Solution to Exercise 10.11

★10.11 Explain how to compute two different 95% confidence intervals for σ^2. One should be based on the covariance matrix estimator obtained in Exercise 10.10, and the other should be based on the original estimator (10.53). Are both of the intervals symmetric? Which seems more reasonable?

A 95% confidence interval for σ^2 based on the covariance matrix estimator obtained in Exercise 10.10 is

$$\big[\hat\sigma^2 - 1.96\hat\sigma^2\sqrt{2/n},\ \ \hat\sigma^2 + 1.96\hat\sigma^2\sqrt{2/n}\,\big]. \tag{S10.23}$$

This interval is evidently symmetric.

A 95% confidence interval for σ^2 based on the covariance matrix (10.53) may be obtained by first finding a 95% confidence interval for σ, which is

$$\big[\hat\sigma - 1.96\hat\sigma/\sqrt{2n},\ \ \hat\sigma + 1.96\hat\sigma/\sqrt{2n}\,\big],$$

and then squaring the two limit points. This yields the interval

$$\big[(\hat\sigma - 1.96\hat\sigma/\sqrt{2n}\,)^2,\ \ (\hat\sigma + 1.96\hat\sigma/\sqrt{2n}\,)^2\big], \tag{S10.24}$$

which is clearly not symmetric.

It seems plausible that the asymmetric interval (S10.24) should yield more reliable inferences than the symmetric interval (S10.23), because it is more likely that the distribution of $\hat\sigma$, rather than that of $\hat\sigma^2$, is approximately normal. This conjecture can be verified by doing a simulation experiment. We generate data from the model

$$y_t = \beta_1 + u_t, \quad u_t \sim \mathrm{N}(0, \sigma^2),$$

with $\beta_1 = 0$ and $\sigma^2 = 1$, for several values of n. The proportion of the time that the two intervals fail to cover the true value, out of 100,000 replications, is shown in Table S10.2.

Table S10.2 Coverage failure proportions for two confidence intervals

Sample size	Symmetric Interval	Asymmetric Interval
10	0.1121	0.0866
20	0.0822	0.0686
40	0.0658	0.0586
80	0.0578	0.0540
160	0.0543	0.0533
320	0.0530	0.0525
640	0.0513	0.0510

These experimental results strongly support the conjecture that the asymmetric interval (S10.24) yields more reliable inferences. The difference is striking in very small samples, where both intervals fail to cover the true value as often as they should, but it becomes negligible for large samples, where both intervals perform very well.