Generating Functions in Probability

Neil Shah

November 22, 2020

Abstract: In this paper, we discuss the probability generating function as well as properties it has related to moment amd radii of convergence. We also discuss the relation to moment generating functions as well as the applications of the probability generating function to topics such as random walks and inversions in permutations.

1 Some Common Probability Generating Functions

Definition 1. [2] Let X be a random variable. The **probability generating function** (PGF) of X is:

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + \ldots = \sum_{i=0}^{\infty} \mathbb{P}[X = i] s^i$$

Some of the most common examples of the probability generating function are for ([4]):

1. Bernoulli Distribution
2. Binomial Distribution
3. Negative Binomial Distribution
4. Poisson Distribution

Fact 2 (Bernoulli Distribution)**.** If X is a random variable with the Bernoulli distribution with parameter p,:

$$G_X(s) = 1 - p + ps$$

Proof. Note that

$$G_X(s) = \sum_{x \geq 0} p_X(x) s^x$$

must be true by the definition of a probability generating function. The Bernoulli distribution reads:

$$P_X(x) = \begin{cases} p & x = a \\ 1 - p & x = b \\ 0 & x \notin a, b \end{cases}$$

Thus:

$$P_X(s) = p_X(0)s^0 + p_X(1)s^1 = (1 - p) + ps$$

□

Fact 3 (Binomial Distribution)**.** If X is a random variable with the Binomial Distribution with parameters n and p:

$$G_X(s) = \sum_{k=0}^{n} p^k q^{n-k} s^k = (1 - p + ps)^n$$

Proof. Note:

$$G_X(s) = \sum_{x \geq 0} p_X(x) s^x$$

Due to how the binomial distribution is defined:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Thus:

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = (ps + 1 - p)^n$$

□

Fact 4 (Negative Binomial Distribution)**.** If X is a random variable with the Negative Binomial Distribution with parameters n and p,:

$$G_X(s) = \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n q^{k-n} s^k = (\frac{ps}{1-(1-p)s})^n$$

if $|s| < q^{-1}$ and where $q = 1 - p$.

Proof. Once again, we have:

$$P_X(s) = \sum_{k \geq 0} p_X(k) s^k$$

Due to how the negative binomial distribution is defined:

$$p_X(k) = \binom{n-1}{k-1} p^n q^{k-n}$$

So:

$$G_X(s) = \sum_{k \geq n} \binom{k-1}{n-1} p^n q^{k-n} s^k = \frac{p^n}{q^n} \sum_{k \geq n} \binom{k-1}{n-1} (qs)^k$$

Simplifying the above yields $(\frac{ps}{1-qs})^n$ as desired. □

Fact 5 (Poisson Distribution)**.** For random variable X with a Poisson Distribution with parameter λ:

$$G_X(s) = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} s^k = e^{\lambda(s-1)}$$

Proof. Yet again, we have:

$$G_X(s) = \sum_{x \geq 0} p_X(x) s^x$$

From how the Poisson Distribution is defined:

$$\forall k \in \mathbb{N}, k \geq 0 : p_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Thus:

$$G_X(s) = \sum_{k\geq 0} \frac{e^{-\lambda}\lambda^k}{k!} \cdot s^k$$
$$= e^{-\lambda} \cdot \sum_{k\geq 0} \frac{\lambda^k \cdot s^k}{k!} = e^{-\lambda(1-s)}$$

□

2 Properties of the PGF

Theorem 6. *For discrete random variable X and $g : \mathbb{R} \to \mathbb{R}$:*

$$\mathbb{E}(g(X)) = \sum_{x\in Im\ X} g(x)\mathbb{P}(X = x)$$

whenever this sum converges absolutely.

Proof. Let I denote the image of the random variable X. If $Y = g(X)$, the image of Y is $g(I)$. So:

$$\mathbb{E}(Y) = \sum_{y\in g(I)} y\mathbb{P}(Y = y) = \sum_{y\in g(I)} y \sum_{x\in I:g(x)=y} \mathbb{P}(X = x)$$
$$= \sum_{x\in I} g(x)\mathbb{P}(X = x)$$

□

Note that we can directly apply this to get a rather interesting result about the probability generating function:

Fact 7. For random variable X, we have:

$$G_X(s) = \mathbb{E}(s^X)$$

where G_X denotes the probability generating function.

Proof. This follows directly from Theorem 6. □

3 Calculating Probabilities with Generating Functions

Theorem 8. *[1]*

$$\mathbb{P}[X = k] = \left(\frac{1}{k!}\right) G_X^k(0)$$

where G_X^k denotes the kth derivative of the generating function.

Proof. Note that:

$$G'_X(s) = p_1 + 2p_2 s + 3p_3 s^2 + 4p_4 s^3 + \dots$$

So, $p_1 = \mathbb{P}(X = 1) = G'_X(0)$. Similarly:

$$G''_X(s) = 2p_2 + 6p_3 s + 12p_4 s^2 + \dots$$

So, $p_2 = \mathbb{P}(X = 2) = \frac{1}{2}G''_X(0)$. From here, we can easily continue with induction and finish the proof. □

Example 9. Let X be a discrete random variable with probability generating function $G_X(s) = \frac{s^2}{6}(2+4s)$. Find the distribution of X.

Note that in this case:

$$G_X(s) = \frac{s^2}{3} + \frac{2s^3}{3} \to G_X(0) = \mathbb{P}(X=0) = 0$$

$$G'_X(s) = \frac{2s}{3} + 2s^2 \to G'_X(0) = \mathbb{P}(X=1) = 0$$

$$G''_X(s) = \frac{2}{3} + 4s \to \frac{1}{2}G''_X(0) = \mathbb{P}(X=2) = \frac{1}{3}$$

$$G'''_X(s) = 4 \to \frac{1}{6}G'''_X(0) = \mathbb{P}(X=3) = \frac{2}{3}$$

$$\forall n \geq 4, G^n_X = 0, \text{ so it is 0 for all others}$$

Thus,

$$X = \begin{cases} 2 & \text{probability of } \frac{1}{3} \\ 3 & \text{probability of } \frac{2}{3} \end{cases}$$

4 Moments

Before we dive into the topic of moments of a random variable, we should probably define what a moment is:

Definition 10 (Moments)**.** For $k \geq 1$, the kth moment of a random variable X is $\mathbb{E}(X^k)$.

Note that it is important to note the relationship between the moments of a random variable and the variance.

Definition 11 (Variance)**.** The variance of a random variable X is:

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

Note that we can use linearity of expectation to show that the variance can be written in terms of moments because:

$$\text{var}(X) = \mathbb{E}(X - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

and both of the above are moments of the random variable X. However, seeing as this is an expository paper about Probability Generating Functions, we can also relate the moments of a random variable to the PGF of the random variable.

Definition 12 (Factorial Moment)**.** The kth factorial moment of a random variable X is:

$$E[X(X-1)\dots(X-k+1)]$$

Theorem 13. *The value of the kth derivative of the probability generating function at $s=1$ is equal to the kth factorial moment. This can be more formally stated for some random variable X and PGF $G_X(s)$ that:*

$$G^{(k)}_X(1) = E[X(X-1)\dots(X-k+1)]$$

Proof. Note that manipulating $G_X^{(k)}(s)$ yields:

$$G_X^{(k)}(s) = \frac{d^k G_X(s)}{ds^k} = \sum_{x=k}^{\infty} x(x-1)\dots(x-k+1)s^{x-k}p_x$$

Plugging in 1 to this yields:

$$G_X^{(k)}(1) = \sum_{x=k}^{\infty} x(x-1)\dots(x-k+1)p_x$$

The above is equal to the desired so we are done. □

5 Moment Generating Function

Now that we've mentioned moments of a random variable, let's take a short detour to discuss the moment generating function:

Definition 14. The **Moment Generating Function** (MGF) for some random variable X is defined as:

$$M_X(s) = \mathbb{E}[e^{sX}]$$

It's important to note that the MGF and the PGF are related as $M_X(s) = G_X(e^s)$.

Example 15. Find the moment generating function of random variable X with a uniform distribution over the interval of 0 to 1.

Note that:

$$M_X(s) = \mathbb{E}[e^{sX}] = \int_0^1 e^{sx}\,dx = \frac{e^s - 1}{s}$$

We know that $M_X(0) = 1$ so this means that the MGF in this example exists at all s in the real numbers.

However, let's now figure out why the "Moment Generating Function" is named after the moment.

Theorem 16. *For some random variable X such that M_X denotes its moment generating function, we have:*

$$M_X(s) = \sum_{k=0}^{\infty} \mathbb{E}[X^k]\frac{s^k}{k!}$$

Proof. To prove this, we'll proceed by using some calculus. Note that the Taylor series for e^x is:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

So:

$$e^{sX} = \sum_{k=0}^{\infty} \frac{s^k X^k}{k!}$$

Using this and relating it back to the MGF yields:

$$M_X(s) = \sum_{k=0}^{\infty} \mathbb{E}[X^k]\frac{s^k}{k!}$$

□

Ultimately, this means that we can use the Taylor series for the moment generating function to get all of the moments of the random variable.

6 Summing Independent Random Variables

Theorem 17. *If X and Y are independent random variables:*

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$

where G denotes the probability generating function.

Proof. Note:

$$G_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X)\mathbb{E}(s^Y) = G_X(s)G_Y(s)$$

□

Combining a bunch of independent random variables $X_1, X_2, \ldots, X_n$ with their sum as S yields:

$$G_S(s) = G_{X_1}(s)\ldots G_{X_n}(s)$$

Theorem 18 (Random Sum Formula)**.** *For independent random variables $N, X_1, \ldots, X_N$, if the X_i values are distributed with G_X, the probability generating function for the sum of the Xs is:*

$$G_S(s) = G_N(G_X(s))$$

Proof. Using the partition theorem with $B_n = \{N = n\}$ we have:

$$G_S(s) = \mathbb{E}(s^S) = \sum_{n=0}^{\infty} \mathbb{E}(s^S|N = n)\mathbb{P}(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(s^S)\mathbb{P}(N = n)$$

$$= \sum_{n=0}^{\infty} G_X(s)^n\mathbb{P}(N = n) = G_N(G_X(s))$$

□

Note that for $s = 1$ we have $G'_S(1) = G'_N(1)G'_X(1)$ which by **Theorem** 13 is the same as $\mathbb{E}(S) = \mathbb{E}(N)\mathbb{E}(X)$.

7 Radius of Convergence

Just as in the section on moments, let's start by actually defining the thing in question: a radius of convergence.

Definition 19. Radius of convergence of a generating function is some $R > 0$ such that:

$$G_X(s) = \sum_{x=0}^{\infty} s^x\mathbb{P}[X = x]$$

converges is $|s| < R$ and diverges if $|s| > R$.

Here are just some facts about the radius of convergence and PGFs:

Fact 20. The radius of convergence of every PGF exists.

Fact 21. The radius of convergence is always ≥ 1 and can be anything from 1 to ∞.

Now that we're on the subject of radii of convergence, let's just use it to deal with proving uniqueness for probability generating functions.

Theorem 22 (Uniqueness)**.** *For random variables X and Y, $G_X(s) = G_Y(s)$ for all s if and only if:*

$$\mathbb{P}[X = k] = \mathbb{P}[Y = k]$$

Essentially, probability generating functions are unique to probability mass functions.

Proof. Note that because G_X and G_Y both have radii of convergence that are at least 1, their power series expansions about the origin are unique if their probability mass functions are distinct. However, if $G_X = G_Y$, the two power series have equal coefficients. □

8 Applications

Now that we've discussed a lot about the theory of probability generating functions, let's talk about how they are applied in different areas of study.

8.1 Inversions of Permutations

Definition 23 (Inversions)**.** A pair of numbers in a permutation of $(1, 2, \ldots, n)$ is considered an inversion if the order of the two numbers is swapped before and after permuting.

For example, if we have $(1, 2, 3)$ and go to $(1, 3, 2)$, the only inversion is $(2, 3)$ as 1 remains in a position before 2 and 3.

Definition 24. Let $I_n(k)$ denote the number of permutations of an n number set with k inversions.

Note that if X_n is the random variable of the number of inversions in a random permutation of n variables:

$$G_{X_n}(s) = \sum_{k=0}^{\binom{n}{2}} \mathbb{P}[X_n = k]s^k = \sum_{k}^{\binom{n}{2}} \frac{I_n(k)}{n!} s^k$$

Note that because you can take a permutation of $n - 1$ and then insert the nth element, you can relate G_{X_n} and $G_{X_{n-1}}$:

Fact 25.

$$G_{X_n}(s) = \frac{1 + s + s^2 + \ldots + s^{n-1}}{n} G_{X_{n-1}}(s)$$

when $n > 1$ and $G_{X_1}(s) = 1$.

Expanding this out yields ([3]):

$$G_{X_n}(s) = \frac{1}{n!} \prod_{k=1}^{n} \frac{1 - s^k}{1 - s}$$

Now, differentiating and evaluating at 1 this probability generating function allows us to find the expected values of the number of inversions in a permutation.

We can actually use this to find the expected value. We have:

$$G_{X_n}(s) = \prod_{k=1}^{n} \frac{1 + s + s^2 + \ldots + s^{k-1}}{k}$$

Note that deriving this and plugging in $s = 1$ yields:

$$\mathbb{E}[X_n] = \frac{n(n-1)}{4}$$

8.2 Random Walks

Definition 26 (One dimensional random walk)**.** The one dimensional random walk is defined as a series of random movements where it starts at the point 0 and increases 1 with probability $\frac{1}{2}$ and decreases 1 with probability $\frac{1}{2}$ at any point.

Let's do an example of finding the PGF for a random walk.

Example 27. Let $G(s)$ be the probability generating function on the number of steps needed to go from 0 to 1 which we will refer to as N. Find $G(s)$.

Note that we have (for first step Y_1):

$$G(s) = \mathbb{E}(s^N) = \mathbb{E}(s^N|Y_1 = 1)\mathbb{P}(Y_1 = 1) + \mathbb{E}(s^N|Y_1 = -1)\mathbb{P}(Y_1 = -1)$$
$$= \frac{1}{2}\left(\mathbb{E}(s^N|Y_1 = 1) + \mathbb{E}(s^N|Y_1 = -1)\right)$$

Note that we have $\mathbb{E}(s^N|Y_1 = 1) = s$ and $\mathbb{E}(s^N|Y_1 = -1) = s(G(s))^2$. This means that:

$$G(s) = \frac{1}{2}\left(s + s(G(s))^2\right)$$

Solving this out yields:

$$G(s) = \frac{1 - \sqrt{1 - s^2}}{2}$$

References

[1] Rachel Fewster. *Chapter 4: Generating Functions*. https://www.stat.auckland.ac.nz/~fewster/325/notes/ch4.pdf. University of Auckland.

[2] Geoffrey Grimmet and Dominic James Anthony Welsh. *Probability: An Introduction*. Oxford University Press, 2017.

[3] Andreas Klappenecker. *Probability Generating Functions*. https://people.engr.tamu.edu/andreas-klappenecker/csce658-s18/genfunc.pdf. Texas Aamp;M university.

[4] *Probability Generating Functions*. https://www.cl.cam.ac.uk/teaching/0708/Probabilty/prob06.pdf. University of Cambridge.