THE STABILIZED PROBABILITY PLOT

By

John R. Michael

Technical Report No. 165
Department of Statistics ONR Contract

June 1982

Research sponsored by the Office of Naval Research
Contract N00014-82-K-0207

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

THE STABILIZED PROBABILITY PLOT

By John R. Michael

Department of Statistics, Southern Methodist University

SUMMARY

The stabilized probability plot or SP plot is introduced. An attractive feature of the SP plot that enhances its interpretability is that the variances of the plotted points are approximately equal. This prompts the definition of a new and powerful goodness-of-fit statistic D_{sp} which, analogous to the standard Kolmogorov-Smirnov statistic D, is defined to be the maximum deviation of the plotted points from their theoretical values. Using either D or D_{sp} it is shown how to construct acceptance regions for QQ, PP, and SP plots. Acceptance regions can help remove much of the subjectivity from the interpretation of these probability plots.

Some Key Words: Goodness-of- fit; Graphical methods; Kolmogorov-Smirnov statistic; Percent-Percent plot; Quantile-Quantile plot; Variance stabilizing transformation.

1. INTRODUCTION

Let $y_1 \le y_2 \le \cdots \le y_n$ be the realization of an ordered random sample of size n from the distribution F. A quantile-quantile plot or QQ plot for a continuous hypothesized location-scale distribution $F_0\{(y-\mu)/\sigma\}$ is constructed by plotting each sample quantile y_i versus a corresponding theoretical standard quantile $x_i=F_0^{-1}(t_i)$, where t_i is an appropriate cumulative proportion. We will choose $t_i = (i-.5)/n$. Similarly, a percent-percent plot or PP plot is constructed by plotting each probability-integral-transformed value $u_i = F_0\{(y_i-\mu)/\sigma\}$ versus the uniform quantile t_i. If μ and σ are unknown, they are replaced by maximum likelihood estimates. See Wilk & Gnanadesikan (1968) for a discussion of QQ and PP plots.

A common occurrance with QQ plots is that certain points, determined by F, are much more variable than others. For example, when F is normal the points nearest the middle of the plot have the smallest variances. The opposite is true for PP plots when $F_0 = F$, regardless of the form of F. A transformation is now described which stabilizes the variances of the plotted points. This enhances the interpretability of the plot and prompts the definition of a new and powerful graphical goodness-of-fit test.

2. THE STABILIZED PROBABILITY PLOT

When $F = F_0$ and μ and σ are known, μ_i can be regarded as the realization of a uniform order statistic. If parameters are efficiently estimated, this is true asymptotically. The arc sine transformation can be used to stabilize the variance of a uniform order statistic just as it does for a binomial random variable.

Suppose we let $S = (2/\pi)\arcsin(U^{\frac{1}{2}})$ where U has the uniform (0,1) distribution. Then the probability density function of S is given by $(\pi/2)\sin(\pi s)$

for $0 \leq s \leq 1$. This distribution will be termed the sine distribution since the density function is proportional to a half-cycle of a sine wave. The sine distribution has the interesting property than its order statistics have the same asymptotic variance: if $S_1 \leq S_2 \leq \ldots \leq S_n$ is an ordered random sample from the sine distribution, then as $n \to \infty$ and $i/n \to p$ the asymptotic variance of nS_i is $1/\pi^2$, independent of p.

The stabilized probability plot or SP plot is now defined as the plot of each $s_i = (2/\pi) \arcsin (u_i^{1/2})$ versus $r_i = (2/\pi) \arcsin (t_i^{1/2})$. If $F_0 = F$ and μ and σ are known, r_i is the mode of S_i. Plotting formulas are summarized in Table 1.

Table 1

Formulas for Constructing Probability Plots

type plot	abcissa	ordinate
QQ	$x_i = F_0^{-1}\{(i-.5)/n\}$	y_i
PP	$t_i = (i-.5)/n$	$u_i = F_0\{(y_i-\mu)/\sigma\}$
SP	$r_i = (2/\pi)\arcsin[\{(i-.5)/n\}^{1/2}]$	$s_i = (2/\pi)\arcsin[F_0^{1/2}\{(y_i-\mu)/\alpha\}]$

3. A NEW GOODNESS-OF-FIT STATISTIC

The standard Kolmogorov-Smirnov statistic can be expressed as $D = \max_{1 \leq i \leq n} |t_i - u_i| + .5/n$. Analogous to D we now define the statistic $D_{sp} = \max_{1 \leq i \leq n} |r_i - s_i|$ which can be used to test the hypothesis that $F = F_0$.

Exact critical points for D_{sp} for testing a simple hypothesis were computed using a recursive algorithm described by Noe (1972). These critical points are given in Table 2 for selected values of $n \leq 100$. Note the peculiar lack of monotonicity in n for $n \leq 3$ and each $\alpha \neq .10$.

For the composite case of normality with both μ and σ unknown, critical points were estimated using Monte Carlo methods. For selected values of

TABLE 2

Critical Points For D_{sp}

	Simple Test of Uniformity						Composite Test of Normality				
n	.50	.25	.10	.05	.01	← α →	.50	.25	.10	.05	.01
1	.167	.270	.356	.399	.455						
2	.198	.273	.353	.406	.495						
3	.195	.260	.333	.377	.461		.143	.212	.249	.261	.271
4	.188	.248	.311	.351	.430		.164	.208	.242	.272	.316
5	.180	.235	.292	.329	.403		.104	.129	.154	.168	.198
6	.173	.224	.277	.311	.380		.102	.126	.149	.163	.193
7	.167	.215	.264	.296	.361		.099	.122	.144	.158	.188
8	.161	.206	.252	.283	.344		.097	.119	.140	.154	.182
9	.156	.198	.242	.271	.330		.095	.116	.136	.149	.177
10	.152	.192	.233	.261	.317		.093	.113	.132	.145	.172
11	.147	.186	.226	.252	.306		.091	.110	.129	.141	.167
12	.143	.180	.218	.244	.296		.090	.108	.126	.138	.163
13	.140	.175	.212	.236	.286		.088	.106	.123	.135	.159
14	.137	.170	.206	.230	.278		.086	.104	.120	.132	.156
15	.133	.166	.201	.224	.270		.085	.102	.118	.129	.152
16	.131	.162	.196	.218	.263		.083	.100	.116	.127	.149
17	.128	.159	.191	.213	.257		.082	.098	.114	.124	.146
18	.125	.155	.187	.208	.251		.081	.096	.112	.122	.144
19	.123	.152	.183	.203	.245		.080	.095	.110	.120	.141
20	.121	.149	.179	.199	.240		.079	.093	.108	.118	.139
21	.119	.146	.176	.195	.235		.078	.092	.107	.117	.137
22	.117	.144	.172	.191	.230		.077	.091	.105	.115	.135
23	.115	.141	.169	.188	.226		.076	.089	.104	.113	.133
24	.113	.139	.166	.185	.222		.075	.088	.102	.112	.131
25	.111	.137	.164	.181	.217		.074	.087	.101	.110	.129
30	.104	.127	.152	.168	.201		.070	.082	.095	.104	.122
35	.098	.120	.142	.157	.188		.066	.078	.090	.098	.115
40	.093	.113	.134	.148	.177		.064	.075	.086	.094	.110
45	.089	.108	.128	.141	.168		.061	.072	.083	.090	.106
50	.085	.103	.122	.134	.160		.059	.069	.080	.087	.102
60	.079	.095	.113	.124	.148		.055	.065	.075	.082	.096
70	.074	.089	.105	.116	.137		.052	.062	.071	.077	.091
80	.070	.084	.099	.109	.129		.050	.059	.068	.074	.087
90	.067	.080	.097	.103	.122		.048	.056	.065	.070	.084
100	.064	.076	.090	.098	.116		.046	.054	.062	.068	.081

$n \leq 100$, 10,000 independent samples were generated, the statistic D_{sp} calculated, and the appropriate sample quantiles recorded. Estimated critical points, smoothed for $n \geq 5$, are also given in Table 2.

4. ACCEPTANCE REGIONS USING D AND D_{sp}

Both D and D_{sp} measure the maximum deviation of plotted points from their theoretical values. Therefore acceptance regions are easily constructed using the appropriate critical points. Formulas are given in Table 3 for three types of plots for each of the two statistics. Of course for a partiular statistic the tests provided by the three acceptance regions are equivalent. One can perform a graphical goodness-of-fit test at the α-level of significance by simply looking to see if all the points fall inside the $100(1-\alpha)\%$ acceptance region.

TABLE 3

Formulas for Constructing $100(1-\alpha)\%$ Acceptance Regions on Probability Plots Using α-level Critical Points d and d_{sp} for D and D_{sp}, Respectively

type plot	statistic	lines defining acceptance region
QQ	D	$y = \mu + \sigma F_0^{-1}\{F_0(x) \pm (d - .5/n)\}$
PP	D	$u = t \pm (d - .5/n)$
SP	D	$s = (2/\pi)\arcsin[\{\sin^2(\pi r/2) \pm (d-.5/n)\}^{\frac{1}{2}}]$
QQ	D_{sp}	$y = \mu + \sigma F_0^{-1}(\sin^2[\arcsin\{F_0^{\frac{1}{2}}(x)\} \pm \pi d_{sp}/2])$
PP	D_{sp}	$u = \sin^2\{\arcsin(t^{\frac{1}{2}}) \pm \pi d_{sp}/2\}$
SP	D_{sp}	$s = r \pm d_{sp}$

5. POWER COMPARISONS

The power comparisons here parallel those reported by Stephens (1974). For testing a simple null hypothesis at the .10 level of significance, exact power for D and D_{sp} was computed using the recursive algorithm of Noe (1972). Following Stephens (1974), three families of alternative distributions were chosen, denoted A, B, and C, each parameterized by k. For k = 1 each distribution is uniform and as k increases each distribution becomes increasingly "nonuniform". For k > 1 family A has more probability near 0 than the uniform distribution, B has more probability near .5, and C has more probability near 0 and 1.

Percentage power is shown in Table 4 for D and D_{sp} for three choices of k and three different sample sizes. For D the critical points .369, .265, and .189 were adapted from Miller (1956). For D_{sp} the critical points .233, .179, and .134 were taken from Table 2. Compared to D, the performance of D_{sp} is slightly better for family A, much better for family B, and somewhat worse for family C. The power figures for D_{sp} can also be compared to those for other statistics reported by Stephens (1974) in his Table 3. The power of D_{sp} is relatively good for families A and B, but poor for family C. Thus when testing a simple null hypothesis the statistic D_{sp} is an improvement over D, but not as good overall as, say, the Watson statistic U^2 as reported by Stephens (1974). The disadvantage of a statistic such as U^2 is that its acceptance region cannot be represented graphically on a probability plot.

TABLE 4

Percentage Power When Testing For Uniformity On The Unit Interval: $\alpha = .10$

		n=10		n=20		n=40	
family	k	D	D_{sp}	D	D_{sp}	D	D_{sp}
	1.5	25	26	40	43	66	68
A	2.0	53	55	81	83	98	99
	3.0	90	91	100	100	100	100
	1.5	9	14	13	27	23	49
B	2.0	12	24	27	56	60	91
	3.0	24	50	67	93	99	100
	1.5	19	16	25	19	36	27
C	2.0	31	27	46	39	73	65
	3.0	53	51	82	78	99	98

For testing the composite hypothesis of normality at the .05 level of significance with both μ and σ unknown, power was estimated using Monte Carlo methods. Percentage power is shown in Table 5 for twelve alternative distributions and three different sample sizes. For D the critical points .262, .192, and .159 were calculated from Stephens (1974). For D_{sp} the critical points .145, .118, and .104 were taken from Table 2. Each pair of entries for D and D_{sp} in Table 5 is based on a different set of 1000 independent samples and represents the percentage of samples that were observed to be significant. In most instances D_{sp} can be seen to be more powerful than D. The power figures for D_{sp} can be compared to those of other statistics reported by Stephens (1974) in his Table 5. The statistic D_{sp} is bettered only by the Anderson-Darling Statistic A^2 and the Shapiro-Wilk statistic W. Even in these cases the differences in power are not great. Again the reader is reminded that the acceptance regions for statistics such as A^2 and W cannot be represented graphically on a probability plot.

TABLE 5

Estimated Percentage Power when Testing for Normality; μ and σ Unknown; $\alpha = .05$

Distribution	n=10		n=20		n=30	
	D	D_{sp}	D	D_{sp}	D	D_{sp}
Chi-squared (1)	58	65	89	99	98	100
Exponential	33	35	60	81	80	98
Chi-squared (3)	25	24	42	61	61	85
Chi-squared (4)	20	20	32	45	48	73
Chi-squared (10)	12	11	16	18	23	30
Lognormal	47	48	79	91	93	99
Uniform	8	7	11	14	15	22
Laplace	13	13	20	23	30	31
Student-t (1)	59	60	86	87	93	93
Student-t (3)	19	18	27	31	35	38
Student-t (4)	12	13	17	20	19	22
Student-t (6)	9	9	9	11	11	14

6. AN ILLUSTRATION WITH DISCUSSION

We now illustrate the different types of plots and acceptance regions using a sample of size 20 to test the composite null hypothesis of normality. The actual sample values are 2.6, 2.7, 2.9, 3.0, 3.0, 3.1, 3.2, 3.4, 3.7, 3.7, 3.9, 4.0, 4.2, 4.3, 4.3, 4.8, 4.8, 5.3, 6.6, and 7.6. The maximum likelihood estimates for μ and σ are 4.055 and 1.259 respectively. Figure 1 shows QQ, PP, and SP plots of the data along with 95 percent acceptance regions using both D and D_{sp}.

The sample was selected from one of several samples simulated from the exponential distribution with location and scale parameters 2.5 and 1.0 respectively. Thus the correct decision here is to reject the hypothesis of normality. Note that we can indeed reject the null hypothesis at the .05 level of significance if we use D_{sp}, but we cannot do so if we use D. This illustrates the increase in power over D that D_{sp} provides when testing

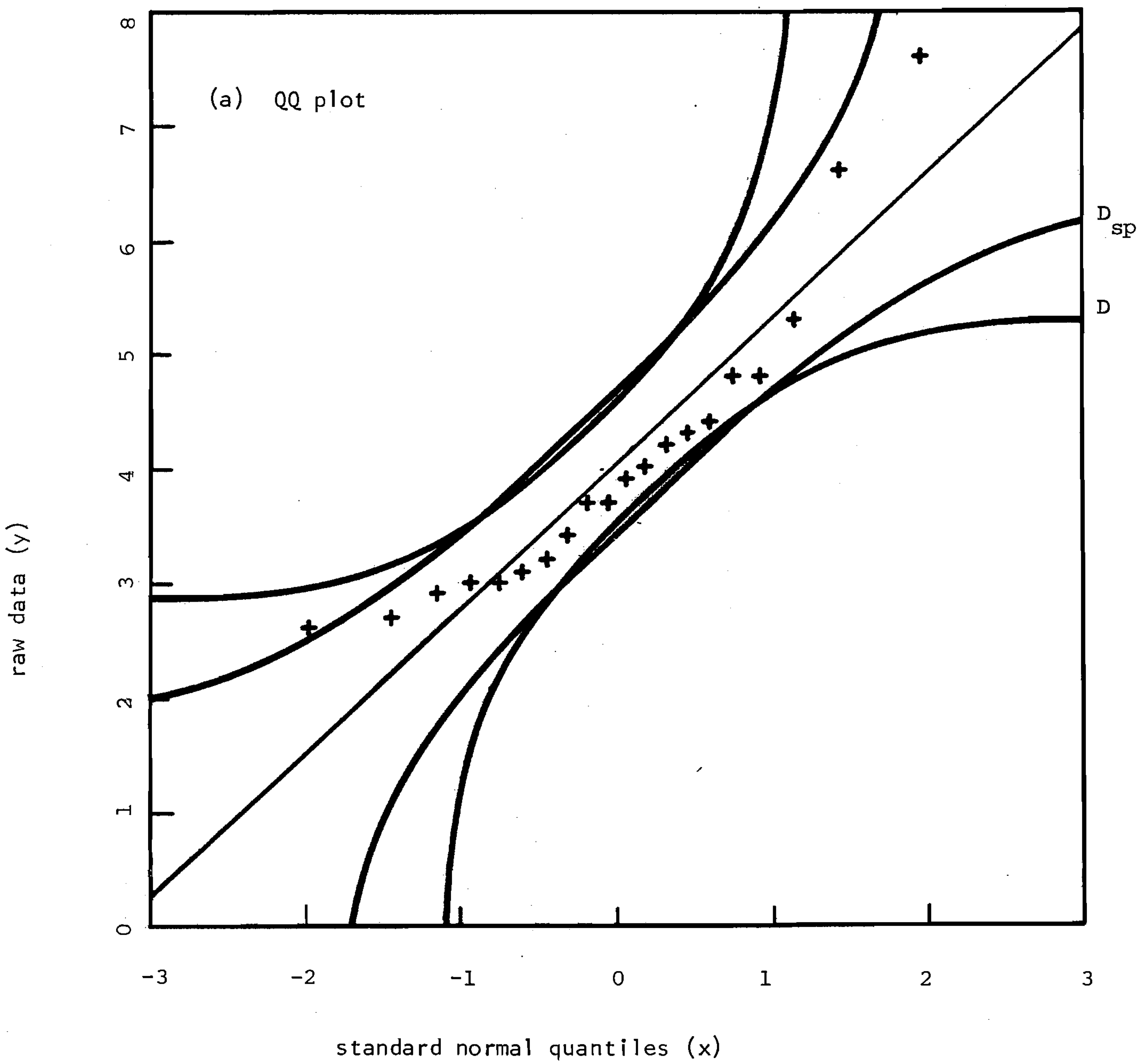
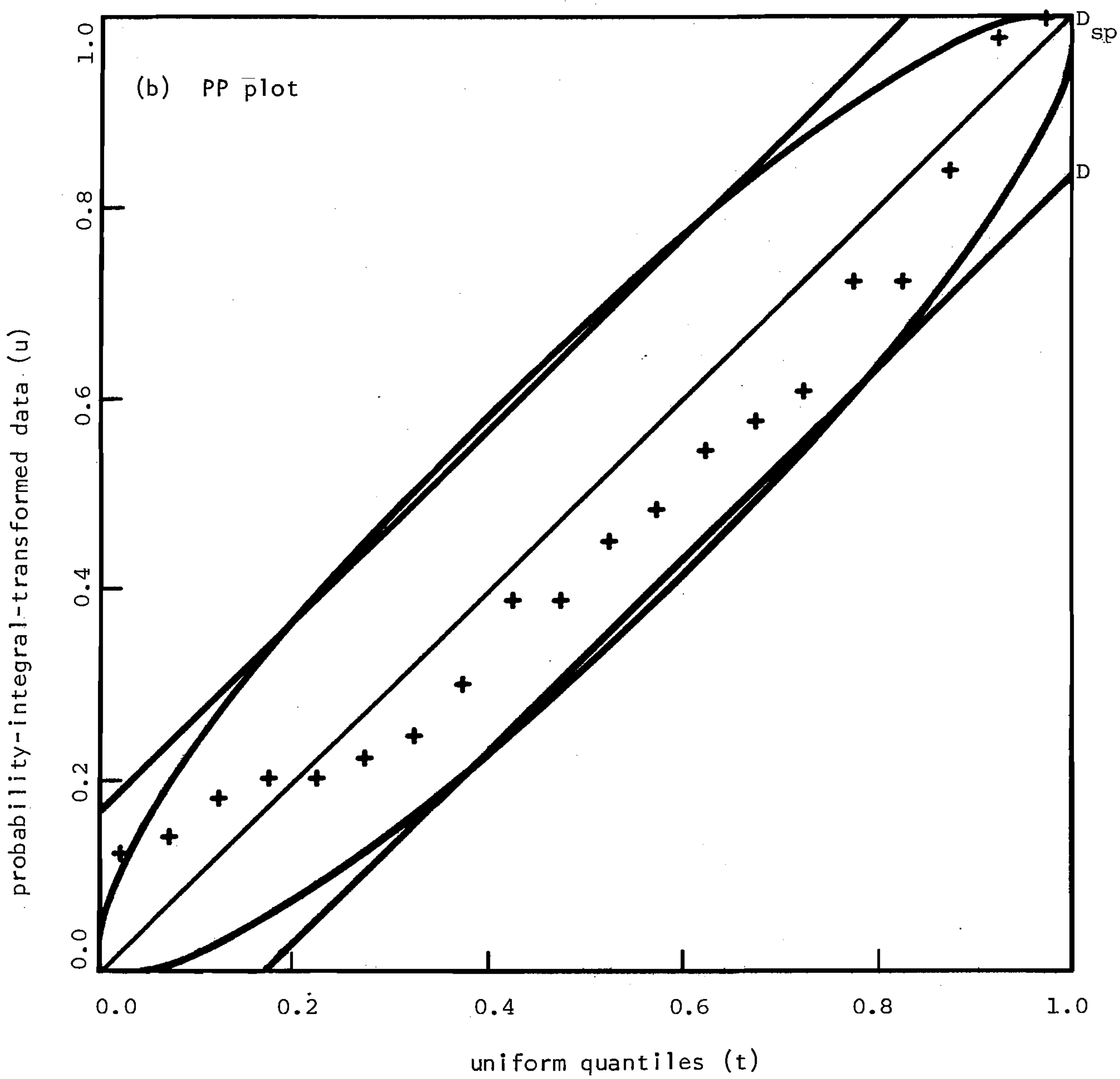
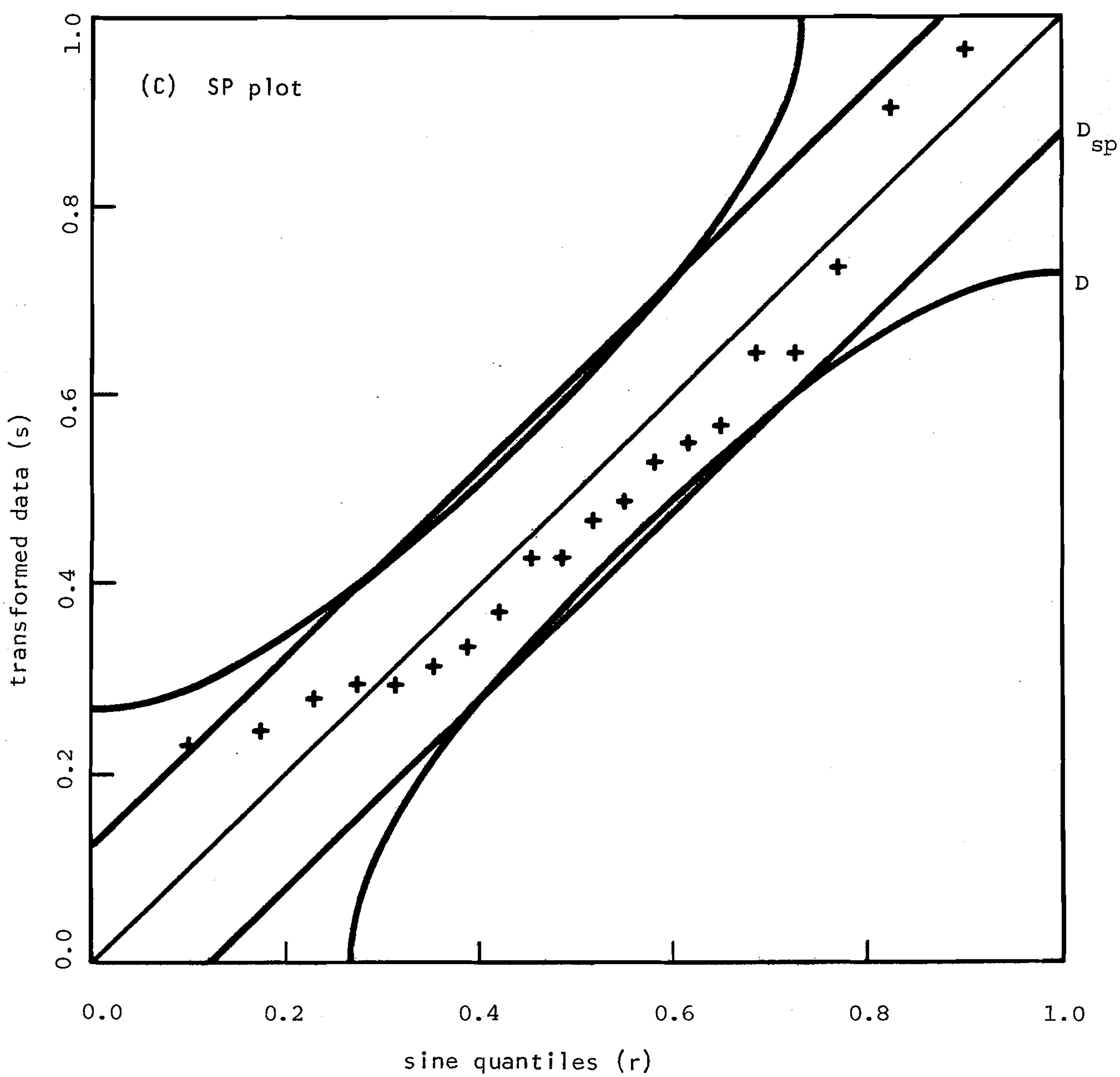


FIGURE 1: Probability plots with 95% acceptance regions based on D and D_{sp}: (a) QQ plot, (b) PP plot, (c) SP plot.


(b) PP plot
probability-integral-transformed data (u)
uniform quantiles (t)
0.0
0.2
0.4
0.6
0.8
1.0
D
sp
D


(C) SP plot
transformed data (s)
sine quantiles (r)
D_{sp}
D
0.0
0.2
0.4
0.6
0.8
1.0

an exponential distribution for normality. From Table 5 in the previous section we see that the estimated percentage powers for D and D_{sp} in this situation are 60 and 81 respectively.

The SP plot and the PP plot share many desirable features. The abcissas of the points depend only upon the sample size and not upon the hypothesized distribution. The points always fall within the unit square and are never bunched closely together as with QQ plots for certain distributions. Plots using different hypothesized distributions can be juxtaposed and the fits easily compared. Once familiar with SP or PP plots, the user need not reorient himself to plots that appear markedly different for different distributions, as can QQ plots.

The SP plot enjoys two advantages over traditional probability plots. The SP plot is easier to interpret since the variances of the ordinates are approximately equal. Also, the acceptance region based upon D_{sp} can be added to the plot by simply drawing two straight lines. Alternatively, the observed value of D_{sp} can be read directly from the plot.

7. RELATED METHODS

It is well known that the standard Kolmogorov-Smirnov statistic D for testing a simple null hypothesis can be used to construct a distribution-free confidence band for F. Letting d be the α-level critical point and F_n be the familiar empirical cumulative distribution function, then a $100(1-\alpha)\%$ confidence band for F is given by $F_n \pm d$. Note that such a confidence band is different in nature from the acceptance regions for probability plots discussed in Section 4, although there is a correspondence: the confidence band contains F_0 if and only if the acceptance region contains all the plotted points. Bickel & Doksum (1977, page 383) discuss this technique and its extension to the case of the normal distribution with unspecified parameters. Iman (1982) constructs acceptance

REFERENCES

Bickel, P. J. & Doksum, K. A. (1977). Mathematical Statistics, San Francisco: Holden-day.

Borokov, A. A. & Sycheva, N. M. (1968). On asymptotically optimal non-parametric criteria. Theory of Probability and its Applications, 13, 359-393.

Canner, P. (1975). A simulation study of one- and two-sample Kolmogorov-Smirnov statistics with a particular weight function. Journal of the American Statistical Association, 70, 209-211.

Doksum, K. A. & Sievers, G. L. (1976). Plotting with confidence: graphical comparisons of two populations. Biometrika, 63, 421-434.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. Technometrics, 17, 111-117.

Iman, R. L. (1982). Graphs for use with the Lillifors test for normal and exponential distributions. The American Statistician, 36, 109-112.

Miller, L. H. (1956). Tables of percentage points of Kolmogorov statistics. Journal of the American Statistical Association, 51, 111-121.

Noe, M. (1972). The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics. Annals of Mathematical Statistics, 43, 58-64.

Stephens, M. A. (1974). EDF statistics for goodness-of-fit and some comparisons. Journal of the American Statistical Association, 69, 730-737.

Wilk, M. B. & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 165
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): THE STABILIZED PROBABILITY PLOT
5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER: 165
7. AUTHOR(s): John R. Michael
8. CONTRACT OR GRANT NUMBER(s): N00014-82-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Southern Methodist University, Dallas, Texas 75275
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-479
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, VA 22217
12. REPORT DATE: June 1982
13. NUMBER OF PAGES: 14
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)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

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

18. SUPPLEMENTARY NOTES

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

Goodness'of'fit; Graphical methods; Kolmogorov-Smirnov statistics; Percent'Percent plot; Quantile-Quantile plot; Variance stabilizing transformation.

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

The stabilized probability plot or SP plot is introduced. An attractive feature of the SP plot that enhances its interpretability is that the variances of the plotted points are approximately equal. This prompts the definition of a new and powerful goodness-of-fit statistic D_{sp} which, analogous to the standard Kolmogorov-Smirnov statistic D, is defined to be the maximum deviation of the plotted points from their theoretical values. Using either D or D_{sp} it is shown how to construct acceptance retions for QQ, PP, and SP plots. Acceptance regions can help remove much of the subjectivity from the interpretation of these probability plots.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)