

An Introduction to Generating Functions in Probability

Undergraduate Mathematics Exposition

April 15, 2025

1 Introduction: Beyond PMFs and PDFs

When studying random variables, the Probability Mass Function (PMF) for discrete variables or the Probability Density Function (PDF) for continuous variables provides a complete description of the distribution. However, sometimes we want different ways to represent or summarize this information. Generating functions provide powerful alternative representations. They "package" the probability distribution into a single function, and operations on this function (like differentiation) can reveal important properties of the original distribution, such as its moments, or simplify problems involving sums of independent random variables.

We will explore three main types:

- **Probability Generating Functions (PGFs):** For discrete variables on non-negative integers.
- **Moment Generating Functions (MGFs):** For discrete or continuous variables, but may not always exist.
- * **Characteristic Functions (CFs):** For any variable; always exists and is fundamental theoretically.

2 Probability Generating Functions (PGFs)

2.1 Motivation and Definition

Consider a discrete random variable X that can only take non-negative integer values, $k \in \{0, 1, 2, \dots\}$. How can we encode its entire PMF, $\Pr(X = k)$, into a single function? The PGF achieves this by using the probabilities as coefficients in a power series.

Definition 2.1 (Probability Generating Function (PGF)). *Let X be a discrete random variable taking values in $\{0, 1, 2, \dots\}$ with PMF $p_k = \Pr(X = k)$. The **Probability Generating Function (PGF)** of X is defined as:*

$$P_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} \Pr(X = k) s^k = p_0 s^0 + p_1 s^1 + p_2 s^2 + \dots \quad (1)$$

where s is a real (or complex) variable. This power series converges at least for $|s| \leq 1$.

Remark 2.2. *The variable s is often treated as a "dummy" variable. The function $P_X(s)$ itself contains the information about the distribution.*

2.2 Key Properties and Uses

Property 2.3 (Basic Properties).

1. $P_X(1) = \sum_{k=0}^{\infty} \Pr(X = k)(1)^k = \sum_{k=0}^{\infty} \Pr(X = k) = 1$. (The sum of all probabilities is 1).
2. $P_X(0) = \Pr(X = 0)(0)^0 + \Pr(X = 1)(0)^1 + \dots = \Pr(X = 0)$ (assuming $0^0 = 1$). The value at $s = 0$ gives the probability of the outcome $k = 0$.

Property 2.4 (Generating Probabilities). The PGF is named because it "generates" the probabilities via differentiation at $s = 0$. Recall the Taylor series expansion of $P_X(s)$ around $s = 0$:

$$P_X(s) = P_X(0) + \frac{P'_X(0)}{1!}s + \frac{P''_X(0)}{2!}s^2 + \dots + \frac{P_X^{(k)}(0)}{k!}s^k + \dots$$

Comparing this to the definition $P_X(s) = \sum_{k=0}^{\infty} \Pr(X = k)s^k$, we see by matching coefficients of s^k :

$$\Pr(X = k) = \frac{P_X^{(k)}(0)}{k!} \quad (2)$$

where $P_X^{(k)}(0)$ is the k -th derivative of $P_X(s)$ evaluated at $s = 0$.

Property 2.5 (Factorial Moments). The derivatives of the PGF at $s = 1$ generate the **factorial moments** of X . The k -th factorial moment is defined as $E[X^{(k)}] = \mathbb{E}[X(X-1)\dots(X-k+1)]$. Differentiating the PGF k times and setting $s = 1$:

$$\begin{aligned} P'_X(s) &= \frac{d}{ds} \sum_{k=0}^{\infty} p_k s^k = \sum_{k=1}^{\infty} k p_k s^{k-1} \\ \implies P'_X(1) &= \sum_{k=1}^{\infty} k p_k = \mathbb{E}[X] = \mathbb{E}[X^{(1)}] \\ P''_X(s) &= \frac{d^2}{ds^2} \sum_{k=0}^{\infty} p_k s^k = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} \\ \implies P''_X(1) &= \sum_{k=2}^{\infty} k(k-1) p_k = \mathbb{E}[X(X-1)] = \mathbb{E}[X^{(2)}] \\ &\vdots \\ P_X^{(k)}(1) &= \sum_{j=k}^{\infty} j(j-1)\dots(j-k+1) p_j = \mathbb{E}[X(X-1)\dots(X-k+1)] = \mathbb{E}[X^{(k)}] \end{aligned}$$

So, we have the important result:

$$\mathbb{E}[X(X-1)\dots(X-k+1)] = P_X^{(k)}(1) \quad (3)$$

Remark 2.6 (Why Factorial Moments?). Factorial moments are useful primarily because:

- They are calculated directly from PGF derivatives at $s = 1$.

- They provide a convenient way to find the variance. Since $E[X^{(2)}] = E[X(X-1)] = E[X^2] - E[X]$, we have $E[X^2] = E[X^{(2)}] + E[X]$. The variance is then:

$$\begin{aligned}\text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= (E[X^{(2)}] + E[X]) - (E[X])^2 \\ &= P_X''(1) + P_X'(1) - (P_X'(1))^2\end{aligned}\tag{4}$$

This allows calculating the variance using only the first two derivatives of the PGF at $s = 1$.

- For some distributions (like Poisson), factorial moments have simpler forms than raw moments ($E[X^k]$).

Example 2.7 (Geometric Distribution PGF and Moments). Let X be the number of Bernoulli trials until the first success, where the probability of success on each trial is p . X takes values $k = 1, 2, 3, \dots$ with PMF $\Pr(X = k) = (1-p)^{k-1}p$. (Note: This is different from the Geometric starting at $k = 0$ which counts failures before success).

The PGF is calculated as:

$$\begin{aligned}P_X(s) &= \mathbb{E}[s^X] = \sum_{k=1}^{\infty} s^k \Pr(X = k) = \sum_{k=1}^{\infty} s^k (1-p)^{k-1} p \\ &= ps \sum_{k=1}^{\infty} s^{k-1} (1-p)^{k-1} \quad (\text{Let } j = k-1) \\ &= ps \sum_{j=0}^{\infty} [s(1-p)]^j \\ &= ps \left(\frac{1}{1-s(1-p)} \right) \quad (\text{Geometric series, requires } |s(1-p)| < 1) \\ &= \frac{ps}{1-s(1-p)}, \quad \text{for } |s| < 1/(1-p)\end{aligned}$$

Now, let's use this PGF to find the mean and variance. **Mean:** We need $P_X'(1)$.

$$P_X'(s) = \frac{p}{(1-s(1-p))^2} \quad (\text{Calculated earlier})$$

$$E[X] = P_X'(1) = \frac{p}{(1-(1-p))^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}$$

Variance: We need $P_X''(1)$ and use Eq. (4).

$$P_X''(s) = \frac{2p(1-p)}{(1-s(1-p))^3} \quad (\text{Calculated earlier})$$

$$E[X(X-1)] = P_X''(1) = \frac{2p(1-p)}{(1-(1-p))^3} = \frac{2p(1-p)}{p^3} = \frac{2(1-p)}{p^2}$$

Using the variance formula:

$$\text{Var}(X) = P_X''(1) + P_X'(1) - (P_X'(1))^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{2-2p+p-1}{p^2} = \boxed{\frac{1-p}{p^2}}$$

These match the known mean and variance for the Geometric(p) distribution starting from $k = 1$.

Property 2.8 (Sums of Independent Random Variables). *If X_1, X_2, \dots, X_n are independent random variables (taking non-negative integer values) with PGFs $P_1(s), \dots, P_n(s)$, and $Y = X_1 + \dots + X_n$, then the PGF of Y is the product of the individual PGFs:*

$$P_Y(s) = \mathbb{E}[s^Y] = \mathbb{E}[s^{X_1 + \dots + X_n}] = \mathbb{E}[s^{X_1} \dots s^{X_n}] = \mathbb{E}[s^{X_1}] \dots \mathbb{E}[s^{X_n}] = P_1(s) \dots P_n(s) \quad (5)$$

The crucial step $\mathbb{E}[s^{X_1} \dots s^{X_n}] = \mathbb{E}[s^{X_1}] \dots \mathbb{E}[s^{X_n}]$ relies on the independence of the X_i .

Example: If X_1, \dots, X_r are i.i.d. Geometric(p) (as defined above, $X_i \geq 1$), then $Y = X_1 + \dots + X_r$ is the number of trials for r successes (Negative Binomial). Its PGF is $P_Y(s) = [P_X(s)]^r = \left(\frac{ps}{1-(1-p)s} \right)^r$.

3 Moment Generating Functions (MGFs)

3.1 Motivation and Definition

Can we generalize the idea of generating functions to continuous random variables, or discrete ones not restricted to non-negative integers? The MGF does this, using an exponential function instead of s^X . It's particularly suited for generating raw moments.

Definition 3.1 (Moment Generating Function (MGF)). *Let X be a random variable (discrete or continuous). The **Moment Generating Function (MGF)** of X is defined as:*

$$M_X(t) = \mathbb{E}[e^{tX}] \quad (6)$$

provided this expectation exists (is finite) for t in some open interval $(-\delta, \delta)$ containing 0. If the expectation does not exist in such an interval, we say the MGF does not exist.

3.2 Existence Condition - A Crucial Point

Unlike PGFs (which always converge for $|s| \leq 1$) and CFs (which always exist), **MGFs are not guaranteed to exist**. The expectation $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$ (for continuous X) or $\sum_k e^{tk} \Pr(X = k)$ (for discrete X) involves the term e^{tx} . If $t \neq 0$, this term grows exponentially as $|x| \rightarrow \infty$. For the expectation to be finite, the tails of the distribution ($f(x)$ or $\Pr(X = k)$ for large $|k|$) must decay sufficiently fast to counteract this exponential growth.

Example 3.2 (Non-existence). *The Cauchy distribution, with PDF $f(x) = \frac{1}{\pi(1+x^2)}$, has "heavy tails" that decay like $1/x^2$. This is not fast enough to overcome the growth of e^{tx} , and its MGF does not exist for any $t \neq 0$.*

3.3 Key Properties and Uses

Property 3.3 (Generating Raw Moments). *If the MGF $M_X(t)$ exists in an interval around $t = 0$, it generates the raw moments $E[X^k]$ via differentiation at $t = 0$. Consider the Taylor expansion of e^{tX} :*

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$$

Taking expectations (assuming we can swap expectation and summation):

$$M_X(t) = \mathbb{E}[e^{tX}] = 1 + t\mathbb{E}[X] + \frac{t^2 \mathbb{E}[X^2]}{2!} + \frac{t^3 \mathbb{E}[X^3]}{3!} + \dots$$

This is the Taylor series for $M_X(t)$ around $t = 0$. By matching coefficients with the general Taylor formula $M_X(t) = \sum_{k=0}^{\infty} \frac{M_X^{(k)}(0)}{k!} t^k$, we find:

$$\mathbb{E}[X^k] = M_X^{(k)}(0) \quad (7)$$

where $M_X^{(k)}(0)$ is the k -th derivative of $M_X(t)$ evaluated at $t = 0$.

Property 3.4 (Sums of Independent Random Variables). *Similar to PGFs, if X_1, \dots, X_n are independent with MGFs $M_1(t), \dots, M_n(t)$, and $Y = X_1 + \dots + X_n$, then (provided all MGFs exist):*

$$M_Y(t) = M_1(t) \dots M_n(t) \quad (8)$$

Remark 3.5. *When the MGF exists, it uniquely determines the distribution (if two RVs have the same MGF in an interval around 0, they have the same distribution). However, this uniqueness property is less general than that of the CF, because the MGF might not exist.*

4 Characteristic Functions (CFs)

4.1 Motivation and Definition

The MGF's potential non-existence is a significant limitation, especially for theoretical work. The Characteristic Function overcomes this by introducing a complex exponential. It is the most general and theoretically important of the three generating functions.

Definition 4.1 (Characteristic Function (CF)). *Let X be any random variable (discrete or continuous). The **Characteristic Function (CF)** of X is defined as:*

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i \sin(tX)] \quad (9)$$

where t is a real variable and $i = \sqrt{-1}$.

4.2 Existence - The Key Advantage

Theorem 4.2 (Existence of CF). *The characteristic function $\phi_X(t)$ exists and is finite for **all** real values of t for **any** random variable X .*

Sketch. We need to show the expectation $\mathbb{E}[e^{itX}]$ converges. Consider the magnitude of the term inside the expectation:

$$|e^{itX}| = |\cos(tX) + i \sin(tX)| = \sqrt{\cos^2(tX) + \sin^2(tX)} = \sqrt{1} = 1$$

The magnitude is always 1, regardless of X or t . Therefore, when taking the expectation (integrating or summing against the PDF/PMF $f(x)$), we have:

$$|\phi_X(t)| = |\mathbb{E}[e^{itX}]| \leq \mathbb{E}[|e^{itX}|] = \mathbb{E}[1] = 1$$

For the integral definition (continuous case): $\int_{-\infty}^{\infty} |e^{itx} f(x)| dx = \int_{-\infty}^{\infty} |e^{itx}| f(x) dx = \int_{-\infty}^{\infty} 1 \cdot f(x) dx = 1$. Since the integral of the absolute value converges, the integral itself converges. A similar argument holds for the discrete case. \square

This universal existence contrasts sharply with the MGF and makes the CF suitable for all distributions, including heavy-tailed ones.

4.3 Key Properties and Uses

Property 4.3 (Relation to Cosine and Sine Averages). *From Euler's formula, we can write:*

$$\phi_X(t) = \mathbb{E}[\cos(tX)] + i \mathbb{E}[\sin(tX)] \quad (10)$$

So, the real part of the CF is the average value of $\cos(tX)$, and the imaginary part is the average value of $\sin(tX)$.

- $\mathbb{E}[\cos(tX)]$ reflects the "even" or symmetric characteristics of the distribution relative to frequency t .
- $\mathbb{E}[\sin(tX)]$ reflects the "odd" or asymmetric characteristics of the distribution relative to frequency t .
- If the distribution of X is symmetric about 0 (i.e., X and $-X$ have the same distribution), then $\mathbb{E}[\sin(tX)] = 0$, and the CF $\phi_X(t)$ is purely real. This is because $\sin(t(-x)) = -\sin(tx)$, causing cancellation in the expectation for symmetric distributions.

The parameter t acts as a frequency, determining how rapidly the angle tX changes relative to X . The term tX itself is a random angle (or phase), and the CF captures the average behavior of e^{itX} over the distribution of these random angles.

Property 4.4 (Generating Raw Moments). *If $E[|X|^k] < \infty$, then the k -th derivative of $\phi_X(t)$ exists, and the raw moments can be found using:*

$$\mathbb{E}[X^k] = \frac{\phi_X^{(k)}(0)}{i^k} \quad (11)$$

This comes from differentiating $\phi_X(t) = E[e^{itX}]$ with respect to t and evaluating at $t = 0$. For example, $\phi_X'(t) = E[iXe^{itX}] \implies \phi_X'(0) = E[iX] = iE[X]$, so $E[X] = \phi_X'(0)/i$.

Property 4.5 (Sums of Independent Random Variables). *As with PGFs and MGFs, if X_1, \dots, X_n are independent with CFs $\phi_1(t), \dots, \phi_n(t)$, and $Y = X_1 + \dots + X_n$, then:*

$$\phi_Y(t) = \phi_1(t) \dots \phi_n(t) \quad (12)$$

Property 4.6 (Uniqueness and Continuity Theorems). • **Uniqueness Theorem:** *The characteristic function $\phi_X(t)$ uniquely determines the probability distribution of X . If $\phi_X(t) = \phi_Y(t)$ for all t , then X and Y have the same distribution. (There are inversion formulas to recover the CDF/PDF/PMF from the CF).*

- **Lévy's Continuity Theorem:** *Convergence of a sequence of distribution functions is equivalent to the pointwise convergence of the corresponding sequence of characteristic functions. This is the cornerstone for proving limit theorems using CFs.*

These properties make the CF the primary tool for rigorous theoretical work, especially concerning convergence of random variables (like the Central Limit Theorem).

5 When to Use Which Function?

The choice depends on the random variable and the goal:

1. **Is the RV discrete taking values $0, 1, 2, \dots$?**

- Use the **PGF** ($E[s^X]$) if you need probabilities ($P(X = k)$) or factorial moments easily, or if dealing with sums/branching processes involving such variables.

2. **Is the RV discrete or continuous (and not covered by PGF)?**

- Does the **MGF** ($E[e^{tX}]$) exist (i.e., tails are sufficiently light)?
 - **Yes:** The MGF is often convenient for finding raw moments ($E[X^k]$) and dealing with sums of independent variables, potentially involving simpler calculations than the CF.
 - **No / Unsure:** Use the CF.
- Do you need guaranteed existence, uniqueness, or are you doing theoretical work (limit theorems, proofs)?
 - **Yes:** Use the **CF** ($E[e^{itX}]$). It always exists, uniquely determines the distribution, and is the standard tool for proving convergence theorems.

In short: PGF for non-negative integers (probabilities/factorial moments), MGF for convenience with moments/sums if it exists, CF for universal applicability and theoretical rigor.

6 Conclusion

Generating functions (PGF, MGF, CF) are indispensable tools in probability theory. They provide alternative ways to represent distributions, often simplifying the calculation of moments and the analysis of sums of independent random variables. While the PGF is specific to non-negative integer variables and the MGF requires an existence condition, the Characteristic Function offers a universally applicable and theoretically robust framework, underpinning many fundamental results in the field. Understanding the properties and appropriate uses of each type of generating function is key to leveraging their power effectively.