Homework 2

DATA604 Simulation and Modeling

Daniel Dittenhafer

February 21, 2016

1

Suppose that X is a discrete random variable having probability function $Pr(X = k) = ck^2$ for $k = 1, 2, 3$. Find c, $Pr(X \leq 2)$, E[X] and Var(X).

This suggests the following:

$$Pr(X = 1) + Pr(X = 2) + Pr(X = 3) = 1$$
$$1^2c + 2^2c + 3^2c = 1$$
$$1c + 4c + 9c = 1$$
$$14c = 1$$
$$c = \frac{1}{14}$$

```
# Define an R function for the probability function
prXk <- function(k)
{
  c = 1/14
  p <- c * k^2

  return (p)
}

prX1 <- prXk(1)
prX1
```

```
## [1] 0.07142857
```

```
prX2 <- prXk(2)
prX2
```

```
## [1] 0.2857143
```

```
prX3 <- prXk(3)
prX3
```

```
## [1] 0.6428571
```

```
ExpVal <- prX1 + (2*prX2) + (3*prX3)

secondMoment <- prX1 + (2^2*prX2) + (3^2*prX3)
```

$Pr(X \leq 2) = 0.3571429$

The expected value $E[X] = 1 \times 0.0714286 + 2 \times 0.2857143 + 3 \times 0.6428571 = 2.5714286$.

The variance $Var(X) = E[X^2] - (E[X])^2 = 7 - 6.6122449 = 0.3877551$

2

Suppose that X is a continuous random variable having p.d.f $f(x) = cx^2$ for $1 \leq x \leq 2$. Find c, $Pr(X \geq 1)$, E[X] and Var(X).

$$\int cx^2 dx = c \int x^2 dx = c \times \frac{x^3}{3} + C$$

$$\int cx^2 dx = c \times \frac{x^3}{3} + C = 1$$

$$c = \frac{3 \times 1}{x^3} = \frac{3}{x^3}$$

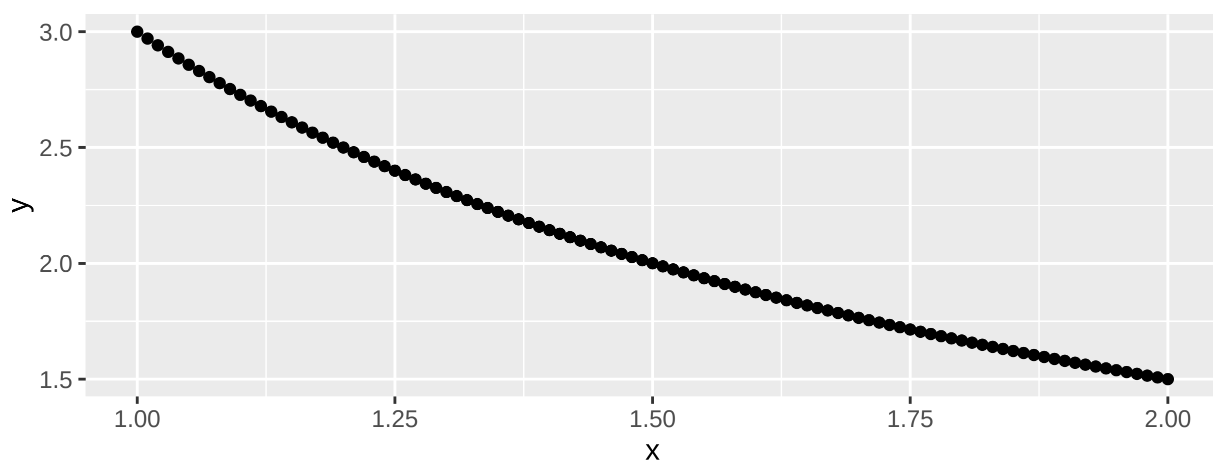
The probabilty $Pr(X \geq 1)$ will equal 1 due to the fact the range is $1 \leq x \leq 2$.

Expected value is... hmmm...

The following visualization shows the probability density function:

```
# Define an R funciton for the PDF
pr2X <- function(x)
{
  c <- 3 / x^3
  return( c * x^2)
}

r <- seq(1, 2, by=0.01)

pr2Xr <- pr2X(r)
```

```
g1 <- ggplot(data=data.frame(x=r, y=pr2Xr)) + geom_point(aes(x=x, y=y))
g1
```



3

Suppose the X and Y are jointly continuous random variables with

$$y - x \text{ for } 0 < x < 1 \text{ and } 1 < y < 2$$
$$0 \text{ otherwise}$$

Assuming $f_{X,Y}(x, y) = y - x$ is the joint density according to the above constraints:

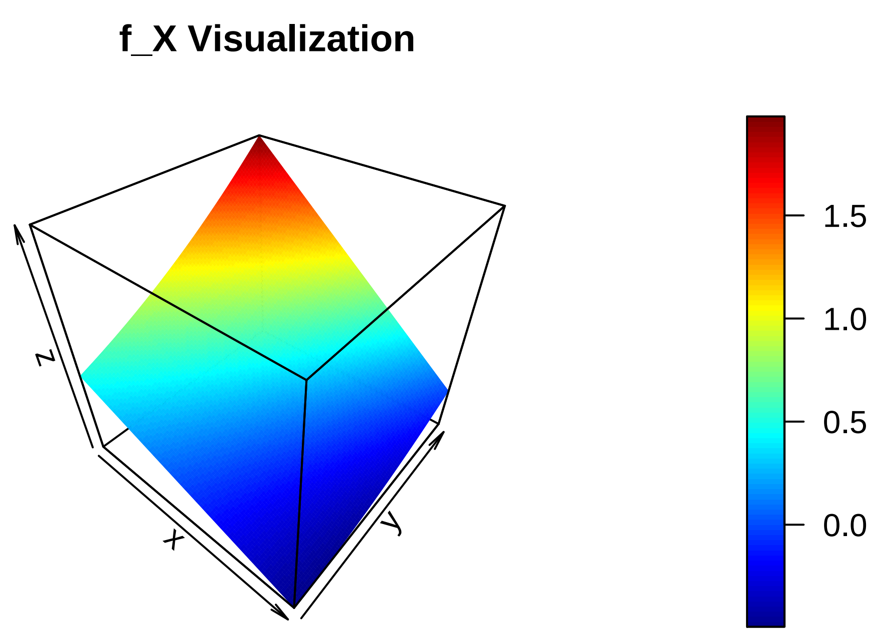
a) Compute and plot $f_X(x)$ and $f_Y(y)$

$f_X(x)$ and $f_Y(y)$ are the marginal densities, therefore

$f_X(x) = \int f_{X,Y}(x,y)dy = \frac{1}{2}y^2 - xy$

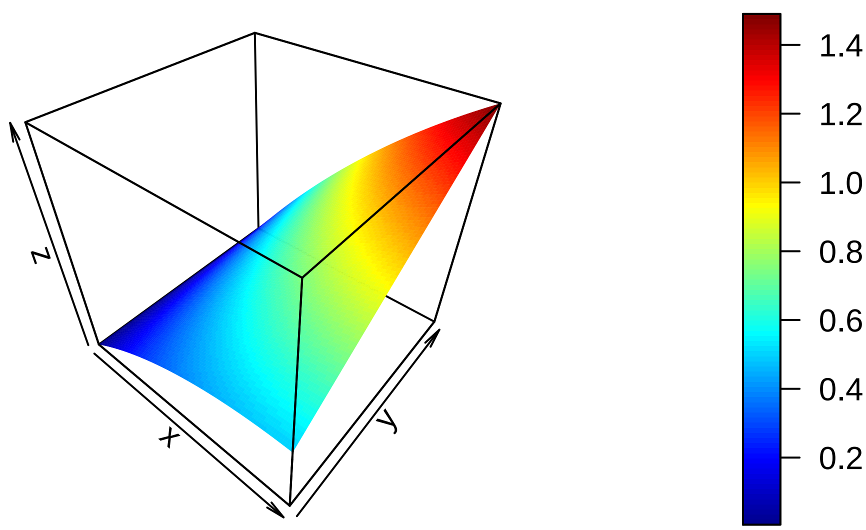
$f_Y(y) = \int f_{X,Y}(x,y)dx = xy - \frac{1}{2}x^2$

```
f_X <- function(x,y)
{
  return ((1/2 * y^2) - (x*y))
}
f_Y <- function(x,y)
{
  return ((x*y) - (1/2 * x^2) )
}

x <- seq(0, 1, by=0.01)
y <- seq(1, 2, by=0.01)
M <- mesh(x, y)
surf3D(M$x, M$y, z=f_X(M$x, M$y),  bty = "f", main="f_X Visualization")
```




```
x <- seq(0, 1, by=0.01)
y <- seq(1, 2, by=0.01)
M <- mesh(x, y)
surf3D(M$x, M$y, z=f_Y(M$x, M$y),  bty = "f", main="f_Y Visualization")
```

f_Y Visualization



b) Are X and Y independent?

if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, we say X, Y are independent.

$f_X(x)f_Y(y) = (\frac{1}{2}y^2 - xy)(xy - \frac{1}{2}x^2) = \frac{1}{2}xy^3 - \frac{5}{4}x^2y^2 + \frac{1}{2}x^3y$

The above is not equal to $y - x$, therefore X and Y are not independent.

c) Compute $F_X(x)$ and $F_Y(y)$

$F_X(x) = \int f_X(x)dx = \int \frac{1}{2}y^2 - xydx = \int \frac{1}{2}y^2dx - \int xydx = \frac{1}{2}xy^2 - \frac{x^2y}{2}$

$F_Y(y) = \int f_Y(y)dy = \int xy - \frac{1}{2}x^2dy = \int xydy - \int \frac{1}{2}x^2dy = \frac{xy^2}{2} - \frac{1}{2}yx^2$

d) Compute E[X], Var(X), E[Y], Var(Y), Cov(X,Y), and Corr(X,Y)

...

4

Suppose that the following 10 obserations come from some distribution (not highly skewed) with unknown mean μ.

```
d <- c(7.3,6.1,3.8,8.4,6.9,7.1,5.3,8.2,4.9,5.8)
kable(t(d))
```

7.3	6.1	3.8	8.4	6.9	7.1	5.3	8.2	4.9	5.8

Compute $\bar{X}, S^2$ and an approximate 95% confidence interval for μ.

```
Xbar <- mean(d)
varD <- var(d)
sdD <- sd(d)
n <- length(d)
seD <- sdD / sqrt(n)
tval <- qt(.975, df=n-1)
```

$\bar{X} = 6.38$

$S^2 = 2.1617778$

$SE = 0.4649492$

t-value $= 2.2621572$

95% CI of $\mu = \bar{X} \pm t \times SE = (5.3282118, 7.4317882)$

5

A random variable X has the memoryless property if, for all s, t > 0,

$$Pr(X > t + s | X > t) = Pr(X > s)$$

Show that the exponential distribution has the memoryless property.

Page 193 of the DES text has an excellent proof of the the memoryless property:

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

$$P(X > t + s | X > t) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

6

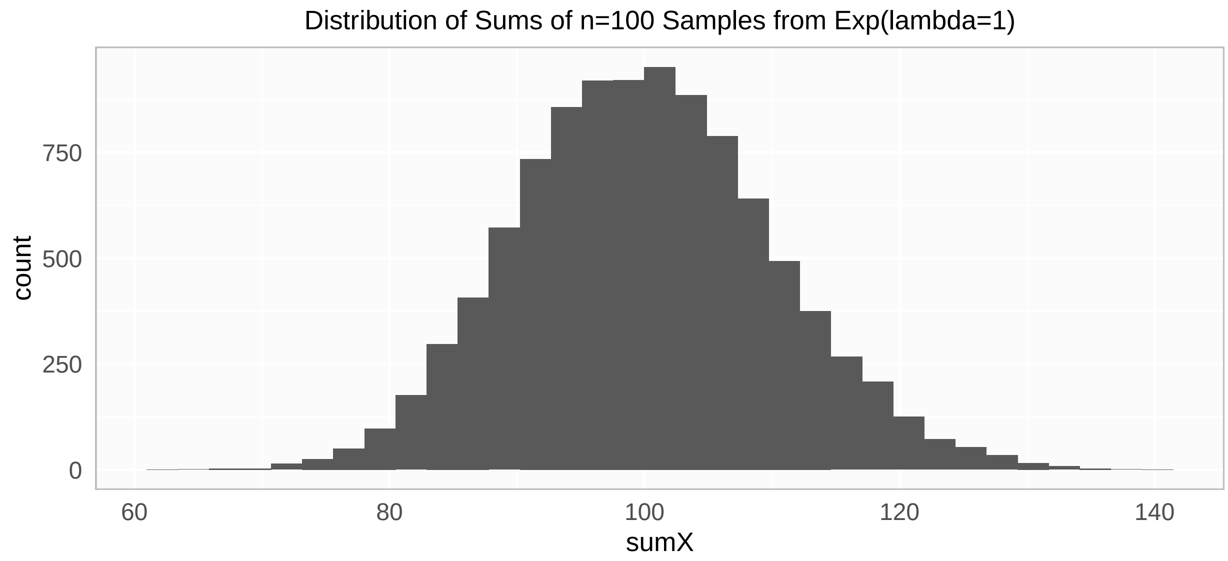
Suppose $X_1, X_2, ..., X_n$ are i.i.d. Exp($\lambda = 1$). Use the Central Limit Theorem to find the approximate value of $Pr(100 \leq \Sigma_{i=1}^{100} X_i \leq 110$

We can take many n=100 samples from the Exp($lambda = 1$) distribution, and then sum them to create a distribution of the sums. Finally, use the number of sum samples that fall in the designated range divided by the total number of the n=100 samples to find an approximate probability.

```
sumX <- c()
for(j in 1:10000)
{
 X <- rexp(100)
 sX <- sum(X)
 sumX <- c(sumX, sX)
}
```

```
g1 <- ggplot(data.frame(sumX=sumX)) +
  geom_histogram(aes(x=sumX)) +
  labs(title="Distribution of Sums of n=100 Samples from Exp(lambda=1)") +
  myTheme
g1
```

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




```
n <- length(sumX)
valInRange <- length(sumX[100 <= sumX & sumX <= 110])

prInRange <- valInRange / n
```

The approximate value of $Pr(100 \leq \Sigma_{i=1}^{100} X_i \leq 110) = 0.3308$

5.13

A random variable X that has pmf given by $p(x) = 1/(n+1)$ over the range $R_X = (0, 1, 2, ..., n)$ is said to have a discrete uniform distribution.

(a) Find the mean and variance of this distribution.

The mean would be $\Sigma_{alli} ip(x)$. In this case $p(x)$ is uniform at $\frac{1}{n+1}$. As a result, this can be factored out of the mean equation resulting in $p(x)\Sigma_{alli} i$. The hint told us that $\Sigma_{alli} i = \frac{n(n+1)}{2}$, therefore the mean would be:

$$\text{mean} \quad = \frac{1}{n+1} \times \frac{n(n+1)}{2} = \frac{n}{2}$$

The variance works the same way for discrete distributions, therefore:

$$\text{variance} \quad = (\frac{1}{n+1} \times \frac{n(n+1)(2n+1)}{6}) - (\frac{n}{2})^2 = \frac{n(2n+1)}{6} - \frac{n^2}{4}$$

$$= \frac{2n(2n+1)}{12} - \frac{3n^2}{12} = \frac{2n(2n+1) - 3n^2}{12} = \frac{n^2 + 2n}{12} = \frac{n(n+2)}{12}$$

(b) If $R_X = (a, a+1, a+2, ..., b)$, compute the mean and variance of X.

Assuming n=b-a and therefore p(x)=1/(n+1) would still apply as in $R_X = (2, 3, 4, 5), a = 2, b = 5, n = 3$ and $p(x) = 1/(3+1) = 1/4$ to retain the uniform distribution. Also refering to http://mathforum.org/library/drmath/view/57166.html, the sum of R_X would be $k(2a + k - 1)/2$ with k=# of items=n+1.

$$\text{mean} \quad = \frac{1}{n+1} \times \frac{n(2a + n - 1)}{2} = \frac{n(2a + n - 1)}{2(n+1)}$$

Variance, hmm. . . .

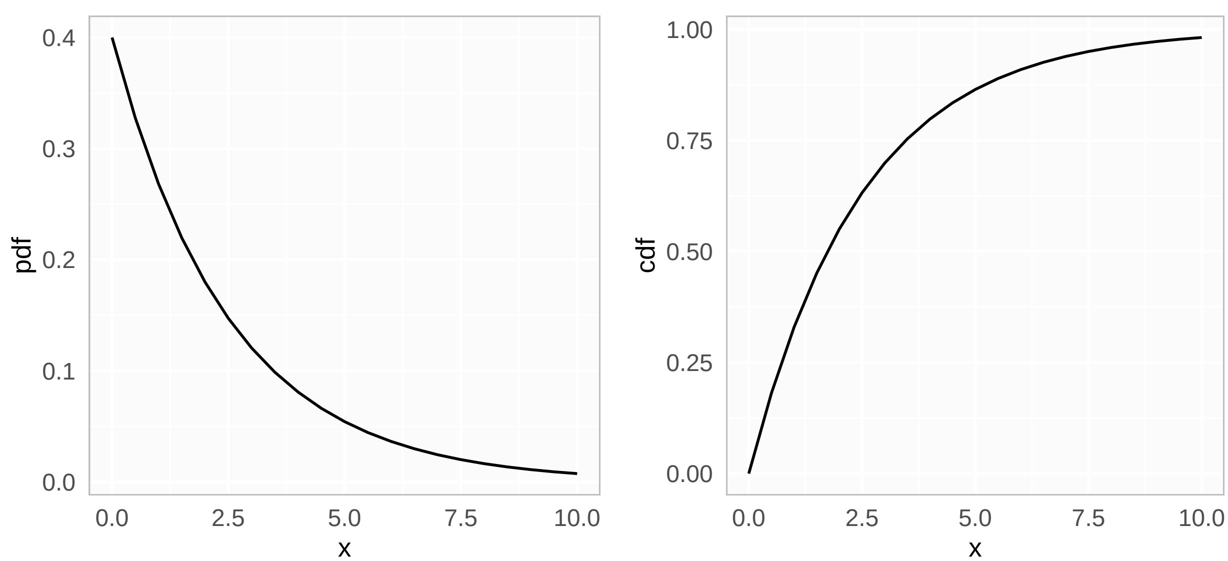
5.14

The lifetime in years of a satellite placed in orbit is given by the following pdf:

$$f(x) = 0.4e^{-0.4x}, x \geq 0$$

(a) What is the probability that this satellite is still "alive" after 5 years?

```
# Define CDF and PDF functions
cdf514 <- function(x)
{

  p <- 1 - exp(-0.4 * x)

  return (p)
}

pdf514 <- function(x)
{

  p <- 0.4 * exp(-0.4 * x)

  return (p)
}

# prepare data set to help us visualize
rx <- data.frame(x=seq(0, 10, by=0.5))
rx$cdf <- cdf514(rx$x)
rx$pdf <- pdf514(rx$x)
# Visualize
g1 <- ggplot(rx) + geom_line(aes(x=x, y=pdf)) + myTheme
g2 <- ggplot(rx) + geom_line(aes(x=x, y=cdf)) + myTheme
grid.arrange(g1, g2, ncol=2)
```




```
p5 <- pdf514(5)
```

The probability that this satellite is still alive after 5 years is 0.0541341.

(b) What is the probability that the satellite dies between 3 and 6 years from the time it is placed in orbit?

```r
# Compute probabilty of satelite's life time is between 3-6 years (i.e. it dies in this range)
p36 <- cdf514(6) - cdf514(3)
```

The probabilty that this satellite dies between 3 and 6 years is 0.2104763.

5.39

(a) Distribution of the length of the linkage?

Normally distributed because the shafts are linked, presumably in a linear manner, to one another. A normal + normal + normal = normal distribution

(b) What is the probability that the linkage will be longer than 150.2 cm

```r
m1 <- 60
s1 <- 0.09

m2 <- 40
s2 <- 0.05

m3 <- 50
s3 <- 0.11

link.m <- m1 + m2 + m3
link.v <- s1^2 + s2^2 + s3^2
link.sd <- sqrt(link.v)
```

The mean of the linkage will be 150 and the standard deviation will be 0.1506652.

```r
z <- (150.2 - 150) / (link.sd)
z
```

```
## [1] 1.327447
```

```r
p <- pnorm(150.2, mean=link.m, sd=link.sd, lower.tail=FALSE)
p
```

```
## [1] 0.09218049
```

```r
pz <- pnorm(z, lower.tail=FALSE)
pz
```

```
## [1] 0.09218049
```

The probabilty that the linkage will be longer than 150.2 cm is 0.0921805.

(c) What proportion of assemblies will be with the tolerance limits of 149.83 to 150.21?

```r
pl <- pnorm(149.83, mean=link.m, sd=link.sd)
pu <- pnorm(150.21, mean=link.m, sd=link.sd)

propInLimits <- pu - pl
propInLimits
```

```
## [1] 0.7887235
```

78.8723% of assemblies will be within the tolerance limits.

Useful Links

Jointly continuous random variables: http://math.arizona.edu/~tgk/464_10/chap6_10_18.pdf

Linear combinations of normal random variables: http://www.statlect.com/probability-distributions/normal-distribution-linear-combina