18.650: Fundamentals of Statistics	October 23, 2023

Lecture 17 — Chi-squared Goodness of Fit Test

Prof. Philippe Rigollet | *Scribe: Anya Katsevich*

We'll talk about the Chi-squared test, also denoted χ^2, or Chi2. But first we need to introduce the χ^2 *distribution*.

1 χ^2 distribution

Definition 1.1: χ^2 distribution

A random variable X has a χ^2 distribution with k degrees of freedom (k is an integer), and we write $X \sim \chi^2_k$, iff

$$X \overset{d}{=} Z_1^2 + \cdots + Z_k^2,$$

where $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

See AoS for the pdf formula, and Figure 1 for a visual depiction of the pdf. As k increases, the mass slouches more and more to the right.

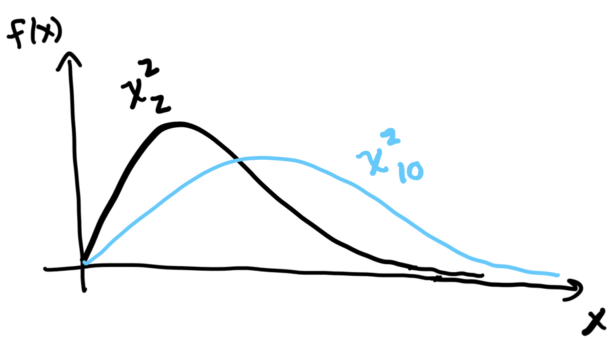

Figure 1: The pdf of the chi squared distribution with $k = 2$ and $k = 10$ degrees of freedom.

Properties:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[Z_1^2 + \cdots + Z_k^2] = \mathbb{E}[Z_1^2] + \cdots + \mathbb{E}[Z_k^2] = k, \\ \mathbb{V}[X] &= k\mathbb{V}[Z_1^2] = k\left(\mathbb{E}[Z_1^4] - \mathbb{E}[Z_1^2]^2\right) = 2k. \end{aligned} \tag{1}$$

So both the average and the spread of the χ^2_k distribution increases as the number of degrees of freedom increases. We let $\chi^2_{k,\alpha}$ denote the αth quantile of the χ^2_k distribution. In other words,

$$\mathbb{P}(X > \chi^2_{k,\alpha}) = \alpha, \quad \text{where} \quad X \sim \chi^2_k.$$

See Figure 2.

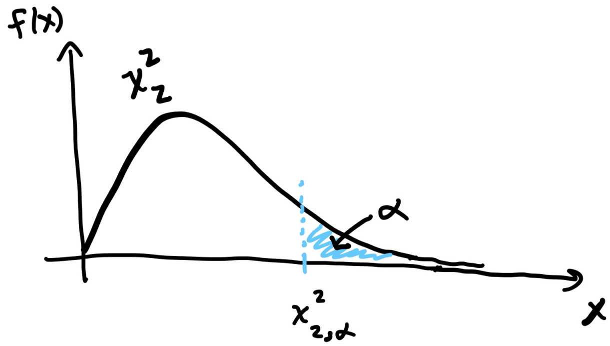

Figure 2: The α-quantile of a χ^2_k distribution is denoted $\chi^2_{k,\alpha}$

2 Goodness of Fit

Suppose we have a hypothesis that a discrete random variable has a certain pmf f_0. Testing this hypothesis is known as a goodness of fit test.

Definition 2.1: Goodness of fit

Let X be a discrete random variable with pmf f that takes k values $\{1, 2, \ldots, k\}$. A *goodness of fit test* is a test to determine whether $H_0 : f = f_0$ or $H_1 : f \neq f_0$.

Remark.

> There's a related test called "χ^2 test for independence" of discrete random variables X and Y. To test whether X and Y are independent, you test the hypothesis that the joint pmf of X and Y is given by the product of the marginal pmfs of X and Y (if this holds, then X is independent of Y). Therefore, this is a special case of testing whether a random variable (now a random vector (X, Y)) has a certain distribution. To read about the χ^2 test for independence, see Chapter 15 in AoS.

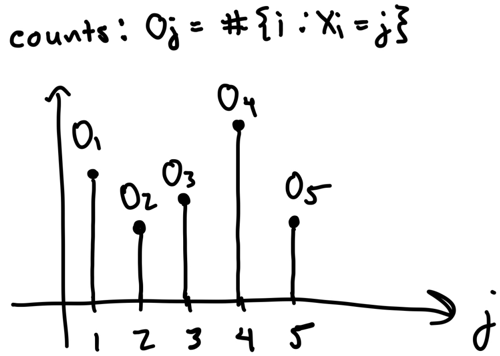

Figure 3: Given observations $X_1, \dots, X_n$ which can take any integer value between 1 and k, we let the count O_j be the number of j's observed in the data.

Suppose we observe $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f$. We tally how many of the X_i equal 1, how many equal 2, and so on. In general, let O_j denote the number of j's we saw in the data; see Figure 3. Now, suppose the null hypothesis is correct, meaning the pmf of $X_1, \dots, X_n$ is $f = f_0$. Then we expect that $O_j \approx nf_0(j)$, i.e. approximately $nf_0(j)$ of the n samples equal j. To prove this, note that we can write O_j as

$$O_j = \sum_{i=1}^{n} \mathbb{1}(X_i = j). \tag{2}$$

Then

$$E_j := \mathbb{E}[O_j] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}(X_i = j)] = \sum_{i=1}^{n} \mathbb{P}(X_i = j) = nf_0(j) \tag{3}$$

under the null. This motivates building a test statistic that aggregates how far all the O_j's are from their corresponding E_j's. The following χ^2 test statistic does this for us:

$$T := \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \tag{4}$$

Clearly, we should reject the null if T is too big. To make this quantitative, we need to know the probability distribution of the random variable T. This is given by the following theorem.

Theorem 2.2: Limit of test statistic T

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_0$, and let O_j be as in (2), $E_j = nf_0(j)$ as in (3), and T as in (4). Then

$$T \rightsquigarrow \chi^2_{k-1} \quad \text{as} \quad n \to \infty.$$

To see why we don't have k degrees of freedom, note that the O_j's sum to n, and this removes one of the degrees of freedom.

Definition 2.3: Pearson's χ^2 test for goodness of fit

Let X be a discrete random variable with pmf f that takes k values $\{1, 2, \ldots, k\}$. Consider the hypothesis test $H_0 : f = f_0$ vs $H_1 : f \neq f_0$. Pearson's χ^2 *goodness of fit test* rejects the null if $T > \chi^2_{k-1,\alpha}$, where T is defined in (4).

We can also compute the p-value given an observed T statistic:

$$\text{p-value} = \mathbb{P}_{H_0}(T > T^{\text{obs}}) = \mathbb{P}(\chi^2_{k-1} > T^{\text{obs}}).$$

Example.

> (Courtesy of ChatGPT) Critics predicted that 50% of Stranger Things viewers would say the typical ending of an episode in the fourth season was a cliffhanger, 30% would say it was a happy ending, and 20% would say it was a sad ending. So the null hypothesis is a pmf f_0 with $f_0(\text{cliff}) = 0.5$, $f_0(\text{happy}) = 0.3$, $f_0(\text{sad}) = 0.2$.
>
> 100 MIT students were surveyed, and 48 said "cliffhanger", 40 said "happy ending", and 12 said "sad ending".
>
> | | expected | observed |
> |---|---|---|
> | cliff | 50 | 48 |
> | happy | 30 | 40 |
> | sad | 20 | 12 |
>
> In this case, $T = (48 - 50)^2/50 + (40 - 30)^2/30 + (12 - 20)^2/20 = 6.61$ and the p-value $= \mathbb{P}(\chi^2_2 \geq 6.61) = 0.037$. Here, the 2 comes from $k - 1 = 3 - 1 = 2$. We would reject at level 5%, and any other level above 0.37%.

Example.

In the Bernoulli kiss example, $H_0 : p = 1/2$ vs $H_1 : p \neq 1/2$. We could also write this as testing whether or not the pmf is $(1/2, 1/2)$ i.e. $f_0(0) = f_0(1) = 1/2$. We observed $n = 124$, with 80 turning heads to the right and 44 to the left.

	expected	observed
right	62	80
left	62	44

The T statistic is then $T = (62 - 80)^2/62 + (62 - 44)^2/62 = 10.46$, and the p value is $\mathbb{P}(Z^2 > 10.46) = \mathbb{P}(\chi_1^2 > 10.46) < 0.001$, using that χ_1^2 has the same distribution as Z^2, where Z is standard Gaussian. But note that $\mathbb{P}(|Z| \geq \sqrt{10.46}) = \mathbb{P}(|Z| \geq 3.23)$, which is the Wald statistic we got in a previous lecture! This is not a coincidence and just reflects the fact that $\chi_1^2 = Z^2$.