7. Transformations of Variables

Basic Theory

The Problem

As usual, we start with a random experiment with probability measure $\mathbb{P}$ on an underlying sample space. Suppose that we have a random variable X for the experiment, taking values in S, and a function $r : S \to T$. Then $Y = r(X)$ is a new random variable taking values in T. If the distribution of X is known, how do we find the distribution of Y? This is a very basic and important question, and in a superficial sense, the solution is easy.

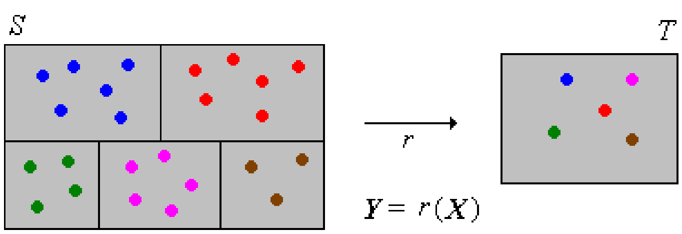
1. Show that $\mathbb{P}(Y \in B) = \mathbb{P}\left(X \in r^{-1}(B)\right)$ for $B \subseteq T$.

However, frequently the distribution of X is known either through its distribution function F or its density function f, and we would similarly like to find the distribution function or density function of Y. This is a difficult problem in general, because as we will see, even simple transformations of variables with simple distributions can lead to variables with complex distributions. We will solve the problem in various special cases.

Transformed Variables with Discrete Distributions

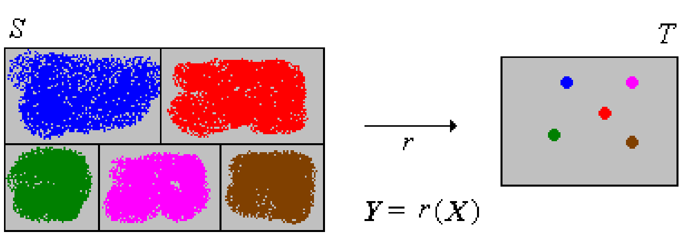
2. Suppose that X has a discrete distribution with probability density function f (and hence S is countable). Show that Y has a discrete distribution with probability density function g given by

$$g(y) = \sum_{x \in r^{-1}(y)} f(x), \quad y \in T$$



3. Suppose that X has a continuous distribution on a subset $S \subseteq \mathbb{R}^n$ with probability density function f, and that T is countable. Show that Y has a discrete distribution with probability density function g given by

$$g(y) = \int_{r^{-1}(y)} f(x)dx, \quad y \in T$$



Transformed Variables with Continuous Distributions

Suppose that X has a continuous distribution on a subset $S \subseteq \mathbb{R}^n$ and that $X = r(X)$ has a continuous distributions on a subset

$T \subseteq \mathbb{R}^m$. Suppose also that X has a known probability density function f. In many cases, the density function of Y can be found by first finding the distribution function of Y (using basic rules of probability) and then computing the appropriate derivatives of the distribution function. This general method is referred to, appropriately enough, as the **distribution function method**.

The Change of Variables Formula

When the transformation r is one-to-one and smooth, there is a formula for the probability density function of Y directly in terms of the probability density function of X. This is known as the **change of variables** formula.

We will explore the one-dimensional case first, where the concepts and formulas are simplest. Thus, suppose that random variable X has a continuous distribution on an interval $S \subseteq \mathbb{R}$. with distribution function F and density function f. Suppose that $Y = r(X)$ where r is a differentiable function from S onto an interval $T \subseteq \mathbb{R}$. As usual, we will let G denote the distribution function of Y and g the density function of Y.

4. Suppose that r is strictly increasing on S. Show that for $y \in T$,

   a. $G(y) = F\left(r^{-1}(y)\right)$
   b. $g(y) = f\left(r^{-1}(y)\right) \dfrac{d r^{-1}(y)}{d y}$

5. Suppose that r is strictly decreasing on S. Show that for $y \in T$,

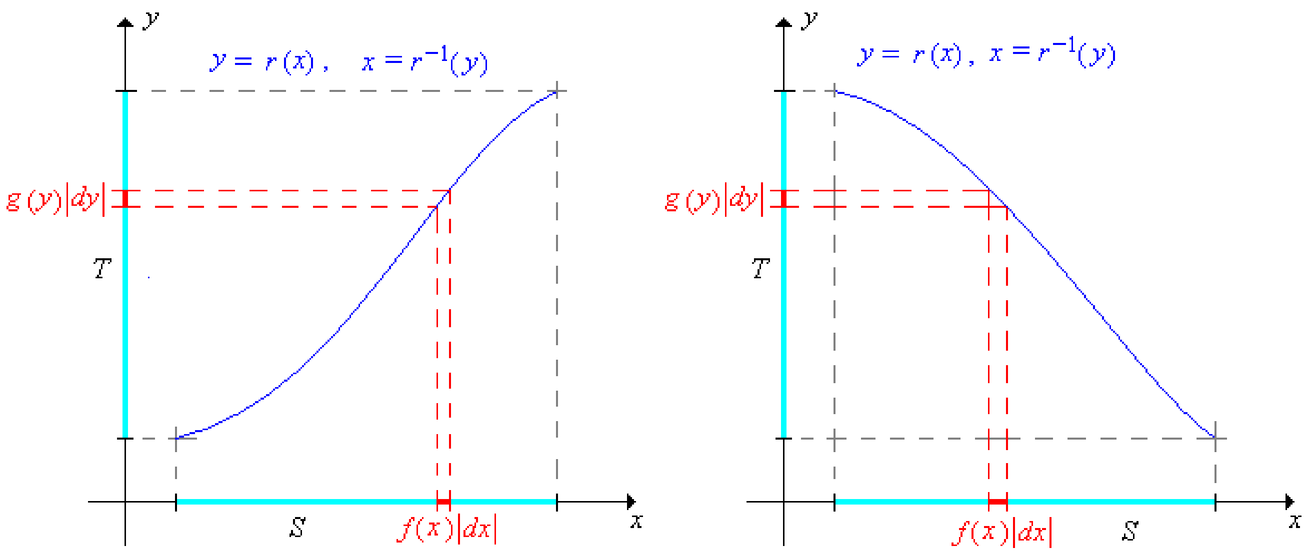
   a. $G(y) = 1 - F\left(r^{-1}(y)\right)$
   b. $g(y) = \left(-f\left(r^{-1}(y)\right)\right) \dfrac{d r^{-1}(y)}{d y}$

The density function formulas in Exercises 4 and 5 can be combined: if r is a strictly monotone on S then the density function g of Y is given by

$$g(y) = f\left(r^{-1}(y)\right) \left|\frac{d r^{-1}(y)}{d y}\right|, \quad y \in T$$

Letting $x = r^{-1}(y)$, the change of variables formula can be written more compactly as

$$g(y) = f(x) \left|\frac{d x}{d y}\right|, \quad y \in T$$



The generalization this result is basically a theorem in multivariate calculus. Suppose that X is a random variable taking values in $S \subseteq \mathbb{R}^n$, and that X has a continuous distribution with probability density function f. Suppose that $Y = r(X)$ where r is a one-to-one, differentiable function form S onto $T \subseteq \mathbb{R}^n$. The **first derivative** of the inverse function $x = r^{-1}(y)$ is the $n \times n$ matrix of first partial derivatives:

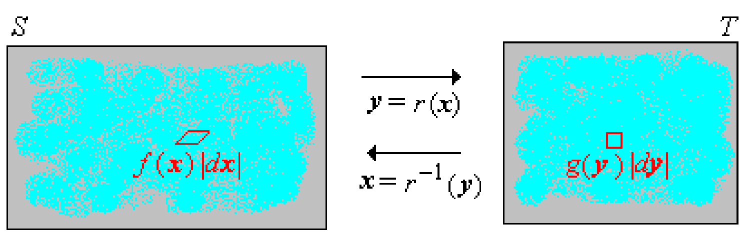
$$\left(\frac{dx}{dy}\right)_{i,j} = \frac{dy_j}{dx_i}$$

The **Jacobian** (named in honor of Karl Gustav Jacobi) of the inverse function is the determinant of the first derivative matrix

$$\det\left(\frac{dx}{dy}\right)$$

With this compact notation, the multivariate change of variables formula states that the density g of Y is given by

$$g(y) = f(x)\left|\det\left(\frac{dx}{dy}\right)\right|, \quad y \in T$$



Special Transformations

Linear Transformations

Linear transformations (or more technically **affine transformations**) are among the most common and important transformations.

Moreover, this type of transformation leads to a simple application of the change of variable theorem. Suppose that X is a random variable taking values in $S \subseteq \mathbb{R}$ and that X has a continuous distribution on S with probability density function f. Let $Y = a + b\,X$ where $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$ Note that and that Y takes values in $T = \{a + b\,x : x \in S\}$.

6. Apply the change of variables theorem to show that Y has probability density function

$$g(y) = \frac{1}{|b|} f\left(\frac{y - a}{b}\right), \quad y \in T$$

When $b > 0$ (which is often the case in applications), this transformation is known as a **location-scale transformation**; a is the location parameter and b is the scale parameter. Location-scale transformations are studied in more detail in the chapter on Special Distributions.

The multivariate version of this result has a simple and elegant form when the linear transformation is expressed in matrix-vector form. Thus suppose that X is a random variable taking values in $S \subseteq \mathbb{R}^n$ and that X has a continuous distribution on S with probability density function f. Let $Y = a + B\,X$ where $a \in \mathbb{R}^n$ and B is an invertible $n \times n$ matrix. Note that and that Y takes values in $T = \{a + B\,x : x \in S\}$.

7. Show that

a. The transformation $y = a + B\,x$ maps $\mathbb{R}^n$ one-to-one and onto $\mathbb{R}^n$.
b. The inverse transformation is $x = B^{-1}(y - a)$.
c. The Jacobian of the inverse transformation is the constant function $\det\left(B^{-1}\right) = \frac{1}{\det(B)}$.

8. Apply the change of variables theorem to show that Y has probability density function

$$g(y) = \frac{f\left(B^{-1}(y - a)\right)}{|\det(B)|}, \quad y \in T$$

Sums and Convolution

Simple addition of random variables is perhaps the most important of all transformations. Suppose that X and Y are real-valued random variables for a random experiment. Our goal is to find the distribution of $Z = X + Y$.

9. Suppose that (X, Y) has a discrete distribution with probability density function f. Show that Z has a discrete distribution with probability density function u given by

$$u(z) = \sum_{x \in \mathbb{R}} f(x, z - x)$$

The sum is actually over the countable set $\{x \in \mathbb{R} : f(x, z - x) > 0\}$.

10. Suppose that (X, Y) has a continuous distribution with probability density function f. Show that the Z has a continuous distribution with probability density function u given by

$$u(z) = \int_{\mathbb{R}} f(x, z - x)\,dx$$

By far the most important special case occurs when X and Y are independent.

11. Suppose that X and Y are independent and have discrete distributions with densities g and h respectively. Show that Z has a probability density function

$$(g * h)(z) = \sum_{x \in \mathbb{R}} g(x)\,h(z - x)$$

The sum is actually over the countable set $\{x \in \mathbb{R} : (g(x) > 0) \text{ and } (h(z - x) > 0)\}$. The probability density function $g * h$ defined in the previous exercise is called the **discrete convolution** of g and h.

12. Suppose that X and Y are independent and have continuous distributions with densities g and h respectively. Show that Z has a probability density function given by

$$(g * h)(z) = \int_{\mathbb{R}} g(x)\, h(z - x) d x$$

The integral is actually over the set $\{x \in \mathbb{R} : (g(x) > 0) \text{ and } (h(z - x) > 0)\}$. The probability density function $g * h$ is called the **continuous convolution** of f and g.

13. Show that convolution (either discrete or continuous) satisfies the properties below (where f, g. and h are probability density functions). Give an analytic proof, based on the definition of convolution, and a probabilistic proof, based on sums of independent random variables

a. $f * g = g * f$ (the **commutative property**)
b. $f * (g * h) = (f * g) * h$ (the **associative property**

Thus, in part (b) we can write $f * g * h$ without ambiguity. Note that if $(X_1, X_2, \ldots, X_n)$ is a sequence of independent and identically distributed random variables, with common probability density function f. then

$$Y_n = X_1 + X_2 + \cdots + X_n$$

has probability density function f^{*n}, the n-fold **convolution power** of f. When appropriately scaled and centered, the distribution of Y_n converges to the standard normal distribution as $n \to \infty$. The precise statement of this result is the **central limit theorem**, one of the fundamental theorems of probability. The central limit theorem is studied in detail in the chapter on Random Samples.

Minimum and Maximum

Suppose that $(X_1, X_2, \ldots, X_n)$ is a sequence of independent real-valued random variables. The **minimum** and **maximum** transformations are very important in a number of applications.

$$U = \min\{X_1, X_2, \ldots, X_n\}$$
$$V = \max\{X_1, X_2, \ldots, X_n\}$$

For example, recall that in the standard model of **structural reliability**, a system consists of n components that operate independently. Suppose that X_i represents the lifetime of component i. Then U is the lifetime of the **series system** which operates if and only if each component is operating. Similarly, V is the lifetime of the **parallel system** which operates if and only if at least one component is operating.

A particularly important special case occurs when the random variables are **identically distributed**, in addition to being independent. In this case, the sequence of variables is a **random sample** of size n from the common distribution. We usually think of the random variables as independent copies of an underlying random variable. The minimum and maximum variables are the extreme examples of **order statistics**. Order statistics are studied in detail in the chapter on Random Samples.

Let F_i denote the distribution function of X_i for each $i \in \{1, 2, \ldots, n\}$, and let G and H denote the distribution functions of U and V respectively.

14. Show that for $x \in \mathbb{R}$,

a. $\{V \le x\} = \{X_1 \le x, X_2 \le x, \ldots, X_n \le x\}$
b. $H(x) = F_1(x) F_2(x) \cdots F_n(x)$

15. Show that for $x \in \mathbb{R}$,

a. $\{U > x\} = \{X_1 > x,\ X_2 > x,\ \ldots,\ X_n > x\}$
b. $G(x) = 1 - (1 - F_1(x))\,(1 - F_2(x))\ \cdots\ (1 - F_n(x))$

From Exercise 14, note that the product of n distribution functions is another distribution function. From Exercise 15, the product of n right-tail distribution functions is a right-tail distribution function. In the reliability setting, where the random variables are nonnegative, the product of n reliability functions is a reliability function. If X_i has a continuous distribution with probability density function f_i for each $i \in \{1, 2, \ldots, n\}$ then U and V also have continuous distributions, and their probability density functions can be obtained by differentiating the distribution functions in Exercises 14 and 15.

The formulas are particularly nice when the random variables X_i are identically distributed, in addition to being independent. Suppose that this is the case, and let F be the common distribution function.

16. Show that for $x \in \mathbb{R}$,

a. $H(x) = F^n(x)$
b. $G(x) = 1 - (1 - F(x))^n$

In particular, it follows that a positive integer power of a distribution function is a distribution. More generally, it's easy to see that every positive power of a distribution function is a distribution function. How could we construct a non-integer power of a distribution function in a probabilistic way? Now, in addition to the independent and identically distributed assumptions, suppose that the common distribution of the variables X_i is continuous, with density function f. Let g and h denote the density functions of U and V. respectively.

17. Show that for $x \in \mathbb{R}$,

a. $h(x) = n\,F^{n-1}\left(x\right)\,f(x)$
b. $g(x) = n\,(1 - F(x))^{n-1}\,f(x)$

Sign and Absolute Value

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

a. $|X|$ has distribution function $G(y) = F(y) - F(-y),\ \ y \ge 0$
b. $|X|$ has density function $g(y) = f(y) + f(-y),\ \ y \ge 0$

Let J denote the **sign** of X, defined by

$$J = \begin{cases} -1, & X < 0 \\ 0, & X = 0 \\ 1, & X > 0 \end{cases}$$

.

19. Suppose that the probability density f of X is symmetric with respect to 0, so that $f(x) = f(-x)$ for all $x \in \mathbb{R}$. Show that

a. $|X|$ has distribution function $G(y) = 2\,F(y) - 1,\ \ y \ge 0$
b. $|X|$ has density function $g(y) = 2\,f(y),\ \ y \ge 0.$

c. J is uniformly distributed on $\{-1, 1\}$
d. $|X|$ and J are independent.

Examples and Applications

This subsection contains computational exercises, many of which involve special parametric families of distributions. It is always interesting when a random variable from one parametric family can be transformed into a variable from another family. It is also interesting when a parametric family is **closed** or **invariant** under some transformation on the variables in the family. Often, such properties are what make the parametric families *special* in the first place. Please note these properties when they occur.

Dice

Recall that a **standard die** is an ordinary 6-sided die. A **fair** die is one in which the faces are equally likely. An **ace six flat die** is a standard die in which faces 1 and 6 occur with probability $\frac{1}{4}$ each and the other faces with probability $\frac{1}{8}$

20. Suppose that two standard dice are rolled and the sequence of scores (X_1, X_2) is recorded. Find the probability density function of $Y = X_1 + X_1$, the sum of the scores, in each of the following cases:

a. fair dice
b. ace-six flat dice

21. In the dice experiment, select two dice and select the sum random variable. Run the simulation 1000 times, updating every 10 runs and note the apparent convergence of the empirical density function to the probability density function for each of the following cases:

a. fair dice
b. ace-six flat dice

22. A fair die and an ace-six flat die are rolled. Find the probability density function of the sum of the scores.

23. Suppose that n standard, fair dice are rolled. Find the probability density function of the following variables:

a. the minimum score
b. the maximum score.

24. In the dice experiment, select fair dice and select each of the following random variables. Vary n with the scroll bar and note the shape of the density function. With $n = 4$, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the probability density function.

a. minimum score
b. maximum score.

The Uniform Distribution

25. Let $Y = X^2$. Find the density function of Y and sketch the graph in each of the following cases:

a. X is uniformly distributed on the interval $[-2, 2]$
b. X is uniformly distributed on the interval $[-1, 3]$
c. X is uniformly distributed on the interval $[2, 4]$.

Compare the distributions in the last exercise. Note that even a simple transformation of a simple distribution can produce a complicated distribution. On the other hand, the uniform distribution is preserved under a linear transformation of the random variable.

26. Suppose that X is uniformly distributed on $S \subseteq \mathbb{R}^n$. Let $Y = a + B\,X$, where $a \in \mathbb{R}^n$ and B is an invertible $n{\times}n$ matrix. Show that Y is uniformly distributed on $T = \left\{a + B\,x : x \in S\right\}$.

27. Suppose that (X, Y) is uniformly distributed on the square $[0, 1]^2$. Let $(U, V) = (X + Y, X - Y)$.

a. Sketch the range of (X, Y) and the range of (U, V)
b. Find the density function of (U, V)
c. Find the density function of U.
d. Find the density function of V.

28. Suppose that (X, Y, Z) is uniformly distributed on the cube $[0, 1]^3$. Find the probability density function of $(U, V, W) = (X + Y, Y + Z, X + Z)$.

29. Suppose that $(X_1, X_2, ..., X_n)$ is a sequence of independent random variables, each uniformly distributed on $[0, 1]$. Find the distribution and density functions of the following variables. Both distributions are beta distributions.

a. $U = \min\{X_1, X_2, ..., X_n\}$
b. $V = \max\{X_1, X_2, ..., X_n\}$

30. In the order statistic experiment, select the uniform distribution.

a. Set $k = 1$ (this gives the minimum U). Vary n with the scroll bar and note the shape of the density function. With $n = 5$, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the true density function.
b. Vary n with the scroll bar, set $k = n$ each time (this gives the maximum V), and note the shape of the density function. With $n = 5$ run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the true density function.

31. Let f denote the density function of the uniform distribution on $[0, 1]$. Compute f^{*2} and f^{*3}. Graph the three density functions on the same set of axes. Note the behavior predicted by the central limit theorem beginning to emerge.

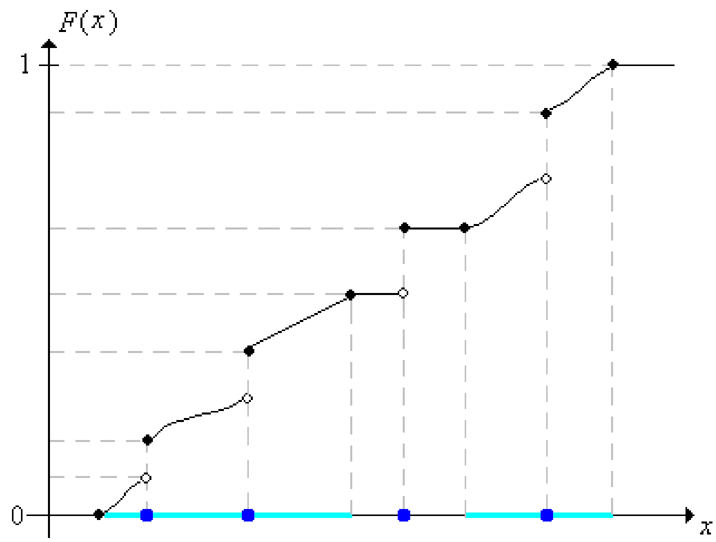
Simulations

A remarkable fact is that the uniform distribution on $(0, 1)$ can be transformed into almost any other distribution on $\mathbb{R}$. This is particularly important for simulations, since many computer languages have an algorithm for generating **random numbers**, which are simulations of independent variables, each uniformly distributed on $(0, 1)$. Conversely, any continuous distribution supported on an interval of $\mathbb{R}$ can be transformed into the uniform distribution on (0, 1).

Suppose first that F is a distribution function for a distribution on $\mathbb{R}$ (which may be discrete, continuous, or mixed), and let F^{-1} denote the quantile function.

32. Suppose that U is uniformly distributed on (0, 1). Show that $X = F^{-1}(U)$ has distribution function F.

Assuming that we can compute F^{-1}, the previous exercise shows how we can simulate a distribution with distribution function F. To rephrase the result, we can simulate a variable with distribution function F by simply computing a *random quantile*. Most of the applets in this project use this method of simulation.



33. Suppose that X has a continuous distribution on an interval $S \subseteq \mathbb{R}$ and that the distribution function F is strictly increasing on S. Show that $U = F(X)$ has the uniform distribution on (0, 1).

34. Show how to simulate the uniform distribution on the interval (a, b) with a random number. Using your calculator, simulate 5 values from the uniform distribution on the interval (2, 10).

Bernoulli Trials

Recall that a **Bernoulli trials sequence** is a sequence $(X_1, X_2, \ldots)$ of independent, identically distributed indicator random variables. In the usual terminology of reliability theory, $X_i = 0$ means **failure** on trial i. while $X_i = 1$ means **success** on trial i. The basic parameter of the process is the probability of success $p = \mathbb{P}(X_i = 1)$. The random process is named for Jacob Bernoulli and is studied in detail in the chapter on Bernoulli trials.

35. Show that the common probability density function of the trial variables X_i is $f(k) = p^k (1 - p)^{1-k}, \quad k \in \{0, 1\}$.

36. Let Y_n denote the number of successes in the first n trials. Use an argument based on combinatorics and independence to show that Y_n has the probability density function f_n given below. This defines the **binomial distribution** with parameters n and p.

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

37. As above, let Y_n denote the number of successes in the first n trials.

a. Show that $Y_n = X_1 + X_2 + \cdots + X_n$
b. Conclude from (a) that $f_n = f^{*n}$
c. Conclude from (b) that $f_m * f_n = f_{m+n}$
d. Prove the result in (c) directly, using the definition of convolution.

In particular, it follows that if Y and Z are independent variables, and that Y has the binomial distribution with parameters n and p while Z has the binomial distribution with parameter m and p. Show that $Y + Z$ has the binomial distribution with parameter $n + m$ and p.

38. Find the probability density function of the difference between the number of successes and the number of failures in n Bernoulli trials.

The Poisson Distribution

Recall that the **Poisson distribution** with parameter $t > 0$ has probability density function

$$f_t(n) = e^{-t}\,\frac{t^n}{n!},\ \ n \in \mathbb{N}$$

This distribution is named for Simeon Poisson and is widely used to model the number of random points in a region of time or space. The Poisson distribution is studied in detail in the chapter on The Poisson Process.

39. Suppose that X and Y are independent variables, and that X has the Poisson distribution with parameter $a > 0$ while Y has the Poisson distribution with parameter $b > 0$. Show that $X + Y$ has the Poisson distribution with parameter $a + b$. Equivalently, show that $f_a * f_b = f_{a+b}$. *Hint*: You will need to use the binomial theorem.

The Exponential Distribution

Recall that the **exponential distribution** with rate parameter $r > 0$ has probability density function $f(t) = r\,e^{-r\,t},\ \ t \ge 0$. These distributions are often used to model random times such as failure times and arrival times. Exponential distributions are studied in more detail in the chapter on Poisson Processes.

40. Show how to simulate, with a random number, the exponential distribution with rate parameter r. Using your calculator, simulate 5 values from the exponential distribution with parameter $r = 3$.

41. Suppose that T has the exponential distribution with rate parameter r. Find the density function of the following random variables:

a. $\lfloor T \rfloor$, the largest integer less than or equal to T.
b. $\lceil T \rceil$, the smallest integer greater than or equal to T.

Note that the distributions in the previous exercise are geometric distributions on $\mathbb{N}$ and on $\mathbb{N}_+$, respectively. In many respects, the geometric distribution is a discrete version of the exponential distribution.

42. Suppose that X and Y are independent random variables, each having the exponential distribution with parameter 1. Let $Z = \frac{Y}{X}$.

a. Find the distribution function of Z.
b. Find the density function of Z.

43. Suppose that X has the exponential distribution with rate parameter $a > 0$, Y has the exponential distribution with rate parameter $b > 0$, and that X and Y are independent. Find the probability density function of $Z = X + Y$ and sketch the graph.

44. Suppose that $(T_1, T_2, \ldots, T_n)$ is a sequence of independent random variables, and that T_i has the exponential distribution with rate parameter $r_i > 0$ for each $i \in \{1, 2, \ldots, n\}$.

a. Find the distribution function of $U = \min\{T_1, T_2, \ldots, T_n\}$.
b. Find the distribution function of $V = \max\{T_1, T_2, \ldots, T_n\}$.
c. Find the density function and reliability function of U and V in the special case that $r_i = r$ for each $i \in \{1, 2, \ldots, n\}$

Note that the minimum U in part (a) has the exponential distribution with parameter $r_1 + r_2 + \cdots + r_n$. In particular, suppose that a series system has independent components, each with an exponentially distributed lifetime. Then the lifetime of the system is also exponentially distributed, and the failure rate of the system is the sum of the component failure rates.

45. In the order statistic experiment, select the exponential distribution.

a. Set $k = 1$ (this gives the minimum U). Vary n with the scroll bar and note the shape of the density function. With $n = 5$, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the true density function.
b. Vary n with the scroll bar, set $k = n$ each time (this gives the maximum V), and note the shape of the density function. With $n = 5$, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the true density function.

46. In the setting of Exercise 44, show that for $i \in \{1, 2, \ldots, n\}$,

$$\mathbb{P}\left(T_i < T_j \quad \text{for all} \quad j \neq i\right) = \frac{r_i}{\sum_{j=1}^{n} r_j}$$

a. Show first that if X and Y have exponential distributions with parameters a and b. respectively, and are independent, then $\mathbb{P}(X < Y) = \frac{a}{a+b}$
b. Next, note that $T_i < T_j$ for all $j \neq i$ if and only if $T_i < \min\left\{T_j : j \neq i\right\}$.
c. Note that he minimum on the right is independent of T_i and, by Exercise 44, has an exponential distribution with parameter $\sum_{j \neq i} r_j$

The result in the previous exercise is very important in the theory of continuous-time Markov chains.

The Gamma Distribution

Recall that for $n \in \{1, 2, \ldots\}$, the **gamma distribution** with shape parameter n has probability density function

$$g_n(t) = e^{-t} \frac{t^{n-1}}{(n-1)!}, \quad t \geq 0$$

This distribution is widely used to model arrival times under certain basic assumptions.. The gamma distribution is studied in detail in the chapter on The Poisson Process.

47. Show that if X has the gamma distribution with shape parameter m and Y has the gamma distribution with shape parameter n. and if X and Y are independent, then $X + Y$ has the gamma distribution with shape parameter $m + n$. Equivalently, show that $g_m * g_n = g_{m+n}$.

48. Suppose that T has the gamma distribution with shape parameter n. Find the density function of $X = \ln(T)$.

The Pareto Distribution

Recall that the **Pareto distribution** with shape parameter $a > 0$ has probability density function

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

The distribution is named for Vilfredo Pareto. It is a heavy-tailed distribution often used for modeling income and other financial variables. The Pareto distribution is studied in more detail in the chapter on Special Distributions.

49. Suppose that X has the Pareto distribution with shape parameter a. Find the density function of $Y = \ln(X)$. Note that Y has the exponential distribution with rate parameter a.

50. Suppose that X has the Pareto distribution with shape parameter a. Find the probability density function of $Y = \frac{1}{X}$. The distribution of Y is the beta distribution with parameters a and $b = 1$.

51. Show how to simulate, with a random number, the Pareto distribution with shape parameter a. Using your calculator, simulate 5 values from the Pareto distribution with shape parameter $a = 2$.

The Normal Distribution

Recall that the **standard normal distribution** has probability density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}, \quad z \in \mathbb{R}$$

52. Suppose that Z has the standard normal distribution, and that $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$. Find the density function of $X = \mu + \sigma Z$ and sketch the graph.

Random variable X has the **normal distribution** with **location parameter** μ and **scale parameter** σ. The normal distribution is perhaps the most important distribution in probability and mathematical statistics. It is widely used to model physical measurements of all types that are subject to small, random errors. The normal distribution is studied in detail in the chapter on Special Distributions.

53. Suppose that Z has the standard normal distribution. Find the density function of $V = Z^2$ and sketch the graph.

Random variable V has the **chi-square distribution** with 1 **degree of freedom**. Chi-square distributions are studied in detail in the chapter on Special Distributions.

54. Suppose that X has the normal distribution with location parameter μ and scale parameter σ, Y has the normal distribution with location parameter ν and scale parameter τ. and that X and Y are independent. Show that $X + Y$ has the normal distribution with location parameter $\mu + \nu$ and scale parameter $\sqrt{\sigma^2 + \tau^2}$

The Cauchy Distribution

55. Suppose that X and Y are independent random variables, each with the standard normal distribution. Show that $T = \frac{X}{Y}$ has probability density function f given below, and sketch the graph.

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

Random variable T has the **Cauchy distribution**, named after Augustin Cauchy; it is a member of the family of **student** t **distributions**. The student t distributions are studied in detail in the chapter on Special Distributions.

56. For $c > 0$, show that the probability density function of c T is

$$f_c(t) = \frac{c}{\pi\left(c^2 + t^2\right)}, \quad t \in \mathbb{R}$$

Of course, this is the Cauchy distribution with scale parameter c.

57. Show that if U has the Cauchy distribution with scale parameter c, V has the Cauchy distribution with scale parameter d. and U and V are independent, then $U + V$ has the Cauchy distribution with scale parameter $c + d$. Equivalently, show that $f_c * f_d = f_{c+d}$

The Beta Distribution

58. Suppose that X has the probability density function $f(x) = 3\,x^2,\ 0 \le x \le 1$. Find the probability density function of $Y = \sqrt[3]{X}$ and sketch its graph.

Random variables X and Y both have **beta distributions**, which are widely used to model random proportions and probabilities. The family of beta distributions is studied in detail in the chapter on Special Distributions.