Lorenzo Naranjo October 28, 2026

Generating Functions

Most of the random variables in this course are built by adding up simpler ones. The position of a random walk is a sum of coin flips, the number of defaults in a loan portfolio is a sum of default indicators, and the log return of a stock over a year is a sum of daily log returns. Computing the distribution of a sum directly is tedious, since it requires adding up the probabilities of all the ways in which the terms can combine. A *generating function* turns this problem into a much simpler one: it encodes the whole distribution of a random variable into a single function, and it converts sums of independent random variables into products of functions.

In these notes we study three closely related transforms. The *probability generating function* (PGF) is the natural tool for random variables that count things, like the number of heads, the number of defaults, or the number of steps until a random walk stops. The *moment generating function* (MGF) extends the idea to random variables with continuous values, such as normal returns. The *characteristic function* (CF) is the version that always exists, and it is the one used to prove limit theorems. All three are expectations of a function of the form $e^{\theta X}$, evaluated at different kinds of arguments.

Generating functions will reappear throughout the course. They provide the characteristic equation that solves the ruin problem, the exponential martingales used to study stopping times, the limit that turns rare Bernoulli events into a Poisson process, and the limit that turns a scaled random walk into a Brownian motion. For more complete treatments, see Blitzstein and Hwang (2019) and Grimmett and Stirzaker (2001). Feller (1968) remains the classic reference on generating functions for integer-valued random variables.

Probability Generating Functions

Definition

Let X be a random variable taking values in the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$, with probability mass function $p_X(k) = \mathrm{P}(X = k)$. The **probability generating function** of X is

$$G_X(z) = \mathrm{E}(z^X) = \sum_{k=0}^{\infty} p_X(k) z^k. \tag{1}$$

The PGF is a power series whose coefficients are the probabilities of X. Since the probabilities add up to one, the series converges at least for $|z| \leq 1$, and we always have $G_X(1) = 1$.

The variable z has no meaning of its own. It is a placeholder that keeps the probabilities in order: the coefficient of z^k is the probability that $X = k$. Because a power series determines its coefficients, we can recover the probability mass function from the PGF by differentiating at $z = 0$,

$$p_X(k) = \frac{G_X^{(k)}(0)}{k!},$$

where $G_X^{(k)}$ denotes the k-th derivative of G_X. Therefore, two random variables with the same PGF have the same distribution. This *uniqueness* property is what makes generating functions useful: if we can compute the PGF of a random variable and recognize it, we know its distribution.

Example 1. A Bernoulli random variable X takes the value 1 with probability p and the value 0 with probability $1 - p$. It is the indicator of a single event, such as a coin landing heads or a borrower defaulting within a year. Its PGF is

$$G_X(z) = (1 - p)z^0 + pz^1 = 1 - p + pz.$$

□

Example 2. Suppose we flip a coin that lands heads with probability p until we see the first head, and let X be the number of flips required. The first head occurs on flip k if the first $k-1$ flips are tails and flip k is a head, so that

$$p_X(k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \ldots$$

We say that X has a *geometric distribution*. Writing $q = 1 - p$, its PGF is a geometric series,

$$G_X(z) = \sum_{k=1}^{\infty} q^{k-1}pz^k = pz\sum_{j=0}^{\infty}(qz)^j = \frac{pz}{1-qz},$$

which converges for $|z| < 1/q$. □

The geometric distribution is the simplest model of a waiting time in discrete time. Its continuous-time counterpart is the exponential distribution from the notes on continuous distributions, which describes the time between arrivals of a Poisson process.

Example 3. A random variable N has a *Poisson distribution* with parameter $\lambda > 0$ if

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

We write $N \sim \text{Poisson}(\lambda)$. Using the power series $e^x = \sum_{k=0}^{\infty} x^k/k!$, the PGF of N is

$$G_N(z) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}z^k = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda z)^k}{k!} = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}.$$

□

The Poisson distribution counts the number of rare events that occur over a period of time, such as defaults in a large portfolio of loans, claims arriving at an insurance company, or large price jumps in a stock. We will see below why it arises so often.

Moments

The PGF also encodes the moments of X. Differentiating (1) term by term gives

$$G_X'(z) = \sum_{k=1}^{\infty} k\, p_X(k) z^{k-1}, \qquad G_X''(z) = \sum_{k=2}^{\infty} k(k-1)\, p_X(k) z^{k-2}.$$

Setting $z = 1$ in each expression we obtain the *factorial moments* of X.

Moments from the PGF

If X takes values in $\mathbb{N}$ and has PGF G_X, then

$$\begin{aligned} \mathrm{E}(X) &= G_X'(1), \\ \mathrm{E}(X(X-1)) &= G_X''(1), \\ \mathrm{V}(X) &= G_X''(1) + G_X'(1) - G_X'(1)^2, \end{aligned}$$

provided these quantities are finite.

Proof

The first two expressions follow from setting $z = 1$ in the derivatives above, since

$$G_X'(1) = \sum_k k\, p_X(k) = \mathrm{E}(X), \qquad G_X''(1) = \sum_k k(k-1)\, p_X(k) = \mathrm{E}(X(X-1)).$$

For the variance, note that $\mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$, so that

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

□

Example 4. If $N \sim \text{Poisson}(\lambda)$, then $G_N(z) = e^{\lambda(z-1)}$, so that $G_N'(z) = \lambda e^{\lambda(z-1)}$ and $G_N''(z) = \lambda^2 e^{\lambda(z-1)}$. Therefore,

$$\mathrm{E}(N) = \lambda, \qquad \mathrm{V}(N) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

The mean and the variance of a Poisson random variable are both equal to λ. □

Example 5. If X has the geometric distribution of Example 2, then

$$G_X'(z) = \frac{p}{(1-qz)^2}, \qquad G_X''(z) = \frac{2pq}{(1-qz)^3}.$$

Evaluating at $z = 1$ and using $1 - q = p$ we find $G'_X(1) = 1/p$ and $G''_X(1) = 2q/p^2$. Hence,

$$\mathrm{E}(X) = \frac{1}{p}, \qquad \mathrm{V}(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

For example, a die shows a six with probability $p = 1/6$, so we expect to wait 6 rolls for the first six. □

Sums of Independent Random Variables

The most important property of generating functions is how they handle sums. If X and Y are independent, then z^X and z^Y are also independent, and, as shown in the notes on conditioning and independence, the expectation of a product of independent random variables is the product of their expectations.

PGF of a Sum

If X and Y are independent random variables taking values in $\mathbb{N}$, then

$$G_{X+Y}(z) = G_X(z)\, G_Y(z).$$

More generally, if $X_1, X_2, \ldots, X_n$ are independent, the PGF of $S_n = X_1 + X_2 + \cdots + X_n$ is

$$G_{S_n}(z) = \prod_{i=1}^{n} G_{X_i}(z).$$

Proof

Since $z^{X+Y} = z^X z^Y$ and the random variables z^X and z^Y are independent,

$$G_{X+Y}(z) = \mathrm{E}(z^X z^Y) = \mathrm{E}(z^X)\mathrm{E}(z^Y) = G_X(z)\, G_Y(z).$$

The general case follows by applying this result repeatedly. □

Compare this with the direct approach. The probability that $X + Y = k$ is

$$\mathrm{P}(X+Y=k)=\sum_{j=0}^{k} p_X(j)\,p_Y(k-j),$$

a sum over all the ways of splitting k between X and Y. The PGF performs all these sums at once: multiplying two power series collects exactly these terms in the coefficient of z^k.

Example 6. Suppose a bank holds n loans, each defaulting independently with probability p. Let X_i be the Bernoulli indicator that loan i defaults, and let $N = X_1 + \cdots + X_n$ be the number of defaults. By Example 1 each X_i has PGF $1 - p + pz$, so that

$$G_N(z) = (1-p+pz)^n.$$

Expanding with the binomial theorem,

$$G_N(z)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}z^k,$$

and reading off the coefficient of z^k gives the *binomial distribution*, $\mathrm{P}(N=k)=\binom{n}{k}p^k(1-p)^{n-k}$. The moments follow just as easily: $G_N'(1)=np$ and $G_N''(1)=n(n-1)p^2$, so that $\mathrm{E}(N)=np$ and $\mathrm{V}(N)=np(1-p)$. □

Example 7. If $N_1 \sim \text{Poisson}(\lambda_1)$ and $N_2 \sim \text{Poisson}(\lambda_2)$ are independent, then

$$G_{N_1+N_2}(z)=e^{\lambda_1(z-1)}e^{\lambda_2(z-1)}=e^{(\lambda_1+\lambda_2)(z-1)},$$

which is the PGF of a Poisson random variable with parameter $\lambda_1+\lambda_2$. By uniqueness, $N_1+N_2 \sim \text{Poisson}(\lambda_1+\lambda_2)$. If defaults in two independent sectors arrive as Poisson counts, total defaults are also a Poisson count. □

Random Sums

In many applications the number of terms in a sum is itself random. An insurance company does not know in advance how many claims it will receive, and a market maker does not know how many orders will arrive during the next minute. Let N be a random variable taking values in $\mathbb{N}$, and let $X_1, X_2, \ldots$ be independent and identically distributed random variables that are also independent of N. The *random sum*

$$S = X_1 + X_2 + \cdots + X_N$$

adds up a random number of terms, with the convention that $S = 0$ when $N = 0$. Its PGF is a composition of functions.

PGF of a Random Sum

If N is independent of the iid sequence $X_1, X_2, \ldots$, all taking values in $\mathbb{N}$, then the random sum $S = X_1 + \cdots + X_N$ has PGF

$$G_S(z) = G_N(G_X(z)),$$

where G_X is the common PGF of the X_i. In particular, $\mathrm{E}(S) = \mathrm{E}(N)\mathrm{E}(X)$.

Proof

We condition on N and use the law of iterated expectations. Given $N = n$, the sum has n independent terms, so its conditional PGF is $G_X(z)^n$. Therefore,

$$G_S(z) = \mathrm{E}(\mathrm{E}(z^S \mid N)) = \mathrm{E}(G_X(z)^N) = G_N(G_X(z)),$$

where the last step uses the definition of G_N evaluated at the point $G_X(z)$. Differentiating with the chain rule and using $G_X(1) = 1$,

$$\mathrm{E}(S) = G_S'(1) = G_N'(G_X(1))\, G_X'(1) = G_N'(1)\, G_X'(1) = \mathrm{E}(N)\mathrm{E}(X).$$

□

Example 8. Suppose the number of orders that arrive at a market maker during a minute is $N \sim \text{Poisson}(\lambda)$, and each order is independently a buy with probability p and a sell with probability $1 - p$. The number of buy orders is the random sum $S = X_1 + \cdots + X_N$, where the X_i are Bernoulli indicators of a buy. Therefore,

$$G_S(z) = G_N(1 - p + pz) = e^{\lambda(1-p+pz-1)} = e^{\lambda p(z-1)},$$

and the number of buy orders is $\text{Poisson}(\lambda p)$. With $\lambda = 30$ orders per minute and $p = 0.6$, buy orders arrive as a Poisson count with mean 18 per minute. □

Selecting each event of a Poisson count independently with probability p is called *thinning*, and the example shows that thinning a Poisson count produces another Poisson count. We will use this property when we study Poisson processes, for example to separate price jumps into upward and downward moves.

The Poisson Limit

Why is the Poisson distribution so common? Consider a portfolio of n loans, each defaulting independently with a small probability p. The number of defaults is binomial, but when n is large and p is small only the product $\lambda = np$, the expected number of defaults, seems to matter. Generating functions make this precise. If we write $p = \lambda/n$, the PGF of the binomial in Example 6 becomes

$$G_N(z) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}z\right)^n = \left(1 + \frac{\lambda(z-1)}{n}\right)^n.$$

Using the limit $(1 + x/n)^n \to e^x$, we find that for every z

$$\lim_{n\to\infty} G_N(z) = e^{\lambda(z-1)},$$

which is the PGF of a Poisson random variable with parameter λ.

To conclude that the probabilities themselves converge we need one more fact, which we state without proof.

Continuity Theorem for PGFs

Let $N_1, N_2, \ldots$ and N be random variables taking values in $\mathbb{N}$. Then

$$\lim_{n\to\infty} \mathrm{P}(N_n = k) = \mathrm{P}(N = k) \quad \text{for all } k$$

if and only if $\lim_{n\to\infty} G_{N_n}(z) = G_N(z)$ for all $0 \le z \le 1$.

Therefore, the number of successes in many independent trials, each with a small probability of success, is approximately Poisson with parameter equal to the expected number of successes. This is sometimes called the *law of rare events*.

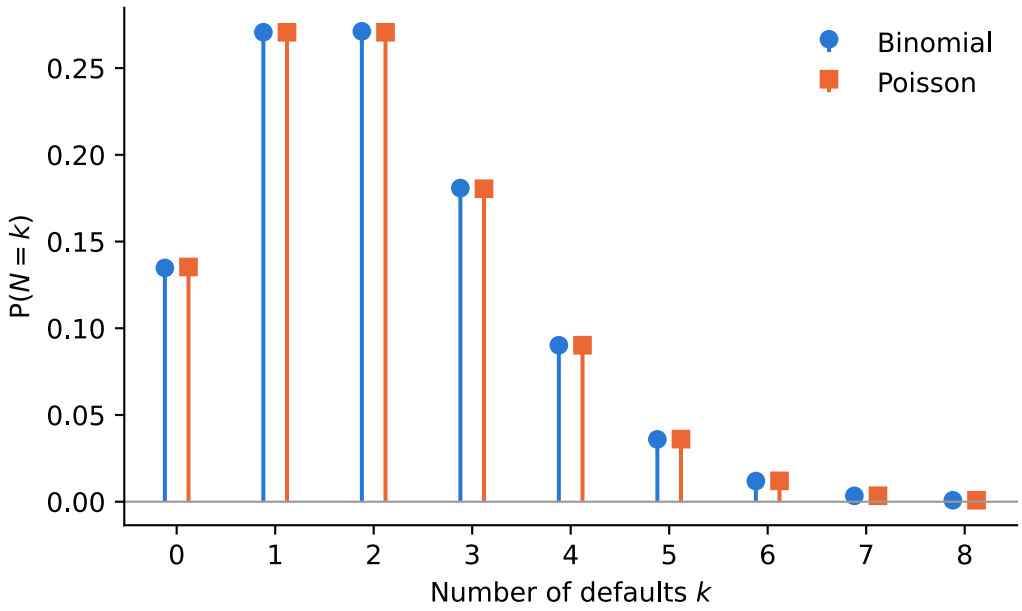


Figure 1: The figure compares the binomial distribution of the number of defaults among 500 loans, each defaulting with probability 0.004, with a Poisson distribution with the same mean $\lambda = 2$.

Example 9. A bank holds $n = 500$ independent loans, each with a default probability of $p = 0.004$ over the next year. The expected number of defaults is $\lambda = np = 2$. The exact probability of no defaults is

$$\mathrm{P}(N = 0) = (1 - 0.004)^{500} = 0.1348,$$

while the Poisson approximation gives $e^{-2} = 0.1353$. Similarly, the probability of exactly three defaults is 0.1808 under the binomial distribution and 0.1804 under the Poisson approximation. Figure 1 shows that the two distributions are nearly indistinguishable. □

The Poisson limit is the first of two limits of sums of Bernoulli random variables that we will study. Here the number of trials grows while each success becomes rare, and the limit is a count. In the notes on Brownian motion, the number of coin flips also grows but each flip is scaled down in size, and the limit is a normal random variable. The first limit leads to the Poisson process and the second to Brownian motion.

Moment Generating Functions

Definition

The PGF only makes sense for random variables that take values in $\mathbb{N}$. To handle random variables with negative or continuous values, such as stock returns, we substitute $z = e^{\theta}$ in the definition of the PGF. The **moment generating function** of a random variable X is

$$M_X(\theta) = \mathrm{E}(e^{\theta X}),$$

defined for all real θ such that the expectation is finite. If X takes values in $\mathbb{N}$, then $M_X(\theta) = G_X(e^{\theta})$. If X has a probability density function f, the expectation is an integral,

$$M_X(\theta) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\, dx.$$

We always have $M_X(0) = 1$.

We have in fact already computed an MGF. In the notes on continuous distributions we showed that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathrm{E}(e^{\alpha X}) = e^{\alpha\mu + \frac{1}{2}\alpha^2\sigma^2}$ for any real α, and used this expression to find the moments of a lognormal random variable. That calculation was the MGF of the normal distribution evaluated at $\theta = \alpha$. Below we see that it is one instance of a general method.

The name comes from the power series of the exponential function. Since $e^{\theta X} = 1 + \theta X + \frac{\theta^2}{2!}X^2 + \frac{\theta^3}{3!}X^3 + \cdots$, taking expectations term by term gives

$$M_X(\theta) = 1 + \theta\, \mathrm{E}(X) + \frac{\theta^2}{2!}\, \mathrm{E}(X^2) + \frac{\theta^3}{3!}\, \mathrm{E}(X^3) + \cdots,$$

so the moments of X are the coefficients of the series.

Properties of the MGF

Suppose $M_X(\theta)$ is finite for all θ in an open interval around zero. Then:

i. The moments of X are the derivatives of M_X at zero, $\mathrm{E}(X^k) = M_X^{(k)}(0)$.

ii. For constants a and b, $M_{a+bX}(\theta) = e^{a\theta} M_X(b\theta)$.

iii. If X and Y are independent, $M_{X+Y}(\theta) = M_X(\theta) M_Y(\theta)$.

iv. The MGF determines the distribution of X uniquely.

The second property follows from $e^{\theta(a+bX)} = e^{a\theta} e^{(b\theta)X}$, and the third from the same argument we used for the PGF. The first and fourth require more work, and we take them as given.

The Exponential and Gamma Distributions

Recall from the notes on continuous distributions that $T \sim \mathrm{Exp}(\lambda)$ has density $\lambda e^{-\lambda t}$ for $t \geq 0$, and that $X \sim \mathrm{Gamma}(\alpha, \lambda)$ has density $\lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)$ for $x > 0$, with $\mathrm{Gamma}(1, \lambda) = \mathrm{Exp}(\lambda)$.

Example 10. For $\theta < \lambda$, the MGF of $T \sim \mathrm{Exp}(\lambda)$ is

$$M_T(\theta) = \int_0^\infty e^{\theta t} \lambda e^{-\lambda t}\, dt = \lambda \int_0^\infty e^{-(\lambda-\theta)t}\, dt = \frac{\lambda}{\lambda - \theta}.$$

For $\theta \geq \lambda$ the integral diverges, so the MGF is only finite on the interval $(-\infty, \lambda)$. Differentiating,

$$M_T'(\theta) = \frac{\lambda}{(\lambda-\theta)^2}, \qquad M_T''(\theta) = \frac{2\lambda}{(\lambda-\theta)^3},$$

so that $\mathrm{E}(T) = 1/\lambda$, $\mathrm{E}(T^2) = 2/\lambda^2$, and $\mathrm{V}(T) = 1/\lambda^2$, without computing any integral by parts. □

Example 11. For $X \sim \mathrm{Gamma}(\alpha, \lambda)$ and $\theta < \lambda$, the integrand of the MGF is proportional to a gamma density with rate $\lambda - \theta$, so that

$$M_X(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-\theta)x}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-\theta)^\alpha} = \left(\frac{\lambda}{\lambda-\theta}\right)^\alpha.$$

□

The MGF of the gamma distribution is the MGF of the exponential distribution raised to the power α. Together with the product rule for sums, this gives three useful results.

First, let $T_1, T_2, \ldots, T_n$ be independent Exp(λ) random variables and let $W_n = T_1 + \cdots + T_n$. Then

$$M_{W_n}(\theta) = \left(\frac{\lambda}{\lambda-\theta}\right)^n, \qquad \theta < \lambda,$$

which is the MGF of a Gamma(n, λ) random variable. By uniqueness, the sum of n independent exponential waiting times has a gamma distribution with shape n. In a Poisson process, W_n is the time at which the n-th event occurs.

Second, the same argument shows that if $X_1 \sim$ Gamma(α_1, λ) and $X_2 \sim$ Gamma(α_2, λ) are independent with the same rate, then $X_1 + X_2 \sim$ Gamma$(\alpha_1 + \alpha_2, \lambda)$.

Third, the notes on continuous distributions show that the square of a standard normal random variable has a Gamma$(1/2, 1/2)$ distribution. If $Z_1, \ldots, Z_k$ are independent standard normal random variables, the second result implies that

$$Z_1^2 + \cdots + Z_k^2 \sim \text{Gamma}(k/2, 1/2),$$

which is the chi-squared distribution with k degrees of freedom. Its MGF is $(1-2\theta)^{-k/2}$ for $\theta < 1/2$, and its mean and variance are k and $2k$.

The Normal Distribution

Recall from the notes on continuous distributions that $X \sim \mathcal{N}(\mu, \sigma^2)$ can be written as $X = \mu + \sigma Z$, where Z has the standard normal density $f(z) = e^{-\frac{1}{2}z^2}/\sqrt{2\pi}$. We now state the MGF of the normal distribution in the language of this chapter. Working with Z first and then using the properties of the MGF shortens the calculation, and the same pattern will give us the characteristic function below.

MGF of the Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then for all real θ

$$M_X(\theta) = e^{\mu\theta + \frac{1}{2}\sigma^2\theta^2}. \tag{2}$$

Proof

We first compute the MGF of $Z \sim \mathcal{N}(0,1)$. Completing the square in the exponent, $\theta z - \frac{1}{2}z^2 = -\frac{1}{2}(z-\theta)^2 + \frac{1}{2}\theta^2$, so that

$$M_Z(\theta) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\theta z - \frac{1}{2}z^2}\, dz = e^{\frac{1}{2}\theta^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\theta)^2}\, dz}_{=1} = e^{\frac{1}{2}\theta^2}.$$

The integral equals one because the integrand is the density of a normal random variable with mean θ and variance one. Since $X = \mu + \sigma Z$, the second property of the MGF gives

$$M_X(\theta) = e^{\mu\theta} M_Z(\sigma\theta) = e^{\mu\theta + \frac{1}{2}\sigma^2\theta^2}.$$

□

Equation (2) has two immediate consequences. First, if $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then

$$M_{X_1+X_2}(\theta) = e^{(\mu_1+\mu_2)\theta + \frac{1}{2}(\sigma_1^2+\sigma_2^2)\theta^2},$$

so by uniqueness $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Sums of independent normal random variables are normal. This is why, in a model where daily log returns are independent and normal, annual log returns are also normal.

Second, we recover the lognormal moments from the notes on continuous distributions. If $Y = e^X$, every moment of Y is a value of the MGF of X,

$$\mathrm{E}(Y^\alpha) = \mathrm{E}(e^{\alpha X}) = M_X(\alpha) = e^{\alpha\mu + \frac{1}{2}\alpha^2\sigma^2},$$

so $\mathrm{E}(Y) = M_X(1)$ and $\mathrm{E}(Y^2) = M_X(2)$. There we found $\mathrm{E}(Y)$ by completing the square inside an integral. Here every moment follows by evaluating one function at different points.

Example 12. Suppose the log return of a stock over the next year is $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0.08$ and $\sigma = 0.20$, so that the stock price in one year is $S_1 = S_0 e^X$. The expected gross return is

$$\mathrm{E}\left(\frac{S_1}{S_0}\right) = M_X(1) = e^{0.08 + \frac{1}{2}(0.20)^2} = e^{0.10} = 1.1052.$$

The expected simple return of 10.52% exceeds $e^{0.08} - 1 = 8.33\%$, the simple return at the mean log return. The difference comes from the term $\frac{1}{2}\sigma^2$, the same correction that will appear in Itô's lemma. □

Cumulant Generating Functions

It is often convenient to work with the logarithm of the MGF,

$$K_X(\theta) = \ln M_X(\theta),$$

called the *cumulant generating function*. Since the MGF of a sum of independent random variables is a product, the cumulant generating function of the sum is a sum, $K_{X+Y}(\theta) = K_X(\theta) + K_Y(\theta)$. Its first two derivatives at zero are the mean and the variance:

$$K_X'(0) = \frac{M_X'(0)}{M_X(0)} = \mathrm{E}(X), \qquad K_X''(0) = \frac{M_X''(0)M_X(0) - M_X'(0)^2}{M_X(0)^2} = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \mathrm{V}(X).$$

For the two distributions that drive this course, the cumulant generating functions are particularly simple:

$$X \sim \mathcal{N}(\mu, \sigma^2): \quad K_X(\theta) = \mu\theta + \frac{1}{2}\sigma^2\theta^2, \qquad N \sim \text{Poisson}(\lambda): \quad K_N(\theta) = \lambda(e^\theta - 1).$$

The derivatives $K_X^{(k)}(0)$ are called the *cumulants* of X. The normal cumulant generating function is a quadratic, so all cumulants beyond the variance are zero. The Poisson one is linear in λ, which reflects the fact that Poisson counts over disjoint periods add up.

The cumulant generating function also produces martingales. If $X_1, X_2, \ldots$ are iid with cumulant generating function K and $S_n = X_1 + \cdots + X_n$, then each factor $e^{\theta X_i - K(\theta)}$ is positive with expectation one, so their product

$$M_n = e^{\theta S_n - nK(\theta)}$$

is a martingale of the type studied in the notes on discrete-time martingales. These *exponential martingales* will allow us to compute the distribution of stopping times and to change probability measures.

When the MGF Does Not Exist

The MGF requires $\mathrm{E}(e^{\theta X})$ to be finite, and for some random variables this fails for every $\theta > 0$. The lognormal distribution is the most important example in finance. If $Y = e^X$ with X normal, all moments $\mathrm{E}(Y^k) = e^{k\mu + \frac{1}{2}k^2\sigma^2}$ are finite, but they grow so fast that the series $\sum_k \theta^k \, \mathrm{E}(Y^k)/k!$ diverges for any $\theta > 0$. Equivalently, $e^{\theta e^x}$ grows faster than the normal density decays, so $\mathrm{E}(e^{\theta Y}) = \infty$. The MGF of a stock price in the Black-Scholes model does not exist, even though the MGF of its logarithm does.

The same problem arises with heavy-tailed distributions, such as the Student t distribution, whose moments are infinite beyond some order. To handle all random variables with a single tool we need a transform that is always finite.

Characteristic Functions

Definition

The **characteristic function** of a random variable X is

$$\phi_X(u) = \mathrm{E}(e^{iuX}), \qquad u \in \mathbb{R},$$

where i is the imaginary unit, $i^2 = -1$. By Euler's formula, $e^{iux} = \cos(ux) + i\sin(ux)$, so the characteristic function is

$$\phi_X(u) = \mathrm{E}(\cos(uX)) + i\,\mathrm{E}(\sin(uX)).$$

Both expectations are finite because the sine and cosine are bounded by one. Therefore, unlike the MGF, the characteristic function exists for every random variable and every real u.

The characteristic function is the MGF evaluated at an imaginary argument. When the MGF is finite around zero, replacing θ by iu in its formula gives $\phi_X(u)$, although justifying this step requires complex analysis. For random variables in $\mathbb{N}$ the relation is direct, $\phi_X(u) = G_X(e^{iu})$. The three transforms are therefore the same object seen from different angles:

$$\phi_X(u) = \mathrm{E}(e^{iuX}), \qquad M_X(\theta) = \mathrm{E}(e^{\theta X}), \qquad G_X(z) = \mathrm{E}(z^X).$$

Properties of the Characteristic Function

For any random variable X:

i. $\phi_X(0) = 1$ and $|\phi_X(u)| \leq 1$ for all u.

ii. For constants a and b, $\phi_{a+bX}(u) = e^{iau}\phi_X(bu)$.

iii. If X and Y are independent, $\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)$.

iv. If $\mathrm{E}(|X|^k) < \infty$, then $\phi_X^{(k)}(0) = i^k\,\mathrm{E}(X^k)$.

v. The characteristic function determines the distribution of X uniquely.

The bound $|\phi_X(u)| \leq \mathrm{E}(|e^{iuX}|) = 1$ holds because $|e^{iux}| = \sqrt{\cos^2(ux) + \sin^2(ux)} = 1$. The second and third properties follow exactly as for the MGF. In particular, $\mathrm{E}(X) = -i\phi_X'(0)$ and $\mathrm{E}(X^2) = -\phi_X''(0)$.

Example 13. If $N \sim \text{Poisson}(\lambda)$, then $\phi_N(u) = G_N(e^{iu}) = e^{\lambda(e^{iu}-1)}$. □

Example 14. If X takes the values 1 and -1 with probability $1/2$ each, then

$$\phi_X(u) = \frac{1}{2}e^{iu} + \frac{1}{2}e^{-iu} = \cos(u).$$

The characteristic function is real because the distribution of X is symmetric around zero, so the sine terms cancel. □

The Normal Characteristic Function

Replacing θ by iu in (2) suggests that the characteristic function of $X \sim \mathcal{N}(\mu, \sigma^2)$ is $e^{i\mu u - \frac{1}{2}\sigma^2 u^2}$. The following argument confirms it without complex analysis.

Characteristic Function of the Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\phi_X(u) = e^{i\mu u - \frac{1}{2}\sigma^2 u^2}.$$

Proof

Let $Z \sim \mathcal{N}(0,1)$ with density $f(z) = e^{-\frac{1}{2}z^2}/\sqrt{2\pi}$, and note that $f'(z) = -zf(z)$. Differentiating $\phi_Z(u) = \int e^{iuz} f(z)\, dz$ under the integral sign,

$$\phi_Z'(u) = \int_{-\infty}^{\infty} ize^{iuz} f(z)\, dz = -i \int_{-\infty}^{\infty} e^{iuz} f'(z)\, dz.$$

Integrating by parts, and using that $f(z) \to 0$ as $|z| \to \infty$,

$$\phi_Z'(u) = -i\left(e^{iuz} f(z)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} iu\, e^{iuz} f(z)\, dz \right) = i^2 u\, \phi_Z(u) = -u\, \phi_Z(u).$$

The solution of this differential equation with $\phi_Z(0) = 1$ is $\phi_Z(u) = e^{-\frac{1}{2}u^2}$. Since $X = \mu + \sigma Z$, the second property of the characteristic function gives $\phi_X(u) = e^{i\mu u}\phi_Z(\sigma u) = e^{i\mu u - \frac{1}{2}\sigma^2 u^2}$. □

Recovering Probabilities

Uniqueness says that the characteristic function determines the distribution, but it is also possible to recover probabilities from it explicitly. For random variables taking integer values the formula is simple.

Inversion for Integer-Valued Random Variables

If X takes values in the integers, then for every integer k

$$\mathrm{P}(X = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iuk} \phi_X(u)\, du.$$

Proof

Writing $\phi_X(u) = \sum_j p_X(j) e^{iuj}$ and integrating term by term,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iuk} \phi_X(u)\, du = \sum_j p_X(j) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iu(j-k)}\, du.$$

If $j = k$ the integrand is one and the integral equals 2π. If $j \neq k$, the integral of $\cos(u(j-k))$ and $\sin(u(j-k))$ over a full period is zero. Therefore only the term $j = k$ survives, and it equals $p_X(k)$. □

A similar formula holds for continuous random variables, where the density is recovered as $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \phi_X(u)\, du$ whenever the integral converges. This is the Fourier inversion formula. It is the basis of the option pricing methods used in practice for models where the density of the stock price is unknown but its characteristic function has a closed form.

Convergence in Distribution

The main use of characteristic functions is to prove limit theorems. We say that a sequence of random variables X_n *converges in distribution* to X if $\mathrm{P}(X_n \leq x) \to \mathrm{P}(X \leq x)$ at every point x where the cumulative distribution function of X is continuous. The following theorem, which we state without proof, reduces convergence in distribution to the convergence of a sequence of functions.

Lévy's Continuity Theorem

The sequence X_n converges in distribution to X if and only if

$$\lim_{n\to\infty} \phi_{X_n}(u) = \phi_X(u) \quad \text{for every real } u.$$

This theorem plays the same role for general random variables that the continuity theorem for PGFs played for counts. Its most important application is the central limit theorem.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be iid random variables with mean zero and variance σ^2, and let $S_n = X_1 + \cdots + X_n$. Then $S_n/\sqrt{n}$ converges in distribution to $\mathcal{N}(0, \sigma^2)$.

Proof

Since $\mathrm{E}(X_1) = 0$ and $\mathrm{E}(X_1^2) = \sigma^2$, the fourth property of the characteristic function gives $\phi'_{X_1}(0) = 0$ and $\phi''_{X_1}(0) = -\sigma^2$. A second-order Taylor expansion around zero then yields

$$\phi_{X_1}(u) = 1 - \frac{1}{2}\sigma^2 u^2 + o(u^2),$$

where $o(u^2)$ denotes terms that vanish faster than u^2 as $u \to 0$. Using the second and third properties,

$$\phi_{S_n/\sqrt{n}}(u) = \phi_{X_1}\left(\frac{u}{\sqrt{n}}\right)^n = \left(1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}\sigma^2 u^2}.$$

The limit is the characteristic function of $\mathcal{N}(0, \sigma^2)$, and Lévy's continuity theorem completes the proof. □

The structure of the proof mirrors the Poisson limit. In both cases we raise the transform of a single term to the n-th power and use $(1 + x/n)^n \to e^x$. In the Poisson limit each term is a rare event with probability λ/n. In the central limit theorem each term is a small shock of size $1/\sqrt{n}$. The notes on Brownian motion apply this argument to a random walk with steps $\pm\sqrt{\Delta t}$, using the characteristic function $\cos(u\sqrt{\Delta t})$ from Example 14.

Summary

The table below collects the transforms of the distributions used in these notes. The MGFs of the geometric distribution and of the exponential and gamma distributions are finite only for $\theta < -\ln(1-p)$ and $\theta < \lambda$, respectively. The characteristic function of each can be obtained from its MGF by setting $\theta = iu$.

Distribution	Support	PGF $G(z)$	MGF $M(\theta)$	Mean	Variance
Bernoulli(p)	$\{0,1\}$	$1-p+pz$	$1-p+pe^{\theta}$	p	$p(1-p)$
Binomial(n,p)	$\{0,\ldots,n\}$	$(1-p+pz)^n$	$(1-p+pe^{\theta})^n$	np	$np(1-p)$
Geometric(p)	$\{1,2,\ldots\}$	$\frac{pz}{1-(1-p)z}$	$\frac{pe^{\theta}}{1-(1-p)e^{\theta}}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	$\{0,1,\ldots\}$	$e^{\lambda(z-1)}$	$e^{\lambda(e^{\theta}-1)}$	λ	λ
Exp(λ)	$[0,\infty)$		$\frac{\lambda}{\lambda-\theta}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(α,λ)	$(0,\infty)$		$\left(\frac{\lambda}{\lambda-\theta}\right)^{\alpha}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\mathcal{N}(\mu,\sigma^2)$	$\mathbb{R}$		$e^{\mu\theta+\frac{1}{2}\sigma^2\theta^2}$	μ	σ^2

Looking Ahead: The Ruin Problem

Consider a random walk whose steps X_i take the value $+1$ with probability p and -1 with probability $q = 1-p$. The steps take negative values, so they have no PGF in the strict sense, but $\mathrm{E}(z^X) = pz + q/z$ is still well defined for $z > 0$. Setting this expression equal to one,

$$pz + \frac{q}{z} = 1 \quad \Leftrightarrow \quad pz^2 - z + q = 0,$$

gives the roots $z = 1$ and $z = q/p$. Whenever $\mathrm{E}(z^X) = 1$, the process z^{S_n} is a martingale, which is why the notes on ruin probabilities try solutions of the form $f(K) = z^K$. When $p \neq q$ the second root q/p gives the ruin probabilities for a biased coin. When $p = q = 1/2$ the equation becomes $(z-1)^2 = 0$ and both roots equal one, so z^K only produces the constant solution. The second solution is $f(K) = K$, which comes from the fact that

the position S_n of a symmetric walk is itself a martingale. We will develop this connection in the notes on ruin probabilities and on martingales.

Practice Problems

Problem 1. Identify the distribution of each random variable from its MGF, and give its mean.

a. $M_A(\theta) = (0.25 + 0.75e^{\theta})^4$.
b. $M_B(\theta) = e^{3\theta + 8\theta^2}$.
c. $M_C(\theta) = (1 - \theta/2)^{-3}$ for $\theta < 2$.
d. $M_D(\theta) = e^{\theta}(0.4 + 0.6e^{\theta})$.

Problem 2. Let U have the uniform density on $[0, 1]$.

a. Show that $M_U(\theta) = (e^{\theta} - 1)/\theta$ for $\theta \neq 0$, with $M_U(0) = 1$.
b. Expand $M_U(\theta)$ as a power series in θ and read off $\mathrm{E}(U^k)$. Compute $\mathrm{E}(U)$ and $\mathrm{V}(U)$.
c. Use the properties of the MGF to find the MGF of $a + (b - a)U$, which has the uniform distribution on $[a, b]$.

Problem 3. An insurer receives $N \sim \text{Poisson}(50)$ claims next year. The claim sizes $X_1, X_2, \ldots$ are independent of each other and of N, each with mean \$20,000 and standard deviation \$30,000. Total claims are $S = X_1 + \cdots + X_N$.

a. Use $\mathrm{E}(S \mid N) = N\,\mathrm{E}(X)$, $\mathrm{V}(S \mid N) = N\,\mathrm{V}(X)$, and the decomposition of the variance

$$\mathrm{V}(S) = \mathrm{E}(\mathrm{V}(S \mid N)) + \mathrm{V}(\mathrm{E}(S \mid N))$$

to show that

$$\mathrm{E}(S) = \mathrm{E}(N)\mathrm{E}(X), \qquad \mathrm{V}(S) = \mathrm{E}(N)\mathrm{V}(X) + \mathrm{V}(N)\mathrm{E}(X)^2.$$

b. Compute $\mathrm{E}(S)$ and $\mathrm{SD}(S)$. Show that $\mathrm{V}(S) = \lambda\,\mathrm{E}(X^2)$ whenever $N \sim \text{Poisson}(\lambda)$.
c. Compare $\mathrm{SD}(S)$ with the standard deviation of total claims if the insurer knew it would receive exactly 50 claims.

Problem 4 (⋆). A random variable X taking values in $\mathbb{N}$ has the probability generating function

$$G_X(z) = \frac{2}{(2-z)(3-z)}.$$

a. Check that $G_X(1) = 1$.
b. Write G_X as a sum of partial fractions and expand each one as a geometric series to show that $\mathrm{P}(X = k) = 2^{-k} - \frac{2}{3}3^{-k}$. Compute $\mathrm{P}(X = 0)$, $\mathrm{P}(X = 1)$, and $\mathrm{P}(X = 2)$.
c. Write G_X as a product of two PGFs and interpret X as a sum of two independent random variables. Hint: if Y has the geometric distribution of Example 2, then $Y - 1$ counts the failures before the first success and has PGF $G_Y(z)/z$.
d. Compute $\mathrm{E}(X)$.

Problem 5 (⋆). The third derivative $\kappa_3 = K_X'''(0)$ of the cumulant generating function equals $\mathrm{E}\left((X - \mathrm{E}\,(X))^3\right)$, and the *skewness* of X is $\kappa_3/\mathrm{V}(X)^{3/2}$.

a. Show that the cumulants of a sum of independent random variables are the sums of their cumulants.
b. Show that every cumulant of $N \sim \text{Poisson}(\lambda)$ equals λ, and compute its skewness for $\lambda = 2$ and $\lambda = 100$.
c. Let S_n be the sum of n iid random variables with variance σ^2 and third cumulant κ_3. Show that the skewness of S_n is the skewness of one term divided by $\sqrt{n}$.
d. Use the cumulant generating function of the $\text{Gamma}(n, \lambda)$ distribution to show that its skewness is $2/\sqrt{n}$.

Problem 6 (⋆). A bank holds 1,000 independent loans. Use a normal table or software for the standard normal CDF Φ.

a. Each loan defaults with probability 0.1. Use the central limit theorem to approximate the probability of at least 120 defaults. The exact probability is 0.0220.
b. Use the previous problem to explain why the approximation falls short of the exact probability.
c. Suppose instead that each loan defaults with probability 0.002. Compute the skewness of the number of defaults, and the probability that the normal approximation

assigns to a negative number of defaults. Which approximation would you use instead?

Problem 7 (★★). The number of orders that arrive at a market maker during a minute is $N \sim \text{Poisson}(\lambda)$, and each order is independently a buy with probability p and a sell with probability $q = 1 - p$. Let B and S be the numbers of buy and sell orders, so $B + S = N$.

a. Show that the joint PGF of B and S is

$$\mathrm{E}(z^B w^S) = \mathrm{E}\left((pz + qw)^N\right) = e^{\lambda p(z-1)}\, e^{\lambda q(w-1)}.$$

b. The coefficient of $z^b w^s$ in the joint PGF is $\mathrm{P}(B = b, S = s)$. Use this to show that B and S are independent, with $B \sim \text{Poisson}(\lambda p)$ and $S \sim \text{Poisson}(\lambda q)$.

c. Suppose instead that exactly n orders arrive. Compute $\mathrm{Cov}(B, S)$ and the correlation of B and S.

Problem 8 (★★). Moments and the MGF are enough to bound tail probabilities.

a. Let $Y \geq 0$ be a random variable and $a > 0$. Show *Markov's inequality*,

$$\mathrm{P}(Y \geq a) \leq \frac{\mathrm{E}(Y)}{a}.$$

b. Apply Markov's inequality to $Y = (X - \mathrm{E}(X))^2$ to obtain *Chebyshev's inequality*,

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq a) \leq \frac{\mathrm{V}(X)}{a^2}.$$

c. Apply Markov's inequality to $Y = e^{\theta X}$ to obtain the *Chernoff bound*

$$\mathrm{P}(X \geq a) \leq e^{-\theta a} M_X(\theta) \qquad \text{for every } \theta > 0.$$

d. Let $N \sim \text{Poisson}(2)$ be the number of defaults in a loan book. Find the θ that gives the smallest Chernoff bound on $\mathrm{P}(N \geq 8)$, compute the bound, and compare it with the exact probability 0.0011 and with the bound from Chebyshev's inequality.

e. For $Z \sim \mathcal{N}(0, 1)$, show that the best Chernoff bound is

$$\mathrm{P}(Z \geq a) \leq e^{-a^2/2},$$

and compare it with the exact tail for $a = 2$ and $a = 3$.

References

Blitzstein, Joseph K., and Jessica Hwang. 2019. *Introduction to Probability*. 2nd ed. Chapman, Hall/CRC.

Feller, William. 1968. *An Introduction to Probability Theory and Its Applications. Volume 1*. 3rd ed. Wiley.

Grimmett, Geoffrey, and David Stirzaker. 2001. *Probability and Random Processes*. 3rd ed. Oxford University Press.