HIGHER ORDER APPROXIMATIONS WITH GENERALIZED LINEAR MODELS

D.A.S. FRASER, G. MONETTE, K.W. NG, A. WONG

*York University, University of Hong Kong,
University of Waterloo*

A real canonical parameter of a generalized linear model can have third order tail probabilities or significance functions by the saddlepoint analysis of Davison (1988). Recent methods using asymptotically modulated densities can produce second or third order tail probabilities for real or vector parameters in the presence of nuisance parameters; the parameters need not be canonical and thus may be based on a noncanonical link function. Examples are given for generalized linear models.

1. Introduction. The saddlepoint methods of Daniels (1954) and Lugannani and Rice (1980) have been applied (Davison, 1988) to obtain tail probabilities for a real canonical parameter of a generalized linear model. The computational aspects have been simplified (Fraser, Reid & Wong, 1991) by working directly with a conditional likelihood (Cox & Reid, 1987); related theory (Fraser & Reid, 1989, 1992a,1992b) shows that the third order asymptotic properties are preserved, and implementations indicate that accuracy is improved.

For a general model consider n independent variables, where a component y_i has the canonical exponential model with density

$$\exp\{y_i\theta_i - c(\theta_i)\}f(y_i) \tag{1}$$

but with canonical parameter $\theta_i = g(X_i\boldsymbol{\beta})$ related by a link function $g(\cdot)$ to a vector $X_i = (x_{i1}, \ldots, x_{ip})$ of concomitant variables having a regression-type parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. The link function describes how the composite effect $X_i\boldsymbol{\beta}$ of the concomitant variables affects the canonical parameter θ_i of the exponential model; note that the definition of the link function, for convenience here, differs slightly from that in McCullagh and Nelder (1989).

For the special case of a canonical link with $\theta_i = X_i\boldsymbol{\beta}$, the likelihood simplifies to

$$\ell(\boldsymbol{\beta}; \mathbf{y}) = a + y'X\boldsymbol{\beta} - \sum_1^n c(X_i\boldsymbol{\beta}) , \tag{2}$$

and has sufficient statistic $\mathbf{y}'X$. For inference concerning β_p with nuisance $\boldsymbol{\beta}_{(1)} = (\beta_1, \ldots, \beta_{p-1})'$, the Cox and Reid (1987) adjusted profile

$$\ell_c(\beta_p) = \ell\big(\widehat{\boldsymbol{\beta}}_{(1)}(\beta_p), \beta_p\big) + \frac{1}{2}\log|j_{\beta_{(1)}\beta_{(1)}}\big(\widehat{\boldsymbol{\beta}}_{(1)}(\beta_p), \beta_p\big)| \tag{3}$$

gives a likelihood function appropriate to the interest parameter β_p in the context of no information concerning the nuisance parameter $\boldsymbol{\beta}_{(1)} = (\beta_1, \ldots, \beta_{p-1})'$. For notation, in the preceding expression, $\widehat{\boldsymbol{\beta}}_{(1)}(\beta_p)$ is the maximum likelihood estimate of $\boldsymbol{\beta}_{(1)}$ for fixed β_p; the first term is thus the profile likelihood; the second term contains the observed information matrix j for the nuisance parameter $\boldsymbol{\beta}_{(1)}$ again with fixed β_p,

$$j_{\beta_{(1)}\beta_{(1)}}\big(\widehat{\boldsymbol{\beta}}_1(\beta_p), \beta_p\big) = -\frac{\partial}{\partial\boldsymbol{\beta}_{(1)}}\frac{\partial}{\partial\boldsymbol{\beta}'_{(2)}}\ell(\boldsymbol{\beta}; y)\Big|_{\boldsymbol{\beta}_{(1)}=\widehat{\boldsymbol{\beta}}_{(1)}(\beta_p)}$$

This adjusted parameter can then be inverted by a procedure (Fraser, Reid & Wong, 1991, Fraser 1991) to give the tail probability

$$p(\beta_p) \; = \; P(\widehat{\beta}_p \leq \widehat{\beta}_p^0; \beta_p) \qquad (4)$$

to third order accuracy (Fraser & Reid, 1989, 1991); this is an extension of the Lugannani & Rice (1980) formula to the more general context with nuisance parameters. The computation by this route is somewhat more direct than in Davison (1988).

For independent observations $y_1, \ldots, y_n$ with a non-canonical link $\theta_i = g(X_i \boldsymbol{\beta})$, there is no longer a sufficient statistic reduction to dimension p; in fact in general there is no reduction from the sample dimension n. Barndorff-Nielsen's (1983) formula for the distribution of the maximum likelihood estimate

$$c(2\pi)^{-p/2} \exp\{\ell(\boldsymbol{\beta}; \mathbf{y}) - \ell(\widehat{\boldsymbol{\beta}}; \mathbf{y})\} |\widehat{j}|^{1/2} d\widehat{\boldsymbol{\beta}} \qquad (5)$$

when renormalized provides $O(n^{-3/2})$ accuracy, but it is in fact a conditional distribution and needs to be calculated conditionally given an ancillary $a(\mathbf{y})$; no general method of determining the appropriate ancillary is available in the literature.

For the present generalized linear models, the method in Fraser (1964) can be adapted to produce an approximate ancillary of first or second order. In the resulting conditional model, the parameters enter nonlinearly in general. The use then of a tangent exponential model (Fraser, 1964, 1990) or the r^* formula of Barndorff-Nielsen (1991) lead to significance with second or third order accuracy for real component parameters. An alternative route (Fraser & Reid, 1992b) based on Fraser (1990) and Cheah, Fraser & Reid (1991) leads to corresponding significance for real or vector parameters.

For the generalized linear models with noncanonical link we develop, as just discussed, the approximate ancillary and obtain significance by the alternate route. A simple example is used for which exact probabilities are available, thus allowing appropriate comparisons.

An example concerning lifetime of leukemia patients is discussed in Section 2; the suggested generalized linear model has two parameters. In Sections 3 and 4 one of the parameters is assumed known a priori and two different types of tangent exponential models are developed and analyzed in the two sections. The general case with nuisance parameter is discussed in Section 5.

2. An example. As illustration we choose Example U from Cox and Snell (1981). The response y is lifetime in weeks for leukemia patients and the concomitant variable x is the logarithm of the initial white blood cell count (Feigl & Zelen, 1965):

x	y	x	y	x	y
3.36	65	4.00	121	4.54	22
2.88	156	4.23	4	5.00	1
3.63	100	3.73	39	5.00	1
3.41	134	3.85	143	4.72	5
3.78	16	3.97	56	5.00	65
4.02	108	4.51	26		

The suggested model is

$$f_i(y_i) \;=\; \exp\{-y_i\theta_i + \ln(\theta_i)\} \tag{6}$$

on $y_i > 0$ where

$$E(y_i) \;=\; \theta_i^{-1} \;=\; \exp\{\alpha + \beta(x_i - \overline{x})\} \; . \tag{7}$$

This is a simple exponential model with link

$$\begin{aligned}
\theta_i \;&=\; g(X_i\boldsymbol{\beta}) \;=\; \exp\{-\alpha - \beta(x_i - \overline{x})\} \\
&=\; \exp\{-(1, x_i - \overline{x})(\alpha, \beta)'\}
\end{aligned} \tag{8}$$

which is nonlinear in the canonical parameters. In Sections 3 and 4 we examine the regression parameter β with an assumed value for α. Then in Section 5 we again examine β but with α treated as a nuisance parameter.

The example has special features that allow an exact analysis for both α known and α as nuisance. For this let $w_i = \log y_i$; then $w_i - \alpha - \beta(x_i - \overline{x}) = t_i$ has the extreme value distribution

$$\exp\{t - e^t\}dt \tag{9}$$

and the model is seen to be a location model. For some background and general discussion, see Lawless (1982).

A natural and fundamental analysis for location and transformation models was initiated by Fisher (for example, 1956); some developments of this with reference to error variable models may be found in Fraser (1968, 1979). For a regression model, the analysis

4

uses the conditional distribution given the error residuals and within the conditional model marginalizes to a pivotal variable for the interest parameter; the computations can be presented succinctly in terms of likelihood.

For the present model in the form preceding (9) we obtain the conditional density for $\widehat{\alpha}$, $\widehat{\beta}$ in terms of likelihood,

$$f(\widehat{\alpha}, \widehat{\beta}; \alpha, \beta) \; = \; cL^0(\alpha - \widehat{\alpha} + \widehat{\alpha}^0, \beta - \widehat{\beta} + \widehat{\beta}^0)$$

where $L^0(\alpha, \beta) = L(\alpha, \beta; \mathbf{y}^0)$ is the observed (non-log) likelihood from the data vector $\mathbf{y}^0$. The significance for β in the case that $\alpha = \alpha_0$ is known is based on a further conditioning which here in fact corresponds to the simple substitution $\alpha = \alpha_0$:

$$
\begin{aligned}
P(\widehat{\beta} \leq \widehat{\beta}^0; \alpha_0, \beta) \; &= \; \int_{-\infty}^{\widehat{\beta}^0} cL(\alpha_0, \beta - \widehat{\beta} + \widehat{\beta}^0) d\widehat{\beta} \\
&= \; \int_{\beta}^{\infty} cL(\alpha_0, \gamma) d\gamma
\end{aligned}
\tag{10}
$$

The significance for β with α unknown is obtained by marginalizing over $\widehat{\alpha}$:

$$P(\widehat{\beta} \leq \widehat{\beta}^0; \beta) \; = \; \int_{\beta}^{\infty} \int_{-\infty}^{\infty} cL(\alpha, \gamma) d\alpha d\gamma \tag{11}$$

where the norming constant and then probabilities can be obtained in general by numerical integration.

3. Tangent model (at parameter value); real parameter. Consider variables $y_1, \ldots, y_n$ with the generalized linear model

$$\exp[y_i g(X_i \beta) - c\{g(X_i \beta)\}] h(y_i) \tag{12}$$

where X_i and β are real. The model in general is a curved $(n, 1)$ exponential model.

For the hypothesis $\beta = \beta_0$ we need the null density for $\beta = \beta_0$ plus some model structure to obtain a test statistic. Standard first order asymptotics would suggest the quantities

$$
\begin{aligned}
q_1(\beta) \; &= \; (\widehat{\beta} - \beta)\widehat{\jmath}^{1/2} \\
q_2(\beta) \; &= \; S(\beta)\widehat{\jmath}^{-1/2} \\
r(\beta) \; &= \; \mathrm{sgn}(\widehat{\beta} - \beta)[2\{\ell(\widehat{\beta}) - \ell(\beta)\}]^{1/2}
\end{aligned}
\tag{13}
$$

based on the maximum likelihood estimate, the score

$$S(\beta) \;=\; (\partial/\partial\beta)\ell(\beta) \; ,$$

and the signed likelihood ratio r. Standard generalized linear model analysis (McCullagh & Nelder, 1989) uses deviance which is equivalent to the use of the likelihood ratio r.

An alternative procedure can be based on approximating the given model by its tangent at β_0, which has density

$$\exp[y_i X_i g_i'(\beta - \beta_0) - c\{g_i + X_i g_i'(\beta - \beta_0)\}]h(y_i) \qquad (14)$$

where $g_i = g(X_i \beta_0)$, $g_i' = g'(X_i \beta_0)$. This is an exponential model with sufficient statistic $\sum y_i X_i g_i'$ and coincides to first derivative with the given model at $\beta = \beta_0$. The likelihood inversion procedure (Fraser, Reid & Wong, 1991, Fraser, 1991) can then be applied to the likelihood from the tangent model to give the significance function

$$p_a(\beta) \;=\; P(\widehat{\beta} \le \widehat{\beta}^0; \beta) \qquad (15)$$

to saddlepoint accuracy. It should be noted that the procedure needs to be repeated for each value of β_0.

For the example in Section 2 with α taken pragmatically at its maximum likelihood value

$$e^{\widehat{\alpha}^0} \;=\; 51.109 \text{ weeks} \; ,$$

we obtain approximations to the significance function for β: the first order approximations $p_1(\beta)$, $p_2(\beta)$, $p_r(\beta)$ and the third order approximation $p_a(\beta)$ obtained from the tangent model. These are plotted in Figure 1 together with the exact $p(\beta)$ as obtained from the numerical integration described in Section 2.

The score and maximum likelihood curves are on either side of the exact, the likelihood ratio is very close to the exact. From experience with the saddlepoint procedure we can find that $p_a(\beta)$ is often accurate to 2, sometimes 3 significant figures. It differs slightly from the exact in the right tail: that it differs from the location-model exact can be attributed to its score based measure-of-departure. The corresponding score test is the locally most powerful test (a marginal test) and can be expected to differ from the location model exact test (a conditional test).

4. Tangent model (at data value); real parameter. In order to use Barndorff-Nielsen (1983) formula (5) for the maximum likelihood estimate we need to determine an approximate ancillary. Following Fraser (1964) we determine the local location relation for a typical coordinate y_i

$$dy_i \;=\; \frac{-\partial F_i(y_i, \theta)/\partial \theta}{\partial F_i(y_i; \theta)/\partial y} d\theta \;=\; d(y_i; \theta) d\theta \;. \tag{16}$$

In the particular case of the example with an exponential life distribution, we obtain

$$d(y; \theta) \;=\; \frac{-y}{\theta} \;, \tag{17}$$

which is easily seen to come from the location model structure mentioned in Section 2.

For the generalized linear model (12) we have then that a change $d\beta$ in β at $\widehat{\beta}^0$ produces the change

$$g_i' X_i d\beta$$

in θ at $\widehat{\theta}^0$ and thus the change

$$d(y_i^0, \widehat{\theta}_i^0) g_i' X_i d\beta \;=\; v_i d\beta \tag{18}$$

in the ith coordinate y_i at y_i^0. For the ancillary direction at $\mathbf{y}^0$ we therefore use $\mathbf{v} = (v_1, \ldots, v_n)'$. In the particular case of the exponential life example we have

$$v_i \;=\; -\frac{y_i^0}{\widehat{\theta}_i^0} \cdot \widehat{\theta}_i^0 (-1) \cdot 1 \;=\; y_i^0 \;. \tag{19}$$

The tangent exponential model (Fraser, 1988, 1990) that coincides with the given conditional model determined at $\mathbf{y}^0$ is

$$\exp\{z\varphi + \ell^0(\beta)\} h(z) \tag{20}$$

where $\ell^0(\beta)$ is the observed likelihood,

$$\varphi \;=\; \frac{d}{d\mathbf{v}} \ell(\beta; \mathbf{y}) \Big|_{\mathbf{y}^0} \tag{21}$$

7

is its canonical parameter, and z is the canonical variable such that $z\mathbf{v}$ coincides with $d\mathbf{y}$ in the direction $\mathbf{v}$ at $\mathbf{y}^0$.

For the exponential life example in Section 2 we have

$$
\begin{aligned}
\ell^0(\beta) &= -\sum_1^n \exp\{-\alpha_0 - \beta(x_i - \overline{x})\}y_i^0 - n\alpha_0 \\
\varphi &= \sum_1^n \exp\{-\alpha_0 - \beta(x_i - \overline{x})\}y_i^0(x_i - \overline{x}) \ .
\end{aligned}
\tag{22}
$$

The left tail probability

$$
p_b(\beta) \;=\; p(\widehat{\beta} \le \widehat{\beta}^0; \beta) \;=\; \Phi(r) + \varphi(r)\left\{\frac{1}{r} - \frac{1}{q}\right\}
\tag{23}
$$

obtained by applying the Lugannani & Rice (1980) formula to the exponential model (20) at the data point $z = 0$ corresponding to $\mathbf{y}^0$. This is asymptotically equivalent to the use of r^* in Barndorff-Nielsen (1991), given the accuracy level of the ancillary.

The approximation $p_b(\theta)$ for the example is also plotted in Figure 1. Its closeness to the exact (essentially superimposed) corresponds to the effectiveness of the generalized Lugannani and Rice approximation for location models (Fraser, 1990; DiCiccio, Field,& Fraser, 1990).

5. Tangent model (at data point); interest parameter with nuisance parameter. Consider the generalized linear model (12) with interest parameter ψ and nuisance parameter λ; for example we could have

$$
\psi \;=\; (\beta_{r+1}, \ldots, \beta_p)' \;=\; \boldsymbol{\beta}_{(2)} \ , \qquad \lambda \;=\; (\beta_1, \ldots, \beta_r)' \;=\; \boldsymbol{\beta}_{(1)} \ .
$$

For the case of a vector interest parameter ψ we can apply Skovgaard (1988) or Cheah, Fraser & Reid (1991) to an adjusted density obtained by the methods in Fraser & Reid (1992a,1992b). For present purposes we restrict our attention to a real interest parameter and real nuisance parameter: $\psi = \beta_2$, $\lambda = \beta_1$, and for convenience follow the pattern for the exponential life example in Section 2 taking $\psi = \beta$, $\lambda = \alpha$.

A change $d\alpha$ at $\widehat{\alpha}^0$ produces by the method outlined in Section 4 the change

$$
d(y_i^0, \widehat{\theta}_i^0) \cdot g_i^0 \cdot 1 \ d\alpha \;=\; y_i^0 d\alpha \;=\; v_i^1 d\alpha
$$

in y_i at $\mathbf{y}^0$; and the change $d\beta$ at $\widehat{\beta}^0$ produces the change

$$d(y_i^0, \widehat{\theta}_i^0) \cdot g_i^0 \cdot (x_i - \overline{x})d\beta \;=\; y_i^0(x_i - \overline{x})d\beta \;=\; v_i^2 d\beta$$

in y_i at $\mathbf{y}^0$. The resulting ancillary directions are $\mathbf{v}^1$, $\mathbf{v}^2$ at $\mathbf{y}^0$:

$$\mathbf{v}^1 \;=\; (v_1^1, \ldots, v_n^1)' , \qquad \mathbf{v}^2 \;=\; (v_1^2, \ldots, v_n^2)' .$$

The tangent exponential model (Fraser, 1988, 1990) that coincides with the given conditional model determined at $\mathbf{y}^0$ is

$$g(z_1, z_2) \;=\; \exp\{\varphi_1 z_1 + \varphi_2 z_2 + \ell^0(\alpha, \beta)\} h(z_1, z_2)$$

where $\ell^0(\alpha, \beta)$ is the observed likelihood

$$\varphi_1 \;=\; \frac{d}{d\mathbf{v}^1}\ell(\alpha, \beta; \mathbf{y})\big|_{\mathbf{y}^0}$$

$$\varphi_2 \;=\; \frac{d}{d\mathbf{v}^2}\ell(\alpha, \beta; \mathbf{y})\big|_{\mathbf{y}^0}$$

are the canonical parameters, and (z_1, z_2) is the canonical variable such that $z_1\mathbf{v}^1 + z_2\mathbf{v}^2$ coincides with $d\mathbf{y}$ in the directions $\mathcal{L}(\mathbf{v}^1, \mathbf{v}^2)$.

For the exponential life case we have

$$\ell^0(\alpha, \beta) \;=\; -\sum \exp\{-\alpha - \beta(x_i - \overline{x})\}y_i^0 - n\alpha ,$$

$$\varphi_1 \;=\; \sum \exp\{-\alpha - \beta(x_i - \overline{x})\}/y_i^0 ,$$

$$\varphi_2 \;=\; \sum \exp\{-\alpha - \beta(x_i - \overline{x})\}y_i^0(x_i - \overline{x}) .$$

A significance function $p_B(\beta)$ for β is obtained by averaging over the conditional distribution of $\widehat{\lambda}_\beta$. The resulting significance function is obtained from (23) with the standardized maximum likelihood quantity q replaced by the adjusted version

$$Q \;=\; \mathrm{sgn}(\widehat{\beta} - \beta) \cdot \left| \frac{(\widehat{\varphi}_1 - \widehat{\varphi}_{\beta 1})z_{\beta 1}^0 + (\widehat{\varphi}_2 - \widehat{\varphi}_{\beta 2})z_{\beta 2}^0}{\{(z_{\beta 1}^0)^2 + (z_{\beta 2}^0)^2\}^{1/2}} \right| .$$

$$\frac{|\widehat{j}|^{1/2}}{|\partial\boldsymbol{\varphi}/\partial\alpha\beta|_{(\widehat{\alpha},\widehat{\beta})}} \cdot \frac{|\partial\boldsymbol{\varphi}/\partial\alpha|_{(\widehat{\alpha}_\beta, \beta)}}{|\Sigma_1^n y_i e^{-\widehat{\alpha}_\beta - \beta(x_i - \overline{x})}|^{1/2}}$$

9

where $\boldsymbol{\varphi}_\beta$ is the pseudo parameter $\boldsymbol{\varphi}$ evaluated at the constrained maximum likelihood estimate $(\widehat{\alpha}_\beta, \beta)$, $\mathbf{z}_\beta^0$ is the corresponding canonical variable value

$$
\mathbf{z}_\beta^0 \;=\; E\left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} : \boldsymbol{\varphi} = \widehat{\boldsymbol{\varphi}}_\beta \right\} \;=\; -\left(\frac{\partial \boldsymbol{\varphi}}{\partial(\alpha,\beta)'} \right)^{-1}_{(\widehat{\alpha}_\beta, \beta)} \cdot \begin{pmatrix} \ell_\alpha^0 \\ \ell_\beta^0 \end{pmatrix}_{(\widehat{\alpha}, \beta)} \;,
$$

and $(\ell_\alpha^0, \ell_\beta^0)$ is the score vector for the observed likelihood function.

For the data in Section 2, we plot $p_B(\beta)$ which allows for the nuisance parameter and also the exact $p(\beta)$ which also allows for the nuisance parameter: they are essentially superimposed. For comparison with the results from Section 4 where the maximum likelihood value is used for α we plot $p_b(\beta)$ with $\alpha = \widehat{\alpha}^0$; this provides a tighter confidence function corresponding to treating the nuisance parameter as known.

References

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika* **70**, 343–65.

Barndorff-Nielsen, O.E. (1991). Modified signed likelihood ratio, *Biometrika* **78**, 557–63.

Cheah, P.K., Fraser, D.A.S., Reid, N. (1991). Multiparameter likelihood testing in exponential models; third order approximations from likelihood, Technical Report, Department of Statistics, University of Toronto.

Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *J.R. Statist. Soc. B* **49**, 1–39.

Cox, D.R., Snell, E.J. (1981). *Applied Statistics.* London: Chapman and Hall.

Daniels, H.E. (1954). Saddlepoint approximations in statistics, *Ann. Math. Statist.* **25**, 631–50.

Davison, A.C. (1988). Approximate conditional inference in generalized linear models, *J.R. Statist. Soc. B* **50**, 445–61.

DiCiccio, T.J., Field, C.A., Fraser, D.A.S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters, *Biometrika* **77**, 77–95.

Feigl, P., Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, *Biometrika* **21**, 826–38.

Fisher, R.A. (1956). Statistical Methods and Scientific Inference, New York: J. Wiley.

Fraser, D.A.S. (1964). Local conditional sufficiency, *J.R. Statist. Soc. B* **26**, 52–62.

Fraser, D.A.S. (1968). *The Structure of Inference.* New York: John Wiley.

Fraser, D.A.S. (1979). *Inference and Linear Models.* New York: McGraw-Hill.

Fraser, D.A.S. (1988). Normed likelihood as saddlepoint approximation, *J. Mult. Anal.* **27,** 181–93.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods, *Biometrika* **77,** 65–76.

Fraser, D.A.S. (1991). Statistical inference: from likelihood to significance, *J. Amer. Stat. Assoc.* **86**, 258–65.

Fraser, D.A.S., Reid, N. (1989). From multiparameter likelihood to tail probabilities for a scalar parameter, Technical Report, Department of Statistics, University of Toronto. Report.

Fraser,D.A.S., Reid, N. (1992a). Simple asymptotic connections between density and cumulant generating functions leading to accurate approximations for distribution functions, *Statist. Sinica,* to appear.

Fraser, D.A.S., Reid, N. (1992b). Significance for real or vector parameters based on adjusted asymptotic densities, Technical report, Department of Statistics, University of Toronto.

Fraser, D.A.S., Reid, N., Wong, A. (1991). Exponential linear models: a two pass procedure for saddlepoint approximations, *J.R. Statist. Soc. B* **52,** 483–92.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: John Wiley.

Lugannani, R., Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables, *Adv. Appl. Prob.* **12,** 475–90.

McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models.* London: Chapman and Hall.

Skovgaard, I.M. (1987). Saddlepoint expansions for directional test probabilities, *J.R. Statist. Soc. B* **50,** 269–80.