Appendix A: Derivation of a Confidence Interval for $\pi_1 c \times \pi_2$

Ken Aho

March 6, 2015

Delta method

Under the Delta method, if:

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sigma} \xrightarrow{d} N(0,1),$$

then

$$\frac{\sqrt{n}[g(\hat{\theta})-g(\theta)]}{g'(\theta)\sigma} \xrightarrow{d} N(0,1).$$

Thus,

$$Var[g(\hat{\theta})] = \sigma^2[g'(\theta)]^2.$$

Delta method application

Variance estimator for $\log(\hat{\pi}_1 c)$

Let $Y_1 \sim BIN(n_1, \pi_1)$, and let $\tau = \pi_1 c$, where c is a constant. Because of the convergence to normality of the sampling distribution of $\hat{\pi}_1$ we have:

$$\frac{\sqrt{n_1}\left(\frac{Y_1 c}{n_1}-\tau\right)}{\sqrt{\pi_1(1-\pi_1)c^2}} \xrightarrow{d} N(0,1).$$

Now, under the delta method, let $g(\tau) = \log(\tau)$. This means $g'(\tau) = \frac{1}{\tau}$, and we have:

$$\left[\log\left(\frac{Y_1 c}{n}\right) - \log(\tau)\right] \xrightarrow{d} N\left(0, \frac{\pi_1(1-\pi_1)c^2}{n_1}(1/\tau)^2\right).$$

Thus, the variance of $\log(\hat{\tau})$ is:

$$\sigma^2_{\log(\hat{\tau})} = \frac{\pi_1(1-\pi_1)c^2}{n_1} \times \frac{1}{\pi_1^2 c^2} = \frac{(1-\pi_1)}{\pi_1 n_1}.$$

The estimator of the variance of $\log(\hat{\tau})$ will be:

$$\hat{\sigma}^2_{\log(\hat{\tau})} = \frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1}.$$

Variance estimator for $\log(\hat{\pi}_2)$

Let $Y_2 \sim BIN(n_2, \pi_2)$. Because of the convergence to normality of the sampling distribution of $\hat{\pi}_2$ we have:

$$\frac{\sqrt{n_2}\left(\frac{Y_2}{n_2} - \pi_2\right)}{\sqrt{\pi_2(1-\pi_2)}} \xrightarrow{d} N(0,1).$$

Let $g(\pi_2) = \log(\pi_2)$. We have $g'(\pi_2) = \frac{1}{\pi_2}$, resulting in:

$$\left[\log\left(\frac{Y_2}{n_2}\right) - \log(\pi_2)\right] \xrightarrow{d} N\left(0, \frac{\pi_2(1-\pi_2)}{n}(1/\pi_2)^2\right).$$

Thus, the variance of $\log(\hat{\pi}_2)$ is:

$$\sigma^2_{\log(\hat{\pi}_2)} = \frac{\pi_2(1-\pi_2)}{n} \times \frac{1}{\pi_2^2} = \frac{(1-\pi_2)}{\pi_2 n}.$$

The estimator of the variance of $\log(\hat{\pi}_2)$ will be:

$$\hat{\sigma}^2_{\log(\hat{\pi}_2)} = \frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}.$$

Variance estimator for $\log(\hat{\pi}_1 c \times \log \hat{\pi}_2)$

We note:

1. $\log(a \times b) = \log(a) + \log(b)$.
2. Let X and Y be independent random variables, then $Var(X+Y) = Var(X) + Var(Y)$.

Thus,

$$Var\Big(\log(\hat{\pi}_1 c \times \hat{\pi}_2)\Big) = Var\Big(\log(\hat{\pi}_1 c)\Big) + Var\Big(\log(\hat{\pi}_2)\Big) = \frac{(1-\pi_1)}{\pi_1 n_1} + \frac{(1-\pi_2)}{\pi_2 n_2}.$$

The variance estimator is:

$$\frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1} + \frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}.$$

Confidence interval for $\pi_1 c \times \pi_2$

Because of the asymptopic normality of $\log(\hat{\pi}_1 c \times \hat{\pi}_2)$, an approximate $(1-\alpha)100\%$ confidence interval for $\pi_1 c \times \pi_2$ is given by:

$$\hat{\theta} \times \exp(\pm z_{1-(\alpha/2)} \hat{\sigma}_{\hat{\theta}}),$$

where $\hat{\theta} = \hat{\pi}_1 c \times \hat{\pi}_2$, and $\hat{\sigma}^2_{\hat{\theta}} = \frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1} + \frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}$.