Notes on Normal Approximation

April 1, 2009

The z score is of some random quantity is

$$Z = \frac{\text{random quantity} - \text{mean}}{\text{standard deviation}}.$$

The **central limit theorem** and extensions like the **delta method** tell us when the z-score has an approximately standard normal distribution. Thus, using R, we can find $P\{Z < z\}$ using `pnorm(z)` and $P\{Z > z\}$ using `1-pnorm(z)`.

If we have a sum S_n of n independent random variables, $X_1, X_2, \ldots X_n$ whose common distribution has mean μ and variance σ^2, then

- the mean $ES_n = n\mu$,
- the variance $\mathrm{Var}(S_n) = n\sigma^2$,
- the standard deviation is $\sigma\sqrt{n}$.

Thus, S_n is approximately normal with mean $n\mu$ and variance $n\sigma^2$. The z-score is

$$Z = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

We can approximate $P\{S_n < x\}$ by noting that this is the same as computing

$$Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} < \frac{x - n\mu}{\sigma\sqrt{n}} = z$$

and finding $P\{Z < z\}$ using the standard normal distribution.

For a sample mean $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$,

- the mean $E\bar{X} = \mu$,
- the variance $\mathrm{Var}(\bar{X}) = \sigma^2/n$,
- the standard deviation is $\sigma/\sqrt{n}$.

Thus, $\bar{X}$ is approximately normal with mean μ and variance σ^2/n. The z-score is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Thus,

$$\bar{X} < x \quad \text{is equivalent to} \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{x - \mu}{\sigma/\sqrt{n}}.$$

For the delta method using $\bar{X}$ and a function g, for a sample mean $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$,

- the mean $Eg(\bar{X}) \approx g(\mu)$,
- the variance $\mathrm{Var}(g(\bar{X})) \approx g'(\mu)^2\sigma^2/n$,
- the standard deviation is $|g'(\mu)|\sigma/\sqrt{n}$.

Thus, $g(\bar{X})$ is approximately normal with mean $g(\mu)$ and variance $g'(\mu)^2\sigma^2/n$. The z-score is

$$Z = \frac{g(\bar{X}) - g(\mu)}{|g'(\mu)|\sigma/\sqrt{n}}.$$

For the two variable delta method, we now have two independent sequences of independent random variables, $X_1, X_2, \ldots X_{n_1}$ whose common distribution has mean μ_1 and variance σ_1^2 and $Y_1, Y_2, \ldots Y_{n_2}$ whose common distribution has mean μ_2 and variance σ_2^2. For a function of the sample means,

- the mean $Eg(\bar{X}, \bar{Y}) \approx g(\mu_1, \mu_2)$,
- the variance
$$\mathrm{Var}(g(\bar{X}.\bar{Y})) \approx \sigma_{g,n}^2 = \left(\frac{\partial}{\partial x} g(\mu_1, \mu_2)\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{\partial}{\partial y} g(\mu_1, \mu_2)\right)^2 \frac{\sigma_2^2}{n_2},$$
- the standard deviation is $|\sigma_{g,n}$.

Thus, $g(\bar{X}, \bar{Y})$ is approximately normal with mean $g(\mu_1, \mu_2)$ and variance $\sigma_{g,n}^2$. the z-score is

$$Z = \frac{g(\bar{X}, \bar{Y}) - g(\mu_1, \mu_2)}{\sigma_{g,n}}.$$

The generalization of the delta method to higher dimensional data will add terms to the variance formula.