Statistics 3657 : Moment Approximations

1 Preliminaries

Suppose that we have a r.v. X and that we wish to calculate the expectation of $g(X)$ for some function g. Of course we could calculate it as $\mathrm{E}(g(X))$ by the appropriate integral (continuous r.v.) or sum (discrete r.v.). This may not always be easy, for example when X itself is a function of several other r.v.'s.

An approximation can be constructed using Taylor's series. In practice usually we are only interested in first or second order approximations. Below we assume that g has enough derivatives for the expressions to be valid.

A first order Taylor's series of a function g, about x_0 is

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + R_1(x)$$

where $R_1(x)$ is the remainder term.

A second order Taylor's series of a function g, about x_0 is

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2 + R_2(x)$$

where $R_2(x)$ is the remainder term.

The remainder terms R_1 and R_2 are of the forms

$$\begin{aligned} R_1(x) &= \frac{1}{2}g''(x^*)(x - x_0)^2 \\ R_2(x) &= \frac{1}{3!}g^{(3)}(x^*)(x - x_0)^3 \end{aligned}$$

where x^* is some number between x_0 and x. More specifically there is a number $\alpha \in [0, 1]$ such that $x^* = \alpha x_0 + (1 - \alpha)x$. Typically it is not easy to calculate x^*, but there are often nice upper bounds on the remainder terms. However in this course we will not be concerned with these beyond noting that they exist. In further probability courses these are studied.

Now consider the Taylor approximation obtained by taking the polynomial part of the Taylor's series, but ignoring the remainder. That is we consider

$$\begin{aligned} g_1(x) &= g(x_0) + g'(x_0)(x - x_0) \\ g_2(x) &= g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2} g''(x_0)(x - x_0)^2 \end{aligned}$$

The function g_1 is a polynomial of degree 1, and the function g_2 is a polynomial of degree 2. We also call g_1 a first order Taylor's approximation of the function g, and call g_2 a second order Taylor's approximation of the function g.

2 Moment Approximation

Let X be a random variable with enough finite moments so that the calculations below are well defined. They will be given in more detail below as needed.

Suppose that $\mathrm{E}(X) = \mu$ is finite, and that $\sigma^2 = \mathrm{Var}(X)$ is also finite. Let g_1 be the first order Taylor's approximation of g about the $x_0 = \mu$, that is

$$g_1(x) = g(\mu) + g'(\mu)(x - \mu) \ .$$

Consider the random variable $g_1(X)$. It is an approximation to the random variable $g(X)$. We can use moments of $g_1(X)$ to approximate moments of $g(X)$. Using this first Taylor's approximation we can easily calculate

$$\begin{aligned} \mathrm{E}(g_1(X)) &= \mathrm{E}\left(g(\mu) + g'(\mu)(X - \mu)\right) \\ &= g(\mu) + g'(\mu)\mathrm{E}\left(X - \mu\right) \\ &= g(\mu) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(g_1(X)) &= \mathrm{Var}\left(g(\mu) + g'(\mu)(X - \mu)\right) \\ &= \left(g'(\mu)\right)^2 \mathrm{Var}\left(X - \mu\right) \\ &= \left(g'(\mu)\right)^2 \sigma^2 \ . \end{aligned}$$

The idea of the moment approximation is to approximate $\mathrm{E}(g(X))$ and $\mathrm{Var}(g(X))$ by moments of the Taylor functions used to approximate g.

Example :

Suppose $X \sim \text{Unif}(1,3)$ and $Y = \sqrt{X}$. Find the first order Taylor approximation to the mean and variance of Y.

$$\mu = E(X) = \int_1^3 x\frac{1}{2}dx = 2$$

Also find the variance of X, as $\text{Var}(X) = \frac{1}{3}$.

Consider $g(x) = \sqrt{x}$ as a mapping from [1,3] to the reals. $g'(x) = \frac{1}{2}x^{-1/2}$. Thus the first order Taylor's approximation about $\mu = 2$ is

$$g_1(x) = g(2) + \frac{1}{2\sqrt{2}}(x-2)\ .$$

Then

$$\begin{aligned} \text{E}(g_1(X)) &= \sqrt{2} \\ \text{Var}(g_1(X)) &= \left(\frac{1}{2\sqrt{2}}\right)^2 \frac{1}{3} \\ &= \frac{1}{24} \end{aligned}$$

Thus we find the first order Taylor's approximation for the mean and variance of Y as

$$\begin{aligned} \text{E}(Y) &\approx \text{E}(g_1(X)) = \sqrt{2} \\ \text{Var}(Y) &\approx \text{Var}(g_1(X)) = \frac{1}{24} \end{aligned}$$

If we approximate $\text{E}(g(X))$ by

$$\text{E}(g_1(X)) = \text{E}(g(\mu) + g'(\mu)(X-\mu)) = g(\mu)$$

we do not have a good approximation. Thus we often use the second order Taylor series to approximate $\text{E}(g(X))$, that is

$$\text{E}(g_2(X)) = \text{E}(g(\mu) + g'(\mu)(X-\mu) + \frac{1}{2}g''(\mu)(X-\mu)^2) = g(\mu) + \frac{g''(\mu)}{2}\sigma^2\ .$$

To use the second order approximation for the variance we obtain

$$\begin{aligned} \text{Var}(g_2(X)) &= \text{Var}\left(g(\mu) + g'(\mu)(X-\mu) + \frac{1}{2}g''(\mu)(X-\mu)^2\right) \\ &= g'(\mu)^2\text{Var}(X) + 2g'(\mu)\frac{1}{2}g''(\mu)\text{Cov}(X-\mu,(X-\mu)^2) + \frac{1}{4}g''(\mu)^2\text{Var}((X-\mu)^2) \end{aligned}$$

Aside : The student should find the mean and variance of $Y = \sqrt{X}$. These are exactly and rounded to 4 decimal places

$$\begin{aligned} \mathrm{E}(Y) &= \frac{\sqrt{27}-1}{3} = 1.3987 \\ \mathrm{Var}(Y) &= \frac{6\sqrt{3}-10}{9} = .0436 \ . \end{aligned}$$

Aside : In general there is no need to carry more digits than required to get a meaningful answer. Here 3 or 4 decimal digits is reasonable, whereas 1 decimal digit is not as it would give a variance of approximately 0.0, and hence not reasonable.

Remark Recall Var$Y \geq 0$. Also $\mathrm{Var}(g_1(X)) \geq 0$ and $\mathrm{Var}(g_2(X)) \geq 0$. However it is **not always true** that

$$\mathrm{E}(g_2(X)^2) - (\mathrm{E}(g_1(X)))^2$$

is greater than or equal to 0. Similarly it is also not guaranteed that

$$\mathrm{E}(g_1(X)^2) - (\mathrm{E}(g_2(X)))^2 \geq 0 \ .$$

Thus when approximating variances with this method we should in general not mix the two different degrees of approximation.

Functions of Several Variables and Moment Approximation

This idea of moment approximation also extends to functions of several variables. Here we only consider two variables.

Suppose (X, Y) are bivariate r.v.s and $Z = g(X, Y)$. We suppose that g has derivatives and hence Taylor approximations.

For ease of writing we use the notation

$$\begin{aligned}
g_x(x,y) &= \frac{\partial g(x,y)}{\partial x} \\
g_y(x,y) &= \frac{\partial g(x,y)}{\partial y} \\
g_{xx}(x,y) &= \frac{\partial^2 g(x,y)}{\partial^2 x} \\
g_{yy}(y,y) &= \frac{\partial^2 g(x,y)}{\partial^2 y} \\
g_{xy}(x,y) &= \frac{\partial^2 g(x,y)}{\partial x\, \partial y} \ .
\end{aligned}$$

First order Taylor's approximation about (x_0, y_0) :

$$g_1(x,y) = g(x_0,y_0) + g_x(x_0,y_0)(x-x_0) + g_y(x_0,y_0)(y-y_0) \ .$$

Second order Taylor's approximation about (x_0, y_0) :

$$\begin{aligned}
g_2(x,y) = \ & g(x_0,y_0) + g_x(x_0,y_0)(x-x_0) + g_y(x_0,y_0)(y-y_0) + \\
& \frac{1}{2} g_{xx}(x_0,y_0)(x-x_0)^2 + g_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \\
& \frac{1}{2} g_{yy}(x_0,y_0)(y-y_0)^2 \ .
\end{aligned}$$

Again we approximate

$$\mathrm{E}(g(X,Y)) \approx \mathrm{E}(g_2(X,Y))$$

and

$$\mathrm{Var}(g(X,Y)) \approx \mathrm{Var}(g_1(X,Y))$$

where the Taylor's approximation are obtained about $(x_0, y_0) = (\mu_X, \mu_Y)$.

Example

Suppose that X and Y are independent and we are interested in the ratio $g(X,Y) = \frac{Y}{X}$.

Then

$$g_1(X,Y) = \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(X-\mu_X) + \frac{1}{\mu_X}(Y-\mu_Y)$$

and

$$\begin{aligned} g_2(X,Y) &= \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(X-\mu_X) + \frac{1}{\mu_X}(Y-\mu_Y) + \\ &\quad \frac{1}{2}\frac{2\mu_Y}{\mu_X^3}(X-\mu_X)^2 + 0 \times (X-\mu_X)(Y-\mu_Y) + 0 \times (Y-\mu_Y)^2 \\ &= \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(X-\mu_X) + \frac{1}{\mu_X}(Y-\mu_Y) + \frac{\mu_Y}{\mu_X^3}(X-\mu_X)^2 \ . \end{aligned}$$

Using the second order Taylor approximation we obtain

$$\begin{aligned} \mathrm{E}(g(X,Y)) &\approx \mathrm{E}(g_2(X,Y)) \\ &= \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^3}\sigma_X^2 \ . \end{aligned}$$

Using the first order Taylor approximation we obtain

$$\begin{aligned} \mathrm{Var}(g(X,Y)) &\approx \mathrm{Var}(g_1(X,Y)) \\ &= \mathrm{Var}\left(\frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(X-\mu_X) + \frac{1}{\mu_X}(Y-\mu_Y)\right) \\ &= \frac{\mu_Y^2}{\mu_X^4}\sigma_X^2 - 2\frac{\mu_Y}{\mu_X^3}\mathrm{Cov}(X,Y) + \frac{1}{\mu_X^2}\sigma_Y^2 \\ &= \frac{\mu_Y^2}{\mu_X^4}\sigma_X^2 + \frac{1}{\mu_X^2}\sigma_Y^2 \ . \end{aligned}$$

QQ Plots

Consider the iid random variables $X_1, X_2, \ldots, X_n$ from a continuous cdf F.

Next consider the ordered data $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, that is the order statistics from these n r.v.s. Notice that $U_i = F(X_i)$ are iid Uniform(0,1), and that $U_{(i)} = F(X_{(i)})$. Thus

$$X_{(i)} = F^{-1}(U_{(i)})$$

We can find the marginal distribution of $U_{(i)}$ from the methods above. Thus $X_{(i)} = F^{-1}(U_{(i)})$

In Chapter 3.7 we found the distribution of $U_{(i)}$, specifically it has the pdf, say f_i given by

$$f_i(x) = \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1} f(x)(1-F(x))^{n-i}$$

where F and f are the cdf and pdf of $U \sim$ Uniform$(0,1)$. In particular we have

$$f_i(x) = \begin{cases} \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Remark : This is a special case of a Beta distribution.

Remark : Since this is a probability density for any positive integer n and $i \leq i \leq n$, we also have (replacing the integers n, i by m, k) that

$$\int_0^1 x^{k-1}(1-x)^{m-k} dx = \frac{(k-1)!(m-k)!}{m!}$$

Thus

$$\begin{aligned} \mathrm{E}(U_{(i)}) &= \int_0^1 x f_i(x) dx \\ &= \int_0^1 x \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} dx \\ &= \frac{n!}{(i-1)!(n-i)!} \int_0^1 x^i (1-x)^{(n+1)-(i+1)} dx \\ &= \frac{n!}{(i-1)!(n-i)!} \cdot \frac{i!(n-i)!}{(n+1)!} \\ &= \frac{i}{n+1} \end{aligned}$$

Thus to a first order Taylor's approximation we have

$$\mathrm{E}(X_{(i)}) = \mathrm{E}(F^{-1}(U_{(i)}) = F^{-1}(\mathrm{E}(U_{(i)})) = F^{-1}\left(\frac{i}{n+1}\right) .$$

Thus when we order the iid data $X_1, X_2, \ldots, X_n$, and plot the pairs $F^{-1}\left(\frac{i}{n+1}\right)$, $X_{(i)}$ on average these points are of the form $\left(F^{-1}\left(\frac{i}{n+1}\right), F^{-1}\left(\frac{i}{n+1}\right)\right)$, that is they fall on a straight line of slope 1. This notion corresponds to plotting on one axis the *theoretical* quantiles and on the other hand the *empirical* quantiles or order statistics. This type of plot is called a quantile-quantile or QQ plot.

Normal QQ Plots

In some cases, namely location and scale families of distributions, it is possible to make QQ plots with respect to a particular member of this family. This is what is done for normal QQ plots.

First we find the relation between non-standard normal and standard normal quantiles.

Recall for a given number $0 < q < 1$, the q-th quantile of a distribution with cdf F is defined as the solution, say x_q of the equation $F(x) = q$. Let x_q be the q-th quantile for a $N(\mu, \sigma^2)$ distribution, and z_q be the q-th quantile for a standard normal distribution. Thus, letting $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$

$$\begin{aligned} P(X \leq x_q) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x_q - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x_q - \mu}{\sigma}\right) \end{aligned}$$

Therefore

$$\frac{x_q - \mu}{\sigma} = z_q$$

or equivalently

$$F^{-1}(q) = x_q = \sigma z_q + \mu$$

where F is the $N(\mu, \sigma^2)$ cdf. Let Φ be the $N(0, 1)$ cdf.

Plotting $X_{(i)}$ against $F^{-1}\left(\frac{i}{n+1}\right)$ will follow approximately a line with intercept 0 and slope 1. However plotting $X_{(i)}$ against $\Phi^{-1}\left(\frac{i}{n+1}\right)$ will follow approximately a straight line of slope σ and intercept μ. In fact this is what is done in software implementations of normal QQ plots, such as qqnorm in the statistical programming language R.

Aside : R actually does something slightly different.

$$\mathrm{E}(X_{(i)}) = \Phi^{-1}\left(\frac{i - \frac{1}{2}}{n}\right) .$$

In R the qqnorm function plots the observed data $x_{(i)}$ against the normal quantiles above. The student might wish to check this by making the corresponding plots in R.

Exponential QQ Plots

One may construct other QQ plots for continuous distributions. Consider the exponential distribution, with cdf

$$F(x) = 1 - e^{-\lambda x} \quad \text{if } x > 0$$

and $F(x) = 0$ for $x \leq 0$. Quantiles are easy to obtain, so that the q-th quantile is the solution of

$$q = F(x) \ .$$

Let x_q be this solution. Therefore

$$q = 1 - e^{-\lambda x_q}$$

or equivalently

$$x_q = -\frac{1}{\lambda} \log(1-q) = \frac{1}{\lambda} \log(1/(1-q)) \ .$$

For data $x_1, \ldots, x_n$ the exponential λ QQ plot would plot the pairs

$$\Big(-\frac{1}{\lambda} \log(1 - \frac{i}{n+1}), x_{(i)}\Big) = \Big(-\frac{1}{\lambda} \log(\frac{n+1-i}{n+1}), x_{(i)}\Big) \ ,$$

$i = 1, \ldots, n$. This plot of course would require knowing λ. Since the exponential λ quantiles are the standard exponential (exponential parameter 1) quantiles divided by λ, one may also plot

$$\Big(-\log(1 - \frac{i}{n+1}), x_{(i)}\Big) = \Big(-\log(\frac{n+1-i}{n+1}), x_{(i)}\Big) \ ,$$

$i = 1, \ldots, n$. Notice that if the X_i are iid exponential, no mater what value for λ, this plot will be approximately a straight line. The departures from the straight line will depend on the sampling variability of the order statistics.

The student might wish to write an R, matlab or other program to make these exponential QQ plots and then try it for some simulated exponential data. You will then see the variability is much more than for normal QQ plots, but useful nonetheless.