**Chapter 3 Disjoint Probability Distribution (Updated: October 6, 2016)**

**Lecture notes**

3.0 Review on what we have learned

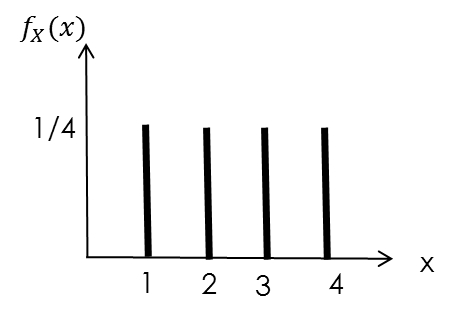
In this chapter, we discuss probabilities that involve two or multiple random variables. We will see the natural extension of notations for a single random variable to those for multiple random variables. Before we proceed, let us review some important notations that we have studied until now.

* A random variable X is a function from the sample space to numerical values.
* PMF for discrete random variables (or PDF for continuous random variables):
* Expectation of a (discrete) random variable X:
* A trial on measuring how far random variable values are from the average:
* Variance, Var(X):
* Standard deviation, :

A new topic that has not been covered is conditional PMFs of a discrete random variable X and expectations:

In chapter 1, we have probabilities P(B), and we also have conditional probabilities P(B|A). What is the difference between the two? Essentially, there is no difference. Probabilities are just an assignment of probability values to given different outcomes. Somebody comes and gives you a new piece of information. And you have new probabilities. We call these new probabilities conditional probabilities. But they behave exactly the same as ordinary probabilities. For example, all the conditional probabilities sum to one as well.

To explain conditional PMF and expectation concretely, suppose we have a PMF described as in the following graph:



Let A be an event that X is greater than or equal to 2. In the graph, the event A consists of these three values: 2, 3, and 4. Then what is the conditional PMF, given that we are told that event A has occurred? It basically tells us that the value 1 has not occurred. There are only three possible outcomes now. Whenever you condition, the relative likelihoods remain the same. They keep the same proportions. They just need to be re-scaled, so that they add up to one. So each one of them is going to get a probability of 1/3 in the conditional situation. Therefore, our conditional pmf is:

You have a random variable and a PMF for it, and so you can talk about the conditional expected value of that random variable.

Similarly, you can safely generalize what we have about (unconditional) probabilities to conditional counterparts. For example,

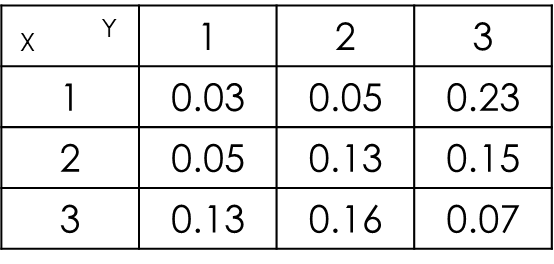
In this chapter 3, all these concepts, PMF/PDF, expectation, variance, conditional PMF/PDF, and conditional expectation for a single random variable, will be generalized to ones for two and multiple random variables.

3.1 Joint Probability Distribution (결합확률분포)

A typical experiment may have several random variables associated with that experiment. For example, a typical student has height and weight. If I give you the PMF of height, it tells me something about distribution of heights in the calls. If I give you the PMF of weight, it tells me something about the different weights in the class. But if I want to ask a question, is there an association between height and weight, then I need to know a little more. And the individual PMFs of height and weight do not tell me anything about those relations. To be able to say something about those relations, I need to know something about joint probabilities. It is how likely is it that certain Xs go together with certain Ys. So these probabilities, essentially, capture associations between these two random variables. And it’s the information I would need to have to do any kind of statistical study that tries to relate the two random variables with each other. This is what is about joint probability(결합확률) or joint probability distribution(결합확률분포).

This is just the notation for what we call the joint PMF(결합확률질량함수). It is the joint probability mass function of the two random variables X and Y that we look at together, jointly. And it gives us the probability that any particular numerical outcome pair does happen.

Activity1(예제2). Here is a joint PMF for two random variables X and Y. In the finite case, you can represent joint PMFS by a table. The following table gives you information on the joint probability of X and Y.



* Is this joint PMF really a PMF? It is so because all probabilities that this joint PMF gives are all non-negative and the total probability is equal to 1.
* What is the joint probability at x=3 and y=2?
* Suppose we are only interested in the distribution of the X. We don’t care about the Y’s. Then what is the probability that X takes on a particular value? Can we find it from the table? For example, how is the probability that X is 3 calculated? We just take a slice of the joint probabilities keeping X fixed as x and adding up all possible different values of Y. We call this marginal PMFs.

Such a function in Activity 1 is call the marginal PMF(주변확률질량함수). We have the joint PMF when we talk about both of X and Y together, and the marginal PMF when we talk about them one at the time.

Once we have a PMF, it is not difficult to make up a distribution function. For Activity1, we can make up a joint distribution function (결합분포함수).

As we have joint distribution functions for joint probability mass/density functions, we can also think of the notion of marginal distribution functions(주변분포함수) for joint distribution functions. Marginal distribution functions of a random variable X also calculate an accumulated sum of probabilities until X is some particular value x but they sum up all probabilities over the other random variable, say, Y.

Activity2 (예제2). Given a disjoint probability mass function as in Activity1, calculate , , and .

For given two continuous random variables X and Y, there are corresponding definitions of

* joint probability density function,
* marginal probability density function,
* joint distribution function
* marginal distribution function.

Note two things. First, since probabilities in an experiment with continuous random variables are defined as some area, the disjoint probability density as a counterpart of a disjoint distribution function is calculated by applying doubly differential operators to the joint distribution function.

Second, is calculated by a disjoint distribution function as follows:

To see why, refer to Figure 3.4 in the main text.

3.2 Conditional Probability Distribution (결합확률분포)

Since we have introduced conditional PMF in the beginning, we also explore the idea of the conditional PMFs with the example of Activity 1.

This conditional PMF is just a conditional probability that we studied in Chapter 1. The conditional PMF f(x|y) is the probability that X takes on a particular value, given Y takes a certain value y. For example, let us take little y to be equal to 2. It means that we are conditioning to live in the smaller world than one shown in the table (Activity1). Conditional probabilities are proportional to original probabilities. So, is equal to the ratio of the joint PMF and the marginal one, as follows:

For the other little x, the same story goes. The last remark is that conditional probabilities are just probabilities, so a conditional PMF must sum to 1, no matter what event you are conditioning on.

What we have done until now is that I give you students an impression that the notion of probabilities for a single random variable is naturally generalized to those for two random variables, and more for multiple random variables.

Activity 1. With the same joint probability mass function in Activity 1 of Section 3.1, answer the following questions.

* Given Y=2, what is the conditional probability mass function of X?
* Given Y=2, what is the probability of X=1, and X=3, respectively?

Note that the difference of P(X,Y) and P(X|Y). P(X,Y) is the probability that both events denoted by X and Y happen simultaneously. P(X|Y) is the probability similar to P(X,Y), but under the condition that Y has happened (that is, in the smaller world). This difference is described as graphs in Figure 3.5 and Figure 3.6.

Two random variables X and Y are said to be independent with each other if

*or*  for all x and y

What is this independence condition telling us? Suppose you know the probability of X=x by . One day, your friend comes to you, and he or she gives you a piece of information that y has occurred. But the information never changes your belief on the probability of X. In other words, the information that y has occurred gives no hint about the probability of x at all.

Note that the independence condition should be satisfied for all x and y. Therefore, it is not a single condition, but a bunch of conditions.

The independence condition above is equivalent, by the definition of the conditional probability mass/density functions, to the following ones:

for all x and y

for all x and y

Activity 2. Verify the independence of X and Y in the following examples.

* 3.3절 예제4-(1), 예제3
* 3.1절 예제2

Until now, we have considered experiments where only two random variables are involved. For experiments with three random variables and more, the same explanation can be made. For example, f(x,y,z) is a probability mass/density function over three random variables X, Y, Z. It says the probability that all of x, y, and z happen simultaneously. No difference from the notation when we have two random variables!

As we see in the multiplication law in Chapter 1(P.26 of the main text), a joint probability mass/density function f(x,y,z) can be stated as:

Since readers have studied the notation for conditional probability, they are familiar with the notation f(y|x). Therefore, they should be familiar with the new but generalized notation for f(z|y,x), which gives us the probability that, given x and y have simultaneously occurred, z occurs.

Three random variables X, Y, and Z are said to be all independent if

Three random variables X, Y, and Z are said to be pairwisely independent if

* ,
* , and
* .

The fact that three random variables X, Y, and Z are pairwisely independent does not necessarily imply the fact that the three random variables are all independent. Can you give an example to explain this?

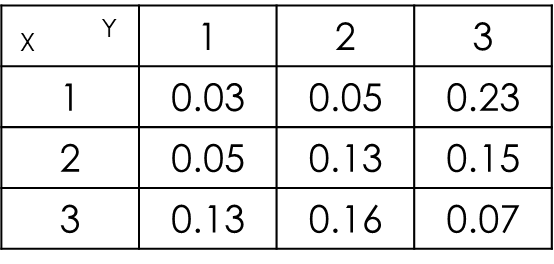
3.3 Expectation in Joint Probability Distribution (결합분포에대한 기댓값)

This section studies the notion of expectation and covariance(공분산) in joint probability distribution. An arbitrary function g(X) for a random variable X is another random variable. We have studied how to compute the expectation of g(X) by the probability distribution of X, not of g(X). Similarly, the expectation of an arbitrary function u(X,Y) is defined with a joint probability distribution f(x,y), as follows:

* + (for discrete random variables)
  + (for continuous random variables)

Note that the expectation of linear functions has the following property as:

Activity1 (3.1절 –예제2의 테이블)



* E(X)
* E(X+Y)
* E(XY)

For E(X+Y), takes two ways to calculate it. First, use the definition of the expectation of X+Y as u(X,Y). Second, use the equation of E(X+Y)=E(X)+E(Y).

Activity2 (3.3절 – 예제2) Given a joint probability density function

* Are X and Y independent with each other?
* E(X)
* E(Y)
* E(XY)
* E(XY)=E(X)E(Y)

In the previous chapter, the variance of a random variable such as Var(X) and Var(Y) has been introduced. For joint probability distribution, people may be interested in such a measure. But they may be also interested in a measure to show the dependence of X and Y is covariance and correlation. The covariance of X and Y is:

One little problem is that the unit of Cov(X,Y) is the product of two units of X and Y. To avoid this awkwardness on the units, the notion of correlation coefficient is introduced.

Activity 3(3.1절 – 예제2 테이블). Using your answer to Activity 1, compute:

* Cov(X,Y)
* Corr(X,Y)