

3.7: Transformations of Random Variables

This section studies how the distribution of a random variable changes when the variable is transformed in a deterministic way. If you are a new student of probability, you should skip the technical details.

Basic Theory

The Problem

As usual, we start with a random experiment modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So to review, Ω is the set of outcomes, \mathcal{F} is the collection of events, and \mathbb{P} is the probability measure on the sample space (Ω, \mathcal{F}) . Suppose now that we have a random variable X for the experiment, taking values in a set S , and a function r from S into another set 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. But first recall that for $B \subseteq T$, $r^{-1}(B) = \{x \in S : r(x) \in B\}$ is the inverse image of B under r .

$$\mathbb{P}(Y \in B) = \mathbb{P}[X \in r^{-1}(B)] \quad \text{for } B \subseteq T.$$

Proof

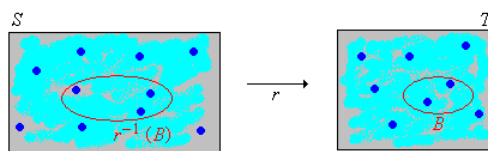


Figure 3.7.1: A function $r : S \rightarrow T$. How is a probability distribution on S transformed by r to a distribution on T ?

However, frequently the distribution of X is known either through its distribution function F or its probability density function f , and we would similarly like to find the distribution function or probability 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

When the transformed variable Y has a discrete distribution, the probability density function of Y can be computed using basic rules of probability.

Suppose that X has a discrete distribution on a countable set S , with probability density function f . Then Y has a discrete distribution with probability density function g given by

$$g(y) = \sum_{x \in r^{-1}\{y\}} f(x), \quad y \in T \quad (3.7.1)$$

Proof

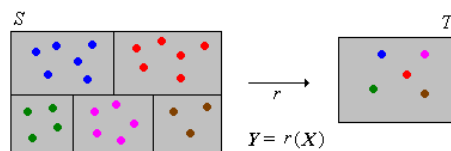


Figure 3.7.2: A transformation of a discrete probability distribution.

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. Then Y has a discrete distribution with probability density function g given by

$$g(y) = \int_{r^{-1}\{y\}} f(x) dx, \quad y \in T \quad (3.7.2)$$

Proof

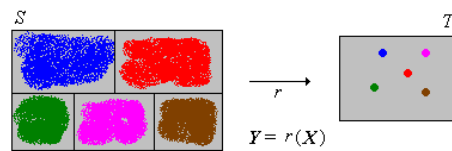


Figure 3.7.3: A continuous distribution on S transformed by a discrete function $r : S \rightarrow T$

So the main problem is often computing the inverse images $r^{-1}\{y\}$ for $y \in T$. The formulas above in the [discrete](#) and [continuous](#) cases are not worth memorizing explicitly; it's usually better to just work each problem from scratch. The main step is to write the event $\{Y = y\}$ in terms of X , and then find the probability of this event using the probability density function of X .

Transformed Variables with Continuous Distributions

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

Suppose that Y is real valued. The distribution function G of Y is given by

$$G(y) = \int_{r^{-1}(-\infty, y]} f(x) dx, \quad y \in \mathbb{R} \quad (3.7.3)$$

Proof

Again, this follows from the definition of f as a PDF of X . For $y \in \mathbb{R}$,

$$G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}[r(X) \in (-\infty, y]] = \mathbb{P}[X \in r^{-1}(-\infty, y]] = \int_{r^{-1}(-\infty, y]} f(x) dx \quad (3.7.4)$$

As in the discrete case, the formula in (4) not much help, and it's usually better to work each problem from scratch. The main step is to write the event $\{Y \leq y\}$ in terms of X , and then find the probability of this event using the probability density function of X .

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. Note that since r is one-to-one, it has an inverse function r^{-1} .

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 probability density function f . Suppose that $Y = r(X)$ where r is a differentiable function from S onto an interval T . As usual, we will let G denote the distribution function of Y and g the probability density function of Y .

Suppose that r is strictly increasing on S . For $y \in T$,

1. $G(y) = F[r^{-1}(y)]$
2. $g(y) = f[r^{-1}(y)] \frac{d}{dy} r^{-1}(y)$

Proof

1. $G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}[r(X) \leq y] = \mathbb{P}[X \leq r^{-1}(y)] = F[r^{-1}(y)]$ for $y \in T$. Note that the inequality is preserved since r is increasing.
2. This follows from part (a) by taking derivatives with respect to y and using the chain rule. Recall that $F' = f$.

Suppose that r is strictly decreasing on S . For $y \in T$,

1. $G(y) = 1 - F[r^{-1}(y)]$
2. $g(y) = -f[r^{-1}(y)] \frac{d}{dy} r^{-1}(y)$

Proof

1. $G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}[r(X) \leq y] = \mathbb{P}[X \geq r^{-1}(y)] = 1 - F[r^{-1}(y)]$ for $y \in T$. Note that the inequality is reversed since r is decreasing.
2. This follows from part (a) by taking derivatives with respect to y and using the chain rule. Recall again that $F' = f$.

The formulas for the probability density functions in the [increasing case](#) and the [decreasing case](#) can be combined:

If r is strictly increasing or strictly decreasing on S then the probability density function g of Y is given by

$$g(y) = f[r^{-1}(y)] \left| \frac{d}{dy} r^{-1}(y) \right| \quad (3.7.5)$$

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

$$g(y) = f(x) \left| \frac{dx}{dy} \right| \quad (3.7.6)$$

Although succinct and easy to remember, the formula is a bit less clear. It must be understood that x on the right should be written in terms of y via the inverse function. The images below give a graphical interpretation of the formula in the two cases where r is increasing and where r is decreasing.

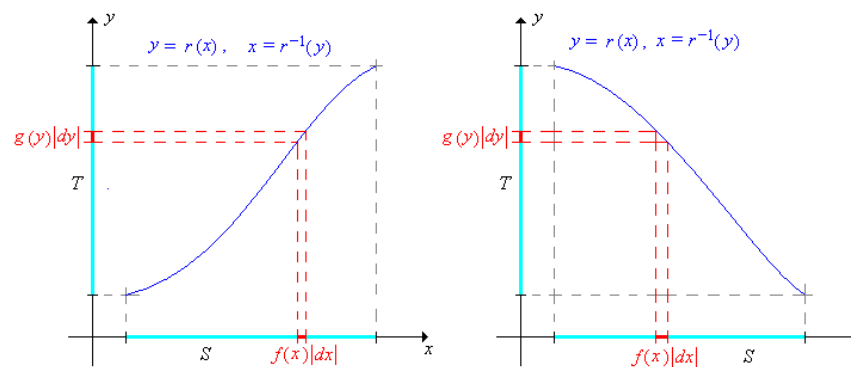


Figure 3.7.4: The change of variables theorems in the increasing and decreasing cases

The generalization of this result from \mathbb{R} to \mathbb{R}^n is basically a theorem in multivariate calculus. First we need some notation. Suppose that r is a one-to-one differentiable function from $S \subseteq \mathbb{R}^n$ onto $T \subseteq \mathbb{R}^n$. The *first derivative* of the inverse function $\mathbf{x} = r^{-1}(\mathbf{y})$ is the $n \times n$ matrix of first partial derivatives:

$$\left(\frac{d\mathbf{x}}{d\mathbf{y}} \right)_{ij} = \frac{\partial x_i}{\partial y_j} \quad (3.7.7)$$

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

$$\det \left(\frac{d\mathbf{x}}{d\mathbf{y}} \right) \quad (3.7.8)$$

With this compact notation, the multivariate change of variables formula is easy to state.

Suppose that \mathbf{X} is a random variable taking values in $S \subseteq \mathbb{R}^n$, and that \mathbf{X} has a continuous distribution with probability density function f . Suppose also $\mathbf{Y} = r(\mathbf{X})$ where r is a differentiable function from S onto $T \subseteq \mathbb{R}^n$. Then the probability density function g of \mathbf{Y} is given by

$$g(\mathbf{y}) = f(\mathbf{x}) \left| \det \left(\frac{d\mathbf{x}}{d\mathbf{y}} \right) \right|, \quad \mathbf{y} \in T \quad (3.7.9)$$

Proof

The result follows from the multivariate change of variables formula in calculus. If $B \subseteq T$ then

$$\mathbb{P}(\mathbf{Y} \in B) = \mathbb{P}[r(\mathbf{X}) \in B] = \mathbb{P}[\mathbf{X} \in r^{-1}(B)] = \int_{r^{-1}(B)} f(\mathbf{x}) d\mathbf{x} \quad (3.7.10)$$

Using the change of variables $\mathbf{x} = r^{-1}(\mathbf{y})$, $d\mathbf{x} = \left| \det \left(\frac{d\mathbf{x}}{d\mathbf{y}} \right) \right| d\mathbf{y}$ we have

$$\mathbb{P}(\mathbf{Y} \in B) = \int_B f[r^{-1}(\mathbf{y})] \left| \det \left(\frac{d\mathbf{x}}{d\mathbf{y}} \right) \right| d\mathbf{y} \quad (3.7.11)$$

So it follows that g defined in the theorem is a PDF for \mathbf{Y} .

The Jacobian is the infinitesimal scale factor that describes how n -dimensional volume changes under the transformation.

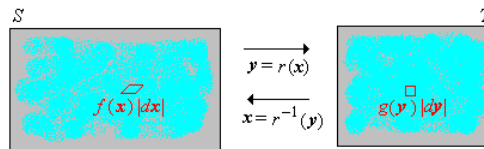


Figure 3.7.5: The multivariate change of variables theorem

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 simple applications of the change of variable theorems. Suppose first that X is a random variable taking values in an interval $S \subseteq \mathbb{R}$ and that X has a continuous distribution on S with probability density function f . Let $Y = a + bX$ where $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$. Note that Y takes values in $T = \{y = a + bx : x \in S\}$, which is also an interval.

Y has probability density function g given by

$$g(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right), \quad y \in T \quad (3.7.12)$$

Proof

The transformation is $y = a + bx$. Hence the inverse transformation is $x = (y - a)/b$ and $dx/dy = 1/b$. The result now follows from the [change of variables theorem](#).

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*. Scale transformations arise naturally when physical units are changed (from feet to meters, for example). Location transformations arise naturally when the physical reference point is changed (measuring time relative to 9:00 AM as opposed to 8:00 AM, for example). The change of temperature measurement from Fahrenheit to Celsius is a location and scale transformation. 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 \mathbf{X} is a random variable taking values in $S \subseteq \mathbb{R}^n$ and that \mathbf{X} has a continuous distribution on S with probability density function f . Let $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$ where $\mathbf{a} \in \mathbb{R}^n$ and \mathbf{B} is an invertible $n \times n$ matrix. Note that \mathbf{Y} takes values in $T = \{\mathbf{a} + \mathbf{B}\mathbf{x} : \mathbf{x} \in S\} \subseteq \mathbb{R}^n$.

\mathbf{Y} has probability density function g given by

$$g(\mathbf{y}) = \frac{1}{|\det(\mathbf{B})|} f[\mathbf{B}^{-1}(\mathbf{y} - \mathbf{a})], \quad \mathbf{y} \in T \quad (3.7.13)$$

Proof

The transformation $\mathbf{y} = \mathbf{a} + \mathbf{B}\mathbf{x}$ maps \mathbb{R}^n one-to-one and onto \mathbb{R}^n . The inverse transformation is $\mathbf{x} = \mathbf{B}^{-1}(\mathbf{y} - \mathbf{a})$. The Jacobian of the inverse transformation is the constant function $\det(\mathbf{B}^{-1}) = 1/\det(\mathbf{B})$. The result now follows from the [multivariate change of variables theorem](#).

Sums and Convolution

Simple addition of random variables is perhaps the most important of all transformations. Suppose that X and Y are random variables on a probability space, taking values in $R \subseteq \mathbb{R}$ and $S \subseteq \mathbb{R}$, respectively, so that (X, Y) takes values in a subset of $R \times S$. Our goal is to find the distribution of $Z = X + Y$. Note that Z takes values in $T = \{z \in \mathbb{R} : z = x + y \text{ for some } x \in R, y \in S\}$. For $z \in T$, let $D_z = \{x \in R : z - x \in S\}$.

Suppose that (X, Y) probability density function f .

1. If (X, Y) has a discrete distribution then $Z = X + Y$ has a discrete distribution with probability density function u given by

$$u(z) = \sum_{x \in D_z} f(x, z - x), \quad z \in T \quad (3.7.14)$$

2. If (X, Y) has a continuous distribution then $Z = X + Y$ has a continuous distribution with probability density function u given by

$$u(z) = \int_{D_z} f(x, z - x) dx, \quad z \in T \quad (3.7.15)$$

Proof

1. $\mathbb{P}(Z = z) = \mathbb{P}(X = x, Y = z - x \text{ for some } x \in D_z) = \sum_{x \in D_z} f(x, z - x)$
2. For $A \subseteq T$, let $C = \{(u, v) \in R \times S : u + v \in A\}$. Then

$$\mathbb{P}(Z \in A) = \mathbb{P}(X + Y \in A) = \int_C f(u, v) d(u, v) \quad (3.7.16)$$

Now use the change of variables $x = u$, $z = u + v$. Then the inverse transformation is $u = x$, $v = z - x$ and the Jacobian is 1. Using the change of variables theorem (8) we have

$$\mathbb{P}(Z \in A) = \int_{D_z \times A} f(x, z - x) d(x, z) = \int_A \int_{D_z} f(x, z - x) dx dz \quad (3.7.17)$$

It follows that Z has probability density function $z \mapsto \int_{D_z} f(x, z - x) dx$.

In the discrete case, R and S are countable, so T is also countable as is D_z for each $z \in T$. In the continuous case, R and S are typically intervals, so T is also an interval as is D_z for $z \in T$. In both cases, determining D_z is often the most difficult step. By far the most important special case occurs when X and Y are independent.

Suppose that X and Y are independent and have probability density functions g and h respectively.

1. If X and Y have discrete distributions then $Z = X + Y$ has a discrete distribution with probability density function $g * h$ given by

$$(g * h)(z) = \sum_{x \in D_z} g(x)h(z - x), \quad z \in T \quad (3.7.18)$$

2. If X and Y have continuous distributions then $Z = X + Y$ has a continuous distribution with probability density function $g * h$ given by

$$(g * h)(z) = \int_{D_z} g(x)h(z - x) dx, \quad z \in T \quad (3.7.19)$$

In both cases, the probability density function $g * h$ is called the *convolution* of g and h .

Proof

Both results follows from the previous result above since $f(x, y) = g(x)h(y)$ is the probability density function of (X, Y) .

As before, determining this set D_z is often the most challenging step in finding the probability density function of Z . However, there is one case where the computations simplify significantly.

Suppose again that X and Y are independent random variables with probability density functions g and h , respectively.

1. In the discrete case, suppose X and Y take values in \mathbb{N} . Then Z has probability density function

$$(g * h)(z) = \sum_{x=0}^z g(x)h(z-x), \quad z \in \mathbb{N} \quad (3.7.20)$$

2. In the continuous case, suppose that X and Y take values in $[0, \infty)$. Then Z has probability density function

$$(g * h)(z) = \int_0^z g(x)h(z-x) dx, \quad z \in [0, \infty) \quad (3.7.21)$$

Proof

1. In this case, $D_z = \{0, 1, \dots, z\}$ for $z \in \mathbb{N}$.
2. In this case, $D_z = [0, z]$ for $z \in [0, \infty)$.

Convolution is a very important mathematical operation that occurs in areas of mathematics outside of probability, and so involving functions that are not necessarily probability density functions. The following result gives some simple properties of convolution.

Convolution (either discrete or continuous) satisfies the following properties, where f , g , and h are probability density functions of the same type.

1. $f * g = g * f$ (the *commutative property*)
2. $(f * g) * h = f * (g * h)$ (the *associative property*)

Proof

An analytic proof is possible, based on the definition of convolution, but a probabilistic proof, based on sums of independent random variables is much better. Thus, suppose that X , Y , and Z are independent random variables with PDFs f , g , and h , respectively.

1. The commutative property of convolution follows from the commutative property of addition: $X + Y = Y + X$.
2. The associative property of convolution follows from the associative property of addition: $(X + Y) + Z = X + (Y + Z)$.

Thus, in part (b) we can write $f * g * h$ without ambiguity. Of course, the constant 0 is the additive identity so $X + 0 = 0 + X = X$ for every random variable X . Also, a constant is independent of every other random variable. It follows that the probability density function δ of 0 (given by $\delta(0) = 1$) is the identity with respect to convolution (at least for discrete PDFs). That is, $f * \delta = \delta * f = f$. The next result is a simple corollary of the [convolution theorem](#), but is important enough to be highlighted.

Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of independent and identically distributed real-valued random variables, with common probability density function f . Then $Y_n = X_1 + X_2 + \dots + X_n$ has probability density function $f^{*n} = f * f * \dots * f$, the n -fold *convolution power* of f , for $n \in \mathbb{N}$.

In statistical terms, \mathbf{X} corresponds to *sampling* from the common distribution. By convention, $Y_0 = 0$, so naturally we take $f^{*0} = \delta$. When appropriately scaled and centered, the distribution of Y_n converges to the standard normal distribution as $n \rightarrow \infty$. 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. Clearly convolution power satisfies the *law of exponents*: $f^{*n} * f^{*m} = f^{*(n+m)}$ for $m, n \in \mathbb{N}$.

Convolution can be generalized to sums of independent variables that are not of the same type, but this generalization is usually done in terms of distribution functions rather than probability density functions.

Products and Quotients

While not as important as sums, products and quotients of real-valued random variables also occur frequently. We will limit our discussion to continuous distributions.

Suppose that (X, Y) has a continuous distribution on \mathbb{R}^2 with probability density function f .

1. Random variable $V = XY$ has probability density function

$$v \mapsto \int_{-\infty}^{\infty} f(x, v/x) \frac{1}{|x|} dx \quad (3.7.22)$$

2. Random variable $W = Y/X$ has probability density function

$$w \mapsto \int_{-\infty}^{\infty} f(x, wx) |x| dx \quad (3.7.23)$$

Proof

We introduce the auxiliary variable $U = X$ so that we have bivariate transformations and can use our [change of variables formula](#).

1. We have the transformation $u = x, v = xy$ and so the inverse transformation is $x = u, y = v/u$. Hence

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{bmatrix} 1 & 0 \\ -v/u^2 & 1/u \end{bmatrix} \quad (3.7.24)$$

and so the Jacobian is $1/u$. Using the change of variables theorem, the joint PDF of (U, V) is $(u, v) \mapsto f(u, v/u) |1/u|$. Hence the PDF of V is

$$v \mapsto \int_{-\infty}^{\infty} f(u, v/u) \frac{1}{|u|} du \quad (3.7.25)$$

2. We have the transformation $u = x, w = y/x$ and so the inverse transformation is $x = u, y = uw$. Hence

$$\frac{\partial(x, y)}{\partial(u, w)} = \begin{bmatrix} 1 & 0 \\ w & u \end{bmatrix} \quad (3.7.26)$$

and so the Jacobian is u . Using the change of variables formula, the joint PDF of (U, W) is $(u, w) \mapsto f(u, uw) |u|$. Hence the PDF of W is

$$w \mapsto \int_{-\infty}^{\infty} f(u, uw) |u| du \quad (3.7.27)$$

If (X, Y) takes values in a subset $D \subseteq \mathbb{R}^2$, then for a given $v \in \mathbb{R}$, the integral in (a) is over $\{x \in \mathbb{R} : (x, v/x) \in D\}$, and for a given $w \in \mathbb{R}$, the integral in (b) is over $\{x \in \mathbb{R} : (x, wx) \in D\}$. As usual, the most important special case of this result is when X and Y are independent.

Suppose that X and Y are independent random variables with continuous distributions on \mathbb{R} having probability density functions g and h , respectively.

1. Random variable $V = XY$ has probability density function

$$v \mapsto \int_{-\infty}^{\infty} g(x) h(v/x) \frac{1}{|x|} dx \quad (3.7.28)$$

2. Random variable $W = Y/X$ has probability density function

$$w \mapsto \int_{-\infty}^{\infty} g(x) h(wx) |x| dx \quad (3.7.29)$$

Proof

These results follow immediately from the previous theorem, since $f(x, y) = g(x)h(y)$ for $(x, y) \in \mathbb{R}^2$.

If X takes values in $S \subseteq \mathbb{R}$ and Y takes values in $T \subseteq \mathbb{R}$, then for a given $v \in \mathbb{R}$, the integral in (a) is over $\{x \in S : v/x \in T\}$, and for a given $w \in \mathbb{R}$, the integral in (b) is over $\{x \in S : wx \in T\}$. As with convolution, determining the domain of integration is often the most challenging step.

Minimum and Maximum

Suppose that (X_1, X_2, \dots, X_n) is a sequence of independent real-valued random variables. The *minimum* and *maximum* transformations

$$U = \min\{X_1, X_2, \dots, X_n\}, \quad V = \max\{X_1, X_2, \dots, X_n\} \quad (3.7.30)$$

are very important in a number of applications. 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 $i \in \{1, 2, \dots, n\}$. 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. The minimum and maximum variables are the extreme examples of *order statistics*. Order statistics are studied in detail in the chapter on Random Samples.

Suppose that (X_1, X_2, \dots, X_n) is a sequence of independent real-valued random variables and that X_i has distribution function F_i for $i \in \{1, 2, \dots, n\}$.

1. $V = \max\{X_1, X_2, \dots, X_n\}$ has distribution function H given by $H(x) = F_1(x)F_2(x) \cdots F_n(x)$ for $x \in \mathbb{R}$.
2. $U = \min\{X_1, X_2, \dots, X_n\}$ has distribution function G given by $G(x) = 1 - [1 - F_1(x)][1 - F_2(x)] \cdots [1 - F_n(x)]$ for $x \in \mathbb{R}$.

Proof

1. Note that since V is the maximum of the variables, $\{V \leq x\} = \{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\}$. Hence by independence,

$$H(x) = \mathbb{P}(V \leq x) = \mathbb{P}(X_1 \leq x)\mathbb{P}(X_2 \leq x) \cdots \mathbb{P}(X_n \leq x) = F_1(x)F_2(x) \cdots F_n(x), \quad x \in \mathbb{R} \quad (3.7.31)$$

2. Note that since U is the minimum of the variables, $\{U > x\} = \{X_1 > x, X_2 > x, \dots, X_n > x\}$. Hence by independence,

$$\begin{aligned} G(x) &= \mathbb{P}(U \leq x) = 1 - \mathbb{P}(U > x) = 1 - \mathbb{P}(X_1 > x)\mathbb{P}(X_2 > x) \cdots \mathbb{P}(X_n > x) \\ &= 1 - [1 - F_1(x)][1 - F_2(x)] \cdots [1 - F_n(x)], \quad x \in \mathbb{R} \end{aligned}$$

From part (a), note that the product of n distribution functions is another distribution function. From part (b), 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 last statement means that the product of n reliability functions is another reliability function. If X_i has a continuous distribution with probability density function f_i for each $i \in \{1, 2, \dots, n\}$, then U and V also have continuous distributions, and their probability density functions can be obtained by differentiating the distribution functions in parts (a) and (b) of [last theorem](#). The computations are straightforward using the product rule for derivatives, but the results are a bit of a mess.

The formulas in [last theorem](#) are particularly nice when the random variables are identically distributed, in addition to being independent

Suppose that (X_1, X_2, \dots, X_n) is a sequence of independent real-valued random variables, with common distribution function F .

1. $V = \max\{X_1, X_2, \dots, X_n\}$ has distribution function H given by $H(x) = F^n(x)$ for $x \in \mathbb{R}$.
2. $U = \min\{X_1, X_2, \dots, X_n\}$ has distribution function G given by $G(x) = 1 - [1 - F(x)]^n$ for $x \in \mathbb{R}$.

In particular, it follows that a positive integer power of a distribution function is a distribution function. 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?

Suppose that (X_1, X_2, \dots, X_n) is a sequence of independent real-valued random variables, with a common continuous distribution that has probability density function f .

1. $V = \max\{X_1, X_2, \dots, X_n\}$ has probability density function h given by $h(x) = nF^{n-1}(x)f(x)$ for $x \in \mathbb{R}$.
2. $U = \min\{X_1, X_2, \dots, X_n\}$ has probability density function g given by $g(x) = n[1 - F(x)]^{n-1}f(x)$ for $x \in \mathbb{R}$.

Coordinate Systems

For our next discussion, we will consider transformations that correspond to common distance-angle based coordinate systems—polar coordinates in the plane, and cylindrical and spherical coordinates in 3-dimensional space. First, for $(x, y) \in \mathbb{R}^2$, let (r, θ)

denote the standard *polar coordinates* corresponding to the Cartesian coordinates (x, y) , so that $r \in [0, \infty)$ is the *radial distance* and $\theta \in [0, 2\pi)$ is the *polar angle*.

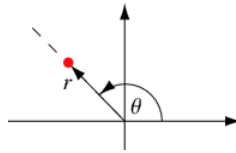


Figure 3.7.6: Polar coordinates. Stover, Christopher and Weisstein, Eric W. "Polar Coordinates." From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/PolarCoordinates.html>

It's best to give the *inverse transformation*: $x = r \cos \theta$, $y = r \sin \theta$. As we all know from calculus, the Jacobian of the transformation is r . Hence the following result is an immediate consequence of our [change of variables theorem](#):

Suppose that (X, Y) has a continuous distribution on \mathbb{R}^2 with probability density function f , and that (R, Θ) are the polar coordinates of (X, Y) . Then (R, Θ) has probability density function g given by

$$g(r, \theta) = f(r \cos \theta, r \sin \theta)r, \quad (r, \theta) \in [0, \infty) \times [0, 2\pi) \quad (3.7.32)$$

Next, for $(x, y, z) \in \mathbb{R}^3$, let (r, θ, z) denote the standard *cylindrical coordinates*, so that (r, θ) are the standard polar coordinates of (x, y) as above, and coordinate z is left unchanged. Given our previous result, the one for cylindrical coordinates should come as no surprise.

Suppose that (X, Y, Z) has a continuous distribution on \mathbb{R}^3 with probability density function f , and that (R, Θ, Z) are the cylindrical coordinates of (X, Y, Z) . Then (R, Θ, Z) has probability density function g given by

$$g(r, \theta, z) = f(r \cos \theta, r \sin \theta, z)r, \quad (r, \theta, z) \in [0, \infty) \times [0, 2\pi) \times \mathbb{R} \quad (3.7.33)$$

Finally, for $(x, y, z) \in \mathbb{R}^3$, let (r, θ, ϕ) denote the standard *spherical coordinates* corresponding to the Cartesian coordinates (x, y, z) , so that $r \in [0, \infty)$ is the *radial distance*, $\theta \in [0, 2\pi)$ is the *azimuth angle*, and $\phi \in [0, \pi]$ is the *polar angle*. (In spite of our use of the word *standard*, different notations and conventions are used in different subjects.)

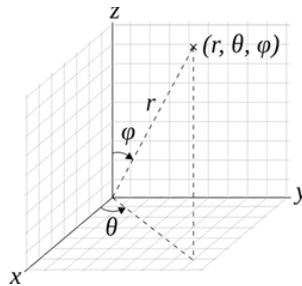


Figure 3.7.7: Spherical coordinates, By Dmcq—Own work, CC BY-SA 3.0, [Wikipedia](#)

Once again, it's best to give the *inverse transformation*: $x = r \sin \phi \cos \theta$, $y = r \sin \phi \sin \theta$, $z = r \cos \phi$. As we remember from calculus, the absolute value of the Jacobian is $r^2 \sin \phi$. Hence the following result is an immediate consequence of the change of variables theorem (8):

Suppose that (X, Y, Z) has a continuous distribution on \mathbb{R}^3 with probability density function f , and that (R, Θ, Φ) are the spherical coordinates of (X, Y, Z) . Then (R, Θ, Φ) has probability density function g given by

$$g(r, \theta, \phi) = f(r \sin \phi \cos \theta, r \sin \phi \sin \theta, r \cos \phi)r^2 \sin \phi, \quad (r, \theta, \phi) \in [0, \infty) \times [0, 2\pi) \times [0, \pi] \quad (3.7.34)$$

Sign and Absolute Value

Our next discussion concerns the sign and absolute value of a real-valued random variable.

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

1. $|X|$ has distribution function G given by $G(y) = F(y) - F(-y)$ for $y \in [0, \infty)$.
2. $|X|$ has probability density function g given by $g(y) = f(y) + f(-y)$ for $y \in [0, \infty)$.

Proof

1. $\mathbb{P}(|X| \leq y) = \mathbb{P}(-y \leq X \leq y) = F(y) - F(-y)$ for $y \in [0, \infty)$.
2. This follows from part (a) by taking derivatives with respect to y .

Recall that the *sign* function on \mathbb{R} (not to be confused, of course, with the *sine* function) is defined as follows:

$$\text{sgn}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \quad (3.7.35)$$

Suppose again that X has a continuous distribution on \mathbb{R} with distribution function F and probability density function f , and suppose in addition that the distribution of X is symmetric about 0. Then

1. $|X|$ has distribution function G given by $G(y) = 2F(y) - 1$ for $y \in [0, \infty)$.
2. $|X|$ has probability density function g given by $g(y) = 2f(y)$ for $y \in [0, \infty)$.
3. $\text{sgn}(X)$ is uniformly distributed on $\{-1, 1\}$.
4. $|X|$ and $\text{sgn}(X)$ are independent.

Proof

1. This follows from the previous theorem, since $F(-y) = 1 - F(y)$ for $y > 0$ by symmetry.
2. This follows from part (a) by taking derivatives.
3. Note that $\mathbb{P}[\text{sgn}(X) = 1] = \mathbb{P}(X > 0) = \frac{1}{2}$ and so $\mathbb{P}[\text{sgn}(X) = -1] = \frac{1}{2}$ also.
4. If $A \subseteq (0, \infty)$ then

$$\mathbb{P}[|X| \in A, \text{sgn}(X) = 1] = \mathbb{P}(X \in A) = \int_A f(x) dx = \frac{1}{2} \int_A 2f(x) dx = \mathbb{P}[\text{sgn}(X) = 1] \mathbb{P}(|X| \in A) \quad (3.7.36)$$

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, with faces labeled from 1 to 6 (usually in the form of dots). 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 $\frac{1}{8}$ each.

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

1. Both dice are standard and fair.
2. Both dice are ace-six flat.
3. The first die is standard and fair, and the second is ace-six flat
4. The dice are both fair, but the first die has faces labeled 1, 2, 2, 3, 3, 4 and the second die has faces labeled 1, 3, 4, 5, 6, 8.

Answer

Let $Y = X_1 + X_2$ denote the sum of the scores.

1.	y	2	3	4	5	6	7	8	9	10	11	12
	$\mathbb{P}(Y = y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.	y	2	3	4	5	6	7	8	9	10	11	12
	$\mathbb{P}(Y = y)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{64}$	$\frac{3}{32}$	$\frac{7}{64}$	$\frac{3}{16}$	$\frac{7}{64}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{1}{16}$	$\frac{1}{16}$

3.	y	2	3	4	5	6	7	8	9	10	11	12
	$\mathbb{P}(Y = y)$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{6}{48}$	$\frac{8}{48}$	$\frac{6}{48}$	$\frac{5}{48}$	$\frac{4}{48}$	$\frac{3}{48}$	$\frac{2}{48}$

4. The distribution is the same as for two standard, fair dice in (a).

In the dice experiment, select two dice and select the sum random variable. Run the simulation 1000 times and compare the empirical density function to the probability density function for each of the following cases:

1. fair dice
2. ace-six flat dice

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

1. the minimum score
2. the maximum score.

Answer

Let U denote the minimum score and V the maximum score.

1. $f(u) = \left(1 - \frac{u-1}{6}\right)^n - \left(1 - \frac{u}{6}\right)^n, \quad u \in \{1, 2, 3, 4, 5, 6\}$
2. $g(v) = \left(\frac{v}{6}\right)^n - \left(\frac{v-1}{6}\right)^n, \quad v \in \{1, 2, 3, 4, 5, 6\}$

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 and note the agreement between the empirical density function and the probability density function.

1. minimum score
2. maximum score.

Uniform Distributions

Recall that for $n \in \mathbb{N}_+$, the standard measure of the size of a set $A \subseteq \mathbb{R}^n$ is

$$\lambda_n(A) = \int_A 1 \, dx \quad (3.7.37)$$

In particular, $\lambda_1(A)$ is the length of A for $A \subseteq \mathbb{R}$, $\lambda_2(A)$ is the area of A for $A \subseteq \mathbb{R}^2$, and $\lambda_3(A)$ is the volume of A for $A \subseteq \mathbb{R}^3$. See the technical details in (1) for more advanced information.

Now if $S \subseteq \mathbb{R}^n$ with $0 < \lambda_n(S) < \infty$, recall that the *uniform distribution* on S is the continuous distribution with constant probability density function f defined by $f(x) = 1/\lambda_n(S)$ for $x \in S$. Uniform distributions are studied in more detail in the chapter on Special Distributions.

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

1. X is uniformly distributed on the interval $[0, 4]$.
2. X is uniformly distributed on the interval $[-2, 2]$.
3. X is uniformly distributed on the interval $[-1, 3]$.

Answer

1. $g(y) = \frac{1}{8\sqrt{y}}, \quad 0 < y < 16$
2. $g(y) = \frac{1}{4\sqrt{y}}, \quad 0 < y < 4$
3. $g(y) = \begin{cases} \frac{1}{4\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{8\sqrt{y}}, & 1 < y < 9 \end{cases}$

Compare the distributions in the last exercise. In part (c), note that even a simple transformation of a simple distribution can produce a complicated distribution. In this particular case, the complexity is caused by the fact that $x \mapsto x^2$ is one-to-one on part of the domain $\{0\} \cup (1, 3]$ and two-to-one on the other part $[-1, 1] \setminus \{0\}$.

On the other hand, the uniform distribution is preserved under a linear transformation of the random variable.

Suppose that \mathbf{X} has the continuous uniform distribution on $S \subseteq \mathbb{R}^n$. Let $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$, where $\mathbf{a} \in \mathbb{R}^n$ and \mathbf{B} is an invertible $n \times n$ matrix. Then \mathbf{Y} is uniformly distributed on $T = \{\mathbf{a} + \mathbf{B}\mathbf{x} : \mathbf{x} \in S\}$.

Proof

This follows directly from the general result on linear transformations in (10). Note that the PDF g of \mathbf{Y} is constant on T .

For the following three exercises, recall that the *standard uniform distribution* is the uniform distribution on the interval $[0, 1]$.

Suppose that X and Y are independent and that each has the standard uniform distribution. Let $U = X + Y$, $V = X - Y$, $W = XY$, $Z = Y/X$. Find the probability density function of each of the follow:

1. (U, V)
2. U
3. V
4. W
5. Z

Answer

1. $g(u, v) = \frac{1}{2}$ for (u, v) in the square region $T \subset \mathbb{R}^2$ with vertices $\{(0, 0), (1, 1), (2, 0), (1, -1)\}$ So (U, V) is uniformly distributed on T .
2. $g_1(u) = \begin{cases} u, & 0 < u < 1 \\ 2 - u, & 1 < u < 2 \end{cases}$
3. $g_2(v) = \begin{cases} 1 - v, & 0 < v < 1 \\ 1 + v, & -1 < v < 0 \end{cases}$
4. $h_1(w) = -\ln w$ for $0 < w \leq 1$
5. $h_2(z) = \begin{cases} \frac{1}{2}, & 0 \leq z \leq 1 \\ \frac{1}{2z^2}, & 1 \leq z < \infty \end{cases}$

Suppose that X , Y , and Z are independent, and that each has the standard uniform distribution. Find the probability density function of $(U, V, W) = (X + Y, Y + Z, X + Z)$.

Answer

$g(u, v, w) = \frac{1}{2}$ for (u, v, w) in the rectangular region $T \subset \mathbb{R}^3$ with vertices $\{(0, 0, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1), (2, 1, 1), (1, 1, 2), (1, 2, 1), (2, 2, 2)\}$ So (U, V, W) is uniformly distributed on T .

Suppose that (X_1, X_2, \dots, X_n) is a sequence of independent random variables, each with the standard uniform distribution. Find the distribution function and probability density function of the following variables.

1. $U = \min\{X_1, X_2, \dots, X_n\}$
2. $V = \max\{X_1, X_2, \dots, X_n\}$

Answer

1. $G(t) = 1 - (1 - t)^n$ and $g(t) = n(1 - t)^{n-1}$, both for $t \in [0, 1]$
2. $H(t) = t^n$ and $h(t) = nt^{n-1}$, both for $t \in [0, 1]$

Both distributions in the last exercise are *beta distributions*. More generally, all of the order statistics from a random sample of standard uniform variables have beta distributions, one of the reasons for the importance of this family of distributions. Beta distributions are studied in more detail in the chapter on Special Distributions.

In the order statistic experiment, select the uniform distribution.

1. Set $k = 1$ (this gives the minimum U). Vary n with the scroll bar and note the shape of the probability density function. With $n = 5$, run the simulation 1000 times and note the agreement between the empirical density function and the true probability density function.
2. Vary n with the scroll bar, set $k = n$ each time (this gives the maximum V), and note the shape of the probability density function. With $n = 5$ run the simulation 1000 times and compare the empirical density function and the probability density function.

Let f denote the probability density function of the standard uniform distribution.

1. Compute f^{*2}
2. Compute f^{*3}
3. Graph f , f^{*2} , and f^{*3} on the same set of axes.

Answer

$$1. f^{*2}(z) = \begin{cases} z, & 0 < z < 1 \\ 2 - z, & 1 < z < 2 \end{cases}$$

$$2. f^{*3}(z) = \begin{cases} \frac{1}{2}z^2, & 0 < z < 1 \\ 1 - \frac{1}{2}(z-1)^2 - \frac{1}{2}(2-z)^2, & 1 < z < 2 \\ \frac{1}{2}(3-z)^2, & 2 < z < 3 \end{cases}$$

In the last exercise, you can see the behavior predicted by the central limit theorem beginning to emerge. Recall that if (X_1, X_2, X_3) is a sequence of independent random variables, each with the standard uniform distribution, then f , f^{*2} , and f^{*3} are the probability density functions of X_1 , $X_1 + X_2$, and $X_1 + X_2 + X_3$, respectively. More generally, if (X_1, X_2, \dots, X_n) is a sequence of independent random variables, each with the standard uniform distribution, then the distribution of $\sum_{i=1}^n X_i$ (which has probability density function f^{*n}) is known as the *Irwin-Hall* distribution with parameter n . The Irwin-Hall distributions are studied in more detail in the chapter on Special Distributions.

Open the Special Distribution Simulator and select the Irwin-Hall distribution. Vary the parameter n from 1 to 3 and note the shape of the probability density function. (These are the density functions in the previous exercise). For each value of n , run the simulation 1000 times and compare the empirical density function and the probability density function.

Simulations

A remarkable fact is that the standard uniform distribution 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 with the standard uniform distribution. Conversely, any continuous distribution supported on an interval of \mathbb{R} can be transformed into the standard uniform distribution.

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.

Suppose that U has the standard uniform distribution. Then $X = F^{-1}(U)$ has distribution function F .

Proof

The critical property satisfied by the quantile function (regardless of the type of distribution) is $F^{-1}(p) \leq x$ if and only if $p \leq F(x)$ for $p \in (0, 1)$ and $x \in \mathbb{R}$. Hence for $x \in \mathbb{R}$, $\mathbb{P}(X \leq x) = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x)$.

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 apps in this project use this method of simulation. The first image below shows the graph of the distribution function of a rather complicated mixed distribution, represented in blue on the horizontal axis. In the second image, note how the uniform distribution on $[0, 1]$, represented by the thick red line, is transformed, via the quantile function, into the given distribution.

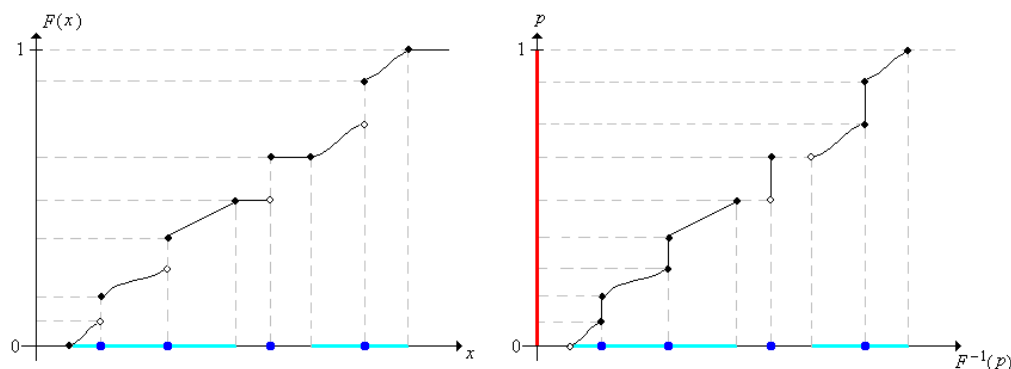


Figure 3.7.8: The random quantile method of simulation

There is a partial converse to the previous result, for continuous distributions.

Suppose that X has a continuous distribution on an interval $S \subseteq \mathbb{R}$. Then $U = F(X)$ has the standard uniform distribution.

Proof

For $u \in (0, 1)$ recall that $F^{-1}(u)$ is a quantile of order u . Since X has a continuous distribution,

$$\mathbb{P}(U \geq u) = \mathbb{P}[F(X) \geq u] = \mathbb{P}[X \geq F^{-1}(u)] = 1 - F[F^{-1}(u)] = 1 - u \quad (3.7.38)$$

Hence U is uniformly distributed on $(0, 1)$.

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 $[2, 10]$.

Answer

$X = a + U(b - a)$ where U is a random number.

Beta Distributions

Suppose that X has the probability density function f given by $f(x) = 3x^2$ for $0 \leq x \leq 1$. Find the probability density function of each of the following:

1. $U = X^2$
2. $V = \sqrt{X}$
3. $W = \frac{1}{X}$

Proof

1. $g(u) = \frac{3}{2}u^{1/2}$, for $0 < u \leq 1$
2. $h(v) = 6v^5$ for $0 \leq v \leq 1$
3. $k(w) = \frac{3}{w^4}$ for $1 \leq w < \infty$

Random variables X , U , and V in the previous exercise have *beta distributions*, the same family of distributions that we saw in the exercise [above](#) for the minimum and maximum of independent standard uniform variables. In general, beta distributions are widely used to model random proportions and probabilities, as well as physical quantities that take values in closed bounded intervals (which after a change of units can be taken to be $[0, 1]$). On the other hand, W has a Pareto distribution, named for Vilfredo Pareto. The family of beta distributions and the family of Pareto distributions are studied in more detail in the chapter on Special Distributions.

Suppose that the radius R of a sphere has a beta distribution probability density function f given by $f(r) = 12r^2(1 - r)$ for $0 \leq r \leq 1$. Find the probability density function of each of the following:

1. The circumference $C = 2\pi R$
2. The surface area $A = 4\pi R^2$
3. The volume $V = \frac{4}{3}\pi R^3$

Answer

1. $g(c) = \frac{3}{4\pi^4} c^2 (2\pi - c)$ for $0 \leq c \leq 2\pi$
2. $h(a) = \frac{3}{8\pi^2} \sqrt{a} (2\sqrt{\pi} - \sqrt{a})$ for $0 \leq a \leq 4\pi$
3. $k(v) = \frac{3}{\pi} \left[1 - \left(\frac{3}{4\pi} \right)^{1/3} v^{1/3} \right]$ for $0 \leq v \leq \frac{4}{3}\pi$

Suppose that the grades on a test are described by the random variable $Y = 100X$ where X has the beta distribution with probability density function f given by $f(x) = 12x(1-x)^2$ for $0 \leq x \leq 1$. The grades are generally low, so the teacher decides to “curve” the grades using the transformation $Z = 10\sqrt{Y} = 100\sqrt{X}$. Find the probability density function of

1. Y
2. Z

Answer

1. $g(y) = \frac{3}{25} \left(\frac{y}{100} \right) \left(1 - \frac{y}{100} \right)^2$ for $0 \leq y \leq 100$.
2. $h(z) = \frac{3}{1250} z \left(\frac{z^2}{10\,000} \right) \left(1 - \frac{z^2}{10\,000} \right)^2$ for $0 \leq z \leq 100$

Bernoulli Trials

Recall that a *Bernoulli trials sequence* is a sequence (X_1, X_2, \dots) 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 $p = \mathbb{P}(X_i = 1)$, so $p \in [0, 1]$. The random process is named for Jacob Bernoulli and is studied in detail in the chapter on Bernoulli trials.

For $i \in \mathbb{N}_+$, the probability density function f of the trial variable X_i is $f(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$.

Proof

By definition, $f(0) = 1 - p$ and $f(1) = p$. These can be combined succinctly with the formula $f(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$.

Now let Y_n denote the number of successes in the first n trials, so that $Y_n = \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$.

Y_n has the probability density function f_n given by

$$f_n(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \dots, n\} \quad (3.7.39)$$

Proof

We have seen this derivation before. The number of bit strings of length n with 1 occurring exactly y times is $\binom{n}{y}$ for $y \in \{0, 1, \dots, n\}$. By the Bernoulli trials assumptions, the probability of each such bit string is $p^y(1-p)^{n-y}$.

The distribution of Y_n is the *binomial distribution* with parameters n and p . The binomial distribution is studied in more detail in the chapter on Bernoulli trials

For $m, n \in \mathbb{N}$

1. $f_n = f^{*n}$.
2. $f_m * f_n = f_{m+n}$.

Proof

Part (a) can be proved directly from the definition of convolution, but the result also follows simply from the fact that $Y_n = X_1 + X_2 + \dots + X_n$.

From part (b) it follows that if Y and Z are independent variables, and that Y has the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$ while Z has the binomial distribution with parameter $m \in \mathbb{N}$ and p , then $Y + Z$ has the binomial distribution with parameter $m + n$ and p .

Find the probability density function of the difference between the number of successes and the number of failures in $n \in \mathbb{N}$ Bernoulli trials with success parameter $p \in [0, 1]$

Answer

$$f(k) = \binom{n}{(n+k)/2} p^{(n+k)/2} (1-p)^{(n-k)/2} \text{ for } k \in \{-n, -2, \dots, n-2, n\}$$

The Poisson Distribution

Recall that the *Poisson distribution* with parameter $t \in (0, \infty)$ has probability density function f given by

$$f_t(n) = e^{-t} \frac{t^n}{n!}, \quad n \in \mathbb{N} \quad (3.7.40)$$

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 parameter t is proportional to the size of the region. The Poisson distribution is studied in detail in the chapter on The Poisson Process.

If $a, b \in (0, \infty)$ then $f_a * f_b = f_{a+b}$.

Proof

Let $z \in \mathbb{N}$. Using the definition of convolution and the binomial theorem we have

$$(f_a * f_b)(z) = \sum_{x=0}^z f_a(x) f_b(z-x) = \sum_{x=0}^z e^{-a} \frac{a^x}{x!} e^{-b} \frac{b^{z-x}}{(z-x)!} = e^{-(a+b)} \frac{1}{z!} \sum_{x=0}^z \frac{z!}{x!(z-x)!} a^x b^{z-x} \quad (3.7.41)$$

$$= e^{-(a+b)} \frac{1}{z!} \sum_{x=0}^z \binom{z}{x} a^x b^{z-x} = e^{-(a+b)} \frac{(a+b)^z}{z!} = f_{a+b}(z) \quad (3.7.42)$$

The last result means that if X and Y are independent variables, and X has the Poisson distribution with parameter $a > 0$ while Y has the Poisson distribution with parameter $b > 0$, then $X + Y$ has the Poisson distribution with parameter $a + b$. In terms of the Poisson model, X could represent the number of points in a region A and Y the number of points in a region B (of the appropriate sizes so that the parameters are a and b respectively). The independence of X and Y corresponds to the regions A and B being disjoint. Then $X + Y$ is the number of points in $A \cup B$.

The Exponential Distribution

Recall that the *exponential distribution* with rate parameter $r \in (0, \infty)$ has probability density function f given by $f(t) = r e^{-rt}$ for $t \in [0, \infty)$. This distribution is often used to model random times such as failure times and lifetimes. In particular, the times between arrivals in the Poisson model of random points in time have independent, identically distributed exponential distributions. The Exponential distribution is studied in more detail in the chapter on Poisson Processes.

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 $r = 3$.

Answer

$$X = -\frac{1}{r} \ln(1 - U) \text{ where } U \text{ is a random number. Since } 1 - U \text{ is also a random number, a simpler solution is } X = -\frac{1}{r} \ln U.$$

For the next exercise, recall that the *floor* and *ceiling* functions on \mathbb{R} are defined by

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}, \quad \lceil x \rceil = \min\{n \in \mathbb{Z} : n \geq x\}, \quad x \in \mathbb{R} \quad (3.7.43)$$

Suppose that T has the exponential distribution with rate parameter $r \in (0, \infty)$. Find the probability density function of each of the following random variables:

1. $Y = \lfloor T \rfloor$
2. $Z = \lceil T \rceil$

Answer

1. $\mathbb{P}(Y = n) = e^{-rn} (1 - e^{-r}) \text{ for } n \in \mathbb{N}$
2. $\mathbb{P}(Z = n) = e^{-r(n-1)} (1 - e^{-r}) \text{ for } n \in \mathbb{N}$

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.

Suppose that T has the exponential distribution with rate parameter $r \in (0, \infty)$. Find the probability density function of each of the following random variables:

1. $X = T^2$
2. $Y = e^T$
3. $Z = \ln T$

Answer

1. $g(x) = re^{-r\sqrt{x}} / 2\sqrt{x}$ for $0 < x < \infty$
2. $h(y) = ry^{-(r+1)}$ for $1 < y < \infty$
3. $k(z) = r \exp(-re^z)e^z$ for $z \in \mathbb{R}$

In the previous exercise, Y has a Pareto distribution while Z has an extreme value distribution. Both of these are studied in more detail in the chapter on Special Distributions.

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

1. Find the distribution function of Z .
2. Find the probability density function of Z .

Answer

1. $G(z) = 1 - \frac{1}{1+z}$, $0 < z < \infty$
2. $g(z) = \frac{1}{(1+z)^2}$, $0 < z < \infty$

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 $Z = X + Y$ in each of the following cases.

1. $a = b$
2. $a \neq b$

Answer

1. $h(z) = a^2 ze^{-az}$ for $0 < z < \infty$
2. $h(z) = \frac{ab}{b-a} (e^{-az} - e^{-bz})$ for $0 < z < \infty$

Suppose that (T_1, T_2, \dots, 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 $i \in \{1, 2, \dots, n\}$.

1. Find the probability density function of $U = \min\{T_1, T_2, \dots, T_n\}$.
2. Find the distribution function of $V = \max\{T_1, T_2, \dots, T_n\}$.
3. Find the probability density function of V in the special case that $r_i = r$ for each $i \in \{1, 2, \dots, n\}$.

Answer

1. $g(t) = ae^{-at}$ for $0 \leq t < \infty$ where $a = r_1 + r_2 + \dots + r_n$
2. $H(t) = (1 - e^{-r_1 t})(1 - e^{-r_2 t}) \dots (1 - e^{-r_n t})$ for $0 \leq t < \infty$
3. $h(t) = nre^{-rt}(1 - e^{-rt})^{n-1}$ for $0 \leq t < \infty$

Note that the minimum U in part (a) has the exponential distribution with parameter $r_1 + r_2 + \dots + r_n$. 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.

In the order statistic experiment, select the exponential distribution.

1. Set $k = 1$ (this gives the minimum U). Vary n with the scroll bar and note the shape of the probability density function. With $n = 5$, run the simulation 1000 times and compare the empirical density function and the probability density function.
2. Vary n with the scroll bar and set $k = n$ each time (this gives the maximum V). Note the shape of the density function. With $n = 5$, run the simulation 1000 times and compare the empirical density function and the probability density function.

Suppose again that (T_1, T_2, \dots, 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 $i \in \{1, 2, \dots, n\}$. Then

$$\mathbb{P}(T_i < T_j \text{ for all } j \neq i) = \frac{r_i}{\sum_{j=1}^n r_j} \quad (3.7.44)$$

Proof

When $n = 2$, the result was shown in the section on joint distributions. Returning to the case of general n , note that $T_i < T_j$ for all $j \neq i$ if and only if $T_i < \min\{T_j : j \neq i\}$. Note that the minimum on the right is independent of T_i and by the [result above](#), 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. If we have a bunch of independent alarm clocks, with exponentially distributed alarm times, then the probability that clock i is the first one to sound is $r_i / \sum_{j=1}^n r_j$.

The Gamma Distribution

Recall that the (standard) *gamma distribution* with shape parameter $n \in \mathbb{N}_+$ has probability density function

$$g_n(t) = e^{-t} \frac{t^{n-1}}{(n-1)!}, \quad 0 \leq t < \infty \quad (3.7.45)$$

With a positive integer shape parameter, as we have here, it is also referred to as the *Erlang distribution*, named for Agner Erlang. This distribution is widely used to model random times under certain basic assumptions. In particular, the n th arrival times in the Poisson model of random points in time has the gamma distribution with parameter n . The Erlang distribution is studied in more detail in the chapter on the Poisson Process, and in greater generality, the gamma distribution is studied in the chapter on Special Distributions.

Let $g = g_1$, and note that this is the probability density function of the exponential distribution with parameter 1, which was the topic of our last discussion.

If $m, n \in \mathbb{N}_+$ then

1. $g_n = g^{*n}$
2. $g_m * g_n = g_{m+n}$

Proof

Part (a) hold trivially when $n = 1$. Also, for $t \in [0, \infty)$,

$$g_n * g(t) = \int_0^t g_n(s)g(t-s) ds = \int_0^t e^{-s} \frac{s^{n-1}}{(n-1)!} e^{-(t-s)} ds = e^{-t} \int_0^t \frac{s^{n-1}}{(n-1)!} ds = e^{-t} \frac{t^n}{n!} = g_{n+1}(t) \quad (3.7.46)$$

Part (b) follows from (a).

Part (b) means 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 $m + n$. In the context of the Poisson model, part (a) means that the n th arrival time is the sum of the n independent interarrival times, which have a common exponential distribution.

Suppose that T has the gamma distribution with shape parameter $n \in \mathbb{N}_+$. Find the probability density function of $X = \ln T$.

Answer

$$h(x) = \frac{1}{(n-1)!} \exp(-e^x) e^{nx} \text{ for } x \in \mathbb{R}$$

The Pareto Distribution

Recall that the *Pareto distribution* with shape parameter $a \in (0, \infty)$ has probability density function f given by

$$f(x) = \frac{a}{x^{a+1}}, \quad 1 \leq x < \infty \quad (3.7.47)$$

Members of this family have already come up in several of the previous exercises. The Pareto distribution, named for Vilfredo Pareto, 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.

Suppose that X has the Pareto distribution with shape parameter a . Find the probability density function of each of the following random variables:

1. $U = X^2$
2. $V = \frac{1}{X}$
3. $Y = \ln X$

Answer

1. $g(u) = \frac{a/2}{u^{a/2+1}}$ for $1 \leq u < \infty$
2. $h(v) = av^{a-1}$ for $0 < v < 1$
3. $k(y) = ae^{-ay}$ for $0 \leq y < \infty$

In the previous exercise, V also has a Pareto distribution but with parameter $\frac{a}{2}$; Y has the beta distribution with parameters a and $b = 1$; and Z has the exponential distribution with rate parameter a .

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 $a = 2$.

Answer

Using the [random quantile method](#), $X = \frac{1}{(1-U)^{1/a}}$ where U is a random number. More simply, $X = \frac{1}{U^{1/a}}$, since $1 - U$ is also a random number.

The Normal Distribution

Recall that the *standard normal distribution* has probability density function ϕ given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \quad (3.7.48)$$

Suppose that Z has the standard normal distribution, and that $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$.

1. Find the probability density function f of $X = \mu + \sigma Z$
2. Sketch the graph of f , noting the important qualitative features.

Answer

1. $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ for $x \in \mathbb{R}$
2. f is symmetric about $x = \mu$. f increases and then decreases, with mode $x = \mu$. f is concave upward, then downward, then upward again, with inflection points at $x = \mu \pm \sigma$. $f(x) \rightarrow 0$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$

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, primarily because of the central limit theorem, one of the fundamental theorems. 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.

Suppose that Z has the standard normal distribution. Find the probability density function of Z^2 and sketch the graph.

Answer

$$g(v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{1}{2}v} \text{ for } 0 < v < \infty$$

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.

Suppose that X and Y are independent random variables, each with the standard normal distribution, and let (R, Θ) be the standard polar coordinates (X, Y) . Find the probability density function of

1. (R, Θ)
2. R
3. Θ

Answer

Note that the joint PDF of (X, Y) is

$$f(x, y) = \phi(x)\phi(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, \quad (x, y) \in \mathbb{R}^2 \quad (3.7.49)$$

From the result above [polar coordinates](#), the PDF of (R, Θ) is

$$g(r, \theta) = f(r \cos \theta, r \sin \theta) r = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2}, \quad (r, \theta) \in [0, \infty) \times [0, 2\pi) \quad (3.7.50)$$

From the factorization theorem for joint PDFs, it follows that R has probability density function $h(r) = r e^{-\frac{1}{2}r^2}$ for $0 \leq r < \infty$, Θ is uniformly distributed on $[0, 2\pi)$, and that R and Θ are independent.

The distribution of R is the (standard) *Rayleigh distribution*, and is named for John William Strutt, Lord Rayleigh. The Rayleigh distribution is studied in more detail in the chapter on Special Distributions.

The standard normal distribution does not have a simple, closed form quantile function, so the random quantile method of simulation does not work well. However, the last exercise points the way to an alternative method of simulation.

Show how to simulate a pair of independent, standard normal variables with a pair of random numbers. Using your calculator, simulate 6 values from the standard normal distribution.

Answer

The Rayleigh distribution in the last exercise has CDF $H(r) = 1 - e^{-\frac{1}{2}r^2}$ for $0 \leq r < \infty$, and hence quantile function $H^{-1}(p) = \sqrt{-2 \ln(1-p)}$ for $0 \leq p < 1$. Thus we can simulate the polar radius R with a random number U by $R = \sqrt{-2 \ln(1-U)}$, or a bit more simply by $R = \sqrt{-2 \ln U}$, since $1-U$ is also a random number. We can simulate the polar angle Θ with a random number V by $\Theta = 2\pi V$. Then, a pair of independent, standard normal variables can be simulated by $X = R \cos \Theta$, $Y = R \sin \Theta$.

The Cauchy Distribution

Suppose that X and Y are independent random variables, each with the standard normal distribution. Find the probability density function of $T = X/Y$.

Answer

As usual, let ϕ denote the standard normal PDF, so that $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ for $z \in \mathbb{R}$. Using the [theorem on quotient](#) above, the PDF f of T is given by

$$f(t) = \int_{-\infty}^{\infty} \phi(x)\phi(tx)|x|dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(1+t^2)x^2/2}|x|dx, \quad t \in \mathbb{R} \quad (3.7.51)$$

Using symmetry and a simple substitution,

$$f(t) = \frac{1}{\pi} \int_0^{\infty} x e^{-(1+t^2)x^2/2} dx = \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R} \quad (3.7.52)$$

Random variable T has the (standard) *Cauchy distribution*, named after Augustin Cauchy. The Cauchy distribution is studied in detail in the chapter on Special Distributions.

Suppose that a light source is 1 unit away from position 0 on an infinite straight wall. We shine the light at the wall an angle Θ to the perpendicular, where Θ is uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Find the probability density function of the position of the light beam $X = \tan \Theta$ on the wall.

Answer

The PDF of Θ is $f(\theta) = \frac{1}{\pi}$ for $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$. The transformation is $x = \tan \theta$ so the inverse transformation is $\theta = \arctan x$. Recall that $\frac{d\theta}{dx} = \frac{1}{1+x^2}$, so by the [change of variables formula](#), X has PDF g given by

$$g(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R} \quad (3.7.53)$$

Thus, X also has the standard Cauchy distribution. Clearly we can simulate a value of the Cauchy distribution by $X = \tan(-\frac{\pi}{2} + \pi U)$ where U is a random number. This is the random quantile method.

Open the Cauchy experiment, which is a simulation of the light problem in the previous exercise. Keep the default parameter values and run the experiment in single step mode a few times. Then run the experiment 1000 times and compare the empirical density function and the probability density function.

This page titled [3.7: Transformations of Random Variables](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.