BIOINFORMATICS & COMPUTATIONAL GENETICS MSc
PROBABILITY AND STATISTICS

Chapter 2. Random Variables and Probability Distributions

2.1 MOTIVATION: Generalization of Notation

The probability definitions, rules, and theorems given previously are all framed in terms of events in a sample space. For example, for an experiment with possible sample outcomes denoted by the *sample space* S, an *event* E was defined as any collection of sample outcomes, that is, any subset of the set S.

$$\begin{array}{ccccccc} \text{EXPERIMENT} & \longrightarrow & \text{SAMPLE OUTCOMES} & \longrightarrow & \text{EVENTS} & \longrightarrow & \text{PROBABILITY FUNCTION} \\ & & \{s_1, s_2, ...\} = S & \longrightarrow & E \subseteq S & \longrightarrow & \mathrm{P}(E).. \end{array}$$

In this framework, it is necessary to consider each experiment with its associated sample space separately - the nature of sample space S is typically different for different experiments.

EXAMPLE 1: Count the number of days in February which have zero precipitation.

SAMPLE SPACE $S = \{0, 1, 2, ..., 28\}$.Let E_i = "i days have zero precipitation"; $E_0, ..., E_{28}$ partition S.

EXAMPLE 2: Count the number of goals in a football match.

SAMPLE SPACE: $S = \{0, 1, 2, 3, ...\}$. Let E_i = "i goals in the match"; $E_0, E_1, E_2, ...$ partition S

In both of these examples, we need a formula to specify each $\mathrm{P}(E_i) = p_i$.

EXAMPLE 3 Measure the operating temperature of an experimental process.

SAMPLE SPACE: $S = \{\ x\ :\ x > T_{min}\ \}$.

Here it is difficult to express

$$\mathrm{P}[\text{ "Measurement is } x \text{ " }]$$

but possible to think about

$$\mathrm{P}[\text{ "Measurement is } \le x \text{ " }] = F(x), \text{ say,}$$

and again we seek a formula for $F(x)$.

A general notation useful for all such examples can be obtained by considering a sample space that is **equivalent** to S for a general experiment, but whose form is more familiar. For example, for a general sample space S, if it were possible to associate a subset of the **integer** or **real** number systems, $\mathbb{X}$ say, with S, then attention could be restricted to considering events in $\mathbb{X}$, whose structure is more convenient, as then S *are collections of sample outcomes* *events in* $\mathbb{X}$ *are intervals of the real numbers*

EXAMPLE : Consider an experiment involving counting the number of breakdowns of a production line in a given month. The experimental sample space S is therefore the collection of sample outcomes $s_0, s_1, s_2, ...$ where s_i is the outcome "there were i breakdowns"; events in S are collections of the s_is. Then a useful equivalent sample space is the set $\mathbb{X} = \{0, 1, 2, ...\}$, and events in $\mathbb{X}$ are collections of non-negative integers.

Formally, therefore, we seek a function or **map** from S to $\mathbb{X}$. This map is known as a **random variable**.

2.2 RANDOM VARIABLES

A **random variable** X is a function from experimental sample space S to some set of real numbers $\mathbb{X}$ that maps $s \in S$ to a unique $x \in \mathbb{X}$

$$\begin{aligned} X: \quad S &\longrightarrow \mathbb{X} \subseteq \mathbb{R} \\ s &\longmapsto x \end{aligned}$$

Interpretation A random variable is a way of describing the outcome of an experiment in terms of real numbers.

	RANDOM VARIABLE	TO BE SPECIFIED
EXAMPLE 1	$X =$"the number of days in Feb. with zero precipitation"	$\mathrm{P}[\ X = x\ ]$ for $x = 0, 1, 2, ..., 28$
EXAMPLE 2	$X =$"the number of goals in a football match"	$\mathrm{P}[\ X = x\ ]$ for $x = 0, 1, 2, 3, ...$
EXAMPLE 3	$X =$"the measured operating temperature"	$\mathrm{P}[\ X \leq x\ ]$ for $x > T_{min}$.

Therefore X is merely the count/number/measured value corresponding to the outcome of the experiment.

Depending on the type of experiment being carried out, there are two possible forms for the set of values that X can take:

A random variable is **DISCRETE** if the set $\mathbb{X}$ is of the form

$$\mathbb{X} = \{x_1, x_2, ..., x_n\} \text{ or } \mathbb{X} = \{x_1, x_2, ...\},$$

that is, a finite or infinite set of **distinct** values $x_1, x_2, ..., x_n, ...$ Discrete random variables are used to describe the outcomes of experiments that involve **counting** or **classification**.

A random variable is **CONTINUOUS** if the set $\mathbb{X}$ is of the form

$$\mathbb{X} = \bigcup_i \{x : a_i \leq x \leq b_i\}$$

for real numbers a_i, b_i, that is, the union of **intervals** in $\mathbb{R}$. Continuous random variables are used to describe the outcomes of experiments that involve **measurement**.

2.3 PROBABILITY DISTRIBUTIONS

A **probability distribution** is a function that assigns probabilities to the possible values of a random variable. When specifying a probability distribution for a random variable, two aspects need to be considered. First, the range of the random variable (that is, the values of the random variable which have positive probability) must be specified. Secondly, the method via which the probabilities are assigned to different values in the range must be specified; typically this is achieved by means of a function or formula.
In summary, we need to find a function or formula via which

$$\mathrm{P}[\ X = x\ ] \quad \text{or} \quad \mathrm{P}[\ X \leq x\ ]$$

can be calculated for each x in a suitable range $\mathbb{X}$. The functions used to specify these probabilities are just real-valued functions of a single real argument, similar to polynomial, exponential, logarithmic or trigonometric functions such as (for example)

$$\begin{aligned} f(x) &= 6x^3 - 3x^2 + 2x - 5 \\ f(x) &= e^x \\ f(x) &= \sin(x) + 2x\cos(2x) \end{aligned}$$

and so on. However, the fundamental rules of probability mean that the functions specifying $\mathrm{P}[\ X = x\ ]$ or $\mathrm{P}[\ X \leq x\ ]$ must exhibit certain properties. As we shall see below, the properties of these functions, and how they are manipulated mathematically, depend crucially on the nature os the random variable.

2.4 DISCRETE PROBABILITY DISTRIBUTIONS

For discrete random variables there are two routes via which the probability distribution can be specified.

2.4.1. PROBABILITY MASS FUNCTION

The probability distribution of a *discrete* random variable X is described by the **probability mass function (pmf)** f_X, specified by

$$f_X(x) = \mathrm{P}[X = x] \qquad \text{for } x \in \mathbb{X} = \{x_1, x_2, ..., x_n, ...\}$$

Because of the probability axioms, the function f_X must exhibit the following properties:

$$\text{(i) } f_X(x_i) \geq 0 \quad \text{for all } i \qquad\qquad \text{(ii)} \sum_i f_X(x_i) = 1.$$

2.4.2. DISCRETE CUMULATIVE DISTRIBUTION FUNCTION

The **cumulative distribution function** or **cdf**, F_X, is defined by

$$F_X(x) = \mathrm{P}[X \leq x] \qquad \text{for } x \in \mathbb{R}$$

The cdf F_X must exhibit the following properties:

(i) $\lim\limits_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim\limits_{x \longrightarrow \infty} F_X(x) = 1$

(iii) $\lim\limits_{h \to 0^+} F_X(x+h) = F_X(x)$ [i.e. F_X is continuous from the right]

(iv) $a < b \Rightarrow F_X(a) \leq F_X(b)$ [i.e. F_X is non-decreasing]

(v) $\mathrm{P}[a < X \leq b] = F_X(b) - F_X(a)$

The cumulative distribution function defined in this way is a "**step function**".

The functions f_X and/or F_X can be used to describe the **probability distribution** of random variable X.

EXAMPLE An electrical circuit comprises six fuses.

let $X =$"number of fuses that fail within one month". Then

$$\mathbb{X} = \{0, 1, 2, 3, 4, 5, 6\}$$

To specify the probability distribution of X, can use the mass function f_X or the cdf F_X. For example,

x	0	1	2	3	4	5	6
$f_X(x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$
$F_X(x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{7}{16}$	$\frac{11}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	$\frac{16}{16}$

as $F_X(0) = \mathrm{P}[\ X \leq 0\ ] = \mathrm{P}[\ X = 0\ ] = f_X(0)$, $F_X(1) = \mathrm{P}[\ X \leq 1\ ] = \mathrm{P}[\ X = 0\ ] + \mathrm{P}[\ X = 1\ ] = f_X(0) + f_X(1)$, and so on. Note also that, for example,

$$\mathrm{P}[\ X \leq 2.5\ ] \equiv \mathrm{P}[\ X \leq 2\ ]$$

as the random variable X only takes values $0, 1, 2, 3, 4, 5, 6$.

EXAMPLE A computer is prone to crashes.

Suppose that P["Computer crashes on any given day"] $= \theta$, for some $0 \leq \theta \leq 1$, independently of crashes on any other day.

Let $X =$"number of days until the first crash". Then

$$\mathbb{X} = \{1, 2, 3, ...\}$$

To specify the probability distribution of X, can use the mass function f_X or the cdf F_X. Now,

$$f_X(x) = \mathrm{P}[\ X = x\ ] = (1-\theta)^{x-1}\theta$$

for $x = 1, 2, 3, ...$ (if the first crash occurs on day x, then we must have a sequence of $x-1$ crash-free days, followed by a crash on day x). Also

$$F_X(x) = \mathrm{P}[\ X \leq x\ ] = \mathrm{P}[\ X = 1\ ] + \mathrm{P}[\ X = 2\ ] + ... + \mathrm{P}[\ X = x\ ] = 1 - (1-\theta)^x$$

as the terms in the summation are merely a geometric progression with first term θ and common term $1-\theta$.

2.4.5 RELATIONSHIP BETWEEN f_X AND F_X

The fundamental relationship between f_X and F_X is obtained by noting that if $x_1 \leq x_2 \leq ... \leq x_n \leq ...$, then

$$\mathrm{P}[X \leq x_i] = \mathrm{P}[X = x_1] + ... + \mathrm{P}[X = x_i],$$

so that

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)\ ,$$

and

$$f_X(x_1) = F_X(x_1) \qquad\qquad f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \quad \text{for } i \geq 2$$

so $\mathrm{P}[c_1 < X \leq c_2] = F_X(c_2) - F_X(c_1)$ for any real numbers $c_1 < c_2$.

Hence, in the discrete case, we can calculate F_X from f_X by **summation**, and calculate f_X from F_X by **differencing**.

2.5. SOME SPECIAL DISCRETE PROBABILITY DISTRIBUTIONS

Discrete probability models are used to model the outcomes of counting experiments. Depending on the experimental situation, it is often possible to justify the use of one of a class of "Special" discrete probability distributions. These are listed in this chapter, and are all motivated from the central concept of a *binary* or 0-1 trial, where the random variable concerned has range consisting of only two values with associated probabilities θ and $1-\theta$ respectively; typically we think of the possible outcomes as "successes" and "failures". All of the distributions in this section are derived by making different modelling assumptions about sequences of 0-1 trials.

Single 0-1 trial - count number of 1s	$\Longrightarrow$ BERNOULLI DISTRIBUTION
n independent 0-1 trials - count number of 1s	$\Longrightarrow$ BINOMIAL DISTRIBUTION
Sequence of independent 0-1 trials - count number of trials until first 1	$\Longrightarrow$ GEOMETRIC DISTRIBUTION
Sequence of independent 0-1 trials - count number of trials until nth 1	$\Longrightarrow$ NEGATIVE BINOMIAL DISTRIBUTION
Limiting case of binomial distribution	$\Longrightarrow$ POISSON DISTRIBUTION

2.5.0 THE DISCRETE UNIFORM DISTRIBUTION $X \sim Uniform(N)$

Range : $\mathbb{X} = \{1, 2, ..., N\}$
Parameter : $N \in \mathbf{Z}^+$
Mass function :

$$f_X(x) = \frac{1}{N} \qquad x \in \{1, 2, ..., N\}$$

Interpretation : This model gives **equal** probability to each of the N possible values of the random variable

2.5.1 THE BERNOULLI DISTRIBUTION $X \sim Bernoulli(\theta)$

Range : $\mathbb{X} = \{0, 1\}$
Parameter : $\theta \in [0, 1]$
Mass function :

$$f_X(x) = \theta^x (1-\theta)^{1-x} \qquad x \in \{0, 1\}$$

Interpretation : A model for "the number of successes in a single 0-1 trial" where the probability of success is θ.

2.5.2 THE BINOMIAL DISTRIBUTION $X \sim Binomial(n, \theta)$

Range : $\mathbb{X} = \{0, 1, ..., n\}$
Parameters : $n \in \mathbf{Z}^+$, $\theta \in [0, 1]$
Mass function :

$$f_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x} \qquad x \in \{0, 1, ..., n\}$$

Recall that $f_X(x) = \mathrm{P}[\ X = x\ ]$. Now if $X = x$, we must have a sequence of n 0-1 trials which contains x 1s and $n-x$ 0s. As the trials are independent, with probability θ of obtaining a 1 on any trial, all such sequences have probability $\theta^x (1-\theta)^{n-x}$ and there are ${}^nC_x$ (that is "n choose x") such sequences, where

$${}^nC_x = \binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{n \times (n-1) \times (n-2) \times ... \times (n-x+1)}{x \times (x-1) \times (x-2) \times ... \times 2 \times 1}$$

(that is we must choose the x positions in the sequence to place the ones). Hence the total probability of observing the event $X = x$ is

$$\binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Interpretation : A model for "the number of successes in n i.i.d. 0-1 trials" where the probability of success on each trial is θ. Note that the sum of n independent and identically distributed (i.i.d.) $Bernoulli(\theta)$ random variables has a Binomial distribution, that is,

$$\text{If } X_1, ..., X_n \sim Bernoulli(\theta) \ i.i.d., \text{ then } X = \sum_{i=1}^{n} X_i \sim Binomial(n, \theta)$$

EXAMPLE : In a multiple choice exam, each of the ten questions has five possible answers. For any single question, the probability of guessing the right answer is 1/5. Therefore the random variable X corresponding to the total number of correct answers assuming that each question is answered by guesswork has a binomial distribution, $X \sim Binomial(10, 1/5)$.

2.5.3 THE GEOMETRIC DISTRIBUTION $X \sim Geometric(\theta)$

Range : $\mathbb{X} = \{1, 2, ...\}$
Parameter : $\theta \in (0, 1]$
Mass function :

$$f_X(x) = (1-\theta)^{x-1}\theta. \qquad x \in \{1, 2, ...\}$$

Interpretation : A model for "the number of 0-1 trials until the first success is obtained " - where the probability of success on each trial is θ. a **discrete waiting-time** model.

Recall that $f_X(x) = \text{P}[\ X = x\ ]$. Now if $X = x$, we must have a sequence of x 0-1 trials which contains $x - 1$ 0s and then a 1. As the trials are independent, with probability θ of obtaining a 1 on any trial, such a sequence has probability $(1-\theta)^{x-1}\theta$ which immediately gives the form of the mass function.

EXAMPLE : A coin is tossed until the first head is obtained. The random variable X corresponding to the number of tosses required to obtain a the first head has a geometric distribution, $X \sim Geometric(1/2)$

2.5.4 THE NEGATIVE BINOMIAL DISTRIBUTION $X \sim NegBin(n, \theta)$

Range : $\mathbb{X} = \{n, n+1, n+2, ...\}$
Parameter : $n \in Z^+$, $\theta \in (0, 1]$
Mass function :

$$f_X(x) = \binom{x-1}{n-1}\theta^n(1-\theta)^{x-n} \qquad x \in \{n, n+1, n+2, ...\}.$$

Interpretation : A model for "the number of 0-1 trials until the nth success is obtained" where the probability of success on each trial is θ.

To compute the mass function, note that if the nth success occurs on the xth trial, then the sequence of 0-1 trials must comprise a sequence of $x - 1$ trials containing $n - 1$ 1s and $x - n$ 0s, followed by a final 1. Thus the mass function for a negative binomial distribution is given by

$$\binom{x-1}{n-1}\theta^{n-1}(1-\theta)^{x-n} \times \theta$$

that is a binomial probability multiplied by a single θ. as all trials are independent.

Note: The sum of n i.i.d. $Geometric(\theta)$ random variables has a Negative Binomial distribution, that is,

$$\text{If } X_1, ..., X_n \sim Geometric(\theta) \qquad \text{with } X_1, ..., X_n \text{ i.i.d, then} \qquad X = \sum_{i=1}^{n} X_i \sim NegBin(n, \theta)$$

that is, the number of trials until the nth 1 is the sum of the number of trials until the first 1, plus the number of trials between the first and second 1, etc. For this reason, the negative binomial distribution is also known as the **GENERALIZED GEOMETRIC** distribution.

EXAMPLE : A coin is tossed until the tenth head is obtained. The random variable X corresponding to the number of tosses required to obtain a the tenth head has a negative binomial distribution, $X \sim NegBin(10, 1/2)$

2.5.5 THE POISSON DISTRIBUTION $X \sim Poisson(\lambda)$

Range : $\mathbb{X} = \{0, 1, 2, ...\}$
Parameter : $\lambda \in \mathbb{R}^+$
Mass function :

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x \in \{0, 1, 2, ...\}$$

Note : The Poisson is the limiting case of the binomial distribution, that is, if we write $\theta = \lambda/n$, and then consider a limiting case as $n \longrightarrow \infty$. Specifically, if $X \sim Bin(n, \lambda/n)$, then

$$\begin{aligned} f_X(x) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{(n-x)!} \left(\frac{1}{n-\lambda}\right)^x \\ &\longrightarrow \frac{\lambda^x}{x!} e^{-\lambda} \end{aligned}$$

as $n \longrightarrow \infty$ with λ constant. That is, we have that

$$Binomial(n, \theta) \longrightarrow Poisson(\lambda).$$

in the limiting case.
Interpretation : The Poisson model is appropriate for count data, where the number of events (accidents, breakdowns etc) that occur in a given time period are being counted.

Alternative derivation : Consider a sequence of events occurring **independently** and **at random** in time at a **constant rate** λ per unit time. Let $X(t)$ be the random variable defined for $t > 0$ by

"$X(t) = x$" if and only if "x events occur in time interval $[0, t)$".

Then $X(t) \sim Poisson(\lambda t)$, so that

$$f_{X(t)}(x) = \mathrm{P}[X(t) = x] = \frac{(\lambda t)^x}{x!} e^{-(\lambda t)} \qquad x \in \{0, 1, 2, ...\}.$$

EXAMPLE : The reliability of a machine is to be monitored. The random variable X corresponding to the number of breakdowns that the machine suffers in one month can be modelled using a Poisson distribution.

2.6 CONTINUOUS PROBABILITY DISTRIBUTIONS

For discrete random variables there are two routes via which the probability distribution can be specified.

2.6.1. CONTINUOUS CUMULATIVE DISTRIBUTION FUNCTION

The probability distribution of a *continuous* random variable X is defined by the continuous **cumulative distribution function** or **c.d.f.**, F_X, specified by

$$F_X(x) = \mathrm{P}[X \leq x] \quad \text{for } x \in \mathbb{X}$$

that is, an identical definition to the discrete case.

The continuous cdf F_X must exhibit the same properties: as for the discrete cdf, except

(iii) $\lim_{h \to 0} F_X(x+h) = F_X(x)$ [i.e. F_X is continuous]

2.6.2. PROBABILITY DENSITY FUNCTION

The **probability density function**, or **pdf**, f_X, is defined by

$$f_X(x) = \frac{d}{dx}\{F_X(x)\}$$

so that, by a fundamental calculus result,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

The pdf f_X must exhibit the following properties:

$$\text{(i) } f_X(x) \geq 0 \text{ for } x \in \mathbb{X} \qquad \text{(ii)} \int_{\mathbb{X}} f_X(x)\, dx = 1.$$

In the continuous case, we calculate F_X from f_X by **integration**, and f_X from F_X by **differentiation**.
In both discrete and continuous cases, $F_X(x)$ id defined for all $x \in \mathbb{R}$, and $f_X(x)$ also defined for all x but may be zero for some values of x. Also, if X is continuous, we have

$$\mathrm{P}[\, a \leq X \leq b \,] = F_X(b) - F_X(a) \longrightarrow 0$$

as $b \longrightarrow a$. Hence, for each x, we must have

$$\mathrm{P}[\, X = x \,] = 0$$

if X is continuous. Therefore must use F_X to specify the probability distribution initially, although it is often easier to think of the "shape" of the distribution via the pdf f_X. Any function that satisfies the properties for a pdf can be used to construct a probability distribution. Note that, for a continuous random variable

$$f_X(x) \neq \mathrm{P}[X = x].$$

EXAMPLE Failure times.
A component is installed at time $x = 0$ and continues to operate until failure. Let $X =$"failure time of the component. then $\mathbb{X} = \{\, x \;:\; x > 0\} = \mathbb{R}^+$. Suppose that

$$\mathrm{P}[\, X > x \,] = \frac{1}{(1+x)^2} \qquad x > 0$$

Then

$$F_X(x) = \mathrm{P}[\, X \leq x \,] = 1 - \mathrm{P}[\, X > x \,] = 1 - \frac{1}{(1+x)^2}$$

for $x > 0$. By differentiation, we have the pdf

$$f_X(x) = \frac{d}{dx}\{F_X(x)\} = \frac{2}{(1+x)^3}$$

2.7 SOME SPECIAL CONTINUOUS DISTRIBUTIONS

Here is a list of probability models are used in standard modelling situations. Unlike the discrete case, there are not really any explicit links between most of them, although some connections can be made by means of "transformation" from one variable to another.

2.7.0 THE CONTINUOUS UNIFORM DISTRIBUTION $X \sim Uniform(a, b)$

Range : $\mathbb{X} = (a, b)$
Parameters : $a, b \in \mathbb{R}$
Density function :

$$f_X(x) = \frac{1}{b-a} \qquad a < x < b$$

Interpretation : A model with **constant** probability density on a region.

2.7.1 THE EXPONENTIAL DISTRIBUTION $X \sim Exponential(\lambda)$

Range : $\mathbb{X} = \mathbb{R}^+$
Parameter : $\lambda \in \mathbb{R}^+$
Density function :

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \in \mathbb{R}^+$$

Interpretation : A **continuous** waiting-time model (that is, a useful model for data that arise at the end of a measurement experiment).

The cdf for the exponential distribution can be calculated easily;

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt = \int_0^x \lambda e^{-\lambda t}\ dt = 1 - e^{-\lambda x} \qquad x \geq 0.$$

and note that

$$\mathrm{P}[\ X > x\ ] = 1 - \mathrm{P}[\ X \leq x\ ] = 1 - F_X(x) = e^{-\lambda x}$$

which may give some motivation for using the Exponential model in practice.

Note: Suppose that a mechanical component fails in such a way that the number of failures in a given month has a Poisson distribution with parameter λ, that is, the failure events occur repeatedly through time at random, but failures are relatively rare. Suppose that a component is installed at time $x = 0$. Then the random variable X corresponding to the *time until the first failure* has an exponential distribution, $X \sim Exponential(\lambda)$.

In fact, it can be shown that *all* the inter-failure event times have an exponential distribution, and that the inter-event times are probabilistically independent. One important property of the Exponential distribution linked to this result is the **lack of memory** property which is defined as follows; consider the **conditional** probability that random variable X is greater than $x_0 + x$, given that X is greater than x_0; such a conditional probability is important if we are trying to assess, say, the probability that, if a component functions without failure for a time x_0, it continues to function without failure until $x_0 + x$. From the conditional probability definition, we have

$$\mathrm{P}\left[X > x_0 + x | X > x_0\right] = \frac{\mathrm{P}\left[(X > x_0 + x) \cap (X > x_0)\right]}{\mathrm{P}\left[X > x_0\right]} = \frac{\mathrm{P}\left[X > x_0 + x\right]}{\mathrm{P}\left[X > x_0\right]} = \frac{e^{-\lambda(x_0+x)}}{e^{-\lambda x_0}} = e^{-\lambda x} = \mathrm{P}\left[X > x\right]$$

so that the component has "forgotten" the time period up to x_0.

2.7.2 <u>THE GAMMA DISTRIBUTION</u> $X \sim Gamma(\alpha, \beta)$

Range : $\mathbb{X} = \mathbb{R}^+$
Parameters : $\alpha, \beta \in \mathbb{R}^+$
Density function :

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x \in \mathbb{R}^+$$

where

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt \qquad \alpha > 0.$$

is known as the **Gamma Function**. This integral cannot by calculated analytically, but can be computed using numerical integration for any $\alpha > 0$.

Interpretation : The Gamma distribution **continuous** waiting-time model. It can be shown the sum of i.i.d. Exponential random variables has a Gamma distribution, that is, if $X_1, X_2, ..., X_n$ are independent and identically distributed $Exponential(\lambda)$ random variables, then

$$X = \sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$$

Notes :

(1) If $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

(2) If $\alpha = 1, 2, ...$, $\Gamma(\alpha) = (\alpha - 1)!$.

(3) $\Gamma(1/2) = \sqrt{\pi}$.

(4) If $\alpha = 1, 2, ...$, then the $Gamma(\alpha/2, 1/2)$ distribution is known as the **Chi-squared distribution** with α **degrees of freedom**, denoted χ^2_α.

EXAMPLE: Suppose that a mechanical component fails in such a way that the number of failures in a given month has a Poisson distribution with parameter λ. Suppose that a component is installed at time $x = 0$. The random variable X corresponding to the time until the nth $(n = 1, 2, ...)$ failure has a gamma distribution, $X \sim Gamma(n, \lambda)$.

2.7.3 <u>THE BETA DISTRIBUTION</u> $X \sim Beta(\alpha, \beta)$

Range : $\mathbb{X} = (0, 1)$
Parameters : $\alpha, \beta \in \mathbb{R}^+$
Density function :

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} \qquad x \in (0, 1).$$

Interpretation : A model for experiments whose outcomes lie in a bounded interval.

EXAMPLE: The percentage content of a particular element in a material is to be measured. The random variable corresponding to the observed content can be modelled using a beta distribution.

2.7.4 <u>THE NORMAL DISTRIBUTION</u> $X \sim N(\mu, \sigma^2)$

Range : $\mathbb{X} = \mathbb{R}$
Parameters : $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$
Density function :

$$f_X(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad x \in \mathrm{R}.$$

Interpretation : A probability model that reflects observed (**empirical**) behaviour of data samples; this distribution is often observed in practice.

The pdf is symmetric about μ, and hence μ is controls the *location* of the distribution and σ^2 controls the *spread* or *scale* of the distribution.

Notes :
(1) The Normal density function is justified by the **Central Limit Theorem**.
(2) Special case: $\mu = 0, \sigma^2 = 1$ - the **standard** or **unit** normal distribution. In this case, the density function is denoted $\phi(x)$, and the cdf is denoted $\Phi(x)$ so that

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\ dt = \int_{-\infty}^{x} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}t^2\right\}\ dt.$$

This integral can only be calculated numerically.
(3) If $X \sim N(0,1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$.
(4) If $X \sim N(0,1)$, and $Y = X^2$, then $Y \sim Gamma(1/2, 1/2) = \chi^2_1$.
(5) If $X \sim N(0,1)$ and $Y \sim \chi^2_\alpha$ are independent random variables, then random variable T, defined by

$$T = \frac{X}{\sqrt{Y/\alpha}}$$

has a **Student-t distribution** with α **degrees of freedom**.
The Student-t distribution plays an important role in certain statistical testing procedures.

THE STANDARD NORMAL CDF