

18.05 Problem Set 7, Spring 2014 Solutions

Problem 1. (10 pts.) (a) $H_0: \theta = 0.5$

H_A : one-sided $\theta > 0.5$, two-sided $\theta \neq 0.5$.

Test statistic: $x =$ number of heads in 250 spins.

Data: $x = 140$.

One-sided data at least as extreme: $x \geq 140$. Using R we compute the one-sided p -value is

$$p = P(x \geq 140|H_0) = 1 - \text{pbinom}(139, 250, 0.5) = 0.03321$$

The one-sided p -value computes the probability in the one-sided tail. Because our null distribution (binomial(250, 0.5)) is symmetric, each tail in the rejection region will have probability $\alpha/2$. In this case the two-sided p -value is computed by doubling the smaller of the one sided values. We computed the right tail p -value just above. This is the smaller of the two p -values so our two-sided p -value is $2 \times 0.03321 = 0.06642$. This rounds to 0.07, so the figure of 7% is the two-sided p -value.

Note: we could use the normal approximation

$$\text{binomial}(250, 0.5) \approx N(125, 250/4)$$

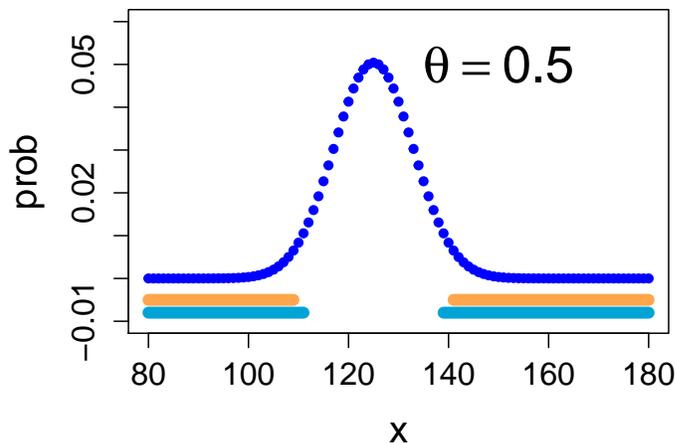
and the z -statistic

$$z = \frac{x - 125}{\sqrt{250/4}} \approx N(0, 1)$$

one-sided: $p = P(z \geq \frac{15}{\sqrt{250/4}}) \approx 0.02889$.

two-sided: $p = P(|z| \geq \frac{15}{\sqrt{250/4}}) \approx 0.05778$.

(b) We saw in part (a) that the quoted 7% was a two-sided p -value. So for the rest of this problem we'll use two-sided tests. The exact p -value was $p = 0.066$. Since $0.05 < p < 0.1$ we reject H_0 at significance $\alpha = 0.1$ and don't reject at $\alpha = 0.05$.



The figure shows the null distribution, the $\alpha = 0.1$ rejection region (blue) and the $\alpha = 0.05$ rejection region (orange). Notice that the data $x = 140$ is in the 0.1 rejection region but not the second.

(c) The problem asks us to find the rejection region for $\alpha = 0.01$. We use R to find the endpoints for the rejection region (called critical values):

```
criticalPoint.left = qbinom(0.005,250,0.5) - 1 = 104
criticalPoint.right = qbinom(0.995,250,0.5) + 1 = 146
```

Note: we added or subtracted one to the value returned by `qbinom` for a discrete distribution like the binomial there is not an exact critical value. So `qbinom(x, n, p)` returns the smallest integer with more than x probability in its left tail. Since the rejection region must have at most $\alpha/2$ in either tail we have to move the R answer by one towards the tail.

Conclusion: we reject for greater than or equal to 146 heads or less than or equal to 104 heads.

(d) (i) For $\alpha = 0.05$ the rejection region is given by the critical points

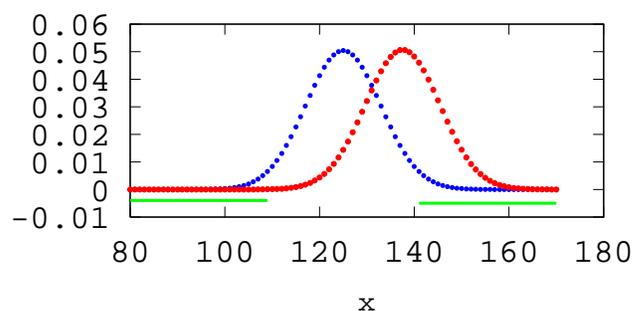
```
criticalPoint.left = qbinom(0.025,250,0.5) - 1 = 109
criticalPoint.right = qbinom(0.975,250,0.5) + 1 = 141
```

$$\begin{aligned} \text{power of } H_A &= P(\text{reject} \mid H_A) \\ &= P(x \leq 109 \text{ or } x \geq 141 \mid H_A) \\ &= \text{sum}(\text{dbinom}(0:109, 250, 0.55)) + \text{sum}(\text{dbinom}(141:250, 250, 0.55)) = 0.35237 \end{aligned}$$

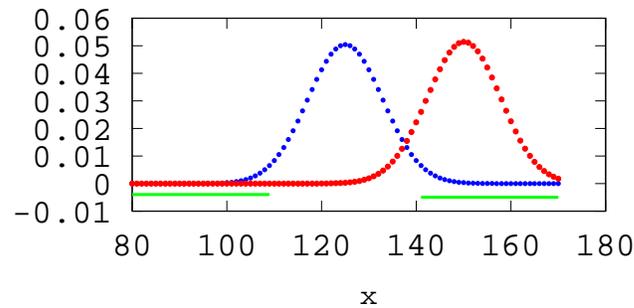
Likewise

$$\begin{aligned} \text{power of } H'_A &= P(\text{reject} \mid H'_A) \\ &= P(x \leq 109 \text{ or } x \geq 141 \mid H'_A) \\ &= \text{sum}(\text{dbinom}(0:109, 250, 0.6)) + \text{sum}(\text{dbinom}(141:250, 250, 0.6)) = 0.88963 \end{aligned}$$

(ii) The two plots below show the null distribution and the distribution of H_A and H'_A . The green line below the graphs shows the rejection region. The greater power of H'_A is explained by its greater separation from H_0 . Most of the probability of H'_A is over the right side of the rejection region.



Distributions of H_0 and H_A

Distributions of H_0 and H'_A

(e) The answer is $n = 1055$ with H_A giving a power of 0.9003.

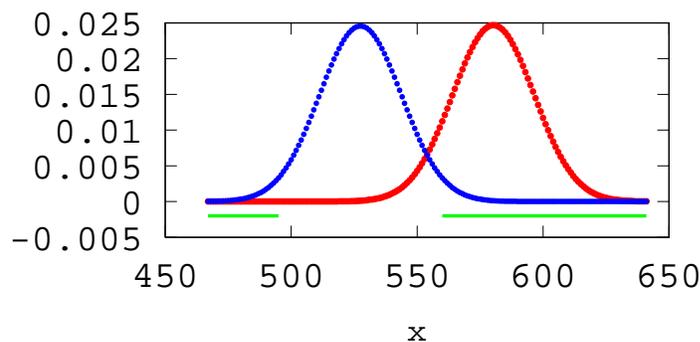
To get this we need to compute the power for various values of n . The steps for each n are :

1. Find the rejection region.
2. Compute the power.

Here is the R-code for one value of n . Creating the loop to check through all values of n until we find the first with power = 0.9 is in the posted code.

```
theta = 0.55
n = 300;
# Find critical points for rejection region (based on theta=0.5)
criticalPoint.left = qbinom(0.025,n,0.5) - 1;
criticalPoint.right = qbinom(0.975,n,0.5) + 1;
rejectionRegion = c(0:criticalPoint.left, criticalPoint.right:n)
power = sum(dbinom(rejectionRegion, n, theta))
print(power)
```

See the two plots with part (d): power increases as n increases because the distributions become more separated.



Plot for $n = 1055$ of the H_0 and $H_A : \theta = 0.55$ distributions. The green lines show the rejection region.

An alternative approach approximating the exact answer with normal distributions is given at the end of these solutions.

(f) We use the usual Bayesian update table.

Hypothesis	prior	likelihood	posterior
$\theta = 0.5$	1/2	$c_1(0.5)^{250}$	$c_2(0.5)^{250} = 0.14757$
$\theta = 0.55$	1/2	$c_1(0.55)^{140}(0.45)^{110}$	$c_2(0.55)^{140}(0.45)^{110} = 0.85243$

The normalizing factor $c_2 = \frac{1}{(0.5)^{250} + (0.55)^{140}(0.45)^{110}}$.

The posterior probability that $\theta = 0.55$ is 0.85.

Note: if we used the beta(1, 1) prior on θ in $[0,1]$ and used the data to update to a posterior of beta(141, 111) then the 90% probability interval is $[0.501, 0.611]$. Although p -values and posterior probability are measure different things. This probability interval and the 7% p -value seem similar in the belief they convey that the coin may be biased.

(g) I'd go with the 90% probability interval discussed in part (f). It appears the coin is biased.

Problem 2. (10 pts.) (a) Type I error is rejecting the null-hypothesis when it is indeed true. This corresponds to thinking someone is lying when they are in fact being truthful. The experiment had $\frac{9}{140}$ type I errors. This is our estimate of the probability of a type I error.

Type II error is not rejecting the null-hypothesis when it is indeed false. This corresponds to thinking someone is telling the truth when they are in fact lying. Based on the data our estimate of the probability of a Type II error is $\frac{15}{140}$.

(b) Significance = $P(\text{type I error}) = P(\text{reject } H_0 \mid H_0)$.
Power = $1 - P(\text{type II error}) = P(\text{reject } H_0 \mid H_A)$.

Problem 3. (10 pts.) (a) This is a two-sided alternative. The t -statistic is

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1}{2/4} = 2.$$

Since we have $n = 16$ our t statistic has 15 degrees of freedom.

We have the two-sided p -value

$$p = P(|t| > 2 \mid H_0) = 2 * (1 - \text{pt}(2, 15)) = 0.063945.$$

Since $p > \alpha = 0.05$ we don't reject the null hypothesis.

Alternatively we could have done the problem in terms of rejection regions. We are given $\bar{x} = 11$, $s^2 = 4$, and $n = 16$. The null hypothesis is $\mu = 10$. Using \bar{x} as our test statistic the rejection region is

$$\left(-\infty, 10 - t_{15,0.025} \frac{s}{\sqrt{n}}\right] \cup \left[10 + t_{15,0.025} \frac{s}{\sqrt{n}}, \infty\right) = (-\infty, 8.93] \cup [11.07, \infty)$$

Here $t_{15,0.025}$ means a *critical value*, i.e. the value with right tail probability 0.025: for $T \sim t(15)$ we have $P(t > t_{15,0.025}) = 0.025$.

Since 11 lies outside the rejection region, we should *not* reject the null-hypothesis.

(b) This is a one-sided alternative. The t -statistic is the same

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1}{2/4} = 2.$$

So we have the one-sided p -value

$$p = P(t > 2 | H_0) = 2*(1-\text{pt}(2, 15)) = 0.031973.$$

Since $p < \alpha = 0.05$ we reject the null hypothesis in favor of the alternative.

Again looking at rejection regions. We use the critical value $t_{15,0.05} \approx 1.753$. The rejection region for \bar{x} is

$$\left[10 + t_{15,0.05} \frac{s}{\sqrt{n}}, \infty\right) = [10.876, \infty).$$

Since 11 lies inside the rejection region, we should reject the null-hypothesis in favor of $H_1 : \mu > 10$.

Problem 4. (10 pts.) (a) Let μ be the actual speed of a given driver. We are given that

$$x_i \sim N(\mu, 5^2) \Rightarrow \bar{x} \sim N(\mu, 5^2/3).$$

The most natural hypotheses are:

H_0 : the driver is not speeding, i.e. $\mu \leq 40$.

H_A : the driver is speeding, i.e. $\mu > 40$.

Both are composite.

Note: we will work with $H_0: \mu = 40$, which is simple.

(b) (i) Giving a ticket to a non-speeder is a type I error (rejecting H_0 when it is true). H_0 is composite, but we can do all our computations with the most extreme value $\mu = 40$ because the one-sided rejection region will have its largest significance level when $\mu = 40$.

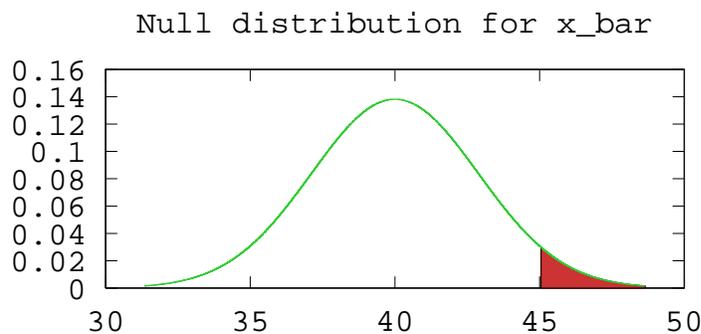
So the null distribution is $\bar{x} \sim N(40, 5^2/3)$. The critical value is

$$c_{0.04} = \text{qnorm}(0.96, 40, 5/\text{sqrt}(3)) = 45.054$$

(Equivalently $c_{0.04} = 40 + z_{0.04} \frac{5}{\sqrt{3}} = 45.054$.)

That is, they should issue a ticket if the average of the three guns is more than 45.054.

(ii) Here is a plot of the null distribution $N(40, 5^2/3)$. The rejection probability of 0.04 is shown.



(c) Power = $P(\text{rejection} | H_A)$. So to find the power we first must find the rejection region. For $n = 3$ this was done in part (b): rejection region = $[45.054, \infty)$. So

$$\text{power} = P(\text{rejection} | \mu = 45) = 1 - \text{pnorm}(45.054, 45, 5/\text{sqrt}(3)) = \boxed{0.493}$$

With n cameras (guns) let's write \bar{x}_n for the sample mean. The null distribution is

$$\bar{x}_n \sim N(40, 5^2/n)$$

The critical value, i.e. the left endpoint of the rejection region, depends on n . Also, in order to do the computations algebraically we need to write everything in terms of standard normal values.

$$c_{0.04} = \text{qnorm}(0.96, 40, 5/\sqrt{n}) = 40 + z_{0.04} \frac{5}{\sqrt{n}}$$

where $z_{0.04}$ is the standard normal critical value

$$z_{0.04} = \text{qnorm}(0.96, 0, 1) = 1.751.$$

We want

$$\text{power} = P(\bar{x} \geq c_{0.04} | \mu = 45) = 0.9$$

Standardizing and doing some algebra we get

$$P\left(\frac{\bar{x} - 45}{5/\sqrt{n}} \geq \frac{c_{0.04} - 45}{5/\sqrt{n}}\right) = 0.9 \quad \Rightarrow \quad P\left(z \geq \frac{-5}{5/\sqrt{n}} + z_{0.04}\right) = 0.9$$

Thus $\frac{-5}{5/\sqrt{n}} + z_{0.04} = z_{0.9}$. We get

$$n = (z_{0.04} - z_{0.9})^2 = (1.7507 - (-1.2816))^2 = 9.1945.$$

Setting n to be the next biggest integer we get $n = 10$.

We could do this all in R. The code for computing the power when $n = 3$ is shown below. Notice that we don't have to phrase everything in terms of standard normal values to compute. You only have to change the first line to compute power for different values of n .

```

n = 3
mu = 40;
sigma = 5/sqrt(n);
alpha = 0.04;
xcrit = qnorm(1-alpha, mu, sigma);
power = 1-pnorm(xcrit, 45, sigma)

```

We could now use R to compute power for increasing values of n and see which value of n gives power more than 0.9.

Problem 5. (10 pts.) (a) $P(\text{type I}) = P(\text{reject } H_0 | H_0) = P(x \leq 0.1 \text{ or } x \geq 1.9 | \theta = 2) = \frac{0.2}{2} = 0.1.$

(b) $P(\text{type II}) = P(\text{don't reject } H_0 | \theta = 2.5) = P(0.1 < x < 1.9 | \theta = 2.5) = \frac{1.8}{2.5} = 0.72.$

Problem 6. (10 pts.) (a) The frequentist has hypotheses $H_0 = \text{'the sun is okay'}$
 $H_A = \text{'the sun has gone nova'}$

The experimental data from the neutrino detector says the sun has gone nova.

The frequentist computes the p -value

$$p = P(\text{sun gone nova} | H_0) = P(\text{computer lied}) = 1/36 = .027$$

(Note: $1/36$ is really 0.028 not 0.027) Since $p < 0.05$ the frequentist rejects H_0 in favor of H_A at significance level 0.05.

This is problematic, the p -value is the probability assuming H_0 of data 'at least as extreme' as the data seen. How is that even defined in this case?

Ignoring the p -value we can explain the comic more clearly. The rejection region = {the detector says yes}. The significance is $1/36 = 0.028$. Therefore the frequentist rejects the null hypothesis and concludes the sun has gone nova at significance level 0.028.

(b) The comic is pointing out the flaw of multiple testing or what's sometimes called data mining. (The bad type of data mining, there is also a good type.) A significance level of 0.05 means that in 20 experiments where H_0 is true we'd expect to reject it once. The scientists test 20 colors. So even if no jelly bean color causes cancer there is a high probability that one of the tests will produce a test statistic in the rejection region.

The fix is to plan on doing n tests and set the significance level for any one test to α/n . Then, assuming H_0 is true for all the tests, the probability that at least one of them will reject is roughly $n * \alpha/n = \alpha$. This is called the Bonferroni correction. (Actually, because of the possibility of multiple rejections the probability at least one will reject is less than or equal to α .)

Normal approximation to problem 1 (e).

We use that $\text{binomial}(n, \theta) \approx N(n\theta, n\theta(1-\theta))$. The normal approximation introduces some error, but we can compute n directly instead of needing to search through a sequence of possible values.

For $\alpha = 0.05$ and $\theta = 0.5$ the critical values are

$$c_{0.025} = n(0.5) + \frac{\sqrt{n}}{2} z_{0.025}$$

$$c_{0.975} = n(0.5) - \frac{\sqrt{n}}{2} z_{0.025}$$

So when $\theta = 0.55$

$$\text{power} = P\left(x \leq \frac{n}{2} - \frac{\sqrt{n}}{2} z_{0.025} \mid \theta = 0.55\right) + P\left(x \geq \frac{n}{2} + \frac{\sqrt{n}}{2} z_{0.025} \mid \theta = 0.55\right)$$

Standardizing assuming $\theta = 0.55$:

$$z = \frac{x - n(0.55)}{\sqrt{n(0.55)(0.45)}}$$

A little algebra gives

$$\begin{aligned} \text{power} &\approx P\left(z \leq \frac{n/2 - \frac{\sqrt{n}}{2} z_{0.025} - n(0.55)}{\sqrt{n(0.55)(0.45)}}\right) + P\left(z \geq \frac{n/2 + \frac{\sqrt{n}}{2} z_{0.025} - n(0.55)}{\sqrt{n(0.55)(0.45)}}\right) \\ &= P\left(z \leq \frac{-0.05n - \frac{\sqrt{n}}{2} z_{0.025}}{\sqrt{n(0.55)(0.45)}}\right) + P\left(z \geq \frac{-0.05n + \frac{\sqrt{n}}{2} z_{0.025}}{\sqrt{n(0.55)(0.45)}}\right) \end{aligned}$$

For n large the left hand probability is essentially 0. So to get power = 0.9 we need

$$P\left(z \geq \frac{-0.05n + \frac{\sqrt{n}}{2} z_{0.025}}{\sqrt{n(0.55)(0.45)}}\right) = 0.9$$

That is

$$\frac{-0.05n + \frac{\sqrt{n}}{2} z_{0.025}}{\sqrt{n(0.55)(0.45)}} = z_{0.9}$$

Using $z_{0.025} = 1.96$ and $z_{0.9} = -1.2816$ and solving for n we get $n = 1047$, which is very close to the exact answer of 1055.

MIT OpenCourseWare
<http://ocw.mit.edu>

18.05 Introduction to Probability and Statistics
Spring 2014

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