IS 604 HW 2

Charley Ferrari

Monday, September 14, 2015

Question 1

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

We know k can only be 1,2,3. And, we know the total probability of the random variable must add up to 1. So, we just need to add up all the probabilities for 1, 2, and 3 and set them equal to 1 to find out c:

$$c * (1)^2 + c * (2)^2 + c * (3)^2 = 1$$
$$14c = 1$$
$$c = \frac{1}{14}$$

$$Pr(X \leq 2) = Pr(X = 1) + Pr(X = 2)$$
$$Pr(X \leq 2) = \frac{1}{14} * 1^2 + \frac{1}{14} * 2^2$$
$$Pr(X \leq 2) = \frac{1}{14} + \frac{4}{14} = \frac{5}{14}$$

This Means

$$Pr(X = 3) = \frac{9}{14}$$

$$E[X] = 1 * Pr(X = 1) + 2 * Pr(X = 2) + 3 * Pr(X = 3)$$
$$E[X] = \frac{1}{14} + \frac{8}{14} + \frac{27}{14} = \frac{36}{14}$$

$$Var(X) = \Sigma Pr(X = k)(x - \mu)^2$$
$$Var(X) = \frac{1}{14}(1 - \frac{36}{14})^2 + \frac{4}{14}(2 - \frac{36}{14})^2 + \frac{9}{14}(3 - \frac{36}{14})^2$$

```r
mu <- 36/14
p1 <- 1/14
p2 <- 4/14
p3 <- 9/14

varx <- p1*(1-mu)^2 + p2*(2-mu)^2 + p3*(3-mu)^2

varx
```

```
## [1] 0.3877551
```

Question 2

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

The area under the curve in this region needs to be 1. So:

$$\int_1^2 cx^2 = 1$$

$$\int cx^2 dx = \frac{1}{3}cx^3$$

$$\frac{1}{3}c(2)^3 - \frac{1}{3}c(1)^3 = 1$$

$$\frac{7}{3}c = 1$$

$$c = \frac{3}{7}$$

For $Pr(X \geq 1)$:

X is defined as $1 \leq x \leq 2$. Therefore, $Pr(X \geq 1)$ must be 1.

For E[X]:

$$E[X] = \int_1^2 xf(x)dx = \int_1^2 \frac{3}{7}x^3$$

$$\int \frac{3}{7}x^3 dx = \frac{3}{28}x^4$$

```
Ex <- (3/28)*2^4 - (3/28)*1^4
Ex
```

```
## [1] 1.607143
```

For Var(X):

$$Var(X) = \int_1^2 x^2 f(x)dx - \mu_2$$

$$E(X^2) = \int_1^2 cx^4 dx$$

$$c\int x^4 dx = c\frac{1}{5}x^5$$

```
c <- 3/7

Ex2 <- c * (1/5) * 2^5 - c^2 * (1/5) * 1^5
Ex2
```

```
## [1] 2.706122
```

```
Varx <- Ex - Ex2

Varx
```

```
## [1] -1.09898
```

Question 3

Suppose that X and Y are jointly continuous random variables with

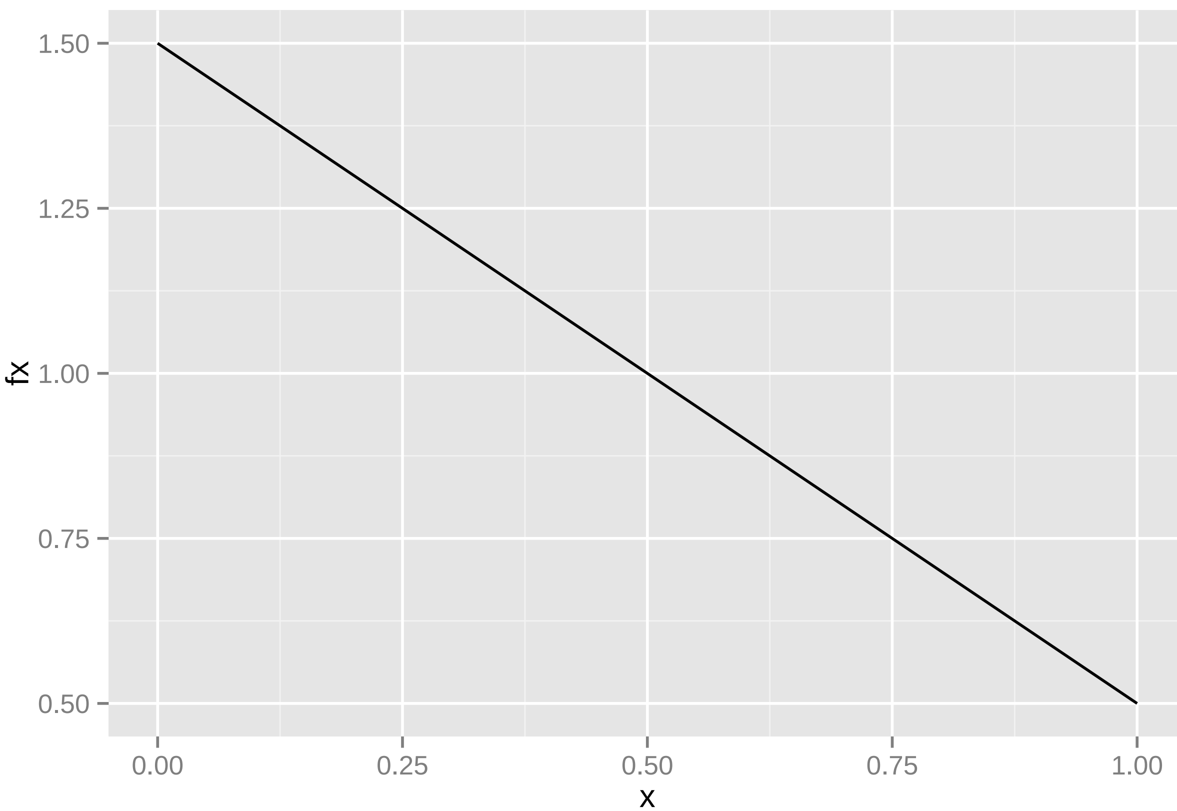
$$f_{xy}(x,y) = y - x for 0 < x < 1 and 1 < y < 2$$

0 otherwise

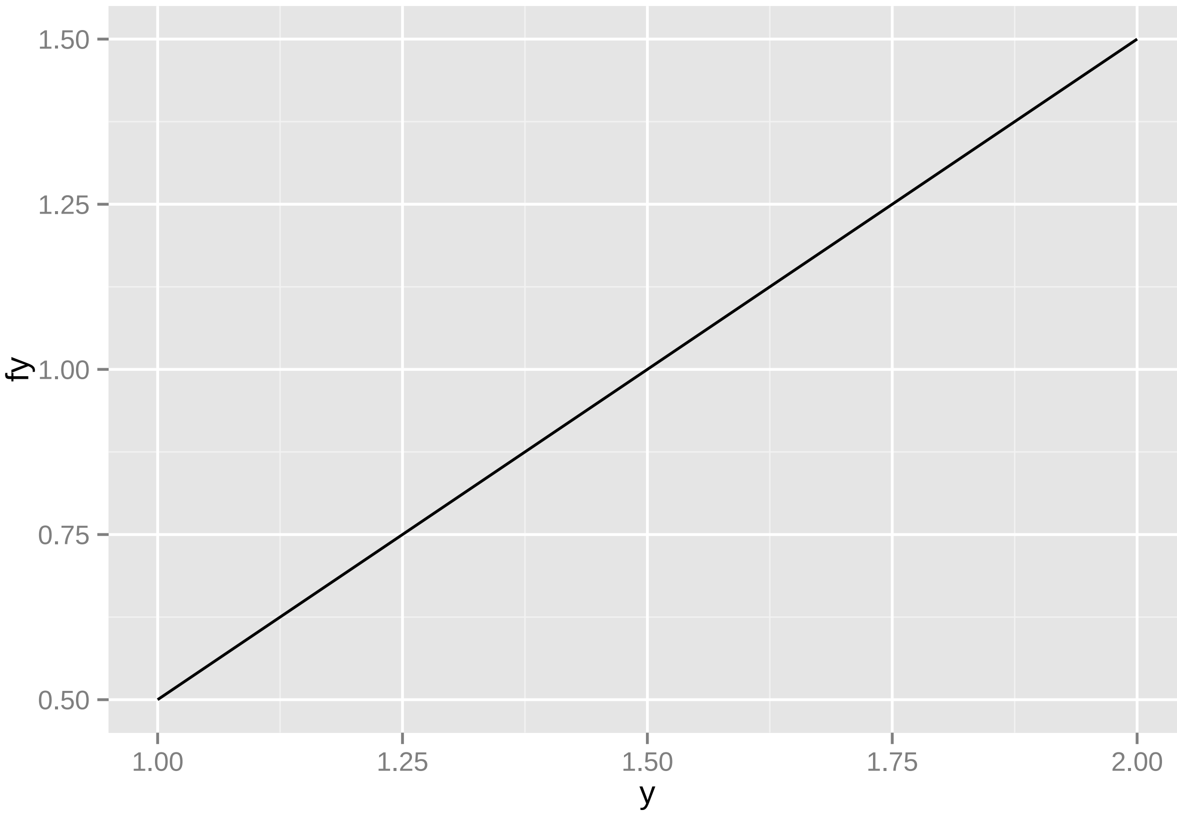
a. Compute and plot $f_x(x)$ and $f_y(y)$.

$$f_x(x) = \int_1^2 (y-x)dy = \frac{3}{2} - x \qquad\qquad 0 < x < 1$$
$$f_y(y) = \int_0^1 (y-x)dx = y - \frac{1}{2} \qquad\qquad 1 < y < 2$$

```
library(ggplot2)

x <- seq(0,1,by=0.1)
fx <- (3/2) - x

xfx <- data.frame(x=x, fx=fx)

ggplot(xfx, aes(x=x, y=fx)) + geom_line()
```



```r
y <- seq(1,2,by=0.1)
fy <- y - (1/2)

yfy <- data.frame(y=y, fy=fy)

ggplot(yfy, aes(x=y, y=fy)) + geom_line()
```



Are X and Y independent?

These are independent iff $f(x,y) = f(x)f(y)$

$$f(x)f(y) = (\frac{3}{2} - x)(y - \frac{1}{2}) = \frac{3}{2}y - xy - \frac{3}{4} + \frac{1}{2}x$$

These are not equal, so X and Y are not independent.

Compute $F_x(x)$ and $F_y(y)$

$$F_x(x) = \int_{-\infty}^{x} f_x(t)dt = \frac{3}{2}x - \frac{1}{2}x^2$$

$$F_y(y) = \int_{-\infty}^{y} f_y(t)dt = \frac{1}{2}y^2 - \frac{1}{2}y$$

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

$$E[X] = \int_0^1 x f_x(x)dx = \int_0^1 \frac{3}{2}x - x^2 = \frac{3}{4}x^2 - \frac{1}{3}x^3$$
$$\frac{3}{4}(1)^2 - \frac{3}{4}(1)^3 - f_x(0)$$

```
Ex <- (3/4) - (1/3)

Ex
```

```
## [1] 0.4166667
```

$$E[Y] = \int_1^2 y f_y(y)dy = \int_1^2 y^2 - \frac{1}{2}y = \frac{1}{3}y^3 - \frac{1}{4}y^2$$

```
Ey <- (1/3)*(2^3) - (1/4)*(2^2) - (1/3) + (1/4)

Ey
```

```
## [1] 1.583333
```

$$Var(X) = \int_0^1 x^2 f(x)dx - \mu^2$$
$$\int x^2(\frac{3}{2} - x) = \frac{3}{2}x^2 - x^3 = \frac{1}{2}x^3 - \frac{1}{4}x^4$$

From 0 to 1 is just calculatig for 1:

```
Varx <- (1/4) - Ex^2
Varx
```

```
## [1] 0.07638889
```

$$Var(Y) = \int_0^1 y^2 f(y)dy - \mu_2$$
$$\int y^2(y - \frac{1}{2})dy = \int y^3 - \frac{1}{2}y^2 = \frac{1}{4}y^4 - \frac{1}{6}y^3$$

```
Ey2 <- (1/4)*(2^4) - (1/6)*(2^3) - (1/4) + (1/6)
Vary <- Ey2 - Ey^2

Vary
```

```
## [1] 0.07638889
```

(copying my work from the test for independence for the Covariance)

$$Cov(X,Y) = E[XY] - E[X]E[X]$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{xy}(x,y)dxdy$$
$$E[XY] = \int_1^2\int_0^1 xy(y-x)dxdy = \int_1^2\int_0^1 xy^2 - x^2ydxdy$$
$$E[XY] = \int_1^2 \frac{x^2y^2}{2} - \frac{x^3y}{3}|_0^1 dy$$
$$E[XY] = \int_1^2 \frac{y^2}{2} - \frac{y}{3}dy = \frac{y^3}{6} - \frac{y^2}{6}|_1^2 = \frac{2}{3}$$

Question 4

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

```
sampl <- c(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 <- sum(sampl)/length(sampl)

S2 <- sum((sampl-xbar)^2)/(length(sampl))

xbar
```

```
## [1] 6.38
```

```
S2
```

```
## [1] 1.9456
```

We're trying to find a 95% confidence interval for μ such that we're 95% sure the population mean is within that interval. We'll use a t distribution, and try to make sure our mean is within 2.5% on either side. This corresponds to a t statistic of 2.16.

$CI = \bar{X} \pm t\frac{S}{\sqrt{n}}$

```
tplusmin <- 2.16*(sqrt(S2)/sqrt(length(sampl)))
```

So μ is within ~0.95 of 6.38.

Question 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.

The exponential distribution has the pdf: $f(x) = \lambda e^{-\lambda x}$

$$Pr(x > s) = e^{-s\lambda}$$
$$Pr(X > t) = e^{-t\lambda}$$
$$Pr(x > t + s) = \frac{Pr(x > t + s, x > t)}{Pr(x > t)} = e^{-s\lambda - t\lambda}$$

$Pr(x > t + s, x > t)$ is the same as $Pr(x > t + s)$, since if x is greater than t+s, it must be also greater than t.

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

Question 6

pdf, $\lambda = 1$

$$f(x) = e^{-x}$$
$$Pr(x < t) = e^{-t}$$
$$Pr(1 < x < 1.1) = Pr(x > 1) - Pr(X > 1.1) = e^{-1} - e^{-1.1}$$

```
aprox <- exp(-1) - exp(-1.1)
aprox
```

```
## [1] 0.03500836
```

The central limit theorem states that if we take a sum 100 times, and work out the probabilities from 100x what I worked out above, $\bar{x}$ should remain the same. The difference will be in the confidence intervals, and variances.

Question 5.13

A random variable X that has a pmf given by $p(x) = \frac{1}{n+1}$ 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.

$$\mu = \sum_{i=1}^{n} xP(x) = \frac{1}{n+1}\frac{n(n+1)}{2} = \frac{n}{2}$$
$$Var(X) = E[X^2] - (E[X])^2$$
$$E[X^2] = \sum_{i=1}^{n} x^2 P(x) = \frac{1}{n+1}\frac{n(n+1)(2n+1)}{6} = \frac{n(2n+1)}{6}$$

b. In this case, we're trying to find the sum from a to b. I'll just subtract the sum from 0 to a-1 from the sum from 0 to b:

$$\sum_{i=a}^{b} i = \frac{b(b+1)}{2} - \frac{a(a-1)}{2}$$
$$\mu = \sum xP(x) = \frac{1}{b-a+1}(\frac{b(b+1)}{2} - \frac{a(a-1)}{2}) = \frac{b^2+b-a^2-a}{2b-2a+2}$$

For some reason I just could not get the math to work out cleanly, and see this will only grow more complicated as I expand this out (I've tried working this out on paper and just got nowhere…) I must have made an arithmitic mistake somewhere but can't find it.

Question 5.14

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

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

a. What is the probability that the satellite is still "alive" after 5 years?

$$Pr(X > 5) = e^{-5\lambda} = e^{-5*0.4}$$

```
palive <- exp(-5*0.4)

palive
```

```
## [1] 0.1353353
```

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

$$Pr(5 > X > 3) = Pr(X > 3) - Pr(X > 5) = e^{-3*0.4} - e^{-5*0.4}$$

```
palive <- exp(-3*0.4) - exp(-5*0.4)

palive
```

```
## [1] 0.1658589
```

Question 5.39

Three shafts are made and assembled into a linkage. The length of each shaft, in centimeters, is distributed as follows:

Shaft 1: N(60,0.09) Shaft 2: N(40,0.05) Shaft 3: N(50,0.11)

a. What is the distribution of the length of the linkage?

When adding normally distributed random variables, you can just add the means and variances.

The three shafts combined will have a distribution of N(150,0.25) in centimeters.

b. What is the probability that the linkage will be longer than 150.2 cm?

$$Z = \frac{X - E[X]}{\sigma(X)}$$

```
z <- (150.2-150)/sqrt(0.25)

# I'll call that 0.4 when I look that up.
```

Looking this up on a left tail z table (the table I was using gave the area from -0.4 to $-infty$, which is the same as what I'd be looking for on the positive side in this case) gave me a probability of 0.3446.

c. The tolerance limits for the assembly are (149.83, 150.21). What proportion of assemblies are within the tolerance limits?

Lets convert these two into z-scores:

```
z1 <- (149.83-150)/sqrt(0.25)
# Lets call that -0.34

z2 <- (150.21-150)/sqrt(0.25)
# Lets call that 0.42
```

The z-score table I'm looking at gives scores to $-\infty$ for both tails, so I'll just get the area from 0.42, and subtract the area from -0.34.

$$p_{z=0.42} = 0.6628$$
$$p_{z=-0.34} = 0.3669$$

```
p <- 0.6628-0.3669
p
```

```
## [1] 0.2959
```