COMPSTAT 2010

*

Score moment estimates

Zdeněk Fabián
Institute of Computer Sciences, Prague

August 17, 2010

Motivation

- Apart from the fact that the ML estimates $\hat{\theta}_{ML}$ are often influenced by outliers, the solution $f(x; \hat{\theta}_{ML})$ of the parametric estimation problem has some other drawbacks:

Motivation

- Apart from the fact that the ML estimates $\hat{\theta}_{ML}$ are often influenced by outliers, the solution $f(x;\hat{\theta}_{ML})$ of the parametric estimation problem has some other drawbacks:

- Instead of $f(x;\hat{\theta}_{ML})$, a few numbers characterizing the data would be useful in further analysis. However, moments $m_k = E(X-m_1)^k, m_1 = EX$ are often queer expressions containing special functions, and moments of heavy-tailed distributions do not exist, so that the approach $\hat{m}_k = m_k(\hat{\theta}_{ML})$ is not used

Motivation

- Apart from the fact that the ML estimates $\hat{\theta}_{ML}$ are often influenced by outliers, the solution $f(x; \hat{\theta}_{ML})$ of the parametric estimation problem has some other drawbacks:

- Instead of $f(x; \hat{\theta}_{ML})$, a few numbers characterizing the data would be useful in further analysis. However, moments $m_k = E(X - m_1)^k, m_1 = EX$ are often queer expressions containing special functions, and moments of heavy-tailed distributions do not exist, so that the approach $\hat{m}_k = m_k(\hat{\theta}_{ML})$ is not used
- Complex problems are solved by using 'pure' data not 'adapted' to the assumed model by an adequate inference function (Pearson correlation coefficient)

Problem

- The reason: The score function $r(x;\theta) = (r_{\theta_1}, ..., r_{\theta_m})$, $r_{\theta_j}(x;\theta) = \frac{\partial}{\partial\theta_j} \log f(x;\theta)$, is a *vector function*, suitable for estimation of parameters, but too complicated to afford useful proposals of sensible numeric characteristics of distributions and too complicated to be used in more complex problems

Problem

- The reason: The score function $r(x;\theta) = (r_{\theta_1}, ..., r_{\theta_m})$, $r_{\theta_j}(x;\theta) = \frac{\partial}{\partial \theta_j} \log f(x;\theta)$, is a *vector function*, suitable for estimation of parameters, but too complicated to afford useful proposals of sensible numeric characteristics of distributions and too complicated to be used in more complex problems
- The problem: To find a relevant scalar inference function $S(x;\theta)$ reflecting basic features of the model distribution, and to use moments

$$M_k(\theta) = \int_{\mathcal{X}} S^k(x;\theta) f(x;\theta)\, dx$$

for generalized moment estimates

Location distributions

- Location distribution $g(y-\mu), \mu \in \mathbb{R}$, g unimodal, regular, with support $\mathbb{R}$
  Scalar score
  $$r_\mu(y;\mu) = \frac{\partial}{\partial\mu} \log g(y-\mu) = S_G(y-\mu)$$
  where function
  $$S_G(y) = -\frac{g'(y)}{g(y)}$$
  is obtained by differentiating according the variable

Location distributions

- Location distribution $g(y-\mu), \mu \in \mathbb{R}$, g unimodal, regular, with support $\mathbb{R}$
  Scalar score

  $$r_\mu(y;\mu) = \frac{\partial}{\partial\mu} \log g(y-\mu) = S_G(y-\mu)$$

  where function

  $$S_G(y) = -\frac{g'(y)}{g(y)}$$

  is obtained by differentiating according the variable
- Scalar score of a distribution with support $\mathbb{R}$

  $$S_G(y;\theta) = -\frac{1}{g(y;\theta)} \frac{d}{dy} g(y;\theta)$$

Log-location distributions - I

The log-location distribution (Lawless 2003) F of random variable $X = \eta^{-1}(Y)$ with support $\mathcal{X} = (0, \infty)$ has density

$$f(x; \tau) = g(u)\eta'(x),$$

where $g(y - \mu)$ is the density of 'prototype' distribution on $\mathbb{R}$,

$$u = \eta(x) - \eta(\tau)$$

and the 'log-location' parameter $\tau = \eta^{-1}(\mu)$ is the 'image' of the location μ of the prototype

Log-location distributions II

- Theorem.

$$\frac{\partial}{\partial \tau} \log f(x; \tau) = S_G(u)\eta'(\tau)$$

Log-location distributions II

- Theorem.
$$\frac{\partial}{\partial \tau} \log f(x;\tau) = S_G(u)\eta'(\tau)$$

- $$T(x;\tau) \equiv S_G(u) = -\frac{1}{f(x;\tau)} \frac{d}{dx}\left(\frac{1}{\eta'(x)} f(x;\tau)\right)$$
transformation-based score (t-score)

Log-location distributions II

- Theorem.

$$\frac{\partial}{\partial \tau} \log f(x;\tau) = S_G(u)\eta'(\tau)$$

- $$T(x;\tau) \equiv S_G(u) = -\frac{1}{f(x;\tau)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x;\tau)\right)$$

  transformation-based score (t-score)

- Scalar score

$$S_\tau(x) = \eta'(\tau)T(x;\tau)$$

Generalizations

- *F* on general interval support $\mathcal{X} \subseteq \mathbb{R}$, $\eta : \mathcal{X} \to \mathbb{R}$

Generalizations

- F on general interval support $\mathcal{X} \subseteq \mathbb{R}$, $\eta : \mathcal{X} \to \mathbb{R}$
- t-score (a general concept)

$$T(x,\theta) = -\frac{1}{f(x;\theta)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x;\theta) \right)$$

where (Johnson, 1949)

$$\eta(x) = \begin{cases} \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log \frac{(x-a)}{(b-x)} & \text{if} \quad \mathcal{X} = (a,b) \end{cases}$$

Generalizations

- F on general interval support $\mathcal{X} \subseteq \mathbb{R}$, $\eta : \mathcal{X} \to \mathbb{R}$
- t-score (a general concept)

$$T(x,\theta) = -\frac{1}{f(x;\theta)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x;\theta)\right)$$

where (Johnson, 1949)

$$\eta(x) = \begin{cases} \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log\frac{(x-a)}{(b-x)} & \text{if} \quad \mathcal{X} = (a,b) \end{cases}$$

- However, to use relation $\frac{\partial}{\partial\tau}\log f(x;\theta) = \eta'(\tau)T(x;\theta)$,

θ has to be in the form $\theta = (\eta^{-1}(\mu), \theta_2, ..., \theta_m)$

Starting point

- $$S_\tau(x;\tau) = \eta'(\tau)T(x;\tau)$$

Starting point

- $$S_\tau(x;\tau) = \eta'(\tau) T(x;\tau)$$
- Example: $f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}$

$$T(x;\tau) = x/\tau - 1 \qquad S_\tau(x;\tau) = \frac{1}{\tau}(x/\tau - 1)$$

Starting point

- $$S_\tau(x;\tau) = \eta'(\tau) T(x;\tau)$$
- Example: $f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}$

$$T(x;\tau) = x/\tau - 1 \qquad S_\tau(x;\tau) = \frac{1}{\tau}(x/\tau - 1)$$

- τ is usually taken as scale parameter, but $\tau = \eta^{-1}(\mu)$ and $T(\tau;\theta) = 0$. Perhaps the most important value is not the parameter, but the 'center' of the distribution itself

Definitions

- Measure of central tendency: t-mean

$$x^*(\theta) : \quad T(x;\theta) = 0$$

Definitions

- Measure of central tendency: t-mean

$$x^*(\theta) : \quad T(x;\theta) = 0$$

- Inference function: Scalar score

$$S(x;\theta) \equiv \eta'(x^*)T(x;\theta)$$

Definitions

- Measure of central tendency: t-mean

$$x^*(\theta): \quad T(x;\theta) = 0$$

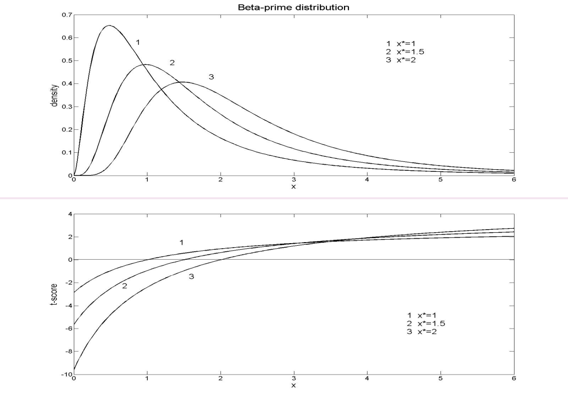
- Inference function: Scalar score

$$S(x;\theta) \equiv \eta'(x^*)T(x;\theta)$$

- $E_\theta S^2$ Fisher information for x^*

Example: Scalar scores of beta-prime distribution

$$f(x) = \frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}} \qquad T(x) = \frac{qx-p}{x+1} \qquad x^* = \frac{p}{q} \qquad S(x) = \frac{q}{p} \frac{qx-p}{x+1}$$



Consequences

- Measure of variability: Score variance: the reciprocal Fisher information

$$\omega^2(\theta) = \frac{1}{E_\theta S^2}$$

Consequences

- Measure of variability: Score variance: the reciprocal Fisher information

$$\omega^2(\theta) = \frac{1}{E_\theta S^2}$$

- 'Center' and 'radius' of the distribution

$$x^*(\theta), \omega(\theta)$$

Consequences

- Measure of variability: Score variance: the reciprocal Fisher information

$$\omega^2(\theta) = \frac{1}{E_\theta S^2}$$

- 'Center' and 'radius' of the distribution

$$x^*(\theta), \omega(\theta)$$

- Estimates: Important are not the estimates of θ, but the sample t-mean $\hat{x}^* = x^*(\hat{\theta}_{ML})$ and sample score standard deviation $\hat{\omega} = \omega(\hat{\theta}_{ML})$, which make possible to compare results for various models with different parameters

Score moment estimators

- $\hat{\theta}_{SM}$ by a generalized moment method

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1, ..., m$$

Score moment estimators

- $\hat{\theta}_{SM}$ by a generalized moment method

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1, ..., m$$

- Scalar score moment estimates are M-estimates, equations are 'simple' ($E_\theta S^k$ is often expressed by simple function of parameters)

Score moment estimators

- $\hat{\theta}_{SM}$ by a generalized moment method

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1, ..., m$$

- Scalar score moment estimates are M-estimates, equations are 'simple' ($E_\theta S^k$ is often expressed by simple function of parameters)
- Scalar scores of heavy-tailed distributions are bounded: estimates are robust

Score moment estimators

- $\hat{\theta}_{SM}$ by a generalized moment method

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1, ..., m$$

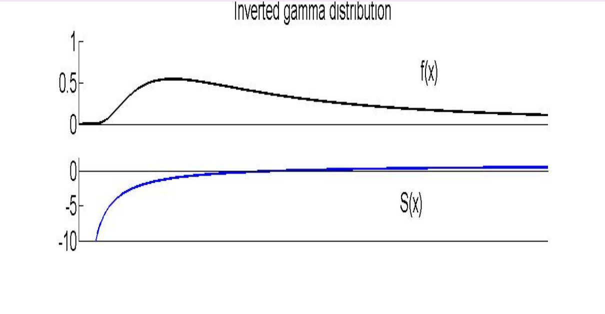
- Scalar score moment estimates are M-estimates, equations are 'simple' ($E_\theta S^k$ is often expressed by simple function of parameters)
- Scalar scores of heavy-tailed distributions are bounded: estimates are robust
- In cases of heavy-tailed distributions, estimates have asymptotic efficiences ~ 0.9.

Inverted gamma distribution

Support $(0, \infty)$, densities and t-scores

$$f(x) = \frac{\gamma^\alpha}{x\Gamma(\alpha)} x^{-\alpha} e^{-\gamma/x} \qquad T(x) = \alpha - \gamma/x$$

$$x^* = \frac{\gamma}{\alpha}, ET^2 = \alpha, \omega^2 = \frac{(x^*)^2}{ET^2} = \frac{\gamma^2}{\alpha^3}, S(x) = \frac{\alpha^2}{\gamma}(1 - x^*/x)$$

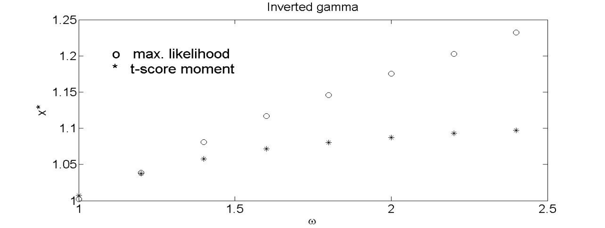


Estimation

$$\sum_{i=1}^{n} (1 - x^*/x_i) = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} (1 - x^*/x_i)^2 = \alpha$$

$\hat{x}^*$ is the harmonic mean



Generalized beta family

Support $\mathcal{X} = (0, \infty)$ and densities

$$f(x;\tau,\alpha,\nu) = \frac{1}{\nu^{\alpha} B(\nu\alpha, \alpha)} \frac{(x/\tau)^{\nu\alpha-1}}{[(x/\tau) + 1/\nu]^{(1+\nu)\alpha}}$$

where B is the beta function. The t-score is

$$T(x;\tau;\alpha,\nu) = \alpha \frac{(x/\tau) - 1}{(x/\tau) + 1/\nu}$$

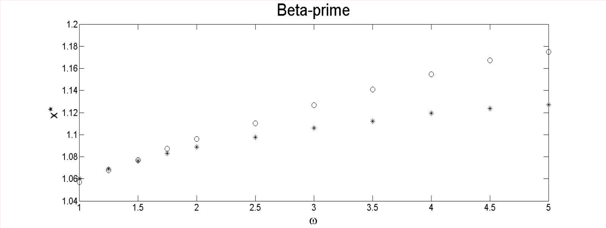
The first three t-score moments $ET = 0$, $ET^2 = \frac{\nu}{(\nu+1)\alpha+1}$ $ET^3 = \frac{2\nu(1-\nu)}{[(\nu+1)\alpha+1][(\nu+1)\alpha+2]}$ are independent of τ

Generalized beta family, $\tau = 1$

By setting $\tau = 1$ we obtain equations

$$\hat{\nu}: \quad \sum_{i=1}^{n} \frac{x_i - 1}{x_i + 1/\nu} = 0$$

and $\hat{\alpha} = (\hat{\nu}/\rho - 1)/(\hat{\nu} + 1)$, where $\rho = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - 1}{x_i + 1/\hat{\nu}} \right)^2$



Estimation of the Threshold Parameter

Uniform distribution on $[0, \gamma]$. ML estimator is $\hat{\gamma}_{ML} = x_{(n)}$. The t-score is

$$T(x) = \frac{2x}{\gamma} - 1,$$

so that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{2x_i}{\gamma} = 1$$

The score moment solution

$$\hat{\gamma}_{SM} = max(x_{(n)}, 2\bar{x})$$

For $n = 5$, 10, 20 and 50 we obtained after 10 000 experiments $\hat{\gamma}_{ML} \approx 0.87$, 0.91, 0.95 and 0.98, respectively, whereas $\hat{\gamma}_{SM} = 1$ with accuracy to three decimal points

Confidence intervals

for $\hat{x}^*_{SM}$ can be established by the modification of the Rao score test or by the use of the distance

$$d(\hat{x}^*_{SM}, x_0) = \frac{|S(\hat{x}^*_{SM}) - S(x_0)|}{ES^2}$$

As

$$\omega^2 = \frac{1}{ES^2} = \frac{(x^*)^2}{ET^2}$$

$$\hat{\omega} = \frac{\hat{x}^*_{SM}}{\left[\frac{1}{n}\sum_{i=1}^n T^2(x_i; \hat{x}^*_{SM})\right]^{1/2}}$$

References

Fabián, Z. (2001). Induced cores and their use in robust parametric estimation. Comm. in Statist. Theory Methods 30, 537-556.

Fabián, Z. (2008). New measures of central tendency and variability of continuous distributions. Comm. Statist. Theory Methods 37, 159-174.

Fabián, Z., Stehlík, M. (2008). A note on favorable estimation when data is contaminated. Comm. Dep. and Quality Management 11, 36-43.

Fabián, Z. (2009). Confidence intervals for a new characteristic of central tendency of distributions. Comm. Statist. Theory Methods 38, 1804-1814.