Semiparametric efficiency bounds for generalized skew-elliptical distributions

YANYUAN MA * MARC G. GENTON & ANASTASIOS A. TSIATIS †

August 20, 2003

SUMMARY

We consider a class of generalized skew-elliptical distributions which is useful for selection modeling and robustness analysis and derive a class of semiparametric estimators for the location and scale parameters of the central part of the model. These estimators are shown to be consistent and asymptotically normal. We present the semiparametric efficiency bound and derive the locally efficient estimator that achieves this bound if the model for the skewing function is correctly specified. The estimators we propose are consistent and asymptotically normal even if the model for the skewing function is misspecified and we compute the loss of efficiency in such cases.

Some key words: Generalized skew-elliptical distributions; Influence function; Nuisance tangent space; Semiparametric efficiency.

1 INTRODUCTION

Consider the model where a p-dimensional random vector X is distributed with density $g(x;\beta)$, where β is a q-dimensional vector of unknown parameters. In order to make inference about β, the usual statistical analysis assumes that a random sample $X_1,\ldots,X_n$ from $g(x;\beta)$ can be observed. However, there are many situations where such a random sample might

*Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8205, USA, yma@unity.ncsu.edu. The work of Yanyuan Ma is supported by grant NIGMS 1 R01 Gm67299-01.

†Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA. genton@stat.ncsu.edu tsiatis@stat.ncsu.edu

not be available, for instance if it is too difficult or too costly to obtain. If the probability density function is distorted by some multiplicative nonnegative weight function $w(x;\beta,\alpha)$, where α denotes some r-dimensional vector of additional unknown parameters, then the observed data is a random sample from a distribution with density

$$f(x;\beta,\alpha) = g(x;\beta)\frac{w(x;\beta,\alpha)}{\mathrm{E}\left\{w(X;\beta,\alpha)\right\}}, \tag{1}$$

where f is said to be the probability density function of a weighted distribution, see Rao (1985) and references therein. In particular, if the observed data are obtained only from a selected portion of the population of interest, then (1) is called a selection model. For example, this can happen if the observation vector X of characteristics of a certain population is measured only for individuals who manifest a certain disease due to cost or ethical reasons, see the survey article by Bayarri and DeGroot (1992) and references therein. For such problems, the goal is to find consistent and asymptotically normal estimators of β in the presence of the nuisance weight function w.

A slightly different point of view is given by a robustness argument. Effectively, if $g(x;\beta)$ is the central model of interest, then the weight function w in (1) can be seen as a contaminating function. For instance, if g is an elliptical probability density function, then w generates asymmetric outliers in the observed sample from f. The goal is then to derive robust estimators of β, that is again to provide consistent and asymptotically normal estimators of β in the presence of a certain class of the nuisance weight function w.

The paper is organized as follows. In Section 2, we describe a class of generalized skew-elliptical distributions which is useful for selection modeling and robustness analysis. We present our main results in Section 3 for a univariate location-scale normal central model. In particular, we derive semiparametric location and scale estimators that are consistent and asymptotically normal regardless of the possible misspecification of the weight function. In addition, we will show that estimators within this class achieve the semiparametric efficiency bound.

2 GENERALIZED SKEW-ELLIPTICAL DISTRIBUTIONS

Generalized skew-elliptical (GSE) distributions have been introduced by Genton and Loperfido (2002). The density of a random vector with a (GSE) distribution is defined through an elliptical density and a skewing function as follows.

Definition 1 *A p-dimensional generalized skew-elliptical (GSE) distribution is a distribution whose probability density function is of the form*

$$f(x) = 2|\Sigma|^{-1/2} g(\Sigma^{-1/2}x^T - \Sigma^{-1/2}\xi^T)\pi(\Sigma^{-1/2}x^T - \Sigma^{-1/2}\xi^T), \ x \in \mathbb{R}^p, \tag{2}$$

where g is the probability density function of a spherical distribution, ξ is the location parameter, $\Sigma^{-1/2}$ is the Cholesky decomposition of the covariance matrix Σ, and the function $\pi : \mathbb{R}^p \to \mathbb{R}$ satisfies $0 \leq \pi(x) \leq 1$ and $\pi(x) + \pi(-x) = 1$. We refer to π as the skewing function.

Note that the location vector ξ and the matrix Σ are not, in general, the expected value and the covariance matrix for f, since GSE distributions may not be symmetric with respect to ξ, but they are for g. In particular, if $g = \phi_p$, the probability density function of the standard multivariate normal distribution, and we choose a parametric model $\pi(x) = \Phi(\alpha x^T)$ for the skewing function, where Φ is the univariate standard normal cumulative distribution function, then (2) is the probability density function of the multivariate skew-normal distribution (Azzalini and Dalla Valle, 1996).

From Definition 1, it is clear that the GSE distributions arise in inference from non-random samples (Copas and Li, 1997) and are therefore selection models of the form in Equation (1). Representation of a GSE distribution as a selection model is straightforward with $g(x;\beta) = |\Sigma|^{-1/2}g(\Sigma^{-1/2}x^T - \Sigma^{-1/2}\xi^T)$, $w(x;\beta,\alpha) = \pi(\Sigma^{-1/2}x^T - \Sigma^{-1/2}\xi^T)$, $\mathrm{E}\{w(X;\beta,\alpha)\} = 1/2$, $\beta = \{\xi, \mathrm{vec}(\Sigma)\}$, and α is embedded in the skewing function π. A weight function with such property can naturally occur when the selection criterion is that a certain component of the measurement is larger than its expected value given the other measurement components, see Arnold and Beaver (2002). One simple example is the distribution of height and weight. Assume they follow a bivariate normal distribution in a general population. Yet in a clinic treating obesity, one would expect that all the samples obtained are the ones whose weight is larger than the expected weight given their height. If we try to estimate the true distribution of the weight in the general population using such data, we are in the presence of a skew-normal distribution $2\phi_1(x;\xi,\sigma)\Phi(\alpha x - \alpha\xi)$ that is a selection model with a weight function $\Phi(\alpha x - \alpha\xi)$ which has the property $\Phi(\alpha x - \alpha\xi) + \Phi(-\alpha x + \alpha\xi) = 1$.

Another application of GSE distributions arises when modeling case-control data in prospective studies (Weinberg and Sandler, 1991; Weinberg and Wacholder, 1993; Wacholder and Weinberg, 1994; Zhang, 2000). Consider a random sample $X_1, \ldots, X_n$ from an elliptical distribution with density g. Let $d_i \in \{0,1\}$, $i = 1, \ldots, n$ be the observed value of a dichotomous random variable D_i associated with the i-th observation, and $P(D_i = 0 | X_i = x_i) = \pi(x_i)$. Prospective studies focus on the conditional distribution of X_i given $D_i = d_i$. From Bayesian inversion we get

$$f(x_i | d_i = 0) = \frac{g(x_i)\,\pi(x_i)}{E\{\pi(X)\}}, \quad f(x_i | d_i = 1) = \frac{g(x_i)\{1 - \pi(x_i)\}}{1 - E\{\pi(X)\}}. \tag{3}$$

If we impose $\pi(x_i) + \pi(-x_i) = 1$, then it follows that $f(x_i | d_i)$ is GSE. This condition on π naturally arises in models such as the logistic regression model where $\pi(x_i) = \exp(\beta x_i^T)/\{1 + \exp(\beta x_i^T)\}$, and the probit regression model where $\pi(x_i) = \Phi(\beta x_i^T)$, when the intercept is assumed to be equal to zero.

3 MAIN RESULTS

As described in the previous section, we are interested in inference on the parameters ξ and Σ in (2), which represent the mean and the covariance matrix of the population of which only samples from a particular subpopulation are available. We make no additional assumptions regarding the skewing function other than the constraints that $0 \le \pi(x) \le 1$ and $\pi(x) + \pi(-x) = 1$. Consequently, we are considering a semiparametric model and therefore use the semiparametric theory as developed by Bickel et al. (1993) and Newey (1990) where regular asymptotically linear (RAL) estimators are represented through their influence function in the Hilbert space consisting of mean zero square integrable random functions with the covariance inner product.

To remain specific and focused, in this paper, all our results are developed in the special case where $g = \phi$, the univariate standard normal probability density function, in which case we use the generalized skew-normal (GSN) distributions

$$f(x) = \frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\pi\left(\frac{x-\xi}{\sigma}\right). \tag{4}$$

The methods we use can be extended in a straightforward manner in more general cases. In the sequel, β represents the vector (ξ, σ). Notice that an arbitrary skewing function $\pi(x)$ can always be written as $H\{m(x)\}$ where H is an arbitrarily chosen symmetric cumulative distribution function and m is an odd function. In particular, an arbitrary $\pi(x)$ can be written as $\Phi\{m(x)\}$ where Φ is the univariate normal cumulative distribution function. Throughout the text, parameters or functions with index 0 refer to the true values of the parameters or the true functions.

All RAL estimators have influence functions which are orthogonal to the nuisance tangent space; that is, the linear subspace corresponding to the mean square closure of all score vectors associated with parametric submodels for the skewing function $\pi(\cdot)$. We begin by deriving the nuisance tangent space and its orthogonal complement.

Proposition 1 *The nuisance tangent space $\Gamma_\pi = \{u(x/\sigma_0 - \xi_0/\sigma_0) : \pi_0(x)u(x)$ is an odd function $\}$.*

Proof. Suppose $f(x;\beta,\alpha) = 2/\sigma\phi(x/\sigma - \xi/\sigma)\pi(x/\sigma - \xi/\sigma, \alpha)$ is a parametric submodel of the GSN model (4), then

$$\left.\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\right|_{(\beta_0,\alpha_0)} = \left.\frac{\partial\pi(x/\sigma_0 - \xi_0/\sigma_0, \alpha)}{\partial\alpha}\right|_{\alpha_0} \Big/ \pi_0\left(\frac{x-\xi_0}{\sigma_0}\right).$$

Since $\pi(x/\sigma_0 - \xi_0/\sigma_0, \alpha) + \pi(-x/\sigma_0 + \xi_0/\sigma_0, \alpha) = 1$, $\partial\pi(x/\sigma_0 - \xi_0/\sigma_0, \alpha)/\partial\alpha + \partial\pi(-x/\sigma_0 + \xi_0/\sigma_0, \alpha)/\partial\alpha = 0$, i.e.

$$\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\left.\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\right|_{(\beta_0,\alpha_0)}$$

is an odd function of $x/\sigma_0 - \xi_0/\sigma_0$. Write

$$\left.\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\right|_{(\beta_0,\alpha_0)}$$

as $u(x/\sigma_0 - \xi_0/\sigma_0)$, we obtain that $\pi_0(x)u(x)$ is an odd function. In fact, for any $r\times 1$ matrix B, where $r = \dim(\alpha)$, $\pi_0(x)u(x)B$ is an odd function. On the other hand, for any $u(x)$ such that $\pi_0(x)u(x)$ is odd, let $h(x) = \pi_0(x)u(x)/[m_0(x)\phi\{m_0(x)\}]$, where $\pi_0(x) = \Phi\{m_0(x)\}$.

Then $h(x)$ is an even function and

$$f(x;\beta,\alpha)=\frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\Phi\left\{m_0\left(\frac{x-\xi}{\sigma}\right)e^{\alpha h^T\left(\frac{x-\xi}{\sigma}\right)}\right\}$$

is a parametric submodel where $\alpha=0$ yields the true model. Notice that

$$\begin{aligned}&\frac{\partial\log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)}\\ =\;& m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\alpha h^T\left(\frac{x-\xi_0}{\sigma_0}\right)}h\left(\frac{x-\xi_0}{\sigma_0}\right)\phi\left\{m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\alpha h^T\left(\frac{x-\xi_0}{\sigma_0}\right)}\right\}\Big/\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\Big|_{\alpha=0}\\ =\;& u\left(\frac{x-\xi_0}{\sigma_0}\right).\end{aligned}$$

In the special case when $\pi_0(x)\equiv 1/2$, thus $m_0(x)\equiv 0$, we can set the parametric submodel to be

$$f(x;\beta,\alpha)=\frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\Phi\left\{\frac{\alpha}{2}u^T\left(\frac{x-\xi}{\sigma}\right)\Big/\phi(0)\right\}.$$

It can be easily verified that

$$\frac{\partial\log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)}=u\left(\frac{x-\xi_0}{\sigma_0}\right),$$

hence $u(x/\sigma_0-\xi_0/\sigma_0)\in\Gamma_\pi$. $\square$

Proposition 2 *The orthogonal complement of the nuisance tangent space is* $\Gamma_\pi^\perp=\{v(x/\sigma_0-\xi_0/\sigma_0):v(x)$ *is an even function that satisfies* $\int v(x)\phi(x)d\mu(x)=0\}$*, where* $\mu(x)$ *is the measure for which densities are defined.*

Proof. Elements in $\Gamma_\pi^\perp$ satisfy

$$\int v\left(\frac{x-\xi_0}{\sigma_0}\right)u\left(\frac{x-\xi_0}{\sigma_0}\right)\frac{2}{\sigma_0}\phi\left(\frac{x-\xi_0}{\sigma_0}\right)\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)d\mu(x)=0 \tag{5}$$

for any $u(x/\sigma_0-\xi_0/\sigma_0)\in\Gamma_\pi$ and

$$\int v\left(\frac{x-\xi_0}{\sigma_0}\right)\frac{2}{\sigma_0}\phi\left(\frac{x-\xi_0}{\sigma_0}\right)\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)d\mu(x)=0. \tag{6}$$

Because $2u(x)\phi(x)\pi_0(x)/\sigma_0$ is an arbitrary odd function, $v(x/\sigma_0-\xi_0/\sigma_0)$ has to be an even function of $x/\sigma_0-\xi_0/\sigma_0$ to ensure Equation (5). Notice that $\pi_0(x)-1/2$ is in fact an odd function, so we get $\int v(x)\phi(x)d\mu(x)=0$ from Equation (6). $\square$

Since influence functions for RAL estimators are orthogonal to the nuisance tangent space derived in Proposition 2, this motivates estimators obtained by solving the following estimating equations.

Proposition 3 *For any even function $v(x)$ s.t. $\int v(x)\phi(x)d\mu(x) = 0$, $\sum_{i=1}^{n} v(X_i/\sigma - \xi/\sigma) = 0$ is a regular asymptotically linear (RAL) estimator for $\beta = (\xi, \sigma)$.*

Proposition 3 provides us a way of constructing RAL estimators as long as we can find a suitable function v. For example, we can take any even function $h(x)$ and construct $v(x) = h(x) - \int h(x)\phi(x)d\mu(x)$. If we take h to be x^{2k}, then the corresponding v functions are $v(x) = x^2 - 1, v(x) = x^4 - 3, v(x) = x^6 - 15$, etc.

The semiparametric efficient estimator is an RAL estimator which has influence function that is proportional to the efficient score. The efficient score is the residual after projecting the score vector with respect to β onto the nuisance tangent space and is derived in the following proposition.

Proposition 4 *The efficient score function is*

$$S_{eff} = \left[\frac{x-\xi_0}{\sigma_0^2}\left\{2\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) - 1\right\} - \frac{2}{\sigma_0}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right), \frac{(x-\xi_0)^2}{\sigma_0^3} - \frac{1}{\sigma_0}\right],$$

where $\pi_{01}(x) = d\pi_0(x)/dx$.

Proof. Calculating $\partial \log f(x;\beta,\alpha)/\partial\xi$ and $\partial \log f(x;\beta,\alpha)/\partial\sigma$, evaluating at ξ_0 and σ_0 yields the score vector

$$S_\beta = \left\{\frac{x-\xi_0}{\sigma_0^2} - \frac{\pi_{01}\left(x/\sigma_0 - \xi_0/\sigma_0\right)}{\sigma_0\pi_0\left(x/\sigma_0 - \xi_0/\sigma_0\right)}, -\frac{1}{\sigma_0} + \frac{(x-\xi_0)^2}{\sigma_0^3} - \frac{(x-\xi_0)\,\pi_{01}\left(x/\sigma_0 - \xi_0/\sigma_0\right)}{\sigma^2\pi_0\left(x/\sigma_0 - \xi_0/\sigma_0\right)}\right\}.$$

We calculate the projection of S_β onto $\Gamma_\pi^\perp$. Assume the projection is $\{v_1(x/\sigma_0 - \xi_0/\sigma_0), v_2(x/\sigma_0 - \xi_0/\sigma_0)\}$ where both v_1 and v_2 are even functions, then

$$\begin{gathered}\left\{\tfrac{x-\xi_0}{\sigma_0^2} - \tfrac{1}{\sigma_0}\pi_{01}\left(\tfrac{x-\xi_0}{\sigma_0}\right)\Big/\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right) - v_1\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\}\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right) + \\ \left[\tfrac{-x+\xi_0}{\sigma_0^2} - \tfrac{1}{\sigma_0}\pi_{01}\left(\tfrac{x-\xi_0}{\sigma_0}\right)\Big/\left\{1-\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\} - v_1\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right]\left\{1-\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\} = 0\end{gathered}$$

and

$$\begin{gathered}\left\{-\tfrac{1}{\sigma_0} + \tfrac{(x-\xi_0)^2}{\sigma_0^3} - \tfrac{(x-\xi_0)}{\sigma_0^2}\pi_{01}\left(\tfrac{x-\xi_0}{\sigma_0}\right)\Big/\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right) - v_2\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\}\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right) + \\ \left[-\tfrac{1}{\sigma_0} + \tfrac{(x-\xi_0)^2}{\sigma_0^3} + \tfrac{(x-\xi_0)}{\sigma_0^2}\pi_{01}\left(\tfrac{x-\xi_0}{\sigma_0}\right)\Big/\left\{1-\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\} - v_2\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right]\left\{1-\pi_0\left(\tfrac{x-\xi_0}{\sigma_0}\right)\right\} = 0.\end{gathered}$$

Notice that we used the fact that $\pi_{01}(x)$ is an even function of x. Solving the two equations yields the result. □

Proposition 5 *A semiparametric efficient estimator of $\beta = (\xi, \sigma)$ is given by*

$$\sum_{i=1}^{n} F_0(X_i; \xi, \sigma) = 0 \tag{7}$$

where

$$F_0(X_i; \xi, \sigma) = \left(\left[\frac{X_i - \xi}{\sigma} \left\{ 2\pi_0 \left(\frac{X_i - \xi}{\sigma} \right) - 1 \right\} - 2\pi_{01} \left(\frac{X_i - \xi}{\sigma} \right) \right], \{(X_i - \xi)^2 - \sigma^2\} \right).$$

Assume the solution to Equation (7) is $\hat{\beta}$, then $n^{1/2}(\hat{\beta} - \beta_0) \to N(0, \{E(S_{eff}^T S_{eff})\}^{-1})$ in distribution. Here, the smallest variance of the estimate given by $\{E(S_{eff}^T S_{eff})\}^{-1}$ has the form

$$A = \sigma_0^2 \begin{pmatrix} \int [\{2\pi_0(x) - 1\}^2 + 4\pi_{01}(x)^2]\phi(x)d\mu(x) & 4\int \pi_{01}(x)\phi(x)d\mu(x) \\ 4\int \pi_{01}(x)\phi(x)d\mu(x) & 2 \end{pmatrix}^{-1}. \tag{8}$$

Remark 1 *Notice that when $\pi(x) \equiv 1/2$, the first component of the efficient score vector is 0, in which case an efficient semiparametric estimator does not exist. Similar phenomena have been observed in Bayesian analysis of selection models, where a constant weight function (corresponding to $\pi(x) \equiv 1/2$ in our case) have to be ruled out a priori to any analysis, see Lee and Berger (2001).*

We omit the proof of Proposition 5 which involves only straightforward algebra. The efficient estimator defined by Equation (7) depends on using the true skewing function π_0, which is unknown to us. However, any choice of skewing function in equation (7) will lead to a consistent asymptotically normal estimator for β. This can be shown by noticing that $v(x) = [x\{2\pi(x) - 1\} - 2\pi_1(x), x^2 - 1]$ satisfies the requirement in Proposition 3, where $\pi_1(x) = d\pi(x)/dx$. In practice, we generally posit a model for $\pi(\cdot)$ in terms of a finite set of parameters α say, $\pi(x/\sigma - \xi/\sigma, \alpha)$ and then estimate α using an estimator $\hat{\alpha}$. We use

$$\sum_{i=1}^{n} F(X_i; \xi, \sigma, \hat{\alpha}) = 0 \tag{9}$$

to denote estimators of the form in Equation (7) with $\pi_0(x/\sigma - \xi/\sigma)$ replaced by $\pi(x/\sigma - \xi/\sigma, \hat{\alpha})$. Notice that $E\{F(X_i; \xi, \sigma, \alpha)\} = 0$ for all values α, hence $E\{\partial F(X_i; \xi, \sigma, \alpha)/\partial\alpha\} = 0$ assuming sufficiently smooth conditions on F to interchange the expectation and the partial derivative. If the true skewing function belongs to this parametric model then $\pi(\cdot, \hat{\alpha})$ will converge to $\pi_0(\cdot)$. But even if the parametric model does not contain the true $\pi_0(\cdot)$, the estimate $\hat{\alpha}$ will generally converge to a constant α^* and $\pi(\cdot, \hat{\alpha})$ will converge to some skewing function $\pi(\cdot, \alpha^*)$. As long as $n^{1/2}(\hat{\alpha} - \alpha^*)$ is bounded in probability, we show in the next proposition that the asymptotic distribution of $\hat{\beta}$ obtained by using $\pi(\cdot, \hat{\alpha})$ is asymptotically the same as that which uses $\pi(\cdot, \alpha^*)$ which we have argued is consistent and asymptotically normal. However, if the parametric model does contain the truth, then the estimator for β in (9) will be semiparametric efficient. Such estimators are referred to as locally efficient.

Proposition 6 *Assume $2\phi(x/\sigma - \xi/\sigma)\pi(x/\sigma - \xi/\sigma, \alpha)/\sigma$ is a parametric model and $n^{1/2}(\hat{\alpha} - \alpha^*)$ is bounded in probability, then the two RAL estimators $\sum_{i=1}^n F(X_i; \xi, \sigma, \alpha^*) = 0$ and $\sum_{i=1}^n F(X_i; \xi, \sigma, \hat{\alpha}) = 0$ are asymptotically equivalent, i.e. if $(\hat{\xi}_1, \hat{\sigma}_1)$ is the solution to the first equation, and $(\hat{\xi}_2, \hat{\sigma}_2)$ is the solution to the second equation, then $n^{1/2}(\hat{\xi}_1 - \hat{\xi}_2) \to 0$ and $n^{1/2}(\hat{\sigma}_1 - \hat{\sigma}_2) \to 0$ in probability.*

Proof. Write (ξ, σ) as β, $F(X_i; \xi, \sigma, \alpha)$ as $F(X_i; \beta, \alpha)$. A Taylor expansion of $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*)$ at $\hat{\alpha}$ yields $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*) = \sum_{i=1}^n F(X_i; \hat{\beta}_2, \hat{\alpha}) + (\alpha^* - \hat{\alpha})\{\sum_{i=1}^n \partial F^T(X_i; \hat{\beta}_2, \tilde{\alpha})/\partial\alpha\}$, where $\tilde{\alpha}$ is between α^* and $\hat{\alpha}$. Denoting $\{\sum_{i=1}^n \partial F^T(X_i; \hat{\beta}_2, \tilde{\alpha})/\partial\alpha\}/n$ by Λ_n, we obtain $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*) = n(\alpha^* - \hat{\alpha})\Lambda_n^T$. Notice that when $n \to \infty$, because of the convergence of $\hat{\alpha}$ to α^* and the consistency property of ξ_2 and σ_2, $\Lambda_n \to \mathrm{E}\{\partial F^T(X_i; \beta_0, \alpha^*)/\partial\alpha\} = 0$ in probability.

A Taylor expansion of $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*)$ at $\hat{\beta}_1$ yields

$$\begin{aligned}
\hat{\beta}_2 - \hat{\beta}_1 &= \left\{\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*) - 0\right\}\left\{\sum_{i=1}^n \frac{\partial F^T(X_i; \tilde{\beta}, \alpha^*)}{\partial\beta}\right\}^{-1} \\
&= (\alpha^* - \hat{\alpha})\,\Lambda_n^T \left\{\frac{1}{n}\sum_{i=1}^n \frac{\partial F^T(X_i; \tilde{\beta}, \alpha^*)}{\partial\beta}\right\}^{-1},
\end{aligned}$$

where $\tilde{\beta}$ is a quantity between $\hat{\beta}_1$ and $\hat{\beta}_2$.

When $n \to \infty$,

$$J_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial F^T(X_i; \tilde{\beta}, \alpha^*)}{\partial\beta} \to \mathrm{E}\left\{\frac{\partial F^T(X_i; \beta_0, \alpha^*)}{\partial\beta}\right\}$$

in probability. For parametric models, $\mathrm{E}\{\partial F^T(X_i; \beta_0, \alpha^*)/\partial\beta\}$ is the matrix related to the Fisher information matrix, which is generally nonsingular and we denote it by J. Combining the results, we have $n^{1/2}(\hat\beta_1 - \hat\beta_2) = n^{1/2}(\hat\alpha - \alpha^*)\Lambda_n^T J_n^{-1}$. Because $n^{1/2}(\hat\alpha - \alpha^*)$ is bounded in probability, $J_n^{-1} \to J^{-1}$ in probability and $\Lambda_n \to 0$ in probability, this implies that $\hat\beta_1 - \hat\beta_2 \to 0$ in probability. $\square$

The efficiency of an estimator depends on how close the true π_0 is to the parametric family $\{\pi(x,\alpha)\}$. One way proposed by Ma and Genton (2002) to construct the parametric submodel is to use $\Phi(P_K(x))$ to approximate $\pi(x,\alpha)$, where $P_K(x)$ is an odd polynomial of order K. In general, the relation between the efficiency loss and the "distance" between π_0 and $\{\pi(x,\alpha)\}$ is given in the following proposition.

Proposition 7 *Let $\nu(x) = \pi(x,\alpha) - \pi_0(x)$, $\theta = \int 4[\partial\{\nu(x)\phi(x)\}/\partial x]^2/\phi(x)d\mu(x)$. The most efficient semiparametric estimator of the form in Equation (9) has efficiency $A + \min_\alpha(\theta)B$, where A is given by Equation (8), and*

$$B = \frac{\theta\sigma_0^2}{[E\{2\pi_0(X)-1+2X\pi_{01}(X)-2\pi_{02}(X)\}-2E(X)^2]^2}\left\{\begin{array}{cc} 1 & -E(X) \\ -E(X) & E(X)^2 \end{array}\right\},$$

which does not depend on the estimator. Here π_{02} denotes $d^2\pi_0(x)/dx^2$, the expectations are taken with respect to $2\phi(x)\pi_0(x)$.

Proof. Assume the estimating equation $\sum_{i=1}^n F(X_i;\beta,\hat\alpha) = 0$ yields the estimate $\hat\beta_1 = (\hat\xi_1,\hat\sigma_1)$, the estimating equation $\sum_{i=1}^n F_0(X_i;\beta) = 0$ yields the estimate $\hat\beta = (\hat\xi,\hat\sigma)$. Then

$$\begin{aligned} 0 &= \sum_{i=1}^n F(X_i;\hat\beta_1,\hat\alpha) \\ &= \sum_{i=1}^n F_0(X_i;\hat\beta) + \sum_{i=1}^n \{F_0(X_i;\hat\beta_1) - F_0(X_i;\hat\beta)\} + \sum_{i=1}^n \{F(X_i;\hat\beta_1,\hat\alpha) - F_0(X_i;\hat\beta_1)\} \\ &= \sum_{i=1}^n (\hat\beta_1 - \hat\beta)\frac{\partial F_0^T(X_i;\tilde\beta)}{\partial\beta} + \sum_{i=1}^n \left\{\frac{X_i-\hat\xi_1}{\hat\sigma_1}2\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right), 0\right\}, \end{aligned}$$

where $\tilde\beta$ is a quantity between $\hat\beta$ and $\hat\beta_1$, $\nu_1(x) = d\nu(x)/dx$. Notice that when $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial F_0^T(X_i;\tilde\beta)}{\partial\beta} \to E\left\{\frac{\partial F_0^T(X_i;\beta_0)}{\partial\beta}\right\}$$

in probability,

$$\frac{1}{n}\sum_{i=1}^n \frac{X_i-\hat\xi_1}{\hat\sigma_1}2\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) \to E\left\{2\frac{X_i-\hat\xi_1}{\hat\sigma_1}\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right)\right\}$$

$$\to E\left\{2\frac{X_i-\xi_0}{\sigma_0}\nu\left(\frac{X_i-\xi_0}{\sigma_0}\right) - 2\nu_1\left(\frac{X_i-\xi_0}{\sigma_0}\right)\right\} = 0$$

in probability due to the consistency of $\hat{\beta}_1$ and $\hat{\sigma}_1$. We calculate the variance of $2(X_i - \xi_0)\nu(X_i/\sigma_0 - \xi_0/\sigma_0)/\sigma_0 - 2\nu_1(X_i/\sigma_0 - \xi_0/\sigma_0)$, which is an even function of $X_i/\sigma_0 - \xi_0/\sigma_0$.

$$E\left[\left\{2\frac{X_i-\xi_0}{\sigma_0}\nu\left(\frac{X_i-\xi_0}{\sigma}\right) - 2\nu_1\left(\frac{X_i-\xi_0}{\sigma_0}\right)\right\}^2\right] = 4\int\{x\nu(x)-\nu_1(x)\}^2 2\phi(x)\pi_0(x)d\mu(x)$$
$$=4\int\{x\nu(x)-\nu_1(x)\}^2\phi(x)d\mu(x) = \int\frac{4}{\phi(x)}\left[\frac{\partial}{\partial x}\{\nu(x)\phi(x)\}\right]^2 d\mu(x) = \theta.$$

Thus,

$$n^{1/2}(\hat{\beta}_1 - \hat{\beta}) \to N\left(0, \left[E\left\{\frac{\partial F_0^T(X_i;\beta_0)}{\partial\beta}\right\}\right]^{-1}\begin{pmatrix}\theta & 0\\ 0 & 0\end{pmatrix}\left[E\left\{\frac{\partial F_0^T(X_i;\beta_0)}{\partial\beta}\right\}\right]^{-T}\right)$$

in distribution. It can be verified that

$$E\left\{\frac{\partial F_0^T(X_i;\beta_0)}{\partial\beta}\right\} = -\begin{bmatrix}\frac{E\{2\pi_0(X)-1+2X\pi_{01}(X)-2\pi_{02}(X)\}}{\sigma_0} & \frac{2E(X)}{\sigma_0}\\ 2\sigma_0 E(X) & 2\sigma_0\end{bmatrix},$$

where expectation E on the right side is taken with respect to $2\phi(x)\pi_0(x)$. Putting these together, we get $n^{1/2}(\hat{\beta}_1 - \hat{\beta}) \to N(0, \theta B)$ in distribution, thus $n^{1/2}(\hat{\beta}_1 - \beta_0) \to N(0, A+\theta B)$ in distribution. With an α that minimizes θ, we will get the most efficient estimator given the parametric model $\pi(x,\alpha)$, the variance of $n^{1/2}(\hat{\beta}_1 - \beta_0)$ is $A + \min_\alpha(\theta)B$. □

In Proposition 7, we deliberately avoided specifying how to find the α that minimizes θ, since this depends on the true π_0 that is unknown to us. Typically in these problems a parametric model is assumed in terms of both β and α and MLE is used to estimate both sets of parameters. However, if the model for the skewing function is not correct then the MLE for β will be biased. What we propose is to use the MLE estimate for $\hat{\alpha}$ in the π function and proceed to estimate ξ and σ using the semiparametric estimating equation with π_0 replaced by π. This estimator will be consistent and asymptotically normal even if the model for $\pi(\cdot)$ was incorrectly specified and will be semiparametric efficient if it is correctly specified.

REFERENCES

Arnold, B. C. & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 7-54.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Bayarri, M. J. & DeGroot, M. (1992). A BAD view of of weighted distributions and selection models. In *Bayesian Statistic 4*. Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, pp. 17-33. Oxford: University Press.

Bickel, P., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and adaptive inference in semiparametric models.* Baltimore: Johns Hopkins University Press.

Copas, J. B. & Li, H. G. (1997). Inference from non-random samples (with discussion). *J. R. Statist. Soc. B* **59**, 55-95.

Genton, M. G. & Loperfido (2002). Generalized skew-elliptical distributions and their quadratic forms. *Institute of Statistics Mimeo Series* **#2539**, *under review.* (*http://www.stat.ncsu.edu/~mggenton/publications.html*)

Lee, J. & Berger, J. O. (2001). Semiparametric Bayesian Analysis of Selection Models. *J. Am. Statist. Assoc.* **96**, 1397-1409.

Ma, Y. & Genton, M. G. (2002). A semiparametric class of generalized skew-elliptical distributions. *Institute of Statistics Mimeo Series* **#2541**, *under review.* (*http://www.stat.ncsu.edu/~mggenton/publications.html*)

Newey, W. K. (1990). Semiparametric Efficiency Bounds. *J. Appl. Econom.* **5**, 99-135.

Rao, C. R. (1985). Weighted distributions arising out of methods of ascertainment: What populations does a sample represent? In *A Celebration of Statistics: The ISI Centenary Volume.* Ed. A. G. Atkinson & S. E. Fienberg, pp. 543-569. New York:Springer-Verlag.

Wacholder, S. & Weinberg, C. R. (1994). Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics* **50**, 350-357.

Weinberg, C. R. & Sandler, D. P. (1991). Randomized recruitment in case-control studies. *Am. J. Epidemiol.* **134**, 421-433.

Weinberg, C. R. & Wacholder, S. (1993). Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika* **80**, 461-465.

Zhang, B. (2000). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statist. Sinica* **10**, 839-865.