Estimators based on ω-dependent generalized weighted Cramér-von Mises distances under censoring - with applications to mixture models

L. Bordes, joint work with E. Beutner

Laboratoire de Mathématiques Appliquées de Pau - CNRS UMR 5142
Université de Pau et des Pays de l'Adour

Groupe de travail *Fiabilité et Domaines Connexes* - 23 janvier 2009

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

- Basic problem in statistics: Fit a parametric family $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ to data.

- Basic problem in statistics: Fit a parametric family $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ to data.
- Possibilities:
  - Maximum Likelihood Method,
  - Moment Method,
  - Minimum Distance Estimators.

- Basic problem in statistics: Fit a parametric family $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ to data.
- Possibilities:
  - Maximum Likelihood Method,
  - Moment Method,
  - Minimum Distance Estimators.
- The idea underlying minimum distance estimators is to minimize 'the distance' between the data and the assumed model $\mathcal{F}$.

To measure the distance between the i.i.d. data and the assumed model $\mathcal{F}$ one usually proceeds as follows:

To measure the distance between the i.i.d. data and the assumed model $\mathcal{F}$ one usually proceeds as follows:

- Summarize the data by the empirical distribution function.

To measure the distance between the i.i.d. data and the assumed model $\mathcal{F}$ one usually proceeds as follows:

- Summarize the data by the empirical distribution function.
- Define a (pseudo)–distance d for distribution functions.

To measure the distance between the i.i.d. data and the assumed model $\mathcal{F}$ one usually proceeds as follows:

- Summarize the data by the empirical distribution function.
- Define a (pseudo)–distance d for distribution functions.
- Estimate θ by the value which minimizes the distance between the empirical distribution function F_n and the assumed model $\mathcal{F} = \{F_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$.

Some applications of minimum distance estimators:

- Estimation in mixture models:
  - Pardo (1997): Chi-square minimum distance,
  - Woodward et al. (1984): Cramer-von-Mises distance.
- Manski (1983) and Brown and Wegkamp (2002) applied them in the context of a reduced-form function: $\varepsilon = \rho(X, Y, \theta)$.
- Lee and Song (2008) applied them to the estimation in GARCH models.

How to proceed if the data come from a random censorship model? Recall that in the random censorship model

How to proceed if the data come from a random censorship model? Recall that in the random censorship model

- In addition to the i.i.d random variables $X_1, \ldots, X_n$ with distribution function (d.f.) F_{θ_0}, there exist censoring variables $C_1, \ldots, C_n$ with d.f. H.

How to proceed if the data come from a random censorship model? Recall that in the random censorship model

- In addition to the i.i.d random variables $X_1, \ldots, X_n$ with distribution function (d.f.) F_{θ_0}, there exist censoring variables $C_1, \ldots, C_n$ with d.f. H.
- We only observe

$$(T_i, \Delta_i) \equiv \left(X_i \wedge C_i, I_{\{X_i \leq C_i\}}\right), \; i = 1, \ldots, n.$$

How to proceed if the data come from a random censorship model? Recall that in the random censorship model

- In addition to the i.i.d random variables $X_1, \ldots, X_n$ with distribution function (d.f.) F_{θ_0}, there exist censoring variables $C_1, \ldots, C_n$ with d.f. H.
- We only observe

$$(T_i, \Delta_i) \equiv \left(X_i \wedge C_i, I_{\{X_i \le C_i\}}\right), \; i = 1, \ldots, n.$$

- F_{θ_0} can be estimated by the Kaplan-Meier estimator

$$\hat{F}_n(t) = 1 - \prod_{\{i:\, T_i \le t\}} \left(1 - \frac{\Delta_i}{Y(T_i)}\right),$$

where $Y(t)$ describes the risk set size at time t^-.

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

- The Cramér-von Mises (CvM) distance is

$$D^{CvM}(G_1, G_2) = \int (G_1(x) - G_2(x))^2 dG_2(x).$$

- The Cramér-von Mises (CvM) distance is

$$D^{CvM}(G_1, G_2) = \int (G_1(x) - G_2(x))^2 dG_2(x).$$

- A minimum CvM distance estimator of θ is any $\hat{\theta}_n$ such that

$$D^{CvM}(F_n, F_{\hat{\theta}_n}) \leq D^{CvM}(F_n, F_\theta) \qquad \forall \theta \in \Theta.$$

- The Cramér-von Mises (CvM) distance is

$$D^{CvM}(G_1, G_2) = \int (G_1(x) - G_2(x))^2 dG_2(x).$$

- A minimum CvM distance estimator of θ is any $\hat{\theta}_n$ such that

$$D^{CvM}(F_n, F_{\hat{\theta}_n}) \leq D^{CvM}(F_n, F_\theta) \qquad \forall \theta \in \Theta.$$

- By the probability integral transform

$$D^{CvM}(F_n, F_\theta) = 1/(12n) + \sum_{i=1}^{n} \left(F_\theta(X_{(i)}) - i/n + 1/(2n)\right)^2,$$

where $X_{(i)}$ denotes the ith order statistic.

- Choi & Bulgren (1968) proposed to estimate the parameter of interest as the argmin of

$$\sum_{i=1}^{n} \left(F_\theta(X_{(i)}) - i/n\right)^2.$$

- Choi & Bulgren (1968) proposed to estimate the parameter of interest as the argmin of

$$\sum_{i=1}^{n} \left(F_\theta(X_{(i)}) - i/n\right)^2 .$$

- Öztürk & Hettmansperger (1997) introduced generalized weighted Cramér-von Mises distance estimators defined by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \int G(F_n(t) - F_\theta(t)) w(t, \theta)\, dt$$

where G is taken from a broad class of distance functions.

Although the Cramér-von Mises distance and the choice of Choi & Bulgren (1968) differ only by $1/(2n)$ in the argument of the function $f(x) = x^2$, MacDonald (1971) provides empirical evidence that the small sample bias of the estimator based on the Cramér-von Mises distance is smaller than the small sample bias of the estimator based on the choice of Choi & Bulgren (1968).

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

Desirable properties of the distance used to calculate $\hat{\theta}$ under the random censorship model:

Desirable properties of the distance used to calculate $\hat{\theta}$ under the random censorship model:

- Easy to calculate.

Desirable properties of the distance used to calculate $\hat{\theta}$ under the random censorship model:

- Easy to calculate.
- Allowing to emphasize or de-emphasize tails.

Desirable properties of the distance used to calculate $\hat{\theta}$ under the random censorship model:

- Easy to calculate.
- Allowing to emphasize or de-emphasize tails.
- Being sensitive to the amount of censoring.

- Define ω-dependent generalized weighted Cramér-von Mises distances under random censorship by:

$$\int_0^\tau G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t) \qquad \theta \in \Theta, \qquad (1),$$

where w_n weight function, and G_n distance function.

- Define ω-dependent generalized weighted Cramér-von Mises distances under random censorship by:

$$\int_0^\tau G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t) \qquad \theta \in \Theta, \qquad (1),$$

  where w_n weight function, and G_n distance function.
- An estimator of θ_0 is then any $\hat{\theta}_n$ which minimizes (1).

- Define ω-dependent generalized weighted Cramér-von Mises distances under random censorship by:

$$\int_0^{\tau} G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t) \qquad \theta \in \Theta, \qquad (1),$$

  where w_n weight function, and G_n distance function.
- An estimator of θ_0 is then any $\hat{\theta}_n$ which minimizes (1).
- Replacing integration w.r.t. F_θ by integration w.r.t. $\hat{F}_n$ transforms the integral into a sum.

- Define ω-dependent generalized weighted Cramér-von Mises distances under random censorship by:

$$\int_0^\tau G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t) \qquad \theta \in \Theta, \qquad (1),$$

  where w_n weight function, and G_n distance function.
- An estimator of θ_0 is then any $\hat{\theta}_n$ which minimizes (1).
- Replacing integration w.r.t. F_θ by integration w.r.t. $\hat{F}_n$ transforms the integral into a sum.
- w_n allows to de-emphasize, for example, the right tail.

- Define ω-dependent generalized weighted Cramér-von Mises distances under random censorship by:

$$\int_0^\tau G_n\left(\hat{F}_n(t) - F_\theta(t), \omega\right) w_n(t) d\hat{F}_n(t) \qquad \theta \in \Theta, \qquad (1),$$

  where w_n weight function, and G_n distance function.
- An estimator of θ_0 is then any $\hat{\theta}_n$ which minimizes (1).
- Replacing integration w.r.t. F_θ by integration w.r.t. $\hat{F}_n$ transforms the integral into a sum.
- w_n allows to de-emphasize, for example, the right tail.
- The distance G_n is allowed to depend on ω, for example the amount of censoring.

Examples

Some examples of ω-dependent generalized weighted Cramér-von Mises distances:

Examples

Some examples of ω-dependent generalized weighted Cramér-von Mises distances:

- Taking $G_n(\cdot,\omega) = (\cdot + 1/(2n))^2$ and $w_n(t) = 1$, we obtain the Cramér-von Mises distance although we are integrating w.r.t. $\hat{F}_n$.

Examples

Some examples of ω-dependent generalized weighted Cramér-von Mises distances:

- Taking $G_n(\cdot, \omega) = (\cdot + 1/(2n))^2$ and $w_n(t) = 1$, we obtain the Cramér-von Mises distance although we are integrating w.r.t. $\hat{F}_n$.
- Given MacDonald's observation an adequate modification under censoring could be $G_n(\cdot, \omega) = (\cdot + 1/c_n)^2$, where $c_n = 2\sum_{i=1}^n I_{\{X_i \leq C_i, X_i \leq \tau\}}$ and $w_n \equiv 1$.

Examples

Some examples of ω-dependent generalized weighted Cramér-von Mises distances:

- Taking $G_n(\cdot,\omega) = (\cdot + 1/(2n))^2$ and $w_n(t) = 1$, we obtain the Cramér-von Mises distance although we are integrating w.r.t. $\hat{F}_n$.
- Given MacDonald's observation an adequate modification under censoring could be $G_n(\cdot,\omega) = (\cdot + 1/c_n)^2$, where $c_n = 2\sum_{i=1}^n I_{\{X_i \leq C_i, X_i \leq \tau\}}$ and $w_n \equiv 1$.
- Taking $w_n(t) = (1 - F_n(t))$ allows us to de-emphasize the right tail. As, under right censoring, we usually have less observations in the right tail this is desirable.

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

Conditions

The following conditions are imposed to obtain the asymptotic results:

Conditions

The following conditions are imposed to obtain the asymptotic results:

A. Let $\tau > 0$ be a real number such that $\tau < \sup\{t > 0; (1 - F_{\theta_0}(t))(1 - H(t)) > 0\}$.

Conditions

The following conditions are imposed to obtain the asymptotic results:

A. Let $\tau > 0$ be a real number such that $\tau < \sup\{t > 0; (1 - F_{\theta_0}(t))(1 - H(t)) > 0\}$.

B. $\sup_{[0,\tau]} |w_n - w_0| \xrightarrow{P} 0$, where w_0 is a bounded deterministic function on $[0, \tau]$.

Conditions

The following conditions are imposed to obtain the asymptotic results:

A. Let $\tau > 0$ be a real number such that $\tau < \sup\{t > 0; (1 - F_{\theta_0}(t))(1 - H(t)) > 0\}$.

B. $\sup_{[0,\tau]} |w_n - w_0| \xrightarrow{P} 0$, where w_0 is a bounded deterministic function on $[0, \tau]$.

C. Let the set Υ consists of all functions G such that
   - (i) $G : [a, b] \to \mathbb{R}^+$, $a \leq -1, 1 \leq b$, is nonnegative,
   - (ii) the restriction of G to the interval $[-1, 1]$ is twice continuously differentiable,
   - (iii) $G(0) = G'(0) = 0$ and $G''(0) > 0$.

   $G_n : [a, b] \times \Omega_n \to \mathbb{R}^+$ is such that $G_n(\cdot, \omega) = G\left(\cdot + o_p(1/\sqrt{n})\right)$ for some $G \in \Upsilon$.

Conditions (continued)

D. If $\theta_n \in \Theta \subset \mathbb{R}^p$, $n \in \mathbb{N}$, then

$$\lim_{n \to +\infty} \int_0^\tau G\left(F_{\theta_0} - F_{\theta_n}\right) w_0 dF_{\theta_0} = 0$$

implies $\lim_{n \to +\infty} \theta_n = \theta_0$.

Conditions (continued)

D. If $\theta_n \in \Theta \subset \mathbb{R}^p$, $n \in \mathbb{N}$, then

$$\lim_{n\to+\infty} \int_0^\tau G\left(F_{\theta_0} - F_{\theta_n}\right) w_0 dF_{\theta_0} = 0$$

implies $\lim_{n\to+\infty} \theta_n = \theta_0$.

E. There exists a measurable function $\eta = (\eta_1, \ldots, \eta_p)^t : (0, q) \equiv (0, F_{\theta_0}(\tau)) \to \mathbb{R}^p$ such that $\Sigma(\tau) = \int_0^\tau \eta(F_{\theta_0}(s))\eta^t(F_{\theta_0}(s))w_0(s)dF_{\theta_0}(s)$ is positive definite, and

$$\sup_{0\leq s\leq\tau} \left|F_\theta(s) - F_{\theta_0}(s) - (\theta - \theta_0)^t \eta \circ F_{\theta_0}(s)\right| = o(\|\theta - \theta_0\|)$$

as $\|\theta - \theta_0\| \to 0$.

Comments

- Condition A ensures that the standard results for the Kaplan-Meier estimator hold true on the interval $[0, \tau]$.

Comments

- Condition A ensures that the standard results for the Kaplan-Meier estimator hold true on the interval $[0, \tau]$.
- The sequence of functions $G_n(\cdot, \omega) = G(\cdot + 1/(2n))$, where $G(\cdot) = (\cdot)^2$, fulfills Condition C.

Comments

- Condition A ensures that the standard results for the Kaplan-Meier estimator hold true on the interval $[0,\tau]$.
- The sequence of functions $G_n(\cdot,\omega) = G(\cdot + 1/(2n))$, where $G(\cdot) = (\cdot)^2$, fulfills Condition C.
- Under Condition A we have that $(1/n)\sum_{i=1}^n I_{\{X_i \le C_i, X_i \le \tau\}}$ converges to $\int_0^\tau (1-H)dF_{\theta_0} \geq (1-H(\tau))F_{\theta_0}(\tau) > 0$. Therefore, $G_n(\cdot,\omega) = (\cdot + 1/c_n)^2$, also fulfills Condition C.

Comments

- Condition A ensures that the standard results for the Kaplan-Meier estimator hold true on the interval $[0, \tau]$.
- The sequence of functions $G_n(\cdot, \omega) = G(\cdot + 1/(2n))$, where $G(\cdot) = (\cdot)^2$, fulfills Condition C.
- Under Condition A we have that $(1/n) \sum_{i=1}^{n} I_{\{X_i \leq C_i, X_i \leq \tau\}}$ converges to $\int_0^\tau (1 - H) dF_{\theta_0} \geq (1 - H(\tau)) F_{\theta_0}(\tau) > 0$. Therefore, $G_n(\cdot, \omega) = (\cdot + 1/c_n)^2$, also fulfills Condition C.
- Condition E allows a first order Taylor expansion of $\theta \mapsto F_\theta(x)$ around θ_0 uniformly in x.

Consistency

Theorem

Any sequence $(\hat{\theta}_n)_{n\geq 1}$ defined by (1) is consistent if Conditions A–D hold.

Consistency

Theorem

Any sequence $(\hat{\theta}_n)_{n\geq 1}$ defined by (1) is consistent if Conditions A–D hold.

Sketch of proof The result would follow if

$$\begin{aligned}
&\int_0^{\tau} G\left(F_{\theta_0} - F_{\hat{\theta}_n}\right) w_0 dF_{\theta_0} \\
&\leq \int_0^{\tau} G_n\left(\hat{F}_n - F_{\hat{\theta}_n}, \omega\right) w_n d\hat{F}_n + o_P(1) \\
&\leq \int_0^{\tau} G_n\left(\hat{F}_n - F_{\theta_0}, \omega\right) w_n d\hat{F}_n + o_P(1) \\
&\leq \sup_{[0,\tau]} G\left(\hat{F}_n - F_{\theta_0} + o_p(1/\sqrt{n})\right) \sup_{[0,\tau]} |w_n| + o_P(1)
\end{aligned}$$

Consistency

Sketch of proof (continued):

$$\sup_{\theta\in\Theta}\left|\int_0^\tau G_n\left(\hat{F}_n - F_\theta,\omega\right) w_n d\hat{F}_n - \int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 dF_{\theta_0}\right|$$

$$\leq \sup_{\theta\in\Theta}\left|\int_0^\tau \left[G\left(\hat{F}_n - F_\theta + o_p(1/\sqrt{n})\right) - G\left(F_{\theta_0} - F_\theta\right)\right] w_n d\hat{F}_n\right|$$

$$+ \sup_{[-1,1]}|G| \times \sup_{[0,\tau]}|w_n - w_0|$$

$$+\sup_{\theta\in\Theta}\left|\int_0^\tau G\left(F_{\theta_0} - F_\theta\right) w_0 d\left(\hat{F}_n - F_{\theta_0}\right)\right|$$

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta)w_0; \theta \in \Theta\}$ *is* $\mathbb{P}$*–Glivenko-Cantelli.*

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta) w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.

Sketch of proof:

- G has finite variations on $[-1, 1]$. Hence, $G = G^+ - G^-$.

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta)w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.

Sketch of proof:

- G has finite variations on $[-1, 1]$. Hence, $G = G^+ - G^-$.
- $\mathcal{W} = \{F_\theta; \theta \in \Theta\} \subset \mathcal{M}$ where $\mathcal{M}$ is the set of monotone increasing functions.

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta)w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.

Sketch of proof:

- G has finite variations on $[-1, 1]$. Hence, $G = G^+ - G^-$.
- $\mathcal{W} = \{F_\theta; \theta \in \Theta\} \subset \mathcal{M}$ where $\mathcal{M}$ is the set of monotone increasing functions.
- The class $\mathcal{M}$ has a finite ε–bracketing number.

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta) w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.

Sketch of proof:

- G has finite variations on $[-1, 1]$. Hence, $G = G^+ - G^-$.
- $\mathcal{W} = \{F_\theta; \theta \in \Theta\} \subset \mathcal{M}$ where $\mathcal{M}$ is the set of monotone increasing functions.
- The class $\mathcal{M}$ has a finite ε–bracketing number.
- $\mathcal{W}' = \{F_\theta - F_{\theta_0}; \theta \in \Theta\}$ has a finite bracketing number.

Consistency

Lemma

Let G fulfilling Condition C. Then, we have that the class of functions $\mathcal{Z} = \{G \circ (F_{\theta_0} - F_\theta) w_0; \theta \in \Theta\}$ is $\mathbb{P}$–Glivenko-Cantelli.

Sketch of proof:

- G has finite variations on $[-1, 1]$. Hence, $G = G^+ - G^-$.
- $\mathcal{W} = \{F_\theta; \theta \in \Theta\} \subset \mathcal{M}$ where $\mathcal{M}$ is the set of monotone increasing functions.
- The class $\mathcal{M}$ has a finite ε–bracketing number.
- $\mathcal{W}' = \{F_\theta - F_{\theta_0}; \theta \in \Theta\}$ has a finite bracketing number.
- Let $[l_i, u_i]$, $i = 1, \ldots, m$, be the brackets covering $\mathcal{W}'$. Then $[w_0 G^+ \circ l_i, w_0 G^+ \circ u_i]$, $i = 1, \ldots, m$, cover $G^+ \circ \mathcal{W}' = \{w_0 G^+ \circ z; z \in \mathcal{W}'\}$.

Asymptotic normality

Theorem

Let $\mathcal{B}$ be a centered Gaussian process on $[0,\tau]$ with covariance function ρ, where for $(s,t) \in [0,\tau]^2$:

$$\rho(s,t) = (1 - F_{\theta_0}(s))(1 - F_{\theta_0}(t)) \int_0^{s \wedge t} \frac{d\Lambda_{\theta_0}(u)}{(1 - F_{\theta_0}(u))(1 - H(u))}$$

If Conditions A–E hold, then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges to a centered normal distribution with variance $\Sigma^{-1}(\tau) C(\tau) \Sigma^{-1}(\tau)$ where $C(\tau) = \mathrm{Var}\left(\int_0^\tau \mathcal{B}\eta \circ F_{\theta_0} w_0 dF_{\theta_0}\right)$.

Sketch of proof:

- Guess a 'limit' of the stochastic process

$$X_n(\xi) = n \int_0^\tau G_n\left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}}, \omega\right) w_n d\hat{F}_n, \quad \xi \in \mathbb{R}^p.$$

which is minimized by $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$.

Sketch of proof:

- Guess a 'limit' of the stochastic process

$$X_n(\xi) = n \int_0^\tau G_n\left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}}, \omega\right) w_n d\hat{F}_n, \quad \xi \in \mathbb{R}^p.$$

  which is minimized by $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$.

- 'Solution' is the process $(\mathcal{B}_n = \sqrt{n}(\hat{F}_n - F_{\theta_0}))$

$$\bar{X}_n(\xi) = \frac{G''(0)}{2} \int_0^\tau \left(\mathcal{B}_n - \xi^t \eta \circ F_{\theta_0}\right)^2 w_0 dF_{\theta_0}, \quad \xi \in \mathbb{R}^p,$$

Sketch of proof:

- Guess a 'limit' of the stochastic process

$$X_n(\xi) = n \int_0^\tau G_n\left(\hat{F}_n - F_{\theta_0+\xi/\sqrt{n}}, \omega\right) w_n d\hat{F}_n, \quad \xi \in \mathbb{R}^p.$$

  which is minimized by $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$.

- 'Solution' is the process $(\mathcal{B}_n = \sqrt{n}(\hat{F}_n - F_{\theta_0}))$

$$\bar{X}_n(\xi) = \frac{G''(0)}{2} \int_0^\tau \left(\mathcal{B}_n - \xi^t \eta \circ F_{\theta_0}\right)^2 w_0 dF_{\theta_0}, \quad \xi \in \mathbb{R}^p,$$

- Indeed, for any $A = \{\xi \in \mathbb{R}^p; \|\xi\| < c\}$ we have

$$\sup_{\xi \in A} \left|X_n(\xi) - \bar{X}_n(\xi)\right| \xrightarrow{P} 0.$$

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- $\bar{\xi}_n$ converges weakly to $\Sigma^{-1}(\tau) \int_0^\tau \mathcal{B} \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- $\bar{\xi}_n$ converges weakly to $\Sigma^{-1}(\tau) \int_0^\tau \mathcal{B} \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- Show that $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is $O_P(1)$.

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- $\bar{\xi}_n$ converges weakly to $\Sigma^{-1}(\tau) \int_0^\tau \mathcal{B} \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- Show that $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is $O_P(1)$.
- Hence, the probability of $E_n = \{\bar{\xi}_n \in A, \hat{\xi}_n \in A\}$ is as large as we want for n large enough.

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- $\bar{\xi}_n$ converges weakly to $\Sigma^{-1}(\tau) \int_0^\tau \mathcal{B} \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- Show that $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is $O_P(1)$.
- Hence, the probability of $E_n = \{\bar{\xi}_n \in A, \hat{\xi}_n \in A\}$ is as large as we want for n large enough.
- Finally, use the uniform convergence of X_n to $\bar{X}_n$ to show that the probability of $\{\hat{\xi}_n \in A \backslash B_n\}$, where $B_n = \{\xi \in \mathbb{R}^p; \|\xi - \bar{\xi}_n\| < \varepsilon\}$ converges to zero.

Sketch of proof (continued):

- The maximizer of $\bar{X}_n$ is $\bar{\xi}_n = \Sigma^{-1}(\tau) \int_0^\tau \mathcal{B}_n \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- $\bar{\xi}_n$ converges weakly to $\Sigma^{-1}(\tau) \int_0^\tau \mathcal{B} \eta \circ F_{\theta_0} w_0 dF_{\theta_0}$.
- Show that $\hat{\xi}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is $O_P(1)$.
- Hence, the probability of $E_n = \{\bar{\xi}_n \in A, \hat{\xi}_n \in A\}$ is as large as we want for n large enough.
- Finally, use the uniform convergence of X_n to $\bar{X}_n$ to show that the probability of $\{\hat{\xi}_n \in A \backslash B_n\}$, where $B_n = \{\xi \in \mathbb{R}^p ; \|\xi - \bar{\xi}_n\| < \varepsilon\}$ converges to zero.
- Thus, $\hat{\xi}_n = \bar{\xi}_n + o_P(1)$.

Outline

1. Introduction
2. Short history of Cramér-von Mises minimum distance estimators
3. ω-dependent generalized weighted Cramér-von Mises distances
4. Asymptotic results
5. Small sample behavior. Some simulations

Small sample behavior

G_n	$(\cdot)^2$		$(\cdot + 1/(2n))^2$		$(\cdot + 1/n)^2$		$(\cdot + 1/c_n)^2$	
	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
0%	0.254	0.132	0.297	0.136	0.341	0.135	0.297	0.135
20%	0.241	0.142	0.283	0.145	0.326	0.147	0.294	0.147
40%	0.218	0.148	0.258	0.154	0.299	0.158	0.288	0.161
60%	0.172	0.153	0.206	0.163	0.243	0.171	0.273	0.191

Table 1: Estimation of the mixture parameter based on **twenty** *observations from a Weibull mixture with cdf* $F_\theta = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3))$. *The values given in the table are based on 10,000 simulations.*

Small sample behavior

G_n	$(\cdot)^2$		$(\cdot + 1/(2n))^2$		$(\cdot + 1/n)^2$		$(\cdot + 1/c_n)^2$	
	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
0%	0.289	0.069	0.300	0.069	0.311	0.069	0.300	0.069
20%	0.284	0.074	0.295	0.074	0.306	0.074	0.297	0.074
40%	0.278	0.082	0.289	0.082	0.300	0.082	0.296	0.083
60%	0.258	0.098	0.269	0.098	0.279	0.098	0.285	0.099

Table 2: Estimation of the mixture parameter based on **eighty** *observations from a Weibull mixture with cdf* $F_\theta = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$. *The values given in the table are based on 10,000 simulations.*

Behavior under contamination

To study the behavior of the estimators under a contamination model when censoring might be present, we took three different contamination models, namely,

- $CM_1 = \{\tilde{F} = (1-\epsilon)F + \epsilon H_1\}$
- $CM_2 = \{\tilde{F} = (1-\epsilon)F + \epsilon H_2\}$,
- $CM_3 = \{\tilde{F} = (1-\epsilon)F + \epsilon H_3\}$.

where $\epsilon = 0.05$, F is the d.f. of the above Weibull mixture, and H_1, H_2 and H_3 are gamma mixtures.

Behavior under contamination

The properties of the three contamination models can be summarized as follows:

Behavior under contamination

The properties of the three contamination models can be summarized as follows:

- CM_1 corresponds to a 'symmetric' contamination model in the sense that $P(X < Y_1) \approx P(X > Y_1)$, where $X \sim F$, $Y_1 \sim H_1$, and that the probability of being censored is approximately equal for X and Y_1.

Behavior under contamination

The properties of the three contamination models can be summarized as follows:

- CM_1 corresponds to a 'symmetric' contamination model in the sense that $P(X < Y_1) \approx P(X > Y_1)$, where $X \sim F$, $Y_1 \sim H_1$, and that the probability of being censored is approximately equal for X and Y_1.
- CM_2 corresponds to a 'left' contamination model, i.e. $P(X < Y_2) \approx 9\%$, $Y_2 \sim H_2$, and the probability for X to be censored much larger than the probability for Y_2 to be censored.

Behavior under contamination

The properties of the three contamination models can be summarized as follows:

- CM_1 corresponds to a 'symmetric' contamination model in the sense that $P(X < Y_1) \approx P(X > Y_1)$, where $X \sim F$, $Y_1 \sim H_1$, and that the probability of being censored is approximately equal for X and Y_1.
- CM_2 corresponds to a 'left' contamination model, i.e. $P(X < Y_2) \approx 9\%$, $Y_2 \sim H_2$, and the probability for X to be censored much larger than the probability for Y_2 to be censored.
- CM_3 corresponds to a 'right' contamination model, i.e. $P(X < Y_3) \approx 93\%$, $Y_3 \sim H_3$, the probability for Y_3 to be censored much larger than the probability for X to be censored.

Symmetric contamination model

G_n	$(\cdot)^2$		$(\cdot + 1/(2n))^2$		$(\cdot + 1/n)^2$		$(\cdot + 1/c_n)^2$	
	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
0%	0.278	0.100	0.300	0.100	0.322	0.100	0.300	0.100
40%	0.258	0.117	0.280	0.117	0.301	0.118	0.295	0.119

Table 3: Estimation of the mixture parameter based on forty observations from a Weibull mixture with cdf $F_\theta = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$ *under a symmetric contamination model. The values given in the table are based on 10,000 simulations.*

Left contamination model

G_n	$(\cdot)^2$		$(\cdot + 1/(2n))^2$		$(\cdot + 1/n)^2$		$(\cdot + 1/c_n)^2$	
	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
0%	0.245	0.099	0.267	0.099	0.289	0.099	0.267	0.099
40%	0.224	0.115	0.245	0.116	0.267	0.117	0.259	0.118

Table 4: Estimation of the mixture parameter based on forty observations from a Weibull mixture with cdf $F_\theta = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$ *under a left contamination model. The values given in the table are based on 10,000 simulations.*

Right contamination model

G_n	$(\cdot)^2$		$(\cdot + 1/(2n))^2$		$(\cdot + 1/n)^2$		$(\cdot + 1/c_n)^2$	
	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE	$\hat{\pi}$	MSE
0%	0.321	0.097	0.343	0.097	0.365	0.096	0.343	0.097
40%	0.299	0.117	0.320	0.117	0.342	0.117	0.343	0.097

Table 5: Estimation of the mixture parameter based on forty observations from a Weibull mixture with cdf $F_\theta = 1 - 0.3\exp(-(x/5)^3) - 0.7\exp(-(x/2)^3)$ *under a left contamination model. The values given in the table are based on 10,000 simulations.*

Behavior under contamination

Is there any heuristic or theoretical explanation for the observed behavior under the different contamination models? Heuristically,

Behavior under contamination

Is there any heuristic or theoretical explanation for the observed behavior under the different contamination models? Heuristically,

- The Weibull mixture is $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ with the first component stochastically larger than the second.

Behavior under contamination

Is there any heuristic or theoretical explanation for the observed behavior under the different contamination models? Heuristically,

- The Weibull mixture is $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ with the first component stochastically larger than the second.
- On average, under CM_2 two small observations. Therefore, by minimizing the distance between the Weibull mixture and the empirical distribution function one puts less weight on the stochastically larger component of the mixture.

Behavior under contamination

Is there any heuristic or theoretical explanation for the observed behavior under the different contamination models? Heuristically,

- The Weibull mixture is $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ with the first component stochastically larger than the second.
- On average, under CM_2 two small observations. Therefore, by minimizing the distance between the Weibull mixture and the empirical distribution function one puts less weight on the stochastically larger component of the mixture.
- On average, under CM_3 two large observations. Thus, more weight on the stochastically larger component.

Behavior under contamination

Theoretically, if there is no censoring:

Behavior under contamination

Theoretically, if there is no censoring:

- For $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ and $w_n \equiv 1$ the influence function IF_{Δ_x} of H at x is, up to a positive constant, given by

$$\int_0^\tau \left(-\exp(-(t/5)^3) + \exp(-(t/2)^3)\right) \cdot (I_{\{t \geq x\}} - H)dH(t)$$

which is

Behavior under contamination

Theoretically, if there is no censoring:

- For $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ and $w_n \equiv 1$ the influence function IF_{Δ_x} of H at x is, up to a positive constant, given by

$$\int_0^{\tau} \left(-\exp(-(t/5)^3) + \exp(-(t/2)^3)\right) \cdot (I_{\{t \geq x\}} - H)dH(t)$$

  which is

- negative for small x,

Behavior under contamination

Theoretically, if there is no censoring:

- For $F_\theta(t) = 1 - \theta \exp(-(t/5)^3) - (1-\theta)\exp(-(t/2)^3)$ and $w_n \equiv 1$ the influence function IF_{Δ_x} of H at x is, up to a positive constant, given by

$$\int_0^\tau (-\exp(-(t/5)^3) + \exp(-(t/2)^3)) \cdot (I_{\{t \geq x\}} - H)dH(t)$$

which is

- negative for small x,
- positive for large x, $x \leq \tau$.

THANK YOU!