LARGE-SAMPLE THEORY

The distribution of a function of several sample means, e.g.

$$g(\bar{X}, \bar{Y})$$

is usually too complicated. The central limit theorem states that this distribution tends, as $N \to \infty$, to a Normal distribution with the mean of

$$g(\mu_x, \mu_y) + ...$$

and a variance given by

$$\left(\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_x}\right)^2 \mathrm{Var}(\bar{X}) + \left(\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_y}\right)^2 \mathrm{Var}(\bar{Y})$$
$$+2\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_x}\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_y}\mathrm{Cov}(\bar{X}, \bar{Y}) + ...$$

When dealing with a random independent sample of X, Y pairs (a special case which does NOT apply to to time-series formulas quoted below), we get

$$\left(\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_x}\right)^2 \frac{\sigma_x^2}{N} + \left(\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_y}\right)^2 \frac{\sigma_y^2}{N} + \frac{\partial g(\mu_x, \mu_y)}{\partial \mu_x}\frac{\partial g(\mu_x, \mu_y)}{\partial \mu_y}\frac{\rho_{xy}\sigma_x\sigma_y}{N} + ...$$

The general formula can be derived from

$$\begin{aligned} g(\bar{X}, \bar{Y}) \quad &\simeq \quad g(\mu_x, \mu_y) + \frac{\partial g(\mu_x, \mu_y)}{\partial \mu_x}\left(\bar{X} - \mu_x\right) + \frac{\partial g(\mu_x, \mu_y)}{\partial \mu_y}\left(\bar{Y} - \mu_y\right) \\ &+\frac{1}{2}\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_x^2}\left(\bar{X} - \mu_x\right)^2 + \frac{1}{2}\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_y^2}\left(\bar{Y} - \mu_y\right)^2 \\ &+\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_x \partial \mu_y}\left(\bar{X} - \mu_x\right)\left(\bar{Y} - \mu_y\right) + ... \end{aligned}$$

by realizing that

$$\begin{aligned} \mathbb{E}\left(\bar{X} - \mu_x\right) \quad &= \quad 0 \\ \mathbb{E}\left[\left(\bar{X} - \mu_x\right)^k\right] \quad &= \quad O\left(\frac{1}{N^{\left[\frac{k+1}{2}\right]}}\right) \end{aligned}$$

The error of this approximation is of the $O\left(\frac{1}{\sqrt{N}}\right)$ type. To remove this error, on has to include an extra term to the mean, thus:

$$g(\mu_x, \mu_y) + \frac{1}{2}\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_x^2}\mathrm{Var}(\bar{X}) + \frac{1}{2}\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_y^2}\mathrm{Var}(\bar{Y})$$
$$+\frac{\partial^2 g(\mu_x, \mu_y)}{\partial \mu_x \partial \mu_y}\mathrm{Cov}(\bar{X}, \bar{Y}) + ...$$

which, in the RIS case, reduces to

$$g(\mu_x,\mu_y)+\frac{1}{2}\frac{\partial^2 g(\mu_x,\mu_y)}{\partial\mu_x^2}\frac{\sigma_x^2}{N}+\frac{1}{2}\frac{\partial^2 g(\mu_x,\mu_y)}{\partial\mu_y^2}\frac{\sigma_y^2}{N}+$$
$$+\frac{\partial^2 g(\mu_x,\mu_y)}{\partial\mu_x\partial\mu_y}\frac{\mu_{1,1}}{N}+...$$

($\mu_{1,1}$ indicates the corresponding *central* moment, namely the covariance between *one* X and the corresponding Y).

It is also necessary to incorporate the $\frac{1}{\sqrt{N}}$ proportional skewness, based on the following third central moment of $g(\bar{X})$ (assuming, for simplicity, that there is only one $\bar{X}$ involved):

$$\mathbb{E}\left[\left(g(\bar{X})-g(\mu_x)-\frac{1}{2}\frac{\partial^2 g(\mu_x)}{\partial\mu_x^2}\mathrm{Var}(\bar{X})+...\right)^3\right]\simeq$$
$$\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^3\mathbb{E}\left[\left(\bar{X}-\mu_x\right)^3\right]+\frac{3}{2}\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^2\frac{\partial^2 g(\mu_x)}{\partial\mu_x^2}\left(\mathbb{E}\left[\left(\bar{X}-\mu_x\right)^4\right]-\mathrm{Var}(\bar{X})^2\right)+...$$

For RIS, this yields

$$\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^3\frac{\mu_3}{N^2}+\frac{3}{2}\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^2\frac{\partial^2 g(\mu_x)}{\partial\mu_x^2}\left(\frac{\mu_4}{N^3}+3\frac{N(N-1)\sigma^4}{N^4}-\frac{\sigma^4}{N^2}\right)+...$$
$$\simeq\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^3\frac{\mu_3}{N^2}+3\left(\frac{\partial g(\mu_x)}{\partial\mu_x}\right)^2\frac{\partial^2 g(\mu_x)}{\partial\mu_x^2}\frac{\sigma^4}{N^2}+...$$

where μ_3 and μ_4 are the *central* moments of the X distribution.

For the **special case of RIS and** $g(\bar{X})$, we get the following formulas

$$\mathbb{E}\left[g(\bar{X})\right]\simeq g(\mu)+\frac{g''(\mu)\sigma^2}{2N}+...$$
$$\mathrm{Var}\left[g(\bar{X})\right]\simeq\frac{g'(\mu)^2\sigma^2}{N}+\frac{g'(\mu)g''(\mu)\kappa_3+\frac{1}{2}g''(\mu)^2\sigma^4+g'(\mu)g'''(\mu)\sigma^4}{N^2}+...$$
$$\mathbb{E}\left\{[g(\bar{X})-\mu_g]^3\right\}\simeq\frac{g'(\mu)^3\cdot\mu_3+3g'(\mu)^2g''(\mu)\sigma^4}{N^2}+...$$
$$\mathbb{E}\left\{[g(\bar{X})-\mu_g]^4\right\}\simeq\frac{3g'(\mu)^4\sigma^4}{N^2}+$$
$$\frac{g'(\mu)^4\kappa_4+18g'(\mu)^3g''(\mu)\kappa_3\sigma^2+15g'(\mu)^2g''(\mu)^2\sigma^6+10g'(\mu)^3g'''(\mu)\sigma^6}{N^3}+...$$

where κ_i are cumulants of the X distribution, and μ_g denotes the expected value of $g(\bar{X})$.

The last expression can be easily converted into the corresponding 4^{th} cumulant of $g(\bar{X})$, getting

$$\frac{g'(\mu)^2\left[g'(\mu)^2\kappa_4+12g'(\mu)g''(\mu)\kappa_3\sigma^2+12g''(\mu)^2\sigma^6+4g'(\mu)g'''(\mu)\sigma^6\right]}{N^3}+...$$