THE AMERICAN STATISTICIAN

Published by the American Statistical Association
Volume 48, Number 4, November 1994

General

271 Estimation Given Conditionals From an Exponential Family *Panagis Moschopoulos and Joan G. Staniswalis*
276 The Shape of Correlation Matrices *Peter J. Rousseeuw and Geert Molenberghs*
280 Training the Next Generation of Biostatisticians *David L. DeMets, Dan Anbar, William Fairweather, Thomas A. Louis, and Robert T. O'Neill*
285 A Core in Statistics for Engineering Students *Robert V. Hogg*
288 A Simple Justification of the Iterative Fitting Procedure for Generalized Linear Models *Stephen L. Hillis and Charles S. Davis*
290 What Some Puzzling Problems Teach About the Theory of Simulation and the Use of Resampling *Julian L. Simon*
294 A Characterization of the Poisson Distribution and the Probability of Winning a Game *Joseph B. Keller*

Teacher's Corner

299 Regression for Longitudinal Data: A Bridge From Least Squares Regression *Dorothy D. Dunlop*
303 A Suggestion for Sunflower Plots *Mark F. Schilling and Ann E. Watkins*
305 **Accent on Teaching Materials**

Statistical Practice

307 Transforming Both Sides of a Tree *A. C. Atkinson*
313 How a Court Accepted an Impossible Explanation *Joseph Gastwirth, Abba Krieger, and Paul Rosenbaum*
315 Simultaneous Confidence Intervals in Multiple Regression *Thomas P. Lane and William H. DuMouchel*

Statistical Computing

322 **Statistical Computing Software Reviews**
322 **Editor's Notes**
324 Review of Five Multilevel Analysis Programs: BMDP-5V, GENMOD, HLM, ML3, and VARCL *Ita G. G. Kreft, Jan de Leeuw, and Rien van der Leeden*
336 Econometric Time Series Analysis Software: A Review *Junghun Kim and Pravin K. Trivedi*

Commentaries

347 **Book Reviews**

Letters to the Editor

351 Letters to the Editor
354 Corrections

355 **Editorial Collaborators**
358 **Index to Volume 48**

Estimation Given Conditionals From an Exponential Family

Panagis MOSCHOPOULOS and Joan G. STANISWALIS

Suppose we are given n independent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a conditionally specified distribution with density $f(x, y)$. The problem of estimating the unknown parameters of $f(x, y)$ is complicated by the presence of an intractable normalizing constant that, in this conditional approach, is chosen so that the density integrates to 1. An approach to estimation is used here that is known to result in asymptotically efficient estimators of the unknown parameters when $f(x, y)$ is from an exponential family. It is an application of a method that has appeared in the literature and is due to J. K. Lindsey. It very conveniently avoids the dependence on the normalizing constants in the joint distribution. The usual maximum-likelihood estimates of the parameters can be obtained using software readily available for Poisson regression. The usefulness of this estimation method is illustrated for a model resulting from specifying that the conditionals are two-parameter, shape and scale, gamma. It is assumed that only the scale parameter depends on the conditioning variable. This model subsumes the BEC class and several characteristics of the model extend those of the BEC class.

KEY WORDS: Bivariate conditionals; Gamma conditionals; Log-linear models.

1. CONDITIONALLY SPECIFIED MODELS

Let (X, Y) be a two-dimensional random variable and denote the marginal and conditional densities of X and Y by $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$. If these densities exist, then

$$f_X(x) f_{Y|X}(y \mid x) = f_Y(y) f_{X|Y}(x \mid y) \tag{1}$$

and the joint density $f_{X,Y}(x, y)$ may be defined from either of the above products.

In conditional specification, we assume that only the conditional densities $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$ are specified and seek proper marginals $f_X(x)$ and $f_Y(y)$ that satisfy (1). If such marginals exist, then the joint density $f_{X,Y}(x, y)$ may be defined again from either product in (1).

Not all pairs of conditional densities lead to legitimate marginals, and therefore a characterization of the joint density. Let $N_a = \{(x, y) : f_{X|Y}(x \mid y) > 0\}$ and $N_b = \{(x, y) : f_{Y|X}(y \mid x) > 0\}$. It can be shown that necessary and sufficient conditions for the existence of $f_X(x)$ and $f_Y(y)$ are that $N_a = N_b$ and that the ratio of the conditionals be factored into the ratio of two integrable functions $u(x)$ and $v(y)$ that is,

$$f_{X|Y}(x \mid y) / f_{Y|X}(y \mid x) = u(x) / v(y). \tag{2}$$

See Arnold, Castillo, and Sarabia (1992). In such a case, the conditional densities are compatible, and the joint density $f_{X,Y}(x, y)$ can be completely characterized although it may not necessarily be uniquely determined.

The usefulness of conditional specification in constructing bivariate models stems from the fact that it is often easier to visualize conditional aspects of a problem rather than characteristics of the marginal or the joint distribution. For example, the lifetime X of a component may be exponentially distributed with a mean depending on the lifetime Y of another component of the system. Such considerations prompted considerable interest in the past few years. Castillo and Galambos (1989) considered the case of normal conditionals, and they completely characterized the resulting class of distributions. Arnold and Strauss (1987) studied the class of all distributions with Pareto conditionals. Arnold and Strauss (1988) considered a bivariate class of models with exponential conditionals (BEC), and Arnold and Strauss (1991) discussed models with conditionals specified in the exponential family. Additional models and related references can be found in Arnold, Castillo, and Sarabia (1992).

One of the main difficulties in conditional specification is that the resulting marginals $f_X(x)$ and $f_Y(y)$ and also the joint density $f_{X,Y}(x, y)$ all involve an intractable constant. This is due to the requirement that the densities should integrate to 1. The integral involved in the process must be evaluated numerically in almost all cases. For example, even in the case of exponential conditionals, the constant involves the exponential integral; see Arnold and Strauss (1988). The consequence is that it becomes difficult to obtain properties of such models; for example, moments, moment-generating functions, and maximum-likelihood estimates are not available in closed mathematical forms. In particular, an optimal method of estimation is still lacking even in the simplest cases. The problem is further complicated by the fact that, in general, the parameters of conditionally specified models are constrained, except in some special cases. Thus it seems best to consider such models on a case-by-case basis.

The estimation method of this paper is illustrated in a model resulting from specifying that the conditionals are two-parameter gamma, shape and scale, but we will assume that only the scale parameter depends on the conditioning variable. As we shall see, this results in a bivariate model that may involve up to five unrestricted parameters. Naturally this model subsumes the BEC class, and several characteristics of the model extend those of the BEC class; for example, the normalizing constant, a function of the exponential integral in the BEC class, now involves a $\Psi(a, b; z)$ hypergeometric function.

We note that the functional form of the model (excluding the normalizing constant) can be derived from regression considerations [see Mardia (1970, chap. 6)], and it is also a special case of the very general model with conditionals in exponential families [see Arnold, Castillo, and

Panagis Moschopoulos and Joan G. Staniswalis are Associate Professors, Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX 79968-0514. This work was supported in part by NIH Grant S06GM08012-22.

Sarabia (1992, chap. 4)]. We develop this model in the next section.

2. A FAMILY OF BIVARIATE DISTRIBUTIONS WITH GAMMA CONDITIONALS

Suppose that (X, Y) is a two-dimensional random vector with density $f_{X,Y}(x, y)$ with support in the positive quadrant, and suppose that $(X \mid Y = y) \sim G(r, \mu)$ and $(Y \mid X = x) \sim G(s, \nu)$ where r and s are fixed parameters, $\mu = \mu(y)$ and $\nu = \nu(x)$. Thus the conditional densities, assuming $r > 0, s > 0, \mu(y) > 0$ and $\nu(x) > 0$, are assumed to be two-parameter gamma densities for $x > 0$, and $y > 0$. Note that only the scale parameters μ and ν are allowed to depend on the conditioning variables. Thus we have

$$f_{Y|X}(y \mid x) = \frac{(\nu(x))^s}{\Gamma(s)} y^{s-1} e^{-\nu(x)\cdot y}, \qquad y > 0,\ \nu(x) > 0, x > 0 \tag{3}$$

and

$$f_{X|Y}(x \mid y) = \frac{(\mu(y))^r}{\Gamma(r)} x^{r-1} e^{-\mu(y)\cdot x}, \qquad x > 0, \mu(y) > 0, y > 0. \tag{4}$$

For the sake of simplicity and completeness but also for illustration purposes, we indicate here how one arrives at a form of the underlying joint density. On substituting (3) and (4) into (1), taking natural logarithms, and equating the second-order partial derivatives $\partial^2(\cdot)/\partial x \partial y$ of both sides of (1), we obtain the necessary conditions

$$\mu(y) = \alpha y + \beta \tag{5}$$

$$\nu(x) = \alpha x + \gamma \tag{6}$$

for some $\alpha > 0, \beta > 0, \gamma > 0$. It can be readily seen that, under (5) and (6), the two conditionals are compatible in the sense of Section 1. It should be noted that the functions $\mu(y)$ and $\nu(x)$ have the same linear form as in the case of the BEC distributions, that $r = s = 1$. Now substituting (3)–(6) in (1), we obtain

$$f_Y(y) \frac{(\alpha y + \beta)^r}{\Gamma(r)} x^{r-1} e^{-(\alpha y+\beta)x} = f_X(x) \frac{(\alpha x + \gamma)^s}{\Gamma(s)} \times y^{s-1} e^{-(\alpha x+\gamma)y}. \tag{7}$$

It follows from the preceding form that except for the normalizing constants, the densities have the forms

$$f_X(x) \propto (\alpha x + \gamma)^{-s} x^{r-1} e^{-\beta x}, \qquad x > 0 \tag{8}$$

$$f_Y(y) \propto (\alpha y + \beta)^{-r} y^{s-1} e^{-\gamma y}, \qquad y > 0 \tag{9}$$

Here $\alpha > 0, \beta > 0, \gamma > 0, r > 0$, and $s > 0$. Substituting (8) and (9) into (7), we obtain

$$f_{X,Y}(x, y) \propto x^{r-1} y^{s-1} e^{-(\beta x+\gamma y+\alpha xy)} \tag{10}$$

for $x > 0, y > 0, \alpha > 0, \beta > 0$, and $\gamma > 0$. Note that the case $r = s = 1$ leads to the BEC class (Arnold and Strauss 1988). Note also that, as in the BEC class, $\alpha = 0$ implies independence.

3. THE NORMALIZING CONSTANTS

In order to write the explicit forms for the marginal densities $f_X(x)$, $f_Y(y)$, and the joint density $f_{X,Y}(x, y)$, we need to evaluate the normalizing constant. Let us note that if $\alpha = 0$, then the marginals are also gamma and clearly independent. If $\alpha > 0$, then it is convenient to reparameterize letting

$$\alpha = \delta\beta\gamma \qquad \text{for some } \delta > 0, \tag{11}$$

in which case,

$$\begin{aligned}\int_0^\infty & (\alpha x + \gamma)^{-s} x^{r-1} e^{-\beta x} dx \\ &= \gamma^{-s} \int_0^\infty (\beta\delta x + 1)^{-s} x^{r-1} e^{-\beta x} dx \\ &= (\beta\delta)^{-r} \gamma^{-s} \int_0^\infty t^{r-1} (1 + t)^{-s} e^{-t/\delta} dt \\ &= (\beta\delta)^{-r} \gamma^{-s} \theta_{r,s}(\delta),\end{aligned} \tag{12}$$

where

$$\theta_{r,s}(\delta) = \int_0^\infty t^{r-1} (1 + t)^{-s} e^{-t/\delta} dt = \Gamma(r) \Psi(r, c; 1/\delta), \tag{13}$$

$c = r - s + 1$, and $\Psi(r, c; 1/\delta)$ is the confluent hypergeometric function (Luke 1969) known as the Ψ-function defined by

$$\Psi(r, c; 1/\delta) = \frac{1}{\Gamma(r)} \int_0^\infty t^{r-1} (1 + t)^{c-r-1} e^{-t/\delta} dt. \tag{14}$$

Thus the marginal densities are given by

$$f_X(x) = \frac{(\beta\delta)^r}{\theta_{r,s}(\delta)} (\beta\delta x + 1)^{-s} x^{r-1} e^{-\beta x}, \qquad x > 0 \tag{15}$$

$$f_Y(y) = \frac{(\gamma\delta)^s}{\theta_{s,r}(\delta)} (\gamma\delta y + 1)^{-r} y^{s-1} e^{-\gamma y}, \qquad y > 0. \tag{16}$$

The joint density is given by

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) \cdot f_{Y|X}(y \mid x) = f_Y(y) \cdot f_{X|Y}(x \mid y) \\ &= \frac{(\beta\delta)^r (\beta\delta x + 1)^{-s}}{\theta_{r,s}(\delta)} x^{r-1} e^{-(\beta x)} \\ &\quad \times \frac{(\beta\gamma\delta x + \gamma)^s}{\Gamma(s)} y^{s-1} e^{-(\beta\gamma\delta x+\gamma)y} \\ &= \frac{(\beta\delta)^r \gamma^s}{\theta_{r,s}(\delta)\Gamma(s)} x^{r-1} y^{s-1} e^{-(\beta x+\gamma y+\beta\gamma\delta xy)} \end{aligned} \tag{17}$$

for $x, y > 0$.

The family of joint densities of the form (17) will be called a bivariate family with gamma conditionals (BGC).

4. SPECIAL CASES

4.1 $\beta = \gamma = 1$

When the scale parameters are assumed equal to one,

$$f_{X,Y}(x, y) = \frac{\delta^r}{\theta_{r,s}(\delta)\Gamma(s)} x^{r-1} y^{s-1} e^{-(x+y+\delta xy)}, \qquad x > 0, y > 0 \tag{18}$$

$$f_X(x) = \frac{\delta^r}{\theta_{r,s}(\delta)} (1 + \delta x)^{-s} x^{r-1} e^{-x}, \qquad x > 0 \tag{19}$$

$$f_Y(y) = \frac{\delta^s}{\theta_{s,r}(\delta)} (1 + \delta y)^{-r} y^{s-1} e^{-y}, \qquad y > 0. \tag{20}$$

4.2 $\beta = \gamma = 1$ and $r = s = 1$

This leads to the one-parameter BEC family. In this case, the normalizing constant is

$$\theta_{1,1} = \int_0^\infty (1+t)^{-1} e^{-t/\delta} = \delta \int_0^\infty (1+\delta u)^{-1} e^{-u} du. \quad (21)$$

Thus $\theta_{1,1}(\delta)$ is expressed in terms of the exponential integral; that is,

$$\theta(\delta) = \delta/\theta_{1,1}(\delta) = \left(\int_0^\infty (1+\delta u)^{-1} e^{-u} \right)^{-1} = \delta e^{-1/\delta}/[-E_i(1/\delta)] \quad (22)$$

where

$$-E_i(u) = \int_u^\infty (e^{-w}/w)\, dw \quad (23)$$

as in Arnold and Strauss (1988).

The BEC family restated in terms of $\theta(\delta)$ as in Arnold and Strauss (1988) is

$$f_{X,Y}(x,y) = \theta(\delta) e^{-(x+y+\delta xy)}, \qquad x > 0, y > 0 \quad (24)$$

$$f_X(x) = \theta(\delta)(1+\delta x)^{-1} e^{-x}, \qquad x > 0 \quad (25)$$

$$f_Y(y) = \theta(\delta)(1+\delta y)^{-1} e^{-y}, \qquad y > 0. \quad (26)$$

5. ESTIMATION

One of the most viable approaches to estimation, especially when conditionals are in exponential families, is a particular form of pseudolikelihood involving conditional densities. In this approach the pseudolikelihood for the sample is taken to be the product of the conditional densities; that is, $\prod_{i=1}^n \{f_{Y|X}(y_i \mid x_i) f_{X|Y}(x_i \mid y_i)\}$. Note that this product does not depend on the normalizing constant. Arnold, Castillo, and Sarabia (1992) considered conditional likelihood estimation in the case of Pareto conditionals and in the BEC class. The resulting estimator is not asymptotically efficient, but it was shown to be more than 88% efficient for the one-parameter BEC distribution.

An approach to estimation is considered here that results in asymptotically efficient estimators. It is due to J. K. Lindsey (see Lindsey and Mersch, 1992 and earlier papers referenced therein). With this approach, parameters in the probability distributions from an exponential family of the form $f(x,y) = \exp[c(x,y) + \theta h(x,y) + s(\theta)]$, where $\theta = (\theta_1, \ldots, \theta_k)$ and $h : \mathbb{R}^2 \to \mathbb{R}^k$, can be estimated as the coefficient in a log-linear Poisson regression model. The explanatory variables in the log-linear Poisson regression models are given by the components of $h(x,y) = (h_1(x,y), \ldots, h_k(x,y))^t$. It very conveniently avoids the dependence on the normalizing constants, in the joint distribution. The usual maximum-likelihood estimates of the parameters can be obtained. The method is now described.

Let $\mathcal{R}(\Delta_1, \Delta_2) = \{R_{ij}\}_{i=1,\infty}^{j=1,\infty}$ denote a partition of the support of f where each $R_{ij} = \{(x,y) \text{ s.t } x \in ((i-1) * \Delta_1, i * \Delta_1] \text{ and } y \in ((j-1) * \Delta_2, j * \Delta_2]\}$. Hence each element R_{ij} in the partition $\mathcal{R}$ is a rectangle with area $\Delta_1 \Delta_2$ and is centered at the point $(g_{1i}, g_{2j}) = ((i-.5)\Delta_1, (j-.5)\Delta_2)$. Let n_{ij} denote the number of observations in the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ falling in R_{ij}.

Rather than maximizing the likelihood directly, consider the multinomial likelihood based on the observed counts n_{ij}. Clearly,

$$P((X,Y) \in R_{ij}) \doteq \Delta_1 \Delta_2 f(g_{1i}, g_{2j})$$

and hence the likelihood function based on the observed counts n_{ij} is

$$L(\theta) \doteq \prod_{i=1}^\infty \prod_{j=1}^\infty \{\Delta_1 \Delta_2 f(g_{1i}, g_{2j})\}^{n_{ij}}.$$

Since $\sum_{i=1}^\infty \sum_{j=1}^\infty \Delta_1 \Delta_2 f(g_{1i}, g_{2j}) \doteq 1$, the likelihood $L(\theta)$ is expressed as

$$L(\theta) \doteq \prod_{i=1}^\infty \prod_{j=1}^\infty \left[\frac{f(g_{1i}, g_{2j})}{\sum_{u=1}^\infty \sum_{v=1}^\infty f(g_{1u}, g_{2v})} \right]^{n_{ij}}$$

and for some large integer m

$$L(\theta) \doteq \prod_{i=1}^m \prod_{j=1}^m \left[\frac{f(g_{1i}, g_{2j})}{\sum_{u=1}^m \sum_{v=1}^m f(g_{1u}, g_{2v})} \right]^{n_{ij}} \quad (27)$$

Note that in contrast with $L(\theta) \prod_{i=1}^n f(X_i, Y_i)$, the preceding does not depend on the normalizing constant for the density.

Next, recognize that this multinomial likelihood is equivalently the likelihood arising from the joint distribution of $n_{11}, \ldots, n_{mm}$ conditional on $\sum_{j=1}^m \sum_{i=1}^m n_{ij} = n$, where n_{ij} are independent Poisson random variates with mean parameter $\mu_{ij} = f(g_{1i}, g_{2j})$ (Bishop, Feinberg, and Holland 1975). Hence the n_{ij} may be modeled as Poisson with mean μ_{ij} having the log-linear model

$$\ln \mu_{ij} = c(g_{1i}, g_{2j}) + \theta h(g_{1i}, g_{2j}) + s(\theta).$$

In this framework, the estimate of the parameter θ is the estimate of the coefficients of the explanatory variables in h, and $s(\theta)$ is the intercept. The term c is an offset, an explanatory variable with known coefficient 1. Note that the normalizing constant for f drops out of the expression for $L(\theta)$ and does not enter into the Poisson regression model. Hence, the asymptotic variance of the log-linear model estimators for the partition $\mathcal{R}(\Delta_1, \Delta_2)$ can be easily computed using the second-order partial derivatives of the logarithm of (27).

Proceeding from the multinomial likelihood in (27), the asymptotic covariance of the estimators of θ_r and θ_s derived from the log-linear model is given by $[nV_{rs}]^{-1}$, where

$$V_{rs} = \frac{\sum_{i=1}^\infty \sum_{j=1}^\infty f(g_{1i}, g_{2j}) h_r(g_{1i}, g_{2j}) h_s(g_{1i}, g_{2j})}{\sum_{i=1}^\infty \sum_{j=1}^\infty f(g_{1i}, g_{2j})} - \prod_{p=r,s} \left(\frac{\sum_{i=1}^\infty \sum_{j=1}^\infty f(g_{1i}, g_{2j}) h_p(g_{1i}, g_{2j})}{\sum_{i=1}^\infty \sum_{j=1}^\infty f(g_{1i}, g_{2j})} \right) \quad (28)$$

for $r, s = 1, \ldots, k$. Knowledge of the normalizing constant for the joint density is not needed to compute (28) since the constant cancels from the numerator and denominator of the expressions involving f.

6. SIMULATIONS FOR THE BEC

Consider estimation for the one-parameter BEC distribution, in which case $f(x, y) = \theta(\delta)\exp[-x - y - \delta xy]$ for $x, y, \delta > 0$. A small simulation study was conducted in order to compare the finite sample variances of the log-linear and conditional-likelihood estimators to the asymptotic expressions for their variance. For this example, $c(x, y) = -x - y, k = 1$, and $h(x, y) = -xy$ in the log-linear model.

Figure 1 provides a comparison between the asymptotic variance given by Equation (28) of the estimator fit using the relation between the multinomial and Poisson distributions (dashed) with $\Delta_1 = \Delta_2 = .05$ and the estimator studied by Arnold, Castillo, and Sarabia (1992, p. 121) that used conditional likelihoods (solid curve). Figure 1 also shows the asymptotic variance of the maximum-likelihood estimator (dotted curve) $[\delta^2 + 2\delta - \theta(\delta)(\theta(\delta) + \theta - 1)]/\delta^4$ (Arnold, Castillo, and Sarabia 1992, p. 119). Note that the dotted and dashed curves are indistinguishable. Figure 2 is similar to Figure 1 except that the estimator fit using the relation between the multinomial and Poisson distributions used $\Delta_1 = .25$ and $\Delta_2 = .2$. This illustrates the fact that a good choice for Δ_1 and Δ_2 depends on the value of δ. Smaller values of Δ_1 and Δ_2 are needed for larger value of δ.

Five blocks of 500 realizations of samples of size $n = 50$ with $\delta = .5$ were generated. The SAS function RANEXP was used to generate the data using the rejection method (Devroye 1986, pp. 41–42). An SAS macro using Proc NLIN (SAS Institute, Inc., 1990, vol. 2, p. 1168) was used for the Poisson regression to fit the log-linear model with $\Delta_1 = .25$ and $\Delta_2 = .2$. It is important that the partitions in $\mathcal{R}$ with zero counts be included as data when computing the coefficients for the log-linear model. For these values of $\delta, \Delta_1, \Delta_2$, the asymptotic efficiency of the estimator derived from the log-linear model is .99 versus .94 for the estimator derived using the conditional likelihood. The estimates using the conditional likelihoods were implemented in Fortran and found by a grid search as suggested by Arnold, Castillo, and Sarabia (1992, p. 120). The simulated means and variances of the two estimators are listed in the following table with their respective asymptotic variances; standard errors are given in parenthesis for the quantities computed by simulation:

	Mean	*n* × *Variance*	*n* × *Asymptotic Variance*
log-linear	.550(.004)	2.81(.17)	2.33
conditional	.561(.004)	2.96(.16)	2.46

The simulated sampling distribution for the conditional-likelihood and the log-linear estimators were very similar, namely, unimodal and skewed right. Only the estimator of δ using the log-linear model is known to be asymptotically efficient for small values of Δ_1 and Δ_2. The effect of Δ_1 and Δ_2 on the estimate of δ was investigated. The estimates of δ eventually stabilize as the grid size is made smaller.

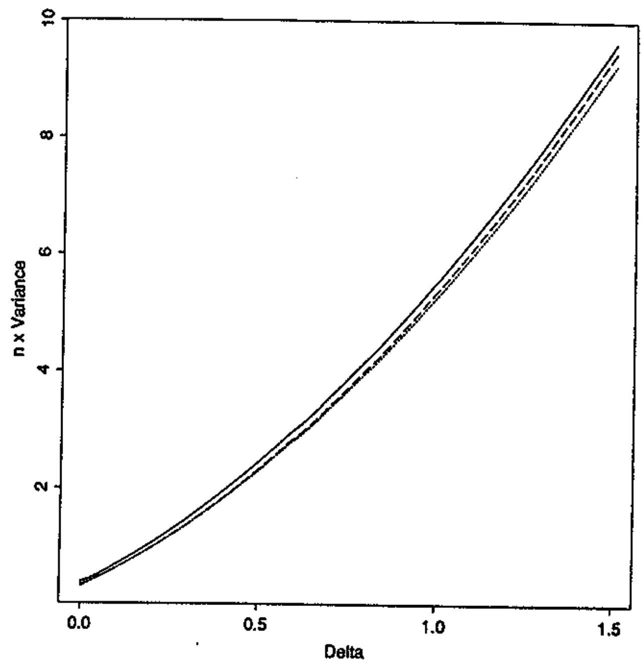

Figure 2. Asymptotic Variances of Competing Estimators. Log-linear model with $\Delta_1 = .25$, and $\Delta_2 = .2$ (dashed curve), conditional likelihood (solid curve), maximum likelihood (dotted curve).

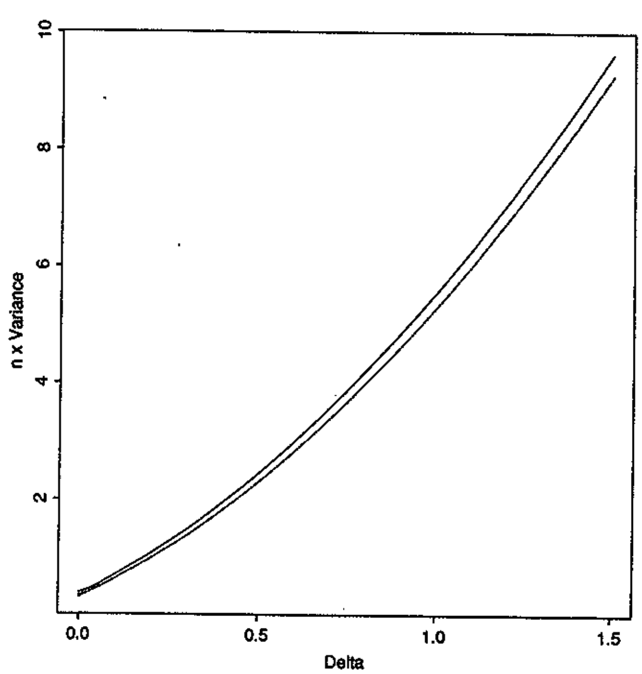

Figure 1. Asymptotic Variances of Competing Estimators. Log-linear model with $\Delta_1 = \Delta_2 = .05$ (dashed curve), conditional likelihood (solid curve), maximum likelihood (dotted curve).

7. SIMULATIONS FOR THE BGC

Consider estimation for the three-parameter BGC distribution that results from setting $\beta = 1$, and $\gamma = 1$ in (17). A small simulation study was conducted to demonstrate the feasibility of estimation via the log-linear Poisson regression model in this example with several unknown parameters, namely δ, r, and s. For this example, $c(x, y) = -x - y - \ln(x) - \ln(y), k = 3, h(x, y) = (-xy, \ln(x), \ln(y))^T$, and $\theta = (\delta, r, s)$ in the log-linear model.

Five blocks of 100 realizations of samples of size $n = 50$ with $\delta = .5$, and $r = 2$, and $s = 3$ were generated. The SAS function RANGAM was used to generate the data using the rejection method (Devroye 1986, pp. 41–42). The same SAS macro based on Proc NLIN used in the BEC simulations described in the last section was used (SAS Institute, Inc., 1990, vol. 2, p. 1168) with $\Delta_1 = .25$ and $\Delta_2 = .2$. The simulated mean and variance of the log-linear estimators of δ, r, and s are listed in the following table with the asymptotic variances computed from (28). Standard errors are given in parenthesis for the quantities computed by simulation.

Parameter	*Mean*	$n \times$ *Variance*	$n \times$ *Asymptotic Variance*
δ	.55(.01)	2.08(.17)	.44
r	2.11(.01)	6.06(.57)	1.33
s	3.08(.02)	7.44(.66)	2.19

These estimators of δ, r, and s using the log-linear model are known to be asymptotically efficient for small values of Δ_1 and Δ_2. Note the ease with which asymptotically efficient estimators for these bivariate models, with conditionals from an exponential family, can be computed using existing software. It is not known how well the conditionally-likelihood estimators perform for the BGC. The conditional-likelihood estimators are not likely to be asymptotically efficient, and computation of the estimates would require writing software specific to the example at hand.

[*Received January 1994.*]

REFERENCES

Arnold, B. C., Castillo, E., and Sarabia, J.-M. (1992), "Conditionally Specified Distributions," *Lecture Notes in Statistics* (No. 73), New York: Springer-Verlag.

Arnold, B. C. (1987), "Bivariate Distributions with Pareto Conditionals," *Statistics and Probability Letters*, 5, 263–266.

Arnold, B. C., and Strauss, D. J. (1988), "Bivariate Distributions with Exponential Conditionals," *Journal of the American Statistical Association*, 83, 522–527.

——— (1991), "Bivariate Distributions with Conditionals in Prescribed Exponential Families," *Journal of the Royal Statistical Society*, 53, 365–375.

Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA: MIT Press.

Castillo, E., and Galambos, J. (1989), "Conditional Distributions and the Bivariate Normal Distribution," *Metrika*, 36, 209–214.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, New York, Springer-Verlag.

Lindsey, J. K., and Mersch, G. (1992), "Fitting and Comparing Probability Distributions With Log-linear Models," *Computational Statistics and Data Analysis*, 13, 373–384.

Luke, Y. L. (1969), *The Special Functions and Their Approximations* (Vol. 1), New York: Academic Press.

Mardia, K. V. (1970), *Families of Bivariate Distributions*, Darien, CT: Hafner.

SAS Institute, Inc. (1990), *SAS/STAT User's Guide* (Version 6, 4th ed.), Cary, NC: Author.