

# STAT 513 fa 2020 Lec 01

## Inference, tests of hypotheses, power and size

Karl B. Gregory

### Inference: tests of hypotheses

- Goal of statistics: learn from random outcomes about the process which generated them.
- Areas of learning:
  - Estimate/describe features of the data generating process.
  - Predict/forecast values of yet-to-be-observed outcomes.
  - Infer/conclude/decide about hypotheses concerning the data generating process.

Inference is the most delicate.

- A statistical inference is a conclusion about a data generating process based on data generated from it.
- More precisely, it is a decision concerning two complementary statements about the data generating process: the *null hypothesis*  $H_0$  and the *alternate hypothesis*  $H_1$ .
- The null and alternate hypotheses make opposing claims. The null hypothesis is formulated such that its rejection represents some finding of interest and failure to reject it represents no findings.
- Formally, a *statistical inference* is a decision to reject  $H_0$  or not to reject  $H_0$  based on the data.
- If we reject the null hypothesis, we conclude that the alternate hypothesis is true. If we fail to reject the null hypothesis, we do NOT conclude that the null hypothesis is true; we say only that we do not have sufficient evidence to reject it.
- The data generating process with which we will be concerned is the random sample: Let  $X_1, \dots, X_n$  be independent identically distributed random variables with a distribution that depends on the parameter  $\theta$ . We will undertake to make inferences about  $\theta$ .
- In general, if the parameter of interest  $\theta$  takes values in some space  $\Theta$ ,  $H_0$  and  $H_1$  take the form

$$H_0: \theta \in \Theta_0 \text{ versus } H_1: \theta \in \Theta_1,$$

where  $\Theta_1 = \Theta \setminus \Theta_0 = \Theta \cap \Theta_0^c$ , so that  $\Theta_0 \cup \Theta_1 = \Theta$ .

- Sometimes we call  $\Theta_0$  the *null space* and  $\Theta_1$  the *alternate space*.

- **Examples:**

- If  $p \in (0, 1)$  is the probability of “heads” when tossing a coin, we might test

$$H_0: p = 1/2 \text{ versus } H_1: p \neq 1/2$$

i.e.  $H_0: p \in \{1/2\}$  versus  $H_1: p \in (0, 1/2) \cup (1/2, 1)$ . If we reject  $H_0$ , we conclude  $p \neq 1/2$ , i.e. that the coin is unbalanced. If we fail to reject  $H_0$ , we do not claim that the coin is balanced; we only say that there is insufficient evidence of unbalancedness.

- If  $\mu \in (0, \infty)$  is the mean delivery time (in days) of packages, we might test

$$H_0: \mu \leq 2 \text{ versus } H_1: \mu > 2$$

i.e.  $H_0: \mu \in (0, 2]$  versus  $H_1: \mu \in (2, \infty)$ . If we reject  $H_0$ , we conclude  $\mu > 2$ , i.e. that the mean delivery time is more than 2 days. If we fail to reject  $H_0$ , we do not claim that the mean delivery time is less than or equal to 2 days; we only say that there is insufficient evidence of a longer mean delivery time.

- If  $\delta \in (-\infty, \infty)$  is the effect of a drug on a clinical response, we might test

$$H_0: \delta = 0 \text{ versus } H_1: \delta \neq 0$$

i.e.  $H_0: \delta \in \{0\}$  versus  $H_1: \delta \in (-\infty, 0) \cup (0, \infty)$ . If we reject  $H_0$ , we conclude  $\delta \neq 0$ , i.e. that the drug has an effect on the clinical response. If we fail to reject  $H_0$ , we do not claim that the drug has no effect; we only say that there is insufficient evidence of an effect.

- **Definition:** If a hypothesis specifies only a single value for a parameter, the hypothesis is called a *simple hypothesis*. Otherwise it is called a *composite hypothesis*.

- **Definition:** A *test of hypotheses* or a *hypothesis test* is a rule for deciding whether or not to reject  $H_0$  based on data. Given a random sample  $X_1, \dots, X_n$  and null and alternate hypotheses  $H_0$  and  $H_1$ , tests of hypotheses take the form

$$\text{Reject } H_0 \text{ iff } T(X_1, \dots, X_n) \in \mathcal{R},$$

where  $T(X_1, \dots, X_n)$  is a function of the sample values called a *test statistic* and  $\mathcal{R}$  is a set called the *rejection region*.

- **Exercises:** For each of the following give (i)  $H_0$  and  $H_1$ , (ii) the test statistic  $T$ , and (iii) the rejection region  $\mathcal{R}$ .

- (a) Based on 10 coin tosses with the probability of “heads” equal to  $p$ , call the coin unbalanced if more than six or fewer than four “heads” are observed.

- (i)  $H_0: p = 1/2$  versus  $H_1: p \neq 1/2$

- (ii)  $T = \#$  “heads” in 10 tosses

- (iii)  $\mathcal{R} = \{0, 1, 2, 3\} \cup \{7, 8, 9, 10\}$

- (b) Conclude that the mean monthly rent  $\mu$  paid by USC students is greater than 600 if mean rent of 20 randomly sampled students exceeds 650.
  - (i)  $H_0: \mu \leq 600$  versus  $H_1: \mu > 600$
  - (ii)  $T$  = average of rents of 20 sampled students
  - (iii)  $\mathcal{R} = (650, \infty)$
- (c) Infer that the standard deviation  $\sigma$  of heights of 2-yr-old trees on a tree farm is less than 3 feet if the standard deviation of the heights of 10 randomly sampled trees is less than 2 feet.
  - (i)  $H_0: \sigma \geq 3$  versus  $H_1: \sigma < 3$
  - (ii)  $T$  = standard deviation of heights of 10 sampled trees
  - (iii)  $\mathcal{R} = (0, 2)$

- Since our decisions are based on random outcomes, inferences may err.
- **Definition:** A *Type I error* is rejecting  $H_0$  when  $H_0$  is true and a *Type II error* is failing to reject  $H_0$  when  $H_0$  is false.
- Possible outcomes of statistical inference:

	$H_0$ true	$H_0$ false
reject $H_0$	Type I error	correct decision
fail to reject $H_0$	correct decision	Type II error

- We want tests for which the probabilities of Type I and Type II errors are small.

## Quality of tests: power and size

- **Definition:** The *power* of a test is the probability that the test will lead to a rejection of  $H_0$ . The power is a function of the true value of the parameter  $\theta$ . For the test

$$\text{Reject } H_0 \text{ iff } T(X_1, \dots, X_n) \in \mathcal{R}$$

we write the power as the function

$$\gamma(\theta) = P(\text{Reject } H_0 \text{ when true value of parameter is } \theta) = P_\theta(T(X_1, \dots, X_n) \in \mathcal{R})$$

of  $\theta$ , where the notation  $P_\theta$  denotes probability computed when the true value of the parameter is equal to  $\theta$ .

- The power function  $\gamma : \Theta \rightarrow [0, 1]$  gives for each value in the parameter space  $\Theta$  the probability that the test will reject  $H_0$  when the parameter is equal to that value.
- For any  $\theta \in \Theta$ , we
  - reject  $H_0$  with probability  $\gamma(\theta)$ .
  - fail to reject  $H_0$  with probability  $1 - \gamma(\theta)$ .

- For  $\theta \in \Theta_0$ ,
  - $\gamma(\theta)$  is the probability of a Type I error.
  - $1 - \gamma(\theta)$  is the probability of a correct decision (fail to reject  $H_0$ ).
- For  $\theta \in \Theta_1$ , the probability of a
  - $\gamma(\theta)$  is the probability of a correct decision (reject  $H_0$ ).
  - $1 - \gamma(\theta)$  is the probability of a Type II error.
- We want  $\gamma(\theta)$  to be small for  $\theta \in \Theta_0$  and large for  $\theta \in \Theta_1$ .
- **Exercise:** Based on 10 coin tosses, call the coin unbalanced if more than six or fewer than four “heads” are observed.
  - In the following, what decision do you make and is it a correct decision, a Type I error, or a Type II error?
    - The true probability of “heads” is 0.6 and you roll 7 heads.
    - The true probability of “heads” is 0.6 and you roll 5 heads.
    - The true probability of “heads” is 0.5 and you roll 7 heads.
    - The true probability of “heads” is 0.5 and you roll 4 heads.
  - Find the power of the test when the true probability of getting “heads” is 0.6.
  - Plot the power  $\gamma(p)$  against  $p$  for  $p = 0.01, 0.02, \dots, 0.99$ .
  - Suppose the coin is balanced. What is the probability that this test will lead to a Type I error?
  - What is the probability of a Type II error if the true probability of “heads” is 1/3?

**Answers:**

- (i) We have

$$\begin{aligned}
 \gamma(0.6) &= P_{0.6}(\text{more than 6 or fewer than 4 “heads”}) \\
 &= P(Y \leq 3) + P(Y \geq 7), \quad Y \sim \text{Binomial}(10, 0.6) \\
 &= P(Y \leq 3) + 1 - P(Y \leq 6) \\
 &= \sum_{y=0}^3 \binom{10}{y} (0.6)^y (1 - 0.6)^{10-y} + 1 - \sum_{y=0}^6 \binom{10}{y} (0.6)^y (1 - 0.6)^{10-y} \\
 &= \text{pbinom}(3, 10, \text{p.seq}) + 1 - \text{pbinom}(6, 10, \text{p.seq}) \\
 &= 0.437.
 \end{aligned}$$

This means that if the true probability of getting “heads” were 0.6, this test would reject the null hypothesis that the coin is balanced with probability 0.437.

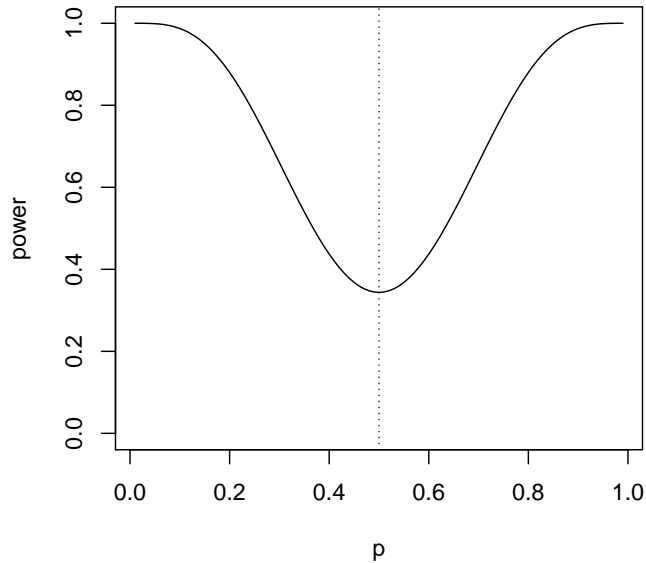
- (ii) Use the following R code:

```

p.seq <- seq(.01,.99,length=99)
power <- pbinom(3,10,p.seq) + 1 - pbinom(6,10,p.seq)

plot(p.seq,power,type="l",ylim=c(0,1),xlab="p")
abline(v=0.5,lty=3) # vert line at null value

```



(iii) If the coin is balanced,  $p = 0.5$ , for which the power is

$$\gamma(0.5) = \text{pbinom}(3,10,.5) + 1 - \text{pbinom}(6,10,.5) = 0.34375,$$

so if the null hypothesis is true the test will lead to a Type I error with probability 0.34375.

(iv) A Type II error is failing to reject the null, so the probability is one minus the power:

$$1 - \gamma(1/3) = 1 - (\text{pbinom}(3,10,1/3) + 1 - \text{pbinom}(6,10,1/3)) = 0.421074.$$

- **Definition:** The *size* of a test, denoted by  $\alpha$ , is defined as

$$\alpha = \sup_{\theta \in \Theta_0} \gamma(\theta),$$

which we read as “the supremum of the power  $\gamma(\theta)$  over all  $\theta \in \Theta_0$ ”. More simply put, the size is the maximum power over the null space. The size is thus the largest probability of a Type I error over all  $\theta \in \Theta_0$ .

- If the null space contains a single point, say  $\Theta_0 = \{\theta_0\}$ , then the size is  $\alpha = \sup_{\theta \in \{\theta_0\}} \gamma(\theta) = \gamma(\theta_0)$ .

- We would like tests to have small size as well as large power when  $H_0$  is false. We will find that there is a size and power trade-off.
- **Exercise (cont):** Based on 10 coin tosses, call the coin unbalanced if more than six or fewer than four “heads” are observed. What is the size of the test?

**Answer:**

Since  $H_0$  specifies a single value  $p \in \{0.5\}$ , the size of the test is

$$\alpha = \sup_{p \in \{0.5\}} \gamma(p) = \gamma(0.5) = \text{pbinom}(3, 10, .5) + 1 - \text{pbinom}(6, 10, .5) = 0.34375.$$

- **Exercises:** For the following tests, (i) find an expression for the power function, (ii) calculate the size, and (iii) make a plot of the power function.

(a) Let  $X_1, \dots, X_{10}$  be a random sample from the Bernoulli( $p$ ) distribution, where  $p \in (0, 1)$  is unknown, and suppose you wish to test  $H_0: p \leq 1/4$  versus  $H_1: p > 1/4$  using the test

$$\text{Reject } H_0 \text{ iff } X_1 + \dots + X_{10} > 5.$$

(b) Let  $X_1, \dots, X_n$  be a random sample from the Normal( $\mu, 1$ ) distribution, where  $\mu \in (-\infty, \infty)$  is unknown, and suppose you wish to test  $H_0: \mu = 0$  versus  $H_1: \mu \neq 0$  using the test

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > 2.$$

(c) Let  $X_1, \dots, X_{25}$  be a random sample from the Exponential( $\lambda$ ) distribution, where  $\lambda \in (0, \infty)$  is unknown, and suppose you wish to test  $H_0: \lambda \geq 2$  versus  $H_1: \lambda < 2$  using the test

$$\text{Reject } H_0 \text{ iff } \bar{X}_{25} < 1.5.$$

**Answers:**

(a) (i) The power function is

$$\begin{aligned} \gamma(p) &= P_p(X_1 + \dots + X_{10} > 5) \\ &= P(Y > 5), \quad Y \sim \text{Binomial}(10, p) \\ &= \sum_{y=6}^{10} \binom{10}{y} p^y (1-p)^{10-y}. \end{aligned}$$

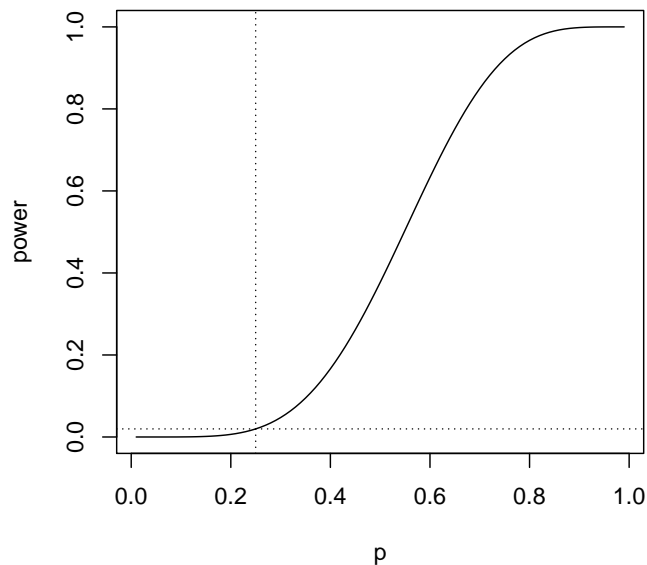
(ii) The size is

$$\begin{aligned}\alpha &= \sup_{p \in (0, 1/4]} \gamma(p) \\ &= \sup_{p \in (0, 1/4]} \sum_{y=6}^{10} \binom{10}{y} p^y (1-p)^{10-y} \\ &= \sum_{y=6}^{10} \binom{10}{y} (1/4)^y (1-1/4)^{10-y} \\ &= 1 - \text{pbinom}(5, 10, 0.25) \\ &= 0.01972771.\end{aligned}$$

(iii) Use the following code:

```
p.seq <- seq(.01, .99, length=99)
power <- 1 - pbinom(5, 10, p.seq)

plot(p.seq, power, type="l", ylim=c(0, 1), xlab="p")
abline(v=0.25, lty=3) # vert line at null boundary
abline(h=0.01972771, lty=3) # horiz line at size
```



(b) (i) The power function is

$$\begin{aligned}\gamma(\mu) &= P_{\mu}(|\sqrt{n}\bar{X}_n| > 2) \\ &= 1 - P_{\mu}(-2 < \sqrt{n}\bar{X}_n < 2) \\ &= 1 - P_{\mu}(-2 - \sqrt{n}\mu < \sqrt{n}(\bar{X}_n - \mu) < 2 - \sqrt{n}\mu) \\ &= 1 - P(-2 - \sqrt{n}\mu < Z < 2 - \sqrt{n}\mu), \quad Z \sim \text{Normal}(0, 1) \\ &= 1 - [\Phi(2 - \sqrt{n}\mu) - \Phi(-2 - \sqrt{n}\mu)]\end{aligned}$$

since  $\sqrt{n}(\bar{X}_n - \mu) = (\bar{X}_n - \mu)/(1/\sqrt{n}) \sim \text{Normal}(0, 1)$ .

(ii) The size is

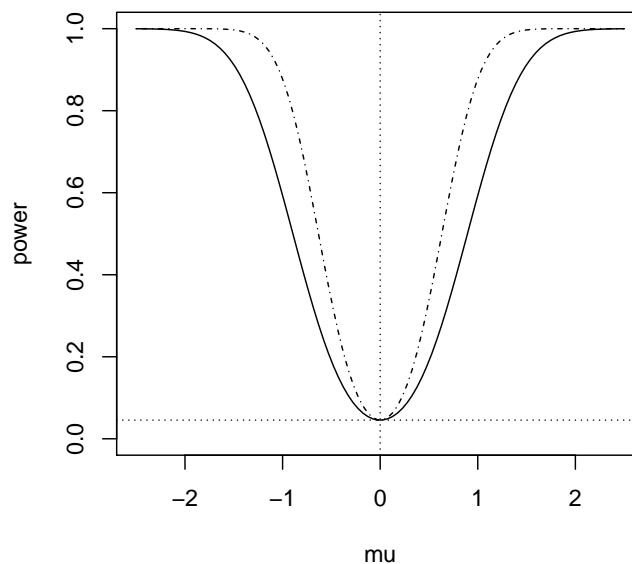
$$\begin{aligned}\alpha &= \sup_{\mu \in \{0\}} \gamma(\mu) \\ &= \gamma(0) \\ &= 1 - [\Phi(2) - \Phi(-2)] \\ &= 1 - (\text{pnorm}(2) - \text{pnorm}(-2)) \\ &= 0.04550026.\end{aligned}$$

(iii) In order to plot the power function, we must choose a sample size  $n$ . The following code plots the power function for  $n = 5$  and  $n = 10$ :

```
mu.seq <- seq(-2.5,2.5,length=100)

power.n5 <- 1-(pnorm(2-sqrt(5)*mu.seq)-pnorm(-2-sqrt(5)*mu.seq))
power.n10 <- 1-(pnorm(2-sqrt(10)*mu.seq)-pnorm(-2-sqrt(10)*mu.seq))

plot(mu.seq,power.n5,type="l",ylim=c(0,1),xlab="mu",ylab="power")
lines(mu.seq,power.n10,ylim=c(0,1),xlab="p",lty=4)
abline(v=0,lty=3) # vert line at null value
abline(h=0.04550026,lty=3) # horiz line at size
```



Note that increasing the sample size increased the power for all  $\mu \neq 0$ ; the size was unaffected, but the Type II error probability was reduced.



(c) (i) The power function is

$$\begin{aligned}\gamma(\lambda) &= P_\lambda(\bar{X}_{25} < 1.5) \\ &= P(Y < 1.5), \quad Y \sim \text{Gamma}(25, \lambda/25), \text{ (show by mgfs)}\end{aligned}$$

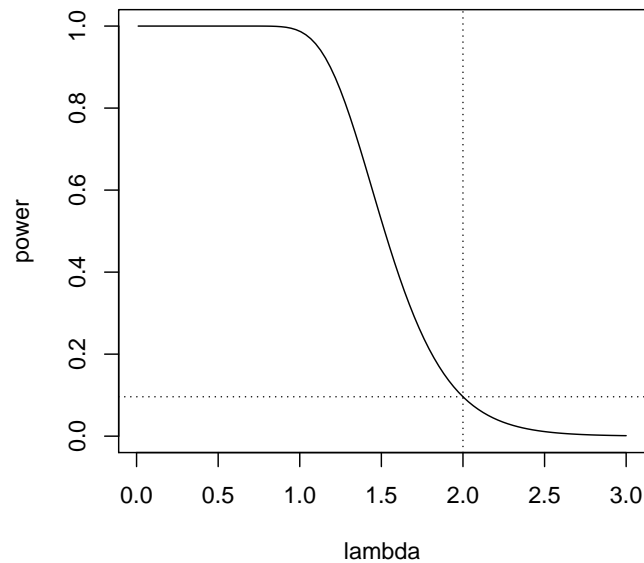
(ii) The size is

$$\begin{aligned}\alpha &= \sup_{\lambda \in [2, \infty)} \gamma(\lambda) \\ &= \gamma(2) \\ &= P(Y < 1.5), \quad Y \sim \text{Gamma}(25, 2/25) \\ &= \text{pgamma}(1.5, \text{shape}=25, \text{scale}=2/25) \\ &= 0.0960497.\end{aligned}$$

(iii) Use the following code:

```
lambda.seq <- seq(.01,3,length=100)
power <- pgamma(1.5,shape=25,scale=lambda.seq/25)

plot(lambda.seq,power,type="l",ylim=c(0,1),xlab="lambda")
abline(v=2,lty=3) # vert line at null boundary
abline(h=0.0960497,lty=3) # horiz line at size
```



## Calibrating the rejection region for desired size

- We often choose the rejection region in order to control the size of the test, that is, to control the Type I error rate (the probability of a Type I error).
- We cannot directly control the Type II error rate, because it depends on the true parameter value, which is unknown to us.
- The smaller the size of a test, the stronger the evidence against  $H_0$  must be in order for the test to reject  $H_0$ .
- **Exercises (cont):** For each of the following tests, (iv) modify the rejection region so that the size is less than or equal to 0.01 and (v) plot the power function of the new test along with the power function of the original test.

(a) Let  $X_1, \dots, X_{10}$  be a random sample from the Bernoulli( $p$ ) distribution, where  $p \in (0, 1)$  is unknown, and suppose you wish to test  $H_0: p \leq 1/4$  versus  $H_1: p > 1/4$  using the test

$$\text{Reject } H_0 \text{ iff } X_1 + \dots + X_{10} \in \{6, 7, 8, 9, 10\}.$$

(b) Let  $X_1, \dots, X_n$  be a random sample from the Normal( $\mu, 1$ ) distribution, where  $\mu \in (-\infty, \infty)$  is unknown, and suppose you wish to test  $H_0: \mu = 0$  versus  $H_1: \mu \neq 0$  using the test

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > 2.$$

(c) Let  $X_1, \dots, X_{25}$  be a random sample from the Exponential( $\lambda$ ) distribution, where  $\lambda \in (0, \infty)$  is unknown, and suppose you wish to test  $H_0: \lambda \geq 2$  versus  $H_1: \lambda < 2$  using the test

$$\text{Reject } H_0 \text{ iff } \bar{X}_{25} < 1.5.$$

### Answers:

(a) (iv) The size of this test, as shown before, is  $0.01972771 > 0.01$ . To make the size smaller, the test needs to require stronger evidence against  $H_0$  in order to reject it. Consider the test

$$\text{Reject } H_0 \text{ iff } X_1 + \dots + X_{10} \in \{7, 8, 9, 10\}.$$

The power function of this test is

$$\gamma(p) = P_p(X_1 + \dots + X_{10} \in \{7, 8, 9, 10\}) = \sum_{y=7}^{10} \binom{10}{y} p^y (1-p)^{10-y},$$

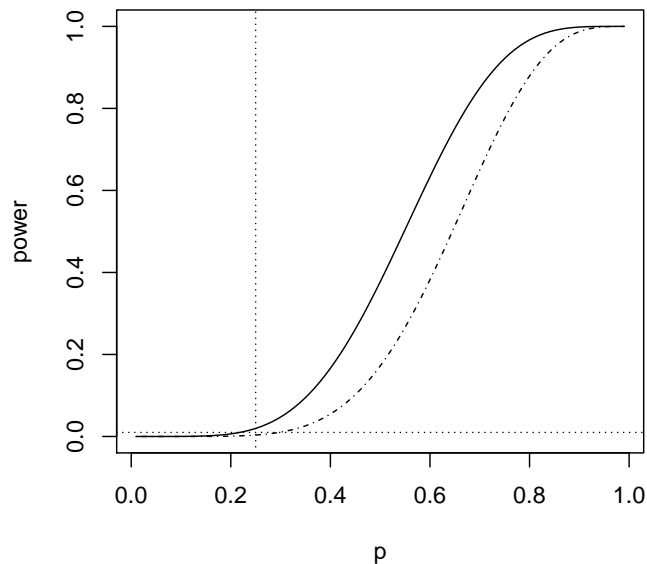
and the size is

$$\begin{aligned}\alpha &= \sup_{p \in (0, 1/4]} \gamma(p) \\ &= \sup_{p \in (0, 1/4]} \sum_{y=7}^{10} \binom{10}{y} p^y (1-p)^{10-y} \\ &= \sum_{y=7}^{10} \binom{10}{y} (1/4)^y (1-1/4)^{10-y} \\ &= 1 - \text{pbinom}(6, 10, 0.25) \\ &= 0.003505707 \\ &< 0.01.\end{aligned}$$

(v) The following R code makes a plot of the power curves of the two tests:

```
p.seq <- seq(.01, .99, length=99)
power.gt5 <- 1 - pbinom(5, 10, p.seq)
power.gt6 <- 1 - pbinom(6, 10, p.seq)

plot(p.seq, power.gt5, type="l", ylim=c(0,1), xlab="p", ylab="power")
lines(p.seq, power.gt6, lty=4)
abline(v=0.25, lty=3) # vert line at null boundary
abline(h=0.01, lty=3) # horiz line at max. allowed size
```



Note that the test with greater size also has greater power across all values of  $p$ ; it is more prone to make Type I errors, but more likely to reject  $H_0$  when it is false, i.e. less prone to make Type II errors. This illustrates the size-power trade-off.

(b) (iv) The size of this test, as shown before, is  $0.04550026 > 0.01$ . To make the size smaller, the test needs to require stronger evidence against  $H_0$  in order to reject it. Suppose we

try

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > 3.$$

This test, following previous work, has power function

$$\begin{aligned}\gamma(\mu) &= P_\mu(|\sqrt{n}\bar{X}_n| > 3) \\ &= 1 - [\Phi(3 - \sqrt{n}\mu) - \Phi(-3 - \sqrt{n}\mu)],\end{aligned}$$

and the size is

$$\begin{aligned}\alpha &= \sup_{\mu \in \{0\}} \gamma(\mu) \\ &= \gamma(0) \\ &= 1 - [\Phi(3) - \Phi(-3)] \\ &= 1 - (\text{pnorm}(3) - \text{pnorm}(-3)) \\ &= 0.002699796 \\ &< 0.01.\end{aligned}$$

So this test has size less than 0.01; however, the smaller the size, the smaller the power of the test under the alternative hypothesis, so we should not make the size smaller than it has to be. We can make the size exactly 0.01 by considering a test of the form

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > C,$$

and choosing  $C$  such that the size of the test is exactly 0.01. The power of the above test, following previous work, is

$$\gamma(\mu) = P_\mu(|\sqrt{n}\bar{X}_n| > C) = 1 - [\Phi(C - \sqrt{n}\mu) - \Phi(-C - \sqrt{n}\mu)],$$

and the size is

$$\alpha = \sup_{\mu \in \{0\}} \gamma(\mu) = \gamma(0) = 1 - [\Phi(C) - \Phi(-C)] = P(Z > C) + P(Z < -C),$$

where  $Z \sim \text{Normal}(0, 1)$ . This equation gives  $C = z_{\alpha/2}$  since  $P(Z > z_{\alpha/2}) = P(Z < -z_{\alpha/2}) = \alpha/2$ . To achieve the size  $\alpha = 0.01$ , we therefore take  $C = z_{.005} = \text{qnorm}(.995) = 2.575829$ .

(v) The following R code plots the power curves of the test

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > C,$$

for  $C = 2$ ,  $C = 2.575829$ , and  $C = 3$  under the sample size  $n = 10$ .

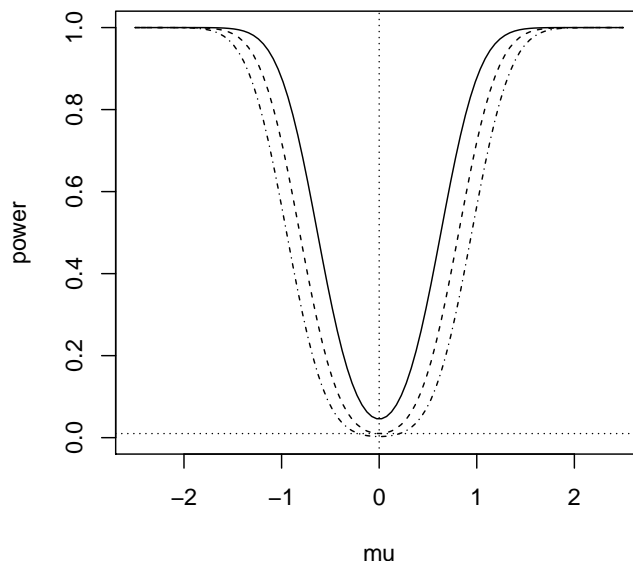
```

mu.seq <- seq(-2.5,2.5,length=100)

power2 <- 1-(pnorm(2-sqrt(10)*mu.seq)-pnorm(-2-sqrt(10)*mu.seq))
power2.575829 <- 1-(pnorm(2.575829-sqrt(10)*mu.seq)
                  -pnorm(-2.575829-sqrt(10)*mu.seq))
power3 <- 1-(pnorm(3-sqrt(10)*mu.seq)-pnorm(-3-sqrt(10)*mu.seq))

plot(mu.seq, power2,type="l",ylim=c(0,1),xlab="mu",ylab="power")
lines(mu.seq, power2.575829,ylim=c(0,1),xlab="p",lty=2)
lines(mu.seq, power3,ylim=c(0,1),xlab="p",lty=4)
abline(v=0,lty=3) # vert line at null value
abline(h=0.01,lty=3) # horiz line at max. allowed size

```



The test with  $C = 2.575829$ , which has size equal to 0.01 has greater power when  $\mu \neq 0$  than the test with  $C = 3$ , which has slightly smaller size. If we are willing to accept a Type I error rate of 0.04550026, then we can use the test with  $C = 2$ , which has the greatest power.

- (c) (iv) The size of this test, as shown before, is  $0.0960497 > 0.01$ . To make the size smaller, the test needs to require stronger evidence against  $H_0$  in order to reject it. We can make the size exactly 0.01 by considering a test of the form

$$\text{Reject } H_0 \text{ iff } \bar{X}_{25} < C$$

and choosing  $C$  such that the size of the test is exactly 0.01. The power of the above test, following previous work, is

$$\gamma(\lambda) = P_\lambda(\bar{X}_{25} < C) = P(Y < C), \quad Y \sim \text{Gamma}(25, \lambda/25) \text{ (show by mgfs) ,}$$

and the size is

$$\alpha = \sup_{\lambda \in [2, \infty)} \gamma(\lambda) = \gamma(2) = P(Y < C), \quad Y \sim \text{Gamma}(25, 2/25).$$

This equation says that if we choose  $C$  to be the  $\alpha$  quantile of the  $\text{Gamma}(25, 2/25)$  distribution, the test will have size  $\alpha$ . For  $\alpha = 0.01$ , we have

$$C = \text{qgamma}(0.01, \text{shape}=25, \text{scale}=2/25) = 1.188267.$$

(v) The following R code plots the power curves of the test

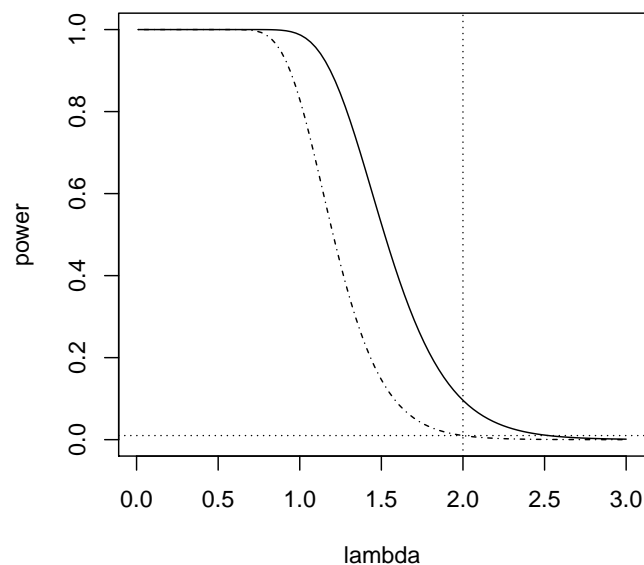
$$\text{Reject } H_0 \text{ iff } \bar{X}_{25} < C,$$

for  $C = 1.5$  and  $C = 1.188267$ .

```
lambda.seq <- seq(.01,3,length=100)

power1.5 <- pgamma(1.5,shape=25,scale=lambda.seq/25)
power1.188267 <- pgamma(1.188267,shape=25,scale=lambda.seq/25)

plot(lambda.seq,power1.5,type="l",ylim=c(0,1),xlab="lambda",ylab="power")
lines(lambda.seq,power1.188267,lty=4)
abline(v=2,lty=3) # vert line at null boundary
abline(h=0.01,lty=3) # horiz line at max. allowed size
```



The test with smaller size has smaller power. This illustrates the size-power trade-off.