Comparison of Bayesian and Frequentist Inference

18.05 Spring 2018

- First discuss two class 19 board questions...

Skipped in Class 19: Board question: genetic linkage

In 1905, William Bateson, Edith Saunders, and Reginald Punnett were examining flower color and pollen shape in sweet pea plants by performing crosses similar to those carried out by Gregor Mendel.

Purple flowers (P) is dominant over red flowers (p).
Long seeds (L) is dominant over round seeds (l).

F_0: PPLL $\times$ ppll   (initial cross)
F_1: PpLl $\times$ PpLl   (all second generation plants were PpLl)
F_2: 2132 plants   (third generation)

$H_0 =$ independent assortment: color and shape are independent.

	purple, long	purple, round	red, long	red, round
Expected	?	?	?	?
Observed	1528	106	117	381

Determine the expected counts for F_2 under H_0 and find the p-value for a Pearson chi-square test. Explain your findings biologically.

Solution

Since every F_1 generation flower has genotype Pp we'd expect F_2 to split 1/4, 1/2, 1/4 between PP, Pp, pp. For phenotype we expect F_2 to have 3/4 purple and 1/4 red flowers. Similarly for LL, Ll, ll. Assuming H_0 that color and shape are independent we'd expect the following probabilities for F_2.

	LL	Ll	ll	
PP	1/16	1/8	1/16	1/4
Pp	1/8	1/4	1/8	1/2
pp	1/16	1/8	1/16	1/4
	1/4	1/2	1/4	1

Genotype

	Long	Round	
Purple	9/16	3/16	3/4
Red	3/16	1/16	1/4
	3/4	1/4	1

Phenotype

Using the total of 2132 plants in F_2, the expected counts come from the phenotype table:

	purple, long	purple, round	red, long	red, round
Expected	1199	400	400	133
Observed	1528	106	117	381

Continued

Using R we compute: $G = 972.0$, $X^2 = 966.6$.

The degrees of freedom is 3 (4 cells - 1 cell needed to make the total work out). The *p*-values for both statistics is too close to zero for R to compute. (For example, $1 - \texttt{pchisq}(70, 3) = 4 \times 10^{-15}$.) With such a small *p*-value we reject H_0 in favor of the alternative that the genes are not independent.

Next, the last board question from Class 19...

(From Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. p.489)

Consider the following contingency table of counts

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

Question asked you to

use a chi-square test with significance 0.01

to test the hypothesis

the number of marriages and education level are independent.

Solution

The null hypothesis is that the cell probabilities are the product of the marginal probabilities. Assuming the null hypothesis we estimate the marginal probabilities in red and multiply them to get the cell probabilities in blue.

Education	Married once	Married mult times	marg probs
College	0.365	0.061	611/1436
No college	0.492	0.082	825/1436
marg probs	1231/1436	205/1436	1

Get expected counts by multiplying cell probabilities by the total number of women surveyed (1436). The table shows the observed and expected counts:

Education	Married once	Married multiple times
College	550 523.8	61 87.2
No college	681 707.2	144 117.8

Solution continued

We then have

$$G = 2\sum O_i \ln(O_i/E_i) = 16.55, \qquad X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 16.01$$

The number of degrees of freedom is $(2-1)(2-1) = 1$. We get

$$p = \texttt{1-pchisq(16.55,1) = } 0.000047$$

Because this is (much) smaller than our chosen significance .01 we reject the null hypothesis in favor of the alternate hypothesis that number of marriages and education level are not independent.

Is this a result you find believable?

Returning to our regularly scheduled programming...

Bayesian inference

- Uses priors
- Logically impeccable
- Probabilities can be interpreted
- Prior is subjective

Frequentist inference

- No prior
- Objective—everyone gets the same answer
- Logically complex
- Conditional probability of error is often misinterpreted as total probability of error
- Requires complete description of experimental protocol and data analysis protocol before starting the experiment. (This is both good and bad)

Concept question

Three different tests are run all with significance level $\alpha = 0.05$.

1. Experiment 1: finds $p = 0.03$ and rejects its null hypothesis H_0.

2. Experiment 2: finds $p = 0.049$ and rejects its null hypothesis.

3. Experiment 3: finds $p = 0.15$ and fails to rejects its null hypothesis.

Which result has the highest probability of being correct?

(Click **4** if you don't know.)

answer: **4.** You can't compute probabilities of hypotheses from p values.

Board question: Stop!

Experiments are run to test a coin that is suspected of being biased towards heads. The significance level is set to $\alpha = 0.1$
Experiment 1: Toss a coin 5 times. Report the sequence of tosses.

Experiment 2: Toss a coin until the first tails. Report the sequence of tosses.

1. Give the test statistic, null distribution and rejection region for each experiment. List all sequences of tosses that produce a test statistic in the rejection region for each experiment.

2. Suppose the data is *HHHHT*.
(a) Do the significance test for both types of experiment.
(b) Do a Bayesian update starting from a flat prior: Beta(1,1).
Draw some conclusions about the fairness of coin from your posterior.
(Use R: `pbeta` for computation in part (b).)

Solution

1. Experiment 1: The test statistic is the number of heads x out of 5 tosses. The null distribution is binomial(5,0.5). The rejection region is $\{x = 5\}$. The sequence of tosses $HHHHH$ is the only one that leads to rejection.
Experiment 2: The test statistic is the number of heads x until the first tails. The null distribution is geom(0.5), the rejection region $\{x \geq 4\}$. The sequences of tosses that lead to rejection are $\{HHHHT, HHHHH**T\}$, where '$**$' means an arbitrary length string of heads.

2a. For experiment 1 and the given data, 'as or more extreme' means 4 or 5 heads. So for experiment 1 the p-value is $P(4 \text{ or } 5 \text{ heads} \mid \text{fair coin}) = 6/32 \approx 0.20$.

For experiment 2 and the given data 'as or more extreme' means at least 4 heads at the start. So $p =$ `1 - pgeom(3,0.5)` $= 0.0625$.

(Solution continued.)

Solution continued

2b. Let θ be the probability of heads, Four heads and a tail updates the prior on θ, Beta(1,1) to the posterior Beta(5,2). Using R we can compute

$$P(\text{Coin is biased to heads}) = P(\theta > 0.5) = 1 \texttt{ -pbeta(0.5,5,2)} = 0.89.$$

If the prior is good then the probability the coin is biased towards heads is 0.89.

Board question: Stop II

For each of the following experiments (all done with $\alpha = 0.05$)
(a) Comment on the validity of the claims.
(b) Find the true probability of a type I error in each experimental setup.

1. By design Ruthi did 50 trials and computed $p = 0.04$.
   She reports $p = 0.04$ with $n = 50$ and declares it significant.
2. Ani did 50 trials and computed $p = 0.06$.
   Since this was not significant, she then did 50 more trials and computed $p = 0.04$ based on all 100 trials.
   She reports $p = 0.04$ with $n = 100$ and declares it significant.
3. Efrat did 50 trials and computed $p = 0.06$.
   Since this was not significant, she started over and computed $p = 0.04$ based on the next 50 trials.
   She reports $p = 0.04$ with $n = 50$ and declares it statistically significant.

Solution

1. (a) This is a reasonable NHST experiment.
(b) The probability of a type I error is 0.05.

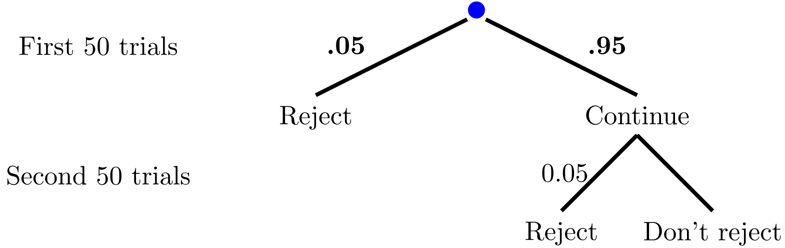
2. (a) The actual experiment run:
(i) Do 50 trials.
(ii) If $p < 0.05$ then stop.
(iii) If not run another 50 trials.
(iv) Compute p again, pretending that all 100 trials were run without any possibility of stopping.

This is not a reasonable NHST experimental setup because the second p-values are computed using the wrong null distribution.

(b) If H_0 is true then the probability of rejecting is already 0.05 by step (ii). It can only increase by allowing steps (iii) and (iv). So the probability of rejecting given H_0 is more than 0.05. We can't say how much more without doing a more complicated computation.

Solution continued

3. (a) See answer to (2a).
(b) The total probability of a type I error is more than 0.05. We can compute it using a probability tree. Since we are looking at type I errors all probabilities are computed assume H_0 is true.



The total probability of falsely rejecting H_0 is $0.05 + 0.05 \times 0.95 = 0.0975$.