

Test for the probability of a binomial distribution

Chao Cheng

August 26, 2022

For an i.i.d sample from a bernoulli distribution

$$x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),$$

The likelihood of the data is

$$f(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}.$$

MLE for p is $\bar{x} = \frac{1}{n} \sum x_i$ and

$$\sum_{i=1}^n x_i \sim \text{Binom}(n, p).$$

So here are mainly two situations: One is to test the probability p against some given value p_0 . The other is to compare the probability between two independent random samples x_1, \dots, x_n and y_1, \dots, y_m .

Case1: One sample x_1, \dots, x_n from $\text{Bernoulli}(p)$, and test p against a given p_0 .

Case2: Two samples: x_1, \dots, x_n from $\text{Bernoulli}(p_1)$ and y_1, \dots, y_m from $\text{Bernoulli}(p_2)$.
And test whether $p_1 = p_2$.

1 Normal approximation

1.1 Case 1

Note that

$$EX = p, \quad \text{Var}X = p(1-p).$$

Then by CLT we have

$$\bar{x} \stackrel{\text{asympt}}{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

For $H_0 : p = p_0$, we propose a test statistic

$$Z = \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Then Z is asymptotically standard normal under H_0 .

Also we know that under H_1 :

$$\begin{aligned} Z &= \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} \cdot \sqrt{\frac{p(1-p)}{p_0(1-p_0)}} + \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &\sim N\left(\frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \frac{p(1-p)}{p_0(1-p_0)}\right). \end{aligned}$$

So the power of the test can be easily computed.

1.2 Case 2

So we have

$$\bar{x} \stackrel{\text{asympt}}{\sim} N\left(p_1, \frac{p_1(1-p_1)}{n}\right), \quad \text{and} \quad \bar{y} \stackrel{\text{asympt}}{\sim} N\left(p_2, \frac{p_2(1-p_2)}{m}\right).$$

A test statistic can be

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}},$$

where $\hat{p} = \frac{n\bar{x} + m\bar{y}}{n+m}$. This test statistic can be found at

<https://stats.stackexchange.com/questions/361015/proof-of-the-standard-error-of-the-distribution-between-two-normal-distributions/361048#361048>

<https://stats.stackexchange.com/questions/113602/test-if-two-binomial-distributions-are-statistically-different-from-each-other>

Here this $\hat{p}(1-\hat{p})$ can be seen as an estimate for the variance $p(1-p)$ when H_0 is true by directly plugging in \hat{p} . This is **NOT** a pooled variance for these two samples, which should always be no greater than $\hat{p}(1-\hat{p})$.

The power of this test statistic is hard to compute under H_1 .

Note: One can also use the same idea in the “t-test.pdf” notes and propose the test statistic

$$T = \frac{\bar{x} - p_0}{\sqrt{S_x/n}},$$

where $S_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ for Case 1.

And for Case 2

$$T = \frac{\bar{x} - \bar{y} - \Delta}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) S_p}},$$

where $S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$ and $\Delta = p_1 - p_2$. But again it is hard to evaluate the testing power of these statistics.

2 Chi-square approximation

See the notes of “chisq_test.pdf” for details.

3 Clopper-Pearson exact test