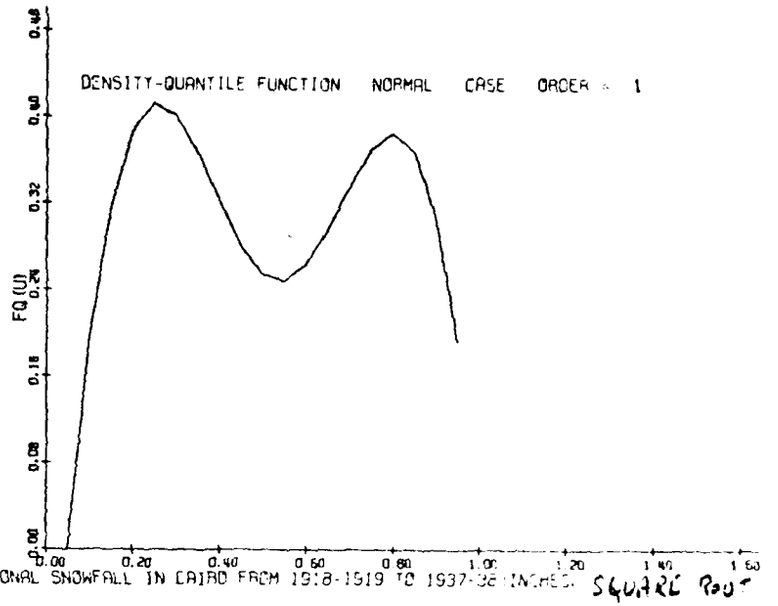
SEASONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES) SQUARE ROOT

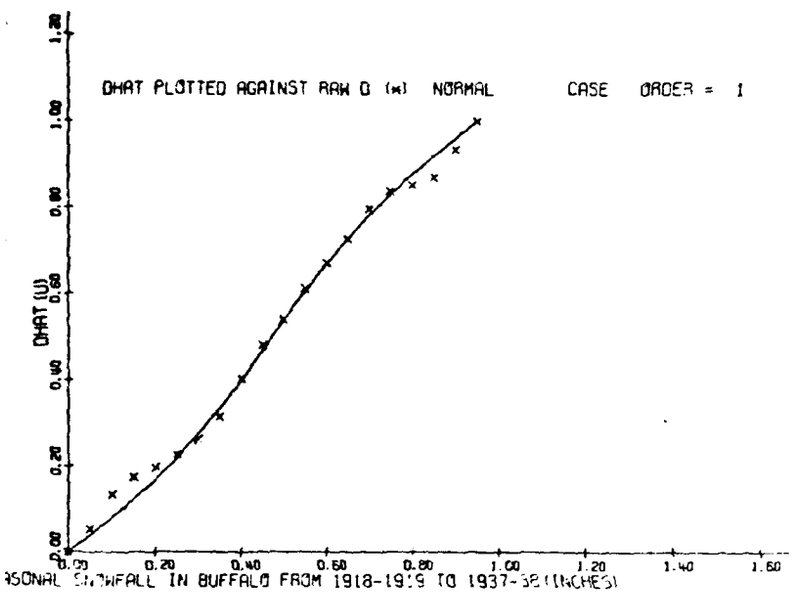


SONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES)

Figure C


DENSITY-QUANTILE FUNCTION NORMAL CASE ORDER = 1
FQ(U) (X10⁻¹)
0.00
0.04
0.08
0.12
0.16
0.20
0.24
0.28
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
SEASONAL SNOWFALL IN BUFFALO FROM 1918-1919 TO 1937-38 (INCHES)


DENSITY-QUANTILE FUNCTION NORMAL CASE ORDER = 1
FQ(U)
0.00
0.08
0.16
0.24
0.32
0.40
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
SEASONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES)
SQUARE ROOT


DHAT PLOTTED AGAINST RAW D (W) NORMAL CASE ORDER = 1
DHAT(U)
0.00
0.20
0.40
0.60
0.80
1.00
1.20
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
SEASONAL SNOWFALL IN BUFFALO FROM 1918-1919 TO 1937-38 (INCHES)

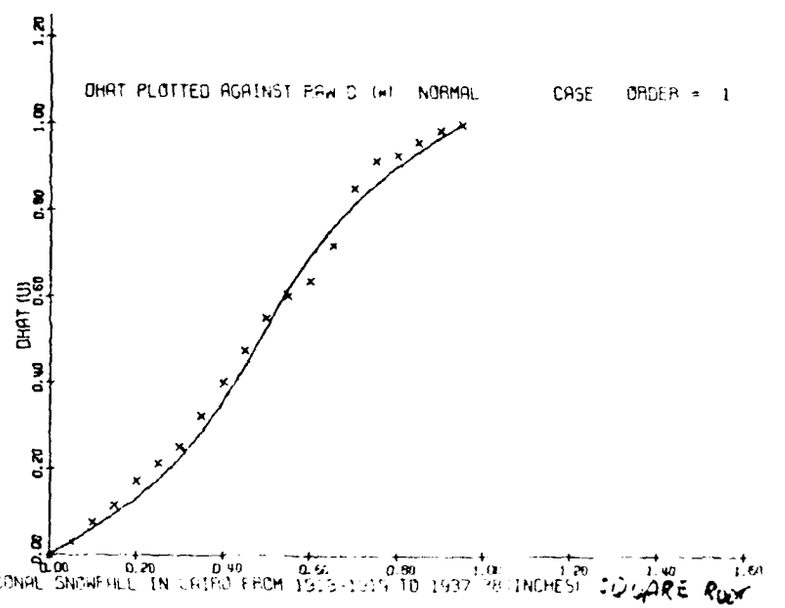
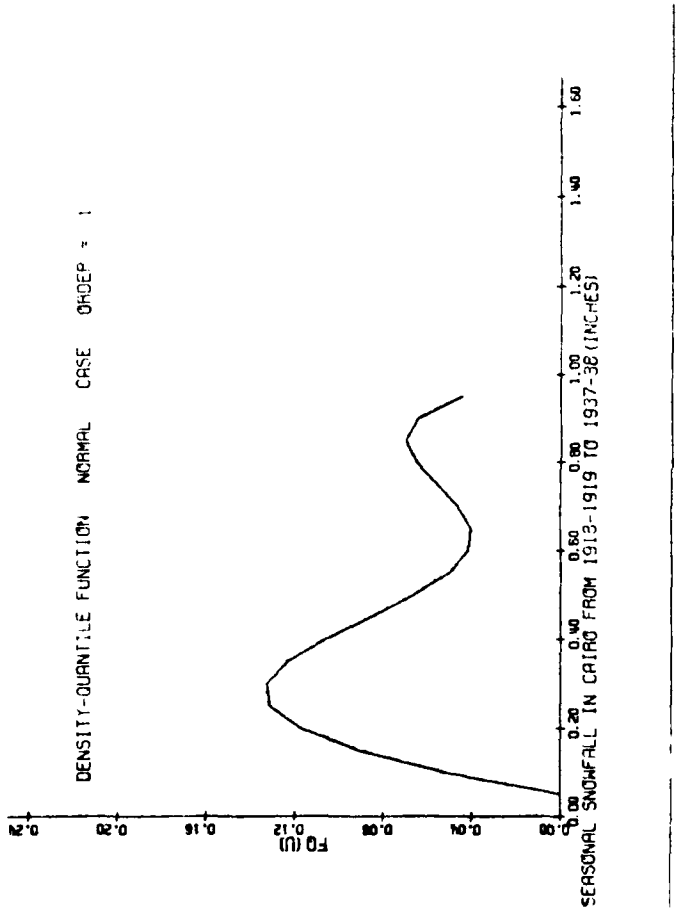
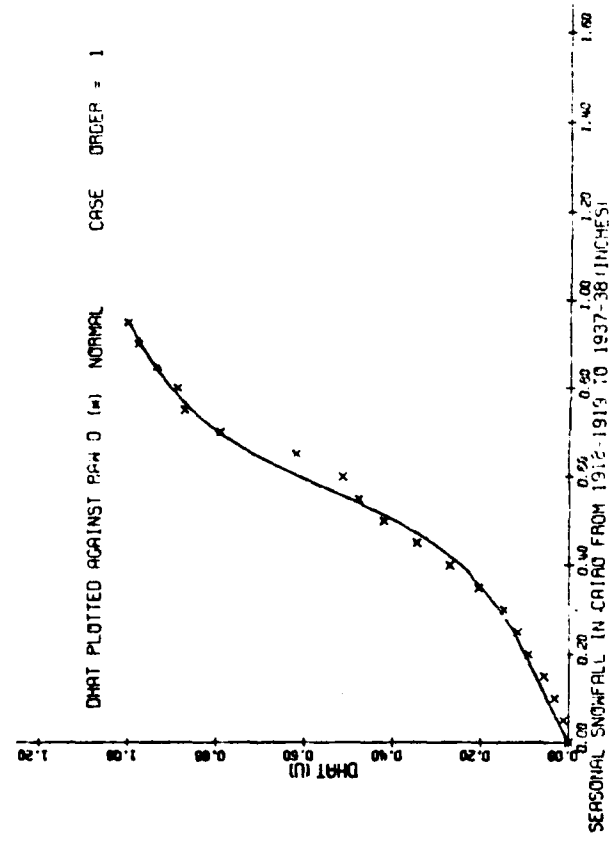

DHAT PLOTTED AGAINST RAW D (W) NORMAL CASE ORDER = 1
DHAT(U)
0.00
0.20
0.40
0.60
0.80
1.00
1.20
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
SEASONAL SNOWFALL IN CAIRO FROM 1918-1919 TO 1937-38 (INCHES)
SQUARE ROOT

Figure D

Figure D (cont)





7. Quantile Simulation and Quantile Boot Strap.

The quantile function $Q(u)$, $0 \leq u \leq 1$ of a random variable X provides a way of simulating a random sample of X. Let U denote a random variable uniform on 0 to 1; then $X \overset{L}{=} Q(U)$. Let $U_1, \ldots, U_n$ be independent uniform random variables; then

$$X_1, \ldots, X_n \overset{L}{=} Q(U_1), \ldots, Q(U_n) .$$

To generate a random sample $X_1, \ldots, X_n$, one generates n random numbers $U_1, \ldots, U_n$ and transforms them to $X_1, \ldots, X_n$.

To obtain by Monte Carlo methods the distribution of a statistic

$$T = g(X_1, \ldots, X_n) ,$$

one would generate a large number N of random samples $X_1, \ldots, X_n$; generate a random sample $T_1, \ldots, T_N$ of the random variable T; and finally form the sample quantile function $\tilde{Q}_T(u)$ which provides an estimator of the true quantile function of T.

When the quantile function $Q(u)$ of X is not known, one can estimate it by the sample quantile function $\tilde{Q}(u)$. Now from random numbers $U_1, \ldots, U_n$, one can generate "boot strap" simulated values [compare Efron (1978)]

$$\tilde{X}_1 = \tilde{Q}(U_1), \ldots, \tilde{X}_n = \tilde{Q}(U_n), \quad \tilde{T} = g(\tilde{X}_1, \ldots, \tilde{X}_n) .$$

One can generate a random sample $\tilde{T}_1, \ldots, \tilde{T}_N$ of T, whose sample quantile function $\tilde{Q}_{\tilde{T}}(u)$ provides an estimator of the true quantile function of T.

Bivariate distributions. An outstanding problem of statistics is the simulation of multivariate distributions. To illustrate the quantile approach to this problem, let (X_1, X_2) have joint distribution function $F(x_1, x_2)$. Denote the marginal distributions functions of X_1 and X_2 by $F_1(x_1)$ and $F_2(x_2)$. Denote the quantile functions of the marginal distributions by $Q_1(u_1)$ and $Q_2(u_2)$. Define

$$D(u_1,u_2) = F(Q_1(u_1) , Q_2(u_2)) ;$$

it is the joint distribution function of the "rank transforms"

$$U_1 = F_1(X_1) , \quad U_2 = F_2(X_2) .$$

One can generate (X_1,X_2) by generating (U_1,U_2) from the distribution $D(u_1,u_2)$ and then forming

$$X_1 = Q_1(U_1) , \quad X_2 = Q_2(U_2) .$$

To generate (U_1, U_2) one chooses U_1 to be uniform on 0 to 1, and then generates U_2 by the conditional distribution $D_{U_2|U_1}(u_2|u_1)$ or its quantile function $Q_{U_2|U_1}(p|u_1)$ by the formula $U_2|U_1 = u_1 \overset{L}{=} Q_{U_2|U_1}(U_2'|u_1)$ where U_2' is uniform on 0 to 1, and independent of U_1 . The conditional quantile function $Q_{U_2|U_1}(p|u_1)$ can be estimated by the sample quantile function of the sample values U_2 corresponding to sample U_1 values near u_1 .

8. Quantile Formulation of Robust Location and Scale Estimators.

Assume a location-scale parameter model for the quantile function of a continuous random variable X: $Q(u) = \mu + \sigma Q_0(u)$

Assume a symmetric distribution, which is equivalent to $Q_0(u)$ being an odd function in the sense that $Q_0(1-u) = -Q_0(u)$

Given a sample $X_1, \ldots, X_n$, the log likelihood function may be written in terms of the sample quantile function [compare Parzen (1979a)]:

$$L = \frac{1}{n} \log f(X_1, \ldots, X_n; \mu, \sigma)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \log \frac{1}{\sigma} f_0\left(\frac{X_i - \mu}{\sigma}\right) \qquad (1)$$

$$= -\log \sigma + \int_{-\infty}^{\infty} \log f_0\left(\frac{x-\mu}{\sigma}\right) d\tilde{F}(x)$$

$$= -\log \sigma + \int_0^1 \log f_0\left(\frac{\tilde{Q}(u)-\mu}{\sigma}\right) du$$

The maximum likelihood estimators $\hat{\mu}$ and $\hat{\sigma}$ satisfy $\frac{\partial L}{\partial \mu} = 0$ and $\frac{\partial L}{\partial \sigma} = 0$. An important role in these equations is played by the Fisher Score function

$$\psi(x) = -\frac{f_0'(x)}{f_0(x)} = -\frac{d}{dx} \log f_0(x) \; ; \qquad (2)$$

Between ψ and the score function $J_0(u) = -(f_0 Q_0(u))'$, there

is an important relation:

$$J_0(u) = \psi(Q_0(u)) \tag{3}$$

The maximum likelihood estimators $\hat{\mu}$ and $\hat{\sigma}$ are the solutions of

$$\int_0^1 \psi\left(\frac{\tilde{Q}(u)-\hat{\mu}}{\hat{\sigma}}\right)du = 0 \tag{4}$$

$$\int_0^1 \psi\left(\frac{\tilde{Q}(u)-\hat{\mu}}{\hat{\sigma}}\right)\{\tilde{Q}(u)-\hat{\mu}\}du = \hat{\sigma}$$

Under the symmetry assumption one seeks robust estimators of location; various standard estimators may be heuristically motivated by approximating (4) in suitable ways.

M-estimators are defined by introducing a window

$$w(x) = \frac{1}{x}\psi(x) \ .$$

Then (4) may be written

$$\int_0^1 w\left(\frac{\tilde{Q}(u)-\hat{\mu}}{\hat{\sigma}}\right)\{\tilde{Q}(u)-\hat{\mu}\}\ du = 0$$

To estimate $\hat{\mu}$, consider the limit of an iterative sequence $\hat{\mu}^{(n)}$ defined by

$$\hat{\mu}^{(n+1)} = \frac{\int_0^1 w\left(\frac{\tilde{Q}(u)-\hat{\mu}^{(n)}}{\hat{\sigma}}\right)\tilde{Q}(u)\ du}{\int_0^1 w\left(\frac{\tilde{Q}(u)-\hat{\mu}^{(n)}}{\hat{\sigma}}\right)\ du} \ . \tag{5}$$

The estimator $\hat{\mu}^{(n)}$ is an M-estimator. For w(x) one could choose a function which corresponds to Student's t distribution with m degrees of freedom [see Parzen (1979a)]:

$$w(x) = \frac{m+1}{m} \frac{1}{1+\frac{1}{m}x^2}$$

Various widely used choices of $w(x)$ are described in Hogg (1979). The most widely used choice for $w(x)$ may be Tukey's bisquare window

$$w(x) = (1-(\frac{x}{c})^2)_+^2$$

where c is a suitable constant, often chosen as 6. The choice of m or c is crucial; it should reflect one's beliefs about how long are the tails of $F_0(x)$.

L-estimators are linear combinations of order statistics which can be written in terms of the quantile function $\tilde{Q}(u)$ as follows, for a suitable weight function $W_\mu(u)$:

$$\hat{\mu} = \frac{\int_0^1 Q(u) W_\mu(u)\,du}{\int_0^1 W_\mu(u)\,du} \tag{6}$$

If the model $Q = \mu + \sigma Q_0$ is assumed to hold, with a symmetric Q_0, and one chooses

$$W_\mu(u) = \psi'(Q_0(u)) \tag{7}$$

then $\hat{\mu}$ is an asymptotically efficient estimator of μ. A rigorous derivation of (6) and of (10) can be obtained from equation (3) of section 9. A heuristic derivation of (6) from (4) is obtained by writing

$$\psi(\frac{\tilde{Q}(u)-\hat{\mu}}{\hat{\sigma}}) = \psi(Q_0(u)) + \psi'(Q_0(u))\{\frac{\tilde{Q}(u)-\hat{\mu}}{\hat{\sigma}} - Q_0(u)\} \tag{8}$$

Since $\psi(Q_0(u))$ and $Q_0(u)$ are odd functions, and $\psi'(Q_0(u))$ is even, the estimation equations for $\hat{\mu}$ are

$$0 = \int_0^1 \psi\left(\frac{\hat{Q}(u)-\hat{\mu}}{\hat{\sigma}}\right) du = \int_0^1 \psi'(Q_0(u))\left\{\frac{\hat{Q}(u)-\hat{\mu}}{\hat{\sigma}}\right\} du \ . \tag{9}$$

From (9) one obtains the estimator defined by (6) and (7).

An estimator of σ which is asymptotically efficient for the model $Q = \mu+\sigma Q_0$, when F_0 is a symmetric distribution, is

$$\hat{\sigma} = \frac{\int_0^1 \tilde{Q}(u) W_\sigma(u) du}{\int_0^1 Q_0(u) W_\sigma(u) du} \tag{10}$$

where

$$W_\sigma(u) = J_0(u) + Q_0(u) W_\mu(u) \ . \tag{11}$$

It is often the case that $W_\sigma(u)$ is approximately equal to $J_0(u)$ (times a constant such as 1 or 2). This helps explain why the following definition works.

Score deviations and minimum residual score deviation estimators. It is a remarkable fact that one can define a universal (and robust) measure of deviation of a sample:

$$\tilde{\sigma}_0 = \int_0^1 f_0 Q_0(u)\ \tilde{q}(u)\ du \ . \tag{12}$$

Assuming that $f_0 Q_0(u)\ \tilde{Q}(u) = 0$ for $u = 0, 1$, we can write $\tilde{\sigma}_0$ in the form

$$\tilde{\sigma}_0 = \int_0^1 J_0(u)\ \tilde{Q}(u)\ du \ , \tag{13}$$

which we call a sample score deviation. To calculate it one has to specify a score function $J_0(u)$. Note that $\tilde{\sigma}_0$ estimates a population quantity defined by

$$\sigma_0 = \int_0^1 J_0(u)\ Q(u)\ du \qquad (14)$$

which we call a score deviation.

Robust estimators of location called R-estimators can be interpreted as minimum "residual score deviation" estimators. More precisely suppose one estimates the location parameter μ by an estimator $\hat{\mu}$ whose residuals $\tilde{Q}(u)-\hat{\mu}$ have smallest score deviation:

$$\int_0^1 J_0(u)\ \{\tilde{Q}(u) - \mu\}du \text{ is minimized;}$$

it can be shown that this is precisely the definition of R-estimators.

M-estimators can also be motivated from this point of view; to avoid specifying $J_0(u) = \psi_0(Q_0(u))$ in (13) one replaces it by $\psi(\frac{\tilde{Q}_0(u)-\mu}{\sigma})$ and the criterion to estimate μ is to minimize

$$\int_0^1 \psi(\frac{\tilde{Q}(u)-\mu}{\sigma})\{\tilde{Q}(u) - \mu\}du = \int_0^1 w(\frac{\tilde{Q}(u)-\mu}{\sigma})\{\tilde{Q}(u)-\mu\}^2\ du$$

whose solution might be sought as the limit of sequences $\hat{\mu}^{(n)}$ of the form of (5)

The fact that R and M estimators of μ can be formulated as minimum residual score deviation estimators seems to explain why these methods can be extended to estimation of regression coefficients. However L estimators do not have a natural generalization to regression. Further R and M estimators yield asymptotically equivalent results when their $J_0(u)$ and ψ functions satisfy (3).

9. Data Summary by a Few Values of the Sample Quantile Function.

To form an estimated quantile function $\hat{Q}(u)$, the simplest approach is to first attempt to fit a parametric family of the form

$$Q(u) = \mu + \sigma\, Q_0(u) \qquad (1)$$

where $Q_0(u)$ is a specified quantile function; μ and σ are called location and scale parameters since $F(x) = F_0\ (\frac{x-\mu}{\sigma})$.

One seeks to form estimators $\hat{\mu}$ and $\hat{\sigma}$ which are asymptotically efficient under the hypothesis that the true quantile function satisfies (1).

Some of the aims for which the quantile function approach to statistical data analysis may provide rigorous, yet simple, methods are as follows:

1. to provide approximately efficient estimators of μ and σ under the hypothesis H_0: $Q(u) = \mu + \sigma\, Q_0(u)$;
2. to perform quick goodness of fit tests of H_0, and/or to find re-expressions (transformations) of the data which satisfy H_0;
3. to perform rigorous goodness of fit tests to identify quantile functions Q_0 for which the data satisfies H_0, and/or to estimate the density-quantile function.

Estimation of μ and σ : Efficient and tractable estimators of μ and σ which are linear functionals in $\tilde{Q}(u)$ can be found using the theory of regression analysis of continuous parameter time series developed by Parzen (1961). The asymptotic distribution theory of $\tilde{Q}(u)$ permits us to write approximately

$$\sqrt{n}\, fQ(u)\, \{\tilde{Q}(u) - Q(u)\} = B(u)$$

Under H_0,

$$Q(u) = \mu + \sigma\, Q_0(u)\ , \quad fQ(u) = \frac{1}{\sigma} f_0 Q_0(u)\ .$$

Consequently, defining $\sigma_B = \sigma/\sqrt{n}$,

$$f_0 Q_0(u)\ \tilde{Q}(u) = \mu f_0 Q_0(u) + \sigma f_0 Q_0(u)\ Q_0(u) + \sigma_B\ B(u)\ . \quad (2)$$

The parameter σ_B is linearly related to σ, but it is here treated as a free parameter. In terms of the inner product of the RKHS of the Brownian Bridge covariance kernel one may express the minimum variance unbiased estimators $\hat{\mu}$ and $\hat{\sigma}$ given $Q(u)$ for $0 \leq u \leq 1$, or $p \leq u \leq q$, or $u = u_1, \ldots, u_k$ as follows

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} = \mathrm{Inf}^{-1} \begin{pmatrix} \langle f_0 Q_0,\ (f_0 Q_0)\tilde{Q} \rangle \\ \langle (f_0 Q_0) Q_0,\ (f_0 Q_0)\tilde{Q} \rangle \end{pmatrix} \quad (3)$$

where Inf is the Information Matrix,

$$\text{Inf} = \begin{pmatrix} \langle f_0 Q_0,\ f_0 Q_0 \rangle & \langle f_0 Q_0,\ (f_0 Q_0) Q_0 \rangle \\ \langle Q_0 (f_0 Q_0),\ f_0 Q_0 \rangle & \langle (f_0 Q_0) Q_0,\ (f_0 Q_0) Q_0 \rangle \end{pmatrix} \quad (4)$$

The variance-covariance matrix of $\hat{\mu}$, σ equals $\sigma_B^2 \{\text{Inf}\}^{-1}$.

It should be emphasized that the foregoing expressions are not valid if $f_0 Q_0(u)$ and $(f_0 Q_0(u))\ Q_0(u)$ do not belong to the RKHS, which can happen in the case of the index set $0 \leq u \leq 1$. Failure to belong to the RKHS seems to be equivalent to the optimal parameter estimator involving a few extreme value order statistics, which implies that the estimators are not asymptotically normal.

If one could accomplish these aims using a "few" (say, 7) selected order statistics, then one could regard these "few" order statistics as an efficient summary of the entire sample of size n. If large samples (as well as small samples) could be effectively represented by a small number of order statistics, then every data set could be published and each reader could easily do "hands on" statistical data analysis.

The problem of choosing order statistics for the estimation of location and scale parameters has an extensive literature. The density-quantile approach has been investigated by Eubank (1979) in his Ph.D. thesis. By using location and scale estimators based on only 7 quantile values for a specified Q_0, one can identify 19 quantile values which are the union of these 7 values over a large number of familiar choices of Q_0. The proposed 19 number universal data summary consists of the median Q(0.5);

the j/16 percentiles $\tilde{Q}(j/16)$, $\tilde{Q}((8+j)/16)$ for $j = 7,6,5,4,3,2,1$; and the .01 and .02 percentiles $\tilde{Q}(0.01)$, $\tilde{Q}(0.02)$, $\tilde{Q}(0.98)$, $\tilde{Q}(0.99)$. We reproduce Tables 31 and 32 of Eubank's thesis which shows which of these order statistics are used to estimate location and scale parameters of familiar probability laws.

Table 32. Order Statistic Selection for Scale Parameter Estimation by Seven Order Statistics

Spacing	Distribution							
	Exponential or Weibull	Pareto ν = .5	Pareto ν = 1	Pareto ν = 2	Pareto ν = 3	Logistic	Normal	Lognormal
.01						✓	✓	
.02							✓	✓
.0625		✓				✓		
.125			✓				✓	✓
.1875		✓		✓	✓	✓		
.25		✓	✓					
.3125	✓	✓		✓				✓
.375			✓		✓			
.4375		✓						
.5			✓	✓		✓	✓	✓
.5625	✓	✓			✓			
.625			✓	✓				
.6875		✓			✓			✓
.75	✓		✓	✓				
.8125					✓	✓		
.875	✓		✓	✓			✓	✓
.9375	✓			✓	✓	✓		
.98	✓				✓		✓	✓
.99	✓					✓	✓	

Table 31. Order Statistic Selection for Location Parameter Estimation by Seven Order Statistics

	Distribution			
Spacing	Normal	Cauchy	Logistic	Extreme Value
.01				✓
.02	✓			✓
.0625				✓
.125	✓	✓	✓	✓
.875				
.25		✓	✓	✓
.3125	✓			
.375		✓	✓	
.4375				✓
.5	✓	✓	✓	
.5625				
.625		✓	✓	
.6875	✓			✓
.75		✓	✓	
.8125				
.875	✓	✓	✓	
.9375				
.98	✓			
.99				

References

Akaike, H. (1973) "Information theory and an Extension of the Maximum Likelihood Principle", 2nd International Symposium on Information Theory, B.H. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, 267-281.

Csorgo, M. and Revesz, P. (1978) "Strong Approximations of the Quantile Process", Annals of Statistics, 6, 882-894.

Efron, B. (1979) "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7, 1-26.

Eubank, R.L. (1979) "A Density-Quantile Function Approach to Choosing Order Statistics for the Estimation of Location and Scale Parameters," Institute of Statistics, Texas A&M University, Technical Report A-10.

Hogg, R.V. (1979) "Statistical Robustness: One View of its Applications Today", American Statistician, 33, 108-115.

Ord, J.K. and Patil, G.K. (1975) "Statistical Modelling: An Alternative View" in Patil, G.P., Kotz, S., Ord, J.K. editors Statistical Distributions in Scientific Work, Reidel: Boston, Vol. 2, 1-9.

Parzen, E. (1961) "Regression Analysis of Continuous Parameter Time Series," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, I, 469-489.

Parzen, E. (1979) "Nonparametric Statistical Data Modeling", Journal of the American Statistical Association, 105-131.

Parzen, E. (1979A)"A Density-Quantile Function Perspective on Robust Estimation", Robustness in Statistics, edited by R. Lanner and G. Wilkinson, Academic Press: New York, 237-258.

Tukey, J. W. (1965) "Which Part of the Sample Contains the Information", Proceedings of the National Academy of Sciences, 53 , 127-134.

Tukey, J.W. (1977) Exploratory Data Analysis. Reading Press: Addison Wesley.

Zacks, S. (1971) The Theory of Statistical Inference, Wiley: New York.