Year 13 Mathematics

EAS 3.14

Probability Distributions

Robert Lakeland & Carl Nugent

Contents

nuLake Ltd
Innovative Publisher of Mathematics Texts

Probability Distributions 3.14

This achievement standard involves applying probability distributions in solving problems.

Achievement	Achievement with Merit	Achievement with Excellence
• Apply probability distributions in solving problems.	• Apply probability distributions, using relational thinking, in solving problems.	• Apply probability distributions, using extended abstract thinking, in solving problems.

- This achievement standard is derived from Level 8 of The New Zealand Curriculum and is related to the achievement objectives:
  - Investigate situations that involve elements of chance:
    - calculating and interpreting expected values and standard deviations of discrete random variables.
    - applying distributions such as the Poisson, binomial, and normal.
- Apply probability distributions in solving problems involves:
  - selecting and using methods
  - demonstrating knowledge of concepts and terms
  - communicating using appropriate representations.
- Relational thinking involves one or more of:
  - selecting and carrying out a logical sequence of steps
  - connecting different concepts or representations
  - demonstrating understanding of concepts

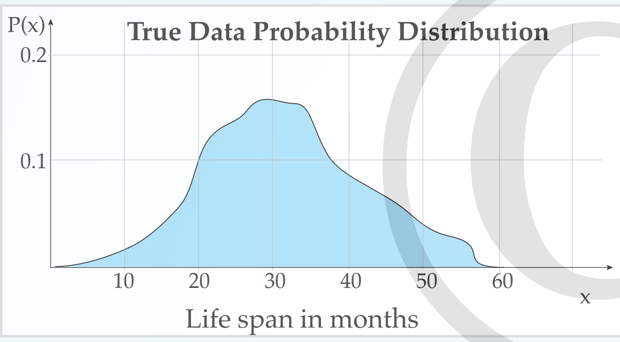
  and also relating findings to a context, or communicating thinking using appropriate statements.
- Extended abstract thinking involves one or more of:
  - devising a strategy to investigate or solve a problem
  - identifying relevant concepts in context
  - developing a chain of logical reasoning
  - making a generalisation

  and also, where appropriate, using contextual knowledge to reflect on the answer.
- Problems are situations that provide opportunities to apply knowledge or understanding of mathematical concepts and methods. Situations will be set in real-life or statistical contexts.
- Methods include a selection from those related to:
  - discrete and continuous probability distributions
  - mean and standard deviation of random variables
  - distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities.

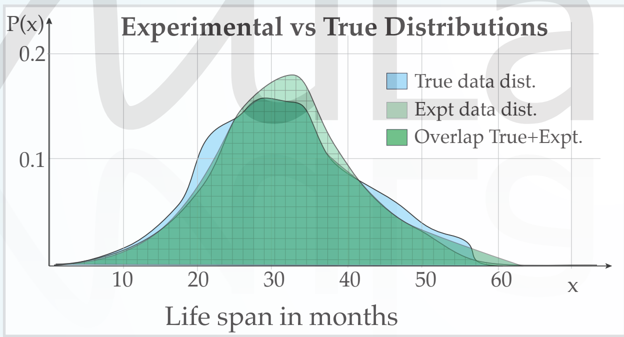
Modelling Real Life Situations

True versus Experimental versus Model Distributions

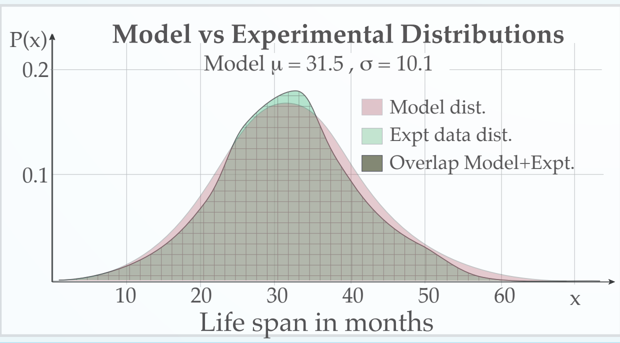
True data is the real life distribution of data. This data is usually unknown to the researcher. It could, for example, be the expected life of computer hard drives. Assuming you cannot run every hard drive until it fails, the researcher will never know the actual distribution. Let us assume the real life or true probability distribution is as per the diagram.



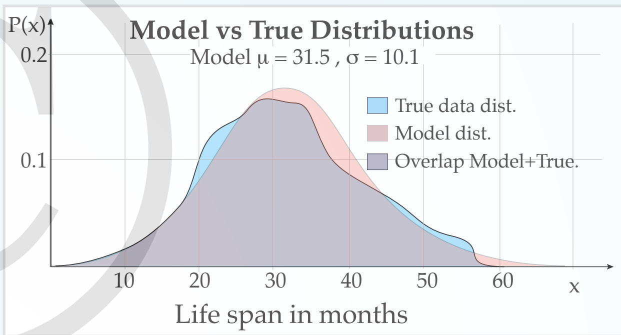
The researcher collects sample or simulation data (called experimental data) on the distribution of life spans and uses these results to develop a model. The researcher collecting sample data on this situation may have ended up with the probability distribution below.



The data from the sample or simulation is likely to be similar to but different from the true data distribution. The researcher calculates the mean and standard deviation from this experimental data and graphs the data. On the basis of the distribution and statistics, the researcher selects the best model for making conclusions about the data, which in this case is a normal distribution.



The theoretical model (normal distribution) is perfectly symmetrical and has the same mean and standard deviation as the experimental data, but you can see that the model is different to the experimental distribution, particularly for results over 45. If we now compare our model to the true data we see it is different in other respects.



Remember the true distribution is not known and if we were to use our model to work out the probability of a result between 45 and 55, this model will underestimate the true distribution whereas the model will overestimate the probability of a result from 30 to 40.

No model will represent a real or true distribution exactly. When we have a distribution which we want to make decisions about, a probability model may describe the distribution well enough to be useful in making these decisions. Even so we need to realise that the probabilities we are calculating are for the theoretical model. This model may approximate the experimental or sample data we have and we are assuming that the experimental or sample data accurately represents the true situation.

In modelling a real situation we should graph the experimental or sample distribution to see how closely it resembles the theoretical model we hope to use. We can compare some theoretical results to our experimental results.

We will also need to consider how closely the experiment fulfills the conditions for the theoretical model.

It could be that we are using a binomial model to represent shots from a free throw line in basketball. Experiments have shown that the probability of a shot going in for a particular player is 0.55. In reality the probability of each shot may be slightly affected by the result of a previous shot (not completely independent) and in modelling this situation with a binomial model we only approximate the real result.

Example

A duck shooter has a probability of 0.4 of hitting any duck that he shoots at during a hunt. In a hunt where he fires 10 shots, find the probability that he

a) shoots exactly 10 ducks.

b) shoots exactly 6 ducks.

c) shoots more ducks than he misses.

We can use the binomial distribution to solve this probability problem, because there are 'n' (10) shots. Each shot is independent of the other. The probability of hitting a duck, 'π', remains constant (0.4) and each shot results in a hit or a miss. The solutions on this and the next page demonstrate the two methods (formula and tables) you can use to solve this problem.

Using the formula

a) with $\pi = 0.4$, $x = 10$ and $n = 10$.

$$P(X = x) = {}^{n}C_{x}\pi^{x}(1-\pi)^{(n-x)}$$

Substituting $\pi = 0.4$, $x = 10$ and $n = 10$.

$$P(X = 10) = {}^{10}C_{10}0.4^{10}(1-0.4)^{(10-10)}$$
$$= 1 \times 0.0001 \times 1$$
$$= 0.0001 \text{ (4 dp)}$$

b) with $\pi = 0.4$, $x = 6$ and $n = 10$.

$$P(X = x) = {}^{n}C_{x}\pi^{x}(1-\pi)^{(n-x)}$$

Substituting $\pi = 0.4$, $x = 6$ and $n = 10$

$$P(X = 6) = {}^{10}C_{6}0.4^{6}(1-0.4)^{(10-6)}$$
$$= 210 \times 0.004096 \times 0.1296$$
$$= 0.1115 \text{ (4 dp)}$$

c) To shoot more than he misses, we calculate $P(X > 5)$ therefore we need to calculate $x = 6, x = 7, \ldots x = 10$ when $n = 10$ and $\pi = 0.4$.

$$P(X > 5) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$
$$= 0.1115 + 0.0425 + 0.0106 + 0.0016 + 0.0001$$
$$= 0.1663$$

We use the binomial distribution table which is supplied in assessments. There is a table at the end of this workbook, part of which is shown here.

n	x	0.05		0.4
10	0	0.5987		0.0060
	1	0.3151	See	0.0403
	2	0.0746	table	0.1209
	3	0.0105	at the	0.2150
	4	0.0010	end of	0.2508
	5	0.0001	this	0.2007
	6		workbook	0.1115
	7			0.0425
	8			0.0106
	9			0.0016
	10			0.0001

a) with $n = 10$, $x = 10$ and $\pi = 0.4$

$P(X = 10) = 0.0001$

b) with $n = 10$, $x = 6$ and $\pi = 0.4$

$P(X = 6) = 0.1115$

c) To shoot more than he misses, we need $P(X > 5)$ so we add the values from the table $x = 6$ to $x = 10$ when $n = 10$ and $\pi = 0.4$.

$$P(X > 5) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$
$$= 0.1115 + 0.0425 + 0.0106 + 0.0016 + 0.0001$$
$$= 0.1663$$

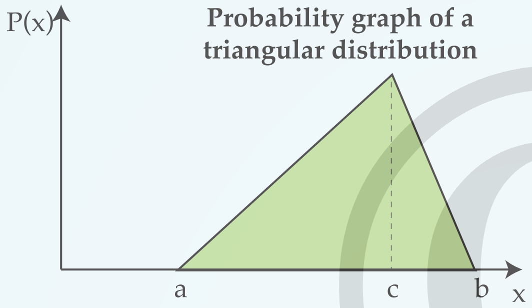
Be careful with the language of each question. It is often a good idea to write out the terms being requested.

If a probability has n = 10 and asks for

- **"less than 4" then you require 0 to 3 i.e. $P(X \le 3)$.**
- **"no more than 4" means 0 to 4. i.e. $P(X \le 4)$.**
- **"more than 4" means 5 to 10 so you calculate $P(X \ge 5) = 1 - P(X \le 4)$.**
- **"over half" again means 6 to 10 so you would calculate $P(X \ge 6) = 1 - P(X \le 5)$.**
- **"no less than 4" means 4 to 10 so you calculate $P(X \ge 4) = 1 - P(X \le 3)$.**
- **"no result" means $P(X = 0)$.**

Triangular Probability Distribution

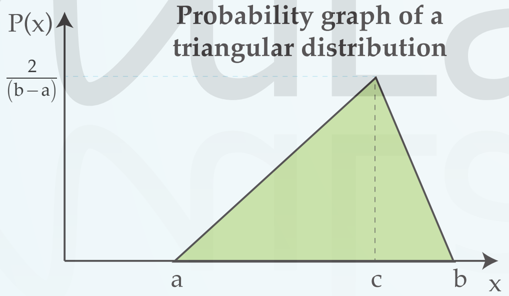
If we only know the minimum, maximum and mode of a distribution we assume the distribution is triangular. For example, if we are expecting a bus to arrive between a and b but typically it arrives at c then we assume the distribution is triangular.



We know the area of the triangle must be one so we can calculate the vertical height. Let the height be h and from the area of a triangle we have

$$\frac{1}{2}(b-a)h = 1$$

$$h = \frac{2}{(b-a)}$$



The height of the line from (a, 0) to (c, h) enables us to calculate the probability from a to c.

$$f(x) = \frac{2(x-a)}{(b-a)(c-a)}$$

The height of the line from (c, h) to (b, 0) enables us to calculate the probability from c to b.

$$f(x) = \frac{2(b-x)}{(b-a)(b-c)}$$

We use a Triangular Distribution when we have

- **a distribution that is approximately Triangular OR**
- **we only have the Maximum, Minimum and Modal x values.**

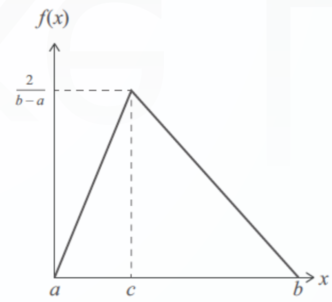
Fortunately you do not have to remember these equations or the graph as both are given to you in the examination tables booklet. See the clippings below.

Get a copy of the tables booklet and inspect it prior to any examination.

Triangular Distribution

The probability density function, $f(x)$, for a triangular distribution is defined as:

$$f(x) = \begin{cases} 0, & x < a \\ \dfrac{2(x-a)}{(b-a)(c-a)}, & a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(b-c)}, & c \le x \le b \\ 0, & x > b \end{cases}$$



Area of a triangle $= \frac{1}{2}$ base $\times$ height

Continuous Uniform Distribution

The probability density function, $f(x)$, for a continuous uniform distribution is defined as:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

64. Salmon being farmed for the export market are weighed to the nearest 50 g (0.05 kg). The mean weight of the salmon is 4.852 kg with a standard deviation of 0.650 kg.

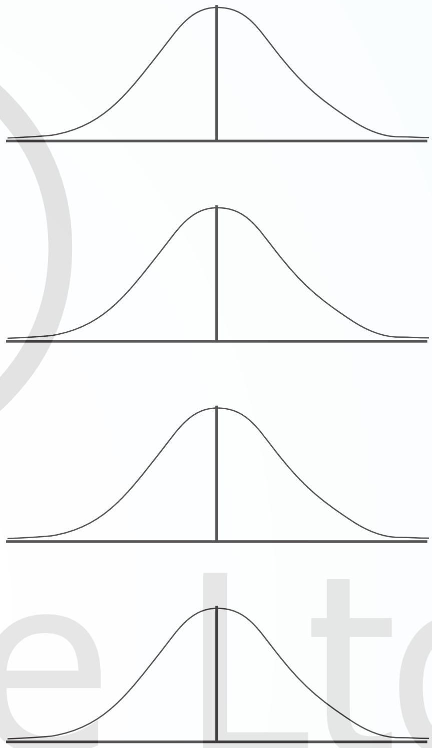
a) Find the probability that a randomly selected salmon weighs more than 5.00 kg.

b) Find the probability that a randomly selected salmon weighs 5.00 kg or more.

c) Find the probability that a randomly selected salmon is recorded as 5.00 kg.

d) The lightest quartile of salmon are sold as baby salmon.
What weight identifies that a salmon will be sold as a baby salmon?

Label the normal curves

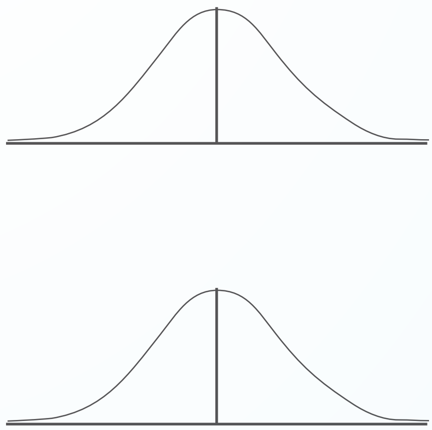


65. The number of passengers wanting to use a bus is normally distributed with a mean of 21.4 and standard deviation of 5.2. The company is considering purchasing new 30 seat buses.

a) Find the probability that over 30 people want to travel on a particular bus.

b) If the bus company reduces the seats to 28 it does not require a conductor.
How many more times is a 28 seat bus likely to have insufficient seats than a 30 seat bus?

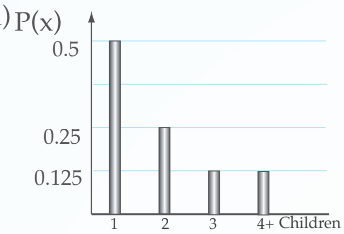
Label the normal curves



Answers

Page 7

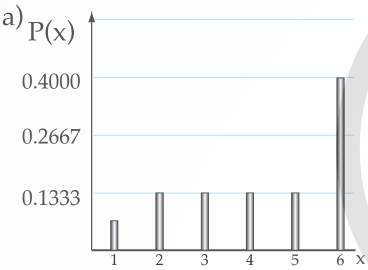
1. a)



b) P(X = 2+) = 0.5

c) P(1 brother) = 0.5 x 0.25
= 0.125

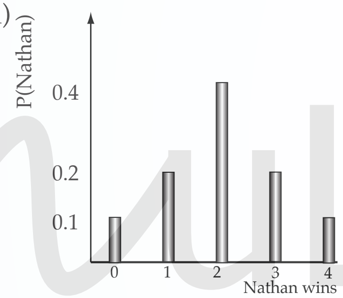
2. a)



b) P(X = 6) = 0.4

c) P(X = odd) = 0.3333

3. a)



b) P(X = 2) = 0.4

c) P(X = 3+) = 0.3

Page 9

4. a)

X	1	2	3	4
P(X = x)	0.25	0.25	0.25	0.25

b) E(X) = 2.5

5. a) E(X) = 1.59 hours
Var(X) = 0.637 (3 sf)

b) P(X < 2) = 0.47

6. a)

X	1	2	3	4	5
P(X = x)	0.45	0.30	0.12	0.10	0.03

b) P(5 days) = 0.03

c) E(X) = 1.96 days

7. a)

X	0	1	2	3
P(X = x)	p	6p	2p	p
P(X = x)	0.1	0.6	0.2	0.1

b) P(3 lambs) = 0.1

c) E(X) = 1.3 lambs
Var(X) = 0.610 (3 sf)

Page 9 cont...

8.

X	⁻6	3	4	5	12
P(X = x)	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

E(X) = 2 squares
Var(X) = 40.3 (3 sf)

9.

X	0	\$5	\$10
P(X = x)	0.375	0.375	0.25

E(Winnings) = 4.375
E(Result) = ⁻0.625
Loss of 62 or 63 cents
Var(X) = 15.2

Page 11

10. E(2X) = 14
Var(2X) = 2^2 x 5.83
= 23.32

11. E(X + X) = 14
Var(X + X) = 5.83 + 5.83
= 11.66

12. a) E(M) = 4.2
Var(M) = 2.676^2
= 7.16
E(C) = 3.75
Var(C) = 2.095^2
= 4.39

b) P(some) = 1 – 0.03
= 0.97

c) E(Tot.) = 7.95
Var(Tot) = 11.55

13. a) About 10.
Distribution fairly symmetrical about 10.

b) E(C) = 9.9
Var(C) = 45.1

c) E(\$) = 123.75
Var(\$) = 7047

14. a) Kaydee E(C) = 15 (symmetrical)
Jenni E(C) = 12 Skewed distribution.

b) Kaydee E(C) = 15
Var(C) = 30.7
Jenni E(C) = 11.85
Var(C) = 18.4.

c) E(K – J) = 3.15
Var(K – J) = 49.1

Page 18

15. a) P(X = 2) = 0.2791

b) P(X < 3) = 0.8217

c) P(X ≤ 2) = 0.8217(6)

d) P(X > 2) = 0.1783(4)

Page 18 cont...

16. a) P(X = 0) = 0.0576

b) P(X = 4) = 0.1361

c) P(X ≤ 3) = 0.8059

d) P(X ≥ 6) = 0.0113

17. a) P(X = 5) = 0.0001

b) P(X = 0) = 0.4437

c) P(X > 3) = 0.0023

d) P(X < 2) = 0.8352

18. a) P(X = 5) = 0.0467

b) P(X ≥ 5) = 0.0580

c) P(X = 1) = 0.1977

d) P(X ≤ 4) = 0.9420

Page 19

19. a) P(X = 6) = 0.1762

b) P(X < 5) = 0.0196

c) P(3 ≤ X ≤ 7) = 0.5636(5)

d) P(X = 9) = 0.1342

20. a) P(X = 2) = 0.0229

b) P(X < 5) = 0.2616

c) P(X = 0) = 0.0003

d) P(X = 8) = 0.0763

21. a) P(X = 4) = 0.2322

b) P(X ≤ 3) = 0.1737

c) P(X > 4) = 0.5941

d) P(X = 0) = 0.0007

22. a) P(X ≥ 5) = 0.8552(1)

b) P(X = 6) = 0.2731

c) P(5 ≤ X ≤ 7) = 0.7121

d) P(X > 0) = 0.9999

Page 20

23. a) Fixed number of events.
Independence assumed.
Probability constant.
Only two outcomes.

b) P(X ≥ 2)
= 1 – [P(X = 0) + P(X = 1)]
= 0.4573(4)

c) P(X ≥ 2) and P(X ≥ 2)
= 0.2091

24. a) P(X = 3) = 0.2668

b) P(X > 2) = 1 – 0.3828
= 0.6172

c) $[P(X>2)]^2$ = 0.3809

25. a) P(X ≥ 5) = 0.9936

b) $[P(X \geq 5)]^5$ = 0.9686

c) Yes. Only two outcomes.
Fixed No. of trials = 10.
Independence assumed.
Constant probability = 0.8.
If a disease or fungi spreads from one bulb to another, results are not independent, for example.