

IMPERIAL COLLEGE LONDON

MATHEMATICS: YEAR 2

Random variables and Distributions

Xin Wang

November 17, 2020

Abstract

A probability distribution is the mathematical function that gives the probabilities of occurrence **of all the different possible outcomes for an experiment**. This is similar to a graph giving a visual picture of the function i.e. it is a visual mathematical description of a random phenomenon in terms of its sample space and the probabilities of events.

Contents

1	Notations	2
2	Random Variables	2
3	Probability distribution functions	3
3.1	Cumulative distribution function	3
3.2	Theoretical Mean and Variance	7
3.3	Probability mass function	10
3.4	Probability density function	11
3.4.1	Properties of continuous probability distribution	11
4	Discrete Probability Distributions	13
4.1	Discrete uniform distribution: $X \sim \text{Unif}(a, b)$	13
4.2	Bernoulli distribution: $X \sim \text{Be}(p)$	14
4.3	Binomial distribution: $X \sim \text{Bin}(n, p)$	15
4.4	Geometric distribution: $X \sim \text{Geom}(p)$	16
4.5	Poisson distribution: $X \sim \text{Po}(\lambda)$	17
5	Continuous Probability Distributions	20
5.1	Continuous uniform distribution: $X \sim \text{Unif}(a, b)$	21
5.2	Exponential distribution: $X \sim \text{Exp}(\lambda)$	21
5.2.1	Memoryless Property	23
5.3	Normal distribution: $X \sim N(\mu, \sigma^2)$	24
5.3.1	Normal distribution table	26
5.3.2	Standardising	27
5.4	Chi-squared distribution: $X \sim \chi^2(k)$	28
5.5	Log-normal distribution: $X \sim \text{lognormal}(\mu, \sigma^2)$	29
6	Change of Variable Technique	29
6.1	Generalisation for an increasing function	29
6.2	Generalisation for a decreasing function	30

1 Notations

The probability of the random variable X taking on a specific value x is denoted as

$$X = x$$

- **Uppercase letters:** Random variables
- **Lowercase letters:** Particular values a random variable can take on

The probability function:

$$P(X = x)$$

interpreted as "the probability that the random variable X takes the value x ".

2 Random Variables

Random variables are a variable in which its values depends on outcome of a random phenomenon.

Random variable X (on a sample space S): A measurable function that assigns a real numbered value to each outcome of S .

Range (of a random variable): The collection of values a random variable can take on.

The **range** of the random variables determines whether it is a:

- **Discrete random variable:** A random variable that can take on values in a specified countable list e.g. coin toss or dice roll.

Has a **probability mass function** that is characteristic of the random variable's probability distribution.

- **Continuous random variable:** A random variable that can take on **any** numerical value **in an interval**.

Has a **probability density function** that is characteristic of the random variable's probability distribution.

3 Probability distribution functions

Often a sample space can be very large and general. By using probability distributions of a specific random variable, focus can be directed to a specific range in order to simplify the description and analysis of a random experiment. Probability theory becomes more useful in practice.

Probability distribution: A list of all of the possible outcomes of a random variable along with their corresponding probability values.

Outcome of die roll	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

Figure 1: An example of a discrete probability distribution

In many scenarios, the number of outcomes can be much larger or infinite and a table would be impractical to write. To get around the problem of writing a table for every distribution, a function is defined. **A function gives the ability to define a probability distribution succinctly.**

As mentioned previously, a function is fundamentally is a box that **takes an input** and **returns an output**. The main benefit of functions is the knowledge of how to transform any input into the output, this knowledge can be used to visualise the function explicitly as a graph.

One of the most important features of functions are **parameters**. Parameters are the most important feature of a probability distribution function because, fundamentally, they define the final output of the function i.e. the likelihood of certain outcomes in a random process. It's often parameters that statisticians try to estimate in data science problems.

Probability distribution function: A function that defines a particular probability distribution e.g. CDF, PDF and PMF.

3.1 Cumulative distribution function

The cumulative distribution function (CDF) of a random variable is a method to describe the distribution of random variables. The advantage of the CDF is that it can be defined for any kind of random variable e.g. discrete, continuous and mixed.

Cumulative Distribution Function (CDF) $F(x)$: The probability that a random variable X taking a value **less than** or **equal to** x .

$$F_X(x) = P(X \leq x) \text{ for all } x \in \mathfrak{R}$$

Example 1: The probability that the random variable has value less than 3 $\{X \leq 3\}$ is an union of probabilities $\{X = 0\}$, $\{X = 1\}$, $\{X = 2\}$ and $\{X = 3\}$.

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.6561 + 0.2916 + 0.0486 + 0.0036 \\ &= 0.9999 \end{aligned}$$

And $P(X = 3)$ can be found with:

$$\begin{aligned} P(X = 3) &= P(X \leq 3) - P(X \leq 2) \\ &= 0.0036 \end{aligned}$$

Note: the CDF completely describes the distribution of a discrete random variable. The CDF can also be used to derive the PMF.

Example 2: Determine the PMF of x from the following CDF:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \leq x < 0 \\ 0.7 & 0 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

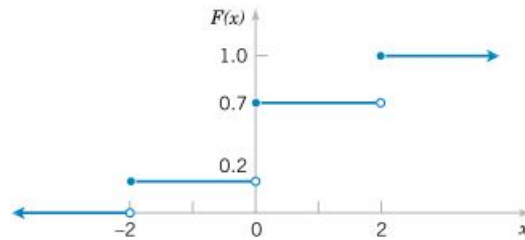


Figure 2: The Cumulative Distribution Function

The PMF at each point is the change in the CMF at that point:

$$\begin{aligned} f(-2) &= 0.2 - 0 = 0.2 \\ f(0) &= 0.7 - 0.2 = 0.5 \\ f(2) &= 1.0 - 0.7 = 0.3 \end{aligned}$$

Example 1: Suppose $X_1 \sim \text{Unif}(0, 1)$ and $X_2 \sim \text{Unif}(0, 1)$, and X_1 and X_2 are independent.

Compute the probability that $P(X_1 \leq \frac{1}{2} \cap X_2 \leq \frac{1}{2})$.

1. Use the CDF:

$$P\left(X_i \leq \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) = \frac{1}{2}$$

2. Note that X_1 and X_2 are independent:

$$\begin{aligned} P\left(X_1 \leq \frac{1}{2} \cap X_2 \leq \frac{1}{2}\right) &= F_X\left(\frac{1}{2}\right) F_X\left(\frac{1}{2}\right) \\ &= \frac{1}{4} \end{aligned}$$

Example 3: Time until chemical reaction is approximated by the continuous CDF:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & 0 \leq x \end{cases}$$

Find the PDF of X and what is the proportion of reactions is completed **within** 200 milliseconds?

1. Find the PDF:

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & 0 \leq x \end{cases}$$

2. Probability that reactions complete within 200 milliseconds:

$$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647$$

When trying evaluate the probabilities of the form $P(a < X \leq b)$:

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b f_X(x) dx \end{aligned}$$

If b is close enough to a :

$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a) \\ &= \int_a^b f_X(x) dx \\ &\approx f_X(a)(b - a) \end{aligned}$$

3.2 Theoretical Mean and Variance

As mentioned previously, parameters are very important in functions and two of the most important parameters when it comes to probability distributions are: **mean** and **variance**.

Expectation operator E : The computation that finds the sum of function X that is weighted by the corresponding probability.

These parameters are often used to summarise a probability distribution of a random variable X . Note that these two numbers **do not uniquely define a distribution**, some distributions could have the same mean and variance.

- **Mean:** A measure of the center of the probability distribution.
- **Variance:** A measure of the dispersion i.e. variability in the distribution.

Mean or Theoretical mean $E(X)$ or μ : Weighted average of the possible values of X where the weights are defined as the probabilities.

$$\mu = E(X) = \sum_x x f(x)$$

or

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

If $f(x)$ is thought of as a probability mass function of a loading on a long, thin beam, $E(X)$ is the point at which the beam balances. This reiterates the definition that $E(X)$ describes the center of the distribution of X .

Properties:

- $E(aX + b) = aE(X) + b$ for any $a, b \in \mathbb{R}$
- $E[g(X)] = \sum_x g(x) f_X(x)$ for any function $g(x)$

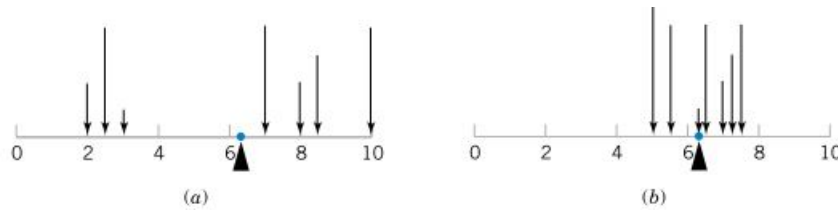


Figure 3: PMF with triangle representing the graphical representation of Mean

Variance or **Theoretical variance** σ^2 or $V(X)$: A measure of the dispersion or the scatter of the distribution.

$$E(X - \mu)^2 = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

or

$$E(X^2) - E(X)^2$$

where $V(X) \geq 0$

Standard deviation σ : A measure of how much the members of a group differ **relative from the mean value** μ .

$$\sigma = \sqrt{\sigma^2}$$

Some distributions may have the same mean but different variances.

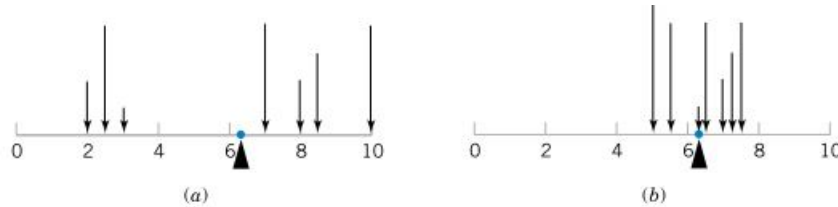


Figure 4: PMF (a) and (b) showing different variance but same mean

Properties:

- $V(aX + b) = a^2 V(X)$ for any $a, b \in \mathfrak{R}$

Example 1: Given the discrete random variable Y with range $[1, 2, \dots, 5]$ with the following probability distribution, find the variance and standard deviation.

y	1	2	3	4	5
$f_Y(y)$	0.1	0.3	0.4	0.1	0.1

1. Find the Mean μ :

$$\begin{aligned} \mu &= E(Y) \\ &= 1(0.1) + 2(0.3) + 3(0.4) + 4(0.1) + 5(0.1) \\ &= 2.8 \end{aligned}$$

2. Find the variance:

$$\begin{aligned} V(Y) &= \sum_y (y - \mu)^2 f_Y(y) \\ &= (1 - 2.8)^2(0.1) + (2 - 2.8)^2(0.3) + (3 - 2.8)^2(0.4) + (4 - 2.8)^2(0.1) + (5 - 2.8)^2(0.1) \\ &= 1.16 \end{aligned}$$

3. Standard deviation:

$$\begin{aligned}\sigma &= \sqrt{1.16} \\ &\approx 1.8\end{aligned}$$

Example 2: Given the following:

$$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 2 \\ 0 & \text{Otherwise} \end{cases}$$

1. Find the value of c for which f_X is a **valid density**.

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

Thus:

$$\begin{aligned}\int_0^2 cx^2 dx &= \left[\frac{cx^3}{3} \right]_0^2 \\ &= \left(\frac{8}{3} \right) c\end{aligned}$$

Simplifying to:

$$\begin{aligned}\frac{8}{3}c &= 1 \\ c &= \frac{3}{8}\end{aligned}$$

2. Obtain the CDF.

$$\begin{aligned}F_X(u) &= \int_0^u \frac{3}{8}x^2 dx \\ &= \frac{3}{8} \int x^2 dx \\ &= \frac{3}{8} \left[\frac{x^3}{3} \right]_0^u \\ &= \frac{u^3}{8}\end{aligned}$$

CDF is:

$$F_X(u) = \begin{cases} 0 & u < 0 \\ \frac{u^3}{8} & 0 \leq u \leq 2 \\ 1 & u > 2 \end{cases}$$

3. Compute $P(1 < X \leq 2)$.

$$\begin{aligned}
 P(1 < X \leq 2) &= F_X(2) - F_X(1) \\
 &= \frac{2^3}{8} - \frac{1^3}{8} \\
 &= \frac{7}{8}
 \end{aligned}$$

3.3 Probability mass function

When a probability function is used to describe a discrete probability distribution, it is called a **probability mass function (PMF)**.

A PMF f returns the probability of an outcome x . Therefore, a probability mass function is written as:

$$f(x) = P(X = x)$$

probability mass function f just returns the probability of the outcome x

Since the PMF returns probabilities, it must obey the axioms of probability that described previously:

- The PMF outputs values between 0 and 1 inclusive.

$$0 \leq f(x) \leq 1$$

- The sum of the PMF over all outcomes is equal to 1.

$$\sum_i f(x_i) = f(x_1) + f(x_2) + \dots = 1$$

A discrete probability distribution can be described as a **table**, as a **function** and, also, graphically.

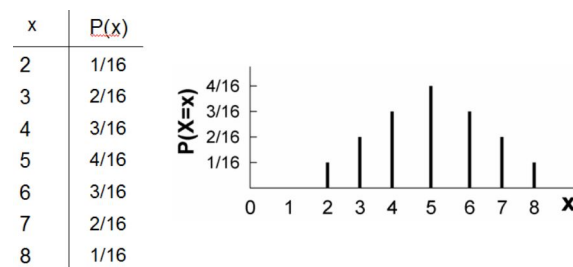


Figure 5: Graphically representing a discrete probability distribution

Some probability distributions crop up so often that they have been extensively studied and have names. This will be covered in later chapters.

- **Bernoulli distribution** $\text{Ber}(p)$: Models an experiment with only two possible outcomes.

- **Binomial distribution** $\text{Bin}(n, p)$: Models the number of successes when someone draws n times with replacement.
- **Geometric distribution** $\text{Geo}(p)$: Models the number of trials needed to get one success.
- **Poisson distribution** $\text{Pois}(\lambda)$: Models the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.

3.4 Probability density function

Some probabilities of random variables have continuous outcomes e.g. the height of an adult picked at random from a population or the amount of time that a taxi driver has to wait before their next job. For these examples, the random variable is better described by a continuous probability distribution.

When a probability function is used to describe a continuous probability distribution, it is called a **probability density function (PDF)**. Probability density functions are slightly more complicated conceptually.

$$P(X \leq x_0) = F_X(x_0) = \int_{-\infty}^{x_0} f_X(x) dx$$

where F_X is monotonically non-decreasing i.e. $F_X(a) \leq F_X(b)$ if $a < b$, and $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$$

where if b is close enough to a then

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx \approx f_X(a)(b - a)$$

3.4.1 Properties of continuous probability distribution

1. Numbers on the vertical axis start at zero and go up. Any output value from a probability density function is **greater than or equal to zero** i.e. the output is non-negative:

$$f(x) \geq 0$$

2. Unlike probability mass functions, the output of a probability density function is **not a probability value**. The probability between two outcomes, a and b is the integral of the probability density function between those two points which is equivalent to finding the area under the curve produced by the probability density function between the points a and b :

$$\int_a^b f(x; \mu, \sigma) dx = P(a < X < b)$$

the integral of the probability density function between 'b' and 'a' (left-hand side) is equal to the probability that the outcome of the random variable is between a and b (right-hand side)

- Remember that rules of probability distributions still apply, namely that the sum of all possible outcomes is equal to 1:

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

- The probability of the random variable being equal **to a specific outcome** is 0. For example, the probability that the outcome is equal to the number 2 is defined as:

$$\int_2^2 f(x) dx = P(2 < X < 2) = 0$$

Problems only deal with:

- Probabilities occurring between two values.
- Probability of an outcome being greater than or less than a specific value.

Example 1: Given

$$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 2 \\ 0 & \text{Otherwise} \end{cases}$$

- Find value of c for which f_X is a valid density:

Require:

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

Thus:

$$\int_0^2 cx^2 dx = \left[\frac{cx^3}{3} \right]_0^2 = \left(\frac{8}{3} \right) c = 1$$

$$c = \frac{3}{8}$$

- Obtain CDF:

For $0 \leq x \leq 2$, the CDF is:

$$F_X(u) = \int_0^u \frac{3}{8} x^2 dx = \frac{3}{8} \int x^2 dx = \frac{3}{8} \left[\frac{x^3}{3} \right]_0^u = \frac{u^3}{8}$$

The CDF is:

$$F_X(u) = \begin{cases} 0 & u < 0 \\ \frac{u^3}{8} & 0 \leq u \leq 2 \\ 1 & u > 2 \end{cases}$$

- Compute $P(1 < X \leq 2)$:

$$P(1 < X \leq 2) = F_X(2) - F_X(1) = \frac{2^3}{8} - \frac{1^3}{8} = \frac{7}{8}$$

4 Discrete Probability Distributions

4.1 Discrete uniform distribution: $X \sim \text{Unif}(a, b)$

The simplest discrete random variable where one assumes **only a finite number of possible values**, each with equal probability. A simple example of the discrete uniform distribution is throwing a fair die. The possible values are 1, 2, 3, 4, 5, 6 and each time the die is thrown the probability of a given score is $\frac{1}{6}$.

Discrete uniform distribution: The distribution of a random variable where each of its n values within the range have equal probability:

$$f(x_i) = \frac{1}{n}$$

Common properties:

- Probability mass unction:

$$f_X(x) = \frac{1}{k} \text{ for } x = 1, 2, \dots, k$$

Equally distributed with even spacing.

- Cumulative distribution function:

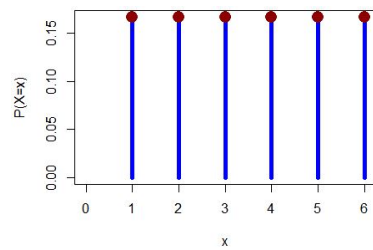
$$F_X(x) = \sum_{i=1}^x \frac{1}{k} = \frac{x}{k} \text{ for } x = 1, 2, \dots, k$$

- Mean:

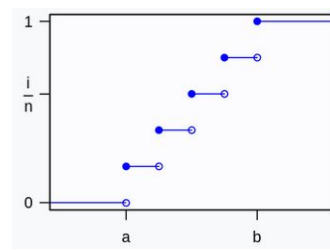
$$E(X) = \frac{k+1}{2} \text{ where } k \text{ is the number of elements in the distribution}$$

- Variance:

$$V(X) = E(X^2) - E(X)^2 = \frac{k^2 - 1}{12}$$



(a) Probability mass function



(b) Cumulative distributive function

4.2 Bernoulli distribution: $X \sim \text{Be}(p)$

The Bernoulli distribution is the **discrete probability distribution** of a random variable which takes a binary, boolean output:

- 1 with probability p
- 0 with probability $1 - p$

Bernoulli trials: Independent trials with a constant probability p of a success on each trial.

Key characteristics of a bernoulli trail:

- Only two possible outcomes e.g. H or T
- n identical trials
- Trials are **independent** - one event does not affect the other event
- Constant probability of success p

The probability function associated with a Bernoulli variable is the following:

$$f(x) = \begin{cases} p^x * (1-p)^{1-x} & \text{if } x = 0,1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

- Expected Value:

$$E(X) = p$$

- Variance:

$$V(X) = p(1-p)$$

The idea behind the Bernoulli distribution is that the experiment is done **only once**. But what happens if there is more than one trial, under the assumption that trials are independent among each other?

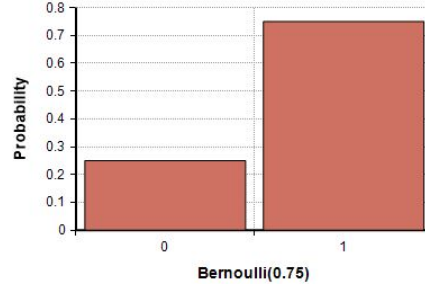


Figure 7: Example of Bernoulli distribution

4.3 Binomial distribution: $X \sim \text{Bin}(n, p)$

This distribution describes the behavior the outputs of n random experiments, each having a Bernoulli distribution with probability p .

Combinatorial notation defined as:

$$C(n, x) = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where:

- n : Number of trials
- x : Number of successes

Running n independent experiments with each having a Bernoulli distribution with parameter p , the probability of having x successes is defined:

$$f(x) = \begin{cases} \binom{n}{x} p^x \times (1-p)^{n-x} & \text{if } x = 0, 1, \dots, n \\ 0 & \text{Otherwise} \end{cases}$$

- Expected Value:

$$E(X) = np$$

- Variance:

$$V(X) = np(1-p)$$

Example 1: A fair coin is tossed 6 times with sample space $S = \{H, T\}$. Note this is a binomial experiment with $n = 6$ and $p = q = \frac{1}{2}$.

1. Probability of exactly two H: $x = 2$ and $n - x = 4$:

$$\begin{aligned} f_X(2; n=6, p=\frac{1}{2}) &= \binom{6}{2} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 \\ &= \frac{6!}{2!(6-2)!} \left(\frac{1}{4}\right) \left(\frac{1}{16}\right) \\ &\approx 0.23 \end{aligned}$$

2. E denotes event where **at least** 4 H is achieved i.e. 4, 5 and 6 H's, find the probability.

$$\begin{aligned}
 P(E) &= \sum_{x \in \{4,5,6\}} f_X \left(x; n=6, p=\frac{1}{2} \right) \\
 &= \binom{6}{4} \left(\frac{1}{2} \right)^4 \left(\frac{1}{2} \right)^2 + \binom{6}{5} \left(\frac{1}{2} \right)^5 \left(\frac{1}{2} \right) + \binom{6}{6} \left(\frac{1}{2} \right)^6 \left(\frac{1}{2} \right)^0 \\
 &= \frac{15}{64} + \frac{6}{64} + \frac{1}{64} \\
 &= \frac{11}{32} \\
 &\approx 0.34
 \end{aligned}$$

3. Find the probability of getting no heads i.e. all failures:

$$\begin{aligned}
 (1-p)^6 &= \left(\frac{1}{2} \right)^6 \\
 &= \frac{1}{64}
 \end{aligned}$$

4.4 Geometric distribution: $X \sim \text{Geom}(p)$

The idea of Geometric distribution is modeling the probability of having a certain number of Bernoulli trials (each with parameter p) before getting the first success. The random variable X denotes the number of trials until first success.

Computing the probability of having first $k-1$ failures, each with probability $1-p$, and then, at the k th trial, having a success (with probability p). Since each Bernoulli trial is independent from the others:

$$f(x) = (1-p)^{x-1} \times p$$

- Mean:

$$E(X) = \frac{1}{p}$$

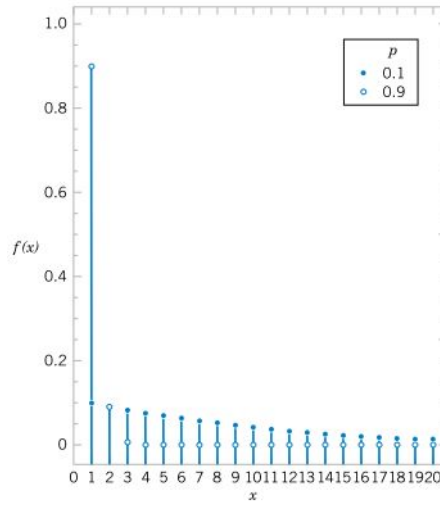
- Variance:

$$V(X) = \frac{(1-p)}{p^2}$$

Example 1: Let random variable X denote the number of bits transmitted up till and including the first error. The probability that a bit transmitted through a channel has an error is 0.1 (Probability of a successful transmission is 0.9).

Find the probability that the 5th bit is an error:

$$\begin{aligned}
 P(X=5) &= P(OOOOE) \\
 &= 0.9^4 \times 0.1 \\
 &= 0.066
 \end{aligned}$$



Note how the probability of an error bit (0.1) and successful bit (0.9) gradually converge as the value of x increases.

4.5 Poisson distribution: $X \sim \text{Po}(\lambda)$

A **Poisson Process** is a model for a series of discrete event where the **average time between events is known, but the exact timing of events is random**. Many physical problems are concerned with events occurring independently of any medium such as time and space e.g. number of aircraft accidents in a set time or counts measured by a Geiger-counter in a certain time interval.

A Poisson Process meets the following criteria:

1. Events are independent of each other. The occurrence of one event does not affect the probability another event will occur (waiting time between events is memoryless).
2. The average rate λ (events per time period) is constant.
3. Two events cannot occur at the same time.

The Poisson distribution gives the probability of observing k events in a time period given the length of the period and the average events per time:

$$P(k \text{ events in time period}) = e^{-\left(\frac{\text{events}}{\text{time}} \times \text{time period}\right)} \times \frac{\left(\frac{\text{events}}{\text{time}} \times \text{time period}\right)^k}{k!}$$

$\frac{\text{events}}{\text{time}} \times \text{time period}$ is usually simplified into **the rate parameter**: lambda λ .

$$P(k \text{ events in time period}) = e^{-\lambda} \times \frac{\lambda^k}{k!}$$

- Mean:

$$E(X) = \lambda$$

- Variance:

$$V(X) = \frac{(1 - p)}{p^2}$$

Lambda can be thought of as the expected number of events in the interval. As we change the rate parameter λ the probability of seeing different numbers of events in one interval changes.

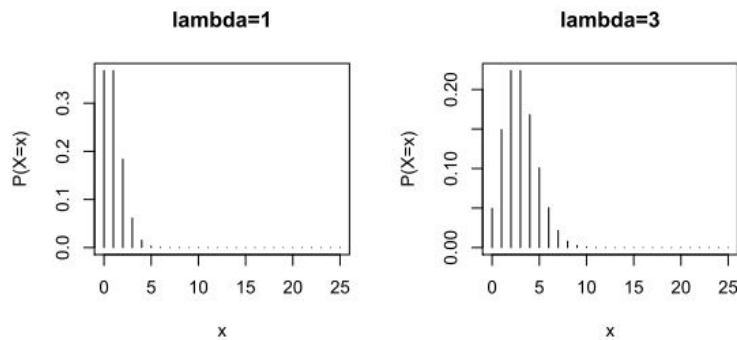


Figure 8: PMF for poisson distributions for selected value of λ

When calculating probabilities, means and variances, it is important to **use consistent units** i.e. if average number of flaws **per millimeter** is 3.4 then average number of flaws in **10 millimeter** is 34.

Example 1: Suppose accidents occur following a Poisson distribution at a **rate of 2 per year**.

1. What is the probability of **more than one accident** in a six-month period?

- (a) Let X denote the number of accidents in a six month period:

$$X \sim \text{Poisson} \left(2 \cdot \frac{1}{2} \right)$$

- (b) Establish the cases:

$$f_X x; \lambda = 1 = \begin{cases} \frac{e^{-1} 1^x}{x!} & x = 0, 1, 2, \dots \\ 0 & \text{Otherwise} \end{cases}$$

- (c) Find the probability that **only one** accident occur and subtract from

1:

$$\begin{aligned}P(X > 1; \lambda = 1) &= 1 - P(X \leq 1; \lambda = 1) \\&= 1 - \{f_X(0; \lambda = 1) + f_X(1; \lambda = 1)\} \\&= 1 - \left\{ \frac{e^{-1}}{0!} + \frac{e^{-1}}{1!} \right\} \\&= 1 - 2e^{-1} \\&\approx 0.264\end{aligned}$$

2. What is the probability of **at least two accidents in a one year period**?

(a) Let Y denote the number of accidents in a year:

$$Y \sim \text{Poisson}(2)$$

(b) Establish the cases:

$$f_Y y = \begin{cases} \frac{e^{-2} 2^y}{y!} & y = 0, 1, 2, \dots \\ 0 & \text{Otherwise} \end{cases}$$

(c) Find the probability that **only one** accident occur and subtract from 1:

$$\begin{aligned}P(Y > 1; \lambda = 2) &= 1 - P(Y \leq 1; \lambda = 2) \\&= 1 - \{f_Y(0; \lambda = 2) + f_Y(1; \lambda = 2)\} \\&= 1 - \left\{ \frac{e^{-2}}{0!} + \frac{e^{-2} \times 2}{1!} \right\} \\&= 1 - 3e^{-2} \\&\approx 0.594\end{aligned}$$

Example 2: Births in a hospital occur randomly at an average rate of 1.8 births per hour. What is the probability of observing **4 births in a given hour** at the hospital?

1. Let X = Number of births in a given hour:

$$\begin{aligned}\text{Mean rate } \lambda &= 1.8 \\ \Rightarrow X &\sim \text{Po}(1.8)\end{aligned}$$

2. Use the formula to calculate the probability of observing exactly 4 births in a given hour:

$$P(X = 4) = e^{-1.8} \frac{1.8^4}{4!} = 0.0723$$

5 Continuous Probability Distributions

A continuous distribution displays the **ranges of probabilities for the outcomes of a random variable with infinite values** and is used to model a continuous random variable. Continuous distributions has a range of values that are infinite, and therefore uncountable. For example, time is infinite since one could count from 0 seconds to a billion, a trillion and so on.

Continuous probability distribution: A probability distribution where the random variable X can take on any value i.e. continuous.

Recall that probability is defined as $\frac{\text{Desired outcome}}{\text{Number of possible outcomes}}$ e.g. a fortune wheel divided into 20 sectors resulting in a probability of $\frac{1}{20}$ of landing in any section. Imagine the fortune is then partitioned into an infinite number of sectors of equal area but each of those would have an area of 0. Due to the infinite nature and the definition of probability mentioned above, the **probability of a continuous random variable taking on any one specific value is 0**.

But remember the total probability must be 1 in order to satisfy one of the axioms of probability:

$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x) dx = 1$$

which creates an apparent contradiction. That is why when working with continuous distributions, probabilities are only assigned to an interval. The probability that X falls between two values (a and b) equals the integral (area under the curve) from a to b .

As mentioned previously with its infinite range, a continuous probability distribution cannot be expressed in tabular form like discrete probability distribution. Instead, an equation or formula is used to describe a continuous probability distribution i.e. **probability density function**.

The probability density function satisfies several condition properties:

- The random variable Y is a function of X i.e. $y = f(x)$
- The value of y is greater than or equal to zero for all values of x
- The total area under the curve of the function is equal to 1

5.1 Continuous uniform distribution: $X \sim \text{Unif}(a, b)$

The uniform distribution for continuous random variables is one of the simplest distribution. In a uniform distribution, all possible values of a continuous random variable inside some interval have the same probability.

Since all values of such random variable inside an interval have the same probabilities, the probability density function is defined as:

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

- Cumulative distribution function:

$$F_X(u) = \begin{cases} 0 & u < a \\ \int_a^u \frac{1}{b-a} dx = \frac{u-a}{b-a} & a \leq u \leq b \\ 1 & u > b \end{cases}$$

- Mean μ :

$$\mu = E(X) = \frac{(a+b)}{2}$$

- Variance σ^2 :

$$\sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

5.2 Exponential distribution: $X \sim \text{Exp}(\lambda)$

The exponential distribution is the probability distribution of the **time between events** in a Poisson point process i.e. a process in which events occur continuously and independently at a constant average rate λ . Exponential distribution and Poisson distribution is closely related. It can be considered the continuous analogue of the geometric distribution and it has the key property of being **memoryless**.

Commonly used to model the lifetime of electronic components. Previously mentioned Poisson distribution defined a random variable as the number of flaws along a length of copper wire. The **distance between the flaws** is another random variable of interest.

For example, let random variable X denote the length from any starting point on the wire until a flaw is found. The distribution of X (distance between flaws) can be obtained from knowledge of the distribution of the number of flaws.

In general if random variable N is defined as the number of flaws in x millimeters of wire and λ (per unit measurement) is defined as the mean number of flaws,

N has a Poisson distribution with **mean** of λx :

$$\begin{aligned} P(X > x) &= P(N = 0) \\ &= \frac{e^{-\lambda x} (\lambda x)^0}{0!} \\ &= e^{-\lambda x} \end{aligned}$$

Common assumption that the wire is longer than the value of x and the x is non-negative $x > 0$.

The CDF of function X is:

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

The PDF of function X is:

$$f(x) = \lambda e^{-\lambda x}$$

- Mean: $\mu = E(X) = \frac{1}{\lambda}$
- Variance: $\sigma^2 = V(X) = \frac{1}{\lambda^2}$

To verify that the PDF is valid, note that the density must be non-negative:

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x; \lambda) dx &= \int_0^{\infty} \lambda e^{-\lambda x} dx \\ &= [-e^{-\lambda x}]_0^{\infty} \\ &= [0 - (-1)] \\ &= 1 \end{aligned}$$

The starting point for measuring X does not matter because the probability of the number of flaws in an interval of a Poisson process depends on the length of the interval, not the starting location.

Example 1: Given a computer network, user log-ins can be modeled as Poisson process with a mean of 25 log-on **per hour** ($\mu = 25$). What is the probability that there are no log-ins in an interval of 6 minutes?

1. Specify the interval investigated:

Let X denote the **time in hours** from the start of the interval until the first log-in.

2. Establish provided data:

X has exponential distribution where $\lambda = 25$ log-in per hour.

3. Convert into same units of measurement:

As we are interested in the probability that X exceeds 6 minutes, 6 minutes in hours is 0.1 hours.

4. Extract data from PDF:

Note $\lambda = 25$ and $u = 0.1$

$$\begin{aligned} P(X > 0.1) &= \int_u^{\infty} \lambda e^{-\lambda x} dx \\ &= \int_{0.1}^{\infty} 25e^{-25x} dx \\ &= e^{-25(0.1)} \\ &= 0.082 \end{aligned}$$

5.2.1 Memoryless Property

Exponential distribution has an important interesting property with conditional probability.

Example 1: Let X denote the time between detections of a particle with a geiger counter and X has an exponential distribution with $\lambda = 1.4$ minutes.

1. The probability that a particle is detected within 30 seconds (0.5 minute) of starting the counter:

$$\begin{aligned} P(X < 0.5 \text{ minute}) &= F(0.5) \\ &= 1 - e^{-\frac{0.5}{1.4}} \\ &= 0.30 \end{aligned}$$

2. If 3 minutes has passed without a particle detected, what are the probability that a particle is detected in the next 30 seconds?

One might assume that the probability should be greater than 0.3 but the nature of exponential distribution means that **it is not true, the probability will remain the same**. The conditional probability of the above scenario after 3 minutes can be defined as:

$$P(X < 3.5 | X > 3) = \frac{P(3 < X < 3.5)}{P(X > 3)}$$

where $P(3 < X < 3.5)$ and $P(X > 3)$ are defined as:

- $P(3 < X < 3.5)$:

$$\begin{aligned} P(3 < X < 3.5) &= F(3.5) - F(3) \\ &= \left[1 - e^{-\frac{3.5}{1.4}} \right] - \left[1 - e^{-\frac{3}{1.4}} \right] \\ &= 0.0035 \end{aligned}$$

- $P(X > 3)$:

$$\begin{aligned} P(X > 3) &= 1 - F(3) \\ &= e^{-\frac{3}{1.4}} \\ &= 0.117 \end{aligned}$$

therefore:

$$P(X < 3.5 | X > 3) = \frac{0.035}{0.117} = 0.30$$

3. After 3 minute without a detection, the probability of particle detection within the next 30 seconds is the **same** probability of detection in the next 30 seconds after starting.

Only exponential distributions have this unique property:

Memoryless property: For an exponential random variable X :

$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2)$$

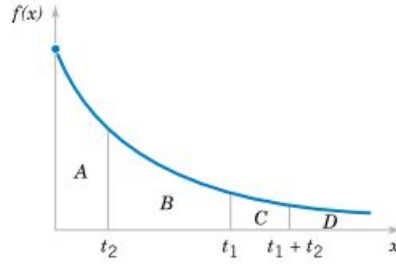


Figure 9: Graphical representation of *memoryless* property

The above figure shows that the proportion of the area of region A divided by total area under the PDF ($A + B + C + D$) is the same as the proportion for region C divided by the total area under the PDF ($C + D$).

The property is intuitive since, with Poisson distributions, it is assumed that intervals can be partitioned into small intervals that are **independent**. This is similar to Bernoulli trials where the knowledge of previous process does not affect the probabilities of events in future sub-intervals.

This type of modelling is seen in reliability studies of electronics for the time until device failure. Due to the memoryless property, regardless of how long the device has been operating, the probability of a failure in the next x hours is the same in the first x hours.

5.3 Normal distribution: $X \sim N(\mu, \sigma^2)$

The normal distribution is the most common distribution in all of probability and statistics. One of the main reasons it crops up so much is due to the Central Limit Theorem.

Originally found by De Moivre called the Central Limit Theorem and also called the Gaussian distribution. The most widely used model for the distribution of

a random variable. Whenever a random experiment is performed, the random variable corresponding to the average tends to have a normal distribution.

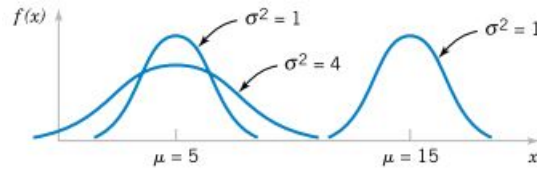


Figure 10: Normal probability distribution with different values of the parameters μ and σ^2

Normal random variable: A random variable X with probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi} \times \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ where: } -\infty < x < \infty$$

with parameters μ where $-\infty < \mu < \infty$ and σ where $\sigma > 0$.

- $E(X) = \mu$
- $V(X) = \sigma^2$

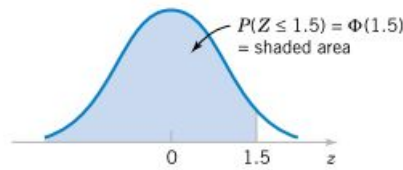


Figure 11: Standard normal distribution, note the symmetrical shape about μ

Standard normal random variable (Z): A normal random variable with $\mu = 0$ and $\sigma^2 = 1$

1. $\phi(z)$ - PDF of a standard normal random variable.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

2. Φ - CDF of a standard normal random variable.

$$\Phi(z) = \int_{-\infty}^z \phi(u) du$$

A standard normal distribution has:

- Mean μ of 0
- Variance σ^2 of 1

Change of variable techniques apply to normal distributions:

Example 1: If $Z \sim N(0, 1)$ and $X = \sigma Z + \mu$ then find $X \sim N(\mu, \sigma^2)$.

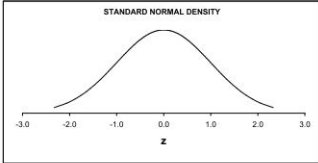
$$\begin{aligned} f_X(x) &= f_Z(z) \left| \frac{dz}{dx} \right| \\ &= \phi\left(\frac{x-\mu}{\sigma}\right) \left| \frac{d}{dx} \frac{x-\mu}{\sigma} \right| \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \end{aligned}$$

which is the PDF of $N(\mu, \sigma^2)$

5.3.1 Normal distribution table

With normal distribution, the CDF cannot be obtained usually. Probability calculations have to be used in conjunction with tables obtained through numeral integration such as:

THE STANDARD NORMAL DISTRIBUTION FUNCTION



Entries in table are probabilities p such that $\Phi(z)=p$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

TABLE OF THE STANDARD NORMAL CDF

The table is a tabulation such that $\Phi(z) = p$ for $z \geq 0$.

Example 1: Find $P(Z \leq 1.6)$.

$$P(Z \leq 1.6) = \Phi(1.6) = 0.9452$$

For $z < 0$, the symmetry of the standard normal distribution about its mean μ is exploited.

$$\begin{aligned}\Phi(z) + \Phi(-z) &= 1 \\ \Rightarrow \Phi(-z) &= 1 - \Phi(z)\end{aligned}$$

Example 2: Find $P(Z \leq -0.75)$.

$$\begin{aligned}P(Z \leq -0.75) &= 1 - \phi(0.75) \\ &= 1 - 0.7734 \\ &= 0.2266\end{aligned}$$

Note:

1. The table is applicable to the first decimal place, any questions can be rounded to the nearest position.
2. The Z is concentrated between -3 and 3 [Three-Sigma Rule].

5.3.2 Standardising

Since calculating probabilities of standard normal random variables requires a table, using the same approach for any arbitrary normal random variable would need a specific separate table for every possible pair of values for μ and σ . Alternatively, any normal distributions are related algebraically and can be transformed into the standard normal random variable.

Any normal distribution can be transformed into a standard normal by:

1. Subtracting the mean μ
2. Dividing by the standard deviation σ

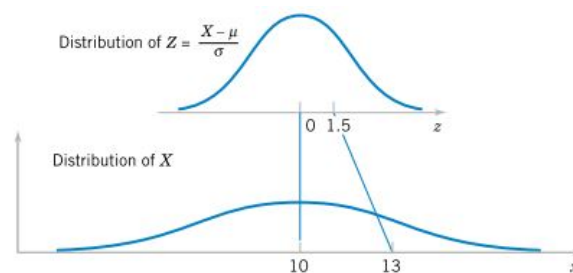


Figure 12: Graphical representation of Standardising

$$Z = \frac{X - \mu}{\sigma}$$

The new variable Z represents the distance of X from its mean (μ) in terms of standard deviations (σ).

The **Three Sigma Rule** still applies and is defined as the following: X is concentrated between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Example 1: Given $X \sim N(3, 5^2)$, standardise it and the following probabilities: $P(X < 5)$ and $P(0.5 < X < 5.5)$.

1. Standardise the expression X :

$$Z = \frac{(X - 3)}{5} \sim N(0, 1)$$

2. Compute $P(X < 5)$:

$$\begin{aligned} P(X < 5) &= P\left(\frac{X - 3}{5} < \frac{5 - 3}{5}\right) \\ &= P(Z < 0.4) \\ &= \Phi(0.4) \\ &= 0.655 \end{aligned}$$

3. Compute $P(0.5 < X < 5.5)$:

$$\begin{aligned} P(0.5 < X < 5.5) &= P\left(\frac{0.5 - 3}{5} < \frac{X - 3}{5} < \frac{5.5 - 3}{5}\right) \\ &= \Phi(0.5) - \Phi(-0.5) \\ &= \Phi(0.5) - [1 - \Phi(0.5)] \\ &= 2\Phi(0.5) - 1 \\ &= 2(0.6915) - 1 \\ &= 0.3830 \end{aligned}$$

5.4 Chi-squared distribution: $X \sim \chi^2(k)$

The chi-square distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. It is a special case of the gamma distribution and is one of the most widely used probability distributions in inferential statistics.

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(Z^2 \leq u) \\ &= \begin{cases} 0, & \text{if } u < 0, \\ P(-\sqrt{u} \leq Z \leq \sqrt{u}) = \int_{-\sqrt{u}}^{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, & \text{if } u \geq 0. \end{cases} \end{aligned}$$

Differentiating both sides w.r.t. u gives

$$f_U(u) = \begin{cases} \frac{1}{2\sqrt{u}\sqrt{2\pi}} e^{-\frac{u}{2}} + \frac{1}{2\sqrt{u}\sqrt{2\pi}} e^{-\frac{u}{2}} = \frac{1}{\sqrt{u}\sqrt{2\pi}} e^{-\frac{u}{2}}, & \text{if } u > 0, \\ 0, & \text{if } u \leq 0. \end{cases}$$

5.5 Log-normal distribution: $X \sim \text{lognormal}(\mu, \sigma^2)$

A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution. Equivalently, if Y has a normal distribution, then the exponential function of Y , $X = \exp(Y)$ has a log-normal distribution.

A random variable which is log-normally distributed takes only positive real values. It is a convenient and useful model for measurements in exact and engineering sciences, as well as medicine, economics and other topics e.g., energies, concentrations, lengths, financial returns and other metrics.

Let $X \sim N(\mu, \sigma^2)$ and take $Y = e^X$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= \begin{cases} 0, & \text{if } y \leq 0, \\ P(e^X \leq y) = P(X \leq \ln y), & \text{if } y > 0 \end{cases} \end{aligned}$$

with

$$P(X \leq \ln y) = \int_{-\infty}^{\ln y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Differentiating both sides w.r.t. y gives

$$f_Y(y) = \begin{cases} \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}, & \text{if } y > 0, \\ 0, & \text{if } y \leq 0. \end{cases}$$

6 Change of Variable Technique

If the probability density function of a random variable X is given as $f_X(x)$, it is possible to calculate the probability density function of some variable $Y = g(X)$.

There is a simple rule to compute the probability density function of the new random variable Y in terms of the probability density function of the original random variable X .

6.1 Generalisation for an increasing function

Let:

- X be a **continuous random variable**.
- $g(x)$ a strictly monotonic function (one-to-one function) defined over the support $c_1 < x < c_2$.
- $Y = g(X)$ be a continuous and increasing function of X with inverse function $X = g^{-1}(Y)$

The blue curve represents the continuous and increasing function $Y = g(X)$

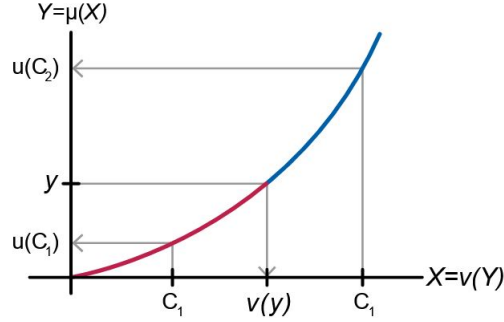


Figure 13: Graphical representation showing that all the inequalities hold

$[Y = g(X)]$. If x values are inputted into the function then y values $\mu(x)$ $[y]$ are received. Since the function is **continuous**, **one-to-one** and **increasing**, an inverse function $X = v(Y)$ $[X = g^{-1}(Y)]$ where y values are inputted then x values $v(y)$ $[x]$ are received.

The distribution of function Y can be derived as:

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq g(x)] = P(X \leq x) = F_X(x)$$

The PDF $f_Y(y)$ is the derivative of the CDF of $F_Y(y)$.

$$f_Y(y) = F_Y'(y) = f_X[v(y)] \cdot v'(y)$$

or

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

6.2 Generalisation for a decreasing function

Let:

- X be a **continuous random variable**.
- $g(x)$ a strictly monotonic function (one-to-one function) defined over the support $c_1 < x < c_2$.
- $Y = g(X)$ be a continuous and decreasing function of X with inverse function $X = g^{-1}(Y)$.

The distribution of function Y can be derived as:

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq g(x)] = P(X \geq x) = 1 - F_X(x)$$

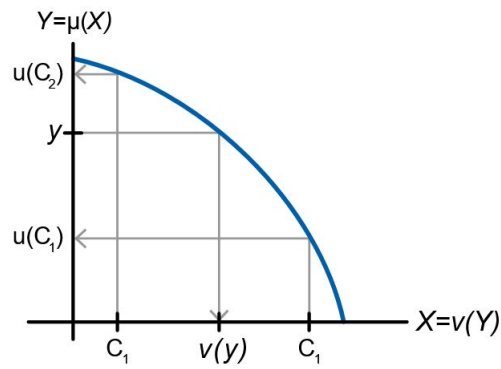


Figure 14: Graphical representation showing that all the inequalities hold

The PDF $f_Y(y)$ is the derivative of the CDF of $F_Y(y)$.

$$f_Y(y) = F'_Y(y) = -f_X[v(y)] \cdot v'(y)$$

or

$$f_Y(y) = \frac{d F_X(x)}{dy} = \frac{d F_X(x)}{dx} \frac{dx}{dy} = -f_X(x) \frac{dx}{dy}$$