SS3858B Tutorial

Feb 3, 2010

1 Normal Distribution

Consider normal distribution with parameter μ and σ^2 and density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Derive MLE, Fisher information (matrix) and determine whether these estimates attain the Cramér-Rao lower bound.

The MLE is

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{pmatrix}.$$

$$\begin{aligned} I(\mu) &= \mathrm{E}\left(\frac{\partial}{\partial\mu}\log f(X)\right)^2 \\ &= \mathrm{E}\left(-\frac{\partial}{\partial\mu}\frac{(x-\mu)^2}{2\sigma^2}\right)^2 \\ &= \mathrm{E}\left(\frac{X-\mu}{\sigma^2}\right)^2 \\ &= \frac{1}{\sigma^2}, \end{aligned}$$

or

$$\begin{aligned} I(\mu) &= -\mathrm{E}\left(\frac{\partial^2}{\partial\mu^2}\log f(X)\right) \\ &= \mathrm{E}\left(\frac{\partial^2}{\partial\mu^2}\frac{(X-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{2\sigma^2}\mathrm{E}\left(-\frac{\partial}{\partial\mu}2(X-\mu)\right) \\ &= \frac{1}{\sigma^2}. \end{aligned}$$

Thus,

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{nI(\mu)},$$

which shows that $\mathrm{Var}(\bar{X})$ attains the CR lower bound.

$$\begin{aligned}
I(\sigma^2) &= \mathrm{E}\left(\frac{\partial}{\partial\sigma^2}\log f(X)\right)^2 \\
&= \mathrm{E}\left(\frac{\partial}{\partial\sigma^2}\left(-\frac{1}{2}\log\sigma^2 - \frac{(X-\mu)^2}{2\sigma^2}\right)\right)^2 \\
&= \frac{1}{4}\mathrm{E}\left(\frac{1}{\sigma^2} - \frac{(X-\mu)^2}{\sigma^4}\right)^2 \\
&= \frac{1}{4\sigma^4}\mathrm{E}\left(\left(\frac{X-\mu}{\sigma}\right)^2 - 1\right)^2 \\
&= \frac{1}{4\sigma^4}(3-2+1) \\
&= \frac{1}{2\sigma^4},
\end{aligned}$$

or

$$\begin{aligned}
I(\sigma^2) &= -\mathrm{E}\left(\frac{\partial^2}{\partial(\sigma^2)^2}\log f(X)\right) \\
&= -\mathrm{E}\left(\frac{\partial^2}{\partial(\sigma^2)^2}\left(-\frac{1}{2}\log\sigma^2 - \frac{(X-\mu)^2}{2\sigma^2}\right)\right) \\
&= \frac{1}{2}\mathrm{E}\left(\frac{\partial}{\partial\sigma^2}\left(\frac{1}{\sigma^2} - \frac{(X-\mu)^2}{\sigma^4}\right)\right) \\
&= \frac{1}{2}\mathrm{E}\left(-\frac{1}{\sigma^4} + \frac{2(X-\mu)^2}{\sigma^6}\right) \\
&= \frac{1}{2\sigma^4}\mathrm{E}\left(\frac{2(X-\mu)^2}{\sigma^2} - 1\right) \\
&= \frac{1}{2\sigma^4}.
\end{aligned}$$

Recall the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

We know from Chapter 6 that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$

i.e.,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Thus,

$$\text{Var}(\hat{\sigma}^2) = \text{Var}\left(\frac{\sigma^2}{n} \cdot \frac{n\hat{\sigma}^2}{\sigma^2}\right) = \frac{2(n-1)}{n^2}\sigma^4 < \frac{1}{nI(\sigma^2)}.$$

Here, $\text{Var}(\hat{\sigma}^2)$ is less than the CR lower bound because $\hat{\sigma}^2$ is a biased estimator of σ^2. CR lower bound is for unbiased estimator ONLY.

Now consider S^2 which is unbiased.

$$\text{Var}(S^2) = \text{Var}\left(\frac{\sigma^2}{n-1} \cdot \frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right) = \frac{2}{n-1}\sigma^4 > \frac{1}{nI(\sigma^2)}.$$

$\text{Var}(S^2)$ does not attain the CR lower bound. But it asymptotically equals to the CR lower bound.

However, if μ is known,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

In this case,

$$\text{Var}(\hat{\sigma}^2) = \frac{1}{n}\text{Var}(X-\mu)^2 = \frac{\sigma^4}{n}\text{Var}\left(\frac{X-\mu}{\sigma}\right)^2 = \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)},$$

which achieves the CR lower bound.

To construct the Fisher information matrix, consider the off-diagonal element

$$\begin{aligned} I_{12}(\mu, \sigma^2) &= \text{E}\left(\frac{\partial}{\partial \mu}\log f(X) \cdot \frac{\partial}{\partial \sigma^2}\log f(X)\right) \\ &= \text{E}\left(\frac{X-\mu}{\sigma^2}\left(-\frac{1}{2\sigma^2} + \frac{(X-\mu)^2}{2\sigma^4}\right)\right) \\ &= 0, \end{aligned}$$

or

$$\begin{aligned} I_{12}(\mu, \sigma^2) &= -\text{E}\left(\frac{\partial^2}{\partial \mu \partial \sigma^2}\log f(X)\right) \\ &= -\text{E}\left(\frac{\partial}{\partial \sigma^2}\frac{X-\mu}{\sigma^2}\right) \\ &= 0 \end{aligned}$$

Therefore the Fisher information matrix is given by

$$I(\theta) = I(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Note that the Cramér-Rao inequality still holds for multivariate case, i.e.,

$$\text{Var}(T) \geq \frac{1}{n}I^{-1}(\theta).$$

For two matrices A and B, we say $A \geq B$ if $A - B$ is a positive semidefinite matrix. $I(\theta)$ is the variance the score function and thus it is positive definite and invertible.

2 Binomial Distribution

Suppose $X_1, \cdots, X_n$ is an iid sample from a Binomial(m,p) population, where m is known.

pmf

$$f(x) = \binom{m}{x} p^x (1-p)^{m-x}.$$

Likelihood

$$L(p) = \prod_{i=1}^{n} \binom{m}{X_i} p^{X_i} (1-p)^{m-X_i}.$$

Log likelihood

$$l(p) = \sum_{i=1}^{n} \log \binom{m}{X_i} + \log(p) \sum_{i=1}^{n} X_i + \log(1-p) \left(mn - \sum_{i=1}^{n} X_i \right).$$

$$\frac{d}{dp} l(p) = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{mn - \sum_{i=1}^{n} X_i}{1-p} \overset{\text{set}}{=} 0 \Rightarrow \hat{p} = \frac{\bar{X}}{m}.$$

$$\begin{aligned}
I(p) &= -\mathrm{E}\left(\frac{\partial^2}{\partial p^2} \log f(X) \right) \\
&= -\mathrm{E}\left(\frac{\partial^2}{\partial p^2} \left(X \log(p) + (m - X) \log(1-p) \right) \right) \\
&= -\mathrm{E}\left(\frac{\partial}{\partial p} \left(\frac{X}{p} - \frac{m-X}{1-p} \right) \right) \\
&= \mathrm{E}\left(\frac{X}{p^2} + \frac{m-X}{(1-p)^2} \right) \\
&= \frac{m}{p} + \frac{m}{1-p} \\
&= \frac{m}{p(1-p)}
\end{aligned}$$

$$\mathrm{Var}(\hat{p}) = \frac{\mathrm{Var}(X)}{n} = \frac{p(1-p)}{mn} = \frac{1}{nI(p)},$$

which shows that $\mathrm{Var}(\hat{p})$ achieves the CR lower bound.