MAT2013: Probability and Statistics [2] [3]	**Spring 2017**

Lecture 2: Random Variables

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

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

2.1 Review

So far, we've learned about how to specify **probability space** from random experiment as follows:

$$(\Omega, \mathcal{A}, \mathbb{P}) \tag{2.1}$$

where Ω is sample space, $\mathcal{A}$ is σ-*algebra*, and $\mathbb{P}$ is probability mass (or dense). We gathered all possible outcomes of random experiment into the sample space, learned how event space is constructed, briefly dealt with the special collection called σ-*algebra* to construct measurable space $(\Omega, \mathcal{A})$. And also, we learned how to assign probabilities on events, defined the probability function $\mathbb{P} : \mathcal{A} \to [0, \infty)$ which completely satisfies Kolmogorov axioms thanks to σ-*algebra*, and learned numerous concepts to deal with the probability function in a mathematical framework. Finally, we got the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Example 2.1.1. From the simple coin tossing experiment, we can construct the probability space as follows

$$\begin{gathered}\Omega = \{H, T\} \\ \mathcal{A} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \\ \mathbb{P} : \mathcal{A} \to [0, 1] \\ \text{with} \\ \mathbb{P}(\emptyset) = 0,\, \mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2},\, \mathbb{P}(\{H, T\}) = 1\end{gathered}$$

$\diamondsuit$

2.2 Random Variable

2.2.1 Mathematical Concept of the Random Variable

But our sample space may quite tedious to describe and inefficient to analyze mathematically if its elements are not numbers. To facilitate mathematical analysis, it is desired to find a means of converting this sample space into one with real numbers. This is achieved via the vehicle of the ***random variable*** defined as follows:

Definition 2.2.1. Given a random experiment with a sample space Ω, let there be a function X, which assigns to each element $w \in \Omega$, one and only one real number $X(w) = x$. This function X is called a random variable.

$$X : \Omega \to \mathbb{R} \tag{2.2}$$

$\square$

In rigorous (measure-theoretic) probability theory, the function X is also required to be measurable (see Appendix. Random Variable). For later use, we will use the definition of random experiment with the concept of probability space which is given by

$$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$$

and also define P_X as

$$\forall B \in \mathcal{B}, \quad P_X(X \in B) \coloneqq \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

on $(\mathbb{R}, \mathcal{B})$ using $\mathbb{P}$ on $(\Omega, \mathcal{A})$. Now, let us look at a few examples to get a better understanding about the concept.

Example 2.2.1. Let's get back to the *Example* 2.1.1, the coin tossing experiment, again.

First, we may define the random variable X as follows

$$\begin{aligned} X : (\Omega, \mathcal{A}) &\to (\mathbb{R}, \mathcal{B}) \\ H &\mapsto 0 \\ T &\mapsto 1, \end{aligned}$$

then we can evaluate the probability of each event in $\mathcal{B}$

$$\begin{aligned} P_X(X = 0) = P_X(X \in \{0\}) &= \mathbb{P}(\{\omega \in \Omega | X(\omega) \in \{0\}\}) = \mathbb{P}(H) = \frac{1}{2} \\ P_X(X = 1) = P_X(X \in \{1\}) &= \mathbb{P}(\{\omega \in \Omega | X(\omega) \in \{1\}\}) = \mathbb{P}(T) = \frac{1}{2} \\ P_X(X \in \mathbb{R} \setminus \{0, 1\}) &= \mathbb{P}(\{\omega \in \Omega | X(\omega) \in \mathbb{R} \setminus \{0, 1\}\}) = \mathbb{P}(\emptyset) = 0 \end{aligned}$$

and then we finally induce the function P_X as follows

$$P_X = \begin{cases} 1/2, & X \in \{0\} \\ 1/2, & X \in \{1\} \\ 0, & \text{otherwise} \end{cases}$$

Therefore, the probability space mapped by random variable X is given by

$$(\mathbb{R}, \mathcal{B}, P_X)$$

$\diamondsuit$

We also may pre-specify the outcomes in natural way to apply a random variable implicitly, to map the "primitive" outcomes into numbers.

Example 2.2.2. Consider the coin tossing experiment again, but now the outcome is the number of H (head) in an experiment. In this example, the outcome implicitly contains the application of random variable which has the form of

$$\begin{aligned} X : (\Omega, \mathcal{A}) &\to (\mathbb{R}, \mathcal{B}) \\ H &\mapsto 1 \\ T &\mapsto 0, \end{aligned}$$

Hence, we can define the probability distribution P_X and so that can define probability space $(\mathbb{R}, \mathcal{B}, P_X)$, directly. $\diamondsuit$

Example 2.2.3. Consider the random experiment that toss a coin three times and the outcome is the "primitive" one.

The number of head is we may Let $X(\omega) = (\# \text{ of } T \in \omega)$ be a random experiment, then

$$\begin{aligned} X(HHH) &= 0 \\ X(HHT) = X(HTH) = X(THH) &= 1 \\ X(HTT) = X(THT) = X(TTH) &= 2 \\ X(TTT) &= 3 \end{aligned}$$

Let $V = \{0, 1, 2, 3\}$ and $\mathcal{F} = 2^V$, then, By X, our probability space $(\Omega, \mathcal{A}, P)$ is mapped to $(V, \mathcal{F}, P_X)$. And we can define

$$P_X = (\{0, 1\}) := P(\{\omega \in \Omega : X(\omega) \in \mathcal{A}\})$$

which is called "induced probability function on $\mathcal{F}$ by X. in short, probability distribution on $\mathcal{F}$." $\diamondsuit$

2.2.2 Induced Probability Function

As we can see in the above *Example* 2.2.1 and 2.2.2, what we want to do with random variable X is to map the sample space Ω into $\mathbb{R}$ (or $\mathbb{R}^n$), so as to deal with the induced probability function P_X on $\mathbb{R}$ where we can take much more advantage, such as numerous existing models (distributions) mathematically well-defined, have been used practically for a long time and therefore qualified successfully.

Formally, we define random variable X, and then map our original probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into the new, advantageous one $(\mathbb{R}, \mathcal{B}, P_X)$.

$$\begin{aligned} X : (\Omega, \mathcal{A}) &\to (\mathbb{R}, \mathcal{B}) \\ \omega &\mapsto x \end{aligned} \tag{2.3}$$

Then, we can define the probability function P_X as follows:

$$\forall B \in \mathcal{B}, \quad P_X(X \in B) \coloneqq \mathbb{P}(\{w \in \Omega : X(w) \in B\}) \tag{2.4}$$

which is called "induced probability function P_X on $\mathcal{B}$ by X." In short, ***probability distribution*** on $\mathcal{B}$ [1].

2.3 Probability Distribution P_X

Suppose that we model a random experiment so as to define probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and that X is a random variable for the experiment taking values in a set $\mathbb{R}$. (Not required to choose $\mathbb{R}$. It's just specific example to make our understanding clear.) Any random variable X for the experiment defines a new probability space $(\mathbb{R}, \mathcal{B}, P_X)$. Now, we will focus on the probability measure (or probability distribution) P_X of X and how they look like so that we will categorize the random variable X.

2.3.1 Classification of Random Variable [4]

A random variable X is called ***continuous*** if and only if the probability measure P_X of the random variable X satisfies $P_X(X = x) = 0$ for each x. By contrast, a ***discrete*** random variable is one that has a finite or countably infinite set of possible values x with $P_X(X = x) > 0$ (i.e., $\exists x_n$ s.t. $P_X(X = x_n) > 0$). Continuous random variable and discrete random variable are dichotomy. [5]

2.3.2 Discrete Random Variable and PMF

For every discrete random variable X, we define a function, ***probability mass function*** (PMF), namely f_X, defined by

$$f_X(x) = \begin{cases} P_X(X = x), & \forall x \in R(X) \\ 0, & \forall x \in \mathbb{R} \setminus R(X) \end{cases} \tag{2.5}$$

Example 2.3.1. **Coin-tossing**

$$\Omega = \{H, T\}$$
$$\mathcal{A} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$$
$$P(\emptyset) = 0, P(\{H\}) = P(\{T\}) = 1/2, P(\{H, T\}) = 1$$

Define Random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ as $X(H) = 0, X(T) = 1$, then a new probability space is given by

$$(\mathbb{R}, \mathcal{B}, P_X)$$
$$\text{where}$$
$$\forall B \in \mathcal{B}, \quad P_X(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

$P_X(B)$ is well defined because X is measurable function satisfying

$$X^{-1}(B) \in \mathcal{A} \Rightarrow B \in \mathcal{B}$$

Then eventually, we can define ***Probability Mass Function*** f_X as

$$f_X(x) = \begin{cases} 1/2, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

$\Diamond$

Properties 2.3.2. Naturally, all PMF's satisfy

1. $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\sum_{x \in \mathbb{R}} f_X(x) = 1$
3. $P_X(X \in B) = \sum_{x \in B} f_X(x), \quad \forall B \in \mathcal{B}$ (σ additivity)

Example 2.3.3. **Coin tossing 3 times**
Consider the random experiment that toss a coin three times, then the sample space Ω is consisted of the outcomes.

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

Find the probability mass function f_X. **D.I.Y.** $\Diamond$

2.3.3 Continuous Random Variable and PDF

A random variable X is said to be an absolutely continuous random variable iff there is a nonnegative Borel measurable function f_X, which is called ***probability density function*** (PDF) of X, such that

$$P_X(X \in B) = \int_B f_X d\mu \quad \text{where} \quad B \in \mathcal{B} \tag{2.6}$$

where μ is a measure which is a function assigining non-negative numbers to measurable subset of A in a way that is additive (i.e., $\mu(A_1 \dot\cup A_2 \dot\cup \cdots) = \mu(A_1) + \mu(A_2) + \cdots$).

Properties 2.3.4. All PDF satisfies

1. $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f_X(x)\mathrm{d}x = 1$
3. $P_X(X \in B) = \int_{x \in B} f_X(x)\mathrm{d}x, \quad \forall B \in \mathcal{B}$ (σ additivity)

Example 2.3.5. Suppose a number is selected at random from $[0,1]$ and let X denote the number obtained. Then the probability space is given by

$$\begin{aligned} \Omega &= \mathbb{R} \quad \text{(Real number)} \\ \mathcal{A} &= \mathcal{B} \quad \text{(borel } \sigma \text{ algebra)} \end{aligned}$$

and

$$\begin{aligned} &\forall B \in \mathcal{B}, \qquad P_X(X \in B) = P_X(X \in B \cap [0,1]) \\ &\forall k \in [0,1], \qquad P_X(X \in [0,k]) = k \end{aligned}$$

Then by definition, we get the probability density function f_X satisfying

$$P_X(X \in B) = \int_{B \cap [0,1]} f_X \text{ dx}$$

then,

$$f_X = 1$$

$\Diamond$

Example 2.3.6. Is there any random variable that is neither discrete nor absolutely continuous? Yes, Cantor function. $\Diamond$

2.3.4 Distribution Function, $F(\cdot)$

Of importance also is a different, but related, funciton, $F(x)$, defined as:

$$F(x) = P(X \leq x) \tag{2.7}$$

the probability that the random variable X takes on values less than or equal to x. In general, we can use the equation for any arbitrary $b \geq a$,

$$P(a \leq X \leq b) = F(b) - F(a) \tag{2.8}$$

Properties 2.3.7. **All Distribution Function satisfies**

- F_X is monotone increasing (not strictly increasing)
- F_X is right continuous
- $\lim_{x\to\infty} F_X(x) = 1$
- $\lim_{x\to-\infty} F_X(x) = 0$

Note that the distribution function $F(x)$ is actually the more fundamental function for determining probabilities. This is because, regardless of whether X is continuous or discrete, $F(\cdot)$ can be used to determine all desired probabilities.

Example 2.3.8. For given probability distribution function $f_X(\cdot)$, find ditribution function $F_X(\cdot)$

$$f_X(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

1. $x < 0$;

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^{x} f_X(t)\mathrm{d}t \\ &= \int_{-\infty}^{x} 0\mathrm{d}t = 0 \end{aligned}$$

2. $x \in [0,1]$;

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^{x} f_X(t)\mathrm{d}t \\ &= 0 + \int_{0}^{x} f_X(t)\mathrm{d}t = \int_{0}^{x} 1\mathrm{d}t \\ &= t|_0^x = x \end{aligned}$$

3. $x > 1$;

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^{x} f_X(t)\mathrm{d}t \\ &= 0 + \int_{0}^{1} f_X(t)\mathrm{d}t + \int_{1}^{x} 0\mathrm{d}t \\ &= t|_0^1 + 0 \\ &= 1 \end{aligned}$$

Then the distribution function $F(\cdot)$ is given by

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

◊

2.3.5 Again, Classification of Random Variable

X is discrete random variable iff F_X is discrete. On the contrary, X is continuous iff F_X is continuous.

2.4 The Fundamental Tools

In summary, the PDF $f(\cdot)$ and the CDF $f(\cdot)$ indicate how the probabilities of events arising from the random phenomenon are distributed over the entire space of the associated random variable X.

That is, with the PDF $f(\cdot)$ and the CDF $f(\cdot)$ we can charaterize the "behavior" of an assicated random variable, say X, of interest from the ensemble aggregate of all possible outcomes. Therefore, the PDF $f(\cdot)$ and the CDF $f(\cdot)$ is a description that can be used to analyze the particular random phenomenon.

However, it may be exhausting and burdensome to develop probabilistic model:

1. clarify sample space and event space
2. assign probabilities to each outcomes with proper assumptions
3. set the random variable to map the exisitng probability space into new one
4. find probability distribution
5. evaluate probability distribution function and distribution function

Fortunately, it turns out that specific (but most) random phenomenon can be generalized for some specific ***class of random phenomenon*** charaterized with sound assumptions which is mathematically well-defined. Such functions provide convenient and compact mathematical representations of the desired ensemble behavior of random variables; they constitute the centerpiece of the probabilistic framework —the fundamental tool used for analyzing random phenomena [3].

2.5 Mathematical Expectation

2.5.1 Expected Value

Definition 2.5.1. Let X be a random variable and g be a function. Then, the expectation of $g(x)$, denoted $E[g(x)]$, is defined by

$$E[g(x)] = \int_\Omega g(x)\,\mathrm{d}\mathbb{P} = \begin{cases} \sum_x g(x) f_X(x) & \text{(discrete)} \\ \int_{\mathbb{R}} g(x) f_X(x) \mathrm{d}x & \text{(continuous)} \end{cases} \tag{2.9}$$

provided the integral exists. □

Properties 2.5.1. Expectation satisfies

- $\mathbb{E}[ag_1(X) + bg_2(X) + c] = a\mathbb{E}[g_1(x)] + b\mathbb{E}[g_2(x)] + c$
- If $g_1(x) \geq 0\, \forall x$, then $\mathbb{E}[g_1(X)] \geq 0$.
- If $g_1(x) \geq g_2(x)\, \forall x$, then $\mathbb{E}[g_1(X)] \geq \mathbb{E}[g_2(X)]$
- If $a \leq g_1(x) \leq b\, \forall x$, then $a \leq \mathbb{E}[g_1(X)] \leq b$.

Definition 2.5.2. Let X be a random varialbe with finite expectation, then we define the mean of X, denoted by $E[X]$, which is given by

$$\mathbb{E}[X] = \int_\Omega X \mathrm{d}\mathbb{P} = \begin{cases} \sum_x x f_X(x) & \text{(discrete)} \\ \int_\mathbb{R} x f_X(x) \mathrm{d}x & \text{(continuous)} \end{cases} \tag{2.10}$$

□

Definition 2.5.3. Let X be a random variable with finite expectation, then we define the variance of X, denoted by $Var(X)$, which is given by

$$\sigma^2 = Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] \tag{2.11}$$

□

Properties 2.5.2. Variance satisfies

- $V[aX + b] = a^2 Var[X]$
- $Var(X) = \mathbb{E}[X^2] - \mu^2$

2.6 Random Variable Transformation

Many problems of practical interest involve a random variable Y that is defined as a function of another random variable X, say according to $Y = \phi(X)$, so that the characteristics of the one arise directly from those of the other via the indicated transformation. In particular if we already know the probability distribution function for X as $f_X(x)$, it will be helpful to know how to determine the corresponding distribution function for Y.

Example 2.6.1. Suppose we know the distribution of a given X and want to find the distribution of Y which is defined as a function of X.

$$Y \coloneqq \phi(X) \quad \text{where} \quad \phi(X) = X^n$$

The probability space induced by X is given by $(\mathbb{R}, \mathcal{B}, P_X)$ where $P_X \coloneqq \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ for any $B \in \mathcal{B}$. And X is uniformly distributed on $(0, 1)$ having the probability distribution function $f_X(x)$ given by

$$f_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & otherwise. \end{cases} \tag{2.12}$$

Find the distribution of Y.

Solution: At first, consider the mapping from primitive probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into the new one induced by Y, which is the form of $(\mathbb{R}, \mathcal{B}, P_Y)$, where the probability distribution for any $B \in \mathcal{B}$ is defined as

$$P_Y[Y \in B] \coloneqq \mathbb{P}[\{\omega \in \Omega : Y(\omega) \in B\}].$$

Then, we want to find the distribution function $F_Y(y)$ given by

$$F_Y(y) = P_Y(Y \leq y) = P_Y[Y \in (-\infty, y)] = \mathbb{P}[\{\omega \in \Omega : Y(\omega) \in (-\infty, y)\}]$$

But, we don't have any knoweledge about random variable Y but only the function $Y = \phi(X)$. Hence, we will substitute X with Y using the function.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}[\{\omega \in \Omega : Y(\omega) \in (-\infty, y)\}] \\
&= \mathbb{P}[\{\omega \in \Omega : Y(\omega) \leq y\}] \\
&= \mathbb{P}[\{\omega \in \Omega : X^n(\omega) \leq y\}] \\
&= \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq y^{1/n}\}]
\end{aligned}$$

Note that, since the funciton $\phi^{-1}(Y) = Y^{1/n}$ is strictly increasing function, the inequality sign is not change. Likewise, if the inverse function has monotonous property, then we only need to maintain (for increasing) or reverse (for decreasing) the sign of the inequality operator. In general, i.e., the inverse function has no such property, we need to determine the transformation more precisely.

Now, we can use the definition of P_X as follows

$$\begin{aligned}
F_Y(y) &= \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq y^{1/n}\}] \\
&= P_X[X \leq y^{1/n}] \\
&= F_X(y^{1/n})
\end{aligned}$$

From the probability distribution function $f_X(x)$, we can evaluate the distribution function $F_X(x)$, given by

$$F_X(x) = \begin{cases} 0, & -\infty < x \leq 0, \\ x, & 0 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

Then the distribution function $F_Y(y)$ is given by

$$F_Y(y) = \begin{cases} 0, & -\infty < y \leq 0, \\ x^{1/n}, & 0 < y < 1, \\ 1, & y \geq 1. \end{cases}$$

■

◊

Example 2.6.2. Suppose we know the distribution of a given X and want to find the distribution of Y which is defined as a function of X.

$$Y \coloneqq \phi(X) \quad \text{where} \quad \phi(X) = X^n$$

The probability space induced by X is given by $(\mathbb{R}, \mathcal{B}, P_X)$ where $P_X := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ for any $B \in \mathcal{B}$. And X is uniformly distributed on $(0,1)$ having the probability distribution function $f_X(x)$ and distribution function $F_X(x)$, which is given by

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

and

$$F_X(x) = \int_{-\infty}^{x} f_X \mathrm{d}x$$

What is the distribution of Y? **D.I.Y.** $\Diamond$

2.7 Appendix

2.7.1 Random Variable

Definition 2.7.1. A function $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ is called a random variable iff

$$\forall B \in \mathcal{B} \quad \text{s.t.} \quad X^{-1}(B) \in \mathcal{A} \tag{2.13}$$

and the probability distribution on $(\mathbb{R}, \mathcal{B})$ can be defined as follows:

$$P_X(X \in B) := \mathbb{P}(\{w \in \Omega : X(w) \in B\}) \quad \forall B \in \mathcal{B} \tag{2.14}$$

$\square$

Details.
Random variables can be defined in a more rigorous manner using the terminology of measure theory. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let X be a function $X : \Omega \to \mathbb{R}$. Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-*algebra* of $\mathbb{R}$ (i.e., the smallest σ-*algebra* containing all the open subsets of $\mathbb{R}$. If, for any $B \in \mathcal{B}(\mathbb{R})$,

$$\{w \in \Omega : X(w) \in B\} \in \mathcal{A} \tag{2.15}$$

then X is a random variable on Ω. As a consequence, if X satisfies the above property, then for any $B \in \mathcal{B}(\mathbb{R})$, $P_X(X \in B)$ can be defined as follows:

$$P_X(X \in B) := \mathbb{P}(\{w \in \Omega : X(w) \in B\}) \tag{2.16}$$

where the probability on the right hand side is well-defined because the set $\{w \in \Omega : X(w) \in B\}$ is measurable by the very definition of random variable.

References

[1] *distinguishing probability measure, function, distribution.* `http://math.stackexchange.com/questions/1073744/distinguishing-probability-measure-function-distribution`. Accessed: 2017-04-05.

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

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

[4] *Probability density function vs. probability mass function.* `http://math.stackexchange.com/questions/23293/probability-density-function-vs-probability-mass-function`. Accessed: 2017-04-05.

[5] *Probability measure.* `http://www.math.uah.edu/stat/prob/Probability.html`. Accessed: 2017-04-05.