NHST Studio

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

You should have downloaded studio9.zip and unzipped it into your 18.05 working directory.

Frequentist vs. Bayesian: likelihood tables

x	0	1	2	3	4	5	6	7	8	9	10
$H_0 : p(x\|\theta = .5)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$H_A : p(x\|\theta = .6)$	.000	.002	.011	.042	.111	.201	.251	.215	.121	.040	.006
$H_A : p(x\|\theta = .7)$	.000	.0001	.001	.009	.037	.103	.200	.267	.233	.121	.028

Likelihood table for test statistic x

Suppose the data gives test statistic $x = 2$.

Frequentist: Look at the entire likelihood table to compute $p = 0.11$.

Bayesian: Use only the $x = 2$ column in the table to update the prior.

hypothesis	prior	likelihood	unnorm. post.	posterior
θ	$P(\theta)$	$P(x = 2 \mid \theta)$	$P(x = 2 \mid \theta)P(\theta)$	$P(\theta \mid x = 2)$
$\theta = .5$	$1/3$	0.044	0.0147	0.7857
$\theta = .6$	$1/3$	0.011	0.0037	0.1964
$\theta = .7$	$1/3$	0.001	0.0003	0.0179
total	1		0.0187	1

Frequentist vs. Bayesian coins

A coin is randomly picked from a drawer.

Experiment: toss the coin 10 times and count the number of heads.

Results: $x = 9$ heads.

(a) Run a significance test with $H_0 =$ 'the coin is fair'.

Use significance level 0.05. Use R to do the computations.

(b) You learn that the drawer contained the following mix of coins with different probabilities of heads:

probability	0.1	0.3	0.5	0.7	0.9
counts	5	5	200	5	5

What is the probability the coin is fair?

(c) Repeat part (b) if the number of fair coins in the drawer was 20 instead of 200.

(d) How are the p-value in part (a) and the probabilities in parts (b) and (c) related?

Solution

All computations were done using `studio9-sol.r`

(a) Let x be the number of heads and θ the probability of heads. The two-sided p-value is

$$p = P(x = 0, 1, 9, 10 \,|\, \theta = 0.5) = 0.021.$$

We reject the null hypothesis at the 0.05 significance level. We conclude that the coin is not fair.

Continued on next slide.

Solution continued

(b) This is a Bayes formula problem:

$$p(\theta = 0.5 \mid x = 9) = \frac{p(x = 9 \mid \theta = 0.5)\, p(\theta = 0.5)}{p(x = 9)}.$$

It was easy to computed the entire update table using R (see `studio9-sol.r`).

Hypothesis θ	prior $p(\theta)$	likelihood $p(x\|\theta)$	unnorm. post.	posterior $p(\theta\|x)$
0.1	0.023	0.000	0.000	0.000
0.3	0.023	0.000	0.000	0.000
0.5	0.909	0.010	0.009	0.434
0.7	0.023	0.121	0.003	0.135
0.9	0.023	0.387	0.009	0.431

The posterior probability that the coin is fair is in blue: 0.434.

Solution continued

(c) This is a repeat of problem (b) with a different prior.

Hypothesis θ	prior $p(\theta)$	likelihood $p(x\|\theta)$	unnorm. post.	posterior $p(\theta\|x)$
0.1	0.125	0.000	0.000	0.000
0.3	0.125	0.000	0.000	0.000
0.5	0.500	0.010	0.005	0.071
0.7	0.125	0.121	0.221	0.221
0.9	0.125	0.387	0.048	0.707

The posterior probability that the coin is fair is in blue: 0.071.

(d) Parts (c) and (d) give actual probabilities that the coin is fair given the data. Their computation depends on having a prior.

The small *p*-value in part (a) is not the probability that the coin is fair. It is computed from the likelihood table. Specifically, it is the probability of seeing such extreme data *given* that the coin is fair.

Board question: Stop

For each of the following experiments (all done with $\alpha = .05$)

(a) Comment on the validity of the claims.

(b) Find the probability of a type I error in each experimental setup.

1. By design Peter did 50 trials and computed $p = .04$.
   He reports $p = .04$ with $n = 50$ and declares it significant.
2. Ruthi did 50 trials and computed $p = .06$.
   Since this was not significant, she started over and computed $p = .04$ based on the next 50 trials.
   She reports $p = .04$ with $n = 50$ and declares it statistically significant.
3. Erika did 50 trials and computed $p = .06$.
   Since this was not significant, she then did 50 more trials and computed $p = .04$ based on all 100 trials.
   She reports $p = .04$ with $n = 100$ and declares it significant.

Hiring and group identity

In an experiment on how group identity affects hiring, a researcher asked HR staff from different companies to evaluate a fictional person's resumé.

- The resumés are identical except for the name of the person.
- The HR staff are asked to give the starting salary they would give this person.
- To analyze the data, the salaries were categorized into four levels.
- The different names were categorized into two groups.
- The dataset also includes publicly available data from the broader economy on the proportion of starting salaries at each level.

R Problem: chi-square test for homogeneity

The dataset is in `studio9Data.tbl` and `studio9.r` has code showing how to load and manipulate this data.

(a) Compare group 1 and group 2 to see if they are assigned to levels in the same proportions. Do this in R by directly coding the test and also by using `chisq.test`.

(b) Test to see if group 1 is assigned levels in the same proportions as starting salaries in the broader economy. Again, code the test directly and using `chisq.test`

18.05 Introduction to Probability and Statistics
Spring 2014