Notes on ratio estimators and the delta-method

Let \((X_i, Y_i)\) be IID pairs of positive random variables, each variable having several moments. Let

\[ R = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}, \]

a ratio estimator. We seek the asymptotic mean and variance of \(R\). Let \(\xi_i = [X_i - E(X_i)]/E(X_i)\) and \(\eta_i = [Y_i - E(Y_i)]/E(Y_i)\), so that \(E(\xi_i) = E(\eta_i) = 0\), while

\[ R = \frac{E(Y_i)}{E(X_i)} \frac{1 + \bar{\eta}}{1 + \bar{\xi}}, \]

where the overline denotes sample average. For instance,

\[ \bar{\eta} = \frac{1}{n} \sum_{i=1}^{n} \eta_i. \]

Thus,

\[ E(R) = \frac{E(Y_i)}{E(X_i)} \left( \frac{1 + \bar{\eta}}{1 + \bar{\xi}} \right) \]

and

\[ \text{var}(R) = \frac{E(Y_i)^2}{E(X_i)^2} \text{var}\left( \frac{1 + \bar{\eta}}{1 + \bar{\xi}} \right). \]

Technically, of course, \(R\) may have an infinite variance, or even an infinite mean, e.g., the denominator might have a positive density near 0. We proceed informally, in this respect among others. For \(n\) large, \(\bar{\xi} \approx \eta \approx 0\). Thus,

\[ \frac{1 + \bar{\eta}}{1 + \bar{\xi}} \approx 1 + \bar{\eta} - \bar{\xi}. \]

This is a one-term Taylor series expansion, called the “delta-method”; for rigor, the remainder term would have to be bounded. Now

\[ E(\bar{\eta} - \bar{\xi}) = 0 \]

while

\[ \text{var}(1 + \bar{\eta} - \bar{\xi}) = \frac{1}{n} \text{var}(\eta_i - \xi_i). \]

Thus, \(R\) is asymptotically unbiased for \(E(Y_i)/E(X_i)\). The asymptotic variance is

\[ \frac{1}{n} \frac{E(Y_i)^2}{E(X_i)^2} \text{var}(\eta_i - \xi_i). \]

To operationalize (9), we can estimate \(E(Y_i)/E(X_i)\) by \(R\), and \(\text{var}(\eta_i - \xi_i)\) by

\[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \bar{Y}}{\bar{Y}} - \frac{X_i - \bar{X}}{\bar{X}} \right)^2. \]
We turn now to asymptotic bias. This requires an additional term in the expansion:

\[
\frac{1 + \eta}{1 + \xi} = (1 + \eta)(1 - \xi + \xi^2) = 1 + \eta - \xi + \xi^2 - \eta\xi.
\]  

(11)

The last two terms in the display are responsible for the asymptotic bias:

\[
E(\overline{\xi^2}) = \text{var}(\overline{\xi}) = \frac{1}{n} \frac{\text{var}(X_i)}{[E(X_i)]^2}
\]

\[
E(\overline{\eta\xi}) = \frac{1}{n} \frac{\text{cov}(X_i, Y_i)}{E(X_i)E(Y_i)}.
\]

(12)

(13)

The asymptotic bias is therefore

\[
\frac{1}{n} \frac{E(Y_i)}{E(X_i)} \left( \frac{\text{var}(X_i)}{[E(X_i)]^2} - \frac{\text{cov}(X_i, Y_i)}{E(X_i)E(Y_i)} \right).
\]

(14)

As before, \( E(X_i) \) can be estimated by the sample mean of the \( X_i \)'s, while \( \text{var}(X_i) \) is estimated by the sample variance, \( \text{cov}(X_i, Y_i) \) by the sample covariance, and so forth.