Bernoulli trials: Plain Jane Version

Let's start with a specific example and generalize, then go back to specifics and generalize some more (SIST p. 33)

4 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.)

observed sample $x_0 = \langle S,S,F,S \rangle$ (or x_{obs})

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,

$x_0 = <X_1=1, X_2=1, X_3=0, X_1=1>$

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

θ a <u>parameter</u>; in the Bernoulli case it’s from 0 to 1

If we knew θ, if we could compute

Pr(x_0; θ) = Pr(observed x_0; assuming prob of success at each trial = θ)

f(x_0; θ)

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

So, $Pr(x_0; \theta)$
$= Pr(X_1 = 1 \text{ and } X_2 = 1 \text{ and } X_3 = 0 \text{ and } X_4 = 1; \theta)$

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)

(e.g., 100 balls, 20 are red and we randomly draw, and success is getting a red ball)

What's Pr(X = 1) assuming the probability of X = 1 is $.2$?

Who is buried in Grant's tomb?

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

Where did .8 come from?
If Pr(S = .2) then Pr(not-S) = .8
(since by the axioms, Pr(S or ~S) = 1 = Pr(S) + Pr (~S))

Note SIST error last line p. 33, it should be 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 Lik($\theta = .2$; x_0) with the likelihood given $\theta = .8$ (measure of comparative "support")

Lik($\theta = .8$; x_0) = Pr(1,1,0,1; .8) = (.)(.)(.)(.)

.0064 vs. .1024

In general, with this x_0,

Lik(θ; x_0) = Pr(1,1,0,1; θ) = (θ)(θ)(1 - θ)(θ) =

θ^3(1 - θ)   order doesn't matter

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

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

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

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

It’s useful to start with the Likelihoodist, because it’s a key example of a logic of (comparative) evidence, and hits one of the big “wars”

Still we don’t usually crank out numbers;
My book does because it’s taking the criticisms in their actual location and the people arguing use numbers

The book asks the reader to find Lik(.75) with the same outcome <1,1,0,1> (note .75 is closer to .8 than to .2 so .8 is more likely)

This is the maximally likely θ as the observed proportion is ¾
What's Lik(.75; x_0)?

.1054

Generalize for 4 Bernoulli trials

More generally, still for 4 trials, say we don't know the result,

Write the result of the kth trial is x_k as $X_k = x_k$
Random variable, capital X_k and lower case x_k is its value

$x_{obs} = (X_1 = x_1$ and $X_2 = x_2$ and $X_3 = x_3$ 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:
$f(x_1; \theta)\, f(x_2; \theta)\, f(x_3; \theta)\, f(x_4; \theta)$

Shortcut abbreviation for multiplying:
$f(x_1; \theta)\ f(x_2; \theta)\ f(x_3; \theta)\ f(x_4; \theta)$

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

Now take the 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,

θ = .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

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

(top p. 36)

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

$(2/9)^9\ (8)^8 \sim 22$ 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

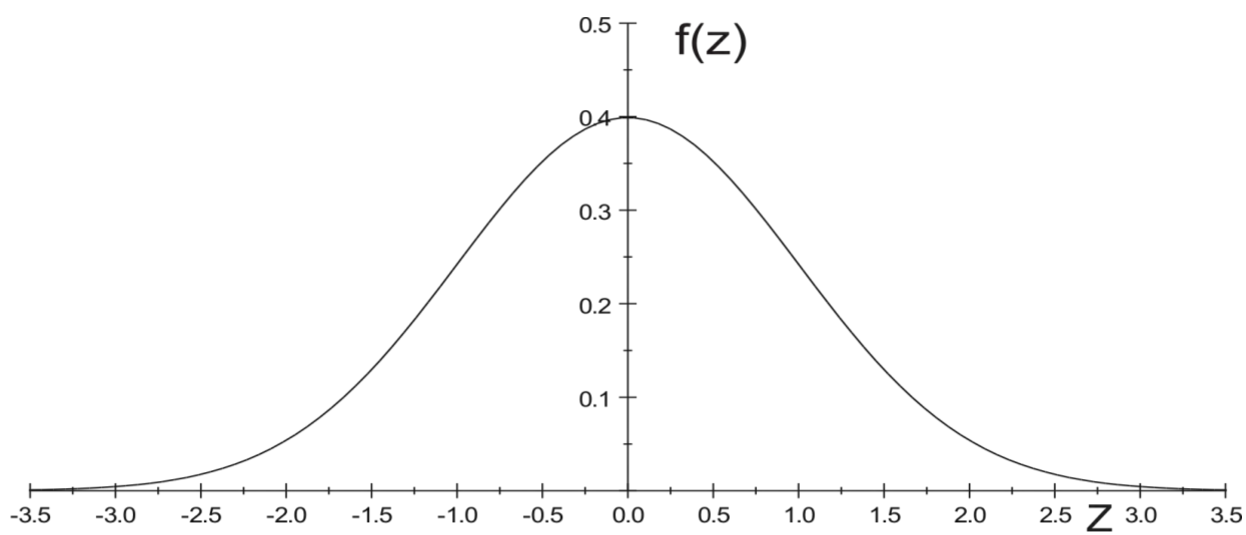
(.53 - .2)/.1 ~ 3.3

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)

$$\text{(area to the right of 3)} \sim 0, \text{ very significant.}$$



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

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

(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 ± 2SE
[$.33 < \theta < .73$]

We'll see how severity also gives a report of discrepancy and has some advantages.

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)

While we're at notation: let's generalize for n Bernoulli trials

x_{obs} a member of the sample space: x_{obs} ε R real numbers)

$x_{obs} = X_1 = x_1$ and $X_2 = x_2$ and $X_3 = x_3$ ….and $X_n = x_n$

$Pr(x_{obs}; \theta) = f(x_1; \theta)\, f(x_2; \theta)\, f(x_3; \theta) \ldots f(x_n; \theta)$

Shortcut abbreviation:

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

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

I’ve run out of letters, let z = number of success out of n, n – k failures Lik(x; θ)

$$\theta^z (1 - \theta)^{n-z}$$

More notation z = $\sum_{k=1}^{n} x_k$

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

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

End Part I of Mayo