Quantitative Risk Management
Case for Week 4

John Dodson

September 28, 2016

Cramér–Rao lower bound

The Cramér–Rao lower bound is a classic result in statistics. I provide below an outline of a proof in the multi-parameter setting. See for example [1] chapter 8.

Consider a parametric model for a random variable X with density function $f_X(\cdot)$ with unknown parameters expressed as the components of θ. Say we have an estimator $\hat{\theta}(\cdot)$. In particular $\hat{\theta}(X)$ is a random variable, but it does not have an explicit dependence on θ. Further, assume the estimator is unbiased. Therefore,

$$0 = \mathrm{E}\left[\hat{\theta}(X) - \theta\right] = \int \left(\hat{\theta}(x) - \theta\right) f_X(x)\, dx$$

Let's further assume that $f_X(x)$ and $\frac{\partial}{\partial\theta} f_X(x)$ are continuous in θ for all x in the support of X. Therefore we can distribute the gradient with respect to the parameters to get

$$0 = \int \left(\hat{\theta}(x) - \theta\right) \frac{\partial f_X(x)}{\partial \theta}\, dx - I \int f_X(x)\, dx$$

or, with some manipulation (noting that $f(x) > 0$ on the support of X by definition),

$$\int \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{f_X(x)}\right) \left(\frac{\partial \log f_X(x)}{\partial \theta} \sqrt{f_X(x)}\right) dx = I$$

Consider any fixed a and b in the parameter space. The previous result means

$$\int \left(a' \left(\hat{\theta}(x) - \theta\right) \sqrt{f_X(x)}\right) \left(\frac{\partial \log f_X(x)}{\partial \theta} \sqrt{f_X(x)}\, b\right) dx = a'b$$

This can be thought of as an inner product in the Hilbert space L^2, which means we can apply Cauchy-Schwarz to get

$$a' \left(\int \left(\hat{\theta}(x) - \theta\right) \left(\hat{\theta}(x) - \theta\right)' f_X(x)\, dx\right) a\, b' \left(\int \frac{\partial \log f_X(x)}{\partial \theta'} \frac{\partial \log f_X(x)}{\partial \theta} f_X(x)\, dx\right) b \;\geq\; \left(a'b\right)^2$$

Fisher Information

Define the Fisher information to be

$$\mathcal{I}(\theta) \triangleq \mathrm{E}\left[\frac{\partial \log f_X(X)}{\partial \theta'}\frac{\partial \log f_X(X)}{\partial \theta}\right] \tag{1}$$

$$= \mathrm{cov}\left[\frac{\partial \log f_X(X)}{\partial \theta'}\right] \tag{2}$$

$$= E\left[-\frac{\partial^2}{\partial \theta' \, \partial \theta}\log f_X(X)\right] \tag{3}$$

if the curvature of the log-likelihood is continuous on its support in the last instance (see Appendix).

With $b \triangleq \mathcal{I}^{-1}(\theta)\, a$, the previous result translates to

$$\left(a' \,\mathrm{cov}\left[\hat{\theta}(X)\right] a\right)\left(a'\mathcal{I}^{-1}(\theta)\, a\right) \geq \left(a'\mathcal{I}^{-1}(\theta)\, a\right)^2$$

So we can conclude that

$$a'\left(\mathrm{cov}\left[\hat{\theta}(X)\right] - \mathcal{I}^{-1}(\theta)\right) a \geq 0$$

for all vectors a.

This conforms to the definition of a positive semi-definite matrix, and can be written as

$$\mathrm{cov}\left[\hat{\theta}(X)\right] \geq \mathcal{I}^{-1}(\theta) \tag{4}$$

Note that the Cramér–Rao lower bound is a special case of the Kullback inequality about the relative entropy of one measure with respect to another.

Appendix

Recalling that we assumed we can apply Leibniz' Rule to distribute partials of θ and integrals over the support of X, so

$$\mathrm{E}\left[\frac{\partial \log f_X(X)}{\partial \theta'}\right] = \int \frac{\partial}{\partial \theta'} f_X(x)\, dx = \frac{\partial}{\partial \theta'}(1) = 0$$

and therefore we can verify (2):

$$\mathrm{cov}\left[\frac{\partial \log f_X(X)}{\partial \theta'}\right] = \mathrm{E}\left[\frac{\partial \log f_X(X)}{\partial \theta'}\frac{\partial \log f_X(X)}{\partial \theta}\right]$$

If we further assume that the second parameter partials of the density are continuous, then

$$\mathrm{E}\left[\frac{\frac{\partial^2}{\partial \theta' \partial \theta} f_X(X)}{f_X(X)}\right] = \frac{\partial^2}{\partial \theta' \partial \theta}(1) = 0$$

and since

$$\frac{\partial^2}{\partial \theta' \partial \theta}\log f_X(X) = \frac{\frac{\partial^2}{\partial \theta' \partial \theta} f_X(X)}{f_X(X)} - \frac{\partial \log f_X(X)}{\partial \theta'}\frac{\partial \log f_X(X)}{\partial \theta}$$

we get the curvature version (3):

$$\mathrm{E}\left[-\frac{\partial^2}{\partial\theta'\partial\theta}\log f_X(X)\right] = \mathrm{E}\left[\frac{\partial \log f_X(X)}{\partial\theta'}\frac{\partial \log f_X(X)}{\partial\theta}\right]$$

From a practical perspective, it is useful to have two options for evaluating the Fisher information. It is sometimes easier to calculate the covariance of the first partials, and sometimes easier to calculate the expected value of the second partials.

References

[1] Morris H DeGroot and Mark J. Schervish. *Probability and Statistics*. Pearson Higher Education, Boston, fourth edition, 2011.