Goodness-of-fit tests for discrete and censored data, based on the empirical distribution function.

by

ANTHONY PETTITT, B.Sc., M.Sc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, October 1973.

ABSTRACT

In this thesis two general problems concerning goodness-of-fit statistics based on the empirical distribution are considered. The first concerns the problem of adapting Kolmogorov-Smirnov type statistics to test for discrete populations. The significance points of the statistics are given and various power comparisons made.

The second problem concerns testing for goodness-of-fit with censored data using the Cramér-von Mises type statistics. The small and large sample distributions are given and the tests are modified so that they can be used to test for the normal and the exponential distributions. The asymptotic theory is developed. Percentage points for the statistics are given and various small sample and large sample power studies are made, for the various cases.

SUMMARY

ACKNOWLEDGEMENTS

§1. INTRODUCTION

Part I Kolmogorov - Smirnov Type statistics for discrete data.

§2. SMALL SAMPLE DISTRIBUTION OF THE STATISTICS

2.1 Introduction

2.2 A result of Steck

2.3 The distribution of D_n, D_n^+ and V_n for discrete populations using Steck's result.

2.4 The distribution of V_n for certain values of n and q.

2.5 The tables of D_n, D_n^+ and V_n.

2.6 Power Study

2.7 Results of power study

§3. ASYMPTOTIC DISTRIBUTION OF THE STATISTICS

3.1 Introduction

3.2 Asymptotic distribution of the statistics for fixed q.

§4. A RESULT CONCERNING THE ASYMPTOTIC POWER OF THE STATISTICS FOR THE SIGNIFICANCE LEVEL TENDING TO 0

4.1 Introduction

4.2 A different formulation of the statistics

4.3 Some extensions of theorems of Hoeffding

4.4 Some examples of the previous results

§5. A COMPARISON OF THE ASYMPTOTIC POWERS

5.1 Introduction

5.2 Previous results

5.3 Results

Part II Goodness-of-fit for censored data using Cramér-von Mises type statistics

§6. THE W^2 STATISTIC FOR TESTING GOODNESS-OF-FIT WITH CENSORED DATA WITH A SIMPLE HYPOTHESIS

6.1 Introduction

6.2 Small sample properties of ${}_pW_n^2$, ${}_rW_n^2$ and ${}_{q,p}W_n^2$

6.3 Exact lower-tail distribution of the statistic ${}_pW_n^2$

6.4 Theory of the asymptotic distribution of the statistics ${}_rW_n^2$, ${}_pW_n^2$ and ${}_{q,p}W_n^2$

6.5 Moments and percentage points of ${}_pW^2$ and ${}_{q,p}W^2$

6.6 Asymptotic power of ${}_pW^2$ and its components against shifts in location and scale

§7. THE W^2 STATISTIC WITH CENSORED DATA FOR A COMPOSITE HYPOTHESIS

7.1 Introduction

7.2 The asymptotic form of the empirical process

7.3 The weighted χ^2 representation of ${}_p\hat{W}^2$

7.4 Testing for normality with ${}_pW^2$

7.5 Testing for the exponential distribution in life testing with ${}_pW^2$

§8. THE ANDERSON-DARLING STATISTIC FOR CENSORED DATA

8.1 Introduction

8.2 Exact moments of ${}_pA_n^2$, ${}_rA_n^2$ and ${}_pA^2$

8.3 Asymptotic distribution of ${}_rA_n^2$, ${}_pA_n^2$

8.4 Significance points of ${}_pA^2$

8.5 Asymptotic power of ${}_pA^2$ against shifts in location and scale

8.6 Testing for normality using ${}_pA^2$

Part III

§9. PRACTICAL APPLICATION OF THE TESTS GIVEN IN PARTS I AND II

9.1 Introduction

9.2 The statistics $D_{n,q}$, $D^{+}_{n,q}$ and $V_{n,q}$ introduced in Part I

9.3 The statistics ${}_{p}W^2$ and ${}_{p}A^2$ introduced in Part II

§10. RELATED AND FURTHER WORK

10.1 Introduction

10.2 The statistic U^2

10.3 Goodness-of-fit statistics for the 2-sample problem and the bivariate sample with censored data

10.4 An alternative method of testing for normality with censored data

10.5 Conclusion

REFERENCES

APPENDIX

ACKNOWLEDGMENTS

I would like to thank Professor M.A. Stephens for his encouragement, guidance and many helpful suggestions throughout the preparation of this work, and without whose enthusiasm this work would never have been started.

I am also grateful to Professor C.W.J. Granger for his assistance in the submission of this thesis; also to the invitation given by Professor E. Parzen to attend the conference 'Uses of the empirical distribution function' at SUNY at Buffalo, New York in September, 1971 and the National Science Foundation for their financial support.

Many thanks are also due to Joan Inns who typed this thesis.

§1. Introduction

In much statistical inference and model building, there is a need to test the validity of assumptions made about the underlying populations, from which observations have been drawn. In its simplest form, this problem can be stated as follows: a random sample of n observations is taken from a population with cumulative distribution function $F(x)$, and it is wished to test the null hypothesis

$$H_o: F(x) \equiv F_o(x)$$

where $F_o(x)$ is a completely specified distribution function. The classical solution to this problem is the χ^2 test of Karl Pearson, described in detail in Kendall and Stuart (1961, p. 419). The criticisms of this test are that its small sample distribution is not tabulated and so is applicable only to large samples, and that if $F(x)$ is assumed not to be discrete then there is a necessary loss of power in its use, due to the fact that the observations have to be grouped. These two criticisms have been overcome by the introduction of the D_n statistic by Kolmogorov (1933). The statistic is based on the empirical distribution function $F_n(x)$, where $F_n(x)$ is defined for a random sample $x_1, x_2, \ldots, x_n$ of size n by

$$F_n(x) = \text{Proportion of observations } (x_1, \ldots, x_n) \leq x$$

The statistic D_n is given by

$$D_n = \sup_x |F_n(x) - F_o(x)|.$$

This statistic is shown to be distribution-free for continuous $F_o(x)$ and Kolmogorov (1933) obtained its asymptotic distribution:

$$\lim_{n\to\infty} P(\sqrt{n}D_n < x) = 1 + 2 \sum_{r=1}^{\infty} (-1)^r \exp(-2r^2x^2) \simeq 1 - 2\exp(-2r^2x^2)$$

He also gave recurrence relations for the probability for finite n, which were used by Birnbaum (1952) to tabulate $P(D_n < k/n)$ for $n = 1(1)\ 100$ and $k = 1(1)\ 15$. Previously Massey (1950) had given these probabilities but for selected values of n and k.

Since D_n is a two-sided statistic, one-sided versions have been considered by Smirnov (1939) and Wald and Wolfowitz (1939),

$$D_n^+ = \sup_x \{F_n(x) - F_o(x)\}$$

$$D_n^- = \sup_x \{F_o(x) - F_n(x)\}$$

Smirnov (1939) gives the asymptotic distribution of these statistics (they are distributed exactly the same):

$$\lim_{n\to\infty} P_r(\sqrt{n}\,D_n^+ \leq x) = 1 - \exp(-2x^2)$$

So $4n(D_n^+)^2$ has the χ^2 distribution with two degree of freedom asymptotically. Birnbaum and Tingey (1951) give the asymptotic points. D_n is known as the Kolmogorov statistic and D_n^+ and D_n^- are known as the Kolmogorov-Smirnov statistics.

Other statistics based on the empirical distribution function to be proposed have been of the form

$$w_n^2 = \int_{-\infty}^{\infty} \{F_n(x) - F_o(x)\}^2\ \psi\{F_o(x)\}\,dF_o(x)$$

known as Cramér-von Mises (CVM) type statistics, which are distribution-free for continuous $F_o(x)$. The statistic with $\psi(x) \equiv 1$ was investigated by Smirnov (1936):

$$W_n^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F_o(x)\}^2\ dF_o(x)$$

who found the asymptotic characteristic function of W_n^2:

$$\phi(t) = \lim_{n\to\infty} E\left[\exp(itW^2)\right] = \left[\frac{(2it)^{\frac{1}{2}}}{\sin\{(2it)^{\frac{1}{2}}\}}\right]^{\frac{1}{2}}$$

This was inverted by Anderson and Darling (1952) to give the limiting distribution of W_n^2 and selected significance points were given. Marshall (1958) found the small sample distribution of W_n^2 for n upto 3 and showed the asymptotic points of Anderson and Darling (1952) to be adequate for n as small as 3. Approximations to the small sample distribution have been made by Pearson and Stephens (1962) and Tiku (1965). Stephens (1970) shows that the modified statistic

$$W_n^{2*} = (W_n^2 - 0.4/n + 0.6/n^2)(1.0 + 1.0/n)$$

can be used with asymptotic significance points with very little error in the actual critical level for small n.

Anderson and Darling (1952) introduced the CVM type statistic with $\psi(x) = \{x(1-x)\}^{-1}$:

$$A_n^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F_o(x)\}^2 [F_o(x)\{1-F_o(x)\}]^{-1} dF_o(x)$$

They found its characteristic function and gave significance points, Anderson and Darling (1954). These were shown to be fairly accurate even for n = 1 by Marshall (1958).

The K-S type and CVM type statistics were modified for tests of goodness-of-fit on the circle by Kuiper (1960) and Watson (1961). The tests have to be invariant of the origin chosen on the circle. Kuiper introduced the statistic

$$V_n = D_n^+ + D_n^-$$

and Watson the statistic

$$U_n^2 = \int_{-\infty}^{\infty} [F_n(x) - F_o(x) - \int_{-\infty}^{\infty} \{F_n(y) - F_o(y)\}\, dF_o(y)]^2\, dF_o(x)$$

Simple computing formulae and simplified significance points for the statistics W_n^2 and U_n^2 are given in Stephens (1970).

The main criticisms of the statistics as they stand, is that they are dependent upon the distribution of the null hypothesis being a continuous population and the null hypothesis being simple. Darling (1955) tackled the latter problem where he gave the limiting distributions of the statistics when the null hypothesis is of the form

$$H_o : F(x) \equiv F_o(x,\theta)$$

where $F_o(x,\theta)$ is a c.d.f. of specified functional form with θ an unknown parameter. Kac, Kiefer and Wolfowitz (1955) investigated the problem of testing for normality, when μ and σ^2 are to be estimated, using the statistic W_n^2. The results of both these papers have been extended by Stephens (1973) and significance points given for large sample tests when testing for exponentiality and normality using the statistics W_n^2, U_n^2 and A_n^2. The small sample distributions of the K-S and CVM type statistics have been found by Monte Carlo methods by various people and the results summarized in Stephens (1970).

The problem of applying the K-S and CVM type goodness-of-fit statistics to discrete and grouped data is tackled by Illyasenko (1952) in the asymptotic case for D_n leaving the c.d.f. as an integral involving the distribution of the order statistics from a k multivariate normal distribution. Noether (1963) in a short note states the significance points of the D_n statistic will be conservative when used with discrete data. The asymptotic distributions of the K-S type statistics remain intractable, but are found by Monte Carlo methods. Also the small sample distributions of the statistics are found in this thesis and extensively tabulated. The power of the statistics is investigated both for small samples and large samples, the former by means of Monte Carlo methods and the latter in two situations: the first where the significance level is kept constant and the second the significance level is allowed to go to 0 as the sample size increases.

In the second part of the thesis we look at the problem of adapting the CVM type statistics to use for goodness-of-fit tests with censored data. The statistics considered are of the type

$$ {}_pS_n^2 = n \int_{-\infty}^{F_o^{-1}} \{F_n(x)-F_o(x)\}^2 \psi(F_o(x))\, dF_o(x) $$

$$ {}_pS_r^2 = n \int_{-\infty}^{x_{(r)}} \{F_n(x)-F_o(x)\}^2 \psi(F_o(x))\, dF_o(x) $$

We take $\psi(x) \equiv 1$ and $\psi(x) = (x(1-x))^{-1}$, the former giving the W^2 statistics ${}_pW^2$ and ${}_rW_n^2$, and the latter the A^2 statistics, ${}_pA_n^2$ and ${}_rA_n^2$. The small sample moments of the W^2 statistics are found and approximate percentage given by fitting a generalized χ^2 distribution to the distribution of W^2 by equating the first three moments.

The asymptotic theory of the statistics is given and a good approximation to the distribution of W^2 and A^2 found. The asymptotic theory is also given for W^2 when used with doublely censored observations. Asymptotic percentage points are given when the censoring is symmetric. The asymptotic powers of the tests based on ${}_pW^2$ and ${}_pA^2$ and statistics derived from them are studied for scale shift of the exponential distribution, and location and variance shift of the normal distribution. The statistics are modified so that they can be used to test a composite hypothesis and the asymptotic theory given. In particular the asymptotic percentage points of testing for normality and the exponential distribution with censored data are given for the W^2 and A^2 statistics. Results of Monte Carlo power studies are given to compare the power of the small sample tests.

Finally, the statistic U^2 is modified to test for censored data and the theory developed.

Part I - Kolmogorov Smirnov Type statistics for discrete data

§2. Small sample distribution of the statistics

2.1 Introduction

In this section, the exact distributions of D_n^+, D_n and V_n are found, and the distributions are tabulated for the discrete uniform population (which can arise from grouping n observations into q mutually exclusive groups which are equally likely) on q points for $n \leq 30$ and $q \leq 12$. The tables are given in Appendix 1. The results of a power study are also given comparing these statistics with the χ^2 statistic and Davids Empty Cell test, see David (1950), showing these latter statistics to be less powerful for certain alternatives.

2.2. A result of Steck

The method used is to apply a result of Steck (1971). He shows that if $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are order statistics from a sample of n independent uniform random variables then

$$\Pr(u_i \leq x_{(i)} \leq v_i,\ i = 1,\ldots,n) = \det R, \text{ where}$$

$$R_{ij} = \binom{j}{j-i+1}(v_i - u_j)_+^{j-i+1} \qquad j-i+1 \geq 0$$

$$= 0 \qquad j-1+1 < 0$$

and $(x)_+ = \max(0,x)$. R is a matrix of Hessenberg form.

This result can be stated in terms of the e.d.f. $F_n(x)$

$$\Pr(g(x) \leq F_n(x) \leq h(x)) = \det(R)$$

where g and h are non negative functions on (0,1), with g continuous to the left and h continuous to the right, and u_i, v_i are defined by $u_i = h^{-1}(i/m)$ $v_i = g^{-1}((i-1)/n)$ $i = 1,\ldots,n$.

2.3. The distribution of D_n, D_n^+ and V_n for discrete populations, using Steck's result

Consider a random variable y to be distributed discretely with $q(<\infty)$ mass points $b_1, < \dots < b_q$. Define $z = G(y)$, where G is the c.d.f. of y, then z is a discrete random variable with mass points $0 < G(b_1) < G(b_2) < \dots G(b_{q-1}) < G(b_q) = 1$. Let $a_i = G(b_i)$ $i = 1,\dots,q$, then $Pr(z \leq a_i) = a_i$ $i = 1,\dots,q$. Consider now the e.d.f. $G_n(y)$ of a random sample $y_1,\dots,y_n$ from the discrete population with c.d.f. $G(y)$, $G_n(y) = \dfrac{\text{Number of } y_i \leq y}{n}$. $G_n(y)$ will have discontinuities only at the points $b_1,\dots,b_q$, where it may have steps of height equal to multiples of $1/n$. Define

$$H_n(z) = \frac{\text{number of } z_i \leq z}{n} = \frac{\text{number of } y_i \leq G^{-1}(z)}{n}$$

where $G^{-1}(z) = \inf\{y : G(y) = z\}$ and $z_i = G(y_i)$ $i = 1,\dots,n$.

Then $H_n(z)$ will be a step function having steps multiples of $1/n$ at the points $a_1,\dots,a_q$. Now $G_n(y) - G(y) = H_n(y) - H(y)$ so

$$\sup_y |G_n(y) - G(y)| = \sup_z |H_n(z) - H(z)|,$$

and it is wished to find the distribution of

$\sup|H_n(z) - H(z)| = \max_i |H_n(a_i) - H(a_i)|$. Now at the points $a_1,\dots,a_q$ $H_n(z)$ behaves like the e.d.f. of a sample of n independent uniform random variables, i.e. $n H_n(z)$ is binomially distributed $Bi(a_i,n)$ at each a_i, $i = 1,\dots,q$.

So $Pr(\max_i |H_n(a_i) - H(a_i)| \leq d)$

$= Pr(\max_i |F_n(a_i) - a_i| \leq d)$ where $F_n(x)$ is the e.d.f. of n independent uniform random variables $x_1,\dots,x_n$. Steck's result can now be used to find $Pr(\max_i|F_n(a_i) - a_i| \leq d)$ and similarly for D_n^+, $Pr(\max_i(F_n(a_i) - a_i) \leq d)$.

Define $h(x) = \min(1, a_i+d)$, $a_{i-1} \leq x < a_i \quad i = 1,\ldots,q$

where $a_o = 0$ and

$$g(x) = \max(0, a_i-d), \quad a_i < x \leq a_{i+1} \quad i = 0,\ldots,q-1$$

then $Pr(\max_i|F_n(a_i) - a_i| \leq d) = Pr(g(x) \leq F_n(x) \leq h(x))$

$$= \det R$$

where

$$R_{ij} = \binom{j}{j-i+1}(v_i-u_j)^{j-i+1} \qquad j-i+1\geq 0$$

$$= 0 \qquad j-i+1<0$$

where v_i and u_j are found by taking inverses of g and h.

Put $n(a_i+d) = \beta_i$ and $[\beta_i] = \gamma_i$

$n(a_i-d) = \alpha_i$ and $[\alpha_i] = \delta_i$

Also if $\gamma_i > n$ then $\gamma_i = n$, and if $\delta_i < 0$ then $\delta_i = 0$.

Define $u_i = 0$, $i = 1,\ldots,\gamma_1$

a_1, $i = \gamma_1+1,\ldots,\gamma_2$

a_2,

.

.

.

a_q, $i = \gamma_{q+1},\ldots,n$

$v_i = a_1$, $i = 1,\ldots,\delta_1$

a_2, $i = \delta_1+1,\ldots,\delta_2$

.

.

.

a_q, $i = \delta_{q-1}+1,\ldots,\delta_k$

1, $i = \delta_q+1,\ldots,n$

2.4. Distribution of V_n for certain values of n and q

If the values of n are restricted so they are multiples of q, then the only possible values of D_n^+ are D_n are multiples of 1/n. We can then find, for example, $Pr(nD_n^+ \leq k_1, nD_n^- \leq k_2)$ by the above method by defining $\beta_i = n(a_i+k_1)$ and $\alpha_i = n(a_i-k_2)$. Now the statistic V_n is defined to be $V_n = D_n^+ + D_n^-$.

So the distribution of V_n is just the convolution of D_n^+ and D_n^-. Thus

$$Pr(nV_n \leq k) = Pr(nD_n^+ + nD_n^- \leq k)$$

Now

$$Pr(nD_n^+ + nD_n^- = k) = \sum_{i=o}^{k} Pr(nD_n^+ = i,\ nD_n^- = k-i)$$

So the p.d.f. of V_n is obtained for this special case, since

$$\begin{aligned} Pr(nD_n^+ = k_1, nD_n^- = k_2) = & Pr(nD_n^+ \leq k_1,\ nD_n^- \leq k_2) \\ & - Pr(nD_n^+ \leq k_1 - 1,\ nD_n^- \leq k_2) \\ & - Pr(nD_n^+ \leq k_1,\ nD_n^- \leq k_2 - 1) \end{aligned}$$

2.5. The tables of D_n, D_n^+ and V_n

For the statistics D_n, D_n^+, V_n the

$$Pr(nS_n \leq k)$$

where S_n is D_n, D_n^+ or V_n is given for values q up to 12 and n up to 30 so that q divides n. Note that for n = q the distribution of D_n and D_n^+ is given by the continuous distribution free results. If D_n^c denotes the statistic for continuous populations and $D_{n,q}$ the one for discrete or grouped populations with q mass points, then

$$Pr(nD_n^c \leq k) = Pr(nD_n^c < k) = Pr(nD_{n,n} < k) = Pr(nD_{n,n} \leq k-1)$$

This follows directly from Massey's argument (1950).

It will be seen from the tables, see Appendix, that grouping has a large effect on the distribution of the statistics.

To note the effect of grouping on D, note that

$Pr(12.D_{12,12} \geq 5) = .02115$, $Pr(12.D_{12,6} \geq 5) = .01422$

$Pr(12.D_{12,4} \geq 5) = .01115$

$Pr(12\ D_{12,12} \geq 4) = .10874$ $Pr(12\ D_{12,6} \geq 4) = .09064$

and $Pr(12\ D_{12,4} \geq 4) = .05974$

$Pr(20\ D_{20,20} \geq 6) = .04307$ $Pr(20\ D_{20,10} \geq 6) = .03273$

$Pr(20\ D_{20,5} \geq 6) = .02203$

$Pr(20\ D_{20,20} \geq 5) = .13763$, $Pr(20\ D_{20,10} \geq 5) = .10910$

$Pr(20\ D_{20,5} \geq 5) = .07617$

Also from Stephens (1965), we have $Pr(V_n \leq \frac{k}{n})$ and the percentage points of the statistic V_n. We see that the effect of grouping is much stronger.

For n = 12, the 10% and 1% significance points are given:

$Pr(12\ D_{12} \geq 5.24) = .10$ and $Pr(12\ V_{12} \geq 6.48) = .1$

we have $Pr(12\ V_{12,12} \geq 6) = .0113$, $Pr(12\ V_{12,12} \geq 7) = .0010$

$Pr(12\ V_{12,6} \geq 6) = .0047$, $Pr(12\ V_{12,6} \geq 7) = .0003$

$Pr(12\ V_{12,12} \geq 5) = .0738$

$Pr(12\ V_{12,6} \geq 5) = .0403$.

The values of det R were calculated using a single precision program written in FORTRAN, using a method derived from Wilkinson (1965). The tables of D_n, D_n^+ and V_n can be taken to be accurate to 5 significant figures for $n \leq 25$, but become inaccurate for larger values of n especially in the upper tail.

2.6. Power Study

A power study was undertaken to compare the power of the statistics D_n, D_n^+ and V_n, when used correctly with discrete data, with the power of the chi-squared statistic and Davids Empty Cell test. The null hypothesis tested, in terms of the multinomial distribution, was that the q sample points were equally likely (or in terms of grouped data, the q cells were equally likely).

$$H_o : p_1 = p_2 = \cdots = p_q = \frac{1}{q} \tag{2.1}$$

The alternatives considered were

1. $\sum_{j=1}^{i} p_j = \frac{(i)^{\delta}}{q}$ $\quad i = 1,\ldots,q$, abbreviated as $A_1(\delta)$

2. $\sum_{j=1}^{i} p_j = \frac{1}{2}\left(\frac{2i}{q}\right)^{\delta}$ $\quad i \leq \frac{q}{2}$

   $\sum_{j=1}^{i} p_j = 1-\frac{1}{2}\left(\frac{2(q-i)}{q}\right)^{\delta}$ $\quad i > \frac{q}{2}$, $A_2(\delta)$

3. $p_i = \frac{1}{q} - \delta$, $\quad i \leq \frac{q}{2}$

   $= \frac{1}{q} + \delta$, $\quad i > \frac{q}{2}$

so $\sum_{j=1}^{i} p_j = \frac{i}{q} - i\delta$, $\quad i \leq \frac{q}{2}$

$= \frac{i}{q} + (1-q)\delta$, $\quad i > \frac{q}{2}$, $A_3(\delta)$

It is to be noticed that all the statistics are discrete. The distributions of the first three statistics are tabulated here and Davids Empty Cell tabulated in David (1950). For the chi-squared statistic, use was made of Kempthorne (1967), where the exact moments of the statistic X^2 are given for the equi-probable cell case. For given n and q, a Pearson curve was fitted to the first four moments of the statistic and percentage points obtained using interpolation of the tables of Johnson, Nixon, Amos and Pearson (1965). The actual chi-squared significant point used was the greatest even integer

$$z_\alpha \leq \frac{n}{q}(P_\alpha + 2n - \frac{n}{q})$$

where P_α is the Pearson curve approximation significant point, since

$$X^2 = \frac{\Sigma(x_i - np_i)^2}{np_i} = \frac{q}{n}\Sigma x_i^2 - 2n + \frac{n}{q}$$

and n and q were chosen so that Σx_i^2 takes only even values. The actual significance of the test is not known, but is at least α% in terms of the approximation, which is reasonably accurate, Kempthorne (1967).

The other statistics are discrete, so no exact α% value could be found. This problem was overcome by using the randomization technique. If S denotes the statistic in question and it takes on integer values 0,1,2,..., then suppose s_u is the smallest integer such that

$$Pr(S \geq s_u) \leq \alpha,$$

and s_L the largest integer such that

$$Pr(S \geq s_L) \geq \alpha$$

and

$$Pr(S \geq s_u) = \alpha_u, \quad Pr(S \geq s_L) = \alpha_L.$$

Then an α% test is defined for the statistic S, if for a given value s of S we reject if $s \geq s_u$ and accept if $s \leq s_L$. Also, reject or accept as U is $\leq \alpha$ or $> \alpha$, where U is a random variable distributed uniformly on (α_u, α_L). This can be applied to nD_n, nD_n^+ and nV_n since they take only integer values.

2.7. Results of power study

The results of the power study are given in Tables 2.1 and 2.2. For the alternatives which have been considered here, which are systematic alternatives to the null, the Kolmogorov type statistics do well compared to Davids Empty Cell test (E) and the Chi-squared test (CH). For the alternatives A_1 and A_3, which could be considered 'trend' alternatives, (since the p_i satisfy either the ordering $p_1 \leq p_2 \leq \cdots \leq p_q$ or $p_1 \geq p_2 \geq \cdots \geq p_q$) the statistics D and $D^{\pm}$ ($D^{\pm}$ refers to either D^+, for alternatives of the second ordering, or D^- for alternatives of the first ordering) are the most powerful. Thus for trend alternatives it seems advantageous to use the statistic D, D^+ and D^- rather than CH or E.

For 'peaked' alternatives which we define to be $p_1 \leq p_2 \leq \cdots \leq p_r \geq p_{r+1} \cdots \geq \cdots \geq p_q$ for some $1 < r < q$ or conversely 'troughed' alternatives $p_1 \geq p_2 \geq \cdots \geq p_r \leq p_{r+1} \leq \cdots \leq p_r$ for some $1 < r < q$, of which A_2 is an example, V should be powerful and this is shown in the results for A_2. V was originally used for goodness-of-fit on the circle, see Kuiper (1960), and so again V can be used for discrete or grouped data on the circle.

For the alternative A_3, which is of the kind $p_1 = p_2 = \cdots = p_r < p_{r+1} = \cdots = p_q$ the statistics D, D^+, V are much more powerful than CH and E. A result concerning the asymptotic power of D in §3 suggests that the power of D_n is optimal for this type of 'dichotanised' alternative.

Thus for small sample sizes, the D, D^+, V statistics are of importance when the alternatives of interest are systematic ones of the kind considered. D and D^+ are especially powerful against trend alternative. However it is known that both E and CH are invariant to the ordering of the p_i's, that is the powers of the tests will be the same for the alternative

$$H_1 : p_1 = p_1^1 , \; p_2 = p_2^1 , \ldots , p_q = p_q^1$$

as for $H_1^\alpha : p_1 = p_{\alpha_1}^1 , \; p_2 = p_{\alpha_2}^1 , \ldots , p_q = p_{\alpha_q}^1$

where $(\alpha_1 , \ldots , \alpha_q)$ is a permutation of $(1,2,\ldots,q)$. For certain alternatives, the $(\alpha_1 , \ldots , \alpha_q)$ permutation was chosen so that

$$\sup_i \left| \sum_{j=1}^{i} p_{\alpha_i}^1 - \frac{i}{q} \right|$$

was a minimum, thus $(\alpha_1 , \ldots , \alpha_q)$ was selected so that the power of D, D^+ was at a minimum, taken over $(\alpha_1 , \ldots , \alpha_q)$. For these alternatives (see Table 2.3, called A_i^* (i = 1,2,3)), the powers of the statistics D, D^+, V are reduced, but usually not by more than half, and often less. Of course the powers of the statistics E and CH remain the same, subject to sampling error. However for these alternatives CH is most powerful.

For the alternative A_3^* D, V lose most power. However this is not surprising since of all the alternatives considered the statistics D, D^+ are comparatively the best for the alternative A_3.

Alternative	D	D^+,D^-	V	E	CH
$A_1(.9)$	64	89^+	61	58	71
$A_1(.8)$	94	145	72	62	93
$A_1(.5)$	433	570	349	199	399
$A_1(.25)$	910	960	858	555	923
$A_1(1.25)$	65	118	73	65	82
$A_1(1.50)$	150	259	95	88	106
$A_1(2.00)$	377	562	281	194	260
$A_1(3.00)$	806	929	707	484	655
$A_2(.8)$	74	64	68	58	82
$A_2(.5)$	178	155	296	165	263
$A_2(.25)$	449	336	757	451	706
$A_2(.1)$	698	521	978	822	968
$A_2(1.5)$	48	44	131	90	119
$A_2(2.0)$	54	79	296	179	253
$A_2(3.0)$	160	180	700	466	600
$A_2(3.5)$	211	255	815	579	744
$A_3(.033)$	77	143^-	76	64	77
$A_3(.066)$	221	352	172	119	170
$A_3(.1)$	498	646	430	260	339
$A_1^*(.5)(1.6.2.5.3.4)$	294	432^+	249	175	379
$A_1^*(2.0)(6,1,5,2,4,3)$	108	167^+	129	217	278
$A_2^*(.5)(3,1,2,4,6,5)$	85	85^-	120	135	255
$A_2^*(3.0)(1,3,2,6,5,4)$	147	177^-	236	454	629
$A_3^*(.1)(1,4,2,5,3,6)$	106	147^-	124	255	346

Table 2.1

Power of Statistics × 1000, ie number of rejections in 1000 replications. +(-) denotes $D^+(D^-)$ statistic. Level of test 5%

$n = 12,\ q = 6$

Alternative	D_n	D_n^+, D_n^-	V_n	E	CH
A_1 (.9)	82	125^+	78	59	70
A_1 (.8)	146	214	105	74	97
A_1 (.5)	684	759	509	314	551
A_1 (.25)	994	998	981	893	994
A_1 (1.25)	98	160^-	76	58	75
A_1 (1.50)	266	371	161	94	117
A_1 (2.00)	649	774	475	234	266
A_1 (3.00)	983	992	922	633	627
A_2 (.8)	94	90	122	67	111
A_2 (.5)	270	237	537	299	456
A_2 (.25)	793	591	984	882	974
A_2 (.1)	997	866	1000	1000	1000
A_2 (1.5)	56	80^+	225	106	116
A_2 (2.0)	112	151	582	249	276
A_2 (3.0)	379	417	969	637	652
A_2 (3.5)	545	514	991	786	792
A_3 (.01)	105	157^-	88	73	83
A_3 (.o2)	333	444	247	97	143
A_3 (.03)	709	796	606	239	275

Table 2.2

Power of statistics, × 1000, ie.number of rejections in 1000 replications. +(−) denotes $D^+(D^-)$ statistic. Level of test 5%

$n = 20$, $q = 20$

§3. Asymptotic distribution of the statistics

3.1 Introduction

In this section, the asymptotic distributions of the statistics are given, but it is found they cannot be given explicitly. A Monte Carlo investigation is made to find them.

3.2 Asymptotic distribution of the Statistics for fixed q.

If $n \to \infty$ then the random variables $\sqrt{n}\{H_n(a_i)-H(a_i)\}$ $(i = 1,\ldots,q-1)$ by the Multi-variate central limit theorem, are distributed as $\underline{Z} = (Z_1,\ldots,Z_{q-1})$ where the $\underline{Z}$ is multivariate normal with $E(\underline{Z}) = \underline{0}$ and $\mathrm{cov}(\underline{Z}) = \underline{V}$ where $(V)_{ij} = a_i(1-a_j)$ $i \leq j$. Then

$$\begin{aligned}
\lim_{n\to\infty} \Pr(\sqrt{n}\, D^+_{n,q} \leq \lambda) &= \Pr(\max_i(Z_i,0) \leq \lambda) \\
\lim_{n\to\infty} \Pr(\sqrt{n}\, D_{n,q} \leq \lambda) &= \Pr(\max_i(Z_i,-Z_i) \leq \lambda) \qquad (3.1) \\
\lim_{n\to\infty} \Pr(\sqrt{n}\, V_{n,q} \leq \lambda) &= \Pr(\max_i(Z_i,0) + \max_i(Z_i,0) \leq \lambda)
\end{aligned}$$

Hence the asymptotic distributions of the statistics involve the order statistics of the multivariate normal. With such a covariance matrix the problem appears unsolvable, so the asymptotic distributions have been simulated. For each value of q, the $a_i = \frac{i}{q}$ $i = 1,\ldots,q-1$ and 10,000 independent values of the vector random variable $\underline{Z}$ were generated using the N.A.G. pseudo-random number generator, and the statistics on the right hand side of (3.1) calculated. The results are given in Tables 3.1-3.3 for D^+_n, D_n, V_n respectively.

It is seen from Table (3.1) that the distributions of the statistics do not approach the distributions of the statistics for the continuous case ($q = \infty$) particularly fast, for $q \rightarrow \infty$. Nor in fact do the distributions of the statistics for fixed q approach the asymptotic distributions for fixed q particularly fast. For example, from Table 3.1

$$\lim \Pr(\sqrt{n}\, D^{+}_{n,5} \leq .821) \simeq .90.$$

However from Appendix 1 we see that

$$\Pr(D^{+}_{30,5} \leq \tfrac{4}{30}) = .906$$

i.e. $$\Pr(\sqrt{30}\, D^{+}_{30,5} \leq .730) = .906$$

so the 10% upper significance points are .821 and .730 for the asymptotic statistics $\sqrt{n}\, D^{+}_{n,q}$ and $D^{+}_{30,q}$ respectively with $q = 5$. The approach of the asymptotic distributions of the other statistics is no better for $q = 5$ and other values of q.

$\sqrt{n}\, D_n^+, q$

q	5	6	7	8	9	10	11	∞
15%	.718	.734	.754	.769	.777	.784	.798	.973
10%	.821	.837	.855	.873	.882	.884	.898	1.073
5%	.975	.989	1.008	1.022	1.040	1.047	1.051	1.224
2½%	1.120	1.129	1.148	1.156	1.176	1.178	1.179	1.358
1%	1.281	1.290	1.297	1.319	1.327	1.342	1.335	1.518

$\sqrt{n}\, D_n, q$

q	5	6	7	8	9	10	11	∞
15%	.885	.906	.917	.938	.941	.947	.957	1.138
10%	.971	.990	1.004	1.020	1.030	1.042	1.044	1.224
5%	1.112	1.131	1.140	1.157	1.166	1.179	1.171	1.358
2½%	1.244	1.254	1.264	1.289	1.287	1.311	1.296	1.480
1%	1.381	1.403	1.430	1.418	1.428	1.455	1.444	1.628

$\sqrt{n}\, V_n, q$

q	5	6	7	8	9	10	11	
15%	1.023	1.066	1.095	1.123	1.151	1.167	1.182	1.537
10%	1.111	1.157	1.180	1.214	1.239	1.251	1.261	1.620
5%	1.250	1.292	1.310	1.341	1.360	1.381	1.394	1.747
2½%	1.371	1.399	1.434	1.456	1.471	1.504	1.511	1.862
1%	1.510	1.560	1.577	1.598	1.644	1.643	1.650	2.001

Table 3.1

Asymptotic significance points of the statistics $\sqrt{n}\, D_n^+, q$, $\sqrt{n}\, D_n, q$, $\sqrt{n}\, V_n, q$, found by simulation using 10,000 random samples.

§4. A result concerning the asymptotic power of the statistics for the significance level tending to 0.

4.1 Introduction

This section is based on an extension of the work of Hoeffding (1965). A test of the simple multinomial hypothesis

$$H_o : p_1 = p_{o1}, p_2 = p_{o2}, \ldots, p_q = p_{oq}$$

based on n observations, when q is fixed, is considered. Hoeffding considers letting the significance level α_n tend to 0 as $n \longrightarrow \infty$, and considers the power of tests for testing H_o when the alternative is not very near to the null. He shows that the likelihood ratio test (LR test) (actually log of the likelihood ratio)

$$L = \sum_{i=1}^{q} x_i \log(x_i/p_{oi})$$

where x_i is the number of observations falling in the i^{th} category, is more powerful than the X^2 test where

$$X^2 = \sum_{i=1}^{q} \left(\frac{x_i - np_{oi}}{np_{oi}}\right)^2$$

except at a particular set of alternatives. We find alternatives where the power of the Kolmogorov-Smirnov type statistics are equivalent to the LR test, and so much more powerful than the X^2 test.

4.2 A different formulation of the statistics

Consider the statistics

$$\mathfrak{D}(x,p) = \max_{M \in \mathfrak{M}} \sum_{i \in M} (x_i - p_i)$$

where $\mathfrak{M}$ is a collection of non-empty proper subsets of the set $\{1,\ldots,q\}$. The statistics D_n, D_n^+ and V_n can be defined by particular choice of $\mathfrak{M}$.

(i) With $\mathfrak{M}_i = \{(1,\ldots,j),\ 1 \le j \le q\}$ then $\mathfrak{D}(x,p)$ gives the statistic D_n^+:

$$\max_{M\in\mathfrak{M}} \sum_{i+M} (x_i - p_i) = \max_j \sum_{i=1}^{j} (x_i - p_i)$$

$$= \max_j n(G_n(a_j) - a_j)$$

where $G_n(a_j) = n \sum_{i=1}^{j} x_i$

and $a_j = \sum_{i=1}^{j} p_i$.

(ii) With $\mathfrak{M}_{ii} = \{(j,\ldots,q),\ 1 \le j \le q\}$ this gives the statistic D_n^-.

(iii) If $\mathfrak{M}_i$ is $\mathfrak{M}$ defined in (i) and $\mathfrak{M}_{ii}$ is $\mathfrak{M}$ defined in (ii) then $\mathfrak{M}_{iii} = \mathfrak{M}_i \cup \mathfrak{M}_{ii}$ will give $\mathfrak{D}$ corresponding to D_n.

(iv) With $\mathfrak{M}_{iv} = \{(1,\ldots,i,\ i',\ldots,j',\ j,\ldots,q)$ for $1 \le i \le i' \le j' \le j \le q\}$

$\mathfrak{D}$ will correspond to V_n. It will be recalled that the statistic V_n is used for goodness-of-fit on the circle. It can be shown that V_n is the value of max D_n, when D_n is calculated on the circle, the max being with respect to the choice of origin, Stephens (1965). If the data is discrete and when the circle is transferred to (0,1) with $a_o, a_1, \ldots, a_q$ ($a_o = 0$, $a_q = 1$) as the mass points, choosing one of the mass points on the circle as origin. Then V_n will be given by

$$\mathfrak{D}(x,p) = \max_{M\in\mathfrak{M}_{iii}} \sum_{r(i)\in M} (x_i - p_i)$$

where $\mathfrak{M}_{iii}$ is $\mathfrak{M}$ given in (iii) for D_n, and r( ) is a cyclic rotation of the points of M. Hence $\mathfrak{M}_{iv}$ consists of $\mathfrak{M}_{iii}$ and all the cyclic rotations of $\mathfrak{M}_{iii}$. By a cyclic rotation of a set M we mean that

$$r(M) = \{i_1,\ldots,i_c\} \text{ where } i_m \equiv j_m + a \pmod q \quad m = 1,\ldots,c.$$

a an integer and

$$M = \{j_1,\ldots,j_c\}$$

and we define $r(i_m) = r(j_m)$.

(v) If $\mathfrak{M}$ contains all the subsets of $\{1,\ldots,k\}$ then

$$\mathfrak{D}(x,p) = \tfrac{1}{2}\,\Sigma|x_i-p_i|,$$

see Hoeffding (1965).

4.3 Some extensions of theorems of Hoeffding

First, some definitions from Hoeffding are quoted. Define

$$I(x,p) = \sum_{i=1}^{q} x_i \log(x_i/p_i)$$

and Ω to be the simplex

$$\Omega = \{(x_1,\ldots,x_q);\ x_1 \geq 0,\ldots,x_q \geq 0,\ x_1+\ldots+x_q = 1\},$$

$$\Omega_o = \{(x_1,\ldots,x_q);\ x_1 > 0,\ldots,x_q > 0,\ x_1+\ldots+x_q = 1\}$$

and

$$\Omega(p) = \{(x_1,\ldots,x_q);\ x_i = 0 \text{ if } p_i = 0\}.$$

If A is a non-empty sub-set of Ω, then $\bar{A}$ denotes the closure of the set A.

This work is based on the idea of large deviations and Sanov's (1961) work and is summarized in Hoeffding's theorem 2.1.

Let A be a proper sub-set of Ω. Define

$$I(A,p) = \inf\{I(x,p);\ x \in A\}$$

and $I(A,p) = \infty$ if $A = \varphi$.

Let $n_1,\ldots,n_q$ be the results of a multinomial experiment with n observations, where $\sum_{i=1}^{q} n_i = n$. Define $z_i = n_i/n$ then

$Pr(Z_{(n)} = z_{(n)}) = [n!/(n_1!,\ldots,n_q!)]p_1^{n_1},\ldots,p_q^{n_q}$ where $Z_{(n)} = Z_1^{(n)},\ldots,Z_q^{(n)})$ is a random variable distributed as a multinomial random variable with $p = (p_1,\ldots,p_q)$, and $z_{(n)} = (z_1,\ldots,z_q) = (n_1/n,\ldots,n_q/n)$.

Let $A_{(n)}$ denote the set of lattice points $z_{(n)}$ containing in A. We quote some of Hoeffding's Theorems, his numbers preceded by an 'H'.

Theorem H2.1. For any set $A \subset \Omega$, and any point $p \in \Omega$ we have

$$C_o n^{-(q-1)/2} \exp\{-nI(A_{(n)},p) \leq Pr(Z_{(n)} \in A/p)$$

$$\leq \binom{n+q-1}{q-1} \exp\{-nI(A_{(n)},p)\}$$

where C_o is a constant not dependent on n. Hence

$$Pr(Z_{(n)} \in A/p) = \exp\{-nI(A_{(n)},p)\} + O(\log n)\}$$

uniformly for $A \subset \Omega$ and $p \in \Omega$.

Also

$$Pr(Z_{(n)} \in A/p) \leq \exp\{-nI(A,p) + O(\log n)\}$$

uniformly for $A \subset \Omega$ and $p \in \Omega$.

This result gives an asymptotic expression for the logarithm of the probability $Pr(Z_{(n)} \in A/p)$. This result is non-trivial if $Z_{(n)}$ is far from its mean p. Hence if A is taken as a critical region for some test this enables the power to be compared with the likelihood ratio test.

We quote Lemmas 4.2 and 4.7 of Hoeffding.

Lemma H4.2 Let A be a non-empty subset of Ω.

(a) Let $p \in \Omega_o$. There is at least one point y such that

$$y \in \bar{A},\ I(y,p) = I(A,p). \qquad (4.1)$$

If $p \in \bar{A}$, then $I(A,p) = 0$ and (4.1) is satisfied only with

$y = p$. If $p \notin \bar{A}$, then $I(A,p) > 0$ and any y which satisfies (4.1) is in the boundary of A.

(b) Let $p \notin \Omega_o$, then $I(A,p) < \infty$ if and only if the intersection $A \cap \Omega(p)$ is non-empty. If this is the case then $I(A,p) = I(A \cap \Omega(p),p)$ and the statements of part (a) are true with A replaced by $A \cap \Omega(p)$.

Lemma H4.7 Let $A = \{x_i\ f(x) > 0\}$ where the function $f(x)$ is continuous in Ω and $\max f(x) > 0$. Let p be a point in Ω_o such that $f(p) < 0$. Suppose further that the derivatives $f_i'(x) = \partial f(x)/\partial x_i$ $i = 1,\ldots,q$ exist and are continuous at all x in Ω_o for which $f(x) = 0$.

Let y be any point in $\bar{A}$ such that $I(y,p) = I(A,p)$. Then if $y \in \Omega_o$ it is necessary that $f(y) = 0$ and

$$\log(y_i/p_i) = a\, f_i'(y) + b \qquad i = 1,\ldots,q$$

where $a > 0$ and b are constants.

Let $p_o \in \Omega_o$ and $0 < \epsilon < \max_x \mathfrak{D}(x,p_o)$. Consider the critical region A for testing the simple hypothesis $H_o : p = p_o$, where

$$A(\epsilon) = \{x;\ \mathfrak{D}(x,p_o) \geq \epsilon\}$$

$A(\epsilon)$ is the union of the sets $A_M(\epsilon) = \{x;\ D_M(x,p_o) \geq \epsilon\}$ i.e.

$$A(\epsilon) = \bigcup_{M \in \mathfrak{M}} A_M(\epsilon)$$

We now state and prove a Theorem which follows Hoeffding's Theorem 8.1.

Theorem 4.1. Suppose that $p_o \in \Omega_o$ and $0 < \epsilon < \max_x \mathfrak{D}(x,p_o)$ then $\exists$ $y \in A(\epsilon)$ such that

$$I(A(\epsilon),p_o) = I(y,p^o)$$

and

$$y_i = \begin{cases} a\, p_{oi} & i \in M^* \\ b\, p_{oi} & i \notin M^* \end{cases}$$

where $\sum_{i \in M^*} (y_i - p_{oi}) = \mathfrak{D}(y, p_o) = \epsilon$

$$a = \frac{\epsilon + h^*}{h^*}, \qquad b = \frac{1 - h^* - \epsilon}{1 - h^*}$$

and
$$I(y, p_o) = \tfrac{1}{2} \frac{\epsilon^2}{h^*(1-h^*)}$$

where $h^* = \sum_{i \in M^*} p_{oi}$

<u>Proof</u> By Lemma H4.2 we must have $\mathfrak{D}(y, p_o) = \epsilon$, and let

$$\mathfrak{D}(y, p_o) = \sum_{i \in M^*} (y_i - p_{oi}) \text{ where } M^* \in \mathfrak{M}. \qquad (4.2)$$

By lemma H4.7, put $f(x) = \mathfrak{D}(x, p_o) - \epsilon$, then $f(y) = 0$ and so if $I(y, p) = I(A(\epsilon), p)$

$$\log(y_i / p_{oi}) = s\delta_i + t \text{ where } \delta_i = \begin{cases} 0 & i \notin M^* \\ 1 & i \in M^* \end{cases} \qquad (4.3)$$

s and t are constants and $s > 0$.

The set M^* may or may not be unique, however there is a unique point y defined by (4.2) and (4.3) given M^*.

$$y_i = \begin{cases} a\, p_{oi} & i \in M^* \\ b\, p_{oi} & i \notin M^* \end{cases}$$

where $a = e^s e^t$ and $b = e^t$. Since $s > 0$, $a > b$.

We also have conditions $\sum y_i = 1$ and $\mathfrak{D}(y, p_o) = \epsilon$. These are equivalent to

$$ah^* + b(1-h^*) = 1$$
$$ah^* - h^* = \epsilon$$

where
$$h^* = \sum_{i \in M^*} p_{oi}$$

So
$$a = \frac{\epsilon + h^*}{h^*}, \qquad b = \frac{1 - \epsilon - h^*}{1 - h^*}$$

We can now find $I(y, p_o)$ explicitly

$$I(y,p^o) = \sum_{i=1}^{k} y_i \log(y_i/p_{oi})$$

$$= \sum_{i\in M^*}' y_i \log(y_i/p_{oi}) + \sum_{i\notin M^*} y_i \log(y_i/p_{oi})$$

$$= h^* a \log a + (1-h^*) b \log b$$

$$= J(H^*), \text{ say.}$$

Now $J(h^*) = (\epsilon+h^*)\log(1+\frac{\epsilon}{h^*}) + (1-\epsilon-h^*)\log(1-\frac{\epsilon}{1-h^*})$

$$= (\epsilon+h^*)(\frac{\epsilon}{h^*} - \tfrac{1}{2}\{\frac{\epsilon}{h^*}\}^2) + (1-\epsilon-h^*)(-\frac{\epsilon}{1-h^*} - \tfrac{1}{2}\{\frac{\epsilon}{1-h^*}\}^2) + O(\epsilon^3)$$

$$= \frac{\epsilon^2}{h^*} + \epsilon - \tfrac{1}{2}\frac{\epsilon^2}{h^*} - \epsilon - \tfrac{1}{2}(\frac{\epsilon^2}{1-h^*}) + \frac{\epsilon^2}{1-h^*} + O(\epsilon^3)$$

$$= \tfrac{1}{2}\frac{\epsilon^2}{h^*(1-h^*)} + O(\epsilon^3)$$

Now $I(A(\epsilon),p^o) = \min_{M\in\mathfrak{M}} I(A_M(\epsilon),p_o)$

$$= \min_{M\in\mathfrak{M}} J(h_M)$$

$$= \min_{M\in\mathfrak{M}} \left\{\tfrac{1}{2}\frac{\epsilon^2}{h_M(1-h_M)}\right\}$$

$$= \min_{M\in\mathfrak{M}} J(h_M)$$

where $J(h_M) = h_M a \log a + (1-h_M) b \log b$

Now $\frac{1}{h(1-h)}$ has a unique minimum at $h = \tfrac{1}{2}$ $(0 < h < 1)$. So the set M^* is chosen such that

$$|h^* - \tfrac{1}{2}| \leq |h_M - \tfrac{1}{2}| \qquad M \in \mathfrak{M}. \qquad (4.4)$$

Note that if $p_{oi} = 1/q$ $i = 1,\ldots,q$, and $\mathfrak{M}$ is chosen to correspond to D_n^+ ($\mathfrak{M}_i$ in §4.2), then there is just one set M^* satisfying (4.4) (or two if q is odd). If $\mathfrak{M}$ is chosen to correspond to D_n ($\mathfrak{M}_{iii}$) then there are two sets M^* satisfying (4.4) (or four if q is odd). If $\mathfrak{M}$ corresponds to V_n then there are 2q sets M^* satisfying (4.4) (or 4q if q is odd).

Theorem 4.2 Let $p_o \in \Omega_o$, $p \in \Omega_o$, $\epsilon > 0$, $\mathfrak{D}(p,p_o) > \epsilon$ then $\exists$ a unique point z such that $z \in \bar{A}'(\epsilon)$,

$$I(A'(\epsilon),p) = I(z,p) \qquad (4.5)$$

The point z is given by

$$\log(z_i/p_i) = \begin{cases} a' & i \in M^+ \\ b' & i \notin M^+ \end{cases}$$

with $a' < b'$ constants and $\mathfrak{D}(z,p_o) = \sum_{i \in M^+} (z_i - p_{oi}) = \epsilon$.

To prove Theorem 4.2 Hoeffding's Lemma 4.8 is needed.

Lemma H4.8 If A is convex and $A \cap \Omega_o$ is not empty and if $p \in \Omega_o$, $p \notin \bar{A}$, then there is exactly one point $y \in \bar{A}$ such that $I(y,p) = I(A,p)$.

A lemma is needed.

Lemma 4.1 The set $A'(\epsilon)$, where

$$A(\epsilon) = \{x;\ \mathfrak{D}(x,p_o) \geq \epsilon\}$$

is convex.

Proof Let $y \in A'(\epsilon)$, i.e. $\mathfrak{D}(y,p_o) < \epsilon$; let $z \in A'(\epsilon)$. Consider $0 \leq a \leq 1$, and define $x = ay + (1-a)z$ then for $M \in \mathfrak{M}$

$$\sum_{i \in M} (x_i - p_{oi}) = \sum_{i \in M} (ay_i + (1-a)z_i - p_{oi})$$

$$= a \sum_{i \in M} (y_i - p_{oi}) + (1-a) \sum_{i \in M} (z_i - p_{oi})$$

$$\leq a \sum_{i \in M_1} (y_i - p_{oi}) + (1-a) \sum_{i \in M_2} (z_i - p_{oi}).$$

where $M_1, M_2 \in \mathfrak{M}$

and $$\mathfrak{D}(y,p_o) = \sum_{i \in M_1} (y_i - p_{oi})$$

$$\mathfrak{D}(z,p_o) = \sum_{i \in M_2} (z_i - p_{oi})$$

$$< \epsilon \quad \text{since } y,z \in A'(\epsilon).$$

Hence $\sum_{i \in M} (x_i - p_{oi}) < \epsilon$ for all $M \in \mathfrak{M}$ and so $\mathfrak{D}(x,p) < \epsilon$, i.e. $x \in A'(\epsilon)$, and so $A'(\epsilon)$ is convex.

<u>Proof of Theorem 4.2</u> By Lemma H4.8 and Lemma 4.1, exactly one point Z which satisfies (4.5). By Lemma H4.7, Z must satisfy

$$\log \frac{z_i}{p_i} = -s'\delta + t', \qquad s' > 0$$

$$\text{where } \delta = \begin{cases} 0 & i \notin M^+ \\ 1 & i \in M^+ \end{cases}$$

where M^+ is chosen so that $\mathfrak{D}(z,p^o) = \sum_{i \in M^+} (z_i - p_{oi}) = \epsilon$ with $f(x) = \epsilon - \sum_{i \in M^+} (z,p_o)$.

So writing $\quad a' = t' - s'$

$\qquad\qquad\quad b' = t'$

we have

$$\log(z_i/p_i) = \begin{cases} a' & i \in M^+ \\ b & i \notin M^+ \end{cases}$$

with $a' < b'$ since $s' > 0$.

Now define

$$\left.\begin{aligned} B(\epsilon) &= \{x;\ I(x,p_o) \geq I(A(\epsilon),p_o)\} \\ \text{where}\quad A(\epsilon) &= \{x;\ \mathfrak{D}(x,p_o) \geq \epsilon\} \\ \text{as before. Then let} & \\ d(p,\epsilon) &= I(B'(\epsilon),p) - I(A'(\epsilon),p) \end{aligned}\right\} \quad (4.6)$$

Note that $A(\epsilon)$ is the critical region of the test based on $\mathfrak{D}$ and the logarithm of the significance level is given by $I(A(\epsilon),p_o)$. $B(\epsilon)$ is the critical region based on the likelihood ratio test, with the logarithm of the significance level given by $I(B(\epsilon),p_o)$, Hoeffding's Theorem 7.1. $d(p,\epsilon)$ is the difference of the logarithm of the probabilities of the type II errors, when the alternative is p. Trivially from Lemma H4.7

$I(B(\epsilon),p_o) = I(A(\epsilon),p_o)$ putting $f(x) = I(x,p_o) - I(A(\epsilon),p_o)$;

Hoeffding's Theorem 6.3 is needed before conditions for $d(p,\epsilon) = 0$ are given, i.e. the powers of the test to be of the same order.

Theorem H6.3 Let A and Λ be non-empty subsets of Ω such that $0 < I(A,\Lambda) < \infty$. Let

$$B = \{x;\ I(x,\Lambda) \geq I(A,\Lambda)\}$$

and for any p such that $I(A',p) < \infty$, let

$$d(p) = I(B',p) - I(A',p)$$

(i) Always $d(p) \geq 0$; $d(p) = 0$ if and only if

$$\{x;\ I(x,p) < I(B',p)\} \subset A$$

(ii) If $d(p) = 0$ and $0 < I(B',p) < \infty$ then $I(B',p)$ $= I(y,p)$ for some common boundary point of A and B.

We now give the next theorem.

Theorem 4.3 Let $p_o \in \Omega_o$ and $0 < \epsilon < \max_x \mathfrak{D}(x,p_o)$ then $d(p,\epsilon) > 0$ unless

$$p_i = p_{oi} \begin{cases} \alpha & i \in M^+ \\ \beta & i \notin M^+ \end{cases}, \text{with } \alpha > \beta$$

and $M^+ \subset \mathfrak{M}$.

Proof With the definitions of (4.6) of A, B and d by Theorem H6.3 with $\Lambda = \{p_o\}$, if $d(p,\epsilon) = 0$ then there exists a common boundary point y of $A(\epsilon)$ and $B(\epsilon)$ such that $I(y,p) = I(B'(\epsilon),p)$. By theorem 4.2 there exists a unique Z such that $z \in \bar{A}'(\epsilon)$ and

$$I(A'(\epsilon),p) = I(z,p) \text{ and}$$

$$\log(z_i/p_i) = \begin{matrix} a' & i \in M^+ \\ b' & i \notin M^+ \end{matrix} \qquad (4.7)$$

where $\qquad \mathfrak{D}(z_i,p_o) = \sum_{i \in M^+} (z_i - p_{oi}) = \epsilon$

If y is a boundary point of $B(\epsilon)$ and $I(y,p) = I(B'(\epsilon),p)$ then by Lemma H4.7 with $f(x) = -I(x,p_o) + I(A(\epsilon),p_o)$

$$\log(y_i/p_i) = -c\log(y_i/p_{oi}) + d \quad c > 0 \qquad (4.8)$$

and $i = 1,\ldots,k$

Also $I(y,p_o) = I(A(\epsilon),p_o)$.

Since $I(z,p) = I(A'(\epsilon),p) = I(B'(\epsilon),p)$ and z must satisfy (4.8).

Eliminating z_i from (4.7) and (4.8) with $y = z$, we have

$$-\log p_i - c\log p_{oi} - d = -(1+c)\{\log p_i + a' + \delta_i(b'-a')\}$$

$$\delta_i = \begin{cases} 0 & i \in M^+ \\ 1 & i \notin M^+ \end{cases}$$

i.e. $$p_i = p_{oi}\ e^{d/c} \times \begin{cases} e^{-\frac{1+c}{c}a'} & i \in M^+ \\ e^{-\frac{1+c}{c}b'} & i \notin M^+ \end{cases}$$

Since $c > 0$, $(1+c)/c > 0$ and $a' < b'$ then $e^{-a'(1+c)/c} > e^{-b'(1+c)/c}$. Hence

$$p_i = p_{oi} \times \begin{cases} \alpha & i \in M^+ \\ \beta & i \notin M^+ \end{cases}$$

with $\alpha > \beta$.

Following Hoeffding's Theorem 8.4, we can say that the likelihood ratio test is 'considerably more powerful' than the $\mathfrak{D}$ test, unless $d(p,\epsilon) = 0$, when the size of the test α_n goes to 0 fast enough. By 'considerably more powerful' we mean that the ratio of the error probabilities of the two tests tends to 0 faster than any power of n. Stating this more precisely in a theorem.

<u>Theorem 4.4</u> Let $p_o \in \Omega_o$ and $0 < \epsilon_n < \max_{x_n} \mathfrak{D}(x_n,p_o)$. The size of the $\mathfrak{D}$-test, which rejects the null hypothesis $p = p^o$ if $z_n \in A_n(\epsilon_n) = \{x_n | \mathfrak{D}(x_n,p_o) \geq \epsilon_n\}$ is given by

$$\Pr(A(\epsilon_n)|p_o) = \exp\{-nI(A(\epsilon_n),p_o) + O(\log n)\}$$

where $I(A(\epsilon_n), p_o)$ is given in in Theorem 4.1.

If $p \in \Omega_o$, $n\epsilon_n \to \infty$ as $n \to \infty$ then (4.9)

$$P(A'(\epsilon_n)|p) = \exp\{-NI(A'(\epsilon_n), p) + O(\log n)\}$$

$I(A'(\epsilon_n), p)$ is given in Theorem 4.3.

Remark From the asymptotic theory of the $\mathfrak{D}$-tests we know that $\sqrt{n}\,\mathfrak{D}$ has a non-trivial limiting distribution, hence $\Pr(\mathfrak{D} > \epsilon_n) \to 0$ if $n\epsilon_n \to \infty$ as $n \to \infty$, i.e. $\alpha_n \to 0$.

Theorem 4.4 (continued)

Consider the likelihood ratio test which rejects the null hypothesis $p = p_o$ if $z_n \in B_n(\epsilon_n) = \{x_n / I(x, p_o) \geq I(A(\epsilon_n), p_o) + \delta_n\}$, where the $\delta_n = O(\log n/n)$, $\delta_n > .0$ are such that the size of the test is given by

$$\Pr(B_n/p_o) \leq \Pr(A(\epsilon_n)|p_o)$$

If the conditions (4.9) are satisfied, then

$$\Pr(B_n'|p) = \exp\{-n\alpha(p,\epsilon_n) + O(\log n | I^{\frac{1}{2}}(A(\epsilon_n), p_o))\} \times P_r(A'(\epsilon_n)|p)$$

where $d(p,\epsilon_n)$ has the properties given in Theorem 4.3.

If $d(p,\epsilon_n) = 0$ then

$$\Pr(B_n'|p) = n^{g/\epsilon} \Pr(A'(\epsilon_n)|p), \quad g \text{ a constant, } g \geq 0,$$

and the $\mathfrak{D}$-test is as least as powerful as the likelihood ratio test for the alternatives p where $d(p,\epsilon_n) = 0$, given in Theorem 4.3.

Proof The proof follows Hoeffding's Theorem 8.4.

4.4. Some examples of the previous results

From Hoeffding's Theorem 8.4 it is seen that the X^2 test is as least as powerful as the LR test when the alternative is given by

$$p_j = 1 - a + ap_{oj}, \quad p_i = ap_{oi} \quad i \neq j \tag{4.10}$$

$$0 < a < 1 \text{ and } p_{oj} = p_{o,min},$$

where $p_{o,min}$ is the minimum of $\{p_{oi}\}$. If there is more than one j such that $p_{o,j} = p_{o,min}$ then there is an alternative of the form (4.10) corresponding to each j such that $p_{o,j} = p_{o,min}$.

From Theorem 4.4, we see that the Ω test is as least as powerful as the LR test when the alternative is given by

$$p_i = p_{oi} \times \begin{cases} \alpha & i \in M^+ \\ \dfrac{1-\alpha h^+}{1-h^+} & i \notin M^+ \end{cases} \qquad (4.11)$$

where $\alpha > 0$, $h^+ = \sum_{i \in M^+} p_{oi}$ and $M^+ \subset \mathfrak{M}$.

The two alternatives (4.10) and (4.11) coincide if $M^+ = \{j\}$ where $P_{oj} = P_{o,min}$. In this case we would exoect the X^2 to be more powerful than the Ω test.

If we consider the case of grouping data into q equi-probable cells then $p_{oi} = \frac{1}{q}$ for $i = 1,\ldots,q$. There are then q alternatives of the form (4.10). For the various sets $\mathfrak{M}$ corresponding to the various Ω statistics (see §4.2) the number of alternatives (4.11) varies. For example if $\mathfrak{M}$ is defined as $\mathfrak{M}_i$ then Ω corresponds to D^+ and there are (q-1) alternatives of the form (4.11); if $\mathfrak{M}$ is defined to be $\mathfrak{M}_{iii}$ and Ω corresponds to D then there are 2(q-1) alternatives of the form (4.11). The Ω tests will be more powerful than the X^2 test except when M^+ of (4.11) is defined such that $M^+ = \{j\}$, $P_{oj} = P_{omin}$. More results involving the power of the statistics are given in the next section for this case when the significance level does not $\rightarrow 0$ or $\rightarrow 1$, with $n \rightarrow \infty$.

§5. A Comparison of the asymptotic powers of the X^2-test and $D_{n,q}$ test with grouped data for a fixed significance level, and a fixed number of cells.

5.1. Introduction

In this section we study the asymptotic powers of the χ^2-test and the $D_{n,q}$ test by use of two approximate results. For the test of the multinomial hypothesis

$$H_o : p_1 = p_{o1},\ p_2 = p_{o2}, \ldots, p_q = p_{oq},$$

the distribution of the X^2 statistic when the alternative

$$H_1 : p_1 = p_{11},\ p_2 = p_{12}, \ldots, p_q = p_{1q}$$

is true is that of a non-central χ^2 distribution with parameter

$$\lambda = \sum_{i=1}^{q} \frac{c_i^2}{p_{oi}} \quad \text{where } p_{1i} - p_{oi} = c_i/\sqrt{n} \qquad (5.1)$$

and $q-1$ degrees of freedom, see Kendall and Stuart (1973, §30.27). Patnaik (1949) has tabulated the power of the X^2 statistic at the .05 significance level for various λ and q. Use is made of these tables.

5.2. Previous results

We found the asymptotic distribution of $\sqrt{n}\,D_{n,q}$ in §3, so providing the asymptotic significance level $\sqrt{n}\,d_\alpha$ for each q, say. Kendall and Stuart (1973, §30.60) provide a lower bound to the power of $\sqrt{n}\,D_n$ for continuous data:

$$\text{Power} = P \geq 1 - \left[\Phi\left(\frac{F_o - F_1 + d}{\{F_1(1-F_1)/n\}^{\frac{1}{2}}}\right) - \Phi\left(\frac{F_o - F_1 - d}{\{F_1(1-F_1)/n\}^{\frac{1}{2}}}\right) \right] \qquad (5.2)$$

where F_o and F_1 are given by the values of $F_o(x_s)$ and $F_1(x_s)$ respectively where $\Delta = |F_o(x_s) - F_1(x_s)| = \sup_x |F_o(x) - F_1(x)|$. For the discrete case

$$F_o(x) = \sum_{i=1}^{r} p_{oi} \quad \text{for} \quad \sum_{i=1}^{r} p_{oi} \le x < \sum_{i=1}^{r+1} p_{oi}$$

and $$F_1(x) = \sum_{i=1}^{r} p_{1i} \quad \text{for} \quad \sum_{i=1}^{r} p_{oi} \le x < \sum_{i=1}^{r+1} p_{oi}$$

The lower bound (5.2) can be used for the discrete case and it will provide an even better bound than for the continuous case.

For this limited study, q was chosen to be equal to 10 and the 5% significance level used. λ, the non-central χ^2 parameter, was chosen to give the X^2 test a power of .50 and .90. The alternatives considered were

(i) $\underline{p}_1 = (\frac{1}{10} + \frac{9\gamma}{\sqrt{n}}, \frac{1}{10} - \frac{\gamma}{\sqrt{n}}, \ldots, \frac{1}{10} - \frac{\gamma}{\sqrt{n}})$

$x_s = \frac{1}{10}$ and $\lambda = 10.90\ \gamma^2$

(ii) $\underline{p}_1 = (\frac{1}{10} + \frac{4\gamma}{\sqrt{n}}, \frac{1}{10} + \frac{4\gamma}{\sqrt{n}}, \frac{1}{10} - \frac{\gamma}{\sqrt{n}}, \ldots, \frac{1}{10} - \frac{\gamma}{\sqrt{n}})$

$x_s = \frac{1}{5}$ and $\lambda = 10.40\ \gamma^2$

(iii) $\underline{p}_1$ such that

$$p_i = \begin{cases} \frac{1}{10} + \frac{\gamma}{\sqrt{n}} & i = 1, \ldots, 5 \\ \frac{1}{10} - \frac{\gamma}{\sqrt{n}} & i = 6, \ldots, 10 \end{cases}$$

then $x_s = \frac{1}{2}$ and $\lambda = 10.10\ \gamma^2$

From Patnaiks (1949) tables for 50% and 90% power we have $\lambda = 8.84$ and $\lambda = 19.85$ respectively. We then solve for γ, and find the lower bound to the power of $\sqrt{n}\, D_{n,q}$ by using (5.2). The value of d_α is found from the M.C. results and Table (3.1). It is found that $d_\alpha = 1.175$ for $q = 10$.

5.3. Results

From the results, Table 5.1, it is seen that the power of the X^2 test is exceeded for alternatives 2. and 3. by the D test. The lower bound to the power (LBP) of D for alternative 1 is less than the X^2 power and one intuitively feels that these lower bounds are fairly close to the actual powers.

For alternatives of the form

$$\underline{p}_1 = \begin{cases} p_i = \frac{1}{q} + a & i = 1,\ldots,q_1 \\ p_i = \frac{1}{q} - b & i = q_1+1,\ldots,q \end{cases}$$

where a and b are chosen so that $\Sigma p_i = 1$, we would expect the power of D to be much greater than that of X^2, provided q_1 is not too near 1 or q. This is the type of alternative for which D was shown to be as powerful as the likelihood test in §4, but under different circumstances. From the Monte Carlo results of §2.6, the results of §4, and the results presented here we see that the D statistic (or D^+ or D^- for the one-sided test) should have good power for this type of alternative, under varying circumstances of sample size and significance level. It must be noted that the power of the X^2 statistic is invariant under orderings of the elements of the alternative $\underline{p}_1$.

X^2 power .50 $\lambda = 8.84$

Alternative	Δ	γ	LBP of D
1	9γ	.099	.17
2	8γ	.149	.52
3	5γ	.297	.73

X^2 power .90 $\lambda = 19.85$

Alternative	Δ	γ	LBP of D
1	9γ	.149	.71
2	8γ	.223	.94
3	5γ	.446	.98

Table 5.1

Comparison of asymptotic power of X^2-test with D test, with discrete data. Testing the null hypothesis that 10 points are equally likely at the 5% level.

Part II Goodness-of-fit for censored data using Cramér-von Mises type statistics

Introduction

In some applications of goodness-of-fit it is necessary only to consider whether a certain proportion of the random sample fits the hypothesised distribution. This may arise in various ways, but the most likely is through censoring of some kind. Statistics of the K-S type were introduced by Renyi (1953) for this problem

$$\sup_{a \leq F(x) \leq b} \left\{ \frac{F_n(x)-F(x)}{F(x)} \right\} .$$

and
$$\sup_{F(x) \geq a} \left\{ \frac{F_n(x)-F(x)}{F(x)} \right\}$$

He found the asymptotic distribution of these statistics for continuous F. Birnbaum and Lientz have introduced many Renyi-type statistics (see their paper (1969) for a review). In particular they have found the exact finite and asymptotic distributions of the statistics

$$\sup_{F_n(x) \leq a} \left\{ F_n(x)-F_o(x) \right\}$$

$$\sup_{F(x) \leq b} \left\{ F_n(x)-F_o(x) \right\}$$

In Part II the extension of the Renyi idea to the CVM-type statistics for continuous F is considered and in particular the statistics below are studied

$$_pW^2_n = n \int_{-\infty}^{F_o^{-1}(p)} \{F_n(x)-F_o(x)\}^2 \, dF_o(x) \qquad (6.1)$$

$$_rW^2_n = n \int_{-\infty}^{x_{(r)}} \{F_n(x)-F_o(x)\}^2 \, dF_o(x) \qquad (6.2)$$

$${}_pA^2_n = n\int_{-\infty}^{F_o^{-1}(p)} \{F_n(x)-F_o(x)\}^2/\{F_o(x)(1-F_o(x))\}\ dF_o(x)$$

$${}_rA^2_n = n\int_{-\infty}^{x'(r)} \{F_n(x)-F_o(x)\}^2/\{F_o(x)(1-F_o(x))\}\ dF_o(x)$$

to test the simple hypothesis $H_o : F(x) \equiv F_o(x)$. The theory is extended for the case of doubly censored data. An approximation is made to the distribution of ${}_pW^2_n$ and ${}_rW^2_n$ for small n. The asymptotic theory of ${}_pW^2_n$ and ${}_rW^2_n$ is given and the tests extended for a composite hypothesis. In particular tests for normality and the exponential distribution are developed, the theory is given and asymptotic percentage points provided, which are checked against empirical percentage points for the statistics. The asymptotic distribution of ${}_pA^2_n$ and ${}_rA^2_n$ is found and percentage points given. Also the asymptotic sample theory of testing composite hypotheses using ${}_pA^2_n$ or ${}_rA^2_n$ is developed providing a test for normality with censored samples having good power.

Various power studies are presented giving the (exact asymptotic) power of ${}_pW^2$ and ${}_pA^2$ for shifts of scale and location, and the small sample empirical power of the tests based on ${}_rW^2_n$ and ${}_rA^2_n$ when testing for normality.

§6. The W^2 statistic for testing goodness-of-fit with censored data with a simple hypothesis

6.1. Introduction

In this section the moments of the statistics ${}_pW^2_n$ and ${}_rW^2_n$ and approximate significance points are found by fitting a distribution of the type of the random variable T, where

$$T = aC_v^2 + b \qquad (6.3)$$

and a, b and v are chosen so that the first three moments of T are equal to those of ${}_pW_n^2$ or ${}_rW_n^2$, and C_v^2 has the χ^2 distribution with v degrees of freedom.

The statistics ${}_pW_n^2$ (6.1) and ${}_rW_n^2$ (6.2) are applicable to testing goodness-of-fit with censored data. The statistic ${}_pW_n^2$ is used with censoring of Type I on the right, where the observations greater than $F_o^{-1}(p)$ are censored, and so the number of observations observed is a random variable. The statistic ${}_rW_n^2$ is used with censoring of Type II an the right, where the values observed are the r smallest, and so the end point of integration $x_{(r)}$ is a random variable. Obviously the statistics can be used with censoring on the left. The statistics

$$n \int_{F_o^{-1}(1-p)}^{\infty} \{F_n(x)-F_o(x)\} \; dF_o(x) \qquad (6.4)$$

$$n \int_{x_{(s)}}^{\infty} \{F_n(x)-F_o(x)\}^2 \; dF_o(x)$$

are distributed the same as ${}_pW_n^2$ and ${}_rW_n^2$, where $s = n-r+1$. The approximation (6.3) is quite accurate for the upper-tail of the distributions of the statistics (see Tiku (1963), and Stephens and Maag (1968) for the application to W_n^2), but is not applicable to the lower tail since, among other reasons, the value of 'a' in (6.3) does not coincide with the minimum value of the statistics and the distribution is adrift at the lower tail. Also found in this section, is the exact lower-tail distribution of the statistic ${}_pW_n^2$, for small n. Also the asymptotic theory of the statistics is found and accurate percentage points provided. The asymptotic theory is extended so that the distribution of the statistics for doubly censored data, e.g.

$$ {}_{p,q}W_n^2 = n \int_{F_o^{-1}(q)}^{F_o^{-1}(p)} \{F_n(x)-F_o(x)\}^2 \; dF_o(x) \qquad 0 \leq q < p \leq 1 $$

can be found.

Also the asymptotic power is given when $F_o(x)$ is considered to be normal and shifts of scale and location are considered.

6.2 Small sample properties of ${}_pW_n^2$, ${}_rW_n^2$ and ${}_{q,p}W_n^2$

6.2.1 Moments of ${}_pW_n^2$

First consider the statistic ${}_pW_n^2$. Upon the transformation $t = F_o(x)$ this statistic becomes

$$ {}_pW_n^2 = n \int_o^p \{F_n(t)-t\}^2 \; dt \tag{6.5} $$

The moments of ${}_pW_n^2$ can be found from (6.5). For example

$$ E[{}_pW_n^2] = n \int_o^p E[\{F_n(t)-t\}^2]\, dt $$

$$ = \int_o^p t(1-t)\, dt $$

$$ = \frac{p^2}{2} - \frac{p^3}{3} \tag{6.6} $$

taking the expectation under the integral sign, and the second moment about the origin:

$$ E[({}_pW_n^2)^2] = n^2 \int_o^p \int_o^p E[\{F_n(t)-t\}^2 \{F_n(s)-s\}^2]\, dt\, ds $$

Now $nF_n(t)$ and $nF_n(s)$ are trinomial random variables. If $t \leq s$ then by tedious expansion it it found that

$$ E[n^2 \{F_n(t)-t\}^2 n^2 \{F_n(s)-s\}^2] $$

$$= nt + 2n(n-2)t^2 + n(n-3)st$$

$$+ 5n(2-n)st^2 + n(2-n)s^2t + n(3n-6)s^2t^2 , \quad \text{for } t \leq s$$

Then upon integrating and noting that

$$\int_o^p \int_o^p f(s,t)\, ds\, dt = 2\int_o^p \int_o^s f(s,t)\, dt\, ds$$

for symmetric $f(s,t)$, it is found that

$$E[({}_pW_n^2)^2] = [\frac{2p^3}{n}[\frac{1}{6} + \{\frac{7n-17}{24}\}p + \{2-n\}\frac{13}{30}p^2 + \frac{n-2}{6}p^3] \qquad (6.7)$$

Note that if $p = 1$ then

$$E[{}_1W_n^2] = \frac{1}{20} - \frac{1}{60n}$$

and so $\text{var}({}_1W_n^2) = \frac{4n-3}{180n}$

in agreement with previous results on W_n^2.

Similarly we can find $E[({}_pW_n^2)^3]$. This expression

$$E[n^2\{F_n(t)-t\}^2 n^2\{F_n(s)-s\}^2 n^2\{F_n(r)-r\}^2] \qquad (6.8)$$

with $r \leq s \leq t$ is used. By tedious algebra again it is found that (6.8) is given by

$$nr + 2n(7n-8)r^2 + 2n(5n-6)rs + n(n-3)rt$$
$$+ 5n(2n^2-19n+18)r^2s + 2n(n^2-19n+20)r^2t + n(2n^2-21n+18)rs^2$$
$$+ n(2-n)rt^2 + n(n^2-27n+30)rst + 4n(3n^2+26n-24)r^2s^2$$
$$+ 2n(-n^2+12n-12)r^2t^2 + 5n(6n^2+35n-42)r^2st$$
$$+ n(6n^2+35n-42)rs^2t + n(-n^2+17n-18)rst^2$$
$$+ n(27n^2-234n+216)r^2s^2t + 5n(3n^2-26n+24)r^2st^2$$
$$+ n(3n^2-26n+24)rs^2t^2 + 5n(-3n^2+26n-24)r^2s^2t^2 .$$

Upon integrating and noting that

$$\int_o^p \int_o^p \int_o^p f(r,s,t)\, dr\, ds\, dt = 4\int_o^p \int_o^t \int_o^s f(r,s,t)\, dr\, ds\, dt$$

it is found that

$$E[({}_pW_n^2)^3] = \frac{6p^4}{n^2}[\frac{1}{24} + \frac{31n-40}{60}p + (\frac{2n^2-19n+18}{60}$$

$$+ \frac{n^2-27n+30}{48} + \frac{5n^2-58n+58}{36})p^2 - \frac{p^3}{7}(\frac{2}{9}\{3n^2-26+24\}$$

$$+ \frac{1}{6}\{n^2-12n+12\} + \frac{13}{30}\{5n^2-45n+42\} + \frac{1}{8}\{n^2-17n+18\})$$

$$+ \frac{7}{60}p^4(3n^2-26n+24) - \frac{5}{162}p^5(3n^2-26n+24)] \qquad (6.9)$$

Thus the first three moments of ${}_pW_n^2$ can be calculated using (6.6), (6.7) and (6.9). As can be seen from (6.9) the expression for the third moment is quite tedious and higher moments were not calculated.

6.2.2 Moments of ${}_rW_n^2$

We derive the moments of ${}_rW_n^2$ by first taking conditional expectations given $T_{(r)} = p$. Now the $T_{(1)}, T_{(2)}, \ldots T_{(r-1)} | T_{(r)} = p$ act like r-1 order statistics from a U(0,p) distribution:

$$f(t_{(1)},\ldots,t_{(r)}) = \frac{n!}{(n-r)!}(1-t_{(r)})^{n-r} \qquad 0 < t_{(1)} < t_{(2)} < \ldots, < t_{(r)} < 1$$

and $$f_r(t_{(r)}) = \frac{n!}{(r-1)!(n-r)!} t_{(r)}^{r-1}(1-t_{(r)})^{n-r}$$

so $$f(t_{(1)},\ldots,t_{(r-1)}|t_{(r)}) = (r-1)!\left(\frac{1}{t_{(r)}}\right)^{r-1}$$

Now $$\left({}_rW_n^2 \,|T_{(r)} = p\right) = n\int_0^p (F_n(t)-t)^2 dt$$

$$= np\int_0^1 (F_n(tp) - tp)^2 dt$$

But $F_n(tp)|T_{(r)} = p$ has the same distribution as $\frac{m}{n}G_m(t)$ where $m = r-1$ and $G_m(t)$ is the e.d.f. of m observations from U(0,1), so ${}_rW_n^2|T_{(r)} = p$ has the same distribution as

$$np\int_0^1 (\frac{m}{n}G_m(t) - tp)^2 dt$$

$$= \frac{p}{n}\int_0^1 (m\,G_m(t) - \frac{np}{m}mt)^2 dt.$$

Now $m\,G_m(t)$ is a Bi(t,m) random variable and so the moments of

$\frac{p}{n}\int_0^1 (m\,G_m(t) - \frac{np}{m}mt)^2$ dt can be found.

We find, with $m_{-i} = m(m-1)\dots(m-i)$,

$$E\left[{}_rW_n^2 \,\middle|\, T_{(r)} = p\right] = \frac{1}{n}\left[\frac{m}{6}(1+2m)p - \frac{2}{3}mnp^2 + \frac{n^2}{3}p^3\right]$$

$$E\left[\left({}_rW_n^2\right)^2 \,\middle|\, T_{(r)} = p\right] = \frac{2}{n^2}\Bigg[p^2\left(\frac{m_{-3}}{18} + m_{-2}\frac{13}{30} + m_{-1}\frac{17}{34} + \frac{m}{6}\right)$$

$$- n\,p^3\left(m_{-2}\frac{2}{9} + m_{-1}\frac{13}{15} + \frac{5}{12}m\right)$$

$$+ n^2p^4\left\{\frac{m_{-1}}{3} + m\frac{13}{30}\right\} - n^3p^5\frac{2m}{9} + \frac{n^4p^6}{18}\Bigg]$$

$$E\left[\left({}_rW_n^2\right)^3 \,\middle|\, T_{(r)} = p\right] = \frac{6}{n^3}\Bigg[p^3\Bigg(m_{-5}\frac{1}{162} + m_{-4}\frac{1}{60} +$$

$$m_{-3}\frac{1667}{2520} + m_{-2}\frac{913}{720} + m_{-1}\frac{2}{3} + \frac{m}{24}\Bigg)$$

$$np^4\left(\frac{m_{-4}}{27} + m_{-3}\frac{7}{15} + m_{-2}\frac{323}{210} + m_{-1}\frac{473}{360} + m\frac{3}{20}\right)$$

$$+ n^2p^5\left(m_{-3}\frac{5}{54} + m_{-2}\frac{7}{10} + m_{-1}\frac{917}{840} + m\frac{43}{180}\right)$$

$$- n^3p^6\left(m_{-2}\frac{10}{81} + m_{-1}\frac{7}{15} + m\frac{271}{1260}\right) + n^4p^7\left(m_{-1}\frac{5}{54} + m\frac{7}{60}\right)$$

$$- n^5p^8\frac{m}{27} + \frac{n^6p^9}{162}\Bigg]$$

We now take expectations with respect to $p = T_{(r)}$.

From Sahran and Greenberg (1962, p.14) we have

$$E\left[T_{(r)}^s\right] = \frac{n!\,(r+s-1)!}{(n+s)!\,(r-1)!} \tag{6.10}$$

and so we replace p^s by (6.10). We have explicitly

$$E\left[{}_rW_n^2\right] = \frac{1}{6}\,\frac{r-1}{n}\,\frac{r}{n+1} + \frac{(r-1)^2}{3n}\,\frac{r}{n+1}$$

$$-\frac{2}{3}\,\frac{r(r-1)(r+1)}{(n+1)(n+2)} + \frac{n}{3}\,\frac{r(r+1)(r+2)}{(n+1)(n+2)(n+3)}$$

and the other moments we leave in their implicit form.

An approximation to the distribution of the statistics used is that of Tiku's (1965), see (6.3). This is a fairly good approximation to the actual distribution (see Durbin and Knott (1972) who find the exact distribution of W_n^2 and compare it with this approximation given by Stephens and Maag (1968a), and show it to be good in the upper tail). Significance points are given in Table 6.1, for selected n, p and r.

§6.3 Exact lower-tail distribution of the statistic ${}_pW_n^2$

§6.3.1 Expressions for calculating ${}_pW_n^2$ and ${}_rW_n^2$ from the data

The statistics (6.1) and (6.2) are easily calculated from the data:

$${}_pW_n^2 = n \int_o^p \{F_n(t)-t)\}^2\,dt$$

$$= n \sum_{i=o}^{R-1} \left[t\frac{i^2}{n^2} - t^2\frac{i}{n} + \frac{t^3}{3}\right]_{t_{(i)}}^{t_{(i+1)}}$$

$$+ n\left[t\frac{R^2}{n^2} - t^2\frac{R}{n} + \frac{t^3}{3}\right]_{t_{(R)}}^{p}, \quad \text{where } t_{(R)} \le p < t_{(R+1)}$$

$$= \sum_{i=1}^{R} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 - \frac{R(4R^2-1)}{12n^2} + n\left(p\frac{R^2}{n^2} - p^2\frac{R}{n} + \frac{p^3}{3}\right) \qquad (6.11)$$

${}_rW_n^2$ is easily calculated; put $p = t_{(R)}$ and $R = r-1$ in the formula for ${}_pW_n^2$:

$${}_rW_n^2 = \sum_{i=1}^{r-1} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 - \frac{(r-1)[4(r-1)^2-1]}{12n^2}$$

$$+ n\left[t_{(r)}\frac{(r-1)^2}{n^2} - t_{(r)}^2\frac{r-1}{n} + \frac{t_{(r)}^3}{3}\right] \qquad (6.12)$$

If $p = 1$ and so $R = n$ in (6.11), ${}_pW_n^2$ then reduces to W_n^2

$${}_{p=1}W_n^2 = \sum_{i=1}^{n} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}$$

Obviously if $r = n$, ${}_rW_n^2$ does not reduce to W_n^2, unless $t_{(r)} = 1$, which occurs with probability 0 for finite n.

6.3.2 Maximum and Minimum values of ${}_pW^2$

It is easily seen that if all the observations are equal to 0 then $F_n(t) = 1 \quad 0 < t \leq p$ and

$$ {}_pW_n^2 = \frac{n}{3}\{1 - (1-p)^3\} \tag{6.13} $$

This gives us the maximum value of ${}_pW_n^2$. The minimum value of ${}_pW_n^2$ is found by minimizing the term

$$ \sum_{i=1}^{R} (t_{(i)} - \frac{2i-1}{2n})^2 $$

This is done by putting $t_{(i)} = \frac{2i-1}{2n}$ and then this term is equal to 0. We then minimize the remaining part of ${}_pW_n^2$ with respect to p. This is done if $p = \frac{R}{n}$ and this value of p is compatible with $t_{(R)} = \frac{2R-1}{2n} = \frac{R}{n} - \frac{1}{2n}$. The value of ${}_pW_n^2$ is then

$$ {}_pW_n^2 = \frac{R}{12n^2} \tag{6.14} $$

$$ = \frac{p}{12n} $$

These formulations help to find the exact distribution of the statistic ${}_pW_n^2$ in the lower tail.

6.3.3 Some previous results of Stephens and Maag

Stephens and Maag (1968a) find the lower tail distribution of the statistic W_n^2. Some of their results are given here.

If P is a random point in R_n and has sample space $0 \leq t_1 \leq \cdots \leq t_n \leq 1$, then $Pr(P \in \delta_v) = \delta_v n!$ Then $Pr(W_n^2 \leq z) = n!V$ where V is the volume of a sphere S inside the simplex T given by $\{0 \leq t_1 \leq \cdots \leq t_n \leq 1\}$. If $a = 1/2n$,

$y = \sqrt{z-\frac{1}{12n}}$, the sphere S has centre $(a, 3a, \ldots, (2n-1)a)$ and radius y. For small enough Z, S is completely inside the simpl(T then

$$\Pr(W_n^2 \leq z) = n!\ \pi^{\frac{n}{2}} y^n / \Gamma(\tfrac{n}{2} + 1) \qquad (6.15)$$

where (6.15) is the volume of a hypersphere (see M.G. Kendall (1961) p. 35, for example). The limits for z are

$$\frac{1}{12n} \leq z \leq \frac{n+3}{12n^2}$$

The lower limit is the minimum value of W_n^2 and the upper value is given by the largest value of y before the sphere S goes outside the simplex T; this occurs at the boundaries $t_1 = 0$ and $t_n = 1$ when $y = a$, i.e. $z = \frac{n+3}{12n^2}$.

6.3.4 Stephens' and Maag's result applied to ${}_pW_n^2$

Now consider the statistic ${}_pW_n^2$, where $pn = k$ i.e. $k \leq pn < k+1$, k an integer, $0 < k < n$. Now ${}_pW_n^2$ if $R = k$ is given by

$${}_pW_n^2 = \sum_{i=1}^{k} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{p}{12n} \qquad (6.16)$$

If $R = k - j$,

$${}_pW_n^2 = \sum_{i=1}^{k-j} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{p}{12n} + \frac{4j^3 - j}{12n^2} \qquad (6.17)$$

and for $R = k + j$, we change the sign of j in (6.17).

For $R = k$ consider the hypersphere S in R_k with radius $y = (z-p/12n)^{\frac{1}{2}}$ and centre $(a, 3a, \ldots, (2k-1)a)$, then

$$\Pr({}_pW_n^2 \leq z \mid R = k) = \frac{k!\,\pi^{k/2} y^k}{\Gamma(k/2+1)}$$

and $0 < y < \frac{1}{2n}$, i.e. $\frac{p}{12n} < z < \frac{3+pn}{12n^2}$.

It can be seen that for a similar formula to hold for $R = k - j$

$$\frac{p}{12n} + \frac{4j^3 - j}{12n^2} < z < \frac{p}{12n} + \frac{1}{4n^2} + \frac{4j^3 - j}{12n^2}$$

For $j > 0$ $4j^3 - j$ is an increasing function of j and it is seen for $j = 1$, z must satisfy $z > \frac{p+3}{12n^2}$, so if $R < k$, ${}_pW_n^2$ cannot attain a value in S.

Now when $R = k + 1$ the minimum value of ${}_pW_n^2$ is $\frac{p}{12n}$, and this is achieved when $t_{(i)} = \frac{2i-1}{2n}$ $i = 1,\ldots,k$ and $t_{(k+1)} = p$. Consider now a hypersphere S' in R_{k+1} with centre $(a, 3a, \ldots, (2k-1)a, (2k+1)a)$ and radius $y = (z-p/12n)^{\frac{1}{2}}$, with a restriction on the t_{k+1} coordinate, $t_{k+1} \leq p$. So we want the volume of hypersphere S' but with $t_{k+1} \leq p$ as y increases. The volume of the hypersphere with no restrictions is:

$$\frac{(k+1)!\ \pi^{\frac{k+1}{2}}\ y^{k+1}}{\Gamma(\frac{K+1}{2}+1)}$$

Using a result of Stephen's and Magg, with slight modification, the volume of the hypersphere with $t_{k+1} \leq p$ is given by

$$(k+1)!\ \pi^{\frac{k+1}{2}}\ y^{k+1} \int_o^{\alpha} \frac{\sin^{k+1} t\, dt}{\Gamma(\frac{k+1}{2}+\frac{1}{2})}$$

where $\alpha = \cos^{-1}(1/2ay)$

For $R > k + 1$, the minumum value of ${}_pW_n^2$ exceeds $\frac{pn+6}{12n^2}$. The hypersphere S' will meet more boundaries when $y > 1/2n$ i.e. $z > \frac{3+pn}{12n^2}$. So

$$\Pr({}_pW_n^2 \leq z | R = k+1) = (k+1)!\ \pi^{\frac{k+1}{2}}\ y^{k+1}$$

$$\times \frac{\int_o^{\alpha} \sin^{k+1} t\, dt}{\Gamma(\frac{k+1}{2}+\frac{1}{2})}$$

for $\frac{pn}{12n^2} \leq z \leq \frac{3+pn}{12n^2}$

So finally we have

$$\Pr({}_pW_n^2 \le z) = \Pr({}_pW_n^2 \le z | R = k)\Pr(R = k)$$

$$+ \Pr({}_pW_n^2 \le z | R = k+1)\Pr(R = k+1)$$

$$= \frac{k!\pi^{\frac{k}{2}} y^k}{\Gamma(\frac{k}{2}+1)} \binom{n}{k} p^k (1-p)^{n-k}$$

$$+ \frac{(k+1)!\pi^{\frac{k+1}{2}} y^{k+1}}{\Gamma(\frac{k+1}{2}+\frac{1}{2})} \int_0^\alpha \sin^{(k+1)} t \, dt$$

$$\times \binom{n}{k+1} p^{k+1} (1-p)^{n-k-1} \qquad (6.18)$$

where $y = (z-p/12n)^{\frac{1}{2}}$

$\alpha = \cos^{-1}(1/2ay), \qquad np = k.$

Now consider the next range of values for z. This is when the hypersphere S only goes outside the simplex at two boundaries viz. $t_1 = 0$ and $t_k = p$.

Following Stephens and Maag (1968a), it is seen that for R = k, the hypersphere S touches the boundaries of the simplex T, except at $t_1 = 0$ and $t_k = p$, if $y < \sqrt{2}/2n$, i.e. if $z < \frac{6+pn}{12n^2}$. Thus

$$\Pr({}_pW_n^2 \le z | R = k) = k!\pi^{\frac{1}{2}(k-1)} y^k$$

$$\times \left\{ \frac{\sqrt{2}}{\Gamma(\frac{k}{2}+1)} - 2\frac{\int_0^\alpha \sin^k t \, dt}{\Gamma(\frac{k}{2}+\frac{1}{2})} \right\} \qquad (6.20)$$

for $\frac{3+pn}{12n^2} \le z \le \frac{6+pn}{12n^2}$. The expression (6.20) is given basically by Stephens and Maag (1968a), where $\alpha = \cos^{-1}(1/2ay)$, $y = \sqrt{z - p/12n}$.

If $R = k - j$, the minimum value of ${}_pW_n^2$ is $\frac{pn+4^2_j - j}{12n^2}$ and so for $j > 1$, ${}_pW_n^2$ lies outside the range of z in (6.20).

For $j = 1$, the minimum value of ${}_pW_n^2$ is $\frac{pn+3}{12n^2}$ and then

$$ {}_pW_n^2 = \sum_{i=1}^{k-1} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{pn+3}{12n^2} $$

Now consider they hypersphere S'', in R_{k-1} space, with radius $y'' = (z-(pn+3)/12n^2)^{\frac{1}{2}}$ and centre $(a, 3a, \ldots, (2k-3)a)$. The first boundary the hypersphere S'' will meet as y'' increase inside the simplex $T'' = \{0 \le t_1 \le \ldots \le t_{k-1} \le p\}$ is at $t_1 = 0$. This will occur if $y'' > \frac{1}{2n}$, i.e. if $z > \frac{pn+6}{12n^2}$. S'' will meet other boundaries if $y'' > \sqrt{2}/2n$, i.e. if $z > \frac{pn+9}{12n^2}$. So for $\frac{pn+3}{12n^2} < z < \frac{pn+9}{12n^2}$ the hypersphere S'' lies completely inside the simplex T''. Thus

$$ \Pr({}_pW_n^2 \le z | R = k-1) = \frac{(k-1)!\,\pi^{\frac{k-1}{2}}\, y''^{k-1}}{\Gamma(\frac{k}{2}-\frac{1}{2})} \qquad (6.21) $$

where $y'' = (z-(pn+3)/12n^2)^{\frac{1}{2}}$ and

for $\frac{pn+3}{12n^2} < z < \frac{pn+9}{12n^2}$.

Now we again consider the hypersphere S' when $R = k+1$. We noted before then $R > k+1$ the minimum value of ${}_pW_n^2$ exceeds $\frac{pn+6}{12n^2}$. We now consider the volume of S' with $t_{k+1} \le p$ and now is restricted by the simplex at the boundaries $t_1 = 0$ and $t_k = p$ with $y' > \frac{1}{2n}$ and $y' < \sqrt{2}/2n$. This basically is the same situation as for $R = k$ with hypersphere

Now the total volume is

$$ (k+1)!\,\pi^{\frac{k+1}{2}}\, y^{k+1} \frac{\int_0^\alpha \sin^{k+1} t\, dt}{\Gamma(\frac{k+1}{2}+\frac{1}{2})} $$

$$ \times \left[1 - 2\int_0^\alpha \sin^{k+1} t\, dt \times \frac{\Gamma(\frac{k+1}{2}+1)}{\Gamma(\frac{k+1}{2}+\frac{1}{2})\pi^{\frac{1}{2}}}\right] \qquad (6.21A) $$

giving $\Pr({}_pW_n^2 \leq z | R = k+1)$.

For $\frac{pn+3}{12n^2} \leq z \leq \frac{pn+6}{12n^2}$, we can combine results (6.20), (6.21) and (6.21A) in the way (6.18) was found, to obtain $\Pr({}_pW_n^2 \leq z)$.

Thus the lower-tail distribution of ${}_pW_n^2$ could be found using (6.18) and (6.20), (6.21) and (6.21A).

6.3.5 <u>The statistics ${}_{q,p}W_n^2$ and ${}_{s,r}W_n^2$</u>

The small sample statistics corresponding to ${}_{q,p}W^2$ are ${}_{q,p}W_n^2$, ${}_{s,r}W_n^2$ and ${}_{s,p}W_n^2$. ${}_{q,p}W_n^2$ is the same as ${}_pW_n^2$ but we put the lower limit of the integral in (6.1) equal $F_o^{-1}(q)$. ${}_{s,r}W_n^2$ corresponds to ${}_rW_n^2$ with the lower limit of the integral put equal to $x_{(s)}$, in (6.2). ${}_{s,p}W_n^2$ is the statistic for mixed censoring with the lower limit equal $x_{(s)}$ and the upper limit equal to $F_o^{-1}(p)$. We will not consider this statistic further.

The statistics ${}_{q,p}W_n^2$ and ${}_{s,r}W_n^2$ can be calculated from the data

$$
\begin{aligned}
{}_{q,p}W_n^2 &= {}_pW_n^2 - {}_qW_n^2 \\
&= n \sum_{i=S}^{R} \left(t_{(i)} - \frac{2i-1}{2n}\right)^2 - \frac{1}{12n}\left[R(4R^2-1)-(S-1)(4(S-1)^2-1)\right] \\
&\quad + n\left(p\frac{R^2}{n^2} - p^2\frac{R}{n} + \frac{p^3}{3}\right) - n\left(q\frac{S^2}{n^2} - q^2\frac{S}{n} + \frac{q^3}{3}\right)
\end{aligned}
$$

where the observations $q \leq t_{(S)} \leq t_{(S+1)} \leq \cdots \leq t_{(R)} \leq p$ are available and the rest censored. ${}_{r,s}W_n^2$ is calculated by substituting $R = r-1$, $S = s+1$, $p = t_{(r)}$ and $q = t_{(s)}$ in the above formula.

The first moment of ${}_{q,p}W_n^2$ can be calculated from

$$_{q,p}W_n^2 = {}_pW_n^2 - {}_qW_n^2$$

and other moments in a similar manner to the way the moments of ${}_pW_n^2$ were calculated. Similarly the moments of ${}_{s,r}W_n^2$ can be calculated using

$$E[{}_{s,r}W_n^2] = E[E[{}_{q,p}W_n^2 | T_{(s)} = q, \; T_{(r)} = p]]$$

and use a result quoted in David (1970) (§3.1):

$$E[T_{(r)}^a T_{(s)}^b] = \{(r-1+a)!(s-1+a+b)!n!\}$$
$$\times \{(r-1)!(s-1+a)!(n+a+b)!\}^{-1}$$

6.4 Theory of the asymptotic distribution of the statistics ${}_rW_n^2$, ${}_pW_n^2$ and ${}_{q,p}W_n^2$

6.4.1. Previous results for W_n^2

It is well known that the empirical process

$$y_n(t) = \sqrt{n}\left\{F_n(t)-t\right\}$$

converges to the Gaussian process y(t) with mean zero, and covariance function given by $\rho(s,t)$, where

$$\rho(s,t) = E[y(s)\ y(t)] = \min(s,t) - st \qquad 0 \leq s,t \leq 1$$

The statistic W_n^2 is shown, by Anderson and Darling (1952), to converge to W^2 where

$$W^2 = \int_0^1 y^2(t)\ dt$$

Anderson and Darling represent the statistic W^2 by a weighted infinite sum of chi-squared random variables with 1 degree of freedom:

$$W^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{\pi^2 j^2}$$

where the Z_j are i.i.d. standard normal random variables. This follows from the form of the characteristic function given by Smirnov (1936)

$$\lim_{n\to\infty} E\left[\exp(itW_n^2)\right] = \left[\frac{(2it)^{\frac{1}{2}}}{\sin\{2it)\}^{\frac{1}{2}}}\right]^{\frac{1}{2}}$$

This representation can also be derived by using the result of Anderson and Darling (1952) that

$$y(t) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} Z_j f_j(t) \qquad (6.23)$$

where the λ's and f's are the eigenvalues and orthonormal eigenfunctions respectively of the integral equation:

$$\lambda \int_0^1 \rho(s,t) f(s)\, ds = f(t) \qquad (6.24)$$

where $\rho(s,t) = \min(s,t) - st \quad 0 \le s,t \le 1$

They show that

$$\rho(s,t) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} f_j(t) f_j(s) \qquad (6.25)$$

and then the characteristic function of W^2, $\phi(t)$, can be expressed in terms of the λ_j:

$$\phi(t) = \prod_{j=1}^{\infty} \left(1 - 2it/\lambda_j\right)^{-\frac{1}{2}}$$

and so the cumulants can be calculated from the λ_j:

$$\kappa_m = 2^{m-1} (m-1)! \sum_{j=1}^{\infty} \left(\frac{1}{\lambda_j}\right)^m, \quad m = 1, 2., , ,$$

Or alternatively, the cumulants of ${}_pW^2$ can be found from $\rho(s,t)$ only:

$$\kappa_m = 2^{m-1} (m-1)! \int_0^1 \rho_m(s,s)\, ds$$

where $\rho_m(s,t) = \int_0^1 \rho_{m-1}(s,u)\, \rho(u,t)\, du$

$$\rho_1(s,t) = \rho(s,t)$$

6.4.2. Previous results applied to ${}_pW^2$

Firstly, it is noted that since Type I and Typed II censoring become equivalent as $n \to \infty$, the statistics ${}_pW_n^2$ and ${}_rW_n^2$ become equivalent and have the same distribution provided $r/n \longrightarrow p$. We denote by ${}_pW^2$, the asymptotic distribution of ${}_pW_n^2$ or ${}_rW_n^2$.

Following Kac and Siegert (1947) and Darling (1957), the asymptotic distribution of ${}_pW^2$, where

$$ {}_pW^2 = \int_0^p y^2(t)\, dt $$

is that of $$\sum_{j=1}^{\infty} \frac{Z_j^2}{\mu_j} \qquad (6.26)$$

where the Z_j are i.i.d. standard normal random variables, the μ_j are eigenvalues of the integral equation:

$$ \mu \int_0^p \rho(s,t)\ g(s)\ ds = g(t) \qquad (6.27) $$

which reduces to

$$ \mu\ g(t) + g''(t) = 0 \qquad 0 \leq t \leq p \qquad (6.28) $$

on differentiation.

So by using Mercer's Theorem

$$ \rho(s,t) = \sum_{j=1}^{\infty} \frac{1}{\mu_j}\ g_j(s)\ g_j(t) \qquad 0 \leq s,t \leq p $$

where the μ_j are eigenvalues and the $g_j(t)$ are orthonormal eigenfunctions of the integral equation (6.27).

Consequently ${}_pW^2$ is a weighted sum of χ_1^2 random variables and its characteristic function is given by

$$ \phi(t) = \prod_{j=1}^{\infty} (1-2it/\mu_j)^{-\frac{1}{2}} \qquad (6.30) $$

and the cumulants of ${}_pW^2$ given by

$$\kappa_m = 2^{m-1}\,(m-1)!\;\sum_{j=1}^{\infty}\left(\frac{1}{\mu_j}\right)^m \qquad 6.31A$$

or by $$\kappa_m = 2^{m-1}\,(m-1)!\;\int_0^p \rho_m(s,s)\,ds \qquad 6.31B \qquad (6.31)$$

where $\rho_m(s,t) = \int_0^p \rho_{m-1}(s,u)\,\rho(u,t)\,du$

$$\rho_1(s,t) = \rho(s,t)$$

6.4.2. Solution of the integral equation for ${}_pW^2$

We can solve the integral equation (6.27) or alternatively the differential equation (6.28)

$$\mu\, g(t) + g''(t) = 0 \qquad 0 \le t \le p < 1$$

A general solution of the differential equation is:

$$g(t) = Ae^{it\mu^{\frac{1}{2}}} + Be^{-it\mu^{\frac{1}{2}}}$$

$$= (A+B)\cos t\mu^{\frac{1}{2}} + i(A-B)\sin t\mu^{\frac{1}{2}}.$$

The boundary conditions are that $g(o) = 0$ thus

$$A + B = 0$$

and the solution reduces to:

$$g(t) = a \sin t\mu^{\frac{1}{2}}$$

To find the eigenvalues μ, substitute the solution $g(t)$ into the original integral equation:

$$g(t) = \mu \int_0^p \{\min(s,t) - st\}\, g(s)\, ds$$

Now writing $\mu = m^2$:

$$m^2 \int_0^p \{\min(s,t) - st\} \sin sm\, ds$$

$$= m^2 \int_0^t s\sin(sm) + m^2 t \int_t^p \sin(sm)\, ds - m^2 t \int_0^p s\sin(sm)\, ds$$

$$= \Big[-ms\cos(ms) + \sin(ms)\Big]_o^t + mt\Big[-\cos(mt)\Big]_t^p$$

$$-t\Big[-ms\cos(ms) + \sin(ms)\Big]_o^p$$

$$= \sin(mt) + t\Big\{-m\cos(mp) + mp\cos(mp) - \sin(mp)\Big\}$$

For the function $g(t) = \sin t\mu^{\frac{1}{2}}$ to be an eigenfunction:

$$-m\cos mp + mp\cos mp - \sin mp = 0$$

Now $g(p) \neq 0$ since $\rho(p,p) \neq 0$ for $p < 1$, consequently $\sin mp = 0$ is not a solution for m. Consequently $\cos mp = 0$ is also not a solution, since this implies that $\sin mp = 0$, so we have that $\sin mp \neq 0$ and $\cos mp \neq 0$, so the general solution is given by

$$\tan mp = -m(1-p)$$

$$\tan \mu^{\frac{1}{2}} p = -\mu^{\frac{1}{2}}(1-p)$$

Now since the roots μ are positive, $\mu^{\frac{1}{2}}p$ must lie in the second or fourth quadrant of the circle, i.e.

$$\mu^{\frac{1}{2}}p = 2k\pi - \delta_k \quad \text{where } 0 < \delta_k < \frac{\pi}{2}$$

$$\text{so} \qquad \mu = \left(\frac{2k\pi - \delta_k}{p}\right)^2$$

for $k = 1,2,\ldots$.

Thus the normalized eigenfunctions are

$$g(t) = \sqrt{2}\sin\mu^{\frac{1}{2}}t \Big/ \left(p - \frac{\sin\mu^{\frac{1}{2}}p\cos\mu^{\frac{1}{2}}p}{\mu^{\frac{1}{2}}}\right)^{\frac{1}{2}} \qquad (6.32)$$

To check that the $g_k(t)$ $k = 1,2,\ldots$ are orthogonal, where

$$g_k(t) = C_k \sin(\mu_k^{\frac{1}{2}} t)$$

$$\mu_k = \left(\frac{2k\pi - \delta_k}{p}\right)^2$$

i.e. μ_k is a solution of

$$\tan \mu_k^{\frac{1}{2}} p = -\mu_k^{\frac{1}{2}}(1-p) \qquad (6.32A)$$

for $a \neq b$ and $a,b \neq 0$:

$$\int_o^p \sin(at)\sin(bt)\,dt = \tfrac{1}{2}\left[\frac{\sin(a-b)t}{a-b} - \frac{\sin(a+b)t}{a+b}\right]_o^p$$

$$= \frac{-1}{a^2-b^2}\left[a\sin(bp)\cos(ap) - b\sin(ap)\cos(bp)\right]$$

$$= \frac{1}{(a^2-b^2)\cos(ap)\cos(bp)}\left[a\tan bp - b\tan ap\right]$$

$$= \frac{1}{(a^2-b^2)\cos(ap)\cos(bp)at}\left[\frac{\tan bp}{b} - \frac{\tan ap}{a}\right].$$

But $g_k(t) \alpha \sin\mu_k^{\frac{1}{2}} t$ and $\dfrac{\tan \mu_k^{\frac{1}{2}} p}{\mu k^{\frac{1}{2}}} = -(1-p)$

thus $\int_o^p g_k(t)\, g_j(t)\, dt = 0$ for $j \neq k$,

putting $a = \mu_k^{\frac{1}{2}}$, $b = \mu_j^{\frac{1}{2}}$.

6.4.3. Solution of the integral equation for doubly censored data, ${}_{q,p}W^2$.

We can extend these results to the case of doubly censored data and solve the integral equation:

$$g(t) = \mu \int_q^p \left\{\min(s,t) - st\right\} g(s)\,ds \quad 0 < q < p < 1 \qquad (6.33)$$

This again becomes on differentiation:

$$\mu g(t) + g''(t) = 0$$

with the general solution:

$$g(t) = a \cos(\mu^{\frac{1}{2}} t) + b \sin(\mu^{\frac{1}{2}} t)$$

For g(t) to satisfy the integral equation (6.33), we find the conditions on a, b and μ. Writing $\mu = m^2$,

$$m^2 \int_q^p \left\{\min(s,t) - st\right\} \cos(ms)\, ds$$

$$= \Big[ms \sin(ms) + \cos(ms)\Big]_q^t + mt\Big[\sin mt\Big]_t^p$$

$$-t\Big[ms \sin(ms) + \cos(ms)\Big]_q^p$$

$$= \cos(mt) - mq \sin(mq) - \cos 9mq)$$

$$+ t\Big\{m \sin mp - mp \sin mp + mq \sin mq$$

$$+ \cos mq - \cos mp\Big\}.$$

$$\equiv \cos mt + G_1 + tH_1$$

where $G_1 = -mq \sin(mq) - \cos mq$ and $H_1 = m \sin(mp) - mp \sin mp$
$+ mq \sin mq + \cos mq - \cos mp.$

Also

$$m^2 \int_q^p \left\{\min(s,t) - st\right\} \sin(sm)\, ds$$

$$= \Big[-ms \cos ms + \sin ms\Big]_q^t + mt\Big[-\cos mt\Big]_t^p$$

$$-t\Big[-ms \cos + \sin ms\Big]_q^p$$

$$= \sin mt + mq \cos mq - \sin mq$$

$$+ t\{-m \cos mp + mp \cos mp - \sin mp$$

$$-mq \cos mq + \sin mq\}$$

$$\equiv \sin(mt) + G_2 + t H_2$$

where G_2 and H_2 are defined similarly.

Thus for g(t) to be an eigenfunction

$$F(t) \equiv aG_1 + at\, H_1 + bG_2 + bt\, H_2 = 0 \quad \text{for all } t: \ q < t < p$$

If a and b are both non-zero then

$$aG_1 + bG_2 = 0$$

and $$aH_1 + bH_2 = 0.$$

Now $$\frac{a}{b} = \frac{-G_2}{G_1} = \frac{-H_2}{H_1}$$

i.e. $G_1 H_2 - G_2 H_1 = 0$

Writing $\sin mp = s_p$, $\sin mq = s_q$, $\cos mp = c_p$, $\cos mq = c_q$

this reduces to

$$(-m^2 q\ s_q\ c_p + m^2 pq\ s_q\ c_p - s_p\ mq\ s_q - m^2 q^2\ c_q\ s_q$$

$$+ mq\ s_q^2 - m\ c_p c_q + mp\ c_p c_q - s_p c_q - mq\ c_q^2 + s_q c_q)$$

$$+ (m^2 q\ c_q\ s_p - m^2 q^2 s_q c_q + c_q^2 mq - c_p\ mq\ c_q$$

$$- ms_q\ s_p + m_p\ s_q s_p - mqs^2 - c_q s_q + c_p s_q)$$

$$= 0$$

On simplification this becomes

$$\tan (p-q)m = \frac{m(p-1-q)}{1+m^2 pq - m^2 q} \qquad (6.34)$$

If $q = 0$, this reduces to the previous result

$$\tan pm = m(p-1)$$

and if $p = 1$

$$\tan(1-q)m = -mq$$

which is symmetric in p and 1 - q.

Therefore a general eigenfunction is

$$g(t) = a\cos(mt) + b\sin(mt)$$

such that

$$\tan(p-q)m = \frac{-m(1+q-p)}{1-m^2 q(1-p)} \qquad (6.35)$$

Now

$$\frac{a}{b} = \frac{-G_2}{G_1} = \frac{-H_2}{H_1}$$

$$= \frac{mq - \tan mq}{mq \tan(mq)+1}$$

so the normalising factor is given by:

$$\int_q^p \left\{\frac{a}{b}\cos(mt) + \sin(mt)\right\}^2 dt$$

$$= \frac{a^2}{2b^2}\left[p-q+\frac{1}{m}\sin(p-q)m\right]$$

$$+\frac{1}{2}\left[p-q-\frac{1}{m}\sin(p-q)m\right]$$

$$-\frac{a}{2bm}\left[\cos 2pm - \cos 2qm\right]$$

6.4.3 ${}_pW^2$ as a quadratic form of normal random variables

Alternatively ${}_pW^2$ can be expressed in terms of the eigenfunctions and eigenvalues of (6.23) and (6.24). It is well known that $\lambda_j = \pi^2 j^2$ and $f_j(t) = \sqrt{2}\sin j\pi t$ (see Anderson and Darling (1952)) so

$$ {}_pW^2 = \int_0^p y^2(t)\,dt $$

$$ = \int_0^p \left(\sum_{j=1}^{\infty} \frac{\sqrt{2}}{j\pi} \sin j\pi t\, Z_j\right)^2 dt $$

$$ = \sum_{j=1}^{\infty} \frac{1}{j^2\pi^2}\left(p - \frac{\sin 2j\pi p}{2j\pi}\right) Z_j^2 $$

$$ + \sum_{j\neq k}^{\infty}\sum^{\infty} \frac{1}{j\pi . k\pi}\left\{\frac{\sin(j-k)\pi p}{(j-k)\pi} - \frac{\sin(j+k)\pi p}{(j+k)\pi}\right\} Z_j Z_k $$

Consequently ${}_pW^2$ can be considered as a quadratic form of infinite dimension:

$$ {}_pW^2 = \underset{\sim}{Z}^T A \underset{\sim}{Z} \qquad (6.36) $$

where $\underset{\sim}{Z} = (Z_1, Z_2, \ldots)^T$ and

$$ (A)_{jj} = \frac{1}{j^2\pi^2}\left(p - \frac{\sin(2j\pi p)}{2j\pi}\right) $$

6.36A

$$ (A)_{jk} = \frac{1}{j\pi . k\pi}\left\{\frac{\sin(j-k)\pi p}{(j-k)\pi} - \frac{\sin(j+k)\pi p}{(j+k)\pi}\right\} $$

$j, k = 1, 2, \ldots$

The characteristic function of ${}_pW^2$ is then given by

$$ \phi(t) = |I - 2it A|^{-\frac{1}{2}} $$

where I is the infinite identity matrix (for example see Hogg and Craig (1970). Consequently if $\alpha_1, \alpha_2, \ldots,$ are the eigenvalues of A, ${}_pW^2$ has characteristic function $\prod_{j=1}^{\infty}(1-2it\alpha_j)^{-\frac{1}{2}}$, and then the cumulants of ${}_pW^2$ are given by

$$\kappa_m = 2^{m-1}(m-1)!\sum_{j=1}^{\infty}\alpha_j^m \qquad m = 1,2,\ldots \tag{6.37}$$

The $\alpha_1^{-1},\alpha_2^{-1},\ldots$ are identically equal to the $\mu_1,\mu_2,\ldots$ of (6.26). The eigenfunctions g(t) can be found in the form of sine Fourier series, by making use of the fact that the eigenfunctions are orthogonal. Consider the infinite vector

$$\underset{\sim}{f}(t) = (f_1(t), f_2(t), \ldots)^T$$

where $$f_j(t) = \sqrt{2}\sin(\pi j t) \qquad j = 1,2,\ldots$$

$f_j(t)$ is an eigenfunction of (6.24) with eigenvalue $\lambda_j = \pi^2 j^2$. The $f_j(t)$ $(j = 1,2,\ldots)$ are known to be orthogonal, i.e.

$$\int_0^1 \underset{\sim}{f}(t)\,\underset{\sim}{f}^T(t)\,dt = I$$

where I is the infinite identity matrix.

Now let

$$\int_0^p \underset{\sim}{f}(t)\,\underset{\sim}{f}^T(t)\,dt = B$$

then it is wished to find $\underset{\sim}{g}(t)$ such that

$$\int_0^p \underset{\sim}{g}(t)\,\underset{\sim}{g}^T(t)\,dt = I$$

and

$$\rho(s,t) = \sum_{j=1}^{\infty}\frac{1}{\mu_j}g_j(s)\,g_j(t)$$

$$= \underset{\sim}{g}^T(s)\,M^{-1}\underset{\sim}{g}(t)$$ where M is a diagonal matrix with $(M)_{jj} = \mu_j$, $j = 1,2,\ldots$.

By methods of linear algebra if $B = PDP^T$ where $P^TP = I$ and D is a diagonal matrix with entries the eigenvalues of B, then

$$g(t) = D^{-\frac{1}{2}} P^T f(t) \qquad (6.39)$$

provides an orthogonal vector:

$$\int_0^p g(t)\, g^T(t)\, dt = \int_0^p D^{-\frac{1}{2}} P^T\, f(t)\, f'(t)\, P D^{-\frac{1}{2}}\, dt$$

$$= D^{-\frac{1}{2}} P^T B P D^{-\frac{1}{2}}$$

$$= D^{-\frac{1}{2}} D D^{-\frac{1}{2}}$$

$$= I$$

The eigenvalues corresponding to $g(t)$ can be found since

$$\rho(s,t) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} f_j(s)\, f_j(t)$$

$$= f^T(s)\, \Lambda^{-1}\, f(t)$$

where Λ is a diagonal matrix with entries $(\Lambda)_{jj} = \lambda_j$ $(= \pi^2 j^2)$.

So $$f^T(s)\, \Lambda^{-1}\, f(t) = g^T(s) M^{-1}\, g(t)$$

$$= f^T(s) P D^{-\frac{1}{2}} M^{-1} D^{-\frac{1}{2}} P^T\, f(t) \qquad 0 \le s,\ t \le p$$

thus $$\Lambda^{-1} = P\, D^{-\frac{1}{2}}\, M^{-1} D^{-\frac{1}{2}} P^T$$

ie. $$M = D^{-\frac{1}{2}} P^T \Lambda\, P\, D^{-\frac{1}{2}} \qquad (6.40)$$

This agrees with the previous result since $A = \Lambda^{-\frac{1}{2}} B\, \Lambda^{-\frac{1}{2}}$.

Thus the eigenvalues μ and the eigenfunctions $g(t)$ corresponding to the integral equation (6.27) can be found giving the representation

$$y_p(t) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\mu_j}} Z_j\, g_j(t) \qquad 0 \le t \le p$$

and hence $${}_pW^2 = \sum_{j=1}^{\infty} \frac{1}{\mu_j} Z_j^2 .$$

Obviously it is easier to find the eigenvalues of the matrix A (6.36). This method leads to approximate eigenvalues, which are close to the exact ones found by solving

$$\tan \mu_p^{\frac{1}{2}} = \mu^{\frac{1}{2}}(1-p)$$

6.5 Moments and percentage points of ${}_pW^2$ and ${}_{q,p}W^2$

6.5.1 Moments of ${}_pW^2$

The first three moments of ${}_pW^2$ were found from the formulae for the moments of ${}_pW^2_n$, be letting $n \to \infty$. The moments are then given by:

$$E[{}_pW^2] = \frac{p^2}{6}(3-2p) \qquad (6.41)$$

$$E[({}_pW^2)^2] = 2p^4\left\{\frac{7}{24} - p\left(\frac{13}{30} - \frac{p}{6}\right)\right\}$$

$$E[({}_pW^2)^3] = 6p^6\left[\frac{23}{144} + \frac{1}{30} + p\left\{-\frac{25}{56} + p \times \left(\frac{7}{20} - \frac{5p}{54}\right)\right\}\right]$$

The central moments derived using these formulae are given in Table 6.2.

The alternative method to find higher moments of ${}_pW^2$ is to use the eigenvalues associated with the integral equation (6.24), and use (6.31).

The eigenvalues are solutions of

$$\tan(\mu^{\frac{1}{2}}p) = -\mu^{\frac{1}{2}}(1-p) \qquad (6.41A)$$

This equation was solved using the Newton-Raphson method of calculating roots with a tolerance of 10^{-5} in subsequent iterations. The eigenvalues, or more exactly the reciprocals of the roots of (6.41A) are given in Table 6.3. κ_4 was calculated using them and is given with the smaller cumulants in Table 6.4. It is seen, comparing the approximations to the first three cumulants with the exact values given in Table 6.2, that they converge to the exact values, as the cumulants become higher. The approximation to κ_3 differs by no more than 1 figure in the sixth significant figure from the exact value, and this only for the values of p near to .5. Thus using formula (6.31):

$$\kappa_m = 2^{m-1}(m-1)! \sum_{j=1}^{\infty} \left(\frac{1}{\mu_j}\right)^m$$

all the higher cumulants of ${}_pW^2$ can be found, using the known eigenvalues, to a very good approximation.

Using the matrix approach to find the eigenvalues of the matrix A defined in (6.36A), good approximations to the eigenvalues can be found by solving (6.41A) for the dimension of A as small as 20. With dimension 45 the eigenvalues of A agree with the solutions of (6.41A) with occasional difference of one digit in the sixth decimal place. For ${}_pW^2$, the matrix approach provides a quick way of generating the eigenvalues, see Table 6.3.

6.5.2 Imhof's method to approximate to the distribution of ${}_pW^2$

We use the exact eigenvalues found from solving (6.41A) to make a good approximation to the distribution of ${}_pW^2$ over its whole range, by a method due to Imhof and developed by Durbin (1970) and Durbin and Knott (1972). The method is to approximate to the distribution of ${}_pW^2$ which has the same distribution as

$$\sum_{j=1}^{\infty} \frac{z_j^2}{\mu_j}$$

where the Z_j are i.i.d. standard normal random variables, by the random variable H,

$$H = \sum_{i=1}^{K} \frac{Z_i^2}{\mu_i} + aC_v^2$$

where K is about 20, μ_i are the eigenvalues obtained from solving the integral equation and tabulated in Table 6.3, C_v^2 is a χ^2 random variable with 'v' degrees of freedom and 'a' is scalar. 'a' and 'v' are chosen so that ${}_pW^2$ and H have the same mean and variance

$$E[H] = \sum_{i=1}^{K} \frac{1}{\mu_i} + av$$

$$\text{var}(H) = 2 \sum_{i=1}^{K} \mu_i^{-2} + 2a^2 v \tag{6.42}$$

Then Imhof (1961) gives the result

$$\Pr(H \leq x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\sin\{\theta(u)\}}{u\rho(u)} du$$

where

$$\theta(u) = \frac{1}{2} \sum_{i=1}^{K} \tan^{-1}(u/\mu_i) + \frac{1}{2} v \tan^{-1}(au) - \frac{1}{2} x u$$

$$\rho(u) = \prod_{i=1}^{K} (1 + u^2/\mu_i^2)^{\frac{1}{4}} \times (1 + a^2 u^2)^{v/4}$$

Imhof shows that when the integral is truncated at U, noting that

$$\left|\frac{\sin\theta(u)}{u\rho(u)}\right| < \left|\frac{1}{u\rho(u)}\right|$$

and $\rho(u) > (au)^{v/2} \prod_{i=1}^{K} (\frac{u}{\mu_i})^{\frac{1}{2}}$

$$= a^{v/2} u^{\frac{K+V}{2}} \prod_{i=1}^{K} \mu_i^{-\frac{1}{2}},$$

$$\text{then } \int^{\infty} \left|\frac{\sin\theta(u)}{u\rho(\theta)}\right| du < \int^{\infty} a^{-v/2} u^{-(\frac{K+V+1}{2})} \prod_{i=1}^{K} \mu_i^{\frac{1}{2}} \, du$$

$$= a^{-v/2} \prod_{i=1}^{K} \mu_i^{\frac{1}{2}} \left[-u^{-(\frac{K+V}{2})} \left(\frac{2}{K+V}\right) \right]^{\infty}$$

$$\doteq a^{-v/2} \prod_{i=1}^{K} \mu_i^{\frac{1}{2}} \frac{1}{(\frac{K+v}{2})} \left(\frac{2}{K+V}\right).$$

U was chosen so that the error was less than 10^{-5}.

It is found using this method that if K = 20 and using the eigenvalues $\mu_i = \pi^2 i^2 \quad i = 1,\ldots,K$, that the significance points of W^2 computed by this method agree with those given by Anderson and Darling (1952) to five significant figures. The integral was calculated using six point Gaussian quadrature and four point Gaussian quadrature with variable interval length until the two quadratures differed by no more than 10^{-5}.

The percentage points of ${}_pW^2$ are given in Table 6.5, obtained by finding the values of $Pr(H \leq x)$ at 100 grid points in each of the upper and lower tails of ${}_pW^2$ and then using quartic inverse interpolation.

6.5.3 The statistic ${}_{q,p}W^2$

The percentage points of ${}_{q,p}W^2$ can be similarly found. The mean and variance can be found by using formula (6.31B). We have

$$E[{}_{q,p}W^2] = p^2\left(\frac{1}{2} - \frac{p}{3}\right) - q^2\left(\frac{1}{2} - \frac{q}{3}\right) \qquad (6.42A)$$

$$\text{and } \operatorname{var}({}_{q,p}W^2) = \frac{4}{3}\left(\frac{p^4}{4} - q^3 p - \frac{2}{5}p^5 + p^2 q^3 + \frac{p^6}{6} - \frac{p^3 q^3}{6} + \frac{3}{4}q^4 - \frac{3}{5}q^5\right)$$

The eigenvalues can be found by solving (6.34). This was done on the computer using the Newton-Raphson method as for ${}_pW^2$, for values of q and p symmetric about $\frac{1}{2}$, taking $p = 1 - q$. The first ten eigenvalues are given in Table 6.3A. The percentage points were found as for ${}_pW^2$ and these are given in Table 6.5A.

6.6 Asymptotic power of ${}_pW^2$ and its components against shifts in location and scale.

6.6.1 Durbin and Knott's results.

We follow Durbin and Knott (1972) to obtain the powers of the tests based on ${}_pW^2$ and its components against shifts in location and scale for normal and exponential populations. Durbin and Knott (1972) show that if the test of the null hypothesis

$$H_o : F \equiv F(x,\Theta_o)$$

against $$H_1 : F \equiv F(x,\Theta_1)$$

where $\Theta_1 = \Theta_o + \gamma n^{-\frac{1}{2}}$, and Θ is a vector parameter, is considered than W^2 under the alternative has distribution that of

$$\sum_{j=1}^{\infty} Z_j^2/(j^2\pi^2)$$

where the Z_j are independent $N(\gamma^T\delta_j, 1)$ where

$$\delta_j = \sqrt{2}\, j\pi \int_o^1 \sin(j\pi t)h(t)\, dt \qquad \text{and}$$

$$h(t) = \frac{\partial F}{\partial\Theta}(x,\Theta) \text{ where } t = F(x,\Theta)$$

6.6.2 Extension of previous results to ${}_pW^2$.

Now consider the Fourier component ${}_pZ_{nj}$ for a finite sample of the empirical process $y_n(t)$:

$${}_pZ_{nj} = \sqrt{\mu j} \int_o^p g_j(t)\, y_n(t)\, dt$$

where $g_j(t)$ is the eigenfunction corresponding to μ_j,

$$g_j(t) = n_j \sin\sqrt{\mu_j}\, t$$

where $\tan(\sqrt{\mu_j}p) = \sqrt{\mu_j}(1-p)$ and n_j is the normalizing constant

$$2 n_j^{-2} = p - \frac{\sin\sqrt{\mu_j}p \cos\sqrt{\mu_j}p}{\sqrt{\mu_j}}$$

The ${}_pZ_{nj}$ $j = 1, 2, \ldots$ are uncorrelated since

$$E\left[{}_pZ_{nj}\;{}_pZ_{nk}\right] = \sqrt{\mu_j\mu_k}\int_o^p\int_o^p E\left[y_n(t)y_n(s)\right]g_j(t)g_k(s)\;dt\;ds$$

$$= \sqrt{\mu_j\mu_k}\int_o^p\int_o^p \{\min(st) - st\}\;g_j(t)g_k(s)\;dt\;ds$$

$$= \sqrt{\frac{\mu_j}{\mu_k}}\int_o^p g_j(t)g_k(t)\;dt$$

$$= \delta_{jk}$$

since $g_k(s)$ satisfies the integral equation (6.27) and the $g_j(t)$ are orthogonal on (o, p).

$$\text{But}\quad E\left[{}_pZ_{nj}\right] = \sqrt{\mu_j}\int_o^p g_j(t)\quad E\left[y_n(t)\right]\;dt$$

$$= 0$$

So the ${}_pZ_{nj}$ are uncorrelated with variance 1.

Also $$\frac{{}_pZ_{nj}}{\sqrt{\mu_j}} = n_j\int_o^p \sin(\sqrt{\mu_j}t)y_n(t)\;dt$$

so $$\frac{{}_pZ_{nj}}{\sqrt{\mu_j}n_j} = -\left[y_n(t)\frac{\cos\sqrt{u_j}\;t}{\sqrt{\mu_j}}\right]_o^p$$

$$+ \int_o^p \frac{\cos\sqrt{\mu_j}\;t}{\sqrt{\mu_j}}\;dy_n(t)$$

$$= -y_n(p)\frac{\cos\sqrt{\mu_j}\;t}{\sqrt{\mu_j}} + \sqrt{n}\int_o^p \frac{\cos\sqrt{\mu_j}t}{\sqrt{\mu_j}}\;dF_n(t)$$

$$- \sqrt{n}\int_o^p \frac{\cos\sqrt{\mu_j}t}{\sqrt{\mu_j}}\;dt$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{r} \frac{\cos(\sqrt{\mu_j}\, t_i)}{\sqrt{\mu_j}} - y_n(p) \frac{\cos\sqrt{\mu_j}\, p}{\sqrt{\mu_j}}$$

$$-\sqrt{n} \quad \frac{\sin \sqrt{\mu_j}\, p}{\left(\sqrt{\mu_j}\,\right)^2}$$

with r observations $t_1, t_2, \ldots, t_r$ less than p.

The first term in the last line immediately above, is a sum of i.i.d. random variables and so is asymptotically normally distributed. The second term is also asymptotically normally distributed with mean zero and variance $p(1-p)\cos^2(\sqrt{\mu_j}\,p)/\mu_j$, and so ${}_pZ_{n_j}$ is asymptotically normally distributed.

Now on H_1, $y_n(t)$ no longer has expectation zero, but can be shown to have expectation such that

$$E[y_n(t)] = \gamma^T h(t)$$

with the covariance function not changed from the null situation, see Durbin and Knott (1972).

It can easily be shown that ${}_pW_n^2$ can be expressed:

$${}_pW_n^2 = \sum_{j=1}^{\infty} \frac{{}_pZ_{n_j}^2}{\mu_j}$$

and so the limiting distribution of ${}_pW_n^2$ on H_1 is that of

$${}_pW^2 = \sum_{j=1}^{\infty} \frac{{}_pZ_j^2}{\mu_j}$$

where the ${}_pZ_j$ are independent $N(\gamma^T\beta_j, 1)$ where

$$\beta_j = \sqrt{\mu_j} \int_0^p g_j(t)h(t)\ dt \qquad (6.45)$$

and $g_j(t)$ is given by (6.37), with the j^{th} smallest eigenvalue, μ_j, satisfying the integral equation (6.27) Following Durbin and Knott (1972), approximate to the distribution of $({}_pW^2 | H_1)$ by

$$W_A^2 = \sum_{j=1}^{K} \frac{{}_pZ_j^2}{\mu_j} + a\ C^2_{v,d^2}$$

where the ${}_pZ_j$ $j = 1,2,\ldots,K$ are independent $N(\gamma^T\beta_j, 1)$ and C^2 is a non-central chi-squared random variable with v degrees of freedom and non-central parameter d^2. a and v are chosen so that W_A^2 has the same mean and variance as ${}_pW^2$ on H_o, (i.e. where the ${}_pZ_j$ are N(0,1)) and d chosen so that W_A^2 has the same mean as ${}_pW^2$ on H_1. Now

$$E[{}_pW^2 | H_1] = \sum_{j=1}^{\infty} \frac{1 + \gamma^T\beta_j\beta_j^T\gamma}{\mu_j}$$

$$= E[{}_pW^2 | H_o] + \sum_{j=1}^{\infty} \frac{\gamma^T\beta_j\beta_j^T\gamma}{\mu_j}$$

and a and v are determined by the original approximation to the distribution of ${}_pW^2$, given by equations (6.42). Then d^2 is determined by

$$\sum_{j=1}^{K} \frac{\gamma^T\beta_j\beta_j^T\gamma}{\mu_j} + ad^2 = \gamma^T\int_0^p h(t)h^T(t)\ dt)\gamma$$

6.6.3 The alternatives for ${}_pW^2$

To see what power is lost by using the censored data rather than the complete sample scale and location shifts were considered under the alternative. The distributions chosen were normal for both kinds of shift and exponential for scale shift. The alternatives and the corresponding function h(t) are given below

i) F is normal with location shift

$$F(x) = \Phi(x - \gamma_n^{-\frac{1}{2}})$$

where $\Phi(\ \ )$ is the standard normal c.d.f. Then

$$h(t) = -\phi(\Phi^{-1}(t)) \text{ where } \phi(t) = \Phi'(t). \qquad (6.46)$$

ii) F is normal with scale shift

$$F(x) = \Phi(x\{1 + \gamma_n^{-\frac{1}{2}}\}^{-\frac{1}{2}})$$

$$\text{then } h(t) = -\tfrac{1}{2}\, \Phi^{-1}(t)\, \phi(\Phi^{-1}(t)) \qquad (6.47)$$

iii) F is exponential with scale shift

$$F(x) = 1 - e^{-x/(1 + \gamma n^{-\frac{1}{2}})}$$

$$\text{then } h(t) = (1 - t)\log(1 - t) \qquad (6.48)$$

(See Dodgson (1972), who considers the power of Durbin and Knott's components for scale shift of an exponential distribution).

The power of the test, for the various alternatives, is compared with the 2-sided test based on the best test for the simple 1-sided alternative.

For alternative (i) the best test is based on $\overline{X}$, for (ii) S^2 and (iii) $\overline{X}$ again. In each case the statistics are asymptotically normally distributed:

i) $X \sim N(\gamma, 1)$

ii) $S^2 \sim N(\gamma, 2)$

iii) $X \sim N(\gamma, 1)$

γ is chosen so that for a given significance level, the two-sided test based on the best test for the one-sided alternative has power equal to .25, .50, .75 and .90. Note that if γ_i is the value of γ for the alternative (i) then $\gamma_{ii} = \sqrt{2}\gamma_i$ and $\gamma_{iii} = \gamma_i$.

Power	1 % level γ'	5% level γ
.25	1.9013	1.2855
.50	2.5758	1.9600
.75	3.2503	2.6345
.90	3.8574	3.3416

Figure 6.1

Values of γ for alternative (i) for various powers of two-sided test based on best one-sided test.

For 1% significance level and 25% power γ is found so that

$$\Pr(Z > 2.5758 - \gamma \text{ and } Z < -2.5758 - \gamma) = .25$$

i.e. $2.5758 - \gamma = 1.9013$

The powers of the test based on best test and for the complete sample ${}_pW^2$ are considered so that each test is based on the same number of observations, i.e. if the best test is based on n observations then ${}_pW^2$ is based on a complete sample of np^{-1} observations. So if the new alternative

$$\theta_1 = \theta_o + \gamma(np)^{-\frac{1}{2}}$$

are considered, the two tests will be based on the same number of observations. So if γ is taken as $\gamma p^{-\frac{1}{2}}$, the two tests are based asymptotically on the same number of observations, rather than the two tests being based on complete samples of the same size.

6.6.4 Results of calculations of power for ${}_pW^2$

The power of ${}_pW^2$ for the various alternatives is given by $Pr(W^2{}_A > w_\alpha)$, where W^2_A is the approximation to ${}_pW^2$ on H_1, and w_α is the α significance point taken from Table 6.5 for ${}_pW^2$. Imhof's method is used for inverting the characteristic function.

$$\text{If } H = \sum_{i=1}^{K} \frac{Z_i^2}{\mu_i} + a\ C^2_{v,d^2}$$

where the Z_i are independent $N(\delta_i, 1)$ and C^2_{v,d^2} is a chi-squared random variable with v degrees of freedom and non-central parameter d^2 then

$$Pr(H \leq x) = \tfrac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin\{\theta(u)\}}{u\rho(u)}\, du$$

where

$$\theta(u) = \tfrac{1}{2}\left\{ \sum_{i=1}^{K} \tan^{-1}(u/\mu i) + v \tan^{-1}(au) - xu + \frac{d^2 au}{1+u^2a^2} + \sum_{i=1}^{K} \delta_i^2 \mu_i^{-1} u(1 + u^2\mu_i^{-2})^{-1} \right\}$$

and $\rho(u) = \prod_{i=1}^{K} (1 + u^2/\mu_i^2)^{\frac{1}{4}} \times (1 + a^2 u^2)^{v/4}$

$$\times \exp\left\{\frac{1}{2} \sum_{i=1}^{K} \frac{\delta_i^2 \mu_i^{-2} u^2}{1+\mu_i^{-2} u^2} + \frac{1}{2} \frac{d^2 a^2 u^2}{1+a^2 u^2}\right\}$$

As in §6.6 the integration can be truncated at u and following Imhof it is found:

$$\int_{\bar{u}}^{\infty} \left|\frac{\sin \theta(u)}{u\rho(u)}\right| \, de < a^{-\frac{v}{2}} \prod_{i=1}^{K} \mu_i \; \bar{u}^{\frac{1}{\frac{K+v}{2}}} \left(\frac{2}{K+v}\right)$$

$$\times \exp\left\{\frac{1}{2} \sum_{i=1}^{K} \frac{\delta_i^2 \mu_i^{2} u^2}{1+\mu_i^{-2} u^2} + \frac{1}{2} \frac{d^2 a^2 \bar{u}^2}{1+a^2 \bar{u}^2}\right.$$

Note that

$$\exp\left\{\frac{1}{2} \sum_{i=1}^{K} \frac{\delta_i \mu_i^{-2} \bar{u}^2}{1+\mu_i^{-2} \bar{u}^{-2}} + \frac{d^2 a^2 \bar{u}^2}{1+a^2 \bar{u}^2}\right.$$

$$< \exp\left\{\frac{1}{2} \sum_{i=1}^{K} \delta_i^2 + \frac{1}{2} d^2\right.$$

$$< \exp\left\{\frac{1}{2}\gamma^2 \int_0^p h^2(t) \, dt + \frac{1}{2} d^2\right\}$$

The value of u was chosen so that the truncation error was less than 10^{-4}.

The results of the intergration are given in Table 6.6. Firstly, it is seen that W^2 has good power with respect to the best test for alternative i, being less than the power of the best test by no more than 20% for 25% power of the best test, for both significance levels.

The percentage loss of power, with respect to the best test, decreases as the power of the best test increases.

It is seen that W^2 has little power against normal scale shift but has reasonable power against exponential scale shift. The percentage loss in power being about 33% at the .25 best test power and 1% significance. The percentage loss of power of W^2 decreases as the power increases.

The power of ${}_pW^2$ with p = .9 for the alternatives (i) and (iii) is fractionally less than the power of W^2. However, for alternative (ii) ${}_pW^2$ is slightly more powerful.

With p = .5, the power for alternatives (i) and (iii) is reduced. The reduction from the power of W^2 for (i) is no more than 10% for alternative (i) but is greater for alternative (iii), the reduction there being about 30% at most at the 5% significance level, and greater at the 1% significance level. Thus for normal location shift the reduction in power is not very much compared with W^2 based on the full sample. For a comparison, the power of the test based on the median, which would be a feasible alternative if we had a proportion p = .5 of the sample, has been calculated for the 5% significance level and γ giving powers of the best test equal to

.25, .50, .75 and .90, and given in Figure 6.2. The statistic based on the median, $\sqrt{n}\tilde{X}_n$, has asymptotic distribution $N(\gamma, \frac{\pi}{2})$, so the power is easily calculated.

Asymptotic power of median for location shift of normal population, at 5% significance level

Power of 'best' test	.25	.50	.75	.90
Power of median	.175	.346	.556	.760
Power of ${}_{.5}W^2$	.199	.397	.627	.826

Figure 6.2

In Figure 6.2 the power of ${}_{.5}W^2$ is also given, taken from Table 6.6. It is seen that ${}_{.5}W^2$ has better power than the median for all four powers of the best test considered.

We suppose also that ${}_pW^2$ will have power for location shift for symmetric distributions, similar to that for the normal distribution location shift.

§6.6.5 The power of the components of ${}_pW^2$

In this section the power of the components ${}_pZ_j$ is found for the case of simultaneous location and variance shift for the normal population. For this case the ${}_pZ_j$ are $N(\gamma^T\beta_j, 1)$ where the β_j is given by (6.45) and now

$$F(x) = \Phi(\{x-\gamma_1 n^{-\frac{1}{2}}\}\{1+\gamma_2 n^{-\frac{1}{2}}\}^{-\frac{1}{2}})$$

So
$$\gamma^T\beta_j = \gamma_1\beta_{1j} + \gamma_2\beta_{2j}$$
$$= \gamma_{12j}, \quad \text{say,}$$

where
$$\beta_{ij} = \sqrt{\mu_j}\int_0^p g_j(t)h_i(t)\,dt \qquad (6.49)$$

and
$$h_1(t) = -\phi(\Phi^{-1}(t))$$
$$h_2(t) = \tfrac{1}{2}\Phi^{-1}(t).h_1(t)$$

Thus if the two-sided test based on ${}_pZ_j$ is considered, then at the α significance level, the power of the test is

$$P_{12j} = 1 - \Phi(Z_{\frac{\alpha}{2}}-\gamma_{12j}) + 1 - \Phi(Z_{\frac{\alpha}{2}}+\gamma_{12j})$$

$$\text{where } \Phi(Z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

If the values of γ_1 and γ_2 are chosen so that the power of ${}_pZ_j$ remains constant at k_β, say then γ_{12j} is chosen so that

$$P_{12j} = k_\beta$$

Thus if the power of ${}_pZ_j$ is plotted in the (γ_1, γ_2) plane, then $P_{12j} = k_\beta$, a constant, will be a straight line.

The only component of interest (against location and variance shift) is the first component, ${}_pZ_1$. For $p = 1.0$, Z_1 has no power against variance shift since $\beta_2 = 0$. However as $p \longrightarrow 0$, the coefficient β_2 increases, and so ${}_pZ_1$ has power against scale shift. In Figure 6.3, the lines

FIGURE 6.3 ISODYNES FOR COMPONENTS ${}_pZ_1$ & ${}_pZ_2$ OF ${}_pW^2$.

$$\gamma_1\beta_{11} + \gamma_2\beta_{21} = \gamma, \text{ a constant}$$

are drawn, where the constant, γ, is chosen so that ${}_pZ_1$ has 50% power against a shift of (γ_1, γ_2) in location and variance, at the 5% level, for the 2-sided test. For this situation from Figure 6.1, it is seen that $\gamma = 1.96$. The values of β_{11} and β_{21} are given below in Figure 6.4. Also in Figure 6.3, are the isodynes for ${}_pZ_2$. These are given by

$$\gamma_1\beta_{12} + \gamma_2\beta_{22} = 1.96$$

for 50% power, where β_{12}, β_{22} are given in Figure 6.4.

It is seen that ${}_pZ_2$ has virtually no power against location shift (i.e. changes in values of γ_1) since β_{12} is or near zero for p = .5(.1)1.0.

The isodynes for 90% power can be found by taking the lines

$$\gamma_1\beta_{12} + \gamma_2\beta_{22} = 3.3416$$

where 3.3416 is given by Figure 6.1. These are the 50% isodynes rescaled.

p	Z_1: β_{11}	Z_1: β_{21}	Z_2: β_{12}	Z_2: β_{22}
1.0	-.948	0.000	0.000	.525
.9	-.946	.008	-.007	.510
.8	-.935	.038	-.018	.476
.7	-.919	.083	-.019	.441
.6	-.897	.139	-.009	.397
.5	-.868	.199	-.007	.354

Figure 6.4. Values of β_{ij} (see (6.49)) for the first 2 components ${}_pZ_1$ and ${}_pZ_2$ of ${}_pW^2$ for location and variance shift of the normal distribution.

Also included in Figure 6.3 is the isodyne for the test based on the statistic

$$L^2 = [\sqrt{n}(\bar{X}-\mu)]^2 + .5[\sqrt{n}(S^2-\sigma^2)]^2$$

which is the likelihood ratio test, see Kendall and Stuart (1973, §24.7). It is shown there, that L^2 has asymptotic distribution χ^2_2 under the null hypothesis. If the null hypothesis is

$$H_o : \mu = 0, \ \sigma^2 = 1$$

then it is shown there that under the alternative

$$H_1 : \mu = \gamma_1 n^{-\frac{1}{2}}, \ \sigma^2 = 1 + \gamma_2 n^{-\frac{1}{2}}$$

L^2 has the non-central chi-squared distribution with 2 degrees of freedom and non-central parameter

$$\delta^2 = \gamma_1^2 + .5\gamma_2^2$$

If the test is carried out at the 5% level, with $\alpha = .05$ then the value of c_α such that $Pr(\chi^2_2 > c_\alpha) = \alpha$ is given by 5.991, from Pearson and Hartley (1966). It is found using Imhof's method to find the distribution of non-central χ^2_{2,δ^2} that

$$Pr(\chi^2_{2,\delta^2} > 5.991) = .5 \quad \text{if } \delta^2 = 4.94$$

and

$$Pr(\chi^2_{2,\delta^2} > 5.991) = .9 \quad \text{if } \delta^2 = 12.65$$

Thus the ellipse

$$\gamma_1^2 + .5\gamma_2^2 = 4.94$$

in Figure 6.3, gives the values of (γ_1, γ_2) such that the power of the LR test based on L^2 is .50, and so for all values inside the ellipse the power of the test based on L^2 is no more than .50

Similarly the ellipse

$$\gamma_1^2 + .5\gamma_2^2 = 12.65$$

gives the isodyne of power .90; the intercepts on the two axes of this ellipse are drawn in Figure 6.3.

Similarly the LR test based on the MLE's for a proportion p of the original sample, can be found. Following Kendall and Stuart (1973, §24.7) if $(\hat{\mu},\hat{\sigma}^2)$ are the maximum likelihood estimates of (μ,σ^2) then asymptotically the LR test is given by

$$_pL^2 = n(\mu_o-\hat{\mu},\sigma^2-\hat{\sigma}^2)V_p^{-1}(\mu_o-\hat{\mu},\sigma^2-\hat{\sigma}^2)^T$$

where V_p is the inverse of the information matrix for a sample singly censored at an end so that a proportion p of the original sample remains. On the null, $_pL^2$ has the χ_2^2 distribution and on the alternative $(\mu_1\sigma_1^2)$, $_pL^2$ has the χ_2^2 distribution with non-central parameter δ^2 given by

$$\delta^2 = (\mu_o-\mu_1,\sigma_o^2-\sigma_1^2)V_p^{-1}(\mu_o-\mu_1,\sigma_o^2-\sigma_1^2)^T$$

Consequently with $(\mu_o\sigma_o^2) = (0,1)$ and $(\mu_1,\sigma_1^2) = (\gamma_1 n^{-\frac{1}{2}},1+\gamma_2 n^{-\frac{1}{2}})$, the non-central parameter is given by

$$\delta^2 = n^{-1}(\gamma_1,\gamma_2)V_p^{-1}(\gamma_1\gamma_2)^T$$

Since V^{-1} is of order n, the non-central parameter is of order 1.

For p = .5 it is found that the 50% isodyne is given by

$$.81831\gamma_1^2 - .39894\gamma_1\gamma_2 + .25\gamma_2^2 = 4.94$$

Figure 6.3 can be extended over the whole (γ_1,γ_2) plane by completing the ellipse and drawing lines

$$\gamma_1\beta_{11}+\gamma_2\beta_2, = -\gamma$$

corresponding to the lines

$$\gamma_1 \beta_{11} + \gamma_2 \beta_{21} = \gamma$$

It is seen from Figure 6.3 that the components ${}_pZ_1$ ($p < 1.0$) have good power for values of γ_1 and γ_2 having different sign, but poor power for γ_1 and γ_2 of the same sign.

Under this alternative y(t) has expectation:

$$E[y(t)] = \gamma_1 h_1(t) + \gamma_2 h_2(t) \equiv G_\gamma(t)$$

where $h_i(t)$ ($i = 1,2$) is defined by (6.49).

Now since $h_1(t)$ is symmetric about .5, and $h_2(t)$ is not, but $|h_2(t)|$ is symmetric about .5, for given values of γ_1 and γ_2 either $|G_\gamma(t)|$ will be large in (0.5) and small in (.5,1) or $|G_\gamma(t)|$ will be large in (.5,1) and small in (0,.5), The larger $|G_\gamma(t)|$ is in (0,p) then the greater the power, of the tests based on ${}_pW^2$ and ${}_pZ_1$ will be. Thus the powers of the tests based on ${}_pW^2$ and ${}_pZ_1$ for this alternative are 1-sided in the sense that the power for the alternative (γ_1, γ_2) is different from the power for $(-\gamma_1, \gamma_2)$.

The component ${}_pZ_1$ for $p = .5$ is seen to have good power for shifts of (γ_1, γ_2), and is better than the LR test for values of γ_1, γ_2 in the neighbourhood of $\gamma_2 = 0$. Also ${}_pZ_2$ for $p = .5$ has good power for shifts of γ_2 and is better than the LR test for γ_1, γ_2 in the neighbourhood of $\gamma_1 = 0$. Consequently the components ${}_pZ_1$ and ${}_pZ_2$ still retain the characteristics of Z_1 and Z_2 for $p = 1.0$, except that as $p \rightarrow .5$, ${}_pZ_1$ has more power against variance shift, with some loss of power against location shift. Thus a test with power against location and variance shift will be given by the statistic ${}_pB^2$ where

$${}_pB^2 = {}_pZ_1^2 + {}_pZ_2^2$$

which has the χ_2^2 distribution, with non-central parameter

$$\delta^2 = (\gamma_1\beta_{11}+\gamma_2\beta_{21})^2 + (\gamma_1\beta_{12}+\gamma_2\beta_{22})^2$$

on the alternative hypothesis.

In Figure 6.5 the isodynes for 50% power, at the 5% significance level, are drawn for the LR tests with p = 1.0 and p = .5 and also the tests based on ${}_pZ_1^2 + {}_pZ_2^2$ for p = 0.5(.1) 1.0. It is seen there is very little difference between ${}_1L^2$, ${}_1B^2$ and ${}_{.9}B^2$; little is lost using ${}_{.8}B^2$ either. ${}_{.5}L^2$ and ${}_{.5}B^2$ are similar except that ${}_{.5}L^2$ has slightly more power against variance shift, and ${}_{.5}B^2$ slightly more power against location shift.

Using Stephens' idea of the area enclosed by the isodyne to measure how good a test is, we have the area of the ellipse

$$ax^2 + bxy + cy^2 = d$$

given by

$$\frac{\pi d}{\sqrt{ac-b^2/4}}$$

These areas are given in Figure 6.6. We see that the tests ${}_pB^2$ compare favourably with ${}_pL^2$.

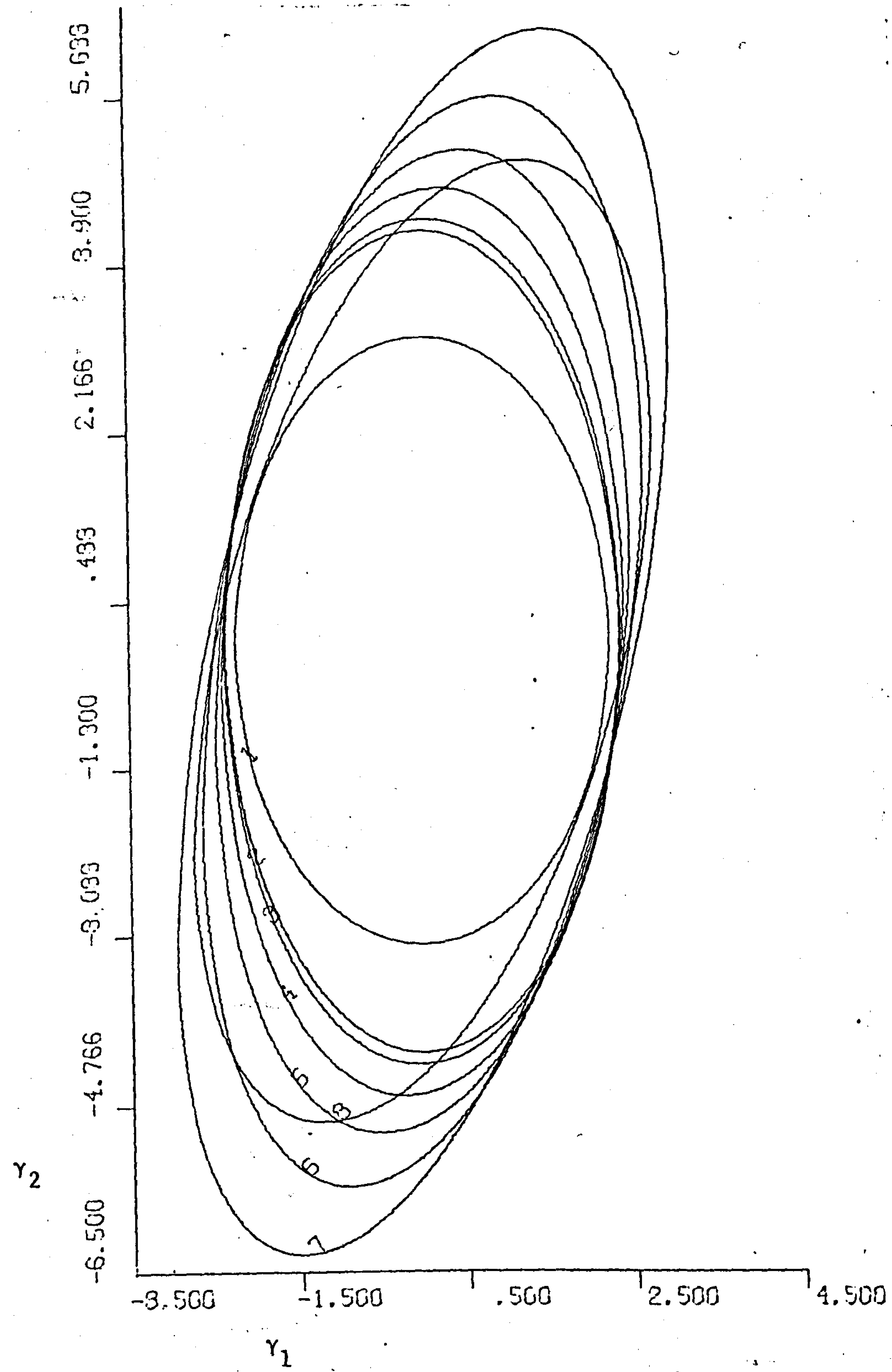


Figure 6.5

50% Isodynes at the 5% level for components of ${}_pW^2$.

Curve 1 Likelihood ratio test ${}_pL^2$, p = 1.0

Curve 8 Likelihood ratio test ${}_pL^2$, p = .5

Curves 2-7 Test based on ${}_pZ_1^2 + {}_pZ_2^2$ p = .9(.1).5 respectively.

Testing for the normal distribution with $\mu = 0$ and $\sigma^2 = 1$, with the alternative $\mu = \gamma_1 n^{-\frac{1}{2}}$, $\sigma^2 = 1 + \gamma_2 n^{-\frac{1}{2}}$.

Figure 6.6. Areas enclosed by isodynes based on ${}_pZ_1^2 + {}_pZ_2^2$ and ${}_pL^2$.

Statistic		Area
${}_{p=1}L^2$		21.95
${}_pZ_1^2 + {}_pZ_2^2$	$p = 1.0$	31.20
	$p = .9$	32.19
	$p = .8$	34.93
	$p = .7$	38.48
	$p = .6$	43.69
	$p = .5$	50.81
${}_{p = .5}L^2$		38.23

TABLE 6.1

Approximate percentage points of ${}_pW_n^2$ and ${}_rW_n^2$, using the 'a C_v^2 + b' approximation.

$p = .9$ or $r = .9n$

% Point	n = 10		n = 20		n = 30	
	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$
50	.0962	.1129	.1032	.1104	.1050	.1095
85	.2456	.2816	.2644	.2815	.2704	.2813
90	.3001	.3419	.3233	.3439	.3311	.3445
95	.3956	.4473	.4263	.4535	.4377	.4557
97.5	.4936	.5545	.5320	.5638	.5468	.5695
99	.6250	.6981	.6738	.7166	.6941	.7229

$p = .75$ or $r = .75n$

% Point	n = 8		n = 16		n = 20		n = 32	
	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$
50	.0680	.0965	.0789	.0933	.0814	.0927	.0851	.0918
85	.1998	.2497	.2194	.2492	.2247	.2492	.2333	.2492
90	.2522	.3044	.2735	.3066	.2794	.3070	.2893	.3075
95	.3467	.3995	.3700	.4070	.3762	.4084	.3883	.4105
97.5	.4454	.4961	.4696	.5098	.4774	.5124	.4906	.5164
99	.5792	.6253	.6048	.6480	.6140	.6515	.6292	.6585

$p = .5$ or $r = .5n$

% Point	n = 10		n = 20		n = 30	
	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$	${}_rW_n^2$	${}_pW_n^2$
50	.0314	.0507	.0373	.0506	.0401	.0505
85	.1374	.1492	.1357	.1504	.1372	.1508
90	.1865	.1873	.1822	.1888	.1822	.1893
95	.2791	.2554	.2704	.2573	.2670	.2579
97.5	.3792	.3260	.3661	.3283	.3588	.3289
99	.5186	.4220	.5002	.4245	.4871	.4254

TABLE 6.2

Exact moments of ${}_pW^2$: derived using the formulae (6.41) for the moments of ${}_pW^2$ about the origin.

p	μ	σ^2	$\mu_3 \times 10^3$	β_1
1.00	.166667	.022222	8.46561	6.531
.99	.166617	.022222	8.46543	6.531
.95	.165458	.022173	8.44383	6.541
.90	.162000	.021870	8.30060	6.587
.85	.156542	.021170	7.95207	6.665
.80	.149333	.020025	7.36778	6.760
.75	.140625	.018457	6.56825	6.861
.70	.130667	.016540	5.61229	6.961
.65	.119708	.014380	4.57980	7.054
.60	.108000	.012096	3.55474	7.140
.55	.095792	.009812	2.61091	7.217
.50	.083333	.007639	1.80225	7.287

TABLE 6.2A

Exact mean and variance of ${}_{q,p}W^2$: derived using formulae (6.42A), taking q = 1 - p

p	μ	σ²
.95	.16425	.022147
.90	.15733	.021680
.85	.14642	.020581
.80	.13200	.018752
.75	.11458	.016211
.70	.09467	.013078
.65	.07275	.009562
.60	.04933	.005952
.55	.02492	.002615

TABLE 6.3

${}_pW^2$. Column(1): eigenvalues of the matrix A defined in (6.36A) with dimension 20. Column(2): solutions of the integral equation (6.27) found by solving $\tan \mu^{\frac{1}{2}} p = - \mu^{\frac{1}{2}}(1-p)$

p = .95 (1)	(2)*	p = .90 (1)	(2)*	p = .5 (1)	(2)*
.101239	-	.100688	-9	.060740	-1
.025251	-	.024780	-1	.010355	-7
.011184	-	.010803	-4	.003926	-7
.006264	-	.005964	-5	.002033	-4
.003990	-1	.003754	-5	.001237	-9
.002758	-	.002571	-	.000830	-2
.002017	-	.001866	-7	.000594	-7
.001537	-8	.001415	-	.000446	-9
.001209	-10	.001108	-9	.000346	-50
.000976	-	.000891	-2	.000272	-80

Exact eigenvalues for p = .6(.1).8, as column (2) above.

p =	.8	.7	.6
	.096865	.88564	.076126
	.022326	.018557	.014379
	.009284	.007407	.005571
	.004981	.003906	.002906
	.003083	.002397	.001775
	.002089	.001616	.001194
	.001506	.001162	.000857
	.001137	.000875	.000645
	.000888	.000683	.000503
	.000712	.000547	.000403

* - indicates a change in least significant digit(s) from value in Column (1).

TABLE 6.3A

Eigenvalues of the equation (6.35), found by Newton-Raphson method, for the statistic ${}_{q,p}W^2$ with symmetric censoring, q = 1-p. Values of $\mu \times 10$.

p	.95	.90	.80	.75
	1.0116	1.0005	.9216	.8444
	.2517	.2422	.1903	.1516
	.1111	.1033	.0717	.0533
	.0620	.0558	.0360	.0259
	.0393	.0345	.0212	.0151
	.0270	.0232	.0139	.0098
	.0197	.0166	.0098	.0069
	.01449	.0125	.0073	.0051
	.0117	.0097	.0056	.0039
	.0094	.0077	.0044	.0031

TABLE 6.4

Cumulants of ${}_pW^2$: calculated using twenty-five eigenvalues found by solving $\tan \mu^{\frac{1}{2}} p = - \mu^{\frac{1}{2}}(1-p)$

p	μ	σ^2	$\mu_3 \times 10^3$	$\kappa_4 \times 10^4$	β_1	β_2
.99	.164414	.022222	8.46543	50.7923	6.531	10.286
.95	.163343	.022173	8.44383	50.6270	6.541	10.298
.90	.159996	.021870	8.30060	49.5242	6.589	10.354
.80	.147552	.020025	7.36777	42.3815	6.760	10.569
.70	.129108	.0165401	5.61228	29.5889	6.961	10.816
.60	.106664	.012096	3.55473	16.1413	7.140	11.032
.50	.082220	.007639	1.80224	6.53931	7.287	11.207

TABLE 6.5

Percentage points of ${}_pW^2$ found by Imhof's method; using about 20 eigenvalues, found by solving

$$\tan \mu^{\frac{1}{2}}p = - \mu^{\frac{1}{2}}(1-p).$$

Percentage Point	p = .50	.55	.60	.65	.70	.75	.80	.85	.90	.95	1.00	Percentage Point
1.0	.0080	.0095	.0112	.0130	.0149	.0167	.0187	.0206	.0211	.0240		1.0
2.5	.0101	.0120	.0141	.0163	.0186	.0210	.0233	.0256	.0250	.0294		2.5
5.0	.0126	.0150	.0176	.0202	.0230	.0258	.0286	.0312	.03140	.0356		5.0
10.0	.0166	.0197	.0230	.0264	.0298	.0333	.0367	.0399	.0426	.0450		10.0
15.0	.0203	.0240	.0279	.0319	.0350	.0400	.0439	.0476	.0506	.0532		15.0
50.0	.0536	.0619	.0707	.0793	.877	.0957	.1030	.1093	.1147	.1178	.1189	50.0
85.0	.1506	.1721	.1929	.2127	.2306	.2465	.2600	.2707	.2784	.2829	.2841	85.0
90.0	.1890	.2153	.2407	.2645	.2861	.3048	.3205	.3327	.3412	.3462	.3473	90.0
95.0	.2579	.2931	.3269	.3581	.3861	.4102	.4298	.4446	.4548	.4599	.4614	95.0
97.5	.3295	.3742	.4167	.4558	.4906	.5201	.5439	.5616	.5733	.5791	.5974	97.5
99.0	.4271	.4847	.5393	.5891	.6330	.6701	.6997	.7212	.7352	.7419	.7434	99.0

TABLE 6.5A

Percentage points of ${}_{q,p}W^2$, for symmetric double censoring, found by Imhof's method using the available eigenvalues found by solving (6.35).

$p = 1 - q$

Percentage Point	.95	.90	.8	.75
1	.0221	.0176	.0113	.0083
2.5	.0278	.0227	.0149	.0112
5	.0342	.0288	.0192	.0147
10	.0438	.0381	.0260	.0202
15	.0521	.0462	.0322	.0254
50	.1166	.1102	.0862	.0718
85	.2814	.2733	.2369	.2084
90	.3445	.3358	.2956	.2621
95	.4584	.4484	.4018	.3592
97.5	.5775	.5661	.5131	.4612
99	.7401	.7269	.6665	.6028

TABLE 6.6

Asymptotic power of ${}_pW^2$ for scale and location shifts for given populations. In parentheses, power based on alternative $\gamma p^{-\frac{1}{2}}$.

5% significant level

p	Alternative	.25	.50	.75	.90
	Power of 'best' test	.25	.50	.75	.90
1.0	i Normal, location shift	.228	.457	.703	.887
	ii Normal scale shift	.063	.085	.129	.213
	iii Exponential scale shift	.191	.382	.612	.817
0.9	i	.227(.247)	.455(.495)	.700(.746)	.884(.915)
	ii	.061(.063)	.081(.085)	.119(.130)	.192(.218)
	iii	.188(.203)	.374(.409)	.601(.647)	.807(.846)
0.5	i	.199(.350)	.397(.672)	.627(.899)	.826(.984)
	ii	.068(.088)	.095(.146)	.136(.240)	.197(.378)
	iii	.131(.216)	.243(.433)	.398(.675)	.580(.866)

1% significant level

p	Alternative				
1.0	i	.219	.447	.695	.862
	ii	.015	.023	.037	.064
	iii	.166	.349	.579	.767
.9	i	.217(.247)	.444(.496)	.691(.747)	.859(.899)
	ii	.015(.016)	.021(.023)	.033(.038)	.055(.066)
	iii	.161(.184)	.339(.383)	.564(.622)	.753(.806)
0.5	i	.177(.405)	.367(.722)	.598(.922)	.781(.985)
	ii	.024(.041)	.038(.080)	.061(.145)	.091(.233)
	iii	.088(.201)	.181(.413)	.316(.657)	.469(.833)

§7. The W^2 statistic with censored data for a composite hypothesis

7.1 Introduction

In this section we consider the composite hypothesis in goodness-of-fit. The null hypothesis:

$$H_o : F(x) \equiv F_o(x,\underset{\sim}{\theta}) \qquad (7.1)$$

is tested, where $F_o(\ ,\underset{\sim}{\theta})$ is a c.d.f. of known form, but contains a parameter $\underset{\sim}{\theta}$ which is unknown. For example, $F_o(\ ,\underset{\sim}{\theta})$ may be of exponential form that is

$$F_o(x,\underset{\sim}{\theta}) = 1 - e^{-\lambda(x-\alpha)}, \qquad x \geq \alpha$$

where $\underset{\sim}{\theta} = (\lambda,\alpha)$

and (λ,α) is unknown, or α is known and λ unknown. If censoring has taken place then $\underset{\sim}{\theta}$ can still be estimated by $\hat{\underset{\sim}{\theta}}$ say. The statistics

$${}_p\hat{W}^2_n = n \int_{-\infty}^{x_p} \left\{F_n(x)-F_o(x,\hat{\underset{\sim}{\theta}}_n)\right\}^2 dF_o(x,\hat{\underset{\sim}{\theta}}_n) \qquad (7.2)$$

where $F_o(x_p,\hat{\underset{\sim}{\theta}}_n) = \hat{p}$

and

$${}_r\hat{W}^2_n = n \int_{-\infty}^{x_{(r)}} \left\{F_n(x)-F_o(x,\hat{\underset{\sim}{\theta}})\right\}^2 dF_o(x,\hat{\underset{\sim}{\theta}}_n) \qquad (7.3)$$

can be constructed as before for the simple hypothesis, corresponding to censoring of types I and II respectively. In this case, however, $\hat{p}$ is a random variable dependent upon $\hat{\underset{\sim}{\theta}}_n$, estimating $p = F_o(x_p,\underset{\sim}{\theta}_o)$ where $\underset{\sim}{\theta}_o$ is the actual value of $\underset{\sim}{\theta}$ when the population has c.d.f. $F_o(\ ,\underset{\sim}{\theta})$.

The small sample distributions of these statistics appear completely intractable. For the asymptotic distribution, we follow the work of Darling (1955), who studied the asymptotic distribution of

$$\hat{W}^2_n = n \int_{-\infty}^{\infty} \{F_n(x) - F_o(x, \hat{\theta}_n)\} dF_o(x, \hat{\theta}_n) ;$$

he omits the cap from W^2. He showed that $\hat{W}^2_n$ converges in distribution to a random variable $\hat{W}^2$ which is distributed as

$$\sum_{j=1}^{\infty} \frac{Z_j^2}{\hat{\lambda}_j} \qquad (7.4)$$

where the Z_j are i.i.d. standard normal random variables and the $\hat{\lambda}_j$ are weights obtained from the solution of an integral equation.

In this section we show that ${}_p\hat{W}^2_n$ and ${}_r\hat{W}^2_n$ both converge to a random variable ${}_p\hat{W}^2$ which is also distributed as a weighted sum of χ^2_1 random variables. For testing for normality and exponentiality, the eigenvalues $\hat{\lambda}_j$ are found, the exact mean and variance for ${}_p\hat{W}^2$ calculated by quadrature and the significance points of ${}_p\hat{W}^2$ found by Imhof's method.

7.2 The asymptotic form of the empirical process

7.2.1 Single censoring and a scalar parameter θ

We first consider the stochastic process

$$\hat{y}_n(t) = \sqrt{n}\{F_n(x)-F_o(x,\hat{\underset{\sim}{\theta}}_n)\} \qquad (7.5)$$

where $t = F_o(x,\hat{\underset{\sim}{\theta}}_n)$, and $\hat{\underset{\sim}{\theta}}_n$ is estimated by an efficient method from the sample when Types I or II censoring have taken place. The likelihood for Type I censoring is

$$L = \frac{n!}{r!N!}\left[1-F(x_p,\theta)\right]^N \prod_{i=1}^{r} f(x_i,\theta) \qquad (7.6)$$

where r values $(x_1,\ldots,x_r) \leq x_p$ are observed and N values $> x_p$ are censored, with a potential sample of size $n = r+n$, if no censoring takes place. For Type II, the likelihood is

$$L = \frac{n!}{N!}\left[1-F(x_{(r)},\theta)\right]^N \prod_{i=1}^{r} f(x_i,\theta) \qquad (7.7)$$

where the smallest r values $(x_1,\ldots,x_r)$ are observed and $x_{(r)}$ is the maximum observed value. N observations are censored of a potential sample size $n = N+r$.

Suppose now θ is scalar valued and θ is estimated by a regular efficient unbiased estimator $\hat{\theta}_n$ based on the observations $(x_1,\ldots,x_r)$. The empirical process with θ unspecified is then defined by

$$\hat{y}_n(t) = \sqrt{n}\left[F_n(x)-F(x,\hat{\theta})\right] \qquad (7.8)$$

$$= \sqrt{n}\left[F_n(x)-F(x,\theta)\right] + \sqrt{n}\left[F(x,\theta)-F(x,\hat{\theta})\right]$$

$$= z_n(t) - \sqrt{n}\left[F(x,\theta)-F(x,\hat{\theta})\right]$$

where $t = F(x,\theta)$ and $z_n(t) = \sqrt{n}\left[F_n(x)-F(x,\theta)\right]$ (7.9)

Now following Darling [2], we write

$$\sqrt{n}\Big[F(x,\hat{\theta})-F(x,\theta)\Big] = \sqrt{n}(\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(x,\theta) + \delta_n \qquad (7.10)$$

expanding by Taylor Series, where $\delta_n \longrightarrow 0$ with probability 1.

Define

$$T_n = \sqrt{n}(\hat{\theta}_n-\theta)$$

and $\gamma(t) = \frac{\partial}{\partial\theta}F(x,\theta)$ with $t = F(x,\theta)$

then

$$\hat{y}_n(t) = z_n(t) - T_n\gamma(t) + \delta_n$$

The covariance function $\rho_n(s,t)$ of $y_n(t)$ is given by

$$\hat{\rho}_n(s,t) = E\Big[z_n(t)z_n(s)\Big] - \gamma(s)E\Big[z_n(t)T_n\Big]$$

$$- \gamma(t)E\Big[z_n(s)T_n\Big]$$

$$+ E\Big[T_n^2\Big]\gamma(t)\,\gamma(s)$$

Now, it is well known that

$$E\Big[z_n(t)z_n(s)\Big] = \min(st) - st$$

and if $v^2 = -n\left\{E\Big[\frac{\partial^2}{\partial\theta^2}\log L\Big]\right\}^{-1}$

then $E\Big[T_n^2\Big] = k^2$, since $\hat{\theta}_n$ is efficient.

Define

$$h_n(t) = E\Big[z_n(t)\,T_n\Big] \qquad (7.11)$$

then

$$\hat{\rho}_n(s,t) = \min(s,t) - st - \gamma(s)\,h_n(t) - \gamma(t)\,h_n(s) + v^2\gamma(t)\gamma(s).$$

Now follow Darling (1955) and express $z_n(t)$ in terms of the indicator function:

$$\psi_t(s) = \begin{cases} 1 & s \leq t \\ 0 & s > t \end{cases}$$

Then

$$z_n(t) = \sqrt{n} \sum_{j=1}^{n} \left[\psi_t(F(X_j,\Theta))-t\right]$$

Suppose $\alpha \subset \{1,2,\ldots,n\}$, then

$$h_n(t) = \sqrt{n}\; E\left[\frac{1}{n}\left(\sum_{j=1}^{n} \psi_t(F(X_j,\Theta))-t\right)T_n\right]$$

$$= \sqrt{n}\; E\left[T_n\psi_t(F(X_\alpha,\Theta))\right] - \sqrt{n}\, t\, E[T_n]$$

$$= \sqrt{n}\, t\, E\left[T_n | F(X_\alpha,\Theta) < t\right] - \sqrt{n}\, t\, E[T_n]$$

and $h_n'(t) = \sqrt{n}\; E\left[T_n | F(X_\alpha,\Theta) = t\right] - \sqrt{n}\, E[T_n]$.

These results are Darling's [2], his lemma 3.3, with $\alpha = 1$.

Now if $p = F(x_p,\Theta)$ for Type I censoring and $p = F(x_{(r)},\Theta)$ for Type II censoring, then we have the following Lemma:

<u>Lemma 7.1</u>. $h_n(t)$, defined by (7.11) satisfies

$$h_n(t) = h_n(p) \quad \text{for } p < t \leq 1$$

<u>Proof</u> By integration,

$$E\left[T_n | F(X_\alpha,\Theta) < t\right] = \frac{p}{t} E\left[T_n | F(X_\alpha,\Theta) < p\right] + \frac{t-p}{t} E\left[T_n | F(X_\alpha,\Theta) \in (p,t)\right]$$

Thus

$$h_n(t) = \sqrt{n}\, p\, E\left[T_n | F(X_\alpha,\Theta) < p\right]$$

$$+ \sqrt{n}\,(t-p)\, E\left[T_n | F(X_\alpha,\Theta) \in (p,t)\right]$$

$$- \sqrt{n}\, p\, E[T_n] - \sqrt{n}\,(t-p)\, E[T_n]$$

$$= h_n(p) + \sqrt{n}\,(t-p) \left\{E\left[T_n | F(X_\alpha,\Theta) \in (p,t)\right] - E[T_n]\right\}$$

Now since T_n is based on the observed variables X_i such that $F(X_i,\Theta) \leq p$, then $E\left[T_n | F(X_\alpha,\Theta) \in (p,t)\right] = E[T_n]$, since T_n is independent of X_α if $F(X_\alpha,\Theta) \in (p,t)$. Thus $h_n(t) = h_n(p)$ since $E[T_n] = 0$.

Let us now consider the form of $h_n(t)$ for $t \leq p$. We consider the function $h_n(t)$ for Type I censoring only, since the two types of censoring are asymptotically equivalent. We have variables $(X_1,\ldots,X_r) \leq x_p$ which are observed and variables $(X_{r+1},\ldots,X_n) > x_p$ which are not observed, and r is a random variable. Let $(X_1',\ldots,X_n')$ be a randomization of $(X_1,\ldots,X_r,\ X_{r+1},\ldots,X_n)$.

From (7.6) we have

$$L = \frac{n!}{r!N!}\left[1-F(x_p,\Theta)\right]^N \prod_{i=1}^{r} f(x_i,\Theta)$$

Now since L is a p.d.f. then $\int_C L\,dx = 1$, where C is the sample space, and so $E\left[\frac{\partial}{\partial\Theta}\log L\right] = 0$. Now since it is assumed $\hat{\Theta}_n$ is efficient then:

$$\frac{n}{v^2}(\hat{\Theta}_n-\Theta) = \frac{\partial}{\partial\Theta}\log L$$

i.e. $T_n = \frac{v^2}{\sqrt{n}}\frac{\partial}{\partial\Theta}\log L$.

Thus $\quad h_n'(t) = v^2 E\left[\frac{\partial}{\partial\Theta}\log L \middle| F(X_\alpha,\Theta) = t\right] - v^2 E\left[\frac{\partial}{\partial\Theta}\log L\right]$

Now Lemma 7.2:

<u>Lemma 7.2</u>. If L is given by (7.6) and $t = F(x(t),\Theta)$

$$E\left[\frac{\partial}{\partial\Theta}\log L \middle| F(X_\alpha,\Theta) = t\right] = \frac{\partial}{\partial\Theta}\log f(x(t),\Theta) \quad \text{for } t \leq p$$

$$= 0 \qquad\qquad \text{for } t > p.$$

<u>Proof</u> Now

$$\frac{\partial}{\partial\Theta}\log L = -N\,\frac{\frac{\partial}{\partial\Theta}F_o(x_p,\Theta)}{1-F_o(x_p,\Theta)} + \sum_{i=1}^{r}\frac{\partial}{\partial\Theta}\log f_o(x_i,\Theta)$$

Now $p = F_o(x,p)$ and write $\frac{\partial F_p}{\partial\Theta} = \frac{\partial}{\partial\Theta}F_o(x_p,\Theta)$

$$\sum_{i=1}^{r} \frac{\partial}{\partial\theta}\log f(x_i,\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(x_i',\theta) - \sum_{i=r+1}^{n} \frac{\partial}{\partial\theta}\log f(x_i,\theta)$$

Thus $\frac{\partial}{\partial\theta}\log L = -N \dfrac{\frac{\partial F_p}{\partial\theta}}{1-p} + \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(x_i',\theta) - \sum_{i=r+1}^{r} \frac{\partial}{\partial\theta} f(x_i,\theta)$

We have for $t \leq p$

$$E\left[\frac{\partial}{\partial\theta}\log L \middle| F(X_\alpha,\theta) = t\right]$$

$$= \frac{-N}{1-p}\frac{\partial F_p}{\partial\theta} + \frac{\partial}{\partial\theta}\log f(x(t),\theta)$$

$$- E\left[\sum_{i=r+1}^{n} \frac{\partial}{\partial\theta}\log f(X_i,\theta)\right]$$

where $t = F_o(x(t),\theta)$, and $t \leq p$.

Since $E\left[\frac{\partial}{\partial\theta}\log f(X_i',\theta)\right] = 0$, the $X_1',\ldots,X_n'$ being equivalent to a complete random sample of size n from the population with c.d.f. $F(x,\theta)$, all that remains is to evaluate

$$E\left[\sum_{i=r+1}^{n} \frac{\partial}{\partial\theta}\log f(X_i,\theta)\right].$$

Since $E\left[\frac{\partial}{\partial\theta}\log L\right] = 0$ we have

$$0 = \frac{-N}{1-p}\frac{\partial F_p}{\partial\theta} + E\left[\sum_{i=r}^{r} \frac{\partial}{\partial\theta}\log f(x_i,\theta)\right]$$

$$= \frac{-N}{1-p}\frac{\partial F_p}{\partial\theta} + E\left[\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(X_i',\theta)\right] - E\left[\sum_{i=r+1}^{n} \frac{\partial}{\partial\theta}\log f(X_i,\theta)\right]$$

But the second term is 0, so

$$E\left[\sum_{i=r+1}^{n} \frac{\partial}{\partial\theta}\log f(X_i,\theta)\right] = \frac{-N}{1-p}\frac{\partial F_p}{\partial\theta}.$$

Thus $E\left[\frac{\partial}{\partial\theta}\log L \mid F(X_\alpha,\theta) = t\right]$

$$= \frac{\partial}{\partial\theta}\log f(x(t),\theta)$$

where $t = F(x(t),\theta)$.

For $t > p$,

$$E\left[\frac{\partial}{\partial\theta}\log L \mid F(X_\alpha,\theta) = t\right] = E\left[\frac{\partial}{\partial\theta}\log L\right]$$

$$= 0 \quad .$$

A similar result holds for Type II censoring, but now $p = F(x_{(r)},\theta)$. However as $n \to \infty$, $x_{(r)} \to x_p$ where $p = \frac{r}{n}$ with probability 1. From Darling (1955)

$$\gamma'(t) = \frac{dx}{dt}\frac{\partial}{\partial\theta}f(x,\theta) = \frac{\partial}{\partial\theta}\log f(x,\theta) = \frac{h_n'(t)}{v^2}$$

and since $h_n(0) = 0$ ($z_n(0)$ is 0 with probability 1) $\gamma(t) = \frac{h_n'(t)}{v^2}$. We can now give the covariance function $\rho(s,t)$ of the limit of the process $y_n(t)$, as $n \to \infty$.

Theorem 7.1

The process $\hat{y}_n(t)$ defined by (7.8) converges in distribution to a Gaussian process $\hat{y}(t)$ with mean 0 and covariance

$$\hat{\rho}(s,t) = \min(s,t) - st \begin{cases} -v^2\gamma(s)\gamma(t) & 0 \le s,\ t \le p \\ +v^2\gamma(s)\gamma(t) - & 0 \le s \le p \le t \le 1 \\ -v^2\gamma(s)\gamma(p) - v^2\gamma(t)\gamma(p) & \\ +v^2\gamma(s)\gamma(t) & p \le s,\ t < 0. \end{cases}$$

Proof The proof follows from Lemma 7.2.

We note that the covariance function $\hat{\rho}(s,t)$ is not symmetric over the interval (0,1), but it is over (0,p), an important result we use later.

7.2.2. Extension of results to vector parameter $\underset{\sim}{\theta}$ and doubly censored data

We can extend Theorem 7.2 to cover the case where $\underset{\sim}{\theta}$ is a q-valued vector parameter $(\theta_1, \theta_2, \ldots, \theta_q)$, and where the sample is doubly censored. We follow closely the results of Sukhatme (1972) and Stephens (1973) for the complete sample.

If $\hat{\underset{\sim}{\theta}}$ is an efficient estimator of $\underset{\sim}{\theta}$ then

$$\underset{\sim}{\theta}_n - \underset{\sim}{\theta}_o = \frac{1}{n} V \frac{\partial \log L}{\partial \underset{\sim}{\theta}}$$

where $$(V^{-1})_{ij} = -E\left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right]$$

and $\frac{\partial \log L}{\partial \underset{\sim}{\theta}}$ is a column vector with $\left(\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right)_i = \frac{\partial \log L}{\partial \theta_i}$

Following (7.9)

$$\sqrt{n}\left\{\hat{F}(x) - F_o(x)\right\} = \sqrt{n}\,(\hat{\underset{\sim}{\theta}}_n - \theta)^T \left\{\frac{\partial}{\partial \underset{\sim}{\theta}} F(x, \theta)\right\} + \delta_n$$

where $\delta_n \longrightarrow 0$ with probability 1, and

$$\left\{\frac{\partial}{\partial \underset{\sim}{\theta}} F(x, \underset{\sim}{\theta})\right\}_i = \frac{\partial}{\partial \theta_i} F(x, \underset{\sim}{\theta}) .$$

Defining $z_n(t)$ as before (7.10) then

$$\hat{y}_n(t) = z_n(t) - T_n \gamma(t) + \delta_n,$$

where $$\underset{\sim}{T}_n = \sqrt{n}(\underset{\sim}{\theta}_n - \underset{\sim}{\theta})$$

$$\underset{\sim}{\gamma}(t) = \frac{\partial}{\partial \underset{\sim}{\theta}} F_o(x, \underset{\sim}{\theta}) \quad \text{where } t = F_o(x, \underset{\sim}{\theta}).$$

The covariance function $\hat{\rho}_n(s,t) = E\left[y_n(s) y_n(t)\right]$ is given by

$$\hat{\rho}_n(s,t) = E\left[y_n(s) y_n^T(t)\right]$$

$$= E\left[z_n(s) z_n(t)\right] - \underset{\sim}{\gamma}^T(s) E\left[z_n(t) T_n\right]$$

$$- E\left[z_n(s)T_n^T\right]\gamma(t) + \gamma^T(s)E\left[T\ T^T\right]\gamma(t)$$

where $E\left[\underset{\sim}{T}\ \underset{\sim}{T}^T\right] = V.$

As before we study the vector function

$$\underset{\sim}{h}_n(t) = E\left[z_n(t)\underset{\sim}{T}_n\right] \qquad (7.13)$$

We have that

$$\underset{\sim}{T}_n = \sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta})$$

$$= \frac{1}{\sqrt{n}} V \frac{\partial \log L}{\partial \underset{\sim}{\theta}}$$

Thus $\underset{\sim}{h}_n(t) = E\left[\frac{1}{n}\left\{\sum_{i=1}^{n} \psi_t(F(X_i,\underset{\sim}{\theta})) - t\right\} V\left\{\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right\}\right]$

and so $V^{-1}\underset{\sim}{h}_n(t) = t\left[\left\{\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right\} \middle| F(X_\alpha,\underset{\sim}{\theta}) < t\right]$

$$- t\,E\left[\left\{\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right\}\right]$$

and $\quad V^{-1}\underset{\sim}{h}_n'(t) = E\left[\left\{\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right\} \middle| F(X_\alpha,\underset{\sim}{\theta}) = t\right]$

$$- E\left[\left\{\frac{\partial \log L}{\partial \underset{\sim}{\theta}}\right\}\right]$$

If we look at the i^{th} element of the RHS, we see by referring to Lemma 7.2 that this is equal to $\frac{\partial}{\partial \theta_i} \log f(x(t),\theta)$ where $x(t) = F(x(t),\underset{\sim}{\theta})$ and $t \leq p$. Thus we have Theorem 7.4.

Theorem 7.4

The process $y_n(t)$ defined (7.8) converges in distribution to a Gaussian process $y(t)$ with mean 0 and covariance

$$\hat{\rho}(s,t) = \min(s,t) - st = \begin{cases} -\underset{\sim}{\gamma}^T(s)\, V\underset{\sim}{\gamma}(t) & 0 \leq s,\ t \leq p \\ +\underset{\sim}{\gamma}^T(s)\, V\underset{\sim}{\gamma}(t) - \underset{\sim}{\gamma}^T(s)\, V\underset{\sim}{\gamma}(p) & \\ -\underset{\sim}{\gamma}^T(p)\, V\underset{\sim}{\gamma}(t) & 0 \leq s \leq p \leq t \leq 1 \\ +\underset{\sim}{\gamma}^T(s)\, V\underset{\sim}{\gamma}(t) & \\ -2\underset{\sim}{\gamma}^T(p)\, V\underset{\sim}{\gamma}(p) & 0 \leq s,\ t \leq 1 \end{cases}$$

If we now consider doubly censored samples, we have the likelihood

$$L \,\alpha \left[F(x_q, \underset{\sim}{\theta})\right]^N \times \prod_{i=1}^{r} f(x_i, \underset{\sim}{\theta}) \times \left[1 - F(x_p, \underset{\sim}{\theta}\right]^N$$

where $n = M + r + N$.

For Type I censoring x_q and x_p are fixed, for Type II censoring $x_q = x_{(1)}$ and $x_p = x_{(r)}$. Asymptotically the two types of censoring are the same. We take $q = \frac{M}{n}$ and $1 - p = \frac{N}{n}$ in the limit. Lemma 7.1 can be extended:

<u>Lemma 7.5</u> $h_n(t)$ defined by (7.11) satisfies

$$\underset{\sim n}{h}(t) = \begin{cases} h_n(q), & 0 \leq t < q \\ h_n(p), & p < t < 1 \end{cases}$$

Also Lemma 7.2 can be extended for this case. Consider the random variables $(X_1, \ldots, X_r), x_q \leq X_1, \ldots, X_r \leq x_p$ which are observed and random variables $(Y_1, \ldots, Y_M)$ and $(Z_1, \ldots, Z_N)$ which are not observed, where $Y_1, \ldots, Y_M < x_p$ and $Z_1, \ldots, Z_n > x_p$. The consider $(X_1', \ldots, X_n')$ which is a randomization of $(X_1, \ldots, X_r, Y_1, \ldots, Y_M, Z_1, \ldots, Z_N)$, we then have Lemma 7.6.

<u>Lemma 7.6</u> For doubly censored data, where L is given by (7.14)

$$E\left[\frac{\partial}{\partial\theta_i} \log L \middle| F(X_\alpha, \underset{\sim}{\theta})\right] = \begin{cases} \frac{\partial}{\partial\theta_i} \log f(x(t)\theta) & \text{for } q \leq t \leq p \\ 0 & t > q,\ t < p \end{cases}$$

where $t = F(x(t), \underset{\sim}{\theta})$.

<u>Proof</u> The proof follows Lemma 7.2.

Finally,

<u>Theorem 7.7</u>

The process $\hat{y}_n(t)$ defined by (7.8) converges in distribution to a Gaussian process $\hat{y}(t)$ with mean 0 and covariance given by $\hat{\rho}(s,t)$:

$$\hat{\rho}(s,t) = \min(s,t) - st$$

$$\left\{ + \gamma^T(s) V\gamma(t) - 2\gamma^T(q) V\gamma(q) \qquad 0 \leq s,\ t < q \right.$$

$$\left\{ \begin{array}{l} + \gamma^T(s) V\gamma(t) - \gamma^T(s) V\gamma(q) \\ \qquad\qquad - \gamma^T(q) V\gamma(t) \qquad 0 < s \leq q \leq t \leq p \end{array} \right.$$

$$\left\{ - \gamma^T(s) V\gamma(t) \qquad\qquad q \leq s,\ t \leq p \right.$$

One major practical problem in the above analysis is that we have assumed the existence of an efficient estimator $\hat{\theta}_n$ of θ. However, following Darling (1955) and Sukhatme (1972), the maximum likelihood estimator, which is asymptotically efficient, when used to estimate θ in $F(x,\theta)$, will provide a function $h_n(t)$ which will converge, with probability one, to $V.\frac{\partial \log L}{\partial \underset{\sim}{\theta}}$. Chernoff et al (1967) present estimators which are linear combinations of function of the order statistics of the sample and are asymptotically efficient. Linear estimators are also given by Plackett (1958) and shown to be asymptotically efficient. We have, then, various methods of estimation which are satisfactory in large samples. The bias of these estimators is usually $O(\frac{1}{n})$ or less, but this bias can be reduced by various methods.

Sarhan and Greenberg (1962) present linear unbiased estimators which are minimum variance. In the circumstances it seems that these are best for small samples.

7.3 The weighted χ^2 representation of $\hat{W}^2_p$

7.3.1. Extension of the theory to $_p\hat{W}^2$.

Darling (1955) shows that $\hat{W}^2$, when estimating one parameter efficiently, has the same distribution as

$$\sum_{j=1}^{\infty} \frac{Z_j^2}{\hat{\lambda}_j}$$

where the Z_j, $j = 1,2,\ldots$ are i.i.d. standard normal random variables, and the $\hat{\lambda}$'s are the eigenvalues of the integral equation

$$f(t) = \hat{\lambda} \int_0^1 \hat{\rho}(s,t)\, f(s)\, ds \qquad (7.15)$$

where $\hat{\rho}(s,t)$ is given by Darling's Lemma 4.1 or by Theorem 7.8 with $p = 1$.

Recently Sukhatme (1972) has extended Darling's work to include the case of when q parameters are estimated efficiently. She shows that if the covariance function $\hat{\rho}(s,t)$ can be written in symmetric form:

$$\hat{\rho}(s,t) = \min(s,t) - st - \sum_{i=1}^{q} \phi_i(s)\, \phi_i(t) \qquad (7.16)$$

then the eigenvalues of the integral equation (7.15), but with $\hat{\rho}(s,t)$ given by (7.16), are given by the solution of $D(\lambda) = 0$, where $D(\lambda)$ is defined by:

$$D(\lambda) = d(\lambda)\Delta(\lambda) \qquad (7.17)$$

where $\Delta(\lambda)$ is the determinant of the symmetric matrix

$$P(\lambda) = (P_{ij}(\lambda)), \text{ where}$$

$$\left.\begin{aligned} P_{ii}(\lambda) &= 1 + \lambda \sum_{j=1}^{\infty} a_{ij}^2 \big/ (1-\lambda/\lambda_j) \\ P_{ij}(\lambda) &= \lambda \sum_{k=1}^{\infty} a_{ik}\, a_{jk} \big/ (1-\lambda/\lambda_k) \end{aligned}\right\} \qquad (7.18)$$

where $$a_{ij} = \int_0^1 \phi_i(s) f_j(s)\, ds \qquad i = 1,..,q$$
$$j = 1,2,..$$

and $d(\lambda)$ is the Fredholm determinant of the integral equation with

$$\rho(s,t) = \min(s,t) - st.$$

and

$$d(\lambda) \equiv \prod_{j=1}^{\infty} (1-\lambda/\lambda_j)$$

Now, when q parameters are estimated by an asymptotically efficient method, we have, by Theorem 7.4 with p = 1,

$$\hat{\rho}(s,t) = \rho(s,t) - \gamma^T(s) V \gamma(t) \qquad (7.19)$$

Now V is the covariance matrix, so since V is non-singular we can diagonalise this matrix:

$$V = R^T W R$$

where W is a diagonal matrix and R an orthogonal matrix. Thus by writing:

$$\underset{\sim}{\phi}(s) = W^{\frac{1}{2}} R \underset{\sim}{\gamma}(t) \qquad (7.20)$$

we have

$$\underset{\sim}{\gamma}^T(s) V \underset{\sim}{\gamma}(t) = \underset{\sim}{\gamma}^T(s) R^T W R \underset{\sim}{\gamma}(t)$$
$$= \phi^T(s)\, \phi(t)$$
$$= \sum_{i=1}^{q} \phi_i(s)\, \phi_i(t)$$

and so the covariance function can always be written in the required form.

We now extend this theory to the case where there are observations censored to (o, p), and then have to find the eigenvalues $\hat{\mu}_j$ of the integral equation

$$g(t) = \hat{\mu} \int_o^p \hat{\rho}(s,t)\, g(s)\, ds$$

where $\hat{\rho}(s,t)$ is given by Theorem 7.4 and g(t) is given by (6.32), then

$$_p\hat{W}^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{\hat{\mu}_j}$$

Following Sukhatme (1972), these eigenvalues can be found by solving D(μ) where D(.) is given by (7.17) and (7.18), where (i), the λ_j's are replaced by the μ_j's, where the μ_j's are given by (6.32A), the eigenvalues of the integral equation

$$g(t) = \mu \int_0^p \rho(s,t)\; g(s)\; ds.$$

and (ii), the a_{ij} are redefined by:

$$a_{ij} = \int_0^p \phi_i(s)\; g_j(s)\; ds \qquad i = 1,2,..,q$$
$$j = 1,2,\ldots .$$

where the $\phi_i(s)$ is given by (7.20) and Theorem 7.4.

Then $\Delta(\mu) = 0$ is solved to find the eigenvalues $\hat{\mu}_j$, $j = 1,2,\ldots$.

7.3.2. The matrix approach to $_p\hat{W}^2$.

Alternatively, from §6 we have that $\rho(s,t)$, where

$$\rho(s,t) = \min(s,t) - st,$$

can be represented by a sum of orthogonal functions (see 6.38)

$$\rho(s,t) = \underset{\sim}{g}^T(s)\; M^{-1}\; \underset{\sim}{g}(t) \qquad 0 \le s,\; t \le p$$

where $\int_0^p g^T(s)\underset{\sim}{g}(s)\; ds = I$, the infinite identity matrix

and M is given by (6.40), $\underset{\sim}{g}(t)$ by (6.39).

Now when q parameters are estimated:

$$\hat{\rho}(s,t) = \rho(s,t) - \underset{\sim}{\gamma}^T(s)\, V\, \underset{\sim}{\gamma}(t)$$

Since the $\underset{\sim}{g}(t)$ are a complete orthonormal system of functions on (o,p) we can express $\underset{\sim}{\gamma}(t)$ as a generalized Fourier series in the functions $\underset{\sim}{g}(t)$ thus:

$$\underset{\sim}{\gamma}(t) = G\, \underset{\sim}{g}(t) \quad \text{where}$$

$$G_{ij} = \int_o^p \gamma_i(s)\, g_j(s)\, ds \qquad i = 1,\ldots,q$$
$$j = 1,2,\ldots,$$

Then $\underset{\sim}{\gamma}^T(s)\, V\, \underset{\sim}{\gamma}(t) = \underset{\sim}{g}^T(s)\, G^T V G\, \underset{\sim}{g}(t)$

and $\hat{\rho}(s,t) = \underset{\sim}{g}^T(s) \left\{M^{-1} - G^T V G\right\} \underset{\sim}{g}(t) \qquad 0 \leq s,\ t \leq p$

(7.21)

Now let $A = M^{-1} - G^T V G$ and suppose A can be diagonalized by finding its eigenvectors and eigenvalues:

$A = Q^T E Q$ where $Q^T Q = I$ and E^{-1} is a diagonal matrix with entries $\varepsilon_1, \varepsilon_2, \ldots$. Then $\hat{\rho}(s,t)$ can be expressed as a sum of orthogonal functions $\underset{\sim}{h}(t)$ where

$$\underset{\sim}{h}(t) = Q^T \underset{\sim}{g}(t)$$

and

$$\hat{\rho}(s,t) = h^T(t)\, E\, h(t)$$

$$= \sum_{j=1}^{\infty} \frac{1}{\varepsilon_j} h_j(s) h_j(t), \quad 0 \leq s,t \leq p$$

Now following Darling (1955) we can represent the Gaussian process $\hat{y}(t)$ thus

$$\hat{y}(t) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\varepsilon_j}} h_j(t) Z_j \quad , \quad 0 \leq t \leq p \qquad (7.22)$$

where the Z_j are i.i.d. normal random variables, and ${}_p\hat{W}^2$ is distributed as $\sum_{j=1}^{\infty} \frac{Z_j^2}{\varepsilon_j}$ with probability 1.

7.4. Testing for normality with ${}_pW^2$

7.4.1. The covariance function for ${}_p\hat{W}^2$.

Given a random sample $X_1,\ldots,X_r$ which has been singly censored, by either type I or II methods, then it may be required to test the observations $x_1,\ldots x_r$ for normality. We can estimate the unknown μ and σ^2 by the maximum likelihood method. Following Cohen (1961), these estimates are found to be complicated and require an auxiliary function $\hat{\lambda}$, which he tabulates. These estimates are given by:

$$\left.\begin{array}{ll} \text{Type I censoring:} & \hat{\mu} = \bar{x} - \lambda(\bar{x}-x_p) \\ & \hat{\sigma}^2 = s^2 + \hat{\lambda}(\bar{x}-x_p)^2 \\ \text{Type II censoring:} & \mu = x - \lambda(x-x_{(r)}) \\ & \hat{\sigma}^2 = s^2 + \hat{\lambda}(x-x_{(r)})^2 \end{array}\right\} \quad 7.26$$

where $\bar{x} = \sum_{i=1}^{r} x_i/n, \quad s^2 = \sum_{i=1}^{r} (x_i-\bar{x})^2/n$

These estimates are biased for μ and σ, and corrections for the bias are given by Saw (1961). These are of the form

$$\hat{\mu}_c = \hat{\mu} - \frac{\hat{\sigma}}{n+1} B(\mu,p)$$

$$\hat{\sigma}_c = \hat{\sigma} - \frac{\hat{\sigma}}{n+1} B(\sigma,p)$$

where $p = r/n$. The bias is considerable for small samples. For example with $p = .5$,

$$B(\mu) = -.960$$

$$B(\sigma) = -1.762$$

Other methods of estimating μ and σ which are linear estimates are summarized in David (1970 §6.3) and Sahren and Greenberg (1962). However asymptotically these methods are equivalent, since the maximum likelihood estimates of μ and σ are asymptotically linear, (see Plackett (1958)).

The covariance matrix for the estimates (7.26) is given by the computation of (7.11). This was found by Cohen (1961) and given in David (1970, p114). If $V^{-1} = W$ then

$$
\begin{aligned}
\omega_{11} &= p\{B(A^{-}+y)+1\} \\
\omega_{12} &= p\{B[1+y(A^{-}+y)]\} \qquad (7.27) \\
\omega_{22} &= (2p+y\omega_{12})
\end{aligned}
$$

where $\quad y = \Phi^{-1}(p), \quad A^{-} = \phi(y)/p, \quad B = \frac{p}{1-p} A^{-}$

Here V_{11} corresponds to the variance of $\hat{\mu}$, etc.

The distribution of ${}_p\hat{W}^2$ can be found now by the methods of §7.3. The function $\underset{\sim}{\gamma}(t)$ defined in (7.10) is given by:

$$
\begin{aligned}
\gamma_1(t) &= \frac{\partial}{\partial\mu}\Phi\left(\frac{x-\mu}{\sigma}\right) \\
&= -\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{where } t = \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

and $\Phi(z)$ is the standard normal c.d.f. and $\phi(z) = \Phi'(z)$

and $\gamma_2(t) = \frac{\partial}{\partial\sigma}\Phi\left(\frac{x-\mu}{\sigma}\right)$

$$
= -\frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right)
$$

Making the substitution $t = \Phi\left(\frac{x-\mu}{\sigma}\right)$, we find

$$
\left.
\begin{aligned}
\gamma_1(t) &= \frac{-1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\Phi^{-1}(t)\right)^2\right] \\
\gamma_2(t) &= \Phi^{-1}(t) \times \gamma_1(t)
\end{aligned}
\right\} \qquad (7.28)
$$

The cumulants of ${}_p\hat{W}^2$ can be calculated by quadrature using formula (6.31) but with $\hat{\rho}(s,t)$ given by (7.19) and $\gamma(t)$ by (7.28). The first two cumulants are given in Table 7.2 for selected values of p.

7.4.2. The eigenvalues of ${}_p\hat{W}^2$ by solving the integral equation.

The eigenvalues of ${}_p\hat{W}^2$ can be found by solving the integral equation, and this is shown in § 7.3 to be equivalent to solving the equation $\Delta(\mu) = 0$, where $\Delta(\mu)$ is the determinant of the matrix defined in (7.18).

Following Sukhatme (1972), we note if we write

$$r = v_{12}(v_{11}v_{22})^{-\frac{1}{2}}$$

then $$\gamma^T(t)\,V\,\gamma(s) = \psi^T(t)\,\psi(s)$$

where $$\psi_1(t) = \sqrt{v_{11}}\,(1-r^2)^{\frac{1}{2}}\,\gamma_1(t)$$

$$\tag{7.29}$$

and $$\psi_2(t) = r\sqrt{v_{11}}\,\gamma_1(t) + \sqrt{v_{22}}\,\gamma_2(t)$$

then $$\Delta(\mu) = P_{11}(\mu)\,P_{22}(\mu) - P_{12}^2(\mu) \tag{7.29A}$$

where $P_{ii}(\mu) = 1 + \mu \sum_{j=1}^{\infty} a_{ij}^2/(1-\mu/\mu_j)$

for i = 1,2 and

$$a_{ij} = \int_0^p \psi_i(t)\,g_j(t)\,dt \qquad i = 1,2 \qquad j = 1,2,\ldots$$

where ψ_i is given by (7.29) and $g_j(t)$ is given by (6.32)

and $$P_{12} = \mu \sum_{j=1}^{\infty} a_{1j}a_{2j}/(1-\mu/\mu_j)$$

We note that, for $p < 1$, the $a_{ij} \neq 0$ $i = 1,2;\ j = 1,2,\ldots$ and so none of the standard eigenvalues μ_j are roots of

$D(\mu)$ where $D(\mu) = d(\mu)\ \Delta\ (\mu)$ and

$$d(\mu) = \prod_{j=1}^{\infty} \left(1 - \frac{\mu}{\mu_j}\right)$$

To solve $\Delta(\mu) = 0$, it is necessary to truncate the summations involved in the P_{ij}. These were truncated at about 20 terms, exactly where depending on the number of μ_j known for each p. This method gives very accurate approximation to the eigenvalues, as shown in Stephens (1973).

The eigenvalues obtained by this method are given in Table 7.3 and they are seen to give reasonable approximations to the first two cumulants. Percentage points were then obtained by utilizing the known eigenvalues by Imhof's method as for ${}_pW^2$ in §6.5. The percentage points are given in Table 7.5 and agree well with the Monte Carlo distribution ${}_r\hat{W}^2_n$. Given that the eigenvalues used are the exact values, then the percentage points are accurate to the number of significant figures given. Also included are the percentage points of ${}_{p=1}\hat{W}^2$, that is the case of the complete sample, computed using Stephens' eigenvalues.

§7.4.3. The matrix approach to find distribution of ${}_p\hat{W}^2$

We can follow the matrix approach to find the distribution of ${}_p\hat{W}^2$ by finding the matrix $A = M^{-1} - G^T VG$ as defined (7.21), and calculating its eigenvalues. However, this method does not seem to provide very accurate values of the eigenvalues of ${}_p\hat{W}^2$. For example, with p = .5

the matrix A is taken to have dimension 18, since for higher dimension the evaluation of the elements $(A)_{ij}$ of A becomes impractical (it involves the quadrature of highly oscillatory functions over (o,p) of the form sin(iπ)sin(jπ). For this example the eigenvalues of the matrix A were

.00767	.00046	.00014
.00445	.00036	.00012
.00815	.00029	.00011
.00131	.00024	.00009
.00081	.00020	.00006
.00062	.00016	-.00015

The negative eigenvalues appears as a result of the inaccuracies in the calculation of the elements of A. The first 17 values give a mean of .01896 and a variance of .00017. When these sigenvalues are compared with those in Table 7.3, it is seen there is a considerable difference between the values.

§7.4.4 Power Study

A small Monte Carlo power study was carried out to compare the powers of ${}_r\hat{W}^2_n$ and ${}_rA^{2}_n$* with $\frac{r}{n}$ = .5, .9 and 1.0, when testing for normality. The alternatives considered gave a variety of values of $(\sqrt{\beta_1}, \beta_2)$. Asymmetric distributions ($\beta_1 \neq 0$) were censored on the left and on the right, and the powers of ${}_p\hat{W}^2$ and ${}_p\hat{A}^2$ comouted. The Table 7.6 gives the number of rejections of the null hypothesis, with 5% risk, under the null in 1000 random samples drawn from the given population. The percentage points of the statistics were found by Monte Carlo

*, introduced in §8,

simulation of the Statistics using 25,000 random samples.

Also computed was the 'power correlation' matrix $R_{(i,j)}$. We let i refer to statistic S_i and j to statistic S_j, then $R_{(i,j)}$ records the number of times S_i and S_j were found significant for that particular sample. We give below in Figure 7.1 the matrix for the null situation.

n = 40

A^2 r = 36 (p = .9)	51	16	14	16	37	40
A^2 r = 20 (p = .5)		52	11	42	13	12
W^2 r = 36 (p = .9)			49	12	36	42
W^2 r = 20 (p = .5)				53	14	13
A^2 Full Sample					54	46
W^2 Full Sample						52

5% Significance Points for Statistics

A^2 r = 36	.605	*
A^2 r = 20	.277	*
A^2 r = 40	.732	**
W^2 r = 36	.112	*
W^2 r = 20	.043	*
W^2 r = 40	.127	**

* Obtained from Monte Carlo empirical distribution using 25,000 samples.

** Obtained from Professor M.A.Stephens' modified test Statistics in a private communication 'Goodness-of-fit tests based on the empirical distribution function, and tests for uniformity'.

Figure 7.1 'Power correlation matrix' for null situation, 1000 random samples.

We can see from figure 7.1 that the statistics based on the full sample and a proportion .9 of the sample are highly correlated, in that the number of samples being rejected by two of the tests taken together, is near to the expected number of 50 out of 1000 rejections. However the tests W^2 and A^2 based on .5 of the sample although highly correlated with one another, are not highly correlated with the tests based on .9 of the sample and the full sample. This indicates that W^2 and A^2, with $p = .5$, are telling us something different about the sample from that which W^2 and A^2, based on .9 of the sample and the full sample tell us.

7.5. Testing for the exponential distribution in life testing with ${}_pW^2$

7.5.1. The mean and variance of ${}_p\hat{W}^2$

If we consider single censoring to the right of observations from an exponential distribiton with p.d.f.

$$G(x,\theta) = \theta^{-1} \exp(-x/\theta), \qquad x > 0$$

then the maximum likelihood estimate of θ is given by:

$$\hat{\theta} = \left(\sum_{i=1}^{r} x_i + N x_p\right)/r$$

for Type I censoring and

$$\hat{\theta} = \left(\sum_{i=1}^{r} x_i + N x_{(r)}\right)/r$$

for Type II censoring. The MLE in this case is the best linear unbiased estimate (B.L.U.E) (see Sarhan and Greenberg (1962, §118.2)).

The variance of the estimate is given by

$$\left\{E\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right]\right\}^{-1} = \theta^2/r = \theta^2/np \qquad (7.30)$$

By following the previous section, the asymptotic distribution of ${}_p\hat{W}^2$ can be found when θ is estimated. The function $\gamma(t)$ is given by:

$$\gamma(t) = \frac{\partial}{\partial\theta} F\left(\frac{x}{\theta}\right) \qquad (7.31)$$

$$= -(1-t)\log(1-t) \text{ when } F\left(\frac{x}{\theta}\right) \text{ is}$$

of exponential form.

We can find the mean and variance of ${}_p\hat{W}^2$ by using formula (6.31) (with $\hat{\rho}(s,t)$ defined with $\gamma(t)$ given by (7.31)) and these upon integration become:

$$\mu({}_p\hat{W}^2) = \mu_p - 2(1-q^3(1+\omega+\omega^2/2))/(27p) \qquad (7.32)$$

$$\sigma^2({}_p\hat{W}^2) = \sigma_p^2 - (8s-4a^2)/p + 2d^2 \qquad (7.33)$$

where for μ:

$$q = 1 - p$$

$$\omega = -3 \log q$$

μ_p is the mean of ${}_pW^2$

and for σ^2

$$s = (1-q^2(1+2f)) \times 5/144$$
$$-\tfrac{1}{4}((1-q^4(1+4f))/16 + \tfrac{1}{16}(1-q^4(1+4f(1+2f))))$$
$$+\tfrac{1}{9}(1-q^5(1+5f)) \times 1/25 + 3(1-q^5(1+5f \times$$
$$(1+2.5f)))2/125)$$

where $f = -\log q$ and

$d = \mu({}_p\hat{W}^2) - \mu_p$

σ^2 is the variance of ${}_pW^2$.

The values of the mean and variance are given in Table 7.7.

7.5.2. The asymptotic distribution of ${}_p\hat{W}^2$

To find the eigenvalues associated with ${}_p\hat{W}^2$, we follow the method of §7.3 and solve the equation $\Delta(\mu) = 0$ where now

$$\Delta(\mu) = P(\mu) \text{ and}$$

$$P(\mu) = 1 + \mu \sum_{j=1}^{\infty} a_j^2/(1-\mu/\mu_j)$$

where $a_j = \int_0^p v\,\gamma(t)\,g_j(t)\,dt$,

where $v^2 = p^{-1}$ and given by (7.30),

$\gamma(t)$ is given by (7.31).

and the μ_j are solutions of (6.32A).

We find none of the a_j's are zero so no standard eigenvalue μ_j is a solution of $D(\mu)$.

The summation in $P(\mu)$ was truncated at about the 20[th] term, the actual point depending on the number of eigenvalues for the particular value of p and solved for μ. These values, or rather μ^{-1}, are given in Table 7.8, and in Table 7.9 the cumulants calculated using these values. One sees, that as before, this approximation to the variance, is quite accurate, so indicating the accuracy of the eigenvalues calculated by solving $P(\mu) = 0$. Percentage points have been calculated, as before, using Imhof's method. These are given in Table 7.10, and should be accurate to at least three decimal places.

7.5.3. An alternative method for testing for goodness of fit.

A simpler way perhaps to test for the exponential distribution, when censored, is to make use of the well known fact that if $X_1, X_2, \ldots, X_n$ are i.i.d. exponential random variables, then the random variables Y_r, $r = 2, \ldots, n$ defined by

$$Y_r = (n-r+1)\left(X_{(r)} - X_{(r-1)}\right); \quad r = 1, 2, \ldots, n$$

with $X_o = 0$,

are also i.i.d. exponential random variables with the same parameter θ as the X's. Consequently if only the observations $X_{(s)}, X_{(s+1)}, \ldots, X_{(t)}$ are available, $1 \leq s < t \leq n$, due to censoring of some kind, then $t-s$ random variables Y_r can be generated from the observed X's.

$$Y_{r-s} = (n-r+1)\left(X_{(r)}-X_{(r-1)}\right), \; r = s+1,\ldots,t.$$

We can then test the $Y_1, \ldots, Y_{t-s}$ as being (t-s) independent observations from an exponential distribution, and use the methods of Seshadri, Csorgo and Stephens (1969) or Stephens (1973). However this method does not make use of the actual number of observations censored, and may lose power for small samples, in a similar way that we would lose efficiency if we estimated θ by using the Y_r's i.e. we effectively lose one observation.

TABLE 7.1

Empirical distribution of ${}_r\hat{W}_n^2$ when testing for normality using Cohen's estimates of μ and σ for Type II single censoring, with 5000 samples.

Significance points of ${}_r\hat{W}_n^2$, where r = pn

p = .5

n	50%	85%	90%	95%	99%
20	.0164	.0282	.0320	.0385	.0549
40	.0168	.0298	.0338	.0414	.0589
60	.0171	.0316	.0365	.0442	.0585
80	.0167	.0322	.0368	.0444	.0622
100	.0170	.0323	.0377	.0454	.0657
Estimated standard error	.0002	.0004	.0004	.0008	

p = .9

n	50%	85%	90%	95%	99%
20	.0456	.0807	.0923	.1122	.1554
40	.0459	.0831	.0947	.1140	.1587
60	.0447	.0815	.0935	.1141	.1651
80	.0454	.0810	.0922	.1124	.1617
100	.0451	.0812	.0943	.1152	.1631
Estimated standard error	.0004	.0009	.0011	.0023	

TABLE 7.2

Mean and variance of ${}_p\hat{W}^2$, testing for normality, found by quadrature.

p	$\mu \times 10^2$	$\sigma^2 \times 10^4$
1.00	5.946	11.669
.95	5.6852	10.986
.90	5.3283	9.858
.85	4.9268	8.646
.80	4.5022	7.394
.75	4.0687	6.177
.70	3.6359	5.042
.65	3.2102	4.015
.60	2.7966	3.113
.55	2.3991	2.341
.50	2.0224	1.700

TABLE 7.3

Eigenvalues for ${}_p\hat{W}^2$ testing for normality with mean and variance unknown, found by solving $\Delta(\mu) = 0$ as defined in (7.29A). Values of $\mu^{-1} \times 10^2$.

p	1.0	.95	.90	.80	.70	.60	.50
	1.834	1.797	1.723	1.494	1.201	.881	.580
	1.344	1.272	1.190	1.009	.814	.618	.434
	.535	.518	.486	.402	.307	.217	.145
	.436	.409	.382	.316	.248	.184	.127
	.252	.242	.224	.181	.136	.096	.065
	.216	.201	.186	.153	.118	.085	.059
	.146	.139	.128	.102	.076	.054	.039
	.129	.121	.111	.090	.048	.050	.034
	.095	.090	.082	.065	.034	.035	
	.085	.080	.073	.045	.025	.032	

TABLE 7.4

Cumulants calculated from the eigenvalues in Table 7.3

p	κ_1	$\kappa_2 \times 10^5$	$\kappa_3 \times 10^6$
.95	.05080	10.986	64.8
.90	.04044	9.407	55.4
.80	.03968	7.192	35.8
.70	.03050	4.613	18.6
.60	.02243	2.523	7.49
.50	.01970	1.450	2.73

TABLE 7.5

Asymptotic percentage points of ${}_p\hat{W}^2$, testing for normality with mean and variance unknown, found by Imhof's method.

Percentage Point	p .5	.6	.7	.8	.9	.95	1.00
1.0	.0037	.0049	.0071	.0103	.0133	.0147	.0161
2.5	.0047	.0063	.0090	.0127	.0161	.0177	.0188
5.0	.0058	.0079	.0111	.0151	.0190	.0207	.0219
10.0	.0073	.0102	.0141	.0186	.0229	.0249	.0262
15.0	.0086	.0121	.0164	.0213	.0261	.0282	.0296
50.0	.0170	.0238	.0311	.0383	.0454	.0486	.0509
85.0	.0324	.0445	.0569	.0695	.0815	.0867	.0905
90.0	.0373	.0510	.0653	.0799	.0936	.0933	.1036
95.0	.0457	.0621	.0797	.0978	.1143	.1211	.1261
97.5	.0539	.0731	.0943	.1159	.1353	.1431	.1487
99.0	.0648	.0877	.1137	.1401	.1633	.1725	.1789

TABLE 7.6

Empirical power of 5% tests for normality, with mean and variance unknown. Based on the number of rejections in 1000 random samples. R denotes censoring on the right, L denotes censoring on the left, for asymmetric distributions. Uncensored samples of all size 40.

Alternative Population		$\sqrt{\beta_1}$	β_2	$\hat{W}^2_n$	$_{.9n}\hat{W}^2_n$	$_{.5n}\hat{W}^2_n$	$\hat{A}^2_n$ *	$_{.9n}\hat{A}^2_n$	$_{.5n}\hat{A}^2_n$
u(0,1)		0.0	1.8	.37	.25	.33	.51	.36	.39
χ^2_2	R	2.0	9.0	.97	.93	.60	.99	.96	.66
	L			.97	.82	.34	.99	.84	.35
χ^2_4	R	1.414	6.0	.74	.64	.29	.81	.71	.30
	L			.76	.56	.23	.83	.57	.23
χ^2_6	R	1.155	5.0	.56	.44	.21	.64	.49	.20
	L			.54	.36	.20	.64	.38	.20
Cauchy		-	-	.98	.94	.91	.99	.94	.91
Student's t 2 d.o.f.		-	-	.75	.54	.54	.78	.57	.57
4 d.o.f.		0	-	.32	.22	.22	.36	.24	.24
6 d.o.f		0	6	.15	.09	.15	.19	.13	.15
10 d.o.f		0	4	.08	.07	.10	.10	.08	.11

* The A^2_n statistics are introruced in § 8

TABLE 7.7

Exact Mean and variance of ${}_p\hat{W}^2$ testing for exponentiality censored on the right, using formulae (7.33) and (7.34)

p	$\mu \times 10^2$	$\sigma^2 \times 10^3$
.99	9.1803	4.2871
.95	8.7977	3.9813
.90	8.2310	3.5481
.85	7.6127	3.0920
.80	6.9669	2.6405
.75	6.3167	2.2114
.50	3.2250	.64338

TABLE 7.8

Values of $\mu^{-1} \times 10^2$ found by solving $\Delta(\mu) = 0$ as defined in § 7.4, when ${}_p\hat{W}^2$ is used for testing for the exponential distribution with scale parameter unknown and censored on the right.

p	1.0	.95	.90	.50
	4.202	4.022	3.804	1.658
	1.712	1.632	1.532	.593
	.815	.775	.722	.262
	.509	.482	.448	.155
	.333	.315	.291	.099
	.242	.228	.211	.070
	.179	.169	.155	.051
	.141	.132	.122	.039
	.112	.105	.096	.031
	.092	.071	.079	.025

TABLE 7.9

First two cumulants calculated from the eigenvalues.

p	κ_1	$\kappa_2 \times 10^3$
.95	.08070	3.978
.90	.07655	3.546
.50	.03079	.643

TABLE 7.10

Percentage points of ${}_p\hat{W}^2$ for the exponential distribution, censored on the right. Found by Imhof's method.

p Percentage Point	.5	.9	.95
1.0	.0058	.0173	.0134
2.5	.0071	.0206	.0174
5.0	.0085	.0243	.0217
10.0	.0105	.0296	.0281
15.0	.0122	.0341	.0335
50.0	.0247	.0652	.0700
85.0	.0531	.01321	.1444
90.0	.0635	.1561	.1701
95.0	.0821	.1986	.2155
97.5	.1015	.2433	.2626
99.0	.1279	.3033	.3251

§8. The Anderson Darling statistic for censored data

8.1 Introduction

Anderson and Darling (1952) introduced the statistic

$$A_n^2 = n\int_{-\infty}^{\infty} \left\{\frac{F_n(x) - F_o(x)}{\sqrt{F_o(x)(1-F_o(x)}}\right\}^2 dF_o(x) \qquad (8.1)$$

This statistic can be modified to test goodness-of-fit with censored data in a similar manner to W_n^2 in §6. For the two types of censoring we have

$${}_pA_n^2 = n\int_{-\infty}^{F_o^{-1}(p)} \left\{\frac{F_n(x) - F_o(x)}{\sqrt{F_o(x)(1-F_o(x))}}\right\}^2 dF_o(x) \qquad (8.2)$$

and

$${}_rA_n^2 = n\int_{-\infty}^{x_{(r)}} \left\{\frac{F_n(x) - F_o(x)}{\sqrt{F_o(x)(1-F_o(x)}}\right\}^2 dF_o(x) \qquad (8.3)$$

The small sample distributions of the statistics ${}_{p.}A_n^2$ and ${}_rA_n^2$ seem intractable as the small-sample distribution of A_n^2 is for $n > 3$. However it is well known that the distribution of A_n^2 converges very quickly to its asymptotic distribution; the asymptotic distribution being reasonable even for $n = 3$, see Lewis (1961). The asymptotic distributions of ${}_pA^2{}_n$ and ${}_rA^2{}_n$, which are denoted by ${}_pA^2$ since the statistics both converge to the same asymptotic distribution with $p = \lim_{n\to\infty} \frac{r}{n}$ for ${}_rA_n^2$, can be approximated to, in a similar manner as ${}_pW^2$.

The statistics can be computed in a similar method as A_n^2. It is found upon integration that

$$ {}_pA_n^2 = \sum_{i=1}^{R} \left(\frac{2i-1}{n}\right)\left[\log(1-t_i) - \log(t_i)\right] $$

$$ - 2 \sum_{i=1}^{R} \log(1-t_i) $$

$$ + n\left[\frac{2R}{n} - \left(\frac{R}{n}\right)^2 - 1\right] \log(1-p) + \frac{R^2}{n} \log p - pn \qquad (8.4) $$

where R observations are less than $F_o^{-1}(p)$ and $t_i = F_o(x_i)$. ${}_rA^2{}_n$ is given by the above formula but substituting $R = r-1$ and $p = F_o(x_{(r)})$.

§ 8.2 <u>Exact moments of</u> ${}_pA^2{}_n$, ${}_rA^2{}_n$ <u>and</u> ${}_pA^2$.

We have following §6.2

$$ E\left[{}_pA^2{}_n\right] = n\int_o^p E\left[\left\{\frac{F_n(t) - t}{\sqrt{t)1-t)}}\right\}^2\right] dt $$

$$ = \int_o^p \frac{t(1-t)}{t(1-t)} dt $$

$$ = p \qquad (8.5) $$

and $$ E\left[{}_pA^2{}_n)^2\right] = n^2 \int_o^p \int_o^p E\left[\left\{\frac{F_n(t) - t}{\sqrt{t(1-t)}}\right\}^2 \left\{\frac{F_n(s) - s}{\sqrt{s(1-s)}}\right\}^2 dtds\right. $$

$$ = \frac{2}{n^2} \int_o^p\int_o^s \frac{1}{t(1-t)s(1-s)} [nt + 2n(n-2)t^2 + n(n-3)st $$

$$ + 5n(2-n)st^2 + n(2-n)s^2t + n(3n-6)s^2t^2] \, ds\, dt \qquad (8.6) $$

which upon itegration will produce analytically unintegrable functions such as log(1-s)/s, which need to be expressed as infinite series in powers of s for an explicit representation. In view of the complexity of $E[({}_pA_n^2)^2]$ and that the small sample distribution converges quickly to ${}_pA^2$, the integral (8.6) has not been calculated.

The mean of ${}_rA_n^2$ can be easily found:

$$E[{}_rA_n^2] = E_{t_{(r)}}[{}_pA_n^2 | p = t_{(r)}]$$

$$= E[t_{(r)}]$$

$$= \frac{r}{n+1}, \qquad \text{from (6.10)}$$

The mean of ${}_pA^2$ is given by (8.5), however the variance can be calculated using (6.31)

$$\kappa_2 = 4\int_0^p\int_0^t \frac{s^2(1-t)^2}{s(1-s)t(1-t)}\, ds\, dt$$

$$= -4\int_0^p \frac{(1-t)}{t}[t + \log(1-t)]\, dt$$

$$= -4\Big[-\frac{p^2}{2} + 2p + (1-p)\log(1-p) + \log(1-p)\log p$$

$$- (1 + \frac{1}{2.2} + \frac{1}{3.3} + \ldots)$$

$$+ \{(1-p) + \frac{(1-p)^2}{2.2} + \frac{(1-p)^3}{3.3} + \ldots\}\Big]$$

$$= -4\{-\frac{p^2}{2} + 2p + (1-p)\log(1-p) + \log(1-p)\ \log(p)$$

$$-\frac{\pi^2}{6} + [(1-p) + \frac{(1-p)^2}{2.2} + \frac{(1-p)^3}{3.3} + \ldots]\} \qquad (8.7)$$

Putting p = 1, this becomes

$$\kappa_2\,({}_{p=1}A^2) = \tfrac{2}{3}\pi^2 - 6$$

in agreement with Anderson and Darling (1955). The variance (κ_2) of ${}_pA^2$ is given in Table 8.1 calculated from formula (8.7) truncating the series at the r^{th} term

$$(1-p) + \frac{(1-p)^2}{2.2} + \frac{(1-p)^3}{3.3} + \ldots$$

where the term $\frac{(1-p)^r}{r^2}$ was less than 10^{-8}. When the series was truncated at the r^{th} term so that this term was less than 10^{-6} then the variance so calculated was in agreement up to 5 significant figures, there being an error of no more than 5 digits in the 6^{th} significant figure.

The integral $\int_0^p \frac{\log(1-t)}{t}\,dt$ which equals the series

$$(1-p) + \frac{(1-p)^2}{2.2} + \frac{(1-p)^3}{3.3} + \ldots$$

found by integrating the series expansion of $\frac{\log t}{1-t}$, is known in the form $\int_0^y \frac{e^x}{1-x}\,dx$ as the Debye function, and is described in Abrawowitz and S egun (1965)(§27. 1.1.) It is tabulated, but the tables would need interpolation for the values of y = log(1-p), so it is easier to compute the function of truncating the series directly on the computer.

§8.3 Asymptotic distribution of ${}_rA^2_n, {}_pA^2_n$.

8.3.1. Previous results

We now consider the new stochastic process

$$y_n(t) = \sqrt{n}\{F_n(t) - t\} . \{t(1-t)\}^{-\frac{1}{2}} \tag{8.8}$$

This is easily shown to converge to a Gaussian process $y(t)$ with mean zero and covariance

$$\rho(s,t) = \frac{\min(s,t) - st}{\{s(1-s)t(1-t)\}^{\frac{1}{2}}} \qquad 0 \leq s,t \leq 1. \tag{8.9}$$

and the statistic A^2_n can be shown to converge in distribution to

$$A^2 = \int_0^1 y^2(t)\ dt.$$

Similarly we can show ${}_rA^2_n$, ${}_pA^2_n$ converge to

$${}_pA^2 = \int_0^p y^2(t)\ dt.$$

A well known representation A^2, see Anderson and Darling (1952), is

$$A^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

where the Z_j are i.i.d. standard normal random variables. This follows from the representation of $y(t)$:

$$y(t) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} Z_j f_j(t)$$

where now λ_j and $f_j(t)$ are the eigenvalues and corresponding eigenfunctions of the integral equation

$$\lambda\int_0^1 \rho(s,t)f(s)\,ds = f(t) \quad 0 \le t \le 1 \tag{8.10}$$

where $\rho(s,t)$ is given by (8.9).

§8.3.2. The distribution of ${}_pA^2$

The theory of §6.4 follows to give a similar representation of ${}_pA^2$:

$${}_pA^2 = \sum_{j=1}^{\infty} \frac{1}{\mu_j} Z_j^2$$

where the μ_j are eigenvalues of

$$\mu\int_0^p \rho(s,t)g(s)\,ds = g(t) \qquad 0 \le t \le p \tag{8.11}$$

The matrix approach could be used to find the μ's. It is well known that the λ_j, $f_j(t)$ of (8.10) are given by

$$\lambda_j = j(j+1) \quad j = 1,2,\ldots$$

and

$$f_j(t) = \sqrt{\frac{2j+1}{j(j+1)}}\, P_j^1(2t-1). \tag{8.12}$$

where $P_j^1(x)$ is the associated Legendre polynomial of the first order:

$$P_j^1(x) = (1-x^2)^{\frac{1}{2}} \frac{d}{dx} P_j(x)$$

and $P_j(x)$ is a series solution of order j in x of the differential equation

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + j(j+1)y = 0$$

(see Whittaker and Watson (1962)).

§8.3.3. A series solution to the integral equation.

We can try to solve the integral equation (8.11). Remembering that with p = 1, the eigenfunctions of (8.11) are $P^1_j(2t-1)$, we try a series solution

$$g(t) = \sqrt{t(1-t)} \sum_{r=0}^{\infty} a_r t^r$$

On substitution into (8.11) this gives

$$a_r = (\mu)^r a_o\left(\frac{1}{\mu} - \frac{1}{(r+1)r}\right) x \ \dots \ \left(\frac{1}{\mu} - \frac{1}{2.1}\right) \tag{8.13}$$

and

$$a_o\left(\frac{1}{\mu} - p + \frac{p^2}{2}\right) = \sum_{r=1}^{\infty} a_r p^{r+1}\left(\frac{1}{r+1} - \frac{p}{r+2}\right)$$

Or substituting for a_r, μ must satisfy

$$\left(\frac{1}{\mu} - p + \frac{p^2}{2}\right) = \sum_{r=1}^{\infty} p^{r+1}\left(\frac{1}{r+1} - \frac{p}{r+2}\right)\left(1 - \frac{\mu}{(r+1)(r)}\right)\dots\left(1 - \frac{\mu}{2.1}\right) \tag{8.14}$$

Consequently if the series $\sum_{r=0}^{\infty} a_r t^r$ is to truncate to a finite series then for some r_o, $\mu = r_o(r_o + 1)$. Consequently p must satisfy the polynomial of degree r_o.

$$\frac{1}{r_o(r_o+1)} - p + \frac{p^2}{2} = \sum_{r=1}^{r_o-1} p^{r+1}\left(\frac{1}{r+1} - \frac{p}{r+2}\right)\left(1 - \frac{r_o(r_o+1)}{(r(r+1))}\right)$$

$$\dots \left(1 - \frac{r_o(r_o+1)}{2.1}\right). \tag{8.15}$$

for $r_o \geq 2$ and $r_o = 1$

$r_o = 1$: $\quad \frac{1}{2} - p + \frac{p^2}{2} = 0$ ie $(p-1)^2 = 0$

ie $p = 1$

For $r_o = 2$ $\quad \frac{1}{6} - p + \frac{p^2}{2} = -2p^2\left(\frac{1}{2} - \frac{p}{3}\right)$

$$\text{ie} \quad 4p^3 - 9p^2 + 6p - 1 = 0$$

$$\text{ie} \quad (p-1)^2(4p-1) = 0 \qquad \text{ie } p = 1 \text{ or } \tfrac{1}{4}.$$

We note $p = 1$ is, of course always, a root of (8.15).

Thus for arbitrary p, $0 < p \leq 1$ there will not be a finite series solution $\sqrt{t(1-t)} \sum_{r=0}^{r_o} t^r$ since p will not in general be a solution of (8.15) for finite r_o.

§8.3.4. <u>The normal quadratic form approach to ${}_pA^2$.</u>

Following §6.4 we have

$$ {}_pA^2 = \int_0^p \sum_{j=1}^{\infty} \left\{ \frac{1}{\sqrt{\lambda_j}} \sqrt{\frac{2j+1}{j(j+1)}}\, P_j^1(2t-1) \right\}^2 dt$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\sqrt{(2j+1)(2k+1)}}{j(j+1)k(k+1)} A_{jk}$$

where for $j \neq k$

$$A_{jk} = \frac{(1-q^2)^2}{k(k+1) - j(j+1)} \left[P_k' P_j'' - P_j' P_k''\right]_{x=q}$$

$$P_k' = \frac{d}{dx} P_k(x), \quad P_k'' = \frac{d^2}{dx^2} P_k(x)$$

and

$$q = 2p - 1.$$

For the elements A_{jj} we have to use the explicit series representation of $P_j(x)$:

$$P_j(x) = \frac{1}{2^n} \sum_{m=0}^{J} (-1)^m \binom{j}{m} \binom{2j-2m}{j} x^{j-2m}$$

where $J = \left[\frac{j}{2}\right]$.

After integration

$$A_{jj} = \Bigg[\sum_{n=1}^{J} \sum_{r=1}^{J} G_{n+1} G_{r+1} (j-2n)(j-2r)$$

$$\times\ x^{2j-2n-2r-1} \left(\frac{1}{2j-2n-2r-1} - \frac{x^2}{2j-2n-2r-1} \right) \Bigg]_{x=-1}^{x=q}$$

and $G_1 = \dfrac{1.3.5 \ldots (2j-1)}{j!}$

$$G_{r+1} = -G_r \times \frac{(j-2r+1)(j-2r+2)}{2r(2j-2r+1)}$$

The eigenvalues of thematrix

$$\underset{\sim}{\ell}^T A \underset{\sim}{\ell} \tag{8.16}$$

where $(\ell)_j = \dfrac{\sqrt{2j+1}}{j(j+1)}$, are then identically equal to μ_j^{-1}, $j = 1, 2, \ldots$.

8.4 Significance points of $_pA^2$

8.4.1. Eigenvalues of the integral equation

The eigenvalues of the integral equation (8.11) were obtained by using the condition (8.14) on μ. This condition involves an infinite sum, and for solution, the sum is truncated at K, say, and the following polynomial in μ solved

$$\left(\frac{1}{\mu} - p + p\frac{2}{2}\right) = \sum_{r=1}^{K} p^{r+1}\left(\frac{1}{r+1} - \frac{p}{r+2}\right)\left(1 - \frac{\mu}{(r+1)r}\right)\cdots\left(1 - \frac{\mu}{2.1}\right) \qquad (8.17)$$

It was found that the sum on the right of (8.17) with $K = 50$ differed from the sum with $K = 80$ only in the seventh significant figure for $\mu < 1000$ and $p \geq 5$, and so K was taken equal to 50 and (8.17) solved for μ. The eigenvalues obtained by solving (8.17) are given in Table 8.2., and the first two cumulants calculated using them. It is seen that the second cumulant calculated from the eigenvalues is very close to the accurate value given in Table 8.1.

The matrix approach to the problem was tried by finding the eigenvalues of the matrix (8.16). However this method did not lead to satisfactory results giving eigenvalues μ^{-1} which were negative or greater than 1. However, the values of $\mu^{-1} \leq .5$ were close to the values given by solving (8.17).

Percentage points of ${}_pA^2$ were then found by Imhof's method as in §6, using the eigenvalues found by solving (8.17), which numbered about 15. The percentage points in Table 8 are believed to be accurate to the number of places given.

8.5 Asymptotic Power of ${}_pA^2$ against shifts in location and scale.

8.5.1. The power of ${}_pA^2$.

We can follow section §6.6 to obtain the asymptotic powers of ${}_pA^2$ against shifts of scale and location.

$y_n(t)$ is taken as the process defined by (8.8) and $g_j(t)$ is given by:

$$g_j(t) = n_j\{t(1-t)\}^{\frac{1}{2}} \sum_{s=0}^{K} a_s t^s \tag{8.18}$$

where $a_s(\mu_j)$ is given by (8.15) n_j the normalizing constant and μ_j the corresponding eigenvalue and K = 50 as in (8.17). Then the components ${}_pZ_{nj}$ are defined:

$${}_pZ_{nj} = \sqrt{\mu_j} \int_0^p g_j(t)y_n(t)\,dt.$$

Once again the ${}_pZ_{nj}, {}_pZ_{nk}$ are uncorrelated and with variance μ_j^1 since the $g_j(t)$ is an eigenfunction of the integral equation (8.11). However the expectation of $y_n(t)$ is changed on the alternative:

$$E[y_n(t)] = \gamma^T h(t)\{t(1-t)\}^{-\frac{1}{2}}.$$

Consequently the limiting distribution of ${}_pA^2$ on H_1 is that of

$${}_pA^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{\mu_j}$$

where the Z_j are independent $N(\gamma^T\beta_j, 1)$ and

$$\beta_j = \sqrt{\mu_j}\int_0^p g_j(t)h(t)\{t(1-t)\}^{-\frac{1}{2}}\,dt. \tag{8.19}$$

Similarly we approximate to the exact distribution of $({}_pA^2/H_1)$ by

$$A_A^2 = \sum_{j=1}^{K} \frac{Z_j^2}{\mu_j} + aC_{v,d^2}$$

where the Z_j are independent $N(\gamma^T\beta_j, 1)$ and C is a non-central chi-squared random variable with v degrees of freedom, and non-centrality parameter d^2, a and v are chosen as in §8.4 and d^2 is chosen so that A_A^2 has mean as ${}_pA^2$ on H_1.

$$E[{}_pA^2/H_1] = \frac{\sum_{j=1}^{\infty} 1+\gamma^T\beta_j\beta_j^T\gamma}{\mu_j}$$

$$= E[{}_pA^2/H_o] + \sum_{j=1}^{\infty} \frac{\gamma^T\beta_j\beta_j^T\gamma}{\mu_j}$$

So

$$\sum_{j=1}^{K} \frac{\gamma^T\beta_j\beta_j^T}{\mu_j}\gamma + ad^2 = \gamma^T\left\{\int_o^P \frac{h(t)h^T(t)}{t(1-t)}\,dt\right\}$$

We choose γ, as before, in §6.6, so that the powers of the best tests are .25, .50, .75 and .90, we have the same alternatives also. The probabilities were calculated using the same approximation as in §6.6, and given in Table 8.3

For alternative (i), ${}_pA^2$ has the same power characteristics as ${}_pW^2$, except that ${}_pA^2$ is more powerful for p = 1.0 and 0.9. For p = .5, ${}_pA^2$ and ${}_pW^2$ have approximately equal power over all significant levels and values of γ.

For alternative (ii), ${}_pA^2$ is not very powerful. However for p = 1.0 the power of ${}_pA^2$ is much great than that of ${}_pW^2$. At the 5% significant level, ${}_pA^2$ does have power .218 and .421 against 'best' test powers of .75 and .90 which is an improvement on the power of ${}_pW^2$. However the power drops for the smaller significant level, 1%, to less than half of the power at the 5% significant level. For p = .9 the power is reduced, but for p = .5 the power is not less than for p = .9 in some cases is more. However for all situations there is little power for ${}_pA^2$ against this alternative, but the power of ${}_pA^2$ is considerably better than that of ${}_pW^2$.

For alternative iii and p = 1.0, ${}_pA^2$ achieves good powers with an improvement over ${}_pW^2$. However as p becomes smaller the difference between the powers of ${}_pW^2$ and ${}_pA^2$ becomes less, for p = .5, ${}_pW^2$ is more powerful than ${}_pA^2$.

Is we base the alternatives on $\gamma p^{-\frac{1}{2}}$ then the power of ${}_pA^2$ increases as before with ${}_pW^2$. The power of ${}_pA^2$ with p = .9 and .5 exceeds the power of the best test for alternative (i).

§8.5.2 <u>The power of the components of ${}_pA^2$.</u>

We can follow §6.6.5 and find the power of ${}_pZ_1$ and ${}_pZ_2$ of ${}_pA^2$ for shifts of location and variance of the normal distribution. The coefficients β_j(8.19) were determined by quadrature using the function $g_j(t)$ defined in (8.18) and h(t) given in (6.49) and defined as

$$h_1(t) = -\phi(\Phi^{-1}(t))$$

$$h_2(t) = \tfrac{1}{2}\Phi^{-1}(t)h_1(t)$$

for normal location and variance shift.

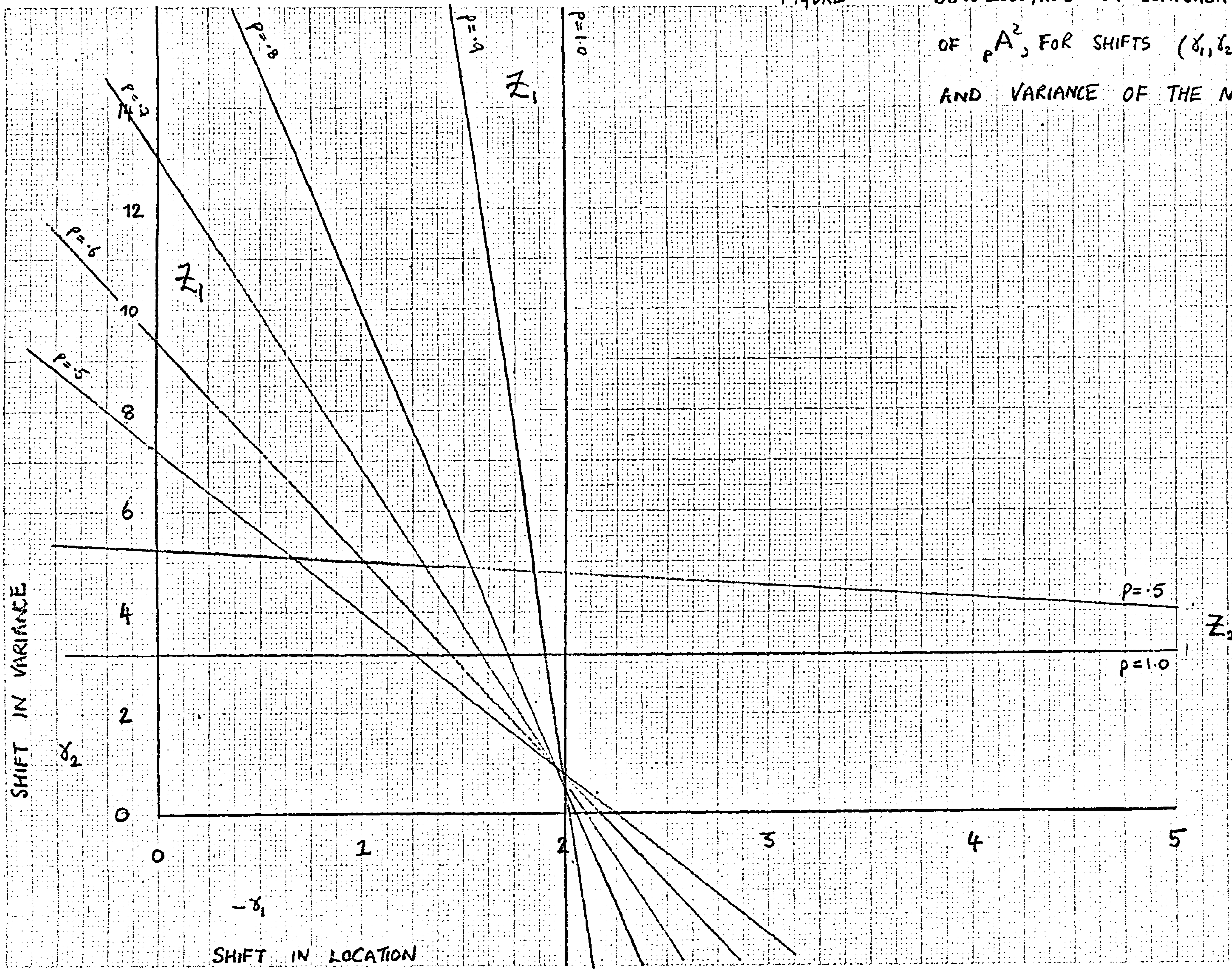
The values of β_j are given in Figure 8.1. Isodynes for ${}_pZ_1$ and ${}_pZ_2$ against location and variance shift are drawn in Figure 8.2, corresponding to Figure 6.3 for ${}_pW^2$. The 50% isodynes are drawn. A comparison of Figure 6.3 with Figure 8.2 reveals that components based on ${}_pA^2$ have better power than those based on ${}_pW^2$. For p = .5. However there isnegligible difference between the powers of ${}_pZ_1$ based on ${}_pA^2$ and ${}_pW^2$.

Also considered, is the power of the statistic ${}_pZ_1^2 + {}_pZ_2^2$ for ${}_pA^2$. The ellipses which are the isodynes for ${}_pZ_1^2 + {}_pZ_2^2$ are drawn in Figure 8.3, and areas enclosed by them given in Figure 8.4. Comparing the isodynes, and areas enclosed by them, of ${}_pW^2$, with those of ${}_pA^2$, we see from Figure 6.6 and Figure 8.4 that ${}_pA^2$ has better power (ie the area enclosed by the isodyne is smaller) than ${}_pW^2$ for values of p = .6(.1)1.0. However for p = .5, ${}_pW^2$ is marginally better than ${}_pA^2$.

p	j = 1, Z_1 Location β_{11}	j = 1, Z_1 Variance β_{21}	j = 2, Z_2 Location β_{12}	j = 2, Z_2 Variance β_{22}
1.0	-.977	0	0	.616
.9	-.968	.034	-.004	.575
.8	-.961	.099	-.009	.528
.7	-.931	.150	-.036	.485
.6	-.903	.209	-.064	.427
.5	-.868	.268	-.072	.372

Figure 8.1 Values of β_{ij} for the first two components ${}_pZ_1$, ${}_pZ_2$ for location and variance shift for the normal population.

FIGURE 8.2 50% ISODYNES FOR COMPONENTS ${}_{\rho}Z_1$ & ${}_{\rho}Z_2$ OF ${}_{\rho}A^2$, FOR SHIFTS (γ_1, γ_2) IN LOCATION AND VARIANCE OF THE NORMAL DISTRIBUTION



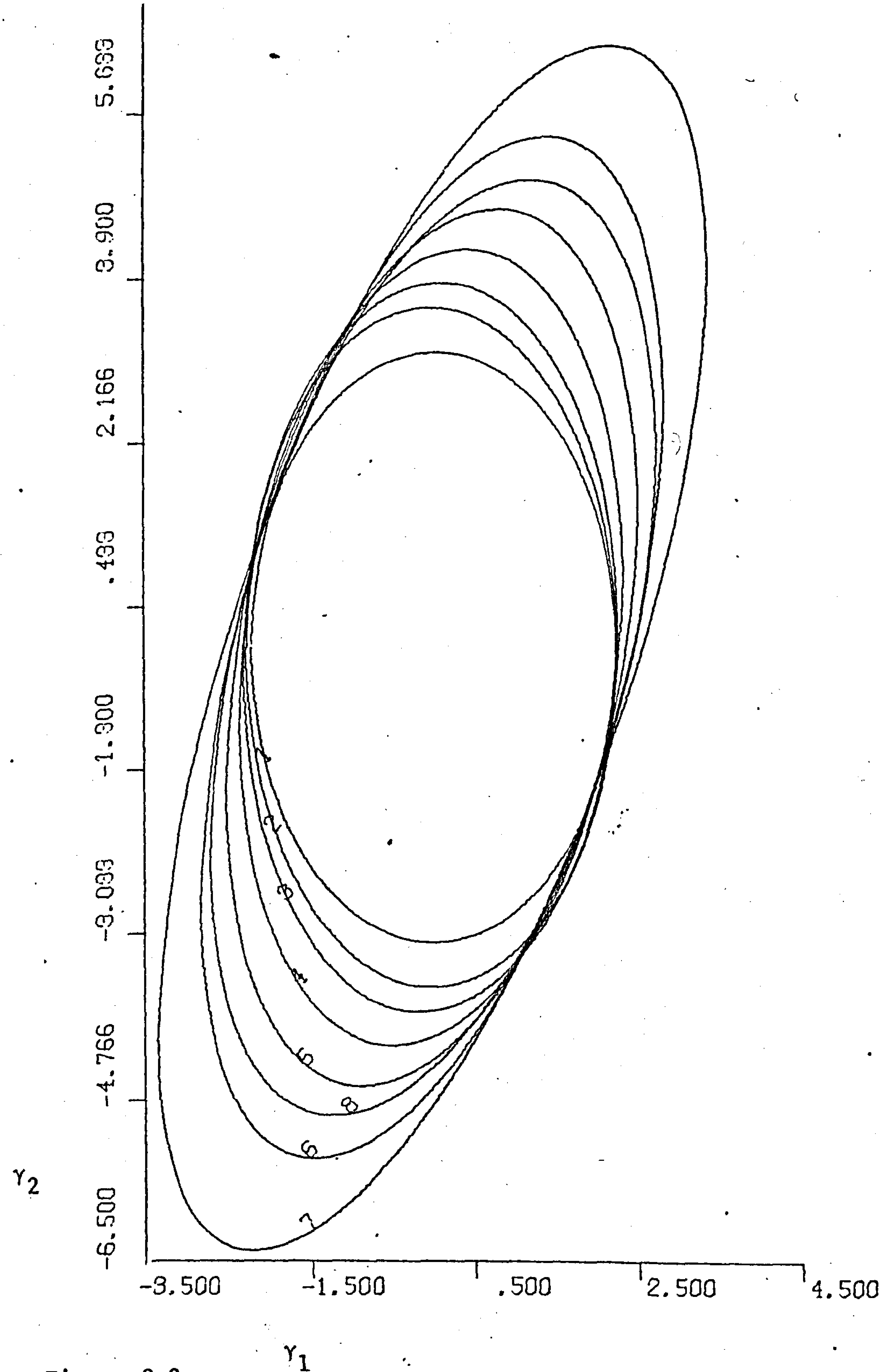


Figure 8.3

50% Isodynes at the 5% level for components of ${}_pA^2$.

Curve 1 Likelihood ratio test ${}_pL^2$, p = 1.0

Curve 8 Likelihood ratio test ${}_pL^2$, p = .5

Curves 2-7 Test based on ${}_pZ_1^2 + {}_pZ_2^2$ p = .9(.1).5 respectively.

Testing for the normal distribution with $\mu = 0$ and $\sigma^2 = 1$, with the alternative $\mu = \gamma_1 n^{-\frac{1}{2}}$, $\sigma^2 = 1 + \gamma_2 n^{-\frac{1}{2}}$.

Statistic	Area
${}_pL^2$ (p=1.0)	21.95
${}_pZ_1^2 + {}_pZ_2^2$ p = 1.0	25.79
p = .9	27.89
p = .8	30.64
p = .7	34.79
p = .6	41.70
p = .5	51.12
${}_pL^2$ (p = .5)	38.23

Figure 8.4 Areas enclosed by isodynes based on ${}_pZ_1^2 + {}_pZ_2^2$ and ${}_pL^2$

8.6 Testing for normality using ${}_pA^2$

The theory follows closely that of §7.4 but with the covariance kernel given by

$$\hat{\rho}(s,t) = \{\min(s,t) - st - \psi_1(s)\psi_1(t) - \psi_2(s)\psi_2(t)\} \times \{s(1-s)t(1-t)\}^{\frac{1}{2}} \qquad 0 \leq s,t \leq p \tag{8.20}$$

The first two cumulants can be calculated by quadrature and are given in Table 8.5. We can use approximate eigenfunctions of the form

$$g_j(t) = \{t(1-t)\}^{\frac{1}{2}} \sum_{S=0}^{K} a_s t^s \tag{8.21}$$

where $a_s(\mu_j)$ is given by (8.17). K is found by trial and error so that an increase in its value makes no appreciable difference to $g_j(t)$. K was taken to be 50, following the calculation pf μ_j from (8.17). The normalizing constant is found by integrating $g_j(t)$ over (o,p). We can find the approximate Fredholm determinant $\Delta(\mu)$ following (7.14), but with the eigenfunctions given by (8.21) and $\hat{\rho}(s,t)$ by (8.20). We can then solve the approximate Fredholm determinant $\Delta(\mu)$ and obtain the eigenvalues μ. These are given in Table 8.6 for p = 1.0, .95, .9 (.1) .5. As p becomes smaller then there is a problem to obtain the eigenvalues due to a lack of eigenvalues from the standard case and an apparent lack of convergence of the approximate Fredholm determinant to the limiting value having infinite terms. The eigenvalues obtained, appear to give percentage points displayed in Table 8.7 in good agreement with the Monte Carlo results given in Table 8.4.

A Monte Carlo power study on ${}_rA^2{}_n$ was carried out and the results have been given in Table 7.6 and described in §7.4. The overall impression is that A^2 seems to be more powerful than W^2 for the various alternatives considered.

Table 8.1

Exact mean and variance of ${}_pA^2$: σ^2 calculated using formula (8.7)

$p = E\left[{}_pA^2\right]$	σ^2
1.00	.57974
.99	.57891
.95	.56668
.90	.53989
.85	.50560
.80	.46673
.75	.42517
.70	.38228
.65	.33908
.60	.29641
.55	.25494
.50	.21526

Table 8.2

Eigenvalues for ${}_pA^2$, found by solving (8.18) and cumulants calculated using them.

p 1.0	.95	.90	.80	.70	.60	.50
.50000	.49680	.48766	.45810	.41778	.37003	.31676
.16667	.16238	.15365	.13329	.11290	.09358	.07549
.08889	.07902	.07242	.06011	.04950	.04023	.03199
.05000	.04614	.04139	.03362	.02737	.02208	.01746
.03333	.03002	.02657	.02134	.01727	.01389	.01093
.02381	.02101	.01843	.01470	.01187	.00952	.00750
.01786	.01550	.01350	.01073	.00865	.00693	.00546
.01389	.01188	.01031	.00817	.00658	.00527	.00414
.01111	.00937	.00812	.00643	.00517	.00414	.00326
.00909	.00755	.00656	.00518	.00416	.00325	.00262

p	.95	.90	.80	.70	.60	.50
κ_1	.88408	.83861	.75167	.66125	.56892	.47561
κ_2	.56572	.53964	.46657	.38207	.29624	.21520

Table 8.3 Percentage points of ${}_pA^2$, found by Imhof's method.

	p						
Percentage point	1.0*	.95	.90	.80	.70	.60	.50
1.0	.201	.169	.152	.125	.103	.084	.067
2.5	.241	.208	.186	.154	.127	.104	.083
5.0	.284	.249	.223	.186	.154	.126	.101
10.0	.346	.308	.280	.234	.194	.160	.128
15.0	.399	.359	.327	.275	.230	.190	.153
50.0	.774	.724	.676	.587	.504	.425	.349
85.0	1.621	1.562	1.494	1.346	1.190	1.029	.864
90.0	1.934	1.872	1.798	1.632	1.451	1.260	1.062
95.0	2.492	2.428	2.344	2.146	1.920	1.676	1.419
97.5	3.077	3.009	2.915	2.684	2.412	2.112	1.792
99.0	3.878	3.804	3.698	3.419	3.083	2.707	2.301

* Calculated using 20 eigenvalues, $\lambda_j = j(j + 1)$.

Table 8.3A Asymptotic powers of ${}_pA^2$ for scale and location shifts for a given population. In parentheses, power based on alternative $\gamma p^{-\frac{1}{2}}$.

5% significance level.

p		Alternative	.25	.50	.75	.90
		Power of 'best' test.	.25	.50	.75	.90
	i	Normal, location shift	.236	.475	.722	.889
1.0	ii	Normal, scale shift	.079	.137	.251	.439
	iii	Exponential	.203	.410	.652	.852
	i		.233(.254)	.467(.508)	.715(.760)	.895(.923)
.9	ii		.073(.076)	.116(.126)	.200(.225)	.349(.396)
	iii		.191(.207)	.383(.418)	.614(.660)	.819(.858)
.5	i		.198(.349)	.395(.670)	.626(.899)	.826(.984)
	ii		.084(.121)	.134(.232)	.212(.401)	.326(.612)
	iii		.123(.200)	.225(.402)	.369(.639)	.545(.840)

1% significance level

p			.25	.50	.75	.90
	i		.235	.475	.724	.883
1.0	ii		.026	.054	.121	.235
	iii		.187	.392	.635	.818
	i		.228(.259)	.463(.517)	.712(.768)	.875(.912)
.9	ii		.021(.023)	.039(.045)	.081(.098)	.155(.191)
	iii		.167(.190)	.352(.398)	.583(.642)	.772(.823)
	i		.177(.404)	.367(.722)	.598(.923)	.782(.985)
.5	ii		.037(.076)	.069(.164)	.121(.309)	.191(.481)
	iii		.080(.179)	.161(.375)	.286(.613)	.427(.798)

Table 8.4 Empirical distribution of ${}_r\hat{A}^2_n$ when testing for normality, using Cohen's estimates of μ and σ for Type II single censoring, with 5000 samples.

Percentage points of ${}_r\hat{A}^2_n$, where r = pn

p = .5

n	50%	85%	90%	95%	99%
20	.1220	.1954	.2174	.2552	.3486
40	.1221	.2052	.2316	.2753	.3750
60	.1212	.2132	.2403	.2852	.3938
80	.1236	.2132	.2432	.2920	.4092
Estimated Standard error	.0011	.0020	.0026	.0036	

p = .9

n	50%	85%	90%	95%	99%
20	.2698	.4517	.5055	.6100	.8365
40	.2678	.4489	.5068	.6050	.8400
60	.2661	.4519	.5179	.6236	.8880
80	.2690	.4518	.5106	.6210	.8725
Estimated Standard error	.0025	.0047	.0057	.0084	

Table 8.5 Exact mean and variance of ${}_p\hat{A}^2$ - testing for normality, found by quadrature.

P	MEAN	VARIANCE σ^2 x 10
1.00	.38445	.36208
.95	.33980	.31272
.90	.30728	.26593
.85	.27998	.22675
.80	.25580	.19348
.75	.23379	.16480
.70	.21341	.13981
.65	.19427	.11784
.60	.17613	.09846
.55	.15878	.08131
.50	.14216	.06619

Table 8.6 Eigenvalues for ${}_p\hat{A}^2$, testing for normality with mean and variance unknown. Values x 10.

p	1.0	.95	.90	.80	.70	.60	.50
	.9836	.9358	.8703	.7448	.6264	.5084	.3927
	.7206	.6621	.6083	.5111	.4325	.3557	.2845
	.3593	.3257	.2938	.2406	.1953	.1542	.1182
	.2897	.2574	.2310	.1423	.1553	.1255	.0990
	.1810	.1640	.1475	.1135	.0934	.0736	.0567
	.1584	.1381	.1106	.0720	.0791	.0635	.0498
	.1148	.0990	.0835	.0657	.0545	.0430	
	.1002	.0884	.0667	.0487	.0476	.0377	
	.0777	.0700	.0548	.0434	.0381	.0319	
	.0692	.0621	.0431	.0342			

Table 8.7 Asymptotic percentage points of ${}_p\hat{A}^2$, testing for normality with unknown mean and variance, found using Imhof's method.

Percentage Point	p = 1.0	.95	.90	.80	.70	.60	.50
1.0	.129	.109	.094	.069	.049	.035	.027
2.5	.148	.125	.110	.084	.061	.046	.036
5.0	.167	.141	.125	.098	.074	.056	.045
10.0	.194	.165	.146	.118	.091	.071	.057
15.0	.215	.183	.163	.133	.104	.083	.067
50.0	.341	.298	.268	.223	.182	.151	.125
85.0	.560	.502	.456	.382	.316	.264	.219
90.0	.631	.568	.517	.434	.360	.300	.248
95.0	.752	.682	.623	.524	.436	.362	.297
97.5	.873	.796	.729	.615	.512	.424	.346
99.0	1.035	.949	.871	.736	.614	.507	.411

PART III

§ 9 Practical Application of the tests given in Parts I and II

§ 9.1 Introduction

In this section we show how to use the tests presented in this thesis. There are two sections, the first considering the statistics introduced in Part I, and the second those statistics introduced in Part II.

9.2 The statistics $D_{n,q}$, $D^+_{n,q}$, $V_{n,q}$. Three hypothetical examples, two small samples and one large sample.

9.3 The statistics ${}_pW^2$ and ${}_pA^2$.

9.2 The statistics $D_{n,q}$, $D^+_{n,q}$, and $V_{n,q}$ introduced in Part I.

These statistics were introduced to test the goodness-of-fit of grouped or discrete data. Percentage points of the statistics are tabulated so that tests of the null hypothesis

$$H_o : p_i = \frac{1}{k} \qquad i = 1, \ldots, k$$

can be made, where p_i is the probability that an observation comes from the i^{th} category. The null hypothesis is that all categories are equally likely. The one-sided statistic $D^+_{n,q}$ and the two-sided statistic $D_{n,q}$ will be powerful against alternatives of the type

$$H_1 : p_i = \frac{1}{k} + \delta \qquad i = 1, \ldots, j \qquad (9.1)$$

$$p_i = \frac{1}{k} - \gamma \qquad i = j+1, \ldots, k,$$

where $j\delta - (k-j)\gamma = 0$ $V_{n,q}$ is designed to test grouped or discrete data which arises from observations on a circle. V_n is powerful against alternatives similar to (9.1), but where p_i $i = 1, \ldots, j$ refer to j adjacent categories or mass points on the circle.

To illustrate the use of $D_{n,q}$ $D^+_{n,q}$ and $V_{n,q}$

Example 1.

We use some hypothetical data in Siegel (1956 p49) Each subject is presented with five photos of himself varying in tone (grades 1-5) and asked to choose the photo he likes best. Ten subjects are chosen and the hypothesis tested is that there is no overall preference for any tone, thus if the hypothesis is true each grade of tone is equally likely to be chosen.

Grade of tone of photo chosen	1	2	3	4	5
Number choosing that grade, x	0	1	0	5	4
$F_o(x)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1
$F_n(x)$	0	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	1
$F_n(x) - F_o(x)$	$-\frac{1}{5}$	$-\frac{3}{10}$	$-\frac{5}{10}$	$-\frac{2}{10}$	0

$$D^+ = 0, \quad D^- = \frac{5}{10}, \quad D = \frac{5}{10}, \quad V = \frac{5}{10}$$

$pr(10\ D^-_{10,5} \geq 5) = 0.0000$

$pr(10\ D_{10,5} \geq 5) = 0.0001$

$pr(10\ V_{10,5} \geq 5) = 0.0002$

Thus there is very strong evidence to suggest that the hypothesis is not true.

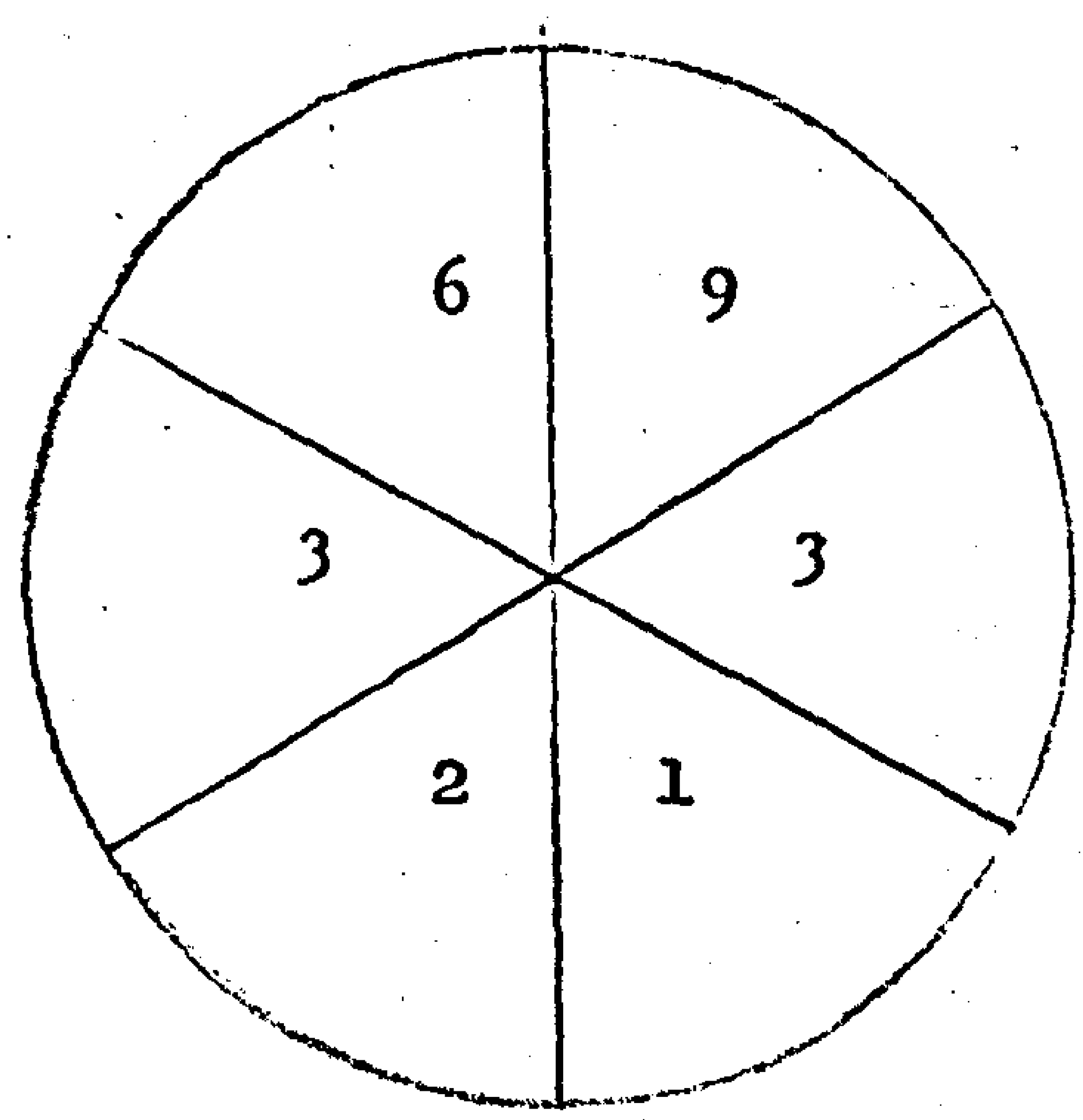
It would not be practical in this situation to use the X^2 statistic, since the small sample distribution is not known exactly, without using Kempthrone's (1967) first four moments of X^2 and fitting, for example, a Pearson curve.

Naively if we use X^2, then we have a value of $X^2 = 11$ for the data, which has significance of about 2.5. using χ^2_4. Using Kempthorne's moments and a Pearson curve fit this probability is conservative, i.e. the significance is > 2.5%.

Thus the D statistics and V show the significance of the data much more plainly than does X^2 : D giving a significance of .001% and V .002% against X^2, which is > 2.5% .

<u>EXAMPLE 2</u>

Here we use $V_{n,q}$ only, in its role to test goodness-of-fit on the circle. Twenty four pigeons were released from a cage, six watchers were employed and the circle divided up into six equal sectors of 60° each. A watcher was assigned to each sector and told to count the number of pigeons flying off in his sector. The data is given below:



It is wished to test the null hypothesis that the pigeons have no preferential direction of flying off. Under this hypothesis each sector is equally likely.

Sector	1	2	3	4	5	6
Number of pigeons	9	3	1	2	3	5
$F_o(x)$	$\frac{4}{24}$	$\frac{8}{24}$	$\frac{12}{24}$	$\frac{16}{24}$	$\frac{20}{24}$	1
$F_n(x)$	$\frac{9}{24}$	$\frac{12}{24}$	$\frac{13}{24}$	$\frac{15}{24}$	$\frac{18}{24}$	1
$F_o(x) - F_n(x)$	$\frac{-5}{24}$	$\frac{-4}{24}$	$\frac{-1}{24}$	$\frac{1}{24}$	$\frac{2}{24}$	0

$$D^+ = \frac{2}{24} \qquad D^- = \frac{5}{24} \quad \text{so } V = \frac{7}{24}$$

From the tables we have $\Pr(24\,V \geq 7) = 1 - .96574$
$= .03426$

Using the X^2 statistic, we have $X^2 = 10$, and from the X^2_5 tables, we have $\Pr(X^2 \geq 10) = .07524$.

EXAMPLE 3

Consider an example to use D_n, D_n^+ or V_n for large samples. A computer program is supposed to generate random digits 0,1,...,9. 1000 digits were generated and one test for randomness was to compare the number of times each digit occured with its expected frequency.

Digit	0	1	2	3	4	5	6	7	8	9
Number of occurrences	91	90	94	105	105	110	107	104	101	93
$F_o(x)$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$F_x(x)$	.091	.181	.275	.380	.485	.595	.702	.806	.907	1.0

$$D^+ = .025, \quad D^- = .007$$

So $\qquad D = .025$ and $V = .032$

Here we use the tables of the asymptotic distributions of the statistics, which are in terms of $\sqrt{n}\ D$ etc. $\sqrt{n} = 31.62$, so $\sqrt{n}\ D = .791$ and $\sqrt{n}\ V = 1.012$

From the Table 3.1, these tables are not significant at the 15% level. However from more extensive tables we have $Pr(\sqrt{n}\ D > .791) \simeq .30$ and $Pr(\sqrt{n}\ V > 1.012) \simeq .29$. The value of the statistic X^2 for this data is 4.82 and from the tables $Pr(\chi^2_9 > 4.82) \simeq .85$. Thus the KS-type statistics give an indication of some non-randomness of the program, whereas the X^2 statistic gives none.

9.3 <u>The statistics ${}_pW^2$ and ${}_pA^2$ introduced in Part II</u>.

In this section we use W^2 and A^2 to test goodness-of-fit with censored data. We give hypothetical examples testing simple and composite hypotheses, and an example from the literature.

<u>EXAMPLE 1</u> <u>Testing a simple hypothesis</u>

It is found that bids made for tender of a contract by firms in a certain magufacturing industry are uniform for part of their range. For a certain contract it is hypothesized that the upper 90% of the distribution is uniform over(1.1, 2.0). Thirty bids were made for the contract; they were:

.68, .88, 1.01, 1.24, 1.27, 1.29, 1.29,
1.32, 1.34, 1.34, 1.40, 1.45, 1.57, 1.67,
1.69, 1.70, 1.74, 1.75, 1.78, 1.85, 1.87,
1.89, 1.90, 1.91, 1.92, 1.92, 1.93, 1.94,
1.95, 1.96.

We test the observations in (1.1, 2.0) for uniformity. We first transform the observations to the interval (0,1). Here

$$F_o(x) = .1 + \frac{x - 1.1}{2.0-1.1} \times (.9) \text{ for } x \in (1.1, 2.0)$$

$$= x - 1.0$$

We transform the 27 observations in (1.1,2.0) to (0,1), by putting $t = F(x)$. We then have
.24, .27, .29, .29, .32, .34, .34, .40, .45, .57
.67, .69, .70, .74, .75, .78, .85, .87, .89, .90, .91,
.92, .92, .93, .94, .95, .96.

In this situation we have censoring of type I on the left. We restrict ourselves to observations greater then $F_o^{-1}(.1)$. To use formulae (6.11) and (8.4) to construct ${}_pW_n^2$ and ${}_pA_n^2$ respectively, we require censoring on the right rather than left. If we make the transformation $t' = 1 - t$ to the observations alone then the t' observations will be censored on the right at $F_o^{-1}(1 - .1)$. We can use formulae (6.11) and (8.4) with the t' values with $R = 27$ and p. .9. We find upon calculation that

$$ {}_pW_n^2 = .489 \text{ and } {}_pA_n^2 = 2.95 $$

From the χ^2 approximation to the small sample distribution of ${}_rW_n^2$ with percentage points given in Table 6.1, we see that the 5% point is .448 and the 2½% is .560, so our value lies between 5 and 2½% significant.

From the asymptotic percentage point of ${}_pA^2$ in Table 8.3 we see that the 2½% point is 2.92 and the 1% point is 3.70, so our value of ${}_rA_n^2$ is about 2½% significant.

Thus the evidence from the statistics ${}_rW_n^2$ and ${}_rA_n^2$ suggests the data has deviated from the hypothesised distribution.

EXAMPLE 2 Testing for normality

An experiment is set up to test the current gains (I_G) of certain transistors. It is known that ln (I_G) has the normal distribution. Forty measurements were to be made, however after taking the twenty smallest values the equipment brokedown and the experiment was unable to continue. It is wished to test the hypothesis that the twenty available observations can be considered to be log-normally distributed. Values of the logarithms of the twenty observations are given below :

- .2671, .0142, .2547, .4237, .4919, .4923, .7065,
1.0047, 1.0317, 1.0598, 1.1647, 1.1759, 1.2130, 1.3446,
1.3448, 1.5620, 1.7712, 1.7756, 1.9014, 1.9500. (9.1)

The mean and standard deviation of the normal population $N(\mu, \sigma^2)$ which these twenty observations were assumed to be drawn from, were estimated using Cohen's (1959) tables for the MLE's of μ and σ.

It was found that

$$\hat{\mu} = 2.0834 \qquad \hat{\sigma} = 1.1022$$

The bias of these estimates was reduced to order n^{-2} using Saw's (1961) corrections

$$\hat{\mu}_c = \hat{\mu} + \frac{\hat{\sigma}}{n+1}\,.96 \qquad \hat{\sigma}_c = \hat{\sigma} + \frac{\hat{\sigma}}{n+1}\,1.762$$

$$= .0576 \qquad\qquad = 1.1496$$

The observations (9.1) are then transformed to (0,1). We define $\hat{t}_i = \Phi((x_i - \hat{\mu}_i)/\hat{\sigma}_c)$, where $\Phi(\cdot)$ is the standard normal c.d.f. These values are given below

.0273, .0467, .0710, .0923, .1035, .1036, .1412
.2073, .2141, .2213, .2494, .2525, .2629, .3015
.3016, .3703, .4408, .4423, .4512, .5027.

We now calculate the statistics ${}_r\hat{W}^2_n$ and ${}_r\hat{A}^2_n$, the statistics for type II censoring, which is the situation here. We use formula (6.12) to calculate ${}_r\hat{W}^2_n$, using the $\hat{t}_i$ and find that ${}_r\hat{W}^2_n = .0141$ with R = 19.

We now calculate ${}_r\hat{A}^2_n$ using formula (8.4) but with R = 19 and $p = \Phi((x_{20)} - \hat{\mu}_c)/\hat{\sigma}_c)$. We find ${}_r\hat{A}^2_n = .0947$.

Comparing these values with asymptotic percentage points we see that the value of ${}_r\hat{W}^2_n$ is about 80% significant from Table 7.5, and ${}_r\hat{A}^2_n$ about 60% significant, so indicating that we can accept the hypothesis of normality.

EXAMPLE 3 Testing for normality

Gupta (1952) presents the following data, showing the number x' of days to death of the first 7 in a sample of 10 mice after inoculation with a uniform culture of human tuberculosis:

x'	41	44	46	54	55	58	60
x = log x'	1.613	1.644	1.663	1.732	1.740	1.763	1.778

Gupta takes log x' to be normally distributed. David (1970, §6.3, Example 6.3.1) finds estimates of μ and σ using Cohen's (1959) MLE and Sarhan and Greenberg's (1962, p222) best linear unbiased estimates. David finds

i) M.L.E. $\hat{\mu} = 1.742$

$\hat{\sigma} = .079$

ii) B.L.U.E. $\mu^* = 1.746$

$\sigma^* = .091$

Using Saw's (1961) corrections to the MLE estimates to reduce the bias to $O\left(\frac{1}{n^2}\right)$:

$$\hat{\mu}_c = \hat{\mu} - \frac{\hat{\sigma}}{n+1} B_p(\mu)$$

$$\hat{\sigma}_c = \hat{\sigma}\left(1 - \frac{1}{n+1} B_p(\sigma)\right)$$

For $p = .7$, $B_p(\mu) = -.342$

$B_p(\sigma) = -1.193$

which gives corrected estimates,

$$\hat{\mu}_c = 1.744$$

$$\hat{\sigma}_c = .0875$$

Using the corrected MLE's it is found that ${}_p\hat{W}^2 = .0230$ with the observations transformed to (0,1) using $t = \Phi((x - \hat{\mu}_c/\hat{\sigma}_c)$ giving

.0672, .1265, .1773, .4455, .4818, .5860, .6512

With the B.L.U.E's the observations transform to :

.1719, .1213, .1809, .4389, .4737, .5741, .6375

giving ${}_p\hat{W}^2 = .0198$. These two values of ${}_p\hat{W}^2$ are quite close, their difference arising from the differing values of $\hat{\sigma}_c$ and σ^*. The values of ${}_p\hat{W}^2$ are about 70% significant, suggesting normality of the $\log x'$ observations and that $(\hat{\mu}_c, \hat{\sigma}_c)$ or (μ^*, σ^*) provide good estimates of (μ, σ).

§10 Related and further work

10.1 Introduction

In this final section we consider a statistic closely related to W_n^2 and A_n^2, Watson's statistic U_n^2, introduced in Watson (1961) to test goodness-of-fit on the circle. The theory of U^2 adapted for censored data is given in §10.2.

In §10.3 and §10.4, we consider further work related to the content of this thesis. In §10.3 the statistic ${}_pW^2$ is extended for the 2-sample problem and the bivariate sample.

In §10.4 an alternative way of testing for normality with censored data is suggested.

10.2 The statistic U^2

10.2.1 Watson's statistic U_n^2

Related to the A^2 and W^2 statistic is the U^2 statistic introduced by Watson (1961) to test goodness-of-fit on the circle. It is a version of W^2 defined by

$$U_n^2 = n\int_{-\infty}^{\infty}\left\{F_n(x)-F_o(x)-\int_{-\infty}^{\infty}\left[F_n(y)-F_o(y)\right]dF_o(y)\right\}^2 dF_o(x)$$

Watson considers the process

$$Z_n(t) = \sqrt{n}\left\{F_n(t)-t-\int_o^1\left[F_n(u)-u\right]du\right\}$$

and shows that

$$E[Z_n(t)] = 0 \text{ and}$$

$$E[Z_n(t)Z_n(s)] = \min(s,t)-\tfrac{1}{2}(s+t)+\tfrac{1}{2}(s-t)^2+\frac{1}{12}, \qquad 0 \leq s,\ t \leq 1$$

He finds the asymptotic distribution of U_n^2 by finding the eigenvalues and eigenfunctions of the integral equation

$$\lambda\int_0^1 \rho(s,t)h(s)ds = h(t), \qquad 0 \le t \le 1$$

where $\rho(s,t) = \lim_{n\to\infty} E[Z_n(t)Z_n(s)]$

§10.2.2 Stephens' modification of U^2 for censored data

Professor M.A. Stephens in a private communication has considered the U_n^2 statistic for censored data. The process

$$Q_n(t) = Y_n(t) - \frac{1}{p}\int_0^p Y_n(u)du$$

where $Y_n(t) = \sqrt{n}\{F_n(t)-t\}$,

is considered. Then ${}_pU_n^2$ is defined as

$${}_pU_n^2 = \int_0^p \{Q_n(t)\}^2\, dt$$

With this development of ${}_pU_n^2$, the covariance function of $Q_n(t)$ can be found.

$$E[Y_n(t)\int_0^p Y_n(u)du] = \int_0^p \{\min(u,t)-ut\}du$$

$$= \int_0^t u(1-t)du + \int_t^p t(1-u)du \ .$$

$$= \frac{t^2}{2}(1-t) + t(p-\frac{p^2}{2}) - t(t-\frac{t^2}{2})$$

$$= tp - \frac{tp^2}{2} - \frac{t^2}{2}$$

Also $E\left[\int_0^p Y_n(u)du \int_0^p Y_n(v)dv\right] = \int_0^p\int_0^p \{\min(u,v)-uv\}du\, dv$

$$= \frac{p^3}{3} - \frac{p^4}{4}$$

Finally

$$E[Q_n(t)Q_n(s)] = E[Y_n(s)Y_n(t)]$$

$$- E\left[Y_n(s)\frac{1}{p}\int_0^p Y_n(u)du\right]$$

$$- E\left[Y_n(t)\frac{1}{p}\int_0^p Y_n(u)du\right]$$

$$+ E\left[\frac{1}{p}\int_0^p Y_n(u)du \, \frac{1}{p}\int_0^p Y_n(v)dv\right]$$

$$= \min(s,t) - st - s + \frac{sp}{2} + \frac{s^2}{2p}$$

$$- t + \frac{tp}{2} + \frac{t^2}{2p} + \frac{p}{3} - \frac{p^2}{4} \qquad (10.1)$$

Denote the covariance of $Q_n(t)$ by $\rho_n(s,t)$.

Following the previous representation of ${}_pW^2$ and ${}_pA^2$, the asymptotic distribution of ${}_pU^2_n$ is given by

$${}_pU^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{\mu_j}$$

where μ_j are eigenvalues of the integral equation

$$\mu\int_0^p \rho(s,t)g(s)ds = g(t), \qquad 0 \le t \le p$$

where $\rho(s,t) = E[Q_n(t)Q_n(s)]$ given by (10.1). As for ${}_pW^2$ we differentiate and find the general solution

$$g(t) = A\sin(mt) + B\cos(mt),$$

where $m^2 = \mu$, to the differential equation. To find the eigenvalues, $g(t)$ is substituted back into the integral equation. It is found upon integration that

$$g(t) \equiv A\sin mt + At(-m\cos mp + mp\cos mp - \sin mp)$$
$$- At(1-p/2)m(1-\cos mp) + A(t^2m/2p)(1-\cos mp)$$
$$+ (A/2p)\{-mp^2\cos mp + 2p\sin mp + 2/m(\cos mp-1)\}$$
$$- A(1-p/2)(-mp\cos mp + \sin mp) + Am(p/3 - p^2/4)(1-\cos mp)$$
$$+ B(\cos mt - 1 + mt\sin mp - mtp\sin mp + t - t\cos mp)$$
$$+ (B/2p)(mp^2\sin mp + 2p\cos mp - 2/m\sin mp)$$
$$- B(1-p/2)(mp\sin mp + \cos mp - 1) + B(t^2/2p)m\sin mp$$

$- B(1-p/2)tm \sin mp + B(p/3-p^2 4)m \sin mp.$

For this to hold the coefficients of the t and t^2 terms must be equal to 0.

For the coefficient of the t^2 term,

$$A(1 - \cos mp) + B \sin mp = 0$$

Then letting $k = m/2$ and provided $\sin kp \neq 0$

$$A \sin kp + B \cos kp = 0 \qquad (10.2)$$

The coefficient of the t term is:

$$A(mp/2 \cos mp - \sin mp - m + mp/2)$$
$$+ B(-mp/2 \sin mp + 1 - \cos mp)$$

which we note is zero if $\sin kp = 0$. So now assume $\sin kp \neq 0$ and substitute for A and B from (10.2) and set equal to 0:

$$- kp \cos kp(1 - 2\sin^2 kp) + 2\cos^3 kp \sin kp$$
$$+ m \cos kp - kp \cos kp - 2kp \sin^2 kp \cos kp + 2\sin^3 kp = 0$$

i.e. $\sin kp + k(1-p) \cos kp = 0$

Thus the eigenvalues are given by solving either

$$\sin(kp) = 0$$

or $\sin(kp) + k(1-p)\cos kp = 0.$

With these conditions on m (or k) the constant term should equal 0 for these values of m to be eigenvalues. The constant term is

$$(A/2p)\{-mp^2 \cos mp + 2p \sin mp + (2/m)(\cos mp-1)\}$$
$$- A(1-p/2)(-mp \cos mp + \sin mp)$$
$$+ Am(p/3-p^2/4)(1 - \cos mp)$$
$$+ (B/2p)(mp^2 \sin mp + 2p \cos mp -2/m \sin mp)$$
$$- B(1-p/2)(mp \sin mp + \cos mp - 1)$$
$$+ B(p/3-p^2/4)m \sin mp$$

If $\sin kp = 0$ then $\sin mp = 0$ and $\cos mp = 1$. The constant term is then:

$$(A/2p)(-mp^2) - A(1-p/2)(-mp)$$
$$+ (B/2p)(2p)$$

i.e. $-\frac{Amp}{2} + Amp - \frac{Amp^2}{2} + B$

or $B = -\frac{Amp}{2}(1+p)$

Thus A and B are determined if $\sin kp = 0$ and so $g(t)$ is determined, up to normalization, i.e.

$$g(t) \alpha \sin mt - \frac{mp}{2}(1+p)\cos mp ,$$

with $m = 2k$, and k determined by $\sin kp = 0$.

If the eigenvalues are given by $\sin kp + k(1-p)\cos kp = 0$ then (10.2) holds since $\sin kp \neq 0$ and $\cos kp \neq 0$, so $A, B \neq 0$. Thus substituting for A and B, the constant term is

$$(-\cos kp/2p)\{-mp^2 \cos mp + 2p \sin mp +$$
$$+ \cos kp(1-p/2)(-mp \cos mp + \sin mp) - \sin kp$$
$$+ (\sin kp)/2p\ (mp^2 \sin mp + 2p \cos mp)$$
$$- \sin kp(1-p/2)(mp \sin mp + \cos mp - 1)$$

i.e. $p(-k \cos kp + kp \cos kp - \sin kp)$

But $\sin kp + k(1-p)\cos kp = 0$, so the constant term is 0, and $g(t)$ is given by

$$g(t) \alpha - \cos kp \sin 2kt + \sin kp \cos 2kt$$

where k satisfies:

$$\tan kp = -k(1-p).$$

Thus the asymptotic distribution of U^2 can be found by finding the two sets of eigenvalues $\{\mu_i\}$ found by solving

(i) $\sin kp = 0$

or (ii) $\sin kp + k(1-p)\cos kp = 0$,

where $k = \frac{1}{2}\mu^{\frac{1}{2}}$.

The first set corresponding to solving (i) are given by

$$k_j p = j\pi \qquad j = 1,2,\ldots$$

or $\mu_j = \frac{4j^2\pi^2}{p^2}$ $\quad j = 1, 2, \ldots$

and the second set, found by solving (ii), correspond to four times the eigenvalues of ${}_pW^2$, found in §6.4.2. The eigenvalues of ${}_pW^2$ are given by solutions of

$$\sin mp + m(1-p)\cos mp = 0$$

where $m = \mu^{\frac{1}{2}}$.

Thus, since $k = \frac{1}{2}\mu^{\frac{1}{2}}$, one set of eigenvalues of ${}_pU^2$ is given by four times the set of eigenvalues of ${}_pW^2$.

The distribution of ${}_pU^2$ can be found in the usual way with these eigenvalues, noting that we can find the cumulants of ${}_pU^2$ from ${}_pW^2$ and W^2. Use

$$\varkappa_m = 2^{m-1}(m-1)! \sum_{j=1}^{\infty} \left(\frac{1}{\mu_j}\right)^m$$

and denote by μ^* the eigenvalues of ${}_pW$ and μ^+ the eigenvalues of W (i.e. $\mu_j^+ = \pi^2 j^2$).

Then $\varkappa_m({}_pU^2) = 2^{m-1}(m-1)! \sum_{j=1}^{\infty} \mu_j^{-m}$

$$= 2^{m-1}(m-1)!\left\{\sum_{j=1}^{\infty} \left(\frac{1}{4\mu_j^*}\right)^m + \left(\frac{p}{4\mu_j}\right)^m\right\}$$

$$= \frac{1}{4^m}\varkappa_m({}_pW^2) + \left(\frac{p}{4}\right)^m \varkappa_m(W^2)$$

The values of $\varkappa_m({}_pW^2)$ $(m = 1, 2, 3)$ are given in §6 in Table 6.2 and the values of $\varkappa_m(W^2)$ are well known (see Anderson and Darling (1952) for example).

§10.2.3 An alternative statistic for the circle

A criticism of Stephens' definition of ${}_pU^2$ keeping in mind that U^2 was introduced to test observations on the circle, is that if we consider the statistic

$$ {}_{q,q+p}U^2 = \int_q^{q+p} R^2(t)dt $$

where R(t) is now defined by

$$ R(t) = Y(t) - \frac{1}{p}\int_q^{q+p} Y(s)ds, $$

and that the choice of origin on the circle was arbitrary, we should like the distribution of ${}_{q,q+p}U^2$ to be the same as ${}_{o,p}U^2$ (i.e. ${}_pU^2$), and for a given set of data, on the circle, ${}_{q,q+p}U^2 = {}_{o,p}U^2$. Consider the covariance of R(t); this is given by:

$$ \begin{aligned} E[R(t)R(s)] &= \min(s,t) - st \\ &- \frac{1}{p}\left\{t(p+q)+\frac{q^2 t}{2} - (p+q)^2\frac{t}{2} - \frac{t^2}{2} - \frac{q^2}{2}\right\} \\ &- \frac{1}{p}\left\{s(p+q)+\frac{q^2 s}{2} - (p+q)^2\frac{s}{2} - \frac{s^2}{2} - \frac{q^2}{2}\right\} \\ &+ \frac{1}{p^2}\left\{\frac{(p+q)^3}{3} - \frac{(p+q)^4}{4} - q^2(p+q)^2 + q^3(p+q) \right. \\ &\left. + \frac{2}{3}q^3 - \frac{3}{4}q^4\right\} \end{aligned} \tag{10.3} $$

The distribution of ${}_{q,q+p}U^2$ is given by

$$ \sum_{j=1}^{\infty} \frac{Z_j^2}{\mu_j} $$

where Z_j are i.i.d. standard normal random variables and the μ are eigenvalues of the integral equation:

$$\mu \int_{q}^{q+p} \tau(s,t)g(s)ds = g(t)$$

where $\tau(s,t) = E[R(s)R(t)]$ (see 10.3). Now this integral equation can be written, by a change of origin,

$$\mu \int_{0}^{p} \tau(s+q,t+q)g(s+q)ds = g(t+q)$$

Consequently if ${}_{q,q+p}U^2$ is to have the same distribution as ${}_{p}U^2$ then the eigenvalues and eigenfunctions of the respective integral equations must be identical, i.e. $\tau(s+q,t+q) = \rho(s,t)$, where $\rho(s,t)$ is given by (10.1). However this is easily seen not to be true and so ${}_{q,q+p}U^2$ and ${}_{p}U^2$ are not identically distributed. There is the practical criticism of not being able to construct the statistic ${}_{q,q+p}U^2$ when only censored observations are available on the circle. Suppose we have observations, in radians, censored so that they lie in the interval $(0, 2\pi p)$ equivalent to $(0,p)$ on the line and we construct the statistic ${}_{p}U^2$ by transforming the observations to $(0,p)$. Then with another choice of origin, say $2\pi(1-q)$ radians on the previous scale, we now have observations in the interval $(2\pi q, 2\pi(q+p))$ and try to construct ${}_{q,p+q}U^2$ by transforming the observations to $(q, q+p)$. However we cannot construct ${}_{q,p+q}U^2$, since we do not know how many observations lie in $(0,2\pi q)$ or equivalently in $(0,q)$ and $F_n(t)$ cannot be known in $(2\pi q, 2\pi(q+p))$ or equivalently $(q, q+p)$. Thus ${}_{q,p+q}U^2$ only has meaning when all observations are available. However if all observations are available, then it would be better to use the statistic

$${}_{p}^{c}U^2 = \int_{0}^{p} \{Y_n(t) - \int_{0}^{1} Y_n(s)ds\}^2 dt$$

for it is seen that ${}_{p}^{c}U^2$ and ${}_{q,q+p}^{c}U^2$, where

$$ {}_{q,q+p}{}^{c}U^2 = \int_q^{q+p} \left\{Y_n(t)-\int_o^1 Y_n(s)ds\right\}^2 dt $$

have identical distributions, by following Watson's argument in his (1961) paper, and also that for a given set of data on the circle ${}_p{}^{c}U^2$ will be identical to ${}_{q,q+p}{}^{c}U^2$ with the origin now at $2\pi(1-q)$; so ${}_p{}^{c}U^2$ is effectively origin free. The distribution of ${}_p{}^{c}U^2$ can be found in a similar way as the distribution of U^2 was found. Now

$$ {}_p{}^{c}U^2 = \int_o^p Z^2(t)dt $$

where the covariance function of Z(t) is given by (10.3), so ${}_p{}^{c}U^2$ has asymptotic distribution that of

$$ \sum_{j=1}^{\infty} \frac{Z_j^2}{\mu_j} $$

where the Z_j are i.i.d. standard normal random variables and the μ are eigenvalues of the integral equation

$$ \mu \int_o^p \rho(s,t)g(s)ds = g(t) \qquad (10.4) $$

where $\rho(s,t) = E[Z(t)Z(s)]$

$$ = \min(s,t) - \tfrac{1}{2}(s+t)+\tfrac{1}{2}(s-t)^2 + \frac{1}{12} \qquad (10.5) $$

given by Watson (1961).

If this equations is differentiated and the differential equation solved then the general solution is

$$ g(t) = A\sin mt + B\cos mt, \quad \text{where } m = \mu^{\frac{1}{2}}. $$

Substituting g(t) into (10.4) we have an expression similar to that for ${}_pU^2$. The t^2 coefficient is given by:

$$ A\frac{m}{2}(1-\cos mp) + B\frac{m}{2}\sin mp . $$

thus with $m \neq 0$, for $g(t)$ to be an eigenfunction

$$A(1 - \cos mp) + B \sin mp = 0$$

or with $2k = m$, then if $\sin kp \neq 0$

$$A \sin kp + B \cos kp = 0$$

as before for ${}_pU^2$.

The t coefficient is given by

$$A\left[-\frac{m}{2} \cos mp + mp \cos mp - \sin mp - \frac{m}{2}\right]$$

$$+ B\left[\frac{1}{2} mp \sin mp - mp \sin mp + 1 - \cos mp\right]$$

which upon simplification substituting for A and B, and assuming $\sin kp \neq 0$, we have that

$$m \sin mp \cos mp\,(1-2p) + 2 \cos mp(1 - \cos mp) + mp \sin mp = 0 \qquad (10.6)$$

The constant term is given by

$$A[-mp^2 \cos mp + 2p \sin mp + \frac{2}{m}(\cos mp - 1) + mp \cos mp - \sin mp]$$

$$+ B[-2 + mp^2 \sin mp + 2p \cos mp - \frac{2}{m} \sin mp - mp \sin mp - \cos mp + 1]$$

which upon substitution for A and B gives

$$m \sin mp(p-1) + 2(\cos mp - 1) = 0 \qquad (10.7)$$

provided $\sin kp \neq 0$ and $p \neq 0$.

If we try and solve (10.6) and (10.7) simultaneously for (mp), it is found that $\sin mp = 0$ is the only solution. However this solution is included in $\sin kp = 0$. We express (10.6) and (10.7) in terms of kp rather than mp, and find that kp must satisfy

$$(1-p)(\sin kp + 1)(1 - 2 \sin^2 kp) = -2p \sin^3 kp$$

for $g(t)$ to be an eigenfunction.

Consequently the eigenvalues of the integral equation (10.4) are given by

$$\sin kp = 0 \qquad \text{i.e. } kp = j\pi \quad j = 1, 2, \qquad \text{i.e. } \mu_j = \left(\frac{2j\pi}{p}\right)^2 \qquad (10.8)$$

and

$$(1-p)(\sin kp + 1)(1 - 2\sin^2 kp) = -2p \sin^3 kp \qquad (10.9)$$

One set of eigenvalues is common to ${}_pU^2$, and the other set found by solving (10.9) for k is a new set.

§10.3 Goodness-of-fit statistics for the 2-sample problem and the bivariate sample with censored data

10.3.1 Introduction

In this section we extend the ${}_pW^2$ statistic for the problem of testing 2 samples given in 10.3.2, and in 10.3.3 ${}_pW^2$ is extended for the bivariate sample.

10.3.2 The 2-sample problem

The asymptotic theory of the censored statistics can be applied to the 2-sample problem. Suppose initially we have 2 random samples both censored on the right so that we have observations $x_1, \ldots, x_r \leq a_p$ from a continuous population with c.d.f. $F(x)$ and $y_1, \ldots, y_s \leq a_p$, from a continuous population with c.d.f. $G(y)$, where the censored x observations number $n - r$, and the censored y observations number $m - s$. The c.d.f.'s can be constructed for both the x and y samples:

$$F_n(x) = \frac{\#\text{ observations} \leqslant x}{n} \qquad \text{for } x \leq a_p$$

$$G_m(y) = \frac{\#\text{ observations} \leqslant y}{m} \qquad \text{for } y \leq a_p$$

We can then construct the 2-sample statistic

$${}_pW^2_{n,m} = \frac{mn}{m+n} \int_{-\infty}^{a_p} \left\{F_n(x) - G_m(x)\right\}^2 dH_{n,m}(x)$$

where

$$H_{n,m}(x) = \frac{nF_n(x) + mG_m(x)}{m+n}$$

and p is defined by $H_{n,m}(a_p) = p$.

Then under the null hypothesis $H_o : F(x) \equiv G(x)$ it is easy to show that ${}_pW^2_{n,m} \longrightarrow {}_pW^2$ as $n,m \longrightarrow \infty$ in some way, following Kiefer (1959).

The 2-sample ${}_pA^2$ and ${}_pU^2$ can be constructed in similar ways, and the doubly censored statistics also constructed.

§10.3.3 The bivariate sample problem

The censored statistics can also be constructed to test the bivariate distribution for (a) independence, and (b) goodness-of-fit, and discuss here the former problem.

Blum, Kiefer and Rosenblatt (1961) considered statistics to test for independence. First define $F_n(x,y)$ for the bivariate sample by $F_n(x,y) = r/n$ when exactly r of the n observations (x_i,y_i) satisfy $x_i \leq x$, $y_i \leq y$, or in terms of the indicator function

$$F_n(x,y) = \frac{1}{n} \sum_{i=1}^{n} \psi_x(x_i)\psi_y(y_i)$$

$$\text{where } \psi_t(s) = \begin{cases} 1 & s \leq t \\ 0 & s > t \end{cases}$$

Then they considered statistics based on the process

$$T_n(x,y) = F_n(x,y) - G_n(x)H_n(y)$$

where $G_n(x)$ is the e.d.f. for the observations $(x_1,\ldots,x_n)$ only and $H_n(y)$ is the e.d.f. for the observations $(y_1,\ldots,y_n)$ only. One such statistic is the CVM type statistic

$$B_n^2 = n \iint \{T_n(x,y)\}^2 \, dF_n(x,y)$$

Obviously such statistics can be adapted for use with censored data:

$${}_rB_n^2 = n \int_{-\infty}^{x_{(r)}} \int_{-\infty}^{y_{(r)}} \{T_n(x,y)\}^2 \, dF_n(x,y) \quad ,$$

where $x_{(r)}$, $y_{(r)}$ are the r^{th} order statistics of the respective samples. The asymptotic distribution of ${}_rB_n^2$ is then found by solving the integral equation

$$\lambda\int_o^p\int_o^p C(s,t;u,v)f(u,v)\,du\,dv = f(s,t) \qquad (10.10)$$

where $r/n \longrightarrow p$ and $C(s,t;u,v)$ is the covariance function of the bivariate process $T_n(x,y)$ when transformed to $(0,1)$ when the null hypothesis is true. Under these conditions

$$C(s,t;u,v) = \{\min(s,u)-su\}\{\min(t,v)-tv\}$$

$$0 \leq s,t,u,v \leq 0.$$

The eigenvalues of the integral equation (10.10) with p=1 are given by $\lambda_{jk} = (\pi j)^2(\pi k)^2$ and so B_n^2 is asymptotically distributed as

$$\sum_{j,k=1}^{\infty} \frac{Z_{jk}^2}{\pi^2 j^2 \pi^2 k^2}$$

where the Z_{jk} are i.i.d. standard normal $j,k = 1,2,\ldots$. Similarly if μ_j are the eigenvalues associated with ${}_pW^2$ (see 6.32A), i.e. the μ_j are eigenvalues of the integral equation

$$\mu\int_o^p \{\min(s,t)-st\}f(s)ds = f(t)$$

then ${}_rB_n^2$ has asymptotic distribution given by ${}_pB^2$ where

$${}_pB^2 = \sum_{j,k=1}^{\infty} \frac{Z_{jk}^2}{\mu_j\mu_k}$$

So the distribution of ${}_pB^2$ can be found by the method that the distribution of ${}_pW^2$ was found.

Blum, Kiefer and Rosenblatt (1961) extended the theory of B_n^2 upto m dimensions, so as to test the joint independence of an m-dimensional random variable. The theory is straightforward if $T_n(x_1,x_2,\ldots,x_m)$ is defined so that the asymptotic covariance function can be factored:

$$E[T_n(x_1,\ldots,x_m)T_n(y_1,\ldots,y_m)]$$

$$= \prod_{i=1}^{m} \{\min(x_i,y_i)-x_iy_i\}$$

so enabling the corresponding integral equation to be solved. Blum, Keifer and Rosenblatt (1961) show how T_n is defined for m = 3, and consequently the 3-dimensional statistic B_n^2 has asymptotic distribution that of

$$\sum_{j,k,\ell} Z_{jk\ell} /(\pi j \, \pi k \, \pi \ell)^2$$

where the $Z_{jk\ell}$ are i.i.d. standard normal $(j,k,\ell = 1,2,\ldots)$. Consequently ${}_pB^2$ has distribution

$$\sum_{j,k,\ell} Z_{jk\ell} /(\mu_j\mu_k\mu_\ell)$$

where the μ_j are defined in (6.32A).

Rothman (1971) considers the version of B_n^2 for observations on a circle. He considers the new process

$$Z_n(x,y) = T_n(x,y) + \iint T_n(x,y)dF_n(x,y) - \int T_n(x,y)dG_n(x)$$

$$- \int T_n(x,y)dH_n(y),$$

and shows that the statistic

$$C_n^2 = n \iint \{Z_n^{*}(x,y)\}^2 \, dF_n(x,y)$$

is asymptotically distributed as

$$\sum_{j,k=1} Z_{jk}/\lambda_{jk}$$

where the λ_{jk} are eigenvalues of the integral equation

$$\lambda\int_0^1\int_0^1 C(s,t;u,v)f(u,v)du\,dv = f(s,t)$$

and C(s,t;u,v) is the asymptotic covariance function of the process $Z_n(x,y)$, given by

$$C(s,t;u,v) = \{\min(s,u-\tfrac{1}{2}(s+u)+\tfrac{1}{2}(s-u)^2+\tfrac{1}{12}\}$$
$$\times \{\min(t,v)-\tfrac{1}{2}(t+v)+\tfrac{1}{2}(t-u)^2+\tfrac{1}{12}\}$$

So once again the covariance kernel factors and the λ_{jk} are given by $\lambda_{jk} = \lambda_j\lambda_k$ where the λ_j are eigenvalues for Watson's statistic U^2 (see §10.2.1). Similarly we can extend C_n^2 for the case of censoring and find the eigenvalues $\mu_{jk} = \mu_j\mu_k$ where the μ_j are eigenvalues for ${}_pU^2$ given by (10.7) and (10.8).

§10.4 An alternative method of testing for normality with censored data

An alternative method for testing for normality not dependent on the c.d.f. is the Shaprio-Wilk statistic W, introduced in Shapiro and Wilk (1965). W is the quotient of the best linear unbiased estimate of σ, say $\hat{\sigma}_V$, and the maximum likelihood estimate of σ, say $\hat{\sigma}_L$, upto a constant term:

$$W \propto \frac{\hat{\sigma}_V}{\hat{\sigma}_L}$$

For the full sample

$$\hat{\sigma}_V = \sum_{i=1}^{n} a_i x_{(i)}$$

where the $\{a_i\}$ are given by Sarhen and Greenberg (1956), §10 under the best linear unbiased coefficients for σ of the normal distribution.

Also

$$\hat{\sigma}_L = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Similarly we could construct a W statistic based on censored data where

$$\hat{\sigma}_V = \sum_{i=s}^{r} b_i x_{(i)}$$

the $\{b_i\}$ are again given by Sahren and Greenberg (1956, §10 when the available observations are $(x_{(s)},\ldots,x_{(r)})$. The maximum likelihood estimate $\hat{\sigma}_L$ is given by Cohen (1959), for singly censored observations only.

An alternative approach is that indicated by Lindgren (1968, §6.4) who approximates to $W^{\frac{1}{2}}$ by the cross-correlation coefficient of the order statistics $(x_{(1)},\ldots,x_{(n)})$ and the expected values of the order statistics $(m_1,\ldots,m_n)$ from an $N(0,1)$ population.

$$W^{\frac{1}{2}} \simeq r(\underset{\sim}{x},\underset{\sim}{m})$$

$$= \frac{\Sigma(x_{(i)}-\bar{x})(m_i-\bar{m})}{\sqrt{\Sigma(x_{(i)}-\bar{x})^2\,\Sigma(m_i-\bar{m})^2}}$$

where $m_i = E[Y_{(i)}]$ where $Y_i \sim N(0,1)$

$$r(\underset{\sim}{x},\underset{\sim}{m}) \;\alpha\; \frac{\Sigma x_{(i)}m_i}{\sqrt{\Sigma(x_{(i)}-\bar{x})^2}}$$

$$\text{since } \Sigma\, m_i = 0.$$

A good approximation to the coefficients $\{a_i\}$ is

$$a_i = 2m_i\,.$$

So W and $r(\underset{\sim}{x},\underset{\sim}{m})$ are equivalent upto this approximation.

Similarly we could construct the statistic $r(\underset{\sim}{x},\underset{\sim}{m})$ for censored data. However, Plackett (1958) shows that the best linear unbiased estimate $\hat{\sigma}_V$ of σ and the maximum likelihood estimate $\hat{\sigma}_L$ of σ are equivalent asymptotically, and $\hat{\sigma}_V$ is given by $\Sigma\, m_i x_{(i)}$ upto a multiplying constant, thus

$r(\underset{\sim}{x},\underset{\sim}{m}) \longrightarrow 1$ with probability 1, as $n \longrightarrow \infty$. Consequently the asymptotic distribution of $W(= \hat{\sigma}_V/\hat{\sigma}_L)$ is given by the constant 1 with probability 1.

However this type of statistic does not lead to analytic results and the whole distribution theory rests on Monte-Carlo sampling.

Although the K-S and CVM statistics have been much maligned in recent years (Shapiro and Wilk (1965), Shapiro, Wilk and Chen (1968), who use the statistics in the wrong context), the CVM statistics do appear to be powerful omnibus statistics with simple computational formulae and elegant asymptotic theory (which is not irrelevant in practice since the statistics are quickly convergent to their asymptotic distributions).

§10.5 Conclusion

In conclusion, we hope that the goodness-of-fit statistics investigated here will be a help to the more efficient analysis of data.

REFERENCES

ABRAMOWITZ, M and SEGUN, I.A. (1964) Handbook of Mathematical Functions, Washington : National Bureau of Standards

ANDERSON, T.W. and DARLING, D.A. (1952) Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. Ann.Math.Statist.v 23, pp 193-212

ANDERSON, T.W. and DARLING, D.A. (1954) A test for goodness-of-fit. J.Amer.Statist.Ass. v49, pp 300-310

BIRNBAUM, Z.W. (1952) Numerical tabulation of the Distribution of Kolmogorov's Statistic for finite sample size. J.Amer.Statist.Ass. v.47 p 431

BIRNBAUM, Z.W. and TINGEY, F.H. (1951) One-sided confidence contours for probability functions' Ann.Math.Statist. v.22 pp.592-6

BIRNBAUM, Z.W. and LIENTZ, B.P. (1969) Tables of critical values of some Renyi type statistics for finite sample sizes, J.Amer.Statist.Ass. pp.871-877

BLUM, J.R, KIEFER, J. and ROSENBLATT, M. (1961) Distribution free test of independence based on sample distribution function. Ann.Math.Statist. v.32, pp.485-492

CHERNOFF, H., GASTWIRTH, J.L., and JOHNS, M.V. (1967) Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. Ann.Math.Statist. v.38 pp.52-72

COHEN, A.C. (1959) Simplified estimates for the normal distribution when samples are singly censored or truncated. Technometrics v.1, pp.217-237

COHEN, A.C. (1961) Tables of MLE's for singly truncated and singly censored samples. TECH. v.3 pp.535-641.

DARLING, D.A., (1955) The Cramér Smirnov test in the parametric case, Ann.Math.Statist. v.26, pp.1-20

DARLING, D.A. (1957) The Kolmorov-Smirnov, Cramér-von Mises Tests. Ann.Math. Statist. v.28, pp.823-838

DAVID, F.N. (1950) Two combinatorial tests of whether a sample has come from a given population BIOKA v.37 pp.97-110

DAVID, H.A., (1970) Order statistics, Wiley, New York..

DODGSON, J. (1972) Components of the Cramér-von Mises Statistic M.Sc. Dissertation, Nottingham University.

DURBIN, J. (1970) Asymptotic distributions of some statistics based on the bivariate sample distribution function in 'Non-parametric techniques in statistical inference' Edited by M.Puri, Cam. Univ.Press.

DURBIN, J. and KNOTT, M. (1972) Components of the Cramér-von Mises Statistic, I. Journ.Roy.Statistics Soc. v.34 pp.290-307

HOEFFDING, W. (1965) 'Asymptotically optimal tests for multinomial distributions' Ann.Math.Statist. v.36, pp.369-405

HOGG, R.V. and CRAIG, A.T. (1970) 'Introduction to Mathematical Statistics' 3rd Ed. Macmillan Co, London.

ILLYASENKO, O.A. (1952) On the influence of grouping of empirical data on A.N.Kolmogcrov's criterion of fit (in Ukranian). Dopovidi, Akad. Nauk Ukrain R.S.R. pp.3-6

IMHOF, J.P., (1961) Computing the distribution of quadratic forms in normal variables BIOKA v.48, pp.419-426

JOHNSON, N.L., NIXON, E. AMOS, D.E. and PEARSON, E.S. (1963) Tables of percentage points of Pearson curves for given β_1 and β_2, expressed in standard measure. BIOKA v.50, pp.459-98

KAC, M., KIEFER, J., and WOLFOWITZ, J. (1955) 'On tests of noramlity and other tests of goodness-on-fit based on distance methods' Ann.Math.Statist. v.26 pp.189-211

KAC, M. and SIEGERT, A.J.F., (1947) An emplicit representation of a stationary Gaussian process. Ann. Math. Statist. v.18, pp.438-442.

KEMPTHORNE, O. (1967) The classical problem of inference - goodness-of-fit in 'Precedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, Los Angeles.

KENDALL, M.G. (1961) A course in the geometry of n-dimensions Griffin, London.

KENDALL, M.G., and STUART, A. (1973) 'The Advanced theory of statistics' Vol.2, 3rd Edn. Griffin, London.

KIEFER, J., (1959) K-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests. Ann.Math.Statist. v.30, pp.420-47

KOLMOGOROV, A. (1933) Sulla determinazione empirica di una legge di Distribuzione, Giorao dell 'Instit. degli. Attuari vol.4, pp.83-91

KUIPER, N.H., (1960) Tests concerning random points on a circle Proc.Konikl Nederl. Akad. van Wetenschappen. Series A. pp.38-49

LEWIS, P.A.W. (1961) Distribution of the Anderson-Darling Statistic. Ann.Math.Statist. v.32. p.118

LINDGREN, B.W. (1968) Statistical Theory, MacMillan, New York.

MAAG, U.R., STREIT, F., DROUILLY, P. (1973) Goodness-of-fit for Grouped Data. J.Amer.Statist.Ass. v.68 pp.462-465

MARSHALL, A.W., (1958). The small sample distribution of nW_n^2 Ann.Math.Statist. v.29, pp.307-9

MASSEY, F.J. (1950) A note on the estimation og a distribution function by confidence limits, Ann.Math.Statist. v.21, pp.116-119.

NOETHER, G.E. (1963) A Note on the Kolmogorov-Statistic in the discrete case. METRIKA v.7, pp.115-116.

PATNAIK, P.B. (1949) The non-central χ^2 and F distributions and their applications BIOKA v.36, pp.202-232.

PEARSON, E.S. and HARTLEY, H.O. (1966) 'Biometrica Tables for Statisticians' Vol.I. Third Ed'n. Cam. Univ. Press.

PEARSON, E.S. and STEPHENS, M.A (1962) The goodness-of-fit tests based on W_N^2 and U_N^2. BIOKA v.49, pp.397-402

PLACKETT, R.L. (1958) Linear estimation from censored data. Ann.Math. Statist. v.29. pp.131-42

RENYI, A. (1953) On the theory of order statistics. Acta.Math. Akad.Sci. Hungar 4. pp.191-231

ROTHMAN, E.D., (1971) Tests of coordinate independence for a bivariate sample on a torus. Ann.Math.Statist. v.42, pp.1962-69

SANOV, I.N. (1961) On the probability of large deviations of random variables. Select.Transl.Math. Statist. and Prob. v.1, pp.213-244

SARHAN, A.E. and GREENBERG, B.G. (1962) 'Contributions to order Statistics' Wiley, New York.

SAW, J.G. (1961) The bias of the maximum likelihood estimates of the location and scale parameters given a type II censored normal sample. Biometrika v.48, pp.448-452.

SESHADRI, V., CSORGO, M., STEPHENS, M.A., (1969) Tests for the esponential distribution using Kolmogorov - type statistics. J.R.Statist.Soc. v.31, pp.499-509

SHAPIRO, S.S. and WILK, M.B. (1965) Analysis of variance test for normality. Biometrika 52, pp.591-611

SIEGEL, S. (1956) Non-parametric statistics for the Behavioral Sciences. McGraw-Hill, New York.

SMIRNOV, N.V. (1936) Sur La distribution de W^2 C.R. Acad.Sci.Paris v.202 pp.449-52

STECK G.P. (1971) Rectangle probabilities for uniform order statistics Ann.Math. Statistic v.42, pp.1-11.

STEPHENS, M.A., (1965) The goodness-of-fit statistic V_n: distribution and significance points. BIOKA v.52, pp.309-21

STEPHENS, M.A. and MAAG, U.R. (1968a) Further percentage points for W_N^2. BIOKA v.55, pp.428-430

STEPHENS, M.A. and MAAG, U.R. (1968b) The V_{NM} two-sample test

Ann. Math. Statist. v.39 pp.923-35

STEPHENS, M.A. (1970) Use of Kolmogorov-Smirnov, Cramér-von Mises, and related statistics without extensive tables. J.R. Statist.Soc. B 32 pp.115-122

STEPHENS, M.A. (1973) 'Asymptotic results and percentage points for goodness-of-fit statistics with unknown parameters.' Submitted to Annals of Statistics.

SUKHATME, S. (1972) Fredholm Determinant of a positive definite kernel of a special type and its application. Ann.Math.Statist. Vol.43, pp.1914-1926.

TIKU, M.L., (1965) Chi-square approximations for the distribution of goodness-of-fit statistics U_n^2 and W_n^2. BIOKA 52 pp.630-33.

TYURIN, YU.N. (1970) 'On parametric hypothesis testing with non-parametric tests' Translation. in 'Theory of Probability and Applications' SIAM No.4 p.722

WHITTAKER, E.T., and WATSON, G.N. A course of modern analysis, 4th Edition Cambridge, Cam. Univ. Press.

WILKINSON, J.H. (1965) The algebraic eigenvalue problem. Clarendon Press, London.

WILKS, S.S. (1962) Mathematical Statistics, John Wiley, New York.

WILKS, S.S. and DALY, J.F. (1939) An optimum property of confidence regions associated with the likelihood function. Ann.Math. Statist. v.10. p.225

Appendix

Distribution of the statistics nD^+, nD, nV for discrete data with n observations, under the null hypothesis that q points (or categories) are equally likely, found using the method of §2 with $q = 3(1)\ 10, 12$ and n a multiple of q and $n \leq 30$. The probabilities are accurate to the number of significant points given; a digit in parentheses indicates some doubt about its accuracy.

$\overset{+}{D}$ Q = 3 Pr $(ND^{+} \leq k)$

	N	3	6	9	12
k	0	.59259	.51852	.48412	.46342
	1	.96296	.84225	.76762	.71694
	2	1.00000	.98217	.93817	.89528
	3		.99863	.99172	.97503
	4		1.00000	.99903	.99614
	5			.99995	.99946
	6			1.00000	.99995

	N	15	18	21	24
k	0	.44928	.43885	.43076	.42426
	1	.67993	.65149	.62881	.61024
	2	.85774	.82553	.79786	.77396
	3	.95408	.93210	.91051	.89013
	4	.98955	.98007	.96848	.95618
	5	.99819	.99564	.99125	.9860
	6	.99971	.99923	.99799	.9954
	7	.99996	.99968	.99938	1.00

	N	27	30
k	0	.4190	.4156
	1	.5949	.5844
	2	.7541	.7354
	3	.8721	.8675
	4	.9496	.946
	5	.9892	.98(9)
	6	.9983	.99(2)

D	Q = 3	Pr(ND ≤ k)		
N	3	6	9	12
k 0	.22222	.12346	.08535	.06520
1	.92593	.68587	.53986	.44423
2	1.00000	.96433	.87639	.79085
3		.99726	.98344	.95006
4		1.00000	.99807	.99229
5			.99990	.99891
6			1.00000	.99991
7				.99995

N	15	18	21	24
k 0	.05274	.04428	.03815	.03351
1	.37713	.32755	.28944	.25926
2	.71627	.65272	.59863	.55232
3	.90819	.86421	.82115	.78031
4	.97911	.95995	.93692	.91176
5	.99639	.99098	.98240	.97103
6	.99943	.99830	.99598	.99208
7	.99993	.99937	.99925	.99787
8			.99990	

N	27	30
k 0	.02988	.02696
1	.2348	.2145
2	.5124	.4777
3	.7422	.7068
4	.8857	.8594
5	.9575	.9429
6	.9865	.9799
7	.9941	

V Q = 3 Pr(N V ≤ k)

k	N	3	6	9	12
	0				
	1	.88889	.61728	.46944	.37816
	2	1.00000	.94650	.83432	.73285
	3		.99588	.97516	.92957
	4		1.00000	.99710	.98843
	5			.99985	.99837
	6			1.00000	.99986
	7				.99999

k	N	15	18	21	24
	0				
	1	.31644	.27198	.23844	.21226
	2	.64920	.58100	.52497	.47839
	3	.87527	.82045	.76866	.72105
	4	.96970	.94345	.91285	.88029
	5	.99458	.98669	.97455	.95885
	6	.99914	.99744	.99403	.98849
	7	.99991	.99956	.99878	.99726
	8	.99999	.99993	.99978	.99941
	9			.99997	.99992

k	N	27	30
	0		
	1	.19125	.17402
	2	.43917	.40575
	3	.67779	.63868
	4	.84735	.81496
	5	.94049	.92035
	6	.98079	.97111
	7	.99473	.99103
	8	.9987	
	9	.9998	

D^+ $Q = 4$ $Pr\ (ND^+ \leq k)$

	N	4	8	12	16
k	0	.48828	.41524	.38277	.36365
	1	.90625	.75032	.66442	.60933
	3	1.00000	.99243	.97013	.93693
	4		.99962	.99442	.98350
	5		.99998	.99938	.99646
	6			.99997	.99946
	7			1.00000	.99993

	N	20	24	28
k	0	.35075	.34133	.33408
	1	.57052	.54143	.51872
	2	.76930	.73031	.69801
	3	.90196	.86874	.83821
	4	.96700	.94747	.92650
	5	.99082	.98252	.97459
	6	.99779	.99505	.99305
	7	.99942	.99923	
	8	.99987	1.00000	
	9			
	10			

Pr (ND ≤ k)

k	N	4	8	12	16
	0	.09375	.03845	.02203	.01468
	1	.81250	.50842	.35248	.26214
	2	.99219	.89209	.75296	.63672
	3	1.00000	.98486	.94026	.87389
	4		.99924	.98885	.96701
	5		.99997	.99876	.99294
	6			.99993	.99894
	7			.99999	.99989

k	N	20	24	28
	0	.01067	.00820	.00656
	1	.20461	.16539	.13725
	2	.54442	.47115	.41235
	3	.80414	.73813	.69805
	4	.93402	.89481	.85319
	5	.98173	.96474	.94312
	6	.99576	.98986	.98089
	7	.99926	.99762	
	8	.99996	.99947	
	9		1.00000	

V Q = 4 Pr (NV ≤ k)

N	4	8	12	16
k 0				
1	.70313	.38025	.24508	.17443
2	.98438	.81818	.63567	.50333
3	1.00000	.97199	.89518	.79556
4		.99847	.97921	.94045
5		1.00000	.99760	.98670
6			.99985	.99795
7			.99999	.99998

N	20	24	28
k 0			
1	.13220	.10462	.08545
2	.40889	.33983	.28787
3	.70081	.61823	.54817
4	.88654	.82687	.76734
5	.96633	.93700	.90148
6	.99193	.98099	.96488
7	.99854	.99522	.98927
8	.99981	.99905	.99721
9	.99998	.99985	.99941
10	1.00000	.99998	

$Q = 5$ $\qquad$ $Pr\ (ND^{+} \leq k)$

	N	5	10	15
k	0	.41472	.34598	.31630
	1	.84544	.67113	.58223
	2	.98496	.90064	.81223
	3	.99968	.97919	.93839
	4	1.00000	.99762	.98399
	5		.99982	.99708
	6		1.00000	.99962
	7			.99995

	N	20	25	30
k	0	.29905	.28752	.27922
	1	.52730	.48948	.46189
	2	.74384	.69114	.65053
	3	.89027	.84470	.80377
	4	.96190	.93466	.90555
	5	.98898	.97648	.95840
	6	.99746	.99270	.98490
	7	.99944	.99808	
	8	.99977	.99926	

D Q = 5 Pr (nD ≶ k)

	N	5	10	15
k	0	.03840	.01161	.00551
	1	.69120	.36535	.22302
	2	.96992	.80135	.62698
	3	.99936	.95838	.87678
	4	1.00000	.99523	.96798
	5		.99965	.99416
	6		.99999	.99924
	7			.99992
	8			

	N	20	25	30
k	0	.00320	.00209	.00147
	1	.14988	.10754	.08089
	2	.49665	.40086	.32943
	3	.78083	.69051	.61098
	4	.92383	.86917	.81129
	5	.97797	.95283	.92076
	6	.99504	.98544	.97072
	7	.99913	.99632	.99057
	8	.99984	.99943	
	9			

V Q = 5 Pr(nV≤k)

k	N	5	10	15
	0			
	1	.51840	.21805	.11917
	2	.93440	.65673	.44608
	3	.99840	.91435	.77200
	4	1.00000	.98888	.93296
	5		.99914	.98668
	6		.99998	.99982
			1.00000	.99999
				1.00000

k	N	20	25	30
	0			
	1	.07497	.05148	.03752
	2	.31761	.23639	.18239
	3	.63102	.51683	.42762
	4	.85085	.76028	.67381
	5	.95285	.90379	.84588
	6	.98867	.96818	.93826
	7	.99792	.99148	.97895
	8	.99970	.99814	.99394
	9	.99997	.99967	.99852
	10	1.00000		

D^+ $Q = 6$

$Pr(ND^+ \leq k)$

	N	6	12	18
k	0	.36023	.29641	.26941
	1	.78712	.60472	.51670
	2	.96721	.85256	.75207
	3	.99811	.95968	.90117
	4	.99998	.99289	.9678(3)
	5	1.00000	.99913	.9920(1)
	6		.99994	.998(5)
	7		1.00000	.999(8)

	N	24	30
k	0	.25387	.24354
	1	.46373	.4278(2)
	2	.67901	.6246
	3	.8414(6)	.7884
	4	.934(2)	.8968
	5	.976(9)	.97
	6	.993(7)	
	7	.999(1)	
	8	1.00000	

D Q = 6 Pr(ND≤k)

	N	6	12	18
k	0	.01543	.00344	.00135
	1	.57656	.25740	.13833
	2	.93441	.70572	.51169
	3	.99623	.91936	.80247
	4	.99996	.98578	.93565
	5	1.00000	.99827	.98401
	6		.99988	.99692
	7			.99955

	N	24	30
k	0	.00069	.00040
	1	.08399	.05540
	2	.37967	.28926
	3	.68400	.58114
	4	.86797	.79391
	5	.95305	.91164
	6	.98625	.96683
	7	.99665	
	8	.99964	

V Q = 6 Pr(NV≤k)

	N	6	12	18
k	0			
	1	.36651	.12009	.05569
	2	.85391	.50056	.29690
	3	.98958	.82748	.63159
	4	.99987	.96420	.85855
	5	1.00000	.99526	.96040
	6		.99966	.99176
	7		.99998	.99871

	N	24	30
k	0		
	1	.03o99	.01928
	2	.19011	.12968
	3	.47398	.36046
	4	.73533	.61930
	5	.89337	.81252
	6	.96574	.92232
	7	.99112	.97296
	8	.99812	.99204

D_n^+ Q=7 $Pr(nD_n^+ \leq k)$

	N	7	14
k	0	.31831	.25921
	1	.73357	.54913
	2	.94469	.80536
	3	.99456	.93571
	4	.99980	.98502
	5	1.00000	.99745
	6		.99971
	7		.99997

	N	21	28
k	0	.23460	.22053
	1	.46376	.41337
	2	.69760	.62272
	3	.86196	.7937
	4	.94719	.9024
	5	.98377	.959(6)
	6	.9959	.986
	7	.999	.996
	8	1.000	.998

D_n Q = 7 $Pr(nD_n \leq k)$

	N	7	14
k	0	.00612	.00100
	1	.47446	.17888
	2	.88937	.61294
	3	.98911	.87144
	4	.99960	.97004
	5	1.00000	.99489
	6		.99942
	7		.99995

	N	21	28
k	0	.00033	.00014
	1	.08461	.04642
	2	.41184	.28623
	3	.72452	.59090
	4	.89440	.80498
	5	.96758	.91953
	6	.99193	.97172
	7	.99837	.99152
	8	.99964	.99771

V_n Q = 7 $Pr(n V_n \leq k)$

	N	7	14
k	1	.25245	.06453
	2	.75581	.36810
	3	.96881	.72439
	4	.99868	.92206
	5	.99999	.98504
	6	1.00000	.99815
	7		.99985

	N	21	28
k	1	.2542	.01251
	2	.19077	.10991
	3	.49814	.34259
	4	.91664	.81373
	6	.97699	.92644
	7	.99502	.97585
	8	.99917	.99335
	9		.99847

Q = 8 D_n $Pr(nD_n \leq k)$

	N	8	16	24
k	0	.00240	.00029	.00008
	1	.38659	.12307	.05124
	2	.83842	.52698	.32811
	3	.97741	.81816	.64753
	4	.99849	.94834	.84717
	5	.99996	.98878	.94523
	6		.99823	.98359
	7		.99980	.9959
	8			.9992

Q=8 V_n $Pr(nV_n \leq k)$

	N	8	16	24
k	1	.17093	.03413	.01142
	2	.65232	.26390	.11960
	3	.93472	.61747	.38205
	4	.99475	.86465	.66473
	5	.99986	.96602	.85771
	6	1.00000	.99405	.95176
	7		.99926	.98681
	8			.99711

Q = 9

D_n^+ $Pr(nD_n^+ \leq k)$

N	9	18	27	
k 0	.25812	.20716	.18638	
1	.64198	.46246	.38411	
2	.89215	.71917	.60583	
3	.98061	.88100	.7852	
4	.99808	.96080	.8981	
5	.99991	.98979	.9588	
6	1.0000	.99796	.9856	
7		.99993	.9957	.9955
8			.9990	

D_n $Pr(nD_n \leq k)$

N	9	18	27
k 0	.00094	.00008	.00002
1	.31261	.08402	.03079
2	.78442	.44956	.25938
3	.96121	.76234	.57422
4	.99615	.92160	.79648
5	.99982	.97957	.91790
6	1.00000	.99593	.97169
7		.99938	.9917
8		.99998	.9978

$Q = 9$ V_n $Pr(nV_n \leq k)$

	N	9	18	27
k	0			
	1	.11439	.01786	.00508
	2	.55206	.18572	.07368
	3	.88869	.51537	.28689
	4	.98637	.79641	.56507
	5	.99926	.93733	.78818
	6	.99999	.98586	.91591
	7	1.00000	.99764	.97248
	8		.99972	.99257
	9			.9983
	10			.9997

Q = 10

D_n^+ $Pr(nD_n^+ \leq k)$

	N	10	20	30
k	0	.23579	.18824	.16900
	1	.60324	.42826	.35352
	2	.86454	.68096	.56748
	3	.97051	.85259	.74958
	4	.99611	.94544	.87200
	5	.99972	.98361	.9433
	6	.99999	.99610	.9789
	7		.9992(5)	.995

D_n $Pr(nD_n \leq k)$

	N	10	20	30
k	0	.00036	.00002	.00000
	1	.25128	.05701	.01839
	2	.72946	.38118	.20380
	3	.94101	.70602	.50608
	4	.99222	.89091	.74433
	5	.99943	.96723	.88667
	6	.99998	.99219	.95635
	7		.99852	.9854
	8			.996

Q = 10 V_n Pr(nVn ≤k)

	N	10	20	30
k	1	.07591	.00927	.00224
	2	.46008	.12890	.04480
	3	.83353	.42303	.21193
	4	.97236	.72221	.47216
	5	.99763	.89957	.71292
	6	.99991	.97249	.87078
	7	1.00000	.99424	.95125
	8		.99909	.98454
	9			.99586

Q = 12

D_n^+ (Pr(nD_n^+ ≤ k)

	N	12	24	N	12	24
k	0	.20100	.15916		.00005	.00000
	1	.53729	.37277		.16014	.02589
	2	.81007	.61387		.62209	.27022
	3	.94563	.79699		.89126	.59711
	4	.98943	.91084		.97885	.82180
	5	.99866	.96721		.99732	.93402
	6	.99990	.9898		.99979	.97968
	7	.99999	.9977		.99999	.9948
	8					.999

V_n Pr(nV_n ≤ k)

	N	12	24
k	1	.03285	.00246
	2	.30887	.06029
	3	.70834	.27449
	4	.92617	.57180
	5	.98869	.80342
	6	.99899	.9286
	7	.99995	.9792
	8		.995
	9		.999