Universität Bonn
Research Module Econometrics & Statistics

Prof. Dominik Liebl
Christopher Walsh

MSE, Bias^2, and Variance of the ML-Estimator of the Standard Deviation

SETUP:

- Assumptions: iid random sample $(X_1, \dots, X_n)$ with $X_i \sim N(\mu, \sigma^2)$ for all $i = 1, \dots, n$.
- Fix values for the mean μ, the variance σ^2, and the sample size n.
  For instance: $\mu = 3$, $\sigma = 1.5$, and $n = \{2, 4, 6, \dots, 30\}$.

MONTE-CARLO ALGORITHM:

1. Simulate a realization from the iid random sample $(X_1, \dots, X_n)$.
2. Compute $s_n = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$

Repeat Steps 1-2 a large number of times, e.g., $R = 10,000$ times and save all estimates $s_{n,1}, \dots, s_{n,R}$. Then approximate the Mean Squared Error (MSE), the squared bias (Bias^2), and the variance (Var) by

$$\begin{aligned}
\text{MSE}(s_n) &\approx \frac{1}{R}\sum_{r=1}^R (s_{n,r} - \sigma)^2 \\
\text{Bias}^2(s_n) &\approx \left(\left(\frac{1}{R}\sum_{r=1}^R s_{n,r}\right) - \sigma\right)^2 \\
\text{Var}(s_n) &\approx \frac{1}{R}\sum_{r=1}^R \left(s_{n,r} - \left(\frac{1}{R}\sum_{r=1}^R s_{n,r}\right)\right)^2
\end{aligned}$$

where $E(s_{n,r}) \approx R^{-1}\sum_{r=1}^R s_{n,r}$. (The *Law of Large Numbers* implies that the approximations become arbitrarily precise for $R \to \infty$.)

PRESENTATION OF THE SIMULATION-RESULTS:

- Plot your results (y-axis: MSE, Bias^2, and Var; x-axis: n)
- Add the corresponding results for the square root of the unbiased variance estimator $\tilde{s}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$.

Additional simulation to illustrate an aspect from estimation theory:

This is an open question. We would like you to think of one simple simulation setting to illustrate a result from estimation theory. Implement the corresponding simulation and then present your result.

Some brief ideas in case nothing comes to mind:

(i) In the setting above it is known that

$$Z_n := (n-1)\frac{s_n^2}{\sigma^2}$$

is Chi-squared distributed with $n-1$ degrees of freedom. For each fixed sample size n in your simulation study plot the empirical distribution of

$$\frac{Z_n - (n-1)}{\sqrt{2(n-1)}}.$$

What happens as you increase n?

(ii) Can you come up with a simple simulation study to illustrate the delta method? For instance you could implement the example from the Section 2.6.3 of the notes.

(iii) Similarly to the previous, simulate some iid data $(X_1, \ldots, X_n)$ from a distribution of your choice with a zero mean and finite variance then illustrate the CLT by looking at the empirical distributions of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{n}\bar{X}_n$$

over your simulations for different sample sizes (n). For a given function $g(x)$, what does the empirical distribution of

$$\sqrt{n}g(\bar{X}_n)$$

look like does this match the "prediction" from the delta method? Take say $g(x) = x^2$. What happens if you choose $g(x) = 1/x$ in this setting?

(iv) Simulate a very simple linear regression model to illustrate the bias, variance and MSE of the OLS estimator – maybe even the asymptotic normality.

(v) Simulate some iid data $(X_1, \ldots, X_n)$ from a distribution of your choice with a mean (μ) and a finite variance (σ^2), then illustrate the CLT by looking at the empirical distributions of

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)}{\sigma} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$$

over your simulations for different sample sizes n. How do the empirical distributions of $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$ compare to those of

$$\sqrt{n}\frac{\bar{X}_n - \mu}{s_n}$$

with $s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$.

The above are all asymptotic in "flavour". This need not be if you want to show a result from your previous econometrics courses. The main point is that it should include a simulation to illustrate a known theoretical result.