PROBABILITY & STATISTICS

Random Variables

* A random variable is a function that assigns numerical values to the outcomes of a random experiment. Formally, if Ω is the outcome space (sample space) of a random experiment, then a function $X : \Omega \rightarrow \mathbb{R}$ is called a random variable.

Key points

* Random variables transform experimental outcomes into numbers.
* They allow us to work with numerical values instead of abstract outcomes.
* The specific values cannot be predicted with certainity before the experiment.
* They bridge the gap between qualitative outcomes and quantitative analysis.

Ex Consider tossing a coin three times.

Outcome space (Ω) = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

* Random variables can be defined as;
  - X = number of heads observed
  - Y = number of tails observed
  - Z = XY (Product of X and Y)

Discrete Random Variables

* A random variable X is discrete if it takes on only a finite (or countably infinite) number of distinct values.

Characteristics

* The sample space Ω_X contains discrete, separate values.
* Common examples include counting outcomes, number of successes, etc.
* Can be represented in tabular form.

Ex Consider the previous coin toss example with the defined random variables.

The functions X, Y and Z are conveniently specified by the table below.

	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X	3	2	2	2	1	1	1	0
Y	0	1	1	1	2	2	2	3
Z	0	2	2	2	2	2	2	0

$\therefore \Omega_X = \Omega_Y = \{0, 1, 2, 3\}$

$\therefore \Omega_Z = \{0, 2\}$

Ex A coin is tossed until two successive tosses give the same result. Let X = number of heads observed and Y = number of tosses made. Give Ω_X, the sample space of X and Ω_Y, the sample space of Y.

$\Omega = \{HH, TT, HTT, THH, HTHH, THTT, HTHTT, THTHH, \ldots\}$

$\therefore \Omega_X = \{0, 1, 2, \ldots\}$

$\Omega_Y = \{2, 3, 4, \ldots\}$

No : ______ Date: __/__/____

Probability Distributions

* The probability function of a discrete random variable X is defined as;

$$P_{(x)} = \{P(X=x) \text{ if } x \in \Omega_X ;\ 0 \text{ otherwise}\}$$

* $P(X=x)$ represents the probability that the random variable X takes the specific value x. The probability function assigns a probability to each possible value of X. For values not in the sample space, the probability is automatically zero.

* The probability distribution describes how probability mass is distributed over all possible values in Ω_X.

Properties

* $P_{(x)} \geqslant 0$ for all x.
* $\Sigma P_{(x)} = 1$ (probabilities sum to 1)
* $P_{(x)} = 0$ for $x \notin \Omega_X$

$\boxed{\text{Ex}}$ Toss a coin three times and observe the sequence of heads and tails. Suppose the results on the tosses of the coin are independent and that any given toss falls heads with probability $1/3$.

Let X = Number of heads observed

Y = Number of tails observed

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$P(\omega)$	$1/27$	$2/27$	$2/27$	$2/27$	$4/27$	$4/27$	$4/27$	$8/27$
$X(\omega)$	3	2	2	2	1	1	1	0
$Y(\omega)$	0	1	1	1	2	2	2	3

$P(X=0) = 8/27$ $\quad P(X=1) = 12/27$ $\quad P(X=2) = 6/27$ $\quad P(X=3) = 1/27$

$P(Y=0) = 1/27$ $\quad P(Y=1) = 6/27$ $\quad P(Y=2) = 12/27$ $\quad P(Y=3) = 8/27$

No:

Date: / /

Functions of Random Variables

* If X is a random variable, then any function of X (like: $2x+5$, x^2, etc.) is also a random variable.

Ex A machine is used to cut rectangular shapes out of sheets of metal. Switch A has 2 settings which are used to select the length L of the rectangle (2 cm or 3 cm). Switch B has three settings which are used to select the width W of the rectangle (1 cm, 2cm or 3cm). A mechanic tests the machine by selecting settings for switches A and B and making a rectangle. The length and width are observed. Suppose the settings are chosen randomly in a way so that in the table below.

	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
$P_{(\omega)}$	0.12	0.13	0.25	0.2	0.15	0.15

a) Find $P_{(L)}$, the probability distribution of length.

b) Find $P_{(W)}$, the probability distribution of width.

c) Find $P_{(X)}$, the probability distribution of perimeter $X = 2L + 2W$.

	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
$P_{(\omega)}$	0.12	0.13	0.25	0.2	0.15	0.15
L	2	2	2	3	3	3
W	1	2	3	1	2	3
X	6	8	10	8	10	12

a) Probability distribution of length;

$$P_{(L=2)} = 0.12 + 0.13 + 0.25 = 0.5$$

$$P_{(L=3)} = 0.2 + 0.15 + 0.15 = 0.5$$

b) Probability distribution of width;

$$P_{(W=1)} = 0.12 + 0.2 = 0.32$$

$$P_{(W=2)} = 0.13 + 0.15 = 0.28$$

$$P_{(W=3)} = 0.25 + 0.15 = 0.4$$

c) Probability distribution of perimeter;

$P(x=6) = 0.12$

$P(x=8) = 0.13 + 0.2 = 0.33$

$P(x=10) = 0.25 + 0.15 = 0.4$

$P(x=12) = 0.15$

Expectation

* Let x be a discrete random variable with range $\Omega_x = \{x_1, x_2, x_3, ..., x_n\}$. The expected value of x, denoted by $E(x)$ is defined by;

$$\boxed{E(x) = \sum_{i=1}^{n} x_i P(x_i)}$$

* This quantity is also referred to as the mean of x, or the mean of the probability distribution of x or the expected value of x.

* Expectation represents the 'average' value we expect over many repetitions. It provides a measure of central tendency or 'location'.

Properties

* Linearity: $E(ax+b) = aE(x) + b$
* Additivity: $E(x+y) = E(x) + E(y)$
* For constants: $E(c) = c$

Ex Calculate the expected value of x of the following probability model.

x_i	x_1	x_2	x_3	x_4	x_5	x_6
$P(x_i)$	0.1	0.2	0.2	0.1	0.3	0.1
$X(x_i)$	3	3	2	1	3	2

$E(x) = (0.1\times3) + (0.2\times3) + (0.2\times2) + (0.1\times1) + (0.3\times3) + (0.1\times2)$

$= 0.3 + 0.6 + 0.4 + 0.1 + 0.9 + 0.2$

$= 2.5$

The random variable X has the probability distribution given by;

x	-2	-1	0	1	2
$P(x)$	0.3	0.1	0.2	0.1	0.3

i. Find $E(x)$.

ii. Find $E(x^2)$.

i. $E(x) = (-2\times0.3)+(-1\times0.1)+(0\times0.2)+(1\times0.1)+(2\times0.3)$

$= 0$

ii.

x	-2	-1	0	1	2
x^2	4	1	0	1	4
$P(x)$	0.3	0.1	0.2	0.1	0.3

$E(x^2) = (4\times0.3)+(1\times0.1)+(0\times0.2)+(1\times0.1)+(4\times0.3)$

$= 2.6$

Variance and Standard Deviation

* The variance of X measures the spread of the distribution;

$$\boxed{\begin{array}{c} Var(x) = \sigma_x^2 = E[(x-\mu)^2] = \sum_{i=1}^{n}(x_i-\mu)^2 P(x_i) \\ \hline Var(x) = E(x^2) - [E(x)]^2 = E(x^2) - \mu^2 \end{array}}$$

; where $\mu = E(x)$

* Then σ_x is called the standard deviation of x. Both the standard deviation and the variance can be used as measures of the spread of a probability distribution. However, the standard deviation is commonly used since it is a measurement in the same unit as x.

Properties

- **Scaling** : $Var(cX) = c^2 Var(X)$
- **Independence** : If X and Y are independent;
  - $Var(X+Y) = Var(X) + Var(Y)$
  - $Var(X-Y) = Var(X) + Var(Y)$
- **Constants** : $Var(X+c) = Var(X)$

Ex Toss an unbiased coin once and win \$k if it is a head and lose \$k if it is a tail. Find the standard deviation of the profit X if the probability distribution of X is given by;

X	$-k$	k
$P(x)$	0.5	0.5

X	$-k$	k
X^2	k^2	k^2
$P(x)$	0.5	0.5

$E(X) = \mu = 0$

$\therefore Var(X) = E(X^2) - \mu^2$

$= k^2$

$\therefore \sigma_X = \underline{k}$

No: Date: / /

Ex Suppose that a game is played with a single die assumed fair. In this game, a player wins \$20 if a 2 turns up, \$40 if a 4 turns up and loses \$30 if a 6 turns up while he neither wins or loses if any other face turns up. Find the expected sum of money won.

x	0	20	0	40	0	-30
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

Expected sum of money won $(E(x)) = 0 + 10/3 + 0 + 20/3 + 0 - 5$

$= 10 - 5$

$= \$5$

Common Mistakes to Avoid

* Forgetting that probabilities must sum to 1.
* Confusing $E(x^2)$ with $[E(x)]^2$
* Not checking if random variables are independent before using additivity properties
* Mixing up variance formulas for sums and difference.

Binomial Distribution

* A binomial distribution arises from experiments with exactly two possible outcomes.
  - Success - The outcome we're interested in counting
  - Failure - The complementary outcome

Key requirements for binomial situations

* Fixed number of trials
* Independent trials (Each trial doesn't affect others)
* Same probability of success (p) for each trial.
* Only two possible outcomes per trial.

Common examples

* Tossing a coin 6 times (success = heads)
* Rolling a dice 10 times (success = even number)
* Testing 100 products (success = defective item)

* Let p be the probability that an event will happen in any single trial. The probability that the event will fail to happen in any single trial is $1-p$. The probability that the event will happen exactly k times in n independent trials is given by the probability function;

$$\boxed{P(X=k) = {}^{n}C_{k}\, p^{k} (1-p)^{n-k}}$$

* When the random variable X denote the number of successes in n independent trials, the above given function is often called the binomial distribution since for $k = 0, 1, 2, \ldots, n$, it corresponds to successive terms in the binomial expansion of $[p + (1-p)]^n$.

Ex What is the probability of getting exactly 2 heads in 6 tosses of a fair coin?

$X \sim Bin(6, 1/2)$

$$P(X=2) = {}^{6}C_{2} (1/2)^2 (1/2)^4$$
$$= 15 \times 1/4 \times 1/16$$
$$= \frac{15}{64}$$

No :

Date: / /

Ex Find the probability that in tossing a fair coin three times, there will appear,

i. 3 heads

ii. 2 tails and 1 head

iii. at least 1 head

iv. not more than 1 tail

i. Let X be number of heads.

$X \sim Bin(3, \frac{1}{2})$

$$P(X=3) = {}^3C_3 (\tfrac{1}{2})^3 (\tfrac{1}{2})^0$$

$$= 1 \times \tfrac{1}{8} \times 1$$

$$= \frac{1}{8}$$

ii. Let Y be number of tails.

$Y \sim Bin(3, \frac{1}{2})$

$$P(Y=2) = {}^3C_2 (\tfrac{1}{2})^2 (\tfrac{1}{2})^1$$

$$= 3 \times \frac{1}{4} \times \frac{1}{2}$$

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

iii. $P(X \geq 1) = 1 - P(X=0)$

$$= 1 - {}^3C_0 (\tfrac{1}{2})^0 (\tfrac{1}{2})^3$$

$$= 1 - \left(1 \times 1 \times \frac{1}{8}\right)$$

$$= \frac{7}{8}$$

iv. $P(X \geq 2) = P(X=2) + P(X=3)$

$= {}^3C_2 (\frac{1}{2})^2 (\frac{1}{2})^1 + {}^3C_3 (\frac{1}{2})^3 (\frac{1}{2})^0$

$= 3 \times \frac{1}{4} \times \frac{1}{2} + 1 \times \frac{1}{8} \times 1$

$= \frac{4}{8}$

$= \frac{1}{2}$

OR

$P(Y \leq 1) = P(Y=0) + P(Y=1)$

$= {}^3C_0 (\frac{1}{2})^0 (\frac{1}{2})^3 + {}^3C_1 (\frac{1}{2})^1 (\frac{1}{2})^2$

$= 1 \times 1 \times \frac{1}{8} + 3 \times \frac{1}{2} \times \frac{1}{4}$

$= \frac{4}{8}$

$= \frac{1}{2}$

$\boxed{\text{Ex}}$ Find the probability that in 5 tosses of a fair die a 3 appears;

i. twice

ii. at most once

iii. at least two times

i. Let x be number of appearances of 3.

$x \sim Bin(5, \frac{1}{6})$

$P(x=2) = {}^5C_2 (\frac{1}{6})^2 (\frac{5}{6})^3$

$= 10 \times \frac{1}{36} \times \frac{125}{216}$

$= \frac{625}{3888}$

ii. $P(x \leq 1) = P(x=0) + P(x=1)$

$= {}^5C_0 (\frac{1}{6})^0 (\frac{5}{6})^5 + {}^5C_1 (\frac{1}{6})^1 (\frac{5}{6})^4$

$= 1 \times 1 \times \frac{3125}{7776} + 5 \times \frac{1}{6} \times \frac{625}{1296}$

$= \frac{3125}{3888}$

No: Date: / /

iii. $P(x>1) = 1-[P(x=0)+P(x=1)]$

$= 1-\frac{3125}{3888}$

$= \frac{763}{3888}$

Ex Find the probability that in a family of 4 children, there will be;

i. at least 1 boy

ii. at least 1 boy and at least 1 girl. (Assume male birth probability = ½)

i. Let x be a child being a boy.

$x \sim Bin(4, \frac{1}{2})$

$P(x>0) = 1-P(x=0)$

$= 1-[{}^4C_0(\frac{1}{2})^0(\frac{1}{2})^4]$

$= 1-\left(1\times1\times\frac{1}{16}\right)$

$= \frac{15}{16}$

ii. P(at least 1 boy and at least 1 girl) = 1-[P(all boys) or P(all girls)]

$= 1-[P(x=4)+P(x=0)]$

$= 1-{}^4C_4(\frac{1}{2})^4(\frac{1}{2})^0 - {}^4C_0(\frac{1}{2})^0(\frac{1}{2})^4$

$= 1-1\times\frac{1}{16}\times1-1\times1\times\frac{1}{16}$

$= \frac{14}{16}$

$= \frac{7}{8}$

Key properties of binomial distribution

i. Mean (Expectation)

* $E(x) = \mu = np$

ii. Variance

* $Var(x) = \sigma^2 = np(1-p) = npq$; where $q = 1-p$

iii. Standard deviation

* $\sigma = \sqrt{npq}$

In 100 tosses of a fair coin, what is the expected number of heads?

$E(x) = np$

$= 100 \times 0.5$

$= \underline{\underline{50}}$

Ex If the probability of a defective bolt is 0.1, find

i. the mean

ii. the standard deviation

for the number of defective bolts in a total of 400 bolts.

i. The mean $= 400 \times 0.1$

$= \underline{\underline{40}}$

ii. The standard deviation $= \sqrt{400 \times 0.1 \times 0.9}$

$= \sqrt{36}$

$= \underline{\underline{6}}$

Ex The random variable X has a binomial distribution with mean 5.76 and standard deviation 1.92. Find $P(X=6)$.

$X \sim Bin(n, p)$

$np = 5.76$

$np(1-p) = (1.92)^2 = 3.6864$

$\therefore (1-p) = \frac{3.6864}{5.76} = 0.64$

$\therefore p = 1 - 0.64 = 0.36$

$\therefore n = \frac{5.76}{0.36} = 16$

$\therefore X \sim Bin(16, 0.36)$

$P(X=6) = {}^{16}C_6 (0.36)^6 (0.64)^{10}$

$= 8008 \times 2.17678 \times 10^{-3} \times 1.152921 \times 10^{-2}$

$= 0.2009735$

Continuous Random Variables

* A continuous random variable is a random variable that can take any value within an interval (or multiple intervals).

* Unlike discrete random variables, the possible values cannot be put in one-to-one correspondence with counting numbers.

* The fundamental difference from discrete random variables is that continuous variables have uncountably infinite possible values within their domain.

Probability Density Function (PDF)

* The probability properties of a continuous random variable X are described by its probability density function $f(x)$, which must satisfy two essential properties;

    i. Non-negativity: $f(x) \geq 0$ for all $x \in \mathbb{R}$

    ii. Total area property: $\int_{-\infty}^{\infty} f(x)\,dx = 1$

* For a continuous random variable, probabilities are represented by areas under the curve of the PDF.

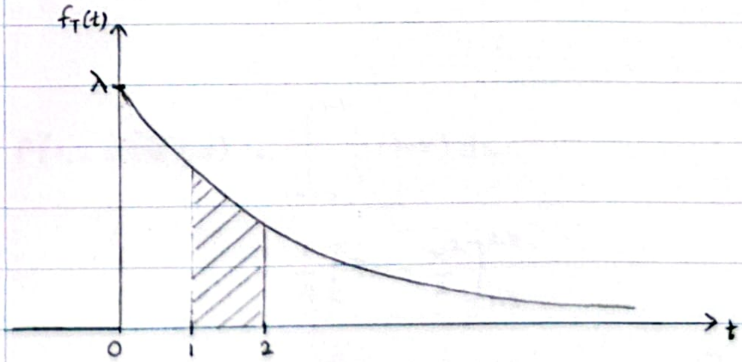
* The probability of the event $\{a < X \leq b\}$ is found from the density function f as follows;

    • $P(a < X \leq b) = \int_a^b f(x)\,dx$

IMPORTANT

* For continuous random variables, $P(X = c) = 0$ for any specific value c. This is why we typically use inequalities like $P(a < X \leq b)$ rather than $P(X = a)$.

$\boxed{Ex}$ Suppose that the random variable T has a density function given by $f_T(t) = 0$ when $t < 0$ and $f_T(t) = \lambda e^{-\lambda t}$ when $t \geq 0$. Show the area corresponding to the $P(1 < T \leq 2)$ in a graph.



* The area corresponding to $P(1 < T \leq 2) = \int_1^2 \lambda e^{-\lambda t}\,dt = e^{-\lambda} - e^{-2\lambda}$ has been shaded in the graph

* This represents an exponential distribution, commonly used to model waiting times or lifespans. The graph shows $P(1 < T \leq 2)$ as the area under the exponential curve between $t = 1$ and $t = 2$.

No :

Date: / /

$\boxed{Ex}$ A continuous random variable has PDF $f(x)$ where $f(x) = kx$, $0 \leq x \leq 4$.

I. Find the value of the constant k.

II. Sketch $y = f(x)$.

III. Find $P(1 \leq x \leq 2.5)$.

I. Since X is a random variable;

$$\int_0^4 f(x)\,dx = 1$$

$$\int_0^4 kx\,dx = 1$$

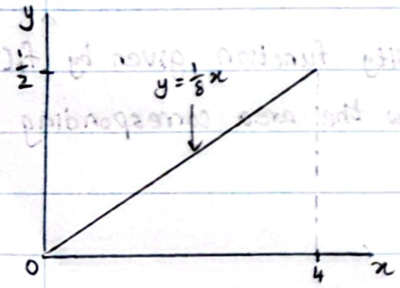
$$k\left[\frac{x^2}{2}\right]_0^4 = 1$$

$$k\left[\frac{16}{2} - \frac{0}{2}\right] = 1$$

$$8k = 1$$

$$k = \frac{1}{8}$$

II. $y = \frac{1}{8}x$; $0 \leq x \leq 4$



III. $P(1 \leq X \leq 2.5) = \int_1^{\frac{5}{2}} \frac{1}{8}x\,dx$

$$= \frac{1}{8}\left[\frac{x^2}{2}\right]_1^{\frac{5}{2}}$$

$$= \frac{1}{16}\left[\frac{25}{4} - 1\right]$$

$$= \frac{1}{16} \times \frac{21}{4}$$

$$= 0.328125$$

Ex The continuous r.v. X has PDF $f_{(x)}$ where $f_{(x)} = k(4-x)$; $1 \leq x \leq 3$.

I. Find the value of constant k.

II. Sketch $y = f_{(x)}$.

III. Find $P(1.2 \leq X \leq 2.4)$.

I. Since X is a random variable;

$$\int_1^3 k(4-x)\,dx = 1$$

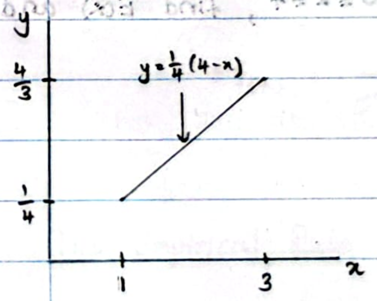
$$k\left[4x - \frac{x^2}{2}\right]_1^3 = 1$$

$$k\left[\frac{15}{2} - \frac{7}{2}\right] = 1$$

$$4k = 1$$

$$k = \frac{1}{4}$$

II. $y = \frac{1}{4}(4-x)$; $1 \leq x \leq 3$



III. $$P(1.2 \leq X \leq 2.4) = \int_{1.2}^{2.4} \frac{1}{4}(4-x)\,dx$$

$$= \frac{1}{4}\left[4x - \frac{x^2}{2}\right]_{1.2}^{2.4}$$

$$= \frac{1}{4}\left[(9.6 - 2.88) - (4.8 - 0.72)\right]$$

$$= \frac{1}{4} \times 2.64$$

$$= 0.66$$

No : Date: / /

Expectation

* If X is a continuous random variable with PDF $f(x)$, then the expectation of X is $E(X)$ where;

$$\boxed{E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx}$$

* $E(x)$ is often denoted by μ and referred to as the mean of X.

* This represents the 'centre of mass' or average value of the distribution.

Variance

* The variance of X is $Var(X)$ where;

$$\boxed{Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2}$$

$$\boxed{Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2}$$

If X is a continuous r.v. with PDF $f(x) = \frac{3x^2}{64}$; $0 \leq x \leq 4$, find $E(x)$ and Var

$$E(x) = \int_0^4 x \cdot \frac{3x^2}{64}\, dx$$

$$= \frac{3}{64}\left[\frac{x^4}{4}\right]_0^4$$

$$= 3$$

$$Var(x) = \int_0^4 x^2\, \frac{3x^2}{64}\, dx - \mu^2$$

$$= \frac{3}{64}\left[\frac{x^5}{5}\right]_0^4 - 9$$

$$= \frac{3}{64} \times \frac{1024}{5} - 9$$

$$= \frac{48}{5} - \frac{45}{5}$$

$$= 0.6$$

No: Date: / /

Normal Distribution

* The normal distribution is the most important continuous distribution in statistics. Many naturally occurring phenomena follow this distribution, including heights, masses and examination results.

Key characteristics

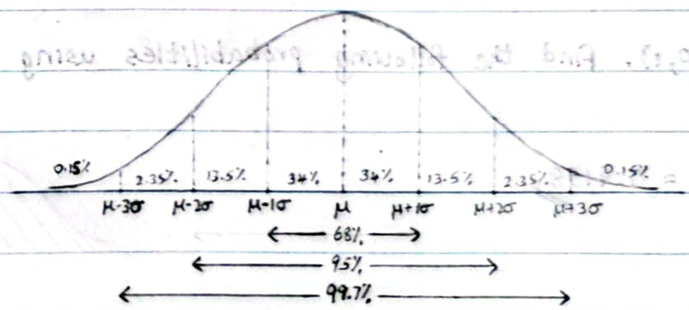
* Bell shaped and symmetrical about the mean (μ)
* Parameters: mean (μ) and variance (σ^2)
* Notation: $X \sim N(\mu, \sigma^2)$
* Expectation: $E(x) = \mu$
* Variance: $Var(x) = \sigma^2$

Probability Density Function

* A continuous random variable x follows a normal distribution with mean μ and variance σ^2 when it has the characteristic bell-shaped probability density function;

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

The Empirical Rule (68-95-99.7 Rule)



* This rule provides a quick way to understand how data is distributed in a normal distribution;

  - Approximately 68% of the distribution lies within $\pm 1\sigma$ of the mean.
  - Approximately 95% of the distribution lies within $\pm 2\sigma$ of the mean.
  - Approximately 99.7% of the distribution lies within $\pm 3\sigma$ of the mean.

* The area under the total curve is 1.

* This rule is particularly useful for quality control and identifying outliers in da

Standard Normal Distribution

* When we standardize a normal variable x with mean μ and variance σ^2 using $z = \frac{x-\mu}{\sigma}$, we get the standard normal distribution.
    - Mean = 0
    - Variance = 1
    - Notation: $Z \sim N(0,1)$

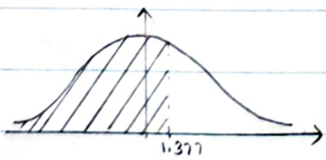
* Standardization allows us to use standard normal tables for any normal distribution problem

* The PDF of the standard normal variable x is denoted by $\phi(x)$ where;
    - $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

* We refer to the tables since it is difficult to evaluate the integral;
    - $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$

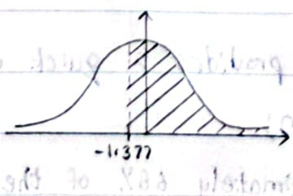
$\boxed{\text{Ex}}$ Let $Z \sim N(0,1)$. Find the following probabilities using standard normal tables.

i. $P(Z<1.377) = 0.9158$



ii. $P(Z>-1.377) = P(Z<1.377)$

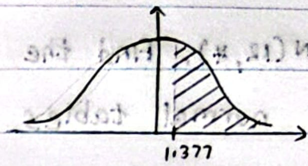
$= 0.9158$



iii. $P(Z > 1.377) = 1 - P(Z < 1.377)$

$= 1 - 0.9158$

$= 0.0842$

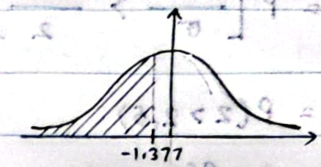


iv. $P(Z < -1.377) = P(Z > 1.377) \mid 1 - P(Z > 1.377)$

$= 1 - P(Z < 1.377)$

$= 1 - 0.9158$

$= 0.0842$

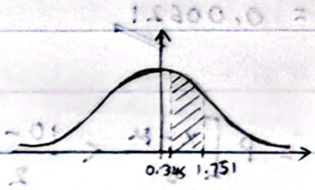


v. $P(0.345 < Z < 1.751)$

$= P(Z < 1.751) - P(Z < 0.345)$

$= 0.9600 - 0.6350$

$= 0.3250$



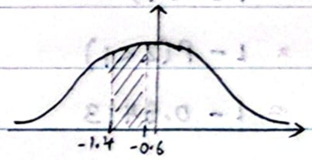
vi. $P(-1.4 < Z < -0.6)$

$= P(Z < -0.6) - P(Z < -1.4)$

$= P(Z < 1.4) - P(Z < 0.6)$

$= 0.9192 - 0.7257$

$= 0.1935$



$\boxed{\text{Ex}}$ Let $X \sim N(300, 25)$. Find $P(X > 305)$.

$P(X > 305) = P\left[\frac{X - \mu}{\sigma} > \frac{305 - 300}{5}\right]$

$= P(Z > 1)$

$= 1 - P(Z < 1)$

$= 1 - 0.8413$

$= 0.1587$

No :   Date: / /

Ex Let $X \sim N(12, 4)$. Find the following probabilities using standardization and standard normal tables.

i. $P(X > 17) = P\left[\frac{X-\mu}{\sigma} > \frac{17-12}{2}\right]$

$= P(Z > 2.5)$

$= 1 - P(Z < 2.5)$

$= 1 - 0.99379$

$= 0.00621$

ii. $P(X < 10) = P\left[\frac{X-\mu}{\sigma} < \frac{10-12}{2}\right]$

$= P(Z < -1)$

$= P(Z > 1)$

$= 1 - P(Z < 1)$

$= 1 - 0.8413$

$= 0.1587$

iii. $P(9 < X < 13) = P\left[\frac{9-12}{2} < \frac{X-\mu}{\sigma} < \frac{13-12}{2}\right]$

$= P(-1.5 < Z < 0.5)$

$= P(Z < 0.5) - P(Z < -1.5)$

$= P(Z < 0.5) - [1 - P(Z < 1.5)]$

$= 0.6915 - 1 + 0.9332$

$= 0.6247$

No : Date: __/__/__

Ex Let $X \sim N(100, 36)$. Find a where $P(X > a) = 0.1093$.

$$P(X > a) = 0.1093$$

$$P\left[Z > \frac{a-100}{6}\right] = 0.1093$$

$$1 - P\left[Z < \frac{a-100}{6}\right] = 0.1093$$

$$P\left[Z < \frac{a-100}{6}\right] = 1 - 0.1093$$

$$P\left[Z < \frac{a-100}{6}\right] = 0.8907$$

$$\frac{a-100}{6} = 1.23$$

$$a - 100 = 7.38$$

$$\underline{\underline{a = 107.38}}$$

Ex Let $X \sim N(\mu, 36)$. Given that $P(X > 82) = 0.0478$, find μ.

$$P(X > 82) = 0.0478$$

$$P\left[\frac{X-\mu}{\sigma} > \frac{82-\mu}{6}\right] = 0.0478$$

$$P\left[Z > \frac{82-\mu}{6}\right] = 0.0478$$

$$1 - P\left[Z < \frac{82-\mu}{6}\right] = 0.0478$$

$$P\left[Z < \frac{82-\mu}{6}\right] = 1 - 0.0478$$

$$P\left[Z < \frac{82-\mu}{6}\right] = 0.9522$$

$$\frac{82-\mu}{6} = 1.667$$

$$82 - \mu = 10.002$$

$$\underline{\underline{\mu = 71.998}}$$

No : ______ Date: __/__/__

Normal Approximation to Binomial Distribution

* As number of trials becomes large, the calculation of binomial probabilities becomes harder. The normal distribution which is continuous and well understood can approximate the binomial distribution in such cases.

* If X is a random variable that follows a binomial distribution with 'n' trial and 'p' probability of success on a given trial, then we can calculate the mean (μ) and standard deviation (σ) of X as follows;
  - $\mu = np$
  - $\sigma = \sqrt{np(1-p)}$

* We can use the normal distribution to approximate the probabilities related to the binomial distribution if n is sufficiently large.
  - If $n > 10$ and p close to $\frac{1}{2}$. OR
  - If $n > 30$ and p moving away from $\frac{1}{2}$

* For n to be 'sufficiently large', it needs to meet the following criteria.
  - $np \geq 5$ AND
  - $n(1-p) \geq 5$

* When both criteria are met, we can use the normal distribution to answer probability questions related to the binomial distribution.
  - If $X \sim Bin(n,p)$, then for large n, $X \sim N(np, npq)$ approximately where $q = 1-p$.

* Since we are approximating a discrete distribution with a continuous one, we need continuity correction (adding or subtracting 0.5 to a discrete x value).
  - $P(X=k)$ becomes $P[(k-0.5) < X < (k+0.5)]$
  - $P(X \leq k)$ becomes $P[X < (k+0.5)]$
  - $P(X<k)$ becomes $P[X < (k-0.5)]$
  - $P(X \geq k)$ becomes $P[X > (k-0.5)]$
  - $P(X>k)$ becomes $P[X > (k+0.5)]$

Ex Find the probability of obtaining between 4 and 7 heads inclusive with 12 tosses of a fair coin;

i. Using the binomial distribution

ii. Using the normal approximation to the binomial distribution.

i. Let X be the random variable 'the number of heads obtained'.

$X \sim Bin(12, 0.5)$

$$P(4 \leq X \leq 7) = P(X=4) + P(X=5) + P(X=6) + P(X=7)$$

$$= {}^{12}C_4\left(\tfrac{1}{2}\right)^4\left(\tfrac{1}{2}\right)^8 + {}^{12}C_5\left(\tfrac{1}{2}\right)^5\left(\tfrac{1}{2}\right)^7 + {}^{12}C_6\left(\tfrac{1}{2}\right)^6\left(\tfrac{1}{2}\right)^6 + {}^{12}C_7\left(\tfrac{1}{2}\right)^7\left(\tfrac{1}{2}\right)^5$$

$$= \left(\frac{1}{2}\right)^{12}\left[\frac{\cancel{12}\times 11\times \cancel{10}^{5}\times 9}{\cancel{4}\times \cancel{3}\times \cancel{2}\times 1} + 2\times\frac{12\times 11\times \cancel{10}\times 9\times 8}{\cancel{5}\times \cancel{4}\times \cancel{3}\times \cancel{2}\times 1} + \frac{\cancel{12}\times 11\times \cancel{10}^{2}\times \cancel{9}^{3}\times \cancel{8}^{2}\times 7}{\cancel{6}\times \cancel{5}\times \cancel{4}\times \cancel{3}\times \cancel{2}\times 1}\right]$$

$$= \frac{495 + 1584 + 924}{4096}$$

$$= 0.733$$

ii. $X \sim N(6, 3)$

$$P(3.5 < X < 7.5) = P(-1.443 < Z < 0.866)$$

$$= P(Z < 0.866) - P(Z < -1.443)$$

$$= 0.8067 - (1 - 0.9255)$$

$$= 0.8067 - 0.0745$$

$$= 0.7322$$

No : Date: / /

Ex It is known that in a sack of mixed grass seeds, 35% are ryegrass. Use the normal approximation to the binomial distribution to find the probability that in a sample of 400 seeds, there are;

i. less than 120 ryegrass seeds

ii. between 120 and 150 ryegrass seeds (inclusive)

iii. more than 160 ryegrass seeds

Let X be the number of ryegrass seeds.

$X \sim Bin(400, 0.35)$ $[np = 140, \; npq = 140 \times 0.65 = 91]$

$X \sim N(140, 91)$

i. $P(X < 119.5) = P(Z < -2.149)$

$= 1 - 0.9842$

$= 0.0158$

ii. $P(119.5 < X < 150.5) = P(-2.149 < Z < 1.101)$

$= P(Z < 1.101) - P(Z < -2.149)$

$= 0.8645 - 0.0158$

$= 0.8487$

iii. $P(X > 160.5) = P(Z > 2.149)$

$= 1 - 0.9842$

$= 0.0158$

Poisson Distribution

* The Poisson probability distribution gives the probability of several events occurring in a fixed interval of time or space if these events happen with a known average rate and independently of the time since the last event.

    - Car accidents on a particular stretch of a road in one day.
    - Accidents in a factory in one week
    - Telephone calls made to a switch board in a given minute.
    - Insurance claims made to a company in a given time.

    } Time

    - Number of typos per page in a book
    - Defective items per square meter of fabric
    - Number of potholes per kilometer on a highway.
    - Trees per hectare in a forest.

    } Space

* A discrete random variable x having the following PDF is said to follow the Poisson distribution.

$$\boxed{P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}} \qquad x = 0, 1, 2, 3, \ldots, \infty \qquad \lambda > 0$$

Properties

* Notation : $X \sim P_o(\lambda)$
* $E(x) = \lambda$
* $Var(x) = \lambda$

Ex If $X \sim Po(2)$, find;

i. $P(X=4)$

ii. $P(X \geq 3)$

i. $P(X=4) = \dfrac{e^{-2} \times 2^4}{4!}$

$= 0.0902$

ii. $P(X \geq 3) = 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left[\dfrac{e^{-2} \times 2^0}{0!} + \dfrac{e^{-2} \times 2^1}{1!} + \dfrac{e^{-2} \times 2^2}{2!}\right]$

$= 0.3233$

Ex The mean number of bacteria per milliliter of a liquid is known to be 4. Assuming that the number of bacteria follows a Poisson distribution, find the probability that in 1ml of liquid, there will be;

i. No bacteria

ii. 4 bacteria

iii. less than 3 bacteria

$X \sim Po(4)$

i. $P(X=0) = \dfrac{e^{-4} \times 4^0}{0!}$

$= 0.0183$

ii. $P(X=4) = \dfrac{e^{-4} \times 4^4}{4!}$

$= 0.1954$

iii. $P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$= \dfrac{e^{-4} \times 4^0}{0!} + \dfrac{e^{-4} \times 4^1}{1!} + \dfrac{e^{-4} \times 4^2}{2!}$

$= 0.2381$

Ex Using the data of previous exercise, find the probability that in 3 ml of liquid, there will be less than 2 bacteria.

Let Y be the number of bacteria in 3ml of liquid.

$Y \sim Po(12)$

$$P(Y<2) = P(Y=0) + P(Y=1)$$

$$= \frac{e^{-12} \times 12^0}{0!} + \frac{e^{-12} \times 12^1}{1!}$$

$$= 7.9875 \times 10^{-5}$$

Normal Approximation to Poisson Distribution

* If $X \sim Po(\lambda)$, for large λ (typically $\lambda \geq 10$), the Poisson distribution can be approximated by $X \sim N(\lambda, \lambda)$ approximately.

* Continuity corrections should be applied here too.

Ex A radioactive disintegration gives counts that follow a Poisson distribution with mean count per second of 25. Find the probability that in 1 second, the count is between 23 and 27 inclusive;

i. Using Poisson distribution

ii. Using normal approximation to the Poisson distribution.

$X \sim Po(25)$

i. $$P(23 \leq X \leq 27) = P(X=23) + P(X=24) + P(X=25) + P(X=26) + P(X=27)$$

$$= \frac{e^{-25} \times 25^{23}}{23!} + \frac{e^{-25} \times 25^{24}}{24!} + \frac{e^{-25} \times 25^{25}}{25!} + \frac{e^{-25} \times 25^{26}}{26!} + \frac{e^{-25} \times 25^{27}}{27!}$$

$$= 0.3827$$

No : Date: / /

ii. Using the normal approximation to Poisson distribution;

$X \sim N(25, 25)$

$P(22.5 < X < 27.5) = P(-0.5 < Z < 0.5)$

$= P(Z < 0.5) - P(Z < -0.5)$

$= 0.6915 - (1 - 0.6915)$

$= 0.6915 - 0.3085$

$= 0.3830$

Sum and Difference of Two Independent Normal Variables

* If X and Y are two independent normal variables such that $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then;

$$\boxed{\begin{array}{l} X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2) \\ X-Y \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2) \end{array}}$$

Ex If $X \sim N(60, 16)$ and $Y \sim N(70, 9)$, find;

i. $P(X+Y < 140)$

ii. $P(120 < X+Y < 135)$

iii. $P(Y-X > 7)$

i. Let $R = X+Y \sim N(130, 25)$.

$P(R < 140) = P(Z < 2)$

$= 0.9772$

ii. $P(120 < R < 135) = P(-2 < Z < 1)$

$= P(Z < 1) - P(Z < -2)$

$= 0.8413 - (1 - 0.9772)$

$= 0.8413 - 0.0228$

$= 0.8185$

No: Date: / /

iii. Let $S = Y - X \sim (10, 25)$

$P(S>7) = P(Z>-0.6)$
$= 0.7257$

Multiples of Normal Variables

* If X is a normal variable such that $X \sim N(\mu, \sigma^2)$, then;

$$\boxed{aX \sim N(a\mu, a^2\sigma^2)}$$

Ex If $X \sim N(50, 25)$, find $P(3X > 160)$.

$3X \sim N(150, 225)$

$P(3X>160) = P(Z>0.667)$
$= 1 - 0.7477$
$= 0.2523$

Ex Each weekday Mr. Jones walks to the local library to read the newspapers. The time he takes to walk to and from the library is a normal variable with mean 15 minutes and standard deviation 2 minutes. The time he spends in the library is a normal variable with mean 25 minutes and standard deviation $\sqrt{12}$ minutes. Find the probability that on a particular day,

i. Mr. Jones is away from the house for more than 45 minutes.
ii. Mr. Jones spends more time travelling than in the library.

Let X be the time he takes to walk to and from the library.
$X \sim N(15, 4)$

Let Y be the time he spends in the library.
$Y \sim N(25, 12)$

No :

Date: / /

i. $X+Y \sim N(40, 16)$

$$P(X+Y>45) = P(Z>1.25)$$
$$= 1-0.8944$$
$$= 0.1056$$

ii. $X-Y \sim N(-10, 16)$

$$P(X>Y) = P(X-Y>0)$$
$$= P(Z>2.5)$$
$$= 1-0.99379$$
$$= 0.00621$$