M3S3/M4S3 STATISTICAL THEORY II

WORKED EXAMPLE: EFFICIENT INFERENCE FOR THE CAUCHY DISTRIBUTION

Suppose that $X_1, \ldots, X_n$ are i.i.d random variables having a Cauchy distribution with pdf

$$f_{X|\theta}(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} \qquad x \in \mathbb{R}$$

for parameter $\theta \in \mathbb{R}$.

(i) Show that the Fisher Information for θ is $I(\theta) = 1/2$.

(ii) Find the an asymptotic approximation to the distribution of the **sample median**, M_n, that is the quantile corresponding to $p = 1/2$, and hence comment on the efficiency of M_n in estimating θ.

(iii) Using the **one-step estimation** approach, construct an **efficient** estimator of θ, stating its asymptotic variance.

(iv) Find the an asymptotic approximation to the distribution of the length, L_n, of the central 95 % sample interval, that is

$$L_n = X_{(k_2)} - X_{(k_1)}$$

where $k_1 = \lceil np_1 \rceil$, $k_2 = \lceil np_2 \rceil$ for $p_1 = 0.025$ and $p_2 = 0.975$, $\lceil x \rceil$ denotes the smallest integer not less than x, and $X_{(k)}$ denotes the kth order statistic.

SOLUTION (i) The Fisher information is computed in the usual way.

$$\begin{aligned}
l(\theta) &= -\log \pi - \log(1+(x-\theta)^2) \\
\dot{l}(\theta) &= \frac{2(x-\theta)}{1+(x-\theta)^2} \\
\ddot{l}(\theta) &= -2\frac{1-(x-\theta)^2}{(1+(x-\theta)^2)^2}
\end{aligned}$$

Thus

$$\begin{aligned}
I(\theta) = E_{f_{X|\theta}}[-\ddot{l}(\theta)] &= 2\int_{-\infty}^{\infty} \frac{1-(x-\theta)^2}{(1+(x-\theta)^2)^2}\frac{1}{\pi}\frac{1}{1+(x-\theta)^2}\,dx \\
&= \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{1-(x-\theta)^2}{(1+(x-\theta)^2)^3}\,dx = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{1-x^2}{(1+x^2)^3}\,dx \\
&= \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1-\tan^2 t}{(1+\tan^2 t)^3}\,\frac{dx}{dt}\,dt \qquad x = \tan t \\
&= \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1-\tan^2 t}{(1+\tan^2 t)^3}\,\sec^2 t\,dt = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1-\tan^2 t}{(\sec^2 t)^3}\,\sec^2 t\,dt \\
&= \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1-\tan^2 t}{(\sec^2 t)^2} t\,dt = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \left(1-\frac{\sin^2 t}{\cos^2 t}\right)\cos^4 t\,dt \\
&= \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} \left(\cos^4 t - \sin^2 t\cos^2 t\right)\,dt \\
&= \frac{8}{\pi}\int_{0}^{\pi/2} \cos^4 t\,dt - \frac{4}{\pi}\int_{0}^{\pi/2} \cos^2 t\,dt
\end{aligned}$$

By a standard result

$$
\begin{aligned}
C(k) &= \int_0^{\pi/2} \cos^k t \, dt = \int_0^{\pi/2} \cos t \, \cos^{(k-1)} t \, dt \\
&= \Big[\sin t \cos^{(k-1)} t \Big]_0^{\pi/2} + (k-1) \int_0^{\pi/2} \sin^2 t \, \cos^{(k-2)} t \, dt \\
&= 0 + (k-1) \int_0^{\pi/2} (1-\cos^2 t) \, \cos^{(k-2)} t \, dt \\
&= (k-1) \int_0^{\pi/2} \cos^{(k-2)} t \, dt - (k-1) \int_0^{\pi/2} \cos^k t \, dt \\
&= (k-1) C(k-2) - (k-1) C(k)
\end{aligned}
$$

and hence

$$
C(k) = \left(\frac{k-1}{k} \right) C(k-2)
$$

Thus, as $C(0) = \pi/2$, $C(2) = \pi/4$ and $C(4) = 3\pi/16$, so

$$
I(\theta) = \frac{8}{\pi} \times \frac{3\pi}{16} - \frac{4}{\pi} \times \frac{\pi}{4} = \frac{1}{2}
$$

as required.

(ii) M_n is the $p = 1/2$ quantile, and by the standard result for sample quantiles

$$
\sqrt{n} \, (M_n - \theta) \xrightarrow{\mathfrak{L}} N(0, \sigma_\theta^2)
$$

as the (true) distribution median is θ (as the pdf is symmetric around θ), and where

$$
\sigma_\theta^2 = \frac{p(1-p)}{\{f_{X|\theta}(\theta)\}^2} = \frac{1/4}{\{1/\pi\}^2} = \frac{\pi^2}{4} \approx 2.467
$$

Thus the estimator M_n is **inefficient**, as the lowest possible variance given by $I(\theta) = 1/2$ is $I(\theta)^{-1} = 2$. Note that an efficient estimator can be found as the solution to the likelihood equations $\dot{l}_n(\theta) = 0$, or equivalently

$$
\sum_{i=1}^n \frac{(x_i - \theta)}{1 + (x_i - \theta)^2} = 0
$$

but this is not straightforward. Note also that the other natural estimator, the sample mean, is not useful, as it has intractable asymptotic behaviour (the expectation of the Cauchy does not exist).

(iii) The one-step approach to efficient estimation computes one of the following two estimators

$$
\widehat{\theta}^{(1)} = \widetilde{\theta}_n - \left(\ddot{\boldsymbol{l}}_n(\widetilde{\theta}_n) \right)^{-1} \dot{\boldsymbol{l}}_n(\widetilde{\theta}_n) \tag{N}
$$

and

$$
\widehat{\theta}^{\star} = \widetilde{\theta}_n + \left(I(\widetilde{\theta}_n) \right)^{-1} \frac{1}{n} \, \dot{\boldsymbol{l}}_n(\widetilde{\theta}_n) \tag{S}
$$

based on the consistent estimators $\widetilde{\theta}_n$. The one-step estimators are asymptotically equivalent to the efficient estimator, and thus have asymptotic variance 2.

Now M_n is a suitable consistent estimator of θ, and

$$\dot{l}_n(\theta) = 2\sum_{i=1}^{n} \frac{(x_i-\theta)}{1+(x_i-\theta)^2} \qquad \ddot{l}_n(\theta) = -2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{(1+(x_i-\theta)^2)^2}$$

so therefore the one-step estimators are

$$\widehat{\theta}^{(1)} = M_n + 4\left(\sum_{i=1}^{n} \frac{1-(x_i-M_n)^2}{(1+(x_i-M_n)^2)^2}\right)^{-1} \sum_{i=1}^{n} \frac{(x_i-M_n)}{1+(x_i-M_n)^2}$$

and

$$\widehat{\theta}^{\star} = M_n + \frac{4}{n}\sum_{i=1}^{n} \frac{(x_i-M_n)}{1+(x_i-M_n)^2}$$

(iv) Using the standard result from lectures, we have that

$$\sqrt{n}\left(\left(\begin{array}{c} X_{(k_1)} \\ X_{(k_2)} \end{array}\right) - \left(\begin{array}{c} x_{p_1} \\ x_{p_2} \end{array}\right)\right) \xrightarrow{\mathfrak{L}} N\left(0, \left[\begin{array}{cc} \dfrac{p_1\,(1-p_1)}{\{f_{X|\theta}\,(x_{p_1})\}^2} & \dfrac{p_1\,(1-p_2)}{f_{X|\theta}\,(x_{p_1})\,f_{X|\theta}\,(x_{p_2})} \\ \\ \dfrac{p_1\,(1-p_2)}{f_{X|\theta}\,(x_{p_1})\,f_{X|\theta}\,(x_{p_2})} & \dfrac{p_2\,(1-p_2)}{\{f_{X|\theta}\,(x_{p_2})\}^2} \end{array}\right]\right) \equiv N(0,\Sigma),$$

say, where here, $p_1 = 0.025$, $p_2 = 0.975$, and x_{p_1} and x_{p_2} are the corresponding true distribution quantiles. Now, here

$$F_{X|\theta}(x) = \int_{-\infty}^{x} f_{X|\theta}(t)\;dt = \int_{-\infty}^{x} \frac{1}{\pi}\frac{1}{1+(t-\theta)^2}\;dt = \frac{1}{\pi}\left[\arctan t\right]_{-\infty}^{x-\theta} = \frac{1}{\pi}(\arctan(x-\theta)+\pi/2)$$

so

$$\frac{1}{\pi}(\arctan(x_{p_1}-\theta)+\pi/2) = 0.025 \qquad \text{and} \qquad \frac{1}{\pi}(\arctan(x_{p_2}-\theta)+\pi/2) = 0.975$$

and hence

$$\begin{aligned} x_{p_1} &= \theta + \tan((0.025\pi)-\pi/2) = \theta - 12.706 \\ x_{p_2} &= \theta + \tan((0.975\pi)-\pi/2) = \theta + 12.706. \end{aligned}$$

Hence

$$\begin{aligned} \Sigma &= \pi^2\left[\begin{array}{cc} 0.025\,(1-0.025)\,(1+12.706^2)^2 & 0.025\,(1-0.975)\,(1+12.706^2)(1+12.706^2) \\ 0.025\,(1-0.975)\,(1+12.706^2)(1+12.706^2) & 0.975\,(1-0.975)\,(1+12.706^2)^2 \end{array}\right] \\ &= \pi^2(1+12.706^2)^2\left[\begin{array}{cc} 0.025\,(1-0.025) & 0.025\,(1-0.975) \\ 0.025\,(1-0.975) & 0.975\,(1-0.975) \end{array}\right] = \left[\begin{array}{cc} 6348.501 & 162.782 \\ 162.782 & 6348.501 \end{array}\right] \end{aligned}$$

Finally, as $L_n = X_{k_2} - X_{k_1}$, using the Delta method with function $g(t_1,t_2) = t_2 - t_1$, we have

$$\sqrt{n}(L_n - \lambda) \xrightarrow{\mathfrak{L}} N(0, \dot{\boldsymbol{g}}(x_{p_1},x_{p_2})\Sigma\dot{\boldsymbol{g}}^{\mathsf{T}})$$

where λ is the true length, that is $2\times 12.706 = 25.412$, and

$$\dot{\boldsymbol{g}}(\boldsymbol{t}) = \left[\frac{\partial g(\boldsymbol{t})}{\partial t_1}\;\frac{\partial g(\boldsymbol{t})}{\partial t_2}\right] = [-1\;1]\,.$$

Thus

$$\dot{\boldsymbol{g}}(x_{p_1},x_{p_2})\Sigma\dot{\boldsymbol{g}}^{\mathsf{T}} = 12371.44.$$

and hence

$$\sqrt{n}(L_n - 25.412) \xrightarrow{\mathfrak{L}} N(0, 12371.44).$$