MATHEMATICS

YEAR 13

NAME:

HIGHER IB

STATISTICS OPTION

BOOK TWO

THE GEOMETRIC DISTRIBUTION

Definition : This occurs in a yes / no situation of constant probabilities. It gives the probability of a string of failures followed by a single success.

ie. $P(FFF....FFS)$

If the probability of success is denoted by p ie. $P(\text{success}) = p$

then

$$\boxed{E(X) = \frac{1}{p} \qquad Var(X) = \frac{q}{p^2}}$$

where $q = 1 - p$

and X stands for the number of goes before you get a success.

Please learn all of this off by heart. It is VERY IMPORTANT

$$\boxed{S_n = 1 - (\text{Fail})^n}$$

NOTATION is $Geo(p)$

Examples : ① A dice is thrown and the game starts as soon as a six is recorded. Find the probability that the game starts on a
(a) first throw (b) second throw (c) third throw (d) n^{th} throw.
(e) What is the chance that the game starts on or before the 3^{rd} throw?
(f) What is the chance that the game starts on or before the 10^{th} throw?
(g) What is the chance that game starts with the 3rd or subsequent throw?
(h) If X stands for the number of throws before a six is thrown, find $E(X)$ and $Var(X)$.

(a) $P(S) = \frac{1}{6}$ $\therefore$ $P(F) = \frac{5}{6}$ $\therefore$ Geometric distribution

$P(\text{starts on first throw}) = P(S) = \underline{\frac{1}{6}}$

(b) $P(\text{starts on second throw}) = P(FS) = \frac{5}{6} \times \frac{1}{6} = \underline{\frac{5}{36}}$

(c) $P(\text{starts on third throw}) = P(FFS) = \left(\frac{5}{6}\right)^2 \frac{1}{6} = \underline{\frac{25}{216}}$

(d) $P(\text{starts on } n^{th} \text{ throw}) = P(\underbrace{FF....F}_{n-1 \text{ of them}} S) = \underline{\left(\frac{5}{6}\right)^{n-1} \times \frac{1}{6}}$

(e) P(on or before 3rd throw) $= P(1) + P(2) + P(3)$

$= (a) + (b) + (c)$

$= \frac{1}{6} + \frac{5}{6}\times\frac{1}{6} + \left(\frac{5}{6}\right)^2 \times \frac{1}{6}$

$= \underline{0.421}$

(f) P(on or before 10th throw) $= P(1) + P(2) + P(3) + \text{-----} + P(10)$

This takes too long to calculate so we use the formula $S_n = 1 - (\text{Fail})^n$

$\therefore \; S_{10} = 1 - \left(\frac{5}{6}\right)^{10} = \underline{0.838}$

(g) P(3rd or subsequent throw) $= P(3, 4, 5, \text{----})$

$= 1 - P(1 \text{ or } 2)$

$= 1 - \left(\frac{1}{6} + \frac{5}{6}\times\frac{1}{6}\right)$

$= \underline{\frac{25}{36}}$

(h) Mean $= E(X) = \frac{1}{p} = \frac{1}{\frac{1}{6}} = \underline{6}$

$\text{Var } X = \frac{q}{p^2} = \frac{\frac{5}{6}}{\left(\frac{1}{6}\right)^2} = \frac{\frac{5}{6}}{\frac{1}{36}} = \underline{30}$

$\therefore$ The average number of throws I would usually expect to throw before I get a six would be <u>six throws</u>.

---

② You stand in 2 in 5 chance of hitting a pheasant. As soon as you've got your first pheasant, conscience strikes, and you stop shooting.

(a) Comment on the probability assumption that enables us to assume a geometric model.

(b) Find the probability that you hit your first pheasant with the 3rd shot.

(c) Find the probability that you'll not be firing any more than 8 time

(d) Find the probability that you stop firing on the 4th or subsequen shot.

(e) If X is the number of shots that you fire, find $E(X)$ and $\text{Var } X$.

(a) We <u>MUST</u> have a <u>CONSTANT PROBABILITY</u> for a <u>geometric</u> model. The probability of hitting ie. $P(S) = \frac{2}{5}$ would probably increase with experience and $\therefore$ not remain constant. Thus a geometric model

(b) P(hit with 3rd shot) = P(FFS) = $\left(\frac{3}{5}\right)^2 \times \left(\frac{2}{5}\right)$ = 0.144

(c) P(hit in first 8 shots) = P(1) + P(2) + P(3) + ----- + P(8)

Again this takes a long time to calculate

$$\therefore \quad S_8 = 1 - (\text{Fail})^8$$
$$= 1 - \left(\frac{3}{5}\right)^8$$
$$= 0.983$$

(d) P(stop firing on 4^{th} or subsequent shot) = P(4, 5, 6, ----)

$$= 1 - P(1, 2, 3)$$
$$= 1 - \left[\left(\frac{2}{5}\right) + \left(\frac{3}{5}\right)\times\left(\frac{2}{5}\right) + \left(\frac{3}{5}\right)^2\times\left(\frac{2}{5}\right)\right]$$
$$= 1 - 0.784$$
$$= 0.216$$

(e) $E(x) = \frac{1}{p} = \frac{1}{0.4} = 2.5$

$\therefore$ You would expect to usually have 2.5 shots on average before you hit the pheasant.

$$Var(x) = \frac{q}{p^2} = \frac{0.6}{0.4^2} = 3.75$$

---

(3) A geometric distribution has a mean of 4. What is its variance

Mean $= \frac{1}{p} = 4 \quad \therefore \quad p = \frac{1}{4} \quad \therefore \quad Geo\left(\frac{1}{4}\right)$

If $p = \frac{1}{4}$ then $q = \frac{3}{4}$.

Thus Variance $= \frac{q}{p^2} = \frac{\frac{3}{4}}{\left(\frac{1}{4}\right)^2} = \frac{\frac{3}{4}}{\frac{1}{16}} = 12$.

---

(4) A geometric distribution has variance 12. What is its mean?

$$\text{Variance} = \frac{q}{p^2} = \frac{(1-p)}{p^2} = 12.$$

$$\therefore \quad 1 - p = 12p^2$$
$$\therefore \quad 0 = 12p^2 + p - 1$$
$$0 = (3p+1)(4p-1)$$
$$\therefore \quad p = -\frac{1}{3} \text{ or } \frac{1}{4}$$

It is impossible for p to be negative $\therefore$ $p = \frac{1}{4}$

$\therefore$ Mean $= \frac{1}{p} = \frac{1}{\frac{1}{4}} = \underline{4}$

---

(5) PAST QUESTION.

In order to decide who starts a game, each player takes it in turn to throw a die. The start is decided by the first person to throw a six.

(i) What assumption about the die must be made in order to apply a geometric model to calculate the probability that the first six occurs on any given throw of the die?

(ii) Find the probability that the game starts with the second throw of the die.

(iii) Find the probability that the game starts with the third or subsequent throw of the die.

(i) We must assume that the die is fair and that the probability of throwing a six remains constant throughout ie. $P(\text{six}) = \frac{1}{6}$

(ii) This is a geometric model with $p = \frac{1}{6}$

$\therefore$ $P(\text{game starts on second throw}) = P(F\ S) = \frac{5}{6} \times \frac{1}{6} = \underline{\frac{5}{36}}$

(iii) $P(3, 4, 5, \ldots) = 1 - P(1 \text{ or } 2)$

$= 1 - \left(\frac{1}{6} + \frac{5}{6} \times \frac{1}{6}\right)$

$= 1 - \frac{11}{36}$

$= \underline{\frac{25}{36} \text{ or } 0.694}.$

---

(6) X is a geometric distribution and $P(X=r) = (1-p)^{r-1} \times p$.

(i) If $E(x) = 5$ find $P(X=4)$

(ii) If $E(x) = 4$ find $P(x < 3)$

(iii) If $Var(X) = 2$ find $P(X=2)$.

(i) $E(x) = 5 = \frac{1}{p}$

$\therefore$ $\underline{p = \frac{1}{5}}$

$\therefore$ $P(x=4) = (1-p)^3 \times p$

$= \left(\frac{4}{5}\right)^3 \times \left(\frac{1}{5}\right)$

$= \underline{0.1024}$

OR $P(x=4) = P(FFFS)$

$= \left(\frac{4}{5}\right)^3 \times \left(\frac{1}{5}\right)$

$= \underline{0.1024}$

(ii) $E(x) = 4 = \frac{1}{p}$

$\therefore$ $\underline{p = \frac{1}{4}}$

$\therefore$ $P(x<3) = P(1 \text{ or } 2)$

$= \left(\frac{1}{4}\right) + \left(\frac{3}{4}\right) \times \left(\frac{1}{4}\right)$

$= \frac{7}{16}$

$= \underline{0.4375}$

(iii) $Var(X) = 2 = \frac{q}{p^2}$.

$\therefore$ $\frac{q}{p^2} = \frac{1-p}{p^2} = 2$

$\therefore$ $1 - p = 2p^2$

$\therefore$ $2p^2 + p - 1 = 0$

$(2p - 1)(p + 1)$

$\therefore$ $\underline{p = \frac{1}{2} \text{ or } -1}$

p must be $\frac{1}{2}$ and not -1.

$\therefore$ $P(x=2) = (1-p)^1 p$

$\therefore$ $P(F\ S) = \left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right)$

$= \frac{1}{4}$

NORMAL APPROXIMATION

Examples ① (a) Roll dice 5 times. Find P(at least 4 fives)

(b) Roll dice 600 times. Find P(at least 110 fives)

(a) $P(\text{at least } 4) = P(4 \text{ or } 5) = P(55555^*) + P(55555)$

$$= \left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)\frac{5!}{4!} + \left(\frac{1}{6}\right)^5$$

$$= \underline{0.00334}.$$

(b) $P(\text{at least } 110 \text{ fives}) = P(110, 111, 112, \ldots\ldots, 600) = 1 - P(0, 1, \ldots\ldots, 109)$

This method would take too long. ∴ Approximation is needed.

This can be approximated by a NORMAL DISTRIBUTION which is valid for $n > 10$ and p is close to $\frac{1}{2}$ but as p moves from $\frac{1}{2}$ we require n to be larger.

$$\boxed{\begin{aligned} \text{Mean} &= np \\ \text{Variance} &= npq \end{aligned}}$$

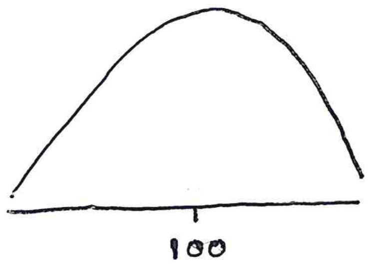
where $q = 1 - p$

In this example, Mean $= 600 \times \frac{1}{6} = \underline{100}$

Variance $= 600 \times \frac{1}{6} \times \frac{5}{6} = \underline{83.33}$

∴ $\underline{\sigma = 9.129}$

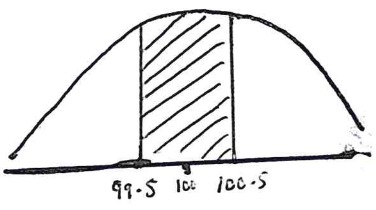
∴



$\sigma = 9.129$

Continuity Correction (HALVES)

eg. Ⓐ P(100 fives)



We assume 100 fives goes between 99.5 and 100.5 so that we can calculate an area on the normal curve.

99 | 100 | 101 fives.

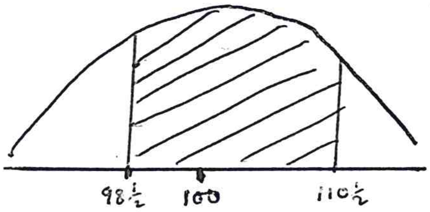
99.5   100.5

(B) P(between 99 and 110 fives)

RING

0 1 2 ---- 98 (99 ... 109 110) 111 ------

98.5 110.5

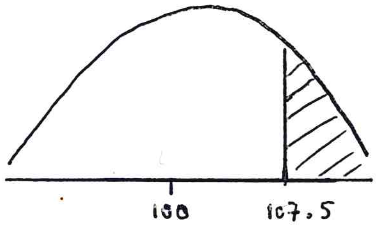


(C) P(at least 108) means 108, 109, 110 ....

$\therefore$ Use $107\frac{1}{2}$

0 1 2 3 ... 107 (108, 109, 110, ...)

107.5

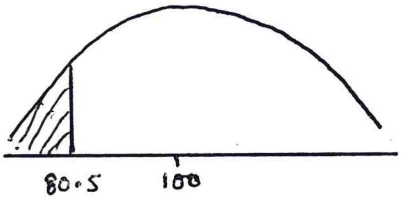


(D) P(no more than 80) means 80, 79, 78, ...

$\therefore$ Use $80\frac{1}{2}$

(0 1 2 .... 79 80) 81 82 ...

80.5

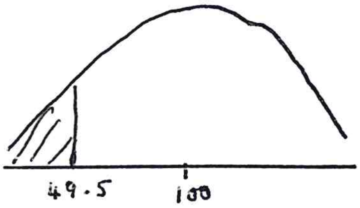


(E) P(less than 50) means 49, 48, 47, ...

$\therefore$ Use $49\frac{1}{2}$

(0 1 2 3 .... 48 49) 50 51 52 ...

49.5

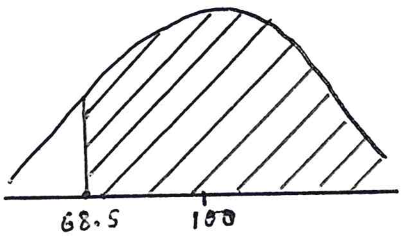


(F) P(more than 68) means 69, 70, 71, ...

$\therefore$ Use $68\frac{1}{2}$

0 1 2 3 ... 67 68 (69 70 ....)

68.5



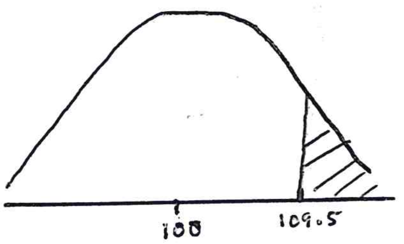
RETURN TO PROBLEM

P(at least 110 fives) means 110, 111, 112 etc

$\therefore$ Use $109\frac{1}{2}$

0 1 2 ... (110 111 112 ...)

109.5



$\sigma = 9.129$

$$\therefore z = \frac{109.5 - 100}{9.129} = 1.041$$

Table gives 85.1 %

$\therefore$ Answer is 14.9 % (or 0.149)

② 40 % of UK are Labour supporters.

(a) Random sample of 5 people. P(less than 4 Labour supporters)

(b) Random sample of 1000 people. P(at least 430 Labour supporters)

(a) $P(L) = 0.4 \qquad P(L^*) = 0.6$

0 1 2 3 4 5

$P(\text{less than } 4) = P(0, 1, 2 \text{ or } 3) = 1 - P(4 \text{ or } 5)$

$\therefore \quad P(LLLLL^*) = (0.4)^4 (0.6) \dfrac{5!}{4!} = \underline{0.0768}$

$P(LLLLL) = (0.4)^5 = \underline{0.01024}$

$\therefore \quad P(\text{less than } 4) = 1 - (0.0768 + 0.01024)$

$= \underline{0.91296}$ (or 91.296 %)

(b) $P(\text{at least } 430) = P(430, 431, \ldots, 1000) = 1 - P(0, 1, \ldots 429)$

This takes TOO LONG. $\therefore$ NORMAL APPROXIMATION.

$\therefore \quad p = 0.4 \qquad q = 0.6 \qquad n = 1000$

$\therefore$ Mean $= np = 1000 \times 0.4 = \underline{400}$

Variance $= npq = 1000 \times 0.4 \times 0.6 = \underline{240}$

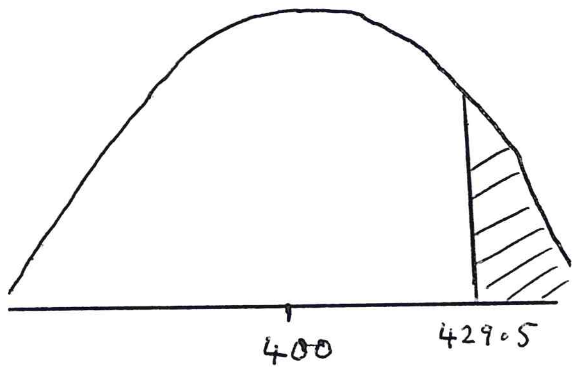
$\therefore \quad \sigma = \underline{15.49}$

USE HALVES $\therefore$ $P(\text{at least } 430) = P(430, 431, \ldots)$

$\therefore$ Use $\underline{429\tfrac{1}{2}}$

0 1 2 . . . . 429 430 431 . . .

429.5



$\sigma = 15.49$

$$z = \frac{429.5 - 400}{15.49} = 1.904$$

Table gives 97.15 %

$\therefore$ Answer is $\underline{2.85\ \%}$

(Look at diagram and THINK about your answer).

③ 85 % of Greshams still have their own calculators. Find

(a) P(less than 10 in a group of 12 still have their own calculators)

(b) P(no more than 80 in a group of 100 still have their own calculators)

(a) 0 1 2 3 4 5 6 7 8 9) 10 11 12

$P(\text{less than } 10) = P(0, \dots, 9) = 1 - P(10, 11, 12)$

$P(10) = P(cccccccccc c^* c^*) = (0.85)^{10}(0.15)^2 \dfrac{12!}{10!\,2!} = 0.292358$

$P(11) = P(ccccccccccc c^*) = (0.85)^{11}(0.15) \dfrac{12!}{11!\,1!} = 0.301218$

$P(12) = P(cccccccccccc) = (0.85)^{12} = 0.142242$

$\therefore P(\text{less than } 10) = 1 - (0.292358 + 0.301218 + 0.142242)$

$= \underline{0.264}$

(b) Standard method **TOO LONG** ∴ **NORMAL APPROXIMATION**.

$p = 0.85 \qquad q = 0.15 \qquad n = 100$

$\text{Mean} = np = 100 \times 0.85 = \underline{85}$

$\text{Variance} = npq = 100 \times 0.85 \times 0.15 = \underline{12.75}$

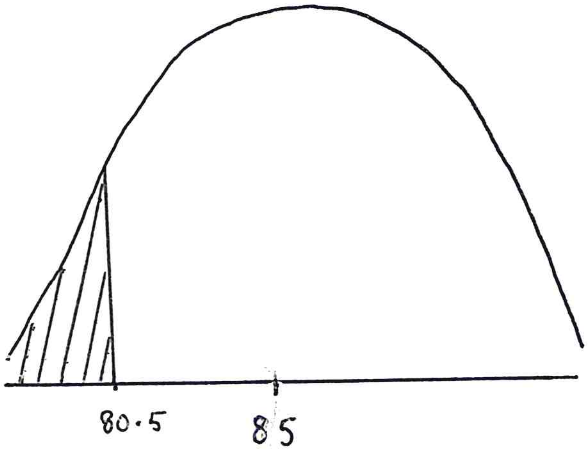
$\therefore \sigma = \sqrt{12.75} = \underline{3.5707}$

USE HALVES ∴ $P(\text{no more than } 80) = P(80, 79, 78, \dots)$

∴ **Use $80\frac{1}{2}$**

0 1 2 . . . . 79 80) 81 82 . . .

80.5



$\therefore z = \dfrac{80.5 - 85}{3.5707} = 1.260$

∴ Table gives 89.62 %

∴ Answer is $\underline{10.38\ \%}$

$\sigma = 3.5707$

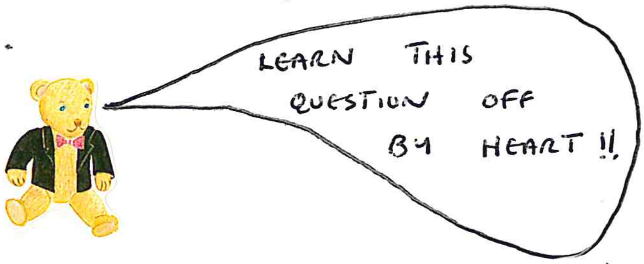
(4) The probability that a silicon chip will fail in less than 1000 hours of use is 0.3. Suppose that 100 random chips are chosen. Let X be the number of chips that fail.

(a) State an assumption that has to be made for X to be modelled on a binomial distribution.

(b) Find $E(X)$ and $Var(X)$.

(c) It is intended to approximate this binomial model by a suitable normal model. State the precise conditions under which the Normal distribution can be used as an approximation to the Binomial distribution. Can this be justified in this case?

(d) Estimate the chance of less than 60 of these chips working after 1000 hours.



(a) The probability of failure must remain constant for the binomial distribution to be appropriate.

(b) $E(X) = np = 100 \times 0.3 = 30$

$Var(X) = npq = 100 \times 0.3 \times 0.7 = 21$

(c) LEARN

$$\boxed{\text{A binomial distribution } B(n,p) \text{ can be approximated by a normal distribution } N(np, npq) \text{ when } n \text{ is large and } p \text{ is not too small and provided } np > 10 \text{ and } nq > 10.}$$

In this case $np = 100 \times 0.3 = 30 > 10$

and $nq = 100 \times 0.7 = 70 > 10$ $\therefore$ Justified.

(d) Less than 60 means 59, 58, 57 ..... $\therefore$ Use $\underline{59.5}$

0 1 2 3 ... 59 | 60 61 ...
59.5

Mean $= np = 100 \times 0.7 = 70$

NOTE: We use 0.7 as we are considering the number working not failing

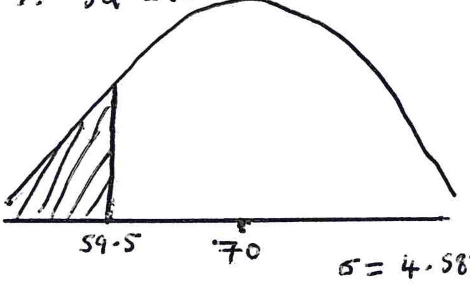
Variance $= npq = 100 \times 0.7 \times 0.3 = 21$

$\therefore$ sd $= \sqrt{21} = 4.583$

$\therefore z = \dfrac{70 - 59.5}{4.583} = 2.291$

$\therefore$ Table gives 98.90 %

$\therefore$ Answer is $\underline{1.1\ \%}$



$\sigma = 4.583$

THE MEAN OF A CONTINUOUS PROBABILITY DISTRIBUTION FUNCTION

For a continuous random variable with pdf $f(x)$ its expectation value, $E(X)$, is given by

$$E(X) = \int_{\text{all } x} xf(x)dx$$

The expectation value is also called the mean, μ.

EXAMPLE

Consider the continuous pdf $f(x) = -4x^3 + 4x, 0 \le x \le 1$. Find $E(X)$.

Solution

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 x(-4x^3 + 4x)dx = \int_0^1 (-4x^4 + 4x^2)dx$$

$$= \left[-\frac{4}{5}x^5 + \frac{4}{3}x^3\right]_0^1$$

$$= \frac{8}{15}$$

Note that this value falls near the middle of the domain of the pdf, as it should. If the pdf is symmetric, the expectation value is in the centre of the domain. If the pdf is asymmetric, the expectation value is not in the centre of the domain, but unless the pdf is very asymmetric, the expectation value should still be near the centre of the domain. If you get an expectation value not near the middle of the domain of the pdf, recheck your math!

THE EXPECTATION VALUE OF ANY FUNCTION

If $g(x)$ is a function of continuous random variable X with pdf $f(x)$ the expectation value of $g(X)$ can be calculated:

$$E[g(X)] = \int_{\text{all } x} g(x)f(x)dx$$

Take care to realise that $g(x)f(x)$ is simply the product of the two functions and not $g(f(x))$.

For our purposes the most important use of $E[g(X)]$ will be to calculate the variance of a continuous function. To do this we will need to use the case of $g(x) = x^2$.

EXAMPLE

Consider the continuous pdf $f(x) = -4x^3 + 4x, 0 \le x \le 1$. Find $E(X^3)$.

Solution

In this case we have that $g(x) = x^3$, therefore,

$$E(X^3) = \int_0^1 x^3 \times f(x)dx = \int_0^1 x^3(-4x^3 + 4x)dx = \int_0^1 (-4x^6 + 4x^4)dx$$

$$= \left[-\frac{4}{7}x^7 + \frac{4}{5}x^5\right]_0^1$$

$$= \frac{8}{35}$$

THE MODE & MEDIAN OF A CONTINUOUS PROBABILITY DISTRIBUTION FUNCTION

The **mode** is the value of X for which $f(x)$ is the greatest over the domain of $f(x)$.

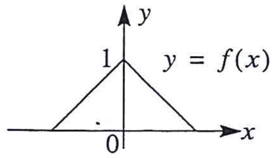
EXAMPLE

Find the mode for each of the following continuous pdfs

(a) $f(x) = -|x| + 1, -1 \leq x \leq 1$. (b) $f(x) = -4x^3 + 4x, 0 \leq x \leq 1$.

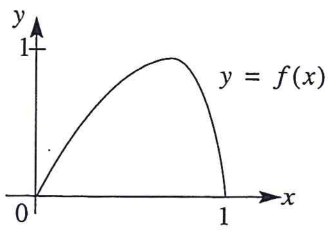
Solution

(a) A graph of the pdf can go a long way in helping determine the mode. From the graph of $f(x) = -|x| + 1, -1 \leq x \leq 1$ it is immediately obvious from symmetry as to where the greatest value of $f(x)$ occurs.



By inspection the mode is at $x = 0$.

(b) This time, sketching $f(x) = -4x^3 + 4x, 0 \leq x \leq 1$ will not immediately reveal where the greatest value of $f(x)$ occurs.



While we could use a graphics calculator to obtain an approximate value for where the greatest value of $f(x)$ occurs, we can obtain an exact value by making use of differential calculus.

$f(x) = -4x^3 + 4x \therefore f'(x) = -12x^2 + 4$

At the stationary point, $f'(x) = 0 \Leftrightarrow -12x^2 + 4 = 0 \Leftrightarrow x = \pm\sqrt{\frac{1}{3}} \approx \pm 0.5774$.

As $0 \leq x \leq 1$, the mode is 0.5774.

The **median** is the value of X for which $P(X \leq x) = 0.5$.

When the random variable X is continuous, this translates to the value of a of the pdf $f(x)$ for which $\int_{-\infty}^{a} f(x)dx = 0.5$.

EXAMPLE

Find the median for each of the following continuous pdfs

(a) $f(x) = -|x| + 1, -1 \leq x \leq 1$. (b) $f(x) = -4x^3 + 4x, 0 \leq x \leq 1$.

Solution

(a) By considering the graph of $f(x) = -|x| + 1, -1 \leq x \leq 1$, again, using the symmetry of the graph, we have that the median occurs at x = 0.

(b) We need to find the value of a for which $\int_{0}^{a}(-4x^3 + 4x)dx = 0.5$.

This gives $\left[-x^4 + 2x^2\right]_0^a = 0.5 \Leftrightarrow -a^4 + 2a^2 = 0.5 \Leftrightarrow 2a^4 - 4a^2 + 1 = 0$

We are now left with a quadratic in a^2. Using the quadratic formula we have:

$a^2 = \frac{4 \pm \sqrt{16-8}}{2 \times 2} = \frac{2 \pm \sqrt{2}}{2}$. From where we have $a^2 \approx 0.2929$ or $a^2 \approx 1.7071$.

Giving $a \approx \pm 0.5411$ or $a \approx \pm 1.3066$.

As $0 \leq x \leq 1$, then we have that the median is 0.5411.

THE VARIANCE OF A CONTINUOUS PROBABILITY DISTRIBUTION FUNCTION

For a continuous random variable with pdf $f(x)$, its variance, $\mathrm{Var}(X) = E[(X-\mu)^2]$, where $\mu = E(X)$. This formula is identical to the formula for the variance of a discrete random variable, but now we must integrate, so that

$$\mathrm{Var}(X) = \int_{\text{all } x} (x-\mu)^2 f(x)dx$$

Just like for discrete variable, there is a "calculation version", which is recommended for actual calculations:

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \int_{\text{all } x} x^2 f(x)dx - \mu^2$$

EXAMPLE

For the continuous pdf $f(x) = -4x^3+4x, 0 \le x \le 1$. Find $\mathrm{Var}(X)$.

Solution

We need to calculate μ^2. From example 1.15 we know that $\mu = \frac{8}{15} \therefore \mu^2 = \frac{64}{225}$.

We also need to calculate $E(X^2) = \int_0^1 x^2 \times f(x)dx = \int_0^1 x^2(-4x^3+4x)dx = \frac{1}{3}$

So $\mathrm{Var}(X) = \frac{1}{3} - \frac{64}{225} = \frac{11}{225} \approx 0.0489$

Remember: always try to conduct a "reality check" on your answer. What can we check here? Since σ must be real and since $\sigma = \sqrt{\mathrm{Var}(X)}$ then $\mathrm{Var}(X)$ must be nonnegative. Furthermore $\mathrm{Var}(X)$ cannot be zero if X is continuous variable. Why?

THE CUMULATIVE DISTRIBUTION FUNCTION OF A CONTINUOUS PDF

For **discrete** random variables the cdf is the **sum** of all the pdf values up to x.
For **continuous** random variables the cdf is the **integral** of all the pdf values up to x.

For a continuous random variable with pdf $f(x)$ its cdf, $F(x)$, is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

Technically so that we can integrate with respect to t, we had to convert $f(x)$ to $f(t)$, where t is a dummy variable, which disappears in the final result, leaving $F(x)$, as we want.

Formally we integrate from negative infinity, but in practice we only integrate from the smallest value of x in the domain of $f(x)$.

EXAMPLE

Consider the random variable X with pdf $f(x) = -4x^3+4x, 0 \le x \le 1$. Determine its cdf.

Solution

The cdf, $F(x)$, is given by $F(x) = \int_0^x (-4t^3 + 4t)\,dt$

$$= \left[-t^4 + 2t^2\right]_0^x$$

$$= -x^4 + 2x^2,\ 0 \le x \le 1$$

The reality check here is that $F(1) = 1$ (why?), which it does.

THE CONTINUOUS UNIFORM (RECTANGULAR) DISTRIBUTION, X~U(a,b)

Now we consider the simplest possible continuous cdf, the continuous **uniform distribution**. Thinking graphically this is always a rectangle with area = 1.

Formally we write X~U(a,b), where a and b are the upper and lower bounds of the domain of $f(x)$. Since the width of the rectangle is $(b - a)$ and since the area of the rectangle must be one, the height of the rectangle is $\frac{1}{b-a}$.

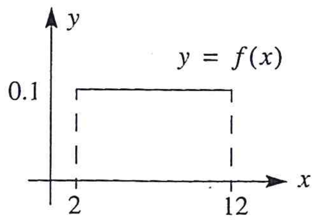
EXAMPLE

For the continuous pdf $f(x) = 0.1,\ 2 \le x \le 12$. Find

(a) $E(X)$ (b) $\mathrm{Var}(X)$ (c) its cdf, $F(x)$.

Solution

(a) Notice that the area under the curve is 1 as it must be. All uniform distributions are symmetric, so $E(X)$ is in the middle of the domain of $f(x)$, that is $E(X) = 7$.



However, let's evaluate the integral just to check.

$$E(X) = \int_2^{12} x \times 0.1\,dx = [0.05x^2]_2^{12} = 0.05 \times 140 = 7$$

(b) $$\mathrm{Var}(X) = \int_2^{12} x^2 f(x)\,dx - \mu^2 = \int_2^{12} 0.1x^2\,dx - 7^2$$

$$= \left[\frac{1}{30}x^3\right]_2^{12} - 49$$

$$= \frac{1720}{30} - 49$$

$$= \frac{25}{3}\ (\approx 8.333)$$

Approximately 70% of the data should be within about $\mu \pm \sigma$, that is, within 7 ± 2.89.

In our case that is $\frac{2.89 \times 2}{10} \times 100\% = 57\%$. Real world pdfs tend to be more peaked in the middle than a uniform distribution and thus closer to the 70% figure.

(c) $f(x) = 0.1,\ 2 \le x \le 12\ \therefore F(x) = \int_2^x 0.1\,dt,\ 0 \le x \le 12$

$$= 0.1(x - 2)$$

This formula should give $F(2) = 0$ and $F(12) = 1$, because 2 and 12 are the lower and upper bounds of the domain of $F(x)$. And indeed they do!

Probability generating functions

This chapter introduces the concept of a probability generating function. When you have completed it you should be able to

- understand the concept of a probability generating function and be able to construct and use the probability generating function for given distributions (including the discrete uniform, binomial, geometric and Poisson)
- use formulae for the mean and variance of a discrete random variable in terms of its probability generating function, and to use these formulae to calculate the mean and variance of probability distributions
- use the result that the probability generating function of the sum of independent variables is the product of the individual probability generating functions of those variables.

3.1 Defining a probability generating function

In certain cases it is convenient to summarise the values, x_i, taken by a discrete random variable, X, and the probabilities, p_i, associated with these values by a function known as the **probability generating function,** often abbreviated as p.g.f. This function, denoted by $\mathrm{G}_X(t)$, involves an arbitrary variable t. $\mathrm{G}_X(t)$ is defined as

$$\mathrm{G}_X(t) = p_1 t^{x_1} + p_2 t^{x_2} + \ldots = \sum p_i t^{x_i}. \qquad (3.1)$$

$\mathrm{G}_X(t)$ only exists if this series converges.

You saw in C2 and C4 that issues of convergence always arise in dealing with infinite series. In this chapter and the next, certain operations on series, such as rearrangement and term-by-term differentiation, are only justified when the series satisfies, often quite strict, convergence conditions. For the purpose of this Statistics module, although you should realise when your solutions depend on assumptions about convergence, you do not need to worry about the details. You can assume that, unless stated otherwise, all the necessary conditions hold.

Consider the probability generating function for a simple distribution. Suppose that the discrete random variable X has the probability distribution shown below.

x	1	4	7
$\mathrm{P}(X = x)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Then, in this case, the probability generating function of X is given by

$$\mathrm{G}_X(t) = \tfrac{1}{3}t^1 + \tfrac{1}{2}t^4 + \tfrac{1}{6}t^7.$$

At first sight this does not appear to be a very useful definition, especially since t does not have any obvious meaning. However, you will see in the course of this chapter that it provides a powerful tool for finding the mean and variance of certain probability distributions and also for finding the probability distribution of a sum of independent random variables.

Look again at the definition of $G_X(t)$ in Equation 3.1. You can see that $G_X(t)$ is constructed by multiplying each value of t^X by the associated probability and then summing. Thus $G_X(t)$ is the expected value of t^X; that is,

$$G_X(t) = E(t^X). \qquad (3.2)$$

Substituting $t = 1$ into Equation 3.2 gives

$$G_X(1) = E(1^X) = E(1) = 1.$$

You can see why this must be so by looking back to Equation 3.1. If you substitute $t = 1$ into this equation, you have

$$G_X(1) = p_1 1^{x_1} + p_2 1^{x_2} + \ldots = p_1 + p_2 + \ldots = \sum p_i.$$

Thus $G_X(1)$ is the sum of the probabilities and for a probability distribution this is equal to 1, that is

$$G_X(1) = \sum p_i = 1. \qquad (3.3)$$

Example 3.1.1

(a) The discrete random variable X is the number of throws taken to throw the first six with a fair dice. Find the probability generating function, $G_X(t)$, of X and verify that $G_X(1) = 1$.

(b) Generalise these results to the situation in which the probability of throwing a six is p.

(a) X has a geometric distribution for which

$$P(X = x) = \left(\tfrac{1}{6}\right) \times \left(\tfrac{5}{6}\right)^{x-1} \text{ for } x = 1, 2, 3, \ldots$$

Thus $G_X(t) = \left(\frac{1}{6}\right)t^1 + \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)t^2 + \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)^2 t^3 + \ldots$

This is an infinite geometric series (see C2 Section 6.3) with first term $\frac{1}{6}t$ and common ratio $\frac{5}{6}t$. Provided that $\left|\frac{5}{6}t\right| < 1$, the series can be summed, giving the probability generating function

$$G_X(t) = \frac{\frac{1}{6}t}{1 - \frac{5}{6}t} = \frac{t}{6 - 5t}.$$

Substituting $t = 1$ gives $G_X(1) = \dfrac{1}{6-5} = 1.$

The value $t = 1$ can be substituted into the expression for $G_X(t)$ since, for this value of t, $\left|\frac{5}{6}t\right| = \frac{5}{6}$, which is less than 1.

Example 3.1.3

The probability generating function of the discrete random variable X is $G_X(t) = k\left(\sqrt{t} + \frac{1}{\sqrt{t}}\right)^4$.

(a) Find the value of k.

(b) Find the probability distribution of X.

(a) Since $G_X(1) = 1$, it follows that $k\left(\sqrt{1} + \frac{1}{\sqrt{1}}\right)^4 = 16k = 1$. So $k = \frac{1}{16}$.

(b) The probability distribution of X is found by expanding $G_X(t)$ as a power series in t.

$$\begin{aligned} G_X(t) &= \tfrac{1}{16}\left(\sqrt{t} + \frac{1}{\sqrt{t}}\right)^4 \\ &= \tfrac{1}{16}\left((\sqrt{t})^4 + 4(\sqrt{t})^3\left(\frac{1}{\sqrt{t}}\right) + 6(\sqrt{t})^2\left(\frac{1}{\sqrt{t}}\right)^2 + 4(\sqrt{t})\left(\frac{1}{\sqrt{t}}\right)^3 + \left(\frac{1}{\sqrt{t}}\right)^4\right) \\ &= \tfrac{1}{16}(t^2 + 4t + 6 + 4t^{-1} + t^{-2}) \\ &= \tfrac{1}{16}t^{-2} + \tfrac{1}{4}t^{-1} + \tfrac{3}{8}t^0 + \tfrac{1}{4}t^1 + \tfrac{1}{16}t^2. \end{aligned}$$

The powers of t in this series give the values taken by X and the coefficients of the terms give the corresponding probabilities. Thus the probability distribution of X is

x	-2	-1	0	1	2
$P(X = x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Here is a summary of the results of this section.

The probability generating function of a discrete random variable, X, is defined by

$$G_X(t) = p_1t^{x_1} + p_2t^{x_2} + \ldots = E(t^X),$$

provided that this series converges.

$G_X(t)$ has the property that $G_X(1) = 1$.

Finding the mean and variance of a discrete random variable

In S1 and S2 you met formulae for the mean and variance of a number of discrete probability distributions, such as the binomial, geometric and Poisson distributions. These formulae were stated but not proved. With the aid of the appropriate probability generating function the proofs of these formulae are relatively straightforward.

The first step is to obtain general results relating the mean and variance of a random variable, X, to its probability generating function, $G_X(t)$.

Consider the definition of $G_X(t)$,

$$G_X(t) = p_1t^{x_1} + p_2t^{x_2} + \ldots$$

Now differentiate with respect to t. This gives

$$G'_X(t) = p_1x_1t^{x_1-1} + p_2x_2t^{x_2-1} + \ldots \qquad (3.4)$$

Such term-by-term differentiation of a series is not always a valid procedure.

Substituting $t = 1$ gives

$$G'_X(1) = p_1x_11^{x_1-1} + p_2x_21^{x_2-1} + \ldots$$

So, $G'_X(1) = p_1x_1 + p_2x_2 + \ldots = E(X)$ (assuming that this series converges).

Thus in order to find the expected value of a random variable it is only necessary to differentiate the probability generating function with respect to t and substitute $t = 1$.

Example 3.2.1
Find the expectation of a random variable, X, such that $X \sim \text{Geo}(p)$.

The probability generating function of a geometric distribution was found in Example 3.1.1. It is

$$\mathrm{G}_X(t) = \frac{pt}{1-qt}, \quad \text{provided that} \quad |qt| < 1.$$

Differentiating with respect to t gives

$$\mathrm{G}'_X(t) = \frac{p(1-qt) - (-q) \times pt}{(1-qt)^2} = \frac{p}{(1-qt)^2}.$$

As $q = 1 - p$ must be less than 1, $|qt|$ will be less than 1 when $t = 1$. Therefore $t = 1$ can be substituted into this expression for $\mathrm{G}'_X(t)$.

Thus $\mathrm{E}(X) = \mathrm{G}'_X(1) = \dfrac{p}{(1-q)^2} = \dfrac{p}{p^2} = \dfrac{1}{p}$.

This is a result which you used without proof in S1.

A similar method can be used to find a variance from a probability generating function. Consider Equation 3.4,

$$\mathrm{G}'_X(t) = p_1x_1t^{x_1-1} + p_2x_2t^{x_2-1} + \ldots$$

Differentiating again with respect to t gives

$$\mathrm{G}''_X(t) = p_1x_1(x_1-1)t^{x_1-2} + p_2x_2(x_2-1)t^{x_2-2} + \ldots,$$

and substituting $t = 1$ gives

$$\mathrm{G}''_X(1) = p_1x_1(x_1-1) + p_2x_2(x_2-1) + \ldots,$$

or, opening the brackets,

$$\begin{aligned}\mathrm{G}''_X(1) &= p_1(x_1^2 - x_1) + p_2(x_2^2 - x_2) + \ldots \\ &= (p_1x_1^2 + p_2x_2^2 + \ldots) - (p_1x_1 + p_2x_2 + \ldots) \\ &= \sum p_ix_i^2 - \mathrm{E}(X).\end{aligned}$$

So $\quad \mathrm{G}''_X(1) = \sum p_ix_i^2 - \mathrm{E}(X). \qquad (3.5)$

Algebraic manipulations on a series, like those above, are only legitimate if the series meets certain convergence conditions.

Now

$$\begin{aligned}\mathrm{Var}(X) &= \sum p_ix_i^2 - (\mathrm{E}(X))^2 \\ &= (\mathrm{G}''_X(1) + \mathrm{E}(X)) - (\mathrm{E}(X))^2 \qquad \text{(using Equation 3.5)} \\ &= \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - (\mathrm{G}'_X(1))^2.\end{aligned}$$

Example 3.2.2

Find the variance of a random variable, X, such that $X \sim \text{Geo}(p)$.

From Example 3.2.1,

$$G'_X(t) = \frac{p}{(1-qt)^2} \quad \text{and} \quad G'_X(1) = \frac{1}{p}.$$

Differentiating $G'_X(t)$ with respect to t gives $G''_X(t) = \dfrac{2qp}{(1-qt)^3}$.

So $$G''_X(1) = \frac{2qp}{(1-q)^3} = \frac{2qp}{p^3} = \frac{2q}{p^2}.$$

Then $\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$

$$= \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{2q+p-1}{p^2} = \frac{2q-q}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

The mean and variance of a discrete random variable, X, with probability generating function $G_X(t)$ are given by

$$E(X) = G'_X(1),$$

and

$$\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

Example 3.2.3

Find the probability generating function of the random variable X where $X \sim \text{Po}(\lambda)$. Hence find the mean and variance of X.

The Poisson probabilities are given by $P(X = x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$ for $x = 0, 1, 2, \ldots$

$$\begin{aligned} G_X(t) &= p_1 t^{x_1} + p_2 t^{x_2} + \cdots + p_n t^{x_n} + \ldots \\ &= e^{-\lambda}\lambda^0 t^0 + e^{-\lambda}\frac{\lambda}{1!}t^1 + e^{-\lambda}\frac{\lambda^2}{2!}t^2 + \cdots + e^{-\lambda}\frac{\lambda^n}{n!}t^n + \ldots \\ &= e^{-\lambda}\left(1 + \frac{\lambda t}{1!} + \frac{\lambda^2 t^2}{2!} + \cdots + \frac{\lambda^n t^n}{n!} + \ldots\right). \end{aligned}$$

The terms in the bracket give the Maclaurin expansion for $e^{\lambda t}$, so

$$G_X(t) = e^{-\lambda}e^{\lambda t} = e^{\lambda(t-1)}$$

for a Poisson distribution with parameter λ.

In order to find the mean and variance you need to calculate $G'_X(1)$ and $G''_X(1)$.

$$\begin{aligned} G'_X(t) &= \lambda e^{\lambda(t-1)}, \quad \text{so} \quad G'_X(1) = \lambda. \\ G''_X(t) &= \lambda^2 e^{\lambda(t-1)}, \quad \text{so} \quad G''_X(1) = \lambda^2. \end{aligned}$$

Thus $E(X) = G'_X(1) = \lambda$,

and

$$\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

You met this result for the variance in S2 Section 3.3, but it was not proved there.

Example 3.2.4

Find the probability generating function for the random variable X where $X \sim \mathrm{B}(n, p)$, and hence find the mean and variance of X.

The binomial probabilities are given by $\mathrm{P}(X = x) = \binom{n}{x} p^x q^{n-x}$, for $x = 0, 1, 2, \ldots, n$, where $q = 1 - p$.

$$\begin{aligned}
\mathrm{G}_X(t) &= p_1 t^{x_1} + p_2 t^{x_2} + \cdots + p_r t^{x_r} + \cdots + p_n t^{x_n} \\
&= \binom{n}{0} p^0 q^n t^0 + \binom{n}{1} p^1 q^{n-1} t^1 + \cdots + \binom{n}{r} p^r q^{n-r} t^r + \cdots + \binom{n}{n} p^n q^0 t^n \\
&= \binom{n}{0} (pt)^0 q^n + \binom{n}{1} (pt)^1 q^{n-1} + \cdots + \binom{n}{r} (pt)^r q^{n-r} + \cdots + \binom{n}{n} (pt)^n q^0.
\end{aligned}$$

But this series is just the binomial series for $(pt + q)^n$, so

$$\mathrm{G}_X(t) = (pt + q)^n.$$

Now find the mean and variance of X.

Since $\mathrm{G}'_X(t) = np(pt + q)^{n-1}$ and $\mathrm{G}''_X(t) = n(n-1)p^2(pt + q)^{n-2}$, putting $t = 1$ and recalling that $p + q = 1$ gives

$$\mathrm{G}'_X(1) = np(p + q)^{n-1} = np$$

and

$$\mathrm{G}''_X(1) = n(n-1)p^2(p + q)^{n-1} = n(n-1)p^2.$$

So $\quad \mathrm{E}(X) = \mathrm{G}'_X(1) = np,$

and
$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - (\mathrm{G}'_X(1))^2 \\
&= n(n-1)p^2 + np - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 = np(1 - p) = npq.
\end{aligned}$$

Again these are results which you have met before without proof, this time in S1 Section 8.3.

Example 3.2.5

Write down the probability generating function of the discrete random variable X for which $\mathrm{P}(X = x) = \frac{1}{5}$ for $x = 1, 2, 3, 4, 5$. Hence find the mean and variance of X.

The probability generating function is

$$\begin{aligned}
\mathrm{G}_X(t) &= p_1 t^{x_1} + p_2 t^{x_2} + \cdots + p_n t^{x_n} \\
&= \tfrac{1}{5}t^1 + \tfrac{1}{5}t^2 + \tfrac{1}{5}t^3 + \tfrac{1}{5}t^4 + \tfrac{1}{5}t^5.
\end{aligned}$$

This is a geometric series which can be summed to give $\frac{t}{5}\left(\frac{1 - t^5}{1 - t}\right)$. However, this result is only valid if $t \neq 1$. Since $t = 1$ is just the value which is needed to find the mean and variance from the probability generating function it is not appropriate to use this summation here. Instead the mean and variance must be calculated using the original form of the probability generating function.

The original function is differentiated to give

$$\mathrm{G}'_X(t) = \tfrac{1}{5} + \tfrac{2}{5}t + \tfrac{3}{5}t^2 + \tfrac{4}{5}t^3 + t^4,$$

and
$$\mathrm{G}''_X(t) = \tfrac{2}{5} + \tfrac{6}{5}t + \tfrac{12}{5}t^2 + 4t^3.$$

Substituting $t = 1$ now gives $\mathrm{G}'_X(1) = 3$ and $\mathrm{G}''_X(1) = 8$.

So $\quad \mathrm{E}(X) = \mathrm{G}'_X(1) = 3,$

and $\quad \mathrm{Var}(X) = \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - (\mathrm{G}'_X(1))^2 = 8 + 3 - 3^2 = 2.$

The method used in Example 3.2.5 to find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ from the probability generating function is not particularly elegant because the summed form of the probability generating function cannot be used as it was for the geometric, Poisson and binomial distributions.

3.3 The probability generating function of the sum of random variables

The probability generating function has the following useful property.

> The probability generating function of the *sum* of independent random variables is equal to the *product* of their individual probability generating functions.

In order to understand this property, first consider the two independent discrete random variables, X and Y, with the following probability distributions.

x	0	2
$P(X = x)$	0.6	0.4

y	0	1	2
$P(Y = y)$	0.3	0.5	0.2

Suppose you wished to find $P(X + Y = 2)$. This probability is given by

$$\begin{aligned} P(X+Y=2) &= P(X=0, Y=2) + P(X=2, Y=0) \\ &= P(X=0)P(Y=2) + P(X=2)P(Y=0) \quad \text{since } X \text{ and } Y \text{ are independent} \\ &= 0.6 \times 0.2 + 0.4 \times 0.3 \\ &= 0.24. \end{aligned}$$

Now consider finding $P(X + Y = 2)$ from the product of the probability generating functions. The probability generating functions of X and Y are

$$G_X(t) = 0.6 + 0.4t^2 \quad \text{and} \quad G_Y(t) = 0.3 + 0.5t + 0.2t^2$$

and the product of these two probability generating functions is

$$\begin{aligned} G_X(t)G_Y(t) &= (0.6 + 0.4t^2)(0.3 + 0.5t + 0.2t^2) \\ &= 0.6 \times 0.3 + 0.6 \times 0.5t + 0.6 \times 0.2t^2 \\ &\quad + 0.4t^2 \times 0.3 + 0.4t^2 \times 0.5t + 0.4t^2 \times 0.2t^2. \end{aligned}$$

There are two terms in the product that contain t^2, and these are the terms for which $X + Y = 2$. The coefficient of the first of these terms is $0.6 \times 0.2 = 0.12$, which is the value of $P(X = 0)P(Y = 2)$; the coefficient of the second is $0.4 \times 0.3 = 0.12$ which is the value of $P(X = 2)P(Y = 0)$. When these two terms are combined the probabilities are summed to give a coefficient of 0.24, which is the value of $P(X + Y = 2)$.

This reasoning may be generalised as follows. Consider the product of two probability generating functions for two independent random variables X and Y,

$$G_X(t)G_Y(t) = \left(p_{x_1}t^{x_1} + p_{x_2}t^{x_2} + \cdots + p_{x_n}t^{x_n} + \ldots\right)\left(p_{y_1}t^{y_1} + p_{y_2}t^{y_2} + \cdots + p_{y_n}t^{y_n} + \ldots\right),$$

where the probability generating functions have been written in series form with $p_{x_i} = P(X = x_i)$ and $p_{y_i} = P(Y = y_i)$.

Consider the term in t^r in the expansion of this product.

Its coefficient will be the sum of terms like $p_{x_i}p_{y_j}$ such that $x_i + y_j = r$.

This coefficient is just what you would calculate if you wanted to find $P(X + Y = r)$.

For each pair of values x_i and y_j such that $x_i + y_j = r$ you would multiply the corresponding probabilities to obtain $p_{x_i}p_{y_j}$ (since X and Y are independent) and then sum these products (since the different pairs are mutually exclusive).

Thus the coefficient of t^r gives $P(X + Y = r)$, so the product of the probability generating functions must be the probability generating function of $X + Y$. This argument can be extended to more than two variables.

The same result can be arrived at more succinctly by using expectation notation. If two variables X and Y are independent, then $E(X)E(Y) = E(XY)$. (This result is proved in Section 6.7.) So $G_X(t)G_Y(t) = E(t^X)E(t^Y) = E(t^X t^Y) = E(t^{X+Y}) = G_{X+Y}(t)$.

This property of probability generating functions gives a powerful method for finding the probability distribution of the sum of independent random variables. In particular it provides a means of proving the results $E(X + Y) = E(X) + E(Y)$ and $Var(X + Y) = Var(X) + Var(Y)$ for independent discrete random variables.

Theorem $E(X+Y) = E(X) + E(Y)$

Proof Let $V = X + Y$.

Then $G_V(t) = G_X(t)G_Y(t)$, and $G'_V(t) = G'_X(t)G_Y(t) + G'_Y(t)G_X(t)$.

So $E(V) = G'_V(1) = G'_X(1)G_Y(1) + G'_Y(1)G_X(1)$.

Recall from Equation 3.3 that $G_X(1) = G_Y(1) = 1$, and substitute these values into the expression for $E(V)$. Then

$$E(V) = G'_X(1) + G'_Y(1) = E(X) + E(Y),$$

so

$$E(X+Y) = E(X) + E(Y).$$

Theorem $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof By differentiating $G'_V(t) = G'_X(t)G_Y(t) + G'_Y(t)G_X(t)$ again,

$$G''_V(t) = G''_X(t)G_Y(t) + 2G'_X(t)G'_Y(t) + G''_Y(t)G_X(t).$$

Evaluating $G''_V(t)$ at $t = 1$ gives

$$\begin{aligned} G''_V(1) &= G''_X(1)G_Y(1) + 2G'_X(1)G'_Y(1) + G''_Y(1)G_X(1) \\ &= G''_X(1) + 2G'_X(1)G'_Y(1) + G''_Y(1). \end{aligned}$$

Then
$$\begin{aligned} \mathrm{Var}(V) &= G''_V(1) + G'_V(1) - (G'_V(1))^2 \\ &= \left(G''_X(1) + 2G'_X(1)G'_Y(1) + G''_Y(1)\right) + \left(G'_X(1) + G'_Y(1)\right) \\ &\quad - \left(G'_X(1) + G'_Y(1)\right)^2 \\ &= G''_X(1) + 2G'_X(1)G'_Y(1) + G''_Y(1) + G'_X(1) + G'_Y(1) \\ &\quad - (G'_X(1))^2 - 2G'_X(1)G'_Y(1) - (G'_Y(1))^2 \\ &= G''_X(1) + G'_X(1) - (G'_X(1))^2 + G''_Y(1) + G'_Y(1) - (G'_Y(1))^2 \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y), \end{aligned}$$

so
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Recall from S3 Section 2.2 that the result $E(X+Y) = E(X) + E(Y)$ is also true for random variables which are not independent, but this result cannot be proved by the method above.

The following example shows how the probability distribution of a sum of random variables can be deduced from the product of the probability generating functions.

Example 3.3.1
Find the probability generating function for the sum, S, of three throws on a fair dice. Hence find $P(S = s)$ for $s = 3, 4, 5$ and 6.

Let X be the result of one throw of the dice. Then X has a uniform distribution with $P(X = x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$.

The probability generating function of X is $G_X(t) = \frac{1}{6}t^1 + \frac{1}{6}t^2 + \cdots + \frac{1}{6}t^6$.

This is a geometric series with first term $\frac{1}{6}t$ and common ratio t. Provided that $t \neq 1$ its sum is given by $\dfrac{t\,(1-t^6)}{6\,(1-t)}$.

The summed form of the probability generating function can be used in this example since it will not be necessary to put $t = 1$.

For S, the sum of three throws on the dice, the probability generating function is found by multiplying the probability generating functions for the individual throws, so

$$\begin{aligned} G_S(t) &= \frac{t\,(1-t^6)}{6\,(1-t)} \times \frac{t\,(1-t^6)}{6\,(1-t)} \times \frac{t\,(1-t^6)}{6\,(1-t)} = \frac{1}{216}t^3\frac{(1-t^6)^3}{(1-t)^3} \\ &= \frac{1}{216}t^3(1 - 3t^6 + 3t^{12} - t^{18})(1 + 3t + 6t^2 + 10t^3 + \ldots) \\ &= \frac{1}{216}(t^3 + 3t^4 + 6t^5 + 10t^6 + \ldots). \end{aligned}$$

The required probabilities are found from the coefficients of the corresponding powers of t. For example, $P(S = 3)$ is given by the power of t^3. So

$$P(S = 3) = \tfrac{1}{216},$$
$$P(S = 4) = \tfrac{3}{216} = \tfrac{1}{72},$$
$$P(S = 5) = \tfrac{6}{216} = \tfrac{1}{36},$$
$$P(S = 6) = \tfrac{10}{216} = \tfrac{5}{108}.$$

If two variables have the same distribution, then they also have the same probability generating function. The reverse is also true. If two variables have the same probability generating function, then they also have the same distribution.

Example 3.3.2
Two random variables X and Y are such that $X \sim B(5, \frac{1}{3})$ and $Y \sim B(10, \frac{1}{3})$.

(a) Write down the probability generating functions of X, Y, and $X + Y$.

(b) Deduce the distribution of $X + Y$.

(a) From Example 3.2.4, the probability generating function of a random variable which is distributed as $B(n, p)$ is $(pt + q)^n$.

Thus the probability generating function of X is $\left(\frac{1}{3}t + \frac{2}{3}\right)^5$ and the probability generating function of Y is $\left(\frac{1}{3}t + \frac{2}{3}\right)^{10}$.

The probability generating function of $X + Y$ is given by the product of these probability generating functions and is thus equal to $\left(\frac{1}{3}t + \frac{2}{3}\right)^5\left(\frac{1}{3}t + \frac{2}{3}\right)^{10} = \left(\frac{1}{3}t + \frac{2}{3}\right)^{15}$.

(b) $X + Y$ has the probability generating function of a variable which has a $B(15, \frac{1}{3})$ distribution. Thus $X + Y \sim B(15, \frac{1}{3})$.

Example 3.3.3*
Show that the probability generating function of X where $X \sim B(n, p)$ tends to $e^{\lambda(t-1)}$, where λ is constant and equal to np, as n tends to infinity. Deduce the distribution of X as n tends to infinity.

The probability generating function of X is $G_X(t) = (pt + q)^n = (1 + p(t - 1))^n$.

Substituting $p = \dfrac{\lambda}{n}$ gives

$$G_X(t) = \left(1 + \frac{\lambda(t-1)}{n}\right)^n.$$

Using the result that $\left(1 + \dfrac{x}{n}\right)^n$ tends to e^x as x tends to infinity, you can see that $G_X(t)$ tends to $e^{\lambda(t-1)}$ as n tends to infinity.

But this is the probability generating function of a random variable which has a $Po(\lambda)$ distribution. Thus the distribution of X tends to the distribution $Po(\lambda)$, where λ is constant and equal to np, as n tends to infinity.

ANOTHER SET OF NOTES FOR

PROBABILITY GENERATING FUNCTIONS

In this section we consider discrete random variables which take values in $\mathbb{N} = \{0, 1, 2, ...\}$.

First we state some key results required for this section:

1 Finite geometric series (GS)

For $x \in \mathbb{R}$, $x \neq 1$, and $n \in \mathbb{N}$, $\sum\limits_{i=0}^{n} x^i = 1 + x + x^2 + ... + x^n = \dfrac{1 - x^{n+1}}{1 - x}$.

2 Sum of an infinite geometric series (GS)

The infinite sum $\sum\limits_{i=0}^{\infty} x^i = 1 + x + x^2 + x^3 + ...$ is finite (or convergent) if and only if $|x| < 1$.

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x} \quad \text{if and only if} \quad |x| < 1.$$

3 Binomial formula

For real constants x, y and for $n \in \mathbb{N}$:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} = y^n + \binom{n}{1} x y^{n-1} + \binom{n}{2} x^2 y^{n-2} + ... + \binom{n}{n-1} x^{n-1} y + x^n$$

4 Exponential series

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \quad \text{for all } x \in \mathbb{R}$$

{Result from the Calculus Option}

5 $\lim\limits_{n \to \infty} \left(1 + \dfrac{a}{n}\right)^n = e^a$ for all $a \in \mathbb{R}$ {Result from HL Core}

6 Binomial series for $r \in \mathbb{Z}^+$ and $|x| < 1$

$$\frac{1}{(1-x)^r} = \sum_{i=0}^{\infty} x^i (-1)^i \frac{(-r)(-r-1)....(-r-(i-1))}{i!}$$

{Result from the Calculus Option}

7 Summation identities

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}$$

PROBABILITY GENERATING FUNCTIONS

Let X be a discrete random variable which takes values in $\mathbb{N} = \{0, 1, 2, 3, ...\}$, and such that $\text{P}(X = k) = p_k$, for $k \in \mathbb{N}$.

The **probability generating function** (PGF), $G(t)$, for X is

$$G(t) = \text{E}(t^X) = \sum_{k=0}^{\infty} p_k t^k$$

$$= p_0 + p_1 t + p_2 t^2 + ... \quad \text{for all values of } t \text{ for which } G(t) \text{ is finite.}$$

We note that:

- $0 \leqslant p_k \leqslant 1$ and $\sum_{k=0}^{\infty} p_k = 1$ by the definition of the (well-defined) probability mass function for a discrete random variable X.
- The PGF $G(t)$ is either a finite series or an infinite series.
  - If $G(t)$ is a finite series then $G(t)$ is defined for all $t \in \mathbb{R}$.
  - If $G(t)$ is an infinite series, then it is a power series and is therefore finite (or convergent) only for t in the interval of convergence of the power series.
- A PGF $G(t)$ defines a discrete random variable X (and its probability distribution) uniquely. Conversely, if X is a discrete random variable which takes values in $\mathbb{N}$, then its PGF $G(t)$ is unique.

Example 27

a Let X be the discrete random variable which takes values 1, 2, 3, and 6, each with probability $\frac{1}{4}$. Find the PGF for X.

b Let $X \sim \mathrm{B}(1, \frac{1}{6})$ be the Bernoulli random variable equal to the number of '6's obtained when an unbiased 6-sided die is rolled once. Find the PGF for X.

c Let $X \sim \mathrm{Geo}(\frac{1}{6})$ be the geometric random variable equal to the number of rolls of an unbiased 6-sided die required to roll a '6'. Find the PGF for X.

---

a $G(t) = p_1 t^1 + p_2 t^2 + p_3 t^3 + p_6 t^6 = \frac{1}{4}(t + t^2 + t^3 + t^6)$

Since $G(t)$ is a finite series, $G(t)$ is finite and therefore defined for all $t \in \mathbb{R}$.

b $\mathrm{P}(X = 0) = p_0 = \frac{5}{6}$

$\mathrm{P}(X = 1) = p_1 = \frac{1}{6}$

$\mathrm{P}(X = k) = 0$ for integers $k \geqslant 2$.

$\therefore$ the PGF for X is $G(t) = p_0 + p_1 t + p_2 t^2 + \ldots$

$= \frac{5}{6} + \frac{1}{6}t$ for $t \in \mathbb{R}$.

c X takes values 1, 2, 3, ... and $\mathrm{P}(X = k) = \frac{1}{6}\left(1 - \frac{1}{6}\right)^{k-1}$ for $k = 1, 2, 3, \ldots$

$= \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}$

$\therefore$ the PGF for X is

$$\begin{aligned}
G(t) &= \sum_{k=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} t^k \\
&= \frac{t}{6}\sum_{k=1}^{\infty}\left(\frac{5t}{6}\right)^{k-1} \\
&= \frac{t}{6}\sum_{i=0}^{\infty}\left(\frac{5t}{6}\right)^{i} && \{\text{Infinite GS}\} \\
&= \frac{t}{6} \times \frac{1}{1 - \frac{5t}{6}} && \text{if and only if } \left|\frac{5t}{6}\right| < 1 \\
&= \frac{t}{6}\left(\frac{6}{6 - 5t}\right) && \text{if and only if } |t| < \frac{6}{5} \\
&= \frac{t}{6 - 5t} && \text{for } t \in \left]-\tfrac{6}{5}, \tfrac{6}{5}\right[.
\end{aligned}$$

Example 28

a Let X be the discrete random variable equal to the outcome of rolling an unbiased 4-sided (tetrahedral) die labelled 1, 2, 3, 4.

i Show that $X \sim \text{DU}(4)$.

ii Find the PGF for X.

b Derive the PGF for the discrete uniform random variable $X \sim \text{DU}(n)$ which takes the values $x = 1, 2, ..., n$.

---

a i X has probability distribution:

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

$\therefore \ X \sim \text{DU}(4)$

ii
$$\begin{aligned} G(t) &= p_0 + p_1 t + p_2 t^2 + ... \\ &= p_1 t^1 + p_2 t^2 + p_3 t^3 + p_4 t^4 \\ &= \tfrac{1}{4}(t + t^2 + t^3 + t^4) \\ &= \frac{t}{4}(1 + t + t^2 + t^3) \qquad \{\text{Finite GS}\} \\ &= \frac{t}{4}\frac{(t^4 - 1)}{(t - 1)}, \quad t \in \mathbb{R} \end{aligned}$$

b Suppose $X \sim \text{DU}(n)$, $n \in \mathbb{Z}^+$.

$\therefore \ P(X = x) = p_x = \dfrac{1}{n}$ for $x = 1, 2, 3, ..., n$.

$$\begin{aligned} \therefore \ G(t) &= p_0 + p_1 t + p_2 t^2 + ... \\ &= p_1 t^1 + p_2 t^2 + ... + p_n t^n \\ &= \frac{1}{n}(t + t^2 + ... + t^n) \\ &= \frac{t}{n}(1 + t + t^2 + ... + t^{n-1}) \qquad \{\text{Finite GS}\} \\ &= \frac{t}{n}\frac{(t^n - 1)}{(t - 1)}, \quad t \in \mathbb{R}. \end{aligned}$$

PROBABILITY GENERATING FUNCTIONS FOR IMPORTANT DISTRIBUTIONS

In **Example 28** we found that for $X \sim \text{DU}(n)$, $n \in \mathbb{Z}^+$, which takes values $k = 1, 2, ..., n$, the PGF for X is
$$\begin{aligned} G(t) &= \frac{1}{n}(t + t^2 + ... + t^n) \\ &= \frac{t}{n}\frac{(t^n - 1)}{(t - 1)} \quad \text{for all } t \in \mathbb{R}. \end{aligned}$$

The PGFs for the important distributions we study in this course are summarised in the following table:

Distribution	Notation	Probability mass function $P(x)$	Probability generating function $G(t)$
Discrete uniform	$X \sim \mathrm{DU}(n)$	$\frac{1}{n}$ for $x = 1, 2, 3, ..., n$	$\frac{t}{n}\frac{(t^n-1)}{t-1}, \ t \in \mathbb{R}$
Bernoulli	$X \sim \mathrm{B}(1, p)$, $0 < p < 1$	$p^x(1-p)^{1-x}$ for $x = 0, 1$	$1 - p + pt, \ t \in \mathbb{R}$
Poisson	$X \sim \mathrm{Po}(m)$	$\frac{m^x e^{-m}}{x!}$ for $x = 0, 1, 2, ...$	$e^{m(t-1)}, \ t \in \mathbb{R}$
Binomial	$X \sim \mathrm{B}(n, p)$, $0 < p < 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, 3, ..., n$	$(1 - p + pt)^n, \ t \in \mathbb{R}$
Geometric	$X \sim \mathrm{Geo}(p)$, $0 < p < 1$	$p(1-p)^{x-1}$ for $x = 1, 2, 3, 4, ...$	$\frac{pt}{1 - t(1-p)}, \ \lvert t \rvert < \frac{1}{1-p}$
Negative binomial	$X \sim \mathrm{NB}(r, p)$, $r \in \mathbb{Z}^+, \ 0 < p < 1$	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$	$\left(\frac{pt}{1 - t(1-p)}\right)^r, \ \lvert t \rvert < \frac{1}{1-p}$

Example 29

Let $X \sim \mathrm{Geo}(p), \ 0 < p < 1$ be a geometric random variable which takes values $x = 1, 2, 3, ...$ with probabilities $\mathrm{P}(X = x) = p(1-p)^{x-1}$.

Prove that the PGF of X is $G(t) = \frac{pt}{1 - t(1-p)}$ for $|t| < \frac{1}{1-p}$.

$$G(t) = \sum_{x=1}^{\infty} \mathrm{P}(X = x) t^x$$

$$= \sum_{x=1}^{\infty} p(1-p)^{x-1} t^x$$

$$= \frac{p}{1-p} \sum_{x=1}^{\infty} [t(1-p)]^x \quad \text{which is a GS with } u_1 = t(1-p) \text{ and } r = t(1-p)$$

$$= \frac{p}{1-p} \times \frac{t(1-p)}{1 - t(1-p)} \quad \text{provided } |t(1-p)| < 1$$

$$= \frac{pt}{1 - t(1-p)} \quad \text{provided } |t| < \frac{1}{1-p}$$

Example 30

Let $X \sim \mathrm{NB}(r, p), \ r \in \mathbb{Z}^+, \ 0 < p < 1$ be a negative binomial random variable which takes values $x = r, r+1, r+2, ...$ with probabilities $\mathrm{P}(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$.

Given that $\sum_{i=0}^{\infty} \binom{r-1+i}{i} [t(1-p)]^i = \frac{1}{(1 - t(1-p))^r}$ provided $|t(1-p)| < 1$, find the PGF for X.

The PGF for X is

$$\begin{aligned}
G(t) &= \sum_{x=r}^{\infty} \mathrm{P}(X=x)t^x \\
&= \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r(1-p)^{x-r}t^x \\
&= (pt)^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} (1-p)^{x-r}t^{x-r} \\
&= (pt)^r \sum_{i=0}^{\infty} \binom{r-1+i}{r-1} (1-p)^i t^i && \{\text{letting } i = x-r\} \\
&= (pt)^r \sum_{i=0}^{\infty} \binom{r-1+i}{i} [t(1-p)]^i && \left\{\text{since } \binom{r-1+i}{r-1} = \binom{r-1+i}{i}\right\} \\
&= (pt)^r \frac{1}{(1-t(1-p))^r} && \text{provided } |t(1-p)| < 1 \quad \{\text{given result}\} \\
&= \left[\frac{pt}{1-t(1-p)}\right]^r && \text{provided } |t| < \frac{1}{1-p}
\end{aligned}$$

MEAN AND VARIANCE

Theorem 11

Let X be a discrete random variable with values in $\mathbb{N}$ and with PGF $G(t) = p_0 + p_1 t + p_2 t^2 + \ldots.$

Then:
1 $G(1) = 1$

2 $\mathrm{E}(X) = G'(1)$

3 $G''(1) = \mathrm{E}(X(X-1)) = \mathrm{E}(X^2) - \mathrm{E}(X)$

4 $\mathrm{Var}(X) = G''(1) + G'(1) - [G'(1)]^2$

Proof:

1 $G(1) = p_0 + p_1 + p_2 + \ldots = \sum p_i = 1$ by definition of the probability mass function for X.

2

$$\begin{aligned}
G(t) &= \sum_{k=0}^{\infty} p_k t^k \\
\therefore \quad G'(t) &= \sum_{k=0}^{\infty} k p_k t^{k-1} \\
\therefore \quad G'(1) &= \sum_{k=0}^{\infty} k p_k \\
&= \sum_{k=0}^{\infty} k\,\mathrm{P}(X=k) \\
&= \mathrm{E}(X) \quad \text{by definition of } \mathrm{E}(X).
\end{aligned}$$

3

$$\begin{aligned}
G''(t) &= \sum_{k=0}^{\infty} k(k-1) p_k t^{k-2} \\
\therefore \quad G''(1) &= \sum_{k=0}^{\infty} k(k-1) p_k \\
&= \sum_{k=0}^{\infty} k(k-1)\,\mathrm{P}(X=k) \\
&= \mathrm{E}(X(X-1)) && \{\text{by } \textbf{Theorem 2} \text{ with } g(X) = X(X-1)\} \\
&= \mathrm{E}(X^2 - X) \\
&= \mathrm{E}(X^2) - \mathrm{E}(X) && \{\text{by Corollary to } \textbf{Theorem 2}\}
\end{aligned}$$

4

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 && \{\textbf{Theorem 3}\} \\
&= \mathrm{E}(X^2) - \mathrm{E}(X) + \mathrm{E}(X) - \{\mathrm{E}(X)\}^2 \\
&= G''(1) + G'(1) - \{G'(1)\}^2
\end{aligned}$$

Example 31

Let $X \sim \mathrm{DU}(n)$ with values $1, 2, ..., n$. X has PGF $G(t) = \frac{1}{n}(t + t^2 + ... + t^n)$.

Use $G(t)$ and differentiation rules to prove that $\mathrm{E}(X) = \frac{n+1}{2}$ and $\mathrm{Var}(X) = \frac{n^2-1}{12}$.

$$G(t) = \frac{1}{n}(t + t^2 + ... + t^n)$$

$$G'(t) = \frac{1}{n}(1 + 2t + 3t^2 + ... + nt^{n-1})$$

$$\therefore \quad G'(1) = \frac{1}{n}(1 + 2 + 3 + ... + n)$$

$$= \frac{1}{n}\left(\frac{n\,(n+1)}{2}\right)$$

$$= \frac{n+1}{2}$$

$$\therefore \quad E(X) = G'(t) = \frac{n+1}{2}$$

$$G''(t) = \frac{1}{n}(2 + 3 \times 2t + 4 \times 3t^2 + ... + n(n-1)t^{n-2})$$

$$\therefore \quad G''(1) = \frac{1}{n}(2 + 6 + 12 + ... + n(n-1))$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n-1} i(i+1)\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n-1} i^2 + \sum_{i=1}^{n-1} i\right\}$$

$$= \frac{1}{n}\left\{\frac{n(n-1)(2n-1)}{6} + \frac{n(n-1)}{2}\right\} \qquad \{\text{well known summation identities}\}$$

$$= \frac{n(n-1)}{n}\left\{\frac{2n-1}{6} + \frac{1}{2}\right\}$$

$$= (n-1)\left\{\frac{2n+2}{6}\right\}$$

$$= \frac{(n-1)(n+1)}{3}$$

$$\therefore \quad \mathrm{Var}(X) = G''(1) + G'(1) - \{G'(1)\}^2$$

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

$$= \frac{(n+1)(4(n-1) + 6 - 3(n+1))}{12}$$

$$= \frac{(n+1)(n-1)}{12}$$

$$= \frac{n^2-1}{12}$$

The mean and variance for some important discrete random variables are summarised below:

Distribution			$\mathrm{E}(X)$	$\mathrm{Var}(X)$
Discrete uniform	$X \sim \mathrm{DU}(n)$	with values $1, 2, ..., n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Binomial	$X \sim \mathrm{B}(n, p)$	with values $0, 1, 2, ..., n$	np	$np(1-p)$
Geometric	$X \sim \mathrm{Geo}(p)$	with values $1, 2, 3, ...$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$X \sim \mathrm{Po}(m)$	with values $0, 1, 2, ...$	m	m
Negative binomial	$X \sim \mathrm{NB}(r, p)$	with values $r, r+1, r+2, ...$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

THE SUM OF INDEPENDENT RANDOM VARIABLES

Theorem 12

Let X and Y be two discrete random variables with values in $\mathbb{N}$ and with probability generating functions $G_X(t)$ and $G_Y(t)$ respectively.
If X and Y are **independent** then the random variable $X+Y$ has probability generating function $G_{X+Y}(t)$ where $G_{X+Y}(t) = G_X(t)G_Y(t)$.

Proof:

Suppose $G_X(t) = p_0 + p_1 t + p_2 t^2 + ...$
and $G_Y(t) = q_0 + q_1 t + q_2 t^2 + ...$

Let $U = X + Y$ be the random variable equal to the sum of the values of X and Y.

$$\begin{aligned}\text{Now}\quad \mathrm{P}(U=r) &= \sum_{k=0}^{r} \mathrm{P}(X=k \text{ and } Y=r-k)\\ &= \sum_{k=0}^{r} \mathrm{P}(X=k)\,\mathrm{P}(Y=r-k) \qquad \{\text{since } X, Y \text{ are independent}\}\\ &= \sum_{k=0}^{r} p_k q_{r-k}\end{aligned}$$

$$\begin{aligned}\therefore\quad G_{X+Y}(t) &= \sum_{r=0}^{\infty}\left(\sum_{k=0}^{r} p_k q_{r-k}\right)t^r\\ &= p_0q_0 + (p_0q_1 + p_1q_0)t + ... + (p_0q_r + p_1q_{r-1} + ... + p_rq_0)t^r + ...\end{aligned}$$

Now consider the product

$$G_X(t)G_Y(t) = (p_0 + p_1 t + p_2 t^2 + ...)(q_0 + q_1 t + q_2 t^2 + ...).$$

By multiplying and collecting like terms we obtain the same function

$$p_0q_0 + (p_0q_1 + p_1q_0)t + ... + (p_0q_r + p_1q_{r-1} + ... + p_rq_0)t^r + ... = G_{X+Y}(t)$$

Corollary:

Suppose $X_1, X_2, ..., X_n$ are independent discrete random variables with values in $\mathbb{N}$ and probability generating functions $G_{X_1}(t), G_{X_2}(t), ..., G_{X_n}(t)$ respectively. The random variable $U = X_1 + X_2 + ... + X_n$ has probability generating function $G_U(t) = G_{X_1}(t)G_{X_2}(t) .... G_{X_n}(t)$.

Proof:

By the above theorem,

$$\begin{aligned}&G_{X_1+X_2+....+X_n}(t)\\ &= G_{X_1+....+X_{n-1}}(t)G_{X_n}(t) && \{\text{letting } X = X_1 + ... + X_{n-1} \text{ and } Y = X_n\}\\ &= (G_{X_1+....+X_{n-2}}(t)G_{X_{n-1}}(t))G_{X_n}(t) && \{\text{letting } X = X_1 + ... + X_{n-1} \text{ and } Y = X_{n-1}\}\\ &\ \vdots\\ &= G_{X_1}(t)G_{X_2}(t) .... G_{X_n}(t) \quad \text{as required.}\end{aligned}$$

Example 32

Consider a binomial random variable $X \sim \mathrm{B}(n, p)$ for constants $n \in \mathbb{Z}^+$, $0 < p < 1$.

Let $X = Y_1 + Y_2 + ... + Y_n$ where $Y_1, ..., Y_n$ are n independent Bernoulli random variables $Y_i \sim \mathrm{B}(1, p)$, $i = 1, ..., n$.

Find the PGF for X in terms of the PGFs for $Y_1, ..., Y_n$.

For each Bernoulli random variable $Y_i \sim \mathrm{B}(1, p)$,
$G_{Y_i}(t) = 1 - p + pt$ and $X = Y_1 + Y_2 + ... + Y_n$

$$\begin{aligned} \therefore \quad G_X(t) &= G_{Y_1}(t)G_{Y_2}(t) \dots . G_{Y_n}(t) \qquad \{\textbf{Theorem 12}\} \\ &= (1 - p + pt)(1 - p + pt) \dots . (1 - p + pt) \\ &= (1 - p + pt)^n \end{aligned}$$

Theorem 13

If X and Y are two independent discrete random variables with PGFs $G(t)$ and $H(t)$ respectively, then:

1 $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$

2 $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

We note that:

- **Theorem 13** is only a special case of the results proved in **Theorem 6** and **Theorem 7**, which apply to any two random variables X and Y (either both discrete or both continuous).
- X and Y are required to be independent for **Theorem 13**, but **Theorem 6** shows that $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ holds also when X and Y are dependent.

Using the results from **Section A**, we obtain:

Corollary:

Let $X_1, X_2, ..., X_n$ be n independent discrete random variables with PGFs $G_1(t)$, $G_2(t)$, ..., $G_n(t)$. Then $\mathrm{E}(X_1 + X_2 + ... + X_n) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + ... + \mathrm{E}(X_n)$
and $\mathrm{Var}(X_1 + X_2 + ... + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + ... + \mathrm{Var}(X_n)$.

Example 33

Let $X \sim \text{DU}(6)$ with values $1, 2, 3, 4, 5, 6,$ be the score obtained when a fair 6-sided die is rolled. Let $Y \sim \text{DU}(4)$ with values $1, 2, 3, 4,$ be the score obtained when a tetrahedral die is rolled.

a Use the method of PGFs to find the probability distribution of $X+Y$.

b Hence find $\text{E}(X+Y)$ and $\text{Var}(X+Y)$.

c Check your answers to b using the formulae for the mean and variance of a discrete uniform distribution.

---

a X has PGF $G_X(t) = \frac{t}{6}\left(\frac{t^6-1}{t-1}\right) = \frac{1}{6}(t+t^2+....+t^6)$.

Y has PGF $G_Y(t) = \frac{t}{4}\left(\frac{t^4-1}{t-1}\right) = \frac{1}{4}(t+t^2+t^3+t^4)$.

Since X and Y are independent random variables, $X+Y$ has PGF

$G(t) = G_X(t)G_Y(t) = \frac{1}{6}(t+t^2+....+t^6) \times \frac{1}{4}(t+t^2+t^3+t^4)$

$= \frac{1}{24}(t^2+2t^3+3t^4+4t^5+4t^6+4t^7+3t^8+2t^9+t^{10})$

$\therefore$ the probability distribution for $X+Y$ is:

$x+y$	2	3	4	5	6	7	8	9	10
$\text{P}(x+y)$	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{24}$

b $G'(t) = \frac{1}{24}(2t+6t^2+12t^3+20t^4+24t^5+28t^6+24t^7+18t^8+10t^9)$

$\therefore$ $G'(1) = \frac{1}{24}(2+6+12+20+24+28+24+18+10)$

$= \frac{1}{24} \times 144$

$= 6$

$\therefore$ $\text{E}(X+Y) = G'(t) = 6$.

$G''(t) = \frac{1}{24}(2+12t+36t^2+80t^3+120t^4+168t^5+168t^6+144t^7+90t^8)$

$\therefore$ $G''(1) = \frac{1}{24} \times 820$

≈ 34.167

$\therefore$ $\text{Var}(X+Y) = G''(1)+G'(1)-[G'(1)]^2$

$\approx 34.167+6-6^2$

≈ 4.17

c $\text{E}(X) = \frac{6+1}{2} = 3.5$ $\qquad$ $\text{Var}(X) = \frac{6^2-1}{12} = \frac{35}{12}$

$\text{E}(Y) = \frac{4+1}{2} = 2.5$ $\qquad$ $\text{Var}(Y) = \frac{4^2-1}{12} = \frac{15}{12}$

$\therefore$ $\text{E}(X+Y) = \text{E}(X)+\text{E}(Y)$ and $\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$

$\text{E}(X+Y) = 3.5+2.5 = 6$

$\text{Var}(X+Y) = \frac{35}{12}+\frac{15}{12} = 4\frac{1}{6} \approx 4.17$