Lab 06

Jennifer Lin

jenniferyjlin@berkeley.edu (mailto:jenniferyjlin@berkeley.edu)

1

From a conceptual standpoint, the expected value and the mean are the same thing, but in which sort of context or scenario would you use each of these terms? Give examples for both.

```
## Expected Value ##
# Expected value is usually used in probability distribution. Given some random varia
bles, we can calcualte the expected value.
# Expected value is the average value of a random variable over a large number of exp
eriments.
# For example, given a weighted dice (the probablility of getting each side one is po
rpotion to the value of that side),
# we can say that the expected value of rolling a large number of dice is (1*1 + 2*2
 + 3*3 + 4*4 + 5*5 + 6*6 ) / (1+2+3+4+5+6)


## Mean ##
# Mean is defined as the sum of all the elements divided by the sum of their frequenc
ies.
# Mean is frequently used to describe a set of values in real life.
# For example, after an exam, every student gets a grade for that.
# If we want to evaluate how these students perform on the exam, we can calculate the
mean of their scores by summing up their scores and then dividing by the numbers of t
he students.
```

2

Why does the cumulative distribution function of the exponential distribution look like a reflection of the PDF?

Do the math by hand to prove why that is the case. What assumption about a parameter must be made in order for the CDF to be an inversion about the x-axis?

CDF of exponential distribution

$$F(x) = P(X \geqslant x)$$

$$= \int_0^x f(w)\, dw$$

$$= \int_0^x \lambda e^{-\lambda w}\, dw$$

$$= \left[-e^{-\lambda w} \right]_0^x$$

$$= 1 - e^{-\lambda x}, \quad x > 0$$

$$CDF \; F(x) = \begin{cases} 0 & , \; x \leqslant 0 \\ 1 - e^{-\lambda x} & , \; x > 0 \end{cases}$$

$$PDF \; f(x) = \lambda \cdot e^{-\lambda x}$$

when $x > 0$ $\begin{cases} CDF \; F(x) = 1 - e^{-\lambda x} \\ PDF \; f(x) = \lambda \cdot e^{-\lambda x} \end{cases}$

$$-\frac{f(x)}{\lambda} + 1 = F(x)$$ (PDF, CDF labeled)

Thus, CDF look like a reflection of the PDF.

if $\lambda = 1 \rightarrow$ The CDF will be an inversion of the PDF about the x-axis.

3

Write a function that creates a logarithmic seq (e.g. 1,10,100,1000,...) and populate a vector up to 10 elements.

Make random draws from the exponential distribution and plot the variance and expected value of each instance. When does it appear to normalize?

```
library(ggplot2)
logSeq_fun <- function(length_N){
  return(10^(seq(from=0,length.out = length_N)))
}
logSeq <- logSeq_fun(10)
print(logSeq)
```
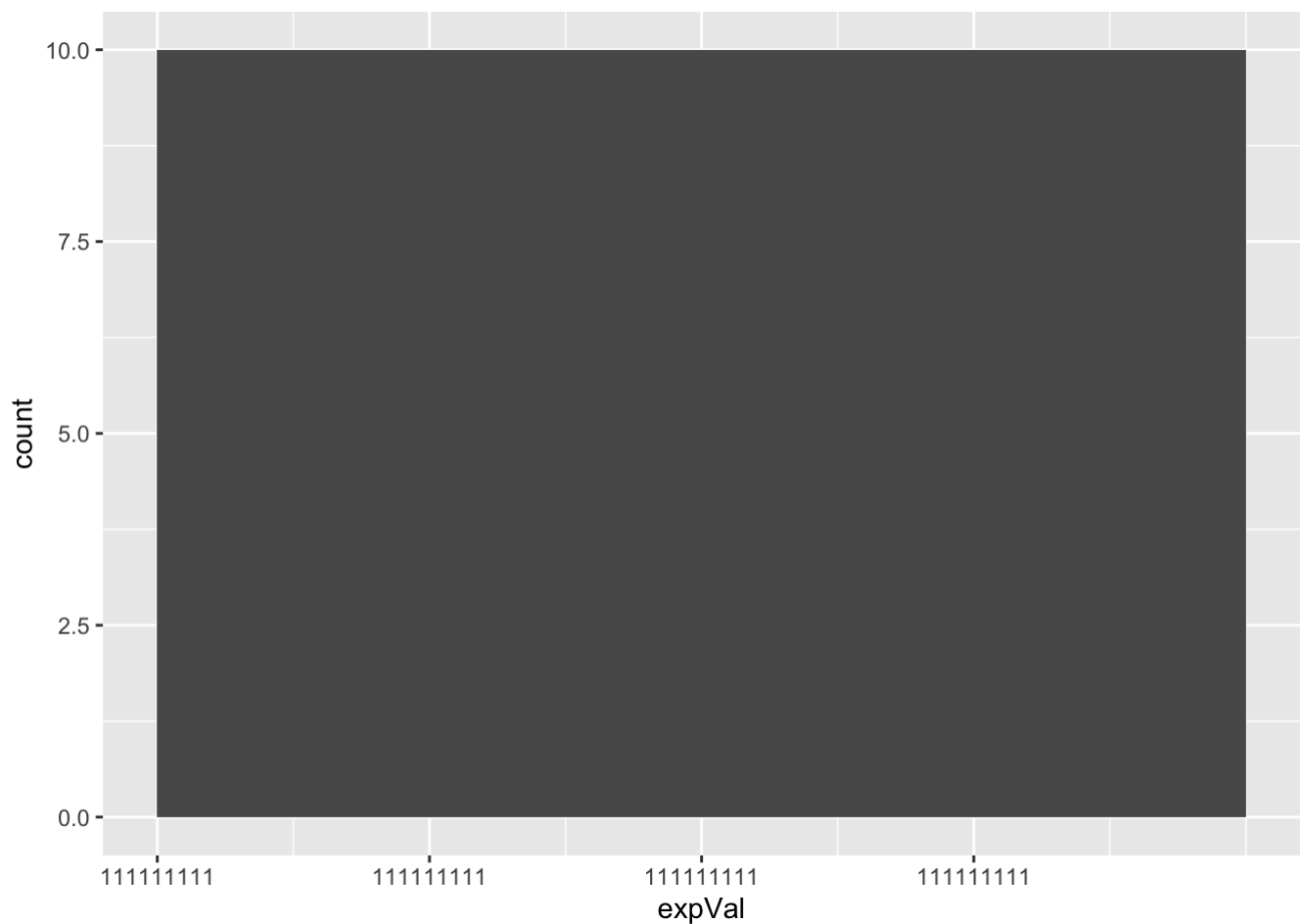
```
##  [1] 1e+00 1e+01 1e+02 1e+03 1e+04 1e+05 1e+06 1e+07 1e+08 1e+09
```

```
# 10 random draws (J=10)
J <- 10
I <- 100
var <- c()
expVal <- c()
for(j in 1:J){
  randomDraws <- c()
  for (i in 1:I){
    randomDraws[i] <- sample(logSeq, 1)
  }
  var[j] <- var(randomDraws)
  expVal[j] <- mean(logSeq)
}
data10 <- data.frame(cbind(xVal=1:J,var,expVal))
# ggplot(data,aes(x=xVal, y = var)) + geom_point()
# ggplot(data,aes(x=xVal, y = expVal)) + geom_point()
ggplot(data10, aes(x=var)) + geom_histogram(bins=20) # VAR: When there are only 10 ra
ndom draws, the result does not appear to normalize.
```
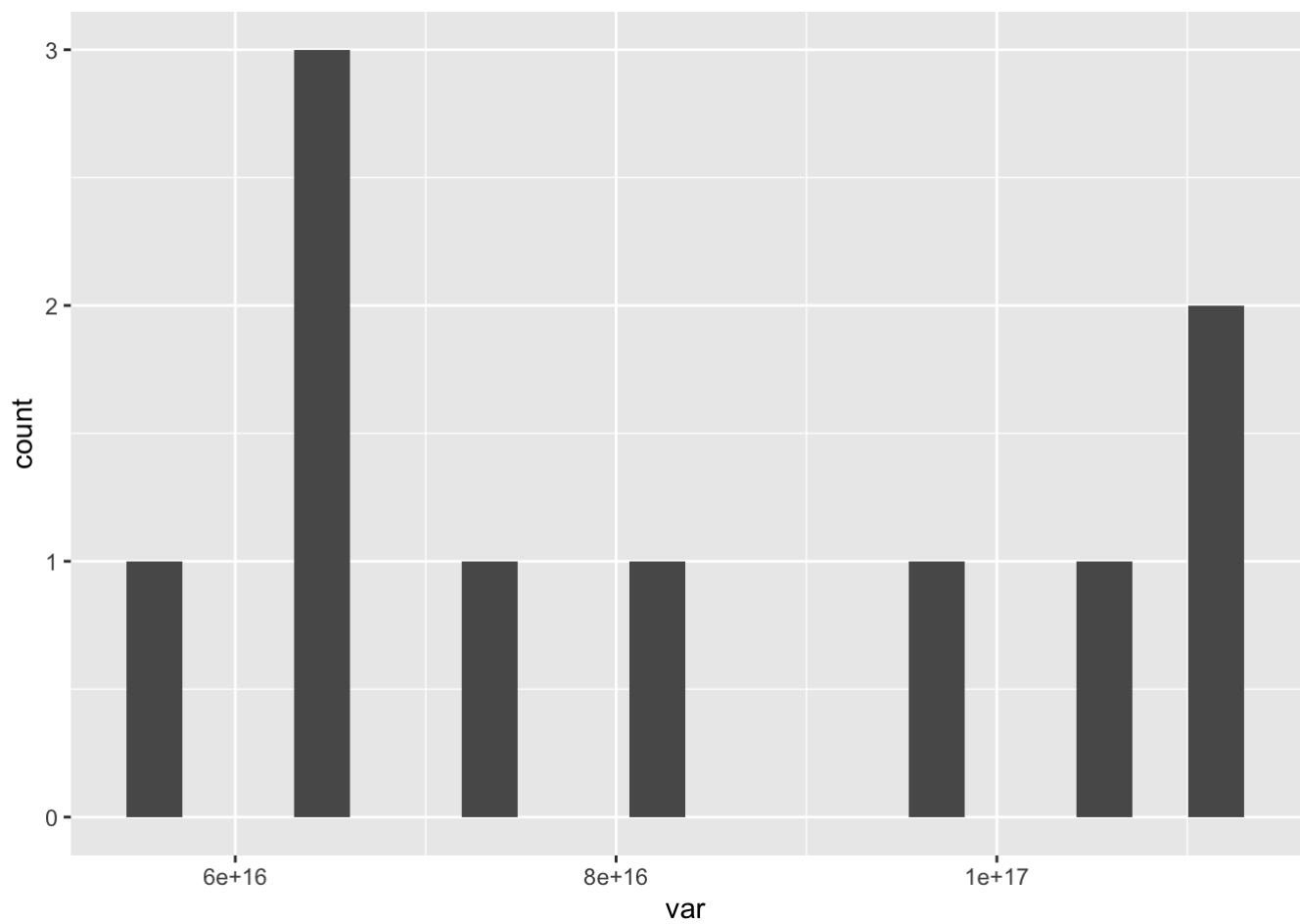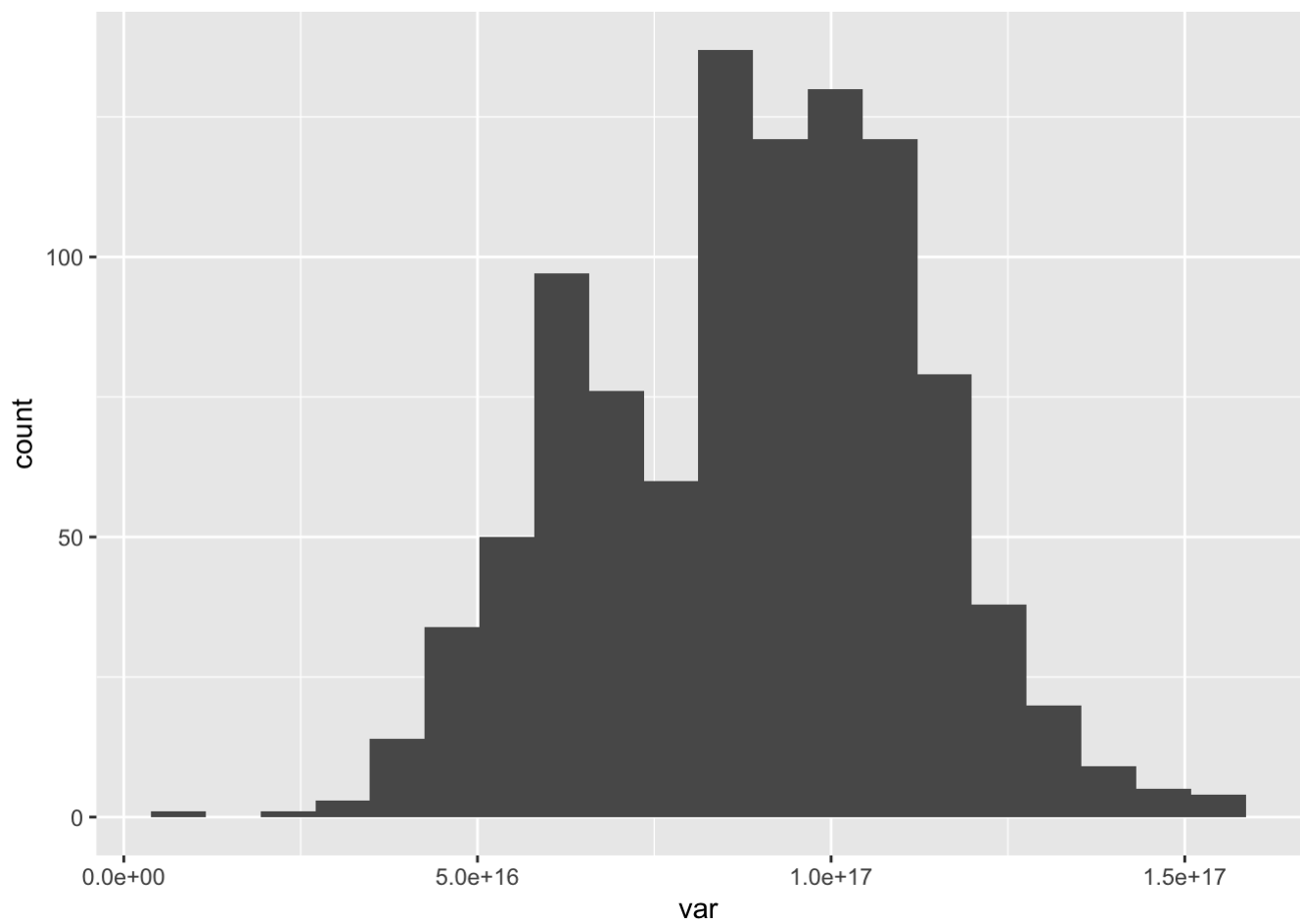
```
ggplot(data10, aes(x=expVal)) + geom_histogram() # EXPECTED VALUE: The result for exp
ected value is always 111111111.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# 1000 random draws (J=10)
J <- 1000
I <- 100
var <- c()
expVal <- c()
for(j in 1:J){
  randomDraws <- c()
  for (i in 1:I){
    randomDraws[i] <- sample(logSeq, 1)
  }
  var[j] <- var(randomDraws)
  expVal[j] <- mean(logSeq)
}
data1000 <- data.frame(cbind(xVal=1:J,var,expVal))
# ggplot(data,aes(x=xVal, y = var)) + geom_point()
# ggplot(data,aes(x=xVal, y = expVal)) + geom_point()
ggplot(data1000, aes(x=var)) + geom_histogram(bins=20) # VAR: When there are 1000 ran
dom draws, the result does  appear to normalize.
```

```
ggplot(data1000, aes(x=expVal)) + geom_histogram() # EXPECTED VALUE: The result for e
xpected value is always 111111111.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
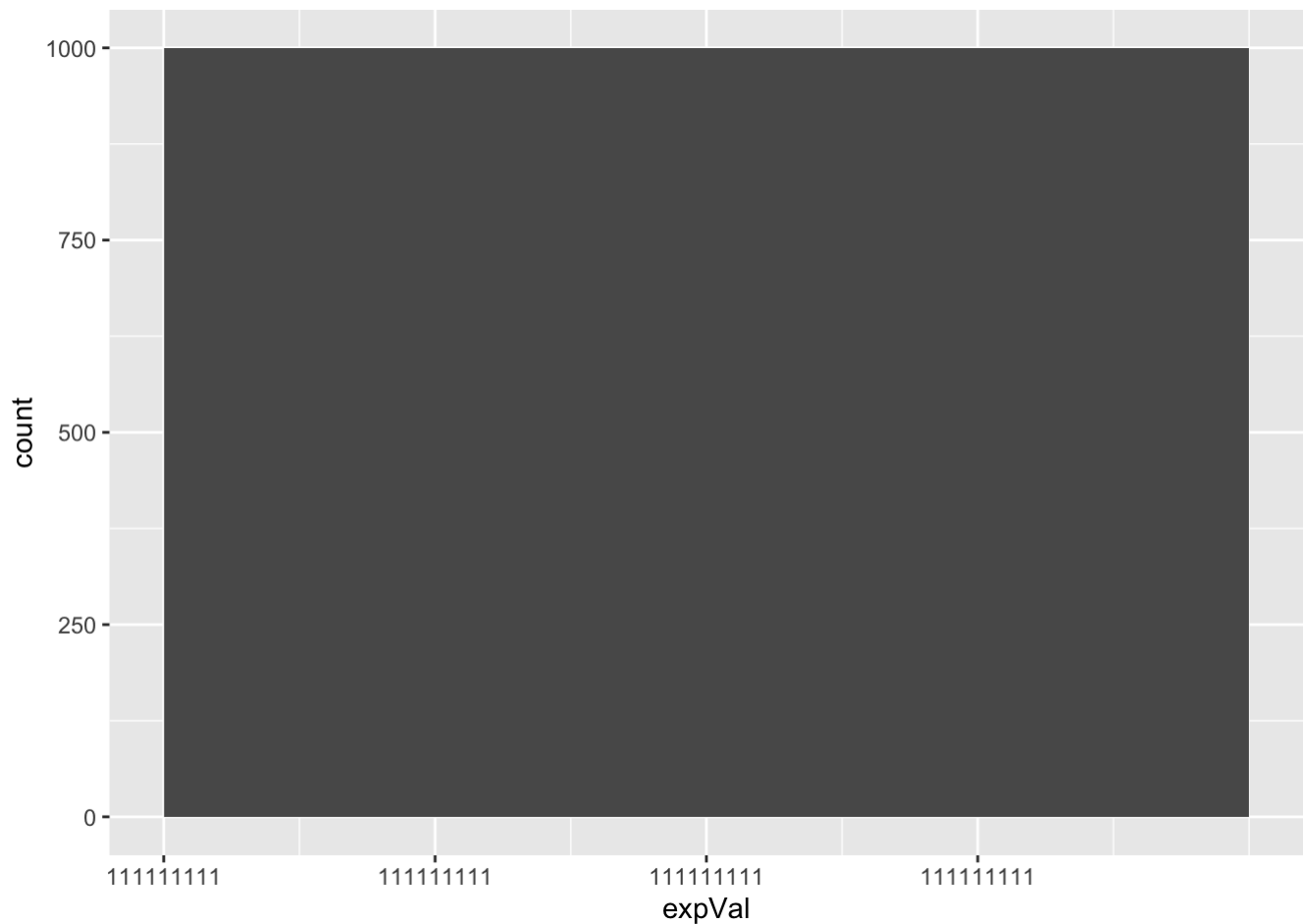
4

Find the variance of the distributions you created in the question above.

What is the difference between the variance and the standard deviation?

```
var(logSeq)
```

```
## [1] 9.851602e+16
```

```
# sd(logSeq)^2 == var(logSeq)
sd(logSeq)^2
```

```
## [1] 9.851602e+16
```
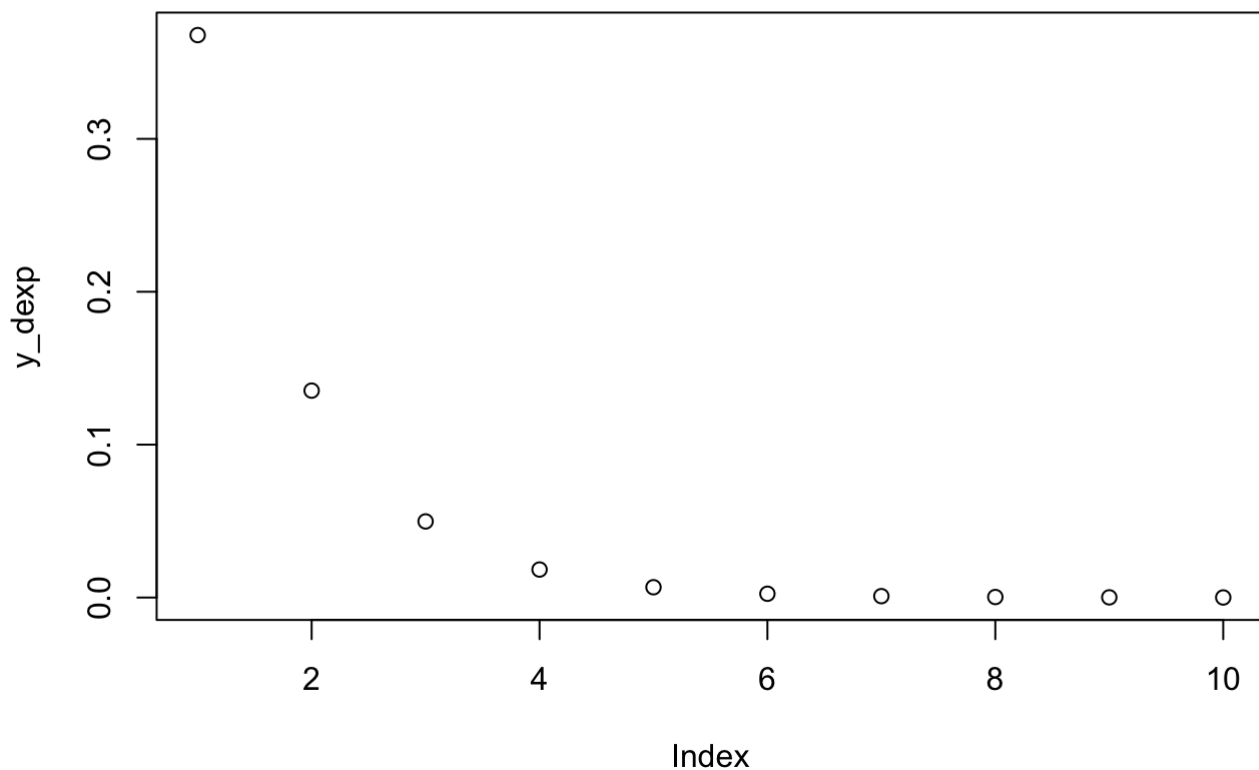
```
sd(logSeq)^2 == var(logSeq)
```

```
## [1] TRUE
```

5

Find the expected value of the exponential distribution using the dexp function and also find the expected value of 1000 random draws.

Why do you think there is a discrepancy between these two values?

```
y_dexp <- dexp(1:10)
plot(y_dexp)
```



```
lower <- floor(qexp(0.001, rate=0.2))
upper <- ceiling(qexp(0.999, rate=0.2))
t <- seq(lower,upper,0.1)
dexp02 <- data.frame(T=t, Density=dexp(t, rate=0.2))

expVal_dexp <- mean(dexp02$Density)
expVal_dexp
```

```
## [1] 0.02875016
```

```
random_draws <- c()
for(i in 1:1000){
  random_draws[i] <- sample(dexp02$Density,1)
}
expVal_1000randomDraws <- mean(random_draws)
expVal_1000randomDraws
```

```
## [1] 0.02955397
```

```
# The difference of expVal_dexp and expVal_1000randomDraws is very small.
expVal_dexp - expVal_1000randomDraws
```

```
## [1] -0.0008038154
```

```
# When we have random draws from the population, only when infinite draws can make ma
ke [the expected value of the random draws] the same as [the expected value of the di
stribution].
```

6

Why does the quantile function for the exponential distribution have an undefined expected value?

```
q <- seq(0.001,0.999,0.001)
qexp02 <- data.frame(Q=q, Quantile=qexp(q, rate=0.2))
head(dexp02)
```

```
##      T    Density
## 1 0.0  0.2000000
## 2 0.1  0.1960397
## 3 0.2  0.1921579
## 4 0.3  0.1883529
## 5 0.4  0.1846233
## 6 0.5  0.1809675
```

```
# The quantile function for the exponential distribution has an undefined expected va
lue
# because the expected value for an absolutely continuous distribution is defined as ∫
xf(x)dx where f is the density function and the integral is taken over the domain of
 f (which is -INF to INF).
# For a random variable X, if we take a draw from its distribution you know it will n
ot be not likely be equal to the average. Therefore we are not "expecting" it to be s
ome particular value. Thus, expected value is undefined.
```

7

What do we mean by saying that exponential distributions do not have a memory?

```
# The exponential distribution is memoryless because the past has no bearing on its f
uture behavior.
# Every instant is like the beginning of a new random period, which has the same dist
ribution regardless of how much time has already elapsed.
# The exponential is the only memoryless continuous random variable.
```

8

Name a few examples of when to use the exponential distribution or when it comes up in nature.

```
# The growth of the population is exponential, given infinite resources of energy and
space.
# Microorganisms in culture
# Compound interest
```

9

Create 3 random number generators in R with different seeds and provide the first ten results from each.

```
# 1st seed = 1
set.seed(1)
rnorm(10)
```

```
##  [1] -0.6264538  0.1836433 -0.8356286  1.5952808  0.3295078 -0.8204684
##  [7]  0.4874291  0.7383247  0.5757814 -0.3053884
```

```
# 2nd seed = 5
set.seed(5)
rnorm(10)
```

```
##  [1] -0.84085548  1.38435934 -1.25549186  0.07014277  1.71144087 -0.60290798
##  [7] -0.47216639 -0.63537131 -0.28577363  0.13810822
```

```
# 3rd seed = 15
set.seed(15)
rnorm(10)
```

```
##  [1]  0.2588229  1.8311207 -0.3396186  0.8971982  0.4880163 -1.2553858
##  [7]  0.0227882  1.0907732 -0.1321224 -1.0750013
```

10 BONUS

Prove how the exponential distribution is a special case of the gamma distribution?

* Gamma density

$$f(x) = \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha) \beta^{\alpha}} , \quad x > 0$$

For parameters $\alpha, \beta > 0$.

* If $\alpha = 1$

$$f(x) = \frac{1}{\beta} e^{\frac{-x}{\beta}} , \quad x > 0$$

$f(x)$ is now also the density of exponential distribution

11 BONUS

Demonstrate and prove (with a script and visual representations) of how the exponential distribution is memoryless.

```
lambda <- 1/3
avrg <- 1/lambda
std.dv <- sqrt(1/lambda^2)

range <- seq(0, avrg+5*std.dv, 0.01)
y <- dexp(range, lambda)

x <- 4
# P(X>4)
1-pexp(x+0,lambda) / 1-pexp(0,lambda)
```

```
## [1] 0.2635971
```

```
# P(X>6+4|X>6) == P(X>4)
(1-pexp(x+6,lambda)) / (1-pexp(6,lambda))
```

```
## [1] 0.2635971
```

```
# Draw the exponential distribution
# In the following two plots,
# the ratio of [ blue part / grey part ] should be the same value.

# P(X>4)
plot(range, y, type="l", ylim=c(0,max(y)+0.01))
cord.a <- c(0, seq(min(range),x,0.01),x)
cord.b <- c(0,dexp(seq(min(range),x,0.01),lambda),0)
polygon(cord.a,cord.b,col="blue")
cord.c <- c(x, seq(max(range),x,-0.01),x)
cord.d <- c(0,dexp(seq(max(range),x,-0.01),lambda),0)
polygon(cord.c,cord.d,col=grey(0.90))
```
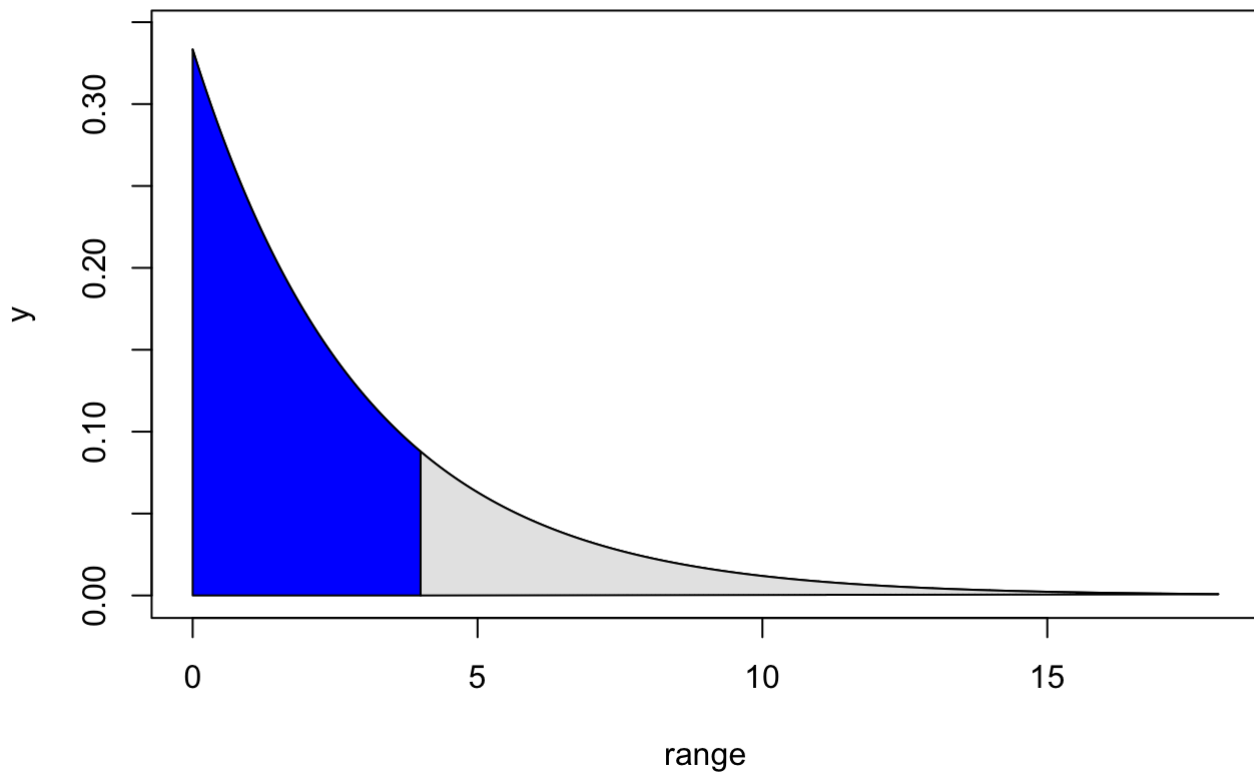
```
# P(X>6+4|X>6) == P(X>4)
plot(range, y, type="l", ylim=c(0,max(y)+0.01))
cord.e <- c(6, seq(6,6+x,0.01),6+x)
cord.f <- c(0,dexp(seq(6,6+x,0.01),lambda),0)
polygon(cord.e,cord.f,col="blue")
cord.g <- c(6+x, seq(max(range),6+x,-0.01),6+x)
cord.h <- c(0,dexp(seq(max(range),6+x,-0.01),lambda),0)
polygon(cord.g,cord.h,col=grey(0.90))
```
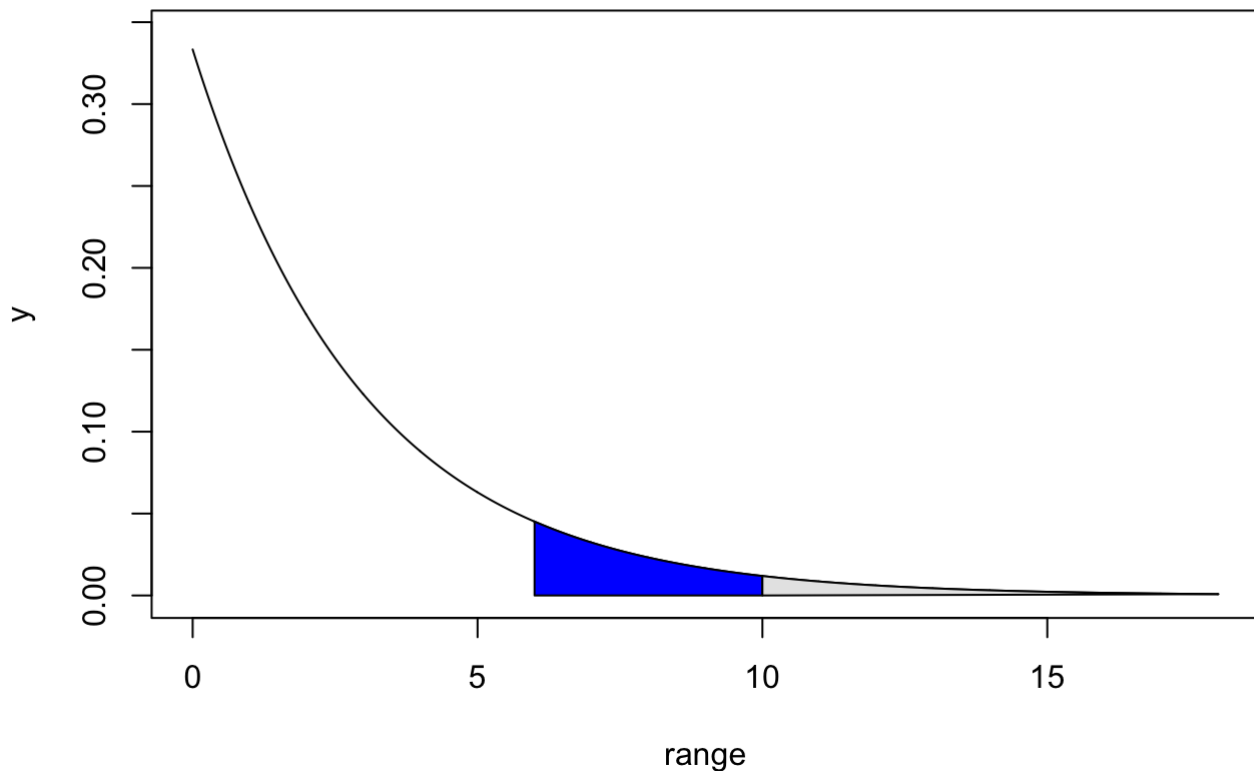
$$X \sim \exp(\lambda), \quad x > 0$$

$$\text{CDF } F(a) = P(X \leq a) = 1 - e^{-\lambda a}$$

LEFT HAND SITE

$$\frac{P(X > s+t \cap x > t)}{P(X > t)} = \frac{P(X > s+t)}{P(X > t)}$$

$$P(X > s+t) = 1 - P(X \leq s+t)$$
$$= 1 - (1 - e^{-\lambda(s+t)})$$
$$= e^{-\lambda(s+t)} \quad \text{——— (1)}$$

$$P(X > t) = e^{-\lambda t} \quad \text{——————— (2)}$$

$$\text{So } (1)/(2) = e^{-\lambda s} \quad \text{——————— (3)}$$

RIGHT HAND SIDE

$$P(X > s) = e^{-\lambda s} \quad \text{——— (4)}$$

$$\text{But } (3) = (4) \quad \#$$