**Joint probability distributions**

More than one random variable at a time may be of interest. For example:

* The white blood cell count for patients at time = 0, 1, and 2 hours after a medicine is administered
* The tire wear for all four tires on an automobile
* The scores for each part of the ACT exam
* The amount of pressure put on both ends of a piece of lumber until it breaks.

To quantify these items with probabilities, we can define a “joint” PDF for discrete or continuous random variables. In order to make the discussion simpler, we will mainly focus only on two random variables at a time.

The joint probability distribution function for discrete random variables X and Y, denoted as f(x,y), has the following characteristics:

* f(x,y) ≥ 0 for all of x and y
* 
* P(X = x, Y = y) = P(X = x ∩ Y = y) = f(x,y)

The joint probability distribution function for continuous random variables X and Y, denoted as f(x,y), has the following characteristics:

* f(x,y) ≥ 0 for all of x and y
* 
* P[(X, Y)∈ A] = where A is some region in the XY plane

Example: Roll a pair of dice

Let X = die #1 result and Y = die #2 result be discrete random variables.



Notice that 

Example: Grades for two courses (GradeTwoCourses.R, JointPDF.ipynb)

Let X be a random variable denoting grade in a math course and Y be a random variable denoting grade in a statistics course. For example, x = 0.90 means a 90% grade in the math course. Suppose the joint PDF is



The 3D equivalent of the curve() function is not available in R by default. Instead, a nice equivalent was presented on Stack Overflow (Q&A website for programming) that uses the rgl package (needs to be installed). Below is the R code:

> library(package = rgl)

> fxy <- function(x,y) {

 x^2 + 2\*y^2

 }

> # Test the function

> fxy(x = 0.90, y = 0.95)

[1] 2.615

> # https://stackoverflow.com/questions/11875941/3d-

 equivalent-of-the-curve-function-in-r

> curve3D <- function(f2, x.range = c(-1, 1), y.range =

 c(-1, 1), col = 1:6, xlab = "x", ylab = "y", zlab =

 "f(x,y)") {

 if (!require(rgl) ) {stop("load rgl")}

 xvec <- seq(x.range[1], x.range[2], len=15)

 yvec <- seq(y.range[1], y.range[2], len=15)

 fz <- outer(xvec, yvec, FUN = f2)

 persp3d(x = xvec, y = yvec, z = fz, col = col, xlab =

 xlab, ylab = ylab, zlab = zlab)

 }

> open3d() # Open graphics window

> curve\_3d(f2 = fxy, x.range = c(0, 1), y.range =c(0, 1),

 col = "red")

Below is a screen capture of the graphics window. The surface plotted can rotated (left click and drag on plot) and zoomed in/out (use middle mouse wheel/area).



Sage can do this plot a little easier:



Volume underneath the surface corresponds to probability. The total volume equals 1. Thus,



We can show this in Sage using



Double integration is not as easy in R, so it is not presented here.

Suppose I wanted to find the probability a randomly selected student receives a B or higher (80% or higher) in BOTH courses. Ignore the possibility of getting + or – grades. How would I represent this in the above plot?

Joint cumulative distribution function (CDF) functions can be defined as one would expect. For example, if X and Y are two continuous random variables X and Y, the joint CDF F(x, y) is

F(x,y) = P(X ≤ x ∩ Y ≤ y) = P(X ≤ x, Y ≤ y)

 = 

The t and s are used here to help avoid confusion between the limits of integration and the random variable.

How can we find the joint PDF from the joint CDF?

Given the joint PDF, we may want to just focus on one of the random variables. To do this, we need to integrate (or sum) out the other random variable.

The marginal PDF for a discrete random variable X is



and for a continuous random variable X is



The marginal probability distribution function (PDF) for a discrete random variable Y is



and for a continuous random variable Y is



Example: Roll a pair of dice

Let X = die #1 result and Y = die #2 result be discrete random variables.



Then



because



Similarly, 

Example: Grades for two courses (GradeTwoCourses.R, JointPDF.ipynb)

Let X be a random variable denoting grade in a math course and Y be a random variable denoting grade in a statistics course. Suppose the joint PDF is



The marginal PDFs are



for 0 < x < 1 and g(x) = 0 for x < 0 and x > 1.



for 0 < y < 1 and h(y) = 0 for y < 0 and y > 1.

> # g(x)

> curve(expr = x^2 + 2/3, xlim = c(0,1), xlab = "x or y",

 ylab = "g(x) or h(y)", col = "red", ylim = c(0,2.5),

 lwd = 2, panel.first = grid())

> # h(y)

> curve(expr = 1/3 + 2\*x^2, xlim = c(0,1), col = "blue",

 add = TRUE, lwd = 2)

> abline(h = 0)

> legend(x = 0, y = 2, legend = c("g(x)", "h(y)"), col =

 c("red", "blue"), lwd = 2, lty = 1, bty = "n")



Which is more likely: receiving greater than an 80% in math or statistics? Explain using the plots.

Sage:



What is the probability a student received a grade of B or better (>80%) in the math course?

Notice that nothing is said at all about the statistics course.



Equivalently, one could have started with .

Sage:



What is the probability a student received a grade of B or better (>80%) in the stat course?



What is the probability of receiving a B or better (>80%) in both classes?

Find P(X>0.8, Y>0.8) = P(X>0.8 ∩ Y>0.8)



Sage:



What is the probability of getting a B or better in at least one of the classes?

P(X > 0.8 ∪ Y > 0.8)

= P(X > 0.8) + P(Y > 0.8) – P(X > 0.8 ∩ Y > 0.8)

= 0.296 + 0.392 – 0.0976

 = 0.5904

Similar to the discussion of conditional probabilities for events earlier in the course, a conditional PDF for random variables can be found. Let X and Y be random variables. The conditional PDF of Y given X = x is

 for g(x) > 0

The conditional PDF of X given Y = y is

 for h(y) > 0

Remember from earlier that P(A|B) = P(A∩B)/P(B).

Example: Grades for two courses (GradeTwoCourses.R, JointPDF.ipynb)

Let X be a random variable denoting grade in a math course and Y be a random variable denoting grade in a statistics course with joint PDF of



Suppose a randomly selected student received a 75% in math, find the PDF for the statistics course grade.

This problem may be more realistic if X and Y denoted final exam grades. Suppose the math final exam is given first and the student receives a 75%. Given the performance on the math final exam, what will happen on the statistics final exam?

The conditional PDF of Y given X = x is

 for 0 < y < 1 and

 otherwise.

Notice that x is not a random variable in the above equation. Instead it is treated as a constant since we are conditioning on its observed value. When 0.75 is observed for X, the conditional PDF of Y given becomes





Please see the R program for code.

Given a 75% in the math class, what is the probability the student gets a B or higher (80% or higher) for statistics?

P(Y > 0.8 | X = 0.75) =







= 0.3562

Sage:



Notice the use of subs method here.

Suppose X was observed to be some other value. The plots show what happens for a few examples.



What happens to the probability of receiving a B or higher (Y > 0.8) as X increases? Would you expect this in a real situation? :)

Often instructors will tell students that their performance in a past course is directly related to their performance in the instructor’s course.

How a student performed in calculus will likely have an effect on how a student will perform in our course. Suppose X was a random variable denoting calculus course grade and Y was a random variable denoting our course’s grade. Based upon passed teaching experience, I could have an approximate joint PDF for X and Y.

Suppose a student received a 75% grade in the calculus course. Given this grade, I could give the student the conditional PDF for the student’s grade for our course. This would help the student determine the probability of receiving a particular grade in our course given information about the calculus course.

Similar to how independence between events was discussed earlier in our course, we can discuss independence between random variables here. Why do you think this is important?

Two random variables X and Y with joint PDF f(x,y) and marginal PDFs of g(x) and h(y), respectively are independent if

f(x,y) = g(x)h(y)

for all possible x and y.

Remember from earlier that P(A∩B) = P(A)P(B) if events A and B are independent

Another definition of independence:

If f(y|x) = h(y), then X and Y are independent.

If f(x|y) = g(x), then X and Y are independent.

Why? Remember that f(y|x) = f(x,y)/g(x). When X and Y are independent, we can replace f(x,y) with g(x)h(y). This results in f(y|x) = g(x)h(y)/g(x) = h(y).

Independence can be generalized to n random variables X1, X2, …, Xn. With a joint PDF of f(x1, x2, …, xn) and marginal PDFs of f(x1), f(x2), …, f(xn), the random variables are independent when

f(x1, x2, …, xn) = f(x1) × f(x2) × × f(xn)

for all possible x1, x2, …, xn.

Example: Roll a pair of dice.

Let X = die #1 result and Y = die #2 result be discrete random variables.



Are the outcomes of rolling a pair of dice independent?

g(x) = 1/6 for x = 1, , 6 and h(y) = 1/6 for y = 1, , 6. Then g(x)×h(y) = 1/36 for all x and y. Thus, X and Y are independent random variables!

Why is this important to know?

* Marginal PDFs are much easier to work with than joint PDFs. For example, look at how much easier it was to examine area underneath a curve in 2D than in 3D.
* Independence helps to simplify the problem. Instead of needing the probabilities for 6×6 = 36 different possible permutations of X and Y outcomes, we need to know only 6 + 6 = 12 probabilities for X and Y individually to get all 36 possible probabilities!

Example: Grades for two courses (GradeTwoCourses.R, JointPDF.ipynb)

Let X be a random variable denoting grade in a math course and Y be a random variable denoting grade in a statistics course with joint PDF of



Are math and statistics course grades independent? From a student’s perspective, why would this be important to know?

Previously, we found

 for 0 < x < 1 and

 for 0 < y < 1

Now,

 g(x)×h(y) = (x2 + 2/3)(1/3 + 2y2)



Thus, X and Y are dependent. Remember that one can also check if either f(y|x) = h(y) or f(x|y) = g(x) are satisfied to determine independence or dependence.

Can we numerically measure the strength and direction of the dependence? Yes – this will be discussed in the next chapter!

Below is a 3D plot comparing the PDF with (blue) and without (red) the independence assumption. If we incorrectly assumed independence, the probability of getting extremely high or low scores in both classes simultaneously would be larger.



