Chapter 1

Random Variables and their Distributions

1.1 Random Variables

Definition 1.1. A random variable $X$ is a function that assigns one and only one numerical value to each outcome of an experiment, that is

$$X : \Omega \to \mathcal{X} \subseteq \mathbb{R},$$

where $\Omega$ is a sample space, $\mathcal{X}$ is a set of all possible values of $X$. □

We will denote rvs by capital letters: $X$, $Y$ or $Z$ and their values by small letters: $x$, $y$ or $z$ respectively.

There are two types of rvs: discrete and continuous. Random variables that take a countable number of values are called discrete. Random variables that take values from an interval of real numbers are called continuous.
Example 1.1. In an efficacy preclinical trial a drug candidate is tested on mice. If the observed response can either be “efficacious” or “non-efficacious”, then we have

\[ \Omega = \{ \omega_1, \omega_2 \}, \]

where \( \omega_1 \) is an outcome of the experiment meaning efficacious response to the drug, \( \omega_2 \) is an outcome meaning non-efficacious response. Assume that \( P(\omega_1) = p \) and \( P(\omega_2) = q = 1 - p \).

It is natural to define a random variable \( X \) as follows:

\[ X(\omega_1) = 1, \quad X(\omega_2) = 0. \]

Then

\[ P_X(X = 1) = P(\{ \omega_j \in \Omega : X(\omega_j) = 1 \}) = P(\omega_1) = p \]

and

\[ P_X(X = 0) = P(\{ \omega_j \in \Omega : X(\omega_j) = 0 \}) = P(\omega_2) = q. \]

Here \( \mathcal{X} = \{ x_1, x_2 \} = \{ 0, 1 \} \).

Here, \( P_X \) is an induced probability function on \( \mathcal{X} \), defined in terms of the probability function on \( \Omega \).

Example 1.2. In the same experiment we might observe a continuous outcome, for example, time to a specific response. Then

\[ \Omega = \{ \omega \in (0, \infty) \}. \]

Many random variables could be of interest in this case. For example, it could be a continuous rv, such as

\[ X : \Omega \to \mathcal{X}, \]
where 
\[ X(\omega) = \ln \omega. \]
Here \( \mathcal{X} = (-\infty, \infty) \). Then, for example,
\[ P(X \in [-1, 1]) = P\{\omega \in \Omega : X(\omega) \in [-1, 1]\} = P(\omega \in [e^{-1}, e]). \]

On the other hand, if we are interested in just two events such as, for example: the time is less than a pre-specified value, say \( t^* \), or that it is more than this value, than we categorize the outcome and the sample space is
\[ \Omega = \{\omega_1, \omega_2\}, \]
where \( \omega_1 \) means that the observed time to a reaction was shorter than or equal to \( t^* \), \( \omega_2 \) means that the time was longer than \( t^* \). Then we can define a discrete random variable as in Example 1.1.

**Definition 1.2.** If \( g : \mathbb{R} \to \mathbb{R} \) is a monotone, continuous function, \( X : \Omega \to \mathcal{X} \subseteq \mathbb{R} \) is a random variable and \( Y = g(X) \), then \( Y \) is a random variable such that \( Y : \Omega \to \mathcal{Y} \subseteq \mathbb{R} \) and \( Y(\omega) = g(X(\omega)) \) for all \( \omega \in \Omega \).

**Example 1.3.** In an experiment we toss two fair coins. Then the sample space is the set of pairs, i.e., \( \Omega = \{tt, th, ht, hh\} \), where \( t \) denotes tail and \( h \) denotes head. Define a rv \( X \) as the number of heads, that is \( X : \Omega \to \mathcal{X} \), where \( \mathcal{X} = \{0, 1, 2\} \), i.e.:
\[ 
X(tt) = 0, \\
X(th) = X(ht) = 1, \\
X(hh) = 2. 
\]
Then \( Y = 2 - X \) denotes the number of tails obtained in the experiment and \( Y : \Omega \to \mathcal{Y} \), where \( \mathcal{Y} = \{0, 1, 2\} \). That is, \( Y(\omega) = 2 - X(\omega) \), for example \( Y(tt) = 2 - X(tt) = 2 \).
1.2 Distribution Functions

**Definition 1.3.** The probability of the event \((X \leq x)\) expressed as a function of \(x \in \mathbb{R}\) and denoted by
\[
F_X(x) = P_X(X \leq x)
\]
is called the **cumulative distribution function** (cdf) of the rv \(X\).

**Example 1.4.** The cdf of the rv defined in Example 1.1 can be written as
\[
F_X(x) = \begin{cases} 
0, & \text{for } x \in (-\infty, 0); \\
q, & \text{for } x \in [0, 1); \\
q + p = 1, & \text{for } x \in [1, \infty). 
\end{cases}
\]

Properties of cumulative distribution functions are given in the following theorem.

**Theorem 1.1.** The function \(F(x)\) is a cdf iff the following conditions hold:

(i) The function is nondecreasing, that is, if \(x_1 < x_2\) then \(F_X(x_1) \leq F_X(x_2)\);

(ii) \(\lim_{x \to -\infty} F_X(x) = 0\);

(iii) \(\lim_{x \to \infty} F_X(x) = 1\);

(iv) \(F(x)\) is right-continuous.

**Example 1.5.** We will show that
\[
F_X(x) = \frac{1}{1 + e^{-x}}
\]
1.3. Density and Mass Functions

is the cdf of a rv $X$. It is enough to show that the function meets the requirements of Theorem 1.1. To check that the function in non-decreasing we may check the sign of the derivative of $F(x)$. It is

$$F'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0,$$

showing that $F(x)$ is increasing.

$$\lim_{x \to -\infty} F_X(x) = 0 \text{ since } \lim_{x \to -\infty} e^{-x} = \infty;$$

$$\lim_{x \to \infty} F_X(x) = 1 \text{ since } \lim_{x \to \infty} e^{-x} = 0.$$ 

Furthermore, it is a continuous function, not only right-continuous.

Now we can define discrete and continuous rvs more formally.

**Definition 1.4.** A random variable $X$ is **continuous** if $F_X(x)$ is a continuous function of $x$. A random variable $X$ is **discrete** if $F_X(x)$ is a step function of $x$.

### 1.3 Density and Mass Functions

#### 1.3.1 Discrete Random Variables

Values of a discrete rv are elements of a countable set $\{x_1, x_2, \ldots\}$. We associate a number $p_X(x_i) = P_X(X = x_i)$ with each value $x_i$, $i = 1, 2, \ldots$, such that:

1. $p_X(x_i) \geq 0$ for all $i$;
2. $\sum_{i=1}^{\infty} p_X(x_i) = 1$. 

Note that
\[ F_X(x_i) = P_X(X \leq x_i) = \sum_{x \leq x_i} p_X(x), \quad (1.1) \]
\[ p_X(x_i) = F_X(x_i) - F_X(x_{i-1}). \quad (1.2) \]

The function \( p_X \) is called the *probability mass function* (pmf) of the random variable \( X \), and the collection of pairs
\[ \{(x_i, p_X(x_i)), i = 1, 2, \ldots\} \quad (1.3) \]
is called the *probability distribution* of \( X \). The distribution is usually presented in either tabular, graphical or mathematical form.

*Example 1.6.* Consider Example 1.1, but now, we have \( n \) mice and we observe efficacy or no efficacy for each mouse independently. We are interested in the number of mice which respond positively to the applied drug candidate. If \( X_i \) is a random variable as defined in Example 1.1 for each mouse, and we may assume that the probability of a positive response is the same for all mice, then we may create a new random variable \( X \) as the sum of all \( X_i \), that is,
\[ X = \sum_{i=1}^{n} X_i. \]

\( X \) denotes \( x \) successes in \( n \) independent trials and it has a binomial distribution, which we denote by
\[ X \sim \text{Bin}(n, p), \]
where \( p \) is the probability of success. The pmf of a binomially distributed rv \( X \) with parameters \( n \) and \( p \) is
\[ P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n, \]
where \( n \) is a positive integer and \( 0 \leq p \leq 1 \).
For example, for \( n = 8 \) and the probability of success \( p = 0.4 \) we obtain the following values of the pmf and the cdf:

<table>
<thead>
<tr>
<th>( k )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = k) )</td>
<td>0.0168</td>
<td>0.0896</td>
<td>0.2090</td>
<td>0.2787</td>
<td>0.2322</td>
<td>0.1239</td>
<td>0.0413</td>
<td>0.0079</td>
<td>0.0007</td>
</tr>
<tr>
<td>( P(X \leq k) )</td>
<td>0.0168</td>
<td>0.1064</td>
<td>0.3154</td>
<td>0.5941</td>
<td>0.8263</td>
<td>0.9502</td>
<td>0.9915</td>
<td>0.9993</td>
<td>1</td>
</tr>
</tbody>
</table>

Here, for example, we could say that the probability of 5 mice responding positively to the drug candidate is \( P(X = 5) = 0.1239 \), but the probability that at least 5 mice will respond positively is \( P(X \leq 5) = 0.8263 \).

### 1.3.2 Continuous Random Variables

Values of a continuous rv are elements of an uncountable set, for example a real interval. A cdf of a continuous rv is a continuous, nondecreasing, differentiable function. We define the *probability density function* (pdf) of a continuous rv as:

\[
    f_X(x) = \frac{d}{dx}F_X(x).
\]

Hence,

\[
    F_X(x) = \int_{-\infty}^{x} f_X(t)dt.
\]

Similarly to the properties of the probability mass function of a discrete rv we have the following properties of the probability density function:

1. \( f_X(x) \geq 0 \) for all \( x \in \mathcal{X} \);
2. \( \int_{\mathcal{X}} f_X(x)dx = 1 \).

Probability of an event that \( X \in (-\infty, a) \), is expressed as an integral

\[
    P_X(-\infty < X < a) = \int_{-\infty}^{a} f_X(x)dx = F_X(a)
\]
or for a bounded interval \((b, c)\) as

\[
P_X(b < X < c) = \int_{b}^{c} f_X(x)\,dx = F_X(c) - F_X(b) .
\] (1.7)

An interesting difference from a discrete rv is that for a \(\delta > 0\)

\[
P_X(X = x) = \lim_{\delta \to 0} (F_X(x + \delta) - F_X(x)) = 0.
\]

**Example 1.7.** Let \(X\) be a random variable with the pdf given by

\[
f(x) = c \sin x I_{[0, \pi]}.
\]

Find the value of \(c\).

Here we use the notation of the indicator function \(I_X(x)\) whose meaning is as follows:

\[
I_X(x) = \begin{cases} 
1, & \text{if } x \in X; \\
0, & \text{otherwise}.
\end{cases}
\]

For a function to be pdf of a random variable, it must meet two conditions: it must be non-negative and integrate to 1 (over the whole region). Hence \(c\) must be a positive number and

\[
\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\pi} c \sin x\,dx = 1.
\]

We have

\[
\int_{0}^{\pi} c \sin x\,dx = -c \cos x \Big|_{0}^{\pi} = -c(-1 - 1) = 2c.
\]

Hence \(c = 1/2\). \(\square\)
1.4  Expected Values

Definition 1.5. The expected value of a function \( g(X) \) is defined by

\[
E(g(X)) = \begin{cases} 
\int_{-\infty}^{\infty} g(x) f(x) dx, & \text{for a continuous r.v.,} \\
\sum_{j=0}^{\infty} g(x_j) p(x_j) & \text{for a discrete r.v.,}
\end{cases}
\]

and \( g \) is any function such that \( E|g(X)| < \infty \). □

Two important special cases of \( g \) are:

- \( g(X) = X \), which gives the mean \( E(X) \) also denoted by \( EX \),
- \( g(X) = (X - EX)^2 \), which gives the variance \( var(X) = E(X - EX)^2 \).

The following relation is very useful while calculating the variance

\[
E(X - EX)^2 = E X^2 - (EX)^2.
\] (1.8)

Example 1.8. Let \( X \) be a random variable such that

\[
f(x) = \begin{cases} 
\frac{1}{2} \sin x, & \text{for } x \in [0, \pi], \\
0 & \text{otherwise.}
\end{cases}
\]

Then the expectation and variance are following

- Expectation

\[
EX = \frac{1}{2} \int_{0}^{\pi} x \sin x dx = \frac{\pi}{2}.
\]
• Variance
\[
\text{var}(X) = E(X^2) - (E(X))^2
\]
\[
= \frac{1}{2} \int_0^\pi x^2 \sin x \, dx - \left(\frac{\pi}{2}\right)^2
\]
\[
= \frac{\pi^2}{4}.
\]

The following useful properties of the expectation follow from properties of integration (summation).

**Theorem 1.2.** Let \(X\) be a random variable and let \(a\), \(b\) and \(c\) be constants. Then for any functions \(g(x)\) and \(h(x)\) whose expectations exist we have:

(a) \(E[ag(X) + bh(X) + c] = a E[g(X)] + b E[h(X)] + c\);

(b) If \(g(x) \geq h(x)\) for all \(x\), then \(E(g(X)) \geq E(h(X))\);

(c) If \(g(x) \geq 0\) for all \(x\), then \(E(g(X)) \geq 0\);

(d) If \(a \geq g(x) \geq b\) for all \(x\), then \(a \geq E(g(X)) \geq b\).

The variance of a random variable together with the mean are the most important parameters used in the theory of statistics. The following theorem is a result of the properties of the expectation function.

**Theorem 1.3.** If \(X\) is a random variable with a finite variance, then for any constants \(a\) and \(b\),

\[
\text{var}(aX + b) = a^2 \text{var} X.
\]
1.5. FAMILIES OF DISTRIBUTIONS

Proof

\[ \text{var}(aX + b) = E((aX + b)^2) - \{E(aX + b)\}^2 \]
\[ = E(a^2X^2 + 2abX + b^2) - (a^2\{E X\}^2 + 2ab E X + b^2) \]
\[ = a^2 E X^2 - a^2\{E X\}^2 = a^2 \text{var}(X) \]

\[ \square \]

1.5 Families of Distributions

Distributions of random variables often depend on some parameters, like \( p, \lambda, \mu, \sigma \) or other. These values are usually unknown and their estimation is one of the most important problems in statistical analyzes. These parameters determine some characteristics of the shape of the pdf/pmf of the random variable. It can be location, spread, skewness etc.

1.5.1 Some families of discrete distributions

1. Uniform \( U(n) \) (equal mass at each outcome): The support set and the pmf are, respectively, \( \mathcal{X} = \{x_1, x_2, \ldots, x_n\} \) and

\[ P(X = x_i) = \frac{1}{n}, \quad x_i \in \mathcal{X}, \]

where \( n \) is a positive integer. In the special case of \( \mathcal{X} = \{1, 2, \ldots, n\} \) we have

\[ E(X) = \frac{n+1}{2}, \quad \text{var}(X) = \frac{(n+1)(n-1)}{12}. \]

Examples:
(a) $X \equiv$ first digit in a randomly selected sequence of length $n$ of 5 digits;
(b) $X \equiv$ randomly selected student in a class of 15 students.

2. Bernoulli($p$) (only two possible outcomes, usually called “success” and “failure”): The support set and the pmf are, respectively, $\mathcal{X} = \{0, 1\}$ and

$$P(X = x) = p^x(1 - p)^{1-x}, \quad x \in \mathcal{X},$$

where $p \in [0, 1]$ is the probability of success.

$$E(X) = p, \quad \text{var}(X) = p(1 - p).$$

Examples:
(a) $X \equiv$ an outcome of tossing a coin;
(b) $X \equiv$ detection of a fault in a tested semiconductor chip;
(c) $X \equiv$ guessed answer in a multiple choice question.

3. Bin($n, p$) (number of success in $n$ independent trials): The support set and the pmf are, respectively, $\mathcal{X} = \{0, 1, 2, \ldots n\}$ and

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}, \quad x \in \mathcal{X},$$

where $p \in [0, 1]$ is the probability of success.

$$E(X) = np, \quad \text{var}(X) = np(1 - p).$$

Examples:
(a) $X \equiv$ number of heads in several tosses of a coin;
(b) $X \equiv$ number of semiconductor chips in several faulty chips in which a test finds a defect;
(c) $X \equiv$ number of correctly guessed answers in a multiple choice test of $n$ questions.
4. Geom\((p)\) (the number of independent Bernoulli trials until first “success”): The support set and the pmf are, respectively, \(\mathcal{X} = \{1, 2, \ldots \}\) and

\[
P(X = x) = p(1 - p)^{x-1} = pq^{x-1}, \quad x \in \mathcal{X},
\]

where \(p \in [0, 1]\) is the probability of success, \(q = 1 - p\).

\[
E(X) = \frac{1}{p}, \quad \text{var}(X) = \frac{1 - p}{p^2}.
\]

Here we will prove the formula for the expectation. By the definition of the expectation we have

\[
E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1}.
\]

Note that for \(x \geq 1\) the following equality holds

\[
\frac{d}{dq}(q^x) = xq^{x-1}.
\]

Hence,

\[
E(X) = p \sum_{x=1}^{\infty} xq^{x-1} = p \sum_{x=1}^{\infty} \frac{d}{dq}(q^x)
\]

\[
= p \frac{d}{dq} \left( \sum_{x=1}^{\infty} q^x \right).
\]

The latter is the sum of a geometric sequence and is equal to \(q/(1 - q)\). Therefore,

\[
E(X) = p \frac{d}{dq} \left( \frac{q}{1 - q} \right) = p \left( \frac{1}{(1 - q)^2} \right) = \frac{1}{p}.
\]

Similar method can be used to show that the \(\text{var}(X) = q/p^2\) (second derivative with respect to \(q\) of \(q^x\) can be applied for this).

Examples include:
(a) $X$ ≡ number of bits transmitted until the first error;
(b) $X$ ≡ number of analyzed samples of air before a rare molecule is detected.

5. Poisson($\lambda$) (a number of outcomes in a period of time or in a part of a space): The support set and the pmf are, respectively, $X = \{0, 1, \ldots\}$ and

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!},$$

where $\lambda > 0$.

$$E(X) = \lambda, \quad \text{var}(X) = \lambda.$$

We will show that $E(X) = \lambda$. The result for the variance can be shown in a similar way. By definition of the expectation we have

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}$$
$$= \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}.$$

Let $z = x - 1$. Then we obtain

$$E(X) = \sum_{z=0}^{\infty} \frac{\lambda^{z+1} e^{-\lambda}}{z!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \lambda$$

as $\frac{\lambda^z e^{-\lambda}}{z!}$ is the pdf of $Z$.

Examples:

(a) count blood cells within a square of a haemocytometer slide;
(b) number of caterpillars on a leaf;
(c) number of plants of a rear variety in a square meter of a meadow;
1.5. FAMILIES OF DISTRIBUTIONS

1.5.2 Some families of continuous distributions

The most common continuous distribution is called Normal and denoted by $\mathcal{N}(\mu, \sigma^2)$. The density function is given by

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

There are two parameters which tell us about the location $E(X) = \mu$ and the spread $\text{var}(X) = \sigma^2$ of the density curve.

There is an extensive theory of statistical analysis for data which are realizations of normally distributed random variables. This distribution is most common in applications, but sometimes it is not feasible to assume that what we observe can indeed be a sample from such a population. In Example 1.2 we observe time to a specific response to a drug candidate. Such a variable can only take nonnegative values, while the normal rv’s domain is $\mathbb{R}$. A lognormal distribution is
often used in such cases. $X$ has a lognormal distribution if $\log X$ is normally distributed.

Other popular continuous distributions include:

1. Uniform, $\mathcal{U}(a, b)$: The support set and the pdf are, respectively, $X = [a, b]$ and

   $$f_X(x) = \frac{1}{b - a} I_{[a,b]}(x).$$

   $$E(X) = \frac{a + b}{2}, \quad \text{var}(X) = \frac{(b - a)^2}{12}.$$ 

2. $\text{Exp}(\lambda)$: The support set and the pdf are, respectively, $X = [0, \infty)$ and

   $$f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x),$$

   where $\lambda > 0$.

   $$E(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}.$$ 

Examples:

- $X \equiv$ the time between arrivals of e-mails on your computer;
- $X \equiv$ the distance between major cracks in a motorway;
- $X \equiv$ the life length of car voltage regulators.

3. Gamma($\alpha, \lambda$): An exponential rv describes the length of time or space between counts. The length until $r$ counts occur is a generalization of such a process and the respective rv is called the Erlang random variable. Its pdf is given by

   $$f_X(x) = \frac{x^{r-1} \lambda^r e^{-\lambda x}}{(r - 1)!} I_{[0,\infty)}(x), \quad r = 1, 2, \ldots.$$
The Erlang distribution is a special case of the \textit{Gamma distribution} in which the parameter $r$ is any positive number, usually denoted by $\alpha$. In the gamma distribution we use the generalization of the factorial represented by the gamma function, as follows.

$$
\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \text{ for } \alpha > 0.
$$

A recursive relationship that may be easily shown integrating the above equation by parts is

$$
\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).
$$

If $\alpha$ is a positive integer, then

$$
\Gamma(\alpha) = (\alpha - 1)!
$$

since $\Gamma(1) = 1$.

The pdf is

$$
f_X(x) = \frac{x^{\alpha-1}\lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} I_{[0,\infty)}(x),
$$

where the $\alpha > 0$, $\lambda > 0$ and $\Gamma(\cdot)$ is the gamma function. The mean and the variance of the gamma rv are

$$
E(X) = \frac{\alpha}{\lambda}, \quad \text{var}(X) = \frac{\alpha}{\lambda^2}.
$$

4. $\chi^2(\nu)$: A rv $X$ has a \textit{Chi-square distribution with $\nu$ degrees of freedom} iff $X$ is gamma distributed with parameters $\alpha = \frac{\nu}{2}$ and $\lambda = \frac{1}{2}$. This distribution is used extensively in interval estimation and hypothesis testing. Its values are tabulated.

\textit{Example 1.9.} Denote by $X$ the times until $\alpha$ responses get on an on-line computer terminal. Assume that rv $X$ has approximately a gamma distribution with mean four seconds and variance eight seconds$^2$. 
(a) Write the probability density function for $X$.

(b) What is the probability that $\alpha$ responses are received on the computer terminal in no more than five seconds?

Solution:

(a) We have

\[ E(X) = \frac{\alpha}{\lambda} = 4 \]

\[ \text{var}(X) = \frac{\alpha}{\lambda^2} = 8 \]

This gives $\lambda = 1/2$ and $\alpha = 2$ and so the pdf is

\[ f_X(x) = \frac{x^{\alpha-1} \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} I_{[0, \infty)}(x) = \frac{1}{4} xe^{-1/2} x I_{[0, \infty)}(x), \]

as $\Gamma(2) = 1$.

(b) This is chi-square distribution with $\nu = 4$ degrees of freedom and the value of $F(5) = P_X(X \leq 5) = 0.7127$ can be found in the statistical tables.

1.6 Moments and Moment Generating Functions

**Definition 1.6.** The $n$th moment ($n \in \mathbb{N}$) of a random variable $X$ is defined as

\[ \mu'_n = E X^n \]

The $n$th central moment of $X$ is defined as

\[ \mu_n = E (X - \mu)^n, \]

where $\mu = \mu'_1 = E X$. \qed
Note, that the second central moment is the variance of a random variable $X$, usually denoted by $\sigma^2$.

Moments give an indication of the shape of the distribution of a random variable. Skewness and kurtosis are measured by the following functions of the third and fourth central moment respectively:

**the coefficient of skewness** is given by

$$\gamma_1 = \frac{\mathbb{E}(X - \mu)^3}{\sigma^3} = \frac{\mu_3}{\sqrt{\mu_2^3}}.$$  

**the coefficient of kurtosis** is given by

$$\gamma_2 = \frac{\mathbb{E}(X - \mu)^4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3.$$  

Moments can be calculated from the definition or by using so called moment generating function.

**Definition 1.7.** The **moment generating function (mgf)** of a random variable $X$ is a function $M_X : \mathbb{R} \rightarrow [0, \infty)$ given by

$$M_X(t) = \mathbb{E} e^{tX},$$

provided that the expectation exists for $t$ in some neighborhood of zero.  

More explicitly, the mgf of $X$ can be written as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \quad \text{if } X \text{ is continuous},$$

$$M_X(t) = \sum_{x \in \mathcal{X}} e^{tx} P(X = x), \quad \text{if } X \text{ is discrete}.$$
CHAPTER 1. RANDOM VARIABLES AND THEIR DISTRIBUTIONS

The method to generate moments is given in the following theorem.

**Theorem 1.4.** If $X$ has mgf $M_X(t)$, then

$$E(X^n) = M_X^{(n)}(0),$$

where

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)|_0.$$

That is, the $n$-th moment is equal to the $n$-th derivative of the mgf evaluated at $t = 0$.

**Proof.** We will prove the theorem for a discrete rv case. It is similar for a continuous rv. For $n = 1$ we have

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} \sum_{x \in \mathcal{X}} e^{tx} P_X(X = x)$$

$$= \sum_{x \in \mathcal{X}} \left( \frac{d}{dt} e^{tx} \right) P_X(X = x)$$

$$= \sum_{x \in \mathcal{X}} (xe^{tx}) P_X(X = x)$$

$$= E(X e^{tX}).$$

Hence, evaluating the last expression at zero we obtain

$$\frac{d}{dt} M_X(t)|_0 = E(X e^{tX})|_0 = E(X).$$

For $n = 2$ we will get

$$\frac{d^2}{dt^2} M_X(t)|_0 = E(X^2 e^{tX})|_0 = E(X^2).$$

Analogously, it can be shown that for any $n \in \mathbb{N}$ we can write

$$\frac{d^n}{dt^n} M_X(t)|_0 = E(X^n e^{tX})|_0 = E(X^n).$$

$\square$
Example 1.10. Find the mgf of $X \sim \text{Exp}(\lambda)$ and use results of Theorem 1.4 to obtain the mean and variance of $X$.

By definition the mgf can be written as

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) \, dx.$$

For the exponential distribution we have

$$f_X(x) = \lambda e^{-\lambda x} I_{(0, \infty)}(x),$$

where $\lambda \in \mathbb{R}_+$. Hence,

$$M_X(t) = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} \, dx = \lambda \int_{0}^{\infty} e^{(t-\lambda)x} \, dx = \frac{\lambda}{\lambda - t} \text{ provided that } |t| < \lambda.$$

Now, using Theorem 1.4 we obtain the first and the second moments, respectively:

$$E(X) = M_X'(0) = \frac{\lambda}{(\lambda - t)^2} \bigg|_{t=0} = \frac{1}{\lambda},$$

$$E(X^2) = M_X^{(2)}(0) = \frac{2\lambda}{(\lambda - t)^3} \bigg|_{t=0} = \frac{2}{\lambda^2}.$$

Hence, the variance of $X$ is

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

Moment generating functions provide methods for comparing distributions or finding their limiting forms. The following two theorems give us the tools.

Theorem 1.5. Let $F_X(x)$ and $F_Y(y)$ be two cdfs whose all moments exist. Then, if the mgfs of $X$ and $Y$ exist and are equal, i.e., $M_X(t) = M_Y(t)$ for all $t$ in some neighborhood of zero, then $F_X(u) = F_Y(u)$ for all $u$. □
THEOREM 1.6. Suppose that \( \{X_1, X_2, \ldots\} \) is a sequence of random
variables, each with mgf \( M_{X_i}(t) \). Furthermore, suppose that

\[
\lim_{i \to \infty} M_{X_i}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of zero},
\]

and \( M_X(t) \) is an mgf. Then, there is a unique cdf \( F_X \) whose moments are determined by \( M_X(t) \) and, for all \( x \) where \( F_X(x) \) is continuous, we have

\[
\lim_{i \to \infty} F_{X_i}(x) = F_X(x).
\]

This theorem means that the convergence of mgfs implies conver-
gence of cdfs.

EXAMPLE 1.11. We know that the Binomial distribution can be ap-
proximated by a Poisson distribution when \( p \) is small and \( n \) is large. Using the above theorem we can confirm this fact.

The mgf of \( X_n \sim \text{Bin}(n, p) \) and of \( Y \sim \text{Poisson}(\lambda) \) are, respec-
tively:

\[
M_{X_n}(t) = [pe^t + (1 - p)]^n, \quad M_Y(t) = e^{\lambda(e^t - 1)}.
\]

We will show that the mgfs of \( X_n \) tend to the mgf of \( Y \), where \( np \to \lambda \).

We will need the following useful result given in the lemma:

**Lemma 1.1.** Let \( a_1, a_2, \ldots \) be a sequence of numbers converging to \( a \), that is, \( \lim_{n \to \infty} a_n = a \). Then

\[
\lim_{n \to \infty} \left( 1 + \frac{a_n}{n} \right)^n = e^a.
\]
Now, we can write

$$M_{X_n}(t) = (pe^t + (1 - p))^n$$

$$= \left( 1 + \frac{1}{n} np(e^t - 1) \right)^n$$

$$= \left( 1 + \frac{np(e^t - 1)}{n} \right)^n$$

$$\xrightarrow{n \to \infty} e^{\lambda(e^t-1)} = M_Y(t).$$

Hence, by Theorem 1.6 the Binomial distribution converges to a Poisson distribution.