ACTUARIAL RESEARCH CLEARING HOUSE
1997 VOL. 1

MINIMUM CRAMÉR-VON MISES ESTIMATORS AND THEIR INFLUENCE FUNCTION

Thierry DUCHESNE

University of Waterloo
Waterloo, ON, Canada, N2L-3G1

Jacques RIOUX[1] and Andrew LUONG

Université Laval
Ste-Foy, QC, Canada, G1K-7P4

ABSTRACT. In this paper, we consider the problem of parametric estimation of loss distributions in a very general context. We use the minimum distance method with the Cramér-von Mises statistic as our particular choice of distance. We show how to compute the influence function for the estimator. We also use the influence function to obtain more information about the variance and the robustness of the estimator. We demonstrate, with an example, how much more resistant to contamination this estimator is compared with the more classical maximum likelihood estimator.

1. INTRODUCTION

If we have a random sample $\{X_1, X_2, \ldots, X_n\}$ coming from a parametric family of distributions $\{F_{\boldsymbol{\theta}} | \boldsymbol{\theta} \epsilon \Theta\}$ and we let F_n be the usual empirical distribution function, then the Cramér-von Mises statistic is commonly known as the following expression

$$W^2(F_n, F_{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} \left[F_n(X_i) - F_{\boldsymbol{\theta}}(X_i)\right]^2 . \qquad (1.1)$$

[1]The second author wishes to thank the "Chaire en Assurance L'Industrielle-Alliance" for the financial support provided for the presentation of this paper.

In this situation, we define the minimum Cramér-von Mises estimator (MCVME) for the sample at hand as that value $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ that minimizes W^2.

Hogg and Klugman (1984) suggested this method for estimating the distribution of the amounts of loss in the context of a property or casualty insurance contract. Wolfowitz (1957) however, presented the more general idea of minimum distance (MD) estimation. The use of such MD estimators is ever increasing mostly because of their good robustness properties, see Donoho and Liu (1988a, 1988b), Beran (1977, 1978, 1984). Moreover, the MCVME is consistent and asymptotically normal under not overly restrictive assumptions on the parametric model, see Duchesne, Rioux and Luong (1997).

The structure of this paper is as follows. In section 2, we set the problem up completely and introduce some notation. We then use the influence function (IF) to derive the asymptotic results for the MCVME in section 3, and we apply those results to a few situations where we can get explicit formulas. We conclude by giving a detailed numerical example of estimation using the MCVME in section 4.

2. THE PROBLEM

The problem we consider is the standard one sample parametric estimation model. A random sample $\{X_1, X_2, \ldots, X_n\}$ coming from $F_{\boldsymbol{\theta}}(\cdot)$ is completely known. We suppose that $\boldsymbol{\theta} \epsilon \Theta$, a compact subspace of $\mathcal{R}^p$, and we think of $\boldsymbol{\theta}^0 \in \Theta$ as the true but unknown value of the parameter $\boldsymbol{\theta}$. One can then write $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_p^0)'$ and we use the MCVM method to estimate $\boldsymbol{\theta}$.

In order to study the asymptotic properties of the MCVME, we use an approach based on the influence function (IF). This device was introduced by Hampel (1973) to study the infinitesimal behavior of real-valued functionals and is defined as folows.

Definition 2.1. *The influence function $IF_{\mathbf{T},F}$ of $\mathbf{T}$ at F is defined as*

$$IF_{\mathbf{T},F}(x) = \lim_{\lambda \downarrow 0} \frac{\mathbf{T}(F_{\lambda,x}) - \mathbf{T}(F)}{\lambda} = \left. \frac{d\mathbf{T}(F_{\lambda,x})}{d\lambda} \right|_{\lambda=0}, \tag{2.1}$$

where $F_{\lambda,x}(u) = F(u) + \lambda\left[\Delta_x(u) - F(u)\right]$ and $\Delta_x(u)$ is the distribution function of a degenerate random variable taking value x with probability 1.

One of the most interesting properties of the IF is that by plotting it, or an approximation of it, you can actually see what are the influential points in the estimation procedure. This may be why it is also referred to as the *influence curve.* You can also compute the asymptotic variance of the estimator (up to a factor of $1/n$) by computing the variance of the IF evaluated at each of the sample points.

For a detailed treatment of the IF, one should refer to Hampel et al (1986) or Staudte and Sheather, (1990). Duchesne, Rioux and Luong (1997) expose the principal results used to obtain the results we present here.

For any differentiable function $g : \mathcal{R}^p \longrightarrow \mathcal{R}^q$, we write its derivative with respect to the vector $\boldsymbol{\theta}'$ as

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left[\frac{\partial g_i(\boldsymbol{\theta})}{\partial \theta_j}\right]_{q\times p}, \tag{2.2}$$

and we write the product of a vector $\mathbf{Z}$ with its transpose, $\mathbf{Z}'$, as

$$\mathbf{Z}^{\otimes 2} = \mathbf{Z}\mathbf{Z}'. \tag{2.3}$$

3. ASYMPTOTIC PROPERTIES OF THE MCVME

The MCVME exhibits most of the asymptotic properties desired in an estimator. It is consistent, asymptotically normal and in many cases, it is robust in the sense of having a bounded IF. Although proofs of the consistency of general classes of estimators including the MD estimators and the MCVME do exist, see Amemiya (1985), a much simpler proof for the consistency of the MCVME is presented in Duchesne, Rioux and Luong (1997). To derive its other asymptotic properties, we need to compute its IF, hence express the estimator as a functional evaluated at F_n, the empirical cumulative distribution function of the sample.

3.1. Functional formulation of the MCVME. As pointed out in the introduction, the estimator $\hat{\boldsymbol{\theta}}_n$ we are looking for is that value of $\boldsymbol{\theta}$ that minimizes $W^2(F_n, F_{\boldsymbol{\theta}})$ as defined in (1.1). If $W^2(F_n, F_{\boldsymbol{\theta}})$ attains its minimum at an interior point of Θ, and $F_{\boldsymbol{\theta}}$ is differentiable with respect to θ_j, $\forall j = 1, \ldots, p$, then $\hat{\boldsymbol{\theta}}_n$ is also a solution of the following p-dimensional system of equations:

$$\frac{\partial}{\partial \theta_j}\left\{\sum_{i=1}^{n}\left(F_n(X_i) - F_{\boldsymbol{\theta}}(X_i)\right)^2\right\} = 0 \quad j = 1, 2, \ldots, p \tag{3.1}$$

which we can rewrite as

$$\int (F_n - F_{\boldsymbol{\theta}})\frac{\partial F_{\boldsymbol{\theta}}}{\partial \theta_j} dF_n = 0 \quad j = 1, 2, \ldots, p. \tag{3.2}$$

Therefore, if we define a p-dimensional functional $\mathbf{T}$ implicitly so that for any distribution G, $\mathbf{T}(G)$ satisfies the system

$$\int (G - F_{\mathbf{T}(G)})\frac{\partial F_{\mathbf{T}(G)}}{\partial \theta_j} dG = 0 \quad j = 1, 2, \ldots, p; \tag{3.3}$$

then $\hat{\boldsymbol{\theta}}_n$, the root of equations (3.2), can be represented implicitly as $\mathbf{T}(F_n)$.

Since the functional $\mathbf{T}$ that concerns us in this paper is defined implicitly, the derivative on the right hand side of (2.1) must also be computed implicitly. In order to do this, we consider $\mathbf{H}(\boldsymbol{\theta}, \lambda) = (H_i(\boldsymbol{\theta}, \lambda), H_2(\boldsymbol{\theta}, \lambda), \ldots, H_p(\boldsymbol{\theta}, \lambda))'$ where $H_j(\boldsymbol{\theta}, \lambda)$ is defined as

$$H_j(\boldsymbol{\theta}, \lambda) = \int (F_{\lambda,x} - F_{\boldsymbol{\theta}^0})\frac{\partial F_{\boldsymbol{\theta}}}{\partial \theta_j} dF_{\lambda,x}. \tag{3.4}$$

From this and (3.3), we get the system $\mathbf{H}(\mathbf{T}(F_{\lambda,x}), \lambda) = \mathbf{0}$. By implicit differentiation, we find that the influence function of $\mathbf{T}$ at F is

$$IF_{\mathbf{T},F}(x) = -\left.\left\{\left[\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}'}\right]^{-1}\frac{\partial \mathbf{H}}{\partial \lambda}\right\}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \lambda=0}}. \tag{3.5}$$

Substituting from (3.4), it is fairly straightforward to apply the results above to obtain

$$\left.\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}'}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \lambda=0}} = -\int \left(\left.\frac{\partial F_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right)^{\otimes 2} \, dF_{\boldsymbol{\theta}^0} \tag{3.6}$$

and

$$\left.\frac{\partial \mathbf{H}}{\partial \lambda}\right|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0\\ \lambda=0}} = \int \left.\frac{\partial F_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} (\Delta_x - F_{\boldsymbol{\theta}^0}) \, dF_{\boldsymbol{\theta}^0}. \tag{3.7}$$

We cannot tell immediately if the IF of the MCVME is bounded for any distribution function $F_{\boldsymbol{\theta}}$. However, we see that (3.6) does not depend on x and that for $IF_{\mathbf{T},F}(x)$ to be bounded, it is sufficient that $\left.\frac{\partial \mathbf{H}}{\partial \lambda}\right|_{\lambda=0}$ should also be bounded. We can then state the following proposition that the reader can easily verify.

Proposition 3.1. *A sufficient condition under which $IF_{\mathbf{T},F}(x)$ is bounded is that each component of* $\mathrm{E}\left[\left\| \left(\left.\frac{\partial F_{\boldsymbol{\theta}}(X)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right)\right\|\right]$ *is finite*

The integrals involved in (3.6) and (3.7) will not lead to explicit results in most cases. Fortunately, it is easy to get a consistent estimate of the IF by computing

$$\widehat{\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}'}} = -\frac{1}{n}\sum_{i=1}^{n} \left(\left.\frac{\partial F_{\boldsymbol{\theta}}(X_i)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} \right)^{\otimes 2} \tag{3.8}$$

instead of (3.6) and

$$\begin{aligned}\widehat{\frac{\partial \mathbf{H}}{\partial \lambda}} = \frac{1}{n}\sum_{X_i \geq x} \left.\frac{\partial F_{\boldsymbol{\theta}}(X_i)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} \left(1 - F_{\hat{\boldsymbol{\theta}}_n}(X_i)\right) \\ - \frac{1}{n}\sum_{X_i < x} \left.\frac{\partial F_{\boldsymbol{\theta}}(X_i)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} \left(F_{\hat{\boldsymbol{\theta}}_n}(X_i)\right)\end{aligned} \tag{3.9}$$

instead of (3.7). We then let $\widehat{IF} = \left\{ \left[\widehat{\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}'}}\right]^{-1} \left[\widehat{\frac{\partial \mathbf{H}}{\partial \lambda}}\right] \right\}$.

3.2. Asymptotic variance. To find the asymptotic variance matrix of the MCVME, all one has to do is compute the variance of the IF. Since the expected value of the IF of the MCVME is zero (see Duchesne, Rioux and Luong (1997), theorem 2.3), we have that

$$
\begin{aligned}
n\text{Var}[\hat{\boldsymbol{\theta}}_n] &\to \text{Var}\left[IF_{\mathbf{T},F_{\boldsymbol{\theta}}}(X)\right] \\
&= \left\{ \left[\frac{\partial H}{\partial \boldsymbol{\theta}'}\right]^{-1} \text{Var}\left[\frac{\partial H}{\partial \lambda}\right] \left(\left[\frac{\partial H}{\partial \boldsymbol{\theta}'}\right]^{-1} \right)' \right\} \Bigg|_{\substack{\boldsymbol{\theta}=\boldsymbol{\theta}^0 \\ \lambda=0}} .
\end{aligned} \tag{3.10}
$$

Once again, in many situations, the integrals involved in equation (3.10) will not lead to explicit results. However, we can derive an estimate of the asymptotic variance matrix of the MCVME using the estimate of the IF described by equations (3.8) and (3.9), and the formula of the usual sample variance. The result thereby obtained is

$$
\widehat{\text{Var}}[\hat{\boldsymbol{\theta}}_n] = \frac{1}{n}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(\widehat{IF}(x_i) - \overline{IF}\right)^{\otimes 2}\right], \tag{3.11}
$$

where

$$
\overline{IF} = (1/n)\sum_{i=1}^{n}\widehat{IF}(x_i). \tag{3.12}
$$

We now apply the results obtained in this section to a series of particular models in order to get more detailed results.

Example 3.1 (One dimensional parameter). When $\theta \in \mathcal{R}$, the influence function of the MCVME is given by

$$
IF_{\mathbf{T},F}(x) = \frac{\int \left(\frac{\partial F_{\theta}}{\partial \theta}\Big|_{\theta=\theta^0} \right) \left[\Delta_x - F_{\boldsymbol{\theta}^0}\right] dF_{\boldsymbol{\theta}^0}}{\int \left(\frac{\partial F_{\theta}}{\partial \theta}\Big|_{\theta=\theta^0} \right)^2 dF_{\boldsymbol{\theta}^0}}. \tag{3.13}
$$

Example 3.2 (A case with 2 parameters ($\boldsymbol{\theta} \in \mathcal{R}^2$)). Rewriting equation (3.5) for the case where $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0)'$, we get the following form for $IC_{\mathbf{T},F}(x)$:

$$
\begin{bmatrix}
\int \left(\frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_1}\right)^2 dF_{\boldsymbol{\theta}^0} & \int \frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_1}\frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_2}\, dF_{\boldsymbol{\theta}^0} \\
\int \frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_1}\frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_2} dF_{\boldsymbol{\theta}^0} & \int \left(\frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_2}\right)^2 dF_{\boldsymbol{\theta}^0}
\end{bmatrix}^{-1}
\begin{bmatrix}
\int \frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_1}(\Delta_x - F_{\boldsymbol{\theta}^0}) dF_{\boldsymbol{\theta}^0} \\
\int \frac{\partial F_{\boldsymbol{\theta}^0}}{\partial \theta_2}(\Delta_x - F_{\boldsymbol{\theta}^0}) dF_{\boldsymbol{\theta}^0}
\end{bmatrix}. \tag{3.14}
$$

Example 3.3 (Location and scale model). Location and scale models are characterized by their cumulative distribution function that looks like

$$F_{\boldsymbol{\theta}^0}(x) = F\left(\frac{x-\theta_1^0}{\theta_2^0}\right), \tag{3.15}$$

where θ_1^0 is the true location parameter and θ_2^0 the true scale parameter. When we apply (3.14) above, we get

$$IF_{\mathbf{T},F}(x) = \begin{bmatrix} \int_{-\infty}^{\infty} \frac{1}{(\theta_2^0)^3} f^3\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\, du & \int_{-\infty}^{\infty} \frac{(u-\theta_1^0)}{(\theta_2^0)^4} f^3\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\, du \\ \int_{-\infty}^{\infty} \frac{(u-\theta_1^0)}{(\theta_2^0)^4} f^3\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\, du & \int_{-\infty}^{\infty} \frac{(u-\theta_1^0)^2}{(\theta_2^0)^5} f^3\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\, du \end{bmatrix}$$
$$\times \begin{bmatrix} \int_{-\infty}^{\infty} \frac{1}{(\theta_2^0)^2} f^2\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\left(F\left(\frac{u-\theta_1^0}{\theta_2^0}\right) - \Delta_x(u)\right)\, du \\ \int_{-\infty}^{\infty} \frac{(u-\theta_1^0)}{(\theta_2^0)^3} f^2\left(\frac{u-\theta_1^0}{\theta_2^0}\right)\left(F\left(\frac{u-\theta_1^0}{\theta_2^0}\right) - \Delta_x(u)\right)\, du \end{bmatrix}. \tag{3.16}$$

Example 3.4 (Exponential distribution). We consider the following parameterization for the exponential distribution:

$$F_\theta(u) = 1 - e^{-\theta u}, \quad \theta > 0,\ u > 0. \tag{3.17}$$

From (3.13), we obtain the influence function as

$$IF_{T,F}(x) = \frac{27\theta^2 x e^{-2\theta x}}{4} + \frac{27\theta e^{-2\theta x}}{8} - \frac{15\theta}{8} \tag{3.18}$$

and we see immediately that it is bounded for a given value of θ. We also get an explicit expression for the asymptotic variance of the MCVME by applying (3.10) :

$$n\mathrm{Var}(\hat{\theta}_n) \to \mathrm{Var}[IF_{\mathbf{T},F}(x)] = \frac{657}{500}\theta^2. \tag{3.19}$$

In this case, the Rao-Cramér bound is θ^2/n, so we have a respectable relative efficiency of $500/657 \approx 76.1\%$ for $\hat{\theta}_n$.

Example 3.5 (Pareto distribution). The Pareto distribution here is of the following form:

$$F_{\boldsymbol{\theta}}(u) = 1 - \left(\frac{\lambda}{\lambda+u}\right)^\alpha, \quad \boldsymbol{\theta} = (\alpha, \lambda), \quad \alpha > 0,\ \lambda > 0,\ u > 0. \tag{3.20}$$

We only consider the case where $\alpha = \alpha_0$ is a known value. We can again get an exact expression for the influence function from (3.13):

$$IF_{T,F}(x) = \frac{-3(3\alpha_0+1)(3\alpha_0+2)\lambda^{2\alpha_0+1}x}{2(2\alpha_0+1)(\lambda+x)^{2\alpha_0+1}} - \frac{3(3\alpha_0+1)(3\alpha_0+2)\lambda^{2\alpha_0+1}}{4\alpha_0(2\alpha_0+1)(\lambda+x)^{2\alpha_0}} + \frac{(15\alpha_0^2+13\alpha_0+2)\lambda}{4\alpha_0(2\alpha_0+1)}. \quad (3.21)$$

Again, this is bounded for fixed values of α_0 and λ and the asymptotic variance of $\hat{\lambda}_n$ we get from (3.10) is

$$n\mathrm{Var}[\hat{\lambda}_n] \to \lambda^2 \left[\frac{(3\alpha_0+2)^2(73\alpha_0^2+27\alpha_0+2)}{20\alpha_0^2(5\alpha_0+1)(5\alpha_0+2)}\right]. \quad (3.22)$$

Example 3.6 (Gamma distribution). We consider the case where the sample comes from a Gamma distribution with $\alpha = \alpha_0$ a known integer. The cumulative distribution function for this model is

$$F_\lambda(u) = 1 - e^{-\lambda u}\sum_{i=0}^{\alpha_0-1}\frac{(\lambda u)^i}{i!}, \quad u>0,\ \lambda>0,\ \alpha_0\in\mathcal{N}. \quad (3.23)$$

Using equation (3.13), we get

$$IF_{T,F}(x) = \left(\frac{27}{4}\right)^{\alpha_0}\frac{\lambda\Gamma(\alpha_0)\Gamma(2\alpha_0)}{\Gamma(3\alpha_0)}\left[e^{-2\lambda x}\sum_{i=0}^{2\alpha_0-1}\frac{(2\lambda x)^i}{i!} - 1\right] + \frac{3^{\alpha_0}\lambda\Gamma(\alpha_0)}{\Gamma(3\alpha_0)}\sum_{i=0}^{\alpha_0-1}\frac{\Gamma(2\alpha_0+i)}{i!\,3^i}. \quad (3.24)$$

As we can see, when α_0 and λ are positive, the influence function given by (3.24) is bounded for all values of x; the MCVME is then robust for this model.

We could not find an explicit general formula to express the asymptotic variance of $\hat{\lambda}_n$. However, were able to obtain the variance exactly from the influence function by fixing α_0 to the successive values $\alpha_0 = 1, 2, 3, 4, \ldots, 10$. We used the symbolic software package Maple V to complete the computations, they are presented in the form of relative efficiencies in Table 1 below.

Table 1

Relative efficiency for the MCVME of λ in the Gamma model.

α_0	1	2	3	4	5	6	7	8	9	10
Efficiency (%)	76.1	83.1	85.8	87.1	87.9	88.5	88.9	89.2	89.4	89.6

4. A NUMERICAL EXAMPLE

We now want to apply our results to a widely used distribution for which we have yet to find an explicit formula for the IF of the MCVME, namely, the log-normal distribution. We will show, using the condition given by proposition (3.1), that the IF of the MCVME is bounded in that case, and we will estimate the IF using formulas (3.8) and (3.9). We will also compare the behavior of the MCVME to that of the method of moments estimator and the maximum likelihood estimator.

4.1. The log-normal distribution. We parameterize the log-normal distribution in the following way:

$$F_{\boldsymbol{\theta}}(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0, \ -\infty < \mu < \infty, \ \sigma > 0, \tag{4.1}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent respectively the standard normal c.d.f. and p.d.f.

The derivatives of $F_{\boldsymbol{\theta}}$ with respect to μ and σ are

$$\frac{\partial F_{\boldsymbol{\theta}}(x)}{\partial \mu} = -\frac{1}{\sigma}\phi\left(\frac{\log x - \mu}{\sigma}\right) \tag{4.2}$$

and

$$\frac{\partial F_{\boldsymbol{\theta}}(x)}{\partial \sigma} = -\frac{(\log x - \mu)}{\sigma}\phi\left(\frac{\log x - \mu}{\sigma}\right) \tag{4.3}$$

Proposition (3.1) says that in order for the MCVME to be robust, the expected value of the absolute value of those last two derivatives must be finite:

$$\begin{aligned} \mathrm{E}\left[\,\left|\frac{\partial F_{\boldsymbol{\theta}}(X)}{\partial \mu}\right|\,\right] &= \frac{1}{\sigma}\,\mathrm{E}\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\log X - \mu}{\sigma}\right)^2}\right] \\ &\leq \frac{1}{\sqrt{2\pi}\sigma^2}\,\mathrm{E}[1] < +\infty \end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[\left| \frac{\partial F_{\theta}(X)}{\partial \sigma} \right| \right] &= \frac{1}{\sigma^2}\,\mathrm{E}\left[\frac{|\log X - \mu|}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\log X - \mu}{\sigma}\right)^2} \right] \\
&\le \frac{1}{\sqrt{2\pi}\sigma^3}\,\mathrm{E}[\,|\log X - \mu|\,] \\
&= \frac{1}{\sqrt{2\pi}\sigma^3}\Big\{ \mathrm{E}\left[(\mu - \log X) I_{\{X \le e^{\mu}\}}\right] \\
&\quad + \mathrm{E}\left[(\log X - \mu) I_{\{X > e^{\mu}\}}\right] \Big\},
\end{aligned}$$

where I_A is the indicator variable of event A. One can solve the integrals for the last two expectations easily, and see that these quantities are finite. Therefore, the IF of the MCVME in the log-normal case is bounded and, hence, the MCVME is a robust estimator in that case.

4.2. The example. We now apply different methods of estimation to a sample coming from a log-normal distribution with parameters $\mu = 7$ and $\sigma = 0.2$. With that choice of parameters, the theoretical mean, standard deviation and 80th percentile are approximately 1119, 226 and 1298. The data were generated with S-Plus, and here is a summary of the sample obtained:

TABLE 1. Summary of the random sample.

Size (n)	Min	1st Quart.	Median	Mean	3rd Quart.	Max
200	575.2	956	1078	1105	1253	1826

Using different methods of estimation, we get different estimates for μ and σ. The results are summarized in table 2.

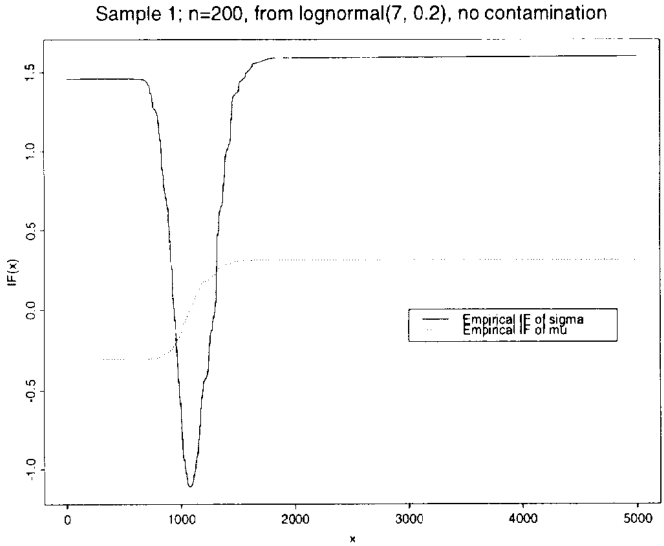
TABLE 2. Estimates of μ and σ (to 3 significant digits).

Estimation method	$\hat{\mu}$	$\hat{\sigma}$	$\mathrm{E}_{\hat{\theta}}(X)$	$\sqrt{\mathrm{Var}_{\hat{\theta}}(X)}$	$\hat{p}_{0.8}$
Moments	6.99	0.205	1100	229	1298
Max. likelihood	6.99	0.208	1110	232	1293
MCVM	6.99	0.204	1100	228	1289

As we can see from table 2, all three methods seem to give similar results. The estimates are very close to the true values, and that was expectable considering the fact that all three estimators are consistent and the sample size is large.

However, our principal interest is to investigate the robustness of the estimators. In order to do so, we append an unusually large outlier to the sample, x_{201}=100,000. If a method is more robust than the others, then the values of the new estimates (including the outlier in the sample) should not differ much from the values of table 2. In the case of the MCVME, we have begun this section by showing that its IF was bounded. Further, we can estimate that IF using the sample and formulas (3.8) and (3.9). The estimate of the IF is plotted in figure 1.

FIGURE 1. Plot of the estimated IF for the log-normal distribution.



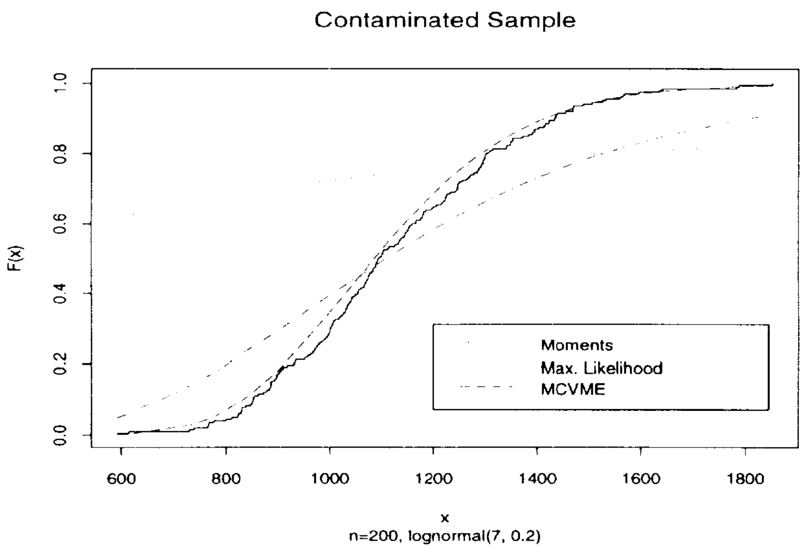
It is fairly obvious from figure 1 that the IF is bounded for the MCVME's of both μ and σ. This leads us to expect that the outlier should not have a significant influence on the MCVME's.

TABLE 3. Estimates of μ and σ (to 3 significant digits), contaminated sample.

Estimation method	$\hat{\mu}$	$\hat{\sigma}$	$E_{\hat{\theta}}(X)$	$\sqrt{\mathrm{Var}_{\hat{\theta}}(X)}$	$\hat{p}_{0.8}$
Moments	5.88	1.731	1600	6970	1536
Max. Likelihood	7.01	0.380	1190	469	1525
MCVM	6.99	0.206	1110	230	1291

Comparing the entries of tables 2 and 3, we see that the outlier had a huge effect on the estimates for both the method of moments and the maximum likelihood (particularly on the estimates of the scale parameter, σ). On the other hand, the influence of the new value on the MCVME was nearly negligible. This is the great value of a robust estimator like the MCVME. Since a picture is still worth a thousand words, we'll cut this paper short by presenting figure 2 that compares the estimated cumulative distribution function for three estimation methods. We believe that it illustrates our point in an extremely convincing manner.

FIGURE 2. CDF's for estimated distributions under contamination



Bibliography

1. Amemiya, T. (1985), "Advanced Econometrics", Harvard University Press Actex, Cambridge, 521 pages.
2. Beran, R. (1977), "Minimum Hellinger Distance Estimates for Parametric Models", *The Annals of Statistics*, vol. 5, pp. 445-463.
3. Beran, R. (1978), "An Efficient and Robust Adaptive Estimator of Location", *The Annals of Statistics*, vol. 6, pp. 292-313.
4. Beran, R. (1984), "Minimum Distance Procedures", *Handbook of Statistics*, vol. 4, pp. 741-754.
5. Char, B.W., Geddes, K.O., Gonnet, G.H., Leong, B.L., Monagan, M.B., Watt, S.M., *Maple V Language Reference Manual*, Springer-Verlag, New York.
6. Donoho, C.L., Liu, R.C. (1988a), "The "Automatic" Robustness of Minimum Distance Functionals", *The Annals of Statistics*, vol. 16, pp. 552-586.
7. Donoho, C.L., Liu, R.C. (1988b), "Pathologies of Some Minimum Distance Estimators", *The Annals of Statistics*, vol. 16, pp. 587-608.
8. Duchesne, T., Rioux, J., Luong, A. (1997), "Minimum Cramér-von Mises Distance Methods for Complete and Grouped Data", *Communications in Statistics, Theory and Methods*, vol. 26, Issue 2.
9. Hampel, F.R. (1974), "The influence curve and its role in Robust estimation", *Journal of the American Statistical Association*, 69, pp. 383-393
10. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986), *Robust Statistics (The Approach Based on Influence Functions)*, Wiley Series in Probability and Statistics, New-York.
11. Hogg, R.V., Klugman, S.A. (1984), *Loss Distributions*, Wiley Series in Probability and Statistics, New-York.
12. Luong, A., Thompson, M.E.(1987), "Minimum-distance methods based on quadratic distances for transforms", *The Canadian Journal of Statistics*, Vol. 15, No. 3, pp. 239-251.
13. Parr, W.C., Schucany, W.R. (1980), "Minimum Distance and Robust Estimation", *The Journal of the American Statistical Association*, vol. 75, pp. 616-624.
14. Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics, New-York.
15. Staudte, R. G. , Sheather, S. J. (1990), *Robust Estimation and Testing, Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons.
16. Wolfowitz, J. (1957), "The Minimum Distance Method", *Annals of Mathematical Statistics*, vol. 28, pp. 75-88.