Quantitative Risk Management
Case for Week 3

John Dodson

September 19, 2018

Cramér–Rao lower bound

The Cramér–Rao lower bound is a classic result in statistics. Below is 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 $f(\cdot)$ in terms of unknown parameters expressed as the components of θ.

Say we have an estimator, $\hat{\theta}(\cdot)$, which recall is a function of a sample that does not have any explicit dependence on θ. If the sample is a single random variable, X, then $\hat{\theta}(X)$ is a random variable. We generally cannot determine a formula for the density of $\hat{\theta}(X)$, but we can put a lower bound on its variance which is useful for characterizing the uncertainty associated with a particular estimate.

Assume that the estimator is unbiased. Therefore,

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

where $\mathcal{X}$ is the support of X.

Further assume that $f(x)$ and $\frac{\partial}{\partial\theta} f(x)$ are continuous in θ for all $x \in \mathcal{X}$. Therefore we can distribute the gradient with respect to the parameters to get

$$0 = \int_{\mathcal{X}} \left(\hat{\theta}(x) - \theta\right) \frac{\partial f(x)}{\partial\theta}\, dx - I \int_{\mathcal{X}} f(x)\, dx$$

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

$$\int_{\mathcal{X}} \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{f(x)}\right) \left(\frac{\partial \log f(x)}{\partial\theta} \sqrt{f(x)}\right) dx = I$$

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

$$\int_{\mathcal{X}} \left(a' \left(\hat{\theta}(x) - \theta\right) \sqrt{f(x)}\right) \left(\frac{\partial \log f(x)}{\partial\theta} \sqrt{f(x)}\, b\right) dx = a'b$$

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

$$a' \left(\int_{\mathcal{X}} \left(\hat{\theta}(x) - \theta\right) \left(\hat{\theta}(x) - \theta\right)' f(x)\, dx\right) a\, b' \left(\int_{\mathcal{X}} \frac{\partial \log f(x)}{\partial\theta'} \frac{\partial \log f(x)}{\partial\theta} f(x)\, dx\right) b \geq \left(a'b\right)^2$$

Fisher Information

Define the Fisher information to be

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

$$= \operatorname{cov}\left[\frac{\partial \log f(X)}{\partial \theta'}\right] \tag{1b}$$

$$= \mathrm{E}\left[-\frac{\partial^2}{\partial \theta'\,\partial \theta}\log f(X)\right] \tag{1c}$$

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

From a practical perspective, it is useful to have options for evaluating the Fisher information. It is sometimes easier to calculate the covariance of the first partials as in (1b), and sometimes easier to calculate the expected value of the second partials as in (1c).

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

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

So we can conclude that

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

for all vectors a.

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

$$\operatorname{cov}\left[\hat{\theta}(X)\right] \geq \mathcal{I}^{-1}(\theta) \tag{2}$$

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

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

and therefore we can verify (1b):

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

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

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

and since

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

we get the curvature version (1c):

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

References

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