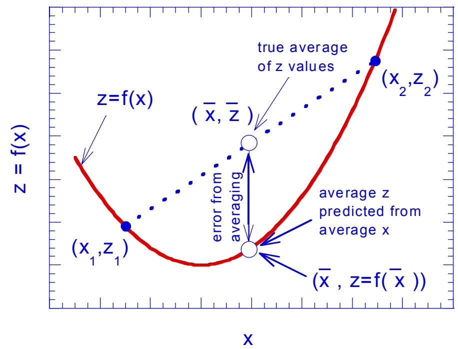
General statement of the problem

If some quantity of interest z is calculated from other quantities x, y, q, w, etc.,

$$z = f(x, y, q, w, \cdots)$$

the average value of z will *NOT*, in general, be equal to the function f evaluated at the averages of x, y, q, w, etc. Whenever f is nonlinear or its input variables are correlated with one another, $\bar{z} \neq f(\bar{x}, \bar{y}, \bar{q}, \bar{w}, \cdots)$, as the following figure illustrates:



Intuition tells us, from the figure above, that the true mean of z (the upper open circle) will differ from f evaluated at the mean of x (the lower open circle) by an amount that depends on two things: the degree of curvature in the function $f(x)$ and the amount of variability in x. Note that $\bar{z}$ deviates from $f(\bar{x})$ because x is variable, not because the mean value $\bar{x}$ is uncertain. That is, increasing the number of measurements may help to estimate $\bar{x}$ more precisely, but it will do nothing to decrease the discrepancy between $f(\bar{x})$ and $\bar{z}$.

Method of moments

<u>Simplest case: single-variable function z=f(x)</u>

Approximate f by a quadratic curve that has a slope of dz/dx and a curvature of d^2z/dx^2, evaluated at the mean of x. Then $\bar{z}$ will be approximately,

$$\bar{z} \approx f(\bar{x}) + \frac{1}{2}\frac{d^2z}{dx^2} Var(x)$$

This result corresponds exactly to the intuitive impression conveyed by the figure above: $\bar{z}$ differs from $f(\bar{x})$ by an amount that is proportional to the curvature of f (measured by the second derivative of z, d^2z/dx^2) and the variability of x, measured by Var(x).

A sketch of the derivation goes like this. Approximate the function $f(x)$ by its second-order Taylor series expansion about the mean of x:

$$z = f(x) \approx f(\bar{x}) + \frac{dz}{dx}(x-\bar{x}) + \frac{1}{2}\frac{d^2z}{dx^2}(x-\bar{x})^2 + \cdots$$

Take the mean of this approximate z:

$$\bar{z} = \frac{\sum z}{n} \approx \frac{\sum f(\bar{x})}{n} + \frac{dz}{dx}\frac{\sum(x-\bar{x})}{n} + \frac{1}{2}\frac{d^2z}{dx^2}\frac{\sum(x-\bar{x})^2}{n} \qquad \textit{noting that:}$$

$$\frac{\sum f(\bar{x})}{n} = f(\bar{x}), \qquad \frac{\sum(x-\bar{x})}{n} = 0, \qquad \frac{\sum(x-\bar{x})^2}{n} = Var(x)$$

So one directly obtains the result given above.

Function of two variables z=f(x,y)

Approximate f by a curved surface whose slope in the x and y dimensions is described by the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$, and whose curvature is described by the four partial second derivatives $\partial^2 z/\partial x^2$, $\partial^2 z/\partial y^2$, $\partial^2 z/\partial x\partial y$, and $\partial^2 z/\partial y\partial x$ (again, these are evaluated at the mean x and mean y). The mean of z is approximately:

$$\bar{z} \approx f(\bar{x},\bar{y}) + \frac{1}{2}\frac{\partial^2 z}{\partial x^2}Var(x) + \frac{1}{2}\frac{\partial^2 z}{\partial y^2}Var(y) + \frac{1}{2}\left(\frac{\partial^2 z}{\partial x\partial y} + \frac{\partial^2 z}{\partial y\partial x}\right)Cov(x,y)$$

The derivation follows that shown above, and will not be shown here. Note that if the surface is saddle-shaped, the variability in x and y can have offsetting effects on $\bar{z}$. If x and y are uncorrelated, then the covariance term vanishes, leaving simply,

$$\bar{z} \approx f(\bar{x},\bar{y}) + \frac{1}{2}\frac{\partial^2 z}{\partial x^2}Var(x) + \frac{1}{2}\frac{\partial^2 z}{\partial y^2}Var(y)$$

Function of many variables z=f(x_1, x_2, x_3...)

Caution: different notation. Here x_1, x_2, x_3...x_m to refer to m different variables (e.g., x_2=y, x_3=q, etc.) rather than different measurements of a single variable. The result for the two-dimensional case shown above can be generalized to,

$$\bar{z} \approx f(\bar{x}_1,\bar{x}_2,\cdots\bar{x}_m) + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 z}{\partial x_i^2}Var(x_i) + \frac{1}{2}\sum_{j=1}^{m}\sum_{i\neq j}^{m}\frac{\partial^2 z}{\partial x_i\partial x_j}Cov(x_i,x_j)$$

Where *all* of the x_i are uncorrelated with one another, the covariance terms vanish, leaving

$$\bar{z} \approx f(\bar{x}_1,\bar{x}_2,\cdots\bar{x}_m) + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 z}{\partial x_i^2}Var(x_i)$$

Exact analytic methods

If the probability density function p(x) is known, the mean of z can be calculated directly, and without approximation, as

$$\bar{z} = \int_{-\infty}^{\infty} f(x)p(x)dx$$

If the integral cannot be solved analytically (as is often the case), it can be evaluated by numerical integration.

Monte Carlo methods

Monte Carlo methods calculate the average value of z by brute force, by evaluating f at many individual values of x, y, etc. that have been randomly generated from distributions that have statistical properties similar to the real-world variables x, y, etc. See the error propagation toolkit for further details.

References: Ang, A. H.-S. and W. H. Tang, *Probability Concepts in Engineering Planning and Design*, 409 pp., John Wiley and Sons, 1975 (Chapter 4).

Rastetter, E. B., A. W. King, B. J. Cosby, G. M. Hornberger, R. V. O'Neill and J. E. Hobbie, Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems, *Ecological Applications*, **2**, 55-70, 1992.