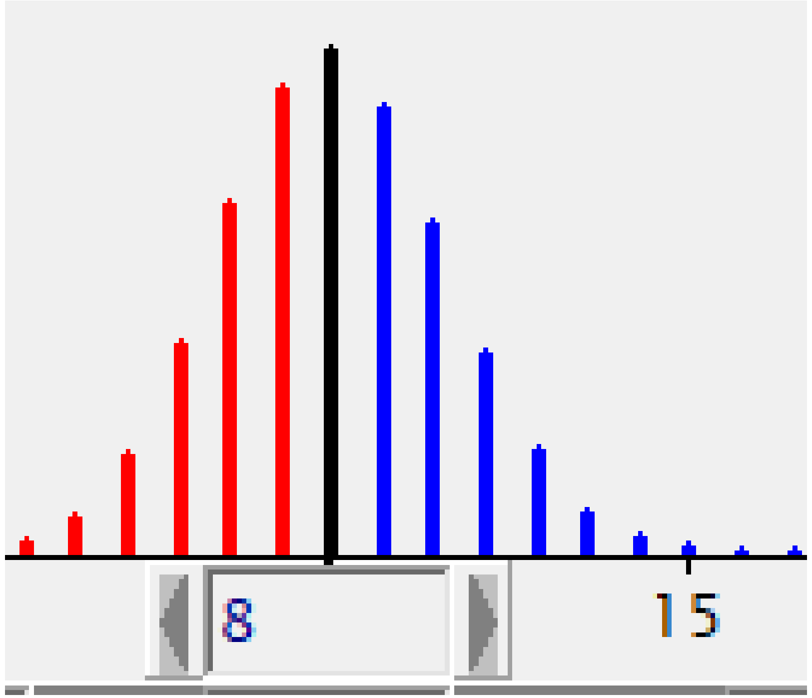
Advanced Mathematics Support Programme®



Continuing Professional Development Standard
National Centre for Excellence in the Teaching of Mathematics

Managed by Mathematics Education Innovation

Using Probability Generating Functions

Paul Chillingworth
Natalie Vernon

Go to student.desmos.com and type in:

G4W 6BZ

Probability Generating Function

A way of representing a probability distribution in a compact form, as a function, that allows key results to be derived in a fairly straightforward way and allows connections to be made.

Representing Probability Distributions

Consider throwing a fair four-sided dice.

There are different ways that we can represent the possible outcomes and their respective probabilities.

Representing Probability Distributions

Consider throwing a fair four-sided dice.

There are different ways that we can represent the possible outcomes and their respective probabilities.

As a table:

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

Representing Probability Distributions

Consider throwing a fair four-sided dice.

There are different ways that we can represent the possible outcomes and their respective probabilities.

As a table:

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

As a distribution function: $P(X = x) = \frac{1}{4} \qquad x = 1, 2, 3, 4$

Probability Generating Function

However there is another representation which may seem a little strange at first

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

$$\frac{1}{4}t$$

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

$$\frac{1}{4}t+\frac{1}{4}t^2$$

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

$$\frac{1}{4}t + \frac{1}{4}t^2 + \frac{1}{4}t^3$$

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

$$\frac{1}{4}t + \frac{1}{4}t^2 + \frac{1}{4}t^3 + \frac{1}{4}t^4$$

Probability Generating Function

However there is another representation which may seem a little strange at first.

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

And that is to form a polynomial whose coefficients are the probabilities and whose powers are the values of x.

We use a dummy variable, usually t for this.

$\frac{1}{4}t+\frac{1}{4}t^2+\frac{1}{4}t^3+\frac{1}{4}t^4$ This is usually referred to as $G_X(t)$.

Probability Generating Function

More generally we have:

$$G_X(t) = \sum P(X = x)t^x$$

summed over all values of the random variable which have non-zero probabilities.

These functions are useful as they contain all the information about the distribution in one function.

G(1)

$$G_X(t) = \sum P(X = x)t^x$$

What is $G(1)$?

Desmos
Screen 2

G(1)

$$G_X(t) = \sum P(X = x)t^x$$

$$G(1) = \sum P(X = x) \times 1^x = \sum P(X = x)$$

G(1)

$$G_X(t) = \sum P(X = x)t^x$$

$$G(1) = \sum P(X = x) \times 1^x = \sum P(X = x) = 1$$

Examples

$$G_X(t) = 0.3 + 0.1t + 0.2t^2 + 0.2t^3 + 0.1t^4 + 0.05t^5 + 0.05t^6$$

How do we know that this is a valid PGF?

What values can the random variable X take?

What is $P(X = 0)$ $P(X = 4)$ $P(X = 7)$?

Desmos Screen 3

Examples

$$G_X(t) = k(1+t)^8$$

Find k.

What is the largest value X can take?

$$G_X(t) = \sum P(X=x)t^x$$

Desmos
Screen 4

$$G(1) = \sum P(X=x) \times 1^x = \sum P(X=x) = 1$$

Examples

$$G_X(t) = \frac{1}{256}(1+t)^8$$

Find $P(X = 4)$

Desmos Screen 5

Examples

$$G_X(t) = \frac{1}{256}(1+t)^8$$

Find $P(X = 4)$

The term in t^4 is $\frac{1}{256} \times \binom{8}{4} 1^4 t^4$

Examples

$$G_X(t) = \frac{1}{256}(1+t)^8$$

Find $P(X = 4)$

The term in t^4 is $\frac{1}{256} \times \binom{8}{4} 1^4 t^4$

$$P(X = 4) = \frac{1}{256} \times \frac{8!}{4!\,4!} = \frac{35}{128}$$

Examples

$$G_X(t) = \frac{1}{256}(1+t)^8$$

What is the distribution of X?

Desmos Screen 6

Examples

$$G_X(t) = (0.5 + 0.5t)^8$$

What is the distribution of X?

Examples

$$G_X(t) = (0.5 + 0.5t)^8$$
$$= 0.5^8 + \binom{8}{1}(0.5)^7(0.5t)^1 + \binom{8}{2}(0.5)^6(0.5t)^2 + \cdots$$

What is the distribution of X?

Examples

$$G_X(t) = (0.5 + 0.5t)^8$$
$$= 0.5^8 + \binom{8}{1}(0.5)^7(0.5t)^1 + \binom{8}{2}(0.5)^6(0.5t)^2 + \cdots$$

What is the distribution of X?

$X \sim B(8,0.5)$

Binomial distribution

More generally for a binomial distribution:

$$G_X(t) = (1 - p + pt)^n$$
$$= (1-p)^n + \binom{n}{1}(1-p)^{n-1}(pt) + \binom{n}{2}(1-p)^{n-2}(pt)^2 + \cdots$$

Binomial distribution

More generally for a binomial distribution:

$$G_X(t) = (1 - p + pt)^n$$
$$= (1-p)^n + \binom{n}{1}(1-p)^{n-1}(pt) + \binom{n}{2}(1-p)^{n-2}(pt)^2 + \cdots$$

The term in t^x is $\binom{n}{x}(1-p)^{n-x}(pt)^x$

Binomial distribution

More generally for a binomial distribution:

$$G_X(t) = (1 - p + pt)^n$$

The term in t^x is $\binom{n}{x}(1-p)^{n-x}(pt)^x$

The coefficient of t^x is $\binom{n}{x}(1-p)^{n-x}p^x$ which is $P(X = x)$

Poisson distribution

A Poisson distribution is used to model events that occur randomly and singly in time or space at fixed rate.

It has probability distribution function:

$P(X = x) = \frac{e^{-\mu} \times \mu^x}{x!}$ where μ is the rate of occurrence in a given interval.

Poisson distribution

A Poisson distribution is used to model events that occur randomly and singly in time or space at fixed rate.

It has probability distribution function:

$P(X = x) = \frac{e^{-\mu} \times \mu^x}{x!}$ where μ is the rate of occurrence in a given interval.

What is the PGF?

Poisson distribution

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

What is the PGF?

$$G_X(t) = \sum_{x=0}^{\infty} \frac{e^{-\mu} \times \mu^x}{x!} \times t^x$$

Poisson distribution

What is the PGF?

$$G_X(t) = \sum_{x=0}^{\infty} \frac{e^{-\mu} \times \mu^x}{x!} \times t^x = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu t)^x}{x!}$$

Poisson distribution

What is the PGF?

$$G_X(t) = \sum_{x=0}^{\infty} \frac{e^{-\mu} \times \mu^x}{x!} \times t^x = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu t)^x}{x!}$$

$$= e^{-\mu}\left(1 + \mu t + \frac{(\mu t)^2}{2!} + \frac{(\mu t)^3}{3!} + \cdots\right)$$

Poisson distribution

What is the PGF?

$$G_X(t) = \sum_{x=0}^{\infty} \frac{e^{-\mu} \times \mu^x}{x!} \times t^x = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu t)^x}{x!}$$

$$= e^{-\mu}\left(1 + \mu t + \frac{(\mu t)^2}{2!} + \frac{(\mu t)^3}{3!} + \cdots\right)$$

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

Some properties

$$G_X(t) = \sum P(X = x)t^x$$

What is $G'_X(t)$?

Some properties

$$G_X(t) = \sum P(X = x)t^x$$

What is $G'_X(t)$?

$$G'_X(t) = \sum xP(X = x)t^{x-1}$$

Some properties

$$G_X(t) = \sum P(X = x)t^x$$

What is $G'_X(t)$?

$$G'_X(t) = \sum xP(X = x)t^{x-1}$$

$$G'_X(1) = \sum xP(X = x)$$

What is this?

Some properties

$$G_X(t) = \sum P(X = x)t^x$$

What is $G'_X(t)$?

$$G'_X(t) = \sum xP(X = x)t^{x-1}$$

$$G'_X(1) = \sum xP(X = x) = E(X)$$

E(X)

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

For a binomial distribution:

$G_X(t) = (1 - p + pt)^n$

For a Poisson distribution:

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

Can you find $E(X)$ for both of these?

Desmos Screen 7

E(X)

$$G_X'(1) = \sum xP(X = x) = E(X)$$

For a binomial distribution:

$$G_X(t) = (1 - p + pt)^n$$
$$G_X'(t) = np(1 - p + pt)^{n-1}$$
$$G_X'(1) = np(1 - p + p)^{n-1} = np$$

E(X)

$$G'_X(1) = \sum xP(X = x) = E(X)$$

For a Poisson distribution:

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

$$G'_X(t) = \mu e^{\mu(t-1)}$$

$$G'_X(1) = \mu e^0 = \mu$$

Using PGFs

If we can show that two random variables have the same PGF then we have shown that the two random variables have the same distribution.

The PGF of X tells us everything there is to know about the distribution of X.

Using PGFs

The power of the PGF is that it gives us an easy way of determining the distribution of the sum of two random variables $X + Y$ when X and Y are independent.

This is difficult to do using probability distribution functions.

The PGF transforms a sum into a product which enables it to be manipulated more readily.

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables $X\ and\ Y$ with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables $X \; and \; Y$ with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

$$G_{X+Y}(t) = E(t^{X+Y})$$

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables X and Y with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

$$G_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y)$$

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables $X\ and\ Y$ with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

$G_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y)$ as $X\ and\ Y$ independent

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables $X\ and\ Y$ with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

$$G_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y) = G_X(t)G_Y(t)$$

Using PGFs

We can do this by considering $G_X(t) = \sum P(X = x)t^x$ as $E(t^X)$

So consider two independent random variables $X\ and\ Y$ with PGFs $G_X(t)$ and $G_Y(t)$ respectively.

$$\boldsymbol{G_{X+Y}(t) = G_X(t)G_Y(t)}$$

The sum of 2 Poisson variables

Suppose that X and Y are independent with $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ).

Find the distribution of $X + Y$

For a Poisson distribution:
$G_X(t) = e^{\mu(t-1)}$

The sum of 2 Poisson variables

Suppose that X and Y are independent with $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ).

Find the distribution of $X + Y$

$$G_{X+Y}(t) = G_X(t)G_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)}$$

For a Poisson distribution:
$G_X(t) = e^{\mu(t-1)}$

The sum of 2 Poisson variables

Suppose that X and Y are independent with $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ).

Find the distribution of $X + Y$

$$G_{X+Y}(t) = G_X(t)G_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}$$

which is a Poisson distribution with parameter $\lambda + \mu$.

For a Poisson distribution:
$G_X(t) = e^{\mu(t-1)}$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p

How many rolls are needed to get the first six?

How many rolls are needed to get the r^{th} six?

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the first six?

Geometric distribution $P(X = x) = p(1-p)^{x-1}$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the first six?

Geometric distribution $P(X = x) = p(1-p)^{x-1}$

$$G_X(t) = \sum p(1-p)^{x-1}t^x$$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the first six?

Geometric distribution $P(X = x) = p(1-p)^{x-1}$

$$G_X(t) = \sum p(1-p)^{x-1}t^x = pt\sum (1-p)^{x-1}t^{x-1}$$

$$= pt\sum (t(1-p))^{x-1}$$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the first six?

Geometric distribution $P(X = x) = p(1-p)^{x-1}$

$$G_X(t) = \sum p(1-p)^{x-1}t^x = pt\sum (1-p)^{x-1}t^{x-1}$$

$$= pt\sum (t(1-p))^{x-1} = \frac{pt}{1-(1-p)t}$$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the r^{th} six?

Negative binomial $P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$

The geometric and negative binomial distributions

Imagine rolling a six-sided dice where getting a six on a single throw has probability p.

How many rolls are needed to get the r^{th} six?

Negative binomial $P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$

Using a similar argument:

$$G_Y(t) = \left(\frac{pt}{1-(1-p)t}\right)^r$$

The geometric and negative binomial distributions

So if take r independent observations of this geometric variable and sum them:

$Y = X_1 + X_2 + \cdots + X_r$

The geometric and negative binomial distributions

So if take r independent observations of this geometric variable and sum them:

$Y = X_1 + X_2 + \cdots + X_r$

So $G_Y(t) = G_{X_1}(t)G_{X_2}(t)\ldots\ldots G_{X_r}(t)$

The geometric and negative binomial distributions

So if take r independent observations of this geometric variable and sum them:

$Y = X_1 + X_2 + \cdots + X_r$

So $G_Y(t) = G_{X_1}(t)G_{X_2}(t)\ldots..\ G_{X_r}(t)$

But these are all equal to $G_X(t)$

The geometric and negative binomial distributions

So if take r independent observations of this geometric variable and sum them:

$Y = X_1 + X_2 + \cdots + X_r$

So $G_Y(t) = G_{X_1}(t)G_{X_2}(t)\ldots..\ G_{X_r}(t)$

But these are all equal to $G_X(t)$

$$G_Y(t) = (G_X(t))^r = \left(\frac{pt}{1-(1-p)t}\right)^r$$

The geometric and negative binomial distributions

So if take r independent observations of this geometric variable and sum them:

$Y = X_1 + X_2 + \cdots + X_r$

So $G_Y(t) = G_{X_1}(t)G_{X_2}(t)\ldots..\ G_{X_r}(t)$

But these are all equal to $G_X(t)$

$$G_Y(t) = (G_X(t))^r = \left(\frac{pt}{1-(1-p)t}\right)^r$$

But this is the PGF for a negative binomial.

In summary

- $G_X(t) = \sum P(X = x)t^x$ is an alternative way of writing a probability distribution.
- It can be thought of as $E(t^X)$
- $G(1) = \sum P(X = x) = 1$
- $E(X) = G_X'(1)$
- $Var(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$
- The common distributions have fairly ‘compact’ PGFs – can be reasonably straightforward to work out $E(X)$ and $Var(X)$
- $G_{X+Y}(t) = G_X(t)G_Y(t)$ if X and Y are independent – this helps to derive the distribution of a sum of random variables.

Other uses

- Expectation and variance of functions of random variables
- Sums of random variables where the number of variables is itself random (a randomly stopped sum)
- Random walks (expected reach times)
- Deciding whether a process will reach a given state.
- Sicherman dice: session I4 at https://mei.org.uk/conference13 .