Curtin College

DIPLOMA OF ENGINEERING

EMTH1019 LASE – WEEK 2

Curtin College

Probability Distributions

Discrete & Continuous Random Variables

Mean & Variance of Random Variables

Probability Functions & Probability Density Functions

KEY CONCEPTS

Discrete random distributions

Continuous random distributions

Normal distributions

Curtin College

RANDOM

Adjective: proceed, made or occurring without definite aim, reason or pattern

Origin: 1275-1325 Middle English or Old French *randon*, derived from *randir – to gallop*

In Statistics we aim to understand the characteristics of a population without measuring every item.

We often study a population by taking a (hopefully) **random** sample.

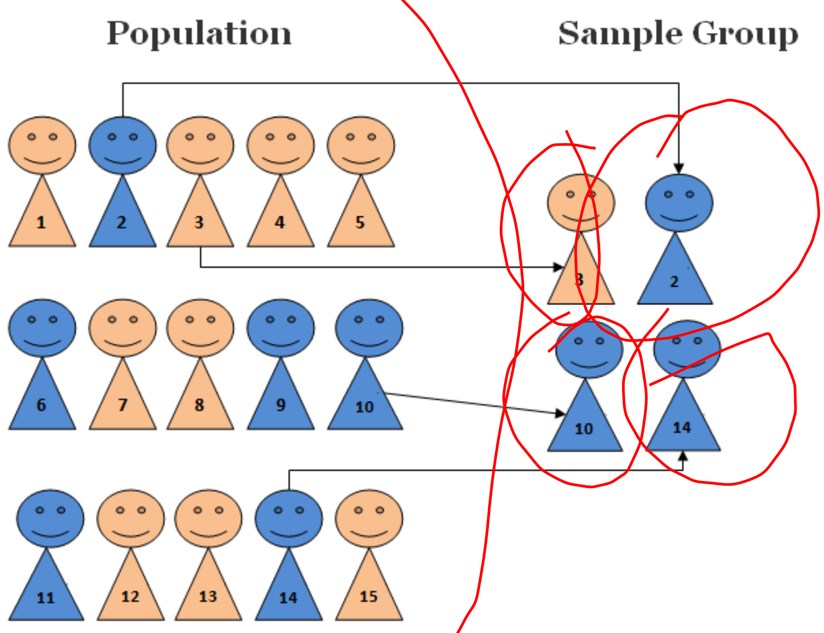
Curtin College

RANDOM SAMPLING

With random sampling each member of the population has an equal chance of being chosen to be in the sample group.

Look at the image.

What are the pros and cons of using sample data versus population data?

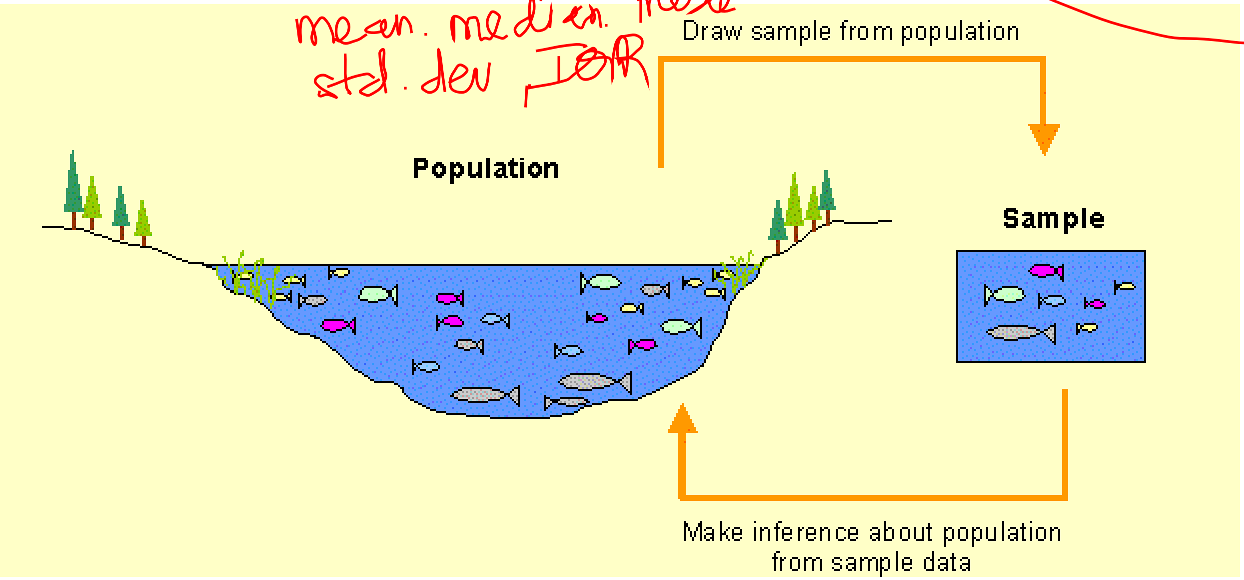


https://www.khanacademy.org/math/math-for-fun-and-glory/math-warmup/random-sample-warmup/p/weak-law-of-large-numbers

STATISTICAL INFERENCE

The process where we use *Sample Statistics* to make inferences about *Populations Parameters*

mean. median. mode
std. dev , IQR



http://labs.geog.uvic.ca/geog226/frLab3.html

Curtin College

RANDOM LANGUAGE

Random experiment

- Is an experiment or a process for which the outcome cannot be predicted with certainty

Outcomes

- Are the result of the experiment

S

1 2 3
4 5 6

Sample space

- Of an experiment is the set of all possible outcomes

Element

- Is what you call each outcome of a sample space

Curtin College

T : tail H : head

S = {Head, Tail}

Examples

Experiment	Outcomes	Sample Space
Toss a coin once	Head, Tail	S = {Head, Tail}
Roll a die once	1, 2, 3, 4, 5, 6	S = {1, 2, 3, 4, 5, 6} any order
Toss a coin twice 2 times	TT, HT, *TH*, HH	S = {*HH*, *HT*, *TH*, *TT*}
Play lottery	Win, Lose	S = {Win, Lose}
Take a test	Pass, Fail	S = {Pass, Fail}
Select a student	Male, Female	S = {Male, Female}

RANDOM VARIABLE

A random variable is the set of all possible values from a *random* experiment

Capital letters

- We use capitals like **X** and **Y**, to avoid confusion with algebra.

$S = \{1, 2, 3, 4, 5, 6\}$

Sample Space

- A random variables set of values is the **Sample Space**

Probability

- P(**X**=value) = probability of that value

$P(X=3)$

Example: Throw a die once

Probability that the outcome = 3

- Random variable **X** = "the score shown on the top face". **X** could be 1,2,3,4,5 or 6.
- Sample Space = {1,2,3,4,5,6}
- $P(\boldsymbol{X} = 3) = 1/6$
- $P(\boldsymbol{X} < 3) = P(X = 1) + P(X = 2) = 1/6 + 1/6 = 1/3$

$P(X > 4) = P(X=5) + P(X=6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

RANDOM VARIABLE – MATHS SPEAK

A random variable is the set of all possible values from a random experiment

If S is a sample space and X is a real valued function *(a function whose values are real) defined over the elements of S, then* X *is called a random variable*

Random variables are denoted by capital letters, X, and their values are denoted by lowercase letters, x.

- x denotes a value of the random variable X
- X= x is interpreted as the set of elements of the sample space for which the random variables take the value of x

EXAMPLE 2 RELAYS IN PARALLEL

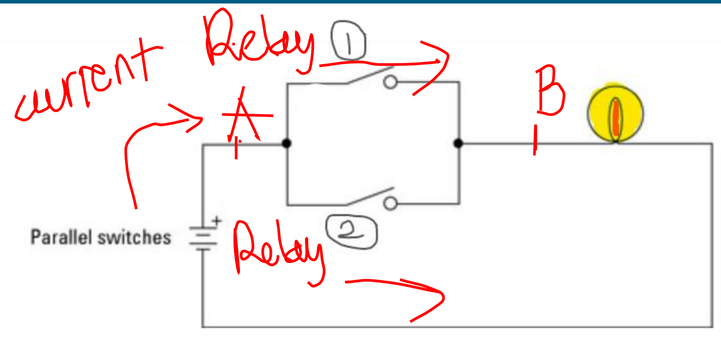
- An electrical circuit has 2 relays operating in parallel numbered 1 & 2
- Current flows when a **switch is thrown** if either **one or both** relays are closed
- The probability of a relay closing properly is 0.8 (same for both relays)
- The relays operate independently
- Let R_i = the outcome that a relay closes properly: R_1 or R_2
- R'_i that it does not close properly. R'_1 or R'_2
- When a switch is thrown an event **X** occurs which is of interest to the operator

Draw the relay and…

- **List all the possibilities of the sample space S= {…..} in terms of R_i and R'_i,**
- **Calculate the probability of each possible outcome**

Curtin College

CONTINUED…



current flows if
Relay ① close, ② open
Relay ② close, ① open
Relay ① and ② close

Solution: Begin by listing the elements of the sample space,

$S = \{R_1 R_2, R_1 R'_2, R'_1 R_2, R'_1 R'_2\}$

$R_1 R_2$: ① and ② close

$R_1 R'_2$: ① close. ② open

$R'_1 R'_2$: ① and ② open

Now construct a table as follows,

$P(\text{close}) = 0.8$

$P(\text{open}) = 1 - 0.8$

$= 0.2$

R_1 and R_2

$= 0.8 \times 0.8$

Sample Space	Probability	Number of relays that close, x
$R_1 R_2$	(0.8)(0.8) = 0.64	2
$R_1 R'_2$	(0.8)(0.2) = 0.16	1
$R'_1 R_2$	(0.2)(0.8) = 0.16	1
$R'_1 R'_2$	(0.2)(0.2) = 0.04	0

$f(x)$

0.64

0.16 + 0.16 = 0.32

0.04

0.64 + 0.16 + 0.16 + 0.04 = 1

TYPES OF RANDOM VARIABLES

Discrete random variable

- You can count the outcomes.
  - The set of possible outcomes is **finite or countable**
- If a coin is thrown 3 times the number of heads, X, can only be 0, 1, 2, 3 so the variable is discrete
  - P(X=0)= 1/2*1/2*1/2=1/8 {TTT}, T and T and T
  - P(X=1)= (1/2*1/2*1/2)*3=3/8 {HTT, THT, TTH}

$P(T) = 0.5 = 1/2$

$P(H) = 0.5 = 1/2$

0 H

TTT

$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$

1H : HTT, THT, TTH

$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$ $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$

Continuous random variable

- You can measure it with a device.
  - Can have **infinitely** many values
  - If we have lots of discrete data we can model it as continuous sometimes
- E.g how tall you are *187.23578453797* cm

$P(X = 2) = \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) \times 3 = \frac{3}{8}$

{HHT, HTH, THH}

DISCRETE PROBABILITY DISTRIBUTIONS

In a discrete probability distribution you make a list of all possible values

- For a die S={1,2,3,4,5,6} but you cannot get 3.46787983214
- All the possible probabilities add up to 1
- Individual probabilities are:
  - Never negative
  - Are always ≥ 0 e.g. P(X=5)=1/6 and P(X=0)=0

Maths speak

- A random variable X is discrete if it can be counted. If X is a discrete random variable then:
  - f(x)=P(X=x) and is called the **probability function (pf)** or **probability mass function (pmf)** of X
  - $f(x) \geq 0$ for all x *Why can't it be negative?*
  - $\sum_x f(x) = 1$ *Why does it have to add up to 1?*

DISCRETE PROBABILITY DISTRIBUTIONS

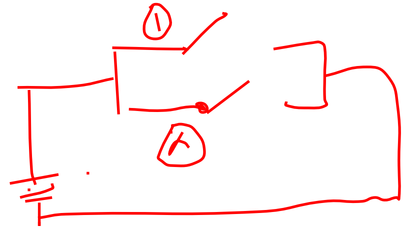
You can express the probabilities associated with the values of a random variable by means of a function f(x)

- f(x)= Probability that X takes on the value x
  - =P(X=x) e.g. P(X=2)

Definition

- A random variable X is discrete if it takes finite or countably infinite values. If X is a discrete random variable then:
  - f(x)=P(X=x) and is called the **probability function (pf)** or **probability mass function (pmf)** of X – it can be an equation or a tale of values
  - f(x)≥ 0 for all x *Why can't it be negative?*
  - $\sum_x f(x) = 1$ *Why does it have to add up to 1?*

sum



USING RELAY EXAMPLE

For the random variable X find the probability function f(x)

Solution: Substituting the values of x = 0, 1, 2 into f(x) we get,

X = no. of relay that is closed

f(0) = P(X = 0) = 0.04
f(1) = P(X = 1) = 0.16+0.16 = 0.32
f(2) = P(X = 2) = 0.64

The probabilities are tabulated below

x	0	1	2
f(x)	0.04	0.32	0.64

THEORY: MEAN (EXPECTED VALUE) & VARIANCE

Definition: If X is a discrete random variable, the mean or EXPECTED VALUE OF X is defined by

$$\mu = E(X) = \sum_x xP(X = x) = \sum_x xf(x)$$

Definition: For a random variable X with mean μ, the variance of X is defined by

$$\sigma^2 = Var(X) = E\left((X - E(X))^2\right) = E\left((X - \mu)^2\right)$$

i.e

$$\sigma^2 = Var(X) = \sum (x - \mu)^2 f(x)$$

Experiment : Throw a die $S = \{1, 2, 3, 4, 5, 6\}$

prob table

X	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

mean = expected value

$= \mu = 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6} = 3.5$

Curtin College

MEAN (EXPECTED VALUE) & VARIANCE

If all outcomes are **equally likely** such as throwing a fair die:

- Expected value= μ = 1*1/6+2*1/6+3*1/6+4*1/6+5*1/6+6*1/6=3.5
- Variance = $Var(X) = \sigma^2 = \sum(x-\mu)^2 f(x) =$
  $$\boxed{(3.5-1)^2 * \frac{1}{6}} + (3.5-2)^2 * \frac{1}{6} + (3.5-3)^2 * \frac{1}{6} + (3.5-4)^2 * \frac{1}{6} + (3.5-5)^2 * \frac{1}{6} (3.5-6)^2 * \frac{1}{6}$$
  $= 2.917$

  mean
- Standard Deviation = Square Root of Variance
- Standard deviation = $\sigma = \sqrt{2.917} = 1.708$

  $= \sqrt{\text{Variance}}$

But what if different outcomes have different probabilities?

Curtin College

DIFFERENT PROBABILITES

Find the mean and variance of the random variable from the relay example

Solution: The probability function, $f(x)$, is given by:

x	0	1	2
f(x)	0.04	0.32	0.64

need to write down the working steps

mean = expected value

$$\mu = E(X) = \sum_{x=0}^{2} xf(x) = 0(0.04) + 1(0.32) + 2(0.64) = 1.6$$

mean

Variance

$$\sigma^2 = Var(X) = \sum_{x=0}^{2} (x-\mu)^2 f(x) = (0-1.6)^2(0.04) + (1-1.6)^2(0.32) + (2-1.6)^2(0.64) = 0.32$$

EASIER WAY OF CALCULATING VARIANCE

I prefer to work with $\boxed{\mathrm{Var(X)} = E(X^2) - (E(X))^2}$

Calculate $E(X)$, $E(X^2)$ and Variance. $\mathrm{Var(X)} = E(X^2) - (E(X))^2$

- Variance is always positive and standard deviation $= \sqrt{Variance}$

1. A car pooling study shows that the number of passengers, X, in a car (excluding the driver) is likely to assume the values 0,1,2,3 and 4 with probabilities given by the table

x	0	1	2	3	4
$P(X = x)$	0.7	0.1	0.1	0.05	0.05

mean $= \mu = E(X)$

$= 0 \times 0.7 + 1 \times 0.1 + 2 \times 0.1 + 3 \times 0.05 + 4 \times 0.05$

$= 0.65$

$E(X^2) = 0^2 \times 0.7 + 1^2 \times 0.1 + 2^2 \times 0.1 + 3^2 \times 0.05 + 4^2 \times 0.05$

$= 1.75$

mean

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

$= 1.75 - 0.65^2$

$= 1.3275$

standard deviation $= \sqrt{1.3275}$

$= 1.15$

Curtin College

MJ AND BASKETBALL

According to WIKI in the 1990-91 season MJ (Michael Jordan) had a 53.9% shooting success

- His probability of making a shot was p=0.539
- The probability MJ does not make the shot is (1-p) = 1-0.539=0.461

prob

What is the chance MJ would make 3 out of the next 4 shots?

300 / 400

There 4 different ways MJ can make 3 out of 4 shots

?

4

✓✓✓✗
✓✓✗✓
✓✗✓✓
✗✓✓✓

hit 3 shots

miss

Each way has a probability of

- 0.539*0.539*0.539*(1-0.539) or Shot*shot*shot*miss

Probability MJ will make 3 of next 4 shots $= 4 * 0.539^3 * (1 - 0.539) = 0.289$

?

Curtin College

CALCULATION

There 4 different ways MJ can make 3 out of 4 shots

✓✓✓✗

✓✓✗✓

✓✗✓✓

✗✓✓✓

Each way has a probability of

- 0.539*0.539*0.539*(1-0.539)
- Shot*shot*shot*miss

Probability MJ will make 3 of next 4 shots =

$$4 * 0.539^3 * (1 - 0.539) = 0.289$$

Curtin College

BINOMIAL DISTRIBUTIONS

Using Combinations to make life easier

$p =$ success
$1-p =$ failure
$n =$ no. of experiment
$x =$ no. of success

Rather than writing out all the possible ways you can write

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

The $\binom{n}{x}$ or ${}^nC_r$ will work out how many different ways there are.

This is a **Binomial Distribution** - $\boldsymbol{Binomial(n, p)}$

Origin: "bi" comes from Latin and means "having two" whilst "nomos" is Greek and means "part" or "portion"

2

So for our question we write:

The probability that MJ will 3 of the next 4 shots is

$$\binom{4}{3} 0.539^3 (1-0.539)^{4-3} = 0.289$$

$\binom{4}{3} 0.539^3 (1-0.539)^1$

Calculator $\boxed{nCr}$

$\binom{4}{3} = 4C3$
$= 4$

Curtin College

Determine the probabilities that MJ makes 0, 1, 2, 3 or 4 of the next 4 shots.

x

$n=4$

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

$\binom{4}{0}=4C0=1$

$$P(X=0)=\binom{4}{0}0.539^0(1-0.539)^{4-0}= 0.045$$

$\binom{4}{1}=4C1=4$

$$P(X=1)=\binom{4}{1}0.539^1(1-0.539)^{4-1}= 0.21$$

$\binom{4}{2}=4C2=6$

$$P(X=2)=\binom{4}{2}0.539^2(1-0.539)^{4-2}=0.37$$

A BIGGER PROBLEM

If I now ask….

$x=20$ $n=30$

What is the probability that MJ will make 20 of the next 30 shots

YOU CAN NOT WRITE EVERY COMBINATIONS AS THERE ARE 30,045,015 WAYS MJ COULD ACHIEVE THIS

You do write

need to write the formula

$$\binom{30}{20} 0.539^{20}(1-0.539)^{30-20} = 0.0558$$

30C20

INDIVIDUAL VERSUS CUMULATIVE PROBABILITY

$x = 20, 21, 22, \dots, 30$

MJ makes at least 20 of the next 30 shots

This is a ***cumulative* probability** as we are calculating $P(20 \le X \le 30)$.

We need to calculate $P(X = 20) + P(X = 21) + P(X = 22) + \cdots P(X = 29) + P(X = 30)$. A tedious but possible calculation.

$$\binom{30}{20} 0.539^{20}(1-0.539)^{30-20} + \binom{30}{21} 0.539^{21}(1-0.539)^{30-21} + \binom{30}{22} 0.539^{22}(1-0.539)^{30-22} +$$

$$\dots\binom{30}{29} 0.539^{29}(1-0.539)^{30-29} + \binom{30}{30} 0.539^{30}(1-0.539)^{30-30} \approx 0.1105 \approx 11\%$$

$x = 30$

An ***individual* discrete probability** would be $\boxed{P(X = 20)}$

CUMULATIVE PROBABILITY

$S = \{0, 1, 2, 3, \ldots, 30\}$

MJ made at least 1 of the next 30 shots $x = 1, 2, 3, 4, 5, \ldots, 30$

$P(X=1) + P(X=2) + \ldots + P(X=30)$

This is cumulative probability as we are calculating $P(X \geq 1)$.

Cumulative probability questions have inequalities.

It might seem you need to work this out for every different variation 1, 2, 3,…29, 30 but you don't.

You only need to exclude $P(X = 0)$ so $P(X \geq 1) = 1 - P(X = 0)$

$$P(X = 0) = \binom{4}{0} 0.539^0 (1 - 0.539)^{4-0} = 8.14 * 10^{-11}$$

The probability Michael Jordan will make at least 1 of the next 30 shots is:

$$\boxed{P(X \geq 1) = 1 - P(X = 0)}$$ Sample space $-$ $x=0$

$$= 1 - 8.14 * 10^{-11} = 1 - 0.0000000000814 \approx 1$$

BERNOULLI AND BINOMIAL

People get these words mixed up.

In the Michael Jordan question the probability of MJ making a shot is the Bernoulli variable. **p=0.539.**

The Binomial part uses that initial probability to work out the probability of making 3 of the next 30 shots or at least 1 of the next 30 shots.

The next few slides present some of the theory behind this and as always some of the jargon.

CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF X

Often we want to know $P(3 \leq X \leq 5)$ or $P(X \leq 4)$ or $P(X < 7)$

The ***cdf*** of a random variable X for any real number x is defined by

$$P(X \leq x)$$

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

What does this mean? Add up all the probabilities up to and including x

Consider the relay question…again. Obtain the ***cdf*** of X

- $F(0) = P(X \leq 0)$ =0.04,
- $F(1) = P(X \leq 1) = f(0) + f(1) = 0.04 + 0.32 = 0.36$
- $F(2) = P(X \leq 2) = f(0) + f(1) + f(2) = 0.04 + 0.32 + 0.64 = 1$

Answer

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.04 & 0 \leq x < 1 \\ 0.36 & 1 \leq x < 2 \\ 1 & x \leq 2 \end{cases}$$

Curtin College

BERNOULLI DISTRIBUTIONS

A Bernoulli Trial is an experiment in which there are only 2 possible outcomes, success or failure. If *P(success)=p*, then P(failure)=1-p

A **Bernoulli Random Variable** measures the outcome of a Bernoulli Trial as 1 or 0

- X=1 if it is a success and X=0 if it is a failure

The probability function of $f(x) = p^x(1-p)^{1-x}$

- Or as a table

x	*0*	*1*
f(x)	*1-p*	*p*

Distribution shorthand… $\boldsymbol{X \sim Bernoulli(p)}$

- If P(success)=p=0.6 then we write $\boxed{\boldsymbol{X \sim Bernoulli(0.6)}}$ and that tells us everything we need to know!

success rate

MEAN & VARIANCE OF THE BERNOULLI RANDOM VARIABLE X

x	0	1
f(x)	1-p	p

The mean or expected value of X

- $E(X) = 0(1-p) + 1p = p$

The Expected value of X^2 $E\left(X^2\right) = 0^2(1-p)+1^2 p = p$

The Variance of X $Var\left(X\right) = p - p^2 = p(1-p)$

A Bernoulli Process that has n trials

- Is a Bernoulli experiment repeated n times
- Has the same probability of success, p, from trial to trial
- Has each trial is independent of the other trials

BINOMIAL RANDOM VARIABLE

When there are $\boxed{\text{n trials}}$….

Binomial Random Variable:

X : number of "successes" in a Bernoulli process of *n* trials

Probability function of the Binomial random variable *X*:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

no. of trials

success rate

In this case we write $\boxed{\boldsymbol{X \sim Binomial(n, p)}}$

The values for pf and for cdf (cumulative) are listed in statistical tables

Curtin College

NOTATION – WE USE COMBINATIONS

Permutations

$^{n}P_{r} = \frac{n!}{(n-r)!} = P(n,r)$

Combinations

$^{n}C_{r} = \frac{n!}{r!(n-r)!} = C(n,r) = \binom{n}{r}$

$\binom{n}{r}$ nCr

Curtin College

BUGS

A certain program has 12 subroutines

When the program is run there is a 10% chance that each subroutine has a bug, which is independent *(does not depend)* of bugs in the other subroutines.

- What information is there? 12 and 10%
- What are they?
- What sort of distribution is this?

Questions

one trial → Bernoulli

1. Probability function of bugs in a single subroutine? *(HINT: what is the probability of a bug in a single sub routine?)*
2. Mean and variance of the number of bugs in ***a*** subroutine? *(Hint: there are formulae)*
3. Probability function of total number of bugs in program *(Hint: remember n=12 so this must be Binomial)*
4. Probability that there are, at most, 6 bugs when the program is run? *(Hint: Binomial and $P(X \le 6) = \cdots$)*
5. Mean number of bugs in the program when it is run once. *(Hint: $E(Y) = np$)*

Curtin College

SOLUTION

1. Probability function of bugs in a single subroutine, e.g. in the 1st, 2nd or… subroutine. We could call that event X_1?
   - If success is having a bug then p=0.1, $\boldsymbol{X \sim Bernoulli(0.1)}$
2. Mean and variance of the number of bugs in ***a*** subroutine?
   - $\boxed{E(X_i)=p \text{=0.1, } Var(X)=p(1-p)}$=0.1(1-0.1)=0.1*0.9=0.09
3. Probability function of total number of bugs in program, we can call that Y
   - $Y \sim Binomial(n = 12, p = 0.1)$
   - $f(y) = P(Y = y) = \binom{12}{y} 0.1^y 0.9^{12-y}, y = 0,1,2,3 \ldots\ldots 12$

max = 6

4. Probability that there are, at most, 6 bugs when the program is run?
   - $P(Y \le 6) = P(Y = 0) + P(Y = 1) + P(Y = 2) + \cdots P(Y = 6) = 0.9999$
5. Mean number of bugs in the program when it is run once.
   - Mean number of bugs $\boxed{E(Y) = np} = 12(0.1) = 1.2$

$\binom{12}{0} 0.1^0 \times 0.9^{12} + \binom{12}{1} 0.1^1 \times 0.9^{11} + \binom{12}{2} 0.1^2 \times 0.9^{10} + \ldots + \binom{12}{6} 0.1^6 \times 0.9^6$

$=$

EXAMPLE COIN TOSSED 10 TIMES

A fair coin is tossed 10 times.

LET X denote the number of heads obtained.

Determine the following:

a) $P(X<3)$

b) $P(X\geq3)$

c) $P(3<X<8)$

- Important: if $X<3$ it does not include 3. You have to pay attention to the inequalities.

SOLUTION

If a head is a success, then p=0.5, n=10, each toss is independent

- X~Binomial(n=10, p=0.5)
- $f(x) = P(X = x) = \binom{10}{x} 0.5^{x} 0.5^{10-x}, y = 0,1,2,3 \ldots\ldots 10$

a) P(X<3)=P(X=0)+P(X=1)+P(X=2)=($P(X \leq 2) = 0.0547$

   a) In the old days you might have looked up a table for ($P(X \leq 2)$ *but now we use computers.*

b) P(X ≥ 3)= $\mathbf{1 - P(X < 3)} = 1 - 0.0547 = 0.9453$

be very careful about the inequalities.

c) *P(3<X<8)= P(X=4)+P(X=5)+P(X=6)+P(X=7)*
   = *0.7734*
   A tedious calculation to do by hand as you need to do the Binomial calculation for each x and then add those probabilities together.

Curtin College

PRUSSIAN OFFICERS KILLED BY HORSEKICK

In the 19th century the Prussian military became concerned about the number of its cavalry officers killed by horse kick during peacetime.

The Prussians had the foresight to collect hard data over 20 years on death by horse kick.

It would have been easy to blame these deaths on what are described as "never events", which are adverse, serious and largely preventable events e.g. blaming the victims for being careless, horses for being vicious or poor leadership and systems.

Instead, the Mathematician Ladislaus Bortkiewicz analyzed the data. In his book "The Law of Small Numbers" (1908) he showed that these fatalities were rare events (low frequency) in a large population and closely followed a Poisson distribution.

POISSON DISTRIBUTION

Named after Simeon Denis Poisson in 1837. It is a discrete probability distribution that gives the probability of a *given number of events occurring in a fixed interval*

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

λ= mean number of occurrences & x=0,1,2…

Applications

- Rare diseases (not infectious diseases because they are not independent)
- Car accidents, traffic flow and ideal gap distance
- Failure of a machine in 1 month

A Poisson Random variable must satisfy the following conditions:

- The number of successes in 2 separate times or places is independent
- The probability of success in a small time interval is PROPORTIONAL TO the entire length of the time interval
- You can replace the time interval with regions of space e.g. number of mice in a field

$X \sim \text{Poisson}(\lambda)$

X~POISSON(λ)

Definition of a Poisson Random Variable

- X= number of certain events occurring in a time interval or region

Probability function of a Poisson Distribution

- $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ where λ= mean number of occurrences and x=0,1,2….
- Sometimes you see μ used instead of λ.

Theorem

- If X~Poisson(λ) then $E(X) = \lambda \; and \; Var(X) = \lambda$ *(how nice is that!)*

Note:

- *Be careful how information is provided. Is the number of events per hour, per day you need to be consistent. Look at the life insurance example.*
- *Maths speak: If the number of events occurring in a time interval or region has Poisson(α) $where \; \alpha = \lambda t$ then $E(X) = \alpha = \lambda t = Var(X)$.*

7

Curtin College

CLASSIC POISSON QUESTION

The mean number of kangaroos/acre on a farm is estimated to be 7.
Assuming a Poisson Distribution determine the probability that there are:

a) 2 kangaroos on an acre.

b) Less than 2 kangaroos on an acre. Poisson

c) Less than 2 kangaroos on 1 of the next 3 acres

binomial with success rate from part (b)

e ! calculator

$n = 3$

e = natural number = 2.718281828

$2! = 2$ factorial $= 2 \times 1$

$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$

SOLUTION

X~Poisson ($\lambda = 7$)

a) $\boxed{P(X=2) = \frac{e^{-\lambda}\lambda^x}{x!}} = \frac{e^{-7}7^2}{2!} = 0.0223$

b) $P(X<2) = P(X=0) + P(X=1) = \frac{e^{-7}7^0}{0!} + \frac{e^{-7}7^1}{1!} = 0.00091 + 0.00638 \approx 0.0073$

c) 1 of the next 3 tells you this is binomial use $P(X=1) = \binom{n}{x} p^x (1-p)^{n-x}$

- $\boxed{p = 0.0073}$ from above, $\boxed{n = 3 \text{ and } x = 1}$
- $\binom{3}{1} 0.0073^1 (1-0.0073)^{3-1} = 0.0216$

(b)

Remember that the meaning of $P(X=x)$ depends on the word question and type of distribution.

Curtin College

POISSON AND LIFE INSURANCE

$\mu = 3$ per week

A life insurance salesman sells on average 3 policies per week.

Assuming a Poisson Distribution calculate the probability that in a week he will sell:

- Some policies $P(X > 0)$
- 2 or more policies but less than 5 $P(2 \leq X < 5)$
- Assuming there are 5 working days in a week what is the probability that on a given day he will sell one policy?

per day $\Rightarrow$ $\boxed{\mu = \frac{3}{5} \text{ per day}}$

https://www.intmath.com/counting-probability/13-poisson-probability-distribution.php

Here, $\mu = 3$ $X \sim \text{Poisson}(\mu = 3)$

College

(a) "Some policies" means "1 or more policies". We can work this out by finding 1 minus the "zero policies" probability: $X > 0$: $X = 1, 2, 3, 4, 5,$

$$P(X > 0) = 1 - \cancel{P(x_0)} \quad P(X=0)$$

Now $\boxed{P(X) = \frac{e^{-\mu}\mu^x}{x!}}$ so $P(\cancel{x_0}) = \frac{e^{-3}3^0}{0!} = 4.9787 \times 10^{-2}$ $X=0$

Therefore the probability of 1 or more policies is given by:

$$\begin{aligned} \text{Probability} &= P(X \geq 0) \\ &= 1 - \cancel{P(x_0)} \quad P(X=0) \\ &= 1 - 4.9787 \times 10^{-2} \\ &= 0.95021 \end{aligned}$$

$\mu = 3$ per week (5 days)

$\mu = \frac{3}{5}$ per day

(b) The probability of selling 2 or more, but less than 5 policies is:

$$\begin{aligned} &P(2 \leq X < 5) \\ &= \cancel{P(x_2) + P(x_3) + P(x_4)} \quad P(X=2) + P(X=3) + P(X=4) \\ &= \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} + \frac{e^{-3}3^4}{4!} \\ &= 0.61611 \end{aligned}$$

need to calculate μ per day

(c) Average number of policies sold per day: $\frac{3}{5} = 0.6$

So on a given day, $P(X) = \frac{e^{-0.6}(0.6)^1}{1!} = 0.32929$

Curtin College

CONTINUOUS PROBABILITY DISTRIBUTIONS

A random variable X is continuous if P(X=x)=0 for all real numbers x

The diameters of a rod can be anywhere from 12 to 15 mm.

What is the probability that a rod selected at random has a diameter of EXACTLY 13 mm? 13.0000000000000000 (forever) mm

The Probability the diameter is exactly 13 mm is ZERO

For continuous random variables we deal with intervals.

$P(X=13)=0$

$$P(X < a) = P(X \leq a)$$

$$P(a < X < b) = P(a \leq X \leq b)$$

NOTE: For discrete random variables the above probability statements are not true.

SOME THEORY

- To calculate the probability we need a probability density function f(x), where the area under f(x) between a and b equals the probability

$$P(a < X < b) = \int_a^b f(x)dx \ \ for\ a < b$$

- The total probability = 1 (cannot be more than 1)

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

- Probability density function is always positive

$$f(x) \geq 0\ for\ all\ x$$

- You can see from Calculus that we cannot get a probability for **continuous** probability density functions where the boundaries are $a = b$, as that integral would be zero.

THEORY - MEAN & VARIANCE

MEAN (μ) OR EXPECTED VALUE $E(X)$

- *If X is a continuous random variable with pdf f(x), the mean or expected value is*

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

VARIANCE OF X

- *For a random variable X with mean μ, the variance of X is*

$$\sigma^2 = Var(X) = E\left(\left(X - E(X)\right)^2\right) = E\left((X-\mu)^2\right)$$

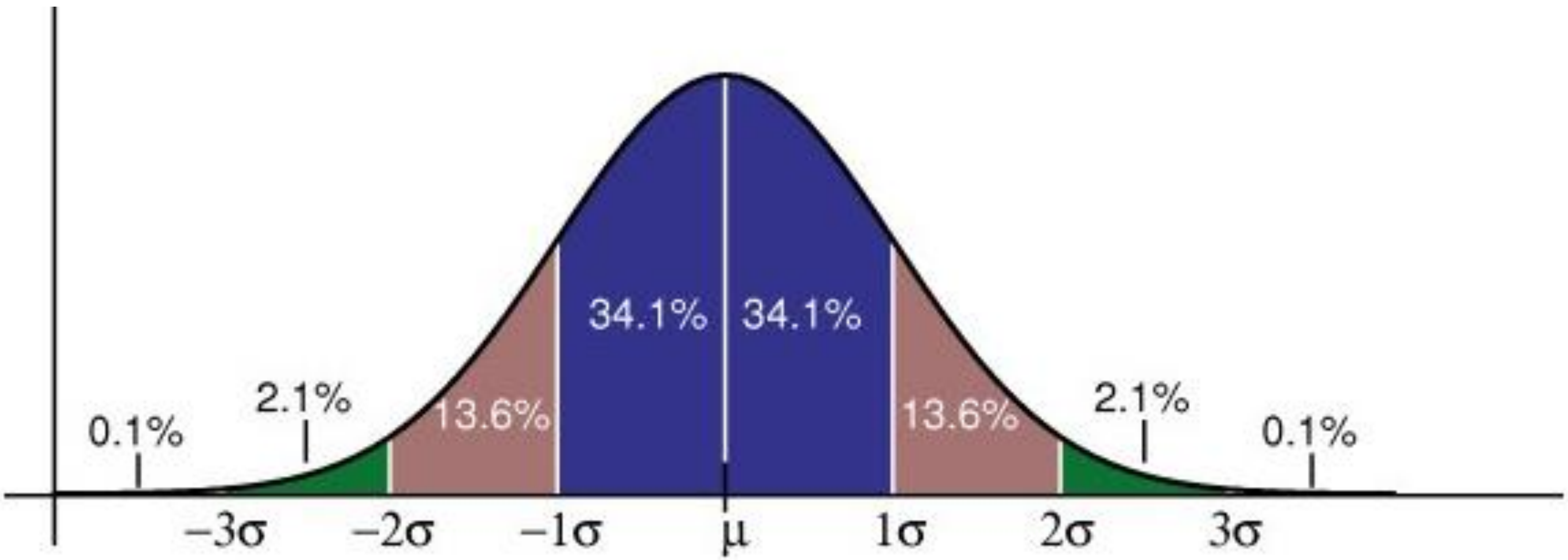
$$\boldsymbol{\sigma^2 = Var(X) = E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx}$$

DO NOT PANIC

We do not do any calculus in this unit so you do not have to be able to calculate these integrals.

We will be using normal distribution tables.

NORMAL DISTRIBUTIONS $X \sim N(\mu, \sigma^2)$



http://www.statisticshowto.com/probability-and-statistics/normal-distributions/

A LITTLE NORMAL HISTORY

Discovered by Abraham de Moivre (1667-1754)

Early applications included: Astronomy, Pierre-Simon Laplace (1749-1827) and Physics, Carl Friedrich Gauss (1777-1855)

A random variable X is said to have a normal distribution if the pdf of X has the form

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ if} -\infty < x < \infty$$

IMPORTANT: μ and σ are constants and the distribution is symmetrical.

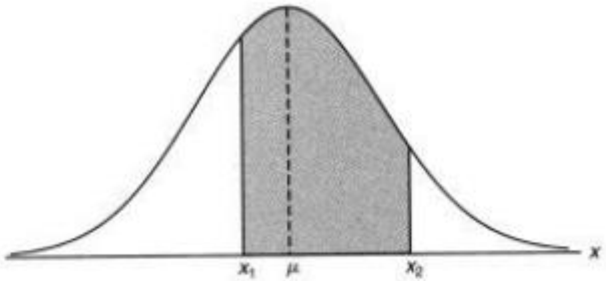
NORMAL DISTRIBUTION THEORY

The pdf of a normal distribution is a difficult function to deal with.

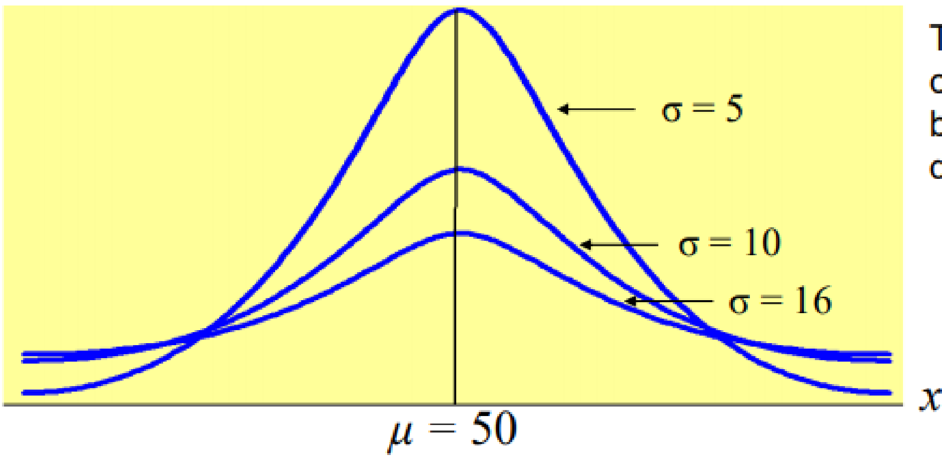
$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \text{ if } -\infty < x < \infty$$

And then to calculate the probability, then is an even *uglier* integral to solve.

$$P(x_1 < X < x_2)=\int_{x_1}^{x_2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

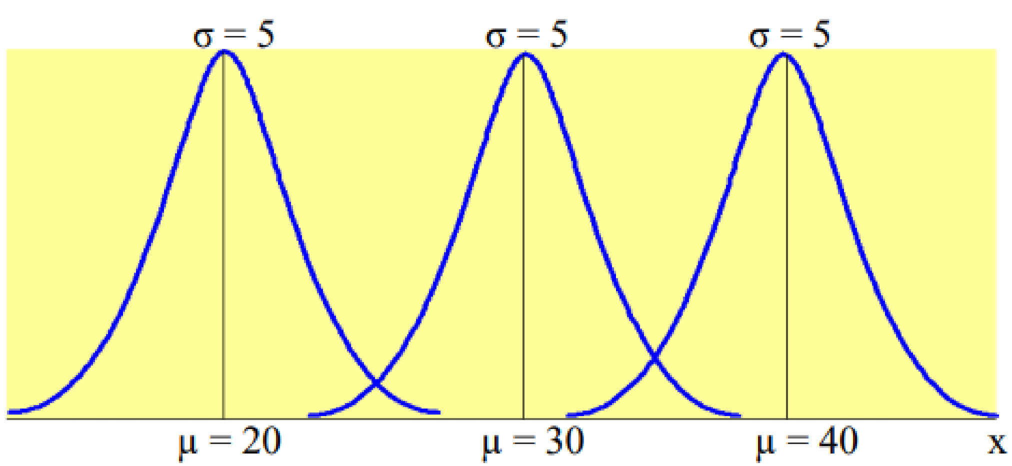


NORMAL DISTRIBUTIONS



Three Normal Distribution curves with the same mean but different standard deviation

Three Normal Distribution curves with different means but same standard deviation



Curtin College

STANDARD NORMAL DISTRIBUTIONS

To get arounds the difficulties of dealing with the equations…

Have tables of values of Standardised Normal Distributions (and Excel!)

When we standardise a normal distribution

- We shift the f(x) so that the $\mu = 0$
- We create a standardised value for the σ which we call z

$$\boldsymbol{z = \frac{(X - \mu)}{\sigma}}$$

This is a VERY IMPORTANT EQUATION

- We started with $\boxed{X \sim N(\mu, \sigma^2)}$, and then standardised using $\boldsymbol{z = \frac{(X-\mu)}{\sigma}}$
- We now have $\boxed{Z \sim N(0,1)}$. The mean is zero and the variance= 1

standard normal distribution

CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION

$P(Z \le z)$ where $Z \sim N(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
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

USING CUMULATIVE NORMAL DISTRIBUTION TABLES

Supose Z has standard normal distribution.

- What is the value of μ and σ?
- P(Z<1.9)
- P(Z>1.9)
- P(-1<Z<1)
- P(-2.1<Z<4.2)

SOLUTION

Solution: $Z \sim N(0, 1)$

(a). $P(Z < 1.9) = 0.9713$

(b). $P(Z > -1.9) = 1 - P(Z \leq -1.9) = 1 - 0.0287 = 0.9713$

or use the fact that the normal curve is symmetric $P(Z < -k) = P(Z > k)$

(c). $P(-1 < Z < 1) = 1 - 2P(Z < -1)$ (by symmetry)

$= 1 - 2(0.1587) = 1 - 0.3174 = 0.6826$

or

$= P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826$

(d). $P(-2.1 < Z < 4.2) = P(Z < 4.2) - P(Z < -2.1) = 1 - 0.0179 = 0.9821$

Curtin College

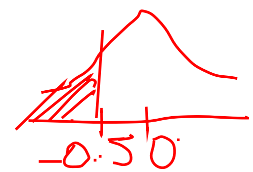
PROBLEM

$X \sim N(\mu = 10, \sigma^2)$. $\Rightarrow Z \sim N(0, 1)$

X has a normal distribution with mean 10 and variance 16

What is:

- The standard deviation $\sigma = \sqrt{16} = 4$
- P(X<8) $= P\left(Z < \frac{8-10}{4}\right) = P(Z < -0.5) = 0.3085$
- P(X>6) $= P\left(Z > \frac{6-10}{4}\right) = P(Z > -1) = 1 - 0.1587$
- P(4<X<12) $= P\left(\frac{4-10}{4} < Z < \frac{12-10}{4}\right) = P(-1.5 < Z < 0.5) = 0.6915 - 0.0668$
- Remember to create your z for each part
- Useful to draw the problem each time
- 4, 0.3085, 0.8413, 0.6247



$P(Z \le z)$ where $Z \sim N(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
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

SOLUTION

Solution: $X \sim N(10, 16) \Rightarrow Z = (X - 10)/4 \sim N(0, 1)$

(a). $P(X < 8) = P(Z < (8 - 10)/4) = P(Z < -0.5) = 0.3085$

(b). $P(X > 6) = P(Z > (6 - 10)/4) = P(Z > -1) = 1 - P(Z \leq -1)$
$= 1 - 0.1587 = 0.8413$

(c). $P(4 < X < 12) = P((4 - 10)/4 < Z < (12 - 10)4) = P(-1.5 < Z < 0.5)$
$= P(Z < 0.5) - P(Z < -1.5) = 0.6915 - 0.0668$
$= 0.6247$

Curtin College

TYPICAL NORMAL DISTRIBUTION QUESTION

With an eye toward improving performance, industrial engineers studied the ability of scanners to read the bar codes of various food and household products. The maximum reduction in power, just before the scanner cannot read the bar code at a fixed distance, is called the maximum attenuation. This quantity, measured in decibels, varies from product to product. After collecting considerable data, the engineers decided to model the variation in maximum attenuation as a normal distribution with mean 10.1 dB and standard deviation 2.7 dB

$\mu = 10.1$ $\sigma = 2.7$

$X \rightarrow Z$ (standard normal distribution)

a) For the next food or product, what is the probability that its maximum attenuation is between 8.5 dB and 13.0 dB?

$P(8.5 < X < 13)$

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

b) What proportion of the products have maximum attenuation greater than 15.1 dB?

$P(X > 15.1)$

Curtin College

SOLUTION

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

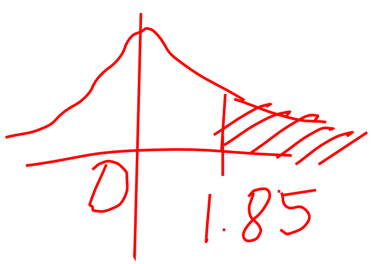
a)

- For x = 8.5: $z = \frac{8.5 - 10.1}{2.7} = -0.59$
- For x = 13.0: $z = \frac{13.0 - 10.1}{2.7} = 1.07$

from the table $z = 1.07$

$z = -0.59$

- $P\,(8.5 \le X \le 13.0)$
  $= P\,(-0.59 \le Z \le 1.07) = 0.8577 - 0.2776$
  $= 0.5801$



b)

- For x = 15.1: $z = \frac{15.1 - 10.1}{2.7} = 1.85$
- $P\,(X > 15.1) = P(Z > 1.85)$
  $= 1 - P(Z < 1.85)$
  $= 1 - 0.9678$
  $= 0.0322$

must change

$X \rightarrow Z$

Curtin College

WORKSHOP QUESTION #15 – WORKING BACKWARDS

15. Given a standard normal distribution, find the value of k such that

(a) $P(Z < k) = 0.0427$ $\quad k = -1.72$ (from z table

(b) $P(Z > k) = 0.2946$ $\quad \Rightarrow P(Z<k) = 1 - 0.2946 = 0.7054$

$\Rightarrow k = 0.54$

(c) $P(-0.93 < Z < k) = 0.7235$

$= P(Z<k) - P(Z<-0.93)$

$= \; ? \; - 0.1762$

$\Rightarrow P(Z<k) = 0.1762 + 0.7235$

$= 0.8997$

$\Rightarrow k = 1.28$

Curtin College

LIMITATIONS OF NORMAL DISTRIBUTIONS

Using this basic normal distribution assumes that you know the mean and variance of the whole population; this can be difficult to know as…

1. **Is all data normally distributed?**
2. **What information do you need to calculate the mean and variance of a large population?**
3. **How much time and $$$$ do you have?**
4. **Think of an organisation that specialises in collecting and using enormous amounts of data.**

THE WORLD IS NOT ALWAYS NORMAL

The trouble with normal distributions is that it sounds so NORMAL

It is nice to think that “all things” has some beautiful inherent symmetry and that if you have enough data it is OK to assume NORMALITY

WRONG

Data is often skewed (not symmetrical) or is symmetrical but not normally distributed.

MJ made the mistake of assuming normality when she had n=500 bits of data. When she graphed the data…….it didn’t even look bell shaped, let alone symmetrical and when tested for normality it wasn’t. *A week wasted…*

So where do we use normal distributions? NEXT WEEK.

Curtin College

HOMEWORK BERNOULLI AND BINOMIAL

The probability of Yasir scoring a goal at football is 27%.

1. What sort of variable is 27%? $X \sim \text{Bernoulli}(p=0.27)$
2. What is the probability he will miss the goal? miss prob $= 1-0.27$
3. If Yasir has 9 shots at goal how many goals would he expect to score? $9 \times 0.27 = 2.43$
4. What is the probability he will score
   a) 4 of the next 9 shots
   b) None of the next 9 shots
   c) Less than 1 shot
   d) 1 of the next 9 shots
   e) At least 1 of the next 9 shots
   f) At least 2 of the next 9 shots
   g) No more than 8 of the next 9 shots

4(a) $P(Y=4) = \binom{9}{4}(0.27)^4 \times (0.73)^5$
$= 0.1388$

(b) $P(Y=0) = \binom{9}{0}(0.27)^0 \times (0.73)^9 = 0.0589$

(c) $P(Y<1) = P(Y=0) = 0.0589$

(d) $P(Y=1) = \binom{9}{1}(0.27)^1 \times (0.73)^8 = 0.1960$

(e) $P(Y \geq 1) = P(Y=1,2,3,\ldots,9)$
$= 1 - P(Y=0) = 1-(b)$

(f) $P(Y \geq 2) = P(Y=2,3,4,\ldots,9)$
$= 1 - P(Y=0) - P(Y=1)$
$= 1 - (b) - (d)$

(g) $P(Y \leq 8) = P(Y=0,1,2,\ldots,7,8)$
$= 1 - P(Y=9)$

Curtin College

ANSWERS BERNOULLI AND BINOMIAL

If the probability of Yasir scoring a goal at football is 27%,		
1	What sort of variable is 27%?	Bernoulli(0.27)
2	What is the probability he will miss?	0.73
3	How many goals would he expect to score after 9 attempts?	2.43
4	What is the probability he will score:	
a	4 of the next 9 shots	0.1388
b	none of the next 9 shots	0.0589
c	less than 1 shot	0.0589
d	1 of the next 9 shots	0.1960
e	at least 1 of the next 9 shots	0.9411
f	at least 2 of the next 9 shots	0.7452
g	no more than 8 of the 9 shots	0.999992

Curtin College

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

HOMEWORK - POISSON AND BINOMIAL

$\mu = 6$

1. On average there are 6 accidents per week on Manning Rd. Assuming a Poisson Distribution determine the probability that in a ***week*** there are:
   a) No accidents $P(X=0)$
   b) Some accidents means $X \geq 1$, $P(X \geq 1) = 1 - P(X=0)$
   c) ~~3~~ 5 accidents $P(X=5) = \frac{e^{-6} \times 6^5}{5!} = 0.1606$
   d) More than 3 accidents but less than 6 $P(3 < X < 6) = P(X=4) + P(X=5) =$
   e) 5 accidents per week in 6 of the next 10 weeks tell you this is a binomial distribution $Y \sim \text{Binomial}(n=10, p=0.1606)$
   f) What is the average number of accidents per day?

2. Still assuming a Poisson Distribution determine the probability that on a given ***day*** there are:
   a) No accidents
   b) Some accidents
   c) 5 accidents
   d) More than 3 accidents but less than 6
   e) No accidents in 6 of the next 10 days

1(e) $P(Y=6) = \binom{10}{6}(0.1606)^6(1-0.1606)^4$

$= 0.0018$

1(f) per day $= 6/7$

2. Recalculate mean per day $= 6/7$

Curtin College

ANSWERS POISSON AND BINOMIAL

1. On average there are 6 accidents per week on Manning Rd.		
Assuming a Poisson distribution determine the probability that in a week there are:		
a	no accidents	0.0025
b	some accidents	0.9975
c	5 accidents	0.1606
d	more than 3 accidents but less than 6	0.2945
e	5 accidents per week in 6 of the next 10 weeks?	0.0018
f	What is the average number of accidents per day?	6/7
2. Still assuming a Poisson distribution determine the probability that on a given day there are		
a	no accidents	0.4244
b	some accidents	0.5756
c	5 accidents	0.0016
d	more than 3 accidents but less than 6	0.0112
e	no accidents in 6 of the next 10 days?	0.1347

Curtin College

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

HOMEWORK NORMAL DISTRIBUTIONS

Assuming a normal distribution, $\mu = 10, \sigma = 3$ determine

a) $P(X < 10) = P(Z < \frac{10-10}{3}) = P(Z < 0) = 0.5.$

b) $P(X > 10) = P(Z > 0) = 0.5.$

c) $P(X = 10) = 0$

d) $P(X < 4.5) = P(Z < \frac{4.5-10}{3}) = P(Z < -1.8) = 0.0336$

e) $P(X > 4.5) = 1 - (d) =$

f) $P(X > 11) = P(Z > \frac{11-10}{3}) = P(Z > 0.33) = 1 - P(Z < 0.33) = 1 - 0.6293 = 0.3707$

g) $P(X > 20) = P(Z > 3.33) = 1 - 0.9996 = 0.0004$

h) $P(4.5 < X < 10) = P(-1.83 < X < 0) = 0.5 - 0.0336 = 0.4664$

i) $P(10 < X < 13) = P(0 < Z < 1) = 0.8413 - 0.5$

j) a such that $P(X < a) = 0.33$

k) b such that $P(X > b) = 0.33$

l) $P(a < X < b)$

(j) $Z = -0.44 = \frac{a-10}{3} \Rightarrow a = 8.68$

(k) $P(X < b) = 1 - 0.33 = 0.67$

$\Rightarrow Z = 0.44 = \frac{b-10}{3} \Rightarrow b = 11.32$

(l) $P(8.68 < X < 11.32) = P(-0.44 < Z < 0.44)$

$=$