Hypothesis Testing

STAT 244
Lecture 15
2/25/16

Recap:

① "Simple" Hypotheses (test one distinct distribution for the data X vs another)

$$H_0: \theta = \theta_0 \quad vs \quad H_1: \theta = \theta_1$$

NP: use likelihood ratio to test:

density or pmf $\rightarrow$ $\dfrac{P(data|\theta_1)}{P(data|\theta_0)}$ Reject H_0 if too large

Procedure: (a) Find form of test from Likelihood Ratio (e.g. "Rej. if $X > c$")

(b) Fix $\alpha = P(\text{reject } H_0 | H_0 \text{ true})$

② "Composite" hypotheses (test one distribution vs a set of distributions)

e.g. $H_0: \theta = \theta_0 \quad vs \quad H_1\ \theta > \theta_0$

Sometimes the test from ① is the same for all $\theta_1 > \theta_0$; then the test is UMP ("uniformly most powerful")

LR test: $\lambda = \dfrac{\max_{H_0\ \theta\text{'s}} L(\theta_1, \ldots \theta_n)}{\max_{\text{all } \theta\text{'s}} L(\theta_1, \ldots, \theta_n)}$

Reject if $\lambda < \lambda_c$

Composite tests like $H_1: \mu > \mu_0$ are great when possible, but the most intellectually revealing examples involve multiple distinct outcomes.

Classic Example: Weldon's Dice

Weldon and assistants rolled 12 dice 26,306 times and counted the number showing 5 or 6 up.

ARE THE DICE FAIR??

Results:

No. of Dice X showing 5 or 6	Observed	Theory	Difference
0	185	203	–18
1	1149	1217	–68
2	3265	3345	–80
3	5475	5576	–101
4	6114	6273	–159
5	5194	5018	176
6	3067	2927	140
7	1331	1254	77
8	403	392	11
9	105	87	18
10	14	13	1
11	4	1	3
12	0	0	0
Total	26,306	26,306	0

"Theory" column assumes $X \sim Bin(12, \frac{1}{3})$,

so $26,306 \cdot \binom{12}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{10} \approx 3,345.366$.

$\theta = \binom{12}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{10} = 0.128$

$\sqrt{n\theta(1-\theta)} = 54$, so

$X=2$ column only 1.48 std. dev

Weldon said agreement good, dice fair.

Karl Pearson said: "No way ~ dice loaded"

need 13 tests!?

Last time, we introduced the

<u>Multinomial</u> <u>Distribution</u>,

a generalization of the Binomial Distribution for K distinct outcomes.

We found that for this dist,

$$L(\theta_1, \dots, \theta_k) = \frac{n!}{X_1!\, X_2! \cdots X_k!}\, \theta_1^{x_1} \cdots \theta_k^{x_k}$$

(k outcomes, n trials)

$$\hat{\theta}_k = \frac{X_i}{n}$$

$H_0 : \theta_1 = a_1, \dots \theta_k = a_k$

H_1 : "otherwise", one equality in H_0 does not hold.

$m_i = n a_i = E(X_i | H_0)$

$$\lambda = \frac{L(a_1, \dots, a_k)}{L(\hat{\theta}_1, \dots, \hat{\theta}_k)} = \frac{L(a_1, \dots, a_k)}{L\left(\frac{X_1}{n}, \dots \frac{X_k}{n}\right)} = \left(\frac{m_1}{X_1}\right)^{X_1} \cdots \left(\frac{m_k}{X_k}\right)^{X_k}$$

$$-\log \lambda = -\sum_{i=1}^{K} \log \left[\frac{m_i}{X_i}\right]^{X_i} = \sum_{i=1}^{k} X_i \log\left(\frac{X_i}{m_i}\right) > c.$$

We further showed that

$$-\log \lambda \cong \frac{1}{2} \sum \frac{(X_i - m_i)^2}{m_i} = \frac{1}{2} \chi^2$$

For "large" n, tests equivalent

("large" means all $m_i \geq 3.5$)

$$\chi^2 = \sum_{\substack{\text{all} \\ \text{categories}}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

How is χ^2 distributed? We'll do the $k=2$ case: $(X_2 = n - X_1;\ a_2 = 1-a_1,\ m_2 = n - m_1)$

$$\chi^2 = \sum_{i=1}^{2} \frac{(X_i - m_i)^2}{m_i}$$

$$= \frac{(X_1 - m_1)^2}{m_1} + \frac{(X_2 - m_2)^2}{m_2}$$

$$= \frac{(X_1 - na_1)^2}{na_1} + \frac{((n - X_1) - n(1-a_1))^2}{n(1-a_1)}$$

$$= \frac{(X_1 - na_1)^2}{na_1} + \frac{(X_1 - na_1)^2}{n(1-a_1)}$$

$$= \frac{(X_1 - na_1)^2}{na_1(1-a_1)}$$

$$= \left(\frac{X_1 - na_1}{\sqrt{na_1(1-a_1)}}\right)^2$$

X_1 is the result of adding n trials, $E(X) = na_1$, $Var(X) = na_1(1-a_1)$ SO WE INVOKE THE CLT which says that approximately,

$$\left(\frac{X_1 - na_1}{\sqrt{na_1(1-a_1)}}\right) \sim N(0,1)$$

Hence χ^2 is distributed as X^2, where $X \sim N(0,1)$, Hence χ^2 is chi-square distributed with 1 degree of freedom for $K=2$.

4

In general, for K outcomes we have $K-1$ degrees of freedom.

Moreover:

$$E(\chi^2) = E\sum_{i=1}^{K} \frac{(X_i - na_i)^2}{na_i}$$

$$= \sum_{i=1}^{K} \frac{E(X_i - na_i)^2}{na_i}$$

$$= \sum_{i=1}^{K} \frac{Var(X_i)}{na_i}$$

$$= \sum_{i=1}^{K} \frac{na_i(1-a_i)}{na_i}$$

$$= \sum_{i=1}^{K} (1-a_i) = K - \sum_{i=1}^{K} a_i$$

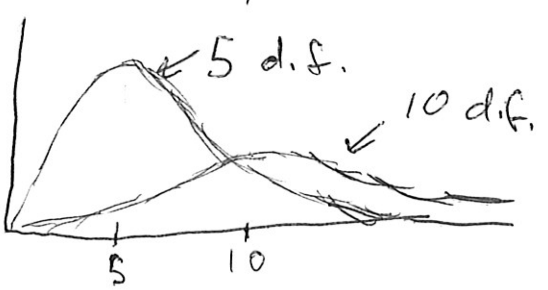
$$\boxed{= K-1}$$

$$\chi^2 = \sum \frac{(obs - Exp)^2}{Exp} = \sum_{i=1}^{n} \frac{(X_i - na_i)}{na_i}$$

Reject H_0 if $\chi^2 > C$

C from χ^2 dist $K-1$ d.f.

[essentially an LR test]



Outcome: A_i

Probability: θ_i

Expected count: $n\theta_i$

H_0 Prob: a_i

H_0 Expected Counts na_i

Observed Counts: X_i

Counts total n

Probs total 1

Multinomial Dist.

Weldon's data have too few counts in $X = 11$ or $X = 12$ to use the χ^2 approximation, so let's group 10, 11, and 12:

Table 7.2. Weldon's dice data with the last three categories grouped together.

No. of Dice X showing 5 or 6	Observed	Theory	Difference
0	185	203	–18
1	1149	1217	–68
2	3265	3345	–80
3	5475	5576	–101
4	6114	6273	–159
5	5194	5018	176
6	3067	2927	140
7	1331	1254	77
8	403	392	11
9	105	87	18
10 – 12	18	14	4
Total	26,306	26,306	0

$$\chi^2 = (-18)^2/203 + \dots + (4)^2/14 = 35.5 \text{ (no roundoff)}$$

$K = 11$, so d.f $= K - 1 = 10$

Table in Rice says $\chi^2_{10} = 25.19$ for $Pr(H_0) = 0.005$

In fact, $\chi^2_{10} = 35.5 \Rightarrow Pr(H_0) \sim 10^{-4}$

Karl Pearson: H_0 is "intrinsically incredible" ... correct!

What if we need to find H_0 from the data?

Instead of $H_0: X \sim Bin(12, \frac{1}{3})$, what if we want to consider

$$H_0: X \sim Bin(12, \theta) \text{ for some } \theta ?$$

<u>General Problem:</u> K cells, $\theta_i = p(i)$

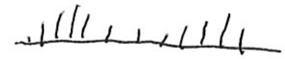
$H_0: \theta = a_i(\theta)$, $i = 1, \ldots, K$

H_1: "otherwise"

<u>Ex:</u> $a_i(\theta) = \binom{12}{i} \theta^i (1-\theta)^{12-i}$

$i = 0, 1, \ldots, 12$ (13 cells)

H_0: <u>Some</u> Binomial 

H_1: "other"  etc etc ...

We procede as before with Weldon's data, but with 2 changes:

1. a_i replaced by $a_i(\hat{\theta})$ and m_i replaced by $n\, a_i(\hat{\theta})$, where $\hat{\theta}$ = MLE of θ, assuming H_0 holds.

2. d. f. $= K - 1 - 1 = K - 2$ ← price of estimating one parameter

Results:

$$\hat{\theta} = \frac{\#\ 5\text{'s or }6\text{'s}}{\#\ \text{trials}} = \frac{(0)(185)+(1)(1149)+(2)(3265)+\cdots}{(12)(26,306)}$$

$$= 0.33769862 \quad \left(\text{from ungrouped data}\right)$$

recompute the table:

Table 7.2. Weldon's Dice Data. The Theory column has been recomputed using the maximum likelihood estimate of the probability of a 5 or 6, namely 0.33769862.

No. of Dice X showing 5 or 6	Observed	Theory	Difference
0	185	187.4	-2.4
1	1149	1146.5	2.5
2	3265	3215.2	49.8
3	5475	5464.7	10.3
4	6114	6269.3	-155.3
5	5194	5114.7	79.3
6	3067	3042.5	24.5
7	1331	1329.7	1.3
8	403	423.8	-20.8
9	105	96.0	9.0
10	14	14.7	-0.7
11	4	1.4	2.6
12	0	0.1	-0.1
Total	26,306	26,306	0.0

group these, as before

$$\chi^2 = \frac{(185-187.4)^2}{187.4} + \frac{(1149-1146.5)^2}{1146.5} + \cdots$$

$$= 8.2, \text{ with } 11-1-1 = 9 \text{ d.f.}$$

close to expectation (and median) of χ^2, Ho strongly supported.

We <u>can</u> compare these hypotheses, as long as degrees of freedom included should have grouped same way, but effect small.

8

Other types of data can sometimes be treated by the multinomial distr, and hence χ^2. (other dists that give rise to contingency tables can also be treated by χ^2 techniques, since we derived χ^2 by taking a Taylor expansion around a likelihood maximum)

Contingency Tables

(a cross classification of data into 2 (in general N) categories)

Ex. Galton's Data

Hus \ Wife	T	M	S	
T	18	28	14	60
M	20	51	28	99
S	12	25	9	46
	50	104	51	205

3 × 3 Table

$n = 205$ counts

$K = 3 \times 3 = 9$ cells

In General

Cell counts X_{ij} $\quad i = 1, \ldots, r$ ← rows, $\quad j = 1, \ldots, c$ ← columns

$r \times c$ table: notation

Factor B

$$
\begin{array}{c|cccc|l}
 & B_1 & B_2 & \cdots & B_c & \\
\hline
A_1 & X_{11} & X_{12} & \cdots & X_{1c} & X_{1+} = \sum_{j=1}^{c} X_{1j} \\
A_2 & X_{21} & \cdots & & & \\
\vdots & \vdots & \cdots & \boxed{X_{ij}} & & \leftarrow \text{row } i, \text{ column } j \\
 & & & & X_{rc} & \\
A_r & X_{r1} & & & & \\
\hline
 & X_{+1} = \sum_{i=1}^{r} X_{i1} & & & &
\end{array}
$$

Factor A

(multinomial contingency tables, cont'd)

$r \times c$ cells.

$\theta_{ij} = P(\text{Trial gives } A_i \cap B_j)$, $\sum_i \sum_j \theta_{ij} = 1$

$X_{ij} = \#$ trials with $A_i \cap B_j$, $\sum_i \sum_j X_{ij} = n$

X_{ij}'s multinomial n trials, probs θ_{ij}

$\theta_{ij} = P(A_i \cap B_j)$

H_0: Factors independent

<u>or</u> $P(A_i \cap B_j) = P(A_i)\,P(B_j)$

<u>or</u> $\theta_{ij} = (\theta_{i+})(\theta_{+j})$ $\quad (\theta_{i+} = \sum_j \theta_{ij})$

(e.g. $P(\text{Hus T} \cap \text{Wf T}) = P(\text{Hus T})\,P(\text{Wf T})$)

H_1: "otherwise". in detail:

H_1 is <u>Composite hypothesis</u>: ($a_{ij}(\theta_{i+}, \theta_{+j})$)

comma

depends on $\begin{cases} \theta_{1+} \; \theta_{2+} \ldots \theta_{r+} \\ \theta_{+1} \; \theta_{+2} \ldots \theta_{+c} \end{cases}$

$(r-1) + (c-1)$ parameters

MLE's $\hat{\theta}_{i+} = \dfrac{X_{i+}}{n}$, $\hat{\theta}_{+j} = \dfrac{X_{+j}}{n}$

Ex. Galton Data

$$\frac{X_{1+}}{n} = \frac{18+28+14}{205} \quad \text{estimates } P(\text{H us T})$$

So, under H_0, MLE of

θ_{ij} is $\hat{\theta}_{i+} \cdot \hat{\theta}_{+j} = \dfrac{X_{i+} X_{+j}}{n \cdot n}$

and the MLE of $m_{ij} = n\,\theta_{ij}$ is

$$n \cdot \frac{X_{i+} X_{+j}}{n \cdot n} = \frac{X_{i+} X_{+j}}{n}$$

$$= \frac{(\text{row total})(\text{col total})}{(\text{total})}$$

$$\chi^2 = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{\left(X_{ij} - \frac{X_{i+} X_{+j}}{n}\right)^2}{\left(\frac{X_{i+} X_{+j}}{n}\right)}$$

for the Galton Data:

$$\chi^2 = \frac{\left(18 - \frac{60 \cdot 50}{205}\right)^2}{\frac{60 \cdot 50}{205}} + \cdots \quad \text{(9 terms)}$$

$$= 2.91$$

$$df = r \cdot c - 1 - [(r-1) + (c-1)]$$

"K" -1 $-$ "m"

$$= (r-1)(c-1) = 2 \cdot 2 = 4$$

χ^2 even less than expected value!