Estimation under L-moment condition models

Alexis Decurninge, Michel Broniatowski

LSTA, Université Pierre et Marie Curie, Paris VI

27 August 2013
Plan

1. L-moments: properties
2. Moment and L-moment equations models
3. Minimum of $\varphi$-divergence estimators
4. Asymptotic properties
Definition

Sample: $X_1, \ldots, X_n$ real random variable of common distribution function $F$

$$
\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathbb{E}[X_{r-k}:r]
$$

$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$: order statistics

- $\lambda_1 = \mathbb{E}[X]$ : measure of location
- $\lambda_2 = \mathbb{E}[X_{2:2} - X_{1:2}]$ : measure of scale
- $\frac{\lambda_3}{\lambda_2} = \frac{\mathbb{E}[X_{3:3} - 2X_{2:3} + X_{1:3}]}{\mathbb{E}[X_{2:2} - X_{1:2}]}$ : measure of skewness
- $\frac{\lambda_4}{\lambda_2} = \frac{\mathbb{E}[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}]}{\mathbb{E}[X_{2:2} - X_{1:2}]}$ : measure of kurtosis

Existence since $\int |x|dF(x) < \infty$
Definition (continuous distributions)

If $F$ is continuous:

$$
\mathbb{E}[X_{j:r}] = \frac{r!}{(j-1)!(r-j)!} \int_{\mathbb{R}} x F(x)^{j-1} (1 - F(x))^{r-j} dF(x)
$$

L-moments can then be written:

$$
\lambda_r = \int_0^1 F^{-1}(t) L_r(t) dt = \int_0^1 F^{-1}(t) dK_r(t)
$$

with

- $L_r$ shifted Legendre polynomials (orthogonal basis in $L_2([0, 1])$)

$$
L_r(t) = \sum_{k=0}^{r} (-1)^k \binom{r}{k} t^{r-k} (1-t)^k = \sum_{k=0}^{r} (-1)^r t^{-k} \binom{r}{k} \binom{r+k}{k} t^k
$$

- $K_r(t) = \int_0^t L_{r-1}(u) du$
Definition (discrete distributions)

L-moments for a multinomial of support $x_1 \leq x_2 \leq \ldots \leq x_n$ and associated weights $\pi_1, \ldots, \pi_n \ (\sum_{i=1}^{n} \pi_i = 1)$

$$\lambda_r = \sum_{i=1}^{n} w_i^{(r)} x_i = \sum_{i=1}^{n} \left[ K_r \left( \sum_{a=1}^{i} \pi_a \right) - K_r \left( \sum_{a=1}^{i-1} \pi_a \right) \right] x_i$$
Estimation of $L$-moments

Unbiased estimator: $U$-statistics

$$l_r^{(u)} = \frac{1}{\binom{n}{r}} \sum_{1 \leq i_1 < \ldots < i_r \leq n} \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}:n}$$

Biased estimator: $V$-statistics (L-moments of the empirical distribution)

$$l_r = \frac{1}{\binom{r+n-1}{n-1}} \sum_{1 \leq i_1 \leq \ldots \leq i_r \leq n} \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}:n}$$
Illustration: asymptotic quality for a Gumbel family
Illustration: preservation of shape information
Moment and L-moment equations models

- Semi-parametric model defined through moments
  \[ \int_{\mathbb{R}} g(x, \theta) dF(x) = \mathbb{E}[g(X, \theta)] = 0 \]
  where \( g : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^l \)

- Semi-parametric model defined through L-moments
  \[ \mathbb{E} \left[ \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{k:n} \right] - \lambda_r(\theta) = 0 \quad 2 \leq r \leq l \]
  equivalent to
  \[ \int_{\mathbb{R}} K(F(x)) dx - f(\theta) = 0 \quad \text{with } f = (f_2, \ldots, f_l) \quad \text{and } K = (K_2, \ldots, K_l) \]
L-moment model example

L-moments of the Weibull family:

\[
\begin{align*}
\int_{\mathbb{R}} K_2(F(x))dx &= f_2(\sigma, \nu) = -\sigma(1 - 2^{-1/\nu})\Gamma(1 + 1/\nu) \\
\int_{\mathbb{R}} K_3(F(x))dx &= f_3(\sigma, \nu) = f_2(\sigma, \nu)[3 - 2\frac{1-3^{-1/\nu}}{1-2^{-1/\nu}}] \\
\int_{\mathbb{R}} K_4(F(x))dx &= f_4(\sigma, \nu) = f_2(\sigma, \nu)[6 + \frac{5(1-4^{-1/\nu})-10(1-3^{-1/\nu})}{1-2^{-1/\nu}}]
\end{align*}
\]
Estimation under L-moment condition models

Minimum of $\varphi$-divergence estimators

$\varphi$-divergence

Let $\varphi : \mathbb{R} \to [0, +\infty]$ be a strictly convex function with $\varphi(1) = 0$
For example: Cressie-Read family $\varphi(u) = \frac{u^k - ku + k - 1}{k(k-1)}$

Divergence between distributions $F$ and $G$:

$$D_\varphi(F, G) = \int_{\mathbb{R}} \varphi \left( \frac{dG}{dF}(x) \right) dF(x)$$

where $\frac{dG}{dF}$ is the Radon-Nikodym derivative.
The model would be:

\[ M_{\theta}^{(0)}(F) = \{ G \in M_+ | G \ll F, \int_{\mathbb{R}} K(G(x))dx = f(\theta) \} \]

Let \( F_n(x) = \sum_{i=1}^{n} \frac{1}{n} \mathbb{1}_{x>x_i} \) be the empirical distribution for a sample \( x_1, ..., x_n \). The plug-in estimator is then:

\[ \hat{\theta}_n^{(0)} = \arg \inf_{\theta \in \Theta} \inf_{G \in M_{\theta}^{(0)}(F_n)} D_\varphi(F_n, G) \]

But:
- existence?
- quick computation?
Minimum of $\varphi$-divergence : illustration
Projection for the L-moment constraint model

We choose to minimize the divergence between quantile measure in order to obtain linearity.
The model become

\[ M_\theta(F) = \{ G \in M_+ | G^{-1} \ll F^{-1}, \int_0^1 K(u) dG^{-1}(t) = f(\theta) \} \]

and the estimator

\[ \hat{\theta}_n = \arg \inf_{\theta \in \Theta} \inf_{G \in M_\theta(F_n)} \int_0^1 \varphi \left( \frac{dG^{-1}}{dF_n^{-1}} \right) dF_n^{-1}(u) \]
Relation with transport

If now \( T_n(x) = G^{-1} \circ F_n(x) \) for \( x \in \mathbb{R} \), we rewrite the estimator

\[
\inf \int_{\mathbb{R}} K(F_n(x)) dT_n(x) = f(\theta) \int_{\mathbb{R}} \varphi \left( T'_n(x) \right) dx =
\]

\[
\inf \int_0^1 K(u) dG^{-1}(u) = f(\theta) \int_0^1 \varphi \left( \frac{dG^{-1}}{dF_n^{-1}} \right) dF_n^{-1}(u)
\]

\( T_n \) is a transport between the random variable \( X \) of distribution function \( F_n \) and \( Y \) of distribution function \( G : T_n(X) \overset{d}{=} Y \)
Dual representation

Fenchel-Legendre transform of $\varphi$

$$\forall t \in \mathbb{R}, \quad \psi(t) = \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\}$$

Proposition

Let $\theta \in \Theta$ and $F$ be fixed. If there exists some $T$ in $M_\theta(F)$ such that $\frac{dT}{d\mu} \in \text{int}(\text{dom}(\varphi)) \, \mu$-a.s.

$$\inf \int \mathbb{R} K(F(x)) dT(x) = f(\theta) \int \varphi \left( \frac{dT}{d\mu} \right) d\mu$$

$$= \sup_{\xi \in \mathbb{R}} \xi^T f(\theta) - \int \mathbb{R} \psi(\xi^T K(F(x))) d\mu$$
Dual representation

If $\psi$ is derivable and there exists a solution $\xi^*$ of the dual problem which is an interior point of
\[ \{ \xi \in \mathbb{R}^l \text{ s.t. } \int_{\mathbb{R}} \psi(\langle \xi, K(F(x)) \rangle) \, d\mu < \infty \} , \]
then $\xi^*$ is the unique maximum checking:

\[ \int \psi'(\xi^T K \circ F(x))K \circ F(x) \, d\mu = f(\theta) \]

and $\theta \mapsto \xi^*(\theta)$ is continuous.
Dual representation for \(\chi^2\)-divergence

With the \(\chi^2\)-divergence \(\varphi(x) = \frac{(x-1)^2}{2}\), \(\psi(t) = \frac{1}{2}t^2 + t\)

The solution \(\xi_1^*\) of the dual problem is

\[
\xi^* = \Omega^{-1} \left( f(\theta) - \int K(F(x))d\mu \right)
\]

with

\[
\Omega = \int K(F(x))K(F(x))^T d\mu
\]

And the estimator is then

\[
\hat{\theta}_n = \arg \inf_{\theta \in \Theta} \left( f(\theta) - \int K(F_n(x))d\mu \right) \Omega_n^{-1} \left( f(\theta) - \int K(F_n(x))d\mu \right)
\]

with \(\Omega_n = \int K(F_n(x))K(F_n(x))^T d\mu\)
Asymptotic properties of the estimators under the model

Theorem

Let $X_1, ..., X_n$ be random samples coming from the same distribution $F_0$. Let suppose that there exists $\theta_0$ such that

- $F_0 \in M_{\theta_0}$, $\theta_0$ is the unique solution of the equation $f(\theta) = f(\theta_0)$
- $f$ is continuous and $\Theta$ is compact
- the matrix $\Omega = \int K(F_0(x))K(F_0(x))^T dx$ is non singular.

Then with probability approaching one,

$$\hat{\theta}_n \xrightarrow{p} \theta_0$$
Asymptotic properties of the estimators under the model

Asymptotic normality of the estimator:

**Theorem**

Let define

- $J_0 = J_f(\theta_0)$ be the Jacobian of $f$ with respect to $\theta$ in $\theta_0$
- $M = (J_0^T \Omega^{-1} J_0)^{-1}$, $H = MJ_0^T \Omega^{-1}$,
  
  $P = \Omega^{-1} - \Omega^{-1} J_0 M J_0^T \Omega^{-1}$
- $\Sigma = \iint [F_0(\min(x, y)) - F_0(x) F_0(y)] K'(F_0(x)).K'(F_0(y)) \, dx \, dy$

Then,

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\xi}_n \end{pmatrix} \xrightarrow{d} N(0, \text{diag}(H\Sigma H^T, P\Sigma P^T))$$
Further work

- Same asymptotic properties under misspecification
- Extension to multivariate case
Radar power data
Thermal noise vs impulsive noise

Graphs showing the log of Gaussian Detector OGD Likelihood Ratio for Thermal Noise and Impulsive Noise. The graphs compare the likelihood, OGD theoretical threshold, and Monte Carlo threshold.
Processing chain for adaptive detection

<table>
<thead>
<tr>
<th>Decision D=0: no target</th>
<th>H(_0) (no target)</th>
<th>H(_1) (target)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision D=1: target</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise ((1 - P_{fa}))</td>
<td>Miss ((1 - P_{det}))</td>
</tr>
<tr>
<td></td>
<td>False alarm (P_{fa})</td>
<td>Detection (P_{det})</td>
</tr>
</tbody>
</table>

**Constraint**: keep the false alarm constant with robustness to
- misspecification
- other targets

**Aim**: estimation of \(H_0\) distribution