Unit 5 Homework: Joint Distributions

Tako Hisada

10/7/2017

1. Unladen Swallows

```
library(MASS)
set.seed(898)
```

We know SD of Wingspan is 4. Since $V(X) = \sqrt{(\sigma_X^2)}$, $V(W) = 4^2 = 16$.

$$V(V) = V[h(W)] = V[\frac{1}{2}w + U] = (\frac{1}{2})^2 * \sigma_W^2 = \frac{1}{4} * 4^2 = 4$$

We know μ_W is 10. From this, we can compute $E(h(W))$:

$$E(V) = E[h(W)] = E[\frac{1}{2}w + U] = \frac{1}{2} * 10 + U = 5 + U$$

Now we want to compute $Cov(W, V)$.

$$Cov(W, V) = E(WV) - \mu_W * \mu_V$$

$$E(WV) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} wv f_{W,V}(w, v) dw dv = \int_{-\infty}^{\infty} w f_W(w) dw \int_{-\infty}^{\infty} v f_V(v) dv$$

```
# Wingspan
w = rnorm(100, mean = 10, sd = 4)
mu_w = 10
sd_w = 4
v_w = sd_w^2

# Velocity
v = 0.5 * w + rnorm(100)

#Sigma = matrix(c(16, 8, 8, 5), 2, 2)
#mvnorm()
```

2. Broken Rulers

3. Great Time to Watch Async

a. Total expected waiting time when you take the Caltrain each morning and each evening for 5 days in a row

```r
i <- 0:5
ex = sum(1/6*i)
j <- 0:10
ey = sum(1/11*j)
exy = ex + ey
```

$$E(X+Y) = E(X) + E(Y)$$
$$E(X) = \sum_{x=0}^{5} x * p(x) = 0 * \frac{1}{6} + 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} = 2.5$$
$$E(Y) = \sum_{x=0}^{10} y * p(y) = 0 * \frac{1}{10} + 1 * \frac{1}{10} + 2 * \frac{1}{10} + 3 * \frac{1}{10} + 4 * \frac{1}{10} + 5 * \frac{1}{10} + 6 * \frac{1}{10} + 7 * \frac{1}{10} + 8 * \frac{1}{10} + 9 * \frac{1}{10} + 10 * \frac{1}{10} = 5$$
$$E(X+Y) = E(X) + E(Y) = 2.5 + 5 = 7.5$$

Since we are being asked for the total expected waiting time when we took the Caltrain for 5 days in a row, we need to multiply this value by 5.

```r
q3a = exy*5
```

$$E(X+Y) = 7.5 * 5 = 37.5$$

37.5 minutes

b. Variance of the total waiting time

$$V(X+Y) = V(X) + V(Y) + 2 * COV(X+Y)$$
$$V(X) = E(X^2) - E(X)^2$$
$$V(Y) = E(Y^2) - E(Y)^2$$

```r
compExUniform <- function(x) {
  sum = 0
  for(i in 0:x) {
    sum = sum + i * 1/(x+1)
  }
  sum
}

ex = compExUniform(5)
ey = compExUniform(10)

compExUniformSq <- function(x) {
  sum = 0
  for(i in 0:x) {
    sum = sum + i^2 * 1/(x+1)
  }
  sum
}
exx_sq <- compExUniformSq(5)
exy_sq <- compExUniformSq(10)
```

```
varx <- exx_sq - ex^2
vary <- exy_sq - ey^2
```

$$COV(X,Y) = E(XY) - \mu_X \mu_Y$$

```
compExXYUniform <- function(x, mux, y, muy) {
  sum = 0
  for(i in 0:x) {
    for(j in 0:y) {
      sum = sum + ((i-mux)*(j-muy))/((x+1)*(y+1))
    }
    #print(sprintf("END SUM: %f", sum))
  }
  sum
}
exxy <- compExXYUniform(5, 2.5, 10, 5)

cov_xy = exxy - ex*ey
```

Now we can compute $V(XY)$

```
varxy = varx + vary - 2*cov_xy
```

$$37\frac{11}{12}$$

However we need to multiple this by 5 as we are interested in finding out the variance of the total waiting time

```
q2b = 5^2 * varxy
```

$$947\frac{11}{12}$$

c. Expected value of the difference between the total evening waiting time and the total morning waiting time over all 5 days

$$E(X - Y)$$

```
i <- 0:5
ex = sum(1/6*i)
j <- 0:10
ey = sum(1/11*j)
exy_diff = abs(ex - ey)
```

$$E(X - Y) = E(X) - E(Y)$$

$$E(X) = \sum_{x=0}^{5} x * p(x) = 0 * \frac{1}{6} + 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} = 2.5$$

$$E(Y) = \sum_{x=0}^{10} y * p(y) = 0 * \frac{1}{10} + 1 * \frac{1}{10} + 2 * \frac{1}{10} + 3 * \frac{1}{10} + 4 * \frac{1}{10} + 5 * \frac{1}{10} + 6 * \frac{1}{10} + 7 * \frac{1}{10} + 8 * \frac{1}{10} + 9 * \frac{1}{10} + 10 * \frac{1}{10} = 5$$

$$E(X-Y) = |E(X) - E(Y)| = 2.5$$

Since we are being asked for the total expected waiting time when we took the Caltrain for 5 days in a row, we need to multiply this value by 5.

```
q3c = exy_diff*5
```

$$E(X+Y) = 2.5 * 5 = 12.5$$

12.5 minutes

d. Variance of the difference between the total evening time and the total morning waiting time over all 5 days

$$V(X-Y) = V(X) + V(Y)$$
$$V(X) = E(X^2) - E(X)^2$$
$$V(Y) = E(Y^2) - E(Y)^2$$

```
compExUniform <- function(x) {
  sum = 0
  for(i in 0:x) {
    sum = sum + i * 1/(x+1)
  }
  sum
}

ex = compExUniform(5)
ey = compExUniform(10)

compExUniformSq <- function(x) {
  sum = 0
  for(i in 0:x) {
    sum = sum + i^2 * 1/(x+1)
  }
  sum
}
exx_sq <- compExUniformSq(5)
exy_sq <- compExUniformSq(10)

varx <- exx_sq - ex^2
vary <- exy_sq - ey^2
```

Now we can compute $V(XY)$

```
varxy = varx + vary
```

$$12\frac{11}{12}$$

However we need to multiple this by 5 as we are interested in finding out the variance of the total waiting time

```r
q3d = 5^2 * varxy
```

$$322\frac{11}{12}$$

4. Maximizing Correlation

Show that if $Y = aX + b$ where X and Y are random variables and a != 0, corr(X, Y) = -1 or +1

$$Corr(X,Y) = P_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sigma_X * \sigma_Y}$$
$$\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt(E((X-\mu_X)^2) * E((Y-\mu_Y)^2))} = \frac{(E[(X-\mu_X)(Y-\mu_Y)])^2}{E((X-\mu_X)^2) * E((Y-\mu_Y)^2)}$$
$$\frac{(\sum_X \sum_Y (X-\mu_X)(Y-\mu_Y)p(x,y))^2}{\sum(X-\mu_X)^2 * p(x)) * \sum(Y-\mu_Y)^2 * p(y))}$$

```r
i <- 0:10
j <- 0:5

compSum <- function(x, y) {
  sum_x = sum(1/length(x)*x)
  print(sprintf("SUM X: %d", sum_x))
  sum_y = 0
  for(idx in 1:length(y)) {
    sum_y = sum_y + (sum(x) + y[idx] * length(x))
    print(sprintf("Y IDX: %d", y[idx]))
    print(sprintf("SUM Y: %d", sum_y))
  }
  sum_y
}
e_xysq <- compSum(i, j)/60
```

```
## [1] "SUM X: 5"
## [1] "Y IDX: 0"
## [1] "SUM Y: 55"
## [1] "Y IDX: 1"
## [1] "SUM Y: 121"
## [1] "Y IDX: 2"
## [1] "SUM Y: 198"
## [1] "Y IDX: 3"
## [1] "SUM Y: 286"
## [1] "Y IDX: 4"
## [1] "SUM Y: 385"
## [1] "Y IDX: 5"
## [1] "SUM Y: 495"
```