Comparison of Bayesian and Frequentist Inference

18.05 Spring 2014

- First discuss last class 19 board question,

Compare

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 know probabilities of hypotheses based just on 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

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

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.

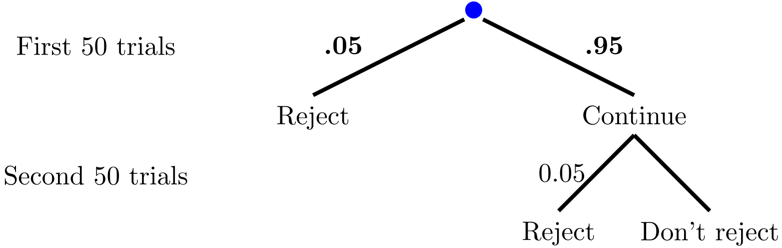
2. (a) The actual experiment run:
(i) Do 50 trials.
(ii) If $p < 0.05$ then stop.
(iii) If not run another 50 trials.
(iv) Compute p again, pretending that all 100 trials were run without any possibility of stopping.

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 more details.

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 $0.05 + 0.05 \times 0.95 = 0.0975$

MIT OpenCourseWare
https://ocw.mit.edu

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.