Math 152 - Statistical Theory - Exam 2

Ethan Ashby

Due: Monday, Nov 16, 2020, 5pm PDT

1. Cookies... Mmmmm... In the mid-1990s a Nabisco marketing campaign claimed that there were at least 1000 chips in every bag of *Chips Ahoy!* cookies. Here is one of their ads: https://www.youtube.com/watch?v=xjM7nDn7XBU.

Let's say we want to investigate whether the 0.001 quantile of the number of chips normal distribution is above 1000. Consider the following two statistics taken on a random sample of size $n = 1000$ bags:

$$\hat{\theta}_1 = \min(X_i)$$
$$\hat{\theta}_2 = \overline{X} - 3.5 \cdot s$$

Let X = number of chips in a bag of *Chips Ahoy!* cookies. Assume $X \sim N(\mu, (115)^2)$.

a. (+3 pts) Determine the value of μ such that the 0.001 quantile (i.e., 0.1%) is 1000.

b. (+3 pts) In your own words (do not try to take expected values) describe the population value that each of $\hat{\theta}_1$ and $\hat{\theta}_2$ estimate.

c. (+10 pts) Run an analysis using bootstrap CIs to investigate whether or not the *Chips Ahoy!* claim will be verified. Follow these steps:

   1. simulate $n = 1000$ observations from $N(\mu, \sigma^2 = (115)^2)$
   2. Find two percentile bootstrap intervals (separately) using $\hat{\theta}_1$ and $\hat{\theta}_2$.
   3. Repeat steps 1 and 2 many times.
   4. For each of the two estimators, record how many times the value of "1000 chips" is completely above the interval.

```
#I'm finding 99% BS percentile CIs for the two test statistics
#I'm generating data from a normal distribution with mean=1355 and sd=115
#Then I'm comparing 1000 to each BS percentile interval and seeing how many times 1000 is above that in
set.seed(11)
above_theta_1<-c()
above_theta_2<-c()
alpha=0.01
for (i in 1:1000){
  #sample from population
  samp<-rnorm(1000, mean=1355.3767, sd=115)
  theta_1<-c()
  theta_2<-c()
  #pull out 1000 bs samples
  for(j in 1:1000){
```

```
    bs_samp<-sample(samp, size=1000, replace=TRUE)
    theta_1<-c(theta_1, min(bs_samp))
    theta_2<-c(theta_2, mean(bs_samp)-3.5*sd(bs_samp))
  }
  theta_1_int<-c(quantile(theta_1, alpha/2), quantile(theta_1, 1-alpha/2))
  theta_2_int<-c(quantile(theta_2, alpha/2), quantile(theta_2, 1-alpha/2))
  #test if above interval
  above_theta_1<-c(above_theta_1, 1000>theta_1_int[2])
  above_theta_2<-c(above_theta_2, 1000>theta_2_int[2])
}

paste("1000 was above the theta_1 intervals", sum(above_theta_1), "out of 1000 CIs", sep=" ")
```

```
## [1] "1000 was above the theta_1 intervals 2 out of 1000 CIs"
```

```
paste("1000 was above the theta_2 intervals", sum(above_theta_2), "out of 1000 CIs", sep=" ")
```

```
## [1] "1000 was above the theta_2 intervals 988 out of 1000 CIs"
```

First estimator has intervals below 1000 in 2 out of 1000 confidence intervals. Second estimator has intervals below 1000 in 988 out of 1000 confidence intervals.

  d. (+3 pts) Given the output from your CIs (in #4) and your answer to b. above, what can you say about the *Chips Ahoy!* claim? In 2-3 sentences, explain how the CIs relate (or don't) to the claim of interest, and how the two different statistics might be producing intervals for different things.

If we consider the *Chips Ahoy!* claim: ALL packages have more than 1000 cookies, we can reject this claim if we find one bag of chips that has less than 1000 cookies. It also suffices to find a confidence interval of a relevant statistic that lies below 1000, as this implies that some of our data (bags) have less than 1000 cookies. The first statistic (sample minimum) is a conservative estimate of the parameter of interest (the population minimum), meaning the population minimum is going to be less than or equal to the sample minimum. The fact that 2 out of 1000 confidence intervals for the sample minimum lie entirely below 1000 indicates that some bags do indeed have less than 1000 chips. So we can reject the claim. The second statistic $(\bar{x} - 3.5 * s)$ is less conservative than the sample minimum, BUT while the sample minimum represent a real data point (capacity of a bag of chips), $\bar{x} - 3.5 * s$ does not necessarily represent a real data point. Thus, I would consider the BS confidence intervals of the sample minima, which indicate that there must exits bags of chips with less than 1000 cookies.

2. Two normal tests Let X_1, X_2 be independently distributed $N(\theta, \sigma^2 = 0.81)$. We'd like to test:

$$H_0 : \theta = 4$$
$$H_1 : \theta \neq 4$$

We have two competing tests:

$$\delta_1 \quad = \quad \{\text{reject } H_0 \text{ if } X_1 > 5.48\}$$
$$\delta_2 \quad = \quad \{\text{reject } H_0 \text{ if } X_1 + X_2 > c\}$$

a. (+5 pts) Find c so that the size of δ_2 is the same as the size of δ_1. We know the size of test 1 is equivalent to $\Pr(X > 5.48|H_0 \text{ true})$. The size of the first test is thus equivalent to pnorm(5.48, mean=4, sd=0.9, lower.tail=FALSE)=`pnorm(5.48, mean=4, sd=0.9, lower.tail=FALSE)`. Since $X_1, X_2 \overset{\text{i.i.d.}}{\sim} N(\theta, \sigma^2 = 0.81)$, then $Y = X_1 + X_2 \sim N(2\theta, 2\sigma^2)$. Under the null, $Y \sim N(8, \sigma^2 = 1.62)$. To identify the appropriate quantile of this distribution, we use qnorm(0.05, mean=8, sd=sqrt(1.62), lower.tail=FALSE)=`qnorm(0.05, mean=8, sd=sqrt(1.62), lower.tail=FALSE)`.

b. (+5 pts) Plot the power function for each test (ideally on the same graph, let me know if I can help with the R code).

```
x = seq(0,14,0.01)
y = pnorm(5.48, mean=x, sd=0.9, lower.tail=FALSE)
plot(x,y, xlab = 'theta', ylab = 'power', 'l', col="red")
abline(v=4, col="magenta", lty=2)
abline(h=0.05004222, col="magenta", lty=2)
abline(v=8, col="magenta", lty=2)
y2 = pnorm(10.09, mean=2*x, sd=sqrt(1.62), lower.tail=FALSE)
lines(x,y2, xlab = 'theta', ylab = 'power', 'l', col="blue")
```
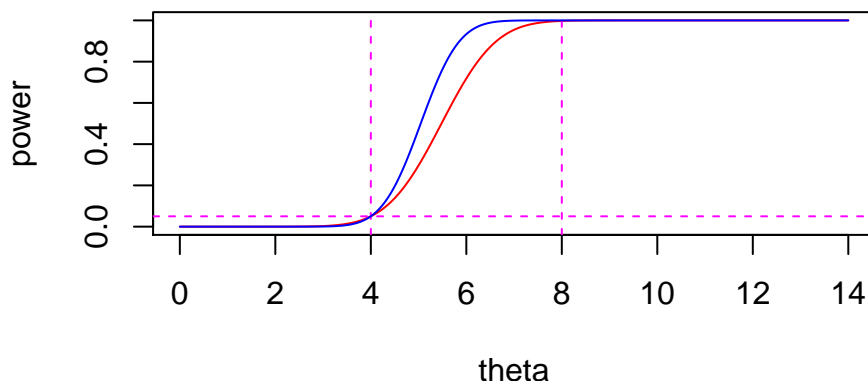

c. (+3 pts) Where are $\beta(\delta_1)$ and $\beta(\delta_2)$ (the type II error rate at a particular value of θ) the largest? Find / approximate the largest the type II error will be for each test.

β is the probability of making a type II error, i.e. failing to reject the null hypothesis when we should. Given the one-sidedness of the test, $\beta(\delta_1)$ and $\beta(\delta_2)$ will both be largest when θ is small. In this case, the null hypothesis will be false, but according to the power function plots, we will almost never reject the null. Thus for $\theta \ll 4$, $\beta(\delta_1)$ and $\beta(\delta_2) \approx 1$.

d. (+4 pts) For what value of θ is the power function highest for each test? Is the power at that "best" value of θ particularly good? Explain.

Both power functions are highest for large values of θ, or $\theta > 8$, and have values of approximately 1. This makes sense: if the data are distributed normally with a large mean, the probability of the data exceeding thresholds will be greater. Power=1 when the null hypothesis is false is great, as we always want to reject H_0 when it isn't true.

e. (+5 pts) Which test is more powerful when H_0 is not true (specify the range of values of interest in order to make the determination of which is more powerful)? Why does your answer make sense?

The second test (reject if $X_1 + X_2 > 10.09$) is more powerful, particularly over the range $\theta \in [4, 8]$. This is evidenced by the blue line (power function of second test) lying above the red line (power function of the first test). This makes sense intuitively, because increasing sample size is a surefire way to improve the power of your test. In this case, we increased the sample size from $n = 1$ in the first test to $n = 2$ for the second test.

3. The absolute value or the squares? Let $X_1, X_2, \ldots, X_{200}$ (that is $n = 200$) be iid with pdf

$$f_X(x) = \frac{\lambda}{2}\exp(-\lambda|x|) \qquad -\infty < x < \infty \qquad \lambda > 0$$

a. (+10 points) Derive a test for the following hypotheses that has size $= 0.05$. It is okay to have a "c" in the test. Please be clear as to what the **test** is:

$$H_0 : \qquad \lambda = 47$$
$$H_1 : \qquad \lambda = 50$$

b. (+4 points) In your own words, explain why your rejection region makes for a good test (answer in terms of intuition not in terms of "because the theorem said so").

c. (+10 pts) Using the asymptotic results from class, find the value of "c" needed in the test, call it c_A. You will need to find the MLE for λ and the Fisher Information in λ. Show your work.

d. (+10 pts) Using a simulation, find the value of "c" needed in the test, call it c_S. Follow these steps:

1. Generate $n = 200$ data values from the distribution above (in looking for "c" you should know whether you are supposed to set H_0 or H_1 to be true).
2. For each dataset, calculate the test statistics (T) from a.
3. Repeat #1 and #2 many times.
4. Find c_S from the many T values in #3.

```
library(nimble)
set.seed(11)
#generate data under null (lambda=47).
Test<-c()
for (i in 1:1000){
  #generate
  samp<-rdexp(200, location = 0, rate = 47)
  #calculate test stats
  Test<-c(Test, sum(abs(samp)))
}
#this gives the 95th quantile under the null for the sum of abs of these random variables
quantile(Test, 0.05)
```

```
##        5%
## 3.781048
```

So our $c_s = 3.781048$.

e. (+10 pts) Estimate the power of your test under Ω_1 given the value of c_s and the same basic idea from part d.

```
set.seed(11)
#generate data under alternative (lambda=50).
Test<-c()
for (i in 1:1000){
  #generate data
  samp<-rdexp(200, location = 0, rate = 50)
  #calculate test stats
  Test<-c(Test, sum(abs(samp)))
}
#how many of these samples under H1 actually reject the null according to c_s
sum(Test<3.781048)/1000
```

```
## [1] 0.247
```

The power of this test under Ω_1 given the value of $c_s = 3.781048$ is 0.247.

    f. (+10 pts) Repeat the entire process of both d. and e. using $T_2 = \sum_{i=1}^{200} x_i^2$. Which is more powerful, the test which uses T (from a.) or the test that uses T_2? Explain using your simulation results.

```r
set.seed(11)
#generate data under null (lambda=47).
Test<-c()
for (i in 1:1000){
  #generate
  samp<-rdexp(200, location = 0, rate = 47)
  #calculate test stats
  Test<-c(Test, sum(samp^2))
}
#this gives the 95th quantile under the null for the sum of abs of these random variables
quantile(Test, 0.05)
```

```
##         5%
## 0.1389298
```

```r
#generate data under alternative (lambda=50).
Test<-c()
for (i in 1:1000){
  #generate data
  samp<-rdexp(200, location = 0, rate = 50)
  #calculate test stats
  Test<-c(Test, sum(samp^2))
}
#how many of these samples under H1 actually reject the null according to c_s
sum(Test<0.1389298)/1000
```

```
## [1] 0.205
```

The power of this new test under Ω_1 given the value of $c_s = 0.1389298$ is 0.205. The power of the new test is worse than the previous test, b/c by the Neyman-Pearson lemma, we know that the first test has the highest power of any test given our particular level of significance. So the second test never stood a chance!

4. Pivot (+10 pts) Consider a single observation $X \sim U[0,\theta]$. Use the observation to find a 99% CI for θ. You should be able to find any constants you need. Hint: consider $Y = X/\theta$.

Evaluate result empirically:

```r
set.seed(11)
a<-c()
for(i in 1:1000){
x=runif(1, 0, 1)
a<-c(a, between(1, 200*x/199, 200*x))}
sum(a)/1000
```

```
## [1] 0.992
```

Yep! Our confidence interval procedure captures the parameter 99.2% of the time! So our pivotal method works!