Estimating Parameters for Discrete Distributions Via the Empirical Probability Generating Function

Mary M. Dowling
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
U.S.A.

Miguel Nakamura*
Centro de Investigación en Matemáticas, A.C.
Apartado Postal 402
Guanajuato, Gto. 36000
México

October 18, 1996

Abstract

We consider parameter estimation for a family of discrete distributions characterized by probability generating functions (pgf's). Kemp and Kemp (1988) suggest estimators based on the empirical probability generating function (epgf); the methods involve solving estimating equations obtained by equating functionals of the epgf and pgf on a fixed, finite set of values. We derive asymptotic theory for these estimators and consider some examples. Graphical techniques based on the theory are shown to be useful for exploratory analysis.

*Research partially supported by CONACYT grant 1858-E9219.

1 Introduction

This paper considers parameter estimation for families of discrete distributions with support on the nonnegative integers or a subset thereof. If X is a random variable possessing one of these distributions for counts, the probability generating function (pgf), $\phi(t) = E[t^X]$, is always defined for $-1 \leq t \leq 1$, and this function of t completely characterizes the distribution of X.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution F having pgf ϕ. In recent years, the empirical probability generating function (epgf), defined by

$$\phi_n(t) = \frac{1}{n} \sum_{i=1}^{n} t^{X_i} \ ,$$

has received attention for use in statistical inference. Methods are based on the fact that $\phi_n(t)$ converges in various modes to $\phi(t)$ as $n \to \infty$; see, for instance, Feuerverger and McDunnough (1984), Marques and Pérez-Abreu (1989), and Csörgó and Teugels (1990). Model specification via the pgf is often simpler than that based on the discrete density. A pgf-viewpoint can also lead to appealing graphical procedures and computational advantages. Most work has focused on goodness of fit for discrete distributions, as in Kocherlakota and Kocherlakota (1986, 1990), Rueda *et al.* (1991), Baringhaus and Henze (1992), and Nakamura and Pérez-Abreu (1993a, 1993b).

The epgf has also been proposed for parameter estimation. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^T \in \Theta$ be an m-dimensional parameter and consider the family of pgf's, $\{\phi(t, \boldsymbol{\theta})\}_{\boldsymbol{\theta} \in \Theta}$. Kemp and Kemp (1988) propose a number of methods for estimating $\boldsymbol{\theta}$. The simplest of these is obtained by fixing $\boldsymbol{t} = (t_1, \ldots, t_m)^T$, with $-1 \leq t_1 < t_2 < \ldots < t_m \leq 1$ and defining $\hat{\boldsymbol{\theta}}_n$ as a solution to the set of m equations $\phi_n(t_j) = \phi(t_j, \boldsymbol{\theta})$, for $1 \leq j \leq m$. These estimating equations may be concisely written as

$$\phi_n(\boldsymbol{t}) = \phi(\boldsymbol{t}, \boldsymbol{\theta}) \ , \tag{1}$$

where we have used the convention that $f(\boldsymbol{x})$ represents the vector or matrix obtained by evaluating function f at each element of the vector or matrix $\boldsymbol{x}$.

Other functionals of the pgf are also possible. For example, upon equating derivatives one obtains

$$\phi_n'(\boldsymbol{t}) = \phi'(\boldsymbol{t}, \boldsymbol{\theta}) \ , \tag{2}$$

or by considering derivatives of logs,

$$\phi_n'(\boldsymbol{t}) \ \phi(\boldsymbol{t}, \boldsymbol{\theta}) = \phi'(\boldsymbol{t}, \boldsymbol{\theta}) \ \phi_n(\boldsymbol{t}) \ . \tag{3}$$

It is also plausible to use a combination of these three types of estimating equations. For example, if $m = 2$, the equations $\phi_n(t_1) = \phi(t_1, \boldsymbol{\theta})$ and $\phi_n'(t_2) = \phi'(t_2, \boldsymbol{\theta})$ may be solved simultaneously for $\boldsymbol{\theta}$.

Kemp and Kemp (1988) show that special cases of these methods result in well-known estimation procedures, namely, the *method of moments*, the *method of mean-and-zero-frequency*, and the *method of even-points*. Each of these methods equates theoretical and empirical versions of different statistical transforms to obtain estimating equations. These methods are useful because they yield "quick" estimates that do not require extensive computation and are potentially useful as starting values for more complex procedures. Also, these quick methods often lead to estimators that are competitive with those from maximum likelihood.

In this paper, we provide an asymptotic theory for these epgf-based methods. This theory allows computation of asymptotic efficiency in a unified setting, and suggests asymptotic estimates of standard errors. Furthermore, and perhaps more importantly, the theory aids in selection of $\boldsymbol{t}$, since this vector's effect on covariances can immediately be assessed, at least in the large-sample setting.

2 Asymptotic Theory

As we will note, any of the proposed estimators, $\hat{\boldsymbol{\theta}}_n$, solves an equation of the form

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(X_i,\boldsymbol{\theta})=\mathbf{0}\ , \tag{4}$$

for some $m-$vector valued function Ψ (which depends on $\boldsymbol{t}$), and is thus an M-estimator. (Here, the symbol $\mathbf{0}$ stands for an m-vector of zeros.) Asymptotic properties of M-estimators have been studied by Huber (1967). If there exists a $\boldsymbol{\theta}_0$ such that $X_1, X_2, \ldots, X_n$ have pgf $\phi(t,\boldsymbol{\theta}_0)$, a necessary condition for consistency of the estimator is that $E_{\boldsymbol{\theta}_0}[\Psi(X,\boldsymbol{\theta})]=\mathbf{0}$ only for $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, where $E_{\boldsymbol{\theta}_0}$ denotes expectation under the distribution determined by $\phi(t,\boldsymbol{\theta}_0)$. When consistent, the estimator is asymptotically normal, that is, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)$ converges to a $N_m(0,\Sigma(\boldsymbol{t}))$ distribution as $n\to\infty$. The form of the limiting covariance matrix is

$$\Sigma(\boldsymbol{t}) = A^{-1}(\boldsymbol{t})\ B(\boldsymbol{t})\ A^{-T}(\boldsymbol{t})\ ,$$

where $A(\boldsymbol{t})$ and $B(\boldsymbol{t})$ are the $m\times m$ matrices defined by

$$A(\boldsymbol{t}) = \left.\frac{\partial}{\partial\boldsymbol{\theta}^T}E_{\boldsymbol{\theta}_0}[\Psi(X,\theta)]\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and

$$B(\boldsymbol{t}) = Cov[\Psi(X,\boldsymbol{\theta}_0)] = E_{\boldsymbol{\theta}_0}[\Psi(X,\boldsymbol{\theta}_0)\ \Psi^T(X,\boldsymbol{\theta}_0)]\ .$$

If $A(\boldsymbol{t})$ and $B(\boldsymbol{t})$ are both continuous functions of $\boldsymbol{\theta}_0$, a consistent estimator of the asymptotic covariance matrix is given immediately by

$$\hat{\Sigma}(\boldsymbol{t}) = \hat{A}^{-1}(\boldsymbol{t})\ \hat{B}(\boldsymbol{t})\ \hat{A}^{-T}(\boldsymbol{t})\ , \tag{5}$$

where $\hat{A}(\boldsymbol{t})$ and $\hat{B}(\boldsymbol{t})$ are $A(\boldsymbol{t})$ and $B(\boldsymbol{t})$ evaluated at $\hat{\boldsymbol{\theta}}_n$.

Estimating equations (1), (2), and (3) may be expressed in form (4) with the following definitions of Ψ :

$$\begin{aligned}
\Psi_1(X_i,\boldsymbol{\theta}) &= \boldsymbol{t}^{X_i} - \phi(\boldsymbol{t},\boldsymbol{\theta})\ , \\
\Psi_2(X_i,\boldsymbol{\theta}) &= X_i\ \boldsymbol{t}^{X_i-1} - \phi'(\boldsymbol{t},\boldsymbol{\theta})\ , \\
\Psi_3(X_i,\boldsymbol{\theta}) &= X_i\ \boldsymbol{t}^{X_i-1}\cdot\phi(\boldsymbol{t},\boldsymbol{\theta}) - \boldsymbol{t}^{X_i}\cdot\phi'(\boldsymbol{t},\boldsymbol{\theta})\ .
\end{aligned}$$

(Here the $\cdot$ operator denotes element-by-element multiplication between two vectors or matrices.) The formulas for the $A(\boldsymbol{t})$ and $B(\boldsymbol{t})$ matrices corresponding to these forms of the estimating equations are developed below. The following relationships will be used to simplify the expressions:

$$E_{\boldsymbol{\theta}_0}[X\ t^{X-1}] = \phi'(t,\boldsymbol{\theta}_0) \tag{6}$$

and

$$E_{\boldsymbol{\theta}_0}[X^2\ t^{X-1}] = t\ \phi''(t,\boldsymbol{\theta}_0)\ +\ \phi'(t,\boldsymbol{\theta}_0)\ . \tag{7}$$

The Ψ associated with a set of equations that combines the forms in (1), (2), and (3) would be composed of the appropriate elements from Ψ_1, Ψ_2, and Ψ_3 above. Expressions for the elements in the corresponding asymptotic covariance matrix are readily obtained by methods such as those described below.

2.1 Equating pgf's

Consider first the estimator described by Ψ_1. Notice that $E_{\boldsymbol{\theta}_0}[\Psi_1(X,\boldsymbol{\theta})] = \phi(\boldsymbol{t},\boldsymbol{\theta}_0) - \phi(\boldsymbol{t},\boldsymbol{\theta})$. One obtains

$$\begin{aligned}
A_1(\boldsymbol{t}) &= -\left.\frac{\partial}{\partial\boldsymbol{\theta}^T}\phi(\boldsymbol{t},\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= -\left.\begin{pmatrix} \frac{\partial}{\partial\theta_1}\phi(t_1,\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial\theta_m}\phi(t_1,\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial\theta_1}\phi(t_m,\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial\theta_m}\phi(t_m,\boldsymbol{\theta}) \end{pmatrix}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \left\{ -\left.\frac{\partial}{\partial\theta_j}\phi(t_i,\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}_{i,j=1,\ldots,m}\ ,
\end{aligned}$$

and since

$$E_{\boldsymbol{\theta}_0}[(t^X - \phi(t,\boldsymbol{\theta}_0))\ (t^X - \phi(t,\boldsymbol{\theta}_0))^T] \ = \\ \phi(tt^T,\boldsymbol{\theta}_0) \ - \ 2\phi(t,\boldsymbol{\theta}_0)\ \phi(t^T,\boldsymbol{\theta}_0) \ + \ \phi(t,\boldsymbol{\theta}_0)\ \phi(t^T,\boldsymbol{\theta}_0)\ ,$$

we have

$$B_1(\boldsymbol{t}) \ = \ \phi(\boldsymbol{t}\boldsymbol{t}^T,\boldsymbol{\theta}_0) \ - \ \phi(\boldsymbol{t},\boldsymbol{\theta}_0)\ \phi(\boldsymbol{t}^T,\boldsymbol{\theta}_0)\ .$$

Hence, the asymptotic covariance matrix has a closed form which is a relatively simple function of the pgf which defines the family under consideration. This form is valid whether the estimating equations have an explicit solution or not.

2.2 Equating Derivatives of pgf's

Next, we consider the estimator based on the function Ψ_2, and we obtain $E_{\boldsymbol{\theta}_0}[\Psi_2(X,\boldsymbol{\theta})] = \phi'(\boldsymbol{t},\boldsymbol{\theta}_0) - \phi'(\boldsymbol{t},\boldsymbol{\theta})$ and

$$\begin{aligned} A_2(\boldsymbol{t}) \ &= \ -\left.\frac{\partial}{\partial\boldsymbol{\theta}^T}\phi'(\boldsymbol{t},\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \ -\left.\begin{pmatrix} \frac{\partial}{\partial\theta_1}\phi'(t_1,\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial\theta_m}\phi'(t_1,\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial\theta_1}\phi'(t_m,\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial\theta_m}\phi'(t_m,\boldsymbol{\theta}) \end{pmatrix}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \ \left\{ -\left.\frac{\partial}{\partial\theta_j}\phi'(t_i,\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}_{i,j=1,\ldots,m} \ . \end{aligned}$$

The $(i,j,)$th element of $\left[\Psi_2(X,\boldsymbol{\theta}_0)\ \Psi_2^T(X,\boldsymbol{\theta}_0)\right]$ is just

$$(Xt_i^{X-1} \ - \ \phi'(t_i,\boldsymbol{\theta}_0))\ (Xt_j^{X-1} \ - \ \phi'(t_j,\boldsymbol{\theta}_0)) \ = \\ X^2\ (t_it_j)^{X-1} \ - \ X\ t_i^{X-1}\ \phi'(t_j,\boldsymbol{\theta}) \ - \ X\ t_i^{X-1}\ \phi'(t_j,\boldsymbol{\theta}) \ + \ \phi'(t_i,\boldsymbol{\theta})\phi'(t_j,\boldsymbol{\theta})\ ,$$

so by (6) and (7) we have

$$E_{\boldsymbol{\theta}_0}\left[\left\{\Psi_2(X,\boldsymbol{\theta}_0)\ \Psi_2^T(X,\boldsymbol{\theta}_0)\right\}_{i,j}\right] \ = \\ (t_it_j)\ \phi''(t_it_j,\boldsymbol{\theta}_0) \ + \ \phi'(t_it_j,\boldsymbol{\theta}_0) \ - \ \phi'(t_i,\boldsymbol{\theta}_0)\phi'(t_j,\boldsymbol{\theta}_0)\ .$$

In matrix notation this becomes

$$B_2(\boldsymbol{t}) \ = \ \boldsymbol{t}\boldsymbol{t}^T\cdot\phi''(\boldsymbol{t}\boldsymbol{t}^T,\boldsymbol{\theta}_0) \ + \ \phi'(\boldsymbol{t}\boldsymbol{t}^T,\boldsymbol{\theta}_0) \ - \ \phi'(\boldsymbol{t},\boldsymbol{\theta}_0)\phi'(\boldsymbol{t}^T,\boldsymbol{\theta}_0)\ .$$

2.3 Equating Derivatives of log pgf's

The function Ψ_3 produces asymptotic results obtained as follows. In this case, $E_{\boldsymbol{\theta}_0}[\Psi_3(X,\boldsymbol{\theta})] = \phi'(\boldsymbol{t},\boldsymbol{\theta}_0)\cdot\phi(\boldsymbol{t},\boldsymbol{\theta}) \; - \; \phi(\boldsymbol{t},\boldsymbol{\theta}_0)\cdot\phi'(\boldsymbol{t},\boldsymbol{\theta})$ and

$$A_3(\boldsymbol{t}) = \left\{ \phi'(t_i,\boldsymbol{\theta}_0)\,\frac{\partial}{\partial\theta_j}\phi(t_i,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \; - \; \phi(t_i,\boldsymbol{\theta}_0)\,\frac{\partial}{\partial\theta_j}\phi'(t_i,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}_{i,j=1,\ldots,m} .$$

The (i,j)th element of $\left[\Psi_3(X,\boldsymbol{\theta}_0)\;\Psi_3^T(X,\boldsymbol{\theta}_0)\right]$ is just

$$(X\; t_i^{X-1}\phi(t_i,\boldsymbol{\theta}_0) \; - \; t_i^X\phi'(t_i,\boldsymbol{\theta}_0))\;(X\; t_j^{X-1}\phi(t_j,\boldsymbol{\theta}_0) \; - \; t_j^X\phi'(t_j,\boldsymbol{\theta}_0)) .$$

Proceeding as in the previous section we find that

$$\begin{aligned}\{B_3(\boldsymbol{t})\}_{i,j} \;=\; & [(t_it_j)\;\phi''(t_it_j,\boldsymbol{\theta}_0) \; + \; \phi'(t_it_j,\boldsymbol{\theta}_0)]\;\phi(t_i,\boldsymbol{\theta}_0)\;\phi(t_j,\boldsymbol{\theta}_0)\\ & -\, t_j\;\phi'(t_it_j,\boldsymbol{\theta}_0)\;\phi'(t_j,\boldsymbol{\theta}_0)\;\phi(t_i,\boldsymbol{\theta}_0)\\ & -\, t_i\;\phi'(t_it_j,\boldsymbol{\theta}_0)\;\phi'(t_i,\boldsymbol{\theta}_0)\;\phi(t_j,\boldsymbol{\theta}_0)\\ & +\, \phi(t_it_j,\boldsymbol{\theta}_0)\;\phi'(t_i,\boldsymbol{\theta}_0)\;\phi'(t_j,\boldsymbol{\theta}_0) .\end{aligned}$$

3 Example: The Zero-inflated Binomial

Kemp and Kemp (1988) use the zero-inflated binomial distribution to highlight some important aspects of epgf estimation. Their example is repeated here to illustrate the additional information provided by the asymptotic theory.

The probability mass function (pmf) of the zero-inflated binomial is

$$P_x = \begin{cases} 1-\alpha+\alpha\; q^n\;, & x=0\\ \alpha\;\begin{pmatrix} n\\ x\end{pmatrix}\;p^x q^{n-x}\;, & 1\le x\le n\;,\end{cases}$$

where $0<\alpha<1$, $0<p<1$, $q=1-p$, and $0<n$, an integer. The mean and variance of this distribution are αnp and $\alpha np(np+q-\alpha np)$, respectively, and the pgf is given by

$$\phi(t,\boldsymbol{\theta}) = 1-\alpha+\alpha\;(q\; + \; pt)^n .$$

Kemp and Kemp show that the maximum likelihood estimators for α and p are found by solving a set of equations of the form in (1), *i.e.*, $\phi_n(\boldsymbol{t}) = \phi(\boldsymbol{t},\boldsymbol{\theta})$, where $\boldsymbol{t} = (t_1,t_2)^T$ with $t_1\to 1$ and $t_2 = 0$ (which is equivalent to the method of mean-and-zero-frequency). The resulting set of estimating equations cannot be solved explicitly. Equations of form (2), *i.e.*, $\phi_n'(\boldsymbol{t}) = \phi'(\boldsymbol{t},\boldsymbol{\theta})$, are considered instead since they do yield closed-form estimators for α and p. Four (t_1,t_2)

pairs are discussed: (i) $t_1, t_2 \to 1$, (ii) $(1, 0)$, (iii) $(1, -1)$, and (iv) $(0, -1)$. The first and second pairs correspond, respectively, to method of moments estimation and the method of mean-and-first-frequency. The third and fourth pairs do not represent typically used estimation methods, but are included for comparison.

The epgf method is demonstrated with two data sets. The first involves the germination of acorns from sixty samples of size $n = 5$. A zero-inflated binomial model is used to account for the fact that acorns may be damaged whilst lying on the ground and only the undamaged nuts will germinate according to a binomial distribution. Data set 2 concerns the number of boys in families with exactly eight children (so $n = 8$). The zero-inflation here explains the fact that some parents carry sex-linked genes that inhibit the birth of boys. The data sets are reproduced in Table 1 for reference.

Table 1 : Two zero-inflated binomial data sets.

Data Set 1		Data Set 2	
No. of acorns germinating	Frequency	No. of boys	Frequency
0	22	0	215
1	23	1	1485
2	11	2	5331
3	2	3	10649
4	1	4	14959
5	1	5	11929
		6	6678
		7	2092
		8	342
Total	60	Total	53680

Estimates for α and p were calculated using each of the (t_1, t_2) choices in (i)–(iv), and maximum likelihood estimates were also found for comparison. Large differences were observed between the various methods. The maximum likelihood estimates for the first data set are $\hat{p}_{MLE} = 0.2307$ and $\hat{\alpha}_{MLE} = 0.8669$. In this case method (iii) is particularly poor in that it results in an estimated $\alpha > 1$ (which suggests a zero-*deflated* distribution). The data in the second set leads to the maximum likelihood estimates $\hat{p}_{MLE} = 0.5152$ and $\hat{\alpha}_{MLE} = 0.9990$. Here, method (ii) results in an estimated $\alpha > 1$ and methods (iii) and (iv) both lead to imaginary values for p. Only method (i) (the method of moments) leads to reasonable estimates in both data sets.

The example is used to show that different choices of (t_1, t_2) can lead to quite different estimates. The best choice (in the sense of highest efficiency) does not remain constant even within a given distributional family, but varies depending on the region of parameter space. Also, some choices of (t_1, t_2) are shown to be extremely inefficient even when the data set is large (as in the second example). As we will see below, the asymptotic theory developed in the previous sections provides further insight into these results.

The estimating equations are $\phi_n'(\boldsymbol{t}) = \phi'(\boldsymbol{t}, (\alpha, p))$, so the $A_2(\boldsymbol{t})$ and $B_2(\boldsymbol{t})$ matrices of Section 2.2 lead to the appropriate asymptotic covariance matrix for $(\hat{\alpha}, \hat{p})$. We obtain

$$A(\boldsymbol{t}) = \begin{pmatrix} -\frac{\partial}{\partial\alpha}\phi'(t_1,(\alpha,p)) & -\frac{\partial}{\partial p}\phi'(t_1,(\alpha,p)) \\ -\frac{\partial}{\partial\alpha}\phi'(t_2,(\alpha,p)) & -\frac{\partial}{\partial p}\phi'(t_2,(\alpha,p)) \end{pmatrix}$$

and

$$B(\boldsymbol{t}) \;=\; \boldsymbol{t}\boldsymbol{t}^T \cdot \phi''(\boldsymbol{t}\boldsymbol{t}^T, (\alpha, p)) \;+\; \phi'(\boldsymbol{t}\boldsymbol{t}^T, (\alpha, p)) \;-\; \phi'(\boldsymbol{t}, (\alpha, p))\phi'(\boldsymbol{t}^T, (\alpha, p)) \;,$$

where

$$\begin{aligned} \phi'(t,(\alpha,p)) &= \alpha n p[1-p(1-t)]^{(n-1)} \;, \\ \phi''(t,(\alpha,p)) &= \alpha n(n-1)p^2[1-p(1-t)]^{(n-2)} \;, \\ \frac{\partial}{\partial\alpha}\phi'(t,(\alpha,p)) &= np[1-p(1-t)]^{(n-1)} \;, \\ \frac{\partial}{\partial p}\phi'(t,(\alpha,p)) &= \alpha n[1-np(1-t)][1-p(1-t)]^{(n-2)} \;. \end{aligned} \tag{8}$$

Surface plots can be used to display the asymptotic variances of $\hat{\alpha}$ and $\hat{p}$, and the determinant of the asymptotic covariance matrix as functions of t_1 and t_2. We denote these quantities by $V[\hat{\alpha}]$, $V[\hat{p}]$, and $|\Sigma(\boldsymbol{t})|$, respectively. Figure 1 shows the plots corresponding to data set 1 with $n = 5$, $p = 0.2307$, and $\alpha = 0.8669$ (*i.e.*, the parameter values used to calculate the asymptotic covariance matrix correspond to the maximum likelihood estimates). A log scale on the response is used for clarity. Figure 2 shows similar plots for the second data set ($n = 8$, $p = 0.5152$ and $\alpha = 0.9990$). The method of moments estimator corresponds to the $(t_1, t_2) = (1, 1)$ corner and in each case we see that the variance here is relatively low. On the other hand, estimators obtained using at least one $t_i = -1$ value have large variances. These observations corroborate the results of Kemp and Kemp.

But indeed, the asymptotic theory can be used to obtain more general results that apply to the entire range of p and α values. The asymptotic covariance matrix depends on $A(\boldsymbol{t})^{-1}$ and we see from (8) that $A(\boldsymbol{t})$ is nearly singular when $[1 - p(1 - t)]$ is near 0. Thus, we can expect problems when epgf estimates are based on t values satisfying $p = 1/(1 - t)$. We saw this in the second data set where $\hat{p}_{MLE} \approx .5$ and the estimators with $t_i = -1$ values resulted in imaginary estimates for p. Estimators based on a $t_i = 0$ value would have problems if the true p was near 1. Problems with poorly chosen t values will be amplified as n increases (*e.g.*, $n = 8$ *vs.* $n = 5$). Kemp and Kemp considered t_i values in $(-1, 1)$ because they fall within the radius of convergence of $\phi(t, \theta)$, and because in some cases $t_i = -1$ leads to a known method (*e.g.*, the method of even points). Our analysis suggests that choices should be restricted to $0 < t_i < 1$ for this distribution.

We illustrate these ideas by comparing epgf estimators based on an expanded set of (t_1, t_2) pairs. These pairs are listed below in Table 2.

Table 2 : Ten settings of (t_1, t_2) values.

pair	1	2	3	4	5	6	7	8	9	0
t_1	–1	–1	–1	–1	0	0	0	.5	.5	1
t_2	–1	0	.5	1	0	.5	1	.5	1	1

Figure 3 shows the log determinants of the asymptotic covariance matrices plotted over the range of p, with $\alpha = .9$ and $n = 5$. Qualitative results were similar for other choices of α and n, and with $\ln(V[\hat{p}])$ or $\ln(V[\hat{\alpha}])$ as the response. Those estimators with at least one $t_i = -1$ (pairs 1, 2, 3, 4) are especially bad for p near 0.5 and those with at least one $t_i = 0$ (pairs 2, 5, 6, 7) perform worse for large p. The asymptotic variances of epgf estimators based on positive t_i values only (pairs 8, 9, 0) are well-behaved throughout the parameter space. The method of moments estimator is included in this group (pair 0). Recall this was the only estimator (among those considered by Kemp and Kemp) that lead to reasonable estimates in both data sets. A closer examination in Figure 4 reveals that pair 9 with $(t_1, t_2) = (.5, 1)$ actually performs better over the entire p range. The asymptotic results for maximum likelihood are shown for comparison in the curve labeled "m".

Our analysis and the experience of Kemp and Kemp indicate that method of moments is a reasonable estimation method over the parameter space of the zero-inflated binomial distribution. Knowing this, we could use method of moments estimates to roughly identify the appropriate region of the parameter space, and then determine (t_1, t_2) values leading to improved estimates (*e.g.*, the (t_1, t_2) that minimize $\ln(|\Sigma(\boldsymbol{t})|)$ in this region). The drawback of such a two-stage procedure is the computational cost in locating improved (t_1, t_2) values. Clearly it is important to choose these values carefully, but a computationally intense search contradicts the original purpose of epgf methods—to obtain quick estimates in cases where maximum likelihood estimates are not readily available. A practical alternative uses $(t_1, t_2) = (.5, 1)$ as suggested in the figures above. Results for the two example data sets are summarized in Table 3. In each case the suggested method obtains estimates that are closer to maximum likelihood with minimal effort.

Table 3 : Summary results for zero-inflated binomial examples.

Estimation Method	Data Set 1		Data Set 2	
	α	p	α	p
$(t_1, t_2) = (1, 1)$	0.7152	0.2797	0.9954	0.5171
(method of moments)	(0.1063)	(0.0463)	(0.0008)	(0.0009)
$(t_1, t_2) = (.5, 1)$	0.8476	0.2360	0.9996	0.5149
	(0.1154)	(0.0397)	(0.0003)	(0.0008)
maximum likelihood	0.8669	0.2307	0.9990	0.5152
	(0.1177)	(0.0391)	(0.0003)	(0.0008)

The asymptotic theory provides a method for estimating the standard errors of the epgf estimators. These values are obtained from the diagonal elements of $\hat{\Sigma}(\boldsymbol{t})$ defined in (5) (divided by n), and are displayed in parentheses above. The standard errors for maximum likelihood are provided by the inverse information matrix. Approximate $(1-\alpha)\,100\%$ confidence intervals, based on asymptotic normality, are then given by $estimate \pm z_{\alpha/2} \times estimated\ error$. Table 4 lists approximate 95% intervals for the epgf estimates based on $(t_1, t_2) = (.5, 1)$.

Table 4 : Asymptotic 95% confidence intervals for $(t_1, t_2) = (.5, 1)$ estimator.

Parameter	Data Set 1	Data Set 2
α	(0.6214,1.0738)	(0.9989,1.0002)
p	(0.1582,0.3138)	(0.5134,0.5164)

The interval for α in the second data set suggests that the appropriate model may not be zero-inflated at all, but rather a binomial model with a success probability greater than 0.5. The wide intervals for data set 1, if they are indeed correct, indicate that the model is difficult to identify in this region of the parameter space, and that a sample size considerably larger than 60 is required to achieve sensible precision. Below, we briefly discuss other methods of interval estimation that were used to verify these results for data set 1.

Bootstrapping is a familiar general-purpose method for obtaining confidence intervals. It is used here for comparison with the asymptotic results. 5000 bootstrapped samples of size 60 were generated from the empirical distribution function of data set 1. Note that this method is nonparametric, as opposed to parametric bootstrapping which would generate samples from a zero-inflated distribution with parameters $\hat{\alpha}$ and $\hat{p}$. Epgf estimates based on $(t_1, t_2) = (.5, 1)$ were calculated for each of the 5000 samples and the resulting intervals given by the 2.5% and 97.5% sample quantiles are $(0.6381, 1.2512)$ for the parameter α and $(0.1355, 0.3321)$ for p. The intervals are similar to, and even wider than, those obtained using the asymptotic theory.

Wald-type asymptotic confidence intervals based on the MLE for data set 1 are $(0.6364, 1.0974)$ and $(0.1541, 0.3073)$ for α and p, respectively. Note again that they are similar in length and location to the intervals produced by the epgf asymptotic theory.

The bootstrap and Wald methods corroborate the wide intervals seen with the asymptotic method in data set 1. We can also quantify how well these intervals perform, regardless of comparison with these other methods. A small-scale simulation was performed in which 5000 simulated samples of size $n = 60$ were drawn from a zero-inflated binomial with parameters $\alpha = .8669$ and $p = .2307$ (*i.e.*, the MLE values). For each generated sample, epgf estimates based on $(t_1, t_2) = (.5, 1)$ and the 95% asymptotic theory confidence intervals were calculated. The resulting observed coverage probabilities were 96.3% for the α intervals and 94.4% for p, indicating that the theory is fairly accurate and useful.

4 Example: The Negative Binomial

An example using the negative binomial distribution shows the usefulness of the epgf asymptotic results in a different context. This distribution arises in a variety of applications and has been studied in detail. For instance, Bogdanoff and Kozin (1985) use the negative binomial pgf and related variants in certain Markov-chain settings in reliability theory. Maximum likelihood and the more traditional method of moments are typical estimation methods used in this application, but the epgf approach is a sensible alternative to consider since the pgf is known explicitly. A common parametrization of the negative binomial, used in Kemp and Kemp and repeated here, has the pgf

$$\phi(t, \boldsymbol{\theta}) = (1 + q - qt)^{-k}, \quad 0 < k, \ 0 < q \ .$$

The mean is kq, the variance is $kq(1 + q)$, and the pmf is given by

$$P_x = \binom{k + x - 1}{x} q^x (1 + q)^{-k-x}, \quad x = 0, 1, \ldots \ .$$

Explicit maximum likelihood estimates for q and k are not possible and typically used quick estimates (method of moments and the method of mean-and-zero-frequency) work well only in limited areas of the parameter space. Kemp and Kemp were thus motivated to develop quick estimators that would be reasonably efficient over the entire parameter space. They arrived at a two-step solution through mathematical approximation of the maximum likelihood equations. The approach is equivalent to equating derivatives of the log pgf and epgf (see Section 2.3) with $t_1 = 1$ and t_2 determined by the data as follows:

$$t_2 = \begin{cases} (\tilde{k}+1)/(\tilde{k}+2) & \text{if } \bar{x} \le 2 \, , \\ \left(\tilde{k}+\frac{\bar{x}}{2}\right) / \left(\tilde{k}+1+\frac{\bar{x}}{2}\right) & \text{if } \bar{x} > 2 \, , \end{cases}$$

where $\tilde{k}$ is given by method of moments (*i.e.*, $\tilde{q} = (s^2 - \bar{x})/\bar{x}$, $\tilde{k} = \bar{x}/\tilde{q}$). A large simulation study was used to verify the method's high efficiency over the entire parameter space. We will show that the asymptotic theory developed earlier in this paper provides quick and easy support for the suggested method.

The estimating equations for (k, q) are of the form $\phi_n'(\boldsymbol{t})\ \phi(\boldsymbol{t}, (k, q)) = \phi'(\boldsymbol{t}, (k, q))\ \phi_n(\boldsymbol{t})$, so the $A_3(\boldsymbol{t})$ and $B_3(\boldsymbol{t})$ matrices of Section 2.3 lead to the appropriate asymptotic covariance matrix for $(\hat{k}, \hat{q})$. The value $t_1 = 1$ is set, and the "best" t_2 (*e.g.*, that which minimizes the determinant of the covariance matrix) can be determined for any (k, q) in the parameter space. Figure 5 plots the optimal t_2 values for various combinations of k and q. We would like to plot Kemp and Kemp's t_2 values for comparison, but they depend on data through $\bar{x}$ and $\tilde{k}$. Instead we use their rule with $E[X] = kq$ and k in place of $\bar{x}$ and $\tilde{k}$, respectively. Figure 6 shows the resulting t_2 values for various (k, q) combinations. The strong similarity between the two graphs gives further evidence that Kemp and Kemp's suggested method will perform well over the parameter space.

5 Concluding Remarks

It is a known fact that there is no single "quick" estimation method which is efficient for all families of distributions. In fact, even for a single distribution there are instances where a given method performs well over some regions of the parameter space but not others. In the epgf method, $\boldsymbol{t}$ is a tuning constant which leads to different estimators of $\boldsymbol{\theta}$ with every choice. It is important to establish a criterion for optimality when estimating a multi-dimensional parameter. We have shown that a standard theory does exist for assessing the asymptotic variance for $\boldsymbol{\theta}$ estimators in this class. The theory can be used for guidance in choosing $\boldsymbol{t}$ values wisely. This was demonstrated for the zero-inflated binomial distribution, where we identified a quick estimator that has improved efficiency relative to the typically used method of moments over the entire parameter space. Graphs based on the asymptotic theory were shown to be useful exploratory tools that can be used in conjunction with simulation studies. The methodology used in the examples can be easily extended to investigate other distributions.

In other situations it may be the case that interest is not centered on estimating $\boldsymbol{\theta}$ itself, but some function $g(\boldsymbol{\theta})$ instead. For example, engineers in reliability problems are more interested in the cumulative distribution function $F(x, \boldsymbol{\theta})$ associated with $\phi(t, \boldsymbol{\theta})$ than with $\boldsymbol{\theta}$ itself, since the parameter $\boldsymbol{\theta}$ may not even have physical meaning. The method of maximum likelihood enjoys an invariance principle; *i.e.*, if $\tilde{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, then $g(\tilde{\boldsymbol{\theta}})$ is an efficient estimator of $g(\boldsymbol{\theta})$. This is not true in general for other methods, including the epgf approach. However, if g is a smooth function of $\boldsymbol{\theta}$, it may be possible to obtain asymptotics for the estimator $g(\hat{\boldsymbol{\theta}})$ based on the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ via the delta-method. One could then proceed to improve the estimate of $g(\boldsymbol{\theta})$ through an appropriate selection of $\boldsymbol{t}$ used in the estimation of $\boldsymbol{\theta}$.

References

Baringhaus, L. and Henze, N. (1992), A goodness of fit test for the Poisson distribution based on the empirical generating function, *Statistics and Probability Letters* **13**, 269-274.

Bogdanoff, J.L. and Kozin, F. (1985), *Probabilistic Models of Cumulative Damage,* John Wiley and Sons, New York.

Csörgó, S. and Teugels, J. L. (1990), Empirical Laplace transform and approximation of compound distributions, *Journal of Applied Probability* **27**, 88-101.

Feuerverger, A. and McDunnough, P. (1984), On statistical transform methods and their efficiency, *Canadian Journal of Statistics* **12**, 303-317.

Huber, P. J. (1967), Behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 221-233.

Kemp, C. D., and Kemp, A. W. (1988), Rapid estimation for discrete distributions, *The Statistician* **37**, 243-255.

Kocherlakota, S. and Kocherlakota, K. (1986), Goodness of fit tests for discrete distributions, *Communications in Statistics: Theory and Methods* **A 15**, 815-829.

Kocherlakota, S. and Kocherlakota, K. (1990), Tests of hypothesis for the weighted binomial distribution, *Biometrics* **46**, 645–656.

Marques, M. and Pérez-Abreu, V. (1989), Law of large numbers and central limit theorem of the empirical probability generating function of stationary random sequences and processes, *Aportaciones Matemáticas, Sociedad Matemática Mexicana* **4**, 100-109.

Nakamura, M. and Pérez-Abreu, V. (1993a), Use of the empirical probability generating function for testing a Poisson Model, *The Canadian Journal of Statistics,* **21**, 149-156.

Nakamura, M. and Pérez-Abreu, V. (1993b), Exploratory data analysis for counts using the empirical probability generating function, *Communications in Statistics: Theory and Methods,* **22**, 827-842.

Rueda, R., Pérez-Abreu, V., and O'Reilly, F. (1991), Goodness of fit for the Poisson distribution based on the probability generating function, *Communications in Statistics: Theory and Methods* **A 20**, 3093-3110.

List of Figures