CRP SANTÉ CENTRE DE RECHERCHE PUBLIC

The variance of the ratio of two random variables

Introduction

We suppose that we have two random variables *X*, *Y* and we are interested in their ratio *Z=Y/X*. Note that it makes no sense to calculate this unless the mean of *X* is high compared to its variance because otherwise the ratio will be unstable. In particular if the denominator, *X*, of *Z* can cross zero the moments are undefined.

We suppose that the random variables have the following moments

$E[X]=\mu_X$

$V[X]=\sigma^2_X$

$E[Y]=\mu_Y$

$V[Y]=\sigma^2_Y$

$Cov[X,Y]=\sigma_{XY}$

Delta method

The approach to use is the so-called *delta method.* This uses a Taylor's expansion to represent a random change in *Z, ΔZ* as a function of random changes in *Y,ΔY* and *X,ΔX*. This technique is often used for derived variables.

$$\frac{d}{dX}\frac{Y}{X} \rightarrow -\frac{Y}{X^2} \qquad \frac{d}{dY}\frac{Y}{X} \rightarrow \frac{1}{X}$$

Thus for ΔZ we have

$$-\frac{Y}{X^2}\Delta X + \frac{1}{X}\cdot\Delta Y$$

The variance of Z however is $E[(\Delta Z)^2]$

$$\left(-\frac{Y}{X^2}\Delta X + \frac{1}{X}\cdot\Delta Y\right)^2 \text{expand} \rightarrow \frac{\Delta Y^2}{X^2} + \frac{Y^2\cdot\Delta X^2}{X^4} - \frac{2\cdot Y\cdot\Delta X\cdot\Delta Y}{X^3}$$

We further assume that $E(X^2)$ is approximately equal to $\mu_X{}^2$ etc

Thus we have finally

$$\mathrm{Var}_Z(\sigma_X,\sigma_Y,\mu_X,\mu_Y,\sigma_{XY}) := \frac{\sigma_Y{}^2}{\mu_X{}^2} + \frac{\mu_Y{}^2\cdot\sigma_X{}^2}{\mu_X{}^4} - \frac{2\cdot\mu_Y\cdot\sigma_{XY}}{\mu_X{}^3}$$

Note. The above formula has to satisfy *dimensional analysis.* The ratio *Z* is in the units of *Y* upon X. The variance of this ratio thus has to be in the units of Y^2 upon X^2. If you study the above formula you will see that for each of the three major terms this is the case:

First term units Y^2/X^2,
Second term units $Y^2X^2/X^4=Y^2/X^2$,
Third term units $YXY/X^3=Y^2/X^2$.

Example

Set parameters

$\mu_X := 30$ $\mu_Y := 50$ $\sigma_{XY} := 2$ $\sigma_Y := 4$ $\sigma_X := 3$

$m := 1000000$ Number of simulations

Some intermediate calculations

Calculate conditional variance of *Y* given *X*. This is the variance of $Y - (\alpha+\beta X)$, where β is the regression of *Y* on *X*. However α is a constant and $\beta=\sigma_{XY}/\sigma^2_X$. Thus we have the variance of $\varepsilon=Y-X\sigma_{XY}/\sigma^2_X$. Call this variance V_ε.

$$V_\varepsilon := \sigma_Y{}^2 + \frac{\sigma_X{}^2\cdot\sigma_{XY}{}^2}{\sigma_X{}^4} - 2\cdot\frac{\sigma_{XY}{}^2}{\sigma_X{}^2}$$

$$V_\varepsilon := \sigma_Y{}^2 + \frac{\sigma_{XY}{}^2}{\sigma_X{}^2} - 2\cdot\frac{\sigma_{XY}{}^2}{\sigma_X{}^2} \qquad \sigma_\varepsilon := \sqrt{V_\varepsilon}$$

$$\beta := \frac{\sigma_{XY}}{\sigma_X{}^2} \qquad \alpha := \mu_Y - \beta\cdot\mu_X$$

Carry out simulation

$X := \mathrm{rnorm}(m, \mu_X, \sigma_X)$ — Simulate X

$\varepsilon := \mathrm{rnorm}(m, 0, \sigma_\varepsilon)$ — Simulate conditional error in *Y* given *X*

$Y := \alpha + \beta \cdot X + \varepsilon$ — Construct *Y* from *X* and the error

Check simulation has produced statistics with reasonable values

$\mu_X = 30$ $\quad \mathrm{mean}(X) = 30.002$

$\mu_Y = 50$ $\quad \mathrm{mean}(Y) = 50.001$

$\sigma_X = 3$ $\quad \sqrt{\mathrm{var}(X)} = 2.997$

$\sigma_Y = 4$ $\quad \sqrt{\mathrm{var}(Y)} = 4.003$

$\sigma_{XY} = 2$ $\quad \mathrm{corr}(X, Y) \cdot \sqrt{\mathrm{var}(X)} \cdot \sqrt{\mathrm{var}(Y)} = 2.003$ — This is the covariance

Check

Now calculate ratio

$Z := \frac{Y}{X}$ $\quad \mathrm{var}(Z) = 0.04$ — Empirical

$\mathrm{Var}_Z(\sigma_X, \sigma_Y, \mu_X, \mu_Y, \sigma_{XY}) = 0.038$ — Theory

Check. The theoretical result is close to the simulation.

Comment

Thus the formula seems to work fairly well. Note however that the formula is approximate and is based on a first order expansion. It will not work well unless the ratio of σ_X to μ_X is small.

General comment

This general approach can be used to obtain the variance of other functions of random variables. It is important to remember that it is only approximate. Nevertheless, it can be quite useful on occasion. For example the (approximate) variance of the empirical log-odds transform is calculated this way.

Another simple example of the method

Find a formula for the logarithm of a random variable *U*=ln(*W*)

$$\frac{d}{dW}\ln(W) \rightarrow \frac{1}{W}$$

Thus for *ΔU* we have

$$\frac{1}{W}\cdot\Delta W$$

The variance of U is, however $E[(\Delta U)^2]$

$$\left(\frac{1}{W}\cdot\Delta W\right)^2 \text{ expand } \rightarrow \frac{\Delta W^2}{W^2}$$

$$\text{Var}_U(\mu_W,\sigma_W) := \frac{{\sigma_W}^2}{{\mu_W}^2}$$

Note that this is the ratio of a term in W^2 and another in W^2 and is thus dimensionless, as is only appropriate for a logarithm.

Numerical example

$U := \ln(X)$ $\quad \text{var}(U) = 0.01024$ $\quad \text{Var}_U(\mu_X,\sigma_X) = 0.01$

Check

$V := \ln(Y)$ $\quad \text{var}(V) = 0.0065142$ $\quad \text{Var}_U(\mu_Y,\sigma_Y) = 0.0064$