Stat 312: Lecture 24
Goodness-of-fit Test for Distributions

Moo K. Chung
mchung@stat.wisc.edu

April 29, 2003

Concepts

1. Knowing the exact distribution of a sample is important for statistical analysis. We want to test if a sample $x_1, \cdots, x_n$ follows a certain probability distribution P. We perform a χ^2 goodness-of-fit test on the null hypothesis

$$H_0 : P(X \in C_i) = p_i \text{ for all } i = 1, \cdots, k$$

For a continuous distribution, C_i are intervals.

In-class problems

Example 1. Let us see if the distribution of output tuft weight (Example 14.11.) follows an exponential distribution with parameter λ.

Interval	0-8	8-16	16-24	24-32	32-40	40-48	48-56	56-64	64-70
Observed #	20	8	7	1	2	1	0	1	0

Solution. The density of the exponential random variable X is given by $f(x) = \frac{1}{\lambda}e^{-x/\lambda}, x \geq 0$. Note that $\int_0^8 \frac{1}{\lambda}e^{-x/\lambda} \doteq 0.5$. Solving this, we estimate $\hat{\lambda} = 11.54$. Based on this, we compute the expected number of elements in each category. For instance, the expected number of element in the interval $[8, 16]$ is $40 \int_0^8 \frac{1}{11.54}e^{-x/11.54}\, dx = 40(\texttt{pexp}(16, 1/11.54) - \texttt{pexp}(8, 1/11.54))$.

```
> interval<-c(8,16,24,32,40,48,56,64,70)
> p<- pexp(interval,1/r)
[1] 0.50 0.75 0.87 0.94 0.97 0.98 0.99 0.996 0.998
> p[2:9]-p[1:8]
[1] 0.25 0.12 0.06 0.03 0.016 0.008 0.004 0.002
> 40*(p[2:9]-p[1:8])
10  5  2.5  1.3  0.6 0.3 0.2 0.06
```

Expected #	20	10	5	2.5	1.3	0.6	0.3	0.2	0.06

Then we compute $\chi^2 = 50$ and the P-value is $\texttt{pchisq}(50, 7) = 0$. So we reject H_0 and conclude that the exponential distribution is not a good fit.

Self-study problems

Example 14.11. This is slightly different from the in-class example setting. Exercise 12.14.

Assignment 8

Due May 8, 11:00am. 14.6., 14.16., 14.18., 14.26., 14.32.