CRAMÉR-VON MISES STATISTICS FOR DISCRETE DISTRIBUTIONS

by

John J. Spinelli

Master of Science, Simon Fraser University, 1980

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Mathematics & Statistics

SIMON FRASER UNIVERSITY
March 1994

ISBN 0-612-01156-9

APPROVAL

Name: John J. Spinelli

Degree: Doctor of Philosophy

Title of thesis: Cramér-Von Mises Statistics for Discrete Distributions

Examining Committee: Dr. S.K. Thomason
Chair

Dr. M.A. Stephens, Senior Supervisor

Dr. R.A. Lockhart

Dr. C. Dean

Dr. D. Eaves

Dr. F. O'Reilly, External Examiner

Date Approved: March 28, 1994

Title of Thesis/Project/Extended Essay

CRAMÉR VON-MISES STATISTICS FOR DISCRETE DISTRIBUTIONS

Author: ______________________

(signature)

JOHN SPINELLI

(name)

28/2/94

(date)

Abstract

Testing the hypothesis that a sample of data arises from a specified distribution, called goodness-of-fit, is an important problem in statistics. To date most of the research has focussed on continuous distributions. Tests based on the empirical distribution function, and in particular the Cramér-von Mises statistics, have been shown to be powerful tests of fit for such distributions.

Discrete distributions are important to many areas of research, and often arise with medical data. In this thesis, the Cramér-von Mises statistics are developed for the Binomial and Poisson distributions. The asymptotic distributions of the test statistics are derived, and the distributions for finite samples are obtained by Monte Carlo methods. They are shown to converge rapidly to their asymptotic distributions. Power studies are given to compare the new tests to other tests which have been proposed for these distributions.

Another important research area is testing goodness-of-fit for regression models. Here the hypothesis is that the data are from a specified distribution, but with mean value dependent on a set of covariates. The regression model for normally distributed observations has been extensively studied. In this thesis, several analogues to the Cramér-von Mises statistics are derived for testing goodness-of-fit for discrete regression models. Asymptotic theory is given and the properties of the test statistics are examined.

Acknowledgements

Many individuals have helped make it possible for me to successfully complete this thesis. Although it is not possible to name them all, I give them my thanks. There are some who merit special recognition and I would like to take this opportunity to acknowledge their contribution.

First, I would like to thank my parents for giving me every opportunity to fulfill my potential. Thanks to my fellow graduate students for their help and in particular, Francois Bellavance, for his support. I most heartily thank Vincent Ng for his endless assistance during my struggles with computer technology.

I want to express my gratitude to the members of the Department of Mathematics and Statistics for their flexibility when faced with the unusual situation that I presented. In particular, I wish to acknowledge the invaluable guidance of Dr. Richard Lockhart.

My colleagues at the B.C. Cancer Agency and the Bone Marrow Transplantion Program, particularly Dr. Gordon Phillips, provided me with essential support in the pursuit of this goal, and for this I thank them.

I especially would like to thank my supervisor, Dr. Michael Stephens. It was an undergraduate course from Professor Stephens which led me to study statistics. As a teacher, he is able to make the most difficult concept understandable, and to make the learning of it an enjoyable experience. Without his care and direction, I am certain that this thesis would not have been completed. I am extremely fortunate to have been his student.

Finally, and most of all, I thank my wife, Monica, for her eternal patience, support and assistance, and most importantly, her love.

Dedication

To Morgan

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In order to model a phenomenon, a researcher will often want to determine if a set of data follows a specific probability distribution. This is done by utilizing a procedure to compare the sample data to the hypothesized distribution. The study of the procedures to perform this comparison is known as *goodness-of-fit.*

The usual situation is to obtain a sample, $y_1, y_2, \ldots, y_n$, of independent and identically distributed observations from a distribution with cumulative distribution function, $F(y)$. Usually, the distribution function contains unknown parameters. Another situation arises when each sample value, y_i, comes from a distribution, $F(y_i, \theta_i)$, with parameter θ_i (or vector of parameters $\boldsymbol{\theta}_i$) which is different for each y_i. The parameters, usually unknown, are related to the observations by an underlying model.

Goodness-of-fit techniques are generally used for one of two reasons. The first reason is to justify the application of specific estimation or hypothesis testing procedures. This rationale has become less important in recent years with the development of robust statistical procedures. The other reason, which remains extremely important, is for prediction of future observations. Although the parameter values may be robustly estimated, information about the extreme tail percentiles for environmental assessment or confidence intervals for medical prediction requires knowledge of the correct probability distribution.

A powerful set of goodness-of-fit procedures, at least for continuous distributions, is based on the empirical distribution function (EDF). The EDF, $F_n(y)$ is defined as:

$$F_n(y) = \frac{\text{the number of observed values } \leq y}{n}$$

The fit is judged by the degree of closeness between the EDF and the cumulative distribution function, $F(y)$. Tests of this type are referred to as EDF tests of fit.

The importance of the EDF to statistics is through the Glivenko-Cantelli Theorem (see e.g. Shorack and Wellner, 1986) which states that

$$\sup_{-\infty<y<\infty} \mid F_n(y) - F(y) \mid \longrightarrow 0$$

almost surely as $n \rightarrow \infty$. The first EDF statistic was proposed by Kolmogorov (1933), and is defined by

$$D = \sup_y \mid F_n(y) - F(y) \mid .$$

Since then, many other similar statistics have been proposed, for example, by Smirnov, (1939) and Kuiper (1960). Another family of EDF statistics for continuous distributions are the Cramér-von Mises statistics first proposed by Cramér (1928). These statistics have been found to be more powerful than test statistics based on the supremum, especially at detecting deviations in the tail of the distribution. (see e.g. Stephens, 1986). The general form of the Cramér-von Mises statistics is

$$Q^2 = n \int_{-\infty}^{\infty} [F_n(y) - F(y)]^2 \psi(y) dF(y)$$

where $\psi(y)$ is a weight function. When the weight function is the identity, the statistic is the Cramér-von Mises statistic, W^2; when the $\psi(y) = \{F(y)[1 - F(y)]\}^{-1}$, the statistic is the Anderson-Darling statistic, A^2.

The classical statistic for examining goodness-of-fit for discrete distributions is the χ^2 test proposed by Pearson (1900). However, with continuous distributions and infinite valued discrete distributions the statistic requires grouping of the data which causes a loss of information and thus a loss of power. Even for finite valued discrete distributions, more powerful tests may exist. Choulakian, Lockhart and Stephens (1994) have discussed Cramér-von Mises statistics for discrete distributions, and have set forth the general definitions and asymptotic theory.

In this thesis, Cramér-von Mises statistics are adapted to give tests for the Poisson and the binomial distributions. Also, the asymptotic and finite sample properties of these tests are examined. Finally, the Cramér-von Mises tests of fit are developed for the important case of independent but not identically distributed discrete variables. Throughout this thesis such variables will be referred to as i.n.i.d.

Chapter 2

Poisson Distribution

2.1 Introduction

In this chapter, the Cramér-von Mises statistics are developed as tests for the Poisson distribution. In section 2.2, the definitions of the Cramér-von Mises statistics, W^2, U^2 and A^2, are given, and the basic theory is presented in section 2.3. In section 2.4, the percentage points to make tests for the Poisson distribution are given for the cases where the mean, μ, is known and also for the case where μ is estimated by $\bar{x}$. In section 2.5 power studies are presented. Comparisons are made with the well known dispersion test which is found to be powerful, as expected, against distributions with larger variance. In many other cases, A^2 is found to have good power and is recommended for use as an omnibus test for the Poisson distribution.

2.2 Definitions

2.2.1 Known Mean

Let p_j be the Poisson probability of observing a count j, defined by

$$p_j = \frac{\mu^j e^{-\mu}}{j!}$$

where the mean, μ, is known. Suppose N independent observations are given; let o_j be the observed number of outcomes j, and let $Np_j = e_j$ be the expected number in cell j. Let $S_j = \sum_{i=0}^{j} o_i$ and $T_j = \sum_{i=0}^{j} e_i$. Then S_j/N and $H_j = T_j/N$ give, respectively, the *cumulative* *observed histogram* and the *cumulative expected histogram* of the data, corresponding to the empirical distribution function $F_N(x)$ and the cumulative distribution function, $F(x)$, in the continuous case. Suppose $Z_j = S_j - T_j$, $j = 1, 2, \ldots$. The Cramér-von Mises statistics W^2, U^2 and A^2 for the Poisson distribution (and any other discrete distribution with infinite support) are then defined by

$$W^2 = N^{-1}\textstyle\sum_{j=0}^{\infty} Z_j^2 p_j, \tag{2.1}$$

$$U^2 = N^{-1}\textstyle\sum_{j=0}^{\infty} (Z_j - \bar{Z})^2 p_j, \tag{2.2}$$

$$A^2 = N^{-1}\textstyle\sum_{j=0}^{\infty} Z_j^2 p_j / \{H_j(1 - H_j)\}, \tag{2.3}$$

where $\bar{Z} = \sum_{j=0}^{\infty} Z_j p_j$.

It is convenient to put these expressions into matrix notation. Note that in the following discussion, all vectors and matrices are infinite dimensional. Let a prime, for example $\mathbf{Z}'$, denote the transpose of a vector or matrix. Let $\mathbf{Z}$ be the vector with jth element Z_j, $\mathbf{I}$ be the identity matrix, and $\mathbf{p}'$ be the vector $(p_1, p_2, \ldots)$. Suppose $\mathbf{D}$ is the diagonal matrix whose j-th diagonal entry is p_j, and let $\mathbf{G}$ be the diagonal matrix whose j-th diagonal element is $H_j(1 - H_j)$. Then

$$W^2 = \mathbf{Z}'\mathbf{D}\mathbf{Z}/N;$$

$$U^2 = \mathbf{Z}'(\mathbf{I} - \mathbf{D}\mathbf{1}\mathbf{1}')\mathbf{D}(\mathbf{I} - \mathbf{1}\mathbf{1}'\mathbf{D})\mathbf{Z}/N;$$

$$A^2 = \mathbf{Z}'\mathbf{D}\mathbf{G}^{-1}\mathbf{Z}/N.$$

The S_j and T_j may be defined in terms of the o_j and e_j. Arrange these quantities into column vectors $\mathbf{S}$, $\mathbf{T}$, $\mathbf{o}$, $\mathbf{e}$ (so that, for example, the j-th component of $\mathbf{S}$ is S_j). Then $\mathbf{Z} = \mathbf{S} - \mathbf{T} = \mathbf{A}\mathbf{d}$ where $\mathbf{d} = \mathbf{o} - \mathbf{e}$ and $\mathbf{A}$ is the partial-sum matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{2.4}$$

In matrix notation, the Pearson χ^2 statistic, defined by $\sum_{i=1}^{K}(o_i - e_i)^2/e_i$ is

$$\chi_P^2 = \mathbf{d}'\mathbf{D}^{-1}\mathbf{d}/N.$$

Since the Poisson distribution has infinite support, the distribution must be divided into K cells to calculate the Pearson χ^2 statistic.

2.2.2 Modified Cramér-von Mises Statistics

The above definitions of the Cramér-von Mises statistics W^2, U^2 and A^2 are chosen to be analogous to the corresponding statistics for testing for continuous distributions. However, various modifications are possible; for example, greater weight may be given to certain parts of the tested distribution (see, for example, de Wet and Venter, 1973). If the p_j are omitted (or equivalently the classical Cramér-von Mises statistics are weighted by the inverse of p_j) in definitions (2.1) - (2.3), greater weight will be given to deviations in the tails, and the new statistics, now called W_m^2 and A_m^2, may give better power against longer-tailed alternative distributions. Then

$$W_m^2 = N^{-1}\sum_{j=0}^{\infty} Z_j^2, \tag{2.5}$$

$$A_m^2 = N^{-1}\sum_{j=0}^{\infty} Z_j^2/\{H_j(1-H_j)\}. \tag{2.6}$$

In matrix form the statistics can be written

$$W_m^2 = \mathbf{Z}'\mathbf{Z}/N$$

$$A_m^2 = \mathbf{Z}'\mathbf{G}^{-1}\mathbf{Z}/N$$

where $\mathbf{G}$ is the diagonal matrix whose j-th diagonal element is $H_j(1-H_j)$.

2.2.3 Estimated Mean

To test for the Poisson distribution with unknown mean, μ, estimate μ by maximum likelihood, that is, $\hat{\mu} = \bar{x}$. The Cramér-von Mises statistics will again be calculated from (2.1) - (2.3) and (2.5) - (2.6), but using $\hat{Z}_j$, $\hat{p}_j$ and $\hat{H}_j$. For example,

$$\hat{p}_j = \frac{\hat{\mu}^j e^{-\hat{\mu}}}{j!},$$

and $\hat{\mathbf{p}}$ is the vector of $\hat{p}_j$ values.

2.3 Distribution Theory

2.3.1 Known mean

For any finite valued discrete distribution, if the null hypothesis is true, the vector $\mathbf{o}$ has a multinomial distribution with parameter $\mathbf{p}$; the mean vector and covariance matrix are $N\mathbf{p}$ and $N(\mathbf{D} - \mathbf{pp}')$ respectively. Thus, under the null hypothesis, $\mathbf{d}/\sqrt{N}$ converges in distribution to a multivariate normal random variable with mean zero and covariance matrix, $\mathbf{\Sigma}_0 = \mathbf{D} - \mathbf{pp}'$, as $N \to \infty$, by application of the central limit theorem. Furthermore, the random variable $\mathbf{Z}/\sqrt{N} = \mathbf{Ad}/\sqrt{N}$ converges in distribution to a random variable with an asymptotic multivariate normal distribution with mean zero and covariance matrix, $\mathbf{\Sigma} = \mathbf{A\Sigma}_0\mathbf{A}'$ with i,jth element $\sigma_{ij} = \min\{H_i, H_j\} - H_iH_j$.

If a random variable, $\mathbf{X}$, has an asymptotic multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}$, and $Y = \mathbf{XQX}$ for positive definite symmetric matrix $\mathbf{Q}$, then Y can be written

$$Y = \textstyle\sum_{i=1}^{K} \lambda_i Z_i^2, \tag{2.7}$$

where Z_i are independent standard normal random variables and λ_i are the eigenvalues of the matrix, $\mathbf{Q}^{1/2}\mathbf{\Sigma}\mathbf{Q}^{1/2}$. All the Cramér-von Mises statistics are of the general form $\mathbf{Z}'\mathbf{MZ}/N$, where $\mathbf{M}$ is positive definite and symmetric. Let $\mathbf{X} = \mathbf{M}^{1/2}\mathbf{Z}/\sqrt{N}$. For an infinite valued discrete distribution, a Cramér-von Mises statistic S has an asymptotic distribution which is a sum of weighted χ^2 random variables with weights equal to the eigenvalues of the matrix, $\mathbf{\Sigma}_X = \mathbf{M}^{1/2}\mathbf{\Sigma}\mathbf{M}^{1/2}$. We can write

$$S = \mathbf{Z}'\mathbf{MZ}/N = \mathbf{X}'\mathbf{X} = \textstyle\sum_{i=1}^{\infty} \lambda_i(\mathbf{w}_i{}'\mathbf{X})^2, \tag{2.8}$$

where the λ_i are the eigenvalues of $\mathbf{\Sigma}_X$, and $\mathbf{w}_i$ are the corresponding eigenvectors, normalized so that $\mathbf{w}_i{}'\mathbf{\Sigma}\mathbf{w}_i = \delta_{ij}$. Here δ_{ij} is Kroneker's δ with $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. In (2.8), the term $s_i = (\mathbf{w}_i{}'\mathbf{X})$ is called the i-th *component* of the statistic. The components, s_i, take different values for different statistics, since they depend on the eigenvectors of $\mathbf{M}$. The normalization of the $\mathbf{w}_i$ makes the variance of each $s_i = 1$. As $N \to \infty$, the s_i have distributions which are independent and each standard normal, and a typical statistic has an asymptotic distribution

$$S = \textstyle\sum_{i=1}^{\infty} \lambda_i s_i^2, \tag{2.9}$$

where the s_i^2 are independent χ_1^2 variables, the same result as shown above.

For practical calculations, an infinite valued discrete distribution, such as the Poisson, must be curtailed after a finite number of terms. It is proposed that this be done at cell K when $p_K < 10^{-3}/N$. Inclusion of further terms does not significantly change the value of the statistic.

2.3.2 Estimated Mean

Let $\boldsymbol{\theta} = [\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2]'$ where $\boldsymbol{\theta}_1$ is a vector of p_1 known parameters and $\boldsymbol{\theta}_2$ is a vector of p_2 parameters estimated from the data, and let $\boldsymbol{\theta}_0$ be the vector of true values of the parameters. Let $\hat{\boldsymbol{\theta}}_2$ be the maximum likelihood estimator of $\boldsymbol{\theta}_2$ and assume suitable regularity conditions which ensure the application of regular maximum likelihood asymptotics (see e.g. Cox and Hinkley, 1974, or Bishop, Fienberg and Holland, 1975). Then, as $N \to \infty$, the variable $\hat{\mathbf{d}}/\sqrt{N} = (\mathbf{o} - \hat{\mathbf{e}})/\sqrt{N}$ converges in distribution to a mean zero multivariate normal random variable with covariance matrix

$$\hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}_0 - \mathbf{B}(\mathbf{B}'\mathbf{D}^{-1}\mathbf{B})\mathbf{B}' \tag{2.10}$$

where $\hat{\mathbf{e}}$ is the vector of expected numbers using the estimated parameters, $\boldsymbol{\Sigma}_0$ is as before, $\mathbf{D}$ is the diagonal matrix whose j-th diagonal entry is p_j and $\mathbf{B}$ is the p_2 by K matrix with i,jth element

$$\frac{\partial p_j}{\partial \theta_i}$$

(Bishop, Fienberg and Holland, 1975). Here, K is the number of cells in the discrete distribution, and θ_i is the ith unknown parameter. When there is only one unknown parameter (2.10) reduces to

$$\hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}_0 - \mathbf{g}\mathbf{g}'/\mathbf{g}'\mathbf{D}^{-1}\mathbf{g},$$

where $\mathbf{g}$ is the vector with jth element,

$$\frac{\partial p_j}{\partial \theta}.$$

For the Poisson distribution, the mean, μ, is replaced by the maximum likelihood estimate, $\bar{x}$ and

$$\begin{aligned} g_j &= (j\mu^{j-1}e^{-\mu} - e^{-\mu}\mu^j)/j! \\ &= p_{j-1} - p_j. \end{aligned}$$

It is easily seen that g_j can also be written

$$g_j = \frac{p_j}{\mu}(j - \mu). \tag{2.11}$$

If follows that

$$\mathbf{g}'\mathbf{D}^{-1}\mathbf{g} = 1/\mu,$$

the inverse of the Poisson variance. Then

$$\hat{\Sigma}_0 = \Sigma_0 - \mu\mathbf{g}\mathbf{g}'. \tag{2.12}$$

Likewise, as $N \to \infty$, $\hat{\mathbf{Z}}/\sqrt{N} = \mathbf{A}\hat{\mathbf{d}}/\sqrt{N}$ converges in distribution to a mean zero multivariate normal random variable with covariance matrix,

$$\begin{aligned} \hat{\Sigma} &= \mathbf{A}\hat{\Sigma}_0\mathbf{A}' \\ &= \mathbf{A}\Sigma_0\mathbf{A}' - \mathbf{A}\mathbf{g}\mathbf{g}'\mathbf{A}'/\mathbf{g}'\mathbf{D}^{-1}\mathbf{g} \\ &= \Sigma - \mu\mathbf{p}\mathbf{p}' \end{aligned} \tag{2.13}$$

since $\mathbf{A}\mathbf{g} = -\mathbf{p}$.

As before, the percentage points for the asymptotic distribution of a typical statistic are determined by finding the eigenvalues λ_i of $\hat{\Sigma}_X = \mathbf{M}^{1/2}\hat{\Sigma}\mathbf{M}^{1/2}$ for the statistic, and using (2.9) and Imhof's method (Imhof, 1961).

2.4 Calculation of Percentage Points

2.4.1 Known Mean

Moments

In order to determine the asymptotic percentage points for each statistic, it is necessary to determine the matrix, $\mathbf{M}$, and compute the eigenvalues of the infinite-dimensional matrix $\Sigma_X = \mathbf{M}^{1/2}\Sigma\mathbf{M}^{1/2}$. From the representation (2.9) the cumulants of the test statistics are given by

$$\kappa_j = 2^{j-1}(j-1)!\textstyle\sum_{i=1}^{\infty}\lambda_i^j.$$

In particular, the mean is $\sum_{i=1}^{\infty}\lambda_i$ and the variance is $2\sum_{i=1}^{\infty}\lambda_i^2$. The mean of each statistic can also be calculated exactly using the multinomial distribution of $\mathbf{o}$, and this calculation can be used to check the accuracy of the eigenvalue decomposition. Recall that $S_j = \sum_{i=0}^{j} o_i$ and $T_j = \sum_{i=0}^{j} e_i$. Then, for statistic W^2

$$\begin{aligned} E(W^2) &= E[\textstyle\sum_{j=0}^{\infty}(S_j - T_j)^2 p_j] \\ &= \textstyle\sum_{j=0}^{\infty} p_j E[(S_j - T_j)^2] \\ &= \textstyle\sum_{j=0}^{\infty} p_j Var[S_j] \\ &= \textstyle\sum_{j=0}^{\infty} p_j H_j(1 - H_j). \end{aligned}$$

The means of the other Cramér-von Mises statistics can be similarly derived and are as follows:

$$\begin{aligned} E(U^2) &= E(W^2) - \textstyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p_i p_j(\min\{H_i, H_j\} - H_i H_j), \\ E(A^2) &= 1, \\ E(W_m^2) &= \textstyle\sum_{j=0}^{\infty} H_j(1 - H_j) \le \mu. \end{aligned}$$

Note that the means of the statistics do not depend on the sample size, N.

The identities given above show that

$$E\left[\sum_{K_n+1}^{\infty} \frac{(S_j - T_j)^2}{\sqrt{N}} p_j w_j\right] \to 0$$

for any sequence $K_n \to \infty$ where w_j are the weights associated with the particular Cramér-von Mises statistic. This can be combined with the well known asymptotic normality of $\mathbf{d}/\sqrt{N}$ for any fixed K, to establish rigorously the asymptotic distributions given in section 2.3, for W^2, U^2, A^2 and W_m^2. See Guttorp and Lockhart (1988) for a similar argument.

The statistic A_m^2, defined in (2.6), however, has infinite expected value (both asymptotic and finite sample) when testing for the Poisson distribution or any other distribution with infinite support; this is because the expectation of each term is one. The expectations of the individual terms of $\mathbf{Z}$ are equal to the variance of the terms of the cumulative observed histogram. Since for A_m^2 the terms are weighted by the inverse of the variance, the expectation is unity. For this reason, the statistic A_m^2 will not be considered further. Note that the Pearson χ^2 statistic also has infinite mean if the data are not categorized into a finite number of cells.

For different values of the known mean μ of the tested Poisson distribution, the means and variances of the asymptotic distributions of the Cramér-von Mises statistics are given in Table 2.1. The means and variances of the statistics when testing for a continuous distribution (Stephens, 1976a, Case 0) are indicated by $\mu = \infty$, and are discussed below.

Percentage Points

The asymptotic distribution of a typical test statistic is an infinite sum of weighted χ_1^2 variables with weights equal to the eigenvalues of the covariance matrix of the statistic. In practice, this sum must be curtailed at a finite number of terms in order to obtain percentage points. This has been done as follows. For various values of the mean, μ, of the Poisson distribution, the λ_i have been calculated for $i = 1, ..., K$ where K has been chosen to make the final values of λ_i sufficiently small that the percentage points do not change with the addition of more eigenvalues. The eigenvalues were found using S-PLUS (S-PLUS, 1991), and the percentage points were then found by Imhof's method.

For W^2, the matrix $\mathbf{M}$ is equal to $\mathbf{D}$, the diagonal matrix with the Poisson probabilities, p_j, on the diagonal. The percentage points for W^2 are recorded in Table 2.2 for selected values of μ.

For the statistics U^2, A^2 and W_m^2, the $\mathbf{M}$ matrices are $(\mathbf{I} - \mathbf{D11'})\mathbf{D}(\mathbf{I} - \mathbf{11'D})$, $\mathbf{DG}^{-1}$ and $\mathbf{I}$, respectively. The percentage points for these statistics have been calculated as above and are recorded in Table 2.2.

Since for the Poisson distribution, μ is neither a location nor a scale parameter, the asymptotic points depend on the parameter μ. There is an interesting connection between these points and the points used for testing a known continuous distribution, given by Stephens (1986), where such a test situation is called Case 0. As $\mu \to \infty$, the percentage points for W^2, U^2 and A^2 tend to those for Case 0. Thus, for large values of μ, the Case 0 points can be used as an approximation for the exact points. However, for smaller values of μ more accurate results will be found by using Table 2.2.

In order to examine the rate of convergence of the percentage points to the asymptotic points, the percentage points of W^2, U^2, A^2 and W_m^2 for known μ have been found by Monte Carlo simulation using 25,000 samples. The standard error of estimation of the level of the pth percentage point is approximately $\sqrt{p(1-p)/n}$ where n is the number of simulations; for the .95 percentage point the standard error is 0.0014.

Percentage points for $\mu = 1, 10$ and various sample sizes, N, are given in Tables 2.3 and 2.4. These points converge rapidly to the asymptotic points, which can be used for samples of size greater than 10. The rapid convergence of points for finite samples to the asymptotic values also occurs for the Cramér-von Mises tests for continuous distributions. For example, suppose the asymptotic point 2.783 is used for a 5% test for A^2 with known mean, $\mu = 1$

when N=10. If the given percentage point for N=10, 2.736, is correct, the actual α level obtained by the test would be $\alpha' = .049$.

An S function (Becker, Chambers and Wilks, 1988; S-PLUS, 1991) has been written to compute the Cramér-von Mises statistics and their asymptotic p-values.

Table 2.1: Asymptotic mean and variance for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ, for selected values of μ.

	W^2		U^2		A^2		W_m^2	
μ	Mean	Var	Mean	Var	Mean	Var	Mean	Var
.1	.078	.0122	.0071	.0001	1.000	1.668	.091	.015
.5	.171	.0454	.0538	.0048	1.000	1.062	.337	.141
1	.172	.0316	.0718	.0053	1.000	.827	.524	.253
1.5	.169	.0275	.0760	.0043	1.000	.738	.660	.350
2	.168	.0262	.0778	.0040	1.000	.694	.772	.448
5	.167	.0237	.0812	.0033	1.000	.623	1.246	1.047
10	.167	.0230	.0823	.0030	1.000	.601	1.773	2.050
20	.167	.0226	.0828	.0029	1.000	.590	2.515	4.056
50	.167	.0225	.0831	.0028	1.000	.584	3.984	10.074
100	.167	.0222	.0832	.0028	1.000	.582	5.638	20.106
∞	.1667	.0222	.0832	.0028	1.000	.580		

Table 2.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ, for selected values of μ.

W^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.104	.164	.209	.300	.338	.518	.614	.841
.5	.222	.333	.427	.597	.769	1.013	1.199	1.435
1	.228	.317	.391	.523	.660	.848	.993	1.339
2	.216	.297	.365	.486	.614	.788	.923	1.243
5	.212	.289	.354	.471	.593	.760	.889	1.196
10	.211	.286	.350	.466	.587	.751	.879	1.181
20	.210	.280	.349	.464	.584	.747	.874	1.175
50	.211	.285	.348	.462	.582	.744	.871	1.171
100	.210	.284	.348	.462	.581	.744	.870	1.169
∞	.209	.284	.347	.461	.581	.743	.869	1.167

U^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.009	.015	.019	.027	.031	.047	.056	.076
.5	.070	.106	.137	.193	.250	.329	.389	.533
1	.097	.133	.163	.215	.269	.342	.399	.533
2	.104	.135	.159	.201	.243	.299	.340	.441
5	.106	.133	.155	.193	.230	.280	.317	.405
10	.106	.132	.154	.190	.226	.274	.310	.394
20	.105	.132	.153	.188	.224	.271	.307	.390
50	.105	.131	.152	.187	.223	.269	.305	.387
100	.105	.131	.152	.187	.222	.269	.304	.387
∞	.105	.131	.152	.187	.222	.268	.304	.385

Table 2.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ, for selected values of μ. (continued)

A^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	1.303	1.982	2.557	3.589	4.664	6.128	7.260	8.436
.5	1.320	1.836	2.265	3.035	3.838	4.936	5.787	7.805
1	1.311	1.758	2.124	2.783	3.467	4.379	5.127	6.849
2	1.284	1.688	2.025	2.627	3.257	4.123	4.787	6.376
5	1.262	1.647	1.968	2.544	3.146	3.970	4.609	6.124
10	1.255	1.634	1.950	2.518	3.111	3.924	4.548	6.046
20	1.252	1.628	1.942	2.505	3.094	3.901	4.525	6.008
50	1.249	1.624	1.936	2.497	3.084	3.887	4.511	5.985
100	1.249	1.623	1.935	2.495	3.081	3.882	4.391	5.977
∞	1.248	1.610	1.933	2.492	3.070	3.857	4.500	6.000

W_m^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.119	.184	.238	.336	.439	.587	.684	.767
.5	.442	.632	.792	1.087	1.378	1.797	2.115	2.769
1	.689	.939	1.146	1.513	1.896	2.421	2.825	3.789
2	.993	1.320	1.593	2.081	2.593	3.266	3.836	5.124
5	1.575	2.077	2.497	3.252	4.042	5.126	5.960	7.943
10	2.229	2.932	3.522	4.582	5.691	7.209	8.382	11.171
20	3.153	4.143	4.974	6.469	8.032	10.171	11.826	15.752
50	4.985	6.549	7.860	10.218	12.685	16.059	18.680	24.868
100	7.051	9.260	11.113	14.445	17.932	22.702	26.403	35.151

Table 2.3: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ=1, for selected sample sizes. The asymptotic points are shown for comparison.

$\mu = 1$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.234	.296	.366	.478	.586	.870
10	.226	.310	.368	.514	.675	.802
15	.239	.321	.393	.531	.626	.811
20	.231	.322	.383	.497	.638	.817
40	.228	.316	.389	.515	.656	.824
50	.230	.313	.387	.521	.655	.840
100	.230	.318	.394	.519	.654	.844
∞	.228	.317	.391	.523	.660	.848

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.106	.146	.167	.182	.233	.282
10	.102	.130	.164	.213	.256	.346
15	.095	.138	.161	.211	.264	.322
20	.098	.132	.162	.209	.267	.327
40	.097	.132	.162	.214	.264	.344
50	.094	.131	.164	.214	.262	.332
100	.099	.134	.162	.213	.267	.338
∞	.097	.133	.163	.215	.269	.342

Table 2.3: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ=1, for selected sample sizes. The asymptotic points are shown for comparison. (continued)

$\mu = 1$

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	1.294	1.701	2.224	2.637	3.908	4.751
10	1.360	1.747	2.191	2.736	3.608	4.551
15	1.365	1.745	2.054	2.732	3.486	4.531
20	1.289	1.758	2.126	2.796	3.574	4.582
40	1.309	1.761	2.115	2.792	3.460	4.448
50	1.294	1.743	2.108	2.769	3.485	4.436
100	1.319	1.770	2.128	2.786	3.448	4.396
∞	1.311	1.758	2.124	2.783	3.467	4.379

W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.731	.940	1.217	1.316	1.915	2.381
10	.716	.909	1.151	1.442	1.913	2.488
15	.704	.919	1.117	1.497	1.933	2.371
20	.696	.950	1.129	1.475	1.878	2.488
40	.690	.926	1.137	1.512	1.900	2.393
50	.684	.923	1.136	1.515	1.889	2.426
100	.694	.944	1.161	1.531	1.905	2.428
∞	.689	.939	1.146	1.513	1.896	2.421

Table 2.4: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ=10, for selected sample sizes. The asymptotic points are shown for comparison.

$\mu = 10$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.215	.289	.348	.459	.565	.696
10	.213	.286	.347	.457	.566	.729
15	.211	.286	.346	.458	.580	.750
20	.211	.288	.351	.460	.576	.722
40	.211	.287	.350	.464	.580	.744
50	.211	.284	.348	.469	.586	.733
100	.210	.286	.351	.462	.584	.743
∞	.211	.286	.350	.466	.587	.751

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.106	.132	.151	.183	.210	.247
10	.105	.132	.152	.186	.218	.264
15	.105	.131	.152	.187	.224	.273
20	.105	.132	.152	.188	.224	.266
40	.105	.132	.152	.188	.222	.267
50	.106	.131	.152	.187	.223	.269
100	.106	.132	.153	.188	.220	.266
∞	.106	.132	.154	.190	.226	.274

Table 2.4: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with known mean μ=10, for selected sample sizes. The asymptotic points are shown for comparison. (continued)

$\mu = 10$

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	1.262	1.653	1.984	2.593	3.233	4.101
10	1.253	1.631	1.954	2.526	3.163	3.990
15	1.255	1.630	1.952	2.557	3.192	4.026
20	1.251	1.644	1.960	2.520	3.107	3.873
40	1.257	1.647	1.942	2.513	3.090	3.956
50	1.254	1.624	1.946	2.536	3.130	3.918
100	1.244	1.632	1.950	2.507	3.132	3.928
∞	1.255	1.634	1.950	2.518	3.111	3.924

W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	2.255	2.942	3.523	4.580	5.737	7.217
10	2.246	2.945	3.532	4.572	5.701	7.122
15	2.230	2.918	3.495	4.594	5.692	7.284
20	2.221	2.948	3.527	4.567	5.639	6.998
40	2.230	2.952	3.505	4.584	5.604	7.244
50	2.224	2.905	3.499	4.608	5.687	7.122
100	2.234	2.986	3.589	4.712	5.761	7.257
∞	2.229	2.932	3.522	4.582	5.691	7.209

2.4.2 Estimated Mean

Moments

The asymptotic percentage points for the various statistics, for testing for the Poisson distribution with mean estimated by $\bar{x}$, may be calculated as for the case where the mean μ is known, except that the matrix Σ is replaced by $\hat{\Sigma}$. It is now necessary to compute the eigenvalues of the matrix $\hat{\Sigma}_X = \mathbf{M}^{1/2}\hat{\Sigma}\mathbf{M}^{1/2}$. The $\mathbf{M}$ matrices for the statistics are those defined in section 2.4.1.

Once again the asymptotic means and variances of the statistics can be determined from the representation (2.9). The asymptotic means can also be written explicitly using (2.13). The means of W^2, A^2 and W_m^2 for testing for the Poisson distribution with estimated mean are as follows:

$$\begin{aligned} E(W^2) &= \textstyle\sum_{j=0}^{\infty} p_j H_j(1 - H_j) - \mu \sum_{j=0}^{\infty} p_j^3, \\ E(A^2) &= 1 - \mu \textstyle\sum_{j=0}^{\infty} p_j^3 / \{H_j(1 - H_j)\}, \\ E(W_m^2) &= \textstyle\sum_{j=0}^{\infty} H_j(1 - H_j) - \mu \sum_{j=0}^{\infty} p_j^2. \end{aligned}$$

Once again the means of the statistics do not depend on the sample size, N. Unlike the case where the mean, μ, is known, the expected value of A^2 is no longer identically one. The mean and variance of the asymptotic distributions of the statistics are found in Table 2.5. For comparison, the means and variances of the statistics when testing for a normal distribution with known variance and unknown mean (Stephens, 1976a, Case 1) are indicated by $\mu = \infty$, and are discussed below.

Percentage Points

The percentage points for the Cramér-von Mises statistics are recorded in Table 2.6 for selected values of μ. As μ tends to infinity the points for W^2, U^2 and A^2 tend to the points for testing for a normal distribution with known variance but unknown mean (Case 1), given by Stephens (1986). The Case 1 points could be used as an approximation for the exact points for large values of μ; however, for smaller μ more accurate results will be found in Table 2.6.

For finite N, the percentage points of W^2, U^2, A^2 and W_m^2 for estimated μ have been found by Monte Carlo simulations using 25,000 samples. Percentage points for $\mu = 1, 10$ and various sample sizes, N, are found in Tables 2.7 and 2.8. These points also converge rapidly to the asymptotic points; these can therefore be used for samples of size greater than 10.

Table 2.5: Asymptotic mean and variance for Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ, for selected values of μ.

	W^2		U^2		A^2		W_m^2	
μ	Mean	Var	Mean	Var	Mean	Var	Mean	Var
.1	.0042	.0000	.0013	.0000	.124	.028	.082	.001
.5	.0450	.0035	.0365	.0024	.347	.146	.104	.015
1	.0658	.0048	.0609	.0045	.432	.148	.215	.043
2	.0700	.0033	.0659	.0030	.480	.120	.357	.072
5	.0727	.0026	.0688	.0024	.504	.099	.606	.152
10	.0738	.0024	.0699	.0022	.512	.092	.875	.284
20	.0743	.0023	.0705	.0021	.516	.089	1.250	.549
50	.0745	.0022	.0708	.0020	.518	.087	1.987	1.342
100	.0747	.0022	.0709	.0020	.519	.086	2.816	2.665
∞	.0748	.0021	.0710	.0020	.519			

Table 2.6: Asymptotic percentage points for Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ, for selected values of μ.

W^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.006	.009	.011	.016	.018	.028	.033	.045
.5	.059	.090	.116	.164	.213	.280	.332	.455
1	.088	.123	.152	.203	.257	.330	.389	.556
2	.093	.121	.144	.182	.221	.273	.315	.408
5	.094	.119	.139	.172	.206	.251	.285	.366
10	.094	.118	.137	.169	.201	.244	.277	.354
20	.094	.117	.135	.167	.199	.241	.273	.349
50	.094	.117	.135	.166	.199	.239	.271	.346
100	.094	.117	.135	.166	.197	.239	.270	.345
∞	.094	.117	.134	.165	.197	.238	.270	.345

U^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.002	.003	.004	.005	.006	.009	.010	.014
.5	.048	.074	.096	.136	.177	.231	.276	.378
1	.081	.115	.143	.194	.246	.319	.375	.507
2	.088	.115	.136	.173	.211	.262	.302	.395
5	.089	.113	.132	.164	.196	.240	.274	.353
10	.089	.112	.130	.160	.192	.234	.266	.343
20	.089	.111	.128	.159	.189	.231	.263	.338
50	.089	.111	.128	.158	.188	.229	.261	.335
100	.089	.110	.128	.157	.188	.229	.260	.335
∞	.088	.110	.127	.157	.187	.228	.259	.334

Table 2.6: Asymptotic percentage points for Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ, for selected values of μ. (continued)

A^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.162	.251	.325	.460	.601	.784	.937	1.286
.5	.456	.649	.811	1.104	1.414	1.828	2.151	2.849
1	.577	.769	.921	1.191	1.465	1.812	2.119	2.762
2	.630	.796	.927	1.151	1.377	1.681	1.913	2.465
5	.640	.790	.908	1.112	1.319	1.598	1.813	2.322
10	.641	.786	.900	1.099	1.301	1.573	1.783	2.281
20	.641	.783	.897	1.093	1.292	1.562	1.769	2.262
50	.641	.782	.894	1.089	1.287	1.555	1.761	2.249
100	.641	.782	.894	1.088	1.286	1.553	1.758	2.245
∞	.644	.782	.894	1.087	1.285	1.551	1.756	2.241

W_m^2	Upper tail significance level α							
μ	.25	.15	.10	.05	.025	.01	.005	.001
.1	.011	.017	.022	.031	.040	.053	.063	.069
.5	.136	.199	.253	.350	.439	.590	.697	.857
1	.287	.391	.475	.624	.778	.988	1.158	1.536
2	.472	.602	.705	.881	1.058	1.295	1.472	1.908
5	.773	.959	1.106	1.359	1.616	1.961	2.227	2.854
10	1.100	1.355	1.557	1.906	2.261	2.738	3.106	3.975
20	1.560	1.914	2.196	2.685	3.181	3.849	4.363	5.581
50	2.468	3.024	3.468	4.235	5.014	6.064	6.872	8.778
100	3.492	4.276	4.902	5.984	7.084	8.566	9.706	12.412

Table 2.7: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ=1, for selected sample sizes. The asymptotic points are shown for comparison.

$\mu = 1$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.086	.110	.146	.207	.207	.237
10	.084	.128	.135	.192	.240	.281
15	.087	.121	.154	.196	.251	.306
20	.086	.121	.148	.200	.258	.324
40	.087	.121	.150	.200	.252	.324
50	.088	.123	.153	.206	.258	.332
100	.089	.125	.154	.206	.261	.336
∞	.088	.123	.152	.203	.257	.330

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.078	.110	.145	.196	.200	.230
10	.079	.120	.129	.189	.233	.264
15	.083	.110	.151	.187	.243	.293
20	.081	.113	.137	.192	.245	.311
40	.080	.112	.141	.191	.242	.312
50	.082	.116	.143	.194	.245	.317
100	.082	.116	.145	.196	.250	.324
∞	.081	.115	.143	.194	.246	.319

Table 2.7: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ=1, for selected sample sizes. The asymptotic points are shown for comparison. (continued)

$\mu = 1$

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.554	.702	.792	1.027	1.184	1.818
10	.574	.697	.884	1.155	1.418	1.644
15	.557	.751	.894	1.175	1.458	1.775
20	.567	.749	.914	1.150	1.430	1.862
40	.564	.748	.901	1.165	1.454	1.840
50	.578	.770	.924	1.191	1.453	1.821
100	.581	.772	.932	1.200	1.486	1.871
∞	.577	.769	.921	1.191	1.465	1.812

W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.247	.435	.505	.551	.636	1.006
10	.292	.386	.464	.631	.750	.997
15	.287	.408	.478	.604	.768	.946
20	.287	.386	.472	.620	.777	.970
40	.281	.388	.463	.606	.770	.971
50	.288	.388	.470	.622	.775	1.008
100	.286	.390	.469	.621	.776	.987
∞	.287	.391	.475	.624	.778	.988

Table 2.8: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ=10, for selected sample sizes. The asymptotic points are shown for comparison.

$\mu = 10$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.095	.119	.138	.165	.195	.227
10	.095	.118	.137	.168	.197	.234
15	.094	.117	.135	.166	.197	.237
20	.094	.118	.137	.169	.199	.240
40	.095	.118	.137	.168	.198	.239
50	.095	.118	.137	.169	.200	.241
100	.094	.118	.136	.169	.202	.245
∞	.094	.118	.137	.169	.201	.244

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.092	.114	.132	.160	.190	.223
10	.091	.113	.131	.162	.190	.226
15	.090	.112	.129	.159	.189	.229
20	.090	.112	.130	.161	.190	.232
40	.090	.112	.131	.160	.189	.229
50	.090	.112	.130	.160	.191	.233
100	.089	.112	.129	.160	.193	.234
∞	.089	.112	.130	.160	.192	.234

Table 2.8: Monte Carlo percentage points for the Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean μ=10. for selected sample sizes. The asymptotic points are shown for comparison. (continued)

$\mu = 10$

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	.603	.744	.845	1.041	1.227	1.525
10	.623	.764	.878	1.063	1.271	1.556
15	.625	.763	.876	1.068	1.261	1.532
20	.634	.777	.891	1.085	1.298	1.586
40	.641	.788	.903	1.095	1.294	1.571
50	.641	.783	.894	1.100	1.296	1.579
100	.641	.786	.902	1.101	1.314	1.570
∞	.641	.786	.900	1.099	1.301	1.573

W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	1.065	1.308	1.492	1.813	2.123	2.547
10	1.081	1.322	1.511	1.841	2.177	2.623
15	1.082	1.324	1.520	1.862	2.184	2.628
20	1.100	1.348	1.543	1.890	2.223	2.721
40	1.104	1.358	1.564	1.891	2.244	2.703
50	1.097	1.354	1.553	1.888	2.232	2.710
100	1.100	1.351	1.556	1.913	2.289	2.757
∞	1.100	1.355	1.557	1.906	2.261	2.738

2.5 Power Comparisons

We now consider the power of the Cramér-von Mises statistics for testing for the Poisson distribution in the more common situation when the mean is estimated from the sample. Where possible. asymptotic power calculations for the Cramér-von Mises statistics and other tests of fit have been made. These have been supplemented by simulations to determine the relative powers for finite samples.

2.5.1 The Test Statistics

The test statistics compared are the following:

1. The Cramér-von Mises statistics defined in Section 2.2.

2. The dispersion test. This is the most commonly used goodness-of-fit test for the Poisson distribution, and was suggested by R. A. Fisher (Kendall and Stuart, Volume 2, 1973). It is defined as follows:

$$D = \frac{\sum_{j=1}^{N}(x_j - \bar{x})^2}{\bar{x}}. \tag{2.14}$$

   This test is often used as a one sided test to detect overdispersed alternatives, however, is used here as a two-sided test to guard against all alternative distributions. Potthoff and Whittinghill (1966b) show that a test based on D is the score (locally most powerful) test against the negative binomial distribution.

3. The k-component smooth test. This is an analogue of the Neyman smooth test for continuous distributions (Neyman, 1937). Such analogues were examined first by Scott (1949) and later by Barton (1955). More recently, they have been developed for testing for the Poisson distribution by Rayner and Best (1989). These tests are constructed to have optimal power against local alternatives whose distributions depart smoothly from the distribution being tested. The alternatives are functions of polynomials orthonormal to the distribution under test. For the Poisson distribution, the orthonormal functions are Poisson-Charlier polynomials, $h_i(j;\mu)$. The ith polynomial is defined as follows:

$$h_i(j;\mu) = \sqrt{\mu^i/i!}\sum_{t=0}^{i}(-1)^{i-t}C_t^j \mu^{-t} C_t^i \tag{2.15}$$

where C_x^n is the binomial coefficient for x successes in n trials.

The test statistic is then defined as

$$\hat{S}_k = N^{-1} \sum_{i=2}^{k+1} V_i^2 \tag{2.16}$$

where $V_i = \sum_{j=1}^{N} h_i(x_j, \bar{x})$. The one-component statistic $\hat{S}_1 = V_2^2$ is a standardized version of the dispersion test, $\hat{S}_1 = (D - N)^2/2N$. The four-component smooth test was recommended by Rayner and Best.

4. Statistics based on the probability generating function (PGF). These have been proposed by several authors (Kocherlakota and Kocherlakota, 1986; Rueda et al, 1991; Nakamura and Perez-Abreu, 1993). Two statistics were examined, called P and T below. They are found as follows.

   Let $\phi(t)$ be the PGF and $\phi_n(t)$ be the empirical probability generating function.

   (a) Rueda et al (1991) proposed the following test statistic:

   $$P = \int_0^1 (\phi_n(t) - \phi(t))^2 dt.$$

   This statistic is an extension of the statistic proposed by Kocherlakota and Kocherlakota (1986); their statistic was the difference between the empirical probability generating function and the PGF at a specific point, t.

   (b) For the Poisson distribution, $\log \phi(t) = \mu(t - 1)$. Nakamura and Perez-Abreu propose a statistic based on the the departure of $\log \phi_n(t)$ from a straight line, using the value of the second derivative. The statistic is referred to as T.

5. Correlation and regression tests of fit. These have been proposed and evaluated for several distributions (Spinelli, 1980). The tests compare the sample order statistics with their expected value or some other asymptotically equivalent value. Let $r(\mathbf{x}, \mathbf{y})$ denote the correlation between vectors $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{m}$ be the vector of expectations of Poisson order statistics. The correlation statistic is defined as

   $$R = N[1 - r^2(\mathbf{x}, \mathbf{m})], \tag{2.17}$$

   where $\mathbf{x}$ is the vector of sample order statistics.

6. The Pearson χ^2 statistic defined in section 2.4.1. As pointed out above, the Poisson distribution must be categorized in order for the Pearson χ^2 test to be used. Since the cell probabilities, and thus the categorization, depend on the value of the mean, μ, a single a priori categorization procedure is impossible to develop. Two categorization procedures were examined for each of $\mu = 1, 10$.

   For $\mu = 1$

   (a) X_1^2 - Three groups $k = 0, 1, 2+$

   (b) X_2^2 - Five groups $k = 0, 1, 2, 3, 4+$

   For $\mu = 10$

   (a) X_1^2 - Five groups $k = 0-6, 7-8, 9-10, 11-12, 13+$

   (b) X_2^2 - Ten groups $k = 0-5, 6, 7, 8, 9, 10, 11, 12, 13, 14+$.

2.5.2 Asymptotic Power

The asymptotic powers of the Cramér-von Mises statistics and the smooth statistics, S_k, with $k = 1, 2, 3, 4$, were examined against the negative binomial alternative. For the purposes of the asymptotic power comparison let the negative binomial distribution be defined as follows:

$$Pr\{Y = y\} = \frac{\Gamma(y+\gamma^{-1})}{y!\Gamma(\gamma^{-1})}\left(\frac{\gamma\mu}{1+\gamma\mu}\right)^y\left(\frac{1}{1+\gamma\mu}\right)^{\gamma^{-1}}, \tag{2.18}$$

for $y = 0, 1, \ldots$ and $\gamma, \mu > 0$. The mean and variance of Y are μ and $\mu(1+\gamma\mu)$, respectively. At $\gamma = 0$, (2.18) reduces to the Poisson distribution. Thus H_0 is: $\gamma = 0$. Let μ be estimated by maximum likelihood, that is, $\hat{\mu} = \bar{x}$. Under H_1, let $\gamma = \delta/\sqrt{N}$, thus $\gamma \to 0$ as $N \to \infty$ and H_1 reduces to H_0.

In section 2.2.1, the i-th component of a test statistic, $s_i = (\mathbf{w}_i{}'\mathbf{X})$, was defined. Under H_0, the s_i^2 are independent and each distributed standard normal. Under H_1, the s_i are independent and normally distributed with variance 1, but with mean ν_i which is not zero. For the k-component smooth statistics the mean ν_i is $\delta\mathbf{w}_i{}'\mathbf{g}$ where $\mathbf{g}$ is the vector with jth element

$$\begin{aligned} g_j &= \left.\frac{\partial Pr\{Y = j\}}{\partial\gamma}\right|_{\gamma=0} \\ &= p_{j-1} - p_j \end{aligned} \tag{2.19}$$

and where p_j is the Poisson probability of observing a count j defined in section 2.2.1 above. For the Cramér-von Mises statistics and W_m^2, the mean of s_i is

$$\delta \mathbf{w}_i{}'\mathbf{M}^{1/2}\mathbf{A}\mathbf{g} = -\delta \mathbf{w}_i{}'\mathbf{M}^{1/2}\mathbf{p} \tag{2.20}$$

where $\mathbf{A}$ is the partial-sum matrix given in (2.4) and $\mathbf{p}$ is the vector of Poisson probabilities.

Equations (2.19) and (2.20) can be derived as follows. Let $\mathbf{p}_0$ and $\mathbf{p}_1$ be the vectors of cell probabilities under the null and alternative hypotheses, respectively. Also let $\mathbf{e}_0$ and $\mathbf{e}_1$ be the vectors of expected numbers in each cell. Then under the alternative hypothesis:

$$\begin{aligned} E(\mathbf{d}/\sqrt{N}) &= E(\mathbf{o} - \mathbf{e}_0)/\sqrt{N} \\ &= [E(\mathbf{o} - \mathbf{e}_1) + (\mathbf{e}_1 - \mathbf{e}_0)]/\sqrt{N} \\ &= \sqrt{N}(\mathbf{p}_1 - \mathbf{p}_0). \end{aligned}$$

The vector $\mathbf{p}_1$ is a function of the parameter γ. A Taylor expansion around $\gamma = 0$, gives

$$\begin{aligned} E(\mathbf{d})/\sqrt{N} &= \sqrt{N}\gamma \left.\frac{\partial \mathbf{p}_1}{\partial \gamma}\right|_{\gamma=0} + \sqrt{N}\gamma^2 \left.\frac{\partial^2 \mathbf{p}_1}{\partial \gamma^2}\right|_{\gamma=0} \\ &= \delta \left.\frac{\partial \mathbf{p}_1}{\partial \gamma}\right|_{\gamma=0} + O(1/\sqrt{N}) \\ &\to \delta \left.\frac{\partial \mathbf{p}_1}{\partial \gamma}\right|_{\gamma=0}. \end{aligned}$$

Equation (2.19) comes from differentiating (2.18) with respect to γ and evaluating it at $\gamma = 0$. For the Cramér-von Mises statistics, the expectation of $\mathbf{X}$ is needed.

$$\begin{aligned} E(\mathbf{X}) &= E(\mathbf{M}^{1/2}\mathbf{Z})/\sqrt{N} \\ &= E(\mathbf{M}^{1/2}\mathbf{A}\mathbf{d}/\sqrt{N}) \\ &= \mathbf{M}^{1/2}\mathbf{A}E(\mathbf{d}/\sqrt{N}). \end{aligned}$$

Since the matrix $\mathbf{A}$ applied to the vector of first differences of probabilities given in (2.19) is equal to $-\mathbf{p}$, the mean of s_i becomes $-\delta \mathbf{w}_i{}'\mathbf{M}^{1/2}\mathbf{p}$. It is easy to show that the covariance of $\mathbf{d}$ under the alternative hypothesis is the same as the covariance under the null hypothesis.

The asymptotic power is compared along the lines proposed by Durbin and Knott (1972) and Durbin, Knott and Taylor (1975), and developed by Stephens (1976b). The test based on the maximum likelihood estimator of the parameter, γ, will be the locally most powerful unbiased test (Cox and Hinkley, 1974). The variance of the maximum likelihood test for the

Poisson distribution against the negative binomial alternative is the inverse of the Cramér-Rao lower bound,

$$E\left(\left.\frac{\partial^2 \log f}{\partial \gamma^2}\right|_{\gamma=0}\right) = \mu^2/2.$$

The parameter, δ, is chosen to make the power for this test a fixed value. Here the value used is 0.50. For a 0.05-level test to give a two-sided power of 0.50, $\delta = 1.96\sqrt{2}/\mu$.

Powers for the Cramér-von Mises statistics were computed by fitting a curve of the form $a + b\chi_p^2$, where a, b, p are chosen so that the first three cumulants match those of the statistics. Powers for smooth tests were determined by evaluating the appropriate non-central χ^2 distribution. The asymptotic powers are given in Table 2.9.

Results and comments

The results of the asymptotic power analysis show that for negative-binomial alternatives, A^2 has the best power among the Cramér-von Mises tests, and is nearly as powerful as the best test, with W_m^2 slightly worse. The results also indicate that adding additional components to the smooth statistic reduces the asymptotic power. The smooth statistic with two components has similar power to A^2 and smooth statistics with additional components have progressively worse power than A^2.

Table 2.9: Asymptotic power of the Cramér-von Mises statistics for testing for the Poisson distribution with estimated mean, μ.

This table gives the asymptotic power (%) of the Cramér-von Mises test for selected values of the mean, μ, against a negative-binomial alternative with parameter (γ) chosen to give the locally most powerful test, the dispersion test, a power of 50%.

μ	Test Statistics						
	W^2	U^2	A^2	W_m^2	S_2	S_3	S_4
.1	49	48	47	48	40	34	31
.5	41	36	40	40	40	34	31
1	33	29	37	36	40	34	31
2	28	27	37	35	40	34	31
5	28	28	37	35	40	34	31
10	28	29	38	36	40	34	31
20	28	29	38	36	40	34	31

2.5.3 Finite Samples

For finite samples, Monte Carlo studies were used to estimate power. Common alternatives to the Poisson distribution can be categorized by the ratio of the variance to the mean; this is equal to one for the Poisson distribution. Distributions with variance larger than the mean are considered *overdispersed*, and with variance smaller than the mean are referred to as *underdispersed*.

The most common overdispersed alternative to the Poisson distribution is the negative binomial. This distribution arises as the number of failures before K successes with probability of failure, p, but can also be regarded as a Poisson-Gamma mixture; that is, the distribution produced when the parameter of a Poisson distribution itself has a Gamma distribution. Another overdispersed alternative examined was a mixture of two Poisson variables. For underdispersed alternatives, the binomial and discrete uniform distributions were examined. Finally, distributions where the parameters could be chosen to give the variance equal to the mean, as for the Poisson distribution, were also investigated. The beta-binomial distribution and the discrete uniform distribution were chosen in this category. Figures 2.1 and 2.2 show the probability functions for the Poisson distribution with mean, $\mu = 1$, and the beta-binomial distribution with parameters $(\alpha = 1, \beta = 2, K = 3)$ which have the same mean and variance.

Comparisons of power for the Cramér-von Mises statistics and the other tests of fit, when used in testing against the above alternatives, are given in Tables 2.10 and 2.11. One thousand samples of size 20 were generated from each alternative distribution with means of $\mu = 1$ and $\mu = 10$. The finite null percentage points of all statistics compared were found by Monte Carlo simulation using 25,000 samples. The maximum standard error of the power results is equal to $.5/\sqrt{1000} \approx 1.6\%$. Random samples were generated using IMSL subroutines (IMSL, 1987).

Results and comments

1. As expected, the dispersion test and the one-component smooth test perform very well for overdispersed alternatives, with the one-component smooth test having slightly better power. The statistics A^2 and W^2_m, the four-component smooth test and the statistics based on the probability generating function also have good power against overdispersed alternatives. The statistics, W^2 and U^2 have lower power than A^2, and

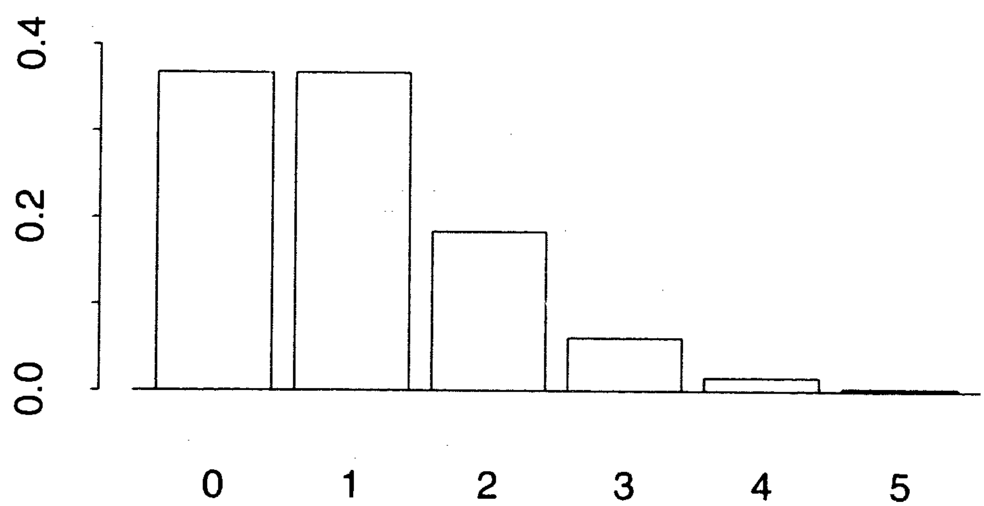

Figure 2.1: Probability function of a Poisson distribution with mean, $\mu = 1$

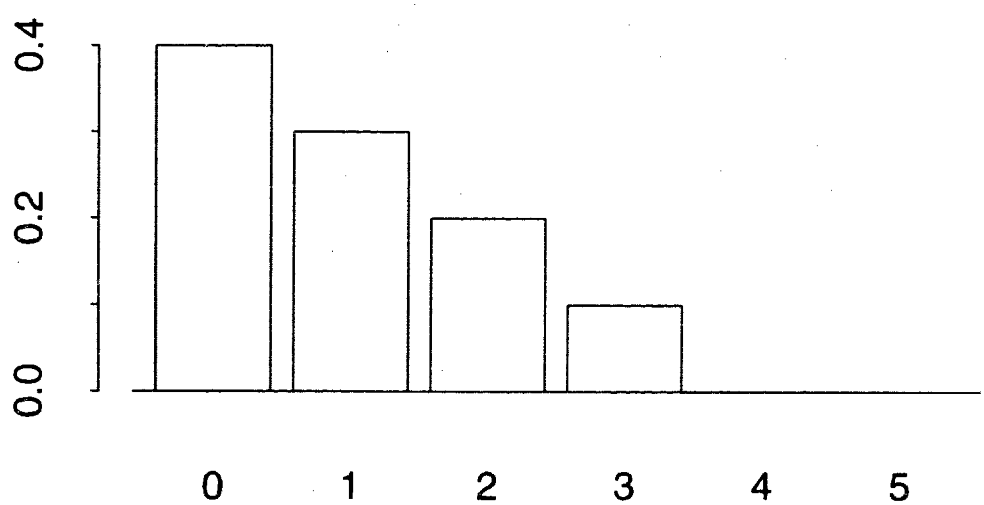

Figure 2.2: Probability function of a beta-binomial distribution ($\alpha = 1, \beta = 2, K = 3$) with mean and variance both equal to 1.

both the Pearson χ^2 statistic and the correlation statistic perform very poorly.

2. For underdispersed alternatives, the dispersion test has the best power. The next most powerful tests are the one-component smooth test and the Cramér-von Mises statistics which have approximately equal power. The four component smooth test and the probability generating function statistic, T, are especially poor. The correlation test was the most powerful at detecting underdispersed distributions with $\mu = 1$, but had very poor power against underdispersed alternatives with $\mu = 10$.

3. Against alternatives with the mean equal to the variance, the Cramér-von Mises statistics have the best power. Since the dispersion test and the one-component smooth test primarily detect differences between the mean and variance, they do very poorly against these alternatives. The four-component smooth test and the probability generating function statistic, P, also have poor power. The power of the correlation statistic was inconsistent, very good against some alternatives and poor against others.

With the exception of the Cramér-von Mises statistics, all statistics gave very poor power against at least one of the classes of alternatives. For over- or under-dispersed alternatives, the dispersion test or the standardized dispersion test (one-component smooth test) are recommended. However, if an omnibus goodness-of-fit test for the Poisson distribution is desired, and in particular, if the alternative is "close" to the Poisson in the sense that the variance is almost equal to the mean, then A^2 is the recommended statistic.

Table 2.10: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20. Alternative distributions with a mean, μ, of 1 were generated. The parameters and variance of the alternative distribution, σ^2, are indicated. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics					
	W^2	U^2	A^2	W_m^2	P	T
Overdispersed						
Neg Bin[$K = 3, p = .25$] (1.33)	88	69	125	118	123	189
Neg Bin[$K = 1, p = .5$] (2)	343	261	457	410	426	530
.5P(.2)+.5P(1.8) (1.64)	338	306	403	382	412	429
.5P(0)+.5P(2.0) (2)	791	756	821	781	789	775
Underdispersed						
Binomial[p=.5,K=2] (.5)	373	370	393	433	468	91
Discrete Uniform[0,2] (.67)	165	158	206	206	83	95
Equal Dispersion						
Beta-Binomial[$\alpha=1, \beta=2, K=3$]	71	73	73	77	34	63

	D	$\hat{S}_1$	$\hat{S}_4$	X_1^2	X_2^2	R
Overdispersed						
Neg Bin[$K = 3, p = .25$] (1.33)	149	198	176	30	93	31
Neg Bin[$K = 1, p = .5$] (2)	490	544	505	160	285	100
.5P(.2)+.5P(1.8) (1.64)	351	437	417	37	113	10
.5P(0)+.5P(2.0) (2)	681	753	783	105	218	19
Underdispersed						
Binomial[p=.5,K=2] (.5)	497	380	211	44	18	604
Discrete Uniform[0,2] (.67)	165	71	104	2	21	477
Equal Dispersion						
Beta-Binomial[$\alpha=1, \beta=2, K=3$]	14	10	45	3	5	25

Table 2.11: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20. Alternative distributions with a mean, μ, of 10 were generated. The parameters and variance of the alternative distribution, σ^2, are indicated. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics					
	W^2	U^2	A^2	W^2_m	P	T
Overdispersed						
Neg Bin[$K = 30, p = .25$] (13.3)	90	86	158	141	68	61
Neg Bin[$K = 10, p = .5$] (20)	347	316	553	501	571	610
.5P(8)+.5P(12) (14)	162	156	269	243	254	295
.5P(7)+.5P(13) (19)	438	425	612	580	620	637
Underdispersed						
Binomial[p=.33,K=30] (6.7)	133	155	120	126	59	24
Binomial[p=.5,K=20] (5)	358	374	344	358	178	90
Discrete Uniform[7,13] (4)	337	375	435	421	292	68
Equal Dispersion						
Beta-Binomial[$\alpha=2, \beta=.6, K=13$]	741	685	748	738	402	745
Discrete Uniform[5,15]	129	132	130	133	3	156

	D	$\hat{S}_1$	$\hat{S}_4$	X_1^2	X_2^2	R
Overdispersed						
Neg Bin[$K = 30, p = .25$] (13.3)	165	205	204	68	61	64
Neg Bin[$K = 10, p = .5$] (20)	608	648	606	224	209	89
.5P(8)+.5P(12) (14)	260	307	254	112	89	33
.5P(7)+.5P(13) (19)	640	688	601	295	249	27
Underdispersed						
Binomial[p=.33,K=30] (6.7)	170	130	19	100	82	57
Binomial[p=.5,K=20] (5)	479	389	62	253	179	108
Discrete Uniform[7,13] (4)	794	641	76	349	484	243
Equal Dispersion						
Beta-Binomial[$\alpha=2, \beta=.6, K=13$]	126	116	551	196	848	906
Discrete Uniform[5,15]	8	6	12	103	80	110

2.6 Examples

2.6.1 Example 1

The data in Table 2.12, taken from Zar (1974), show the number of sparrow nests found in a one hectare area. The cumulative observed and expected histograms are given in Figure 2.3, and the standardized difference is plotted in Figure 2.4. The standardized difference is the difference between the observed and expected values divided by the standard deviation to give asymptotic standard normal values. The sample mean and variance are 1.1 and .810, respectively, indicating a small amount of underdispersion. The values and significance levels of the Cramér-von Mises statistics and other test statistics are found in Table 2.13. The Cramér-von Mises statistics were calculated by stopping at the first seven terms, and the Pearson χ^2 statistics was calculated after grouping the data for cells three or greater. The Cramér-von Mises statistics and the Pearson χ^2 all suggested evidence against the Poisson hypothesis, with significance levels around 0.05. The dispersion test and standardized dispersion test did not reject the Poisson hypothesis, as each had a significance level greater than 0.10.

Table 2.12: Sparrow Nest Data

No. of nests	Frequency	Cumulative Frequency	Cumulative Expected	Standardized Difference	Pr(X=x)
0	9	9	13.32	-1.45	.3328
1	22	31	27.96	1.05	.3661
2	6	37	36.02	0.52	.2013
3	2	39	38.97	0.03	.0738
4	1	40	39.96	0.47	.0203
5	0	40	39.99	0.20	.0044
6	0	40	40.00	0.08	.0008
7	0	40	40.00	0.03	.0001
8	0	40	40.00	0.01	.0000

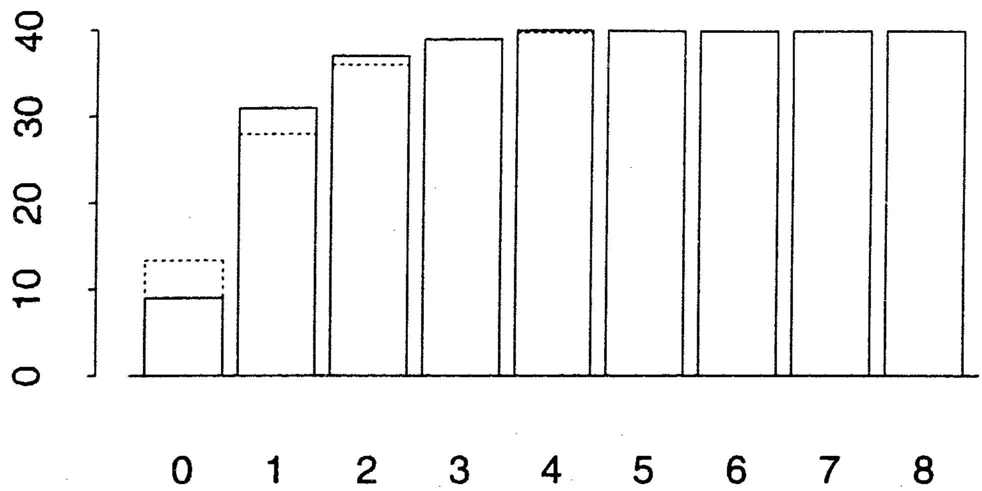

Figure 2.3: Cumulative observed (—) and expected (- -) histograms for the sparrow nest data.

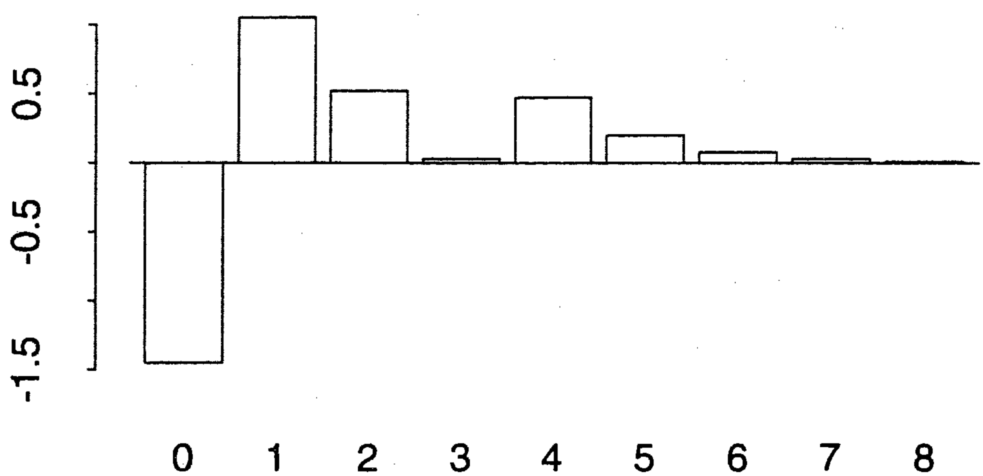

Figure 2.4: Standardized difference between the observed and expected histograms for the sparrow nest data.

Table 2.13: Test statistics for the sparrow nest data

Test Statistic	Value	Significance Level
W^2	.24	.027
U^2	.24	.024
A^2	1.16	.054
W_m^2	.72	.038
D	28.59	.185
S_1	1.59	.208
χ_P^2	5.99	.050

2.6.2 Example 2

The data in Table 2.14 show the frequency of radioactive decay counts of Polonium, taken from Hoaglin (1980), and reproduced in Rayner and Best (1989). The cumulative observed and expected histograms are found in Figure 2.5, and the difference is plotted in Figure 2.6. The standardized difference is the difference between the observed and expected values divided by the standard deviation to give asymptotic standard normal values. The sample mean and variance are 3.87 and 3.70, respectively, indicating Poisson dispersion. The values and significance levels of the Cramér-von Mises statistics and other test statistics are found in Table 2.15. The Cramér-von Mises statistics were calculated by stopping after the first fourteen terms, and the Pearson χ^2 statistic was calculated after grouping the data for cells eleven or greater. The Cramér-von Mises statistics, A^2 and W_m^2 rejected the Poisson hypothesis, and the significance levels for W^2 and U^2 were just larger than 0.05. The dispersion test and standardized dispersion test each had a significance level around 0.10, and the Pearson χ^2 test accepted the Poisson hypothesis.

Table 2.14: Radioactive Decay Counts of Polonium

Count	Frequency	Cumulative Frequency	Cumulative Expected	Standardized Difference	Pr(X=x)
0	57	57	54.31	0.37	.0282
1	203	260	264.59	-0.30	.0806
2	383	643	671.65	-1.28	.1561
3	525	1168	1196.97	-1.14	.2014
4	532	1700	1705.41	-0.22	.1950
5	408	2108	2099.10	0.44	.1510
6	273	2381	2353.14	1.84	.0974
7	139	2520	2493.64	2.52	.0539
8	45	2565	2561.63	0.50	.0261
9	27	2592	2590.88	0.27	.0112
10	10	2602	2602.20	-0.08	.0043
11	4	2606	2606.19	-0.14	.0153
12	0	2606	2607.47	-2.04	.0005
13	1	2607	2607.86	-2.28	.0001
14	1	2608	2607.96	0.19	.0001

Table 2.15: Test statistics for the Polonium count data

Test Statistic	Value	Significance Level
W^2	.16	.060
U^2	.15	.069
A^2	1.25	.033
W_m^2	1.26	.042
D	2488.92	.096
S_1	2.71	.099
χ_P^2	12.96	.226

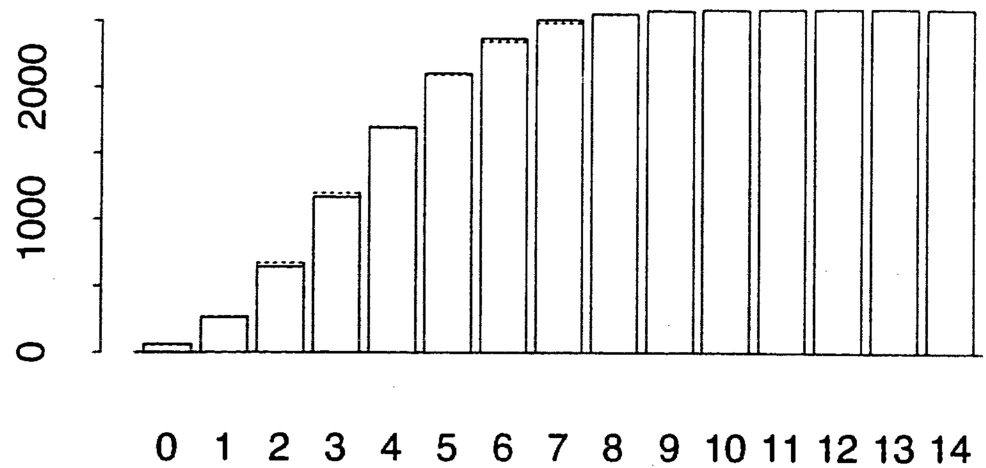

Figure 2.5: Cumulative observed (—) and expected (- -) histograms for the Polonium count data.

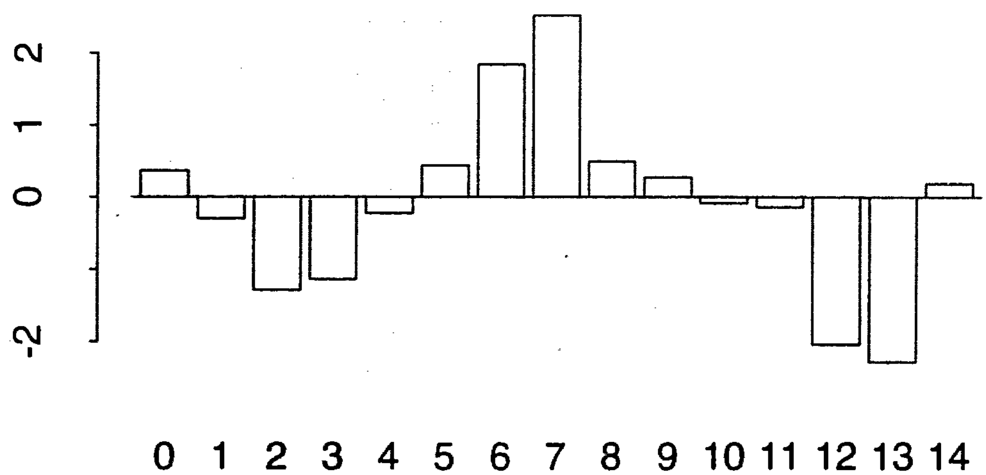

Figure 2.6: Standardized difference between the observed and expected histograms for the Polonium count data.

Chapter 3

Binomial Distribution

3.1 Introduction

In this chapter, the Cramér-von Mises statistics are developed as tests for the binomial distribution. In section 3.2, the definitions of the Cramér-von Mises statistics and the basic theory are given. In section 3.3, the percentage points to make tests for the binomial distribution are given for the cases where the probability of success, θ, is known and also for the case where θ is estimated. In section 3.4 power studies are presented. Comparisons are made with the Pearson χ^2 test and the score test for the beta-binomial distribution which is found to be powerful against distributions with larger variance. In many other cases, A^2 is found to have good power and is recommended for use as an omnibus test for the binomial distribution.

3.2 Cramér-von Mises Statistics

Let X be a binomial random variable, with parameters θ and K; the probability that $X = j$, p_j is given by

$$p_j = p_j(\theta, K) = C_j^K \theta^j (1-\theta)^{K-j}$$

where $j = 0, 1, \ldots, K$ and K and θ are known. Suppose the random sample to be tested is $x_1, x_2, \ldots, x_N$. When θ is not known, and is estimated by maximum likelihood, the estimate is $\hat{\theta} = \bar{x}/K = \sum_{j=1}^{N} x_j/KN$. The Cramér-von Mises statistics will again be calculated as for the Poisson distribution, using $\hat{Z}_j$, $\hat{p}_j$ and $\hat{H}_j$. Here,

$$\hat{p}_j = C_j^K \hat{\theta}^j (1-\hat{\theta})^{K-j},$$

where $j = 0, 1, \ldots, K$, and $\hat{\mathbf{p}}$ is the vector of $\hat{p}_j$ values.

From section 2.3.2 the asymptotic covariance matrix of the variable $\hat{\mathbf{d}}/\sqrt{N} = (\mathbf{o}-\hat{\mathbf{e}})/\sqrt{N}$ is

$$\hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}_0 - \mathbf{g}\mathbf{g}'/\mathbf{g}'\mathbf{D}^{-1}\mathbf{g},$$

where $\boldsymbol{\Sigma}_0$ is defined in section 2.3.1. The vector $\mathbf{g}$ has jth component

$$\begin{aligned} g_j &= C_j^K[j\theta^{j-1}(1-\theta)^{K-j} - (K-j)(1-\theta)^{K-j-1}\theta^j] \\ &= K(p_{j-1}(\theta, K-1) - p_j(\theta, K-1), \end{aligned} \tag{3.1}$$

where $p_j(\theta, K-1)$ is the binomial probability of observing a count j given a probability of success, θ, in $K-1$ trials (with $p_{-1}(\theta, K-1) = p_K(\theta, K-1) = 0$). By combining terms in a different way, g_j can be written

$$g_j = \frac{(j-K\theta)}{\theta(1-\theta)} p_j. \tag{3.2}$$

Using (3.2) it is easily seen that $\mathbf{g}'\mathbf{D}^{-1}\mathbf{g} = K/[\theta(1-\theta)]$. Then

$$\hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}_0 - \frac{\theta(1-\theta)}{K}\mathbf{g}\mathbf{g}', \tag{3.3}$$

and as $N \to \infty$, $\hat{\mathbf{Z}}/\sqrt{N} = \mathbf{A}\hat{\mathbf{d}}/\sqrt{N}$ converges in distribution to a mean zero multivariate normal random variable with covariance matrix,

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \mathbf{A}\hat{\boldsymbol{\Sigma}}_0\mathbf{A}' \\ &= \mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}' - \mathbf{A}\mathbf{g}\mathbf{g}'\mathbf{A}'/\mathbf{g}'\mathbf{D}^{-1}\mathbf{g} \\ &= \boldsymbol{\Sigma} - K\theta(1-\theta)\mathbf{r}\mathbf{r}', \end{aligned} \tag{3.4}$$

where $\mathbf{r}$ is the vector with jth element $r_j = p_j(\theta, K-1)$; the result follows since $\mathbf{A}\mathbf{g} = -K\mathbf{r}$.

3.3 Moments and Percentage Points

3.3.1 Known θ

The mean value of the W^2 statistic becomes, for known θ;

$$E(W^2) = E[\textstyle\sum_{j=0}^K (S_j - T_j)^2 p_j]$$

$$\begin{aligned} &= \textstyle\sum_{j=0}^{K} p_j E[(S_j - T_j)^2] \\ &= \textstyle\sum_{j=0}^{K} p_j Var[S_j] \\ &= \textstyle\sum_{j=0}^{K} p_j H_j(1 - H_j). \end{aligned}$$

The means of the other test statistics can be similarly derived and are as follows:

$$\begin{aligned} E(U^2) &= E(W^2) - \textstyle\sum_{i=0}^{K}\sum_{j=0}^{K} p_i p_j (\min\{H_i, H_j\} - H_i H_j) \\ E(A^2) &= 1 - p_K \end{aligned}$$

For the modified statistics of section 2.2.2;

$$\begin{aligned} E(W_m^2) &= \textstyle\sum_{j=0}^{K} H_j(1 - H_j) \leq \theta K \\ E(A_m^2) &= K \end{aligned}$$

Note that the means of the test statistics do not depend on the sample size N. For different values of the number of trials, K, and for different values of the known probability of success, θ, the means and variances of the asymptotic distributions for all these statistics are given in Table 3.1 for $K = 5$ and 20 and a range of values of θ.

Tables

For Table 3.1 and in all other tables occurring in this chapter, for reasons of space, only a small selection will be given of the various values which have been calculated.

Percentage Points

As before, the asymptotic distribution of a typical test statistic is a sum of weighted χ_1^2 variables, with weights equal to the eigenvalues of the covariance matrix, $\mathbf{\Sigma}_X$, of the statistic. The λ_i have been calculated for various values of the number of trials, K, and the success probability, θ, using S-PLUS (S-PLUS, 1991). The percentage points were then found by Imhof's method (Imhof, 1961).

For W^2, the matrix $\mathbf{M}$ is equal to $\mathbf{D}$, the diagonal matrix with the binomial probabilities, p_j, on the diagonal. For statistics U^2, A^2, W_m^2 and A_m^2, the $\mathbf{M}$ matrices are $(\mathbf{I}-\mathbf{D}\mathbf{1}\mathbf{1}')\mathbf{D}(\mathbf{I}-\mathbf{1}\mathbf{1}'\mathbf{D})$, $\mathbf{D}\mathbf{G}^{-1}$, $\mathbf{I}$ and $\mathbf{G}^{-1}$, respectively. The asymptotic percentage points for all these statistics are recorded in Table 3.2 for $K = 5$ and 20 and a range of values of θ.

Since for the binomial distribution K and θ are neither location nor scale parameters, the asymptotic points depend on these parameters. There is a connection between these points and the points used for testing for a known continuous distribution, given by Stephens (1986), where such a test situation is called Case 0. As K tends to infinity, the percentage points for W^2, U^2 and A^2 tend to those for Case 0. These points were given in the preceding chapter.

For finite samples, percentage points for $K = 20$, $\theta = .5$ and for various N, are given in Table 3.3. With the exception of the statistic, A_m^2, the points converge rapidly to the asymptotic points, which can be used for samples of size greater than 10. This is typical for all the extensive tables which have been produced. Even for A_m^2, additional simulations for sample size 500 and 1000 show that the points do in fact converge to the asymptotic points, but extremely slowly.

An S function (Becker, Chambers and Wilks, 1988; S-PLUS, 1991) has been written to compute the Cramér-von Mises statistics and their asymptotic p-values.

Table 3.1: Asymptotic mean (M) and variance (V) for the Cramér-von Mises test statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ.

$K = 5$

	W^2		U^2		A^2		W_m^2		A_m^2	
θ	M	V	M	V	M	V	M	V	M	V
.01	0.044	0.004	0.002	0.000	1.000	1.817	0.048	0.004	5	10.146
.05	0.140	0.037	0.029	0.002	1.000	1.332	0.198	0.063	5	10.685
.10	0.168	0.044	0.056	0.005	1.000	1.034	0.326	0.138	5	11.279
.20	0.163	0.028	0.071	0.005	1.000	0.805	0.476	0.217	5	12.248
.30	0.159	0.026	0.074	0.005	0.998	0.728	0.558	0.264	5	12.942
.40	0.157	0.024	0.074	0.004	0.990	0.694	0.601	0.294	5	13.360
.50	0.155	0.024	0.073	0.004	0.969	0.675	0.615	0.304	5	13.500
.60	0.152	0.023	0.069	0.004	0.922	0.656	0.601	0.294	5	13.360
.70	0.146	0.022	0.062	0.004	0.832	0.614	0.558	0.264	5	12.942
.80	0.133	0.022	0.051	0.003	0.672	0.493	0.476	0.217	5	12.248
.90	0.085	0.013	0.039	0.003	0.409	0.239	0.326	0.138	5	11.279
.95	0.036	0.003	0.023	0.001	0.226	0.085	0.198	0.063	5	10.685
.99	0.002	0.000	0.002	0.000	0.049	0.005	0.048	0.004	5	10.146

$K = 20$

	W^2		U^2		A^2		W_m^2		A_m^2	
θ	M	V	M	V	M	V	M	V	M	V
.01	0.125	0.030	0.021	0.001	1.000	1.435	0.167	0.046	20	41.392
.05	0.170	0.031	0.072	0.005	1.000	0.821	0.513	0.244	20	46.597
.10	0.166	0.026	0.077	0.004	1.000	0.692	0.734	0.410	20	52.486
.20	0.165	0.024	0.080	0.003	1.000	0.635	0.996	0.692	20	62.234
.30	0.165	0.023	0.081	0.003	1.000	0.618	1.146	0.894	20	69.213
.40	0.164	0.023	0.081	0.003	1.000	0.609	1.228	1.015	20	73.403
.50	0.164	0.023	0.081	0.003	1.000	0.605	1.254	1.055	20	74.800
.60	0.163	0.022	0.081	0.003	1.000	0.602	1.228	1.015	20	73.403
.70	0.162	0.022	0.081	0.003	0.999	0.602	1.146	0.894	20	69.213
.80	0.159	0.022	0.079	0.003	0.989	0.604	0.996	0.692	20	62.234
.90	0.150	0.021	0.069	0.003	0.878	0.588	0.734	0.410	20	52.486
.95	0.128	0.021	0.052	0.003	0.642	0.436	0.513	0.244	20	46.597
.99	0.025	0.001	0.017	0.001	0.182	0.056	0.167	0.046	20	41.392

Table 3.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ.

$K = 5$

W^2 — Upper tail significance level α

θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.059	0.091	0.120	0.170	0.191	0.294	0.358	0.481
.05	0.184	0.285	0.373	0.523	0.686	0.894	1.074	1.179
.10	0.219	0.328	0.420	0.586	0.759	0.996	1.178	1.417
.20	0.218	0.301	0.370	0.493	0.621	0.794	0.930	1.251
.30	0.208	0.288	0.355	0.477	0.604	0.778	0.913	1.232
.40	0.206	0.284	0.348	0.465	0.587	0.753	0.882	1.190
.50	0.203	0.279	0.342	0.457	0.577	0.741	0.869	1.176
.60	0.199	0.274	0.336	0.449	0.567	0.728	0.854	1.151
.70	0.193	0.267	0.327	0.437	0.551	0.707	0.827	1.114
.80	0.176	0.251	0.314	0.425	0.543	0.707	0.832	1.184
.90	0.111	0.171	0.222	0.313	0.408	0.537	0.642	0.872
.95	0.048	0.074	0.098	0.137	0.180	0.229	0.230	0.234
.99	0.003	0.005	0.006	0.009	0.010	0.015	0.018	0.020

U^2 — Upper tail significance level α

θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.003	0.004	0.006	0.008	0.009	0.014	0.017	0.019
.05	0.038	0.059	0.077	0.109	0.142	0.183	0.223	0.310
.10	0.073	0.112	0.144	0.203	0.264	0.347	0.412	0.564
.20	0.097	0.133	0.163	0.214	0.267	0.339	0.394	0.536
.30	0.099	0.133	0.161	0.207	0.256	0.321	0.371	0.490
.40	0.100	0.134	0.160	0.205	0.250	0.311	0.356	0.461
.50	0.098	0.130	0.156	0.200	0.245	0.305	0.352	0.462
.60	0.094	0.126	0.151	0.194	0.238	0.300	0.338	0.436
.70	0.083	0.114	0.139	0.183	0.229	0.291	0.339	0.453
.80	0.068	0.096	0.119	0.159	0.201	0.259	0.304	0.432
.90	0.051	0.077	0.099	0.139	0.181	0.238	0.281	0.337
.95	0.030	0.047	0.062	0.087	0.114	0.146	0.178	0.245
.99	0.003	0.004	0.005	0.008	0.009	0.013	0.016	0.018

Table 3.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 5$

A^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	1.310	2.024	2.625	3.707	4.856	6.338	7.570	0.357
.05	1.312	1.899	2.397	3.295	4.320	5.510	6.512	8.443
.10	1.328	1.837	2.260	3.006	3.788	4.853	5.679	7.648
.20	1.318	1.756	2.114	2.755	3.420	4.348	5.034	6.711
.30	1.300	1.714	2.054	2.661	3.294	4.155	4.832	6.382
.40	1.286	1.690	2.022	2.613	3.230	4.064	4.730	6.241
.50	1.263	1.661	1.988	2.570	3.176	3.993	4.650	6.159
.60	1.212	1.607	1.931	2.505	3.103	3.879	4.555	6.071
.70	1.099	1.487	1.809	2.371	2.959	3.759	4.385	5.865
.80	0.887	1.239	1.533	2.059	2.610	3.362	3.940	5.955
.90	0.534	0.785	0.999	1.383	1.773	2.331	2.753	3.363
.95	0.295	0.448	0.578	0.811	1.054	1.385	1.641	1.908
.99	0.065	0.100	0.131	0.183	0.242	0.318	0.367	0.413

W_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.063	0.098	0.127	0.178	0.235	0.307	0.368	0.410
.05	0.258	0.390	0.502	0.703	0.912	1.197	1.417	1.673
.10	0.428	0.616	0.776	1.066	1.367	1.772	2.104	2.662
.20	0.631	0.863	1.051	1.392	1.745	2.226	2.598	3.485
.30	0.731	0.983	1.189	1.564	1.954	2.496	2.901	3.882
.40	0.783	1.049	1.269	1.664	2.076	2.630	3.077	4.114
.50	0.800	1.070	1.294	1.695	2.114	2.687	3.133	4.190
.60	0.783	1.049	1.269	1.664	2.076	2.630	3.077	4.114
.70	0.731	0.983	1.189	1.564	1.954	2.496	2.901	3.882
.80	0.631	0.863	1.051	1.392	1.745	2.226	2.598	3.485
.90	0.428	0.616	0.776	1.066	1.367	1.772	2.104	2.662
.95	0.258	0.390	0.502	0.703	0.912	1.197	1.417	1.673
.99	0.063	0.098	0.127	0.178	0.235	0.307	0.368	0.410

Table 3.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 5$

A^2_m	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	6.623	8.125	9.258	11.116	12.910	15.212	16.918	20.883
.05	6.614	8.159	9.336	11.287	13.191	15.668	17.524	22.004
.10	6.605	8.196	9.421	11.472	13.493	16.155	18.165	22.915
.20	6.589	8.254	9.552	11.756	13.963	16.907	19.158	24.473
.30	6.577	8.294	9.643	11.955	14.289	17.425	19.838	25.549
.40	6.570	8.317	9.697	12.072	14.481	17.730	20.210	26.172
.50	6.568	8.324	9.715	12.111	14.545	17.830	20.357	26.381
.60	6.570	8.317	9.697	12.072	14.481	17.730	20.210	26.172
.70	6.577	8.294	9.643	11.955	14.289	17.425	19.838	25.549
.80	6.589	8.254	9.552	11.756	13.963	16.907	19.158	24.473
.90	6.605	8.196	9.421	11.472	13.493	16.155	18.165	22.915
.95	6.614	8.159	9.336	11.287	13.191	15.668	17.524	22.004
.99	6.623	8.125	9.258	11.116	12.910	15.212	16.918	20.883

$K = 20$

W^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.164	0.255	0.332	0.466	0.616	0.803	0.964	1.073
.05	0.226	0.314	0.386	0.515	0.650	0.835	0.978	1.317
.10	0.215	0.294	0.362	0.482	0.609	0.782	0.915	1.232
.20	0.211	0.288	0.352	0.469	0.591	0.758	0.887	1.193
.30	0.210	0.285	0.349	0.465	0.585	0.750	0.878	1.180
.40	0.209	0.284	0.348	0.462	0.582	0.746	0.872	1.173
.50	0.208	0.283	0.346	0.460	0.579	0.742	0.868	1.167
.60	0.207	0.282	0.344	0.458	0.577	0.738	0.864	1.161
.70	0.206	0.280	0.342	0.455	0.573	0.733	0.859	1.154
.80	0.203	0.277	0.338	0.450	0.566	0.726	0.849	1.141
.90	0.195	0.266	0.327	0.435	0.548	0.703	0.823	1.108
.95	0.169	0.241	0.302	0.413	0.533	0.687	0.814	1.132
.99	0.033	0.052	0.067	0.095	0.124	0.160	0.160	0.209

Table 3.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 20$

U^2				Upper tail significance level α				
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.027	0.043	0.056	0.079	0.103	0.131	0.132	0.134
.05	0.097	0.133	0.163	0.215	0.269	0.342	0.398	0.540
.10	0.103	0.135	0.160	0.202	0.245	0.302	0.347	0.446
.20	0.105	0.134	0.157	0.196	0.235	0.286	0.325	0.415
.30	0.106	0.134	0.156	0.194	0.231	0.281	0.319	0.408
.40	0.106	0.133	0.155	0.193	0.230	0.280	0.317	0.404
.50	0.106	0.133	0.155	0.192	0.230	0.279	0.317	0.404
.60	0.106	0.133	0.155	0.193	0.230	0.280	0.317	0.404
.70	0.105	0.134	0.156	0.193	0.231	0.281	0.319	0.407
.80	0.104	0.133	0.155	0.194	0.232	0.283	0.321	0.411
.90	0.092	0.121	0.144	0.184	0.224	0.276	0.316	0.408
.95	0.069	0.096	0.118	0.157	0.197	0.252	0.295	0.397
.99	0.023	0.036	0.046	0.066	0.086	0.110	0.112	0.112

A^2				Upper tail significance level α				
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	1.304	1.921	2.446	3.390	4.331	5.716	6.753	8.238
.05	1.313	1.757	2.121	2.775	3.454	4.370	5.105	6.812
.10	1.285	1.689	2.024	2.624	3.252	4.110	4.776	6.374
.20	1.269	1.657	1.979	2.557	3.162	3.989	4.632	6.152
.30	1.264	1.646	1.965	2.536	3.133	3.951	4.584	6.087
.40	1.262	1.642	1.958	2.525	3.119	3.931	4.565	6.053
.50	1.261	1.640	1.955	2.520	3.111	3.919	4.548	6.032
.60	1.262	1.639	1.953	2.516	3.105	3.910	4.533	6.016
.70	1.263	1.640	1.953	2.514	3.100	3.903	4.523	6.001
.80	1.258	1.634	1.946	2.503	3.086	3.884	4.501	5.971
.90	1.146	1.520	1.829	2.378	2.950	3.705	4.340	5.724
.95	0.842	1.173	1.451	1.946	2.465	3.175	3.724	5.026
.99	0.237	0.362	0.467	0.656	0.853	1.121	1.329	1.820

Table 3.2: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with known success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 20$

W_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.217	0.329	0.424	0.595	0.772	1.015	1.202	1.438
.05	0.676	0.921	1.124	1.485	1.862	2.377	2.777	3.722
.10	0.947	1.260	1.520	1.988	2.477	3.110	3.666	4.897
.20	1.267	1.673	2.013	2.625	3.264	4.134	4.817	6.406
.30	1.452	1.915	2.302	2.999	3.727	4.724	5.497	7.354
.40	1.553	2.046	2.459	3.202	3.980	5.043	5.867	7.820
.50	1.585	2.088	2.510	3.267	4.060	5.148	5.985	7.980
.60	1.553	2.046	2.459	3.202	3.980	5.043	5.867	7.820
.70	1.452	1.915	2.302	2.999	3.727	4.724	5.497	7.354
.80	1.267	1.673	2.013	2.625	3.264	4.134	4.817	6.406
.90	0.947	1.260	1.520	1.988	2.477	3.110	3.666	4.897
.95	0.676	0.921	1.124	1.485	1.862	2.377	2.777	3.722
.99	0.217	0.329	0.424	0.595	0.772	1.015	1.202	1.438

A_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	23.86	26.59	28.55	31.64	34.50	38.03	40.57	46.16
.05	23.97	26.90	29.05	32.46	35.68	39.73	42.70	49.41
.10	24.07	27.22	29.56	33.33	36.94	41.58	45.04	53.00
.20	24.21	27.70	30.33	34.67	38.90	44.44	48.64	58.48
.30	24.30	28.02	30.85	35.56	40.20	46.34	51.02	62.04
.40	24.35	28.20	31.15	36.07	40.95	47.43	52.38	64.06
.50	24.36	28.26	31.24	36.24	41.20	47.78	52.82	64.72
.60	24.35	28.20	31.15	36.07	40.95	47.43	52.38	64.06
.70	24.30	28.02	30.85	35.56	40.20	46.34	51.02	62.04
.80	24.21	27.70	30.33	34.67	38.90	44.44	48.64	58.48
.90	24.07	27.22	29.56	33.33	36.94	41.58	45.03	53.00
.95	23.97	26.91	29.05	32.46	35.68	39.73	42.70	49.41
.99	23.86	26.59	28.55	31.64	34.50	38.03	40.57	46.16

Table 3.3: Monte Carlo percentage points for the Cramér-von Mises test statistics testing for the binomial distribution with known success probability $\theta = .5$, and number of trials $K = 20$. The asymptotic points are shown for comparison.

$K = 20$ $\theta = .5$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.219	0.281	0.341	0.454	0.564	0.700
10	0.211	0.284	0.345	0.453	0.563	0.714
15	0.207	0.284	0.348	0.462	0.579	0.724
20	0.207	0.285	0.346	0.459	0.569	0.729
40	0.211	0.283	0.345	0.457	0.571	0.735
50	0.206	0.282	0.350	0.462	0.582	0.739
100	0.209	0.279	0.344	0.465	0.586	0.739
∞	0.208	0.283	0.346	0.460	0.579	0.742

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.106	0.132	0.150	0.182	0.220	0.256
10	0.105	0.132	0.154	0.190	0.226	0.268
15	0.105	0.132	0.154	0.191	0.222	0.265
20	0.105	0.132	0.153	0.189	0.226	0.269
40	0.106	0.134	0.156	0.192	0.228	0.274
50	0.104	0.132	0.153	0.190	0.224	0.277
100	0.105	0.133	0.154	0.191	0.228	0.272
∞	0.106	0.133	0.155	0.192	0.230	0.279

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	1.243	1.637	1.994	2.651	3.284	4.178
10	1.263	1.645	1.985	2.558	3.164	4.034
15	1.243	1.645	1.969	2.573	3.205	3.975
20	1.253	1.648	1.962	2.570	3.172	3.979
40	1.268	1.646	1.956	2.519	3.114	3.914
50	1.254	1.648	1.978	2.548	3.154	3.910
100	1.258	1.620	1.944	2.543	3.144	3.976
∞	1.261	1.640	1.955	2.520	3.111	3.919

Table 3.3: Monte Carlo percentage points for the Cramér-von Mises test statistics testing for the binomial distribution with known success probability $\theta = .5$, and number of trials $K = 20$. The asymptotic points are shown for comparison. (continued)

	$K = 20$	$\theta = .5$				
W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	1.588	2.112	2.513	3.243	3.991	5.032
10	1.600	2.082	2.505	3.250	4.026	5.086
15	1.570	2.085	2.521	3.304	4.074	5.074
20	1.572	2.084	2.521	3.296	4.056	5.142
40	1.600	2.102	2.505	3.243	4.019	5.140
50	1.574	2.099	2.538	3.269	4.092	5.091
100	1.584	2.062	2.489	3.283	4.110	5.190
∞	1.585	2.088	2.510	3.267	4.060	5.148

A_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	13.044	18.285	25.506	46.752	58.148	199.391
10	14.229	21.973	26.699	39.431	97.077	115.409
15	15.511	20.660	25.991	46.502	70.282	94.160
20	15.451	20.859	27.195	48.676	59.152	95.150
40	17.032	24.074	28.657	38.458	54.522	152.088
50	18.110	23.575	28.305	38.071	65.781	127.135
100	17.859	23.108	28.498	45.954	66.182	86.277
∞	24.360	28.260	31.240	36.240	41.200	47.780

3.3.2 Estimated θ

The mean values of W^2, A^2, W_m^2 and A_m^2 for testing for the binomial distribution with estimated θ are as follows:

$$\begin{aligned} E(W^2) &= \textstyle\sum_{j=0}^{K} p_j H_j(1-H_j) - K\theta(1-\theta)\sum_{j=0}^{K} r_j^2 p_j \\ E(A^2) &= 1 - p_K - K\theta(1-\theta)\textstyle\sum_{j=0}^{K} r_j^2 p_j / \{H_j(1-H_j)\} \\ E(W_m^2) &= \textstyle\sum_{j=0}^{K} H_j(1-H_j) - K\theta(1-\theta)\sum_{j=0}^{K} r_j^2 \\ E(A_m^2) &= K - 1 - K\theta(1-\theta)\textstyle\sum_{j=0}^{K} r_j^2 / \{H_j(1-H_j)\} \end{aligned}$$

where $p_j = p_j(\theta, K)$ and $r_j = p_j(\theta, K-1)$.

Once again the means of the statistics do not depend on the sample size, N. The mean and variance of the asymptotic distribution of each statistic are given in Table 3.4.

Percentage Points

The percentage points for the Cramér-von Mises statistics are recorded in Table 3.5 for selected values of $K = 5, 20$ and a range of values of θ. As K tends to infinity the points for W^2, U^2 and A^2 tend to the points for testing for a normal distribution with known variance but estimated mean (Case 1), given by Stephens (1986).

For finite N, the percentage points of W^2, U^2, A^2, W_m^2 and A_m^2 for estimated success probability, θ, have been found by Monte Carlo simulation using 25,000 samples. Percentage points for $K = 20$, $\theta = .5$ and various sample sizes, N, are given in Table 3.6. These points converge rapidly to the asymptotic points (except for A_m^2); these can therefore be used for samples of size greater than 10.

Table 3.4: Asymptotic mean (M) and variance (V) for the Cramér-von Mises test statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ.

$K = 5$

	W^2		U^2		A^2		W_m^2		A_m^2	
θ	M	V	M	V	M	V	M	V	M	V
.01	0.001	0.000	0.000	0.000	0.064	0.008	0.002	0.000	3.94	7.84
.05	0.017	0.001	0.011	0.000	0.238	0.093	0.034	0.002	3.74	7.31
.10	0.041	0.003	0.036	0.002	0.350	0.157	0.093	0.014	3.55	6.79
.20	0.060	0.004	0.058	0.004	0.441	0.162	0.188	0.038	3.27	6.01
.30	0.062	0.003	0.059	0.003	0.474	0.148	0.240	0.046	3.09	5.46
.40	0.063	0.003	0.059	0.003	0.485	0.145	0.265	0.049	2.98	5.13
.50	0.063	0.003	0.059	0.003	0.478	0.148	0.273	0.051	2.95	5.02
.60	0.062	0.003	0.057	0.003	0.447	0.151	0.265	0.049	2.98	5.13
.70	0.060	0.003	0.052	0.003	0.379	0.138	0.240	0.046	3.09	5.46
.80	0.049	0.004	0.039	0.003	0.263	0.087	0.188	0.038	3.27	6.01
.90	0.019	0.001	0.013	0.000	0.108	0.019	0.093	0.014	3.55	6.79
.95	0.004	0.000	0.003	0.000	0.036	0.002	0.034	0.002	3.74	7.31
.99	0.000	0.000	0.000	0.000	0.002	0.000	0.002	0.000	3.94	7.84

$K = 20$

	W^2		U^2		A^2		W_m^2		A_m^2	
θ	M	V	M	V	M	V	M	V	M	V
.01	0.125	0.030	0.021	0.001	1.000	1.435	0.167	0.046	18.75	38.00
.05	0.064	0.005	0.060	0.004	0.433	0.150	0.209	0.042	18.05	38.48
.10	0.068	0.003	0.065	0.003	0.481	0.122	0.338	0.067	17.47	39.10
.20	0.071	0.003	0.067	0.003	0.502	0.107	0.477	0.104	16.71	40.12
.30	0.071	0.003	0.068	0.002	0.509	0.103	0.555	0.131	16.27	40.87
.40	0.072	0.003	0.068	0.002	0.513	0.101	0.597	0.147	16.02	41.31
.50	0.072	0.003	0.068	0.002	0.517	0.102	0.611	0.152	15.95	41.47
.60	0.072	0.003	0.068	0.002	0.520	0.104	0.597	0.147	16.02	41.31
.70	0.071	0.003	0.067	0.002	0.524	0.109	0.555	0.131	16.27	40.87
.80	0.070	0.003	0.065	0.002	0.522	0.120	0.477	0.104	16.71	40.12
.90	0.066	0.003	0.058	0.003	0.438	0.134	0.338	0.067	17.47	39.10
.95	0.049	0.003	0.039	0.002	0.260	0.077	0.209	0.042	18.05	38.48
.99	0.025	0.001	0.017	0.001	0.182	0.056	0.167	0.046	18.75	38.00

Table 3.5: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ.

$K = 5$

W^2 — Upper tail significance level α

θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.001	0.002	0.003	0.004	0.004	0.006	0.007	0.008
.05	0.022	0.034	0.045	0.063	0.083	0.105	0.107	0.108
.10	0.054	0.083	0.107	0.152	0.199	0.259	0.310	0.423
.20	0.080	0.114	0.142	0.192	0.244	0.315	0.371	0.492
.30	0.084	0.113	0.136	0.176	0.217	0.268	0.312	0.409
.40	0.084	0.112	0.136	0.176	0.218	0.273	0.318	0.422
.50	0.085	0.113	0.136	0.175	0.214	0.265	0.307	0.398
.60	0.083	0.111	0.135	0.174	0.215	0.269	0.315	0.417
.70	0.081	0.111	0.135	0.176	0.218	0.274	0.318	0.419
.80	0.064	0.096	0.122	0.170	0.219	0.288	0.340	0.464
.90	0.024	0.038	0.050	0.070	0.091	0.121	0.163	0.196
.95	0.005	0.008	0.010	0.015	0.017	0.026	0.030	0.041
.99	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

U^2 — Upper tail significance level α

θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001
.05	0.014	0.022	0.029	0.040	0.045	0.070	0.083	0.114
.10	0.047	0.073	0.095	0.133	0.175	0.228	0.274	0.300
.20	0.077	0.110	0.138	0.188	0.239	0.311	0.368	0.477
.30	0.080	0.108	0.131	0.169	0.209	0.263	0.302	0.397
.40	0.079	0.107	0.130	0.169	0.210	0.265	0.308	0.409
.50	0.079	0.106	0.128	0.166	0.204	0.255	0.294	0.380
.60	0.076	0.103	0.124	0.162	0.200	0.254	0.292	0.387
.70	0.071	0.098	0.120	0.159	0.199	0.253	0.294	0.394
.80	0.051	0.078	0.100	0.140	0.181	0.238	0.282	0.333
.90	0.017	0.026	0.035	0.049	0.064	0.083	0.100	0.111
.95	0.004	0.006	0.008	0.011	0.014	0.018	0.022	0.031
.99	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.5: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 5$

A^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.085	0.133	0.172	0.244	0.319	0.410	0.500	0.681
.05	0.310	0.470	0.606	0.849	1.102	1.448	1.714	2.010
.10	0.461	0.661	0.831	1.133	1.442	1.883	2.262	2.955
.20	0.595	0.796	0.960	1.236	1.518	1.914	2.182	2.895
.30	0.636	0.825	0.973	1.227	1.482	1.824	2.070	2.705
.40	0.647	0.831	0.977	1.228	1.480	1.817	2.040	2.684
.50	0.641	0.829	0.977	1.230	1.485	1.826	2.083	2.707
.60	0.602	0.794	0.947	1.212	1.480	1.930	2.112	2.761
.70	0.507	0.692	0.844	1.111	1.386	1.762	2.057	2.738
.80	0.346	0.495	0.621	0.849	1.080	1.413	1.657	2.325
.90	0.141	0.214	0.276	0.386	0.502	0.659	0.780	0.936
.95	0.047	0.073	0.095	0.134	0.175	0.235	0.273	0.375
.99	0.002	0.004	0.005	0.007	0.008	0.012	0.015	0.016

W_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.002	0.004	0.005	0.007	0.008	0.012	0.014	0.020
.05	0.044	0.068	0.089	0.125	0.164	0.213	0.252	0.284
.10	0.120	0.181	0.233	0.325	0.422	0.553	0.591	0.899
.20	0.251	0.349	0.430	0.574	0.723	0.927	1.085	1.460
.30	0.325	0.432	0.519	0.662	0.809	1.008	1.148	1.491
.40	0.356	0.465	0.551	0.700	0.850	1.054	1.213	1.583
.50	0.365	0.477	0.565	0.717	0.869	1.072	1.227	1.586
.60	0.356	0.465	0.551	0.700	0.850	1.054	1.213	1.583
.70	0.325	0.432	0.519	0.662	0.809	1.008	1.148	1.491
.80	0.251	0.349	0.430	0.574	0.723	0.927	1.085	1.460
.90	0.120	0.181	0.233	0.325	0.422	0.553	0.591	0.899
.95	0.044	0.068	0.089	0.125	0.164	0.213	0.252	0.284
.99	0.002	0.004	0.005	0.007	0.008	0.012	0.014	0.020

Table 3.5: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 5$

A_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	5.30	6.65	7.68	9.37	11.02	13.15	14.74	18.37
.05	5.03	6.33	7.33	8.99	10.62	12.74	14.33	18.01
.10	4.76	6.02	6.99	8.61	10.21	12.32	13.91	17.61
.20	4.37	5.55	6.47	8.03	9.58	11.63	13.20	16.86
.30	4.12	5.25	6.13	7.63	9.12	11.11	12.60	16.19
.40	3.99	5.08	5.93	7.38	8.84	10.77	12.21	15.72
.50	3.94	5.02	5.87	7.30	8.74	10.65	12.06	15.54
.60	3.99	5.08	5.93	7.38	8.84	10.77	12.21	15.72
.70	4.12	5.25	6.13	7.63	9.12	11.11	12.60	16.19
.80	4.37	5.55	6.47	8.03	9.58	11.63	13.20	16.86
.90	4.76	6.02	6.99	8.61	10.21	12.32	13.91	17.61
.95	5.03	6.33	7.33	8.99	10.62	12.74	14.33	18.01
.99	5.30	6.65	7.68	9.37	11.02	13.15	14.74	18.37

$K = 20$

W^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.164	0.255	0.332	0.466	0.616	0.803	0.964	1.073
.05	0.086	0.121	0.149	0.200	0.253	0.326	0.383	0.567
.10	0.091	0.119	0.141	0.180	0.219	0.272	0.313	0.407
.20	0.093	0.118	0.138	0.173	0.207	0.253	0.288	0.370
.30	0.093	0.118	0.137	0.170	0.204	0.248	0.282	0.362
.40	0.093	0.117	0.137	0.169	0.202	0.246	0.280	0.359
.50	0.093	0.117	0.136	0.169	0.202	0.246	0.279	0.358
.60	0.093	0.117	0.136	0.169	0.202	0.247	0.280	0.359
.70	0.093	0.117	0.137	0.170	0.204	0.249	0.283	0.363
.80	0.092	0.118	0.138	0.172	0.207	0.253	0.289	0.372
.90	0.089	0.117	0.139	0.177	0.215	0.265	0.305	0.397
.95	0.065	0.095	0.120	0.166	0.214	0.279	0.329	0.417
.99	0.033	0.052	0.067	0.095	0.124	0.160	0.160	0.209

Table 3.5: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 20$

U^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.027	0.043	0.056	0.079	0.103	0.131	0.132	0.134
.05	0.080	0.114	0.142	0.193	0.246	0.318	0.374	0.493
.10	0.086	0.113	0.135	0.172	0.211	0.262	0.303	0.398
.20	0.088	0.113	0.132	0.165	0.199	0.244	0.278	0.360
.30	0.088	0.112	0.131	0.163	0.195	0.239	0.272	0.351
.40	0.088	0.111	0.130	0.161	0.193	0.236	0.269	0.347
.50	0.088	0.111	0.129	0.161	0.193	0.235	0.268	0.345
.60	0.088	0.111	0.129	0.160	0.192	0.235	0.267	0.344
.70	0.087	0.110	0.129	0.160	0.192	0.235	0.267	0.344
.80	0.086	0.109	0.128	0.160	0.192	0.236	0.268	0.346
.90	0.078	0.103	0.123	0.157	0.191	0.240	0.271	0.351
.95	0.050	0.075	0.095	0.132	0.172	0.224	0.265	0.361
.99	0.023	0.036	0.046	0.066	0.086	0.110	0.112	0.112

A^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	1.304	1.921	2.446	3.390	4.331	5.716	6.753	8.238
.05	0.580	0.773	0.928	1.198	1.472	1.833	2.127	2.770
.10	0.633	0.801	0.933	1.159	1.387	1.693	1.928	2.484
.20	0.644	0.799	0.922	1.132	1.346	1.633	1.853	2.378
.30	0.648	0.800	0.920	1.127	1.336	1.619	1.837	2.352
.40	0.651	0.802	0.921	1.127	1.336	1.618	1.835	2.353
.50	0.655	0.806	0.925	1.132	1.342	1.625	1.843	2.359
.60	0.660	0.812	0.933	1.142	1.354	1.640	1.860	2.385
.70	0.667	0.823	0.947	1.160	1.377	1.669	1.894	2.427
.80	0.672	0.837	0.966	1.190	1.416	1.722	1.957	2.516
.90	0.580	0.760	0.904	1.156	1.413	1.759	2.027	2.661
.95	0.342	0.482	0.599	0.810	1.030	1.331	1.564	2.098
.99	0.237	0.362	0.467	0.656	0.853	1.121	1.329	1.820

Table 3.5: Asymptotic percentage points for the Cramér-von Mises statistics for testing for the binomial distribution with estimated success probability θ, for selected number of trials K and selected values of θ. (continued)

$K = 20$

W_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	0.217	0.329	0.424	0.595	0.772	1.015	1.202	1.438
.05	0.279	0.381	0.465	0.613	0.765	0.973	1.136	1.517
.10	0.448	0.573	0.672	0.841	1.011	1.239	1.419	1.826
.20	0.617	0.771	0.892	1.101	1.313	1.597	1.814	2.330
.30	0.711	0.884	1.020	1.254	1.492	1.811	2.056	2.636
.40	0.762	0.945	1.089	1.338	1.590	1.929	2.189	2.805
.50	0.779	0.964	1.111	1.365	1.621	1.967	2.232	2.859
.60	0.762	0.945	1.089	1.338	1.590	1.929	2.189	2.805
.70	0.711	0.884	1.020	1.254	1.492	1.811	2.056	2.636
.80	0.617	0.771	0.892	1.101	1.313	1.597	1.814	2.330
.90	0.448	0.573	0.672	0.841	1.011	1.239	1.419	1.826
.95	0.279	0.381	0.465	0.613	0.765	0.973	1.136	1.517
.99	0.217	0.329	0.424	0.595	0.772	1.015	1.202	1.438

A_m^2	Upper tail significance level α							
θ	.25	.15	.10	.05	.025	.01	.005	.001
.01	22.44	25.06	26.95	29.92	32.66	36.06	38.50	43.87
.05	21.69	24.36	26.29	29.37	32.24	35.84	38.46	44.30
.10	21.06	23.77	25.76	28.94	31.95	35.75	38.53	44.84
.20	20.25	23.03	25.09	28.43	31.62	35.70	38.73	45.67
.30	19.76	22.60	24.71	28.15	31.45	35.71	38.89	46.21
.40	19.50	22.36	24.50	28.00	31.37	35.73	38.99	46.51
.50	19.42	22.29	24.44	27.95	31.35	35.74	39.02	46.61
.60	19.50	22.36	24.50	28.00	31.37	35.73	38.99	46.51
.70	19.76	22.60	24.71	28.15	31.45	35.71	38.89	46.21
.80	20.25	23.03	25.09	28.43	31.62	35.70	38.73	45.67
.90	21.06	23.77	25.76	28.94	31.95	35.75	38.53	44.84
.95	21.69	24.36	26.29	29.37	32.24	35.84	38.46	44.30
.99	22.44	25.06	26.95	29.92	32.66	36.06	38.50	43.87

Table 3.6: Monte Carlo percentage points for the Cramér-von Mises test statistics testing for the binomial distribution with estimated success probability $\theta = .5$, and number of trials $K = 20$. The asymptotic points are shown for comparison.

$K = 20$ $\theta = .5$

W^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.096	0.121	0.137	0.157	0.196	0.225
10	0.094	0.118	0.137	0.168	0.201	0.241
15	0.093	0.116	0.135	0.167	0.197	0.235
20	0.093	0.117	0.135	0.167	0.199	0.240
40	0.094	0.118	0.137	0.171	0.204	0.244
50	0.093	0.116	0.134	0.166	0.198	0.242
100	0.093	0.116	0.135	0.166	0.201	0.243
∞	0.093	0.117	0.136	0.169	0.202	0.246

U^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.091	0.115	0.133	0.152	0.186	0.222
10	0.089	0.112	0.130	0.161	0.193	0.232
15	0.088	0.111	0.129	0.159	0.188	0.226
20	0.088	0.111	0.129	0.159	0.191	0.231
40	0.090	0.112	0.130	0.162	0.194	0.234
50	0.088	0.110	0.127	0.158	0.189	0.230
100	0.088	0.110	0.128	0.158	0.192	0.232
∞	0.088	0.111	0.129	0.161	0.193	0.235

A^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.610	0.757	0.836	1.081	1.284	1.645
10	0.636	0.783	0.905	1.120	1.337	1.641
15	0.637	0.788	0.906	1.105	1.313	1.608
20	0.639	0.789	0.902	1.110	1.326	1.609
40	0.652	0.805	0.921	1.128	1.347	1.614
50	0.646	0.797	0.911	1.120	1.347	1.613
100	0.652	0.795	0.916	1.122	1.339	1.600
∞	0.655	0.806	0.925	1.132	1.342	1.625

Table 3.6: Monte Carlo percentage points for the Cramér-von Mises test statistics testing for the binomial distribution with estimated success probability $\theta = .5$, and number of trials $K = 20$. The asymptotic points are shown for comparison. (continued)

$K = 20$ $\theta = .5$

W_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.755	0.941	1.034	1.282	1.531	1.842
10	0.769	0.952	1.090	1.339	1.590	1.914
15	0.765	0.945	1.084	1.324	1.560	1.870
20	0.770	0.945	1.086	1.337	1.584	1.892
40	0.782	0.967	1.104	1.363	1.621	1.934
50	0.771	0.948	1.094	1.345	1.602	1.923
100	0.777	0.955	1.094	1.350	1.611	1.922
∞	0.779	0.964	1.111	1.365	1.621	1.967

A_m^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	5.916	7.631	10.068	16.586	28.239	55.986
10	7.575	10.257	13.578	24.027	41.820	84.035
15	8.349	11.523	15.222	26.157	43.964	80.888
20	8.984	12.156	16.014	26.474	44.892	86.868
40	10.452	14.166	18.795	28.542	46.034	110.188
50	10.848	15.134	19.613	29.580	47.834	111.561
100	12.130	15.733	19.247	32.628	55.947	85.529
∞	19.420	22.290	24.440	27.950	31.350	35.740

3.4 Power Comparisons

We now consider the power of the Cramér-von Mises statistics for testing for the binomial distribution in the more common situation when the probability of success is estimated from the sample. Where possible, calculations have been made of asymptotic power for the Cramér-von Mises statistics and other tests of fit. These have been supplemented by simulations to determine the relative powers for finite samples.

3.4.1 The Test Statistics

The test statistics compared are the following:

1. The Cramér-von Mises statistics.

2. The Kolmogorov-Smirnov statistic. The Kolmogorov-Smirnov statistic is a popular goodness-of-fit statistic for continuous distributions although it has been shown to have poor power relative to the Cramér-von Mises statistics. The statistic has been developed for discrete distributions by Pettitt and Stephens (1977) and for the Poisson distribution by Campbell and Oprian (1979). The statistic is the maximum discrepancy between the cumulative observed and the cumulative expected histogram.

3. The Pearson χ^2_P statistic using $K+1$ cells. This statistic is the most common test of fit for discrete distributions.

4. The likelihood ratio statistic, G^2. This statistic arises as the likelihood ratio test for the multinomial distribution. It is defined

$$G^2 = 2\sum_{i=0}^{K} o_i \ln(o_i/N\hat{p}_i),$$

   where o_i is the observed number in cell i and $\hat{p}_i$ is the estimated probability in cell i (Bishop et al, 1975).

5. The dispersion test. This is the analogue of the Poisson dispersion test and is attributed to Fisher (Kendall and Stuart, Volume 2, 1973).

$$D = \frac{\sum_{j=1}^{N}(x_j - K\hat{\theta})^2}{K\hat{\theta}(1-\hat{\theta})}.$$

This test is often used as a one sided test to detect overdispersed alternatives, but is used here as a two-sided test to guard against all alternative distributions. It has been shown that a test based on D is the score test against the beta-binomial distribution by Potthoff and Whittinghill (1966a) for known probability of success, θ, and by Tarone (1979) when θ is estimated by $\bar{x}/K$. The score test is

$$S = (D - N)/\sqrt{2N(K-1)}. \tag{3.5}$$

6. The k-component smooth test. This has been developed for testing for the binomial distribution by Rayner and Best (1989). For the binomial distribution, the orthonormal functions used in the test are Krawtchouk polynomials. The ith polynomial is defined as follows:

$$h_i(j;\theta,K) = \sum_{t=0}^{i} C_t^j C_{i-t}^{K-j} \left(\frac{-\theta}{(1-\theta)}\right)^{i-t} \left(\frac{(1-\theta)}{\theta}\right)^t / \sqrt{C_i^K} \tag{3.6}$$

where C_x^n is the binomial coefficient for x successes in n trials.

The test statistic is then defined as

$$\hat{S}_k = N^{-1} \sum_{i=2}^{k+1} V_i^2 \tag{3.7}$$

where $V_i = \sum_{j=1}^{N} h_i(x_j;\hat{\theta},K)$. The one-component statistic $\hat{S}_1 = V_2^2$ is similar to the score test against the beta-binomial alternative;

$$V_2 = \sum_{j=1}^{N} \left\{(x_j - K\hat{\theta})^2 + (2\hat{\theta}-1)x_j + K\hat{\theta}(1-\hat{\theta})\right\} / \hat{\theta}(1-\hat{\theta})\sqrt{2K(K-1)},$$

where K is the number of trials. The k-component statistic is equivalent to the Pearson χ^2 statistic. The tests based on the first component, the second component and the sum of the first two components were examined.

7. Generating function statistics.

As was noted for the Poisson distribution, test statistics can be based on the probability generating function. We have adapted the statistic, P, given by Rueda et al (1991) for the binomial distribution. For this distribution, $P(t) = ((1-\theta)+\theta t)^K$, and the computing formula for P becomes:

$$\begin{aligned} P &= 1/n\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{x_i + x_j + 1} - \\ &\quad 2\sum_{i=1}^{N}\sum_{j=0}^{K} C_j^K \theta^j (1-\theta)^{K-j}/(x_i + j + 1) + \frac{1-(1-\theta)^{2K+1}}{\theta(2K+1)} \end{aligned}$$

A statistic, analogous to P, can be based on the moment generating function (MGF). Let $\phi(t)$ be the MGF and $\phi_n(t)$ be the empirical moment generating function. The suggested test statistic is:

$$M = \int_0^1 (\phi_n(t) - \phi(t))^2 dt.$$

For the binomial distribution, $\phi(t) = ((1-\theta) + \theta e^t)^K$, and the computing formula for M is as follows:

$$\begin{aligned} M &= 1/n \sum_{i=1}^N \sum_{j=1}^N \frac{e^{x_i + x_j} - 1}{x_i + x + j} - \\ & 2\sum_{i=1}^N \sum_{j=0}^K C_j^K \theta^j (1-\theta)^{K-j} \frac{e^{x_j + j} - 1}{x_j + j} + \\ & \sum_{j=0}^{2K} C_j^{2K} (1-\theta)^{2K-j} \theta^j (e^j - 1)/j \end{aligned}$$

3.4.2 Asymptotic Power

Calculations of asymptotic power can be made in a similar way to those for the Poisson distribution. These powers were evaluated against the beta-binomial alternative. Let the beta-binomial distribution be defined as follows:

$$Pr\{Y = j\} = C_j^K \frac{\prod_{r=0}^{j-1}(\theta + \gamma r) \prod_{r=0}^{K-j-1}(1 - \theta + \gamma r)}{\prod_{r=0}^{K-1}(1 + \gamma r)} \tag{3.8}$$

for $j = 0, 1, \ldots, K, 0 < \theta < 1, \gamma > 0$. The mean and variance of Y are $K\theta$ and $K\theta(1-\theta)(1 + K\gamma)/(1+\gamma)$, respectively. At $\gamma = 0$, (3.8) reduces to the binomial distribution. Thus H_0 is: $\gamma = 0$. Let θ be estimated by maximum likelihood, that is, $\hat{\theta} = \bar{x}/K$. Under H_1, let $\gamma = \delta/\sqrt{N}$, thus $\gamma \to 0$ as $N \to \infty$ and H_1 reduces to H_0.

For the Pearson χ^2 test the mean of a typical component, s_i is now $\delta \mathbf{w}_i{}'\mathbf{g}$ where $\mathbf{g}$ is the vector with jth element

$$\begin{aligned} g_j &= \frac{\partial Pr\{Y = j\}}{\partial \gamma} \\ &= \frac{K(K-1)}{2}\{\theta p_{j-2}(\theta, K-2) + (1-\theta)p_j(\theta, K-2) - p_j(\theta, K)\} \end{aligned} \tag{3.9}$$

and where $p_j(\theta, K)$ is the binomial probability of observing a count j defined in section 3.2 above, $p_{j-2}(\theta, K-2) = 0$ for $j < 2$ and $p_j(\theta, K-2) = 0$ for $j > K-2$. For the Cramér-von Mises and the modified Cramér-von Mises test statistics the mean of s_i is

$$\delta \mathbf{w}_i{}'\mathbf{M}^{-1/2}\mathbf{A}\mathbf{g} \tag{3.10}$$

where **A** is the partial-sum matrix as before. Again the covariance of **d** under the alternate hypothesis is the same as the covariance under the null hypothesis.

The parameter, δ, is chosen to make the power for the test based on the m.l.e. equal to 0.50, as was done for the Poisson distribution. For an 0.05-level test to give a two-sided power of 0.50, $\delta = 1.96/\sqrt{K(K-1)/2}$. Powers for the Pearson χ^2 test were determined by evaluating the appropriate non-central χ^2 distribution, and powers for the other statistics were obtained by fitting $a + b\chi_p^2$. The asymptotic powers are given in Tables 3.7.

Results and comments

The results of the asymptotic power analysis show that for beta-binomial alternatives, A^2 has the best power among the Cramér-von Mises tests, and is nearly as powerful as the best test. Compared to the two-component smooth test, S_2, A^2 has slightly better power for a small number of trials and slightly lower power for a large number of trials. The modified W^2 test statistic has slightly lower power than A^2, but better than W^2 or U^2. The modified A^2 statistic, A_m^2, has poor power relative to A^2, particularly when the probability of success is different from 0.5. As expected, the power of the Pearson χ^2 test is very high for binomial families with a small number of trials, but is very poor for binomial families with a large number of trials.

Table 3.7: Asymptotic power of the Cramér-von Mises test statistics for testing for the binomial distribution with estimated success probability, θ.

This table gives the asymptotic power (%) of the Cramér-von Mises test for selected values of the number of trials, K, and probability of success, $\theta = .1$ and $\theta = .5$, against a beta-binomial alternative with parameter (γ) chosen to give the locally most powerful test a power of 50%.

$\theta = .1$

Test Statistics

K	W^2	U^2	A^2	W_m^2	A_m^2	χ_P^2	S_2
2	50	50	50	50	50	40	40
3	48	47	47	48	41	34	40
4	46	44	45	45	36	31	40
5	43	41	43	43	33	28	40
6	41	38	41	41	31	26	40
8	36	34	40	38	29	23	40
10	33	30	39	37	27	21	40
12	31	30	38	36	26	19	40
20	28	28	37	35	24	15	40
40	28	28	37	35	23	14	40
50	28	28	37	35	23	11	40

$\theta = .5$

Test Statistics

K	W^2	U^2	A^2	W_m^2	A_m^2	χ_P^2	S_2
2	50	50	50	50	50	40	40
3	34	30	42	40	46	34	40
4	30	29	39	36	44	31	40
5	29	28	38	36	42	28	40
6	29	28	38	35	41	26	40
8	28	28	38	35	39	23	40
10	28	28	38	35	38	21	40
12	28	28	38	35	37	19	40
20	28	28	38	35	34	15	40
40	28	29	38	36	32	14	40
50	28	29	38	36	31	11	40

3.4.3 Finite Samples

Power studies were undertaken for finite samples using simulation. Common alternatives to the binomial distribution can be categorized by the ratio of the variance to the mean; this is equal to $K\theta(1-\theta)$ for the binomial distribution. Distributions with variance larger than the binomial variance are considered *overdispersed*, and those with smaller variance are *underdispersed*.

The most common overdispersed alternative to the binomial distribution is the beta-binomial (BB). This distribution, included in the power study, is a mixture of binomial distributions with common number of trials, K, and with the probability of success, θ, sampled from a beta distribution. This distribution is also referred to as the Polya-Eggenberger or binomial-beta distribution. Other overdispersed distributions examined were the mixture of two binomial random variables, the discrete uniform (DU) and the truncated Poisson distribution (TP). For underdispersed alternatives, the discrete uniform, Hypergeometric (H) and a "subnormal" binomial mixture (SB) (Johnson, Kotz and Kemp, 1992) were examined. The "subnormal" binomial distribution arises when each of the K trials has a different probability of success, and these probabilities are fixed for all samples. Finally, distributions which could have variance equal to that of the binomial distribution (binomial dispersion) were also investigated. The discrete uniform distribution was chosen where possible; otherwise, distributions with dispersion equal to the binomial were constructed.

Comparisons of power for the Cramér-von Mises tests and the other tests of fit, when used in testing against the above alternatives are given in Tables 3.8 - 3.11 for values of the number of trials, K, equal to 5 or 20. One thousand samples of size 20 were generated from each alternative distribution with mean equal to $.1K$ and $.5K$. The finite percentage points of all statistics compared were found by Monte Carlo simulation using 25,000 samples. The maximum standard error of the power results is equal to $.5/\sqrt{1000} \approx 1.6\%$. Random samples were generated using IMSL subroutines (IMSL, 1987).

Results and Comments

1. As expected, the binomial dispersion test, D, and the one-component smooth test, $\hat{S}_1$, perform very well for overdispersed alternatives, with A^2 and χ^2_P only marginally worse.

2. For underdispersed alternatives, D once again has the best power with A^2 close behind.

For these alternatives χ_P^2 has lower power.

3. For alternatives with binomial dispersion, all three Cramér-von Mises statistics have more power than the $\hat{S}_1$ or χ_P^2.

4. Overall, A^2 performs very well as an omnibus test statistic.

Table 3.8: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20 for testing for the binomial distribution with 5 trials. Alternative distributions with a mean, $\mu = K\theta = .5$, corresponding to θ equal to .1 were generated. The binomial variance is .45. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics						
	W^2	U^2	A^2	W_m^2	A_m^2	KS	R
Overdispersed							
BB[$\alpha = .7, \beta = 6.3$] (.675)	267	237	366	299	383	299	24
BB[$\alpha = 1, \beta = 1$] (.9)	511	439	626	532	633	.531	104
.5B(.01)+.5B(.19) (.612)	186	183	252	215	252	210	14
.9B(.05)+.1B(.19) (.9)	441	352	590	458	648	465	98
Underdispersed							
DU[0,1] (.25)	407	407	407	407	124	407	584
SB[.01*4,.46] (.288)	261	261	235	261	55	261	375

	$\hat{S}_1$	$\hat{S}_2$	D_b	X_P^2	G^2	P	M
Overdispersed							
BB[$\alpha = .7, \beta = 6.3$] (.675)	366	382	333	363	313	299	287
BB[$\alpha = 1, \beta = 1$] (.9)	640	649	602	620	574	542	528
.5B(.01)+.5B(.19) (.612)	250	263	217	246	208	195	192
.9B(.05)+.1B(.19) (.9)	627	632	613	641	575	489	463
Underdispersed							
DU[0,1] (.25)	249	249	407	124	407	407	407
SB[.01*4,.46] (.288)	120	120	238	55	214	261	261

Table 3.9: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20 for testing for the binomial distribution with 5 trials. Alternative distributions with a mean, $\mu = K\theta = 2.5$, corresponding to θ equal to .5 were generated. The binomial variance is 1.25. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics						
	W^2	U^2	A^2	W^2_m	A^2_m	KS	R
Overdispersed							
BB[$\alpha = 3.5, \beta = 3.5$] (1.875)	192	182	298	257	377	199	8
BB[$\alpha =, \beta =$] (2.5)	528	482	729	673	823	539	4
.5B(1/3)+.5B(2/3) (1.806)	202	188	292	259	353	201	334
.75B(.4)+.25B(.8) (1.850)	648	604	776	732	822	639	2
DU[0,5] (2.92)	151	158	143	163	66	149	51
TP[$\mu = 3.272, K = 4$] (1.342)	195	182	338	272	428	203	11
Underdispersed							
H[$M=8, m=5, X=4$] (.536)	520	509	540	542	344	490	586
MB[.1,.1,.5,.9,.9] (.61)	447	441	443	464	250	396	522
Binomial Dispersion							
DU[1,4] (1.25)	763	720	913	872	953	779	4
C[.083,0,.417,.417,0,.083] (1.25)	579	567	655	495	345	501	770

	$\hat{S}_1$	$\hat{S}_2$	D_b	X^2_P	G^2	P	M
BB[$\alpha = 3.5, \beta = 3.5$] (1.875)	370	360	352	324	206	334	261
BB[$\alpha =, \beta =$] (2.5)	825	805	819	755	574	746	643
.5B(1/3)+.5B(2/3) (1.806)	334	318	316	298	196	297	234
.75B(.4)+.25B(.8) (1.850)	835	808	818	756	628	723	604
DU[0,5] (2.92)	954	947	951	917	828	900	830
TP[$\mu = 3.272, K = 4$] (1.342)	405	401	391	341	223	310	234
Underdispersed							
H[$M=8, m=5, X=4$] (.536)	715	375	729	275	424	170	5
MB[.1,.1,.5,.9,.9] (.61)	568	216	581	200	358	132	7
Binomial Dispersion							
DU[1,4] (1.25)	1	1	2	174	196	11	39
C[.083,0,.417,.417,0,.083] (1.25)	206	239	205	999	999	237	307

Table 3.10: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20 for testing for the binomial distribution with 20 trials. Alternative distributions with a mean, $\mu = K\theta = 2$, corresponding to θ equal to .1 were generated. The binomial variance is 1.8. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics						
	W^2	U^2	A^2	W^2_m	A^2_m	KS	R
Overdispersed							
BB[$\alpha = 3.7, \beta = 3.7$] (2.7)	153	140	264	237	349	180	43
BB[$\alpha = 7/6, \beta = 63/6$] (4.5)	397	355	589	536	632	420	76
.5B[.05]+.5B[.15] (2.75)	216	200	309	288	327	225	25
.9B[.075]+.1B[.325] (3.94)	370	264	549	455	709	345	285
DU[0,4] (2.0)	257	252	282	290	94	244	119
Underdispersed							
DU[1,3] (2/3)	452	457	671	584	73	498	825
H[$M=40, m=20, X=4$](.923)	740	741	773	766	245	606	917
SB[.05*19,.95] (.902)	262	285	291	313	22	297	278
Binomial Dispersion							
C[.15,.3,.1,.3,.15,0*16] (1.8)	401	408	310	318	19	354	230

	$\hat{S}_1$	$\hat{S}_2$	D_b	X^2_P	G^2	P	M
Overdispersed							
BB[$\alpha = 3.7, \beta = 3.7$] (2.7)	349	339	301	319	230	243	205
BB[$\alpha = 7/6, \beta = 63/6$] (4.5)	663	641	630	599	507	540	442
.5B[.05]+.5B[.15] (2.75)	358	353	317	295	220	300	268
.9B[.075]+.1B[.325] (3.94)	649	684	623	691	573	349	214
DU[0,4] (2.0)	39	203	25	157	232	219	255
Underdispersed							
DU[1,3] (2/3)	899	589	948	68	917	876	582
H[$M=40, m=20, X=4$](.923)	835	630	875	121	661	608	246
SB[.05*19,.95] (.902)	369	302	422	21	291	569	516
Binomial Dispersion							
C[.15,.3,.1,.3,.15,0*16] (1.8)	10	69	12	91	316	66	85

Table 3.11: Power Comparison
This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20 for testing for the binomial distribution with 20 trials. Alternative distributions with a mean, $\mu = K\theta = 10$, corresponding to θ equal to .5 were generated. The binomial variance is 5. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics						
	W^2	U^2	A^2	W^2_m	A^2_m	KS	R
Overdispersed							
BB[$\alpha = 18.5, \beta = 18.5$] (7.5)	147	139	282	222	335	136	32
BB[$\alpha = 35/6, \beta = 35/6$] (12.5)	581	565	821	764	849	554	15
.5B[.4]+.5B[.6] (10.93)	336	328	497	444	450	301	18
.75B[.45]+.25B[.65] (7.85)	181	181	340	256	367	173	28
DU[6,14] (20/3)	296	298	341	353	76	280	71
DU[5,15] (10.0)	684	668	824	801	649	629	45
TP[$\mu = 10.019, K = 20$] (9.81)	319	311	596	493	635	302	31
Underdispersed							
DU[8,12] (2.0)	329	357	446	447	0	376	357
H[$M=30, m=20, X=15$] (1.72)	708	723	706	731	4	527	220
SB[.1*10,.9*10] (1.8)	711	719	699	724	5	548	251
Binomial Dispersion							
C[0*7,.2,.15,.1,.1,.1,.15,.2,0*7] (5.0)	323	338	312	338	5	284	432

Table 3.11: Power Comparison (continued)

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20 for testing for the binomial distribution with 20 trials. Alternative distributions with a mean, $\mu = K\theta = 10$, corresponding to θ equal to .5 were generated. The binomial variance is 5. All tests are at the 5% level.

Alternative Distribution (σ^2)	Test Statistics					
	$\hat{S}_1$	$\hat{S}_2$	D_b	X_P^2	G^2	P
Overdispersed						
BB[$\alpha = 18.5, \beta = 18.5$] (7.5)	336	352	293	300	230	309
BB[$\alpha = 35/6, \beta = 35/6$] (12.5)	879	876	861	813	752	856
.5B[.4]+.5B[.6] (10.93)	555	514	506	392	360	509
.75B[.45]+.25B[.65] (7.85)	397	410	359	330	258	314
DU[6,14] (20/3)	122	79	89	105	250	61
DU[5,15] (10.0)	815	770	778	534	606	743
TP[$\mu = 10.019, K = 20$] (9.81)	656	663	623	606	514	570
Underdispersed						
DU[8,12] (2.0)	700	65	800	1	197	342
H[$M=30, m=20, X=15$] (1.72)	825	305	879	3	177	597
SB[.1*10,.9*10] (1.8)	764	264	818	3	197	548
Binomial Dispersion						
C[0*7,.2,.15,.1,.1,.1,.15,.2,0*7] (5.0)	1	0	1	104	391	5

3.5 Example

Table 3.12 records the data from a dice throwing experiment due to Weldon, discussed by Pearson (1900) and presented in Rayner and Best (1989). The data show the number of occurrences of a 5 or 6 on any die from a throw of 12 dice repeated 26,306 times. The sample mean and variance are 4.044 and 2.698, respectively, and the estimated probability of success is .3377. Clearly, the probability of success is 1/3 for a fair die, so this data could be tested for a binomial distribution with known probability of success equal to 1/3. For this reason expected frequencies are given for both known probability of success and for the probability of success equal to the estimated value, .3377. We define the *standardized difference* as the difference between the observed and expected values divided by the standard deviation. Asymptotically, the standardized difference has a standard normal distribution. The cumulative observed and expected histograms are found in Figures 3.1 and 3.3 for known and estimated success probabilities, respectively. The standardized differences are plotted in Figures 3.2 and 3.4. The values and significance levels of the Cramér-von Mises statistics and other test statistics are given in Table 3.13.

The Cramér-von Mises statistics and the Pearson χ^2 test reject the hypothesis of a binomial distribution with probability of success equal to 1/3, but do not reject the binomial hypothesis when the parameter can be estimated from the data.

Table 3.12: Weldon Dice Data

No. of 5 or 6	Frequency	Cum. Frequency	Cum. Expected p=1/3	Std. Difference	Cum. Expected p=.3377	Std. Difference
0	185	185	202.75	-1.251	187.38	-0.174
1	1149	1334	1419.25	-2.326	1333.89	0.003
2	3265	4599	4764.61	-2.651	4549.13	0.813
3	5475	10074	10340.22	-3.361	10013.83	0.764
4	6114	16188	16612.78	-5.429	16283.18	-1.208
5	5194	21382	21630.83	-4.013	21397.83	-0.251
6	3067	24449	24558.03	-2.699	24440.37	0.207
7	1331	25780	25812.54	-1.479	25770.10	0.432
8	403	26183	26204.58	-2.147	26193.86	-1.028
9	105	26288	26291.69	-0.977	26289.89	-0.471
10	14	26302	26304.76	-2.483	26304.58	-2.166
11	4	26306	26305.95	0.222	26305.94	0.241
12	0	26306	26306.00		26306.00	

Table 3.13: Test statistics for the Weldon dice data

	Known Parameter		Estimated Parameter	
Test Statistic	Value	Significance Level	Value	Significance Level
W^2	2.85	<.001	0.12	0.13
U^2	0.58	<.001	0.12	0.12
A^2	14.64	<.001	0.60	0.30
W_m^2	13.75	<.001	0.60	0.22
A_m^2	92.14	<.001	9.05	0.42
D			26646	0.54
χ_P^2	41.31	<.001	13.16	0.36

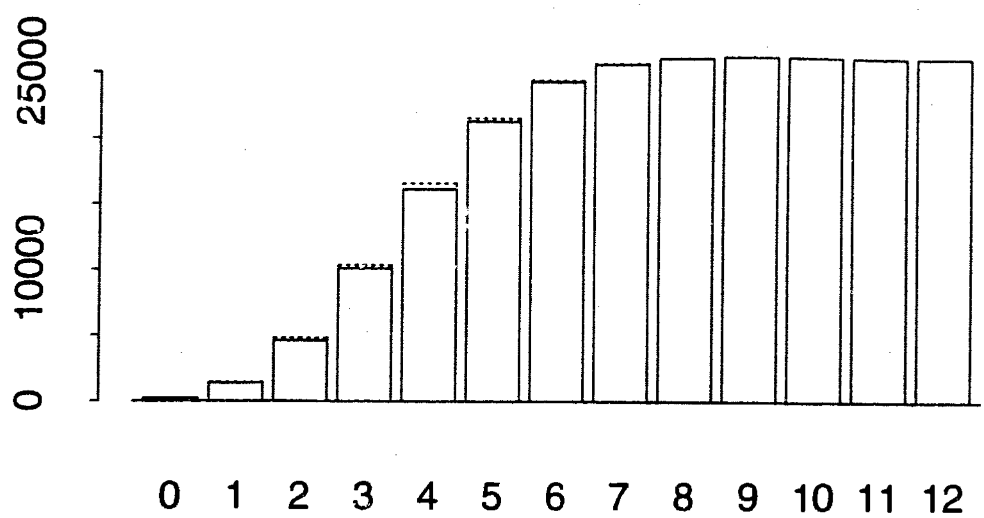

Figure 3.1: Cumulative observed (—) and expected (- -) histograms for Weldon's dice data with (p=1/3).

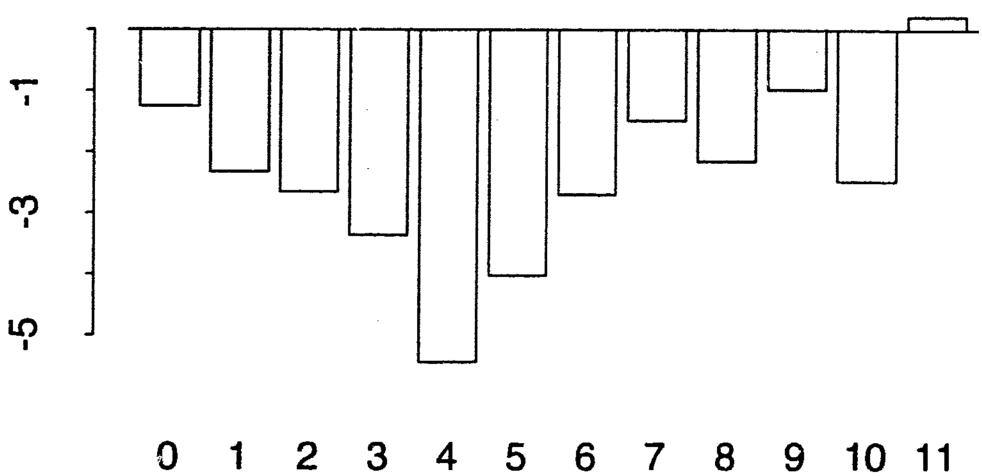

Figure 3.2: Standardized difference between the observed and expected histograms for Weldon's dice data with (p=1/3).

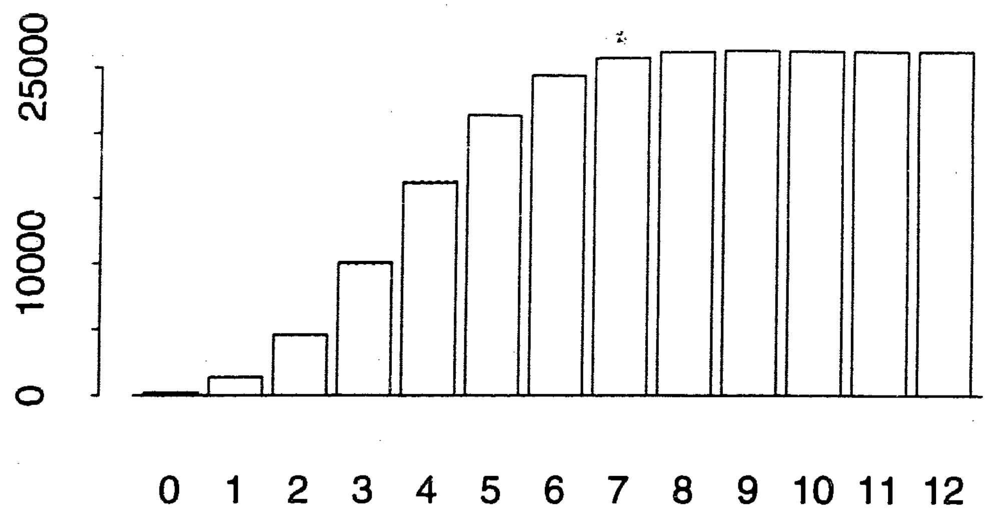

Figure 3.3: Cumulative observed (—) and expected (- -) histograms for Weldon's dice data with ($\hat{p}$=.3377).

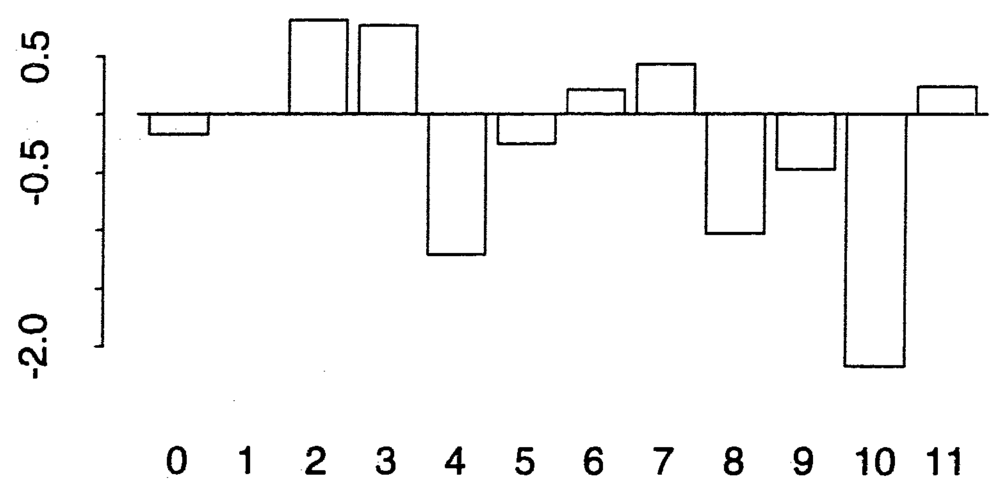

Figure 3.4: Standardized difference between the observed and expected histograms for Weldon's dice data with ($\hat{p}$=.3377).

Chapter 4

Discrete Uniform Distribution

4.1 Introduction

The discrete uniform distribution with K cells is the distribution for which $p_j = 1/K$ for $j = 1, \ldots, K$. Choulakian, Lockhart and Stephens (1994) have discussed Cramér-von Mises statistics for the discrete uniform distribution, and have found analytically the eigenvalues and eigenvectors needed for the asymptotic distributions. The authors also discuss the components of the various statistics and give the asymptotic percentage points for each test statistic. In this chapter, asymptotic power comparisons are made following the lines of the comparisons for the Poisson and binomial distributions.

4.2 Power Comparisons

4.2.1 Lehmann Alternative

The asymptotic powers of the Cramér-von Mises test statistics were examined against the Lehmann alternative to the discrete uniform distribution. The Lehmann alternative is defined as follows:

$$Pr\{Y \leq j\} = (j/K)^{\gamma} \tag{4.1}$$

for $j = 1, \ldots, K, \gamma > 0$. At $\gamma = 1$, (4.1) reduces to the discrete uniform distribution. Thus H_0 is: $\gamma = 1$. Under H_1, let $\gamma = 1 + \delta/\sqrt{N}$; thus as $N \rightarrow \infty$, H_1 becomes H_0.

The i-th component of a test statistic, $s_i = (\mathbf{w}_i'\mathbf{X})$, now has mean $\delta\mathbf{w}_i'\mathbf{Ag}$ where $\mathbf{A}$ is the partial-sum matrix, and $\mathbf{g}$ is the vector with jth element

$$g_j = \frac{\partial Pr\{Y=j\}}{\partial\gamma} = \frac{j}{K}\ln\left(\frac{j}{K}\right) - \frac{j-1}{K}\ln\left(\frac{j-1}{K}\right).$$

The variance of the maximum likelihood test for the discrete uniform distribution against the Lehmann alternative is the inverse of the Cramér-Rao lower bound. As before, the parameter, δ, is chosen to make the power for the test based on the m.l.e. equal to 0.5. For a 0.05-level test to give a two-sided power of 0.50, $\delta = 1.96/\sqrt{J}$, where J is the Cramér-Rao lower bound

$$E\left(\left.\frac{\partial^2\log f}{\partial\gamma^2}\right|_{\gamma=0}\right) = K^{-1}\textstyle\sum_{j=1}^{K} j^2(j-1)[\log j - \log(j-1)]^2.$$

The asymptotic powers for the Cramér-von Mises test statistics, the first two components and the sum of the first two components of A^2, and the Pearson χ^2 statistic, are given in Table 4.1. Also included in the table are the results for the continuous uniform distribution corresponding to $K = \infty$.

Results and comments

Against Lehmann alternatives, A^2 has the best power among the Cramér-von Mises tests, and is nearly as powerful as the best test. As expected, the power of the Pearson χ^2 test is very high for distributions with a small number of cells, but is very poor for distributions with a large number of cells. The power of the first component of A^2 is slightly lower than that of A^2. The power of the second component is negligible reflecting the fact the Lehmann alternative is primarily a shift in the mean. The power of the sum of the first two components gives power less than the first component reflecting the utilization of a component with very little power.

4.2.2 One Parameter Beta-Binomial Alternative

The asymptotic powers of the Cramér-von Mises test statistics were also examined against the one parameter beta-binomial alternative. Let the one parameter beta-binomial distribution be defined as follows:

$$Pr\{Y=j+1\} = \frac{\Gamma(l+1)}{\Gamma(j+1)\Gamma(l-j+1)}\frac{\Gamma(j+\gamma)\Gamma(l-j+\gamma)}{\Gamma(l+2\gamma)}\frac{\Gamma(2\gamma)}{[(\Gamma(\gamma)]^2} \qquad (4.2)$$

for $j = 0, 1, \ldots, l, \gamma > 1, l = K - 1$ where K is the total number of cells for which the distribution has non-zero probability. Thus, the distribution of Y is over the same range as that of the discrete uniform distribution over $1, 2, \ldots, K$. The mean and variance of Y are $l/2 + 1 = (K+1)/2$ and $(K-1)[2\gamma + (K-1)]/4(2\gamma+1)$, respectively. At $\gamma = 1$, (4.2) reduces to the discrete uniform distribution. Thus H_0 is: $\gamma = 1$. Under H_1, let $\gamma = 1 + \delta/\sqrt{N}$, so that in the limit H_1 approaches H_0.

For a typical component the mean is $\delta \mathbf{w}_i{}'\mathbf{A}\mathbf{g}$ where $\mathbf{A}$ is the partial-sum matrix, and $\mathbf{g}$ is the vector with jth element

$$\begin{aligned} g_j &= \frac{\partial Pr\{Y = j\}}{\partial \gamma} \\ &= \{2[\Psi(2) - \Psi(1) - \Psi(K)] + \Psi(K - i) + \Psi(i+1)\}/K \end{aligned}$$

where $\Psi(a)$ is the digamma function defined $\Psi(a) = \Gamma(a)'/\Gamma(a)$.

Again the parameter, δ, is chosen to make the power of the test based on the m.l.e. equal to 0.50. For a 0.05-level test to give a two-sided power of 0.50, $\delta = 1.96/\sqrt{J}$, where J is the Cramér-Rao lower bound

$$E\left(\left. \frac{\partial^2 \log f}{\partial \gamma^2} \right|_{\gamma=0} \right) = K^{-1} \sum_{j=0}^{l} \{\Psi'(j+1) + \Psi'(l-j+1) + C\}.$$

Here Ψ' is the trigamma function, and

$$C = 4\Psi'(2) - 4\Psi'(l+2) - 2\Psi'(1).$$

The asymptotic powers for the Cramér-von Mises test statistics, the first two components and the sum of the first two components of A^2, and the Pearson χ^2 statistic, are given in Table 4.2. Also included in the table are the results for the continuous uniform distribution corresponding to $K = \infty$, for which value the beta-binomial alternative becomes a one-parameter beta distribution.

Results and comments

The results of the asymptotic power analysis show that against one-parameter beta-binomial alternatives, U^2 has the best power among the Cramér-von Mises tests, and clearly greater power than Pearson χ^2. None of the test statistics examined had power approaching that of the best test. The power of the Pearson χ^2 test is highest for discrete uniform distributions with a moderate number of cells (5 or 6), but is very poor for discrete uniform distributions with a large number of cells. The power of the first component of A^2 gives no power as the mean of the alternative distribution is identical to that of the null distribution. The power of the second component is nearly as large as that of the best test reflecting the fact that the beta-binomial alternative gives primarily a shift in the variance. The power of the sum of the first two components is less than that of the second component reflecting the inclusion of a component with very little power.

The asymptotic power of the Cramér-von Mises statistics is overall very poor for one-parameter beta-binomial alternatives. The eigenvalues for W^2 and A^2, and thus the relative weights given to each component are decreasing with increasing i. Therefore, the largest weight is given to the first component, the next largest weight to the next component, etc. The first component offers no power, since the alternative distribution has the same mean as the null distribution. Any statistic which gives relatively higher weight to the second component (variance) will have higher power. U^2 gives identical weight to pairs of components, thus giving the second component a higher weight than A^2.

The examination of the asymptotic power of tests of fit for the uniform distribution against the one-parameter beta-binomial distribution suggests that no one test statistic will be most powerful against all alternatives.

Table 4.1: Asymptotic power of the Cramér-von Mises test statistics for testing for the discrete uniform distribution.

This table gives the asymptotic power (%) of the Cramér-von Mises statistics, the first component of A^2, S_1, the second component of A^2, S_2, and the Pearson χ^2 statistic for selected values of the number of cells, K, against a Lehmann alternative with parameter (γ) chosen to give the locally most powerful test a power of 50%.

	Test Statistics						
K	W^2	U^2	A^2	S_1	S_2	S_1+S_2	χ^2_P
2	50	50	50	50	–	–	50
3	48	40	48	48	08	40	40
4	46	34	47	47	09	40	34
5	45	31	46	45	10	40	31
6	44	28	45	44	10	39	28
8	43	25	45	43	11	39	24
10	42	23	44	43	11	38	21
12	42	23	44	43	11	38	20
20	40	21	43	41	11	38	16
40	40	20	42	40	11	37	11
50	39	20	42	40	11	37	11
∞	39	19	41	40	11	36	–

Table 4.2: Asymptotic power of the Cramér-von Mises test statistics for testing for the discrete uniform distribution.

This table gives the asymptotic power (%) of the Cramér-von Mises statistics, the first component of A^2, S_1, the second component of A^2, S_2, and the Pearson χ^2 statistic for selected values of the number of cells, K, against a one parameter beta-binomial alternative with parameter (γ) chosen to give the locally most powerful test a power of 50%.

	Test Statistics						
K	W^2	U^2	A^2	S_1	S_2	S_1+S_2	χ^2_P
3	08	14	08	05	19	14	14
4	10	25	11	05	33	26	22
5	11	30	14	05	41	32	24
6	12	33	15	05	44	35	24
8	11	34	16	05	47	37	23
10	11	33	16	05	47	38	21
12	11	33	16	05	47	38	19
20	10	31	16	05	46	37	16
40	10	29	16	05	45	36	12
50	10	29	15	05	44	35	11
∞	09	25	14	05	41	32	–

Chapter 5

Regression Models

5.1 Introduction

In this chapter, the empirical process and Cramér-von Mises statistics are introduced and developed for tests on variables from a discrete distribution which are independent but not identically distributed. (i.n.i.d.). In section 5.2 the definitions of the empirical processes are given. The distributions of the empirical processes and the respective Cramér-von Mises statistics are shown in section 5.3 for known parameters and in section 5.4 for unknown parameters. The theory is illustrated for Poisson regression, logistic regression and complementary log-log regression and some percentage points are given in section 5.6. In section 5.7, power comparisons are given for testing for Poisson regression. Finally, examples are presented to illustrate the techniques.

5.2 Definitions

The definitions of the empirical process and Cramer-von Mises statistics for continuous i.i.d. variables, continuous i.n.i.d. variables and discrete i.i.d. variables will first be reviewed.

I.I.D. Continuous Variables

Let $y_1, y_2, \ldots, y_N$ be a sample of independent and identically distributed observations with continuous distribution function $G(y)$. The hypothesis to be tested is that $G(y) = F(y)$

where $F(y)$ is completely specified. The *empirical process* of the sample is defined as

$$Y_N(y) = \sqrt{N}[F_N(y) - F(y)], \quad -\infty < y < \infty$$

where $F_N(y)$ is the proportion of y_i less than or equal to y, the *empirical distribution function*, and $F(y)$ is the distribution function. For continuous distributions, if Y has distribution $F(y)$, then $U = F(Y)$ is distributed as a uniform random variable on [0,1]. This is referred to as the *probability integral transformation.* Also, let $u_i = F(y_i)$ and $U_N(t)$ be the proportion of u_i less than or equal to t. Statistics based on the empirical process $Y_N(y)$ are equivalent to statistics based on the process

$$Z_N(t) = \sqrt{N}\{U_N(t) - t\} \qquad 0 \le t \le 1.$$

In particular, Cramér-von Mises statistics for testing fully specified continuous distributions are defined as follows:

$$W^2 = \int_0^1 [Z_N(t)]^2 dt, \tag{5.1}$$

$$U^2 = \int_0^1 [Z_N(t) - \overline{Z_N}]^2 dt, \tag{5.2}$$

$$A^2 = \int_0^1 [Z_N(t)]^2/[t(1-t)]\} dt, \tag{5.3}$$

where $\overline{Z_N} = \int_0^1 Z_N(t)dt$. The weight function for A^2, $1/t(1-t)$, is the inverse of the variance of the process $Z_N(t)$, at t.

Let $u_{(1)}, u_{(2)}, \ldots, u_{(N)}$ be the u_i arranged in ascending order, and $\bar{U} = \sum_{i=1}^N u_i/N$. The computing formulas of the Cramér-von Mises statistics can be written

$$W^2 = N\textstyle\sum_{i=1}^N \{u_{(i)} - (2i-1)/(2N)\}^2 + 1/(12N), \tag{5.4}$$

$$U^2 = W^2 - N(\bar{U} - .5)^2, \tag{5.5}$$

$$A^2 = -N - (1/N)\textstyle\sum_{i=1}^N (2i-1)[\ln u_{(i)} + \ln\{1 - u_{(N+1-i)}\}]. \tag{5.6}$$

Independent Non-Identically Distributed Continuous Variables

To extend the discussion to the i.n.i.d. case, let $y_1, y_2, \ldots, y_N$ be a sample of independent observations with continuous distribution functions. Suppose the null hypothesis is that $F(y_i; \gamma_i)$ is the distribution of y_i. We refer to $F(y; \gamma_i)$ as $F_i(y)$.

Let $1\{L\}$ be the indicator function taking the value 1 when L is true and 0 otherwise. The empirical distribution function, $F_N(y)$ can be written as a sum of indicator functions.

$$F_N(y) = (1/N)\sum_{i=1}^{N} 1\{y_i \leq y\}.$$

The expected value of $1\{y_i \leq y\}$ is $Pr(Y_i \leq y) = F_i(y)$. Let $U_i = F_i(Y_i)$; then each random variable, U_i, is distributed uniformly on [0,1]. Also define $u_i = F_i(y_i)$; then

$$\begin{aligned} Y_N(t) &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{u_i \leq t\} - Pr(U_i \leq t) \\ &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{u_i \leq t\} - t]\}, \\ &= \sqrt{N}\{U_N(t) - t\}, \qquad 0 \leq t \leq 1. \end{aligned} \tag{5.7}$$

The process, $Y_N(t)$, is referred to as the residual process.

The definitions and computing formulas of the Cramér-von Mises statistics for testing for i.n.i.d. continuous distributions are identical to those given in the previous section with the uniform process $Z_N(t)$ replaced by the residual process.

I.I.D. Discrete Variables

Let $y_1, y_2, \ldots, y_N$ be a sample of independent and identically distributed observations with discrete distribution function $G(y)$, and let the null hypothesis be: $G(y) = F(y)$, where $F(y)$ is completely specified. For discrete distributions, the empirical process is the same as for continuous distributions but the range of the variable is now discrete; thus

$$Y_N(j) = \sqrt{N}[F_N(j) - F(j)], \quad j = 0, 1, 2, \ldots$$

The transformation $U = F(Y)$ can be made but U will not now be distributed as a uniform random variable. Let $U_N(t)$ be defined as above; then

$$Z_N(t) = \sqrt{N}\{U_N(t) - Pr(U \leq t)\}.$$

Note that $Pr(U \leq t) \leq t$, with equality if and only if t is in the closure of the range of F.

It will be useful to recall the notation introduced in 2.2. Let p_j be the probability of observing a count j; for simplicity, the sample space will be assumed to be the integers from 1 to K, where K can be infinite.

Suppose N independent observations are given; let o_j be the observed number of outcomes j, and let $Np_j = e_j$ be the expected number in cell j. Let $S_j = \sum_{i=1}^{j} o_i$, $T_j = \sum_{i=1}^{j} e_i$

and $H_j = \sum_{i=1}^{j} p_i$, and define $Z_j = S_j - T_j$, $j = 1, 2, \ldots, K$. The Cramér-von Mises statistics W^2, U^2, A^2 and W_m^2 for discrete distributions are then

$$\begin{aligned} W^2 &= N^{-1}\sum_{j=1}^{K} Z_j^2 p_j, & (5.8)\\ U^2 &= N^{-1}\sum_{j=1}^{K} (Z_j - \bar{Z})^2 p_j, & (5.9)\\ A^2 &= N^{-1}\sum_{j=1}^{K} Z_j^2 p_j / \{H_j(1 - H_j)\}, & (5.10)\\ W_m^2 &= N^{-1}\sum_{j=1}^{K} Z_j^2, & (5.11) \end{aligned}$$

where $\bar{Z} = \sum_{j=1}^{K} Z_j p_j$.

These statistics can also be expressed as a weighted sum of the empirical process:

$$\begin{aligned} W^2 &= \sum_{j=1}^{K} Y_N^2(j) p_j, & (5.12)\\ U^2 &= \sum_{j=1}^{K} (Y_N(j) - \overline{Y_N})^2 p_j, & (5.13)\\ A^2 &= \sum_{j=1}^{K} Y_N^2(j) p_j / \{H_j(1 - H_j)\}, & (5.14)\\ W_m^2 &= \sum_{j=1}^{K} Y_N^2(j), & (5.15) \end{aligned}$$

where $\overline{Y_N} = \sum_{j=1}^{K} Y_N(j) p_j$.

I.N.I.D. Discrete Variables- Empirical Processes

There are four possible residual processes which can be examined for i.n.i.d. discrete variables. Let $F(y_i; \gamma_i)$ be the distribution function of observation y_i, $i = 1, \ldots, N$. It will be supposed that these distributions are from the same parametric family (for example, the Poisson family), but each y_i has a different γ_i. However, several random variables could have the same distribution (e.g. the Poisson with the same mean). Let M be the number of different distributions and N_i the number of observations from the ith distribution. Let $y_{il}, l = 1, \ldots, N_i$ be the observations with distribution i, $i = 1, \ldots, M$. The four residual processes will now be defined.

1. Untransformed. Each individual y_i is compared to F_i on the original scale.

$$\begin{aligned} Y_{1,N}(j) &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{y_i \le j\} - F_i(j)\},\\ &= \frac{\sqrt{N}}{N}\sum_{i=1}^{M}\{[\sum_{l=1}^{N_i} 1\{y_{il} \le j\}] - F_i(j)\},\\ &= \sqrt{N}\{F_N(j) - \bar{F}(j)\} \qquad j = 1, 2, \ldots, K \end{aligned}$$

where $\bar{F}(j) = N^{-1}\sum_{i=1}^{M} N_i F_i(j)$. This process effectively compares the empirical distribution function with the average distribution function.

2. F-transformed. Let $U_i = F_i(Y_i)$, and $u_i = F_i(y_i)$, where $F_i(j) = Pr(Y_i \leq j)$. Then

$$\begin{aligned} Y_{2,N}(t) &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{u_i \leq t\} - Pr(U_i \leq t)\} \\ &= \sqrt{N}\{U_N(t) - \overline{P_{U_i}(t)}\}, \qquad 0 \leq t \leq 1 \end{aligned}$$

where $\overline{P_{U_i}(t)} = N^{-1}\sum_{i=1}^{M} N_i Pr[U_i \leq t]$. This process transforms each observation to the uniform scale and compares it to its expected value, and is the natural analogue of the residual process for continuous random variables.

3. G-transformed. Let $V_i = G_i(Y_i)$, and $v_i = G_i(y_i)$, where $G_i(j) = Pr(Y_i < j)$. Then

$$\begin{aligned} Y_{3,N}(t) &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{v_i \leq t\} - Pr(V_i \leq t)\}, \\ &= \sqrt{N}\{V_N(t) - \overline{P_{V_i}(t)}\}, \qquad 0 \leq t \leq 1 \end{aligned}$$

where $\overline{P_{V_i}(t)} = N^{-1}\sum_{i=1}^{M} N_i Pr[V_i \leq t]$. The reason for examining this process will become apparent when the test statistics are defined.

4. Random-transformed. Let $U_i^* = F_i^*(Y_i) = F_i(Y_i) - \delta_i p_i(Y_j)$, where δ_i is a random variable distributed as uniform on [0,1], and $p_i(j) = Pr(y_i = j)$. Then, the variable $U^* = F^*(Y)$ is distributed uniformly on [0,1]. Also let $u_i^* = F_i^*(y_i) = F_i(y_i) - \delta_i p_i(y_i)$. Define

$$\begin{aligned} Y_{4,N}(t) &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{u_i^* \leq t\} - Pr(U_i^* \leq t) \\ &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{u_i^* \leq t\} - t]\}, \\ &= \sqrt{N}\{U_N^*(t) - t\}, \qquad 0 \leq t \leq 1. \end{aligned}$$

This process is equivalent to the residual process $Y_N(t)$ for continuous random variables.

Note that when all the F_i are continuous, $Y_{2,N}(j), Y_{3,N}(t)$ and $Y_{4,N}(t)$ are equivalent.

I.N.I.D. Discrete Variables- Cramér-von Mises Statistics

A set of Cramér-von Mises statistics, W^2, U^2, A^2, and W_m^2, were developed based on each of the above processes as follows.

1. The following statistics come from applying the definition of the discrete Cramér-von Mises statistics given in (5.12 - 5.15) to the untransformed process $Y_{1,N}(j)$:

$$W_u^2 = \sum_{j=1}^{K}[Y_{1,N}(j)]^2\overline{p(j)}, \quad (5.16)$$

$$U_u^2 = \sum_{j=1}^{K}[Y_{1,N}(j) - \overline{Y_{1,N}})]^2\overline{p(j)}, \quad (5.17)$$

$$A_u^2 = \sum_{j=1}^{K}[Y_{1,N}(j)]^2\overline{p(j)}/\{\overline{H(j)}(1 - \overline{H(j)})\}, \quad (5.18)$$

$$W_{mu}^2 = \sum_{j=1}^{K}[Y_{1,N}(j)]^2, \quad (5.19)$$

where $\overline{p(j)} = \sum_{i=1}^{M} p_i(j)$, $\overline{Y_{1,N}} = \sum_{j=1}^{K} Y_{1,N}(j)\overline{p(j)}$ and $\overline{H(j)} = \sum_{i=1}^{j}\overline{p(i)}$. When $F_i = F$ for all i these test statistics are identical to those defined for a common discrete distribution.

2. The F-transformed process leads to statistics similar to the Cramér-von Mises statistics for continuous distributions. The test statistics are the integral of the squared process over [0,1].

$$W_f^2 = \int_0^1 [Y_{2,N}(t)]^2 dt, \quad (5.20)$$

$$U_f^2 = \int_0^1 [Y_{2,N}(t) - \overline{Y_{2,N}}]^2 dt, \quad (5.21)$$

$$A_f^2 = \int_0^1 [Y_{2,N}(t)]^2/[\{w(t)[1 - w(t)\}]\}dt, \quad (5.22)$$

where $\overline{Y_{2,N}} = \int_0^1 Y_{2,N}(t)dt$. For A^2, the weight function, $1/w(t)[1-w(t)]$, is the inverse of the variance of the process at t, with

$$w(t) = \overline{P_{U_i}(t)}.$$

3. The G-transformed process leads to statistics similar to that of the F-transformed process.

$$W_g^2 = \int_0^1 [Y_{3,N}(t)]^2 dt, \quad (5.23)$$

$$U_g^2 = \int_0^1 [Y_{3,N}(t) - \overline{Y_{3,N}}]^2 dt, \quad (5.24)$$

$$A_g^2 = \int_0^1 [Y_{3,N}(t)]^2/[\{w(t)[1 - w(t)\}]\}dt, \quad (5.25)$$

where $\overline{Y_{3,N}} = \int_0^1 Y_{3,N}(t)dt$, and

$$w(t) = \overline{P_{V_i}(t)}.$$

In order to compute the test statistics based on the F- and G-transformed processes, numerical integration must be used.

4. The random-transformed process is equivalent to the continuous empirical process on [0,1], and the test statistics are identical to the Cramér-von Mises statistics for continuous distributions.

$$W_r^2 = \int_0^1 [Y_{4,N}(t)]^2 dt, \tag{5.26}$$

$$U_r^2 = \int_0^1 [Y_{4,N}(t) - \overline{Y_{4N}}]^2 dt, \tag{5.27}$$

$$A_r^2 = \int_0^1 [Y_{4,N}(t)]^2/[t(1-t)]\}dt, \tag{5.28}$$

where $\overline{Y_{4,N}} = \int_0^1 Y_{4,N}(t)dt$.

The G-transformed process was examined because when $F_i = F$, for all i, and F is discrete, $W_g^2 = W^2$, where W^2 is the statistic defined for i.i.d. discrete distributions given in (5.8); however, $W_f^2 \neq W^2$. Similar results hold for U_g^2 and A_g^2. To see this, let $S_j = \sum_{i=1}^{j} o_i/N$ and $j = 1, 2, \ldots, K$, and observe the following identities:

1.

$$U_N(t) = N^{-1}\sum_{i=1}^{N} 1\{F(y_i) \leq t\} = \begin{cases} S_{j-1}/N & \text{if } F(j-1) \leq t < F(j) \\ 1 & \text{if } F(y_{(N)}) \leq t \leq 1 \end{cases}$$

for $j = 1, 2, \ldots, K$, where $S_0 = F(0) = 0$, and $y_{(N)}$ is the largest observation.

2.

$$V_N(t) = N^{-1}\sum_{i=1}^{N} 1\{G(y_i) \leq t\} = \begin{cases} S_j/N & \text{if } G(j) \leq t < G(j+1) \\ 1 & \text{if } G(y_{(N)}) \leq t \leq 1 \end{cases} = \begin{cases} S_j/N & \text{if } F(j-1) \leq t < F(j) \\ 1 & \text{if } F(y_{(N)}-1) \leq t \leq 1 \end{cases}$$

since $G(j) = F(j-1)$.

3.

$$\begin{aligned} P_U(t) &= Pr[F(Y) \leq t] \\ &= \begin{cases} 0 & \text{if } 0 \leq t < F(1) \\ F(j-1) & \text{if } F(j-1) \leq t < F(j) \\ 1 & \text{if } t = 1. \end{cases} \end{aligned}$$

4.

$$\begin{aligned} P_V(t) &= Pr[G(Y) \leq t] \\ &= \begin{cases} F(j) & \text{if } F(j-1) \leq t < F(j) \\ 1 & \text{if } t = 1. \end{cases} \end{aligned}$$

Using the above identities, we obtain

$$\begin{aligned} W_g^2 &= \int_0^1 [Y_{3,N}(t)]^2 dt \\ &= N \int_0^1 \{V_N(t) - P_V(t)\}^2 dt \\ &= N \textstyle\sum_{j=1}^K \displaystyle\int_{F(j-1)}^{F(j)} [S_j/N - F(j)]^2 dt \\ &= N \textstyle\sum_{j=1}^K [S_j/N - F(j)]^2 [F(j) - F(j-1)] \\ &= N \textstyle\sum_{j=1}^K [S_j/N - F(j)]^2 p(j) \\ &= W^2. \end{aligned}$$

In contrast, we have

$$\begin{aligned} W_f^2 &= \int_0^1 [Y_{2,N}(t)]^2 dt \\ &= N \int_0^1 \{U_N(t) - P_U(t)\}^2 dt \\ &= N \textstyle\sum_{j=1}^K \displaystyle\int_{F(j-1)}^{F(j)} [S_{j-1}/N - F(j-1)]^2 dt \\ &= N \textstyle\sum_{j=1}^K [S_{j-1}/N - F(j-1)]^2 [F(j) - F(j-1)] \\ &= N \textstyle\sum_{j=1}^K [S_{j-1}/N - F(j-1)]^2 p(j) \\ &= N \textstyle\sum_{j=1}^{K-1} [S_j/N - F(j)]^2 p(j+1). \end{aligned}$$

Thus, in the case where the sample consists of i.i.d. random variables, the statistics W_f^2, U_f^2 and A_f^2 do not reduce to the Cramér-von Mises statistics for discrete data defined in (5.8-5.11).

5.3 Distribution Theory - Known Parameters

5.3.1 Introduction

In this section, it will be shown that when the number of different distributions is fixed and finite, the untransformed empirical process $Y_{1,N}(j)$, defined in section 5.2, converges to a mean zero multivariate normal distribution with a given covariance matrix. As before, let M be the number of different distributions, N_i be the number of observations from the ith distribution, and $N = \sum_{i=1}^{M} N_i$. Also, let

$$N_i/N \rightarrow c_i \quad \text{as } N \rightarrow \infty \tag{5.29}$$

where $0 < c_i < 1$.

For the ith distribution, let $\mathbf{o}_i$, $\mathbf{p}_i$, $\mathbf{d}_i$ and $\mathbf{Z}_i$ be the vectors $\mathbf{o}$, $\mathbf{p}$, $\mathbf{d}$ and $\mathbf{Z}$ defined as in section 2.2, and let $\mathbf{D}_i$ be the corresponding $\mathbf{D}$. Also, let $F_i(j) = \sum_{k=0}^{j} p_i(k)$. Then $\mathbf{d}_i/\sqrt{N_i} = (\mathbf{o}_i - N_i\mathbf{p}_i)/\sqrt{N_i}$ has an asymptotic multivariate normal distribution with mean zero and covariance matrix, $\mathbf{\Sigma}_{0i} = \mathbf{D}_i - \mathbf{p}_i\mathbf{p}_i'$. Also, the statistic $\mathbf{Z}_i/\sqrt{N_i}$ has an asymptotic multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}_i = \mathbf{A}\mathbf{\Sigma}_{0i}\mathbf{A}'$ with j,kth element $\sigma_{i,jk} = \min\{F_i(j), F_i(k)\} - F_i(j)F_i(k)$.

The empirical process $Y_{1,N}(j)$ can now be written as a finite sum of K dimensional vectors, where K is the number of cells of the discrete distribution. Let $\mathbf{Y}_{1,N}$ be the vector with jth element $Y_{1,N}(j)$. Then

$$\begin{aligned}
\mathbf{Y}_{1,N} &= \frac{\sqrt{N}}{N}\sum_{i=1}^{M}\mathbf{Z}_i \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{M}\sqrt{N_i}\mathbf{Z}_i/\sqrt{N_i} \\
&= \sum_{i=1}^{M}\sqrt{\frac{N_i}{N}}\mathbf{Z}_i/\sqrt{N_i} \\
&\approx \sum_{i=1}^{M}\sqrt{c_i}\mathbf{Z}_i/\sqrt{N_i}.
\end{aligned}$$

Let $\mathbf{Y}_1 = \lim_{N\rightarrow\infty}\mathbf{Y}_{1,N}$. Then since $\mathbf{Z}_i/\sqrt{N_i}$ is asymptotically multivariate normal with mean zero and since $\mathbf{Z}_i, \mathbf{Z}_{i'}$ are independent for $i \neq i'$, $\mathbf{Y}_1$ is distributed multivariate normal with mean zero and covariance matrix, $\mathbf{\Sigma}_1$, with j,kth element

$$\sum_{i=1}^{M} c_i[\min\{F_i(j), F_i(k)\} - F_i(j)F_i(k)].$$

The four test statistics defined in terms of $Y_{1,N}$ given in (5.16 - 5.19) are of the general form $\mathbf{Z}'\mathbf{M}\mathbf{Z}$, where $\mathbf{M}$ is positive definite and symmetric. As before, the test statistics can be written asymptotically as a weighted sum of independent χ_1^2 variables, and percentage points can be found.

5.4 Estimated Parameters

5.4.1 Introduction

The more important problem where parameters must be estimated before testing fit will now be discussed, with reference to $Y_{1,N}(j)$. The specific situation considered is when parameters are given by a generalized linear model (McCullagh and Nelder, 1989). The parameter γ_i in $F(y_i;\gamma_i)$ becomes the mean μ_i of y_i. The vector of means, $\boldsymbol{\mu}$, depends on a matrix of known covariates, $\mathbf{X}$, and a vector $\boldsymbol{\theta}$ of parameters. This is done through the relationship $\boldsymbol{\mu} = g(\boldsymbol{\eta})$, and $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta}$. The function, $g(\cdot)$, or sometimes $g^{-1}(\cdot)$ is the *link* between the random and systematic components. The distributions are assumed to come from a member of the exponential family, but each with a different mean.

Let $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2']'$ where $\boldsymbol{\theta}_1$ is a vector of p_1 known parameters and $\boldsymbol{\theta}_2$ is a vector of p_2 parameters estimated from the data, and let $\boldsymbol{\theta}_0$ be the vector of true values of the parameters. Let $\hat{\boldsymbol{\theta}}_2$ be the maximum likelihood estimator of $\boldsymbol{\theta}_2$ and $\hat{F}(y) = F(y;\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2)$ be the estimated distribution function.

Regularity conditions are assumed such that the maximum likelihood estimator can be written

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{N}}\mathcal{J}^{-1}\sum_{i=1}^{N}\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} + \varepsilon_N, \tag{5.30}$$

where $\lim_{N\to\infty}\varepsilon_N = 0$, and

$$\mathcal{J} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left[\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]. \tag{5.31}$$

The regularity conditions include the following:

1. For all i, $F_i(y,\boldsymbol{\theta})$ has a density $f_i(y,\boldsymbol{\theta})$ such that $\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}$ exists, and

$$E\left[\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] = 0. \tag{5.32}$$

2. For all N, the matrices

$$\mathcal{J}_N = N^{-1}\sum_{i=1}^{N} E\left[\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\frac{\partial \ln f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]. \tag{5.33}$$

exist and converge to the finite positive-definite matrix $\mathcal{J}$ as $\lim_{N\to\infty}$.

3. Since y_i is discrete, it follows that for all j,

$$\left.\frac{\partial F(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \sum_{i=1}^{j} \left.\frac{\partial f(i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{5.34}$$

4. For each j, the function $g_2(j)$ exists such that

$$g_2(j) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} \left.\frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{5.35}$$

5.4.2 I.N.I.D. Discrete Variables

The process $\hat{Y}_{1,N}(j)$

Suppose $\hat{\mathbf{Z}}_i$ is $\mathbf{Z}_i$ with $\boldsymbol{\theta}$ replaced by the m.l.e. $\hat{\boldsymbol{\theta}}$. Similarly, let $Y_{1,N}(j)$ become $\hat{Y}_{1,N}(j)$. Let $\hat{\mathbf{Y}}_{1,N}$ be the vector with jth element $\hat{Y}_{1,N}(j)$. The vector can be written

$$\begin{aligned}\hat{\mathbf{Y}}_{1,N} &= \frac{\sqrt{N}}{N}\sum_{i=1}^{M}\hat{\mathbf{Z}}_i \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{M}\sqrt{N_i}\hat{\mathbf{Z}}_i/\sqrt{N_i} \\ &= \sum_{i=1}^{M}\sqrt{\frac{N_i}{N}}\hat{\mathbf{Z}}_i/\sqrt{N_i} \\ &= \sum_{i=1}^{M}\sqrt{c_i}\hat{\mathbf{Z}}_i/\sqrt{N_i}. \\ &\approx \sum_{i=1}^{M}\sqrt{c_i}\mathbf{A}\hat{\mathbf{d}}_i/\sqrt{N_i}. \end{aligned} \tag{5.36}$$

where $\mathbf{A}$ is the partial sum matrix. Let $\hat{\mathbf{Y}}_1 = \lim_{N\to\infty}\hat{\mathbf{Y}}_{1,N}$. Each $\hat{\mathbf{d}}_i/\sqrt{N_i}$ is asymptotically multivariate normal with mean zero, but the vectors are not independent and the distribution of $\hat{\mathbf{Y}}_1$ is not immediately obvious. In order to show that the asymptotic distribution of $\hat{\mathbf{Y}}_{1,N}$ is multivariate normal the following is required.

1. Let $\mathcal{D}$ be the vector of length KM formed by appending the vectors, $\mathbf{d}_i/\sqrt{N_i}$, in a column. Since each $\mathbf{d}_i/\sqrt{N_i}$ is independent and asymptotically K dimensional multivariate normal with covariance matrix $\boldsymbol{\Sigma}_{0i}$, $\mathcal{D}$ is asymptotically MK dimensional multivariate normal with covariance matrix made up of M partitions, the ith partition having the matrix $\boldsymbol{\Sigma}_{0i}$ on the diagonal.

2. Let $\mathcal{E}$ be the vector of length KM formed by appending the vectors, $\hat{\mathbf{d}}_i/\sqrt{N_i}$ in a column.

3. Suppose $\mathbf{C}_i$ is a K by K diagonal matrix, with all diagonal elements equal to the constant $\sqrt{c_i}$ defined in (5.29).

4. Let C be the K by MK matrix formed by placing the M matrices $\mathbf{C}_i$ side by side. Then define $\mathbf{d} = \mathcal{C}\mathcal{D}$ and $\hat{\mathbf{d}} = \mathcal{C}\mathcal{E}$. From (5.36) it can be seen that $\hat{\mathbf{Y}}_{1,N} \approx \mathbf{A}\hat{\mathbf{d}}$.

5. Let $\mathbf{R}_i$ be the p_2 by K matrix with j,kth element

$$\frac{\partial p_i(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_j}},$$

where $\boldsymbol{\theta}_{2_j}$ is the jth component of $\boldsymbol{\theta}_2$.

6. Let R be the p_2 by MK matrix formed by placing the matrices, $\mathbf{R}_i$, side by side.

7. Suppose $\mathbf{p}_i$ and $\hat{\mathbf{p}}_i$ are the vectors of length K with jth element $p_i(j)$ and $\hat{p}_i(j)$.

8. Let $\mathbf{p}$ and $\hat{\mathbf{p}}$ be the vectors of length MK formed by stacking the vectors, $\mathbf{p}_i$ and $\hat{\mathbf{p}}_i$.

9. Let P be the MK by MK diagonal matrix with the vector $\mathbf{p}$ on the diagonal.

10. Suppose $\mathbf{N}_i$ is a vector of length K with each element $\sqrt{N_i}$, and $\mathbf{N}$ be the vector of length MK formed by stacking the vectors $\mathbf{N}_i$. Then define $\mathcal{N}$ as the MK by MK diagonal matrix with the components of vector $\mathbf{N}$ on the diagonal.

11. Let M be the MK by MK diagonal matrix with diagonal elements $\mathbf{N}/\sqrt{N}$. The diagonal elements are thus the $\sqrt{c_i}$ each repeated K times.

Suppose $\mathbf{L}$, a vector of length p_2, is $\mathcal{R}\mathcal{P}^{-1}\mathcal{N}\mathcal{D}$. The kth element of $\mathbf{L}$ is given by

$$\begin{aligned}
&\textstyle\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{o_i(j)-N_i p_i(j)}{p_i(j)}\frac{\partial p_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}} \\
&= \textstyle\sum_{i=1}^{M}\sum_{j=1}^{K}\frac{o_i(j)}{p_i(j)}\frac{\partial p_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}} - \sum_{i=1}^{M}\sum_{j=1}^{K}N_i\frac{\partial p_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}} \\
&= \textstyle\sum_{i=1}^{M}\sum_{j=1}^{K}o_i(j)\frac{\partial \ln p_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}} - \sum_{i=1}^{M}N_i E[\frac{\partial \ln p_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}}] \\
&= \textstyle\sum_{i=1}^{N}\frac{\partial \ln p_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{2_k}}
\end{aligned} \tag{5.37}$$

since the second term is 0 from (5.32). Therefore, combining (5.30) and (5.37) we have,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = \frac{1}{\sqrt{N}}\mathcal{J}^{-1}\mathcal{R}\mathcal{P}^{-1}\mathcal{N}\mathcal{D} + \varepsilon_N, \tag{5.38}$$

where $\lim_{N\to\infty} \varepsilon_N = 0$. Also, using the assumed regularity conditions, it can be shown that

$$\begin{aligned}\sqrt{N}[\hat{\mathbf{p}} - \mathbf{p}] &= \sqrt{N}\mathcal{R}'(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + \varepsilon_N, \\ &= \frac{1}{\sqrt{N}}\mathcal{R}'\mathcal{J}^{-1}\mathcal{R}\mathcal{P}^{-1}\mathcal{N}\mathcal{D} + \varepsilon_N. \end{aligned} \tag{5.39}$$

where $\lim_{N\to\infty} \varepsilon_N = 0$. Thus,

$$\begin{aligned}\hat{\mathbf{d}} &= \mathcal{C}\mathcal{E} \\ &= \mathcal{C}[\mathcal{D} - \mathcal{N}(\hat{\mathbf{p}} - \mathbf{p})] \\ &= \mathcal{C}[\mathcal{D} - \frac{\mathcal{N}}{\sqrt{N}}\mathcal{R}'\mathcal{J}^{-1}\mathcal{R}\mathcal{P}^{-1}\frac{\mathcal{N}}{\sqrt{N}}\mathcal{D} \\ &= \mathcal{C}\left\{I - \mathcal{M}\mathcal{R}'\mathcal{J}^{-1}\mathcal{R}\mathcal{P}^{-1}\mathcal{M}\right\}\mathcal{D} \end{aligned} \tag{5.40}$$

as $N \to \infty$. Since (5.40) is a linear combination of asymptotic multivariate normal random variables, it is asymptotically multivariate normally distributed. Finally since $\hat{Y}_{1,N}$ can be written as a linear combination of $\hat{\mathbf{d}}$, it is asymptotically multivariate normally distributed.

The vector $\hat{Y}_{1,N}$ is easily shown to have mean zero; its covariance can be derived as follows. Let $F_i(j) = \sum_{k=1}^{j} p_i(k)$. Then

$$\begin{aligned}&Cov[\hat{Y}_{1,N}(j), \hat{Y}_{1,N}(k)] \\ &= E[\hat{Y}_{1,N}(j)\hat{Y}_{1,N}(k)] \\ &= E[\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{y_i \le j\} - \hat{F}_i(j)\}\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\{1\{y_i \le k\} - \hat{F}_i(k)\}] \\ &= N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N} E[\sqrt{N}\{1\{y_i \le j\} - \hat{F}_i(j)\}\sqrt{N}\{1\{y_{i'} \le k\} - \hat{F}_{i'}(k)\}] \\ &= N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N} E[\sqrt{N}\{1\{y_i \le j\} - F_i(j) + F_i(j) - \hat{F}_i(j)\} \\ &\quad \sqrt{N}\{1\{y_{i'} \le k\} - F_{i'}(k) + F_{i'}(k) - \hat{F}_{i'}(k)\}] \\ &= N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N} E[N\{1\{y_i \le j\} - F_i(j)\}\{1\{y_{i'} \le k\} - F_{i'}(k)\} - \\ &\quad 2E[N\{1\{y_i \le j\} - F_i(j)\}]\{\hat{F}_{i'}(k) - F_{i'}(k)\}] + \\ &\quad E[\{F_i(j) - \hat{F}_i(j)\}\{F_{i'}(k) - \hat{F}_{i'}(k)\}] \\ &= N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\{A - 2B + C\}. \end{aligned}$$

Notice that

$$\begin{aligned} N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N} A &= Cov[Y_{1,N}(j), Y_{1,N}(k)] \\ &= \textstyle\sum_{i=1}^{M} c_i[\min\{F_i(j), F_i(k)\} - F_i(j)F_i(k)]. \end{aligned}$$

This is the covariance of the discrete empirical process with known parameters. From (5.39), C can be written

$$\begin{aligned} & E[\{\frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)] + \varepsilon_N\}\{(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)'\frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}] + \varepsilon_N\} \\ = & \frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}E[(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)']\frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} + \varepsilon_N \end{aligned}$$

which has a limit

$$\frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}$$

as $N \to \infty$, where $\mathcal{J}$ is given in (5.31). The term B can be derived as follows,

$$\begin{aligned} B &= E[N\{1\{y_i \le j\} - F_i(j)\}]\{\hat{F}_{i'}(k) - F_{i'}(k)\}] \\ &= E[\sqrt{N}\{1\{y_i \le j\} - F_i(j)\}\{\frac{1}{\sqrt{N}}\frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + \varepsilon_N\}] \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\{E[1\{y_i \le j\}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + \varepsilon_N] + F_i(j)E[(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)] + \varepsilon_N \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\{E[1\{y_i \le j\}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + \varepsilon_N]\} \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\{E[1\{y_i \le j\}\sum_{i=1}^{N}\frac{\partial \ln f_i(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}] + \varepsilon_N \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\{\sum_{l=1}^{j}\frac{\partial \ln f_i(l,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}f_i(l)\} + \varepsilon_N \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\{\sum_{l=1}^{j}\frac{\partial f_i(l,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\} + \varepsilon_N \\ &= \frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} + \varepsilon_N \end{aligned}$$

using the results (5.30 - 5.32). The limiting value of B is the same as that of C. Combining the result for each term, and letting $\hat{\boldsymbol{\Sigma}}_1 = Cov[\hat{Y}_{1,N}(j), \hat{Y}_{1,N}(k)]$, we have

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_1 = & \textstyle\sum_{i=1}^{M} c_i[\min\{F_i(j), F_i(k)\} - F_i(j)F_i(k)] - \\ & \textstyle\sum_{i=1}^{M}\sum_{i'=1}^{M} c_i c_{i'}\displaystyle\frac{\partial F_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'}\mathcal{J}^{-1}\frac{\partial F_{i'}(k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} \end{aligned}$$

As before, the test statistics can be written asymptotically as a weighted sum of independent χ_1^2 variables and percentage percentage points found.

The process $\hat{Y}_{4,N}(t)$

Since the process, $\hat{Y}_{4,N}(t)$, is equivalent to that of $\hat{Y}_n(t)$ for continuous distributions, the convergence of $\hat{Y}_{4,N}(t)$ to a mean zero Gaussian process with covariance function

$$\hat{\rho}(s,t) = \min(s,t) - st - g_2(s)' J^{-1} g_2(t),$$

where

$$g_2(t) = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \left. \frac{\partial F_i^*[F_i^{*-1}(t,\boldsymbol{\theta}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}_2} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

follows from the application of the result from Loynes (1980) for non-i.i.d. continuous distributions.

The processes $\hat{Y}_{2,N}(t)$ and $\hat{Y}_{3,N}(t)$

The covariance functions for these two processes can be found in a similar manner to the covariance matrix for $\hat{Y}_{1,N}(j)$. Define $F^-(t) = \max\{j : F(j) \leq t\}$ and $G^-(t) = \max\{j : G(j) \leq t\}$, so that $Pr[F(Y) \leq t] = F[F^-(t)]$ and $Pr[G(Y) \leq t] = G[G^-(t)]$. Then

$$\begin{aligned}\hat{\rho}_2(s,t) &= \sum_{i=1}^M c_i[\min\{F_i[F_i^-(s)], F_i[F_i^-(t)]\} - F_i[F_i^-(s)]F_i[F_i^-(t)] - \\ &\quad \sum_{i=1}^M \sum_{i'=1}^M c_i c_{i'} \frac{\partial F_i[F_i^-(s,\boldsymbol{\theta}),\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}_2'} \mathcal{J}^{-1} \frac{\partial F_{i'}[F_{i'}^-(t,\boldsymbol{\theta}),\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}_2},\end{aligned}$$

and

$$\begin{aligned}\hat{\rho}_3(s,t) &= \sum_{i=1}^M c_i[\min\{G_i[G_i^-(s)], G_i[G_i^-(t)]\} - G_i[G_i^-(s)]G_i[G_i^-(t)] - \\ &\quad \sum_{i=1}^M \sum_{i'=1}^M c_i c_{i'} \frac{\partial G_i[G_i^-(s,\boldsymbol{\theta}),\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}_2'} \mathcal{J}^{-1} \frac{\partial G_{i'}[G_{i'}^-(t,\boldsymbol{\theta}),\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}_2}.\end{aligned}$$

The convergence of statistics based on these processes will be discussed in section 5.6. We next turn to a discussion of three commonly used regression models.

5.5 Covariances - Specific Models

5.5.1 Poisson Regression

Poisson regression is a commonly used method to relate an observed count with a set of explanatory variables. Let $y_1, y_2, \ldots, y_N$ be a sample of observed counts and $\mathbf{X}$ be a matrix of explanatory covariates consisting of row vectors, $\mathbf{X}_i$, which contain the covariates for the ith count. Each count, y_i is assumed to be Poisson distributed with mean, μ_i. The mean, μ_i, is related to the parameters and explanatory variables by the following link:

$$\mu_i = m_i g(\eta_i) = m_i \exp\{\mathbf{X}_i \boldsymbol{\theta}\}$$

where η_i is the linear predictor, $\mathbf{X}_i\boldsymbol{\theta}$. A Poisson regression model is also referred to as a log-linear model, since $g^{-1}(\cdot)$ is the logarithmic function. The term, m_i, is referred to as an *offset*, and is often modeled as an additional term in the linear predictor with parameter equal to 1. In this case the covariate, $\log(m_i)$, would be included in the linear predictor.

The density for the ith count, y_i is

$$\begin{aligned} f_i(j) &= \frac{\mu_i^j \exp(-\mu_i)}{j!}, \\ &= \frac{m_i^j \exp\{j\mathbf{X}_i'\boldsymbol{\theta}\} \exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\})}{j!}, \end{aligned}$$

and the log-density at y_i is

$$\begin{aligned} \log f_i(y_i) &= y_i \log \mu_i - \mu_i - \log(y_i!) \\ &= y_i \log m_i + y_i \mathbf{X}_i'\boldsymbol{\theta} - m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\} - \log(y_i!). \end{aligned}$$

Now,

$$\begin{aligned} &\left.\frac{\partial f_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \frac{m_i^j}{j!}\left[\exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) \exp\{j\mathbf{X}_i'\boldsymbol{\theta}\} j\mathbf{X}_i' - \right. \\ &\quad \left. \exp\{j\mathbf{X}_i'\boldsymbol{\theta}\} j\mathbf{X}_i' \exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}\mathbf{X}_i'\right] \\ &= \frac{\exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) m_i^j}{j!}\left[\exp\{j\mathbf{X}_i'\boldsymbol{\theta}\} j - \exp\{(j+1)\mathbf{X}_i'\boldsymbol{\theta}\} m_i\right] \mathbf{X}_i' \\ &= m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}\left[\frac{\exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) \exp\{(j-1)\mathbf{X}_i'\boldsymbol{\theta}\}}{(j-1)!} - \right. \\ &\quad \left. \frac{\exp(-m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) \exp\{(j)\mathbf{X}_i'\boldsymbol{\theta}\}}{j!}\right] \mathbf{X}_i' \\ &= m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}\left[f_i(j-1) - f_i(j)\right] \mathbf{X}_i' \\ &= \mu_i \left[f_i(j-1) - f_i(j)\right] \mathbf{X}_i'. \end{aligned}$$

Therefore,

$$\left.\frac{\partial F_i(j,\boldsymbol{\theta}\}}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\mu_i f_i(j)\mathbf{X}_i'.$$

Also,

$$\left.\frac{\partial \log f_i(y_i,\boldsymbol{\theta}\}}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = y_i\mathbf{X}_i' - m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}\mathbf{X}_i',$$

and

$$\begin{aligned}\left.\frac{\partial^2 \log f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -m_i \exp\{\mathbf{X}_i'\boldsymbol{\theta}\}(\mathbf{X}_i'\mathbf{X}_i), \\ &= -\mu_i(\mathbf{X}_i'\mathbf{X}_i).\end{aligned}$$

Thus,

$$\mathcal{J} = \lim_{N\to\infty} N^{-1}\textstyle\sum_{i=1}^N \mu_i(\mathbf{X}_i'\mathbf{X}_i).$$

5.5.2 Logistic Regression

Logistic regression is used to relate an observed binomial random variable with a set of explanatory variables. Let $y_1, y_2, \ldots, y_N$ be a sample of observed counts, and $\mathbf{X}$ be a matrix of explanatory covariates consisting of row vectors, $\mathbf{X}_i$, which contain the covariates for the ith count. Each count, y_i is assumed to be binomially distributed with probability, π_i, of success, and number of trials, m_i. Let the means, $\mu_i = m_i\pi_i$ be related to the parameters and explanatory variables by the following link:

$$\mu_i = m_i g(\eta_i) = m_i \frac{\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{1+\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}$$

where η_i is the linear predictor, $\mathbf{X}_i\boldsymbol{\theta}$. Thus, $g^{-1}(\pi) = \log[\pi/(1-\pi)]$, the logit function. An important property of the logistic regression model is that the parameters have the same interpretation whether the data are sampled prospectively or retrospectively; for this reason it is often used in epidemiological research.

The density for the ith count, y_i is

$$\begin{aligned} f_i(j,m_i) &= C_j^{m_i}\pi_i^j(1-\pi_i)^{m_i-j} \\ &= C_j^{m_i} g(\eta_i)^j[1-g(\eta_i)]^{m_i-j}, \end{aligned}$$

and the log-density at y_i is

$$\log f_i(y_i,m_i) = y_i \log \pi_i + (m_i - y_i)\log(1-\pi_i) + C$$

$$= y_i\mathbf{X}_i'\boldsymbol{\theta} - y_i\log(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) - m_i\log(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) + y_i\log(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) + C$$

$$= y_i\mathbf{X}_i'\boldsymbol{\theta} - m_i\log(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}) + C$$

where C is a term that does not involve the parameters. Now,

$$\left.\frac{\partial f_i(j,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$= C_j^{m_i}\left[jg(\eta_i)^{j-1}\frac{\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i\boldsymbol{\theta}\})^2}[1-g(\eta_i)]^{m_i-j}\mathbf{X}_i' - (m_i-j)[1-g(\eta_i)]^{m_i-j-1}\frac{\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i\boldsymbol{\theta}\})^2}g(\eta_i)]^j\mathbf{X}_i'\right]$$

$$= C_j^{m_i}g(\eta_i)^{j-1}[1-g(\eta_i)]^{m_i-j-1}\frac{\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i\boldsymbol{\theta}\})^2}[j[1-g(\eta_i)]-(m_i-j)g(\eta_i)]\,\mathbf{X}_i'$$

$$= C_j^{m_i}g(\eta_i)^{j-1}[1-g(\eta_i)]^{m_i-j-1}\frac{\exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i\boldsymbol{\theta}\})^2}[j-m_ig(\eta_i)]\,\mathbf{X}_i'$$

$$= m_ig(\eta_i)[1-g(\eta_i)]\,[f_i(j-1,m_i-1)-f_i(j,m_i-1)]\,\mathbf{X}_i'.$$

Therefore,

$$\left.\frac{\partial F_i(j,\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -m_ig(\eta_i)[1-g(\eta_i)]f_i(j,m_i-1)\mathbf{X}_i'.$$

Also,

$$\left.\frac{\partial\log f_i(y_i,\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = y_i\mathbf{X}_i' - m_i\frac{\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\})}\mathbf{X}_i',$$

and

$$\left.\frac{\partial^2\log f_i(y_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -m_i\frac{\exp\{\mathbf{X}_i'\boldsymbol{\theta}\}}{(1+\exp\{\mathbf{X}_i'\boldsymbol{\theta}\})^2}(\mathbf{X}_i'\mathbf{X}_i),$$

$$= -m_ig(\eta_i)[1-g(\eta_i)](\mathbf{X}_i'\mathbf{X}_i).$$

Thus,

$$\mathcal{J} = \lim_{N\to\infty} N^{-1}\textstyle\sum_{i=1}^N m_ig(\eta_i)[1-g(\eta_i)](\mathbf{X}_i'\mathbf{X}_i).$$

5.5.3 Complementary Log-Log Regression

Complementary log-log regression is an alternate method to relate an observed binomial random variable to a set of explanatory variables. The situation is as for logistic regression but with the following link:

$$\mu_i = m_ig(\eta_i) = m_i[1-\exp(-\exp\{\mathbf{X}_i\boldsymbol{\theta}\})]$$

where η_i is the linear predictor, $\mathbf{X}_i\boldsymbol{\theta}$. Notice that $1 - g(\eta_i) = \exp(\eta_i)$. Once again, the name of the procedure refers to the inverse link function,

$$g^{-1}(\pi_i) = \log[-\log(1-\pi_i)].$$

where $\pi_i = \mu_i/m_i$. The complementary log-log regression model is often used in the analysis of limited dilution assays.

Let $\zeta_i = \exp\eta_i$, then the density for the ith count, y_i is

$$\begin{aligned} f_i(j, m_i) &= C_j^{m_i}\pi_i^j(1-\pi_i)^{m_i-j} \\ &= C_j^{m_i} g(\eta_i)^j[1-g(\eta_i)]^{m_i-j}, \end{aligned}$$

and the log-density at y_i is

$$\begin{aligned} \log f_i(y_i, m_i) &= y_i \log\pi_i + (m_i - y_i)\log(1-\pi_i) + C \\ &= y_i \log[g(\eta_i)] + (m_i - y_i)\zeta_i + C \end{aligned}$$

where C is a term that does not involve the parameters. Now, let

$$g'(\eta_i) = \frac{\partial g(\eta_i)}{\partial \boldsymbol{\theta}_2}.$$

Then

$$\begin{aligned} &\left.\frac{\partial f_i(j,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= C_j^{m_i}\left[jg(\eta_i)^{j-1}g'(\eta_i)[1-g(\eta_i)]^{m_i-j} - \right. \\ &\qquad \left.(m_i-j)[1-g(\eta_i)]^{m_i-j-1}g'(\eta_i)g(\eta_i)]^j\right] \\ &= C_j^{m_i}g'(\eta_i)g(\eta_i)^{j-1}[1-g(\eta_i)]^{m_i-j-1}\left[jg(\eta_i) - (m_i-j)g(\eta_i)\right] \\ &= m_i g'(\eta_i)\left[f_i(j-1, m_i-1) - f_i(j, m_i-1)\right] \\ &= m_i[1-g(\eta_i)]\zeta_i^2\left[f_i(j-1,m_i-1) - f_i(j,m_i-1)\right]\mathbf{X}_i' \end{aligned}$$

since $g'(\eta_i) = [1-g(\eta_i)]\zeta_i$. Therefore,

$$\left.\frac{\partial F_i(j,\boldsymbol{\theta}\}}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -m_i[1-g(\eta_i)]\zeta_i^2 f_i(j, m_i-1)\mathbf{X}_i'.$$

Also,

$$\left.\frac{\partial \log f_i(y_i,\boldsymbol{\theta}\}}{\partial \boldsymbol{\theta}_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = y_i\zeta_i[1-g(\eta_i)]/g(\eta_i)\mathbf{X}_i' - (m_i-j)\zeta_i\mathbf{X}_i',$$

and

$$\left.\frac{\partial^2 \log f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$= \left\{\frac{y_i}{g(\eta_i)^2}\left[g(\eta_i)[1-g(\eta_i)]\zeta_i - g(\eta_i)[1-g(\eta_i)]\zeta_i^2 - [1-g(\eta_i)]^2\zeta_i^2\right] - (m_i - y_i)\zeta_i\right\}(\mathbf{X}_i'\mathbf{X}_i)$$

$$= \left[\frac{y_i\zeta_i[1-g(\eta_i)]}{g(\eta_i)^2}(g(\eta_i) - g(\eta_i)\zeta_i - [1-g(\eta_i)]\zeta_i) - m_i\zeta_i - y_i\zeta_i\right](\mathbf{X}_i'\mathbf{X}_i).$$

Now $E[y_i] = \mu_i = m_i g(\eta_i)$, and

$$E\left[\left.\frac{\partial^2 \log f_i(y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$$

$$= \left\{\frac{n_i\zeta_i[1-g(\eta_i)]}{g(\eta_i)}[g(\eta_i) - g(\eta_i)\zeta_i - [1-g(\eta_i)]\zeta_i] - m_i\zeta_i - m_i g(\eta_i)\zeta_i\right](\mathbf{X}_i'\mathbf{X}_i),$$

$$= \left\{\frac{n_i\zeta_i[1-g(\eta_i)]}{g(\eta_i)}[g(\eta_i) - \zeta_i] - m_i\zeta_i[1-g(\eta_i)]\right\}(\mathbf{X}_i'\mathbf{X}_i)$$

$$= \frac{n_i\zeta_i^2[1-g(\eta_i)]}{g(\eta_i)}(\mathbf{X}_i'\mathbf{X}_i)$$

$$= \frac{n_i\zeta_i^2}{[\exp(\zeta_i)-1]}(\mathbf{X}_i'\mathbf{X}_i)$$

since $[1-g(\eta_i)]/g(\eta_i) = [\exp(\zeta_i)-1]$. Thus,

$$\mathcal{J} = \lim_{N\to\infty} N^{-1}\textstyle\sum_{i=1}^{N} \dfrac{n_i\zeta_i^2}{[\exp(\zeta_i)-1]}(\mathbf{X}_i'\mathbf{X}_i).$$

5.6 Calculation of Percentage Points

The Cramér-von Mises statistics based on the untransformed process, $\hat{Y}_{1,N}(j)$, were defined in (5.16 - 5.19). Asymptotic percentage points have then been found in the usual way.

For the Poisson regression model, the number of cells was truncated at K, where K was chosen to make the final values of λ_i sufficiently small that the percentage points do not change with the addition of more eigenvalues. For binomial models (logistic regression, complementary log-log regression), K is the maximum number of trials for any one observation.

For W_u^2, the matrix $\mathbf{M}$ is equal to $\overline{\mathbf{D}}$, the diagonal matrix with the average probability of falling into cell j, $\overline{p(j)}$, on the diagonal, where $\overline{p(j)} = \sum_{i=1}^{M} p_i(j)$. For the statistics U_u^2,

A_u^2 and W_{mu}^2 the **M** matrices are the matrices corresponding to those in the i.i.d. case, but averaged over all distributions.

For the statistics based on the empirical processes, $\hat{Y}_{2,N}(t)$ and $\hat{Y}_{3,N}(t)$, it is assumed without proof that the empirical process converges to a Gaussian process. Then, for a statistic based on the process $\hat{Y}_{2,N}(t)$, it is necessary to find the eigenvalues of the covariance function, $\hat{\rho}_x(s,t)$. where $\hat{\rho}_x(s,t) = \hat{\rho}_2(s,t)\sqrt{\Psi(s)\Psi(t)}$ and $\Psi(s)$ is the weight function of the appropriate Cramér-von Mises statistic. These are the weights in the usual asymptotic distribution of the statistic and percentage can then be found. Similarly, percentage points can be obtained for the statistics based on $\hat{Y}_{3,N}(t)$.

The eigenvalues above were approximated as follows. The interval $[0,1]$ was discretized into K points, the covariance function was evaluated at each of the points. The eigenvalues of the resulting matrices were then found. For the percentage points presented below, a discretization of $K = 50$ was used. The eigenvalues were found using S-PLUS (S-PLUS, 1991), and the percentage points were then found by Imhof's method.

In order to examine the rate of convergence of percentage points for finite samples to the asymptotic points, percentage points were generated by Monte Carlo simulation using 10,000 samples. The results are given in Tables 5.1 through 5.8 for a variety of Poisson regression models. The standard error of estimation of the level of the pth percentage point is approximately $\sqrt{p(1-p)/n}$ where n is the number of simulations; for the .95 percentage point the standard error is .22%.

All the points given are for models with one estimated parameter, the overall mean. Points are given for the statistics, W_u^2, A_u^2, W_f^2, A_f^2, W_g^2, A_g^2, for each of the following models:

Model 1 Two equally proportioned distributions with $\mu_1 = .5$, $\mu_2 = 1.5$;

Model 2 Two equal proportioned distributions with $\mu_1 = 5$, $\mu_2 = 15$;

Model 3 Two unequal proportioned distributions with $\mu_1 = .9$ (with sampling proportion, p=.9), $\mu_2 = 1.9$ (p=.1);

Model 4 Two unequal proportioned distributions with $\mu_1 = 9$ (p=.9), $\mu_2 = 19$ (p=.1);

Model 5 Five equally proportioned distributions with $\mu_1 = .2$, $\mu_2 = .6$, $\mu_3 = 1.0$, $\mu_4 = 1.4$ and $\mu_5 = 1.8$;

Model 6 Five equally proportioned distributions with $\mu_1 = 2$, $\mu_2 = 6$, $\mu_3 = 10$, $\mu_4 = 14$ and $\mu_5 = 18$;

Model 7 Ten equally proportioned distributions with $\mu_1 = .1$, $\mu_2 = .3$, …, $\mu_9 = 1.7$, $\mu_{10} = 1.9$;

Model 8 Ten equally proportioned distributions with $\mu_1 = 1$, $\mu_2 = 3$, …, $\mu_9 = 17$ and $\mu_{10} = 19$.

In the table, N refers to the total sample size. For example, when generating from Model 2, a sample of size 20 consists of 10 observations from a Poisson distribution with mean, $\mu = 5$, and 10 observations from a Poisson distribution with mean, $\mu = 15$.

In all cases the points converge rapidly to the asymptotic points, which can be used for samples of size greater than 20, and in some cases for sample sizes of greater than 10. Similar results were found for the other statistics defined in section 5.2. The convergence of the Monte Carlo points to the asymptotic points gives strong empirical evidence that the limiting processes for $\hat{Y}_{2,N}(t)$ and $\hat{Y}_{3,N}(t)$ are Gaussian.

An S function (Becker, Chambers and Wilks, 1988; S-PLUS, 1991) has been written to compute the statistics and their asymptotic p-values, for the Poisson and logistic regression models.

Table 5.1: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 1. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.097	0.126	0.159	0.205	0.250	0.322
20	0.094	0.129	0.159	0.215	0.273	0.343
30	0.093	0.131	0.159	0.213	0.276	0.351
50	0.094	0.134	0.164	0.217	0.276	0.362
100	0.093	0.131	0.163	0.220	0.279	0.362
∞	0.093	0.130	0.161	0.216	0.274	0.352

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.510	0.744	0.854	1.095	1.352	1.637
20	0.557	0.738	0.910	1.136	1.436	1.732
30	0.560	0.735	0.886	1.145	1.442	1.815
50	0.564	0.756	0.901	1.172	1.469	1.848
100	0.555	0.740	0.903	1.171	1.454	1.825
∞	0.558	0.743	0.893	1.158	1.431	1.778

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.068	0.087	0.101	0.131	0.154	0.186
20	0.068	0.088	0.106	0.137	0.164	0.205
30	0.066	0.087	0.104	0.132	0.162	0.205
50	0.068	0.091	0.108	0.139	0.170	0.215
100	0.067	0.089	0.107	0.138	0.172	0.221
∞	0.068	0.090	0.108	0.139	0.171	0.213

Table 5.1: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 1. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.482	0.617	0.717	0.884	1.098	1.366
20	0.482	0.655	0.790	1.008	1.217	1.579
30	0.474	0.640	0.774	0.997	1.211	1.541
50	0.483	0.664	0.798	1.046	1.314	1.671
100	0.482	0.655	0.803	1.051	1.322	1.736
∞	0.474	0.645	0.787	1.036	1.298	1.656

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.091	0.119	0.141	0.181	0.222	0.266
20	0.089	0.119	0.145	0.188	0.232	0.291
30	0.090	0.123	0.146	0.193	0.241	0.301
50	0.092	0.121	0.147	0.197	0.253	0.320
100	0.091	0.123	0.151	0.199	0.242	0.313
∞	0.089	0.120	0.145	0.191	0.239	0.304

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.556	0.700	0.815	1.015	1.184	1.466
20	0.553	0.712	0.837	1.069	1.328	1.588
30	0.559	0.718	0.854	1.091	1.353	1.677
50	0.568	0.723	0.865	1.122	1.385	1.758
100	0.567	0.739	0.871	1.123	1.366	1.728
∞	0.548	0.709	0.842	1.079	1.327	1.666

Table 5.2: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 2. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.078	0.098	0.114	0.143	0.173	0.207
20	0.079	0.099	0.115	0.142	0.169	0.214
30	0.078	0.097	0.115	0.145	0.173	0.212
50	0.078	0.097	0.113	0.139	0.169	0.208
100	0.077	0.095	0.111	0.139	0.170	0.212
∞	0.078	0.097	0.113	0.142	0.172	0.214

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.546	0.674	0.779	0.963	1.153	1.407
20	0.558	0.683	0.786	0.959	1.159	1.419
30	0.553	0.681	0.784	0.968	1.164	1.408
50	0.559	0.686	0.780	0.963	1.127	1.373
100	0.552	0.674	0.768	0.950	1.149	1.416
∞	0.560	0.685	0.787	0.967	1.156	1.415

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.100	0.123	0.141	0.170	0.203	0.240
20	0.103	0.127	0.147	0.181	0.213	0.253
30	0.101	0.127	0.145	0.179	0.213	0.261
50	0.102	0.126	0.145	0.177	0.210	0.256
100	0.100	0.126	0.145	0.179	0.212	0.256
∞	0.101	0.125	0.145	0.178	0.212	0.256

Table 5.2: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 2. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.652	0.793	0.913	1.115	1.337	1.626
20	0.692	0.844	0.952	1.166	1.385	1.691
30	0.691	0.849	0.961	1.173	1.407	1.701
50	0.689	0.842	0.960	1.159	1.382	1.666
100	0.688	0.845	0.960	1.169	1.368	1.684
∞	0.675	0.826	0.946	1.152	1.361	1.642

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.101	0.125	0.144	0.173	0.207	0.245
20	0.104	0.129	0.149	0.182	0.218	0.260
30	0.102	0.126	0.147	0.182	0.219	0.263
50	0.103	0.128	0.147	0.180	0.218	0.257
100	0.102	0.127	0.146	0.182	0.212	0.261
∞	0.102	0.126	0.146	0.179	0.213	0.258

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.643	0.778	0.887	1.077	1.279	1.533
20	0.673	0.821	0.937	1.146	1.365	1.624
30	0.668	0.825	0.929	1.140	1.347	1.629
50	0.672	0.822	0.933	1.146	1.324	1.568
100	0.667	0.817	0.928	1.131	1.342	1.651
∞	0.656	0.803	0.919	1.119	1.323	1.596

Table 5.3: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 3. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.093	0.126	0.151	0.196	0.264	0.299
20	0.093	0.131	0.157	0.212	0.260	0.328
30	0.090	0.125	0.154	0.203	0.260	0.332
50	0.093	0.128	0.157	0.208	0.263	0.329
100	0.091	0.127	0.155	0.209	0.261	0.333
∞	0.091	0.127	0.157	0.210	0.265	0.340

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.582	0.744	0.852	1.114	1.407	1.816
20	0.571	0.756	0.906	1.163	1.486	1.844
30	0.559	0.759	0.913	1.150	1.431	1.802
50	0.576	0.762	0.917	1.189	1.460	1.785
100	0.571	0.754	0.911	1.187	1.447	1.773
∞	0.575	0.765	0.917	1.185	1.459	1.811

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.063	0.088	0.107	0.149	0.188	0.238
20	0.064	0.092	0.114	0.154	0.192	0.252
30	0.063	0.087	0.110	0.152	0.191	0.247
50	0.064	0.091	0.113	0.153	0.191	0.253
100	0.064	0.090	0.112	0.153	0.193	0.253
∞	0.064	0.091	0.114	0.156	0.200	0.259

Table 5.3: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 3. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.484	0.725	0.924	1.422	1.896	2.401
20	0.542	0.723	0.888	1.165	1.538	2.160
30	0.507	0.681	0.843	1.162	1.558	2.171
50	0.544	0.724	0.869	1.118	1.464	1.985
100	0.585	0.743	0.878	1.132	1.410	1.918
∞	0.592	0.786	0.939	1.206	1.472	1.824

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.089	0.123	0.139	0.191	0.238	0.303
20	0.089	0.125	0.153	0.203	0.256	0.314
30	0.088	0.120	0.148	0.200	0.248	0.331
50	0.088	0.124	0.152	0.203	0.252	0.315
100	0.089	0.122	0.152	0.203	0.253	0.315
∞	0.087	0.121	0.150	0.201	0.254	0.327

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.545	0.736	0.856	1.129	1.296	1.703
20	0.569	0.759	0.900	1.149	1.414	1.762
30	0.563	0.743	0.887	1.147	1.394	1.799
50	0.567	0.754	0.912	1.156	1.404	1.764
100	0.571	0.755	0.900	1.166	1.403	1.770
∞	0.558	0.741	0.890	1.149	1.415	1.773

Table 5.4: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 4. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.099	0.122	0.140	0.172	0.205	0.239
20	0.101	0.123	0.143	0.172	0.208	0.245
30	0.099	0.123	0.141	0.173	0.205	0.249
50	0.098	0.122	0.142	0.174	0.207	0.252
100	0.099	0.123	0.141	0.178	0.212	0.265
∞	0.098	0.122	0.141	0.174	0.207	0.251

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.605	0.733	0.834	1.011	1.195	1.407
20	0.617	0.750	0.852	1.013	1.200	1.456
30	0.610	0.737	0.834	1.010	1.193	1.442
50	0.606	0.738	0.845	1.017	1.196	1.462
100	0.615	0.750	0.850	1.032	1.233	1.497
∞	0.611	0.743	0.846	1.023	1.199	1.434

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.096	0.119	0.138	0.168	0.198	0.236
20	0.098	0.122	0.140	0.170	0.202	0.245
30	0.094	0.119	0.139	0.167	0.201	0.244
50	0.096	0.120	0.139	0.171	0.205	0.248
100	0.097	0.121	0.141	0.175	0.208	0.257
∞	0.096	0.120	0.139	0.171	0.204	0.248

Table 5.4: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 4. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.646	0.789	0.904	1.117	1.310	1.615
20	0.669	0.816	0.944	1.146	1.370	1.658
30	0.650	0.801	0.921	1.108	1.324	1.668
50	0.659	0.813	0.939	1.138	1.344	1.619
100	0.670	0.823	0.944	1.171	1.403	1.680
∞	0.648	0.797	0.915	1.120	1.329	1.612

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.096	0.120	0.138	0.169	0.196	0.240
20	0.099	0.122	0.141	0.171	0.203	0.245
30	0.095	0.119	0.139	0.171	0.201	0.240
50	0.096	0.121	0.139	0.173	0.204	0.250
100	0.097	0.121	0.140	0.175	0.211	0.263
∞	0.096	0.120	0.139	0.172	0.204	0.248

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.631	0.768	0.881	1.066	1.259	1.502
20	0.649	0.793	0.907	1.109	1.308	1.601
30	0.633	0.779	0.891	1.082	1.284	1.570
50	0.643	0.797	0.905	1.102	1.324	1.557
100	0.651	0.795	0.915	1.133	1.343	1.642
∞	0.628	0.772	0.886	1.084	1.285	1.556

Table 5.5: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 5. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.093	0.119	0.134	0.211	0.224	0.251
10	0.095	0.127	0.154	0.192	0.262	0.302
20	0.089	0.123	0.153	0.210	0.259	0.333
30	0.087	0.125	0.151	0.203	0.256	0.334
50	0.089	0.125	0.155	0.205	0.261	0.339
100	0.091	0.126	0.159	0.211	0.263	0.336
∞	0.090	0.126	0.155	0.208	0.263	0.338

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.519	0.692	0.811	1.088	1.195	1.530
10	0.585	0.740	0.835	1.118	1.379	1.650
20	0.544	0.750	0.898	1.178	1.483	1.819
30	0.565	0.747	0.910	1.148	1.424	1.768
50	0.569	0.750	0.907	1.163	1.444	1.808
100	0.580	0.773	0.927	1.204	1.470	1.875
∞	0.572	0.761	0.912	1.180	1.453	1.812

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.066	0.084	0.099	0.126	0.133	0.172
10	0.068	0.087	0.101	0.126	0.153	0.184
20	0.068	0.087	0.103	0.130	0.158	0.198
30	0.069	0.089	0.105	0.132	0.161	0.199
50	0.069	0.090	0.105	0.133	0.160	0.195
100	0.069	0.089	0.105	0.134	0.164	0.206
∞	0.070	0.090	0.107	0.136	0.165	0.205

Table 5.5: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 5. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.415	0.503	0.613	0.716	0.892	1.102
10	0.431	0.539	0.625	0.762	0.930	1.141
20	0.430	0.548	0.638	0.793	0.961	1.191
30	0.437	0.556	0.647	0.816	0.988	1.200
50	0.444	0.566	0.659	0.819	0.989	1.167
100	0.435	0.556	0.649	0.819	1.012	1.233
∞	0.431	0.550	0.647	0.815	0.989	1.223

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.080	0.106	0.132	0.160	0.180	0.246
10	0.087	0.113	0.134	0.165	0.206	0.250
20	0.086	0.113	0.135	0.177	0.222	0.280
30	0.087	0.114	0.137	0.177	0.215	0.270
50	0.088	0.116	0.136	0.176	0.215	0.276
100	0.086	0.113	0.137	0.182	0.226	0.280
∞	0.086	0.113	0.136	0.175	0.217	0.273

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.480	0.626	0.732	0.882	1.148	1.255
10	0.525	0.664	0.775	0.952	1.164	1.413
20	0.528	0.676	0.805	1.017	1.261	1.574
30	0.533	0.675	0.797	1.018	1.237	1.505
50	0.546	0.689	0.802	1.015	1.241	1.557
100	0.538	0.684	0.809	1.055	1.276	1.541
∞	0.529	0.678	0.799	1.011	1.231	1.531

Table 5.6: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 6. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.088	0.108	0.125	0.154	0.179	0.212
10	0.088	0.109	0.125	0.152	0.179	0.213
20	0.089	0.110	0.129	0.160	0.188	0.226
30	0.089	0.109	0.125	0.154	0.185	0.221
50	0.089	0.111	0.129	0.158	0.191	0.230
100	0.088	0.109	0.127	0.157	0.186	0.224
∞	0.089	0.110	0.127	0.156	0.186	0.225

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.573	0.700	0.808	0.993	1.191	1.446
10	0.589	0.716	0.812	0.994	1.179	1.415
20	0.599	0.729	0.837	1.033	1.233	1.515
30	0.599	0.727	0.833	1.011	1.197	1.454
50	0.599	0.729	0.838	1.027	1.229	1.496
100	0.600	0.734	0.836	1.019	1.201	1.439
∞	0.605	0.736	0.839	1.017	1.198	1.442

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.096	0.119	0.135	0.162	0.187	0.222
10	0.096	0.117	0.135	0.162	0.189	0.220
20	0.095	0.119	0.136	0.165	0.196	0.235
30	0.097	0.118	0.138	0.168	0.199	0.242
50	0.096	0.120	0.138	0.172	0.205	0.252
100	0.095	0.118	0.137	0.167	0.199	0.238
∞	0.095	0.119	0.138	0.170	0.202	0.246

Table 5.6: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 6. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.618	0.750	0.868	1.057	1.288	1.594
10	0.635	0.775	0.889	1.088	1.288	1.584
20	0.648	0.794	0.906	1.100	1.304	1.615
30	0.659	0.805	0.917	1.113	1.327	1.599
50	0.662	0.807	0.928	1.145	1.366	1.689
100	0.658	0.800	0.915	1.119	1.321	1.582
∞	0.644	0.789	0.904	1.105	1.309	1.586

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.097	0.119	0.135	0.163	0.191	0.227
10	0.095	0.118	0.134	0.163	0.191	0.222
20	0.096	0.119	0.137	0.167	0.196	0.237
30	0.097	0.119	0.138	0.169	0.200	0.240
50	0.096	0.121	0.140	0.171	0.207	0.251
100	0.095	0.119	0.136	0.168	0.199	0.243
∞	0.096	0.119	0.138	0.170	0.202	0.245

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
5	0.607	0.734	0.835	1.012	1.194	1.461
10	0.621	0.756	0.868	1.042	1.217	1.480
20	0.631	0.771	0.876	1.063	1.263	1.584
30	0.638	0.779	0.894	1.076	1.295	1.546
50	0.644	0.788	0.904	1.116	1.334	1.599
100	0.635	0.777	0.890	1.068	1.272	1.501
∞	0.624	0.765	0.877	1.072	1.270	1.536

Table 5.7: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 7. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.091	0.120	0.150	0.201	0.246	0.319
20	0.093	0.126	0.154	0.205	0.260	0.327
30	0.091	0.127	0.161	0.218	0.275	0.343
50	0.090	0.128	0.156	0.210	0.275	0.350
100	0.091	0.127	0.155	0.210	0.269	0.338
∞	0.091	0.126	0.156	0.209	0.265	0.341

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.525	0.682	0.830	1.092	1.277	1.719
20	0.528	0.714	0.826	1.105	1.336	1.673
30	0.534	0.721	0.860	1.130	1.395	1.763
50	0.531	0.720	0.869	1.132	1.410	1.755
100	0.539	0.711	0.861	1.114	1.385	1.734
∞	0.537	0.713	0.857	1.109	1.369	1.712

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.064	0.081	0.094	0.116	0.137	0.162
20	0.066	0.082	0.096	0.120	0.141	0.174
30	0.066	0.082	0.095	0.118	0.141	0.175
50	0.065	0.082	0.095	0.117	0.139	0.173
100	0.065	0.082	0.097	0.120	0.145	0.175
∞	0.066	0.083	0.098	0.122	0.147	0.180

Table 5.7: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 7. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.478	0.596	0.685	0.855	1.038	1.339
20	0.499	0.627	0.725	0.897	1.088	1.354
30	0.498	0.617	0.721	0.912	1.088	1.360
50	0.500	0.623	0.724	0.899	1.074	1.361
100	0.503	0.636	0.738	0.920	1.108	1.381
∞	0.503	0.635	0.742	0.931	1.125	1.390

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.085	0.111	0.130	0.161	0.195	0.235
20	0.086	0.111	0.130	0.167	0.199	0.251
30	0.086	0.112	0.134	0.169	0.201	0.253
50	0.086	0.112	0.133	0.168	0.205	0.254
100	0.086	0.112	0.134	0.167	0.205	0.255
∞	0.085	0.110	0.129	0.164	0.199	0.248

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.527	0.653	0.761	0.943	1.109	1.329
20	0.534	0.661	0.776	0.946	1.138	1.410
30	0.538	0.675	0.777	0.954	1.153	1.406
50	0.541	0.674	0.777	0.966	1.164	1.381
100	0.539	0.677	0.794	0.967	1.152	1.388
∞	0.530	0.660	0.764	0.945	1.129	1.378

Table 5.8: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 8. The asymptotic points are shown for comparison.

W_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.064	0.076	0.085	0.100	0.115	0.137
20	0.064	0.076	0.085	0.100	0.116	0.135
30	0.063	0.076	0.086	0.103	0.120	0.143
50	0.063	0.075	0.085	0.102	0.120	0.142
100	0.063	0.075	0.084	0.100	0.116	0.140
∞	0.063	0.076	0.086	0.102	0.119	0.140

A_u^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.436	0.512	0.570	0.676	0.777	0.948
20	0.441	0.516	0.572	0.673	0.773	0.902
30	0.442	0.522	0.583	0.693	0.785	0.939
50	0.442	0.519	0.582	0.684	0.783	0.923
100	0.446	0.518	0.576	0.672	0.767	0.893
∞	0.446	0.525	0.585	0.687	0.787	0.919

W_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.101	0.125	0.143	0.174	0.201	0.246
20	0.100	0.124	0.143	0.177	0.210	0.249
30	0.101	0.126	0.144	0.177	0.206	0.244
50	0.101	0.124	0.144	0.174	0.207	0.252
100	0.102	0.127	0.147	0.178	0.212	0.254
∞	0.101	0.126	0.146	0.180	0.213	0.258

Table 5.8: Monte Carlo percentage points for selected sample sizes are given for the Cramér-von Mises statistics for testing for Model 8. The asymptotic points are shown for comparison. (continued)

A_f^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.663	0.810	0.925	1.126	1.351	1.709
20	0.677	0.824	0.939	1.152	1.365	1.674
30	0.688	0.847	0.972	1.159	1.364	1.581
50	0.688	0.836	0.950	1.150	1.337	1.630
100	0.702	0.850	0.969	1.164	1.354	1.633
∞	0.680	0.829	0.947	1.151	1.356	1.631

W_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.106	0.130	0.151	0.182	0.211	0.254
20	0.106	0.131	0.151	0.183	0.219	0.266
30	0.106	0.131	0.150	0.185	0.219	0.260
50	0.107	0.131	0.150	0.185	0.219	0.264
100	0.108	0.133	0.153	0.188	0.220	0.271
∞	0.106	0.132	0.153	0.187	0.223	0.269

A_g^2	Upper tail significance level α					
N	.25	.15	.10	.05	.025	.01
10	0.657	0.799	0.911	1.101	1.302	1.583
20	0.674	0.812	0.915	1.107	1.321	1.612
30	0.674	0.808	0.926	1.132	1.314	1.556
50	0.673	0.816	0.928	1.129	1.324	1.611
100	0.680	0.824	0.941	1.136	1.333	1.593
∞	0.663	0.810	0.926	1.127	1.330	1.603

5.7 Power Comparisons

For the test of the Poisson regression model with one estimated parameter, the overall mean, the power of the Cramér-von Mises statistics has been examined. The following tests of fit were included in the comparison.

1. The Cramér-von Mises statistics defined in Section 5.2.

2. The Deviance. The deviance is the log-likelihood ratio statistic comparing the proposed model against a fully parameterized model (McCullagh and Nelder, 1989). The deviance statistic for Poisson regression is defined

$$\mathbf{D} = 2\sum_{j=1}^{N}[x_j log(x_j/\hat{\mu}_j) - (y_j - \hat{\mu}_j)],$$

   where $\hat{\mu}_j$ is the estimated mean for that observation. This statistic has also been referred to as G^2 (Bishop, Fienberg and Holland, 1975). The test is asymptotically distributed as χ^2_{N-p} where p is the number of estimated parameters, but the chi-square approximation is not considered adequate for small sample sizes (McCullagh and Nelder, 1989). This limits its accuracy as a goodness-of-fit test.

3. Generalized Pearson χ^2 statistic. The generalized Pearson statistic is an extension of the dispersion test defined in section 2.5. The test is defined as

$$\chi^2_P = \sum_{j=1}^{N}\frac{(x_j - \hat{\mu}_j)^2}{\hat{\mu}_j} \qquad (5.41)$$

   (McCullagh and Nelder, 1989).

4. The score test against the negative binomial distribution. This test was proposed by Dean and Lawless (1989) and Dean (1992), generalizing the work of Collings and Margolin (1985). The statistic is

$$P_B = \frac{\{\sum_{j=1}^{N}(x_j - \hat{\mu}_j)^2 - x_j\}}{\sqrt{2\sum_{j=1}^{N}\hat{\mu}_j^2}}. \qquad (5.42)$$

Common alternatives to the Poisson distribution can be distinguished by the ratio of the variance to the mean; this is equal to one for the Poisson distribution. Distributions with variance larger than the mean are considered ***overdispersed***, and with variance smaller than the mean are referred to as ***underdispersed***. The same alternatives used in the power comparisons presented in section 2.5 are examined. The overdispersed alternatives examined were the negative binomial and the Poisson mixture, the binomial was included as an underdispersed alternative and the beta-binomial and the discrete uniform were used as alternatives with dispersion approximately equal to the mean.

Comparisons of power for the Cramér-von Mises statistics and the other tests of fit, when used in testing against the above alternatives, are given in Tables 5.9 and 5.11 for the three Poisson regression models numbered 1, 2 and 8 in the previous section.

Model 1 Two equally proportioned distributions with $\mu_1 = .5$, $\mu_2 = 1.5$;

Model 2 Two equal proportioned distributions with $\mu_1 = 5$, $\mu_2 = 15$;

Model 8 Ten equally proportioned distributions with $\mu_1 = 1$, $\mu_2 = 3$, ..., $\mu_9 = 17$ and $\mu_{10} = 19$.

Random samples of size 20 from a common alternative distribution, with mean equal to the hypothesized Poisson mean, were generated using IMSL subroutines (IMSL, 1987). For example, for a sample of size 20 from a negative binomial alternative with mean structure given by model 2, 10 observations were generated from a negative binomial with a mean 5 and 10 observations were generated from a negative binomial distribution with mean equal to 15.

The critical values (percentage points of the null distribution) for all the test statistics used for comparison were found by Monte Carlo simulation using 10,000 samples. The number of Monte Carlo samples used for the power studies was 1000. The maximum standard error of the power results is equal to $.5/\sqrt{1000} \approx 1.6\%$.

Results and comments

1. The Cramér-von Mises statistics based on the untransformed empirical process, W_u^2 and A_u^2 have generally worse power than the other Cramér-von Mises statistics particularly when the overall mean is large, such as when $\mu = 10$.

2. As expected, the dispersion-based score tests and the deviance statistic perform very well for overdispersed alternatives, with the deviance having slightly better power. The A^2 statistics, A_f^2, A_g^2 and A_r^2, also have good power against overdispersed alternatives. The W^2 statistics have lower power than A^2 statistics.

3. For underdispersed alternatives, the dispersion-based tests have the best power. The Cramér-von Mises statistics have lower power than the dispersion based tests. The W^2 statistics had generally higher power than the A^2 statistics for these alternatives.

4. Against alternatives with the mean equal to the variance, the Cramér-von Mises statistics have the best power. Since the dispersion-based tests primarily detect differences between the mean and variance, they perform very poorly against these alternatives.

5. The gain in power to detect alternatives with similar mean and variance by the use of the Cramér-von Mises statistics over other test statistics is somewhat offset by the greater computational difficulty in calculating the test statistics and their p-values. The computational difficulty increases with the number of unique estimated means.

The Cramér-von Mises statistics A_f^2 and A_g^2 are shown to be powerful statistics for testing for Poisson regression models, particularly if the alternative is "close" to the Poisson in the sense that the variance is almost equal to the mean. The statistic, A_g^2 has slightly better power than A_f^2, and is the recommended statistic for testing for Poisson regression models.

Table 5.9: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20. Alternative distributions were generated with the Model 1 mean structure. The variance of the alternative distribution relative to the Poisson variance is indicated for each distribution. All tests are at the 5% level.

Alternative Distribution	Test Statistics					
	W_u^2	A_u^2	W_f^2	A_f^2	W_g^2	A_g^2
Overdispersed						
Negative Binomial (2)	291	349	322	332	316	437
Poisson Mixture (1.64)	262	324	286	332	248	343
Underdispersed						
Binomial (.5)	77	79	61	70	95	81
Equal Dispersion						
Discrete Uniform	77	60	242	153	264	258

	W_r^2	A_r^2	χ_P^2	Dev	P_B
Overdispersed					
Negative Binomial (2)	267	341	444	457	443
Poisson Mixture (1.64)	209	259	287	393	385
Underdispersed					
Binomial (.5)	79	80	119	121	123
Equal Dispersion					
Discrete Uniform	149	134	144	88	76

Table 5.10: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20. Alternative distributions were generated with the Model 2 mean structure. The variance of the alternative distribution relative to the Poisson variance is indicated for each distribution. All tests are at the 5% level.

Alternative Distribution	Test Statistics					
	W_u^2	A_u^2	W_f^2	A_f^2	W_g^2	A_g^2
Overdispersed						
Negative Binomial (2)	144	295	352	552	342	517
Poisson Mixture (1.64)	166	304	335	568	317	516
Underdispersed						
Binomial (.5)	102	97	390	344	376	346
Equal Dispersion						
Beta-Binomial	710	740	940	929	973	978
Discrete Uniform	72	62	94	68	96	85

	W_r^2	A_r^2	χ_P^2	Dev	P_B
Overdispersed					
Negative Binomial (2)	348	540	604	616	536
Poisson Mixture (1.64)	335	531	544	575	607
Underdispersed					
Binomial (.5)	102	97	484	461	691
Equal Dispersion					
Beta-Binomial	947	940	231	375	346
Discrete Uniform	92	63	18	15	16

Table 5.11: Power Comparison

This table gives the percentage of 1000 samples rejected by the statistics for a sample of size 20. Alternative distributions were generated with the Model 8 mean structure. The variance of the alternative distribution relative to the Poisson variance is indicated for each distribution. All tests are at the 5% level.

Alternative Distribution	Test Statistics					
	W_u^2	A_u^2	W_f^2	A_f^2	W_g^2	A_g^2
Overdispersed						
Negative Binomial (2)	155	250	335	524	318	500
Poisson Mixture (1.64)	213	308	354	580	326	490
Underdispersed						
Binomial (.5)	19	14	448	387	442	426
Equal Dispersion						
Beta-Binomial	98	120	874	872	949	957
Discrete Uniform	58	43	91	73	82	69

	W_r^2	A_r^2	χ_P^2	Dev	P_B
Overdispersed					
Negative Binomial (2)	307	511	591	503	503
Poisson Mixture (1.64)	308	335	547	571	621
Underdispersed					
Binomial (.5)	403	364	504	804	511
Equal Dispersion					
Beta-Binomial	894	896	223	373	342
Discrete Uniform	84	71	10	12	11

5.8 Examples

5.8.1 Example: non-I.I.D. Binomial

The data listed below, taken from Kupper and Haseman (1978), show the numbers (here called successes) of an unspecified laboratory event in pregnant mice. There are ten pregnant mice in each of a treatment and a control group. The data may represent number of fetal abnormalities (events) out of a number of live births (trials). The number of events/trials is given below:

CONTROL GROUP:	0/5, 2/6, 0/7, 0/7, 0/8,
	0/8, 0/8, 1/9, 2/9, 1/10.
TREATMENT GROUP:	0/5, 2/5, 1/7, 0/8, 2/8,
	3/8, 0/9, 4/9, 1/10, 6/10.

Two different models were examined: No treatment difference (common probability of success), and a model with two parameters (separate probabilities of success). In addition, the two treatment groups were individually tested for the binomial distribution.

Treatment Group

The estimated success probability of the treatment group data is 0.241 and the estimated residual variance is 3.374. Let the expected residual binomial variance be defined as

$$(N-p)^{-1}\textstyle\sum_{i=1}^{N} K_i\hat{\theta}_i(1-\hat{\theta}_i),$$

where K_i is the number of trials, $\hat{\theta}_i$ is the estimated success probability for the i observation and p is the number of estimated parameters. Then, the expected residual variance is 1.603, indicating that this data set has greater than binomial dispersion. Figures 5.1-5.6 show plots of the residual empirical distribution function with the average residual distribution function, and plots of the standardized residual empirical process for each of the three empirical processes, labeled $\hat{Y}_{1,N}(j)$, $\hat{Y}_{2,N}(t)$ and $\hat{Y}_{3,N}(t)$, and defined in 5.4.2. The standardized residual empirical process is the value of the residual process divided by its standard deviation to give pointwise asymptotic standard normal values, and is the process used in the calculation of the A^2 test statistics.

The values and significance levels of the Cramér-von Mises statistics and a score test, N_A, with its small sample correction (Dean, 1992), are found in Table 5.12. The score test strongly rejects the binomial hypothesis, whereas the Cramér-von Mises statistics give mixed results. Only A_f^2 and A_g^2 (based on the F- and G-transformed residual processes ,$Y_{2,N}$ and $Y_{3,N}$) reject the binomial hypothesis, although A_u^2 and W_f^2 give near-significant results.

Table 5.12: Test statistics and significance levels for the laboratory data from the treatment group only.

Test Statistic	Value	Significance Level
W_u^2	.08	.292
U_u^2	.08	.293
A_u^2	1.07	.066
W_m^2	.59	.120
W_f^2	.15	.065
U_f^2	.12	.091
A_f^2	1.14	.039
W_g^2	.11	.148
U_g^2	.11	.125
A_g^2	.91	.079
N_A	2.57	.010
N_A(corrected)	2.82	.005

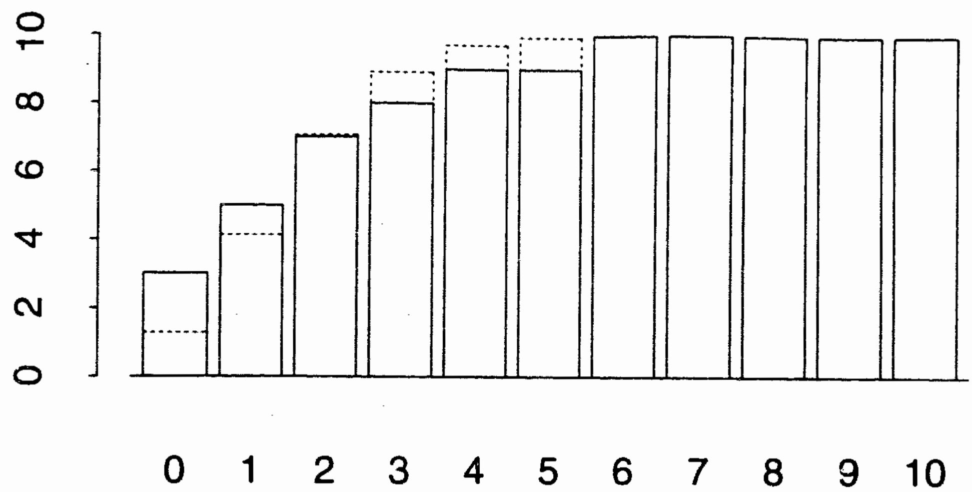

Figure 5.1: Cumulative observed (—) and average expected (- -) histogram for the treatment group laboratory data.

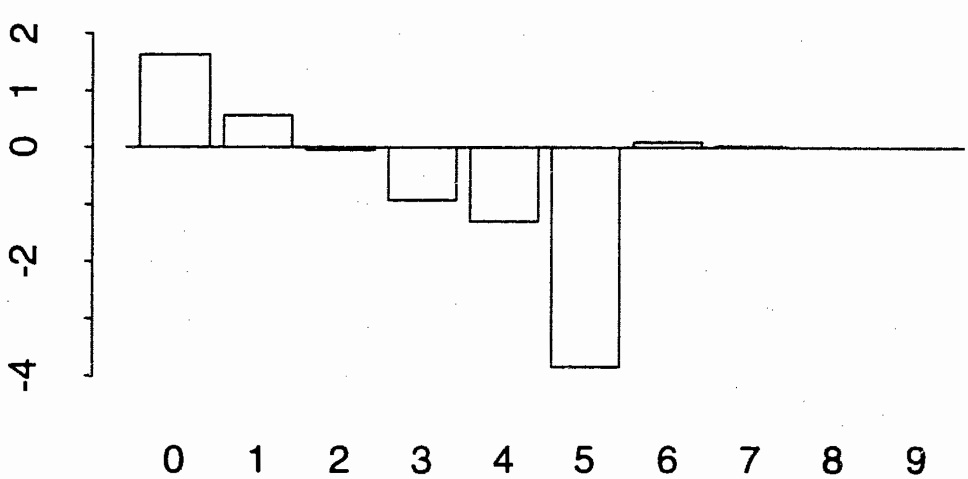

Figure 5.2: Standardized difference between the observed and average expected histogram for the treatment group laboratory data.

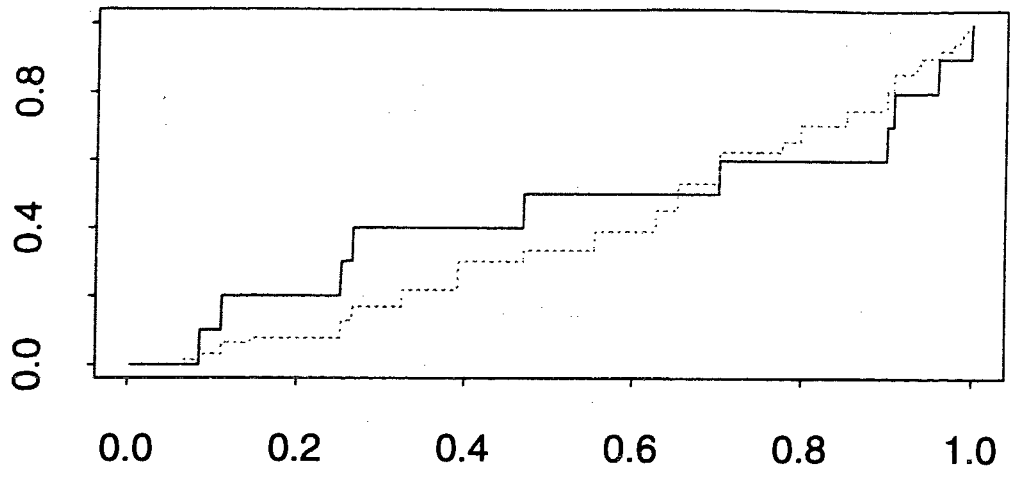

Figure 5.3: F-transformed empirical distribution function (—) and average F-transformed distribution function (- -) for the treatment group laboratory data.

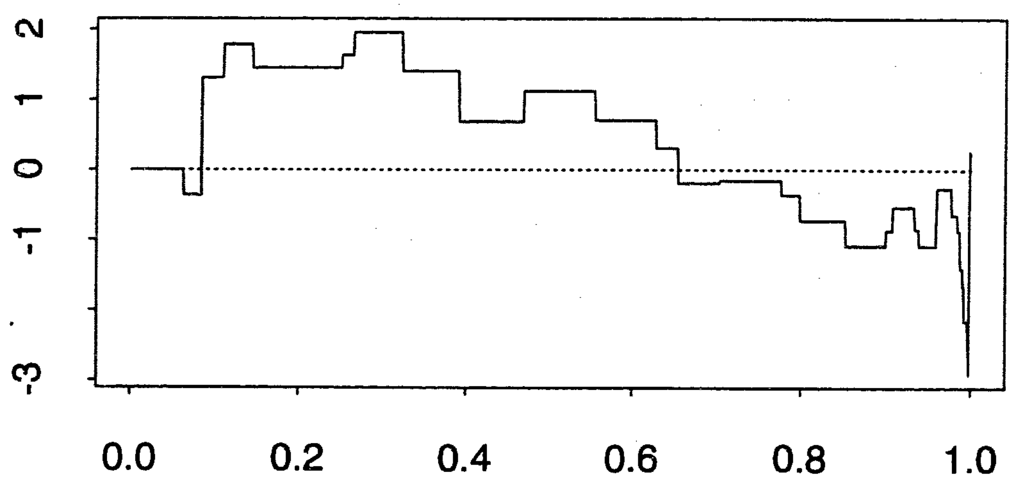

Figure 5.4: Standardized F-transformed residual empirical process plot for the treatment group laboratory data.

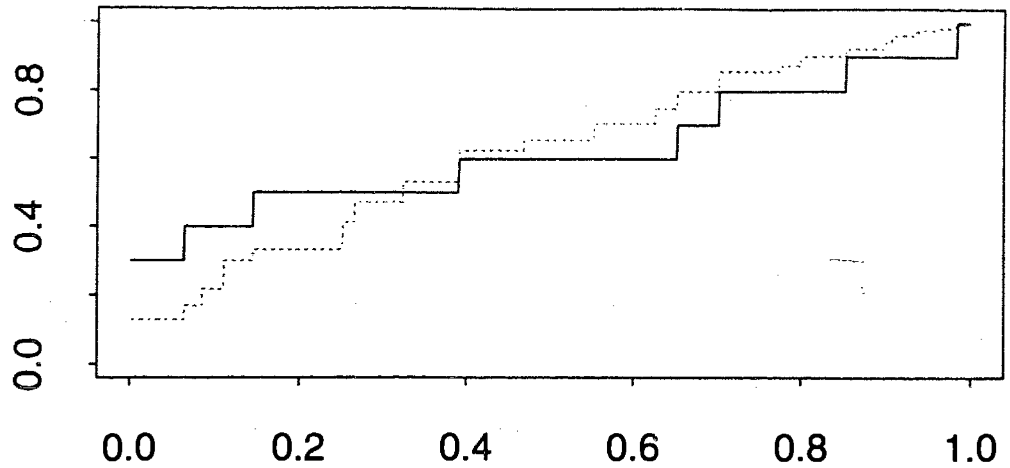

Figure 5.5: G-transformed empirical distribution function (—) and average G-transformed distribution function (- -) for the treatment group laboratory data.

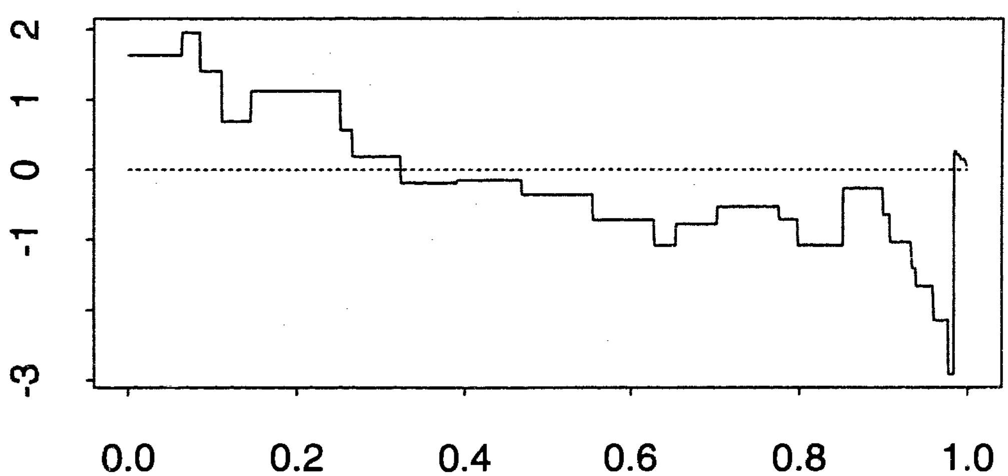

Figure 5.6: Standardized G-transformed residual empirical process plot for the treatment group laboratory data.

Control Group

The plots of the (F-transformed) residual empirical distribution function with the average residual distribution function, and the standardized residual empirical process are shown in Figures 5.7 and 5.8. The estimated probability of success is 0.078 and the estimated residual variance is 0.676. The expected residual binomial variance is 0.615, which supports binomial dispersion. The values and significance levels of the Cramér-von Mises statistics and the score test for overdispersion are found in Table 5.13. All test statistics failed to reject the Binomial hypothesis.

Table 5.13: Test statistics and significance levels for the laboratory data from the control group only.

Test Statistic	Value	Significance Level
W_u^2	0.04	.36
U_u^2	0.04	.33
A_u^2	0.32	.40
W_m^2	0.11	.37
W_f^2	0.03	.34
U_f^2	0.03	.29
A_f^2	0.21	.30
W_g^2	0.05	.34
U_g^2	0.05	.31
A_g^2	0.35	.36
N_A	0.24	.81
N_A(corrected)	0.48	.63

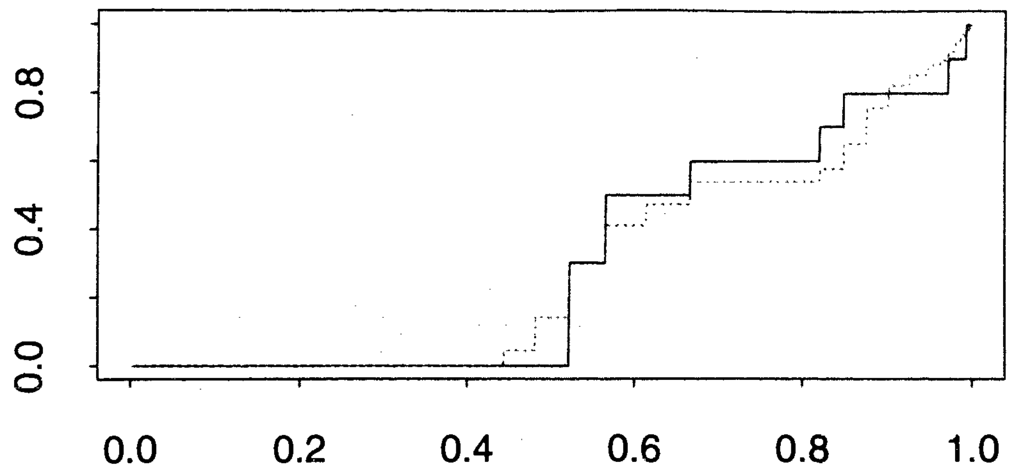

Figure 5.7: F-transformed empirical distribution function (—) and average F-transformed distribution function (- -) for the control group laboratory data.

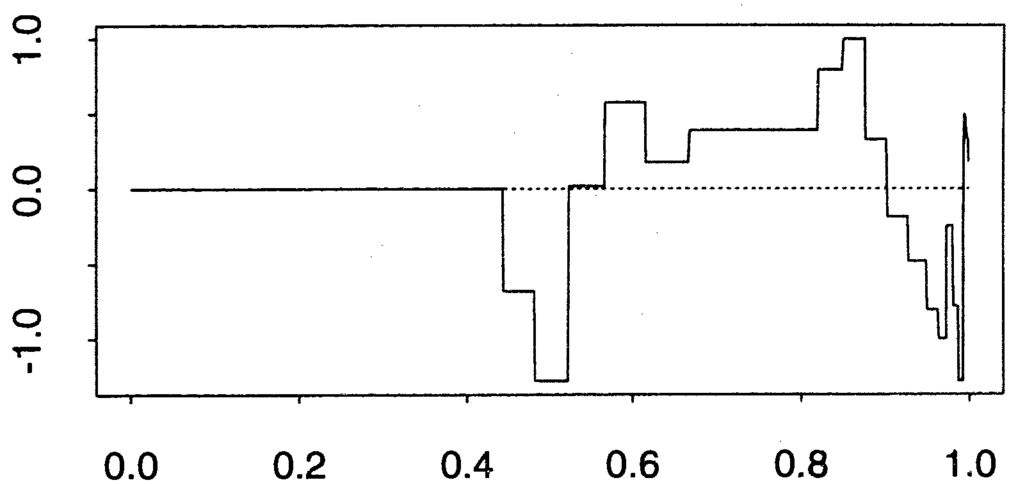

Figure 5.8: Standardized residual empirical process plot (F-transformed) for the control group.

Total Group

Two models were examined:

1. The model with a single probability of success, estimated to be 0.160.
2. The model with separate probabilities of success for each of the treatment and control groups, estimated to be 0.02405 and 0.078, respectively.

The residual plots of the standardized empirical process (F-transformation) for each of the two groups are found in Figures 5.9 and 5.10, respectively. The values and significance levels of the Cramér-von Mises statistics and the score test for overdispersion, for each of the two models, are found in Table 5.14. For the model with a single probability of success, all the tests, with the exception of W_u^2 and U_u^2, strongly reject the binomial model. The statistics, W_u^2 and U_u^2 show only weak evidence against this model. For the model with separate probabilities of success, the results are not as consistent, although most statistics give weak evidence against the model. The corrected score statistic gives the most significant result.

Table 5.14: Test statistics and significance levels for the laboratory data - single success probability and separate success probability models.

Test Statistic	Common Success Prob. Value	Common Success Prob. Significance Level	Separate Success Prob. Value	Separate Success Prob. Significance Level
W_u^2	0.20	.045	0.10	.159
U_u^2	0.17	.065	0.07	.229
A_u^2	1.84	.009	0.73	.092
W_m^2	0.88	.020	0.38	.150
W_f^2	0.29	.003	0.07	.196
U_f^2	0.20	.009	0.05	.315
A_f^2	1.62	.005	0.94	.051
W_g^2	0.25	.008	0.12	.100
U_g^2	0.23	.011	0.11	.084
A_g^2	1.79	.004	0.78	.089
N_A	3.88	<.001	2.01	.044
N_A(corrected)	4.05	<.001	2.80	.005

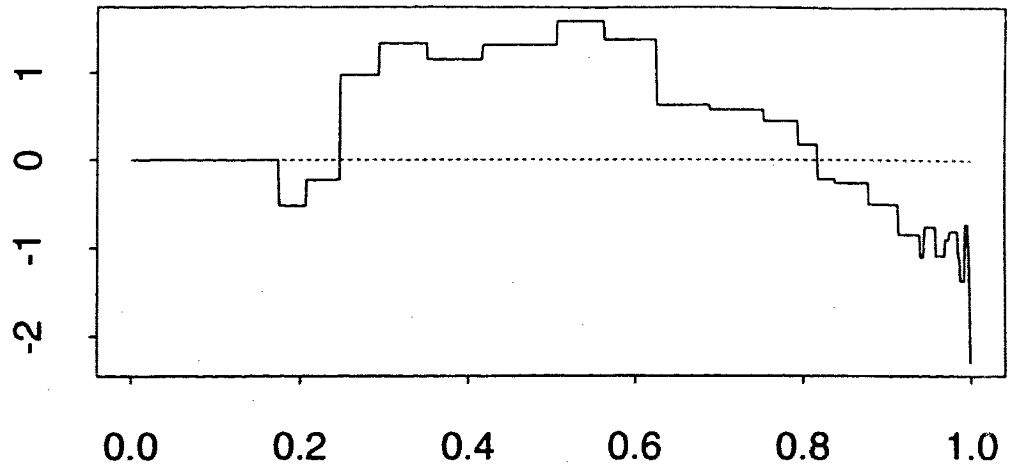

Figure 5.9: Standardized residual empirical process plot (F-transformed) for the model with a single probability of success.

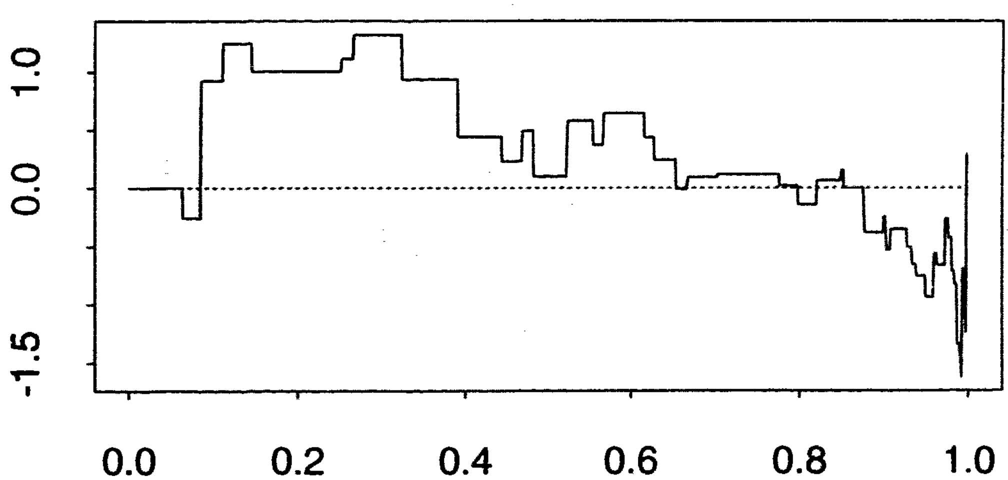

Figure 5.10: Standardized residual empirical process plot (F-transformed) for the model with separate probabilities of success.

5.8.2 Example: Poisson Regression

The data in Table 5.15 come from a study of cancer in 4213 male aluminum workers (Spinelli et al, 1991). The data were combined into 44 subgroups created by a cross-classification of exposure and age. The variables are

1. Exposure to Coal Tar Pitch Volatiles (1=<1 BSM-year of exposure, 2=1-5 BSM-years of exposure, 3=5-10 BSM-years of exposure, 4=10+ BSM-years of exposure)

2. Age (1=20-29, 2=30-34, ... 10=70-74, 11=75+).

3. Number of bladder cancer cases.

4. Person-years at risk in that sub-group.

Table 5.15: Bladder cancer in aluminum workers data.

Subgroup	Exposure Level	Age	No. of Cases	Person-Years at Risk
1	1	1	0	1332.11
2	1	2	0	1388.75
3	1	3	1	1696.86
4	1	4	0	2061.61
5	1	5	0	2183.38
6	1	6	0	2068.96
7	1	7	0	1711.92
8	1	8	1	1246.69
9	1	9	1	773.91
10	1	10	0	412.83
11	1	11	1	225.61
12	2	1	0	1615.40
13	2	2	0	1866.16
14	2	3	0	2009.76
15	2	4	0	2081.41
16	2	5	0	1966.08
17	2	6	1	1557.57
18	2	7	0	1048.04
19	2	8	0	635.21
20	2	9	0	371.64
21	2	10	0	204.39
22	2	11	0	83.41
23	3	1	0	200.91
24	3	2	0	678.42
25	3	3	0	1190.94
26	3	4	0	1482.89
27	3	5	0	1535.35
28	3	6	1	1362.64
29	3	7	0	851.45
30	3	8	0	456.51
31	3	9	0	215.72
32	3	10	0	99.26
33	3	11	1	31.79
34	4	1	0	3.12
35	4	2	0	102.99
36	4	3	0	420.73
37	4	4	0	1136.63
38	4	5	0	1564.33
39	4	6	1	1587.80
40	4	7	2	1102.41
41	4	8	3	663.04
42	4	9	0	337.23
43	4	10	3	102.19
44	4	11	0	11.55

The main purpose of the study was to determine if the risk of bladder cancer increased with increasing exposure to coal tar pitch volatiles. A Poisson regression model (Poisson error, log link) was fitted. Age was treated as a factor and exposure was analyzed as a continuous covariate to assess trend. The following models were fitted to the data.

1. Constant + Person-Years(Offset)
2. Constant + Person-Years(Offset) + Age
3. Constant + Person-Years(Offset) + Exposure
4. Constant + Person-Years(Offset) + Age + Exposure

The residual plots of the standardized empirical process (F-transformation) for each of the four models are found in Figures 5.11 to 5.17, respectively. Table 5.16 shows the significance levels of the Cramér-von Mises test statistics and a score test for overdispersion (against the beta-binomial distribution), P_B (Dean, 1992). The score test and a small sample corrected version of the score test are presented.

Table 5.16: Test statistics and significance levels for the laboratory data - single success probability and separate success probability models.

	Significance Level			
Test Statistic	Model 1	Model 2	Model 3	Model 4
W_u^2	.336	.893	.690	.871
U_u^2	.419	.816	.718	.789
A_u^2	.267	.763	.626	.804
W_m^2	.290	.715	.602	.757
W_f^2	.002	.344	.003	.512
U_f^2	.010	.414	.021	.451
A_f^2	<.001	.386	.026	.428
W_g^2	.227	.771	.105	.817
U_g^2	.122	.649	.070	.755
A_g^2	<.001	.343	<.001	.731
P_B	.002	.227	.026	.335
P_B(corrected)	<.001	.029	.012	.871

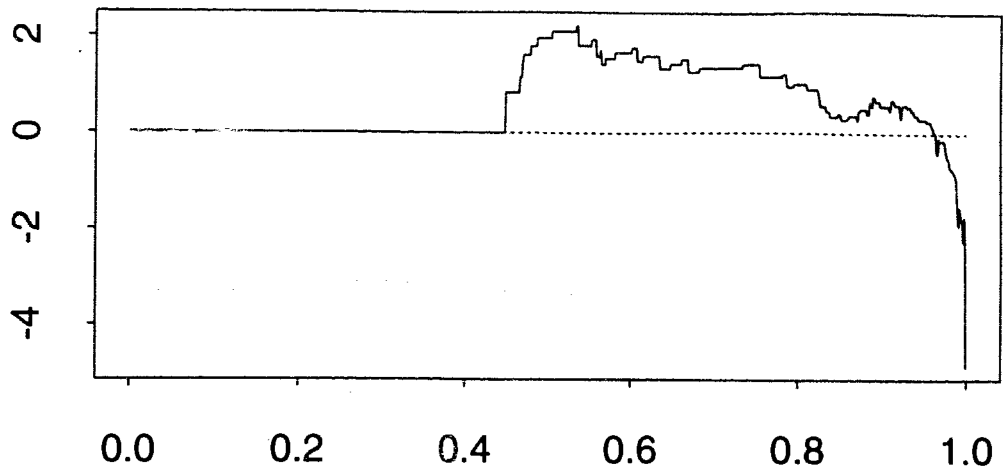

Figure 5.11: Standardized residual empirical process plot (F-transformed) for Model 1, Offset only.

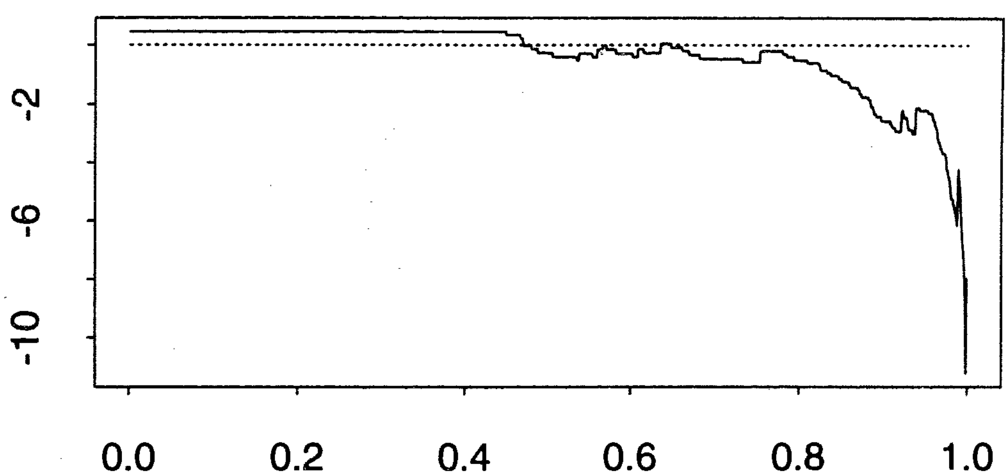

Figure 5.12: Standardized residual empirical process plot (G-transformed) for Model 1, Offset only.

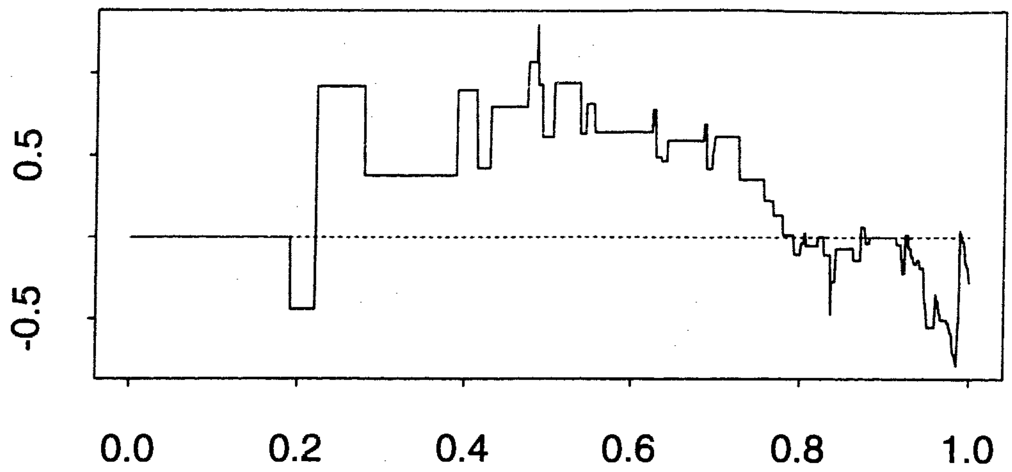

Figure 5.13: Standardized residual empirical process plot (F-transformed) for Model 2, Offset + Age.

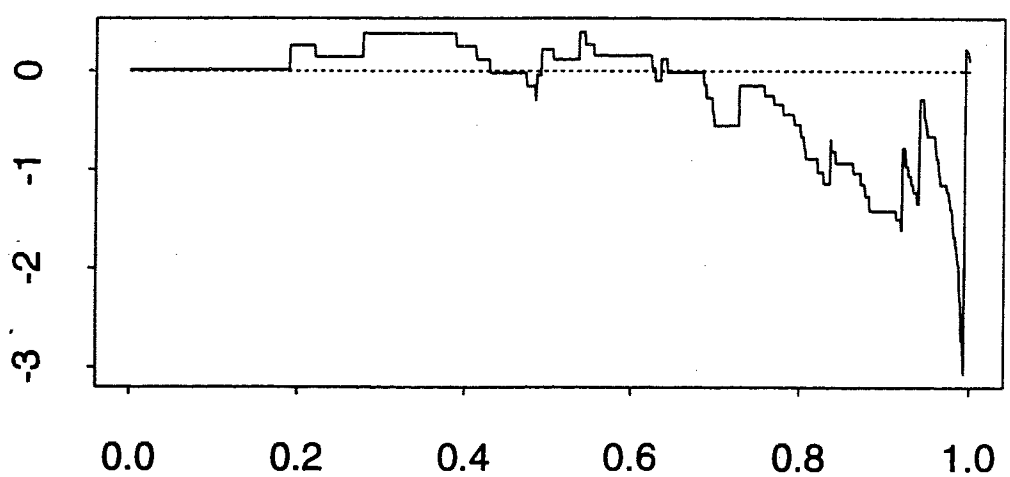

Figure 5.14: Standardized residual empirical process plot (G-transformed) for Model 2, Offset + Age.

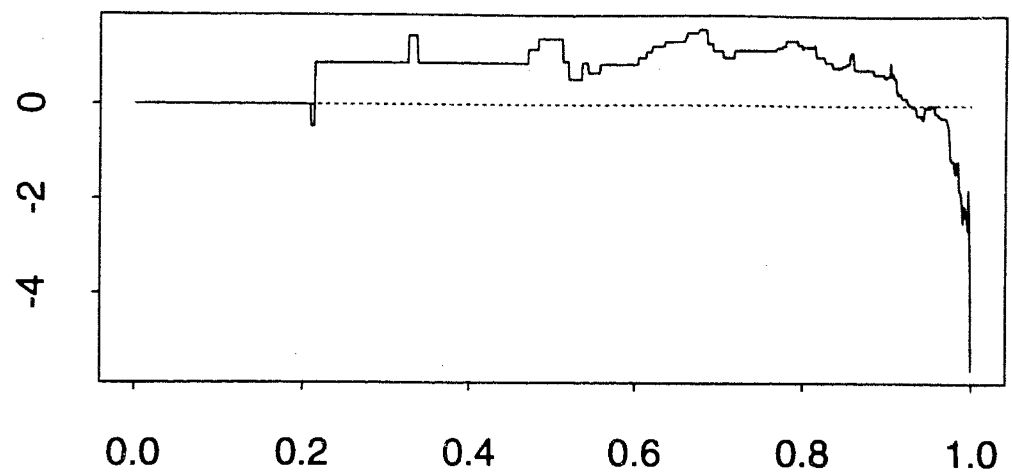

Figure 5.15: Standardized residual empirical process plot (F-transformed) for Model 3, Offset + Exposure.

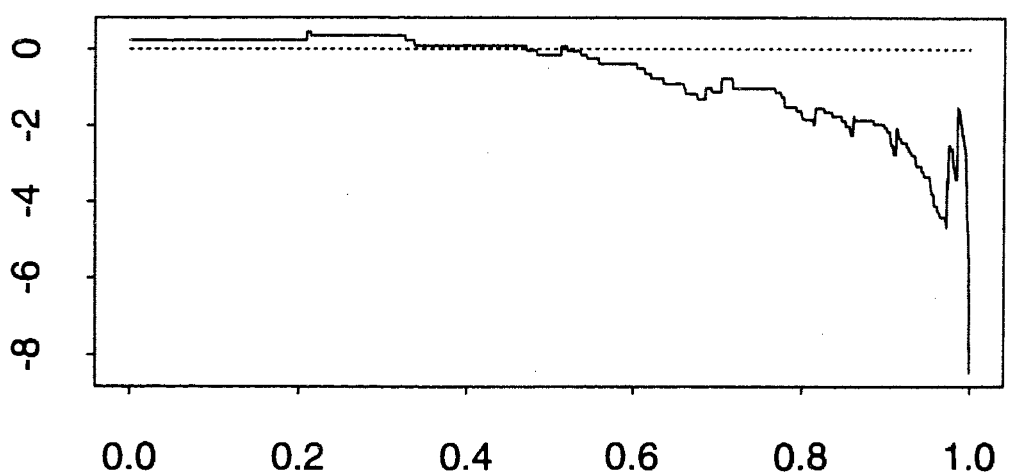

Figure 5.16: Standardized residual empirical process plot (G-transformed) for Model 3, Offset + Exposure.

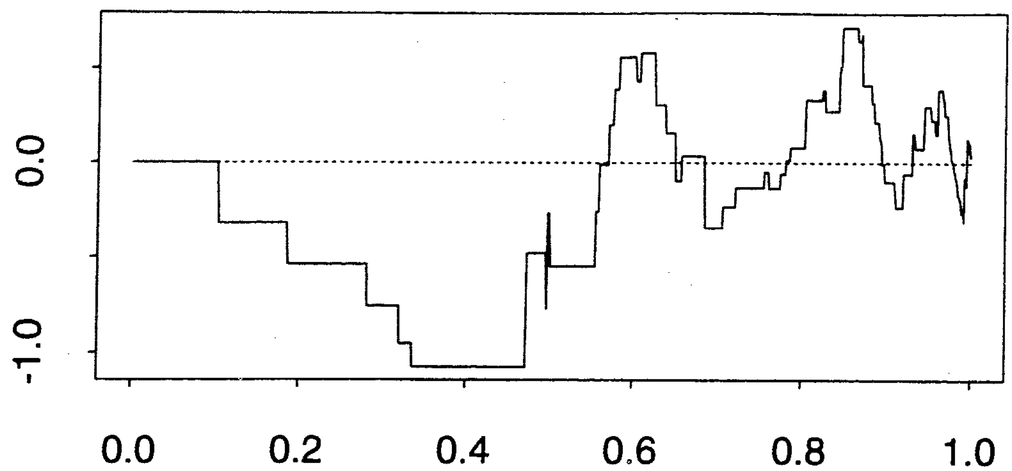

Figure 5.17: Standardized residual empirical process plot (F-transformed) for Model 4, Offset + Age + Exposure.

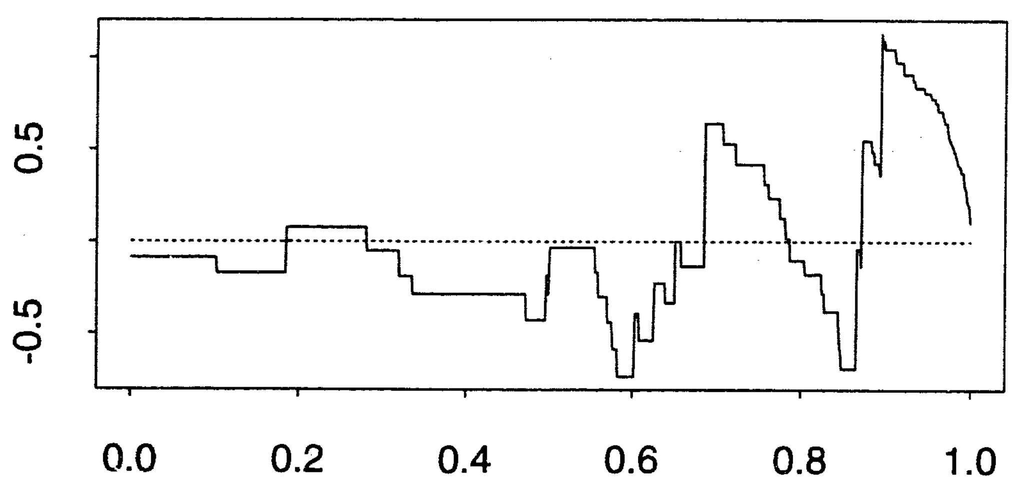

Figure 5.18: Standardized residual empirical process plot (G-transformed) for Model 4, Offset + Age + Exposure.

As would be expected from the Monte Carlo power simulations in section 5.7, the Cramér-von Mises statistics based on the untransformed process have little power to detect departures from the Poisson model. The statistics, A_f^2, A_g^2, W_f^2 and P_B clearly reject Models 1 and 3. Similarly, all the test statistics accept the Poisson hypothesis for Model 4, the final model proposed. There is some disagreement between the tests for Model 2. All the Cramér-von Mises statistics accept the Poisson hypothesis for this model. The uncorrected score test also accepts the model, whereas the corrected score test rejects the Poisson model. For model 2, the estimated residual variance is 0.676. The estimated Poisson variance, defined as $(N-p)^{-1}\sum_{i=1}^{N}\hat{\mu}_i$, where $\hat{\mu}$ is the estimated mean for the ith observation, is 0.485, indicating slight overdispersion. An examination of the residual processes in Figures 5.13 and 5.14 indicate that the data fit the Poisson model fairly well.

Bibliography

[1] Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.*, 23, 193-212.

[2] Barton, D.E. (1955). A form of Neyman's psi-2 test of goodness-of-fit applicable to grouped and discrete data. *Skand. Aktuartdskr.*, 39, 1-17.

[3] Becker, R.A., Chambers, J.M. and Wilks, A.R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics.* Wadsworth and Brooks/Cole, Pacific Grove, California.

[4] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[5] Bishop, Y.M., Feinberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis*, MIT Press, Cambridge.

[6] Campbell, D.B. and Oprian, C.A. (1979). On the Kolmogorov-Smirnov test for the Poisson distribution with unknown mean. *Biom. J.*, 21, No. 1, 17-24.

[7] Choulakian, V., Lockhart, R.A. and Stephens, M.A. (1994). Cramér-von Mises statistics for discrete distributions. *Canadian Journal of Statistics*, in print.

[8] Collings, B.J. and Margolin, B.H. (1985). Testing goodness-of-fit for the Poisson assumption when observations are not identically distibuted. *J. Amer. Statist. Assoc.*, 80, No. 390, 411-418.

[9] Conover, W.J. (1972). A Kolmogorov goodness of fit test for discontinuous distributions. *J. Amer. Statist. Assoc.*, 67, No. 339, 591-596.

[10] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

[11] Cramér, H. (1928). On the composition of elementary errors. Second paper: Statistical Applications. *Skand. Aktuartdskr.*, 11, 141-180.

[12] Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *J. Amer. Statist. Assoc.*, 87, No. 418, 451-457.

[13] Dean, C. and Lawless, J.F. (1989). Tests for detecting overdispersion in Poisson regression models. *J. Amer. Statist. Assoc.*, 84, No. 406, 467-472.

[14] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.*, 23, 277-281.

[15] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.*, 20, 393-403.

[16] Durbin, J. and Knott, M. (1972). Components of Cramér-Von Mises Statistics, I. *J. Roy. Statist. Soc.*, B 34, 216-237.

[17] Durbin, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, 1, 279- 209.

[18] Durbin, J. (1973b). *Distribution Theory for Tests Based on the Sample Distribution Function*, Regional Conference Series in Applied Mathematics, 9, SIAM, Philadelphia.

[19] Durbin, J., Knott, M. and Taylor, C.C. (1975). Components of Cramér-von Mises statistics. *J. Roy. Statist. Soc.*, B 34, 290-307.

[20] Guttorp, P. and Lockhart, R.A. (1988). On the asymptotic distribution of quadratic forms in uniform order statistics, *Ann. Statist.*, 16, No. 1, 433-449.

[21] Hoaglin, D.C. (1980). A Poissonness plot. *Amer. Statistician*, 34, 146-149.

[22] Imhoff, J.P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419-426.

[23] *IMSL, Problem-Solving Software Systems*, soft cover edition, (1987), IMSL, Inc, Houston, Texas.

[24] Johnson, N.L., Kotz, S and Kemp, A. (1992). *Univariate Discrete Distributions*, Second Edition, Wiley, New York.

[25] Kac, M. and Siegert, A.J.F. (1947). An explicit representation of a stationary Gaussian process. *Ann. Math. Statist.*, 18, 438-442.

[26] Kendall, M.G. and Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th edition, Charles Griffin & Company Limited, London.

[27] Kocherlakota, S. and Kocherlakota, K. (1986). Goodness of fit tests for discrete distributions. *Comm. Statist. Theory Methods*, 15, 815-829.

[28] Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari.*, 4, 83-91.

[29] Kuiper, N.H. (1960). Tests concerning random points on a circle. *Koninkl. Neder. Aka. van. Wetenschappen.*, A 63, 38-47.

[30] Kupper, L.L. and Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, 34, 69-76.

[31] Loynes, R.M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.*, 8, 285 -298.

[32] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, second edition, Chapman and Hall, London.

[33] Mukantseva, L.A. (1977). Testing normality in one dimensional and multidimensional linear regression. *Theor. Probability Appl.*, 19, 516-519.

[34] Nakamura, M. and Perez-Abreu, V. (1993). An empirical probability generating function approach for testing a Poisson model. *Canadian Journal of Statistics*, 21,149-156.

[35] Neyman, J. (1937). "Smooth" tests for goodness-of-fit. *Skand. Aktuarietidskr.*, 20, 149-199.

[36] Pearson, Karl. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in a random sampling. *Phil. Mag.*, (5) 50, 157.

[37] Pettitt, A.N. and Stephens, M.A. (1977). The Kolmogorov-Smirnov goodness of fit statistic with discrete grouped data. *Technometrics*, 19, No.2, 205-210.

[38] Pierce, D.A. and Kopecky, K.J. (1979) Testing goodness-of-fit for the distribution of errors in regression models. *Biometrika*, 66, 977-986.

[39] Potthoff, R.F. and Whittinghill, M. (1966a). Testing for homogeneity: I. The binomial and multinomial distribution, *Biometrika*, 53, 167-182.

[40] Potthoff, R.F. and Whittinghill, M. (1966b). Testing for homogeneity: II. The Poisson distribution, *Biometrika*, 53, 183-190.

[41] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley and Sons, New York.

[42] Rayner, J.C.W. and Best, D.J. (1989). *Smooth Tests of Goodness of Fit*, Oxford University Press, New York.

[43] Rueda, R., Perez-Abreu, V. and O'Reilly, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Comm. Statist. Theory Methods*, A20. 3093-3110.

[44] *S-Plus Reference Manual 3.0.* (1991). Statistical Sciences Inc., Seattle, Washington.

[45] Scott, E.L. (1949). Distribution of the longitude of periastron of spectroscopic binaries. *Astrophysical Journal*, 109, 194-207.

[46] Shorak, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[47] Smirnov, N.V. (1939). Sur les escartes de la courbe de distribution empirique. *Mat. Sbornik (N.S.)*, 6, (48), 3-26.

[48] Smirnov, N.V. (1936). Sur la distribution de (Criterium de R.v.Mises). *C.R. Acad. Sci. Paris*, 202, 449-452.

[49] Solomon, H and Stephens, M.A. (1977). Approximations to density functions using Pearson curves. *J. Amer. Statist. Assoc.*,73, No. 361, 153-160.

[50] Solomon, H. and Stephens, M.A. (1978). Distribution of a sum of weighted chi-square variables. *J. Amer. Statist. Assoc.*, 72, No. 360, 881-885.

[51] Spinelli, J.J. (1980). Regression and EDF Tests of Fit. M.Sc. Thesis, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia.

[52] Spinelli, J.J., Band P.R., Svirchev, L.M. and Gallagher, R.P. (1991). Mortality and cancer incidence in aluminum reduction plant workers. *J. Occup. Med.*, 33, 1150-1155.

[53] Stephens, M.A. (1976b). Asymptotic power of EDF statistics for exponentiality against gamma and Weibull alternatives. Technical Report No. 297, Department of Statistics, Stanford University.

[54] Stephens, M.A. (1976a). Asymptotic results and percentage points for goodness-of-fit statistics with unknown parameters. *Ann. Statist.*, 4, No. 2, 357-369.

[55] Stephens, M.A. (1986). Tests based on EDF statistics, *Goodness of Fit Techniques*, Chapter 4, edited by D'Agostino and Stephens, M.A., Marcel Dekker Inc., New York.

[56] Stuart, A. and Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1, 5th ed.*, Oxford University Press, New York.

[57] Tarone, R.E. (1979). Testing the goodness-of-fit of the binomial distribution. *Biometrika*, 66, 585-590.

[58] Watson, G.S. (1966). Goodness of fit tests on a circle. *Biometrika*, 48, 109-114.

[59] de Wet, T. and Venter, J.H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.*, 1, No. 2, 380-387.

[60] Zar, J. (1974). *Biostatistical Analysis*, Prentice-Hall, Englewood Cliffs, NJ.