Ann. Inst. Statist. Math.
Vol. 43, No. 1, 167–179 (1991)

ESTIMATING MEANS FROM A NON-I.I.D. MIXTURE OF POISSON SAMPLES

STEVEN G. FROM

Department of Mathematics and Computer Science, University of Nebraska-Omaha, Omaha, NB 68182-0243, U.S.A.

(Received May 23, 1989; revised May 15, 1990)

Abstract. In this paper, a family of estimators for estimating means when mixing two independent Poisson samples is proposed. This family is based on the probability-generating function of the Poisson distribution and is offered as an alternative to the maximum likelihood estimators, which have some drawbacks. These estimators include the method of moments estimators as a special limiting case.

Key words and phrases: Poisson distribution, probability-generating function, generalized variance.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be independent random variables such that $X_{v_1}, \dots, X_{v_m}$ have mean θ_2 and $X_{v_{m+1}}, \dots, X_{v_n}$ have mean θ_1, where $1 \leq m < n$ and $0 \leq \theta_1 < \theta_2 < \infty$. It is assumed that m and n are known and that $(v_1, \dots, v_m)$, the vector of indices of observations with mean θ_2, is a nuisance parameter. The problem is to estimate (θ_1, θ_2), the parameter of interest. This kind of problem arises in real-life situations where confidentiality of a person's particular group membership is extremely important. This problem has been considered before, usually assuming that θ_1 and θ_2 are both exponential or both normal means. Usually, an m-outlier model is assumed. See Kale (1975), Shaked and Tran (1982), Gather and Kale (1988) and From and Saxena (1989). Kale presented the maximum likelihood estimators under the assumption that the X_i's have distributions with monotone likelihood ratio in x. These estimators are sometimes inadequate because of inconsistency and extreme bias.

It is emphasized that X_i's do not form an i.i.d. collection of random variables so that we are not talking about an i.i.d. mixture. There is an exhaustive literature in the i.i.d. case, none of which will be given here, with one exception. It should be mentioned that the approach taken in Section 2 is similar to that used by Quandt and Ramsey (1978) to estimate the parameters of an i.i.d. mixture of normal distributions. The major difference is that the estimators proposed in Section 2 are of closed form.

In this paper, it is assumed throughout that θ_1 and θ_2 are Poisson means, i.e.,

$$P[X_i = x] = \frac{\theta^x e^{-\theta}}{x!}, \qquad x = 0, 1, 2, \ldots,$$

where $\theta = \theta_1$ or θ_2, $i = 1, \ldots, n$. The probability-generating function (pgf) of X_i is

$$P_{X_i}(\alpha) = E(\alpha^{X_i}) = e^{\theta(\alpha-1)}. \tag{1.1}$$

Under the m-outlier model given in Gather and Kale (1988), the maximum likelihood estimators are given by the trimmed means

$$\hat{\theta}_1^{\mathrm{mle}} = \frac{1}{n-m}\sum_{i=1}^{n-m} X_{(i)} \quad \text{and} \quad \hat{\theta}_2^{\mathrm{mle}} = \frac{1}{m}\sum_{i=n-m+1}^{n} X_{(i)} \tag{1.2}$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics.

In Section 2, a family of estimators of θ_1 and θ_2 is proposed as an alternative to the maximum likelihood estimators, and the relationship of this family to the method of moments estimators is discussed. In Section 3, some numerical studies are presented.

It should be mentioned that in outlier problems, one considers (n, m) relatively small. In this paper, the asymptotic properties of the estimators are considered as $m \to \infty$, $n \to \infty$ in such a way that $m = np + O(n^\delta)$, where $0 < p < 1$ is fixed and $\delta < 1/2$.

In the sequel, θ_1 and θ_2 are used both as dummy variables to be solved for in defining the estimators of Section 2 as well as the true values of these parameters. The same is true for the variables u_1 and u_2 in Section 2. The particular usage will be clear from the context.

2. Estimators of θ_1 and θ_2

It is clear that if $T(X_1, \ldots, X_n)$ is any symmetric function of the observations then the distribution of T depends only on (θ_1, θ_2) and not on the nuisance parameter v.

In this section, the method of moments estimators and the proposed probability-generating function-based estimators of θ_1 and θ_2 are presented. Properties of these estimators are also discussed.

2.1 *Method of moments estimators*

Let

$$W_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad W_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

The method of moments estimators $\hat{\theta}_1^{\mathrm{mom}}$ and $\hat{\theta}_2^{\mathrm{mom}}$ satisfy the equations

$$W_1 = \left(1 - \frac{m}{n}\right)\theta_1 + \frac{m}{n}\theta_2 \tag{2.1}$$

and

$$W_2 = \left(1 - \frac{m}{n}\right)(\theta_1 + \theta_1^2) + \frac{m}{n}(\theta_2 + \theta_2^2). \tag{2.2}$$

In closed form,

$$\hat{\theta}_1^{\text{mom}} = W_1 - \sqrt{\frac{p}{1-p}(W_2 - W_1 - W_1^2)} \tag{2.3}$$

and

$$\hat{\theta}_2^{\text{mom}} = W_1 + \sqrt{\frac{1-p}{p}(W_2 - W_1 - W_1^2)} \tag{2.4}$$

provided $W_2 - W_1 - W_1^2 \geq 0$.

It is clear that the variances of W_1, W_2 and $W_2 - W_1 - W_1^2$ are all $O(n^{-1})$. Since all non-negative integer moments of a Poisson distribution exist, the weak law of large numbers or Tchebychev's theorem implies that

$$W_1 \xrightarrow{p} (1-p)\theta_1 + p\theta_2 \quad \text{and} \quad W_2 \xrightarrow{p} (1-p)(\theta_1 + \theta_1^2) + p(\theta_2 + \theta_2^2).$$

Consequently, $W_2 - W_1 - W_1^2 \xrightarrow{p} p(1-p)(\theta_2 - \theta_1)^2 > 0$, upon replacing W_1 and W_2 by their means given in (2.1) and (2.2). Therefore, the estimators can be found by studying their properties when $W_2 - W_1 - W_1^2 > 0$ since it is easily seen that the joint distribution of $(\hat{\theta}_1^{\text{mom}}, \hat{\theta}_2^{\text{mom}})$ is asymptotically normal. The variance covariance matrix is given in the Appendix.

Next, the method of moments estimators are imbedded in a more general family of estimators. This will allow one to produce more efficient estimators than $\hat{\theta}_i^{\text{mom}}$, $i = 1, 2$, for many parameter combinations of p, θ_1 and θ_2.

2.2 *A pgf-based family of estimators of θ_1 and θ_2*

Let $\alpha > 1/2$, $\alpha \neq 1$. Let

$$A_1 = A_1(\alpha;\ X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} \alpha^{X_i}$$

and

$$A_2 = A_2(\alpha;\ X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} (2\alpha - 1)^{X_i}.$$

Define a family of estimators $\hat{\theta}_i(\alpha)$, $i = 1, 2$ indexed by α as follows. Let $\hat{\theta}_1(\alpha)$ and $\hat{\theta}_2(\alpha)$ be the values of θ_1 and θ_2 satisfying the system

$$A_1 = \left(1 - \frac{m}{n}\right)u_1 + \frac{m}{n}u_2 \tag{2.5}$$

and

$$A_2 = \left(1 - \frac{m}{n}\right)u_1^2 + \frac{m}{n}u_2^2 \tag{2.6}$$

where

$$u_1 = e^{\theta_1(\alpha - 1)} \quad \text{and} \quad u_2 = e^{\theta_2(\alpha - 1)}.$$

The motivation for (2.5) and (2.6) comes from the fact that

$$E(A_1) = \left(1 - \frac{m}{n}\right) u_1 + \frac{m}{n} u_2$$

and

$$E(A_2) = \left(1 - \frac{m}{n}\right) u_1^2 + \frac{m}{n} u_2^2.$$

The system defined by (2.5) and (2.6) has a unique solution satisfying $\hat{\theta}_1(\alpha) \leq \hat{\theta}_2(\alpha)$ provided $A_2 \geq A_1^2$. Since the variance of $A_2 - A_1^2$ is $O(n^{-1})$, either Tchebychev's theorem or the weak law of large numbers gives

$$A_2 - A_1^2 \xrightarrow{P} p(1-p)[e^{\theta_1(\alpha-1)} - e^{\theta_2(\alpha-1)}]^2 > 0.$$

In order to define the pgf-based estimators, we need to distinguish two cases.

Suppose $1/2 < \alpha < 1$. Let

$$\hat{u}_1 = A_1 + \sqrt{\frac{p}{1-p}(A_2 - A_1^2)}$$

and

$$\hat{u}_2 = A_1 - \sqrt{\frac{1-p}{p}(A_2 - A_1^2)}.$$

Again, since the variances of $\hat{u}_1$ and $\hat{u}_2$ are $O(n^{-1})$,

$$\hat{u}_1 \xrightarrow{P} e^{\theta_1(\alpha-1)} \leq 1 \quad \text{and} \quad \hat{u}_2 \xrightarrow{P} e^{\theta_2(\alpha-1)} \leq 1.$$

Clearly $\hat{u}_1 \geq \hat{u}_2$ when $A_2 - A_1^2 > 0$. Thus, $P(\hat{u}_2 \leq \hat{u}_1 \leq 1) \to 1$ as $n \to \infty$. Similarly, if $\alpha > 1$,

$$\hat{u}_1 = A_1 - \sqrt{\frac{p}{1-p}(A_2 - A_1^2)},$$

and

$$\hat{u}_2 = A_1 + \sqrt{\frac{1-p}{p}(A_2 - A_1^2)},$$

then $P(1 \leq \hat{u}_1 \leq \hat{u}_2) \to 1$ as $n \to \infty$. In either of the above two cases for α, we have

$$\hat{\theta}_1(\alpha) = \frac{\ln \hat{u}_1}{\alpha - 1} \quad \text{and} \quad \hat{\theta}_2(\alpha) = \frac{\ln \hat{u}_2}{\alpha - 1},$$

with probability tending to 1 as $n \to \infty$.

By standard large sample theory given in Serfling (1980), it is clear that $\hat{\theta}_1(\alpha)$ and $\hat{\theta}_2(\alpha)$ are consistent (for $0 < \theta_1 \leq \theta_2$) and asymptotically normal (for $0 < \theta_1 < \theta_2$), for all $\alpha > 1/2$, $\alpha \neq 1$. The asymptotic variance covariance matrix of $\sqrt{n}(\hat{\theta}_i(\alpha) - \theta_i)$, $i = 1, 2$, is presented in the Appendix.

The following theorem describes the relationship between $\hat{\theta}_i(\alpha)$ and $\hat{\theta}_i^{\mathrm{mom}}$, $i = 1, 2$.

THEOREM 2.1. *For* $0 < \theta_1 < \theta_2$, *we have*

$$P\left[\lim_{\alpha\to 1^-} \hat{\theta}_i(\alpha) = \hat{\theta}_i^{\text{mom}},\ i = 1,\ 2\right] \to 1$$

as $n \to \infty$, $m \to \infty$ *and* $m/n \to p \in (0,\ 1)$.

PROOF. By the law of large numbers,

$$P[W_2 - W_1 - W_1^2 > 0,\ A_2 - A_1^2 > 0] \to 1.$$

It suffices to show that $W_2 - W_1 - W_1^2 > 0$ and $A_2 - A_1^2 > 0$ imply $\hat{\theta}_i(\alpha) \to \hat{\theta}_i^{\text{mom}}$ as $\alpha \to 1^-$, $i = 1, 2$. The proof is for $i = 1$. Suppose $W_2 - W_1 - W_1^2 > 0$ and $A_2 - A_1^2 > 0$. Then

$$\sum_{j=1}^{n} X_j > 0 \quad \text{or} \quad X_j \geq 1 \tag{2.7}$$

for some j. By L'Hôpital's rule,

$$\begin{aligned}
\lim_{\alpha\to 1^-} \hat{\theta}_1(\alpha) &= \lim_{\alpha\to 1^-} \frac{\ln \hat{u}_1}{\alpha - 1} \\
&= \lim_{\alpha\to 1^-} \frac{\dfrac{\partial \hat{u}_1}{\partial \alpha}}{\hat{u}_1} \\
&= \lim_{\alpha\to 1^-} \left[\frac{\dfrac{\partial A_1}{\partial \alpha} + \dfrac{\left(\dfrac{\partial A_2}{\partial \alpha} - 2A_1\dfrac{\partial A_1}{\partial \alpha}\right)}{2\sqrt{\dfrac{p}{1-p}(A_2 - A_1^2)}}\left(\dfrac{p}{1-p}\right)}{\hat{u}_1}\right] \\
&= \lim_{\alpha\to 1^-} \left(\frac{1}{n}\sum_{i=1}^{n} X_i \alpha^{X_i - 1}\right) + \sqrt{\frac{p}{1-p}} L,
\end{aligned} \tag{2.8}$$

where

$$L = \lim_{\alpha\to 1^-} \frac{\dfrac{1}{2}\dfrac{\partial f}{\partial \alpha}(\alpha;\ X_1, \ldots,\ X_n)}{\sqrt{f(\alpha;\ X_1, \ldots,\ X_n)}} \tag{2.9}$$

and

$$f(\alpha;\ X_1, \ldots,\ X_n) = A_2(\alpha;\ X_1, \ldots,\ X_n) - (A_1(\alpha;\ X_1, \ldots,\ X_n))^2,$$

provided L exists. By (2.7), f and $\partial f/\partial \alpha$ are both non-constant polynomials in α, given $X_1, \ldots,\ X_n$. Thus, eigher $|L| = +\infty$ or L exists. Since

$$\lim_{\alpha\to 1^-} \sqrt{f(\alpha;\ X_1, \ldots,\ X_n)} = \lim_{\alpha\to 1^-} \frac{1}{2}\frac{\partial f}{\partial \alpha}(\alpha;\ X_1, \ldots,\ X_n) = 0,$$

we apply L'Hôpital's rule once more to show L exists. Now L can be written as

$$L = \lim_{\alpha \to 1^-} \frac{\partial}{\partial \alpha} \left(\sqrt{f(\alpha;\ X_1, \ldots, X_n)} \right). \tag{2.10}$$

Since

$$f(\alpha;\ X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (2\alpha - 1)^{X_i} - \left(\frac{1}{n} \sum_{i=1}^{n} \alpha^{X_i} \right)^2, \tag{2.11}$$

we have

$$\begin{aligned} \frac{1}{2} \frac{\partial f}{\partial \alpha}(\alpha;\ X_1, \ldots, X_n) &= \frac{1}{n} \sum_{i=1}^{n} X_i (2\alpha - 1)^{X_i - 1} \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^{n} \alpha^{X_i} \right) \left(\frac{1}{n} \sum_{i=1}^{n} X_i \alpha^{X_i - 1} \right) \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} \frac{1}{2} \frac{\partial^2 f}{\partial \alpha^2}(\alpha;\ X_1, \ldots, X_n) &= \frac{2}{n} \sum_{i=1}^{n} X_i (X_i - 1)(2\alpha - 1)^{X_i - 2} \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^{n} \alpha^{X_i} \right) \left(\frac{1}{n} \sum_{i=1}^{n} X_i (X_i - 1) \alpha^{X_i - 2} \right) \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \alpha^{X_i - 1} \right) \left(\frac{1}{n} \sum_{i=1}^{n} X_i \alpha^{X_i - 1} \right). \end{aligned} \tag{2.13}$$

By (2.13),

$$\begin{aligned} \lim_{\alpha \to 1^-} \frac{1}{2} \frac{\partial^2 f}{\partial \alpha^2}(\alpha;\ X_1, \ldots, X_n) &= \frac{1}{n} \sum_{i=1}^{n} 2X_i (X_i - 1) \\ &\quad - \frac{1}{n} \sum_{i=1}^{n} X_i (X_i - 1) - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \sum_{i=1}^{n} X_i - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \\ &= W_2 - W_1 - W_1^2 > 0. \end{aligned}$$

Thus,

$$\lim_{\alpha \to 1^-} \frac{1}{2} \frac{\partial^2 f}{\partial \alpha^2}(\alpha;\ X_1, \ldots, X_n) = W_2 - W_1 - W_1^2. \tag{2.14}$$

Now if $|L| = +\infty$, then (2.8), (2.10), (2.14) and L'Hôpital's rule imply $M = 0$ where

$$M = \lim_{\alpha \to 1^-} \frac{\frac{1}{2}\frac{\partial^2 f}{\partial \alpha^2}(\alpha;\ X_1, \ldots, X_n)}{\frac{\partial}{\partial \alpha}\left(\sqrt{f(\alpha;\ X_1, \ldots, X_n)}\right)}. \tag{2.15}$$

Since $L = M$, $L = 0$, a contradiction to $|L| = +\infty$. Thus, L exists. By (2.10), (2.14) and (2.15),

$$M = L = \frac{W_2 - W_1 - W_1^2}{L} \quad \text{or} \quad L^2 = W_2 - W_1 - W_1^2$$

and so $L = \pm\sqrt{W_2 - W_1 - W_1^2}$. The sign of L is now determined. Since

$$\lim_{\alpha \to 1^-} \frac{1}{2}\frac{\partial f}{\partial \alpha}(\alpha;\ X_1, \ldots, X_n) = 0$$

and

$$\lim_{\alpha \to 1^-} \frac{1}{2}\frac{\partial^2 f}{\partial \alpha^2}(\alpha;\ X_1, \ldots, X_n) > 0,$$

$\exists \alpha_0 = \alpha_0(X_1, \ldots, X_n) \in (1/2, 1) \ni (1/2)(\partial^2 f/\partial \alpha^2)(\alpha;\ X_1, \ldots, X_n) > 0$ on $(\alpha_0, 1)$. So $(1/2)(\partial f/\partial \alpha)$, as a function of α, is increasing on $(\alpha_0, 1)$ and approaches zero at 1. This gives

$$\frac{1}{2}\frac{\partial f}{\partial \alpha}(\alpha;\ X_1, \ldots, X_n) \leq 0$$

on $(\alpha_0, 1)$. So $L \leq 0$ and $L = -\sqrt{W_2 - W_1 - W_1^2}$. Thus, by (2.15),

$$\begin{aligned}
\lim_{\alpha \to 1^-} \hat{\theta}_1(\alpha) &= \frac{1}{n}\sum_{i=1}^{n} X_i + \left(\sqrt{\frac{p}{1-p}}\right)\left(-\sqrt{W_2 - W_1 - W_1^2}\right) \\
&= W_1 - \sqrt{\frac{p}{1-p}(W_2 - W_1 - W_1^2)} \\
&= \hat{\theta}_1^{\text{mom}}.
\end{aligned}$$

Similarly,

$$\lim_{\alpha \to 1^-} \hat{\theta}_2(\alpha) = W_1 + \sqrt{\frac{1-p}{p}(W_2 - W_1 - W_1^2)} = \hat{\theta}_2^{\text{mom}}. \qquad \square$$

Remark 1. A similar theorem holds as $\alpha \to 1^+$.

Remark 2. Let "AV" stand for asymptotic variance. It is *not* true that

$$\lim_{\alpha \to 1^-} \text{AV}(\sqrt{n}(\hat{\theta}_i(\alpha) - \theta_i)) = \text{AV}(\sqrt{n}(\hat{\theta}_i^{\text{mom}} - \theta_i)), \qquad i = 1, 2.$$

In fact, it can be shown that

$$\lim_{\alpha \to 1^-} \mathrm{AV}(\sqrt{n}(\hat{\theta}_i(\alpha) - \theta_i)) = +\infty, \qquad i = 1,\, 2.$$

This is not surprising since $\alpha = 1$ is a boundary case.

Remark 3. The above theorem states that $\hat{\theta}_i^{\mathrm{mom}}$ can be thought of as special limiting cases of $\hat{\theta}_i(\alpha)$, $i = 1,\, 2$.

3. The selection of optimal α-values

In this section, the asymptotic Wilks generalized variances of the method of moments estimators and the pgf-based estimators are compared. Let W_{mom} and W_{pgf} denote the asymptotic Wilks generalized variances of the method of moments and the pgf-based estimators, respectively. From the appendix,

$$W_{\mathrm{mom}} = \text{determinant of } G_{\mathrm{mom}} \Sigma_{\mathrm{mom}} G_{\mathrm{mom}}^T$$

and

$$W_{\mathrm{pgf}} = \text{determinant of } G_{\mathrm{pgf}} \Sigma_{\mathrm{pgf}} G_{\mathrm{pgf}}^T.$$

Define the asymptotic relative efficiency of $(\hat{\theta}_1(\alpha),\, \hat{\theta}_2(\alpha))$ relative to $(\hat{\theta}_1^{\mathrm{mom}},\, \hat{\theta}_2^{\mathrm{mom}})$ by

$$e = \frac{W_{\mathrm{mom}}}{W_{\mathrm{pgf}}}.$$

This was calculated for various parameter combinations using the α-value in $[.51,\, .99]$ which minimizes W_{pgf}. This optimal α, along with e, is presented in Table 1. The parameter combinations in Table 1 are just a small fraction of all the parameter combinations the author considered, however, these are representative.

From our detailed calculations, several overall conclusions can be made.

(i) Given θ_1 and θ_2, as p increases, the optimal α decreases and e increases. Thus, the larger p is, the better the pgf-based estimators perform in comparison to the method of moments estimators.

(ii) For a given value of p, the larger θ_2 is in comparison to θ_1, the greater the efficiency of the pgf-based estimators relative to the method of moments estimators.

Next, $(\hat{\theta}_1^{\mathrm{mom}},\, \hat{\theta}_2^{\mathrm{mom}})$ and $(\hat{\theta}_1(\alpha),\, \hat{\theta}_2(\alpha))$ are compared to the trimmed means maximum likelihood estimators of Gather and Kale (1988) and also to the maximum likelihood estimates obtained under the classical mixture model where $X_1, \ldots, X_n$ are i.i.d. with probability distribution function

$$g(x,\, p,\, \theta_1,\, \theta_2) = P[X_i = x] = (1-p)\frac{\theta_1^x e^{-\theta_1}}{x!} + p\frac{\theta_2^x e^{-\theta_2}}{x!}, \qquad x = 0,\, 1,\, 2, \ldots,$$

Table 1.

p	.01	.10	.10	.30	.30	.50
θ_1	.50	.50	.50	.50	1.00	1.00
θ_2	2.00	2.00	5.00	5.00	2.00	2.00
α	.99	.94	.90	.84	.95	.88
e	.97	1.02	1.12	1.41	1.01	1.09

p	.50	.70	.90	.99	.01	.10
θ_1	1.00	1.00	2.00	2.00	.10	.10
θ_2	5.00	5.00	3.00	3.00	.20	.20
α	.85	.80	.86	.84	.99	.99
e	1.45	1.97	1.20	1.28	$.99^+$	$.99^+$

p	.10	.30	.30	.50	.50	.70
θ_1	.10	.10	.10	.10	.10	.10
θ_2	.50	.50	.90	.90	3.00	3.00
α	.82	.68	.62	.59	.62	.58
e	1.04	1.14	1.39	1.52	3.55	5.04

p	.90	.99	.01	.10	.10	.30
θ_1	.50	.50	2.00	2.00	2.00	2.00
θ_2	1.00	1.00	6.00	6.00	10.00	10.00
α	.77	.75	.99	.99	.97	.94
e	1.18	1.20	.93	$.99^+$	1.02	1.12

p	.30	.50	.50	.70	.90	.99
θ_1	5.00	5.00	5.00	5.00	.10	.10
θ_2	6.00	6.00	10.00	10.00	20.00	20.00
α	.99	.99	.95	.92	.88	.84
e	$.99^+$	1.00^+	1.10	1.29	39.36	369.01

where p is known and $0 < \theta_1 < \theta_2 < \infty$. Let $(\hat{\theta}_1^{\text{IID}}, \hat{\theta}_2^{\text{IID}})$ denote the values of (θ_1, θ_2) which maximize the traditional likelihood function $L = \prod_{i=1}^{n} g(X_i, m/n, \theta_1, \theta_2)$ subject to the restriction $\hat{\theta}_1^{\text{mle}} \le \theta_1 \le \theta_2 \le \hat{\theta}_2^{\text{mle}}$. Since it is unknown to the author whether or not the classical asymptotic theory of i.i.d. maximum likelihood estimators is applicable here, empirical Wilks generalized variances based on 200 replications of each parameter combination of m, n, θ_1 and θ_2 were found. The actual values of θ_1 and θ_2 were used to find these generalized variances. Table 2 presents a small representative fraction of the actual results generated. The optimal value of α was used to find $(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$.

Table 2. Generalized variances.

$(m, n, \theta_1, \theta_2)$	$(\hat{\theta}_1^{\text{mom}}, \hat{\theta}_2^{\text{mom}})$	$(\hat{\theta}_1^{\text{mle}}, \hat{\theta}_2^{\text{mle}})$	$(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$	$(\hat{\theta}_1^{\text{IID}}, \hat{\theta}_2^{\text{IID}})$
(10, 100, 1.0, 2.0)	.0073	.0249	.0073	.0067
(10, 100, 1.0, 3.0)	.0082	.0113	.0082	.0079
(10, 100, 1.0, 5.0)	.0048	.0053	.0047	.0079
(10, 100, 1.0, 10.0)	.0120	.0097	.0112	.0183
(15, 50, 1.0, 2.0)	.0110	.0264	.0108	.0111
(15, 50, 1.0, 3.0)	.0135	.0213	.0130	.0121
(15, 50, 1.0, 5.0)	.0165	.0138	.0148	.0185
(15, 50, 1.0, 10.0)	.0227	.0172	.0201	.0265
(50, 100, 1.0, 2.0)	.0037	.0135	.0035	.0037
(50, 100, 1.0, 3.0)	.0038	.0109	.0033	.0034
(50, 100, 1.0, 5.0)	.0033	.0030	.0023	.0041
(50, 100, 1.0, 10.0)	.0067	.0034	.0043	.0061
(35, 50, 1.0, 2.0)	.0175	.0339	.0171	.0151
(35, 50, 1.0, 3.0)	.0426	.0266	.0329	.0372
(35, 50, 1.0, 5.0)	.0278	.0139	.0171	.0195
(35, 50, 1.0, 10.0)	.0352	.0140	.0180	.0325
(90, 100, 1.0, 2.0)	.0133	.0251	.0123	.0109
(90, 100, 1.0, 3.0)	.0432	.0284	.0329	.0314
(90, 100, 1.0, 5.0)	.0509	.0089	.0114	.0229
(90, 100, 1.0, 10.0)	.0689	.0101	.0159	.0355
(10, 100, 2.0, 3.0)	.0227	.1242	.0230	.0199
(30, 100, 5.0, 10.0)	.0323	.0855	.0322	.0406
(50, 100, .1, .5)	.000165	.000116	.000149	.000140
(70, 100, 2.0, 10.0)	.0211	.0106	.0137	.0181
(90, 100, .1, 3.0)	.0193	.0003	.0023	.0060

From Table 2, it is observed that

i) The trimmed means estimators are superior when θ_1 is much smaller than θ_2.

ii) When θ_1 is near θ_2, only the trimmed means estimators appear to perform less well than the others, which appear to perform equally well.

iii) Again, we see that the larger the value of m/n, the better the performance of $(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$, relative to $(\hat{\theta}_1^{\text{mom}}, \hat{\theta}_2^{\text{mom}})$.

iv) The estimators $(\hat{\theta}_1^{\text{IID}}, \hat{\theta}_2^{\text{IID}})$ are inferior to $(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$ when θ_2 is much larger than θ_1. They are also much more difficult to compute than the other three sets of estimators and are not recommended.

Overall, in non-outlier models, the pgf-based estimators $(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$ are recommended. Table 1 should allow for a wise choice of α since one usually has some idea about the values of θ_1 and θ_2 in applications. If θ_1 is near θ_2, the three sets of estimators $(\hat{\theta}_1^{\text{mom}}, \hat{\theta}_2^{\text{mom}})$, $(\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha))$ and $(\hat{\theta}_1^{\text{IID}}, \hat{\theta}_2^{\text{IID}})$ appear equally efficient. Since the optimal value of α in this case is near 1.0, the method of moments and pgf-based estimators are virtually the same, by Theorem 2.1. In this special case, the method of moments estimators are recommended. In m-outlier models with p near 0 and θ_2 much larger than θ_1, it appears that the trimmed mean estimators of Gather and Kale (1988) cannot be improved upon. This seems reasonable since we virtually have "complete separation" of two random samples and the trimmed means are asymptotically efficient sample means.

Appendix

A.1 *Asymptotic variances of* $\sqrt{n}(\hat{\theta}_i^{\text{mom}} - \theta_i)$

Let

$$E(W_1) = (1-p)\theta_1 + p\theta_2,$$
$$E(W_2) = (1-p)(\theta_1 + \theta_1^2) + p(\theta_2 + \theta_2^2),$$

$$\Sigma_{\text{mom}} = \begin{bmatrix} (1-p)\theta_1 + p(\theta_2) & (1-p)(\theta_1 + 2\theta_1^2) + p(\theta_2 + 2\theta_2^2) \\ (1-p)(\theta_1 + 2\theta_1^2) + p(\theta_2 + 2\theta_2^2) & (1-p)V_{\theta_1}(X^2) + pV_{\theta_2}(X^2) \end{bmatrix},$$

where

$$V_{\theta_i}(X^2) = 4\theta_i^3 + 6\theta_i^2 + \theta_i, \qquad i = 1, 2,$$

and let

$$G_{\text{mom}} = \begin{bmatrix} \partial\hat{\theta}_1^{\text{mom}}/\partial W_1 & \partial\hat{\theta}_1^{\text{mom}}/\partial W_2 \\ \partial\hat{\theta}_2^{\text{mom}}/\partial W_1 & \partial\hat{\theta}_2^{\text{mom}}/\partial W_2 \end{bmatrix}$$

evaluated at $W_i = E(W_i)$, $i = 1, 2$. The partial derivatives are given by:

$$\partial\hat{\theta}_1^{\text{mom}}/\partial W_1 = 1 + \frac{\sqrt{\dfrac{p}{1-p}}(1+2W_1)}{2\sqrt{W_2 - W_1 - W_1^2}},$$

$$\partial\hat{\theta}_1^{\text{mom}}/\partial W_2 = \frac{-\sqrt{\dfrac{p}{1-p}}}{2\sqrt{W_2 - W_1 - W_1^2}},$$

$$\partial\hat{\theta}_2^{\text{mom}}/\partial W_1 = 1 - \frac{\sqrt{\dfrac{1-p}{p}}(1+2W_1)}{2\sqrt{W_2 - W_1 - W_1^2}} \qquad \text{and}$$

$$\partial\hat{\theta}_2^{\text{mom}}/\partial W_2 = \frac{\sqrt{\dfrac{1-p}{p}}}{2\sqrt{W_2 - W_1 - W_1^2}}.$$

Then the asymptotic covariance matrix of $(\sqrt{n}(\hat{\theta}_1^{\text{mom}} - \theta_1), \sqrt{n}(\hat{\theta}_2^{\text{mom}} - \theta_2))^T$ is $G_{\text{mom}}\Sigma_{\text{mom}}G_{\text{mom}}^T$.

A.2 *Asymptotic variances of* $\sqrt{n}(\hat{\theta}_i(\alpha) - \theta_i)$

Let

$$\Sigma_{\text{pgf}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

where

$$\begin{aligned}
\sigma_{11} &= (1-p)(e^{\theta_1(\alpha^2-1)} - e^{\theta_1(2\alpha-2)}) + p(e^{\theta_2(\alpha^2-1)} - e^{\theta_2(2\alpha-2)}), \\
\sigma_{12} = \sigma_{21} &= (1-p)(e^{\theta_1(2\alpha^2-\alpha-1)} - e^{\theta_1(3\alpha-3)}) + p(e^{\theta_2(2\alpha^2-\alpha-1)} - e^{\theta_2(3\alpha-3)}) \\
\sigma_{22} &= (1-p)(e^{\theta_1(4\alpha^2-4\alpha)} - e^{\theta_1(4\alpha-4)}) + p(e^{\theta_2(4\alpha^2-4\alpha)} - e^{\theta_2(4\alpha-4)}).
\end{aligned}$$

Let

$$G_{\text{pgf}} = \begin{bmatrix} \partial\hat{\theta}_1(\alpha)/\partial A_1 & \partial\hat{\theta}_1(\alpha)/\partial A_2 \\ \partial\hat{\theta}_2(\alpha)/\partial A_1 & \partial\hat{\theta}_2(\alpha)/\partial A_2 \end{bmatrix}$$

evaluated at

$$A_1 = E(A_1) = (1-p)e^{\theta_1(\alpha-1)} + pe^{\theta_2(\alpha-1)}$$

and

$$A_2 = E(A_2) = (1-p)e^{2\theta_1(\alpha-1)} + pe^{2\theta_2(\alpha-1)}.$$

There are two cases:

Case (1) $\alpha > 1$: We have

$$\frac{\partial\hat{\theta}_1(\alpha)}{\partial A_1} = \left((\alpha-1)\cdot\left(A_1 - \sqrt{\frac{p}{1-p}(A_2 - A_1^2)}\right)\right)^{-1}\cdot\left(1 + \frac{\sqrt{\frac{p}{1-p}}A_1}{\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial\hat{\theta}_1(\alpha)}{\partial A_2} = \left((\alpha-1)\cdot\left(A_1 - \sqrt{\frac{p}{1-p}(A_2 - A_1^2)}\right)\right)^{-1}\cdot\left(\frac{-\sqrt{\frac{p}{1-p}}}{2\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial\hat{\theta}_2(\alpha)}{\partial A_1} = \left((\alpha-1)\cdot\left(A_1 + \sqrt{\frac{1-p}{p}(A_2 - A_1^2)}\right)\right)^{-1}\cdot\left(1 - \frac{\sqrt{\frac{1-p}{p}}A_1}{\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial\hat{\theta}_2(\alpha)}{\partial A_2} = \left((\alpha-1)\cdot\left(A_1 + \sqrt{\frac{1-p}{p}(A_2 - A_1^2)}\right)\right)^{-1}\cdot\left(\frac{\sqrt{\frac{1-p}{p}}}{2\sqrt{A_2 - A_1^2}}\right).$$

Case (2) $1/2 < \alpha < 1$: We have

$$\frac{\partial \hat{\theta}_1(\alpha)}{\partial A_1} = \left((\alpha - 1) \cdot \left(A_1 + \sqrt{\frac{p}{1-p}(A_2 - A_1^2)}\right)\right)^{-1} \cdot \left(1 - \frac{\sqrt{\frac{p}{1-p}}A_1}{\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial \hat{\theta}_1(\alpha)}{\partial A_2} = \left((\alpha - 1) \cdot \left(A_1 + \sqrt{\frac{p}{1-p}(A_2 - A_1^2)}\right)\right)^{-1} \cdot \left(\frac{\sqrt{\frac{p}{1-p}}}{2\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial \hat{\theta}_2(\alpha)}{\partial A_1} = \left((\alpha - 1) \cdot \left(A_1 - \sqrt{\frac{1-p}{p}(A_2 - A_1^2)}\right)\right)^{-1} \cdot \left(1 + \frac{\sqrt{\frac{1-p}{p}}A_1}{\sqrt{A_2 - A_1^2}}\right),$$

$$\frac{\partial \hat{\theta}_2(\alpha)}{\partial A_2} = \left((\alpha - 1) \cdot \left(A_1 - \sqrt{\frac{1-p}{p}(A_2 - A_1^2)}\right)\right)^{-1} \cdot \left(\frac{-\sqrt{\frac{1-p}{p}}}{2\sqrt{A_2 - A_1^2}}\right).$$

The asymptotic covariance matrix of $(\sqrt{n}(\hat{\theta}_1(\alpha) - \theta_1), \sqrt{n}(\hat{\theta}_2(\alpha) - \theta_2))^T$ is $G_{\text{pgf}}\Sigma_{\text{pgf}}G_{\text{pgf}}^T$.

References

From, S. G. and Saxena, K. M. L. (1989). Estimating parameters from mixed samples using sample fractional moments, *J. Statist. Plann. Inference*, **21**, 231–244.

Gather, U. and Kale, B. K. (1988). Maximum likelihood estimation in the presence of outliers, *Comm. Statist. A—Theory Methods*, **17**, 3767–3784.

Kale, B. K. (1975). Trimmed means and the method of maximum likelihood when spurious observations are present, *Applied Statistics* (ed. R. P. Gupta), 175–185, North Holland, Amsterdam.

Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions, *J. Amer. Statist. Assoc.*, **73**, 730–738.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shaked, M. and Tran, L. T. (1982). Estimating parameters from mixed samples, *J. Amer. Statist. Assoc.*, **77**, 196–203.