Kullback–Leibler 情報量を変換し境界値問題と見て重み付き残差法で母数推定する

Takuya KAWANISHI

Kanazawa University

September 7, 2022

Introduction

- When the support of the distribution function depend on the parameter, MLE may not exist or is not consistent. For example, for the three-parameter Weibull (TPW) distribution, when the shape parameter is less than unity, no MLE exists (e.g., Smith, 1985).
- In such cases, the maximum product of spacings (MPS) method (Cheng and Amin, 1983; Ranneby 1984) can be used.
- Jiang (2013) proposed a modification of MPS (Jiang's modified MPS, JMMPS) which reduces the bias of the shape and scale parameters of the TPW. But the method has not attracted much attention.
- Kawanishi (2020) found that the MLE, MPS and JMMPS can be derived from the same framework: Transforming KLD to integral over the closed interval, (then minimizing the KLD can be seen a boundary value problem), and then applying the method of weighted residuals.

Why transformed KLD? Why MWR?

Transformed KLD

- Integral over a compact set $[0, 1]$ is easier to handle than the integral over entire $\mathbb{R}$.
- homogeneous partition $p_0 = 0, p_1, \ldots, p_{n+1} = 1$ of the interval $[0, 1]$.

MWR

- Various method under the theory of MWR can be applied;
- the Galerkin method leads to the formation of Jiang's modified MPS (JMMPS);
- the ML, MPS and JMMPS can be constructed under one theory.

Continuous empirical quantile function

We define $Q^{\mathrm{E}}(p, \theta) \mapsto x$ as follows.

(a) As the values of Q^{E} at points $0, p_1, \ldots, p_n, 1$,

$$
\begin{aligned}
Q^{\mathrm{E}}(0, \theta) &= {}_*x; \\
Q^{\mathrm{E}}(p_i, \theta) &= X_{(i)}, \quad p_i = \frac{i}{n+1}, \quad i = 1, \ldots n; \\
Q^{\mathrm{E}}(1, \theta) &= x_*;
\end{aligned}
$$

where ${}_*x = \inf\{x : F(x, \theta) > 0\}$, $x_* = \sup\{x : F(x; \theta) < 1\}$.

(b) the function Q^{E} is continuous with respect to p;

(c) on each interval (p_{i-1}, p_i), $i = 1, \ldots, n+1$, Q^{E} is twice continuously differentiable with respect to p and θ;

(d) fix θ, Q^{E} is an increasing function of p.

Transformed KLD

- Transformed continuous edf

$$\tilde{F}^{\mathrm{E}} := F \circ (Q^{\mathrm{E}}, \pi^{\theta}), \quad (\pi^{\theta} : (p, \theta) \mapsto \theta \text{ is a projection.}) \tag{1}$$

- Let F_0 be the true distribution and let

$$F^{\mathrm{E}} := \tilde{F}^{\mathrm{E}} \circ (F_0, \pi^{\theta}), \quad f^{\mathrm{E}} = \frac{\partial F^{\mathrm{E}}}{\partial x}(x, \theta), \tag{2}$$

we have

$$\frac{f_0}{f^{\mathrm{E}}} = \frac{1}{\frac{\partial \tilde{F}^{\mathrm{E}}}{\partial p}}$$

- Transformed KLD

$$-D_{\mathrm{KL}}(\theta, \theta_0) = -\int_{\infty}^{\infty} \log\left(\frac{f_0}{f^{\mathrm{E}}}\right) f_0\, dx \boxed{= \int_0^1 \log\left(\frac{\partial \tilde{F}^{\mathrm{E}}}{\partial p}\right) dp} \tag{3}$$

Minimizing KLD as a BVP

- To find the minimizer, we look for the stationary points.

$$0 = -\frac{\partial}{\partial\theta} D_{\mathrm{KL}} = \int_0^1 \frac{\tilde{F}^{\mathrm{E}}_{p\theta}}{\tilde{F}^{\mathrm{E}}_{p}}\, dp = \sum_{i=1}^{n+1} \int_{p_{i-1}}^{p_i} \frac{\tilde{F}^{\mathrm{E}}_{\theta p}}{\tilde{F}^{\mathrm{E}}_{p}}\, dp. \tag{4}$$

- Boundary conditions:

$$\tilde{F}^{\mathrm{E}}_{\theta}(0;\theta) = \tilde{F}^{\mathrm{E}}_{\theta}(1;\theta) = 0$$

  - But $\lim_{p\downarrow 0} F^{\mathrm{E}}_{\theta}(p;\theta)$ or $\lim_{p\uparrow 1} F^{\mathrm{E}}_{\theta}(p;\theta)$ can be nonzero, including infinity.

KLDMWR–collocation estimator (CE)

- In the collocation method of MWR, we use the function itself as the trial function and use Dirac's delta function as the weighting function.

$$(R, w) = \sum_{i=1}^{n} \left. \frac{\tilde{F}^{\mathrm{E}}_{\theta p}}{\tilde{F}^{\mathrm{E}}_{p}} \right|_{p_i} = \sum_{i=1}^{n} \frac{f_\theta(X_{(i)})}{f(X_{(i)})} = 0, \tag{5}$$

which is nothing but the likelihood equation.

  - $X_{(i)}$ represents the order statistics of the sample X.

- **The collocation method leads to the ML.**

KLDMWR–subdomain estimator (SE)

- In the subdomain method, the trial function is a simple function: constant over each interval; the weighting functions are indicator function of the interval. The trial functions are

$$\frac{\tilde{F}^{\mathrm{E}}_{\theta p}}{\tilde{F}^{\mathrm{E}}_{p}} = \frac{F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})}{F(X_{(i)}) - F(X_{(i-1)})}$$

- Making the weighted residual zero $(R, w) = 0$ yields

$$\begin{aligned}(R, w) &= \frac{1}{n+1}\sum_{i=1}^{n+1} \frac{F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})}{F(X_{(i)}) - F(X_{(i-1)})} \\ &= \frac{1}{n+1}\sum_{i=1}^{n+1} \frac{D_{\theta,i}}{D_i} = 0 \qquad (6)\end{aligned}$$

- **The subdomain method leads to the maximum product of spacings (MPS) method.**

KLDMWR–Galerkin estimator (GE) 1

- We use the same basis function for trial and weighting functions.
- Basis function $i = 0, \ldots, n+1$,

$$\phi_i = \begin{cases} \dfrac{y - p_{i-1}}{p_i - p_{i-1}} & p_{i-1} \le y \le p_i, \ (i \ge 1); \\ \dfrac{p_{i+1} - y}{p_{i+1} - p_i} & p_i \le y \le p_{i+1}, \ (i \le n). \end{cases} \tag{7}$$

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}p} \right)_i = \begin{cases} \dfrac{1}{p_i - p_{i-1}} & p_{i-1} \le y \le p_i, \ (i \ge 1); \\ \dfrac{-1}{p_{i+1} - p_i} & p_i \le y \le p_{i+1}, \ (i \le n). \end{cases} \tag{8}$$

KLDMWR–Galerkin estimator (GE) 2

- Trial functions

$$\left(\tilde{F}^{\mathrm{E}}_{p}\right)^{h} = \sum_{i=0}^{n+1} \tilde{F}^{\mathrm{E}}(p_i)\left(\frac{\partial \phi_i}{\partial p}\right)$$

$$\left(\tilde{F}^{\mathrm{E}}_{\theta p}\right)^{h} = \sum_{i=0}^{n+1} \tilde{F}^{\mathrm{E}}_{\theta}(p_i)\left(\frac{\partial \phi_i}{\partial p}\right)$$

- Weighting function

$$w(p) = \sum_{i=1}^{n} \phi_i(p) \tag{9}$$

enditemize

- This weighting function satifies the boundary conditions $w(0) = w(1) = 0$. (Note $\tilde{F}^{\mathrm{E}}_{\theta}(0;\theta) = \tilde{F}^{\mathrm{E}}_{\theta}(1;\theta) = 0$.)

KLDMWR–Galerkin estimator (GE) 3

- Let the weighting function be a linear combination of ϕ_i, then

$$\int_{p_{i-1}}^{p_i} \left(\frac{\partial \tilde{F}_\theta^{\mathrm{E}}}{\partial p} \frac{\partial p}{\partial \tilde{F}^{\mathrm{E}}} \right)^h w \, \mathrm{d}t$$

$$= \int_{p_{i-1}}^{p_i} \frac{\tilde{F}_{\theta,i}^{\mathrm{E}} - \tilde{F}_{\theta,i-1}^{\mathrm{E}}}{p_i - p_{i-1}} \frac{p_i - p_{i-1}}{\tilde{F}_i^{\mathrm{E}} - \tilde{F}_{i-1}^{\mathrm{E}}} \left\{ \frac{t - p_{i-1}}{p_i - p_{i-1}} + \frac{p_i - t}{p_i - p_{i-1}} \right\} dt$$

$$= (p_i - p_{i-1}) \frac{\tilde{F}_{\theta,i}^{\mathrm{E}} - \tilde{F}_{\theta,i-1}^{\mathrm{E}}}{\tilde{F}_i^{\mathrm{E}} - \tilde{F}_{i-1}^{\mathrm{E}}} = \frac{1}{n+1} \frac{F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})}{F(X_{(i)}) - F(X_{(i-1)})};$$

$$\int_0^{p_1} \left(\frac{\partial \tilde{F}_\theta^{\mathrm{E}}}{\partial p} \frac{\partial p}{\partial \tilde{F}^{\mathrm{E}}} \right)^h w \, dt = \frac{1}{2(n+1)} \frac{F_\theta(X_{(1)})}{F(X_{(1)})};$$

$$\int_{p_n}^{1} \left(\frac{\partial \tilde{F}_\theta^{\mathrm{E}}}{\partial p} \frac{\partial p}{\partial \tilde{F}^{\mathrm{E}}} \right)^h w \, dt = \frac{1}{2(n+1)} \frac{-F_\theta(X_{(n)})}{1 - F(X_{(n)})}.$$

KLDMWR–Galerkin estimator (GE) 4

- Making the weighted residual zero,

$$\left(\left(\frac{\partial \tilde{F}_\theta^{\mathrm{E}}}{\partial p} \frac{\partial p}{\partial \tilde{F}^{\mathrm{E}}} \right)^h, w \right) = 0,$$

  yields

$$(R, w) = \frac{1}{n+1} \left(\frac{1}{2} \frac{D_{\theta,1}}{D_1} + \sum_{i=2}^{n} \frac{D_{\theta,i}}{D_i} + \frac{1}{2} \frac{D_{\theta,n+1}}{D_{n+1}} \right) = 0 \qquad (10)$$

  - where $D_i := F(X_{(i)}) - F(X_{(i-1)})$, $D_{\theta,i} := F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})$
  - and $F(X_{(0)}) = F_\theta(X_{(0)}) = F_\theta(X_{(n+1)}) = 0$, $F(X_{(n+1)}) = 1$,
- **The Galerkin method leads to Jiang's modified maximum product of spacings (JMMPS) method.**

Advantages of MPS and JMMPS over ML

- Applicable for nonregular cases.
  - Let us assume that Q^{E} is linear in each interval $[p_{i-1}, p_i)$, a modest assumption. Then for a fixed θ,

$$\int_0^{p_1} \frac{\tilde{F}^{\mathrm{E}}_{\theta p}}{\tilde{F}^{\mathrm{E}}_{p}}\, dp = \frac{1}{\tilde{F}^{\mathrm{E}}_{p}} \left\{ \tilde{F}^{\mathrm{E}}_{\theta}(p_1;\theta) - \underbrace{\tilde{F}^{\mathrm{E}}_{\theta}(0;\theta)}_{=0} \right\}$$

$$\int_{p_n}^{1} \frac{\tilde{F}^{\mathrm{E}}_{\theta p}}{\tilde{F}^{\mathrm{E}}_{p}}\, dp = \frac{1}{\tilde{F}^{\mathrm{E}}_{p}} \left\{ \underbrace{\tilde{F}^{\mathrm{E}}_{p}(1;\theta)}_{=0} - \tilde{F}^{\mathrm{E}}_{\theta}(p_1;\theta) \right\}$$

  - Even if f_θ/f is unbounded, $\int_{p_{i-1}}^{p_i} \tilde{F}^{\mathrm{E}}_{\theta p}/\tilde{F}^{\mathrm{E}}_{p}\, dp$ can be calculated: $\tilde{F}^{\mathrm{E}}_{\theta}$ is usually bounded.

As boundary value problems

- Problem of the dependence of distribution function on the parameter is converted to the singularity in boundary values.
- For example, for the three parameter Weibull (TPW) distribution, ML does not work when the shape parameter $\xi < 1$. In that case, the boundary value is

$$\tilde{F}^{\mathrm{E}}_{\theta}(0, \theta) = 0$$
$$\left| \lim_{p \downarrow 0} \tilde{F}^{\mathrm{E}}_{\theta}(p, \theta) \right| = \infty$$

- BVP in statistics? The boundary conditions are the properties of distribution functions.

Conclusions and future work

- With the KLDMWR framework we deal with the estimators, (at least) MLE, MPSE, and JMMPSE under the same theory.
- The problem of the dependence of the support of distribution on the parameter is converted to the singularity of the boundary conditions.
- Effects of the scheme of solving BVP sample performance of estimators for small samples are to be investigated.
- Advantages and disadvantages of the estimators under the light of the methods for BVP are to be explored.

References

R. C. H. Cheng and N. a. K. Amin, Journal of the Royal Statistical Society: Series B (Methodological) 45, 394 (1983).

M. Ekström, Journal of Statistical Planning and Inference 138, 1778 (2008).

T. Kawanishi, Comp. Appl. Math. 39, (2020).

B. Ranneby, Scandinavian Journal of Statistics 11, 93 (1984).

R. L. SMITH, Biometrika 72, 67 (1985).