EE 4TM4: Digital Communications II
Probability Theory

I. Random Variables

A random variable is a real-valued function defined on the sample space.

Example: Suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0,1,2 with respective probabilities

$$P\{Y=0\} = P\{(T,T)\} = \frac{1}{4},$$
$$P\{Y=1\} = P\{(T,H),(H,T)\} = \frac{1}{2},$$
$$P\{Y=2\} = P\{(H,H)\} = \frac{1}{4}.$$

Example: Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \cdots$, with respective probabilities

$$P\{N=1\} = P\{H\} = p,$$
$$P\{N=2\} = P\{(T,H)\} = (1-p)p,$$
$$P\{N=3\} = P\{(T,T,H)\} = (1-p)^2 p,$$
$$\vdots$$
$$P\{N=n\} = P\{(T,T,\cdots,T,H)\} = (1-p)^{n-1}p, \quad n \geq 1.$$

Note that

$$\sum_{n=1}^{\infty} P\{N=n\} = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1.$$

Random variables that take on either a finite or a countable number of possible values are called discrete. Random variables that take on a continuum of possible values are known as continuous random variables.

The cumulative distribution function (cdf) (or more simply the distribution function) $F(\cdot)$ of the random variable X is defined for any real number b, $-\infty < b < \infty$, by

$$F(b) = P\{X \leq b\}.$$

In words, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b. Some properties of the cdf F are

1) $F(b)$ is a nondecreasing function of b,
2) $\lim_{b\to\infty} F(b) = F(\infty) = 1$,
3) $\lim_{b\to-\infty} F(b) = F(-\infty) = 0$.
   Note that $P\{a < X \le b\} = F(b) - F(a)$ for all $a < b$. Also note that $P\{X < b\} = \lim_{h\to 0^+} P\{X \le b - h\} = \lim_{h\to 0^+} F(b-h)$, which does not necessarily equal $F(b)$ since $F(b)$ also includes the probability that X equals b.

A. Discrete Random Variables

For a discrete random variable X, we define the probability mass function $p(a)$ of X by $p(a) = P\{X = a\}$. The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \cdots$, then $p(x_i) > 0$, $i = 1, 2, \cdots$, $p(x) = 0$ for all other values of x. Since X must take on one of the values x_i, we have $\sum_{i=1}^{\infty} p(x_i) = 1$.

The cumulative distribution function F can be expressed in terms of $p(a)$ by $F(a) = \sum_{x_i \le a} p(x_i)$. For instance, suppose X has a probability mass function given by $p(1) = \frac{1}{2}$, $p(2) = \frac{1}{3}$, $p(3) = \frac{1}{6}$, then, the cumulative distribution function of F of X is given by $F(a) = 0$ for $a < 1$, $F(a) = \frac{1}{2}$ for $1 \le a < 2$, $F(a) = \frac{5}{6}$ for $2 \le a < 3$, $F(a) = 1$ for $3 \le a$.

- The Bernoulli Random Variable

  Suppose that a trial whose outcome can be classified as either a "success" or as a "failure" is performed. If we let X equal 1 if the outcome is a success and 0 if it is failure, then the probability mass function of X is given by $p(0) = P\{X = 0\} = 1 - p$, $p(1) = P\{X = 1\} = p$, where p, $0 \le p \le 1$, is the probability that the trial is a "success." The random variable X is said to be a Bernoulli random variable.
- The Binomial Random Variable

  Suppose that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1-p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p).

  The probability mass function of a binomial random variable having parameters (n, p) is given by

  $$p(i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \cdots, n,$$

  where $\binom{n}{i} = \frac{n!}{(n-i)!i!}$.

  Note that, by the binomial theorem, the probabilities sum to one, that is,

  $$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1.$$
- The Geometric Random Variable

  Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success, then X is said to be a geometric random variable with parameter p. Its probability mass function is given by

$$p(n) = P\{X = n\} = (1-p)^{n-1}p, \quad n = 1, 2, \cdots.$$

Note that

$$\sum_{n=1}^{\infty} p(n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = 1.$$

- The Poisson Random Variable

  A random variable X, taking on one of the values $0, 1, 2, \cdots$, is said to be a Poisson random variable with parameter λ, if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, \cdots.$$

  Note that

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

  An important property of the Poisson random variable is that it may be used to approximate a binomial random variable when the binomial parameter n is large and p is small. To see this, suppose that X is a binomial random variable with parameters (n, p), and let $\lambda = np$. Then

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} (\frac{\lambda}{n})^i (1-\frac{\lambda}{n})^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}. \end{aligned}$$

  Now, for n large and p small

$$(1-\frac{\lambda}{n})^n \approx e^{-\lambda}, \quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1, \quad (1-\frac{\lambda}{n})^i \approx 1.$$

  Hence, for n large and p small,

$$P\{X = i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}.$$

B. Continuous Random Variables

We say that X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx.$$

The function $f(x)$ is called the probability density function of the random variable X. Note that

$$\begin{aligned} &1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx, \\ &P\{a \le X \le b\} = \int_a^b f(x)dx, \\ &P\{X = a\} = \int_a^b f(x)dx = 0. \end{aligned}$$

Something is of probability zero does not mean it's impossible.

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx.$$

Differentiating both sides of the preceding yields

$$\frac{d}{da}F(a) = f(a).$$

- The Uniform Random Variable

  A random variable is said to be uniformly distributed over the interval $(0, 1)$ if its probability density function is given by $f(x) = 1$ for $0 < x < 1$ and $f(x) = 0$ otherwise. Note that

  $$P\{a \leq X \leq b\} = \int_{a}^{b} f(x)dx = b - a.$$

  In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by $f(x) = \frac{1}{\beta-\alpha}$ if $\alpha < x < \beta$ and $f(x) = 0$ otherwise.

- Exponential Random Variables

  A continuous random variable whose probability density function is given, for some $\lambda > 0$, by $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$ and $f(x) = 0$ if $x < 0$ is said to be an exponential random variable with parameter λ. Note that the cumulative distribution function F is given by

  $$F(a) = \int_{0}^{a} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}, \quad a \geq 0.$$

  In particular, $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1$.

- Normal Random Variables

  We say that X is a normal random variable (or simply that X is normally distributed) with parameters μ and σ^2 if the density of X is given by

  $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

  An important fact about normal random variables is that if X is normally distributed with parameters μ and σ^2 then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$. To prove this, suppose first that $\alpha > 0$ and note that $F_Y(\cdot)$, the cumulative distribution function of the random variable Y, is given by

  $$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{\alpha X + \beta \leq a\} \\ &= P\{X \leq \frac{a-\beta}{\alpha}\} \\ &= F_X(\frac{a-\beta}{\alpha}) \\ &= \int_{-\infty}^{(a-\beta)/\alpha} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp(\frac{-(y-(\alpha\mu+\beta))^2}{2\alpha^2\sigma^2})dy, \end{aligned}$$

where the last equality is obtained by the change in variables $y = \alpha x + \beta$. However, since $F_Y(a) = \int_{-\infty}^{a} f_Y(y)dv$, it follows that the probability density function $f_Y(\cdot)$ is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\alpha\sigma}\exp(\frac{-(y-(\alpha\mu+\beta))^2}{2\alpha^2\sigma^2}), \quad -\infty < y < \infty.$$

Hence, Y is normally distributed with parameters $\alpha\mu+\beta$ and $(\alpha\sigma)^2$. A similar result is also true when $\alpha < 0$. One implication of the preceding result is that if X is normally distributed with parameters μ and σ^2 then $Y = (X-\mu)/\sigma$ is normally distributed with parameters 0 and 1. Such a random variable Y is said to be have the standard or unit normal distribution.

Finally, we shall verify that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

Note that

$$\begin{aligned}\Big(\int_{-\infty}^{\infty} e^{-x^2/2}dx\Big)^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}dx \int_{-\infty}^{\infty} e^{-y^2/2}dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy.\end{aligned}$$

Let $x = r\cos\theta$ and $y = r\sin\theta$. The Jacobian determinant is given by

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Therefore,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy = \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2} r dr d\theta = 2\pi.$$

C. Expectation and Variance of a Random Variable

Expectation of a random variable X: $E[X] = \sum_x xp(x)$ (discrete) and $E[X] = \int xf(x)dx$ (continuous).

- Expectation of a Bernoulli Random Variable

$$E[X] = 0(1-p) + 1(p) = p.$$

- Expectation of a Binomial Random Variable

$$\begin{aligned} E[X] &= \sum_{i=0}^{n} ip(i) \\ &= \sum_{i=0}^{n} i\binom{n}{i}p^i(1-p)^{n-i} \\ &= \sum_{i=1}^{n} \frac{in!}{(n-i)!i!}p^i(1-p)^{n-i} \\ &= \sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!}p^i(1-p)^{n-i} \end{aligned}$$

$$
\begin{aligned}
&= np\sum_{i=1}^{n}\frac{(n-1)!}{(n-i)!(i-1)!}p^{i-1}(1-p)^{n-i} \\
&= np\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k}(1-p)^{n-1-k} \\
&= np[p+(1-p)]^{n-1} \\
&= np,
\end{aligned}
$$

where the second from the last equality follows by letting $k = i-1$. Note that this result can also be obtained by writing the binomial random variable as the sum of n independent Bernoulli random variables.

- Expectation of a Geometric Random Variable

$$
\begin{aligned}
E[X] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} \\
&= p\sum_{n=1}^{\infty} nq^{n-1} \\
&= p\sum_{n=1}^{\infty}\frac{d}{dq}(q^n) \\
&= p\frac{d}{dq}\Big(\sum_{n=1}^{\infty} q^n\Big) \\
&= p\frac{d}{dq}\Big(\frac{q}{1-q}\Big) \\
&= \frac{p}{(1-q)^2} \\
&= \frac{1}{p},
\end{aligned}
$$

where $q = 1-p$.

- Expectation of a Poisson Random Variable

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty}\frac{ie^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=1}^{\infty}\frac{e^{-\lambda}\lambda^i}{(i-1)!} \\
&= \lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda.
\end{aligned}
$$

- Expectation of a Uniform Random Variable

$$
\begin{aligned}
E[X] &= \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha} dx \\
&= \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} \\
&= \frac{\beta+\alpha}{2}.
\end{aligned}
$$

- Expectation of an Exponential Random Variable

$$
\begin{aligned}
E[X] &= \int_{0}^{\infty} x\lambda e^{-\lambda x} dx \\
&= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\
&= 0 - \frac{e^{-\lambda x}}{\lambda}\Bigg|_0^{\infty} \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

- Expectation of a Normal Random Variable

$$
\begin{aligned}
E[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} xe^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)e^{-(x-\mu)^2/2\sigma^2} dx + \mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} ye^{-y^2/2\sigma^2} dy + \mu \\
&= \mu.
\end{aligned}
$$

Proposition (Expectation of a Function of a Random Variable): (a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

$$E[g(X)] = \sum_x g(x)p(x).$$

(b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Corollary: If a and b are constants, then $E[aX+b] = aE[X]+b$.

Variance of a random variable X: $\mathrm{Var}(X) = E[(X-E[X])^2]$. A useful identity: $\mathrm{Var}(X) = E[X^2]-(E[X])^2$.

- Variance of a Bernoulli Random Variable

$$\begin{aligned}\text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= p - p^2.\end{aligned}$$

- Variance of a Normal Random Variable

$$\begin{aligned}\text{Var}(X) &= E[(X-\mu)^2] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\Big(-ye^{-y^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}dy\Big) \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}dy \\ &= \sigma^2.\end{aligned}$$

D. Jointly Distributed Random Variables

We define, for any two random variables X and Y, the joint cumulative probability distribution function X and Y by

$$F(a,b) = P\{X \le a, Y \le b\}, \quad -\infty < a,b < \infty.$$

In the case where X and Y are both discrete random variables, it is convenient to define the joint probability mass function of X and Y by

$$p(x,y) = P\{X = x, Y = y\}.$$

Let $f(x,y)$ be the joint probability density function of X and Y. We have

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)dxdy.$$

Because

$$F(a,b) = P\{X \le a, Y \le b\} = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dydx,$$

differentiation yields

$$\frac{d^2}{dadb}F(a,b) = f(a,b).$$

An important identity:

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n].$$

The random variables X and Y are said to be independent if, for all a, b,

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}.$$

In terms of the joint distribution function F of X and Y, we have that X and Y are independent if $F(a,b) = F_X(a)F_Y(b)$ for all a, b. When X and Y are discrete, the condition of independence reduces to $p(x,y) = p_X(x)p_Y(y)$ while if X and Y are jointly continuous, independence reduces to $f(x,y) = f_X(x)f_Y(y)$.

One can extend the definition of independence to multiple random variables. For example, X, Y, Z are independent if

$$P\{X \leq a, Y \leq b, Z \leq c\} = P\{X \leq a\}P\{Y \leq b\}P\{Z \leq c\}$$

for a, b, c.

Proposition: If X and Y are independent, then for any functions h and g $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

The covariance of any two random variables X and Y, denoted by $\mathrm{Cov}(X,Y)$, is defined by

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y].\end{aligned}$$

Note that if X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$. However, $\mathrm{Cov}(X,Y) = 0$ does not X and Y are independent. For example, $P\{X = 1\} = P\{X = -1\} = 1/2$, $P\{Y = 1|X = 1\} = P\{Y = -1|X = 1\} = P\{Y = 2|X = -1\} = P\{Y = -2|X = -1\}$. It is clear that X and Y are not independent, but it can be verified that $\mathrm{Cov}(X,Y) = 0$.

E. Moment Generating Functions

The moment generating function $\phi(t)$ of the random variable X is defined for all values t by

$$\begin{aligned}\phi(t) &= E[e^{tX}] \\ &= \begin{cases} \sum\limits_x e^{tx}p(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx}f(x)dx, & \text{if } X \text{ is continuous} \end{cases}\end{aligned}$$

We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\begin{aligned}\phi'(t) &= \frac{d}{dt}E[e^{tX}] \\ &= E\Big[\frac{d}{dt}e^{tX}\Big] \\ &= E[Xe^{tX}].\end{aligned}$$

Hence, $\phi'(0) = E[X]$. Similarly,

$$\begin{aligned}
\phi''(t) &= \frac{d}{dt}\phi'(t) \\
&= \frac{d}{dt}E[Xe^{tX}] \\
&= E\Big[\frac{d}{dt}(Xe^{tX})\Big] \\
&= E[X^2 e^{tX}].
\end{aligned}$$

And so $\phi''(0) = E[X^2]$. In general, the nth derivative of $\phi(t)$ evaluated at $t = 0$ equals $E[X^n]$, that is, $\phi^n(0) = E[X^n]$, $n \geq 1$. Note that if $\phi(t)$ is an analytic function, then

$$\phi(t) = \phi(0) + \phi'(0)t + \phi''(0)t^2/2 + \cdots,$$

where $\phi(0) = 1$, and $\phi^{(n)}(0) = E[X^n]$ for $n \geq 1$. Therefore, knowing the moments of X, we can reconstruct $\phi(t)$ and further reconstruct the distribution of X.

- The binomial Distribution with Parameters n and p

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \\
&= (pe^t + 1 - p)^n.
\end{aligned}$$

Hence, $\phi'(t) = n(pe^t + 1 - p)^{n-1}pe^t$ and so $E[X] = \phi'(0) = np$. Differentiating a second time yields

$$\phi''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1}pe^t$$

and so $E[X^2] = \phi''(0) = n(n-1)p^2 + np$. Thus, the variance of X is given by

$$\begin{aligned}
\text{Var}(X) &= E[X^2] - (E[X])^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= np(1-p).
\end{aligned}$$

- The Poisson Distribution with Mean λ

$$\begin{aligned}\phi(t) &= E[e^{tX}] \\ &= \sum_{n=0}^{\infty} \frac{e^{tn}e^{-\lambda}\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=\infty}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= \exp\{\lambda(e^t - 1)\}.\end{aligned}$$

Differentiating yields

$$\begin{aligned}\phi'(t) &= \lambda e^t \exp\{\lambda(e^t - 1)\}, \\ \phi''(t) &= (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}.\end{aligned}$$

and so

$$\begin{aligned}E[X] &= \phi'(0) = \lambda, \\ E[X^2] &= \phi''(0) = \lambda^2 + \lambda, \\ \text{Var}(X) &= E[X^2] - (E[X])^2 = \lambda.\end{aligned}$$

- The Exponential Distribution with Parameter λ

$$\begin{aligned}\phi(t) &= E[e^{tX}] \\ &= \int_0^{\infty} e^{tx} e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx \\ &= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda.\end{aligned}$$

Differentiating of $\phi(t)$ yields

$$\begin{aligned}\phi'(t) &= \frac{\lambda}{(\lambda - t)^2} \\ \phi''(t) &= \frac{2\lambda}{(\lambda - t)^3}.\end{aligned}$$

Hence

$$E[X] = \phi'(0) = \frac{1}{\lambda}, \quad E[X^2] = \phi''(0) = \frac{2}{\lambda^2}.$$

The variance of X is thus given by

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}.$$

- The Normal Distribution with Parameters μ and σ^2

  The moment generating function of a standard normal random variable Z is obtained as follows.

$$\begin{aligned}
E[e^{tZ}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2tx)/2} dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\
&= e^{t^2/2}.
\end{aligned}$$

  If Z is a standard normal, then $X = \sigma Z + \mu$ is normal with parameters μ and σ^2; therefore,

$$\phi(t) = E[e^{tX}] = E[e^{t(\sigma Z+\mu)}] = e^{t\mu} E[e^{t\sigma Z}] = \exp(\frac{\sigma^2 t^2}{2} + \mu t).$$

  By differentiating we obtain

$$\begin{aligned}
\phi'(t) &= (\mu + t\sigma^2) \exp(\frac{\sigma^2 t^2}{2} + \mu t), \\
\phi''(t) &= (\mu + t\sigma^2)^2 \exp(\frac{\sigma^2 t^2}{2} + \mu t) + \sigma^2 \exp(\frac{\sigma^2 t^2}{2} + \mu t),
\end{aligned}$$

  and so

$$\begin{aligned}
E[X] &= \phi'(0) = \mu, \\
E[X^2] &= \phi''(0) = \mu^2 + \sigma^2,
\end{aligned}$$

  implying that

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

An important property of moment generating functions is that the moment generating function of the sum of independent random variables is just the product of the individual moment generating functions. To see this, suppose that X and Y are independent and have moment generating function $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then $\phi_{X+Y}(t)$, the moment generating function of $X + Y$, is given by

$$\begin{aligned}
\phi_{X+Y}(t) &= E[e^{t(X+Y)}] \\
&= E[e^{tX} e^{tY}] \\
&= E[e^{tX}] E[e^{tY}] \\
&= \phi_X(t) \phi_Y(t).
\end{aligned}$$

Another important result is that the moment generating function uniquely determines the distribution. That is, there exists a one-to-one correspondence between the moment generating function and the distribution function of a random variable.

F. Limit Theorems

Markov Inequality: If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}.$$

Proof: Note that

$$\begin{aligned} E[X] &\geq E[X1_{\{X \geq A\}}] \\ &\geq E[a1_{\{X \geq A\}}] \\ &= aP\{X \geq a\}, \end{aligned}$$

and the result is proven.

As a corollary, we obtain the following.

Chebyshev's Inequality: If X is a random variable with mean μ and variance σ^2, then, for any value $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Proof: Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2}.$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, the preceding is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2},$$

and the proof is complete.

Weak Law of Large Numbers: Let $X_1, X_2, \cdots$ be a sequence of independent random variables having a common distribution, and let $E[X_i] = \mu$. Then for any $\epsilon > 0$

$$P\left\{ \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| > \epsilon \right\} \to 0 \quad \text{as } n \to \infty.$$

Let $Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu$. Note that $E[Y_n] = 0$ and $\mathrm{Var}(Y_n) = \frac{\sigma^2}{n}$, where $\sigma^2 = \mathrm{Var}(X_i)$. Therefore, by Chebyshev's inequality

$$P\{|Y_n| > \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as } n \to \infty.$$

Note that in this proof we actually only need $X_1, X_2, \cdots$ to be uncorrelated.

Strong Law of Large Numbers: Let $X_1, X_2, \cdots$ be a sequence of independent random variables having a common distribution, and let $E[X_i] = \mu$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as } n \to \infty.$$

Central Limit Theorem: Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$P\Big\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\Big\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx$$

as $n \to \infty$.

We now present a heuristic proof of the central limit theorem. Suppose first that the X_i have mean 0 and variance 1, and let $E[e^{tX}]$ denote their common moment generating function. Then, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ is

$$E\Big[\exp\Big\{t\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big\}\Big] = E[e^{tX_1/\sqrt{n}}e^{tX_2/\sqrt{n}}\cdots e^{tX_n/\sqrt{n}}] = (E[e^{tX/\sqrt{n}}])^n.$$

Now, for n large, we obtain from the Taylor series expansion of e^y that

$$e^{tX/\sqrt{n}} \approx 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n}.$$

Taking expectations shows that when n is large

$$\begin{aligned} E[e^{tX/\sqrt{n}}] &\approx 1 + \frac{tE[X]}{\sqrt{n}} + \frac{t^2E[X^2]}{2n} \\ &= 1 + \frac{t^2}{2n} \quad \text{because } E[X]=0, E[X^2]=1. \end{aligned}$$

Therefore, we obtain that when n is large

$$E\Big[\exp\Big\{t\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big\}\Big] \approx (1 + \frac{t^2}{2n})^n.$$

When n goes to ∞ the approximation can be shown to become exact and we have

$$\lim_{n\to\infty} E\Big[\exp\Big\{t\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big\}\Big] = e^{t^2/2}.$$

Thus, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the moment generating function of a (standard) normal random variable with mean 0 and variance 1. Using this, it can be proven that the distribution function of the random variable $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the standard normal distribution function Φ.

When the X_i have mean μ and variance σ^2, the random variables $\frac{X_i-\mu}{\sigma}$ have mean 0 and variance 1. Thus, the preceding shows that

$$P\Big\{\frac{X_1 - \mu + X_2 - \mu + \cdots + X_n - \mu}{\sigma\sqrt{n}} \leq a\Big\} \to \Phi(a)$$

which proves the central limit theorem.

References

[1] S. M. Ross, *Introduction to Probablity Models*. Tenth Edition. Academic Press, 2009.