The null distribution for the difference between two sample proportions

Suppose we have two populations with sizes n_1 and n_2, and for these populations the proportion of individuals with a characteristic of interest is given by p_1 and p_2, respectively. Then from the CLT (assuming at least 15 successes and failures), the sample proportions $\hat{p}_1$ and $\hat{p}_2$ are distributed as follows:

$$\hat{p}_1 \sim N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right) \qquad \hat{p}_2 \sim N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$$

We are interested in the distribution of $\hat{p}_1 - \hat{p}_2$.

From our previous theorem,

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

But we will operate under a null hypothesis where $p_1 - p_2 = 0$ (i.e., the proportions in each group are the same). So let's call the common proportion p.

Then

$$\hat{p}_1 - \hat{p}_2 \sim N\left(0, \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

Convert $\hat{p}_1 - \hat{p}_2$ to the Z distribution (by subtracting off its mean, and dividing by its standard deviation). This gives

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

and Z ~ N(0,1) if the null hypothesis of equal proportions is true.

This allows us to calculate p-values based on the difference in sample proportions