1: PROBABILITY DISTRIBUTIONS

1.1: DISCRETE RANDOM VARIABLES

- The results of some experiments can take different values on different occasions - the result is said to be a **variable**.
- If the experiment involves random selection, the variable itself is called a **random variable**.
- We usually denote random variables with capital letters. The set of possible values that a random variable X can take is the **range** of X.
- Hence, the range of a random variable X is analogous to the sample space of an unstructured experiment. An outcome of an experiment is analogous to an element of the range of X, and an event of an experiment is analogous to a subset of the range of X.
- A random variable X is said to be discrete if its range consists of a countable number of values. In other words, if we form a one-to-one correspondence between the elements of the range of X and the natural numbers, X is discrete
- For example, let X be the number of heads in the first 5 tosses of a coin, and let Y be the number of tosses before the first head appears. X has range $\{0,1,2,3,4,5\}$, but Y has range $\{0,1,2,3,4 \ldots\} = \mathbb{N}$. Regardless, both random variables are discrete.
- The **probability function** of a discrete random variable X is the function $p(x)$ satisfying $P(X = x)$, where x is an element of the range of X.
- The probability function satisfies the following axioms:

1: $0 \leq P(X = x) \leq 1 \; \forall x \in \text{range of X}, : P(X = x) = 0 \; \forall x \notin \text{range of X}$

2: $\sum_{i=1}^{n} P(X = x_i) = 1$ if the range of X is $\{x_1, x_2, x_{3,\ldots,} x_n\}$.

- An alternative to the probability function is the **(cumulative) distribution function** of X $F_X(x_i)$, defined as:

$$F_X(x) = P(X \leq x) \; \forall x \in \text{range of X}$$

$$F_X(x) = \sum_{y \leq x} p(y)$$

- The distribution function is monotonic increasing, and must satisfy the conditions $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.

1.2: EXPECTATION AND VARIANCE OF DISCRETE RANDOM VARIABLES

- The **expectation** (mean) of a random variable X is denoted by $E[X]$ where:

$$E[X] = \sum_{x} xP(X = x)$$

- We often denote the expectation by μ, i.e. $\mu = E[X]$

- **Variance** is our preferred measure of dispersion. The variance of a random variable X is denoted by $\text{Var}[X]$ and is defined as:

$$\text{Var}[X] = E[(X - E(X)^2] = E[(X - \mu)^2]$$

- This simplifies to:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

- We often denote the expectation by σ^2, i.e. $\sigma^2 = \text{Var}[X]$. The square root of the variance, σ, is called the **standard deviation**.

1.3: INTRODUCTION TO DISTRIBUTIONS

- **Distributions** refer to how the probabilities are allocated amongst the elements of the range of X.
- Distributions can be arbitrarily defined, or they can be determined by parameters of a certain statistical distribution.
- Consider the following distribution

x	1	2	3	4	5
$P(X = x)$	0.1	0.1	0.2	0.4	0.2

- This distribution is arbitrary. Its expectation and variance can be calculated with ease:

$$E[X] = \sum_x xP(X = x) = 1(0.1) + 2(0.1) + 3(0.2) + 4(0.4) + 5(0.2) = 3.5$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 1(0.1) + 4(0.1) + 9(0.2) + 16(0.4) + 25(0.2) = 13.7$$

- There are many distributions which we can use to assign probabilities to the range of a random variables. We shall explore some of them now.
- Statistical distributions are typically defined by **parameters**. One may describe the distribution of a random variable as belonging to a family of probability distributions, distinguished from each other by the values of a finite number of parameters

1.4: DISCRETE UNIFORM

- A random variable is said to have a **discrete uniform distribution** if each element of its range is equiprobable.
- Suppose the range of a random variable X is $\{1,2,3,4, \ldots n\}$. If X follows a discrete uniform distribution, then:

$$P(X = x) = \begin{cases} \dfrac{1}{n} & x = 1,2,3,4, \ldots, n \\ 0 & \text{Otherwise} \end{cases}$$

- The distribution function of X is given as:

$$P(X \leq x) = \begin{cases} 0 & x < 1 \\ \dfrac{1}{n}\lfloor x \rfloor & 0 \leq \text{x} \leq \text{n} \\ 1 & x > n \end{cases}$$

- The expectation of a random variable X with a discrete uniform distribution with range $\{1,2,3,4,...n\}$ is

$$\mathrm{E}[X] = \sum_{x=1}^{n} xP(X = x) = \frac{1}{n}\sum_{x=1}^{n} x = \frac{1}{n}\left[\frac{n(n+1)}{2}\right] = \frac{n+1}{2}$$

- The variance of a random variable X with a discrete uniform distribution with range $\{1,2,3,4,...n\}$ is

$$\mathrm{Var}[X] = \sum_{x=1}^{n} x^2 P(X = x) - (\mathrm{E}[X])^2 = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{4(n+1)(2n+1)}{24} - \frac{6(n+1)^2}{24} = \frac{4[2n^2+3n+1] - 6[n^2+2n+1]}{24}$$

$$= \frac{2n^2-2}{24} = \frac{n^2-1}{12} \qquad \text{So } \mathrm{Var}[X] = \frac{n^2-1}{12}$$

- A general discrete uniform distribution can take any discrete value in the interval $[a, b]$. The examples above are for the common case $[1, n]$, whilst $[a, b]$ is more general.
- $[a, b]$ is regarded as the general parameters for the distribution. The expectation of a random variable X with a discrete uniform distribution with parameters $[a, b]$ is:

$$E[X] = \frac{a+b}{2}$$

- The variance of a random variable a random variable X with a discrete uniform distribution with parameters $[a, b]$ is:

$$\frac{(b-a+1)^2 - 1}{12}$$

1.5: BERNOULLI DISTRIBUTION

- A Bernoulli trial is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success p is the same every time the experiment is conducted.
- We usually denote the event $\{X = 1\}$ as success and the event $\{X = 0\}$ as failure.
- Hence, $P(X = 1) = p$ and $P(X = 0) = 1 - p$. This is a very simple probability distribution.

1.6: GEOMETRIC DISTRIBUTION

- Bernoulli trials are continued until the first success occurs. A random variable X denotes the number of failures.
- If the constant probability of success is p, then X is said to have a geometric distribution with parameter p, and X has probability function:

$$P(X = x) = (1-p)^x p$$

- The range of X is the natural numbers.
- The expectation of X is given as such:

$$E[X] = pQ \Rightarrow \mathrm{E}[X] = \frac{1-p}{p}$$

- By a similar method, $\mathrm{Var[X]} = \frac{1-p}{p^2}$

1.7: BINOMIAL DISTRIBUTION

- n *independent* Bernoulli trials are conducted (and this number must remain fixed), the *constant* probability of success in each one being p. Success or failure must be the only two outcomes.
- The random variable X is defined as the total number of successes in these n trials.
- X is said to have a **binomial distribution** with index n and parameter p.
- We write this is as $X{\sim}B(n, p)$.
- The probability function of a binomial distribution is as such:

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

- The binomial (combinatorial) coefficient $\binom{n}{x}$ has to be incorporated into the probability function as there may be many outcomes favourable to $\{X = x\}$.
- Does $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ satisfy the condition that $\sum_{i=1}^{n} P(X = x_i) = 1$?
- Yes, it does:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \{p + (1-p)\}^n = 1$$

- The expectation of a random variable X which follows a binomial distribution with index n and parameter p is:

$$E[X] = np$$

- The variance of a random variable X which follows a binomial distribution with index n and parameter p is:

$$\mathrm{Var}[X] = np(1-p)$$

- It is important to note that, for all discrete probability distributions $P(X \leq x) \neq P(X < x)$ (in general). This is because $P(X < x) = P(X \leq x - 1)$ in discrete systems.
- The cumulative probabilities $P(X \leq x)$ for the binomial distribution, with varying values of n, are tabulated.
- Note that $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = P(X \leq b) - P(X \leq a - 1)$

1.8: POISSON DISTRIBUTION

- A random variable X is said to have a Poisson distribution with parameter $\mu > 0$ if the probability function of X is:

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

- If X follows a Poisson distribution with parameter $\mu > 0$, we write $X{\sim}Po(\mu)$.
- For a random variable X to be modelled by a Poisson distribution, the following conditions must be satisfied:
    - Events happen singly in space and time.
    - Events are independent.

- The probability that a event will occur is proportional to the size of the region.
  - The probability that a event will occur in an extremely small region is virtually zero.
- Suppose X is the number of red cars passing an outlook in a period of time 1 hour long. Suppose the rate of red cars is 42 in this hour. Then X could be modelled be a Poisson distribution with parameter 42 (i.e. $X \sim Po(42)$)
- Suppose Y is the number of red cars passing an outlook in a period of time 30 minutes long. Then, assuming the rate of red cars remains unchanged, $Y \sim Po(21)$.
- The expectation of a random variable X which follows a Poisson distribution with parameter μ is:

$$E[X] = \mu$$

- The variance of a random variable X which follows a Poisson distribution with parameter μ is:

$$\mathrm{Var}[X] = \mu$$

- The variance is equal to the mean if X follows a Poisson distribution.

Summary

Discrete distribution	**Parameters**	**Probability function $\boldsymbol{P(X = x)}$**	**Expectation $\boldsymbol{E[X]}$**	**Variance $\mathbf{Var}[\boldsymbol{x}]$**
Discrete uniform distribution	$[a, b]$ $n = b - a + 1$	$\begin{cases} \frac{1}{n} & x \in [a, b]; x \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{n^2 - 1}{12}$
Bernoulli distribution	$0 < p < 1$	$P(X = 0) = 1 - p$ $P(X = 1) = p$	p	$p(1 - p)$
Geometric distribution	$0 < p < 1$	$(1 - p)^x p$	$\frac{1 - p}{p}$	$\frac{1 - p}{p^2}$
Binomial distribution	$X \sim B(n, p)$ $n \in \mathbb{N}$ $0 < p < 1$	$\binom{n}{x} p^x (1 - p)^{n-x}$	np	$np(1 - p)$
Poisson distribution	$X \sim Po(\mu)$ $\mu > 0$	$\frac{e^{-\mu}\mu^x}{x!}$	μ	μ

1.9: CONTINUOUS DISTRIBUTIONS

- A random variable is continuous if its cumulative distribution function is a continuous function.
- A continuous random variable does not possess a probability function, as:

$$F_X(x) = P(X \le x) \text{ is continuous} \Rightarrow P(X \le x) = P(X < x) \Rightarrow P(X = x) = 0$$

- Like the discrete analogue, the continuous distribution function must be monotonic increasing, and must satisfy the conditions $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
- Instead, probabilities are assigned to continuous intervals of the range.
- The events $\{X \le a\}$ and $\{a < x \le b\}$ are mutually exclusive, and $\{X \le a\} \cup \{a < x \le b\} = \{X \le b\}$, so:

$$P(a < x \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a)$$

- We do not distinguish between open and closed intervals for continuous random variables.

1.10: PROBABILITY DENSITY FUNCTION

- We define $f_X(x)$ as the **probability density function** of the continuous random variable X. It is the derivative of the cumulative distribution function of the continuous random variable X, i.e.:

$$f_X(x) = \frac{dF_X(x)}{dx} \Leftrightarrow F_X(x) = \int f_X(x)dx$$

- More precisely:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

- $f_X(x) \ge 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$ must also be satisfied.
- In practice, ∞ is typically replaced with the upper bound of the range (unless, of course, the range *is* infinite), whilst $-\infty$ is typically replaced with the lower bound of the range.
- Finally:

$$P(a < x \le b) = \int_{a}^{b} f_X(x)dx = F_X(b) - F_X(a)$$

1.11: EXPECTATION AND VARIANCE OF CONTINUOUS RANDOM VARIABLES

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

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} x^2 f_X(x)dx - (E[X])^2$$

1.12: CONTINUOUS UNIFORM

- If a continuous random variable X follows a continuous uniform distribution with parameters $[a, b]$, X is equally likely to take any value in the interval $[a, b]$. We write $X \sim U[a, b]$.
- More precisely, if $[c, d] \subseteq [a, b]$ where $[c, d]$ has width $h \ne 0$, and if $[c, d] \ne [m, n] \subseteq [a, b]$ where $[m, n]$ has width h, then:

$$P(X \in [c, d]) = P(X \in [m, n])$$

- What is the probability density function of the continuous uniform distribution? Well, the probability remains constant over the interval. Let this probability be k. Then:

$$\int_{a}^{b} kdx = 1 \Rightarrow k = \frac{1}{b - a}$$

- So if $X \sim U[a,b]$

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

- What is the continuous distribution function of $X \sim U[a,b]$?:

$$x < a \Rightarrow F_X(x) = 0$$

$$x \in [a,b] \Rightarrow F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{a}^{x} \frac{1}{b-a} dy = \frac{x-a}{b-a}$$

$$x > b \Rightarrow F_X(x) = 1$$

- Note how we use the dummy variable t in the integral above. Collating these results:

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

- The expectation and variance of the continuous uniform distribution are given as follows:

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

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\frac{a+b}{2}\right)^2 = \int_{a}^{b} \frac{x^2}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \left[\frac{x^3}{3[b-a]}\right]_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{1}{12}(b-a)^2$$

1.13: EXPONENTIAL DISTRIBUTION

- The exponential distribution with parameter μ has the following probability density function:

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \mu e^{-\mu x} & x \ge 0 \end{cases}$$

- The cumulative distribution function can be calculated as such:

$$x < 0 \Rightarrow F_X(x) = 0$$

$$x \ge 0 \Rightarrow F_X(x) = \int_{0}^{x} \mu e^{-\mu y} dy = [-e^{-\mu y}]_0^x = -e^{-\mu x} - -1 = 1 - e^{-\mu x}$$

- Hence:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\mu x} & x \ge 0 \end{cases}$$

- The expectation and variance of an exponential distribution are as such:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^{\infty} x\mu e^{-\mu x}\, dx = \mu \int_0^{\infty} x e^{-\mu x} = \frac{\mu}{\mu^2} = \frac{1}{\mu}$$

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - (E[X])^2 = \int_0^{\infty} x^2 \mu e^{-\mu x}\, dx - \frac{1}{\mu^2} = \mu \int_0^{\infty} x^2 e^{-\mu x}\, dx - \frac{1}{\mu^2} = \frac{1}{\mu^2}$$

- These results require integration by parts.

1.14: NORMAL DISTRIBUTION

- The normal distribution is regarded as the most important distribution in all statistics.
- Suppose a continuous random variable X has the following probability density function:

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

- If this is the case, we say X has a normal distribution with parameters μ and σ^2 (i.e. $X \sim N(\mu, \sigma^2)$).
- The cumulative distribution function is given as:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- The **standard normal distribution Z** is defined as $Z \sim N(0,1)$.
- The probability density function of the standard normal distribution is denoted by $f_Z(z) = \varphi(z)$.
- The cumulative distribution function of the standard normal distribution is denoted by $F_Z(z) = \Phi(z)$.

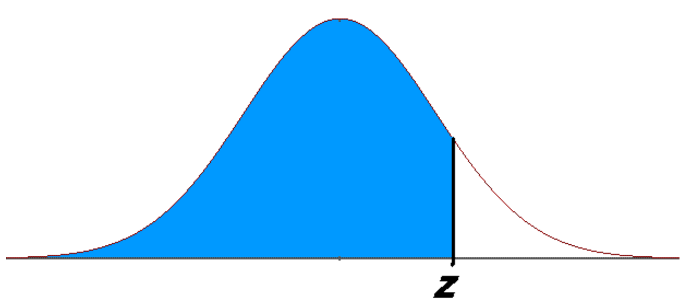
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < \infty$$

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = 1$$

- $\Phi(z)$ returns the probability that a continuous random variable with a standard normal distribution will take a value between $-\infty$ and z.
- As $\Phi(z)$ is an integral with no elementary anti-derivative, it can only be evaluated using numerical methods.
- To make life easier, $\Phi(z)$ is fully tabulated, usually in the region of $0 \le z \le 4$.
- To find $\Phi(-z)$, use the following identity:

$$\Phi(-z) = 1 - \Phi(z)$$



- This identity comes from the fact that the standard normal distribution is symmetrical in the line $z = 0$. Also useful is $z\Phi(z) = -(-z)\Phi(-z)$, as this implies $z\Phi(z)$ is an odd function.
- Also, as a result of the symmetrical nature of $\Phi(z)$:

$$\Phi(0) = 0.5$$

- Furthermore:

$$\Phi(-\infty) = 0$$

$$\Phi(\infty) = 1$$

- We can convert any normal distribution $X \sim N(\mu, \sigma^2)$ into the standard normal distribution $Z \sim N(0,1)$ using the following statistic:

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

- Hence:

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

- Thus the probability of any event for any normal distribution can be defined using this function. However, the tabulations only define $\Phi(z)$ for P(Z<z), and z is only tabulated for positive values, so manipulations may be in order.

> For example, if $X \sim N(1,2^2)$, what is the value of P(X<3)?
>
> Well $Z = \frac{3-1}{2} = 1 \Rightarrow P(X < 3) = P(Z < 1) = \Phi(1) = 0.8413.$
>
> What is the value of P(X<0)?
>
> Well $Z = \frac{0-1}{2} = -\frac{1}{2} \Rightarrow P(X < 0) = P\left(Z < -\frac{1}{2}\right) = \Phi\left(-\frac{1}{2}\right) = 1 - \Phi\left(\frac{1}{2}\right) = 1 - 0.6915 = 0.3085.$

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

- The expectation and variance of $X \sim N(\mu, \sigma^2)$ are as such:

$$E[X] = \mu$$

$$\mathrm{Var}[X] = \sigma^2$$