Probability generating functions

Introduction

We will introduce the idea of a probability generating function, by first considering a simple example of a discrete probability distribution.

<u>Example</u>

A discrete random variable X has the following probability distribution.

X	1	2	3	4
$P(X = x_i)$	$1/8$	$1/8$	$1/4$	$1/2$

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

$$E(X) = \sum x.P(X = x)$$
$$= 1 \times \frac{1}{8} + 2 \times \frac{1}{8} + 3 \times \frac{1}{4} + 4 \times \frac{1}{2}$$
$$= \frac{1 + 2 + 6 + 16}{8} = \frac{25}{8}$$

$$E(X^2) = \sum x^2.P(X = x)$$
$$= 1^2 \times \frac{1}{8} + 2^2 \times \frac{1}{8} + 3^2 \times \frac{1}{4} + 4 \times \frac{1}{2}$$
$$= \frac{1 + 2 + 6 + 16}{8} = \frac{87}{8}$$

$$Var(X) = E(X^2) + \left[E(X)\right]^2 = \frac{87}{8} - \left(\frac{25}{8}\right)^2 = \frac{696 - 625}{16} = \frac{71}{16}$$

These are calculations that the student should have encountered at an earlier stage. We now introduce, however, a function $G_x(t)$, called the probability generating function, which is

$$G_x(t) = E(t^x) = \sum t^x.P(X = x)$$

We can form a table for our current example as follows:

x	1	2	3	4
$P(X = x_i)$	$1/8$	$1/8$	$1/4$	$1/2$
t^x	t	t^2	t^3	t^4

Thus

$$\begin{aligned} Gx(t) &= E(t^x) \\ &= \sum t^x.p(x) \\ &= \frac{1}{8}t + \frac{1}{8}t^2 + \frac{1}{4}t^3 + \frac{1}{2}t^4 \end{aligned}$$

Then, suppose we differentiate $Gx(t)$

$$G_x'(t) = \frac{1}{8} + \frac{1}{8}.2t + \frac{1}{4}.3t^2 + \frac{1}{2}.4t^3$$

The coefficients of the series t, $2t, 3t^2$ and $4t^3$ are the same as those we met in the definition of $E(X)$. That is,

$$E(X) = G_x'(1)$$

That is, when we substitute $t = 1$ into $G_x'(t)$ we obtain

$$E(t) = G_x'(1) = \frac{1}{8} + \frac{2}{8} + \frac{3}{4} + \frac{4}{2} = \frac{25}{8}$$

Differentiating $G_x'(t)$ gives

$$G_x''(\mathrm{t}) = \frac{1}{8}.2 + \frac{1}{4}.3.2\mathrm{t} + \frac{1}{2}.4.3\mathrm{t}^2$$

Then $G_x''(1) = \frac{1}{8}.2 + \frac{1}{4}.3.2 + \frac{1}{2}.4.3$

Now consider $G_x''(1) + G_x'(1)$

$$G_x''(1) + G_x'(1) = \frac{1}{8}.2 + \frac{1}{4}.3.2 + \frac{1}{2}.4.3 + \frac{1}{8} + \frac{1}{8}.2 + \frac{1}{4}.3 + \frac{1}{2}.4$$

This could be written

$$G_x''(1) + G_x'(1) = \frac{1}{8} + \frac{1}{8}.2^2 + \frac{1}{4}.3^2 + \frac{1}{2}.4^2$$

This is equal to $E(X^2)$. That is

$$G_x''(1) + G_x'(1) = E(X^2)$$

Hence

$$Var(X) = G_x''(1) + G_x'(1) - \left[G_x'(1)\right]^2$$

This illustrates that probability generating functions could be useful short-cuts to finding *E*(*X*) and *VarX*. But in our current example the "short-cut" seems far from short! The real advantage from probability generating functions comes from the ability to write certain probability distributions as powers of a generating series. We will show this later, but firstly, we offer a formal definition of a probability generating function.

Probability generating functions

Let X be a discrete random variable. Let $P(X = x_i)$ be the probability that X takes the value x_i . Then the probability generating function defined for X is the function

$$G_x(t) = \sum_{i=o}^{n} t^{xi}.P(X = x_i) = E(t^x)$$

This is a function of two variables. The first is x_i the values that the random variable X can take. The second is t, which is introduced to create the idea of a family of related series.
The main properties of probability generating functions are as follows.

First Property

$$G_x(1) = 1$$

Proof

$$\begin{aligned} G_x(1) &= E(t^x)\big|_{t=1} \\ &= p_1 t^1 + p_2 t^2 + p_3 t^3 + \ldots + p_n t^n \big|_{t=1} \\ &= p_1 + p_2 + p_3 + \ldots p_n \\ &= 1 \end{aligned}$$

Here we use the abbreviation $p_i = P(X = x_i)$ for the probability that X takes the value x_i.
The symbol

$$\big|_{t=1}$$

means that we evaluate the expression by substituting t=1

The last line follows since the sum of a probability distribution is 1. This first result could be said to state the obvious, but it is a property of a probability generating function, and we must include it.

Second Property

$$G_x^{'}(1) = E(X)$$

Proof

$$\begin{aligned} G_x^{'}(1) &= G_x^{'}(t)\big|_{t=1} \\ &= \frac{d}{dt} G \times (t)\big|_{t=1} \\ &= \frac{d}{dt}\left\{ p_1 t + p_2 t^2 + p_3 t^3 + \ldots p_k t^k + \ldots + p_n t^n \right\}\big|_{t=1} \\ &= p_1 + 2p_2 t + 2p_2 t^2 + \ldots k p_k t^{k-1} + \ldots + n p_n t^{n-1}\big|_{t=1} \\ &= p_1 + 2p_2 + 3p_3 + \ldots + k p_k + n p_n \\ &= \sum_{i-1}^{n} x_i p_i = E(X) \end{aligned}$$

Third Property

$$G_x''(1) = E\left(X^2\right) - E(X)$$

Proof

We will write this as $E\left(X^2\right) = G_x''(1) + E(X)$

Then

$$\begin{aligned} G_x''(1) &= \frac{d}{dt^2} G_x(t)\big|_{t=1} \\ &= \frac{d}{dt}\left\{ p_1 + 2p_2 t + 3p_3 t^2 + \ldots + p_k t^{k-1} + \ldots + p_n t^{n-1} \right\}\big|_{t=1} \\ &= 2p_2 + 6p_3 t + \ldots + k(k-1) p_k t^{k-2} + \ldots + n(n-1) p_n t^{n-2}\big|_{t=1} \\ &= 2p_2 + 6p_3 + \ldots + k(k-1) p_k + \ldots + n(n-1) p_n \end{aligned}$$

Therefore

$$
\begin{aligned}
G_x''(1)+E(X) &= 2p_2+6p_3+\ldots+k(k-1)p_k+\ldots+n(n-1)p_n \\
&= P_1+4P_2+9P_3+\ldots+KP_k+\ldots+NP_n \\
&= P_1+4P_2+9P_3+\ldots+P_k\left(K(K-1)+K\right)+\ldots+P_n\left(n(n-1)+n\right) \\
&= p_1+4p_2+9p_3+\ldots+k^2p_k+\ldots+n^2p_n \\
&= \sum_{i=1}^{n} x_i p_i \\
&= E\left(X^2\right)
\end{aligned}
$$

<u>Fourth Property</u>

$$\text{Var(X)} = G_x''(1)+G_x'(1)-\left[G_x'(1)\right]^2$$

Proof

$$Var(X) = E(X^2)-[E(X)]^2$$
$$= G_x''(1) + G_x'(1) - [G_x'(1)]^2$$

We now proceed to apply these properties to the uniform (uniform), Binomial, Geometric and Poisson distributions to find the mean and variance in each case.

Uniform, discrete distribution

Let X be a random variable with a uniform discrete distribution. Then

$$p_i = P(X=i) = \begin{cases} \dfrac{1}{n} & \text{for } i = 1,2,3,\ldots,\ n \\ 0 & \text{otherwise} \end{cases}$$

The probability generating function is

$$
\begin{aligned}
G_x(t) &= \sum_{i=0}^{n} t^i.\frac{1}{n} \\
&= t\frac{1}{n}+t^2\frac{1}{n}+t^3\frac{1}{n}+\ldots+t^n\frac{1}{n} \\
&= \frac{1}{n}(t+t^2+t^3+\ldots+t^n)
\end{aligned}
$$

$t, t^2, t^3, \ldots, t^n$ is a geometric series with first term t, ratio t, and n terms. Hence

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

by the result for the sum of a geometric series.

Then

$$\begin{aligned}
E(X) &= G_x^{'}(1) \\
&= \frac{1}{n}.\frac{d}{dt}(t + t^2 + t^3 + ... + t^n)\Big|_{t=1} \\
&= \frac{1}{n}(1 + 2t + 3t^2 + ... + nt^{n-1})\Big|_{t=1} \\
&= \frac{1}{n}(1 + 2 + 3 + ... + n) \\
&= \frac{1}{n}\frac{(n+1)n}{2} \qquad \text{since } \sum_{i=1}^{n}\frac{n(n+1)}{2} \\
&= \frac{n+1}{2}
\end{aligned}$$

$$\begin{aligned}
G_x''(1) &= \frac{1}{n}\cdot\frac{d}{dt}(1 + 2t + 3t^2 + ... + nt^{n-i})\Big|_{t=1} \\
&= \frac{1}{n}(2 + 6t + ... + n(n-1)t^{n-2})\Big|_{t=1} \\
&= \frac{1}{n}(2 + 6 + ...n(n-1))
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= G_x^{''}(1) + G_x^{'}(1) \\
&= \frac{1}{n}(2 + 6 + ... + n(n-1)) + \frac{1}{n}(1 + 2 + 3 + ... + n) \\
&= \frac{1}{n}(1 + 4 + ... + n^2) \\
&= \frac{1}{n}.\frac{1}{6}n(n+1)(2n+1) \qquad \text{since } \sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(n+2) \\
&= \frac{1}{6}(n+1)(2n+1)
\end{aligned}$$

$$
\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{1}{6}(n+1)(2n+1) - \frac{(n+1)^2}{4} \\
&= \frac{1}{12}[2(n+1)(2n+1) - 3(n+1)(n+1)] \\
&= \frac{1}{12}[(n+1)(4n+2-3n-3)] \\
&= \frac{1}{12}[(n+1)(n-1)] \\
&= \frac{n^2-1}{12}
\end{aligned}
$$

Binomial Distribution

Recall that the discrete random variable X follows the Binomial distribution when it arises from the application of n successive trials to a situation where at each trial there is a probability p of a success and $q = 1-p$ of a failure. Then the probability of r successes is given by

$$P(X = r) = {}^nC_r p^r q^{n-r}$$

where ${}^nC_r = \dfrac{n!}{r!(n-r)!}$ is the Binomial coefficient.

We write $X \sim B(n, p)$

Then the probability generating function for X is

$$G_x(t) = \sum_{r=0}^{n} {}^nC_r \; p^r \; q^{n\text{-}r} \; t^r$$

The series $\sum_{r=0}^{n} {}^nC_r p^r q^{n\text{-}r}$ comes from expanding the expression $(p+q)^n$ by the Binomial Theorem. Hence, the probability generating function for $B(n, p)$ is

$$G_x(t) = (pt+q)^n$$

Thus, the Binomial distribution is that distribution whose probability generating function is $(pt+q)^n$

We now use the properties of a probability generating function to find *E(X)* and *Var(X)* for $B(n,p)$.

$$
\begin{aligned}
E(X) &= G_x^{'}(t)\Big|_{t=1} \\
&= \frac{d}{dt}(pt+q)^n\Big|_{t=1} \\
&= n(pt+q)^{n-1}.p\Big|_{t=1} \\
&= n(pt+q).p \\
&= np \qquad \text{since } p+q = 1
\end{aligned}
$$

$$
\begin{aligned}
G_x''(t)\Big|_{t=1} &= \frac{d}{dt}np(pt+q)^{n-1}\Big|_{t=a} \\
&= n(n-1)p^2(p+q) \\
&= n(n-1)p^2
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= G_x''(1) + G_x^{'}(1) - [G_x^{'}(1)]^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= n.p.(1-p) \\
&= npq \qquad \text{since } q = 1-p
\end{aligned}
$$

Geometric Distribution

The discrete random variable *X* follows the Geometric distribution when it arises from the application of possibly infinite trials to a situation where at each trial there is a probability p of a success and $q = 1-p$ of a failure. The trials are repeated until a success is obtained.

The probability that *r* trials will be necessary until a success is obtained is given by

$$P(X=r) = pq^{r-1}$$

We write

$$X \sim Geo(p)$$

The probability generating function for the Geometric distribution is

$$G_x(t) = \sum_{r=1}^{n} pq^{r-1}.t^r$$
$$= pt + pqt^2 + pq^2t^3 + ... + pq^{r-1}t^r + ...$$

This is, not surprisingly, a geometric series with first term *pt* and ratio *qt*.

The sum to infinity is

$$G_x(t) = S_\infty = \frac{pt}{1-qt}$$

Then

$$E(X) = G_x'(t)\Big|_{t=1}$$
$$= p(1-qt)^{-1} + pt(1-qt)^{-2}.q\Big|_{t=1}$$
$$= \frac{p}{1-qt} + \frac{pqt}{(1-qt)^2}\Bigg|_{t=1}$$
$$= \frac{p}{1-q} + \frac{pq}{(1-q)^2}$$
$$= \frac{p}{p} + \frac{pq}{p^2} \qquad \text{since } p = 1-q$$
$$= \frac{p+q}{p}$$
$$= \frac{1}{p} \qquad \text{since } p+q = 1$$

$$G_X''(t)\Big|_{t=1} = \frac{d}{dt}\left(\frac{p}{1-qt} + \frac{pqt}{(1-qt)^2}\right)\Bigg|_{t=1}$$
$$= \frac{d}{dt}\left(\frac{p(1-qt)+pqt}{(1-qt)^2}\right)\Bigg|_{t=1}$$
$$= \frac{d}{dt}\left(\frac{p - pqt + pqt}{(1-qt)^2}\right)\Bigg|_{t=1}$$
$$= \frac{d}{dt}p(1-qt)^{-2}\Bigg|_{t=1}$$
$$= 2pqt(1-qt)^{-3}\Big|_{t=1} = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}$$

$$Var(X) = G_x''(1) + G_x'(1) - [G_x'(1)]^2$$

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

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

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

$$= \frac{q}{p^2}$$

Poisson Distribution

A discrete random variable X follows the Poisson distribution when X is such that it can take values $r = 0, 1, 2, \ldots$ and the probability that X takes the value is given by

$$p(r) = P(X = r) = \frac{\lambda^r e^{-\lambda}}{r!} \qquad r = 0, 1, 2, \ldots$$

We write $X \sim Po(\lambda)$.

The probability generating function for the Poisson distribution is given by:-

$$G_x(t) = \sum_{r=0}^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} t^r$$

Then

$$E(X) = G_x'(t)\Big|_{t=1}$$

$$= \frac{d}{dt}\left(e^{-\lambda} + \lambda e^{-\lambda} t + \frac{\lambda^2}{2!} e^{-\lambda} t^2 + \ldots\right)\Bigg|_{t=1}$$

$$= \left(\lambda e^{-\lambda} + \lambda^2 e^{-\lambda} t + \frac{\lambda^3}{2!} e^{-\lambda} t^2 + \ldots\right)\Bigg|_{t=1}$$

$$= e^{-\lambda}.\left(\lambda + \lambda^2 + \frac{\lambda^3}{2!} + \frac{\lambda^4}{3!} + \ldots\right) = \lambda e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right)$$

$$= \lambda e^{-\lambda} e^{\lambda} \qquad \text{since } e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots$$

$$= \lambda$$

$$G_x''(t) = \frac{d}{dt}(\lambda e^{-\lambda} + \lambda^2 e^{-\lambda} t + \frac{\lambda^3}{2!} e^{-\lambda} t^2 + \frac{\lambda^4}{3!} e^{-\lambda} t^3 + ...)\Big|_{t=1}$$
$$= \lambda^2 e^{-\lambda} + \lambda^3 e^{-\lambda} t + \frac{\lambda^4}{2!} e^{-\lambda} t^2 + \frac{\lambda^5}{3!} e^{-\lambda} t^3 + ...\Big|_{t=1}$$
$$= \lambda^2 e^{-\lambda} + \lambda^3 e^{-\lambda} + \frac{\lambda^4}{2!} e^{-\lambda} + \frac{\lambda^5}{3!} e^{-\lambda}$$
$$= \lambda^2 e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} = ...)$$
$$= \lambda^2 e^{-\lambda} e^{\lambda}$$
$$= \lambda^2$$

Then

$$Var(X) = G_x''(1) + G_x'(1) - [G_x'(1)]^2$$
$$= \lambda^2 + \lambda - \lambda^2 = \lambda$$

The probability generating function of the sum of two independent variables

We begin our discussion of the sum of two independent variables by introducing an example.

Example

In a simulation of the Second World War London will be "successfully" blitzed when a five or a six has been obtained twice from the repeated throw of a six-sided, fair dice. Let Y be the random variable representing the number of throws up to and including the occurrence of the second successful throw.

i. Find the probability generating function $G_x(t)$ for the geometric series that arises with parameter $^1/_3$
ii. Hence, find the probability generating function of *X*.
iii. Find the mean of *Y*.

Answer

(i) The geometric series generated by parameter $^1/_3$ is $X \sim$ Geo(1/3).

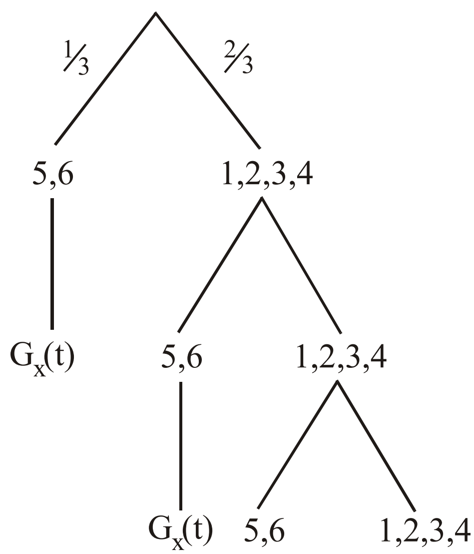
Then

$$P(X=r)=pq^{r-1}=\frac{1}{3}\left(\frac{2}{3}\right)^{r-1}$$

The probability generating function is

$$G_x(t)=\frac{pt}{1-qt}=\frac{t/3}{1-2t/3}$$

(ii) We begin by drawing a probability tree for Y:



Once a 5 or 6 has been thrown once the next 5 or 6 terminates the process, so after any 5, 6 the probability tree follows the geometric series with probability generating function $G_x(t)$ of the first part of the question. Hence

$$\begin{aligned}G_y(t)&=ptG_x(t)+pqt^2G_x(t)+pq^2t^3G_x(t)+....\\&=G_x(t)(pt+pqt^2+pq^2t^3+....)\\&=G_x(t).G_x(t)\\&=\frac{t/3}{1-2t/3}.\frac{t/3}{1-2t/3}\\&=\frac{(t/3)^2}{(1-2t/3)^2}\end{aligned}$$

(iii) $$G_y'(t)\Big|_{t=1} = \frac{d}{dt}\frac{t^2}{9}\left(1-\frac{2t}{3}\right)^{-2}\Bigg|_{t=1}$$

$$= \frac{2t}{9}\left(1-\frac{2t}{3}\right)^{-2} + \frac{2t^2}{9}\left(1-\frac{2t}{3}\right)^{-3}.\frac{2}{3}\Bigg|_{t=1}$$

$$= \frac{2}{9}\left(\frac{1}{3}\right)^{-2} + \frac{4}{27}\left(\frac{1}{3}\right)^{-3}$$

$$= 2+4 = 6$$

The second part of this question could have been solved more efficiently without recourse to a diagram of the probability tree. The first and second throws of the dice are each independent of the other. Both follow the geometric distribution $X\sim$ Geo(1/3) with probability generating function

$$G_x(t) = \frac{t/3}{1-2t/3}.$$

We require the probability distribution of a success in the first and second throws. Hence, we seek the probability generating function corresponding to the variable $Y = X + X$.

The following general result would enable us to write this down immediately:

Result

The probability generating function of the sum of the two independent random variables is the product of the probability generating functions of these two variables. That is

$$G_{(X+Y)}(t) = G_x(t)G_y(t)$$

Hence, in the above example, since $Y = X + X$

$$G_y(t) = G_{(X+X)}(t) = G_X(t).G_x(t)$$

$$= \frac{(t/3)^2}{(1-2t/3)^2}$$

We will now seek to justify this result that

$$G_{(X+Y)}(t) = G_x(t)G_y(t)$$

Firstly, we ask, what is the probability distribution of the variable $X+Y$? It is the distribution of probabilities of each pair of events $X=x_i$, $Y=y_i$. That is, $P(x_i, y_i)$ is the probability that X takes the value x_i and Y takes the value y_i. If x_i and y_i are independent

$$P(x_i \text{ and } y_i) = P(x_i)\cdot P(y_i)$$

Then, let

$$G_x(t) = \sum_{i=0}^{\infty} p(x_i).t^i = p(x_0).t^0 + p(x_1).t^1 + p(x_2).t^2 + \ldots$$

$$G_y(t) = \sum_{j=0}^{\infty} p(y_i).t^i = p(y_0).t^0 + p(y_1).t^1 + p(y_2).t^2 + \ldots$$

So

$$\begin{aligned} G_x(t).G_y(t) &= \sum_{i=0}^{\infty} p(x_i).t^i \\ &= (p(x_0).t^0 + p(x_1).t^1 + p(x_2).t^2 + \ldots).(p(y_0).t^0 + p(y_1).t^1 + p(y_2).t^2 + \ldots) \\ &= p(y_0)p(x_0)t^0t^0 + p(y_1)p(x_1)t^1t^1 + p(y_2)p(x_2)t^2t^2 + \ldots \end{aligned}$$

The following diagram illustrates this process

	$p(x_0)t^0$	$p(x_1)t^1$	$p(x_2)t^2$	. . . .
$p(y_0)t^0$	$p(x_0)t^0 \cdot p(y_0)t^0$ ↓	↗ $p(x_1)t^1 \cdot p(y_0)t^0$ →	$p(x_2)t^2 \cdot p(y_0)t^0$ ↙	. . . .
$p(y_1)t^1$	$p(x_0)t^0 \cdot p(y_1)t^1$ ↗	$p(x_1)t^1 \cdot p(y_1)t^1$ ↙	. . . .	. . . .
$p(y_2)t^2$	$p(x_0)t^0 \cdot p(y_2)t^2$ ↓	. . . .	. . . .	. . . .
. . . .	. . . .	. . . .	. . . .	. . . .

Following the arrows ensures that all the terms are visited once. This is an example of a diagonalisation process.

The entries in the table are the entries for the distribution $X+Y$; they arise from the product of $G_x(t)$ and $G_y(t)$; hence the table demonstrates that, provided X and Y are independent

$$G_{(X+Y)}(t) = G_x(t)G_y(t)$$