Lecture 4

Discrete Random Variables and Moments of Distributions

By
Dr Sean Maudsley-Barton and Abdul Ali

Manchester Metropolitan University

Aims

- Understand the concept of a **discrete** random variable and its probability distribution.
- Be able to use a discrete probability distribution to calculate probabilities.
- Understand the concept of **expected value** and **variance**.
- Derive the expected value and variance for random variables.
- Derive the expected value and variance for different distributions.

Discrete Random Variables (Chapter 5)

Random Variables

- In many, if not most, practical situations the random outcome of interest is a **numerical** value.

Definition 5.1 (*Random Variable (R.V)*)

*A random variable is the result of any **statistical experiment** which results in a numerical outcome (numerical value).*

- R.V are usually denoted as **capital roman letters** such as T, X, Y, Z.
- The **sample space (S)** of a random variable will **always be a set of numbers**.

Types of random variables:

The **two** important ones are,

- **Discrete Random variable**
- **Continuous Random variable**

Discrete Random Variables

Discrete Random variable:

The variable can take on a set of **discrete**, usually **integer** (**whole number**) values and arise from **counting process**.

Example 5.1: Let X = **score when rolling a die**. This is clearly an experiment with a random numerical outcome with sample space S = **{1, 2, 3, 4, 5, 6}** and the score, X, is a discrete random variable.

Continuous Random variable:

The variable can take on any value in a given **range** or **interval** and arise from **measurement process**.

Example 5.1a: Let T = **Time spent waiting** for the next bus. Here T could be any positive number (***Integer or decimal*** *from a range or interval*), so T would be regarded as a continuous random variable. Similarly,

- **Height** of an individual
- **Weight** of an individual

- We will firstly consider **discrete random** variables.

Discrete Probability Distributions

Manchester Metropolitan University

In order to understand how a random variable is likely to behave, and thus be able to **predict its possible future values**, we clearly need to consider the probability with which it will take on particular values. **This set of probability values is known as a distribution or Probability distribution**.

Some simple examples will help to clarify the idea.

Definition 5.2 (*Mass Function*)

We denote the **mass function** of a discrete random variable, X say, as,

$$\boxed{f(x) = P(X = x)}$$

Name of the R.V

Observed Value of the R.V

Discrete Probability Distributions

Example 5.2: (***A Uniform Distribution***)

Consider **rolling a fair die** and let the discrete random variable, **X**, be the score observed on the die.

We know that the probability of getting any of the values in the set {1,2,3,4,5,6} is $\frac{1}{6}$ and this is the probability distribution. We can **represent the distribution** in various ways.

By **tabulating** the values:

$$P(X=1) = \frac{1}{6}$$
$$P(X=2) = \frac{1}{6}$$
$$P(X=3) = \frac{1}{6}$$
$$P(X=4) = \frac{1}{6}$$
$$P(X=5) = \frac{1}{6}$$
$$P(X=6) = \frac{1}{6}$$

- obviously a bit long-winded and only really feasible if there are a small number of possible values for the random variable.

Discrete Probability Distributions

Example 5.2: (*A Uniform Distribution*)

- By means of a **formula**:

$$P(X = x) = \begin{cases} \frac{1}{6} & x = 1,2,3,4,5,6 \\ 0, & otherwise \end{cases}$$

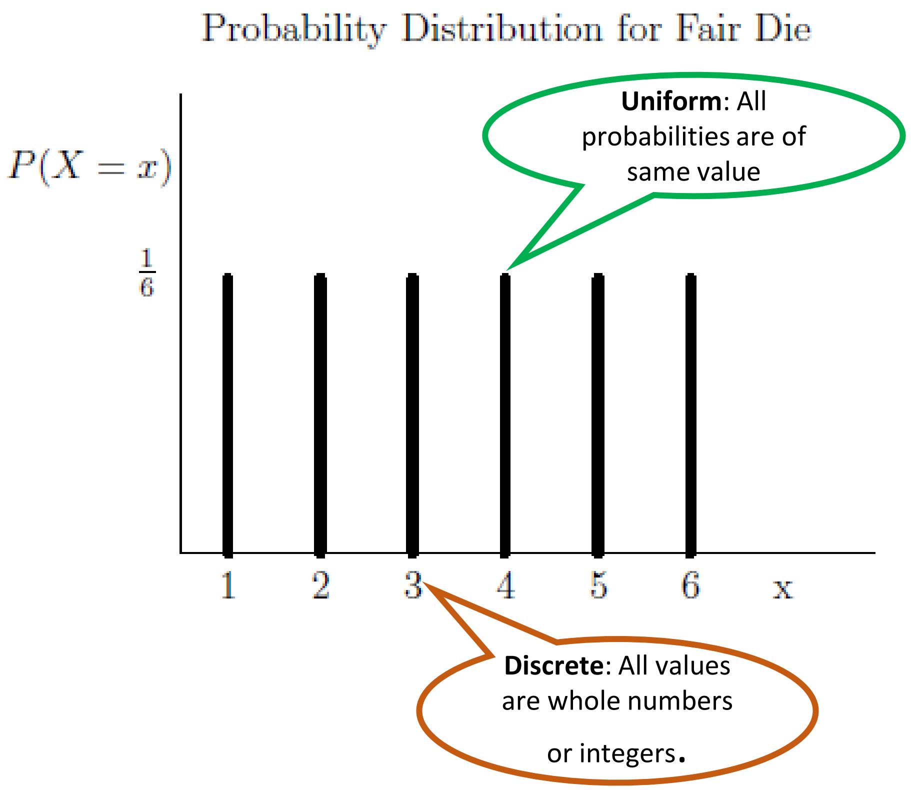
Note:

- There's a subtle difference between a capital X and a small x. The former represents the random variable, i.e. the random numerical outcome of the experiment, the latter represents a value that is observed for the variable.
- Thus, in this case, we read **P(X = x)** as "the probability that the score on the die is x", where x can be any value in the set {1,2,3,4,5,6}.

Discrete Probability Distributions

Example 5.2: (*A Uniform Distribution*)

- By means of a **graph**:



Clearly a very **useful way to visualise** how the **probability is distributed**. You can also see why this is called a (**discrete**) **uniform distribution** - it's because the values are all the same.

Discrete Probability Distributions

Example 5.2: (*A Uniform Distribution*)

Some questions to consider:

1. Suppose your **die** had **n** sides, where **n** is some **whole number greater than 1** (if **n=2** you've basically got a **coin**) and the faces are numbered $1, \ldots, n$. What does the distribution of the score look like now?

2. Can you represent the distribution in each of the three ways suggested above (i.e. table, formula and graph)?

Discrete Probability Distributions

Example 5.3: (*An Urn problem*)

An urn contains **five balls** numbered **1 to 5**. **Two balls** are drawn simultaneously.

1. Let X be the larger of the two numbers.
2. Let Y be the sum of the two numbers.

Find the probability distributions of X and Y.

Discrete Probability Distributions

Example 5.3: (*An Urn problem*)

We proceed as follows by enumerating all the possibilities and noting that there are $\boldsymbol{C_2^5}$**= 10** ways of drawing the **2 balls** from the urn:

1. To find the distribution of X,

X	Outcome				$P(X=x)$
2	(2,1)				1/10
3	(3,1)	(3,2)			2/10
4	(4,1)	(4,2)	(4,3)		3/10
5	(5,1)	(5,2)	(5,3)	(5,4)	4/10

To check that this is a valid distribution we note that:

1. $0 \leq \mathrm{P}(\mathrm{X} = x) \leq 1 \ \forall\ x$ and
2. $\sum_{\mathrm{x}} \mathrm{P}(\mathrm{X} = x) = 1$.

We also note that (**Mass function**) $f(x) = 0$ for any other values of X, since it is **impossible to observe** any other values.

Discrete Probability Distributions

Example 5.3: (*An Urn problem*)

2. To find the distribution of Y,

The Proof/Demonstration is:

Y	Outcome		$P(Y=y)$
3	(2,1)		1/10
4	(3,1)		1/10
5	(4,1)	(3,2)	2/10
6	(5,1)	(4,2)	2/10
7	(5,2)	(4,3)	2/10
8	(5,3)		1/10
9	(5,4)		1/10

Again, to check that this is a valid distribution we note that

1. $0 \leq P(Y = y) \leq 1 \ \forall\, y$ and
2. $\sum_y P(Y = y) = 1$

Discrete Probability Distributions

Example 5.4: (*A geometric Distribution*)

A **learner driver** never seems to remember anything from one lesson or test to the next. Let's suppose that each time he takes the test $P(\text{Pass}) = \frac{1}{4} \Rightarrow P(\text{Fail}) = \frac{3}{4}$, by the law of complements.

Let the random variable Y be the number of attempts he requires until he passes his test.

- We can construct the **probability distribution of Y** as follows:

	1st pass	
$P(Y = 1) \equiv$ Pass		$= \frac{1}{4}$
	1st fail, 2nd pass	
$P(Y = 2) \equiv$ Fail $\rightarrow$ Pass		$= 3/4 \times 1/4 = \frac{3}{16}$
	1st fail, 2nd fail, 3rd pass	
$P(Y = 3) \equiv$ Fail $\rightarrow$ Fail $\rightarrow$ Pass		$= 3/4 \times 3/4 \times 1/4 = \frac{9}{64}$
	1st fail, 2nd fail, 3rd fail, 4th pass	
$P(Y = 4) \equiv$ Fail $\rightarrow$ Fail $\rightarrow$ Fail $\rightarrow$ Pass		$= 3/4 \times 3/4 \times 3/4 \times 1/4 = \frac{27}{256}$ …

And we could, of course, carry on indefinitely (there's no guarantee that George will ever pass!).

- What patterns do we notice?

Discrete Probability Distributions

Example 5.4: (*A geometric Distribution*)

A formula for the probability distribution is derived as follows.

- If he passes on the y^{th} attempt, he must have had $y-1$ failures each with probability $\frac{3}{4}$ followed by his successful pass with probability $\frac{1}{4}$.
- If we assume **each attempt is independent** of the others, the probability of this is:

$$\overbrace{\underbrace{\frac{3}{4}\times\frac{3}{4}\times\cdots\times\frac{3}{4}}_{y-1\ \text{times}}}^{y-1\ \text{fails}}\times\overbrace{\frac{1}{4}}^{y^{th}\ \text{pass}}=\left(\frac{3}{4}\right)^{y-1}\frac{1}{4}$$

Thus we can write a **function**,

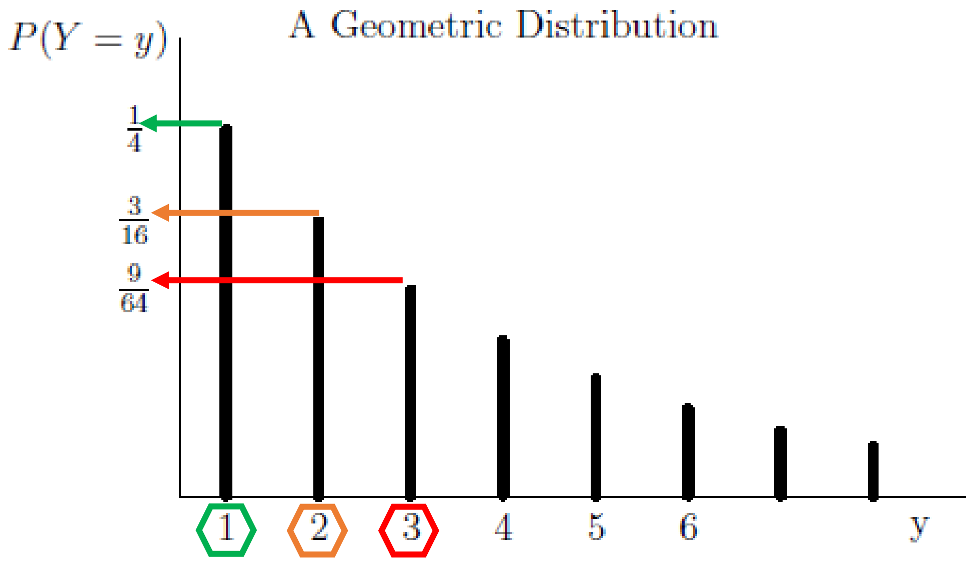
$$\boxed{P(Y=y)=\begin{cases}\left(\frac{3}{4}\right)^{y-1}\frac{1}{4} & y=1.2.3\ldots\ldots\\ 0 & Otherwise\end{cases}}$$

Clearly, these probabilities are **quickly getting very small** - you may recognise these terms as being in a **geometric sequence** with **common ratio** $\frac{3}{4}$.

Discrete Probability Distributions

Example 5.4: (*A geometric Distribution*)

- A **graph** of the probability distribution looks like



- The choice of $\frac{1}{4}$ to represent the **probability of passing** his test at any attempt was clearly **arbitrary**.
- What **range of values** would it be possible to choose in this situation?

The answer is clearly that we can define **two events**, say "**success**" and "**failure**" with probabilities π and $1-\pi$, respectively.

Discrete Probability Distributions

If the first **success** occurs on the x^{th} trial, then we must have previously observed $(x-1)$ **failures**. Assuming all, **trials are independent**, this leads to,

Definition 5.3 (*The geometric Distribution*)

A random variable $X \sim$ **Geom(π)** representing the number of independent trials until the first success where **P(success) = π** has **probability mass function**,

Mass function of geometric Probability distribution.

$$\boxed{f(x;\ \pi) = \begin{cases} (1-\pi)^{x-1}\,\pi & x = 1,2,3,\ldots\ldots \\ 0 & Otherwise \end{cases}}$$

The symbol $\sim$ is taken to mean "**has the distribution**".

Properties of Probability Distributions

Because random variables and their associated distributions are a special case of a more general definition of probability, they **must follow the same rules** as before.

For example, for any **probability mass function**,

1. $0 \leq f(x) \leq 1$ - any probability is always between 0 and 1.

2. $\sum_x f(x) = 1$ - the sum of all the probabilities equals 1.

As we have seen, probability distributions can be represented in a variety of ways.

- In practice, we use **tables of distributions** or use **computer functions** to evaluate them.
- However, the **underlying principle is the same** - the *probability distribution gives us information about the chance of the random variable taking on particular values*.

Properties of Probability Distributions

Definition 5.4: (*Cumulative Distribution Function (CDF)*)

Suppose a random variable, X, has **probability mass function** defined by the function $f(x)$. The **cumulative distribution function (CDF)**, $F(x)$ is defined as,

$$\boxed{F(x) = P(X \leq x) = \sum_{t \leq x} f(t)}$$

i.e. it's the **sum** of all the probabilities corresponding to all values **up to** and **including x**.

Because of it's definition, the **CDF has 3 immediate properties**,

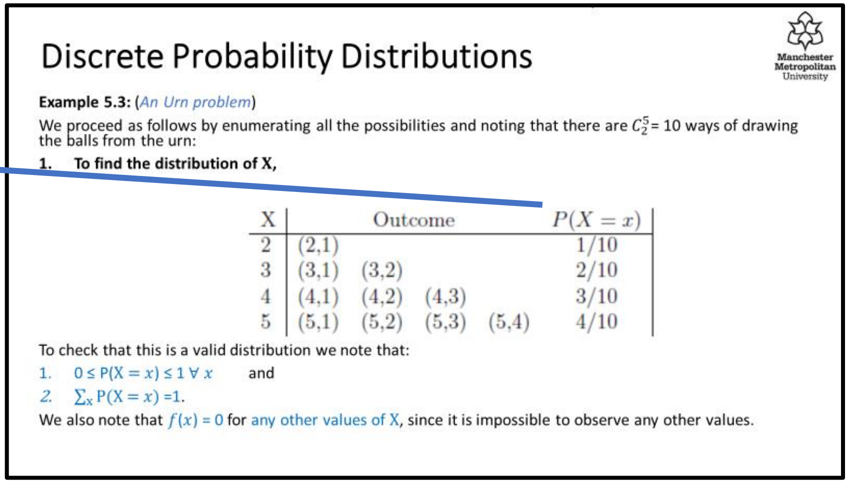
1. $F(-\infty) = 0$
2. $0 \leq F(x) \leq 1 \; \forall \; x$
3. $F(\infty) = 1$

Properties of Probability Distributions

Example 5.5: (*An Urn problem*)

Consider the setup from **Example 5.3**, i.e. the **maximum of two numbers** drawn from an urn. We found the probability distribution to be,

X	$f(x) = P(X = x)$
2	1/10
3	2/10
4	3/10
5	4/10



Discrete Probability Distributions

Example 5.3: (*An Urn problem*)

We proceed as follows by enumerating all the possibilities and noting that there are $C_2^5 = 10$ ways of drawing the balls from the urn:

1. **To find the distribution of X,**

X	Outcome				$P(X = x)$
2	(2,1)				1/10
3	(3,1)	(3,2)			2/10
4	(4,1)	(4,2)	(4,3)		3/10
5	(5,1)	(5,2)	(5,3)	(5,4)	4/10

To check that this is a valid distribution we note that:

1. $0 \le P(X = x) \le 1 \ \forall \ x$ and
2. $\sum_x P(X = x) = 1$.

We also note that $f(x) = 0$ for any other values of X, since it is impossible to observe any other values.

So that the **CDF function** can be written as,

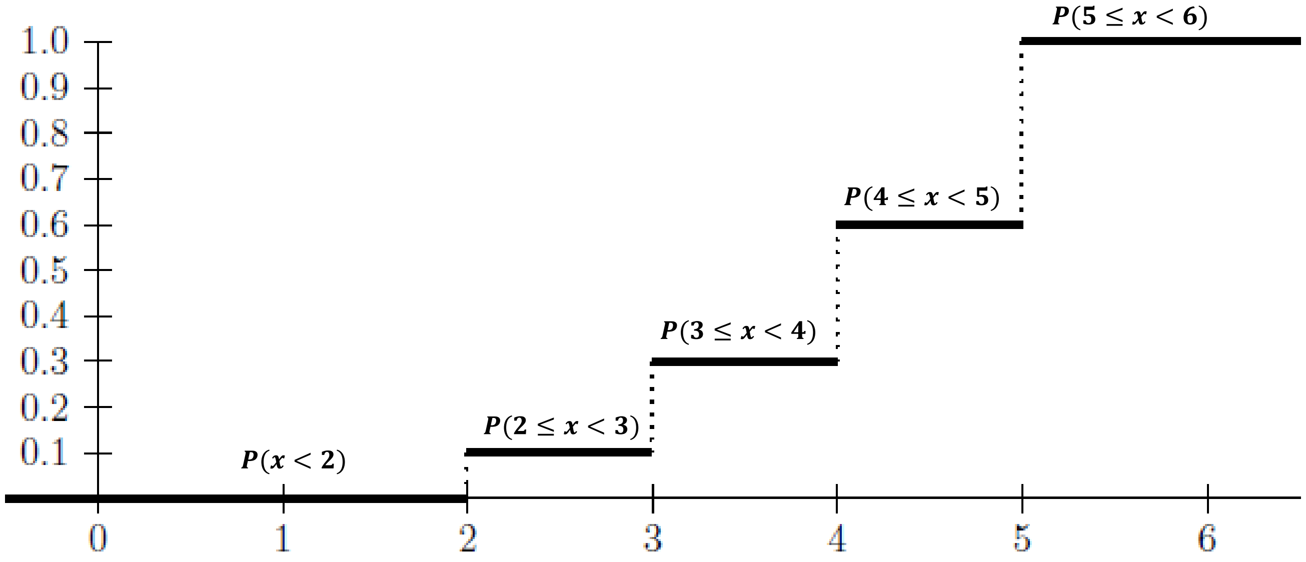
$$F(x) = \begin{cases} & & 0 & x < 2 \\ 1/10 & = & 1/10 & 2 \le x < 3 \\ 1/10 + 2/10 & = & 3/10 & 3 \le x < 4 \\ 1/10 + 2/10 + 3/10 & = & 6/10 & 4 \le x < 5 \\ 1/10 + 2/10 + 3/10 + 4/10 & = & 1 & 5 \le x \end{cases} \qquad \boxed{\begin{array}{c} 0.0 \\ 0.1 \\ 0.3 \\ 0.6 \\ 1.0 \end{array}}$$

This type of function is known as a **step function** and is illustrated in the following figure.

Properties of Probability Distributions

Example 5.5: (*An Urn problem*)

(Step function)



CDF function for urn problem distribution

Properties of Probability Distributions

Example 5.6: (*Geometric Cumulative Distribution Function*)

The **CDF** of the **geometric** distribution is given by,

$$\boxed{F(x) = 1 - (1-\pi)^x}$$

The Proof/Demonstration is:

We have,

$$F(x) = \sum_{t \le x} f(t)$$
$$= \pi + \pi(1-\pi) + \pi(1-\pi)^2 + \ldots\ldots + \pi(1-\pi)^{x-1}$$

This is $\boldsymbol{x}$ **terms** of a **geometric series** with **first term $\boldsymbol{\pi}$** and **common ratio 1− $\boldsymbol{\pi}$**.

Thus, the **sum** is,

$$F(x) = \pi \frac{1-(1-\pi)^x}{1-(1-\pi)} = 1 - (1-\pi)^x$$

- The **(cumulative) distribution function** is **more useful** than the **mass function** since we can always write (*for a discrete distribution*),

$$\boxed{f(x) = F(x) - F(x-1)}$$

Properties of Probability Distributions

Example 5.7: (*Passing the Test*)

Assuming the learner driver's attempts at passing the driving test follow the geometric distribution given in **example 5.4**, find the probability he,

1. passes on the 10th attempt. {=}
2. takes fewer than 4 attempts to pass the test. {Less than or <}
3. takes at least 8 attempts to pass the test. {Greater than or equal or ≥}
4. takes between 4 and 8 attempts inclusive to pass $\{ \leq Y \leq \}$

Properties of Probability Distributions

Example 5.7: (*Passing the Test*)

Let the (discrete) random variable **Y** be the number of attempts to pass the test with,

$$P(Y=y)=\begin{cases}\left(\frac{3}{4}\right)^{y-1}\times\frac{1}{4} & y=1.2.3\ldots\ldots\\ 0 & Otherwise\end{cases}$$

If we note that $Y\sim$ Geom$\left(\frac{1}{4}\right)$, we have,

Mass function

$$f(y)=\left(\frac{3}{4}\right)^{y-1}\times\frac{1}{4}$$

CDF function

$$F(y)=1-\left(1-\frac{1}{4}\right)^{y}$$

$$=1-\left(\frac{3}{4}\right)^{y}$$

Properties of Probability Distributions

Example 5.6: (*Geometric Cumulative Distribution Function*)

The **CDF** of the **geometric** distribution is given by,

$$\boxed{F(x)=1-(1-\pi)^{x}}$$

The Proof/Demonstration is:

We have,

$$F(x)=\sum_{t\le x}f(t)$$
$$=\pi+\pi(1-\pi)+\pi(1-\pi)^2+\ldots..+\pi(1-\pi)^{x-1}$$

This is x **terms** of a **geometric series** with **first term** π and **common ratio** $1-\pi$.

Thus, the **sum** is,

$$F(x)=\pi\frac{1-(1-\pi)^{x}}{1-(1-\pi)}=1-(1-\pi)^{x}$$

- The **(cumulative) distribution function** is **more useful** than the **mass function** since we can always write (*for a discrete distribution*),

$$f(x)=F(x)-F(x-1)$$

Properties of Probability Distributions

Example 5.7: (*Passing the Test*)

The solutions are,

Mass function

1. $P(Y = 10) = \boldsymbol{f(10)} = \left(\frac{3}{4}\right)^9 \times \frac{1}{4} = \mathbf{0.0188}$, i.e. he has 9 failures and then passes.

2. $P(Y < 4) \equiv P(Y \leq 3) = \boldsymbol{F(3)} = 1 - \left(\frac{3}{4}\right)^3 = \frac{37}{64} = \mathbf{0.578}$ ← CDF function

3. We use the **law of complements**. We have,

$$P(Y \geq 8) = 1 - P(Y \leq 7) \qquad \text{\{the complementary event\}}$$

$$= 1 - F(7) \quad \text{← CDF function}$$

$$= 1 - \left(1 - \left(\frac{3}{4}\right)^7\right) = \mathbf{0.1335}$$

Note that, the law of complements says that for a **discrete random** variable, Y , say, we must have, for any value a.

$$\boxed{\boldsymbol{P(Y \leq a - 1) + P(Y \geq a) = 1}}$$

Properties of Probability Distributions

Example 5.7: (*Passing the Test*)

4. The definition of the CDF gives,

$$P(4 \le Y \le 8) = P(Y \le 8) - P(Y \le 3)$$

$$= F(8) - F(3)$$

CDF function

$$= \left[1 - \left(\frac{3}{4}\right)^8\right] - \left[1 - \left(\frac{3}{4}\right)^3\right]$$

$$= \left(\frac{3}{4}\right)^3 - \left(\frac{3}{4}\right)^8 = \mathbf{0.3218}$$

In general, for a discrete random variable X say, we have

$$\boxed{\boldsymbol{P(a \le X \le b) = P(X \le b) - P(X \le a - 1)}}$$

Note:

- As mentioned earlier, probability distributions are **often presented in tabular form**.
- The exact **format will vary** depending on the publisher.

Mean, Variance and Other Moments (Chapter 6)

Mean and Variance

- The **mean** and **variance** of a random variable essentially *mirror* the definitions of mean and variance for samples.

- The ***mean or expected value*** is the ***average*** value of the variable if it were observed repeatedly.

- The ***variance*** indicates the likely ***spread of values*** of the variable.

Mean, Variance and Other Moments

Manchester Metropolitan University

Mean and Variance

Definition 6.1: (*Expected Value*)

- The **expected value (mean) of a discrete random variable**, **X** say, is defined as,

$$\boxed{\mathrm{E}[X] = \sum_{x} x \times P(X = x) = \mu}$$

where the summation extends over all possible values of the variable X.

- The **expected value** of any **function of a discrete random variable**, $\boldsymbol{g(X)}$, say is defined as,

$$\boxed{\mathrm{E}[g(X)] = \sum_{x} g(x) \times P(X = x)}$$

Mean, Variance and Other Moments

Mean and Variance

Definition 6.1: (*Expected Value*)

- The **variance** of a discrete random variable is defined as,

$$\boxed{\mathrm{var}(\mathrm{X}) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2}$$

where $\mu = \mathrm{E}[X]$, defined above.

(**Note:** Proof is shown later.)

Mean, Variance and Other Moments

Example 6.1: (*A loaded die*)

A discrete random variable, X, representing the score on a loaded die has the following **probability mass function**,

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

Find,

1. $\mathrm{E}[X]$
2. $\mathrm{E}[X^2]$
3. var (X)
4. $\mathrm{E}[e^X]$

Mean, Variance and Other Moments

$$\mathrm{E}[X] = \sum_x x \times P(X = x) = \mu$$

Example 6.1: (*A loaded die*)
Using the definitions above,
1. The **expected value** of X is,

$$E[X] = 1 \times \frac{1}{21} + 2 \times \frac{2}{21} + 3 \times \frac{3}{21} + 4 \times \frac{4}{21} + 5 \times \frac{5}{21} + 6 \times \frac{6}{21}$$

$$= \mathbf{4.3333} = \mu$$

As might be expected, ***the value of the mean is towards the upper end of the range of*** $\boldsymbol{X}$ because higher scores are more likely.

What do you think the mean score would be on a fair die?

Mean, Variance and Other Moments

$$\mathrm{E}[g(X)] = \sum_x g(x) \times P(X = x)$$

Example 6.1: (*A loaded die*)
Using the definitions above,

2. We need to find $\mathrm{E}[X^2]$ in order to **find the variance**. We have, using the definition,

$$\mathrm{E}[X^2] = 1^2 \times \frac{1}{21} + 2^2 \times \frac{2}{21} + 3^2 \times \frac{3}{21} + 4^2 \times \frac{4}{21} + 5^2 \times \frac{5}{21} + 6^2 \times \frac{6}{21}$$

$$= \frac{441}{21} = 21$$

$$\mathrm{var(X)} = \mathrm{E}[(X - \mu)^2] = \mathrm{E}[X^2] - \mu^2$$

3. The **variance** is then found as,

$$\mathrm{var(X)} = \mathrm{E}[X^2] - \mu^2$$
$$= 21 - (4.33332)^2 = \mathbf{2.2222}$$

Note: $\mathrm{E}[X^2] \neq E[X]^2 = \mu^2$

Note that a variance, by definition is ***always positive***.

Mean, Variance and Other Moments

Example 6.1: (*A loaded die*)

4. e^X is just a function of X. Using the definition, we have,

$$\mathrm{E}[e^X] = e^1 \times \frac{1}{21} + e^2 \times \frac{2}{21} + e^3 \times \frac{3}{21} + e^4 \times \frac{4}{21} + e^5 \times \frac{5}{21} + e^6 \times \frac{6}{21}$$

$$= \mathbf{164.622}$$

Mean, Variance and Other Moments

Example 6.2: (*The Geometric Distribution*)

In order to derive the ***mean*** *and* ***variance*** of the **Geometric distribution**, we use the following results about geometric series. If **$|r| < 1$**, then

$$g(r) = \sum_{k=0}^{\infty} ar^k = \frac{a}{1-r} \qquad \{r \text{ is a Common Ratio}\}$$

$$g'(r) = \sum_{k=1}^{\infty} akr^{k-1} = \frac{a}{(1-r)^2}$$

$$g''(r) = \sum_{k=2}^{\infty} ak(k-1)r^{k-2} = \frac{2a}{(1-r)^3}$$

The results are obtained by ***successive differentiation*** - ***you're allowed to do that to an infinite series as long as it's convergent***.

Mean, Variance and Other Moments

Example 6.2: (*The Geometric Distribution*)

We can write the **mass function for the Geometric distribution** as,

$$\boxed{f(x;\pi) = \begin{cases} (1-\pi)^{x-1}\,\pi & x = 1,2,3,\ldots \\ 0 & Otherwise \end{cases}}$$

The Proof/Demonstration is:

To find $\mathbf{E}[\boldsymbol{X}]$ we write,

$$\mathrm{E}[X] = \sum_{x=1}^{\infty} x(1-\pi)^{x-1}\,\pi \qquad \text{\{Definition 6.1\}}$$

$$= (\pi + 2\pi(1-\pi) + 3\pi(1-\pi)^2 + 4\pi(1-\pi)^3 + \cdots$$

This is **clearly $\boldsymbol{g'(1-\pi)}$** where $a = \pi$. Hence,

$$\boxed{\mathbf{E}[\boldsymbol{X}] = \frac{\pi}{[1-(1-\pi)]^2} = \frac{\mathbf{1}}{\boldsymbol{\pi}}}$$

Mean, Variance and Other Moments

Example 6.2: (*The Geometric Distribution*)

To find **var(X)** we use a **factorial moment**, defined as $\mathbf{E\,[X\,(X-1)]}$.

For the **geometric distribution**,

$$\mathrm{E}[X(\mathrm{X}-1)] = \sum_{x=1}^{\infty} x(x-1)\pi(1-\pi)^{x-1}$$

$$E[X(X-1)] = (1-\pi)\sum_{x=2}^{\infty} x(x-1)\pi(1-\pi)^{x-2}$$

The infinite series is **clearly** $\boldsymbol{g''(1-\pi)}$ where $a=\pi$. Hence,

$$\mathrm{E}[X(X-1)] = (1-\pi)\frac{2\pi}{[1-(1-\pi)]^3} = \frac{\mathbf{2(1-\pi)}}{\boldsymbol{\pi^2}}$$

Mean, Variance and Other Moments

Example 6.2: (*The Geometric Distribution*)

Now,

$$\boxed{\mathrm{var}(\mathrm{X}) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2}$$

$$\mathrm{var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

$$= \mathrm{E}[X(X-1)] + \mathrm{E}[X] - \mathrm{E}[X]^2$$

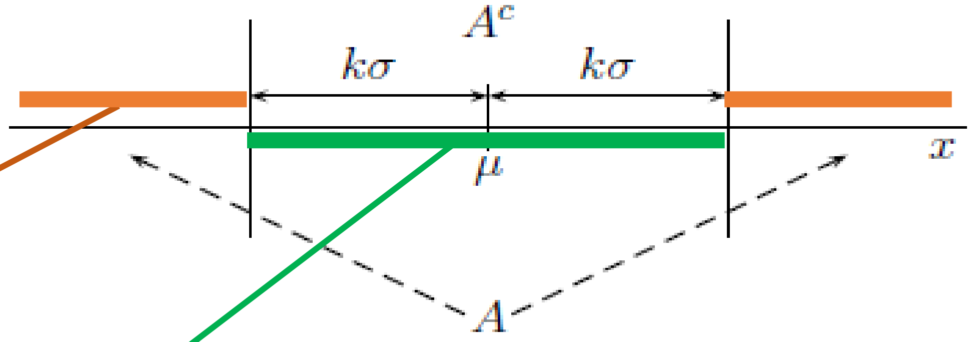
$$= \frac{\mathbf{2(1-\pi)}}{\boldsymbol{\pi^2}} + \frac{1}{\pi} - \frac{1}{\pi^2}$$

$$\boxed{\mathrm{var}(X) = \frac{\mathbf{1-\pi}}{\boldsymbol{\pi^2}}}$$

Further Properties of Moments

The way in which the variance is related to the spread of a random variable is illustrated in the following useful theorem.

Theorem 6.1: (*Chebychev's Inequality*)



If $\mathrm{E}[X] = \mu$ and $\mathrm{var}(X) = \sigma^2$, then

Standard Deviation = $\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2}$
And **k** is *any positive number*.

1. $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ for all $k > 0$

and it follows that,

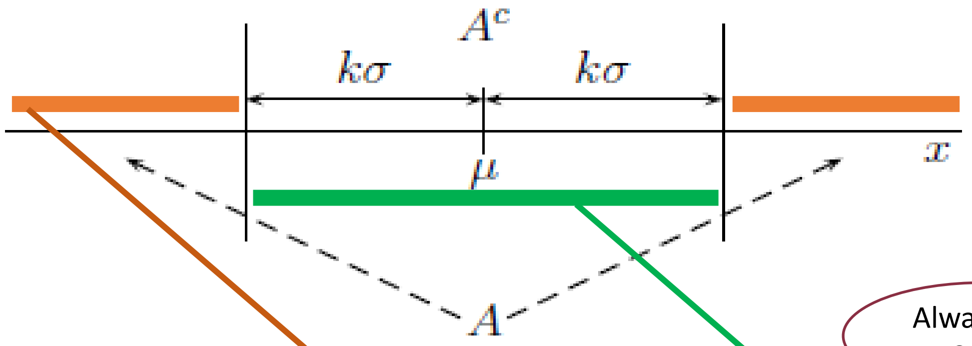
2. $P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$ for all $k > 0$

Note: *Chebychev's Inequality is applicable to any probability distribution.*

Further Properties of Moments

Manchester Metropolitan University

Proof 6.1: Let $p_j = P(X = x)$, denote the probability distribution and define the set $A = \{x : |x - \mu| \geq k\sigma\}$.



Now,

Always **+ve** or **0** value.

$$\sigma^2 = \sum_j (x - \mu)^2 p_j = \sum_A (x_j - \mu)^2 p_j + \sum_{A^c} (x_j - \mu)^2 p_j$$

$$\geq \sum_A \underbrace{(x_j - \mu)^2}_{} p_j$$

This is the sum of squared distances between 'x_j' and μ. The figure above shows the distances are always equal to or more than $k\sigma$.

$$\sigma^2 \geq k^2\sigma^2 \sum_A p_j = k^2\sigma^2 \times P(|X - \mu| \geq \boldsymbol{k}\sigma)$$

Therefore, $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ for all $k > 0$

Further Properties of Moments

Proof 6.1:

Note:

For example, the probability that any random variable takes on a value **more** than **two standard deviations** away from the mean, i.e. $K = 2$ is at most $\frac{1}{2^2} = \mathbf{\frac{1}{4}}$ (but might be much less than this, the **inequality gives an upper bound on the probability**).

Further Properties of Moments

Theorem 6.2 (*The Weak law of large Numbers **(WLLN)***)

It's important to understand the difference between the **sample mean,** $\bar{X}$, and the **population mean** μ. https://onlinestatbook.com/stat_sim/sampling_dist/index.html

The former (i.e. $\bar{X}$) is a feature of a finite sample of realisations from a random variable, the latter (i.e. μ) is a theoretical value that tells us the mean value of random variable were we able to observe an infinite number of outcomes.

However, it is easy to show that, ***as the sample size increases*, the sample mean should get *closer* and closer to the theoretical population mean** - the ***WLLN***. Thus,

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n} \qquad \forall\, \varepsilon > 0,$$

Where $\bar{X} = \frac{1}{n}(X_1 + X_1 + \cdots + X_n)$

Note that we have **n random variables** and we assume that each one is *independent* of the other as well as *identically distributed*, e.g.

$$\boldsymbol{var(X_1) = var(X_2) = \sigma^2}$$

Further Properties of Moments

Proof 6.2:

Note that

$$var\,(\bar{X}) = var\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) = \frac{1}{n^2} var(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n^2}\big(var(X_1) + var(X_2) + \cdots + var(X_n)\big) = \frac{\mathbf{1}}{\boldsymbol{n}^{\mathbf{2}}}\boldsymbol{n\sigma}^{\mathbf{2}} = \frac{\boldsymbol{\sigma}^{\mathbf{2}}}{\boldsymbol{n}}$$

because of their **independence** (*see next theorem*) and **identical** distribution.

- The derivation of this result is a straightforward application of **Chebychev's inequality**.
- As $\boldsymbol{n \to \infty}$, the probability that the value of the sample mean is *arbitrarily close* to the population **converges to one**, i.e. **it's a certainty**.
- Much work in applied statistics revolves around **how big a sample** we need to take in order to assure ourselves that the **sample mean is** ***close enough*** to the population mean.

The following important result tells us how to find the ***mean*** *and* ***variance*** *of a* ***linear combination*** of random variables.

Further Properties of Moments

Theorem 6.3 (*Linear Combinations*)

For any random variables, X and Y, and constants a and b, we have,

$$\mathrm{E}[aX \pm b] = a\mathrm{E}[X] \pm b$$

$$\mathrm{E}[aX \pm bY] = a\mathrm{E}[X] \pm b\mathrm{E}[Y]$$

Property of the Expected value and variance

$$var(aX \pm b) = a^2 var(X)$$

$$var(aX \pm bY) = a^2 var(X) + b^2 var(Y),$$

if X and Y are **independent** random variables.

Further Properties of Moments

Proof 6.3

Straightforward using ***properties of summations*** and ***mass functions***.

For example, the first result is proved as follows,

$$\mathrm{E}[aX \pm b] = \sum_x (ax \pm b)P(X = x)$$

$$= \sum_x [ax\, P(X = x) \pm b\, P(X = x)]$$

$$= a \sum_x x\, P(X = x) \pm b \sum_x P(X = x)$$

$$= a\mathrm{E}[X] \pm b$$

We can use these results, for example, to prove the usual form for calculating the variance as shown in next theorem.

Further Properties of Moments

Theorem 6.4 (*Variance of a Random Variable*)

Proof 6.4

We have,

$$
\begin{aligned}
var(X) = \mathrm{E}[(X-\mu)^2] &= \mathrm{E}[X^2 - 2\mu X + \mu^2] \\
&= \mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2 \\
&= \mathrm{E}[X^2] - 2\mu^2 + \mu^2 \\
&= \mathrm{E}[X^2] - \mu^2
\end{aligned}
$$

Further Properties of Moments

Example 6.3: (*A loaded die again*)

Using the same probability distribution as in **Example 6.1**, assume that a gambling game pays out winnings, £W, as the following functions of the score, X,

1. $W_1 = 2X$

2. $W_2 = 3X - 10$

What are the **expected** winnings of the game and what would the **variance** of the winnings be?

Further Properties of Moments

Example 6.3 (*A loaded die again*)

Since both are linear functions of the random variable, X, we have

1. E $[W_1]$ = 2 × E $[X]$ = 2 × 4.333 = £8.666.

Also,

$$var(W_1) = 2^2\, var(X) = 4 \times 2.2222 = 8.8888.$$

2. E $[W_2]$ = 3 × E $[X] - 10 = 3 \times 4.3333 - 10 = £3$

Also,

$$var(W_2) = 3^2\, var(X) = 9 \times 2.2222 = \mathbf{19.9999}$$

- It must be emphasised that **we can only use this theorem if the functions are linear**, i.e. E $[g(X)]$ = $g($E$[X])$ $\Longleftrightarrow$ $g(.)$ is linear.
- As a counter example, suppose **E $[X^2]$ did equal E $[X]^2$** - this would imply that all random variables had a **variance of zero**!