

# 随机变量及其分布

July 31, 2017

A random variable is a real-valued function defined on a sample space. For each random variable  $X$  and each set  $C$  of real numbers, we could calculate the probability that  $X$  takes its value in  $C$ . The collection of all of these probabilities is the distribution of  $X$ .

## 1 离散型随机变量

Discrete distributions are those that assign positive probability to at most countably many different values. A discrete distribution can be characterized by its **probability function (p.f.)**, which specifies the probability that the random variable takes each of the different possible values. A random variable with a discrete distribution will be called a discrete random variable.

Let  $S$  be the sample space for an experiment. A real-valued function that is defined on  $S$  is called a **random variable**.

Let  $X$  be a random variable. The distribution of  $X$  is the collection of all probabilities of the form  $\Pr(X \in C)$  for all sets  $C$  of real numbers such that  $\{X \in C\}$  is an event.

A random variable  $X$  has a discrete distribution or that  $X$  is a discrete random variable if  $X$  can take only a finite number  $k$  of different values  $x_1, \dots, x_k$  or, at most, an infinite sequence of different values  $x_1, x_2, \dots$ .

If a random variable  $X$  has a discrete distribution, the **probability function** (abbreviated **p.f.**) of  $X$  is defined as the function  $f$  such that for every real number  $x$ ,

$$f(x) = \Pr(X = x) . \quad (1)$$

The closure of the set  $\{x : f(x) > 0\}$  is called the **support of (the distribution of)  $X$** .

If  $X$  has a discrete distribution, the probability of each subset  $C$  of the real line can be determined from the relation

$$\Pr(X \in C) = \sum_{x_i \in C} f(x_i) .$$

## 1.1 (0 – 1) 分布

A random variable  $Z$  that **takes only two values 0 and 1 with  $\Pr(Z = 1) = p$**  has the Bernoulli distribution with parameter  $p$ . We also say that  $Z$  is a **Bernoulli random variable with parameter  $p$** .

设随机变量  $X$  只能取 0 和 1 两个值,

$$P\{X = k\} = p^k(1 - p)^{1-k}, \quad k = 0, 1 \quad (0 < p < 1) \quad (2)$$

## 1.2 二项分布

$$X \sim b(n, p)$$

**伯努利试验:**

试验  $E$  只有两个可能的结果:  $A$  及  $\bar{A}$ ;

$n$  重伯努利试验:

设  $P(A) = p$  ( $0 < p < 1$ ), 则  $P(\bar{A}) = 1 - p$ ,  $E$  独立地重复进行  $n$  次

$$P\{X = k\} = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \dots, n \quad (3)$$

这里  $q = 1 - p$ ;

$$\sum_{k=0}^n P\{X = k\} = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1 \quad (4)$$

[1] The binomial distribution typically occurs in the study of **repeated independent trials of random events**. The  $S = s$  successes in  $n$  trials is given by the **binomial probability distribution**:

$$P(S = s) = \frac{n!}{s!(n-s)!} p^s q^{n-s} = \binom{n}{s} p^s q^{n-s}. \quad (5)$$

### 1.3 泊松分布

$X \sim \pi(\lambda)$

设随机变量  $X$  所有可能取的值为  $0, 1, 2, \dots$ , 而各个值的概率为

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots \quad (6)$$

这里  $\lambda > 0$ ;

$$\sum_{k=0}^{\infty} P\{X = k\} = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1 \quad (7)$$

[1] Poisson noise, where fluctuations in a low rate of arrival of particles at a detector cause statistically predictable fluctuations in the detector signal. The Poisson distribution can

be developed by considering the probabilities that varying numbers of events are detected over an interval during which events occur at a constant rate of probability. The essential features of the development are that it assumes that (1) the event rate is small enough that there will be observationally accessible intervals in which at most one event occurs (i.e., one can consider intervals containing either zero or one event), and (2) the total number of events is small enough that it is useful to model their occurrence by a discrete probability distribution.

Define the probability  $P_n(t)$  that exactly  $n$  events occur in a time  $t$ , and that the probability of one event occurring in a short time interval  $dt$  will be  $\mu dt$ , where  $\mu$  is a constant such that  $\mu dt \ll 1$ . This time interval  $dt$  is short enough that the possibility that more than one event occurs within it can be neglected. Based on this hypothesis, set up a recursion relation for  $P_n(t)$  by considering the two following mutually exclusive possibilities for the occurrence of  $n$  events in a time  $t + dt$ : (1) that  $n$  events occur during a time  $t$  and no events occur in a subsequent time interval  $dt$ , and (2) that  $n - 1$  events occur during the time  $t$  and one event occurs during the subsequent interval  $dt$ .

$$P_n(t + dt) = P_n(t)P_0(dt) + P_{n-1}(t)P_1(dt) .$$

Insert  $P_1(dt) = \mu dt$  and  $P_0(dt) = 1 - P_1(dt)$  and dividing through by  $dt$ ,

$$\frac{dP_n(t)}{dt} = \frac{P_n(t + dt) - P_n(t)}{dt} = \mu P_{n-1}(t) - \mu P_n(t) . \quad (8)$$

For  $n = 0$ , it simplifies (because the possibility involving  $P_{n-1}$  does not exist) to

$$\frac{dP_0(t)}{dt} = -\mu P_0(t) .$$

This equation, with initial condition  $P_0(0) = 1$  (meaning that it is certain that no events are observed in an interval of zero length), has solution  $P_0(t) = e^{-\mu t}$ . The probability that

no events have occurred before time  $t$  decays exponentially with  $t$ , at a rate dependent on the magnitude of  $\mu$ . From this starting point and the further initial conditions  $P_n(0) = 0$  for  $n \geq 1$  (again, no detection of events occurs during an interval of zero length), the recursion relation can be solved to yield

$$P_n(t) = \frac{(\mu t)^n}{n!} e^{-\mu t} \quad (9)$$

It can be checked by substituting it into the recursion formula and by verifying that it satisfies the initial conditions  $P_n(0) = \delta_{n0}$ . The Poisson-distribution probabilities given for a discrete random variable  $X$  in the standard form,

$$P(n) = \frac{\mu^n}{n!} e^{-\mu}, \quad X = n = 0, 1, 2, \dots \quad (10)$$

### 1.3.1 Relation to Binomial Distribution

A Poisson distribution becomes a good approximation of the binomial distribution for a large number  $n$  of trials and small probability  $p \sim \mu/n$ , with  $\mu$  held constant.

#### Theorem

In the limit  $n \rightarrow \infty$  and  $p \rightarrow 0$  so that the mean value  $np \rightarrow \mu$  stays finite, the binomial distribution becomes a Poisson distribution.

### 1.3.2 Poisson Processes

[2] A Poisson process with rate  $\lambda$  per unit time is a process that satisfies the following two properties:

- i. The number of arrivals in every fixed interval of time of length  $t$  has the Poisson distribution with mean  $\lambda t$ .
- ii. The numbers of arrivals in every collection of disjoint time intervals are independent.

A Poisson process can be used to model occurrences in any region that can be subdivided into arbitrarily small pieces. There are **three assumptions** that lead to the Poisson process model.

The first assumption is that the **numbers of occurrences in any collection of disjoint intervals of time must be mutually independent**.

The second assumption is that the **probability of an occurrence during each very short interval of time must be approximately proportional to the length of that interval**. To express this condition more formally, use the standard mathematical notation in which  $o(t)$  denotes any function of  $t$  having the property that

$$\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0 . \quad (11)$$

$o(t)$  must be a function that approaches 0 as  $t \rightarrow 0$ , and, furthermore, this function must approach 0 at a rate faster than  $t$  itself. The second assumption can now be expressed as follows: There exists a constant  $\lambda > 0$  such that for every time interval of length  $t$ , the probability of at least one occurrence during that interval has the form  $\lambda t + o(t)$ . Thus, for every very small value of  $t$ , the probability of at least one occurrence during an interval of length  $t$  is equal to  $\lambda t$  plus a quantity having a smaller order of magnitude.

One of the consequences of the second assumption is that the **process being observed must be stationary over the entire period of observation**, i.e. the **probability of an occurrence must be the same over the entire period**. There can be neither busy intervals, during which we know in advance that occurrences are likely to be more frequent, nor quiet intervals, during which we know in advance that occurrences are likely to be less frequent. This condition is reflected in the fact that the same constant  $\lambda$  expresses the probability of an occurrence in every interval over the entire period of observation. The second assumption

can be relaxed at the cost of more complicated mathematics.

The third assumption is that, for each very short interval of time, the probability that there will be two or more occurrences in that interval must have a smaller order of magnitude than the probability that there will be just one occurrence. In symbols, the probability of two or more occurrences in a time interval of length  $t$  must be  $o(t)$ . Thus, the probability of two or more occurrences in a small interval must be negligible in comparison with the probability of one occurrence in that interval. Of course, it follows from the second assumption that the probability of one occurrence in that same interval will itself be negligible in comparison with the probability of no occurrences.

Under the preceding three assumptions, it can be shown that the process will satisfy the definition of a Poisson process with rate  $\lambda$ .

## 2 连续型随机变量

A random variable  $X$  has a continuous distribution or that  $X$  is a continuous random variable if there exists a nonnegative function  $f$ , defined on the real line, such that for every interval of real numbers (bounded or unbounded), the probability that  $X$  takes a value in the interval is the integral of  $f$  over the interval. The function  $f$  is called the **probability density function** (abbreviated **p.d.f.**) of  $X$ . The closure of the set  $\{x : f(x) > 0\}$  is called the support of (the distribution of)  $X$ .

Continuous Distributions Assign Probability 0 to Individual Values.

## 2.1 Nonuniqueness of the p.d.f.

## 2.2 均匀分布

$$X \sim U(a, b)$$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \end{cases} \quad (12)$$

对任意实数,  $x_1, x_2, (x_1 \leq x_2)$

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx \quad (13)$$

## 2.3 指数分布

$$f(x) = \begin{cases} \frac{1}{\theta} \exp \left[ -\frac{x}{\theta} \right], & x > 0 \\ 0, & \end{cases} \quad (14)$$

## 2.4 正态分布

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right], \quad -\infty < x < \infty, \quad (15)$$

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp \left[ -\frac{(t-\mu)^2}{2\sigma^2} \right] dt, \quad (16)$$



标准正态分布  $\mu = 0, \sigma = 1$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{t^2}{2} \right], \quad (17)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left[ -\frac{t^2}{2} \right] dt, \quad (18)$$

$$\Phi(-x) = 1 - \Phi(x) \quad (19)$$

若  $X \sim N(\mu, \sigma^2)$ , 则

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \quad (20)$$

#### 2.4.1 Limits of Poisson and Binomial Distributions

[1] In a special limit, the discrete Poisson probability distribution is closely related to the continuous Gauss distribution.

##### Theorem

For large  $n$  and mean value  $\mu$ , the Poisson distribution approaches a Gauss distribution.

##### Theorem

In the limit  $n \rightarrow \infty$ , with  $p$  a finite trial probability such that the mean value  $np \rightarrow \infty$ , the binomial distribution becomes a Gauss normal distribution. When  $np \rightarrow \mu < \infty$ , the binomial distribution becomes a Poisson distribution.

### 3 The Cumulative Distribution Function

The distribution function or **cumulative distribution function** (abbreviated **c.d.f.**)  $F$  of a random variable  $X$  is the function

$$F(x) = \Pr(X \leq x) \text{ for } -\infty < x < \infty . \quad (21)$$

The function  $F(x)$  is nondecreasing as  $x$  increases; that is, if  $x_1 < x_2$ , then  $F(x_1) \leq F(x_2)$ .

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ and } \lim_{x \rightarrow \infty} F(x) = 1.$$

A c.d.f. is always continuous from the right; that is,  $F(x) = F(x^+)$  at every point  $x$ .

For every value  $x$ ,

$$\Pr(X > x) = 1 - F(x) . \quad (22)$$

For all values  $x_1$  and  $x_2$  such that  $x_1 < x_2$ ,

$$\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1) . \quad (23)$$

For each value  $x$ ,

$$\Pr(X < x) = F(x^-) . \quad (24)$$

For every value  $x$ ,

$$\Pr(X = x) = F(x) - F(x^-) . \quad (25)$$

#### 3.1 The Quantile Function

Let  $X$  be a random variable with c.d.f.  $F$ . For each  $p$  strictly between 0 and 1, define

$F^{-1}(p)$  to be the smallest value  $x$  such that  $F(x) \geq p$ . Then  $F^{-1}(p)$  is called the  $p$  quantile of  $X$  or the  $100p$  percentile of  $X$ . The function  $F^{-1}$  defined here on the open interval  $(0, 1)$  is called the **quantile function of  $X$** .

Quantiles of Continuous Distributions : When the c.d.f. of a random variable  $X$  is continuous and one-to-one over the whole set of possible values of  $X$ , the inverse  $F^{-1}(p)$  of  $F$  exists and equals the quantile function of  $X$ .

Quantiles of Discrete Distributions :

The  $1/2$  quantile or the 50th percentile of a distribution is called its **median**. The  $1/4$  quantile or 25th percentile is the **lower quartile**. The  $3/4$  quantile or 75th percentile is called the **upper quartile**.

## 4 多维随机变量

设  $E$  是一个随机试验，它的样本空间是  $S = \{e\}$ ，设  $X = X(e)$  和  $Y = Y(e)$  是定义在  $S$  上的随机变量，由它们构成的一个向量  $(X, Y)$ ，叫做**二维随机变量**或**二维随机向量**。

设  $(X, Y)$  是二维随机变量，对于任意实数  $x, y$ ，二元函数：

$$F(x, y) = P\{(X \leq x) \cap (Y \leq y)\} \stackrel{\text{记成}}{=} P(X \leq x, Y \leq y) \quad (26)$$

称为二维随机变量  $(X, Y)$  的分布函数，或称为随机变量  $X$  和  $Y$  的**联合分布函数**。

Let  $X$  and  $Y$  be random variables such that  $X$  is discrete and  $Y$  is continuous. Suppose that there is a function  $f(x, y)$  defined on the  $xy$ -plane such that, for every pair  $A$  and  $B$  of subsets of the real numbers

$$\Pr(X \in A \text{ and } Y \in B) = \int_B \sum_{x \in A} f(x, y) dy \quad (27)$$

if the integral exists. Then the function  $f$  is called the joint p.f./p.d.f. of  $X$  and  $Y$ .

Let  $X$  and  $Y$  have a joint c.d.f.  $F$ . The c.d.f.  $F_1$  of just the single random variable  $X$  can be derived from the joint c.d.f.  $F$  as  $F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$ . Similarly, the c.d.f.  $F_2$  of  $Y$  equals  $F_2(y) = \lim_{x \rightarrow \infty} F(x, y)$ , for  $0 < y < \infty$ .

If  $X$  and  $Y$  have a continuous joint distribution with joint p.d.f.  $f$ , then the joint c.d.f. at  $(x, y)$  is

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(r, s) dr ds .$$

The joint p.d.f. can be derived from the joint c.d.f. by

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

at every point  $(x, y)$  at which these second-order derivatives exist.

## 4.1 边缘分布

Suppose that  $X$  and  $Y$  have a joint distribution. The c.d.f. of  $X$

$$F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$$

is called the **marginal c.d.f. of  $X$** . Similarly, the p.f. or p.d.f. of  $X$  associated with the marginal c.d.f. of  $X$  is called the **marginal p.f.** or **marginal p.d.f. of  $X$** .

If  $X$  and  $Y$  have a discrete joint distribution for which the joint p.f. is  $f$ , then the marginal p.f.  $f_1$  of  $X$  is

$$f_1(x) = \sum_{\text{All } y} f(x, y) . \quad (28)$$

Similarly, the marginal p.f.  $f_2$  of  $Y$  is

$$f_2(y) = \sum_{\text{All } x} f(x, y) . \quad (29)$$

If  $X$  and  $Y$  have a continuous joint distribution with joint p.d.f.  $f$ , then the marginal p.d.f.  $f_1$  of  $X$  is

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{for } -\infty < x < \infty . \quad (30)$$

Similarly, the marginal p.d.f.  $f_2$  of  $Y$  is

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad \text{for } -\infty < y < \infty. \quad (31)$$

## 4.2 条件分布

## 4.3 相互独立的随机变量

设  $F(x, y)$  及  $F_X(x), F_Y(y)$  分别是二维随机变量  $(X, Y)$  的分布函数及边缘分布函数。

若对于所有  $x, y$  有

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}, \quad (32)$$

即

$$F(x, y) = F_X(x)F_Y(y) \quad (33)$$

称随机变量  $X$  和  $Y$  是相互独立的。

## 4.4 两个随机变量的函数的分布

## 4.5 Multivariate Distributions

The joint c.d.f. of  $n$  random variables  $X_1, \dots, X_n$  is the function  $F$  whose value at every point  $(x_1, \dots, x_n)$  in  $n$ -dimensional space  $R^n$  is specified by the relation

$$F(x_1, \dots, x_n) = \Pr(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n). \quad (34)$$

Every multivariate c.d.f. satisfies properties similar to those given earlier for univariate and bivariate c.d.f.'s.

It is said that  $n$  random variables  $X_1, \dots, X_n$  have a discrete joint distribution if the random vector  $(X_1, \dots, X_n)$  can have only a finite number or an infinite sequence of different possible values  $(x_1, \dots, x_n)$  in  $R^n$ . The joint p.f. of  $X_1, \dots, X_n$  is then defined as the function  $f$  such that for every point  $(x_1, \dots, x_n) \in R^n$ ,

$$f(x_1, \dots, x_n) = \Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) .$$

$n$  random variables  $X_1, \dots, X_n$  are independent if, for every  $n$  sets  $A_1, A_2, \dots, A_n$  of real numbers,

$$\Pr(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \Pr(X_1 \in A_1) \Pr(X_2 \in A_2) \cdots \Pr(X_n \in A_n) .$$

If  $X_1, \dots, X_n$  are independent, it follows easily that the random variables in every nonempty subset of  $X_1, \dots, X_n$  are also independent.

Let  $F$  denote the joint c.d.f. of  $X_1, \dots, X_n$ , and let  $F_i$  denote the marginal univariate c.d.f. of  $X_i$  for  $i = 1, \dots, n$ . The variables  $X_1, \dots, X_n$  are independent if and only if, for all points  $(x_1, x_2, \dots, x_n) \in R^n$ ,

$$F(x_1, \dots, x_n) = F_1(x_1) F_2(x_2) \cdots F_n(x_n) .$$

$X_1, \dots, X_n$  are independent if and only if their joint c.d.f. is the product of their  $n$  individual marginal c.d.f.'s.

If  $X_1, \dots, X_n$  have a continuous, discrete, or mixed joint distribution for which the joint p.d.f., joint p.f., or joint p.f./p.d.f. is  $f$ , and if  $f_i$  is the marginal univariate p.d.f. or p.f. of  $X_i (i = 1, \dots, n)$ , then  $X_1, \dots, X_n$  are independent if and only if the following relation is satisfied at all points  $(x_1, x_2, \dots, x_n) \in R^n$ :

$$f(x_1, \dots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) .$$

Consider a given probability distribution on the real line that can be represented by either a p.f. or a p.d.f.  $f$ . It is said that  $n$  random variables  $X_1, \dots, X_n$  form a random sample from this distribution if these random variables are independent and the marginal p.f. or p.d.f. of each of them is  $f$ . Such random variables are also said to be **independent and identically distributed**, abbreviated **i.i.d.** We refer to the number  $n$  of random variables as the sample size.

$X_1, \dots, X_n$  form a random sample from the distribution represented by  $f$  if their joint p.f. or p.d.f.  $g$  is specified as follows at all points  $(x_1, x_2, \dots, x_n) \in R^n$ :

$$g(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n) .$$

Clearly, an i.i.d. sample cannot have a mixed joint distribution.

## References

- [1] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists, Seventh Edition: A Comprehensive Guide*. Academic Press, 7 edition, January 2012.
- [2] M.H. DeGroot and M.J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.