Joint Distributions, Independence, Conditional Expectations and Further Topics

IE231 - Lecture Notes 11

May 2, 2017

Joint Distributions

So far, we learned about joint probabilities in Bayesian context such as $P(A|B) = P(A, B)/P(B)$. Now, we are going to expand this concept into discrete and continuous distributions. Define $P(X = x, Y = y) = f(x, y)$ as the probability mass function (discrete) or probability density function (continuous).

Same probability laws apply to joint distributions as well.

- $f(x, y) \geq 0$ for all (x, y).
- $\sum_x \sum_y f(x, y) = 1$ or $\int_x \int_y f(x, y) dx dy = 1$

Example (Discrete): Suppose there are 10 balls in a box; 3 white, 4 black and 3 red. Two balls are randomly selected. Let's say random variable X is the number of white balls picked and r.v. Y is the number of black balls picked. (a) Find the joint probability function and (b) find the probabilities.

(a) Let's first enumerate the alternatives. (x, y) pair can be either of $(0, 0), (0, 1), (0, 2), (1, 1), (2, 0), (1, 0)$. Total number of alternatives are $\binom{10}{2}$. To calculate, number of ways of getting 1 white and 1 black ball is $\binom{3}{1}\binom{4}{1}\binom{3}{0}$. So, the probability will be $\dfrac{\binom{3}{1}\binom{4}{1}\binom{3}{0}}{\binom{10}{2}}$. We can generalize it to a function.

$$f(x, y) = \frac{\binom{3}{x}\binom{4}{y}\binom{3}{2-x-y}}{\binom{10}{2}}$$

Let's also make it into an R function

```
f_xy_ballpick <- function(x,y,picked=2,n_balls=10,n_x=3,n_y=4){
    #picked is the number of balls picked
    #n_balls is the total number of balls
    #n_x is the number of balls belonging to rv X (white)
    #n_y is the number of balls belonging to rv Y (black)
    #x and y are values to our random variables their total cannot exceed picked

    #If the sum of x and y is greater than picked, then its probability is zero.
    if(x+y > picked){
        return(0)
    }

    #Remember choose is the R function of binomial coefficient (or combination)
    (choose(n_x,x)*choose(n_y,y)*choose(n_balls - n_x - n_y,picked-x-y))/choose(n_balls,picked)

}

f_xy_ballpick(x=1,y=1)
```

```
## [1] 0.2666667
```

(b) Using the above formula we can calculate all the probabilities within the specified region $x + y \leq 2$.

```r
#First create an empty probability matrix.
#Let's say that columns are x = 0,1,2 and rows are y = 0, 1, 2
prob_matrix<-matrix(0,ncol=3,nrow=3)

#Indices in R start from 1 so 1,1 is actually x=0,y=0
prob_matrix[1,1]<-f_xy_ballpick(x=0,y=0)
prob_matrix[1,2]<-f_xy_ballpick(x=1,y=0)
prob_matrix[1,3]<-f_xy_ballpick(x=2,y=0)
prob_matrix[2,1]<-f_xy_ballpick(x=0,y=1)
prob_matrix[2,2]<-f_xy_ballpick(x=1,y=1)
prob_matrix[2,3]<-f_xy_ballpick(x=2,y=1)
prob_matrix[3,1]<-f_xy_ballpick(x=0,y=2)
prob_matrix[3,2]<-f_xy_ballpick(x=1,y=2)
prob_matrix[3,3]<-f_xy_ballpick(x=2,y=2)

#Let's also define the colnames and rownames of the matrix.
#paste0 is an R command which just appends statements
colnames(prob_matrix) <- paste0("x_",0:2)
rownames(prob_matrix) <- paste0("y_",0:2)

round(prob_matrix,2)
```

```
##      x_0  x_1  x_2
## y_0 0.07 0.20 0.07
## y_1 0.27 0.27 0.00
## y_2 0.13 0.00 0.00
```

Example (continuous): *(This is from the textbook, Example 3.15)* A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of time that the drive-in and the walk-in facilities are in use and suppose that the joint density function of these random variables is

$$f(x,y) = \frac{2}{5}(2x + 3y), 0 \leq x \leq 1, 0 \leq y \leq 1$$

and 0 for other values of x and y.

(a) Verify $\int_x \int_y f(x,y)dxdy = 1$

(b) Find $P[(X,Y) \in A]$, where $A = \{(x,y)|0 < x < 1/2, 1/4 < y < 1/2\}$

(c) (see the book for the full calculations)

$$\int_x \int_y f(x,y)dxdy = \int_0^1 \int_0^1 \frac{2}{5}(2x + 3y)dxdy = 1$$

(a) (see the book for the full calculations)

$$\int_x \int_y f(x,y)dxdy = \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5}(2x + 3y)dxdy = 13/160$$

Marginal Distributions

You can get the marginal distributions by just summing up or integrating the other random variable such as $P(Y = y) = \sum_x f(x,y)$ or $f(y) = \int_x f(x,y)dx$. Let's calculate the marginal distribution of black balls (rv Y)

2

in the above example.

```
#Let's recall the prob_matrix
round(prob_matrix,2)
```

```
##      x_0  x_1  x_2
## y_0 0.07 0.20 0.07
## y_1 0.27 0.27 0.00
## y_2 0.13 0.00 0.00
```

```
#rowSums is an R function that calculates the sum of each row.
#It is equivalent to y_0 = prob_matrix[1,1] + prob_matrix[1,2] + prob_matrix[1,3]
rowSums(prob_matrix)
```

```
##       y_0       y_1       y_2
## 0.3333333 0.5333333 0.1333333
```

Marginal distribution of y in the second example is calculated as follows.

$$\int_x \frac{2}{5}(2x+3y)dx = \frac{2(1+3y)}{5}$$

Conditional Distribution

Similar to Bayes' Rule, it is possible to calculate conditional probabilities of joint distributions. Let's denote g(x) as the marginal distribution of x and h(y) as the marginal distribution of y. The formula of conditional distribution of x given y is as follows.

$$f(x|y) = f(x,y)/h(y)$$

Note that conditional distribution function is useless if x and y are independent. $(f(x|y) = f(x))$

Conditional Expectation

We learned about conditional distributions, but what about expectations? $(E[X|Y=y])$

$$E[X|Y=y] = \sum_x xP(X=x|Y=y) \quad E[X|Y=y] = \int_x xf(x|y)dx \quad E[E[X|Y]] = \sum_y E[X|Y=y]P(Y=y) = E[X]$$

Example: A mouse is put into a labyrinth with 3 passages, at the end of the labyrinth there is cheese. First passage leads to the cheese in 3 mins. Second passage delays the mouse for 5 minutes and returns the mouse to the starting point. Third is the same as the second but the travel time is 10 minutes. It is equally likely that the mouse chooses any of those passages. What is the expected amount of time that the mouse will get to cheese?

Say T is time and Y is the passage chosen.

$$E[T] = E[E[T|Y]] = 1/3E[T|Y=1] + 1/3E[T|Y=3] + 1/3E[T|Y=3]$$

$$E[T|Y=1] = 3$$

$$E[T|Y = 2] = 5 + E[T]$$

$$E[T|Y = 3] = 10 + E[T]$$

$$E[T] = 1/3(3 + 5 + E[T] + 10 + E[T]) = 18$$

Further Topics

Moment Generating Function (MGF)

If we define a function $g(X) = X^r$ of r.v. X, the expected value $E[g(X)]$ is called the rth moment about the origin.

$$E[X^r] = \sum_x x^r f(x)$$

$$E[X^r] = \int_x x^r f(x) dx$$

The first moment gives us the expectation $E[X^1]$. With the second moment $E[X^2]$ we can calculate the variance $V(X) = E[X^2] - E[X]^2$.

The moment generating functon $M_X(t)$ is defined as follows.

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$$

$$M_X(t) = E[e^{tX}] = \int_x e^{tx} f(x) dx$$

If the sum or interval above converges, then MGF exists. If MGF exists then all moments can be calculated using the following derivative.

$$\frac{d^r M_X(t)}{dt^r} = E[X^r], at\ t = 0$$

For instance, the MGF of binomial distribution is $M_X(t) = \sum_0^n e^{tx} \binom{n}{x} p^x q^{n-x}$.

Covariance

We know about the variance ($V(X) = \sigma_x^2 = $ E[(X-E[X])^2]). But what about the variance of two dependent random variables? Then we talk about the **covariance** of the joint distribution ($V(X,Y) = E[(X - E[X])(Y - E[Y])]$) or ($E[XY] - E[X]E[Y]$).

Correlation

Simply put, it is the magnitude of (linear) relationship between random processes X and Y. Correlation coefficient can be found by using covariance and variances of the marginal distributions. $(\frac{\sigma_{XY}}{\sigma_X \sigma_Y})$.

Correlation is frequently used to indicate the similarity between two processes. Though, there is a popular saying that 'correlation does not imply causation', meaning seemingly correlated processes might actually be independent. Ask your instructor (or Google) about 'spurious correlations'.