MAT2013: Probability and Statistics [1] [2] **Spring 2017**

Lecture 4: Multidimensional Random Variables

Lecturer: Jae Guk, Kim *Scribes: Jisung Lim*

Scribes: Jisung Lim, *B.S. Candidate of Industrial Engineering in Yonsei University, South Korea.*

4.1 Notation

This note use some different notation than before. The brief changes are here.

a	A scalar
$\boldsymbol{a}$	A vector
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable

4.2 Introduction

We are going to investigate situations in which we deal simultaneously with several random variables on the same sample space. That is, for more than two variables, let's say two of X and Y, we want to find the distribution of $P_{XY}[(X,Y) \in A]$ where $A \in \mathcal{B}_2$ (a σ-*algebra* consisting of sbusets of $\mathbb{R}^2$). We go through finite or countable random valued vector with an example.

Example 4.2.1. **Tossing a coin 2 times**
At first, let us consider the primitive probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For each coin tossing, there are two disjoint outcomes H (for head), and T (for tail), so that the sample space is given by $\Omega = \{HH, HT, TH, TT\}$ and the event space is given by $\mathcal{A} = 2^\Omega$ which is the σ-*algebra* of Ω. Also, the probability measure $\mathbb{P}$ is given as a equiprobable model for each single outcomes, (i.e., $\mathbb{P}(HH) = \mathbb{P}(HT) = \mathbb{P}(TH) = \mathbb{P}(TT) = 1/4$).

Now let X_1 and X_2 be random variables that denote the number of heads obtained in the first and second toss, respectively. Thus, $X(\omega) = (X_1(\omega), X_2(\omega))$, and

$$\begin{aligned} X(HH) &= (1,1), \quad X(HT) = (1,0) \\ X(TH) &= (0,1), \quad X(TT) = (0,0). \end{aligned}$$

By the vector valued random variable X, the existing probability space is mapped to the new one $(\mathbb{R}^2, \mathcal{B}_2, P_X)$, where induced distribution $P_X(X_1 = x_1, X_2 = x_2)$ is given by

$$P_X(X_1 = x_1, X_2 = x_2) = \mathbb{P}[\{\omega \in \Omega : X_1(\omega) = x_1, X_2(\omega) = x_2\}]$$

and the probability distribution function $f_X(x_1, x_2)$ is given by

$$f_X(x_1,x_2) = \begin{cases} 1/4, & x_1 = 1, x_2 = 1 \\ 1/4, & x_1 = 1, x_2 = 0 \\ 1/4, & x_1 = 0, x_2 = 1 \\ 1/4, & x_1 = 0, x_2 = 0 \\ 0, & \text{otherwise.} \end{cases}$$

also we can see that f_X satisfies

1. $f_X(x_1, x_2) \geq 0, \quad \forall (x_1, x_2) \in \mathbb{R}^2$
2. $\sum_{x \in \mathbb{R}^2} f_X(x) = 1$

$\Diamond$

4.2.1 Vector-valued Random Variable and Induced Distribution

In a more general sense, what we've seen in the example is to define *random variable*, says $\mathbf{x}$, to *map the primitive probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ (which is tedious and inefficient to analyze mathematically) *into the new one on* $\mathbb{R}^n$ $(\mathbb{R}^n, \mathcal{B}_n, P_{\mathbf{x}})$ (which is appropriate to facilitate mathematical analysis). In that process, a new probability distribution is induced according to the random variable.

The whole things are achived via the vehicle of the random variable defined as follow:

$$\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots, \mathrm{x}_\mathrm{n}) \tag{4.1}$$

which maps the measurable space $(\Omega, \mathcal{A})$ into $(\mathbb{R}^n, \mathcal{B}_n)$. That is,

$$\begin{aligned} \mathbf{x} : (\Omega, \mathcal{A}) &\to (\mathbb{R}^n, \mathcal{B}_n) \\ \omega &\mapsto \boldsymbol{x} \end{aligned} \tag{4.2}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, \ldots, x_n)$. And also $P_{\mathbf{x}} : \mathcal{B}_n \to \mathbb{R}$ is defined as follows

$$\forall B \in \mathcal{B}_n, \quad P_{\mathbf{x}}(\mathbf{x} \in B) \coloneqq \mathbb{P}(\{\omega \in \Omega : \mathbf{x}(\omega) \in B\}) \tag{4.3}$$

wihch is called probability distribution of $\mathbf{x}$ on the sample space $\mathbb{R}^n$ with *σ-algebra* $\mathcal{B}_n$.

4.2.2 Distribution Function and PDF

The story here goes the same as the case of 1-d random variable. Thus, we have a natural extension of distribution function of random variable for the vector-valued random variable:

$$F_{\mathbf{x}}(\boldsymbol{x}) = P_{\mathbf{x}}(\mathbf{x} \leq \boldsymbol{x}) = P_{\mathbf{x}}(\mathrm{x}_1 \leq x_1, \mathrm{x}_2 \leq x_2, \ldots, \mathrm{x}_n \leq x_2) \tag{4.4}$$

4.2.2.1 Discrete Case

For discrete random vector, $F_{\mathbf{x}}(\boldsymbol{x})$ is given by

$$F_{\mathbf{x}}(\boldsymbol{x}) = \sum_{\mathbf{x} \le \boldsymbol{x}} f_{\mathbf{x}}(\boldsymbol{x}) = \sum_{\mathrm{x}_1 \le x_1, \ldots, \mathrm{x}_n \le x_n} f_{\mathbf{x}}(\boldsymbol{x}). \tag{4.5}$$

and the mass function $f_{\mathbf{x}}(\boldsymbol{x})$ can be obtained by

$$f_{\mathbf{x}}(\boldsymbol{x}) = P_{\mathbf{x}}(\mathbf{x} = \boldsymbol{x}) = P_{\mathbf{x}}(\mathrm{x}_1 = x_1, \mathrm{x}_2 = x_2, \ldots, \mathrm{x}_n = x_n) \tag{4.6}$$

and satisfies

1. $f_{\mathbf{x}}(\boldsymbol{x}) \ge 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^n$
2. $\sum_{\boldsymbol{x} \in \mathbb{R}^n} f_{\mathbf{x}}(\boldsymbol{x}) = 1$

4.2.2.2 (Absolutely) Continuous Case

For continuous random vector, $F_{\mathbf{x}}(\boldsymbol{x})$ is given by

$$\forall B \in \mathcal{B}_n, \quad F_{\mathbf{x}}(\boldsymbol{x}) = \int_{\mathbb{R}^n} f_{\mathbf{x}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{x}}(x_1, x_2, x_3, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{4.7}$$

and the density $f_{\mathbf{x}}(\boldsymbol{x})$ can be obtained by partial differentiation, given by

$$f_{\mathbf{x}}(\boldsymbol{x}) = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x})}{\partial x_1 \partial x_2 \cdots \partial x_n} \tag{4.8}$$

and satisfies

- $f_{\mathbf{x}}(\boldsymbol{x}) \ge 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^n$
- $\int_{\mathbb{R}^n} f_{\mathbf{x}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = 1$

4.2.3 Marginal Distributions

Now, at this point, it is natural to ask if we get individual P_{x_i}'s from the joint probability distribution $P_{\mathbf{x}}$. Let's consider the coin-tossing example to get an intuitive idea. (From this section, we only consider two random variable case for notational convenience)

Example 4.2.2. **Tossing a coin 2 times**
Reconsider the example tossing coin twice. First, the sample space is given by

$$\Omega = \{HH, HT, TH, TT\}$$

and let x_1 and x_2 be random variables that denote the number of heads obtained in the first and second toss, respectively. Thus, $\mathbf{x}(\omega) = (\mathrm{x}_1(\omega), \mathrm{x}_2(\omega))$. Then, the probability that the outcome of first coin has one head is

$$\begin{aligned}
f_{\mathrm{x}_1}(1) &= P_{\mathrm{x}_1}(\mathrm{x}_1 = 1) \\
&= \mathbb{P}[\{HH, HT\}] \\
&= \mathbb{P}[\{HH\} \cap \{HT\}] \\
&= \mathbb{P}[\{HH\}] + \mathbb{P}[\{HT\}] \\
&= P_{\mathbf{x}}(\mathrm{x}_1 = 1, \mathrm{x}_2 = 1) + P_{\mathbf{x}}(\mathrm{x}_1 = 1, \mathrm{x}_2 = 0) \\
&= f_{\mathbf{x}}(1, 1) + f_{\mathbf{x}}(1, 0)
\end{aligned}$$

So,

$$\begin{aligned}
f_{\mathrm{x}_1}(1) &= \sum_{x_2} f_{\mathbf{x}}(1, x_2) = \frac{1}{2} \\
f_{\mathrm{x}_1}(0) &= \sum_{x_2} f_{\mathbf{x}}(0, x_2) = \frac{1}{2}
\end{aligned}$$

$\Diamond$

In conclusion, for a random vector $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2)$, the probability distribution of x_1 and x_2 are

$$\begin{aligned}
f_{\mathrm{x}_1}(x_1) &= \sum_{x_2} f_{\mathbf{x}}(x_1, x_2) \\
f_{\mathrm{x}_2}(x_2) &= \sum_{x_1} f_{\mathbf{x}}(x_1, x_2)
\end{aligned} \tag{4.9}$$

and we call $f_{\mathrm{x}_1}(x_1)$ and $f_{\mathrm{x}_2}(x_2)$ marginal probability mass function of $\mathbf{x}$.

For the case of (absolutely) continuous random vector, $f_{\mathrm{x}_1}(x_1)$ and $f_{\mathrm{x}_2}(x_2)$ is given by

$$\begin{aligned}
f_{\mathrm{x}_1}(x_1) &= \int_{\mathbb{R}} f_{\mathbf{x}}(x_1, x_2)\, \mathrm{d}x_2 \\
f_{\mathrm{x}_2}(x_2) &= \int_{\mathbb{R}} f_{\mathbf{x}}(x_1, x_2)\, \mathrm{d}x_1
\end{aligned} \tag{4.10}$$

4.3 Independent Random Variables

We discussed about the independence of multiple events before. In this chapter, since we've deal with more than two random variables simultaneously, we may inclined to discuss about the independence between multiple random variables. Let's see the coin-tossing example again.

Example 4.3.1. **Tossing a coin 2 times**
Consider tossing a coin two times where the sample sapce is $\Omega = \{HH, HT, TH, TT\}$. Let E_1 and E_2 be events that head is obtained as an outcome of the first and second toss, respectively.

Then, E_1 and E_2 is given by

$$E_1 = \{HH, HT\} \quad \text{and} \quad E_2 = \{HH, TH\}.$$

Since $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(\{HH\}) = 1/4 = 1/2 \cdot 1/2 = \mathbb{P}(E_1)\mathbb{P}(E_2)$, then the event E_1 and E_2 is independent. To consider the case of tail occurred, let $E_1' = \{TT, TH\}$ and $E_2' = \{TT, HT\}$, then

$$\begin{aligned}
\mathbb{P}(E_1 \cap E_2') &= \mathbb{P}(E_1)P(E_2') \\
\mathbb{P}(E_1' \cap E_2) &= \mathbb{P}(E_1')P(E_2) \\
\mathbb{P}(E_1' \cap E_2') &= \mathbb{P}(E_1')P(E_2')
\end{aligned}$$

$\Diamond$

At this point, someone may notice that each event is corresponding to the random variable x_1 and x_2, which was defined in *Example 4.2.2.* i.e.,

$$\begin{aligned}
\{\omega \in \Omega : \omega \in \mathrm{x}_1^{-1}(1)\} &= E_1 \\
\{\omega \in \Omega : \omega \in \mathrm{x}_1^{-1}(0)\} &= E_1' \\
\{\omega \in \Omega : \omega \in \mathrm{x}_2^{-1}(1)\} &= E_2 \\
\{\omega \in \Omega : \omega \in \mathrm{x}_2^{-1}(0)\} &= E_2'
\end{aligned}$$

We know that the outcome of first and second coin is independent, therefore intuitively we want to say that random variable x_1 and x_2 are independent *if and only if* for all events $E_1(x_1)$ and $E_2(x_2)$ such that

$$\begin{aligned}
E_1(x_1) &= \{\omega \in \Omega : \omega \in \mathrm{x}_1^{-1}(x_1)\} \quad \forall x_1 \in \mathbb{R} \\
E_2(x_2) &= \{\omega \in \Omega : \omega \in \mathrm{x}_2^{-1}(x_2)\} \quad \forall x_2 \in \mathbb{R}
\end{aligned}$$

the following equality is satisfied

$$\mathbb{P}(E_1(x_1) \cap E_2(x_2)) = \mathbb{P}(E_1(x_1))\mathbb{P}(E_2(x_2))$$

Definition 4.3.1. Random variables $\mathrm{x}_1, \ldots, \mathrm{x}_n$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ are said to be mutually independent *if and only if* for all $B_1, B_2, \ldots, B_n \in \mathcal{B}_n$,

$$P_{\mathbf{x}}(\mathrm{x}_1 \in B_1, \mathrm{x}_2 \in B_2, \ldots, \mathrm{x}_n \in B_n) = P_{\mathrm{x}_1}(\mathrm{x}_1 \in B_1)P_{\mathrm{x}_2}(\mathrm{x}_2 \in B_2) \cdots P_{\mathrm{x}_n}(\mathrm{x}_n \in B_n) \tag{4.11}$$

$\square$

Theorem 4.3.2. Suppose that $\mathrm{x}_1, \ldots, \mathrm{x}_n$ are random variables, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then $\mathrm{x}_1, \ldots, \mathrm{x}_n$ are mutually independent *if and only if*

$$\forall \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, \quad F_{\mathbf{x}}(\boldsymbol{x}) = F_{\mathrm{x}_1}(x_1)F_{\mathrm{x}_2}(x_2) \cdots F_{\mathrm{x}_n}(x_n) \tag{4.12}$$

$\square$

Corollary 4.3.3. Let (x, y) be a random vector with probability distribution function $f_{\mathrm{x}}(x)$and $f_{\mathrm{y}}(y)$. Then, x and y are mutually independent (i.e., $F_{\mathrm{xy}}(x,y) = F_{\mathrm{x}}(x)F_{\mathrm{y}}(y)$), *if and only if*

$$\forall x, y \in \mathbb{R}, \quad f_{\mathrm{xy}}(x,y) = f_{\mathrm{x}}(x)f_{\mathrm{y}}(y) \tag{4.13}$$

□

Corollary 4.3.4. Let (x, y) be an independent random variable, then

$$\mathbb{E}_{\mathrm{xy}}[g(\mathrm{x})h(\mathrm{y})] = \mathbb{E}_{\mathrm{x}}[g(\mathrm{x})]\mathbb{E}_{\mathrm{y}}[h(\mathrm{y})] \tag{4.14}$$

□

4.4 Distributional Characteristics

4.4.1 Expectation

For random vector $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, the expectation $\mathbb{E}_{\mathbf{x}}[\cdot]$ is defined by

Definition 4.4.1. Let $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ is a random vector of which each component is defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the expectation of the function $g_{\mathbf{x}}(\boldsymbol{x})$ is

$$\begin{aligned} E[g_{\mathbf{x}}(\boldsymbol{x})] &:= \int_{\Omega} g_{\mathbf{x}}(\mathbf{x}(\omega))\, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}^n} g_{\mathbf{x}}(\boldsymbol{x})\, \mathrm{d}P_{\mathbf{x}}(\boldsymbol{x}) \\ &= \int_{\mathbb{R}^n} g_{\mathbf{x}}(\boldsymbol{x}) f_{\mathbf{x}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \end{aligned} \tag{4.15}$$

and for discrete case

$$E[g_{\mathbf{x}}(\boldsymbol{x})] = \sum_{\boldsymbol{x} \in \mathbb{R}^n} g_{\mathbf{x}}(\boldsymbol{x}) f_{\mathbf{x}}(\boldsymbol{x}) \tag{4.16}$$

□

Theorem 4.4.1. If x and y are independent, then

$$\mathbb{E}[g(\mathrm{x})h(\mathrm{y})] = \mathbb{E}[g(\mathrm{x})]\mathbb{E}[h(\mathrm{y})] \tag{4.17}$$

□

Proof: Use Fubini's theorem. **D.I.Y.** ■

4.4.2 Covariance

As we saw before, 'independence' indicates that there is no probabilistic relationship between random variables. Then, naturally, we can be inclined to consider the case where there **is** a probabilistic realtionship.

For our consideration, we investigate two aspects of the relationship: the type and the intensity. Let us consider the relationship between two random variables x : *height* and y : *weight* of male students at Yonsei university.

Intuitively, the large values of x's tends to be observed with large values of y's and small values of x's with smal values of y's. i.e., if $x > \mu_{\mathrm{x}}$, then $y > \mu_{\mathrm{y}}$ is likely to be true and the product $(x - \mu_{\mathrm{x}})(y - \mu_{\mathrm{y}})$ will be positive, and if $x < \mu_{\mathrm{x}}$, then $y < \mu_{\mathrm{y}}$ is likely to be true and the product $(x - \mu_{\mathrm{x}})(y - \mu_{\mathrm{y}})$ will also be positive.

Thus, the sign of $(x - \mu_{\mathrm{x}})(y - \mu_{\mathrm{y}})$ gives us information regarding the type of relationship between X and Y. Hence, the covariance $Cov[\mathrm{x}, \mathrm{y}] = \mathbb{E}[(\mathrm{x} - \mu_{\mathrm{x}})(\mathrm{y} - \mu_{\mathrm{y}})]$ indicates the tendency; the type of the relationship.

Definition 4.4.2. Covariance between two random variable x and y is given by

$$Cov[\mathrm{x}, \mathrm{y}] = \mathbb{E}[(\mathrm{x} - \mu_{\mathrm{x}})(\mathrm{y} - \mu_{\mathrm{y}})] \tag{4.18}$$

□

Now we need to consider the other tendency, called intensity, which decides how strong or weak is the relationship. Although the covariance provides an information about what the type of the relationship is, it can not give us any consistent information about the intensity. Hence, we need to normalize the quantitiy with standard deviation of each random variables σ_{x} and σ_{y}, respectively. Therefore, the correlation $\rho_{\mathrm{xy}} = Cov(\mathrm{x}, \mathrm{y})/\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}$ indicates the tendency; the intensity of the relationship.

Definition 4.4.3. Correlation between two random variable x and y is given by

$$\rho_{\mathrm{xy}} = \frac{Cov(\mathrm{x}, \mathrm{y})}{\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}} \tag{4.19}$$

□

4.5 Conditioning

This chapter is not yet completed.

- Formal definition of conditional expectation [3]
- Sigma algebra generated from random variable [4]

We once discussed about the conditional probability between events, let's say two A and B, in the same probability space $(\mathbb{R}, \mathcal{B}, P_X)$, let's say induced by X, which has the form of

$$P_X[A|B] = \frac{P_X[A \cap B]}{P_X[A]}$$

and this is also valid in the 'primitive' probablity space $(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

Sometimes, there may exist a strong need to conditioning some random variables upon other variables. Now to develop an idea for the conditioning, let us first see the coin-tossing example to illustrate the process to evaluate some function under some conditions using conditional probability.

Example 4.5.1. **Add up the coins with head**
Consider you tossed three coins 10, 50, and 100 won. Then, the sample space is

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Define X by $X(\omega)$: The sum of the (monetary) values of those coins that land with heads up.
What is the expected value of X given that two coins have landed with heads up?

Solution: The given condition can be denoted by a subset $B \subset \Omega$ such that

$$B = \{HHT, HTH, THH\}$$

then, we can define the conditional expectation of X given B as follows

$$\begin{aligned}
\mathbb{E}_X[X|B] &\coloneqq \sum_{\omega_0 \in B} X(\{\omega_0\})\mathbb{P}[\{X = X(\{\omega_0\})\}|B] \\
&= \sum_{\omega_0 \in B} X(\{\omega_0\}) \frac{\mathbb{P}[\{X = X(\{\omega_0\})\} \cap B]}{\mathbb{P}[B]} \\
&= \frac{1}{\mathbb{P}[B]} \sum_{\omega_0 \in B} X(\{\omega_0\})\mathbb{P}[\{\omega_0\}] \\
&= \frac{1}{\mathbb{P}[B]} \sum_{\omega_0 \in B} X(\{\omega_0\}) P_X(X = X(\{\omega_0\})) \\
&= \frac{1}{\mathbb{P}[B]} \sum_{x \in X(\omega_0)} x P_X(X = x)
\end{aligned}$$

or, for (absolutely) continuous random variable,

$$\mathbb{E}_X[X|B] = \frac{1}{\mathbb{P}[B]} \int_{X(\omega_0)} x P_X(X = x)\, \mathrm{d}x$$

$\blacksquare$ $\diamond$

In this example, we used the given information to find the events of interest $B \subset \Omega$ and then we only considered those events to evalute the function (i.e., $\mathbb{E}_X[X|B]$). Note that the evaluated function is conditioned by a subset of Ω, not another random variable. Hence, now we develop this concept into the one for random variable.

Example 4.5.2. **(Continued from 4.5.1)**
Let Y be a random variable defined by

$$Y(\omega) : \text{the sum of the values of 10 and 50 when they show heads,}$$

then we can find associated event

$$Y(\omega) = \begin{cases} 0, & \omega \in \{TTH, TTT\} \\ 10, & \omega \in \{HTH, HTT\} \\ 50, & \omega \in \{THH, THT\} \\ 60, & \omega \in \{HHH, HHT\} \end{cases}$$

and a collection of the events assiciated with Y is given by

$$\mathbb{C} = \{\{TTH, TTT\}, \{HTH, HTT\}, \{THH, THT\}, \{HHH, HHT\}\}.$$

Note that the collection of the events is the subset of σ-*algebra* of primitive sample space Ω. Now, we can evaluate each expectation values of X given those events, respectively.

$$\begin{aligned} \mathbb{E}[X|\{Y=0\}] &= a \\ \mathbb{E}[X|\{Y=10\}] &= b \\ \mathbb{E}[X|\{Y=50\}] &= c \\ \mathbb{E}[X|\{Y=60\}] &= d \end{aligned}$$

From this results, we are inclined to sum up those four results to construct a new function as follows

$$\mathbb{E}[X|Y](\omega) = \begin{cases} a & Y(\omega) = 0 \\ b & Y(\omega) = 10 \\ c & Y(\omega) = 50 \\ d & Y(\omega) = 60 \end{cases}$$

which is called conditional expectation of X given Y. [3] $\Diamond$

As we can see, the conditional expectation of X given Y is the function $\mathbb{E}[X|Y] : \mathbb{C} \to \mathbb{R}$ where $\mathbb{C}$ is the subset of σ-*algebra* of Ω. Hence, we can see this set function a random variable. Furthermore, we can always find a function, says $\phi(\cdot)$, which coposite with function Y so that it can represent the function $\mathbb{E}[X|Y]$, i.e.,

$$\exists \phi \quad \text{s.t.} \quad \mathbb{E}[X|Y](\omega) = (\phi \circ Y)(\omega) = \phi(Y(\omega)) \tag{4.20}$$

Hence, the conditional expectation of X given Y is the form of

$$\mathbb{E}[X|Y=y] \coloneqq \phi(y) \tag{4.21}$$

Generally, it is too difficult to find the function ϕ, but only in some specific cases we can find ϕ easily. For discrete random variable X and Y,

$$\mathbb{E}[X|Y=y] \coloneqq \phi(y) = \sum_x x f_{X|Y}(x|y) \tag{4.22}$$

and for (absolutely) continuous random variable X and Y,

$$\mathbb{E}[X|Y=y] \coloneqq \phi(y) = \int_{\mathbb{R}} x f_{X|Y}(x|y)\, dx \tag{4.23}$$

where

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_Y(x,y)}{f_Y(y)}, & f_Y(y) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.24}$$

For example, let B be a subset of Ω and let $\mathbf{1}_B : \Omega \to \mathbb{R}$ with support $\{0, 1\}$ be the random variable given by

$$\mathbf{1}_B = \begin{cases} 1, & \omega \in B \\ 0, & \omega \notin B \end{cases} \tag{4.25}$$

4.6 Random Variable Transformation

We discussed about the distributional approach to the transformation of random variables. Now, we shall discuss the calculus-based approach. First, we will deal with an univariate transformation and then extend the concept to a multivariate transformation.

4.6.1 Univariate Transformation

First of all, recall the basic calculus "Integration of Substitution", given by

$$\int_{\phi(a)}^{\phi(b)} f(x)\,dx = \int_a^b f(\phi(t))\,|\phi^i(t)|\,dt \tag{4.26}$$

where $\phi(t) : [a, b] \to 1$. This formula is used to transform one integral into another integral. Therefore, from this formula, we can find the continuous probability distribution $P[Y \in A]$ of which integrand is a density:

$$P[Y \in A] = \int_{\phi^{-1}(A)} f_X(x)\,dx = \int_A f_Y(\phi^{-1}(y))\,|(\phi^{-1})'(y)|\,dy \tag{4.27}$$

4.6.2 Multivariate Transformation

Let us begin the discussion about the case of multivariate transformation in some rigorous sense.

Let G be an open set in $\mathbb{R}^n$ and let $\phi : G \to \mathbb{R}^n$ be continuously differentiable. Suppose ϕ is injective on G and its Jacobian never vanishes then

$$\int_{\phi(G)} f(y)\,dy = \int_G f(\phi(x))\,||J_\phi(x)||\,dx \tag{4.28}$$

where

$$J_\phi(x) = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{bmatrix}. \tag{4.29}$$

4.6.3 Bivariate transformation

A specific example is given by the $\mathbb{R}^2 \to \mathbb{R}^2$ transformation, which is called bivariate transformation.

Let X be a random vector and ϕ be a transformation from $\mathbb{R}^n \to \mathbb{R}^n$ we want to find the distribution of $Y = \phi(X)$. Suppose ϕ is good enough to satisfy any regularity needed to apply the theorem. Note that, from a random vector $Y = (Y_1, Y_2)$, given by

$$\begin{aligned} Y_1 &= \phi_1(X_1, X_2) \\ Y_2 &= \phi_2(X_1, X_2), \end{aligned} \tag{4.30}$$

we can find ϕ^{-1} such that

$$\begin{aligned} X_1 &= \phi_1^{-1}(Y_1, Y_2) \\ X_2 &= \phi_2^{-1}(Y_1, Y_2). \end{aligned} \tag{4.31}$$

Then, by the Change of Variables

$$\begin{aligned} P[Y \in A] &= \iint_{\phi^{-1}(A)} f_X(x_1, x_2)\ dx_1 dx_2 \\ &= \iint_A f_Y(\phi^{-1}(y_1, y_2))\ ||J||\ dy_1 dy_2 \end{aligned} \tag{4.32}$$

where

$$J(x) = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} \tag{4.33}$$

From this specific example, we can understand the following theorem of $\mathbb{R}^n \to \mathbb{R}^n$ generalization.

Theorem 4.6.1. Let $X = (x_1, x_2, \ldots, x_n)$ have joint distributions f. Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and intjective with non-vaninsh Jacobian. Then $Y = \phi(x)$ has density:

$$f_Y(y) = \begin{cases} f_X(\phi^{-1}(y))\ ||(J_{\phi^{-1}}(y))||, & y \in \phi(\mathbb{R}^n) \\ 0 & \text{otherwise} \end{cases} \tag{4.34}$$

□

Example 4.6.2. Let X and Y be i.i.d. normal R.V's with $\mu = 0$ and $\sigma^2 = 1$. What is the joint distribution of $(U, V) = (X + Y, X - Y)$?
Sketch:

1. Define $\phi : (U, V) \to (X, Y)$ and find ϕ^{-1}
2. Evaluate Jacobian $J_{\phi^{-1}}(U, V)$ and it's determinant $||J||$
3. Use the theorem to find $f_{(U,V)}(u, v)$.

Solution: $\frac{1}{\sqrt{4\pi}} \exp\{-u^2/4\} \frac{1}{\sqrt{4\pi}} \exp\{-v^2/4\}$ ■ ◊

4.7 Moment Generating Function

Definition 4.7.1. Let X be a random variable, then

$$M_X(t) = \begin{cases} \sum_i e^{tx_i} f(x_i) & x\text{: distrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx & x\text{: continuous} \end{cases} \tag{4.35}$$

when it exists and call it the *moment generating function* as a function fo t. □

By the definition of mathematical expectation $\mathbb{E}[\cdot]$, MGF can be denoted by

$$M_X(t) = \mathbb{E}_X[e^{tX}] \tag{4.36}$$

and by differentiating the function with respect to t, we obtain,

$$\begin{aligned} M'_X(t) &= \frac{d}{dt}\mathbb{E}[e^{tX}] \\ &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dt} e^{tx} f(x)\, dx \quad \text{(Not always commutable)} \\ &= \mathbb{E}[\frac{d}{dt}(e^{tX})] \\ &= \mathbb{E}[Xe^{tX}] \end{aligned} \tag{4.37}$$

and let take $t = 0$ then,

$$M'_X(0) = \mathbb{E}[X] \tag{4.38}$$

which is called the first moment. Similarly, the second derivative is given by

$$\begin{aligned} M''_X(t) &= \frac{d}{dt} M'_X(t) \\ &= \frac{d}{dt}\mathbb{E}[Xe^{tX}] \\ &= \frac{d}{dt}\int_{-\infty}^{\infty} xe^{tx} f(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dt} xe^{tx} f(x)\, dx \quad \text{(Not always commutable)} \\ &= \mathbb{E}[\frac{d}{dt}(Xe^{tX})] \\ &= \mathbb{E}[X^2 e^{tX}] \end{aligned} \tag{4.39}$$

and when $t = 0$, the second moment is obtained as follows

$$M''_X(0) = \mathbb{E}[X^2] \tag{4.40}$$

In general, the kth derivative is given by

$$M_X^{(k)}(t) = \mathbb{E}[X^k e^{tX}] \tag{4.41}$$

and take $t = 0$, then

$$M_X^{(k)}(0) = \mathbb{E}[X^k] \tag{4.42}$$

which is called the k-th moment of random variable X.

Example 4.7.1. Let X be a random variable, the PDF of which is given by

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad \text{where} \quad 0 < x < \infty,\ \alpha > 0,\ \beta > 0$$

Find MGF of X.
Solution:

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(t)\, dx \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_{-\infty}^{\infty} e^{tx} x^{\alpha-1} e^{-x/\beta}\, dx \\
&= \frac{(\beta/(1-t\beta))^\alpha}{(\beta/(1-t\beta))^\alpha} \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_{-\infty}^{\infty} e^{-x(1-t\beta)/\beta} x^{\alpha-1}\, dx \\
&= \frac{(\beta/(1-t\beta))^\alpha}{\beta^\alpha} \\
&= \left\{ \frac{1}{1-t\beta} \right\}^\alpha
\end{aligned}$$

■

◊

Properties 4.7.2. *MFG has following properties:*

- **Uniqueness:**
  If two random variables have the same MGF, than they have the same distribution.
- **Linear Transformation:**

If $Y = aX + b$, then

$$M_Y(t) = e^{bt} M_X(at).$$

- **Independent sums:**
  For independent random variable, says X and Y, with $M_X(t), M_Y(t)$, respectively, the MGF of $Z = X + Y$ is given by

$$M_Z(t) = M_{X+Y}(t) = M_X(t)M_Y(t) = \iint e^{t(x+y)} f_X(x) f_Y(y)\, dxdy.$$

References

[1] Youngstown State University G. Jay Kerns. *Introduction to Probability and Statistics Using R.* `http://ipsur.org/index.html`. G. Jay Kerns, 2010.

[2] Babatunde A. Ogunnaike. *Random Phenomena: Fundamentals of Probability and Statistics for Engineers.* CRC Press, 2009.

[3] *WIKI, Conditional Expectation, Formal Def.* `https://en.wikipedia.org/wiki/Conditional_expectation`. Accessed: 2017-04-16.

[4] *WIKI, Sigma algebra and random variable.* `https://en.wikipedia.org/wiki/Sigma-algebra`. Accessed: 2017-04-16.