Continuity Correction for the Score Statistic in Discrete Regression Models

By John E. Kolassa[1]

Rutgers University Department of Statistics

This paper corrects the usual chi–squared approximation to the distribution function of the conditional score statistic in a generalized linear model, when the underlying distribution is discrete. The proposed method corrects by a multiple of the difference between the number of sufficient statistics lying in the acceptance region for the test and the volume of this region. The multiplier is calculated from the multivariate Edgeworth approximation to the distribution of a lattice random vector.

1. Introduction. This paper addresses the problem of hypothesis testing in generalized linear models in the presence of a canonical nuisance parameter. Marginal approaches involving the score statistic are fully efficient [8], but have the drawback that the sampling distribution of the test statistic depends, at least weakly, on the nuisance parameter. Conditional inference avoids problems arising from this dependence, often at a cost in efficiency that is not particularly severe [7]. This paper applies a continuity correction to the standard χ^2 approximation to the distribution of the conditional score statistic.

When the distribution of raw responses is continuous, standard Edgeworth techniques may be employed to improve on the usual normal theory approximation to the test statistic sampling distribution. When the distribution of raw responses is discrete, standard Edgeworth series results do not apply. This paper applies a continuity correction to the estimation of probabilities associated with the conditional scores statistic arising in generalized linear models. Approximations to the conditional expectation and variance, as presented by Waterman and Lindsay [18], are used in conjunction with a first Edgeworth correction term calculated by Yarnold [17], to accurately approximate p-values.

Section 2 reviews the multivariate Edgeworth series, and section 3 discusses estimation of probabilities for ellipses that arise from score testing, and reviews an adjustment to standard approximations that accounts for the lattice nature of certain regression models. Section 4 reviews generalized linear model notation. Section 5 presents an artificial multinomial example, and section 6 presents an example concerning cancer remissions.

2. Multivariate Edgeworth Series. Suppose $\boldsymbol{X} = (X^r)$ is a random vector in $\boldsymbol{R}^m$, such that

$$\mathrm{E}[\boldsymbol{X}] = \mathbf{o} \text{ and } \mathrm{Var}[\boldsymbol{X}] = \boldsymbol{\Sigma}. \tag{1}$$

[1]This research was supported by grant R01 CA 63050 from the National Cancer Institute.

Suppose that

(2) $$\mathrm{E}\left[\|\boldsymbol{X}\|^{p}\right] < \infty, \text{ for some } p \geq 2.$$

Let κ_ν be the generic multivariate cumulant with indices given by entries in the vector ν of elements of $\{1, \ldots, m\}$. For instance, κ_j is the expectation of X^j, κ_{jk} is the covariance of X^j and X^k, and κ_{jkl} is the third order mixed cumulant of X^j, X^k, and X^l. Let $\boldsymbol{\kappa}$ be the collection of these cumulants. Suppose $\{\boldsymbol{X}_i\}$ is an independent and identically distributed collection of copies of $\boldsymbol{X}$. Let

(3) $$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n),$$

and let $\boldsymbol{\kappa}^n$ be its cumulants. Then $\kappa^n_\nu = n^{1-|\nu|/2}\kappa_\nu$. Here $|.|$ applied to a vector denotes its length. These cumulants are generated by expanding the cumulant generating function $\mathcal{K}_n(\boldsymbol{\beta}) = \log E[e^{\boldsymbol{\beta}^\top \boldsymbol{T}_n}]$ as

(4) $$\mathcal{K}_n(\boldsymbol{\beta}) = \sum_{r=1}^{\infty} \frac{1}{r!} \sum_{|\boldsymbol{\nu}|=r} \kappa^n_{\boldsymbol{\nu}} \boldsymbol{\beta}^{\boldsymbol{\nu}}.$$

The above inner sum over $\boldsymbol{\nu}$ includes all vectors of integers in $\{1, \ldots, m\}$ of length r. The cumulant generating function always exists for arguments $\boldsymbol{\beta}$ such that $i\boldsymbol{\beta} \in \mathfrak{R}^m$; the following exposition requires no further condition on $\mathcal{K}_n$.

The following is similar to the treatment of series expansions by Bhattacharya and Rao [2] and Chambers [3]. McCullagh [13] presents an alternative treatment.

The Edgeworth series consists of terms in the term-wise Fourier inversion of (4), after exponentiating non–quadratic terms, and discarding terms that are sufficiently small in n. Since the power of n is only indirectly related to the order in the β_r, introduce the extra variable τ to account for it. Define polynomials $P_r(\boldsymbol{\beta}; \boldsymbol{\kappa}^n)$ by the power series

(5) $$\sum_{r=0}^{k-2} P_r(\boldsymbol{\beta}; \boldsymbol{\kappa}^n)\tau^r = \exp\left[\sum_{r=3}^{\infty} \frac{1}{r!} \sum_{|\boldsymbol{\nu}|=r} \boldsymbol{\kappa}^n_{\boldsymbol{\nu}} \boldsymbol{\beta}^{\boldsymbol{\nu}} \tau^{r-2}\right].$$

For each r, let $P_r(\Phi_{0,\boldsymbol{\Sigma}}; \boldsymbol{\kappa}^n)$ be the polynomial $P_r(\boldsymbol{\beta}; \boldsymbol{\kappa}^n)$ with $(-1)^{|\boldsymbol{\nu}|} D^{\boldsymbol{\nu}} \Phi_{0,\boldsymbol{\Sigma}}$ substituted for $\boldsymbol{\beta}^{\boldsymbol{\nu}}$ for all vectors ν, where $\Phi_{0,\boldsymbol{\Sigma}}$ is the distribution function for a normal random variable with mean 0 and covariance matrix $\boldsymbol{\Sigma}$. Define the Edgeworth series approximation to the distribution function F_n of T_n using cumulants up to the order k as follows:

(6) $$E_k(\boldsymbol{t}; \boldsymbol{\kappa}^n) = \sum_{r=0}^{k-2} P_r(\Phi_{0,\boldsymbol{\Sigma}}; \boldsymbol{\kappa}^n).$$

Bhattacharya and Rao [2] show that

(7) $$F_n(\boldsymbol{t}) = E_k(\boldsymbol{t}; \boldsymbol{\kappa}^n) + o(n^{-\frac{k-2}{2}}),$$

uniformly in $\boldsymbol{t}$, as long as Cramér's condition

(8) $$\limsup_{\|\boldsymbol{\beta}\| \to \infty, \boldsymbol{\beta} \in \mathfrak{R}^m} |\exp(\mathcal{K}(i\boldsymbol{\beta}))|$$

holds.

When $\{\boldsymbol{X}_i\} = \{(X_i^r)\}$ is a collection of independent and identically distributed random vectors in $\mathfrak{R}^m$, with cumulative distribution function F and cumulants $\boldsymbol{\kappa}$, and such that

$$\mathrm{P}\left[\boldsymbol{X}_i \in \boldsymbol{y} + \mathfrak{Z}^m\right] = 1, \tag{9}$$

for $\mathfrak{Z}$ the integers, define T_n, F_n, and $\boldsymbol{\kappa}^n$ as before. Condition (8) fails, and Bhattacharya and Rao [2] derive the following alternative approximation to F_n:

$$A_{n,k}(\boldsymbol{t};\boldsymbol{\kappa}) = \sum_{|\boldsymbol{\alpha}|\leq k-2} S_{\boldsymbol{\alpha}}(\boldsymbol{t} - \sqrt{n}\boldsymbol{y})(-1)^{|\boldsymbol{\alpha}|} n^{-|\boldsymbol{\alpha}|/2} D^{\boldsymbol{\alpha}} E_k(\boldsymbol{t};\boldsymbol{\kappa}^n); \tag{10}$$

here

$$S_{\boldsymbol{\alpha}}(\boldsymbol{x}) = (-1)^{|\boldsymbol{\alpha}|} \prod_{j=1}^{m} g_{\alpha_j} Q_{\alpha_j}(\sqrt{n}x_j)$$

and

$$Q_\nu(x) = \begin{cases} \sum_{j=1}^{\infty} \cos(2j\pi x)/(2^{\nu-1}(j\pi)^\nu) & \text{if } \nu \text{ even} \\ \sum_{j=1}^{\infty} \sin(2j\pi x)/(2^{\nu-1}(j\pi)^\nu) & \text{if } \nu \text{ odd} \end{cases}$$
$$g_\nu = \begin{cases} +1 & \text{if } \nu = 4k+1 \text{ or } \nu = 4k+2 \\ -1 & \text{if } \nu = 4k-1 \text{ or } \nu = 4k \end{cases}.$$

The Q_ν are piecewise polynomial on intervals of the form $[z, z+1)$ for integer z, and their versions on $[0,1)$ are multiples of the Bernoulli polynomials. All are continuous except for the first. This generalizes the univariate result of Esseen [4].

3. Multivariate Testing. Consider the problem of testing a hypothesis about a model having the sufficient statistic vector $\boldsymbol{T} \in \mathfrak{R}^m$, of form (3), satisfying (2), and with the null hypothesis implying (1). Suppose that $\boldsymbol{\Sigma}$ is invertible. Use the test statistic $V(\boldsymbol{T}) = \boldsymbol{T}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{T}$, under (2). This statistic is a trivial example of a Wald statistic, and in a full exponential family is the score statistic. When $\boldsymbol{T}$ is approximately multivariate normal, V has a distribution that is approximately χ^2 on m degrees of freedom. This may be seen by expressing $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}^\top \boldsymbol{\Omega}$, and noting that $\boldsymbol{\Omega}(\boldsymbol{T} - \boldsymbol{\mu})$ is approximately multivariate normal with mean $\mathbf{o}$ and all components independent with unit variance.

Let $\mathfrak{E}_n = \{\boldsymbol{t} | \boldsymbol{t}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{t} \leq v\}$ be the elliptical set of $\boldsymbol{t}$ giving rise to $V \leq v$, which may depend on n through the support for $\boldsymbol{T}$. The multivariate Edgeworth series allows the approximation of probabilities of sets like $\mathfrak{E}_n$, as long as $\boldsymbol{T}$ satisfies regularity conditions. Kolassa [10] presents a discussion of these.

When $\boldsymbol{T}$ has a lattice distribution, approximating probabilities of elliptical regions becomes tricky. Kolassa and McCullagh [12] show that in the unidimensional lattice case, the Edgeworth series gives approximations valid to $O(n^{-1})$ if evaluated at continuity corrected points and using the third cumulant. The Edgeworth series gives higher order approximations if the cumulants are adjusted and fourth and higher order cumulants are used. In the multivariate case Esseen [4] showed as Theorem 1 of §7 that $\mathrm{P}[\boldsymbol{T}_n \in \mathfrak{E}_n] = G_m(v) + O(n^{-m/(m+1)})$, assuming only finite fourth moments for general $\boldsymbol{T}$. Here G_m represents the χ^2 cumulative distribution function with m degrees of freedom.

Sharper results are based on careful expansions for the distribution function of $\boldsymbol{T}$. Rao [15, 16] develops an analogue to the Esseen's series for the cumulative distribution function of a lattice distribution in the multivariate case. Error bounds here contain factors of $\log(n)$ later proved unnecessary by Bhattacharya and Rao [2]. Evaluation of this series is, however, difficult for non-rectangular and non-elliptical sets. Kolassa [9] shows that this Rao series, when evaluated at midpoints of lattice cubes, is equivalent to the Edgeworth series at the same points, with cumulants adjusted by Sheppard's corrections, to the same order of error. Yarnold [17] addresses the problem of evaluating the Rao series for convex sets, and in particular for standardized ellipses. The Yarnold approximation is the χ^2 approximation plus the difference between the actual number of points in the ellipse and the volume of the ellipse divided by the volume of a unit cube of the lattice, times the normal approximation to the density at each point on the ellipse boundary. Specifically, suppose that $\boldsymbol{X}_i$ are independent and identically distributed vectors satisfying (9) and (1), and that $\boldsymbol{T}_n$ arises as in (3). Then

$$\mathrm{P}\left[V(\boldsymbol{T}) \leq v\right] = G_m(v) + \left(N(nv) - \frac{(\pi n v)^{m/2} \det \boldsymbol{\Sigma}^{1/2}}{\Gamma(m/2+1)}\right) \times \tag{11}$$
$$\frac{\exp(-v/2)}{(2\pi n)^{m/2} \det \boldsymbol{\Sigma}^{1/2}} + O(n^{-1}),$$

where $N(nv)$ is the number of vectors of integers $\boldsymbol{m}$ such that $\sqrt{n}\boldsymbol{m} + n\boldsymbol{y} \in \mathfrak{E}_n$.

In order to gain a heuristic understanding of (11), consider the univariate case. Feller [5] notes as part of Theorem XVI.4.2 that the $F(t) = E_3(t; \boldsymbol{\kappa}^n) + o(1/\sqrt{n})$, as long as t lies midway between support points for T; furthermore, it is not difficult to show that $o(1/\sqrt{n})$ may be replaced by $O(1/n)$ when a fourth cumulant for T exists. Hence $\mathrm{P}\left[|T - \kappa_1^n| \leq w\right] = E_3(\kappa_1^n + w; \boldsymbol{\kappa}^n) - E_3(\kappa_1^n + w; \boldsymbol{\kappa}^n) + o(1/n)$, as long as $\kappa_1^n + w$ and $\kappa_1^n - w$ are both half way between support points. Furthermore, because of cancellation, $E_3(\kappa_1^n + w; \boldsymbol{\kappa}^n) - E_3(\kappa_1^n - w; \boldsymbol{\kappa}^n) = E_2(\kappa_1^n + w; \boldsymbol{\kappa}^n) - E_2(\kappa_1^n - w; \boldsymbol{\kappa}^n)$, the Gaussian approximation to the probability of the interval. Hence the χ^2 approximation to the distribution of $(T - \kappa_1^n)^2$ holds with error $o(1/n)$, and no correction term is needed, as long as it is evaluated at the squares of lattice midpoints. Also in this case, the length of the interval $(\kappa_1^n - w, \kappa_1^n + w)$ is exactly the same as the number of points in the interval, and the correction term in (11) is zero.

When either $\kappa_1^n + w$ or $\kappa_1^n - w$ or both are not half way between support points, the points at which the tail probabilities are evaluated ought to be moved. The combined distance of this shift is the difference between the number of points in the original interval and the number of points in the interval, and impact of the shift on the probability approximation is approximately the length of the shift times the derivative of the cumulative distribution function approximation at the original interval end point, exactly giving the second term in (11).

The quantity $N(nv)$ is calculated recursively. For each $j \leq m$ let

$$\mathcal{S}(nv, j) = \{(t_j, \ldots, t_m) | (t_1, \ldots, t_j, t_{j+1}, \ldots, t_m) \in \mathfrak{E} \text{ for some } t_{j+1}, \ldots, t_m\}.$$

Then $\mathcal{S}(nv, j)$ might be determined recursively, by noting that for each entry in $\mathcal{S}(nv, j)$, the corresponding minimum and maximal points in $\mathcal{S}(nv, j-1)$ may be calculated by solving a quadratic equation, and $N(nv)$ is the number of elements of $\mathcal{S}(nv, 1)$.

Kolassa [11] performs similar calculations. Let

$$\mathcal{S}^*(j) = \{(t_j, \ldots, t_m) | (t_1, \ldots, t_j, t_{j+1}, \ldots, t_m) \in \mathfrak{T} \text{ for some } t_{j+1}, \ldots, t_m\}.$$

Then one might recursively generate $\mathcal{S}^*(j)$ using linear programming, and as an alternative to using (11), summing the saddlepoint approximations to the probability atoms for points in $\mathcal{S}(nv, 1)$. These approximate probabilities may be normalized by dividing by the sum of all the probabilities associated with the points in $\mathcal{S}^*(j)$. This procedure, while potentially more accurate, is more computationally intensive, in that a nonlinear saddlepoint equation must be solved for all of the points in $\mathcal{S}^*(1)$. Alternatively, one might sum Edgeworth approximations to these probability atoms, possibly re-normalizing by summation over $\mathcal{S}^*(1)$. Kolassa [10] argues that the resulting approximation is also accurate to $O(1/n)$. Since the Edgeworth series only requires the evaluation of elementary functions, the associated calculations are only slightly more intensive than those described above.

Alternatively, one might calculate the exact probabilities associated with elements of $\mathcal{S}(nv, 1)$. Naive calculation of these probabilities is often infeasible, since while in interesting examples $\mathcal{S}(nv, 1)$ might contain manageable numbers of entries, the number of entries in the associated set of $\boldsymbol{Y}$ grows exponentially. Sophisticated algorithms for specific cases of logistic and Poisson regression exist [14], but even these algorithms fail for even moderately sized examples.

Methods proposed in this paper are intended to work for conditional inference, but the error term in (11) was derived under the assumption that the distribution considered was a marginal distribution. Barndorff–Nielsen and Cox [1] provide regularity conditions that guarantee that derivatives of (6) approximate conditional probabilities associated with $\boldsymbol{T}$ to the same order as in (7). Hence the same arguments used to justify (11) hold for this conditional distribution.

4. Generalized Linear Models. Suppose that the independent responses $Y_1, \ldots, Y_m$ have a density or mass function of the form

$$f_{Y_j}(y_j; \eta_j) = \exp\left(\eta_j y_j - \mathcal{H}_{Y_j}(\eta_j) - \mathcal{G}_{Y_j}(y_j)\right) \text{ for } y_j \in \mathfrak{Y}_j \text{ and } \eta_j \in \mathfrak{R}. \tag{12}$$

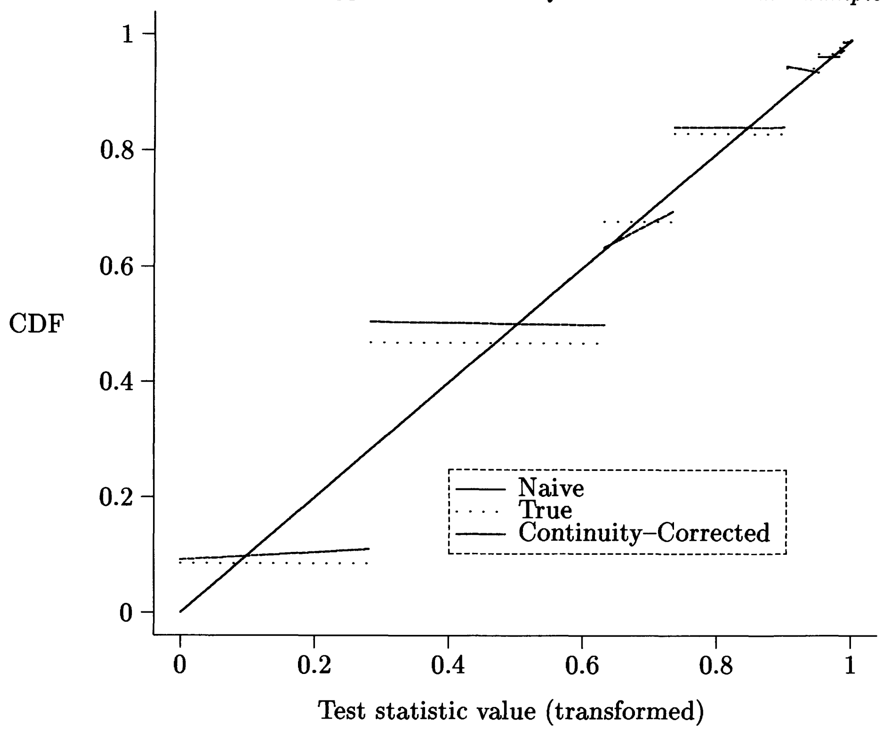
The distribution of

$$\boldsymbol{T} = \boldsymbol{Z}^\top \boldsymbol{Y} \tag{13}$$

has the form (12) with $\mathcal{H}_Y$ replaced by $\mathcal{H}_{\boldsymbol{T}}(\boldsymbol{\theta}) = \sum_{i=1}^m \mathcal{H}_Y \left(\sum_{j=1}^m z_j^i \theta^j\right)$.

Often tests and confidence regions are desired for some but not all components of $\boldsymbol{\theta}$. Suppose $\boldsymbol{\theta}$ may be partitioned as $(\boldsymbol{\theta}^*, \boldsymbol{\theta}^\dagger)$, with $\boldsymbol{\theta}^* \in \mathfrak{R}^{\mathrm{m}}$ and $\boldsymbol{\theta}^\dagger \in \mathfrak{R}^{m-\mathrm{m}}$ and we desire inference on $\boldsymbol{\theta}^*$. In this case the test statistic $V(\boldsymbol{t}, \boldsymbol{\theta})$ is constructed specifically to test the components of $\boldsymbol{\theta}$ of interest, and to depend on $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}^*$. Hence $\mathcal{J}(\boldsymbol{t}, \boldsymbol{\theta}) = \{\boldsymbol{s} \in \mathfrak{T} | V(\boldsymbol{s}, \boldsymbol{\theta}) \geq V(\boldsymbol{t}, \boldsymbol{\theta})\}$ depends only on $\boldsymbol{t}$ and $\boldsymbol{\theta}^*$. If $\boldsymbol{T} = (\boldsymbol{U}, \boldsymbol{V})$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are canonical sufficient statistics for $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^\dagger$ respectively, then define the conditional p-value

Figure 1: True and approximate CDFs for the Multinomial Example



to be $p(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\theta}^*) = \sum_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}^*)} \mathrm{P}_{\boldsymbol{\theta}^*}[\boldsymbol{U}=\boldsymbol{u}|\boldsymbol{V}=\boldsymbol{v}]$ for $\mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}^*) = \{\boldsymbol{s} \in \mathfrak{T}(\boldsymbol{v})|V(\boldsymbol{s},\boldsymbol{\theta}^*) \geq V(\boldsymbol{t},\boldsymbol{\theta}^*)\}$, and $\mathfrak{T}(\boldsymbol{v})$ be the sample space of $\boldsymbol{T} = (\boldsymbol{U},\boldsymbol{V})$ conditional on $\boldsymbol{V}$. Then $\mathcal{T}(\boldsymbol{t}) = \{\boldsymbol{\theta}^*|p(\boldsymbol{u}|\boldsymbol{v},\boldsymbol{\theta}^*) > \alpha\}$ forms a $1-\alpha$ confidence region. In these circumstances one often uses the conditional scores test defined by letting $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{U}|\boldsymbol{V}]$ and $\boldsymbol{\Sigma} = \mathrm{Var}[\boldsymbol{U}|\boldsymbol{V}]$, and setting $V(\boldsymbol{t},\boldsymbol{\theta}) = (\boldsymbol{u}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}-\boldsymbol{\mu})$. Waterman and Lindsay [18] provide an asymptotic approximation to the conditional score function as their equation (3); this approximation is of the form $\boldsymbol{Z}^\top(\boldsymbol{Y} - \boldsymbol{\omega}(\boldsymbol{\theta}^*))$. Then $\boldsymbol{Z}^\top\boldsymbol{\omega}(\boldsymbol{\theta}^*)$ approximates $\mathbf{o}$, and this quantity may be numerically differentiated to approximate $\boldsymbol{\Sigma}$.

5. An Artificial Example. Consider the multinomial distribution arising from randomly assigning 9 objects to one of three equally likely bins; this example also arises by conditioning the sufficient statistics from a Poisson regression model on the sufficient statistic associated with the constant term. One might test the null hypothesis that all three bins are equally likely vs. the alternative that at least one of the bins has a probability not equal to one third. If T^1 and T^2 are the number of observations in bins 1 and 2 respectively, minus 3, then the null expectation of

Table 1: Sarcoma Data

LI	SEX	AOP	Group Size	Number of Successes
0	0	0	3	3
0	0	1	2	2
0	1	0	4	4
0	1	1	1	1
1	0	0	5	5
1	0	1	5	3
1	1	0	9	5
1	1	1	17	6

$\boldsymbol{T} = (T^1, T^2)$ is $(0,0)$, and the null variance matrix is

$$\Sigma = \begin{pmatrix} 9 \times \frac{1}{3} \times \frac{2}{3} & -9 \times \frac{1}{3} \times \frac{2}{3}/2 \\ -9 \times \frac{1}{3} \times \frac{2}{3}/2 & 9 \times \frac{1}{3} \times \frac{2}{3} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Hence $\Sigma^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$. Hence the score statistic is $V = \frac{2}{3}[(T^1)^2 + T^1T^2 + (T^2)^2]$. There are $3^9 = 19683$ configurations, if the objects are identifiable, yielding 55 different configurations of $\boldsymbol{T}$, and only 10 different possible values for V. Thus, even though the underlying distribution is close to continuous, the distribution of the test statistic is highly discrete. Figure 1 exhibits the cumulative distribution function of $G_2^{-1}(V)$.

6. Sarcoma Example. Goorin, *et al.* [6] present results of a study of 46 patients treated for nonmetastatic osteogenic sarcoma. The treatment was considered successful if the patient was disease-free for at least 3 years. The investigators fit a logistic model with a constant term, and indicators for gender (SEX) taking the values 1 for men and 0 for women, presence of lymphocytic infiltration (LI), and of any osteoid pathology (AOP). Variables LI and AOP take the values 1 if the pathology is present and 0 if not.

The indicators for SEX, LI, and AOP split the sample into eight groups (Table 1). The four parameters are θ_1 corresponding to the constant term, θ_2 corresponding to LI, θ_3 corresponding to SEX, and θ_4 corresponding to AOP. Let $\boldsymbol{T}$ be the corresponding vector of canonical sufficient statistics. We compare this method to asymptotic calculations adjusted to account for the number of points in the acceptance region. This cuts the error in naive chi–square approximation by half (Table 2). Intermediate results for these calculations are given in Table 3.

Exact results like those in Table 2 are often criticized as being too conservative, in that the observed significance level contains the entire probability of the observed sufficient statistic vector, and the same criticism might be applied to the approximate values as well. A common alternative summary of information against the null hypothesis is the mid-p value, found by averaging the observed significance level and the next next smaller attainable significance level. With some additional

Table 2: Results of Various Tests involving the Sarcoma Example

SEX and AOP		LI and SEX	
Exact	4.20	Exact	0.75
Uncorrected	4.41	Uncorrected	1.24
Corrected	4.10	Corrected	1.02

p-values are represented in *per cent.*

Table 3: Intermediate Results for Tests of Various Hypotheses for Sarcoma Example

	SEX and AOP	LI and SEX
$\boldsymbol{\Sigma}^{-1}$	$\begin{pmatrix} 0.724 & -0.122 \\ -0.122 & 0.494 \end{pmatrix}$	$\begin{pmatrix} 0.551 & -0.072 \\ -0.072 & 0.465 \end{pmatrix}$
Statistic value	24.796	8.772
Volume of Ellipse	39.098	47.082
Number of points in ellipse	40	49

effort, one could determine the point in the sample space whose statistic value is the next larger value than that observed, repeat the use of (11), and average the two results to obtain an approximation to the mid-p value. Figure 2 exhibits the resulting cumulative distribution function for the transformed statistic $G_2^{-1}(V)$.

7. Conclusion. An easily–applied approximation due to Yarnold [17] applies a continuity correction to tail probabilities of multivariate score test statistics V for canonical exponential families, when sufficient statistics $\boldsymbol{T}$ are distributed on a multivariate lattice. These families include those specified by many popular generalized linear models. This correction may also be applied to conditional scores tests. The correction to the approximation of a probability of the form $\mathrm{P}[V \leq v]$ is calculated using the standard chi–square approximation to tail probabilities, plus a term calculated from the difference between the number of lattice points $\boldsymbol{t}$ satisfying $\{V(\boldsymbol{t}) \leq v\}$ and the volume of this ellipse. Application of this correction frequently reduces the error of approximation by half or more.

Acknowledgements The author is grateful to two anonymous referees for helpful comments. I wish to express my appreciation to Jack Hall and the organizers of this conference for the opportunity to take part in this celebration of Jack's work.

REFERENCES

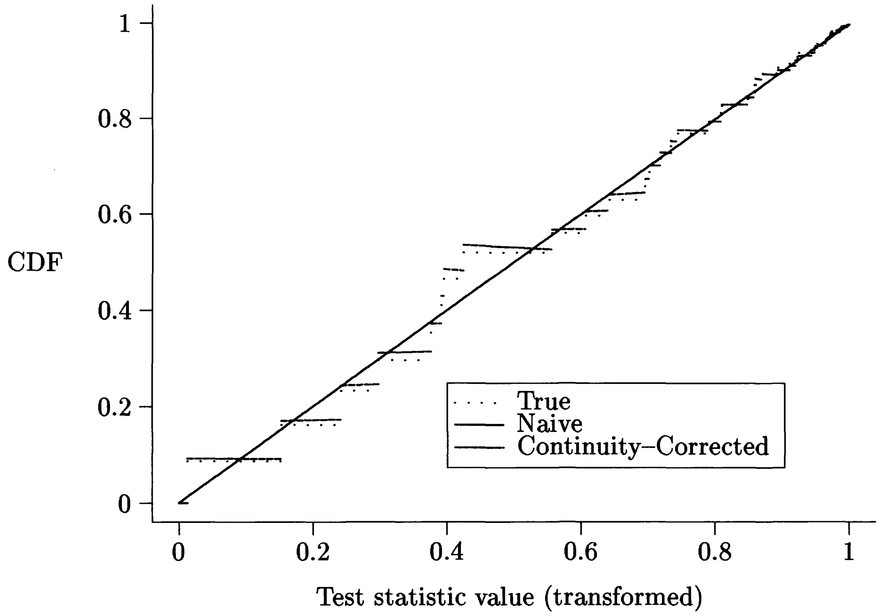
[1] O.E. Barndorff-Nielsen and D.R. Cox. Edgeworth and saddlepoint approximations with statistical applications. *Journal of the Royal Statistical Society Series B*, 41:279–312, 1979.

[2] R.N. Bhattacharya and R.R. Rao. *Normal Approximation and Asymptotic Expansions.* Wiley, 1976.

[3] J.M. Chambers. On methods of asymptotic approximation for multivariate distributions. *Biometrika*, 54:367–383, 1967.

[4] C.G. Esseen. Fourier analysis of distribution functions. *Acta Mathematica*, 77:1–125, 1945.

Figure 2: True and approximate CDFs for the Score Statistic for testing SEX and LI



[5] W. Feller. *An Introduction to Probability Theory and its Applications.* Wiley, 1971.

[6] A.M. Goorin, A. Perez-Atayde, M. Gebhardt, and J. Andersen. Weekly high–dose methotrexate and doxorubicin for osteosarcoma: The Dana–Farber Cancer Institute/The Children's Hospital - study III. *Journal of Clinical Oncology,* 5:1178–1184, 1987.

[7] W.J. Hall. Some sequential tests for matched pairs: A sequential permutation test. In P.K. Sen, editor, *Contributions to Statistics: Essays in Honour of Norman L. Johnson,* pages 211–228. North–Holland, 1983.

[8] W.J. Hall and D.J. Mathiason. On large–sample estimation and testing in parametric models. *Review of the International Statistical Institute,* 58:77–97, 1990.

[9] J.E. Kolassa. *Topics in Series Approximations to Distribution Functions.* Unpublished doctoral dissertation, 1989.

[10] J.E. Kolassa. *Series Approximation Methods in Statistics, 2nd Edn.* Springer - Verlag, 1997.

[11] J.E. Kolassa. Algorithms for approximate conditional inference. *Statistics and Computing,* 2002.

[12] J.E. Kolassa and P. McCullagh. Edgeworth series for lattice distributions. *Annals of Statistics,* 18:981–985, 1990.

[13] P. McCullagh. *Tensor Methods in Statistics.* Chapman and Hall, 1987.

[14] C. Mehta, N. Patel, and P. Senchaudhuri. Efficient monte carlo methods for conditional logistic regression. *Journal of the American Statistical Association,* 95:99–108, 2000.

[15] R.R. Rao. *Some Problems in Probability Theory.* Unpublished doctoral thesis, 1960.

[16] R.R. Rao. On the central limit theorem in R_k. *Bulletin of the American Mathematical Society,* 67:359–361, 1961.

[17] R.P. Waterman and B.G. Lindsay. A simple and accurate method for approximate conditional inference in linear exponential family models. *Journal of the Royal Statistical Society Series B,* 58:177–188, 1996.

[18] J. Yarnold. Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Annals of Mathematical Statistics*, 43:1566–1580, 1972.

Department of Statistics
110 Frelinghuysen Drive
Piscataway, NJ 08854 USA
kolassa@stat.rutgers.edu