1 Normal approximation

The underlying process is constructed as a sequence of N Bernoulli trials with success rate $\tilde{p}$. Under the normal approximation, the PDF of the observed success rate $f(p)$ is given by:

$$f(p) \sim N\left(\tilde{p},\ \sqrt{\frac{\tilde{p}(1-\tilde{p})}{N}}\right) \tag{1}$$

2 One proportion test

Let's assume that we prepared an actual realization of N Bernoulli trials. Our objective is to assign a p-value telling us the probability that the success rate of the observed realization is in fact higher than some test rate p_0. As we saw above (TODO) the first step is to calculate the z-score according to:

$$z = \frac{p - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{N} \tag{2}$$

where p_0 is the rate that we are testing against. A small calculation show that the PDF of the observed z-scores should behave (under the normal approximation) as:

$$f(z) \sim N\left(\mu_z,\ \sigma_z\right) \quad \text{with} \quad \begin{cases} \mu_z & = (\tilde{p} - p_0)\sqrt{\frac{N}{p_0(1-p_0)}} \\ \sigma_z & = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{p_0(1-p_0)}} \end{cases} \tag{3}$$

3 One tailed p-value

One can then extract the one tailed p-value from the z-score as follows:

$$p = 1 - \Phi\left(z\right) \tag{4}$$

where Φ stands for the cumulative distribution function of the standard normal distribution $N(0,1)$. Since there are no closed forms for Φ and its inverse the quantile function, one has to rely on error functions (and their inverses). Instead, We can simplify (and maybe gain more insight) by using the Pólya approximation:

$$\Phi(z) \approx \frac{1}{2}\left[1 + \text{sign}(z)\sqrt{1 - \exp\left(-\frac{2z^2}{\pi}\right)}\right] \tag{5}$$

In addition to its simplicity, this closed form approximation has the advantage that it can easily be inverted such that we eventually get:

$$z(p) \approx \text{sign}\left(\frac{1}{2} - p\right)\sqrt{\frac{\pi}{2}\log\frac{1}{4p(1-p)}} \tag{6}$$

In the limit $p \ll 1$, we can simplify the inverse relation to:

$$z(p) \approx \sqrt{\frac{\pi}{2} \log \frac{1}{4p}} \quad ; \quad \left| \frac{dz}{dp} \right| \approx \sqrt{\frac{\pi}{2}} \, \frac{1}{2p\sqrt{\log 1/4p}} \tag{7}$$

The "presumptive" PDF of p-values ($p \ll 1$) would then be:

$$PDF(p) \approx 2^{\frac{\pi}{2\sigma_z^2}-2} \frac{\exp\left[-\frac{1}{2}\left(\frac{\mu_z}{\sigma_z}\right)^2\right]}{\sigma_z} \frac{\exp\left(\frac{\mu_z}{\sigma_z^2}\sqrt{\frac{\pi}{2}\log\frac{1}{4p}}\right)}{p^{1-\frac{\pi}{4\sigma_z^2}}\sqrt{\log\frac{1}{4p}}} \tag{8}$$

Note that this expansion is valid only for $p < 1/4$.