

6 — PROBABILITY GENERATING FUNCTIONS

Certain derivations presented in this course have been somewhat heavy on algebra. For example, determining the expectation of the Binomial distribution (page 5.1) turned out to be fairly tiresome. Another example of hard work was determining the set of probabilities associated with a sum, $P(X + Y = t)$. Many of these tasks are greatly simplified by using *probability generating functions*...

Probability Generating Functions — Introduction

A polynomial whose coefficients are the probabilities associated with the different outcomes of throwing a fair die was given on page 5.8:

$$G(\eta) = 0\eta^0 + \frac{1}{6}\eta^1 + \frac{1}{6}\eta^2 + \frac{1}{6}\eta^3 + \frac{1}{6}\eta^4 + \frac{1}{6}\eta^5 + \frac{1}{6}\eta^6$$

The coefficients of the square of this polynomial give rise to a higher-order polynomial whose coefficients are the probabilities associated with the different sums which can occur when two dice are thrown.

There is nothing special about a fair die. The set of probabilities associated with almost any discrete distribution can be used as the coefficients of $G(\eta)$ whose general form is:

$$G(\eta) = P(X = 0)\eta^0 + P(X = 1)\eta^1 + P(X = 2)\eta^2 + P(X = 3)\eta^3 + P(X = 4)\eta^4 + \dots \quad (6.1)$$

This is a power series which, for any particular distribution, is known as the associated probability generating function. Commonly one uses the term *generating function*, without the attribute probability, when the context is obviously probability. Generating functions have interesting properties and can often greatly reduce the amount of hard work which is involved in analysing a distribution.

The crucial point to notice, in the power series expansion of $G(\eta)$, is that the coefficient of η^r is the probability $P(X = r)$.

Properties of Generating Functions

It is to be noted that $G(0) = P(X = 0)$ and, rather more importantly, that:

$$G(1) = P(X = 0) + P(X = 1) + P(X = 2) + \dots = \sum_r P(X = r) = 1$$

In the particular case of the generating function for the fair die:

$$G(1) = 0 + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

Next, consider $G(\eta)$ together with its first and second derivatives $G'(\eta)$ and $G''(\eta)$ (the differentiation is with respect to η of course):

$$\begin{aligned} G(\eta) &= P(X = 0)\eta^0 + P(X = 1)\eta^1 + P(X = 2)\eta^2 + P(X = 3)\eta^3 + P(X = 4)\eta^4 + \dots \\ G'(\eta) &= 1P(X = 1)\eta^0 + 2P(X = 2)\eta^1 + 3P(X = 3)\eta^2 + 4P(X = 4)\eta^3 + \dots \\ G''(\eta) &= 2.1P(X = 2)\eta^0 + 3.2P(X = 3)\eta^1 + 4.3P(X = 4)\eta^2 + \dots \end{aligned}$$

Now, consider $G(1)$, $G'(1)$ and $G''(1)$:

$$\begin{aligned} G(1) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + \dots \\ G'(1) &= 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + 4P(X = 4) + \dots \\ G''(1) &= 2.1P(X = 2) + 3.2P(X = 3) + 4.3P(X = 4) + \dots \end{aligned}$$

At this stage, recall the general formula for the expectation of an arbitrary function of a random variable:

$$E(f(X)) = \sum_r f(r).P(X = r)$$

Then, express $G(1)$, $G'(1)$ and $G''(1)$ in sigma notation and derive the following results:

$$\begin{aligned} G(1) &= \sum_r P(X = r) = 1 \\ G'(1) &= \sum_r r.P(X = r) = E(X) \\ G''(1) &= \sum_r r(r - 1).P(X = r) = E(X(X - 1)) \end{aligned} \tag{6.2}$$

By differentiating the generating function one can directly obtain the expectation $E(X)$. By differentiating again one can obtain $E(X(X - 1))$. These two results lead to the rapid derivation of $V(X)$. Note:

$$\begin{aligned} G''(1) + G'(1) - (G'(1))^2 &= E(X(X - 1)) + E(X) - (E(X))^2 \\ &= E(X^2) - E(X) + E(X) - (E(X))^2 \\ &= E(X^2) - (E(X))^2 \\ &= V(X) \end{aligned} \tag{6.3}$$

A generating function is particularly helpful when the probabilities, as coefficients, lead to a power series which can be expressed in a simplified form. With many of the commonly-used distributions, the probabilities do indeed lead to simple generating functions. Often it is quite easy to determine the generating function by simple inspection. Some examples will be considered...

The Binomial Distribution

The set of probabilities for the Binomial distribution can be defined as:

$$P(X = r) = \binom{n}{r} p^r q^{n-r} \quad \text{where } r = 0, 1, \dots, n$$

Accordingly, from (6.1), the generating function is:

$$\begin{aligned} G(\eta) &= \binom{n}{0} p^0 q^n \eta^0 + \binom{n}{1} p^1 q^{n-1} \eta^1 + \binom{n}{2} p^2 q^{n-2} \eta^2 + \binom{n}{3} p^3 q^{n-3} \eta^3 + \dots \\ &= \binom{n}{0} (p\eta)^0 q^n + \binom{n}{1} (p\eta)^1 q^{n-1} + \binom{n}{2} (p\eta)^2 q^{n-2} + \binom{n}{3} (p\eta)^3 q^{n-3} + \dots \end{aligned}$$

This last expression is easily recognised as the expansion of $(q + p\eta)^n$ so the generating function and its first two derivatives are:

$$\begin{aligned} G(\eta) &= (q + p\eta)^n \\ G'(\eta) &= n(q + p\eta)^{n-1} p \\ G''(\eta) &= n(n-1)(q + p\eta)^{n-2} p^2 \end{aligned}$$

Accordingly:

$$\begin{aligned} G(1) &= (q + p)^n = 1^n = 1 \\ G'(1) &= n(q + p)^{n-1} p = n \cdot 1^{n-1} p = np \\ G''(1) &= n(n-1)(q + p)^{n-2} p^2 = n(n-1) \cdot 1^{n-2} p^2 = n(n-1)p^2 \end{aligned}$$

By (6.2), $G'(1) = E(X) = np$. This result was derived very much more tediously on page 5.1.

The variance is derived from (6.3):

$$\begin{aligned} V(X) &= G''(1) + G'(1) - (G'(1))^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= n^2 p^2 - np^2 + np - n^2 p^2 \\ &= np(1-p) \\ &= npq \end{aligned}$$

This result was derived on page 5.2.

The Geometric Distribution

The set of probabilities for the Geometric distribution can be defined as:

$$P(X = r) = q^r p \quad \text{where } r = 0, 1, \dots$$

Remember, this represents r successive failures (each of probability q) before a single success (probability p). Note that r is unbounded; there can be an indefinite number of failures before the first success.

From (6.1), the generating function is:

$$\begin{aligned} G(\eta) &= q^0 p \eta^0 + q^1 p \eta^1 + q^2 p \eta^2 + q^3 p \eta^3 + \dots \\ &= p((q\eta)^0 + (q\eta)^1 + (q\eta)^2 + (q\eta)^3 + \dots) \end{aligned}$$

The item in brackets is easily recognised as an infinite geometric progression, the expansion of $(1 - q\eta)^{-1}$, so the generating function and its first two derivatives are:

$$\begin{aligned} G(\eta) &= p(1 - q\eta)^{-1} \\ G'(\eta) &= p(1 - q\eta)^{-2} q \\ G''(\eta) &= 2p(1 - q\eta)^{-3} q^2 \end{aligned}$$

Accordingly:

$$\begin{aligned} G(1) &= p(1 - q)^{-1} = p \cdot p^{-1} = 1 \\ G'(1) &= p(1 - q)^{-2} q = \frac{p}{p^2} q = \frac{q}{p} \\ G''(1) &= 2p(1 - q)^{-3} q^2 = 2 \frac{p}{p^3} q^2 = 2 \frac{q^2}{p^2} \end{aligned}$$

By (6.2), $E(X) = \frac{q}{p}$.

The variance is derived from (6.3):

$$\begin{aligned} V(X) &= G''(1) + G'(1) - (G'(1))^2 \\ &= 2 \frac{q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} \\ &= \frac{q}{p} \left(\frac{q}{p} + 1 \right) \\ &= \frac{q}{p} \left(\frac{q + p}{p} \right) \\ &= \frac{q}{p^2} \end{aligned}$$

Both the expectation and the variance of the Geometric distribution are difficult to derive without using the generating function.

The Poisson Distribution

The set of probabilities for the Poisson distribution can be defined as:

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad \text{where } r = 0, 1, \dots$$

This was introduced as the probability of r murders in a year when the average over a long period is λ murders in a year. As with the Geometric distribution r is unbounded; there can, in principle, be an indefinite number of murders in a year.

From (6.1), the generating function is:

$$\begin{aligned} G(\eta) &= \frac{\lambda^0}{0!} e^{-\lambda} \eta^0 + \frac{\lambda^1}{1!} e^{-\lambda} \eta^1 + \frac{\lambda^2}{2!} e^{-\lambda} \eta^2 + \frac{\lambda^3}{3!} e^{-\lambda} \eta^3 + \dots \\ &= \left(\frac{(\lambda\eta)^0}{0!} + \frac{(\lambda\eta)^1}{1!} + \frac{(\lambda\eta)^2}{2!} + \frac{(\lambda\eta)^3}{3!} + \dots \right) e^{-\lambda} \end{aligned}$$

The item in brackets is easily recognised as an exponential series, the expansion of $e^{(\lambda\eta)}$, so the generating function and its first two derivatives are:

$$\begin{aligned} G(\eta) &= e^{(\lambda\eta)} e^{-\lambda} \\ G'(\eta) &= \lambda e^{(\lambda\eta)} e^{-\lambda} \\ G''(\eta) &= \lambda^2 e^{(\lambda\eta)} e^{-\lambda} \end{aligned}$$

Accordingly:

$$\begin{aligned} G(1) &= e^\lambda e^{-\lambda} = 1 \\ G'(1) &= \lambda e^\lambda e^{-\lambda} = \lambda \\ G''(1) &= \lambda^2 e^\lambda e^{-\lambda} = \lambda^2 \end{aligned}$$

By (6.2), $E(X) = \lambda$.

The variance is derived from (6.3):

$$\begin{aligned} V(X) &= G''(1) + G'(1) - (G'(1))^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned}$$

The expectation of the Poisson distribution was derived without great difficulty on page 4.5 and it was pointed out that λ was the obvious result since the analysis of the Poisson distribution began by taking λ as the expectation. The variance has not been derived before and it is interesting to note that the variance and the expectation are the same.

The Uniform Distribution

In all the illustrations so far, a crucial step has been to recognise some expansion so that the generating function can be reduced to a simple form. In the case of the distribution for a fair die, the generating function cannot be simplified and little benefit accrues from using it to determine the expectation and variance.

Nevertheless the generating function *can* be used and the following analysis is a final illustration of the use of generating functions to derive the expectation and variance of a distribution.

The generating function and its first two derivatives are:

$$\begin{aligned}G(\eta) &= 0\eta^0 + \frac{1}{6}\eta^1 + \frac{1}{6}\eta^2 + \frac{1}{6}\eta^3 + \frac{1}{6}\eta^4 + \frac{1}{6}\eta^5 + \frac{1}{6}\eta^6 \\G'(\eta) &= 1 \cdot \frac{1}{6}\eta^0 + 2 \cdot \frac{1}{6}\eta^1 + 3 \cdot \frac{1}{6}\eta^2 + 4 \cdot \frac{1}{6}\eta^3 + 5 \cdot \frac{1}{6}\eta^4 + 6 \cdot \frac{1}{6}\eta^5 \\G''(\eta) &= 2 \cdot \frac{1}{6}\eta^0 + 6 \cdot \frac{1}{6}\eta^1 + 12 \cdot \frac{1}{6}\eta^2 + 20 \cdot \frac{1}{6}\eta^3 + 30 \cdot \frac{1}{6}\eta^4\end{aligned}$$

Accordingly:

$$\begin{aligned}G(1) &= 0 + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1 \\G'(1) &= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2} \\G''(1) &= \frac{2 + 6 + 12 + 20 + 30}{6} = \frac{70}{6} = \frac{35}{3}\end{aligned}$$

By (6.2), $E(X) = \frac{7}{2}$.

The variance is derived from (6.3):

$$\begin{aligned}V(X) &= G''(1) + G'(1) - (G'(1))^2 \\&= \frac{35}{3} + \frac{7}{2} - \frac{49}{4} \\&= \frac{140 + 42 - 147}{12} \\&= \frac{35}{12}\end{aligned}$$

The expectation and variance for a fair die were derived on pages 3.9 and 4.2 and the present calculations are almost identical. This is not a fruitful use of generating functions. Nevertheless, the generating function for a fair die can be very well worth exploiting in a quite different way...

The Generating Function as a Special Expectation

The general form of $G(\eta)$, given in (6.1), is:

$$G(\eta) = P(X = 0)\eta^0 + P(X = 1)\eta^1 + P(X = 2)\eta^2 + P(X = 3)\eta^3 + P(X = 4)\eta^4 + \dots$$

In sigma notation this can be rewritten:

$$G(\eta) = \sum_r P(X = r)\eta^r$$

Recall again the general formula for the expectation of an arbitrary function of a random variable:

$$E(f(X)) = \sum_r f(r) \cdot P(X = r)$$

Taking $f(X) = \eta^X$ as a function of X :

$$E(\eta^X) = \sum_r \eta^r \cdot P(X = r)$$

This is $G(\eta)$ and the identity:

$$G(\eta) = E(\eta^X) \tag{6.4}$$

is the starting point for some important theory...

The Sum of two Independent Random Variables — A Theorem

Suppose X and Y are two independent random variables whose values are r and s where $r, s = 0, 1, 2, \dots$

Let $G_X(\eta)$ be the generating function associated with X and $G_Y(\eta)$ be the generating function associated with Y . From (6.4):

$$G_X(\eta) = E(\eta^X) \quad \text{and} \quad G_Y(\eta) = E(\eta^Y)$$

Next, consider the generating function associated with $X + Y$ and call this G_{X+Y} . Again from (6.4):

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

From the general formula for the expectation of a function of random variables:

$$G_{X+Y}(\eta) = E(\eta^{X+Y}) = \sum_r \sum_s \eta^{r+s} P(X = r, Y = s) = \sum_r \sum_s \eta^r \eta^s P(X = r) \cdot P(Y = s)$$

Note that $P(X = r, Y = s)$ may be split only because X and Y are independent.

The function after the double-sigma sign can be separated into the product of two terms, the first of which does not depend on s and the second of which does not depend on r . Hence:

$$\begin{aligned} G_{X+Y}(\eta) &= \left(\sum_r \eta^r \cdot \text{P}(X = r) \right) \left(\sum_s \eta^s \cdot \text{P}(X = s) \right) \\ &= \text{E}(\eta^X) \cdot \text{E}(\eta^Y) \\ &= G_X(\eta) \cdot G_Y(\eta) \end{aligned}$$

In words: the generating function of the sum of two independent random variables is the product of the separate generating functions of the random variables.

This is a most useful theorem.

Two Dice

The theorem $G_{X+Y}(\eta) = G_X(\eta) \cdot G_Y(\eta)$ brings the analysis back to the starting point. When the generating function for a fair die was introduced on page 5.8, it was shown without explanation that squaring this generating function gives the generating function for the sum of two dice.

The explanation is now clear. The two dice were identical so $G_X = G_Y = G$ and the square, G^2 , gave the generating function for $X + Y$. In summary:

$$G_X(\eta) = G_Y(\eta) = G(\eta) = 0\eta^0 + \frac{1}{6}\eta^1 + \frac{1}{6}\eta^2 + \frac{1}{6}\eta^3 + \frac{1}{6}\eta^4 + \frac{1}{6}\eta^5 + \frac{1}{6}\eta^6$$

and:

$$\begin{aligned} G_{X+Y}(\eta) = (G(\eta))^2 &= 0\eta^0 + 0\eta^1 + \frac{1}{36}\eta^2 + \frac{2}{36}\eta^3 + \frac{3}{36}\eta^4 + \frac{4}{36}\eta^5 + \frac{5}{36}\eta^6 + \frac{6}{36}\eta^7 \\ &\quad + \frac{5}{36}\eta^8 + \frac{4}{36}\eta^9 + \frac{3}{36}\eta^{10} + \frac{2}{36}\eta^{11} + \frac{1}{36}\eta^{12} \end{aligned}$$

Note that although both X and Y are distributed Uniform(1,6), the distribution of $X + Y$ is emphatically *not* uniform. The distribution of $X + Y$ is triangular with a peak in the middle.

If the number of dice is increased, inspection of G^2 , G^3 , G^4 and so on shows that the shape of the distribution continues to change and tends to that of a symmetrical Binomial distribution (one in which $p = q = \frac{1}{2}$).

Two Binomial Distributions

On reflection, there has never been any reason to suppose that combining two Uniform distributions (in the manner of throwing two dice and adding their scores) would produce another Uniform distribution. By contrast, combining two Binomial distributions *can* produce another Binomial distribution. . .

Imagine that you have a pile of identical biased coins, each of which has a probability p of landing heads when tossed. Suppose you take 3 of these coins for yourself and hand 5 others to a friend.

You then toss your coins and count the number of heads, clearly a number in the range 0 to 3. Your friend (independently) tosses the other coins and again counts the number of heads, a number in the range 0 to 5. You then add your head count to your friend's head count.

Let X be the random number whose value r is the number of heads you obtain and Y be the random number whose value s is the number of heads your friend obtains. Clearly, X is distributed Binomial(3, p) and Y is distributed Binomial(5, p). It will be shown that $X + Y$ is distributed Binomial(8, p).

Given that X and Y are Binomially distributed:

$$P(X = r) = \binom{3}{r} p^r q^{3-r} \quad \text{and} \quad P(Y = s) = \binom{5}{s} p^s q^{5-s}$$

where r is in the range 0 to 3 and s is in the range 0 to 5.

From page 6.3, the generating functions associated with X and Y are:

$$G_X(\eta) = (q + p\eta)^3 \quad \text{and} \quad G_Y(\eta) = (q + p\eta)^5$$

From the theorem:

$$\begin{aligned} G_{X+Y}(\eta) &= G_X(\eta) \cdot G_Y(\eta) \\ &= (q + p\eta)^3 (q + p\eta)^5 \\ &= (q + p\eta)^8 \end{aligned}$$

Now $(q + p\eta)^8$ is the generating function for a random variable which is distributed Binomial(8, p). Accordingly, combining the two random variables X and Y , which are distributed Binomial(3, p) and Binomial(5, p), produces a random variable $X + Y$ which is distributed Binomial(8, p).

On reflection, this result is hardly surprising. A third party watching you and your friend could simply regard each trial as the tossing of 8 independent coins. The fact that the tossing employs two people is neither here nor there. The total number of heads is clearly distributed Binomial(8, p).

There is nothing special about the values 3 and 5 and the analysis readily generalises to the case where you have n_1 coins and your friend has n_2 coins. The only restrictions are that all $n_1 + n_2$ coins must be independent and identically distributed.

Each coin individually is, of course, distributed Binomial(1, p) and its generating function is $G(\eta) = (q + p\eta)^1$. For an arbitrary n such coins the generating function is:

$$G_n(\eta) = (G(\eta))^n = (q + p\eta)^n$$

This is the generating function of a random variable distributed Binomial(n , p).

Glossary

The following technical term has been introduced:

Probability Generating Function

Exercises — VI

1. A board game requires you to ‘throw a 6 to start’. How many throws do you expect to have to make before the throw which delivers you the required 6?
2. Consider again the battling Knights of Exercises II, question 7. Supposing battling Knights became a popular form of entertainment, promoters of such tournaments would wish to inform spectators of the expected length of a battle. Taking a round as the unit of measure, just what is the expected length of a battle?
3. If a random variable X is distributed $\text{Poisson}(\lambda)$, show that:

$$E(X(X-1)(X-2)\dots(X-k)) = \lambda^{k+1}$$

4. Use generating functions to find the distribution of $X + Y$, where X and Y are independent random variables distributed $\text{Binomial}(n_X, p_X)$ and $\text{Binomial}(n_Y, p_Y)$ respectively, for the case where n_X and n_Y are arbitrary positive integers (they may or may not be equal) but $p_X = p_Y$. In the alternative case, where $n_X = n_Y$ but $p_X \neq p_Y$, find the mean and variance of $X + Y$. Is the distribution Binomial?
5. [From Part IA of the Computer Science Tripos, 1993] X and Y are independent random variables having Poisson distributions with parameters α and β respectively. By using probability generating functions, or otherwise, prove that $X+Y$ has a Poisson distribution and give its parameter.

Find the conditional distribution for X given that $X + Y = n$, and give its mean and variance. Explain your result in words.