

ON THE HALF-SAMPLE METHOD FOR GOODNESS-OF-FIT

BY

MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 20

JULY 12, 1978

PREPARED UNDER GRANT

DAAG29-77-G-0031

FOR THE U.S. ARMY RESEARCH OFFICE

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



On the Half-Sample Method for Goodness-of-Fit

By

Michael A. Stephens

TECHNICAL REPORT NO. 20

July 12, 1978

Prepared under Grant DAAG29-77-G-0031

For the U.S. Army Research Office

Herbert Solomon, Project Director

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under Office of Naval Research Contract N00014-76-C-0475
(NR-042-267) and issued as Technical Report No. 260.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

1. Introduction. Suppose a random sample $x_1, x_2, \dots, x_n$ is given, assumed to be drawn from a population with continuous distribution $F(x; \theta)$, where θ is a vector of parameters. Some or all of the components of θ may be unknown, and when they are estimated by efficient estimators, say by maximum likelihood, we write $\hat{\theta}$ for the vector θ , with estimated components where necessary. A well-known class of goodness-of-fit tests is based on the EDF (empirical distribution function); the test statistics are measures of the discrepancy $y(x) = \sqrt{n}(F_n(x) - F(x; \theta))$, for the case where θ is known. Important statistics are $D^+ = \sup_x y(x)$, $D^- = \sup_x (-y(x))$, $D = \max(D^+, D^-)$; $V = D^+ + D^-$, $W^2 = \int y^2(x) dF$; $U^2 = \int (y(x) - \bar{y})^2 dF$; $A^2 = \int y^2(x) \omega^2(x) dF$, where $F = F(x; \theta)$ and $\omega^2(x) = \{F(1-F)\}^{-1}$ and integrals are from $-\infty$ to ∞ .

When θ is known, as above, we refer to the test situation as Case 0. When θ is replaced by $\hat{\theta}$, the distributions (including asymptotic distributions) of EDF statistics are greatly changed. When unknown components of θ are location and scale parameters, the distributions do not depend on the correct values of these parameters but only on the functional form of F . Two important examples are Case 3, where F is the normal distribution with $\theta = (\mu, \sigma^2)$, both components unknown, and Case 4, where $F(x; \theta) = 1 - \exp(-x/\theta)$ and θ is unknown. The Case numbers follow those in Stephens (1974) where the statistics above, and their power properties in the above situations, are discussed in some detail; also discussed are Cases 1 and 2, where $F(x; \theta)$ is the normal distribution with, respectively, only μ or σ^2 unknown. For W^2 , U^2 and A^2 asymptotic points can be derived theoretically (Stephens, 1976; Durbin, Knott and Taylor, 1975); for the D -group and V this has so far not been done. For finite n , a number of authors have produced percentage points for Cases 3 and 4 by Monte Carlo methods (for references see Stephens, 1974),

and for Case 4, some theoretical advances have been made by Durbin (1975b) and by Margolin and Maurer (1976). Stephens (1970, 1974) used the exact and Monte Carlo points to make modifications to the above test statistics, so that for each of Cases 0, 3 and 4, they can be used with asymptotic percentage points only.

2. The half-sample method. Rao (1972) produced an adaptation of the empirical process, using an estimate for θ , which would asymptotically behave as in Case 0, but it is quite tedious to calculate. Durbin (1973) then pointed out a much simpler process with the same property. Suppose a randomly-chosen half of the x -sample is used to estimate unknown components of θ , by efficient methods. This estimate $\tilde{\theta}$ is used to construct $F(x; \tilde{\theta})$ and EDF test statistics are calculated in the usual way, using the whole sample. Asymptotically, the behaviour of $y(x)$ (with $F(x; \tilde{\theta})$ instead of $F(x; \theta)$) is as though θ were known completely, and EDF statistics will have the distributions of Case 0 (Durbin 1973, p.59).

Two obvious applications would be to testing of normality and exponentiality as discussed above. For practical calculations, suppose $\tilde{x}$ and $\tilde{s}$ are the mean and standard deviation of the random half-sample; these replace $\bar{x}$ and s in the calculations for z_i set forth in Section 3.1 of Stephens (1974). (Note that there is a misprint in Case 4: z_i should be given by $z_i = 1 - \exp(-x_i/\bar{x})$, and in the half-sample method z_i would be $1 - \exp(-x_i/\tilde{x})$). The set z_i is calculated from the x_i for the whole sample, and then the test statistics are calculated as given by Stephens. Their asymptotic percentage points will be those in his Table 1.0, and we shall see below that the finite- n modifications also work well for most statistics (but importantly not A^2).

3. The method of random substitution. Another technique for making a test with unknown parameters into a Case 0 test has also been given by Durbin (1961). When the method is applied to goodness-of-fit tests, a combination of the given sample and an arbitrarily chosen value of the unknown parameter(s) is used to construct a new set, S , of sample values. Durbin calls this procedure the method of random substitution, and shows that the distribution function of the values x in S is $F(x, \theta)$, with θ known; thus the original null hypothesis can be tested using S and Case 0 points. In subsequent work Durbin (1975a, 1976) has further shown that the half-sample device and the method of random substitution have empirical processes which are asymptotically equivalent even for true distributions alternative to the null; thus we can expect these two techniques to possess similar power properties, at least for large samples. Durbin (1976) discusses the two methods from a practical point of view, and suggests that the half-sample technique is easier and more attractive in practice. The existence of this straightforward method then implies that the effort expended by many authors in finding percentage points for Cases 3 and 4, as described in Section 1, might have been unnecessary; one merely splits the sample, estimates parameters, and refers to existing Case 0 points, at least for large samples! An objection, for many statisticians, to the half-sample device (also to the method of random substitution) is that two statisticians can obtain different values for test criteria from the same data. This objection does not apply to standard Case 3 and Case 4 procedures; we now show that there is also a considerable difference in power.

4. Comparison of the half-sample method and the standard Case 3 and Case 4 methods.

(a) Percentage points for finite n . If the half-sample device is followed, the distributions of EDF statistics behave asymptotically as in Case 0: it would be a remarkable bonus if the distributions for finite n were not too different also. This has been investigated as follows. For normal samples of size n , the half-sample method for testing normality (called HS3) was followed, and modified forms of EDF statistics calculated as in Stephens (1974, Table 1.0). Hopefully these would then have their Case 0 asymptotic distributions, with the percentage points in that Table. Results based on 10,000 Monte Carlo samples are given in Table 1. It can be seen that the results are very good, except for A^2 . The experiment was repeated for exponential samples, using the split-sample method for the exponential test (HS4), with results in Table 2. Again A^2 gives poor results, and also D^+ and D^- . Note that the A^2 points are those for the unmodified statistic, because in Case 0 A^2 requires no modification for finite samples; "asymptotic" begins at $n = 3$. In the power studies which follow, the Monte Carlo points in Tables 1 and 2 have been used for all the modified statistics.

(b) Power studies. We next examine the power of HS3 and HS4 against the rival 'standard' Case 3 and Case 4 techniques, also by Monte Carlo methods. From various alternative distributions, at least 1000 samples of size n were taken and tested by the two techniques. Table 3 gives the results for testing normality, both parameters unknown, and Table 4 gives them for the test for exponentiality, θ unknown. The tables record the percentage of Monte Carlo samples declared significant by the statistics when tests were made at the 10% level. Tables are available also for 5% tests.

5. Discussion. (a) In Tables 3 and 4 it can be seen that the half-sample method produces uneven powers for the various test statistics; these results throw some light on the effect of the method. The null hypothesis is that x has distribution $F(x;\theta)$, and the test statistics are calculated from values $z_i = F(x_i;\theta^*)$, $i = 1, 2, \dots, n$, where $\theta^* = \theta$, or θ^* is an estimate when θ is unknown. For Case 0, $\theta^* = \theta$ and if the null hypothesis is true the z_i are uniformly distributed between 0 and 1; this is also true for the z_i found from the half-sample method, or by the method of random substitution. When the null hypothesis is not true, for any of these techniques, the z_i will not be uniform, and their pattern can be roughly determined by observing which test statistics are large. When D^+ or D^- are large, z -values have moved towards 0 or 1 respectively; large D and W^2 imply a shift of the mean to either right or left, and large V or U^2 imply a clustering of z -values, or a division into two groups towards 0 and 1. Large values of A^2 also imply values approaching either extreme. As an illustration, consider the test for normality, on a sample of size $n = 10$, when in fact the true distribution is χ_1^2 . When the half-sample method was used, it was found that in most samples the z -values were almost all between 0.1 and 0.4; thus D^+ will be large, and the clustering makes V and U^2 large also, leading to high powers for these statistics in Table 3. The relative powers thus give some indication of the z -pattern. When standard Case 3 techniques are used in the same problem, the z -values do not cluster together as above; the fitted normal curve is in fact close enough to the EDF that no one statistic is overwhelmingly superior to the others. This levelling of power in standard Case 3 and Case 4 techniques was noted previously (Stephens, 1974); it can be seen again throughout Tables 3 and 4.

(b) Turning now to the main comparison, we see that the half-sample method is nowhere superior to use of standard Case 3 or Case 4 techniques, and the work done in providing percentage points for these Cases has after all been useful. Where tables of percentage points for EDF statistics have been prepared for other distributions under test, with unknown parameters, we might expect to find similar comparisons. Such tables exist for the Gamma distribution (Pettitt and Stephens (1976)) and the extreme value distribution (Stephens, 1976b).

Among the EDF statistics it is clear from the tables that on the whole, for a test for normality, the Kolmogorov D is a poor statistic, and that A^2 has a slight superiority over the others. This bears out results in Stephens (1974, Tables 5 and 6). In Case 4, D is not relatively so weak; also, if one were sure of the direction of skewness of the alternative distribution, statistics D^+ or D^- will give good power for this Case. Overall, however, A^2 still scores best when this information is not available. The evidence of both tables gives good support to A^2 as an omnibus test statistic for both Cases 3 and 4.

Finally we emphasise that this work does not obviate the usefulness of the split-sample method for occasions where parameters to be estimated are not location and/or scale parameters or where the necessary tables have not been provided. Some comparisons are being made between this method and Pearson's chi-square, its obvious rival in these situations.

Acknowledgements. This work was supported in part by the National Research Council of Canada, and in part by the U.S. Office of Naval Research Contract N00014-76-C-0475 and by the U.S. Army Research Office Grant DAAG29-77-G-0031. The author thanks Mr. Al Byrd for much help with computations.

Table 1

Upper tail percentage points for modified EDF statistics:

half-sample test for normality.

The points are found from 10000 Monte Carlo samples for each sample size.

Statistic	Sample Size	$\alpha(\%)$:	10	5	2.5	1.0
D^+	10		1.087	1.237	1.376	1.520
	20		1.070	1.223	1.349	1.512
	50		1.076	1.240	1.380	1.550
	∞		1.073	1.224	1.358	1.518
D^-	10		1.078	1.232	1.363	1.485
	20		1.064	1.231	1.368	1.540
	50		1.067	1.216	1.358	1.529
	∞		1.073	1.224	1.358	1.518
D	10		1.235	1.370	1.476	1.585
	20		1.222	1.358	1.479	1.636
	50		1.230	1.371	1.493	1.635
	∞		1.224	1.358	1.480	1.628
V	10		1.595	1.705	1.787	1.867
	20		1.613	1.740	1.842	1.968
	50		1.615	1.738	1.854	1.992
	∞		1.620	1.747	1.862	2.001
W^2	10		.351	.456	.566	.711
	20		.341	.457	.580	.779
	50		.351	.480	.604	.767
	∞		.347	.461	.581	.743
U^2	10		.149	.182	.207	.230
	20		.152	.184	.218	.256
	50		.154	.188	.222	.267
	∞		.152	.187	.221	.267
A^2	10		3.00	4.55	6.68	10.34
	20		2.27	3.13	4.27	6.09
	50		2.09	2.76	3.52	4.44
	∞		1.93	2.49	3.07	3.86

Table 2

Upper tail percentage points for modified EDF statistics:

half-sample test for exponentiality

The points are found from 10000 Monte Carlo samples for each sample size.

Statistic	Sample size	$\alpha(\%)$:	10	5	2.5	1.0
D^+	10		.943	1.088	1.215	1.362
	20		.982	1.132	1.257	1.4
	50		1.025	1.167	1.284	1.4
	∞		1.073	1.224	1.358	1.518
D^-	10		1.179	1.316	1.422	1.568
	20		1.140	1.298	1.423	1.566
	50		1.111	1.270	1.421	1.581
	∞		1.073	1.224	1.358	1.518
D	10		1.230	1.354	1.465	1.596
	20		1.228	1.351	1.471	1.619
	50		1.222	1.363	1.499	1.635
	∞		1.224	1.358	1.480	1.628
V	10		1.617	1.738	1.837	1.954
	20		1.687	1.735	1.851	1.988
	50		1.609	1.735	1.851	1.988
	∞		1.620	1.747	1.862	2.001
W^2	10		.352	.466	.579	.717
	20		.345	.464	.594	.758
	50		.349	.476	.598	.774
	∞		.347	.461	.581	.743
U^2	10		.152	.187	.218	.253
	20		.151	.184	.219	.260
	50		.151	.185	.220	.267
	∞		.152	.187	.221	.267
A^2	10		2.18	2.99	3.92	5.60
	20		2.05	2.69	3.50	4.50
	50		1.99	2.66	3.29	4.31
	∞		1.93	2.49	3.07	3.86

Table 3

Comparison of power, test for normality:

half-sample method (HS) versus standard Case 3 (C3).

The table gives the percentage of Monte Carlo samples declared significant by the test statistics. The test is at the 10% level.

True distribution	n	Method	Statistic:	D^+	D^-	D	V	W^2	U^2	A^2
uniform	10	HS		9	10	11	9	10	10	10
	10	C3		9	7	14	18	16	18	16
	20	HS		11	13	12	16	10	15	11
	20	C3		16	16	21	28	29	32	31
χ_1^2	10	HS		33	7	22	48	21	50	18
	10	C3		68	33	67	76	77	76	79
	20	HS		55	39	51	84	50	82	58
	20	C3		95	87	94	97	98	97	98
χ_2^2	10	HS		23	7	18	29	15	30	13
	10	C3		45	9	43	52	53	52	56
	20	HS		40	12	31	59	29	56	31
	20	C3		75	40	69	77	79	76	84
χ_4^2	10	HS		16	7	13	18	12	18	10
	10	C3		28	4	27	32	32	32	34
	20	HS		28	7	20	33	19	33	19
	20	C3		50	8	42	46	53	49	58
χ_6^2	10	HS		15	8	13	15	12	16	13
	10	C3		25	3	25	25	26	26	28
	20	HS		23	6	16	26	14	26	15
	20	C3		43	4	35	34	40	36	44
Cauchy	10	HS		23	25	35	42	31	44	23
	10	C3		50	48	65	67	67	67	67
	20	HS		47	47	61	71	58	72	66
	20	C3		79	80	88	90	91	91	91
log-normal	10	HS		28	7	21	38	19	38	17
	10	C3		62	20	59	66	67	66	69
	20	HS		52	22	44	71	43	70	49
	20	C3		90	65	87	90	92	90	93

Table 4

Comparison of power, test for exponentiality:half-sample method (HS) versus standard Case 4 (C4).

The table gives the percentage of Monte Carlo samples declared significant by the test statistics. The test is at the 10% level.

True distribution	n	Method	Statistic: D^+ D^- D V W^2 U^2 A^2							
uniform	10	HS	3	24	20	44	18	41	10	
	10	C4	1	55	43	48	51	48	45	
	20	HS	10	50	40	74	43	68	35	
	20	C4	14	83	70	79	82	75	80	
χ_1^2	10	HS	35	8	22	29	21	29	39	
	10	C4	42	2	32	29	35	30	53	
	20	HS	54	8	37	49	40	50	61	
	20	C4	69	1	56	49	60	51	77	
χ_4^2	10	HS	3	22	18	28	17	29	10	
	10	C4	1	46	35	32	40	36	33	
	20	HS	3	38	29	47	31	49	26	
	20	C4	1	70	54	50	61	54	60	
χ_6^2	10	HS	2	36	30	51	29	56	15	
	10	C4	0	74	62	60	71	65	65	
	20	HS	5	74	64	88	65	90	59	
	20	C4	3	96	91	88	95	91	95	
half-Cauchy	10	HS	38	7	30	29	30	30	41	
	10	C4	49	4	43	39	47	41	49	
	20	HS	55	6	49	50	48	50	60	
	20	C4	73	2	68	59	70	61	70	
log-normal	10	HS	13	11	14	13	14	13	13	
	10	C4	12	15	18	17	18	17	15	
	20	HS	19	14	19	22	20	22	18	
	20	C4	18	18	21	22	23	25	23	
half-normal	10	HS	3	16	13	19	13	19	9	
	10	C4	1	25	20	19	22	21	18	
	20	HS	3	24	17	25	17	25	14	
	20	C4	1	45	28	26	33	28	29	

References

- Durbin, J. (1961) Some methods of constructing exact tests. Biometrika, 48, 41-55.
- Durbin, J. (1973) Distribution theory for tests based on the sample distribution function. S.I.A.M. Publications No. 9: Philadelphia.
- Durbin, J. (1975a) Tests of model specification based on residuals. A survey of Statistical Design and Linear Models. Ed. J.N. Srivastava, Rotterdam, North Holland.
- Durbin, J. (1975b) Kolmogorov-Smirnov tests where parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika, 62, 5-22.
- Durbin, J. (1976) Kolmogorov-Smirnov tests when parameters are estimated. Empirical Distributions and Processes. Lecture Notes in Mathematics No. 566. Ed. P. Gaenssler and P. Revesz, Berlin, Springer-Verlag.
- Durbin, J., Knott, M., and Taylor, C.C. (1975) Components of Cramér-von Mises Statistics, II, Jour. Roy. Statist. Soc., B, 37, 216-237.
- Margolin, Barry H. and Maurer, Willi (1976), Tests of the Kolmogorov-Smirnov type for exponential data with unknown scale, and related problems, Biometrika, 63, 149-160.
- Pettitt, A.N. and Stephens, M.A. (1976) EDF Statistics for Testing for the Gamma Distribution, with Application to Testing for Equal Variances, submitted to Jour. Amer. Statist. Assoc.
- Rao, K.C. (1972) The Kolmogorov, Cramér-von Mises, chi-square statistics for goodness-of-fit tests in the parametric case. Abstract 133-6, Bull. Inst. Math. Statist. 1, 87.
- Stephens, M.A. (1970) Use of Kolmogorov, von Mises, and periodogram statistics without extensive tables, Jour. Roy. Statist. Soc. B, 32, 115-122.
- Stephens, M.A. (1974) EDF statistics for goodness-of-fit and some comparisons, Jour. Amer. Statist. Assn., 69, 730-737.
- Stephens, M.A. (1976a) Asymptotic results for goodness-of-fit statistics with unknown parameters, Annals. Statist., 4, 357-369.
- Stephens, M.A. (1976b) Goodness-of-fit for the extreme-value distribution. Technical Report, Stanford University, Dept. of Statistics, To appear in Biometrika.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 20	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Half-Sample Method for Goodness-of-Fit		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael A. Stephens		8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-G-0031
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE July 12, 1978
		13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		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 findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents. This report partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 260.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Empirical process; goodness-of-fit; half-sample method; test for exponentiality; test for normality.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Please see reverse side.		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Summary

Two interesting processes, related to the empirical distribution function, have been pointed out by Rao (1972) and by Durbin (1973). The second, in particular, leads to the half-sample method, an elegant and simple technique for dealing with unknown parameters in goodness-of-fit testing, without the necessity of new tables of percentage points for each distribution tested. To those who have spent some effort in producing such tables, it is natural to wonder whether this effort has been in vain. In this paper we show that the answer is no.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)