Probability Generating Functions and Moment Generating Functions

Contents

1 Overview

A random variable can be described in several equivalent or complementary ways. The cumulative distribution function (CDF) describes accumulated probability, a probability mass function (PMF) describes point probabilities for a discrete random variable, and a probability density function (PDF) describes probability through area for a continuous random variable.

Probability generating functions (PGFs) and moment generating functions (MGFs) give another viewpoint: they encode information about the distribution inside a function. For nonnegative integer-valued random variables, the PGF encodes probabilities as coefficients. The MGF, when it exists around zero, encodes raw moments through derivatives.

The main ideas are:

$$\boxed{\text{PGF: probabilities as coefficients.}} \qquad \boxed{\text{MGF: moments by derivatives.}}$$

This note is written to be self-contained: it starts from CDFs, PMFs, PDFs, and expectation, then develops PGFs and MGFs, examples, advanced remarks, and worked problems.

2 Why do we need PGF and MGF?

Suppose we have a random variable X. The most direct way to describe X is to list all probabilities or give a density. For simple examples, this is fine. But in probability theory, actuarial science, statistics, branching processes, random sums, and limit theorems, listing probabilities can become painful.

Generating functions help because they turn a probability problem into an algebra problem.

2.1 Three big reasons

1. **They compress a distribution into one function.** Instead of carrying around infinitely many probabilities $p_0, p_1, p_2, \ldots$, the PGF stores them inside one power series.
2. **They make sums easier.** If X and Y are independent, then

   $$M_{X+Y}(t) = M_X(t)M_Y(t), \qquad G_{X+Y}(s) = G_X(s)G_Y(s).$$

   Convolution in probability becomes multiplication in generating-function language.
3. **They produce moments.** The MGF gives $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\mathbb{E}[X^3]$, and so on through derivatives. The PGF gives factorial moments, which are very useful for counting distributions.

Remark 2.1 (Important warning)**.** *PGFs and MGFs are not probabilities themselves. A probability must be between* 0 *and* 1*, but an expectation, a moment, an MGF value, or a variance can be larger than* 1*. For example, if* X *is the number of claims in an insurance portfolio, then* $\mathbb{E}[X]$ *could be* 1000*, not because probability is larger than* 1*, but because expectation is an average value, not a probability.*

3 CDF, PMF, and PDF before PGF and MGF

3.1 CDF: the universal description

Definition 3.1 (Cumulative distribution function)**.** *For any real-valued random variable* X*, the cumulative distribution function, or CDF, is*

$$F_X(x) = \mathbb{P}(X \le x), \qquad x \in \mathbb{R}.$$

The CDF works for every random variable: discrete, continuous, mixed, or unusual. It always answers the same question:

$$\boxed{F_X(x) = \text{probability that } X \text{ is at most } x.}$$

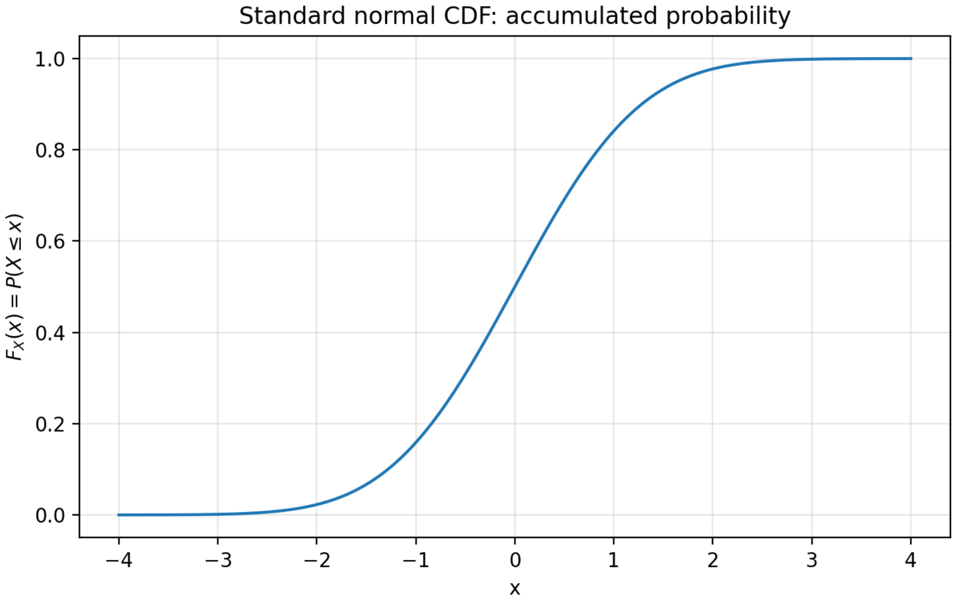

Figure 1: A CDF accumulates probability from the left. It is always nondecreasing and ranges from 0 to 1.

3.2 PMF: for discrete random variables

Definition 3.2 (Probability mass function)**.** *If X is discrete, its probability mass function, or PMF, is*

$$p_X(k) = \mathbb{P}(X = k).$$

For example, if $X \sim \text{Poisson}(3)$, then

$$\mathbb{P}(X = k) = e^{-3}\frac{3^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

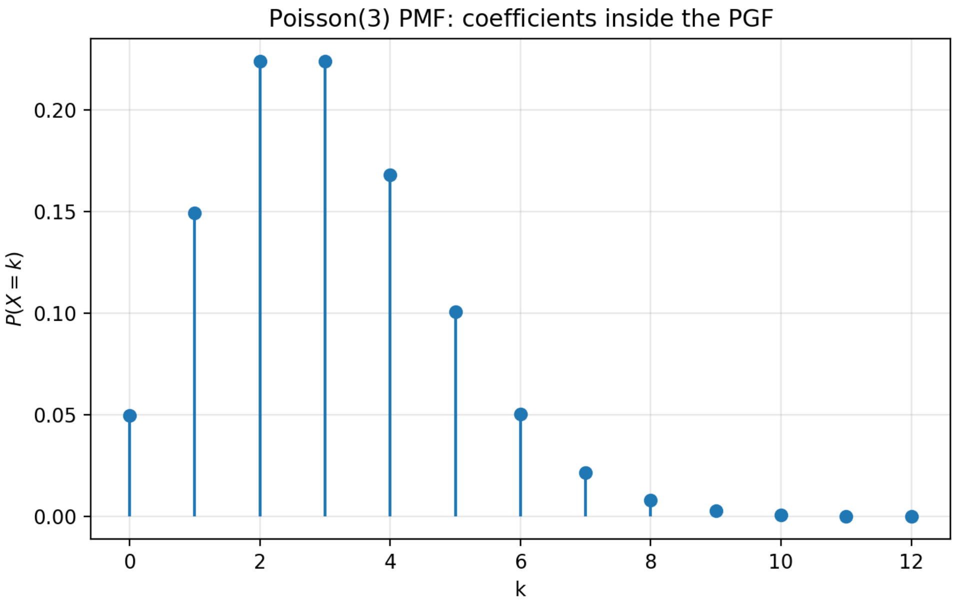

Figure 2: A PMF gives probability at each possible discrete value. In a PGF, these probabilities become coefficients.

3.3 PDF: for continuous random variables

Definition 3.3 (Probability density function)**.** *If X is continuous with density f_X, then probabilities are computed by integrals:*

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,dx.$$

A common mistake is to think $f_X(x)$ itself is a probability. It is not. The probability at exactly one point is usually zero:

$$\mathbb{P}(X = x) = 0$$

for continuous random variables. Probability comes from area:

$$\boxed{\text{probability} = \text{area under the PDF.}}$$

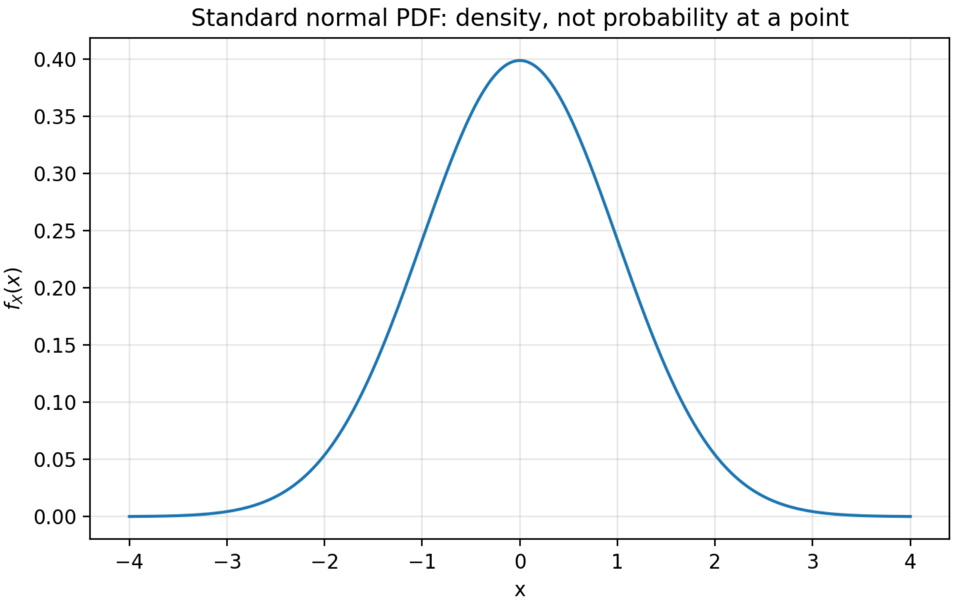

Figure 3: A PDF can be larger than 1 in some distributions. That is allowed because density is not probability. Probability is area.

4 Expectation and moments

4.1 Expectation

For a function g, the expectation of $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_k g(k) p_X(k)$$

for a discrete random variable, and

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

for a continuous random variable.

More generally, in measure-theoretic probability,

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega).$$

This general definition includes both sums and integrals as special cases.

4.2 Moments

Definition 4.1 (Raw and central moments)**.** *The nth raw moment of X is*

$$\mu_n' = \mathbb{E}[X^n].$$

The nth central moment is

$$\mu_n = \mathbb{E}[(X - \mathbb{E}[X])^n].$$

The first raw moment is the mean:

$$\mathbb{E}[X].$$

The second central moment is the variance:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Moments describe shape:

Moment	Formula	Meaning
First raw moment	$\mathbb{E}[X]$	average / center
Second central moment	$\mathrm{Var}(X)$	spread
Third central moment	$\mathbb{E}[(X-\mu)^3]$	skewness direction
Fourth central moment	$\mathbb{E}[(X-\mu)^4]$	tail heaviness / kurtosis

5 Probability generating function

5.1 Definition

Definition 5.1 (Probability generating function)**.** *Let X be a nonnegative integer-valued random variable:*

$$X \in \{0, 1, 2, \dots\}.$$

The probability generating function, or PGF, of X is

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

Using the PMF, this becomes

$$G_X(s) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) s^k.$$

This is why PGF is called a probability generating function: it is a power series whose coefficients are probabilities.

$$\boxed{\text{coefficient of } s^k \text{ in } G_X(s) = \mathbb{P}(X = k).}$$

5.2 Recovering probabilities from a PGF

If

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \cdots ,$$

then

$$p_k = \mathbb{P}(X = k).$$

Equivalently, by Taylor's formula,

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

5.3 Basic properties

Proposition 5.1 (Basic PGF facts)**.** *If X is a nonnegative integer-valued random variable with PGF G_X, then:*

1. $G_X(1) = 1$.
2. $G_X(0) = \mathbb{P}(X = 0)$.
3. $G'_X(1) = \mathbb{E}[X]$, *if the expectation is finite.*
4. $G''_X(1) = \mathbb{E}[X(X-1)]$, *if the second factorial moment is finite.*

Proof. Starting from

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k,$$

we plug in $s = 1$:

$$G_X(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Also,

$$G_X(0) = p_0.$$

Differentiating,

$$G'_X(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}.$$

Thus,

$$G'_X(1) = \sum_{k=1}^{\infty} k p_k = \mathbb{E}[X].$$

Differentiating again,

$$G''_X(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2},$$

so

$$G''_X(1) = \sum_{k=2}^{\infty} k(k-1) p_k = \mathbb{E}[X(X-1)].$$

□

From these,

$$\mathbb{E}[X] = G'_X(1),$$

and

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = G''_X(1) + G'_X(1).$$

Therefore,

$$\boxed{\operatorname{Var}(X) = G''_X(1) + G'_X(1) - \big(G'_X(1)\big)^2.}$$

5.4 PGF and sums

Proposition 5.2 (PGF of a sum)**.** *If X and Y are independent nonnegative integer-valued random variables, then*

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

Proof. By definition,

$$G_{X+Y}(s) = \mathbb{E}[s^{X+Y}].$$

Since $s^{X+Y} = s^X s^Y$,

$$G_{X+Y}(s) = \mathbb{E}[s^X s^Y].$$

Independence gives

$$\mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y].$$

Thus,

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

□

This is one of the main reasons PGFs are useful. Normally, finding the distribution of a sum requires convolution:

$$\mathbb{P}(X+Y=k) = \sum_{j=0}^{k} \mathbb{P}(X=j)\mathbb{P}(Y=k-j).$$

PGFs turn this convolution into multiplication.

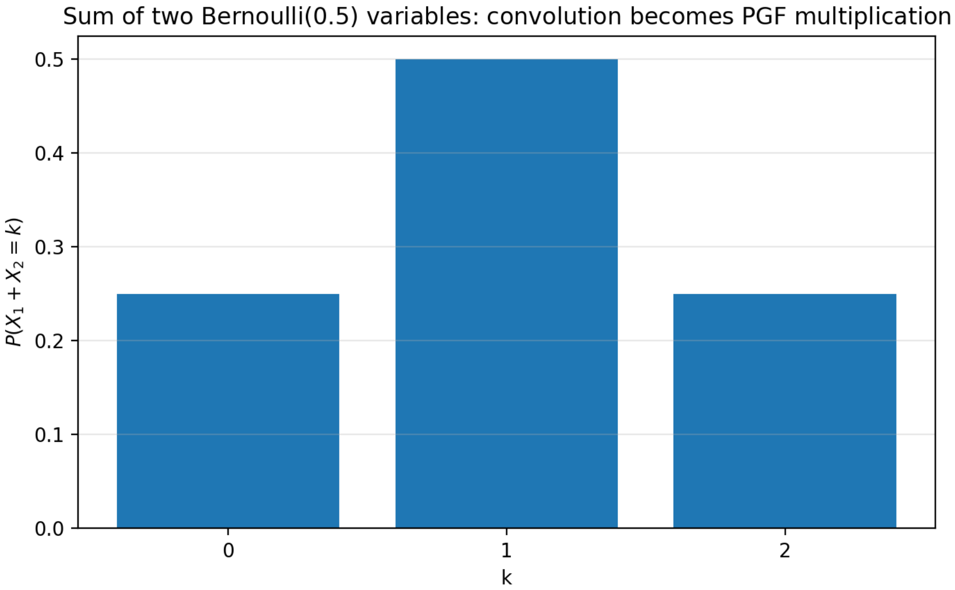

Figure 4: For independent discrete variables, adding variables corresponds to multiplying PGFs. This is why PGFs simplify counting and convolution problems.

6 Moment generating function

6.1 Definition

Definition 6.1 (Moment generating function)**.** *The moment generating function, or MGF, of a real-valued random variable X is*

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

for all real t where the expectation exists and is finite.

The MGF is called "moment generating" because derivatives at $t = 0$ generate moments.

6.2 Why derivatives give moments

The exponential function has the Taylor expansion

$$e^{tX} = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots .$$

Taking expectation formally,

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{n=0}^{\infty} \frac{\mathbb{E}[X^n]}{n!} t^n.$$

Therefore, the coefficient of $t^n/n!$ is $\mathbb{E}[X^n]$.

Thus,

$$\boxed{M_X^{(n)}(0) = \mathbb{E}[X^n].}$$

In particular,

$$M_X'(0) = \mathbb{E}[X]$$

and

$$M_X''(0) = \mathbb{E}[X^2].$$

Then

$$\boxed{\operatorname{Var}(X) = M_X''(0) - \big(M_X'(0)\big)^2.}$$

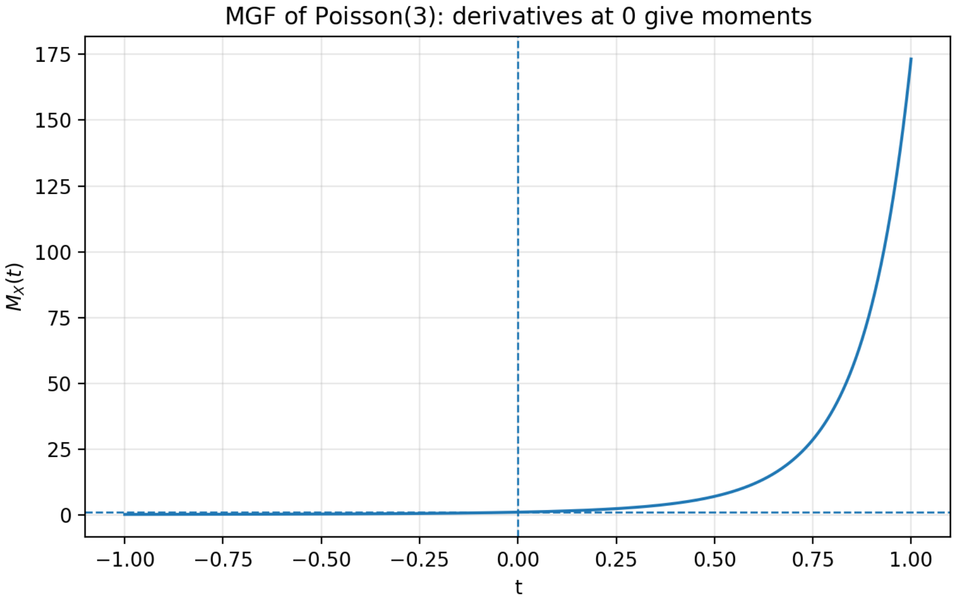

Figure 5: The MGF of a Poisson(3) variable is $M_X(t) = \exp(3(e^t - 1))$. Its derivatives at $t = 0$ give moments.

6.3 MGF and sums

Proposition 6.1 (MGF of a sum)**.** *If X and Y are independent random variables, then*

$$M_{X+Y}(t) = M_X(t)M_Y(t),$$

for all t where the expectations exist.

Proof.

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}e^{tY}].$$

By independence,

$$\mathbb{E}[e^{tX}e^{tY}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}].$$

Thus,

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

□

6.4 Existence matters

Not every random variable has an MGF in an open interval around 0.

- The Cauchy distribution has no finite mean and no MGF around 0.
- The lognormal distribution has all positive integer moments finite, but its MGF is infinite for every $t > 0$.
- The characteristic function, discussed later, always exists because $|e^{itX}| = 1$.

This is a advanced subtlety: moments alone do not always determine a distribution, but if the MGF exists on an open interval around 0, then it uniquely determines the distribution.

Theorem 6.1 (Uniqueness from MGF)**.** *If $M_X(t)$ exists and is finite for all t in an open interval around 0, then the MGF uniquely determines the distribution of X.*

7 Relationship between PGF and MGF

If X is nonnegative integer-valued, then both PGF and MGF can be used. Since

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

we get

$$\boxed{M_X(t) = G_X(e^t).}$$

Equivalently,

$$\boxed{G_X(s) = M_X(\log s), \qquad s > 0.}$$

So for counting random variables, PGF and MGF are closely connected. The PGF is often more natural because it directly stores probabilities as coefficients.

8 Examples by distribution

8.1 Bernoulli distribution

Let

$$X \sim \text{Bernoulli}(p),$$

so

$$\mathbb{P}(X = 1) = p, \qquad \mathbb{P}(X = 0) = 1 - p.$$

The PGF is

$$G_X(s) = \mathbb{E}[s^X] = (1 - p)s^0 + ps^1 = 1 - p + ps.$$

The MGF is

$$M_X(t) = \mathbb{E}[e^{tX}] = (1 - p)e^0 + pe^t = 1 - p + pe^t.$$

Now

$$M_X'(t) = pe^t,$$

so

$$M_X'(0) = p.$$

Thus,

$$\mathbb{E}[X] = p.$$

Also,

$$M_X''(t) = pe^t,$$

so

$$M_X''(0) = p.$$

Therefore,

$$\text{Var}(X) = p - p^2 = p(1 - p).$$

8.2 Binomial distribution

Let

$$X \sim \text{Binomial}(n, p).$$

We can think of X as a sum of independent Bernoulli variables:

$$X = X_1 + \cdots + X_n, \qquad X_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p).$$

Because PGFs multiply under independent sums,

$$G_X(s) = \prod_{i=1}^{n} G_{X_i}(s) = (1 - p + ps)^n.$$

Similarly,

$$M_X(t) = (1 - p + pe^t)^n.$$

This is much faster than deriving everything from scratch.

8.3 Geometric distribution

Let X be the number of failures before the first success, where success probability is p. Then

$$\mathbb{P}(X = k) = (1-p)^k p, \qquad k = 0, 1, 2, \ldots.$$

The PGF is

$$G_X(s) = \sum_{k=0}^{\infty} p(1-p)^k s^k.$$

Using the geometric series formula,

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}, \qquad |r| < 1,$$

we get

$$G_X(s) = \frac{p}{1 - (1-p)s}.$$

The MGF is

$$M_X(t) = G_X(e^t) = \frac{p}{1 - (1-p)e^t},$$

valid when

$$(1-p)e^t < 1.$$

8.4 Poisson distribution

Let

$$X \sim \text{Poisson}(\lambda).$$

Then

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

The PGF is

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

Since

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

we obtain

$$G_X(s) = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}.$$

The MGF is

$$M_X(t) = G_X(e^t) = e^{\lambda(e^t-1)}.$$

Now

$$\log M_X(t) = \lambda(e^t - 1).$$

This will be useful later for cumulants.

8.5 Exponential distribution

Let

$$X \sim \text{Exponential}(\lambda),$$

where $\lambda > 0$ is the rate. Its PDF is

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

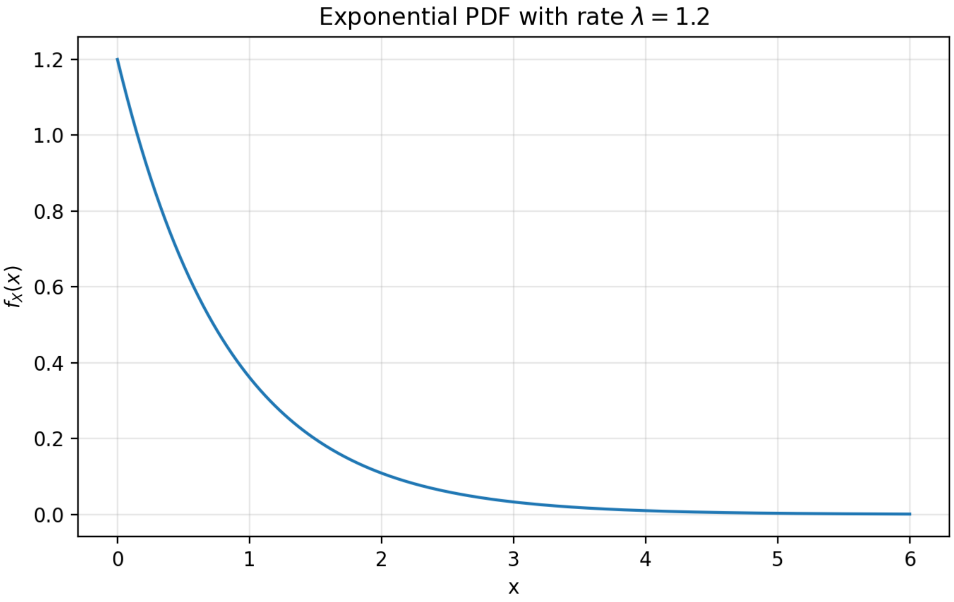

Figure 6: An exponential distribution is continuous, so it has a PDF and an MGF, but not a PGF in the usual counting-variable sense.

The MGF is

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} \, dx.$$

Combining the exponentials,

$$M_X(t) = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx.$$

This integral is finite only when $t < \lambda$. Then

$$M_X(t) = \frac{\lambda}{\lambda - t}, \qquad t < \lambda.$$

8.6 Gamma distribution

If

$$X \sim \text{Gamma}(\alpha, \lambda),$$

where α is shape and λ is rate, then

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \qquad t < \lambda.$$

This generalizes the exponential distribution, which is the case $\alpha = 1$.

8.7 Normal distribution

If

$$X \sim N(\mu, \sigma^2),$$

then

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

The normal distribution is continuous, so we do not usually use a PGF for it. But the MGF is very useful.

For example,

$$K_X(t) = \log M_X(t) = \mu t + \frac{\sigma^2 t^2}{2}.$$

Then

$$K_X'(0) = \mu, \qquad K_X''(0) = \sigma^2.$$

This shows how the cumulant generating function directly gives mean and variance.

9 Table of common PGFs and MGFs

Distribution	PGF $G_X(s)$	MGF $M_X(t)$
Bernoulli(p)	$1 - p + ps$	$1 - p + pe^t$
Binomial(n, p)	$(1 - p + ps)^n$	$(1 - p + pe^t)^n$
Geometric(p), failures before first success	$\frac{p}{1 - (1 - p)s}$	$\frac{p}{1 - (1 - p)e^t}$
Poisson(λ)	$e^{\lambda(s-1)}$	$e^{\lambda(e^t-1)}$

Distribution	PGF $G_X(s)$	MGF $M_X(t)$
Exponential(λ)	Not usually used	$\dfrac{\lambda}{\lambda - t}$, $t < \lambda$
Gamma(α, λ)	Not usually used	$\left(\dfrac{\lambda}{\lambda - t}\right)^{\alpha}$, $t < \lambda$
Normal(μ, σ^2)	Not usually used	$\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$

10 Cumulant generating functions

Definition 10.1 (Cumulant generating function). *If $M_X(t)$ exists, the cumulant generating function, or CGF, is*

$$K_X(t) = \log M_X(t).$$

The derivatives of K_X at 0 give cumulants:

$$\kappa_n = K_X^{(n)}(0).$$

The first two cumulants are especially important:

$$K_X'(0) = \mathbb{E}[X], \qquad K_X''(0) = \mathrm{Var}(X).$$

Why is this useful? Because for independent sums,

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Taking logs,

$$K_{X+Y}(t) = K_X(t) + K_Y(t).$$

So cumulants add for independent sums. This is why CGFs are common in advanced probability, asymptotic statistics, large deviations, and exponential families.

11 Characteristic functions: an advanced safety tool

Definition 11.1 (Characteristic function). *The characteristic function of X is*

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \qquad t \in \mathbb{R},$$

where $i^2 = -1$.

Unlike MGFs, characteristic functions always exist because

$$|e^{itX}| = 1.$$

So

$$|\varphi_X(t)| \le \mathbb{E}[|e^{itX}|] = 1.$$

Characteristic functions are central in advanced probability because:

- they always exist;

- they uniquely determine the distribution;
- they turn independent sums into products;
- they are used in proofs of the central limit theorem.

A simple relationship is:

$$\varphi_X(t) = M_X(it),$$

whenever that expression is meaningful.

12 Compound sums and actuarial motivation

One strong reason to learn PGFs and MGFs is random sums.

Suppose

$$S = Y_1 + Y_2 + \cdots + Y_N,$$

where N is a random count and $Y_1, Y_2, \dots$ are iid severities independent of N.

This appears in insurance:

- N = number of claims;
- Y_i = amount of claim i;
- S = total claim amount.

Condition on N:

$$M_S(t) = \mathbb{E}[e^{tS}] = \mathbb{E}\left[\mathbb{E}[e^{t(Y_1+\cdots+Y_N)} \mid N]\right].$$

Given $N = n$,

$$M_{Y_1+\cdots+Y_n}(t) = (M_Y(t))^n.$$

Thus,

$$M_S(t) = \mathbb{E}[(M_Y(t))^N].$$

This is the PGF of N evaluated at $M_Y(t)$:

$$\boxed{M_S(t) = G_N(M_Y(t)).}$$

If

$$N \sim \text{Poisson}(\lambda),$$

then

$$G_N(s) = e^{\lambda(s-1)}.$$

Therefore,

$$\boxed{M_S(t) = \exp\{\lambda(M_Y(t) - 1)\}.}$$

This is the MGF of a compound Poisson sum.

This formula is extremely useful because deriving the full density or PMF of S directly can be very difficult.

13 A deeper view: transforms of probability measures

From a more advanced viewpoint, PGFs and MGFs are examples of transforms of probability measures.

Let μ_X be the distribution of X. Then

$$M_X(t) = \int_{\mathbb{R}} e^{tx}\, d\mu_X(x).$$

The characteristic function is

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx}\, d\mu_X(x).$$

The Laplace transform for a nonnegative random variable is

$$L_X(t) = \mathbb{E}[e^{-tX}] = \int_0^\infty e^{-tx}\, d\mu_X(x).$$

For nonnegative integer-valued X, the PGF is

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k.$$

So the philosophy is:

$$\boxed{\text{turn a probability distribution into a function that is easier to manipulate.}}$$

14 Common confusions

14.1 Is a PDF a probability?

No. A PDF is density. Probability is area:

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\, dx.$$

A PDF value can be bigger than 1 as long as the total area is 1.

14.2 Can an expectation be larger than 1?

Yes. Expectation is an average value, not a probability. If X is a count, $\mathbb{E}[X]$ can be 10, 100, or more.

14.3 Can an MGF be larger than 1?

Yes. $M_X(0) = 1$, but for other t, the MGF can be larger than 1. It is not a probability.

14.4 Do PGFs work for continuous random variables?

Not in the usual way. PGFs are designed for nonnegative integer-valued random variables. Continuous random variables are usually described using PDFs, CDFs, MGFs, Laplace transforms, or characteristic functions.

14.5 Does the MGF always exist?

No. Some distributions have no finite MGF near 0. If the MGF exists in an open interval around 0, it is very powerful because it uniquely determines the distribution.

15 Practice problems with detailed solutions

Problem 1: Bernoulli PGF and MGF

Let $X \sim \text{Bernoulli}(p)$. Find $G_X(s)$ and $M_X(t)$, then use the MGF to find $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution

A Bernoulli random variable has only two possible values:

$$\mathbb{P}(X = 0) = 1 - p, \qquad \mathbb{P}(X = 1) = p.$$

The PGF is

$$G_X(s) = \mathbb{E}[s^X] = (1 - p)s^0 + ps^1 = 1 - p + ps.$$

The MGF is

$$M_X(t) = \mathbb{E}[e^{tX}] = (1 - p)e^0 + pe^t = 1 - p + pe^t.$$

Differentiate once:

$$M_X'(t) = pe^t.$$

Therefore,

$$\mathbb{E}[X] = M_X'(0) = p.$$

Differentiate again:

$$M_X''(t) = pe^t,$$

so

$$M_X''(0) = p.$$

Since $M_X''(0) = \mathbb{E}[X^2]$,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p).$$

Problem 2: Recovering probabilities from a PGF

Suppose a nonnegative integer-valued random variable has PGF

$$G_X(s) = 0.2 + 0.5s + 0.3s^2.$$

Find the distribution of X, then compute $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution

For a PGF,

$$G_X(s) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)s^k.$$

Therefore, the coefficient of s^k is $\mathbb{P}(X = k)$. Hence

$$\mathbb{P}(X = 0) = 0.2, \qquad \mathbb{P}(X = 1) = 0.5, \qquad \mathbb{P}(X = 2) = 0.3.$$

Now compute the expectation directly:

$$\mathbb{E}[X] = 0(0.2) + 1(0.5) + 2(0.3) = 1.1.$$

Next,

$$\mathbb{E}[X^2] = 0^2(0.2) + 1^2(0.5) + 2^2(0.3) = 0.5 + 1.2 = 1.7.$$

Thus,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1.7 - (1.1)^2 = 1.7 - 1.21 = 0.49.$$

This example shows that a PGF stores the full discrete distribution as coefficients.

Problem 3: Using PGF derivatives

For the same PGF

$$G_X(s) = 0.2 + 0.5s + 0.3s^2,$$

recover $\mathbb{P}(X = 2)$ using derivatives.

Solution

Taylor's formula gives

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

First differentiate:

$$G_X'(s) = 0.5 + 0.6s.$$

Differentiate again:

$$G_X''(s) = 0.6.$$

Therefore,

$$\mathbb{P}(X = 2) = \frac{G_X''(0)}{2!} = \frac{0.6}{2} = 0.3.$$

This matches the coefficient of s^2.

Problem 4: Poisson mean and variance using the PGF

Let $X \sim \mathrm{Poisson}(\lambda)$. Use the PGF to find $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Solution

The PGF of a Poisson(λ) random variable is

$$G_X(s) = e^{\lambda(s-1)}.$$

Differentiate once:

$$G'_X(s) = \lambda e^{\lambda(s-1)}.$$

Then

$$\mathbb{E}[X] = G'_X(1) = \lambda e^0 = \lambda.$$

Differentiate again:

$$G''_X(s) = \lambda^2 e^{\lambda(s-1)}.$$

Therefore,

$$G''_X(1) = \lambda^2.$$

For PGFs,

$$\operatorname{Var}(X) = G''_X(1) + G'_X(1) - \big(G'_X(1)\big)^2.$$

Substitute the values:

$$\operatorname{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

So for a Poisson random variable,

$$\mathbb{E}[X] = \lambda, \qquad \operatorname{Var}(X) = \lambda.$$

Problem 5: Binomial PGF from Bernoulli sums

Let $X \sim \text{Binomial}(n, p)$. Derive its PGF using Bernoulli random variables, then compute $\mathbb{E}[X]$ and $\operatorname{Var}(X)$.

Solution

A Binomial(n, p) random variable can be written as

$$X = X_1 + X_2 + \cdots + X_n,$$

where

$$X_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p).$$

The PGF of each Bernoulli variable is

$$G_{X_i}(s) = 1 - p + ps.$$

Because PGFs multiply for independent sums,

$$G_X(s) = \prod_{i=1}^{n} G_{X_i}(s) = (1 - p + ps)^n.$$

To compute the mean, differentiate:

$$G'_X(s) = n(1 - p + ps)^{n-1} p.$$

Thus,

$$\mathbb{E}[X] = G'_X(1) = np(1-p+p)^{n-1} = np.$$

For the variance, first compute

$$G''_X(s) = n(n-1)p^2(1-p+ps)^{n-2}.$$

So

$$G''_X(1) = n(n-1)p^2.$$

Therefore,

$$\operatorname{Var}(X) = G''_X(1) + G'_X(1) - \big(G'_X(1)\big)^2.$$

Substitute:

$$\operatorname{Var}(X) = n(n-1)p^2 + np - n^2p^2.$$

Simplify:

$$\operatorname{Var}(X) = np - np^2 = np(1-p).$$

Problem 6: Identifying a distribution from a PGF

Suppose

$$G_X(s) = (0.7 + 0.3s)^4.$$

Identify the distribution of X and compute $\mathbb{P}(X = 2)$.

Solution

The PGF of a Binomial(n, p) random variable is

$$G_X(s) = (1-p+ps)^n.$$

Comparing

$$(0.7 + 0.3s)^4$$

with

$$(1-p+ps)^n,$$

we get

$$1 - p = 0.7, \qquad p = 0.3, \qquad n = 4.$$

Therefore,

$$X \sim \text{Binomial}(4, 0.3).$$

Then

$$\mathbb{P}(X = 2) = \binom{4}{2}(0.3)^2(0.7)^2.$$

Compute:

$$\mathbb{P}(X = 2) = 6(0.09)(0.49) = 0.2646.$$

Problem 7: Sum of independent Poisson variables

Let

$$X \sim \text{Poisson}(\lambda_1), \qquad Y \sim \text{Poisson}(\lambda_2),$$

with X and Y independent. Use PGFs to find the distribution of $X + Y$.

Solution

The PGF of X is

$$G_X(s) = e^{\lambda_1(s-1)}.$$

The PGF of Y is

$$G_Y(s) = e^{\lambda_2(s-1)}.$$

Since X and Y are independent,

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

Therefore,

$$G_{X+Y}(s) = e^{\lambda_1(s-1)}e^{\lambda_2(s-1)} = e^{(\lambda_1+\lambda_2)(s-1)}.$$

This is the PGF of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Hence

$$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

This is a major benefit of generating functions: they turn sums into multiplication.

Problem 8: Convolution and PGF multiplication

Let X and Y be independent nonnegative integer-valued random variables. Explain why

$$\mathbb{P}(X + Y = k) = \sum_{j=0}^{k} \mathbb{P}(X = j)\mathbb{P}(Y = k - j).$$

Then explain how this appears in PGF multiplication.

Solution

The event $X + Y = k$ can happen in several mutually exclusive ways:

$$(X, Y) = (0, k), (1, k-1), (2, k-2), \ldots, (k, 0).$$

So we add the probabilities of all possible ways:

$$\mathbb{P}(X + Y = k) = \sum_{j=0}^{k} \mathbb{P}(X = j, Y = k - j).$$

Because X and Y are independent,

$$\mathbb{P}(X = j, Y = k - j) = \mathbb{P}(X = j)\mathbb{P}(Y = k - j).$$

Therefore,

$$\mathbb{P}(X + Y = k) = \sum_{j=0}^{k} \mathbb{P}(X = j)\mathbb{P}(Y = k - j).$$

Now write the PGFs as

$$G_X(s) = \sum_{a=0}^{\infty} \mathbb{P}(X = a)s^a, \qquad G_Y(s) = \sum_{b=0}^{\infty} \mathbb{P}(Y = b)s^b.$$

Multiplying gives

$$G_X(s)G_Y(s) = \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} \mathbb{P}(X = a)\mathbb{P}(Y = b)s^{a+b}.$$

The coefficient of s^k collects all terms where $a + b = k$. Thus the coefficient of s^k is

$$\sum_{j=0}^{k} \mathbb{P}(X = j)\mathbb{P}(Y = k - j),$$

which is exactly $\mathbb{P}(X + Y = k)$.

Problem 9: Geometric distribution using the PGF

Let X be the number of failures before the first success, with success probability p. Its PGF is

$$G_X(s) = \frac{p}{1 - (1 - p)s}.$$

Use the PGF to find $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Solution

Let $q = 1 - p$. Then

$$G_X(s) = \frac{p}{1 - qs} = p(1 - qs)^{-1}.$$

Differentiate:

$$G'_X(s) = pq(1 - qs)^{-2}.$$

Thus

$$\mathbb{E}[X] = G'_X(1) = \frac{pq}{(1 - q)^2}.$$

Since $1 - q = p$,

$$\mathbb{E}[X] = \frac{pq}{p^2} = \frac{q}{p} = \frac{1 - p}{p}.$$

Differentiate again:

$$G''_X(s) = 2pq^2(1 - qs)^{-3}.$$

Thus

$$G''_X(1) = \frac{2pq^2}{(1 - q)^3} = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2}.$$

Now use

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - \big(G'_X(1)\big)^2.$$

Substitute:

$$\mathrm{Var}(X) = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}.$$

Combine terms:

$$\mathrm{Var}(X) = \frac{q^2}{p^2} + \frac{q}{p} = \frac{q^2}{p^2} + \frac{qp}{p^2} = \frac{q(q + p)}{p^2}.$$

Since $q + p = 1$,

$$\mathrm{Var}(X) = \frac{q}{p^2} = \frac{1 - p}{p^2}.$$

Problem 10: Exponential distribution using the MGF

Let $X \sim \text{Exponential}(\lambda)$ with rate $\lambda > 0$. Its MGF is

$$M_X(t) = \frac{\lambda}{\lambda - t}, \qquad t < \lambda.$$

Use the MGF to find $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution

Rewrite the MGF as

$$M_X(t) = \lambda(\lambda - t)^{-1}.$$

Differentiate:

$$M_X'(t) = \lambda(\lambda - t)^{-2}.$$

Therefore,

$$\mathbb{E}[X] = M_X'(0) = \lambda\lambda^{-2} = \frac{1}{\lambda}.$$

Differentiate again:

$$M_X''(t) = 2\lambda(\lambda - t)^{-3}.$$

Thus,

$$M_X''(0) = 2\lambda\lambda^{-3} = \frac{2}{\lambda^2}.$$

Since $M_X''(0) = \mathbb{E}[X^2]$,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Problem 11: Sum of independent normal variables

Let

$$X \sim N(\mu_1, \sigma_1^2), \qquad Y \sim N(\mu_2, \sigma_2^2),$$

and suppose X and Y are independent. Use MGFs to find the distribution of $X + Y$.

Solution

The MGF of X is

$$M_X(t) = \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right).$$

The MGF of Y is

$$M_Y(t) = \exp\left(\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right).$$

For independent variables,

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Thus,

$$M_{X+Y}(t) = \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right) \exp\left(\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right).$$

Combine exponents:

$$M_{X+Y}(t) = \exp\left((\mu_1 + \mu_2)t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right).$$

This is the MGF of a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Therefore,

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Problem 12: Compound Poisson sum

Suppose

$$S = Y_1 + \cdots + Y_N,$$

where $N \sim \text{Poisson}(\lambda)$, the Y_i are iid, and N is independent of the Y_i. If Y_i has MGF $M_Y(t)$, find $M_S(t)$. Then find $\mathbb{E}[S]$ and $\text{Var}(S)$ in terms of λ, $\mathbb{E}[Y]$, and $\text{Var}(Y)$.

Solution

Condition on N. If $N = n$, then

$$S = Y_1 + \cdots + Y_n.$$

Because the Y_i are iid,

$$M_{S|N=n}(t) = (M_Y(t))^n.$$

Therefore,

$$M_S(t) = \mathbb{E}[(M_Y(t))^N].$$

This is the PGF of N evaluated at $M_Y(t)$:

$$M_S(t) = G_N(M_Y(t)).$$

For $N \sim \text{Poisson}(\lambda)$,

$$G_N(s) = e^{\lambda(s-1)}.$$

Hence

$$M_S(t) = \exp\{\lambda(M_Y(t) - 1)\}.$$

Now let

$$\mu = \mathbb{E}[Y], \qquad \sigma^2 = \text{Var}(Y).$$

The cumulant generating function of S is

$$K_S(t) = \log M_S(t) = \lambda(M_Y(t) - 1).$$

Differentiate:

$$K_S'(t) = \lambda M_Y'(t).$$

Thus

$$\mathbb{E}[S] = K_S'(0) = \lambda M_Y'(0) = \lambda\mu.$$

Differentiate again:

$$K_S''(t) = \lambda M_Y''(t).$$

For cumulant generating functions,

$$\text{Var}(S) = K_S''(0) = \lambda M_Y''(0).$$

But

$$M_Y''(0) = \mathbb{E}[Y^2] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2 = \sigma^2 + \mu^2.$$

Therefore,

$$\mathrm{Var}(S) = \lambda(\sigma^2 + \mu^2).$$

This is a standard actuarial formula for compound Poisson sums.

16 Matplotlib code for the figures

The following Python code generates the figures used in this note. It uses Matplotlib and can be adapted for other visual explanations.

```
import numpy as np
import math
import matplotlib.pyplot as plt
from pathlib import Path

out = Path("figures")
out.mkdir(exist_ok=True)

# Poisson PMF: coefficients in a PGF
lam = 3.0
k = np.arange(0, 13)
pmf = np.exp(-lam) * lam**k / np.array([math.factorial(int(i)) for i in k])
plt.figure(figsize=(7, 4.5), dpi=200)
plt.stem(k, pmf, basefmt="␣")
plt.xlabel("k")
plt.ylabel(r"$P(X=k)$")
plt.title(r"Poisson$(3)$␣PMF:␣coefficients␣inside␣the␣PGF")
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.savefig("figures/poisson_pmf_coefficients.png", bbox_inches="tight")
plt.close()

# Standard normal PDF
x = np.linspace(-4, 4, 500)
pdf = 1 / np.sqrt(2*np.pi) * np.exp(-x**2/2)
plt.figure(figsize=(7, 4.5), dpi=200)
plt.plot(x, pdf)
plt.xlabel("x")
plt.ylabel(r"$f_X(x)$")
plt.title("Standard␣normal␣PDF:␣density,␣not␣probability␣at␣a␣point")
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.savefig("figures/normal_pdf.png", bbox_inches="tight")
plt.close()

# Standard normal CDF
cdf = 0.5 * (1 + np.vectorize(math.erf)(x/np.sqrt(2)))
plt.figure(figsize=(7, 4.5), dpi=200)
plt.plot(x, cdf)
plt.xlabel("x")
plt.ylabel(r"$F_X(x)=P(X\leq␣x)$")
```

```
plt.title("Standard␣normal␣CDF:␣accumulated␣probability")
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.savefig("figures/normal_cdf.png", bbox_inches="tight")
plt.close()

# Poisson MGF
t = np.linspace(-1, 1, 500)
mgf = np.exp(lam * (np.exp(t) - 1))
plt.figure(figsize=(7, 4.5), dpi=200)
plt.plot(t, mgf)
plt.axvline(0, linestyle="--", linewidth=1)
plt.axhline(1, linestyle="--", linewidth=1)
plt.xlabel("t")
plt.ylabel(r"$M_X(t)$")
plt.title(r"MGF␣of␣Poisson$(3)$:␣derivatives␣at␣0␣give␣moments")
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.savefig("figures/poisson_mgf.png", bbox_inches="tight")
plt.close()
```

17 Final summary

Object	What it does
CDF	Accumulates probability: $F_X(x) = \mathbb{P}(X \leq x)$
PMF	Gives point probabilities for discrete variables
PDF	Gives density for continuous variables; probability is area
PGF	Stores probabilities as coefficients for nonnegative integer variables
MGF	Generates raw moments through derivatives at 0
CGF	Generates cumulants; especially useful for sums and asymptotics
Characteristic function	Always exists and uniquely determines the distribution

The main summary is:

$$\boxed{\text{PGF encodes probabilities as coefficients.}} \qquad \boxed{\text{MGF encodes moments through derivatives.}}$$