

Contrast between Likelihoodists & Significance tester for Bernoulli trials

Background on Bernoulli trials: These have 2 possible outcomes, “success” or “failure”, S or F (heads or tails, correct guess if milk put in first, winning ticket, etc.) (SIST, p. 33)

Consider 4 Bernoulli trials observed sample $x_0 = \langle S, S, F, S \rangle$

We can use a random variable, which takes value 1 whenever the trial is S, 0 when it's F.

$$x_0 = \langle 1, 1, 0, 1 \rangle$$

equivalently,

$$x_0 = \langle X_1=1, X_2=1, X_3=0, X_4=1 \rangle$$

Let $\Pr(X = 1) = \theta$ for any trial, and that trials are independent

θ a parameter; in the Bernoulli case it's from 0 to 1

If we knew θ , if we could compute

$\Pr(x_0; \theta) = \Pr(\text{observed } x_0; \text{assuming prob of success at each trial} = \theta)$

$$f(x_0; \theta)$$

The **joint** outcome involves series of “ands”

x_0 = the 1st trial is 1 and 2nd trial is 1 and 3rd trial is 0 and 4th trial is 1

Because the trials are *independent*, the probability multiplies

$$\Pr(x_0; \theta) = \Pr(X_1 = 1; \theta) \Pr(X_2 = 1; \theta) \Pr(X_3 = 0; \theta) \Pr(X_4 = 1; \theta)$$

Suppose $\theta = .2$ (as in Royall's example)

Therefore, $\text{Lik}(\theta = .2; x_0) = \text{Pr}(1, 1, 0, 1; .2) = (.2)(.2)(.8)(.2)$

Where did .8 come from?

If $\text{Pr}(S = .2)$ then $\text{Pr}(\text{not-}S) = .8$

(since by the axioms, $\text{Pr}(S \text{ or } \sim S) = 1 = \text{Pr}(S) + \text{Pr}(\sim S)$)

Note SIST error last line p. 33, it should be $\text{Lik}(.2)$ because Royall is about to use $H_0: \theta \leq .2$ vs $H_1: \theta > .2$ to compare his likelihoodist inference with the frequentist significance test

We want to compare $\text{Lik}(\theta = .2; x_0)$ with the likelihood given $\theta = .8$ (measure of comparative “support”)

So $\text{Lik}(\theta = .2; x_0) = \text{Pr}(1, 1, 0, 1; .2) = (.2)(.2)(.8)(.2)$
and $\text{Lik}(\theta = .8; x_0) = \text{Pr}(1, 1, 0, 1; .8) = (.8)(.8)(.2)(.8)$

$\text{LR}(\theta = .2 \text{ over } \theta = .8) = .0064 / .1024$

$(.2)^3(.8) / (.8)^3(.2) = (.25)^3(4) \sim .06$

Can also write the LR reverse $\text{LR}(\theta = .8 \text{ over } \theta = .2) = 16.6$

Generalize for 4 Bernoulli trials

Write the result of the k th trial is x_k as $X_k = x_k$

Random variable, capital X_k and lower case x_k is its value

$$x_{\text{obs}} = (X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } X_3 = x_3 \text{ and } X_4 = x_4)$$

$$\Pr(x; \theta) = \Pr(x_1; \theta) \Pr(x_2; \theta) \Pr(x_3; \theta) \Pr(x_4; \theta)$$

These should really be frequency distributions:

$$f(x_1; \theta) f(x_2; \theta) f(x_3; \theta) f(x_4; \theta)$$

$$\prod_{k=1}^4 f(x_k; \theta)$$

Comparing the likelihoodist and the significance tester

Use Royall example on p. 34, $n = 17$, there are 9 successes and 8 failures (ugly numbers, they're his)

$$\text{Lik}(x; \theta) = \theta^9 (1 - \theta)^8$$

Observed proportion of successes = .53

Even without calculating,

$\theta = .53$ makes the observed outcome most probable, it's the maximally likely θ value

He fixes $\theta = .2$ and considers the Likelihood ratio of .2 and various alternatives

Since the sample proportion is .53, any value of θ further from .53 than .2 is will be *less well* supported than .2

Start with .2, .33 more takes us to .53, another .33 goes to .86
So any $\theta > .86$ is less likely than is .2

Likelihood ratio of .2 and .9

$$\text{LR} (\theta = .2 \text{ over } \theta = .9) = [.2^9 (.8)^8] / [.9^9 (.1)^8] = 22.2$$

(top p. 36)

$\theta = .2$ is 22 times more likely than $\theta = .9$, given the outcome

Both are too hideously small, we would never be computing them. But we can group

$$(2/9)^9 (8)^8 \sim 22 \text{ top of p. 36}$$

Royall:

“Because $H_0: \theta \leq .2$ contains some simple hypotheses that are better supported than some hypotheses in H_1 (e.g., $\theta = .2$ is better supported than $\theta = .9$)...the law of likelihood does not allow the characterization of these observations as strong evidence for H_1 over H_0 .

The significance tester tests $H_0: \theta \leq .2$ vs. $H_1: \theta > .2$

So he rules out composite hypotheses.

The significance tester tests $H_0: \theta \leq .2$ vs. $H_1: \theta > .2$

She would reject H_0 and infer some (pos) discrepancy from .2

(observed mean M – expected mean under H_0) in standard deviation or standard error units

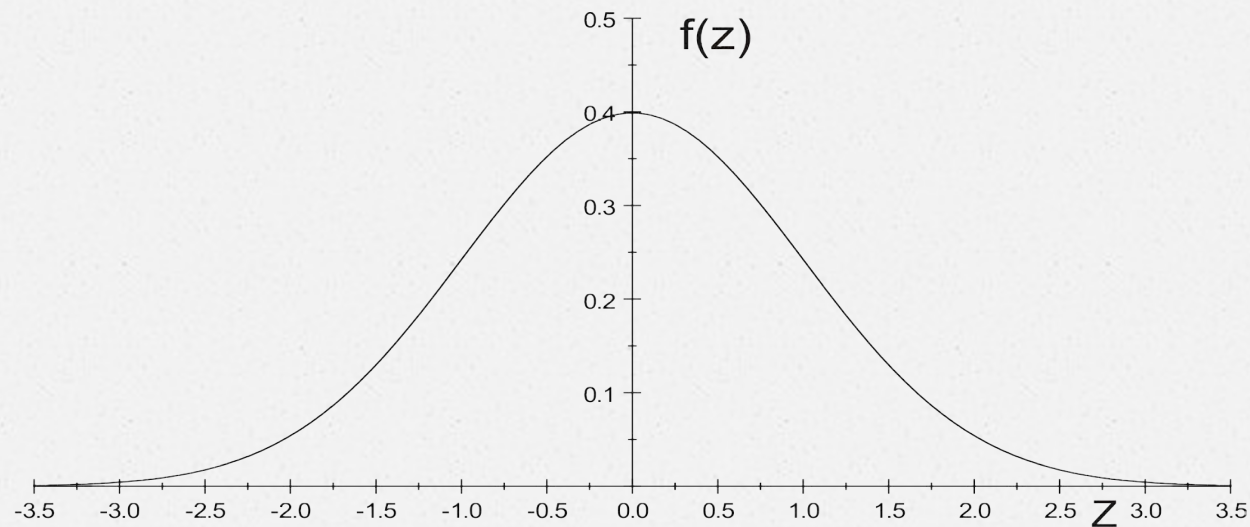
$$(.53 - .2)/.1 \sim \underline{3.3}$$

Here 1 SE is .1

Test Statistic $d(x_0)$ is $(.53 - .2)/.1$

Lets us use the Standard Normal curve (we're using a Normal approximation)

(area to the right of 3) ~ 0 , very significant.



$$\Pr(d(X) \geq d(x_0); H_0) \sim .003$$

$\Pr(d(X) < d(x_0); H_0) \sim .997$

(see p. 35)

Admittedly, just reporting there's evidence $H_1: \theta > .2$, as our significance tester, doesn't seem so informative either.

In inferring H_1 , she is only inferring *some* positive discrepancy from .3

A 95 % confidence interval estimate, which we have not discussed, would be $.53 \pm 2SE$
 $[.33 < \theta < .73]$

The Likelihoodist gives a series of comparisons: this is better supported than that, less strongly than some other value.

If you give enough comparisons, maybe our inferences aren't so different.

Is this really a statistical inference? Or just a report of the data? For the Likelihoodist it is, and the fact that a significance test is not comparative even precludes it from being a proper measure of evidence.

One Stat War Explained

Likelihoodists maintain that any genuine test or “rule of rejection” should be restricted to comparing the likelihood of H versus some point alternative H' relative to fixed data x

No wonder the Likelihoodist disagrees with the significance tester.

Elliott Sober: “The fact that significance tests don’t contrast the null with alternatives suffices to show that they do not provide a good rule for rejection” (Sober 2008, p. 56).

The significance test has an alternative $H_1: \theta > 0.2!$ (not a point) (STINT p. 35)

See the comparison in Souvenir B, likelihood vs error statistical p. 41

Spanos manuscript chapter 2: 19-21, set theoretic observations