STAT 249
3/11/16
Lecture 16

Review

Testing Composite Hypotheses

H_0: Group of θ_0's \
H_1: Group of θ_1's / Mutually Exclusive

Likelihood Ratio Test

Reject H_0 if $\lambda < \lambda_c$ $(0 \le \lambda \le 1$ always$)$

$$\lambda = \frac{\max_{\theta_0\text{'s}} L(\theta)}{\max_{\text{all } \theta\text{'s}} L(\theta)}$$

$\max_{\theta_0\text{'s}} L(\theta)$ ← Restricted to H_0

$\max_{\text{all } \theta\text{'s}} L(\theta)$ ← No Restrictions

[Alternative, equivalent version:
Reject H_0 if $-2 \log \lambda > k$]

Multinomial Distrib. k dimensional

$X = (X_1, X_2, \ldots, X_k)$, $X_1 + \cdots + X_k = n$

(n indep trials, $A_1, \ldots, A_n$ outcomes,
$X_i = \# A_i$'s, $\theta_i = P(X_i)$, $\sum \theta_i = 1$)

Ex: Roulette
n spins $\quad k = 38$

Marginal Dist. X_i is Binomial(n, θ_i)
X_i's NOT indep!

Testing Roulette Wheel

χ^2 test is approx LR test.

k cells

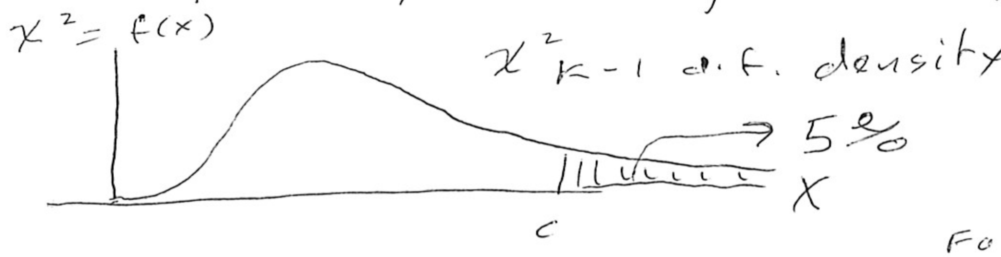
$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}} = \sum_{i=1}^{k} \frac{(X_i - na_i)^2}{na_i}$$

$a_i = H_0$ prob of cell i

$na_i = m_i =$ expected counts under H_0

Reject if $\chi^2 > C$

C from χ^2 table, $k-1$ d.f.



Contingency Tables

X_{ij} multinomial n trials probs θ_{ij}

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

H_0: Factors indep

$\theta_{ij} = (\theta_{i+})(\theta_{+j})$

H_1: "otherwise"

We can use the multinomial dist because we can regard the $r \times c$ cells as K outcomes. What if margins are fixed?

$r \times c$ table

$n = \sum_i \sum_j$ trials

$i = 1, \ldots, r$
$j = 1, \ldots, c$

cell counts X_{ij}

$X_{i+} = \sum_{j=1}^{c} X_{ij}$

$X_{+j} = \sum_{i=1}^{r} X_{ij}$

Tests of Homogeneity – Draft Lottery

row totals fixed

Table 7.4.

Drawing numbers	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Totals
	Months												
1-122	9	7	5	8	9	11	12	13	10	9	12	17	122
123-244	12	12	10	8	7	7	7	7	15	15	12	10	122
245-366	10	10	16	14	15	12	12	11	5	7	6	4	122
Totals	31	29	31	30	31	30	31	31	30	31	30	31	366

Example 7.F (Right-Handedness). To what degree is the propensity to be right-handed socially determined? Is it the same in different cultures? In different historical epochs? Two psychologists addressed this question by examining works of art that portrayed activities that could be judged as being done right- or left-handedly. (Stanley Coren and Clare Porac, "Fifty Centuries of Right-Handedness: The Historical Record" Science (1977), Vol. 198, pp. 631-632.) The following tables summarize their findings, looking at the data in two different ways.

Table 7.5. Counts of 1180 art works showing activity that can be categorized as left- or right-handed, (a) by geographical area, and (b) by historical epoch.

(a)	Right	Left	Total	% Right
Central Europe	312	23	335	93%
Medit. Europe	300	17	317	95%
Middle East	85	4	89	96%
Africa	105	12	117	90%
Central Asia	93	8	101	92%
Far East	126	13	139	91%
Americas	72	10	82	88%
Total	1093	87	1180	92.6%

Left and Right-Handedness by Historical Epoch or Location

(b)				
Pre 3000 BC	35	4	39	90%
2000 BC	44	7	51	86%
1000 BC	89	10	99	90%
500 BC	134	8	142	94%
~0 BC	130	4	134	97%
AD 500	39	3	42	93%
AD 1000	57	7	64	89%
AD 1200	40	1	41	98%
AD 1400	44	6	50	88%
AD 1500	63	5	68	93%
AD 1600	68	4	72	94%
AD 1700	66	5	71	93%
AD 1800	95	6	101	94%
AD 1850	38	1	39	97%
AD 1900	71	6	77	92%
AD 1950	80	10	90	89%
Total	1093	87	1180	92.6%

column totals fixed

Ex Adaptation in Evolution
(& The "MK" Test)

Adaptive protein evolution at the *Adh* locus in *Drosophila*

John H. McDonald & Martin Kreitman

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

PROTEINS often differ in amino-acid sequence across species. This difference has evolved by the accumulation of neutral mutations by random drift, the fixation of adaptive mutations by selection, or a mixture of the two. Here we propose a simple statistical test of the neutral protein evolution hypothesis based on a comparison of the number of amino-acid replacement substitutions to synonymous substitutions in the coding region of a locus. If the observed substitutions are neutral, the ratio of replacement to synonymous fixed differences between species should be the same as the ratio of replacement to synonymous polymorphisms within species. DNA sequence data on the *Adh* locus (encoding alcohol dehydrogenase, EC 1.1.1.1) in three species in the *Drosophila melanogaster* species subgroup do not fit this expectation; instead, there are more fixed replacement differences between species than expected. We suggest that these excess replacement substitutions result from adaptive fixation of selectively advantageous mutations.

TABLE 2 Number of replacement and synonymous substitutions for fixed differences between species and polymorphisms within species

	Fixed	Polymorphic
Replacement	7	2
Synonymous	17	42

	Fixed (DNA changes between species)	Polymorphic (DNA changes within species)	Total
replacement (DNA changes, protein changes)	7	2	$9 = X_{1+}$
synonymous (DNA changes, protein stays the same)	17	42	$59 = X_{2+}$

row totals fixed

Fixed Margins:

Product - Multinomial, or
Testing Homgeneity of proportions.

Ex 1st Word Usage - James Mill and John Stuart Mill

	I	II	
James	X_{11}	X_{12}	$X_{1+} = 1075$
J. S.	X_{21}	X_{22}	$X_{2+} = 451$
			$n = 1626$

X_{1+}, X_{2+} fixed, given
X_{ij} multinomial <u>within rows</u>.

Ex: 200 applic. to UC (Grad)
100 male, 100 female

	Adm	Deny	Adm no aid	
M	X_{11}	X_{12}	X_{13}	$X_{1+} = 100$
F	X_{21}	X_{22}	X_{23}	$X_{2+} = 100$

H_0: $P(I / \text{James}) = P(I / J.S.)$
or $P(\text{Adm} \mid M) = P(\text{Adm} / F)$
or $P(\text{synonymous} / \text{fixed}) = P(\text{synonymous} / \text{polymorphic})$

OR $P(B_j / A_i) = P(B_j)$ all i, j

Test of Homogeneity

H_0: $(X_{i1}, \ldots, X_{ic})$ Multinomial
X_{i+} trials
probs $(b_1, \ldots, b_c)$ the <u>same</u>
for all rows i

Under H_0:

$$L(b_1, b_2, \ldots, b_c) = \prod_{i=1}^{r} \left(\frac{X_{i+}!}{\prod_{j=1}^{c} X_{ij}!} b_1^{X_{i1}} \cdots b_c^{X_{ic}} \right)$$

$$= \begin{pmatrix} \text{involves} \\ X\text{'s} \end{pmatrix} b_1^{X_{+1}} \cdots b_c^{X_{+c}}$$

$$\hat{b}_j = X_{+j}/n$$

Under H_1:

$$\max L = \begin{pmatrix} \text{involves} \\ X\text{'s} \end{pmatrix} \prod_{i=1}^{r} \left(\frac{X_{i1}}{X_{i+}}\right)^{X_{i1}} \cdots \left(\frac{X_{ic}}{X_{i+}}\right)^{X_{ic}}$$

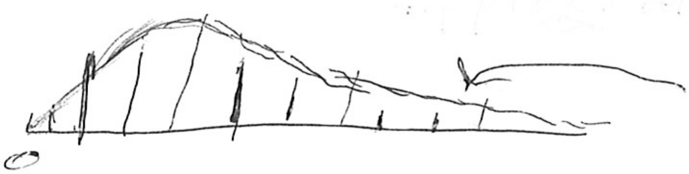
$$\lambda = \prod_{i=1}^{r} \prod_{j=1}^{c} \left(\frac{m_{ij}}{X_{ij}}\right)^{X_{ij}}, \quad m_{ij} = \frac{X_{i+} X_{+j}}{n}$$

Same test as in full multinomial situation! Same χ^2 statistic, same d.f.

Key Points

χ^2 is for large samples—
χ^2 distribution is an approx
to the distribution given Ho
(the actual distribution is
discrete)

("Expected" ≥ 3.5)



χ^2's were originally designed to avoid being misled by deviation selected as "large"

Ex: Roulette

Ex: Authorship

Ex: Evolutionary Adaptation

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

$H_o = E(\text{count} \mid H_o)$ or, when this is incompletely specified, the MLE of $E(\text{count} \mid H_o)$

Uses	d.f.
① Test fit, H_o completely specified [Ex: Roulette] [Ex: Weldon's dice, $P(5 \cup 6) = 1/3$]	$K - 1$
② Test fit, H_o not completely specified [Ex: Weldon's dice as test of Binomial dist, $P(5 \cup 6) = \hat{\theta}$]	$K - 1 - \binom{\#\text{ params}}{\text{estimated}}$
③ Test independence [Ex: Galton, tall/short spouse selection]	$(r-1)(c-1)$
④ Test homogeneity (fixed margin) [Ex. MK test for selection, draft lottery]	$(r-1)(c-1)$

22 d.f.

$\chi^2 = 32.16$

Table 7.4.

Drawing numbers	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Totals
	Months												
1-122	9	7	5	8	9	11	12	13	10	9	12	17	122
123-244	12	12	10	8	7	7	7	7	15	15	12	10	122
245-366	10	10	16	14	15	12	12	11	5	7	6	4	122
Totals	31	29	31	30	31	30	31	31	30	31	30	31	366

Example 7.F (Right-Handedness). To what degree is the propensity to be right-handed socially determined? Is it the same in different cultures? In different historical epochs? Two psychologists addressed this question by examining works of art that portrayed activities that could be judged as being done right- or left-handedly. (Stanley Coren and Clare Porac, "Fifty Centuries of Right-Handedness: The Historical Record" Science (1977), Vol. 198, pp. 631-632.) The following tables summarize their findings, looking at the data in two different ways.

Table 7.5. Counts of 1180 art works showing activity that can be categorized as left- or right-handed, (a) by geographical area, and (b) by historical epoch.

(a)

	Right	Left	Total	% Right
Central Europe	312	23	335	93%
Medit. Europe	300	17	317	95%
Middle East	85	4	89	96%
Africa	105	12	117	90%
Central Asia	93	8	101	92%
Far East	126	13	139	91%
Americas	72	10	82	88%
Total	1093	87	1180	92.6%

6 d.f.

$\chi^2 = 8.14$

(b)

Pre 3000 BC	35	4	39	90%
2000 BC	44	7	51	86%
1000 BC	89	10	99	90%
500 BC	134	8	142	94%
~0 BC	130	4	134	97%
AD 500	39	3	42	93%
AD 1000	57	7	64	89%
AD 1200	40	1	41	98%
AD 1400	44	6	50	88%
AD 1500	63	5	68	93%
AD 1600	68	4	72	94%
AD 1700	66	5	71	93%
AD 1800	95	6	101	94%
AD 1850	38	1	39	97%
AD 1900	71	6	77	92%
AD 1950	80	10	90	89%
Total	1093	87	1180	92.6%

15 d.f.

$\chi^2 = 17.4$

But how to interpret?

P - values

Classical Testing:

- fix level of α (say, .05)
- choose test (preferably high power)
- look at data, accept or reject α

[works well with tables, e.g. in Rice]

Ex.: H_0: Factors in 3×4 table indep.

Test: Reject if $\chi^2 > 12.59$ $(2 \times 3 = 6 \text{ d.f})$

Observe $\chi^2 = 12.50$ accept indep.

Observe $\chi^2 = 12.60$ reject indep.

Are these two situations really different??

P-value: The smallest level α at which you would reject H_0

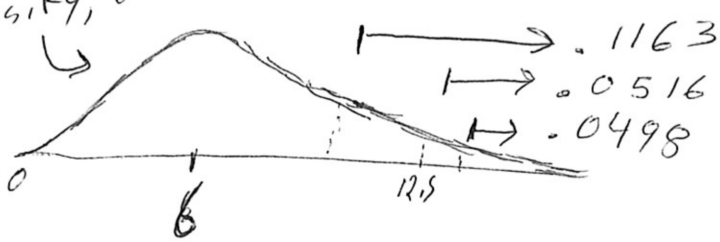
Ex: (as above)

$\chi^2 = 12.50 \rightarrow P = .0516$

$\chi^2 = 12.60 \rightarrow P = .0498$

$\chi^2 = 10.20 \rightarrow P = .1163$

Chi-square density, 6 d.f.



Can report P, or if only tables are available, interval: $P > .1$ or $.01 < P < .05$ or $P < .01$ or $P \ll .01$

d.f. = 22
$\chi^2 = 32.16$
p-value 2%

Table 7.4.

Drawing numbers	Months												Totals
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
1-122	9	7	5	8	9	11	12	13	10	9	12	17	122
123-244	12	12	10	8	7	7	7	7	15	15	12	10	122
245-366	10	10	16	14	15	12	12	11	5	7	6	4	122
Totals	31	29	31	30	31	30	31	31	30	31	30	31	366

Example 7.F (Right-Handedness). To what degree is the propensity to be right-handed socially determined? Is it the same in different cultures? In different historical epochs? Two psychologists addressed this question by examining works of art that portrayed activities that could be judged as being done right- or left-handedly. (Stanley Coren and Clare Porac, "Fifty Centuries of Right-Handedness: The Historical Record" Science (1977), Vol. 198, pp. 631-632.) The following tables summarize their findings, looking at the data in two different ways.

Table 7.5. Counts of 1180 art works showing activity that can be categorized as left- or right-handed, (a) by geographical area, and (b) by historical epoch.

(a)	Right	Left	Total	% Right
Central Europe	312	23	335	93%
Medit. Europe	300	17	317	95%
Middle East	85	4	89	96%
Africa	105	12	117	90%
Central Asia	93	8	101	92%
Far East	126	13	139	91%
Americas	72	10	82	88%
Total	1093	87	1180	92.6%

6 d.f.
$\chi^2 = 8.14$
$p > 10\%$

(b)

Pre 3000 BC	35	4	39	90%
2000 BC	44	7	51	86%
1000 BC	89	10	99	90%
500 BC	134	8	142	94%
~0 BC	130	4	134	97%
AD 500	39	3	42	93%
AD 1000	57	7	64	89%
AD 1200	40	1	41	98%
AD 1400	44	6	50	88%
AD 1500	63	5	68	93%
AD 1600	68	4	72	94%
AD 1700	66	5	71	93%
AD 1800	95	6	101	94%
AD 1850	38	1	39	97%
AD 1900	71	6	77	92%
AD 1950	80	10	90	89%
Total	1093	87	1180	92.6%

15 d.f.
$\chi^2 = 17.9$
$p > 10\%$

MK Test

Expected:

	Fixed	Polymorphic
Replacement	3.16	5.81
Synonymous	20.84	38.19

Observed:

	Fixed	Polymorphic	row marginals
	7	2	9
	17	42	59
column marginals	24	44	68

$\chi^2 = 8.27$ 1 d.f.

p - value 0.0042

The authors actually used a "G-test" to get a p-value of 0.006. The G-test uses an <u>exact</u> value of $-\log\lambda = -\sum x_i \log\left(\frac{m_i}{x_i}\right)$ rather than $-\log\lambda \approx \frac{1}{2}\chi^2$. Unfortunately, the G-test still has to approximate the distribution of $-\log\lambda$ by the distribution of χ^2, so either way there is an approximation.

Confidence Intervals

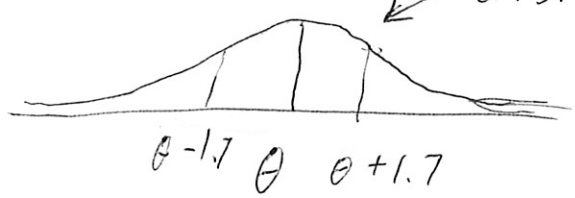
Point estimation

$\hat{\theta}$ estimate (may be from MLE)
(accompanied by SE = "standard error"
= estimated standard deviation)

Ex: $\hat{\theta} = 14.3$, SE = 1.7

If normal:

dist. of $\hat{\theta}$



Confidence intervals (C.I.)

"$[\hat{L}, \hat{U}]$, a 95% C.I."
→ an interval estimate

Ex: "The interval (11.0, 17.6) is an approx 95% CI for θ"

Questions: How to find?
What does the statement mean?
(What is "Confidence")

Finding CI's

Ex: (by Pivotal Method)

A pivotal quantity is a function of observations and unobservable parameters whose distribution does not depend on the parameters:

eg $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Data: $X_1, \ldots, X_n$

Model: X_i's indep. $N(\theta, 25)$

So MLE $\bar{X}$ is $N(\theta, \frac{25}{n})$

So $P(-1.96 \cdot \frac{5}{\sqrt{n}} \le \bar{X} - \theta \le 1.96 \cdot \frac{5}{\sqrt{n}}) = .95$

$P(-\bar{X} - 1.96 \frac{5}{\sqrt{n}} \le -\theta \le -\bar{X} + 1.96 \frac{5}{\sqrt{n}}) = .95$

$P(\bar{X} - 1.96 \cdot \frac{5}{\sqrt{n}} \le \theta \le \bar{X} + 1.96 \frac{5}{\sqrt{n}}) = .95$

The interval $[\hat{L}, \hat{U}]$ is a 95% CI

$\hat{L} = \bar{X} - 1.96 \cdot \frac{5}{\sqrt{n}}$

$\hat{U} = \bar{X} + 1.96 \frac{5}{\sqrt{n}}$

Interpretation: This is a random interval that includes θ with prob. .95

Interval is random

θ is not random

adopt a Bayesian approach as in Chapter 4. The virtue of confidence intervals is that they combine an assessment of accuracy with the estimate; their drawback is the propensity of the statement to be misinterpreted as a Bayesian statement when it is in fact somewhat weaker than that.

Confidence Interval Example. Eighty samples of size n = 25 were taken from an N(350, 25) distribution and the eighty 95% confidence intervals for the mean θ = 350 were computed. In this example, 5 of the 80 missed the target, about as expected.

$\bar{X}$

X-bar	Lower	Upper	Covers?
348.17	346.21	350.13	
351.21	349.25	353.17	
350.15	348.19	352.11	
350.69	348.73	352.65	
348.51	346.55	350.47	
350.69	348.73	352.65	
352.94	350.98	354.90	No
350.35	348.39	352.31	
349.11	347.15	351.07	
348.77	346.81	350.73	
349.88	347.92	351.84	
349.40	347.44	351.36	
349.60	347.64	351.56	
349.39	347.43	351.35	
350.82	348.86	352.78	
350.38	348.42	352.34	
349.62	347.66	351.58	
349.77	347.81	351.73	
350.02	348.06	351.98	
349.81	347.85	351.77	
349.14	347.18	351.10	
349.10	347.14	351.06	
348.47	346.51	350.43	
349.73	347.77	351.69	
348.79	346.83	350.75	
350.43	348.47	352.39	
350.65	348.69	352.61	
349.29	347.33	351.25	
349.17	347.21	351.13	
350.00	348.04	351.96	
349.97	348.01	351.93	
349.60	347.64	351.56	
351.41	349.45	353.37	
350.86	348.90	352.82	
351.28	349.32	353.24	
351.14	349.18	353.10	
349.54	347.58	351.50	
350.59	348.63	352.55	
351.58	349.62	353.54	
350.93	348.97	352.89	

X-bar	Lower	Upper	Covers?
350.44	348.48	352.40	
349.52	347.56	351.48	
347.75	345.79	349.71	No
349.10	347.14	351.06	
349.44	347.48	351.40	
348.47	346.51	350.43	
348.60	346.64	350.56	
349.37	347.41	351.33	
351.37	349.41	353.33	
350.10	348.14	352.06	
349.15	347.19	351.11	
350.97	349.01	352.93	
350.46	348.50	352.42	
350.16	348.20	352.12	
351.29	349.33	353.25	
350.37	348.41	352.33	
348.92	346.96	350.88	
349.25	347.29	351.21	
349.31	347.35	351.27	
351.23	349.27	353.19	
349.99	348.03	351.95	
350.29	348.33	352.25	
350.88	348.92	352.84	
347.41	345.45	349.37	No
349.91	347.95	351.87	
348.53	346.57	350.49	
350.03	348.07	351.99	
352.13	350.17	354.09	No
349.99	348.03	351.95	
350.81	348.85	352.77	
350.14	348.18	352.10	
350.39	348.43	352.35	
349.50	347.54	351.46	
351.29	349.33	353.25	
349.74	347.78	351.70	
351.14	349.18	353.10	
349.89	347.93	351.85	
350.80	348.84	352.76	
347.98	346.02	349.94	No
349.07	347.11	351.03	

$X_1, \ldots, X_{25}$ indep $N(350, 25)$

$\bar{X} = \frac{1}{25} \sum X_i$ $\quad N(350, 1)$

Lower $= \bar{X} - 1.96$ $\quad$ Upper $= \bar{X} + 1.96$

More on the meaning of CIs

Suppose we calculate $\bar{X}$ from data and find

$$\hat{L} = 11.0 \quad \hat{U} = 17.6.$$

Does this mean

$$P(11.0 \leq \theta \leq 17.6) = .95?$$

No → only the interval is random.

Can't make prob. statements about θ without a prior distribution for θ. Our confidence is in the procedure being used - it is 95% reliable, but in a given application that's all we can say.

Could have

$$P(11.0 \leq \theta \leq 17.6 \mid \bar{X}) > .95$$

or $< .95$

even " $= 0$!

(if prior prob $P(\theta > 10) = 0$