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:

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

where Φ stands for the cumulative distribution function of the standard normal distribution $N(0,1)$. The inverse function is defined as the quantile function:

$$z = \sqrt{2}\,\mathrm{erfi}(2p-1) \tag{5}$$

We can simplify (and maybe gain more insight) by using the Pólya approximation:

$$\Phi(z) \approx \frac{1}{2}\left[1 + \mathrm{sign}(z)\sqrt{1 - \exp\left(-\frac{2z^2}{\pi}\right)}\right] \tag{6}$$

Maybe need to use that:

$$\mathrm{sign}(z) \approx \lim_{k\to+\infty} \tanh kz \approx \lim_{\varepsilon\to 0} \frac{x}{\sqrt{x^2+\varepsilon^2}} \tag{7}$$