Review of elementary Stochastic Processes

1.1. Random Variables and Probability Theory

In order to understand the theory of stochastic processes, it is necessary to have a firm grounding in the basic concepts of probability theory. As a result, we shall briefly review some of these concepts and at the same time establish some useful notation.

The *distribution function* $F(\ )$ of the random variable X is defined for any real number x by

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

A random variable X is said to be *discrete* if its set of possible values is countable. For discrete random variables, the probability mass function $p(x)$ is defined by

$$p(x) = P\{X = x\}$$

Clearly,

$$F(x) = \sum_{y \leq x} p(y)$$

A random variable is called *continuous* if there exists a function $f(x)$, called the *probability density function*, such that

$$P\{X \text{ is in } B\} = \int_B f(x)\, dx$$

for every Borel set B. Since $F(x) = \int_{-\infty}^{x} f(x)\, dx$, it follows that

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

The *expectation* or *mean* of the random variable X, denoted by EX, is defined by

$$EX = \int_{-\infty}^{\infty} x\, dF(x) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} xf(x)\, dx & \text{if } X \text{ is continuous} \\ \displaystyle\sum_{x:\, p(x) > 0} xp(x) & \text{if } X \text{ is discrete} \end{cases} \tag{1}$$

provided the above integral exists.

Equation (1) also defines the expectation of any function of X, say $h(X)$. Since $h(X)$ is itself a random variable, it follows from (1) that

#

$$Eh(X) = \int_{-\infty}^{\infty} x\, dF_h(x)$$

where F_h is the distribution function of $h(X)$. However, it can be shown that this is identical to $\int_{-\infty}^{\infty} h(x)\, dF(x)$, i.e.,

$$Eh(X) = \int_{-\infty}^{\infty} h(x)\, dF(x) \tag{2}$$

The above equation is sometimes known as the *law of the unconscious statistician* [as statisticians have been accused of using the identity (2) without realizing that it is not a definition].

The *variance* of the random variable X is defined by

$$\begin{aligned}\operatorname{Var} X &= E(X - EX)^2 \\ &= EX^2 - (EX)^2\end{aligned}$$

Jointly Distributed Random Variables

The *joint distribution* F of two random variables X and Y is defined by

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

The distributions $F_X(x) = \lim_{y \to \infty} F(x, y)$ and $F_Y(y) = \lim_{x \to \infty} F(x, y)$ are called the *marginal distributions* of X and Y. X and Y are said to be *independent* if

$$F(x, y) = F_X(x) F_Y(y)$$

for all real x and y. It can be shown that X and Y are independent if and only if

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

for all functions g and h for which the above expectations exist.

Two jointly distributed random variables X and Y are said to be *uncorrelated* if their covariance, defined by

$$\begin{aligned} \operatorname{Cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= EXY - EXEY \end{aligned}$$

is zero. It follows that independent random variables are uncorrelated. However, the converse is not true. (Give an example.)

The random variables X and Y are said to be *jointly continuous* if there exists a function $f(x, y)$, called the *joint probability density function*, such that

$$P\{X \text{ is in } A,\ Y \text{ is in } B\} = \int_A \int_B f(x, y)\, dy\, dx$$

for every two Borel sets A and B.

The joint distribution of any collection $X_1, X_2, \ldots, X_n$ of random variables is defined by

$$F(x_1, \ldots, x_n) = P\{X_1 \le x_1, \ldots, X_n \le x_n\}$$

Furthermore, the n random variables are said to be independent if

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \ldots F_{X_n}(x_n)$$

where

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to \infty \\ j \neq i}} F(x_1, \ldots, x_n)$$

Characteristic Functions and Laplace Transforms

The *characteristic function* $\phi(\;)$ of X is defined, for any real number u, by

$$\begin{aligned}\phi(u) &= E[e^{iuX}] \\ &= \int_{-\infty}^{\infty} e^{iux}\, dF(x)\end{aligned}$$

A random variable always possesses a characteristic function (that is, the above integral always exists) and, in fact, it can be proven that there is a one-to-one correspondence between distribution functions and characteristic functions. This result is quite important as it enables us to characterize the probability law of a random variable by its characteristic function. (See Tables 1 and 2.)

#

TABLE I

Discrete Probability Laws	Probability Mass Function $p(x)$	Characteristic Function	Mean	Variance
Poisson with parameter $\lambda > 0$	$p(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$e^{\lambda(e^{iu}-1)}$	λ	λ
Binomial with parameters n and p	$p(x) = \dbinom{n}{x} p^x(1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^{iu} + q)^n$ where $q = 1 - p$	np	npq
Geometric $0 \leq p \leq 1$, $q = 1 - p$	$p(x) = pq^x$, $x = 0, 1, \ldots$	$\dfrac{p}{1 - qe^{iu}}$	$\dfrac{q}{p}$	$\dfrac{q}{p^2}$
Negative Binomial with parameters $r = 1, 2, \ldots$ and p, $0 \leq p \leq 1$, $q = 1 - p$	$p(x) = \dbinom{r+x-1}{x} p^r q^x$, $x = 0, 1, 2, \ldots$	$\left(\dfrac{p}{1 - qe^{iu}}\right)^r$	$\dfrac{rq}{p}$	$\dfrac{rq}{p^2}$

#

TABLE 2

Continuous Probability Laws	Probability Density Function	Characteristic Function	Mean	Variance
Exponential $\lambda > 0$	$f(x) = \lambda e^{-\lambda x}$ $x \geq 0$	$\dfrac{\lambda}{\lambda - iu}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Gamma $r > 0, \lambda > 0$	$f(x) = \dfrac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}$ $x \geq 0$	$\left(\dfrac{\lambda}{\lambda - iu}\right)^r$	$\dfrac{r}{\lambda}$	$\dfrac{r}{\lambda^2}$
Uniform over $[a, b]$	$f(x) = \dfrac{1}{b - a}$ $a < x < b$	$\dfrac{e^{iub} - e^{iua}}{iu(b - a)}$	$\dfrac{a + b}{2}$	$\dfrac{(b - a)^2}{12}$
Normal μ, σ^2	$f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$	$e^{iu\mu - u^2\sigma^2/2}$	μ	σ^2

#

A further useful result is that if $X_1, \ldots, X_n$ are independent, then the characteristic function of their sum $X_1 + \cdots + X_n$ is just the product of the individual characteristic functions. This result is quite useful, as it often enables us to determine the distribution of the sum of independent random variables by first calculating the characteristic function and then attempting to identify it.

EXAMPLE. Let X and Y be independent and identically distributed normal random variables having mean μ and variance σ^2. Then,

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)$$

where the subscript indicates the random variable associated with the characteristic function. Hence (see Table 2),

$$\begin{aligned}\phi_{X+Y}(u) &= (e^{i\mu u - \sigma^2 u^2/2})^2 \\ &= e^{i2\mu u - \sigma^2 u^2}\end{aligned}$$

which is the characteristic function of a normal random variable having mean 2μ and variance $2\sigma^2$. Therefore, by the uniqueness of the characteristic function, this is the distribution of $X + Y$.

#

We may also define the joint characteristic of the random variables $X_1, \ldots, X_n$ by

$$\phi(u_1, u_2, \ldots, u_n) = E\left[\exp\left(i \sum_{i=1}^{n} u_i X_i\right)\right]$$

It may be proven that the joint characteristic function uniquely determines the joint distribution.

When dealing with random variables which only assume nonnegative values, it is sometimes more convenient to use *Laplace transforms* rather than characteristic functions. The Laplace transform of the distribution F (or, more precisely, of the random variable having distribution F) is defined by

$$\tilde{F}(s) = \int_0^\infty e^{-sx}\, dF(x)$$

This integral exists for a complex variable $s = a + bi$ where $a \geq 0$. As in the case of characteristic functions, the Laplace transform uniquely determines the distribution.

We may also define Laplace transforms for arbitrary functions in the following manner: The Laplace transform of the function g, denoted $\tilde{g}$, is defined by

$$\tilde{g}(s) = \int_0^\infty e^{-sx}\, dg(x)$$

#

provided the integral exists. It can be shown that $\tilde{g}$ determines g up to an additive constant.

Convolutions

If X and Y are independent random variables, with X having distribution F and Y having distribution G, then the distribution of $X+Y$ is given by $F * G$, where

$$(F * G)(t) = \iint_{X+Y \leq t} dF(x)\, dG(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{t-y} dF(x)\, dG(y)$$

$$= \int_{-\infty}^{\infty} F(t-y)\, dG(y) = \int_{-\infty}^{\infty} G(t-x)\, dF(x)$$

$F * G$ is called the *convolution* of F and G. If $G = F$, then $F * F$ is denoted by F_2. Similarly, we denote by F_n the n-fold convolution of F with itself. That is,

$$F_n = F * \underbrace{(F * F * \cdots * F)}_{n-1}$$

It is easy to show that the characteristic function (Laplace transform) of a convolution is just the product of the characteristic functions (Laplace transforms).

Limit Theorems

Some of the most important results in probability theory are in the form of limit theorems. The two most important are:

#

Law of Large Numbers. If $X_1, X_2, \ldots$ are independent and identically distributed with mean μ, then with probability one,

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \qquad \text{as } n \to \infty$$

Central Limit Theorem. If $X_1, X_2, \ldots$ are independent and identically distributed with mean μ and variance σ^2, then

$$\lim_{n \to \infty} P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

1.2. Conditional Expectation

If X and Y are discrete random variables, then the conditional probability mass function of Y, given X, is defined, for all x such that $P\{X = x\} > 0$, by

$$p_{Y|X}(y \mid x) = P\{Y = y \mid X = x\} = \frac{P\{Y = y, X = x\}}{P\{X = x\}}$$

Similarly, the conditional distribution function $F_{Y|X}(y \mid x)$ of Y, given X, is defined for all x such that $P\{X = x\} > 0$, by

$$F_{Y|X}(y \mid x) = P\{Y \leq y \mid X = x\} = \sum_{y' \leq y} p_{Y|X}(y' \mid x)$$

If X and Y have a joint probability density function $f_{X,Y}(x, y)$, the conditional probability density function of Y, given X, is defined for all x such that $f_X(x) > 0$ by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

and the conditional probability distribution function of Y, given X, by

$$F_{Y|X}(y|x) = P\{Y \leq y | X = x\} = \int_{-\infty}^{y} f_{Y|X}(y|x)\, dy$$

The conditional expectation of Y, given X, is defined for all x such that $f_X(x) > 0$, by

$$E[Y|X = x] = \int_{-\infty}^{\infty} y\, dF_{Y|X}(y|x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy$$

Let us denote by $E[Y|X]$ the function of X whose value at $X = x$ is $E[Y|X = x]$. An extremely important property of conditional expectation is that for all random variables X and Y,

$$EY = E[E[Y|X]] = \int_{-\infty}^{\infty} E[Y|X = x]\, dF(x) \tag{3}$$

#

provided the relevant expectations exist. Hence, (3) states that the expectation of Y may be obtained by first conditioning on X (to obtain $E[Y \mid X]$) and then taking the expectation (with respect to X) of this quantity.

EXAMPLE. Suppose that the number of accidents occurring in a factory in a month is a random variable with distribution F. Suppose also that the number of workmen injured in each accident are independent and have a common distribution G. What is the expected number of workmen injured each month?

Example. A prisoner is placed in a cell containing three doors. The first door leads immediately to freedom. The second door leads into a tunnel which returns him to the cell after one day's travel. The third door leads to a similar tunnel which returns him to his cell after three days. Assuming that the prisoner is at all times equally likely to choose any one of the doors, what is the expected length of time until the prisoner reaches freedom?

Let Y denote the time until the prisoner reaches freedom, and let X denote the door that he initially chooses. We first note that

$$
\begin{aligned}
E[Y \mid X = 1] &= 0 \\
E[Y \mid X = 2] &= 1 + EY \qquad (5) \\
E[Y \mid X = 3] &= 3 + EY
\end{aligned}
$$

#

To see why this is so, consider $E[Y \mid X = 2]$, and reason as follows: If the prisoner chooses the second door, then he spends one day in the tunnel and then returns to his cell. But once he returns to his cell the problem is as before, and hence his expected time until freedom from that moment on is just EY. Hence, $E[Y \mid X = 2] = 1 + EY$. Therefore, from (5) we obtain

$$EY = \tfrac{1}{3}[0 + 1 + EY + 3 + EY]$$

or

$$EY = 4$$

Functional Equations and Lack of Memory of the Exponential Distribution

The following two functional equations occur quite frequently in the theory of applied probability:

$$f(s + t) = f(s) \cdot f(t) \qquad \text{for all } s, t \geq 0 \tag{6}$$

$$f(s + t) = f(s) + f(t) \qquad \text{for all } s, t \geq 0 \tag{7}$$

It turns out that the only (measurable) solution to these functional equations are of the respective forms

$$f(t) = e^{-\lambda t}$$

and

$$f(t) = ct$$

We shall now use (6) to prove that the exponential is the unique distribution without memory.

A random variable X is said to be without memory, or *memoryless*, if

$$P\{X > s + t \mid X > t\} = P\{X > s\} \qquad \text{for all } s, t \geq 0 \tag{8}$$

If we think of X as being the lifetime of some instrument, then (8) states that the probability that the instrument lives for at least $s + t$ hours given that it has survived t hours is the same as the initial probability that it lives for at least s hours. That is, the instrument does not deteriorate.

Suppose now that X is memoryless, and let $\bar{F}(x) = P\{X > x\}$. Now from (8) we obtain

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

or

$$\bar{F}(s + t) = \bar{F}(s) \cdot \bar{F}(t)$$

implying that

$$\bar{F}(t) = e^{-\lambda t}$$

which is the distribution of the exponential random variable. Also, by considering the argument in reverse, it follows that the exponential distribution is memoryless.

1.3. Stochastic Processes

A *stochastic process* $\{X(t),\ t \in T\}$ is a family of random variables. That is, for each t contained in the index set T, $X(t)$, is a random variable. The variable t is often interpreted as time, and hence $X(t)$ represents the *state* of the process at time t. For instance, $X(t)$ may represent the amount of inventory in a retail store at time t or the number of people in a bank at time t or the position of a particle at time t, etc.

The set T is called the *index set* of the stochastic process. If T is a countable set, then the stochastic process is said to be a *discrete time* process. If T is an open or closed interval of the real line, then we say that the stochastic process is a *continuous time* process.

The set of possible values which the random variables $X(t)$, $t \in T$ may assume is called the *state space* of the process.

#

A continuous time stochastic process $\{X(t), t \in T\}$ is said to have *independent increments* if for all choices of $t_0 < t_1 < t_2 < \cdots < t_n$, the n random variables

$$X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$$

are independent. The process is said to have *stationary independent increments* if in addition $X(t_2 + s) - X(t_1 + s)$ has the same distribution as $X(t_2) - X(t_1)$ for all $t_1, t_2 \in T$ and $s > 0$.

Examples of Stochastic Processes

1. The General Random Walk

Let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables, and let $X_n = \sum_{i=1}^{n} Y_i$. The stochastic process

$$\{X_n, n = 0, 1, 2, \ldots\}$$

is called the *general random walk process*. If Y_i represents the number of items sold by a retail store during the ith week, then X_n would be the total number of items sold during the first n weeks.

#

2. The Wiener Process

A stochastic process $\{X(t),\ t \geq 0\}$ is said to be a *Wiener process* if

(i) $\{X(t),\ t \geq 0\}$ has stationary, independent increments
(ii) for every $t > 0$, $X(t)$ is normally distributed with mean 0
(iii) $X(0) = 0$.

Both the general random walk process and the Wiener process are examples of a class of stochastic processes known as *Markov processes*. A Markov process is a stochastic process with the property that given the value of $X(t)$, the probability of $X(s + t)$, where $s > 0$, is independent of the values of $X(u)$, $u < t$. That is, the conditional distribution of the future $X(s + t)$, given the present $X(t)$ and the past $X(u)$, $u < t$, is independent of the past. More formally, the process $\{X(t),\ t \in T\}$ is said to be a Markov process if

$$P\{X(t) \leq x \mid X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_n) = x_n\}$$
$$= P\{X(t) \leq x \mid X(t_n) = x_n\}$$

whenever $t_1 < t_2 < \cdots < t_n < t$.

#

Problems

1. Show that $\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\, \mathrm{Cov}(X, Y)$.

2. If X has characteristic function $\phi(u)$, show that

$$EX = \frac{1}{i}\frac{d}{du}\phi(u)|_{u=0},$$

and in general,

$$EX^n = \left(\frac{1}{i}\right)^n \frac{d^n}{du^n}\phi(u)|_{u=0}$$

whenever the moments exist.

3. If X and Y are independent Poisson random variables, find the distribution of $X+Y$.

4. Let X and Y be Poisson with respective means λ_1 and λ_2. Show that the conditional distribution of X, given $X+Y$, is binomial.

5. Suppose X is distributed as a Poisson random variable with mean λ. The parameter λ is itself a random variable whose distribution law is exponential with mean $1/\mu$. Find the distribution of X.

6. An urn has n chips. Chips are drawn one at a time and then put back in the urn. Let N denote the number of drawings required until some chip is drawn more than once. Find the probability distribution of N.

#

7. A man with n keys wants to open his door. He tries the keys in a random manner. Let N be the number of trials required to open the door. Find EN and $\text{Var}\, N$ if (a) unsuccessful keys are eliminated from further selection, (b) if they are not.

8. The conditional variance of Y, given X, is defined by

$$\text{Var}(Y|X) = E[(Y - E(Y|X))^2|X]$$

Show that $\text{Var}\, Y = E[\text{Var}(Y|X)] + \text{Var}\, E[Y|X]$.

9. Let N denote the number of customers arriving at a store in a given day. Suppose that the amounts spent by the customers are independent and have a distribution F. Find the mean and variance of the total amount of money spent in the store.

10. Show that every stochastic process with independent increments is a Markov process

11. Let $Y_{1n}, Y_{2,n}, \ldots, Y_{n,n}$ be n independent random variables with the identical uniform distribution on $(0,t)$. Let $Z_n = \min(Y_{1,n}, \ldots, Y_{n,n})$.

(a) Find $P\{Z_n > x\}$.

(b) Let t be a function of n such that $\lim_{n\to\infty} n/t = \lambda$.

Show that

$$\lim_{n\to\infty} P\{Z_n > x\} = e^{-\lambda x}.$$

12. Show that the only continuous solutions of the functional Equations (6) and (7) are respectively $f(t) = e^{-\lambda t}$ and $f(t) = ct$.

13. Show for the Wiener process that $\text{Var}\, X(t) = \sigma^2 t$ for some $\sigma^2 > 0$.

#