Module:	Probability and Stochastic Modelling
Term:	Autumn 2024
Lecturer:	Richard Pymar

2 Univariate Distribution Theory for Discrete and Continuous Random Variables

Distribution functions are a good tool for describing random variables. In this lecture, we will study some important and useful distribution functions. Meanwhile, we will introduce some other concepts such as expectation, variance, and, more generally, moments which describe random variables in the view of the average behaviour. We start with two important subclasses of random variables: the collection of 'discrete' random variables and the collection of 'continuous' random variables.

2.1 Discrete and Continuous Random Variables

Definition 2.1 (Discrete random variable, mass function, range) The random variable X is called discrete if it takes values in some countable subset $\{x_1, x_2, \ldots\}$, only, of $\mathbb{R}$. The discrete random variable has (probability) mass function $p_X : \mathbb{R} \longrightarrow [0,1]$ given by $p_X(x) = \mathbb{P}(X = x)$ and range given by $R_X = \{x \in \mathbb{R} : p_X(x) > 0\}$.

The distribution and mass function are related by

$$F_X(x) = \sum_{i:x_i \le x} p_X(x_i), \quad p_X(x) = F_X(x) - \lim_{y \nearrow x} F_X(y).$$

We can see that the distribution function of a discrete variable has jump discontinuities at each $x \in R_X$ and is constant in between; such a distribution is called **atomic**.

Lemma 2.1 The probability mass function $p_X : \mathbb{R} \longrightarrow [0,1]$ satisfies:

1. the set of x such that $p_X(x) > 0$ is countable,
2. $\sum_i p_X(x_i) = 1$, where $x_1, x_2, \cdots$ are the values of x such that $p_X(x) > 0$.

Lemma 1 in fact characterizes the probability mass function.

Definition 2.2 (Continuous random variable, density function, range) The random variable X is called continuous if its distribution function can be expressed as

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds \qquad \text{for } x \in \mathbb{R},$$

for some integrable function $f_X : \mathbb{R} \longrightarrow [0, \infty)$ called the (probability) density function (pdf) of X and range $R_X = \{x \in \mathbb{R} : f_X(x) > 0\}$.

The distribution function of a continuous random variable is certainly continuous (actually it is 'absolutely continuous'). Note that when we refer to a continuous random variable X we are asserting that the distribution function of the random variable rather than the random variable (function) X itself is continuous.

The density function, $f_X(\cdot)$, is not prescribed uniquely by the above since two integrable functions which take identical values except at some specific point have equal integrals. However, if $F_X(\cdot)$ is differentiable at s then we shall normally set $f_X(s) = F_X'(s)$.

Lemma 2.2 If X has density function $f_X(\cdot)$ then

1. $f_X(x) \geq 0$,
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$,
3. $\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)dx$,
4. $\mathbb{P}(X = x) = 0,$ for all $x \in \mathbb{R}$.

Proof (Proof of 4.) Noting that

$$\mathbb{P}(X = x) \leq \mathbb{P}\{x - h < X \leq x\} = \int_{x-h}^{x} f_X(s)ds,$$

and taking the limit, we get

$$0 \leq \mathbb{P}(X = x) \leq \lim_{h \to 0} \int_{x-h}^{x} f_X(s)ds = 0.$$

Remark 2.1

1. Properties 1 and 2 are also the characterization of a density function, that is, if an integrable function $f_X(\cdot)$ satisfies properties 1 and 2, then $F_X(x) = \int_{-\infty}^{x} f_X(s)ds$ is a distribution function of some continuous random variable.
2. Property 4 tells us that one of the great differences between discrete and continuous random variables is that continuous random variables satisfy $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$. So it is impossible and meaningless to describe continuous random variables by defining $\mathbb{P}(X = x)$ for all $x \in \mathbb{R}$ as we did for discrete random variables.
3. Property 4 also shows that null events need not be the impossible event $\varnothing$ and events with probability 1 need not be the certain event Ω.

The numerical value $f_X(x)$ is not a probability. However we can think of $f_X(x)\Delta x$ as the 'element of probability' $\mathbb{P}(x < X \le x + \Delta x)$ since

$$\mathbb{P}(x < X \le x + \Delta x) = \int_x^{x+\Delta x} f_X(s)ds \simeq f_X(x)\Delta x.$$

These are heuristic, and not rigorous, considerations.

In many cases proofs of results for discrete random variables can be written for continuous random variables by replacing the summation sign by an integral sign, and the probability mass at x, namely $p_X(x)$, by $f_X(x)dx$ in its stead.

Example 2.1 (Continuous variables) A straight rod is flung down at random onto a horizontal plane and the angle ω between the rod and true north is measured. The result is a number in $\Omega = [0, 2\pi)$. Let's dismiss concerns about what $\mathcal{F}$ should be for the moment; we can just suppose that $\mathcal{F}$ contains all 'nice' subsets of Ω, including the collection of open subintervals such as (a, b), where $0 \le a < b < 2\pi$. The implicit symmetry suggests the probability measure $\mathbb{P}$ satisfies $\mathbb{P}((a, b)) = (b - a)/(2\pi)$; that is to say, the probability that the angle lies in some interval is directly proportional to the length of the interval. Here are two random variables X and Y:

$$X(\omega) = \omega, \quad Y(\omega) = \omega^2.$$

Notice that Y is a function of X in that $Y = X^2$. The distribution functions of X and Y are

$$F_X(x) = \begin{cases} 0 & x < 0, \\ x/2\pi & 0 \le x < 2\pi, \\ 1 & x \ge 2\pi, \end{cases} \qquad F_Y(y) = \begin{cases} 0 & y < 0, \\ \sqrt{y}/2\pi & 0 \le y < 4\pi^2, \\ 1 & y \ge 4\pi^2. \end{cases}$$

To see this, let $0 \le x < 2\pi$ and $0 \le y < 4\pi^2$. Then

$$\begin{aligned} F_X(x) &= \mathbb{P}(\{\omega \in \Omega : 0 \le X(\omega) \le x\}) \\ &= \mathbb{P}(\{\omega \in \Omega : 0 \le \omega \le x\}) = x/2\pi \\ F_Y(y) &= \mathbb{P}(\{\omega : Y(\omega) \le y\}) \\ &= \mathbb{P}(\omega : \omega^2 \le y) = \mathbb{P}(\{\omega : 0 \le \omega \le \sqrt{y}\}) \\ &= \mathbb{P}(X \le \sqrt{y}) = \sqrt{y}/2\pi. \end{aligned}$$

The random variables X and Y are continuous because

$$F_X(x) = \int_{-\infty}^x f_X(u)du, \quad F_Y(y) = \int_{-\infty}^y f_Y(u)du,$$

where

$$f_X(u) = \begin{cases} \frac{1}{2\pi} & \text{if } 0 \le u < 2\pi \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(u) = \begin{cases} \frac{u^{-\frac{1}{2}}}{4\pi} & \text{if } 0 \le u < 4\pi^2 \\ 0 & \text{otherwise} \end{cases}$$

Example 2.2 (A random variable which is neither continuous nor discrete) A coin is tossed, and a head turns up with probability $p\ (= 1-q)$. If a head turns up then a rod is flung on the ground and the angle measured as in Example 2.1. Then $\Omega = \{T\} \cup \{(H, x) : 0 \leq x < 2\pi\}$, in the obvious notation. Let $X : \Omega \to \mathbb{R}$ be given by

$$X(T) = -1, \quad X((H, x)) = x.$$

The random variable X takes values in $\{-1\} \cup [0, 2\pi)$. We say that X is continuous with the exception of a point mass (or atom) at -1.

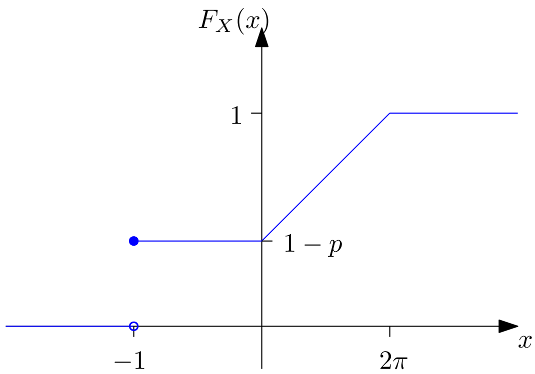

Figure 1: the distribution function of a random variable which is neither discrete nor continuous, see Example 2.2

2.2 Expectation, Variance and Moments

A concept that seems to be second nature to us is that of the 'average'. Moreover, when we know the probabilities associated with certain events happening, we talk about weighted averages. A more formal concept is that of expectation.

Definition 2.3 (Expectation of discrete random variable) The mean value, or expectation, or expected value of the random variable X with probability mass function $p_X(\cdot)$ is defined to be

$$\mathbb{E}[X] = \sum_x x p_X(x)$$

whenever this sum is absolutely convergent.

Expectation is a measure of location. It is the average of all the possible outcomes, weighted by the probability of getting that outcome.

If X is a random variable and $g : \mathbb{R} \longrightarrow \mathbb{R}$, then $Y = g(X)$, given formally by $Y(\omega) = g(X(\omega))$, is a random variable also. A natural question to ask is how could we calculate the expected value of the random variable $Y = g(X)$. To calculate its expectation we need first to find its probability mass function $p_Y(\cdot)$. This process can be complicated, and it is avoided by the following lemma (called by some the 'law of the unconscious statistician'!).

Lemma 2.3 (LOTUS) If X is a random variable with mass function $p_X(\cdot)$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$, then

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$$

whenever the sum is absolutely convergent.

Example 2.3 Suppose that X takes values $-2, -1, 1, 3$ with probabilities $\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}$ respectively. The random variable $Y = X^2$ takes values $1, 4, 9$ with probabilities $\frac{3}{8}, \frac{1}{4}, \frac{3}{8}$ respectively, and so

$$\mathbb{E}[Y] = \sum_x x\mathbb{P}(Y = x) = 1 \cdot \frac{3}{8} + 4 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = \frac{19}{4}.$$

Alternatively, use the law of the unconscious statistician to find that

$$\mathbb{E}[Y] = \mathbb{E}[X^2] = \sum_x x^2\mathbb{P}(X = x) = 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = \frac{19}{4}.$$

Definition 2.4 (Variance of discrete random variables) The variance of the random variable X with probability mass function $p_X(\cdot)$ is defined to be

$$\text{var}(X) = \sum_x (x - \mathbb{E}[X])^2 p_X(x).$$

Variance is a measure of dispersion. It is the weighted sum of the squared deviations from $\mathbb{E}[X]$ and takes the value zero when all of the xs are equal to $\mathbb{E}[X]$.

Definition 2.5 (Moments of discrete random variables) If k is a positive integer, the kth moment μ_k of X is defined to be

$$\mu_k = \mathbb{E}[X^k] = \sum_x x^k p_X(x),$$

and the kth central moment σ_k as

$$\sigma_k = \mathbb{E}[X - \mu_1]^k = \sum_x (x - \mu_1)^k p_X(x).$$

In general the first moment μ_1 is known as the mean and is denoted by μ and the second (central) moment σ_2 is known as the variance var(X) and is usually denoted by σ^2. The variance is, by construction, a non-negative number. The standard deviation is the positive root of the variance, i.e. $\sigma = \sqrt{\text{var}(X)}$.

Recall that the expectation of a discrete random variable X is $\mathbb{E}[X] = \sum_x x\mathbb{P}(X = x)$. It is the average of the possible values of X, each value being weighted by its probability. For continuous random variables instead of summation we have integration.

Definition 2.6 (Expectation of continuous random variables) The expectation of a continuous random variable X with density function f_X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

whenever this integral is absolutely convergent.

Example 2.4 (Example 2.1 revisited) The random variables X and Y have mean values

$$\mathbb{E}[X] = \int_0^{2\pi} \frac{x}{2\pi}dx = \pi, \quad \mathbb{E}[Y] = \int_0^{4\pi^2} \frac{\sqrt{y}}{4\pi}dy = \frac{4}{3}\pi^2.$$

Lemma 2.4 (LOTUS) Suppose that X is a random variable with density function $f_X(\cdot)$. Let $g : \mathbb{R} \mapsto \mathbb{R}$ and $Y = g(X)$. Then Y is a random variable, with p.d.f. $f_Y(\cdot)$, say, and

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_{-\infty}^{\infty} g(x) f_X(x)dx.$$

This lemma enables us to find $\mathbb{E}[Y]$ without calculating $f_Y(\cdot)$.

Example 2.5 (Example 2.1 revisited)

$$\mathbb{E}[Y] = \mathbb{E}[X^2] = \int_0^{2\pi} x^2 f_X(x)dx = \int_0^{2\pi} \frac{x^2}{2\pi}dx = \frac{4}{3}\pi^2.$$

The ideas of variance and moments for discrete random variables can be extended to the continuous case, where integrals are used instead of summations.

Definition 2.7 (Variance of continuous random variables) The variance of the random variable X with density function $f_X(\cdot)$ is defined to be

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x)dx.$$

Definition 2.8 (Moments of continuous random variable) If k is a positive integer, the kth moment μ_k of random variable X is defined to be

$$\mu_k = \mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)dx,$$

and the kth central moment σ_k as

$$\sigma_k = \mathbb{E}[(X - \mu_1)^k] = \int_{-\infty}^{\infty} (x - \mu_1)^k f_X(x)dx.$$

Note that the moments of random variables may not exist because the integral does not converge. Later on we will see examples of random variables for which none of the moments exist or for only moments up to a given power k' exist.

The following are some useful properties of expectation and variance. They hold for any random variable.

Lemma 2.5

1. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
2. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ for $a, b \in \mathbb{R}$;
3. var$(X) = 0$ if and only if $\mathbb{P}(X = c) = 1$ for $c \in \mathbb{R}$;
4. var$(aX + b) = a^2$var(X) for $a, b \in \mathbb{R}$;
5. var$(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

2.3 Examples of discrete random variables

2.3.1 Bernoulli distribution

A Bernoulli trial is an experiment with two possible outcomes: success and failure. Let $X = 1$ for success and $X = 0$ for failure. Suppose $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ with $0 < p < 1$, so that

$$\mathbb{P}(X = x) = p^x(1 - p)^{1-x}, \quad x \in \{0, 1\}.$$

The expectation and variance of X are

$$\mathbb{E}[X] = \sum_x x\mathbb{P}(X = x) = p,$$

$$\text{var}(X) = \sum_x (x - \mathbb{E}(X))^2\mathbb{P}(X = x) = p(1 - p).$$

2.3.2 Binomial distribution

Many experiments can be modelled as a sequence of Bernoulli trials, the simplest being the repeated tossing of a coin; $p =$ probability of a head,

$X = 1$ if the coin shows heads. Other examples include gambling games (for example, in roulette let $X = 1$ if red occurs, so $p =$ probability of red), election polls ($X = 1$ if candidate A gets a vote), and incidence of a disease ($p =$ probability that a "randomly selected" person gets infected).

If n identical Bernoulli trials are performed define the events

$$A_i = \{X = 1 \text{ on } i\text{th trial}\}, \qquad i = 1, 2, \ldots, n.$$

If we assume that the events $A_1, A_2, \ldots, A_n$ are a collection of independent events (as in the case of coin tossing) it is then easy to derive the distribution of the total number of successes in n trials.

Define a random variable Y by $Y =$ total number of successes in n trials. The event $Y = k$ will occur only if out of the events $A_1, A_2, \ldots, A_n$, exactly k of them occur, and necessarily $n - k$ of them do not occur. One particular outcome of the Bernoulli trials might be $A_1 \cap A_2 \cap A_3^c \cap \ldots \cap A_{n-1} \cap A_n^c$. The probability of this event is

$$\mathbb{P}\left(A_1 \cap A_2 \cap A_3^c \cap \ldots \cap A_{n-1} \cap A_n^c\right) = pp(1-p)\ldots p(1-p) = p^k(1-p)^{n-k},$$

where we have used the independence of the A_i's in this calculation. Notice that the calculation is not dependent on which set of k of the A_i occur. Putting all this together, we see that a particular sequence of n trials with exactly k successes has probability $p^k(1-p)^{n-k}$ of occurring. Since there are $\binom{n}{k}$ such sequences (the number of orderings of k ones and $n - k$ zeros), we have

$$\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

and Y is called a Binomial Bin(n, p) random variable.

The random variable Y can alternatively, and equivalently, be defined in the following way: in a sequence of n identical, independent Bernoulli trials, each with success probability p, define the random variables $X_1, X_2, \ldots, X_n$ by

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

The random variable

$$Y = \sum_{i=1}^{n} X_i$$

has the Bin(n, p) distribution.

The expectation of Y is

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&= np(p + (1-p))^{n-1} = np
\end{aligned}$$

noting that the $k = 0$ term makes no contribution.

It can be shown that

$$\mathbb{E}[Y^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = np(1-p) + n^2p^2,$$

thus the variance of Y is

$$\text{var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = np(1-p) + n^2p^2 - (np)^2 = np(1-p).$$

Noting the identity $Y = X_1 + \cdots + X_n$ and independence of $X_1, \cdots, X_n$, these results can be obtained directly by

$$\begin{aligned}
&\mathbb{E}[Y] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np, \\
&\text{var}(Y) = \text{var}(X_1) + \cdots + \text{var}(X_n) = np(1-p).
\end{aligned}$$

2.3.3 Poisson distribution

The random variable Y is said to have a Poisson distribution with parameter λ if

$$\mathbb{P}(Y = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

The parameter λ is sometimes called the intensity parameter.

The Poisson distribution is a widely applied discrete distribution, and can serve as a model for a number of different types of experiments. For example, if we are modelling a phenomenon in which we are waiting for a bus, waiting for customers to arrive in a bank, or the number of occurrences in a given time interval, the the Poisson distribution can sometimes be used to fulfill that purpose. One of the basic assumptions upon which the Poisson distribution is built, is that, for small time intervals, the probability of an arrival is proportional to the length of the interval. This makes it a reasonable model for situations like those indicated above.

Another area of application is for spatial distributions, where, for example, the Poisson distribution may be used to model the distribution of rainfall in an area or distribution of fish in a lake.

The expectation of Y is

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Since

$$\mathbb{E}[Y^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 + \lambda,$$

the variance of Y

$$\text{var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \lambda.$$

2.3.4 Geometric distribution

The binomial distribution counts the number of successes in a fixed number of Bernoulli trials. The geometric distribution instead counts the number of trials until the first success.

Suppose that independent trials, each having a probability of success p, are performed until a success occurs. Let the random variable X be the number of trials until the first success. Clearly X has probability mass function (pmf) given by

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

since the first $k-1$ trials must result in failure and the last in success. We say that X has the geometric distribution with parameter p. The expectation and variance of X are

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{var}(X) = \frac{q}{p^2},$$

where $q = 1 - p$.

It is easy to see that

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

Suppose we take any fixed integer $n > 0$. Then

$$\mathbb{P}(X > k + n \mid X > n) = \mathbb{P}(X > k), \quad k = 0, 1, 2, \ldots$$

since

$$\mathbb{P}(X > k+n \mid X > n) = \frac{(1-p)^{k+n}}{(1-p)^n} = (1-p)^k = \mathbb{P}(X > k), \quad k = 0, 1, 2, \ldots.$$

This is called the ‘lack of memory’ property. Can you see why? The geometric distribution is the only discrete memoryless random distribution.

2.4 Examples of continuous random variables

2.4.1 Uniform distribution

X is uniform on $(a, b]$, $a < b$, if

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{otherwise} \end{cases}$$

so that

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & b < x. \end{cases}$$

We can verify that $\int_{-\infty}^{\infty} f_X(x)dx = \int_a^b 1/(b-a)dx = 1$. The uniform distribution gets its name from the fact that its density is uniform, or constant, over the interval $(a, b]$. The expectation and variance of X are

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{b+a}{2},$$

and

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}[(X-\mathbb{E}[X])^2] \\
&= \int_{-\infty}^{\infty} \left(x - \frac{b+a}{2}\right)^2 f_X(x)dx \\
&= \int_a^b \frac{\left(x-\frac{b+a}{2}\right)^2}{b-a}dx = \frac{(b-a)^2}{12}.
\end{aligned}$$

2.4.2 Normal distribution

The normal distribution, also known as the Gaussian distribution, plays a central role in a large body of statistics. There are three main reasons for this. First, the normal distribution, and distributions associated with it, are very tractable. Second, the normal distribution has the familiar bell shape, whose symmetry makes it an appealing choice for many population models. Although there are many other distributions that are also bell-shaped, most do not possess the analytic tractability of the normal. Third, there is the Central Limit theorem which shows that (as we shall see later in the course), under mild conditions, the normal distribution can be used to approximate a large variety of distributions in large samples.

The normal distribution has two parameters, usually denoted by μ and σ^2, which happen to correspond to its mean and variance (as will be shown later). The pdf of a random variable X with the normal distribution with mean μ and variance σ^2, usually denoted by $X \sim N(\mu, \sigma^2)$, is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

for $x \in \mathbb{R}$.

If the normal random variable has mean 0 and variance 1, it is called a *standard* normal random variable. Traditionally we denote the probability density and cumulative distribution functions of the standard normal random variable by $\phi(\cdot)$ and $\Phi(\cdot)$. That is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad \Phi(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} \phi(s)\,ds.$$

Example 2.6 Show that if $X \sim N(\mu, \sigma^2)$ then the random variable

$Z = \frac{X-\mu}{\sigma}$ has a $N(0,1)$ distribution.

$$\begin{aligned}
\mathbb{P}(Z \leq z) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq z\right) \\
&= \mathbb{P}\left(X \leq z\sigma + \mu\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z\sigma+\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt,
\end{aligned}$$

by substituting $t = (x-\mu)/\sigma$.

It therefore follows that all normal probabilities can be calculated in terms of the standard normal.

Example 2.7 If $X \sim N(\mu, \sigma^2)$, then

$$\mathbb{P}(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

It is verified by the fact that

$$\begin{aligned}
\mathbb{P}\left(a < X \leq b\right) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathbb{P}\left(Z \leq \frac{b-\mu}{\sigma}) - \mathbb{P}(Z \leq \frac{a-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).
\end{aligned}$$

Furthermore, expected values can be simplified by carrying out the details in the $N(0,1)$ case, then transforming the result to the $N(\mu, \sigma^2)$ case. For example, if $Z \sim N(0,1)$, then

$$\begin{aligned}
\mathbb{E}[Z] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz \\
&= \lim_{b\to\infty, a\to-\infty} \frac{1}{\sqrt{2\pi}} \int_{a}^{b} z e^{-\frac{1}{2}z^2} dz \\
&= \lim_{b\to\infty, a\to-\infty} \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\right]_{z=a}^{z=b} = 0,
\end{aligned}$$

and

$$\begin{aligned}
\text{var}(Z) &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz - 0^2 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z.ze^{-\frac{1}{2}z^2} dz \\
&= -\frac{1}{\sqrt{2\pi}} \left(ze^{-\frac{1}{2}z^2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = 1.
\end{aligned}$$

Since $X \sim N(\mu, \sigma^2)$ and (by definition of Z) $X = \mu + \sigma Z$, then

$$\begin{aligned}
&\mathbb{E}(X) = \mathbb{E}(\mu + \sigma Z) = \mu, \\
&\text{var}(X) = \text{var}(\mu + \sigma Z) = \sigma^2.
\end{aligned}$$

Among the many uses of the Normal distribution, an important one is its use as an approximation to other distributions (which is partially justified by the Central Limit Theorem). For example, if $X \sim \text{Bin}(n, p)$, then $\mathbb{E}[X] = np$ and $\text{var}(X) = np(1-p)$, and under suitable conditions the distribution of X can be approximated by a normal random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. A conservative rule to follow is that the approximation will be good if $np \geq 5$ and $n(1-p) \geq 5$. As with most approximations, however, there are no absolute rules, and each application should be checked to decide whether the approximation is good enough for its intended use.

2.4.3 Gamma distribution

The gamma family of continuous random variables X is given by

$$f_X(x) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} & 0 < x < \infty \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$, $\beta > 0$ and

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \tag{1}$$

We write the distribution as $X \sim \text{Gamma}(\alpha, \beta)$.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_0^\infty x \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} dx,
\end{aligned}$$

now make the substitution $y = \beta x$

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \frac{y^\alpha e^{-y}}{\beta\Gamma(\alpha)} dy \\
&= \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} \\
&= \frac{\alpha}{\beta}\frac{\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= \frac{\alpha}{\beta}.
\end{aligned}$$

The variance is given by

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}[X^2] - E[X]^2 \\
&= \int_0^\infty x^2 \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} dx - \left(\frac{\alpha}{\beta}\right)^2 \\
&= \int_0^\infty \frac{y^{\alpha+1} e^{-y}}{\beta\Gamma(\alpha)} \frac{1}{\beta} dy - \left(\frac{\alpha}{\beta}\right)^2 \\
&= \frac{\Gamma(\alpha+2)}{\beta^2\Gamma(\alpha)} - \left(\frac{\alpha}{\beta}\right)^2 \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\beta^2\Gamma(\alpha)} - \left(\frac{\alpha}{\beta}\right)^2 \\
&= \frac{\alpha}{\beta^2}.
\end{aligned}$$

We have used the relations $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$. The parameter α is known as the shape parameter, and $1/\beta$ is called the scale parameter since most of its influence is on the spread of the distribution. The Gamma distribution appears in reliability and in survival studies where it is used to model survival times of subjects and waiting times of “customers” in a queueing system. We shall see that one special case of the gamma distribution is the exponential distribution when $\alpha = 1$ and the sum of independent identically distributed exponential random variables is gamma-distributed.

2.4.4 Chi squared distribution

There are two important special cases of the gamma distribution. If we set $\beta = 1/2$, $\alpha = d/2$, where d is a positive integer, then the gamma pdf becomes

$$f_X(x) = \begin{cases} \frac{1}{2^{d/2}} \frac{1}{\Gamma(d/2)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} & 0 < x < \infty \\ 0 & \text{otherwise,} \end{cases}$$

which is known as the **chi squared** pdf with d degrees of freedom, written $X \sim \chi^2_d$. In this case $\mathbb{E}[X] = d$ and Var$[X] = 2d$.

The chi squared distribution plays an important role in statistical inference, especially when sampling from a normal distribution, and also the chi squared test.

2.4.5 Exponential distribution

Another important special case of the gamma distribution is obtained when we set $\alpha = 1$ and $\beta = \lambda$. We then have

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

the exponential pdf with scale parameter $1/\lambda$. We write $X \sim \text{Exp}(\lambda)$. In this case $\mathbb{E}[X] = 1/\lambda$ and $\text{var}(X) = 1/\lambda^2$.

The exponential distribution can be used to model lifetimes, analogous to the case of the geometric distribution in the discrete case. In fact, the exponential distribution shares the 'memoryless' property of the geometric distribution. If $X \sim \text{Exp}(\lambda)$, then for $y > 0$

$$\mathbb{P}(X > x + y | X > x) = \mathbb{P}(X > y)$$

since

$$\begin{aligned} \mathbb{P}(X > x + y | X > x) &= \frac{\mathbb{P}(X > x + y \text{ and } X > x)}{\mathbb{P}(X > x)} \\ &= \frac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > x)} \\ &= \frac{\int_{x+y}^{\infty} \lambda e^{-\lambda u} du}{\int_{x}^{\infty} \lambda e^{-\lambda u} du} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} \\ &= e^{-\lambda y} \\ &= P(X > y). \end{aligned}$$

Note that $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}$.

2.4.6 Cauchy distribution

X has a Cauchy distribution if

$$f_X(x) = \frac{1}{\pi(1 + x^2)},$$

for $x \in \mathbb{R}$. Since

$$\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1 + x^2)} dx = \infty,$$

the mean of X does not exist. This distribution is notable for having no moments and for its frequent appearance in counter-examples.