ANLY 511: Assignment #3

Brody Vogel

September 20th

Preparation

```
set.seed(1234)
```

Problem 17

$E[X] \approx .384$;

$s[X] \approx .321$;

$E[X^{1/3}] \approx 2.878$

```
betas <- rbeta(100000, .5, .8)
mean(betas)
```

```
## [1] 0.383107
```

```
standardDev <- sd(betas)
standardDev
```

```
## [1] 0.3203473
```

```
lastOne <- mean(betas^(-1/3))
lastOne
```

```
## [1] 2.727044
```

Problem 18

$E[\sqrt(X)] \approx .5437$, which can be shown with ≈ 1.4 million simulations.

The standard error is the standard deviation of a statistic's sampling distribution, so it can be used to find the value at which we can be fairly sure that our estimation of $E[\sqrt(X)]$ is within .001 of the true statistic.

To be safe, I ran the script below so that it found the number of simulations at which the standard error is .00025. I did this so that the probability of getting a sample mean that is more than .001 from the average is $< 5\%$, which the Central Limit Theorem tells us will be the case.

```
n <- 0
stE <- 2
while (stE > .00025) {
  n <- n + 20000
  stE <- sd(rbeta(n, .5, .8)^(1/2))/sqrt(n)
}
print(n)
```

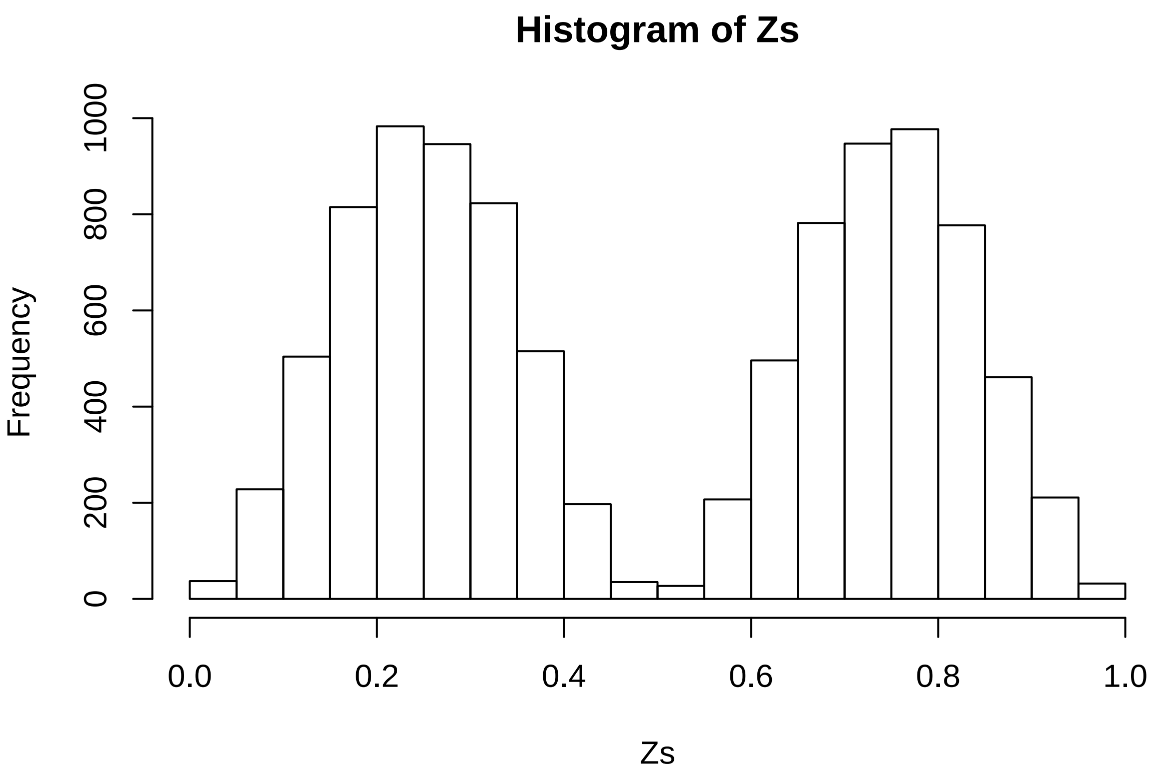
```
## [1] 1440000
```

```
print(mean(rbeta(n, .5, .8)^(1/2)))
```

```
## [1] 0.543976
```

Problem 19

```
counter <- 0
Zs <- c()
while (counter < 10000) {
  x <- runif(1)
  y <- runif(1)
  if (y <= sin(2*pi*x)^2) {
    Zs <- c(Zs, x)
    counter <- counter + 1
  }
}
hist(Zs)
```



It looks like Z is bimodal with relative means on the intervals (.2, .3) and (.7, .8). This is because the values of $sin(2\pi X)^2$ on (.2, .3) and (.7, .8) are all at or above .9. Since Y is ~U(0,1), it makes sense that the value of Z would most often be X values that, after the $sin(2\pi X)^2$ transformation, are well above E[Y] = .5.

Problem 20

Pr(+|Drunk) = .99 ; Pr(+|Sober) = .02

a) sensitivity = .99 ; specificity = .98

A test's "sensitivity" is its ability to correctly identify those who should test positive, i.e., fit the conditions.

A test's "specificity" is its ability to correctly identify those who should test negative, i.e., do not fit the conditions.

b)

$Pr(Drunk|+) = Pr(+|Drunk)\frac{Pr(Drunk)}{Pr(+)} = .99\frac{.002}{\frac{499(.02)+1(.99)}{500}} = .0902$, assuming everyone is equally likely to be pulled over.

Problem 21

(1)

$Pr(X < 20) \approx .9991$;

$Pr(X > 10|X < 20) = \frac{Pr(X>10)-Pr(X\geq 20)}{Pr(X<20)} \approx .4159$

(2) and (3)

$Pr(X < 60) \approx .9192$;

$Pr(X < 60|X > 30) \approx .9191$;

$Pr(X > 30|X < 60) \approx .9990$

```
##(1)
pbinom(19, 50, prob = .2)
```

```
## [1] 0.9990676
```

```
(1 - pbinom(10, 50, prob = .2) - (1 - pbinom(19, 50, prob = .2)))/pbinom(19, 50, prob = .2)
```

```
## [1] 0.4158959
```

```
##(2) and (3)
binoms <- rbinom(100000, 500, prob = .1)
length(binoms[binoms < 60])/100000
```

```
## [1] 0.91927
```

```
p <- binoms[binoms > 30]
length(p[p < 60])/length(p)
```

```
## [1] 0.9191876
```

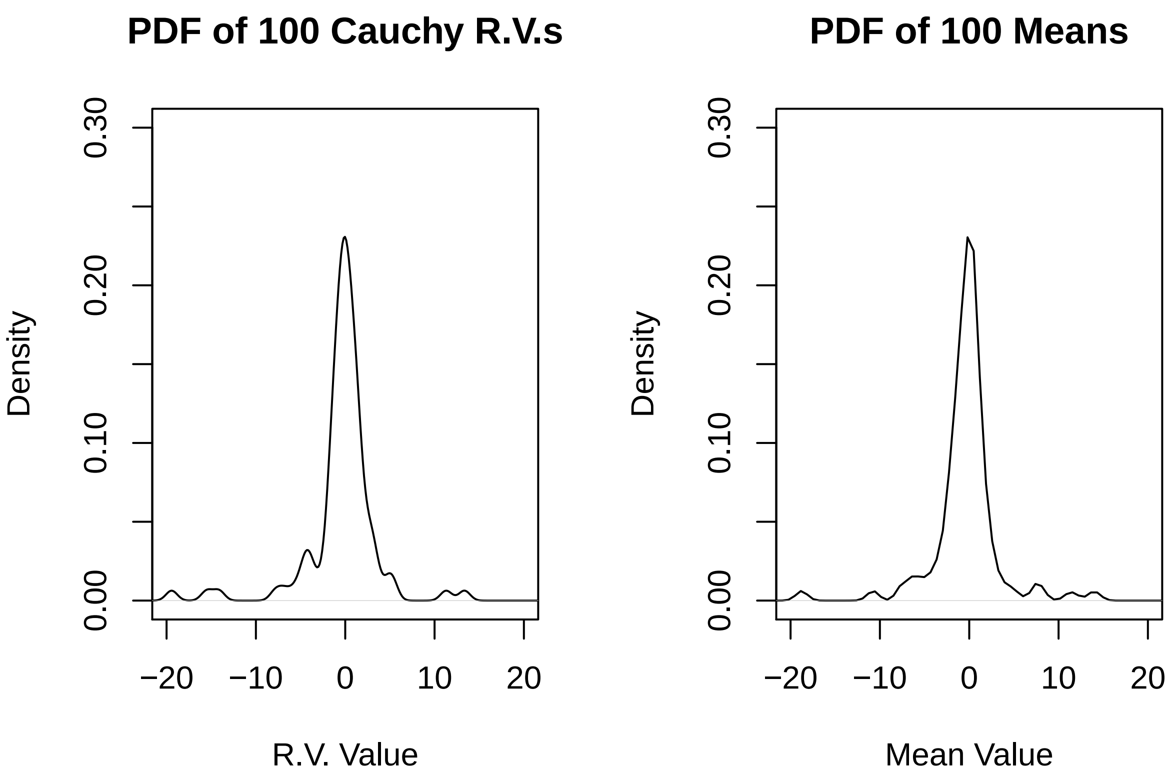
```
q <- binoms[binoms < 60]
length(q[q > 30])/length(q)
```

```
## [1] 0.9988904
```

Problem 22

```r
#for (x in seq(100, 5000, 500)) {
#cauchys1 <- rcauchy(x)
#plot(density(cauchys1), xlim = c(-20, 20))
#cauchy1means <- replicate(x, mean(rcauchy(x)))
#plot(density(cauchy1means), xlim = c(-20, 20))
#}

cauchys1 <- rcauchy(100)
par(mfrow = c(1,2))
plot(density(cauchys1), xlim = c(-20, 20), ylim = c(0, .3), main = 'PDF of 100 Cauchy R.V.s', xlab = 'R
cauchy1means <- replicate(100, mean(rcauchy(100)))
plot(density(cauchy1means), xlim = c(-20, 20), ylim = c(0, .3), main = 'PDF of 100 Means', xlab = 'Mean
```




```r
par(mfrow = c(1,1))
```

Because the PDF of a Cauchy R.V. with location 0 and scale 1 centers around 0, both the above PDFs for a sample of 100 such random variables and for a sample of 100 *means* of 100 *samples* of 100 such random variables center around 0, too. Neither are perfectly normal, though, because the Cauchy R.V. has an undefined variance. This means the distribution actually gets *less* stable as the number of samples increases, which is why I've chosen to highlight such a small sample size, here.

Problem 23

a) $F(x) = 1 - e^{-\lambda x}$, because X~Exp($\lambda$).

b) $Pr(X > A) = 1 - Pr(X \leq A) = 1 - (1 - e^{-\lambda A}) = e^{-\lambda A}$

c) $Pr(X > z \cap X > A) = Pr(X > z|X > A) * Pr(X > A)$, by Bayes' Theorem.

d) From (c), $Pr(X > z|X > A) = \frac{Pr(X>z \cap X>A)}{Pr(X>A)}$. If A > z, the right side of this equation is equal to 1. In the context of the problem, this isn't useful. So, assuming z > A, $\frac{Pr(X>z \cap X>A)}{Pr(X>A)} = \frac{Pr(X>z)}{Pr(X>A)}$. Thus, $Pr(X > z|X > A) = \frac{Pr(X>z)}{Pr(X>A)} = \frac{e^{-\lambda A}}{e^{-\lambda z}} = e^{-\lambda(A-z)}$, for a given z.

e) With respect to the R.V. Y, $Pr(Y > y) = Pr([X|X > A] - A > y) = Pr([X|X > A] > y + A) = Pr(X > y + A|X > A)$.

f) Using what we know from (d) and (e), if we substitute y + A for z, we know $Pr(Y > y) = Pr(X > y+A|X > A) = e^{-\lambda(y+A-A)} = e^{-\lambda y}$. Rearranging things, then, $Pr(Y \leq y) = 1-Pr(Y > y) = 1-e^{-\lambda y}$.

Furthermore, from (b), we know $Pr(X > y) = e^{-\lambda y}$. So $Pr(X \leq y) = 1 - Pr(X > y) = 1 - e^{-\lambda y}$.

So X and Y have the same cumulative distribution function.