4. Generating Functions

As usual, our starting point is a random experiment with probability measure $\mathbb{P}$ on an underlying sample space. A **generating function** of a random variable is an expected value of a certain transformation of the variable. Most generating functions share four important properties:

1. Under mild conditions, the generating function completely determines the distribution.
2. The generating function of a sum of independent variables is the product of the generating functions
3. The moments of the random variable can be obtained from the derivatives of the generating function.
4. Ordinary (pointwise) convergence of a sequence of generating functions corresponds to the special convergence of the corresponding distributions.

Property 1 is most important. Often a random variable is shown to have a certain distribution by showing that the generating function has a certain form. The process of recovering the distribution from the generating function is known as **inversion**. Property 2 is frequently used to determine the distribution of a sum of independent variables. By contrast, recall that the probability density function of a sum of independent variables is the convolution of the individual density functions, a much more complicated operation. Property 3 is useful because often computing moments from the generating function is easier than computing the moments directly from the definition. The last property is known as the **continuity theorem**. Often it is easer to show the convergence of the generating functions than to prove convergence of the distributions directly.

The Probability Generating Function

Definition

Suppose that X is a random variable taking values in $\mathbb{N}$. The **probability generating function** G of X is defined as follows, for all values $t \in \mathbb{R}$ for which the expected value exists:

$$G(t) = \mathbb{E}\left(t^X\right)$$

Let f denote the probability density function of X, so that $f(n) = \mathbb{P}(X = n)$, $n \in \mathbb{N}$.

1. Show that

$$G(t) = \sum_{n=0}^{\infty} f(n)\, t^n$$

Thus, $G(t)$ is a **power series** in t, with the values of the probability density function as the coefficients. In the language of combinatorics, G is the **ordinary generating function** of f. Recall from calculus that there exists $r \in [0, \infty]$ such that the series converges absolutely for $|t| < r$ and diverges for $|t| > r$. The number r is the **radius of convergence** of the series.

2. Show that $G(1) = 1$ and hence $r \geq 1$.

Inversion

Recall from calculus that a power series can be differentiated term by term, just like a polynomial. Each derivative series has the same radius of convergence as the original series. We denote the derivative of order n by $G^{(n)}$.

3. Show that $f(n) = \frac{G^{(n)}(0)}{n!}$ for $n \in \mathbb{N}$. Thus, the probability generating function G completely determines the distribution of X.

4. Show that $\mathbb{P}(N \text{ is even}) = \frac{1+G(-1)}{2}$. This is not a particularly important result, but has a certain curiosity.

Moments

Recall that if n and k are nonnegative integers with $k \le n$, then the number of permutations of size k chosen from a population of n objects is

$$n^{(k)} = n\,(n-1)\,\cdots\,(n-(k-1))$$

5. Suppose that the radius of convergence satisfies $r > 1$. Show that $G^{(k)}(1) = \mathbb{E}\left(X^{(k)}\right)$ for $k \in \mathbb{N}$; these moments are called **factorial moments**. In particular, X has finite moments of all orders.

6. In particular, show that

   a. $\mathbb{E}(X) = G'(1)$
   b. $\text{var}(X) = G''(1) + G'(1)\,(1 - G'(1))$

Sums

7. Suppose that X_1 and X_2 are independent random variables with probability generating functions G_1 and G_2 respectively. Show that the probability generating function of $X_1 + X_2$ is

$$G(t) = G_1(t)\,G_2(t)$$

The Moment Generating Function

Definition

Let X be a real-valued random variable. The **moment generating function** of X is the function M defined by

$$M(t) = \mathbb{E}\left(e^{t\,X}\right), \quad t \in \mathbb{R}$$

Note that since $e^{t\,X} \ge 0$ with probability 1, $M(t)$ exists, as a real number or ∞, for any $t \in \mathbb{R}$.

8. Suppose that X has a continuous distribution on $\mathbb{R}$ with probability density function f. Show that

$$M(t) = \int_{-\infty}^{\infty} e^{t\,x}\, f(x) dt$$

Thus, the moment generating function of X is closely related to the **Laplace transform** of the density function f. The Laplace transform is named for Simeon Laplace, and is widely used in many areas of applied mathematics.

Inversion

The basic inversion theorem for moment generating functions states that if $M(t) < \infty$ for t is some open interval about 0, then M completely determines the distribution of X. Thus, if two distributions on $\mathbb{R}$ have moment generating functions that are equal (and finite) in an open interval about 0, then the distributions are the same.

Moments

9. Suppose that X has moment generating function M and that M is finite in some interval about 0. Then X has moments of all orders. Show formally the following result (the interchange of sum and expected value is justified by the finiteness assumption).

$$M(t) = \sum_{n=0}^{\infty} \frac{\mathbb{E}(X^n)}{n!} t^n$$

10. Show that $M^{(n)}(0) = \mathbb{E}(X^n)$ for $n \in \mathbb{N}$. Thus, the derivatives of the moment generating function at 0 determine the moments of the variable (hence the name). In the language of combinatorics, the moment generating function is the **exponential generating function** of the sequence of moments.

Thus, a random variable that does not have finite moments of all orders cannot have a finite moment generating function. Even when a random variable does have moments of all orders, the moment generating function may not exist. A counterexample is given below.

Transformations

11. Suppose that X is a real-valued random variable with moment generating function M and that a and b are constants. Show that the moment generating function of $Y = a\,X + b$ is

$$N(t) = e^{b\,t}\, M(a\,t)$$

12. Suppose that X_1 and X_2 are independent, real-valued random variables with moment generating functions M_1 and M_2 respectively. Show that the moment generating function of $Y = X_1 + X_2$ is

$$M(t) = M_1(t)\, M_2(t)$$

13. Suppose that X is a random variable taking values in $\mathbb{N}$ with probability generating function G. Show that the moment generating function of X is

$$M(t) = G\left(e^t\right)$$

The Chernoff Bounds

14. Suppose that X is a real-valued random variable with moment generating function M. Prove the **Chernoff**

bounds:

a. $\mathbb{P}(X \ge x) \le e^{-t\,x}\, M(t), \quad t > 0$
b. $\mathbb{P}(X \le x) \le e^{-t\,x}\, M(t), \quad t < 0$

Hint: Show that $\mathbb{P}(X \ge x) = \mathbb{P}\left(e^{t\,X} \ge e^{t\,x}\right)$ if $t > 0$ and $\mathbb{P}(X \le x) = \mathbb{P}\left(e^{t\,X} \ge e^{t\,x}\right)$ if $t < 0$. Then use Markov's inequality.

Naturally, the best Chernoff bound (in either (a) or (b)) is obtained by finding t that minimizes $e^{-t\,x}\, M(t)$.

The Characteristic Function

Definition

From a mathematical point of view, the nicest of the generating functions is the **characteristic function** which is defined for a real-valued random variable X by

$$\chi(t) = \mathbb{E}\left(e^{i\,t\,X}\right) = \mathbb{E}\left(\cos\left(t\,X\right)\right) + i\,\mathbb{E}\left(\sin\left(t\,X\right)\right), \quad t \in \mathbb{R}$$

Note that χ is a complex valued function, and thus this subsection requires knowledge of complex analysis, at the undergraduate level. Note that χ is defined for all $t \in \mathbb{R}$ because the random variable in the expected value is bounded in magnitude. Indeed, $\left|e^{i\,t\,X}\right| = 1$ for all $t \in \mathbb{R}$. Many of the properties of the characteristic function are more elegant than the corresponding properties of the probability or moment generating functions, because the characteristic function always exists.

15. Suppose that X has a continuous distribution on $\mathbb{R}$ with probability density function f. Show that

$$\chi(t) = \int_{-\infty}^{\infty} e^{i\,t\,x}\, f(x)\,d\,x$$

Thus, the characteristic function of X is closely related to the **Fourier transform** of the density function f. The Fourier transform is named for Joseph Fourier, and is widely used in many areas of applied mathematics.

Inversion

The characteristic function completely determines the distribution. That is, random variables X and Y have the same distribution if and only if they have the same characteristic function. Indeed, the general **inversion formula** is a formula for computing certain combinations of probabilities from the characteristic function: if $a < b$ then

$$\int_{-n}^{n} \frac{e^{-i\,a\,t} - e^{-i\,b\,t}}{2\,\pi\,i\,t}\, \chi(t)\,d\,t \to \mathbb{P}(a < X < b) + \frac{1}{2}\left(\mathbb{P}(X = b) - \mathbb{P}(X = a)\right) \text{ as } \quad n \to \infty$$

The probability combinations on the right side completely determine the distribution of X. Suppose that X has a continuous distribution with probability density function f. A special inversion formula states that at every point x

where f is differentiable,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i t x} \chi(t) dt$$

Moments

As with the other generating functions, the characteristic function can be used to find the moments of X. Moreover, this can be done even when only some of the moments exist. If $\mathbb{E}(|X|^n) < \infty$ then

$$\chi(t) = \sum_{k=0}^{n} \frac{\mathbb{E}\left(X^k\right)}{k!} (i t)^k + o\left(t^n\right)$$

and therefore $\chi^{(n)}(0) = i^n \mathbb{E}(X^n)$

Transformations

16. Suppose that X is a real-valued random variable with characteristic function χ and that a and b are constants. Show that the characteristic function of $Y = a X + b$ is

$$\psi(t) = e^{i b t} \chi(a t)$$

17. Suppose that X_1 and X_2 are independent, real-valued random variables with characteristic functions χ_1 and χ_2 respectively. Show that the characteristic function of $Y = X_1 + X_2$ is

$$\chi(t) = \chi_1(t) \chi_2(t)$$

The characteristic function of a random variable can be obtained from the moment generating function, under the basic existence condition that we saw earlier. Specifically, suppose that X is a real-valued random variable with moment generating function M that satisfies $M(t) < \infty$ for t in some interval I about 0. Then the characteristic function χ of X satisfies $\chi(t) = M(i t)$ for $t \in I$.

Convergence in Distribution

The final important property of characteristic functions that we will discuss relates to convergence in distribution. Suppose that $(X_1, X_2, \ldots)$ is a sequence of real-valued random with characteristic functions $(\chi_1, \chi_2, \ldots)$ respectively. The random variables need not be defined on the same probability space.

The **continuity theorem** states that if the distribution of X_n converges to the distribution of a random variable X as $n \to \infty$ and X has characteristic function χ, then $\chi_n(t) \to \chi(t)$ as $n \to \infty$ for all t. Conversely, if $\chi_n(t) \to \chi(t)$ as $n \to \infty$ for t in some open interval about 0, and χ is continuous at 0, then χ is the characteristic function of a random variable X, and the distribution of X_n converges to the distribution of X as $n \to \infty$.

The continuity theorem can be used to prove the central limit theorem, one of the fundamental theorems of probability. Also, the continuity theorem has a straightforward generalization to distributions on $\mathbb{R}^n$.

The Joint Characteristic Function

Suppose now that (X, Y) is a random vector for an experiment, taking values in $\mathbb{R}^2$. The (joint) **characteristic function** of (X, Y) is defined by

$$\chi(s, t) = \mathbb{E}\left(\exp\left(i\,s\,X + i\,t\,Y\right)\right), \quad (s, t) \in \mathbb{R}^2$$

Once again, the most important fact is that χ completely determines the distribution: two random vectors taking values in $\mathbb{R}^2$ have the same characteristic function if and only if they have the same distribution.

The joint moments can be obtained from the derivatives of the characteristic function. Suppose that $m \in \mathbb{N}$ and $n \in \mathbb{N}$. If $\mathbb{E}\left(\left|X^m\,Y^n\right|\right) < \infty$ then

$$\chi^{(m,n)}\left(0, 0\right) = i^{m+n}\,\mathbb{E}\left(X^m\,Y^n\right)$$

Now let χ_1, χ_2, and χ_+ denote the characteristic functions of X, Y, and $X + Y$, respectively.

18. Show that

a. $\chi(s, 0) = \chi_1(s)$
b. $\chi(0, t) = \chi_2(t)$
c. $\chi(t, t) = \chi_+(t)$

19. Show that X and Y are independent if and only if $\chi(s, t) = \chi_1(s)\,\chi_2(t)$ for all (s, t).

Naturally, the results for bivariate characteristic functions have analogies in the general multivariate case. Only the notation is more complicated.

Examples and Applications

Bernoulli Trials

20. Suppose X is an indicator random variable with $\mathbb{P}(X = 1) = p$, where $p \in [0, 1]$ is a parameter. Show that X has probability generating function $G(t) = 1 - p + p\,t$ for $t \in \mathbb{R}$

Recall that a **Bernoulli trials process** is a sequence $(X_1, X_2, \ldots)$ of independent, identically distributed indicator random variables. In the usual language of reliability, X_i denotes the outcome of trial i, where 1 denotes success and 0 denotes failure. The probability of success $p = \mathbb{P}(X_i = 1)$ is the basic parameter of the process. The process is named for James Bernoulli. A separate chapter on the Bernoulli Trials explores this process in more detail.

The number of successes in the first n trials is $Y_n = \sum_{i=1}^n X_i$. Recall that this random variable has the binomial distribution with parameters n and p, which has density function

$$\mathbb{P}(Y_n = k) = \binom{n}{k}\,p^k\,(1 - p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}$$

21. Show that Y_n has probability generating function $G_n(t) = \left(1 - p + p\,t\right)^n$, $t \in \mathbb{R}$ in two ways:

a. Using the definition and the probability density function.
b. Using the previous exercise and the representation as a sum of indicator variables.

22. Show that

a. $\mathbb{E}\left(Y_n^{(k)}\right) = n^{(k)}\ p^k$
b. $\mathbb{E}(Y_n) = n\,p$
c. $\text{var}(Y_n) = n\,p\,(1-p)$
d. $\mathbb{P}(Y_n \text{ is even}) = \frac{1}{2}\left(1 - \left(1 - 2\,p\right)^n\right)$

23. Suppose that W has probability density function $\mathbb{P}(N = n) = p\,(1-p)^{n-1}\quad n \in \mathbb{N}_+$ where $p \in (0, 1]$ is a parameter. Thus, W has the geometric distribution on $\mathbb{N}_+$ with parameter p, and models the trial number of the first success in a sequence of Bernoulli trials. Let H denote the probability generating function of W. Show that

a. $H(t) = \frac{p\,t}{1-(1-p)\,t},\quad t < \frac{1}{1-p}$
b. $\mathbb{E}\left(W^{(k)}\right) = k!\,\frac{(1-p)^{k-1}}{p^k}$
c. $\mathbb{E}(W) = \frac{1}{p}$
d. $\text{var}(W) = \frac{1-p}{p^2}$
e. $\mathbb{P}(W \text{ is even}) = \frac{1-p}{2-p}$

24. Suppose that U has the binomial distribution with parameters m and p, V has the binomial distribution with parameters n and q, and that U and V are independent.

a. Find the probability generating function of $U + V$
b. Show that if $p = q$ then $U + V$ has the binomial distribution with parameters $m + n$ and p.
c. Show that if $p \neq q$ then $U + V$ does not have a binomial distribution.

The Poisson Distribution

Recall that the **Poisson distribution** has density function

$$f(n) = e^{-a}\,\frac{a^n}{n!},\quad n \in \mathbb{N}$$

where $a > 0$ is a parameter. The Poisson distribution is named after Simeon Poisson and is widely used to model the number of “random points” in a region of time or space; the parameter is proportional to the size of the region of time or space. The Poisson distribution is studied in more detail in the chapter on the Poisson Process.

25. Suppose that X has Poisson distribution with parameter a. Let G denote the probability generating function of X. Show that

a. $G(t) = e^{a(t-1)}, \; t \in \mathbb{R}$
b. $\mathbb{E}\left(X^{(k)}\right) = a^k$
c. $\mathbb{E}(X) = a$
d. $\text{var}(X) = a$
e. $\mathbb{P}(X \text{ is even}) = \frac{1}{2}\left(1 + e^{-2a}\right)$

26. Suppose that X has the Poisson distribution with parameter a, Y has the Poisson distribution with parameter b, and that X and Y are independent. Show that $X + Y$ has the Poisson distribution with parameter $a + b$.

27. Suppose that X has the Poisson distribution with parameter $a > 0$. Use the Chernoff bounds to show that

$$\mathbb{P}(X \ge n) \le e^{n-a} \left(\frac{a}{n}\right)^n, \quad n > a$$

28. Let G_n denote the probability generating function of the binomial distribution with parameters n and p_n, where $n\, p_n \to a$ as $n \to \infty$ (and where $a > 0$). Let G denote the probability generating function of the Poisson distribution with parameter a. Show that $G_n(t) \to G(t)$ as $n \to \infty$ for any $t \in \mathbb{R}$. Thus conclude that the binomial distribution with parameters n and p_n converges to the Poisson distribution with parameter a as $n \to \infty$.

The Exponential Distribution

Recall that the **exponential distribution** is a continuous distribution with probability density function

$$f(t) = r\, e^{-r\,t}, \; t \ge 0$$

where $r > 0$ is the **rate parameter**. This distribution is widely used to model failure times and other "arrival times". The exponential distribution is studied in more detail in the chapter on the Poisson Process.

29. Suppose that X has the exponential distribution with rate parameter r. Let M denote the moment generating function of X. Show that

a. $M(s) = \frac{r}{r-s}, \; s < r.$
b. $\mathbb{E}(X^n) = \frac{n!}{r^n}$

30. Suppose that $X = (X_1, X_2, \ldots, X_n)$ is a sequence of independent random variables, each having the exponential distribution with rate parameter r. Find the moment generating function of $T = \sum_{i=1}^n X_i$. Random variable T has the gamma distribution with parameters n and a.

Uniform Distributions

31. Suppose that X is uniformly distributed on the interval $[a, b]$. Let M denote the moment generating function of X. Show that

a. $M(t) = \begin{cases} \frac{e^{b t} - e^{a t}}{(b - a) t}, & t \ne 0 \\ 1, & t = 0 \end{cases}$

b. $\mathbb{E}(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b - a)}$

32. Suppose that (X, Y) is uniformly distributed on the triangle $T = \left\{(x, y) \in \mathbb{R}^2 : 0 \le x \le y \le 1\right\}$.

a. Find the joint moment generating function of (X, Y).
b. Find the moment generating function of X.
c. Find the moment generating function of Y.
d. Find the moment generating function of $X + Y$

33. Suppose that (X, Y) has probability density function $f(x, y) = x + y$, $0 \le x \le 1$, $0 \le y \le 1$

a. Find the joint moment generating function (X, Y).
b. Find the moment generating function of X.
c. Find the moment generating function of Y.
d. Find the moment generating function of $X + Y$

The Normal Distribution

Recall that the **standard normal distribution** is a continuous distribution with density function $\phi(z) = \frac{1}{\sqrt{2 \pi}} e^{-\frac{1}{2} z^2}$, $z \in \mathbb{R}$. Normal distributions are widely used to model physical measurements subject to small, random errors and are studied in more detail in the chapter on Special Distributions.

34. Suppose that Z has the standard normal distribution. Let M denote the moment generating function of Z. Show that

a. $M(t) = e^{\frac{1}{2} t^2}$, $t \in \mathbb{R}$
b. $\mathbb{E}\left(Z^{2 n}\right) = \frac{(2 n)!}{2^n n!}$, $n \in \mathbb{N}$
c. $\mathbb{E}\left(Z^{2 n + 1}\right) = 0$, $n \in \mathbb{N}$

35. Suppose again that Z has the standard normal distribution. Recall that $X = \mu + \sigma Z$ has the normal distribution with mean μ and standard deviation σ. Show that the moment generating function of X is $M(t) = \exp\left(\mu t + \frac{1}{2} \sigma^2 t^2\right)$ for $t \in \mathbb{R}$.

36. Suppose that X and Y are independent random variables, each with a normal distribution. Show that $X + Y$ has a normal distribution.

The Pareto Distribution

Suppose that X has the **Pareto distribution**, which is a continuous distribution with probability density function

$$f(x) = \frac{a}{x^{a+1}}, \quad x \geq 1$$

where $a > 0$ is a parameter. The Pareto distribution is named for Vilfredo Pareto. It is a heavy-tailed distribution that is widely used to model financial variables such as income. The Pareto distribution is studied in more detail in the chapter on Special Distributions.

37. Let M denote the moment generating function of X. Show that

a. $\mathbb{E}(X^n) = \begin{cases} \frac{a}{a-n}, & n < a \\ \infty, & n \geq a \end{cases}$
b. $M(t) = \infty$ for any $t > 0$.

The Cauchy Distribution

Suppose that X has the **Cauchy distribution**, a continuous distribution with probability density function

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}, \quad x \in \mathbb{R}$$

.

This distribution is named for Augustin Cauchy and is a member of the family of **student t distributions**. The t distributions are studied in more detail in the chapter on Special Distributions. The graph of f is known as the **Witch of Agnesi**, named for Maria Agnesi.

38. Let M denote the moment generating function of X. Show that

a. $\mathbb{E}(X)$ does not exist.
b. $M(t) = \infty$ for $t > 0$.

39. Let χ denote the characteristic function of X. Show that $\chi(t) = e^{-|t|}$ for $t \in \mathbb{R}$.

Counterexample

For the Pareto distribution, only some of the moments are finite; naturally, the moment generating function was infinite. We will now give an example of a distribution for which all of the moments are finite, yet still the moment generating function is infinite. Furthermore, we will see two different distributions that have the same moments of all orders.

Suppose that Z has the standard normal distribution and let $X = e^Z$. The distribution of X is known as a lognormal distribution.

40. Use the change of variables formula to show that X has probability density function

$$f(x) = \frac{\exp\left(-\ln(x)^2/2\right)}{\sqrt{2\pi}\, x}, \quad x > 0$$

41. Use the moment generating function of the standard normal distribution to show that $\mathbb{E}(X^n) = e^{\frac{1}{2}n^2}$ for $n \in \mathbb{N}$. Thus, X has finite moments of all orders.

42. Show that $\mathbb{E}\left(e^{tX}\right) = \infty$ for $t > 0$. Thus, the moment generating function of X is infinite at any positive value of t.

43. Now let $h(x) = \sin\left(2\pi \ln(x)\right)$, $x > 0$. Show that for $n \in \mathbb{N}$,

$$\int_0^\infty x^n f(x) h(x) dx = e^{-\frac{1}{2}n^2}\, \mathbb{E}\left(\sin\left(2\pi U\right)\right)$$

where U is normally distributed with mean n and standard deviation 1. *Hint:* Use the change of variables $u = \ln(x)$ and then complete the square in the exponential factor.

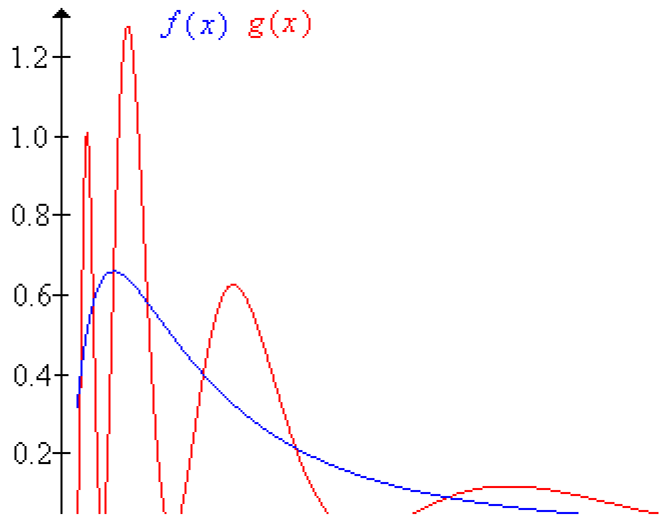
44. Use the result in the previous exercise and symmetry to conclude that for $n \in \mathbb{N}$,

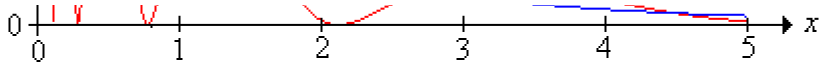
$$\int_0^\infty x^n f(x) h(x) dx = 0$$

45. Let $g(x) = f(x)(1 + h(x))$, $x > 0$. Use the result in the previous exercise to show that g is a probability density function.

46. Let Y have probability density function g. Use the result in Exercise 40 to show that Y has the same moments as X. That is, $\mathbb{E}(Y^n) = e^{\frac{1}{2}n^2}$ for $n \in \mathbb{N}$.

The graphs of f and g are shown below, in blue and red, respectively.




0
0
1
2
3
4
5
x