The Chi-square Distribution

Background: If Z has a standard normal distribution then by definition Z^2 has a χ_1^2 distribution, i.e. a chi-square distribution with one degree of freedom.

If $Z_1, \ldots, Z_k$ are independent random variables, each with a standard normal distribution, then by definition $Z_1^2 + \cdots + Z_k^2$ has a χ_k^2 distribution.

What is the expected value of a χ_k^2 random variable?

Here is a result that is relevant to our test statistics. Suppose $(M_1, \ldots, M_k)$ follow a multinomial distribution with parameters $(n, p_1, \ldots, p_k)$. Then for n large,

$$\frac{(M_1 - np_1)^2}{np_1} + \ldots + \frac{(M_k - np_k)^2}{np_k}$$

has an approximate $\chi_(^2 k-1)$ distribution. One rule of thumb says that when $np_j >= 5$ for $j = 1, \ldots, k$, then the approximation holds.

This is where the χ^2 test gets its name.

The degrees of freedom are $k-1$ instead of k because of the constraint that $M_1, \ldots, M_k$ sum to n, so we don't really have k freely varying random variables.

We will establish this result for the special case of $k = 2$, i.e. the binomial case.

$$\frac{(M_1 - np_1)^2}{np_1} + \frac{(M_2 - np_2)^2}{np_2}$$

Find a common denominator, and use the fact that $p_2 = 1 - p_1$ to show that the above equals

$$\frac{(M_1 - np_1)^2}{np_1 p_2}$$

Explain why the result now follows.