Stat 312: Lecture 24
Chi-square Goodness-of-fit Test

Moo K. Chung
mchung@stat.wisc.edu

April 24, 2003

Concepts

1. The null hypothesis of interest is

$$H_0 : p_1 = c_1, \cdots, p_k = c_k,$$

$$H_1 : p_j \neq c_j \text{ for some } j.$$

2. The test statistic

$$\chi^2 = \frac{(N_1 - EN_1)^2}{EN_1} + \cdots \frac{(N_k - EN_k)^2}{EN_k}.$$

Assuming $EN_j \geq 5$ for every j,

$$\chi^2 \sim \chi^2_{k-1}.$$

Proof. When $k = 2$ (binomial data), it is easy to prove. Note that $\chi^2 = \frac{(N_1 - np_1)^2}{np_1(1-p_1)}$. We have shown that for large n,

$$\frac{N_1/n - p_1}{\sqrt{p_1(1-p_1)/n}} \sim N(0,1)$$

3. χ^2 goodness-of-fit test can be used to test whether sample data follows a particular distribution.

In-class problems

Example 1. There are four blood types (ABO system). It is believed that 34, 15, 23 and 28 % of people have blood type A, B, AB and O respectively. The blood samples of 100 students were collected: A(12), B(56), AB(2), O(30). Test if the collected blood sample contradicts the previous belief. *Solution.* The null hypothesis would be

$$H_0 : p_A = 0.34, p_B = 0.15, p_{AB} = 0.23, p_O = 0.28.$$

```
> chi <-(12-34)^2/34+(56-15)^2/15+
(2-23)^2/23+(30-28)^2/28
[1] 145.6187
> 1-pchisq(chi,3)
[1] 0
```

Since the P-value is almost 0, we reject H_0.

Example 2. 90 rats proceed down a ramp to one of three doors. Test if the rats have no preference concerning the choice of a door when the observed frequencies for the three doors are $n_1 = 23, n_2 = 36, n_3 = 31$. *Solution.* We compute $\chi^2 = 2.87$. The P-value the goodness-of-fit test is $1 - \texttt{pchisq}(2.87, 2) = 0.24$. Since this is somewhat large, we do not reject H_0.

Example 3. The number of accidents X per week at an intersection was checked for 50 weeks: 32weeks with no accident, 12 weeks with 1 accident, 6 weeks with 2 accidents. Test if X follows a Poission distribution.
Solution. $H_0 : P(X = x) = \lambda^x e^{-\lambda}/x!$ for all $x = 0, 1, 2, \cdots$. Since λ unknown, we estimate it with MLE. $\hat{\lambda} = \bar{x} = (12 + 2 \times 6)/50 = 0.48$. Define three categories by $X = 0, X = 1$ and $X \geq 2$. The expected probabilities are estimated to give $\hat{p}_1 = 0.619, \hat{p}_2 = 0.297, \hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2 = 0.084$. Then the test statistic $\chi^2 \sim \chi^2_1$ (the additional degree of freedom is lost due to the estimation of λ). $\chi^2 = 1.354$ and the P-value would be 0.25. We do not reject H_0.

Self-study problems

Example 14.2. 14.8, 14.9.