Notes for Probability and Stochastic Process

©Author: Zhan Cheng E-Mail: zhanc113003@gmail.com Textbook: ISBN 978-7-5635-4537-7

Chapter 1:Events and Their Probabilities

1.1 Experiment, Sample Space and Random Event

1.1.1 Basic Definations

Random experiment has three characteristics: Reapeatability, Predictability, Uncertainty. And it can be denoted by E , here is an example:

$$E_1: \text{Determintation of the sex of a newborn child.}$$

Each possible outcome is called an ***sample point*** or ***elementary event*** ω . The set of all possible outcomes of E is known as the ***sample space*** Ω . Here is an example about Ω_1 corresponding to E_1:

$$\text{For } E_1\text{: } \Omega_1 = \{g, b\}, \text{ where the outcome } g \text{ means that the child is a girl and } b \text{ that it is a boy}$$

Any subset of the Ω is known as an ***random event*** or ***event*** $A, B, C, \cdots$. We say that A occurs when the outcome of the experiment lies in A. Those events must occur in the experiment are called the ***inevitable events*** S. Those could not happen anytime are said to be ***impossible events*** $\emptyset$.

1.1.2 Events as Sets

The relationships and operations between random events could be described in term of ***set theory***.

Let Ω be the sample space of the random experiment E and $A, B, A_i (i = 1, 2, \cdots)$ be the random events of E.

1. $A \subset B$: if event A occurs, then B occurs
2. $A \cup B$: either A or B occurs

3. $A \cap B$ or AB : both A and B occur
4. $A - B$: A occurs but B does not occur

Common operations of the events are not listed, but there are two important formulas to remember:

$$A - B = A - AB = A\overline{B}$$
$$A \cup B = A \cup \overline{A}B$$

1.2 Probabilities Defined on Events

1.2.1 Classical Probability

A random experiment E is ***classical*** if:

(i) : E contains only different limited basic events.

(ii) : all outcomes are equally likely to occur.

If $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, we define the probability of event A as

$$P(A) = \frac{\#A}{\#\Omega}$$

where $\#$A means the number of all possible outcomes of event A, $\#\Omega$ means the number of all possible outcomes of sample space

and $P(\emptyset) = 0$

For classical random experiment E, the probability has the following properties:

(1) for every event $A, P(A) \geq 0$,

(2)$P(\Omega) = 1$,

(3)for every finite sequence of n disjoint events $A_1, A_2, \cdots, A_n$,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

We just skip that proof.

1.2.2 Geometric Probability

A random experiment E is ***geometric*** if:

(i) : the sample space is a measurable region, i.e. $0 < L(\Omega) < \infty$, and

(ii) : the probability of every event has nothing to do with its position and shape.

In this case, we define the probability of event A as

$$P(A) = \frac{L(A)}{L(\Omega)}$$

and $P(\emptyset) = 0$

For geometrical random experiment E, the probability has the following properties:

(1) for every event $A, P(A) \geq 0$,

(2)$P(\Omega) = 1$,

(3)for every finite sequence of n disjoint events $A_1, A_2, \cdots, A_n$,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

We skip that proof as well.

1.2.3 The Frequency Interpretation of Probability

Let $f_n(A)$ be the times that A occurs. The ration

$$F_n(A) = \frac{f_n(A)}{n}$$

is said to be the ***frequency*** of event A in the n trials.

If n is large enough, the probability of event A will be approximated by $F_n(A)$

For a random experiment E, the frequency has the following properties:

(1) for every event $A, F_n(A) \geq 0,$

(2)$F_n(\Omega) = 1,$

(3)for every finite sequence of n disjoint events $A_1, A_2, \cdots, A_n,$

$$F_n(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} F_n(A_i).$$

We skip the part of σ-algebra because it is too hard 😳

1.3 Conditional Probabilities

1.3.1 The Defination of Conditional Probability

Given two events A and B with $P(B) > 0$, the ***conditional probability of*** A ***given*** B is

$$P(A|B) = \frac{P(AB)}{P(B)}$$

If $P(B) > 0,$ then $P(A|B)$ is also a probability, that is

(1) for every event $A, P(A|B) \geq 0,$

(2)$P(\Omega|B) = 1,$

(3)for every finite sequence of countable disjoint events $A_1, A_2, \cdots, A_n,$

$$P(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B).$$

Some fomulas:

$$P(\overline{A}|B) = 1 - P(A|B)$$
$$\text{if } A \subset B, \text{ then } P(C - A|B) = P(C|B) - P(A|B) \text{ and } P(A|B) \leq P(C|B)$$
$$P(A \cup C|B) = P(A|B) + P(C|B) - P(AC|B)$$

1.3.2 The Multiplication Rule

Assume that $P(B) > 0$. Then

$$P(AB) = P(B) \cdot P(A|B)$$

Or if $P(A) > 0$, Then

$$P(AB) = P(A) \cdot P(B|A)$$

Suppose that $A_1, A_2, \cdot\cdot\cdot, A_n$ are events satisfying $P(A_1A_2 \cdots A_{n-1}) > 0$.Then

$$P(A_1A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$$

1.3.3 Total Probability Formula

Let Ω denote the sample space of some experiment. n events $P(B_1B_2 \cdots B_n)$ are said to form a ***partition*** of Ω if these events satisfy:

(1)$B_1, B_2, \cdots, B_n$ are disjoint and

(2)$\bigcup_{i=1}^{n} B_i = \Omega$

Suppose that the events $B_1, B_2, \cdot\cdot\cdot, B_n$ form a partition of the sample space Ω and $P(B_i) > 0$ for $i = 1, 2, \cdots, n$. Then, for every event A in Ω

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

1.3.4 Baye's Theorem

Let the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω such that $P(B_i) > 0$ for $i = 1, 2, \cdots, n$, and let A be an event such that $P(A) > 0$. Then for $i = 1, 2, \cdots, n$,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

1.4 Independence of Two Events

1.4.1 Independence of Two Events and Several Events

Two events A and B are ***independent*** if

$$P(AB) = P(A)P(B)$$

Three events A, B and C in the sample space S of a random experiment are said to be ***mutually independent*** if

$$\begin{gathered} P(AB) = P(A)P(B) \\ P(BC) = P(B)P(C) \\ P(AC) = P(A)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{gathered}$$

Events A, B and C are said to be ***pairwise independent*** if the first three equations hold.

1.4.2 Bernoulli Trials

A ***Bernoulli experiment*** E is such kind of random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, mainly, *success* and *failure*

A sequence of Bernoulli trials E_n occurs when a Bernoulli experiment is *performed serveral independent times so that the probability of success, say, p, remains the same from trial to trial*

The outcomes of E_n are the 2^n sequences of lenth n. The number of outcomes of E_n that contain a exactly k times is given by the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The probability that the outcome of an experiment that consists of n Bernoulli trials has k success and $n-k$ failures is given by:

$$P_n(k) = \binom{n}{k} p^k q^{n-k}$$

Chapter 2: Random Variable

2.1 The Definition of a Random Variable

A ***random variable*** X is a function that assgins a real number to each outcome in the sample space Ω, that is

$$X : X(\omega) \text{ for each } \omega \in \Omega$$

It is said to be of ***discrete type*** if the number of different values it can take is finite or countably infinite.

2.2 The Distribution Function of a Random Variable

2.2.1 The Definition of Random Events and its Probability

We could use $\{X \in I\}$ denote ***random events*** where I is a subset of the real line, i.e.,

$$\{X \in I\} = \{\omega | X(\omega) \in I, \omega \in \Omega\}$$

We can denote the ***probability*** that the values of X belong to the subset I by $P(X \in I)$, then

$$P(X \in I) = P(\omega : X(\omega) \in I)$$

2.2.2 The Definition of Distribution Function

The function $F(x)$ that associates with each real number x the probability $P(X \leq x)$ that the random variable X takes on a value smaller than or equal to this number is called the ***distribution function*** of X, with its abbreviation d.f.. That is

$$F(x) = P(X \leq x), \forall x \in \mathbb{R}$$

It should be emphasized that the distribution function is defined in this way for each random variable X.

Some author use the term ***cumulative distribution function***, instead of distribution function, and use the abbreviation c.d.f..

2.2.3 The Properties of Distribution Function

The d.f. $F(x)$ of every random variable X has the following properties:

(1) if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$

(2) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$

$(3) F(x) = F(x^+)$ at every point x

$(4) P(X > x) = 1 - F(x)$ for every value x

$(5) P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$ for all values x_1 and x_2 such that $x_1 < x_2$

$(6) P(X < x) = F(x^-)$ for each value x

$(7) P(X = x) = F(x) - F(x^-)$

$(8) P(X \geq x) = 1 - F(x^-)$

2.2.4 The Discrete Case of Distribution Function

For a discrete random variable X, we define the ***probability (mass) function*** $p(x)$ of X by $p(x) = P(X = x)$. The abbreviation for probability function is p.f..

Let $\{x_1, x_2, \cdots\}$ be the set of possible values of the discrete random variable X. The function p has the following properties:

$(1) p(x_k) \geq 0$ for all x_k; $p(x) = 0$ for all other values of x

(2) $\sum_{k=1}^{\infty} p(x_k) = 1$

Since $p(x)$ is an function, it can be presented as a formula, just like $p(x_k) = P(X = x_k), k = 1, 2, \cdots, n, \cdots$

2.2.5 The Continuous Case of Distribution Function

For a continuous random variable X, if there is a function f defined for all $x \in \mathbb{R}$ and have the following properties:

$(1) f(x) \geq 0$ for any real number x

(2)if $I$is any subset of $\mathbb{R}$, then

$$P(x \in I) = \int_I f(x)dx$$

where $f(x)$ is called the ***probability density function*** of r.v. X. The abbreviation for it is p.d.f..

If f is the p.d.f of the r.v. X, then

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

If $I = (-\infty, x]$, then the distribution function F of X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx$$

We also deduce that

$$\frac{d}{dx} F(x) = f(x)$$

for any x where $F(x)$ is differentiable.

Also,

(1) The d.f. $F(x)$ of continuous r.v. X is continuous since $P(X = x) = 0$

(2) $\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1$

(3) $f(x) = \frac{d}{dx}F(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x<X\leq x+\Delta x)}{\Delta x}$

2.3 The Distribution Function of Function of a Random Variable

Because a random variable is a real-valued function and the composition of two functions is also a function, we can assert that if X is a random variable, then $Y := g(X) = g(X(\omega))$, where g is a real-valued function defined on the real line, is a random variable as well.

2.3.1 The Distribution Function Method

This method is the usual way to obtain p.d.f. of a function of a continuous random variable. It can be divided into four steps:

$(1)\text{transform the event } \{Y \leq y\} \text{ to } \{g(X) \leq y\}$

$(2)\text{transform the event } \{g(X) \leq y\} \text{ to } \{X \in I_y\}, \text{where } I_y = \{x : g(x) \leq y\}$

$(3)\text{calculate the probability } P(X \in I_y), \text{ which is just the d.f. } F(y)(= P(Y \leq y)) \text{ of } Y$

$(4)\text{obtain p.d.f. } f(y) \text{ of } Y \text{ by } f(x) = F^{'}(x)$

2.3.2 The Density Function Method

Let X be a continuous random variable having p.d.f f_X. Suppose that $g(x)$ is a strictly monotonic (increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined by $Y = g(X)$ has a p.d.f. given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left|\frac{d}{dy}g^{-1}(y)\right| \text{if y } = g(x) \text{ for some } x \\ 0 \text{ if y } \neq g(x) \text{ for all } x \end{cases}$$

Where $x = g^{-1}(y)$ is denfined to equal that value of x such that $g(x) = y$

2.4 Mathematical Expectation and Variance

2.4.1 Expectation of a Random Variable

2.4.2 Expectation of Functions of a Random Variable

2.4.3 Variance of a Random Variable

2.4.4 The Application of Expectation and Variation