

Lecture Notes on Probability and Statistics

*for BSc Health Data Science, KMUTT community
who would like to know more advanced
than our taughts*

Kittapat Ratanaphupha
Last update on 12/12/2022 at 15:14

Contents

1	Axioms on Probability	2
1.1	What is a Random Variable?	3
1.2	Measure Spaces	3
1.2.1	How can we measure things mathematically?	3
1.2.2	Probability spaces	6
1.3	Formal Definition on a Random Variable	7
1.3.1	Why we would like to mention a random variable formally? . . .	7
1.3.2	Necessary functions for interpreting the probability and statistics	7
1.3.3	Examples and counterexamples for random variables	8
1.4	What is a Probability Distribution?	8
1.5	Lebesgue Integration and Its Consequences	9
1.5.1	Definition of Lebesgue integration	9
1.5.2	Almost everywhere and almost surely properties	9
1.5.3	Convergence theorems	9
1.5.4	Fubini's theorem	9
1.5.5	Radon-Nikodym theorem	9
1.6	Formal Definition of Probability Distribution	9
1.7	Expected Value	9
1.8	Conditional Probability and Bayes' Rule	9
1.9	Interpretation of Probability	9

Chapter 1

Axioms on Probability

1.1 What is a Random Variable?	3
1.2 Measure Spaces	3
1.2.1 How can we measure things mathematically?	3
1.2.2 Probability spaces	6
1.3 Formal Definition on a Random Variable	7
1.3.1 Why we would like to mention a random variable formally?	7
1.3.2 Necessary functions for interpreting the probability and statistics	7
1.3.3 Examples and counterexamples for random variables	8
1.4 What is a Probability Distribution?	8
1.5 Lebesgue Integration and Its Consequences	9
1.5.1 Definition of Lebesgue integration	9
1.5.2 Almost everywhere and almost surely properties	9
1.5.3 Convergence theorems	9
1.5.4 Fubini's theorem	9
1.5.5 Radon-Nikodym theorem	9
1.6 Formal Definition of Probability Distribution	9
1.7 Expected Value	9
1.8 Conditional Probability and Bayes' Rule	9
1.9 Interpretation of Probability	9

This lecture is for revising *Probability and Statistics* related courses concerning with Health Data Science program, KMUTT-PCCMS. Here is a list of courses which are related with:

- MTH 10x: Mathematics X (Prerequisites for all math courses)
- CHHD 301: Biostatistics
- CPE 353: Design of Experiments

Hereinafter, we will discuss about a basic concept of probability and statistics ~~without mentioning any rigorous mathematical analytic approaches~~. The lecture we will follow for this chapter was provided to everyone before we have started this.

1.1 What is a Random Variable?

A **random variable** is a mathematical formalization of a quantity or object which depends on random events. — *Wikipedia English*

As we say, random variable,

1. **Random** is some kind of doings which cannot be determined exactly from time to time. The result will be given *by chances*
2. **Variable** is some kind of a placeholder which can be more than one possibility *by assumption*.

Then, when we say about a random variable is then a kind of a placeholder of something which can be varied *by chances*, but how do we measure its chances concerning the event.

First, we list a set of possibilities into numbers by focusing a relation between its event to transform the possibility into an encoded number. For example, in the event that we would like to toss a coin, a set of possibilities will be like {Head, Tail, Middle}. We can do transformation a set of possibilities to be mathematical *formalized* as $\{-1, 0, 1\}$ respectively.

As we see by the previous example,

- {Head, Tail, Middle} is a sample space.
- $\{-1, 0, 1\}$ is a space which we can measure the value (which will be later called *probability*.)

In simple manner, a random variable is a function which transforms an event from sample space into some space that we can measure its probability.¹

We see that, it is necessary to construct a framework before we discuss *a random variable by formally*. The mathematical tool for handling this is called **measure theory with the Kolmogorov axioms**.

1.2 Measure Spaces

We will have some doubt about this mathematical approach for probability and statistics that how can this overcome the informal description for random variables or any terminology in probability and statistics. We will show then by the following.

1.2.1 How can we measure things mathematically?

First, we will discuss and pursue everyone to believe/not believe a sense of measurement.

¹Formally, a random variable is a measurable function between a sample space which is a σ -algebra into a measurable space which we will discuss later term by term.

Checklist for giving some intuition sense on measurement a volume which we will be interested in a specific region.

- ☐ We know that how to measure a such region that how volume is (if we could.)
- ☐ We can measure nothing by doing nothing.
- ☐ If we can measure an object, we can measure its compliment by subtract its value from the value we have measured a such region.
- ☐ If we can measure a set of objects, we can measure the combined object.
- ☐ If we can measure a set of objects, we can measure the mutual things that are included in all objects we are interested.

Then, we transcribe it mathematically like this.

Definition 1.1 (σ -algebra on X). *A collection \mathfrak{M} of subsets of a set X is called a σ -algebra if*

- $X \in \mathfrak{M}$
(it is as we know that a such region is measurable.)
- if $A \in \mathfrak{M}$, then $A^c \in \mathfrak{M}$.
(it is as if we can measure an object, we can measure its compliment.)
- if $A_1, A_2, \dots \in \mathfrak{M}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{M}$.
(it is as if we can measure a set of objects, we can measure its combined object.)

By then, we can imply that:

Remark 1.1.1. $\emptyset \in \mathfrak{M}$ (it is that we can measure nothing by doing nothing.)

Remark 1.1.2. if $A_1, A_2, \dots \in \mathfrak{M}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathfrak{M}$ (it is that if the set of objects is measurable by each object, then we can measure the mutual things that are included in all objects we are interested.)

Remark 1.1.3. If $A, B \in \mathfrak{M}$, $A \setminus B \in \mathfrak{M}$.

Proof. Left blank intentionally □

How to know that how big is σ -algebra
(for who are interested in measure theory.)

1. Say that if we have $D = \{A, B, C\}$ as a set of measurable sets, so what is a smallest σ -algebra generated by D then?
2. If we say X is a set, $\mathcal{P}(X)$ can be a σ -algebra on X .
3. Also, suppose that X is a set, $\{X, \emptyset\}$ is a σ -algebra on X .
4. The σ -algebra which is generated from a collection of open sets on a topological space X is called a *Borel σ -algebra* on X .

After this, we will call a pair (X, \mathfrak{M}) as a measurable space as each symbol is defined as in Definition 1.1.

The second component, a measure, will be discussed by, again, some checklist.

Checklist for giving some intuition sense on a value of measurement

- If we know that a value of measurement of objects such that they pairwise disjoint, a value of measurement for the combined object is then a summation of such each object.
- A value of measurement on an empty set must be zero.

By formally, the measure can be defined as:

Definition 1.2 (Measure on a measurable space (X, \mathfrak{M})). *Let (X, \mathfrak{M}) be a measurable space. A measure on \mathfrak{M} is a function $\mu : \mathfrak{M} \rightarrow [0, \infty]$ such that*

- $\mu(\emptyset) = 0$
(A value of measurement on an empty set is zero.)
- for any countable collection of disjoint sets $\{A_n\}_{n=1}^{\infty}$ in \mathfrak{M} then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

(A value of measurement on its union of disjoint objects is a summation of a value of measurement on each object.)

Note that, we will call (X, \mathfrak{M}, μ) as a measure space. By this, here are consequences for this definition,

Remark 1.2.1. *Let (X, \mathfrak{M}, μ) be a measure space.*

- If $A_1, A_2, \dots, A_n \in \mathfrak{M}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$$

- If $A, B \in \mathfrak{M}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$; if $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.
- If $A_n \in \mathfrak{M}$ for each $n \in \mathbb{N}$, then

$$- \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

- If $A_1 \subseteq A_2 \subseteq \dots$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$$

- If $A_1 \supseteq A_2 \supseteq \dots$,

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$$

Proof. Left blank intentionally

□

The last component, a measurable function, in layman's term, is a conversion of measurement.

Definition 1.3 (Measurable function between two measure spaces). *Let (X, \mathfrak{M}) and (Y, \mathfrak{N}) be measurable spaces. A function $f : X \rightarrow Y$ is said to be $(\mathfrak{M}, \mathfrak{N})$ -measurable if $f^{-1}(E) \in \mathfrak{M}$ for any $E \in \mathfrak{N}$. Informally, we can simply call f as a measurable function if X, Y are well-understood.*

1.2.2 Probability spaces

In probability, there is an important concept that is a probability space. As we were in high school, a definition of a probability space is well-defined as we will say that.

- a probability of whole sample space is 1.
- There are no events such that its probability is more than 1 or less than 0.
- The separated events have probability equal to their probability on each event combined.

As we discussed about some concept on the measure theory, we will call this a probability space formally by the following definition.

Definition 1.4 (Probability space; Kolmogorov Axioms). *Let (Ω, F, P) be a measure space as said to be a probability space if,*

- (Unitary axiom) $P(\Omega) = 1$
- (Non-negativity) if $E \in F$, then $P(E) \geq 0$.
- (σ -additivity) if $E_n \in F$ for each $n \in \mathbb{N}$ and $E_i \cap E_j = \emptyset$ for each $i \neq j$, then

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n).$$

So, its consequences can be written as this,

Remark 1.4.1. *Let (Ω, F, P) be a probability space.*

- (Monotonicity of probability measure) if $A, B \in F$ and $A \subseteq B$, then $P(A) \leq P(B)$.
- (Probability of no event) $P(\emptyset) = 0$
- (Probability of compliment event) If $A \in F$, then $P(A^c) = 1 - P(A)$
- (Probability measure has a bounded image) For each $A \in F$, $0 \leq P(A) \leq 1$.

Proof. Left blank intentionally

□

1.3 Formal Definition on a Random Variable

Definition 1.5 (Random variable on (Ω, F, P) into (E, \mathfrak{M})). Let (Ω, F, P) be a probability space, says a sample space, and (E, \mathfrak{M}) be a measure space.

A random variable is a function $X : \Omega \rightarrow E$ such that a probability of X in S which $S \in \mathfrak{M}$ is defined as

$$P(X \in S) = P(\{\omega \in \Omega | X(\omega) \in S\}).$$

In layman's term, the probability of random variable in a certain set is equal to the probability of event in the sample space which maps to the measure space that is included in the interested set.

Interesting note: consider an event of rolling a die, there are values $\{1, 2, 3, 4, 5, 6\}$ which is easily determined that is a natural number. When we interpret a probability, we do not think a random variable as X directly, but its inverse function since we assign a probability in a sample space $\Omega = \{\square, \square, \square, \square, \square, \square\}$.

For example, $P(X < 3)$ is said to be a probability that a die shows a number that is less than 3. It means $P(X < 3) = P(\{\square, \square\})$. For instance, $X^{-1}(1) = \square$. Say that it is an abstraction of sample space which can use mathematical tools to analyze it.

1.3.1 Why we would like to mention a random variable formally?

- There are theorems in measure theory that provide well-definition of many familiar terminologies, such as, probability distribution, expected value, joint distribution, etc.
- Major theorems in probability and statistics are derived from measure-theoretic approach, e.g. the Laws of large number, the Central limit theorem.
- Not even in fundamental theorems and terminology we mentioned, in advanced concepts on statistics, i.e. sampling theory (bias-variance tradeoff, unbiased estimator as examples), high-dimensional statistics (which is a crucial base for machine learning in dimensional reduction tasks) and statistical learning theory (which is a fundamental of machine learning in classification and regression tasks), also using this definition to derive elegant properties in them.

1.3.2 Necessary functions for interpreting the probability and statistics

Definition 1.6 (Indicator function). Let X, A be sets such that $A \subseteq X$. Define $\mathbf{1}_A : X \rightarrow \{0, 1\}$ be an indicator function of A that,

$$\mathbf{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Definition 1.7 (Simple function). Let (X, \mathfrak{M}) be a measurable space. Suppose that $A_1, A_2, \dots, A_k \in \mathfrak{M}$. Then $f : X \rightarrow \mathbb{C}$ is a simple function such that

$$f(x) = \sum_{i=1}^k a_i \mathbf{1}_{A_i}(x)$$

where $\{a_i\}_{i=1}^k$ is a sequence in \mathbb{C} .

1.3.3 Examples and counterexamples for random variables

Here are the examples and counterexamples to show that are random variables or not, but we have to verify these before affirm the conclusion.

- Let X be a measurable function such that $\text{im } X = C$ which that $|C| \leq |\mathbb{N}|$ (set theory: C is countable².) For example, $X : \Omega \rightarrow \{1, 5, 9\}$ where Ω is a sample space in a probability space (Ω, F, P) . X is called *discrete random variable*.
- Let X be a measurable function such that $X : \Omega \rightarrow C$ as the same as the above bullet, but C is uncountable infinite.³ We will call then *continuous random variable*. E.g. $X : \Omega \rightarrow \mathbb{R}$ or $X : \Omega \rightarrow [0, 1]$ as if we interpret X as a bijective function.
- Let $X(\omega) = \mathbf{1}_A(\omega) - \mathbf{1}_{A^c}(\omega)$ where A is a non-measurable set. Say that X is not a random variable.
- Let X be the same definition as the last bullet. $|X|$ is a random variable.

Proof. Leave it blank intentionally. □

1.4 What is a Probability Distribution?

A **probability distribution** is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.

— *Wikipedia English*

In simple manner (which means in undergraduate class), a probability distribution is a function from random variable to some real number which we will discuss later that it can be extended to an extended real number. To be simpler, a probability distribution is a probability assignment for each representative value of event in sample space which probability that happening at least an event is equal to 1.

As in undergraduate course, a probability distribution can be classified as two types corresponding to discrete and continuous variables, *probability mass function* and *probability distribution function* respectively which are defined by,

Let $X : (\Omega, F, P) \rightarrow (E, \mathfrak{M})$ be a random variable.

1. Probability mass function: say that X is a discrete random variable. $p_X : E \rightarrow [0, 1]$ is a probability mass function (PMF) as defined as

$$p_X(x) = P(X = x)$$

$$\text{as } \sum_{x \in E} p_X(x) = 1.$$

²say that C is countable, by set theoretically, by the existence of an injective function between C into \mathbb{N} .

³says that there are no injective functions $f : C \rightarrow \mathbb{N}$ by the opposite of definition of countable set in the set theory in cardinality.

2. Probability density function: say that X is a continuous random variable. $f_X : E \rightarrow [0, \infty]$ is a probability distribution function (PDF) as defined as

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

as $\int_E f(x) dx = 1$.

but there is a set of restriction such that not every function which has domain and range as described above be any probability distribution, but Lebesgue integration and its important theorems must be mentioned which will be discussed in the next section as the following.

- Definition of Lebesgue integration
- Almost everywhere and almost surely properties
- Convergence theorems for Lebesgue integration
- Fubini's theorem
- Absolutely continuous on measures and Radon-Nikodym theorem

1.5 Lebesgue Integration and Its Consequences

1.5.1 Definition of Lebesgue integration

1.5.2 Almost everywhere and almost surely properties

1.5.3 Convergence theorems

1.5.4 Fubini's theorem

1.5.5 Radon-Nikodym theorem

1.6 Formal Definition of Probability Distribution

1.7 Expected Value

1.8 Conditional Probability and Bayes' Rule

1.9 Interpretation of Probability