Chapter 2: Random Variables and Their Moments

Definition 1 Suppose that \((\Omega, \mathcal{F}, P)\) is a probability space. A function \(X : \Omega \mapsto R\) is called a random variable on \((\Omega, \mathcal{F}, P)\) if

\[
\{X \leq a\} \in \mathcal{F}, \quad \forall a \in R.
\]

It is easy to see that, if \(X\) is a random variable on \((\Omega, \mathcal{F}, P)\), then for any real numbers \(a \leq b\),

\[
\{X < b\}, \quad \{a < X \leq b\}, \quad \{a \leq X \leq b\}, \quad \{a < X < b\}
\]

all belong to \(\mathcal{F}\), so we can talk about their probabilities.

Definition 2 Suppose that \(X\) is a random variable on \((\Omega, \mathcal{F}, P)\), then the function defined by

\[
F(x) = P(X \leq x), \quad x \in R
\]

is called the distribution function of \(X\).

It is easy to prove the following result.

Proposition 1 Suppose that \(F\) is the distribution function of a random variable \(X\) on \((\Omega, \mathcal{F}, P)\), then

(i). \(F\) is nondecreasing;

(ii). \(\lim_{x \to \infty} F(x) = 1, \lim_{x \to -\infty} F(x) = 0\);

(iii). \(F\) is right continuous.

Proof. The proof is routine and we omit the details. \(\blacksquare\)

It can be proven that, whenever a function \(F\) defined on \(R\) satisfies the three conditions in the proposition above, one can construct a probabilistic space and a random variable \(X\) on this probabilistic space such that \(F\) is the distribution function of \(X\). We are not going to pursue this.

The distribution function \(F\) of a random variable \(X\) determines all the statistical properties of \(X\). Once we know the distribution function \(F\) of \(X\), we can find the probability of any event defined in terms of \(X\). For instance, for any real numbers \(a < b\),

\[
\begin{align*}
P(X < b) &= F(b-) \\
P(a < X \leq b) &= F(b) - F(a) \\
P(a < X < b) &= F(b-) - F(a) \\
P(a \leq X \leq b) &= F(b) - F(a-).\end{align*}
\]

We will be mainly interested in two kinds of random variables: discrete random variables and absolutely random variables.
**Definition 3** A random variable $X$ is said to be discrete if it can take on at most a countable number of values.

For a discrete random variable $X$, we define the probability mass function $p(x)$ of $X$ by

$$p(x) = P(X = x), \quad \forall x \in R.$$ 

The probability mass function $p(x)$ of a discrete random variable is positive for at most a countable number of values of $x$. That is, if the range of $X$ is $\{x_1, x_2, \cdots \}$, then $p(x) = 0$ for every $x$ outside this range and obviously we have $\sum_i p(x_i) = 1$.

Obviously, the probability mass function $p(x)$ of a discrete random variable $X$ also completely determines the statistical properties of $X$. It has the following relations with the distribution function $F$ of $X$:

$$F(x) = \sum_{y \leq x} F(y), \quad p(x) = F(x) - F(x-).$$

**Definition 4** A random variable $X$ on $(\Omega, \mathcal{F}, P)$ is said to be continuous if

$$P(X = x) = 0, \quad \forall x \in R,$$

or equivalently, the distribution function of $X$ is continuous.

**Definition 5** A random variable $X$ on $(\Omega, \mathcal{F}, P)$ is said to be absolutely continuous if there exists a nonnegative function $f$ on $R$ such that

$$P(X \leq x) = \int_{-\infty}^{x} f(t) dt, \quad \forall x \in R.$$ 

The function $f$ is called the probability density function of the random variable $X$.

Obviously, an absolutely continuous random variable is continuous. It is also clear that

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

The density function $f(x)$ of an absolutely continuous random variable $X$ also completely determines the statistical properties of $X$. If $X$ is an absolutely continuous random variable with density $f$, then for any real numbers $a \leq b$,

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

**Definition 6** Suppose that $X$ is a random variable on $(\Omega, \mathcal{F}, P)$.

(i) If $X$ is a discrete random variable such that

$$\sum_{x} |x| P(X = x) < \infty,$$

we say that $X$ has finite expectation and we define the expectation of $X$ to be

$$EX = \sum_{x} x P(X = x);$$

2
(ii) If $X$ is an absolutely continuous random variable with density $f$ such that

$$
\int_{-\infty}^{\infty} |x| f(x) dx < \infty,
$$

we say that $X$ has finite expectation and we define the expectation of $X$ to be

$$
EX = \int_{-\infty}^{\infty} x f(x) dx.
$$

Of course, there are random variables which are neither discrete and absolutely continuous. To define the expectation of these general random variables, we need to use some measure theory.

**Definition 7** Suppose that $X$ is a random variable on $(\Omega, \mathcal{F}, P)$ with distribution $F$. If

$$
\int_{\Omega} |X| dP = \int_{-\infty}^{\infty} |x| dF(x) < \infty,
$$

we say that $X$ has finite expectation and we define the expectation of $X$ to be

$$
EX = \int_{\Omega} X dP = \int_{-\infty}^{\infty} x dF(x)
$$

The following result is very useful in calculating the expectation of a function of a random variable.

**Theorem 2** Suppose that $X$ is a random variable on $(\Omega, \mathcal{F}, P)$, $\phi$ is a function defined on $\mathbb{R}$ and that $Y = \phi(X)$.

(i) If $X$ is discrete, then $Y$ has finite expectation if and only if

$$
\sum_{x} |\phi(x)| P(X = x) < \infty. \tag{1}
$$

And when (1) holds,

$$
EY = \sum_{x} \phi(x) P(X = x).
$$

(ii) If $X$ is absolutely continuous with density $f$, then $Y$ has finite expectation if and only if

$$
\int_{-\infty}^{\infty} |\phi(x)| f(x) dx < \infty. \tag{2}
$$

And when (2) holds,

$$
EY = \int_{-\infty}^{\infty} \phi(x) f(x) dx.
$$
**Proof.** We are only going to prove (i). The proof of (ii) goes along the same line of argument. Let $y_1, y_2, \ldots$ be the possible values of $Y$ and let $x_1, x_2, \ldots$ be the possible values of $X$. For each $y_j$, let

$$A_j = \{x_i : \phi(x_i) = y_j\}.$$  

Then

$$P(Y = y_j) = P(X \in A_j) = \sum_{x \in A_j} P(X = x).$$

Therefore

$$\sum_j |y_j| P(Y = y_j) = \sum_j |y_j| \sum_{x \in A_j} P(X = x) = \sum_j \sum_{x \in A_j} |y_j| P(X = x)$$

$$= \sum_j \sum_{x \in A_j} |\phi(x)||P(X = x) = \sum_x |\phi(x)||P(X = x).$$

Thus $Y$ has finite expectation if and only if (1) holds.

If $Y$ has finite expectation, then by repeating the above argument, we can get

$$EY = \sum_x \phi(x)P(X = x).$$

\[\blacksquare\]

**Definition 8** Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$ and let $r > 0$ be an integer. We say that $X$ has a finite moment of order $r$ if $X^r$ has finite expectation. In that case we define the $r$-th moment of $X$ as $E(X^r)$. If $X$ has a finite moment of order $r$, then the $r$-th moment of $X - EX$ is called the $r$-th central moment of $X$.

The theorem above provides a method for calculating the moments of $X$ using the mass function or the density function of $X$.

Among the central moments, we will be mainly interested in the second central moment. We call the second central moment of $X$ the variance of $X$:

$$\text{Var}X = E(X - EX)^2$$

The variance of a random variable measures how spread out the random variable is from its expectation. One can easily derive the following formula for the variance

$$\text{Var}X = EX^2 - (EX)^2.$$

Now we briefly review the most important types random variables.

**Binomial random variable.** Suppose $n$ is a positive integer and $p \in (0, 1)$. A random variable $X$ on $(\Omega, \mathcal{F}, P)$ is called a binomial random variable with parameters $n$ and $p$, if

$$P(X = x) = \left\{ \begin{array}{ll} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n, \\
0, & \text{otherwise} \end{array} \right.$$
Binomial random variables occur frequently in applications. Suppose that independent trials, each results in a success with probability \( p \) and a failure with probability \( 1 - p \), are performed. Then the total number \( S_n \) of successes in the first \( n \) trials is a binomial random variable with parameters \( n \) and \( p \).

It is easy to check that if \( X \) is a binomial random variable with parameters \( n \) and \( p \), then

\[
EX = np, \quad \text{Var}X = np(1 - p).
\]

**Geometric random variable.** Suppose \( p \in (0, 1) \). A random variable \( X \) on \((\Omega, \mathcal{F}, P)\) is called a geometric random variable with parameter \( p \) if

\[
P(X = x) = \begin{cases} 
  p(1 - p)^{x-1}, & x = 1, 2, \ldots, \\
  0, & \text{otherwise} 
\end{cases}
\]

Geometric random variables arise naturally in the following situation. Suppose that independent trials, each results in a success with probability \( p \) and a failure with probability \( 1 - p \), are performed. Let \( X \) be the number of trials needed to get a success, then \( X \) is a geometric random variable with parameter \( p \).

One can check that if \( X \) is a geometric random variable with parameter \( p \), then

\[
EX = \frac{1}{p}, \quad \text{Var}X = \frac{1 - p}{p^2}.
\]

**Poisson random variable.** Suppose that \( \lambda > 0 \). A random variable \( X \) on \((\Omega, \mathcal{F}, P)\) is called a Poisson random variable with parameter \( \lambda \) if

\[
P(X = x) = \begin{cases} 
  \frac{e^{-\lambda} \lambda^x}{x!}, & x = 0, 1, 2, \ldots, \\
  0, & \text{otherwise} 
\end{cases}
\]

The Poisson random variable is very important in applications because it can be used as an approximation for a binomial random variable with parameters \( n \) and \( p \) when \( n \) is large and \( p \) is small so that \( np \) is of moderate size. To see this, suppose that \( X \) is a binomial random variable with parameters \( n \) and \( p \) and let \( \lambda = np \). Then

\[
P(X = i) = \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!} \left( \frac{\lambda}{n} \right)^i \left( 1 - \frac{\lambda}{n} \right)^{n-i} = \frac{n(n-1) \cdots (n-i+1) \lambda^i (1 - \lambda/n)^n}{n^i i! (1 - \lambda/n)^i}.
\]

Now, for \( n \) large and \( \lambda \) moderate,

\[
\left( 1 - \frac{\lambda}{n} \right)^n \approx e^{-\lambda}, \quad \frac{n(n-1) \cdots (n-i+1)}{n^i} \approx 1, \quad \left( \frac{\lambda}{n} \right)^i \approx 1.
\]

Hence, for \( n \) large and \( \lambda \) moderate,

\[
P(X = i) \approx e^{-\lambda} \frac{\lambda^i}{i!}.
\]
In other words, if \( n \) independent trials, each results in a success with probability \( p \) and a failure with probability \( 1 - p \), are performed, then, when \( n \) is large and \( p \) is small so that \( np \) is moderate, the number of successes is approximately a Poisson random variable with parameter \( \lambda = np \). Some examples of random variables which are approximately Poisson are as follows: (a) the number of misprints on a page (or a group of pages); (b) the number of people in a community who are over 95; (c) the number \( \alpha \)-particles discharged from some radioactive material in a fixed period of time. It is easy to see that if \( X \) is a Poisson random variable with parameter \( \lambda \), then

\[
EX = \lambda, \quad \text{Var}X = \lambda.
\]

**Negative binomial random variable.** Suppose \( r > 0 \) is an integer and \( p \in (0, 1) \). A random variable \( X \) on \( (\Omega, \mathcal{F}, P) \) is called a negative binomial random variable with parameters \( r \) and \( p \) if

\[
P(X = x) = \begin{cases} 
p^r \left( \frac{x - 1}{r - 1} \right) (1 - p)^{x - r}, & x = r, r + 1, \ldots, \\
0, & \text{otherwise}
\end{cases}
\]

Negative binomial random variables occur naturally in the following situation. Suppose that independent trials, each results in a success with probability \( p \) and a failure with probability \( 1 - p \), are performed. If \( X \) is the number of trials needed to get \( r \) successes, then \( X \) is a negative binomial random variable with parameter \( r \) and \( p \).

We are going to check later that if \( X \) is a negative binomial random variable with parameter \( r \) and \( p \), then

\[
EX = \frac{r}{p}, \quad \text{Var}X = \frac{r(1 - p)}{p^2}.
\]

**Uniform random variable.** Suppose that \( a < b \) are real numbers. A random variable \( X \) on \( (\Omega, \mathcal{F}, P) \) is called a uniform random variable on \( (a, b) \) if its density function is given by

\[
f(x) = \begin{cases} 
\frac{1}{b - a}, & a < x < b \\
0, & \text{otherwise}
\end{cases}
\]

It is elementary to check that if \( X \) is a uniform random variable on \( (a, b) \), then

\[
EX = \frac{a + b}{2}, \quad \text{Var}X = \frac{(b - a)^2}{12}.
\]

**Normal random variable.** Suppose that \( \mu \) and \( \sigma > 0 \) are two positive numbers. A random variable \( X \) on \( (\Omega, \mathcal{F}, P) \) is called a normal random variable with parameters \( \mu \) and \( \sigma^2 \) if its density is given by

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.
\]

Normal random variables occur very often in applications. The reason for that is the central limit theorem which we are going to cover later.

One can check that if \( X \) is a normal random variable with parameters \( \mu \) and \( \sigma^2 \), then

\[
EX = \mu, \quad \text{Var}X = \sigma^2.
\]
**Exponential random variable.** Suppose $\lambda > 0$. A random variable $X$ on $(\Omega, \mathcal{F}, P)$ is called an exponential random variable with parameter $\lambda$ if its density is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

It is easy to see that if $X$ is an exponential random variable with parameter $\lambda$, then its distribution function is given by

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Therefore we have

$$P(X > t) = e^{-\lambda t}, \quad t > 0.$$

From this we can easily check that the exponential random variable $X$ satisfies the following memoryless property

$$P(X > s + t | X > t) = P(X > s), \quad s, t > 0.$$

Actually the inverse result is true. If a nonnegative random variable $X$ satisfies the memoryless property, then it is must be an exponential random variable. But we are not going to prove this. Therefore exponential random variables are very useful in describing lifetimes of equipments, such as electric fuses, that neither deteriorates nor improves in the course of time but can fail due sporadic chance happening. Exponential random variables are the continuous versions of geometric random variables.

It is easy to show that

$$EX = \frac{1}{\lambda}, \quad \text{Var}X = \frac{1}{\lambda^2}.$$

**Gamma random variable.** Suppose $\alpha$ and $\lambda$ are two positive real numbers. A random variable $X$ on $(\Omega, \mathcal{F}, P)$ is called a Gamma random variable with parameters $\alpha$ and $\lambda$ if its density is given by

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

It is easy to see that an exponential random variable with parameter $\lambda$ is a Gamma random variable with parameters $1$ and $\lambda$. One can also check that if $X$ is a normal random variable with parameters $\mu = 0$ and $\sigma^2$, then $Y = X^2$ is a Gamma random variable with parameters $1/2$ and $1/(2\sigma^2)$.

Gamma random variables are the continuous versions of the negative binomial random variables.

One can check that when $X$ is a Gamma random variable with parameters $\alpha$ and $\lambda$,

$$EX = \frac{\alpha}{\lambda}, \quad \text{Var}X = \frac{\alpha}{\lambda^2}.$$

Now we are going to talk about random vectors.

**Definition 9** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $n > 0$ is an integer. A function $X = (X_1, \ldots, X_n)$ from $\Omega$ into $\mathbb{R}^n$ is called an $n$-dimensional random vector on $(\Omega, \mathcal{F}, P)$ if $X_1, \ldots, X_n$ are random variables on $(\Omega, \mathcal{F}, P)$. 
If \( \mathbf{X} = (X_1, \ldots, X_n) \) is a random vector on \( \mathbf{X} = (X_1, \ldots, X_n) \), the following function

\[
F(\mathbf{x}) = F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n), \quad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n
\]
is called the distribution function of \( \mathbf{X} \), or the joint distribution function of \( X_1, \ldots, X_n \). As in the one dimensional case, the distribution function \( F \) determines all the statistical properties of \( \mathbf{X} \). For instance if \( F \) is the distribution function of \( \mathbf{X} = (X_1, X_2) \), then for any \( a_1 < b_1 \) and \( a_2 < b_2 \),

\[
P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1).
\]

The distribution function \( F_{X_i}(x_i) \) of \( X_i \), \( i = 1, \ldots, n \), is called the \( i \)-th marginal distribution function of \( \mathbf{X} \). It is easy to see that the joint distribution function determines the marginal distribution functions. For instance,

\[
F_{X_i}(x_i) = \lim_{x_2 \to \infty, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n).
\]

But in general, the marginal distribution functions do not determine the joint distribution function.

**Definition 10** An \( n \)-dimensional random vector \( \mathbf{X} \) is said to be discrete if the range of \( \mathbf{X} \) is at most countable.

For an \( n \)-dimensional random vector \( \mathbf{X} \) on \( (\Omega, \mathcal{F}, P) \), the following function

\[
p(\mathbf{x}) = p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n), \quad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n
\]
is called the mass function of \( \mathbf{X} \), or the joint mass function of \( X_1, \ldots, X_n \). The mass function \( p_{X_i}(x_i) \) of \( X_i \), \( i = 1, \ldots, n \), is called the \( i \)-th marginal mass function of \( \mathbf{X} \). Similar to the distribution case, the joint mass function determines the marginal mass functions, but not the other way around.

**Definition 11** An \( n \)-dimensional random vector \( \mathbf{X} \) on \( (\Omega, \mathcal{F}, P) \) is said to be absolutely continuous if there is a nonnegative function \( f \) on \( \mathbb{R}^n \) such that

\[
P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n) du_1 \cdots du_n, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.
\]

The function \( f \) is called the density function of \( \mathbf{X} \), or the joint density function of \( X_1, \ldots, X_n \). The density \( f_{X_i}(x_i) \) of \( X_i \), \( i = 1, \ldots, n \), is called the \( i \)-th marginal density of \( \mathbf{X} \).

Similar to the distribution case, the joint density function determines the marginal density functions, but not the other way around.

One can show that if \( \mathbf{X} \) is an \( n \)-dimensional absolutely continuous random vector with density \( f \), then for any Borel subset of \( \mathbb{R}^n \)

\[
P(\mathbf{X} \in A) = \int_A f(x_1, \ldots, x_n) dx_1 \cdots dx_n.
\]


**Definition 12** Suppose that $X_1, \ldots, X_n$ are random variables on $(\Omega, \mathcal{F}, P)$. We say that $X_1, \ldots, X_n$ are independent if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

or equivalently, if the joint distribution function is equal to the product of the marginal distributions.

If $X_1, \ldots, X_n$ are independent random variables on $(\Omega, \mathcal{F}, P)$, then one can check that, for any Borel subsets $A_1, \ldots, A_n$ of $\mathbb{R}$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

One can also check that, in the discrete case, $X_1, \ldots, X_n$ are independent if and only if the joint mass function is equal to the product of the marginal mass functions and that, in the absolutely continuous case, $X_1, \ldots, X_n$ are independent if and only if the joint density function is equal to the product of the marginal density functions.

**Theorem 3** Suppose that $(X, Y)$ is a discrete 2-dimensional random vector on $(\Omega, \mathcal{F}, P)$.

(i) If $(X, Y)$ is discrete with mass function $p_{X,Y}$, then the mass function of $X + Y$ is given by

$$p_{X+Y}(z) = \sum_x p_{X,Y}(x, z-x).$$

In particular, when $X$ and $Y$ are independent discrete random variables,

$$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x).$$

(ii) If $(X, Y)$ is absolutely continuous with density $f_{X,Y}$, then the density of $X + Y$ is given by

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx.$$  

In particular, when $X$ and $Y$ are independent discrete random variables,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx.$$  

**Proof.** We only give the proof of part (ii). The proof of (i) is similar. For any $z \in \mathbb{R}$, set

$$A_z = \{(x, y) : x + y \leq z\}.$$  

Then

$$P(X + Y \leq z) = \int_{A_z} f_{X,Y}(x, y)dxdy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{z-x} f_{X,Y}(x, y)dy \right)dx.$$  

Make the change of variable $y = v - x$ in the inner integral we get

$$P(X + Y \leq z) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{z} f_{X,Y}(x, v-x)dv \right)dx = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f_{X,Y}(x, v-x)dx \right)dv.$$  

The proof is now complete.

Similar to Theorem 2, we have the following result.
Theorem 4 Suppose that \( \mathbf{X} \) is an \( n \)-dimensional random vector on \( (\Omega, \mathcal{F}, P) \), \( \phi \) is a Borel function defined on \( \mathbb{R}^n \) and that \( Y = \phi(\mathbf{X}) \).

(i) If \( \mathbf{X} \) is discrete with joint mass function \( p_{\mathbf{X}} \), then \( Y \) has finite expectation if and only if
\[
\sum_{\mathbf{x}} |\phi(\mathbf{x})| p_{\mathbf{X}}(\mathbf{x}) < \infty.
\]
And when (3) holds,
\[
EY = \sum_{\mathbf{x}} \phi(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}).
\]

(ii) If \( \mathbf{X} \) is absolutely continuous with density \( f_{\mathbf{X}} \), then \( Y \) has finite expectation if and only if
\[
\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\phi(x_1, \ldots, x_n)| f(x_1, \ldots, x_n) dx_1 \cdots dx_n < \infty.
\]
And when (4) holds,
\[
EY = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \cdots dx_n
\]

Proof. The proof is similar to that of Theorem 2, we omit the details. \( \blacksquare \)

Theorem 5 Let \( X_1, \ldots, X_n \) be random variables on \( (\Omega, \mathcal{F}, P) \) with finite expectations and \( c_1, \ldots, c_n \) are real numbers, then
\[
E \sum_{k=1}^{n} c_k X_k = \sum_{k=1}^{n} c_k EX_k.
\]

Theorem 6 Let \( X \) and \( Y \) be random variables on \( (\Omega, \mathcal{F}, P) \) with finite expectations. Then \( XY \) has finite expectation and
\[
E(XY) = (EX)(EY).
\]

Definition 13 Let \( X \) and \( Y \) be random variables on \( (\Omega, \mathcal{F}, P) \) with finite second moments. We define the covariance of \( X \) and \( Y \) to be
\[
\text{Cov}(X,Y) = E[(X - EX)(Y - EY)] = E(XY) - (EX)(EY).
\]

From the linearity of expectations, we immediately get the following result.

Theorem 7 Let \( X_1, \ldots, X_n \) be random variables on \( (\Omega, \mathcal{F}, P) \) with finite second moments. Then
\[
\text{Var} \left( \sum_{k=1}^{n} X_k \right) = \sum_{k=1}^{n} \text{Var}(X_k) + 2 \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{Cov}(X_k, X_j).
\]

In particular, if \( X_1, \ldots, X_n \) are independent, then
\[
\text{Var} \left( \sum_{k=1}^{n} X_k \right) = \sum_{k=1}^{n} \text{Var}(X_k).
\]
Example 1 Find the expectation and variance of a binomial random variable with parameters $n$ and $p$.

Example 2 Suppose that we have a population of $r$ objects, $r_1$ of which are of type one and $r - r_1$ of type two. A sample of size $n$ is drawn without replacement from this population. Let $S_n$ be the number of objects of type one in the sample. Find $ES_n$ and $VarS_n$.

Solution For $i = 1, 2, \ldots, r_1$, define $X_i$ to be lif the $i$-th object of type one is in the sample and to be zero otherwise. Then $S_n = \sum_{i=1}^{r_1} X_i$. Now

$$EX_i = P(X_i = 1) = \frac{\binom{r-1}{n-1}}{\binom{r}{n}} = \frac{n}{r},$$
hence

$$ES_n = \frac{n r_1}{r}.$$ 

Also

$$Var(X_i) = \frac{n}{r} \left(1 - \frac{n}{r}\right).$$

For $1 \leq i < j \leq r_1$,

$$E[X_i X_j] = P(X_i = 1, X_j = 1) = \frac{n(n-1)}{r(r-1)},$$

thus

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - (EX_i)(EX_j)$$

$$= \frac{n(n-1)}{r(r-1)} - \left(\frac{n}{r}\right)^2$$

$$= \frac{n(n-r)}{r r(r-1)}.$$ 

Consequently

$$Var(S_n) = \frac{r_1}{r} \frac{n}{r(1 - \frac{n}{r})} - r_1(1 - \frac{r_1}{r}) \frac{n}{r} \frac{r - n}{r(r-1)}$$

$$= n \frac{r_1}{r} \frac{1 - \frac{r_1}{r}}{1 - \frac{n-1}{r-1}}.$$ 

Example 3 Suppose that there are $N$ different types of coupons and each time one obtains a coupon it is equally likely to be any one of the $N$ types. Let $S_n$ be the number of coupons one needs to get $n \leq N$ distinct coupons. Find $ES_n$ and $VarS_n$.

Solution For $i = 0, 1, \ldots, n_1$, define $X_i$ to be the number of additional coupons that need to be obtained after $i$ distinct types have been collected in order to obtain another distinct type. Then

$$S_n = X_0 + X_1 \cdots + X_{n-1}.$$
Obviously \( X_0 = 1 \). We can see that, for \( i = 1, 2, \ldots, n - 1 \), \( X_i \) is a geometric random variable with parameter \( \frac{N-i}{N} \), and that \( X_1, X_2, \ldots, X_{n-1} \) are independent. Thus

\[
ES_n = 1 + \sum_{i=1}^{n-1} \frac{N}{N - i} = N\left(\frac{1}{N} + \frac{1}{N - 1} + \cdots + \frac{1}{N - n + 1}\right),
\]

and

\[
VarS_n = \sum_{i=1}^{n-1} \text{Var}X_i = \sum_{i=1}^{n-1} \frac{N}{N(1 - i/N)^2}.
\]

**Example 4** Suppose that each of \( N \) men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. Let \( S_N \) be the number of men who get their own hats. Find \( ES_N \) and \( VarS_N \).

**Solution** For each \( i = 1, 2, \ldots, N \), define \( X_i \) to be 1 if the \( i \)-th person gets his own hat and to be zero otherwise. Then \( S_N = X_1 + X_2 + \cdots + X_N \). Obviously for each \( i \),

\[
EX_i = P(X_i = 1) = \frac{1}{N}.
\]

Thus

\[
ES_N = \sum_{i=1}^{N} EX_i = 1.
\]

We know that, for each \( i \),

\[
\text{Var}X_i = \frac{1}{N} (1 - \frac{1}{N}) = \frac{N-1}{N^2}.
\]

Also, for \( i \neq j \),

\[
E[X_iX_j] = P(X_i = 1, X_j = 1) = \frac{1}{N^2(N-1)},
\]

thus

\[
\text{Cov}(X_i, X_j) = E[X_iX_j] - (EX_i)(EX_j) = \frac{1}{N^2(N-1)}.
\]

Consequently

\[
\text{Var}S_N = \text{Var}\left(\sum_{i=1}^{N} X_i\right) = \frac{N-1}{N} + 2 \left(\frac{N}{2}\right) \frac{1}{N^2(N-1)} = \frac{N-1}{N} + \frac{1}{N} = 1.
\]
The **quick-sort algorithm.** Suppose we are presented with a set of $n$ distinct real numbers $x_1, \ldots, x_n$ and that we want to sort them, that is, put them in increasing order. An efficient way for accomplishing this is the quick-sort algorithm defined as follows. When $n = 2$, the algorithm compares the two values and put then in the appropriate order. When $n > 2$, one of the element is randomly chosen, say it is $x_i$, and then all of the other values are compared with $x_i$. Those smaller that $x_i$ are put in a bracket to the left of $x_i$ and those larger than $x_i$ in a bracket to the right of $x_i$. The algorithm then repeat itself on these brackets and continues until all values have been sorted. Let $S_n$ denote the number of comparisons that it takes the quick-sort algorithm to sort $n$ distinct values, then $ES_n$ is a measure of the effectiveness of this algorithm. To find $ES_n$, we first express $S_n$ as the sum of other random variables as follows. First, give the following names to the numbers to be sorted: let 1 stands for the smallest, let 2 stands for the next smallest, and so on. Then, for $1 \leq i < j \leq n$, Let $X_{i,j}$ equal to 1 if $i$ and $j$ are ever directly compared and equal to 0 otherwise. Then

$$S_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_{i,j}.$$  

consequently 

$$ES_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n E X_{i,j} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(\ i \text{ and } j \text{ are ever compared}).$$

To determine $P(\ i \text{ and } j \text{ are ever compared})$, note that the values $i, i+1, \ldots, j-1, j$ will remain in the same bracket if the number chosen for the first comparison is not between $i$ and $j$. Thus all the values will remain in the same bracket until the first time that one of them is chosen as a comparison value. At that point all the values between $i$ and $j$ will be compared with this comparison value. Now, if this comparison value is neither $i$ nor $j$, then upon comparison, $i$ will go to a left bracket and $j$ to a right bracket, and thus $i$ and $j$ will never be compared. On the other hand, if comparison value of the set $i, i+1, \ldots, j-1, j$ is either $i$ or $j$, then there will be a direct comparison between $i$ and $j$. Now, given that the comparison value is one of the values between $i$ and $j$, it is equally likely to be any of these $j-i+1$ values, and thus the probability that it is either $i$ or $j$ is $2/(j-i+1)$. Therefore

$$P(\ i \text{ and } j \text{ are ever compared}) = \frac{2}{j-i+1}.$$  

Therefore 

$$ES_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j-i+1}$$

which is approximately $2n \log n$.

To deal with the next example, we need the following elementary result.

**Lemma 8** For arbitrary real numbers $x_1, X - 2, \ldots, x_n$,

$$\max_i x_i = \sum_i x_i - \sum_{i<j} \min(x_i, x_j) + \sum_{i<j<k} \min(x_i, x_j, x_k) + \cdots + (-1)^{n+1} \min(x_1, \ldots, x_n).$$

13
Proof. This can proved directly using induction. Here we sketch a proof using the inclusion-exclusion formula for probabilities. In the case when all \( x_i \) are in \([0, 1]\), we can define \( A_i = \{ U \leq x_i \} \), where \( U \) is a uniform random variable on \([0, 1]\). Then the result follows easily from the inclusion-exclusion formula for probabilities. The general case follows easily from this special case.

Example 5. Suppose that there are \( n \) different types of coupons, and that each time one collects a coupon it is, independent of the previous coupons collected, a type \( i \) coupon with probability \( p_i \), \( \sum_{i=1}^{n} p_i = 1 \). Let \( X \) be the number of coupons one needs to collect to get a complete set of at least one of each type.

Solution. Let \( X_i \) be the number of coupons one needs to collect to get a type \( i \), then \( X = \max(X_1, X_2, \ldots, X_n) \). Because each new coupon obtained is a type \( i \) coupon with probability \( p_i \), \( X_i \) is a geometric random variable with parameter \( p_i \). Because \( \min(X_i, X_j) \) is the number of coupons one needs to collect to get either a type \( i \) or a type \( j \), we know that, for \( i \neq j \), \( \min(X_i, X_j) \) is a geometric random variable with parameter \( p_i + p_j \). Similarly, \( \min(X_i, X_j, X_k) \) is a geometric random variable with parameter \( p_i + p_j + p_k \), and so on. Therefore

\[
E[X] = \sum_i \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \sum_{i<j<k} \frac{1}{p_i + p_j + p_k} + \cdots + (-1)^{n+1} \frac{1}{p_1 + \cdots + p_n}.
\]

The Probabilistic Method The probabilistic method is a technique for analyzing the properties of the elements of a set by introducing probabilities on the set, and then studying an element chosen according to those probabilities. We have already talked about this technique in Chapter 1. Now we are going to see how it can sometimes be used to bound complicated functions.

Let \( f \) be a function defined on a finite set \( S \), and suppose that we are interested in

\[
M = \max_{s \in S} f(s) \quad m = \min_{s \in S} f(s).
\]

useful bounds for \( m \) can often be obtained by letting \( S \) be a random element of \( S \) for which the \( Ef(S) \) is computable and noting that

\[
M \geq Ef(S), \quad m \leq Ef(S)
\]

with strict inequalities if \( f(S) \) is not a constant. That is, \( Ef(S) \) is a lower bound for \( M \) and an upper bound for \( m \).

Now we look at a concrete example.

The maximum number of Hamiltonian paths in a tournament. A round-robin tournament of \( n \geq 2 \) contestants is one in which each one of the \( \binom{n}{2} \) pairs of contestants play each other exactly once, with the outcome of any play being that one of the contestants wins and the other loses. Suppose that the players are labeled \( 1, 2, \ldots, n \). The permutation \( i_1, i_2, \ldots, i_n \) is said to be a Hamiltonian path if \( i_1 \) beats \( i_2 \), \( i_2 \) beats \( i_3 \), \ldots, and \( i_{n-1} \) beats \( i_n \). A problem of some interest is to determine the maximum number of Hamiltonian paths.

We claim that there is an outcome of the tournament that results in more than \( n! / 2^{n-1} \) Hamiltonian paths. To begin, let the outcome of the tournament specify the results of each of the \( \binom{n}{2} \) games played, and let \( S \) denote the set of all \( 2\binom{n}{2} \) possible tournament outcomes. Then, with

14
$f(s)$ defined as the number of Hamiltonian paths that result when the outcome in $s \in S$, we are asked to prove that

$$\max_{s \in S} f(s) \geq \frac{n!}{2^{n-1}}.$$  

To show this, consider the randomly chosen outcome $S$ that is obtained when the results of the $\binom{n}{2}$ games are independent, with each contestant being equally likely to win each encounter. To determine $Ef(S)$, number the $n!$ permutations, and let for $i = 1, 2, \ldots, n!$,

$$X_i = \begin{cases} 1, & \text{if the permutation } i \text{ is a Hamiltonian path} \\ 0, & \text{otherwise} \end{cases}$$

Then $f(S) = \sum_i X_i$, hence

$$Ef(S) = \sum_i EX_i.$$  

Since

$$EX_i = P(X_i = 1) = 2^{n-1},$$

we have

$$Ef(S) \geq \frac{n!}{2^{n-1}}.$$  

Because $f(S)$ is not a constant, the inequality above implies that there is an outcome of the tournament that results in more than $n!/2^{n-1}$ Hamiltonian paths.