

3.2: Continuous Distributions

In the previous section, we considered discrete distributions. In this section, we study a complementary type of distribution. As usual, if you are a new student of probability, you may want to skip the technical details.

Basic Theory

Definitions and Basic Properties

As usual, our starting point is a random experiment modeled by a probability space $(S, \mathcal{S}, \mathbb{P})$. So to review, S is the set of outcomes, \mathcal{S} the collection of events, and \mathbb{P} the probability measure on the sample space (S, \mathcal{S}) . We use the terms *probability measure* and *probability distribution* synonymously in this text. Also, since we use a general definition of random variable, every probability measure can be thought of as the probability distribution of a random variable, so we can always take this point of view if we like. Indeed, most probability measures naturally have random variables associated with them.

In this section, we assume that $S \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$.

Details

Technically, S is a measurable subset of \mathbb{R}^n and \mathcal{S} is the σ -algebra measurable subsets of S . Typically in applications, S is defined by a finite number of inequalities involving elementary function.

Here is our first fundamental definition.

The probability measure \mathbb{P} is *continuous* if $\mathbb{P}(\{x\}) = 0$ for all $x \in S$.

The fact that each point is assigned probability 0 might seem impossible or paradoxical at first, but soon we will see very familiar analogies.

If \mathbb{P} is a continuous distribution then $\mathbb{P}(C) = 0$ for every countable $C \subseteq S$.

Proof

Since C is countable, it follows from the additivity axiom of probability that

$$\mathbb{P}(C) = \sum_{x \in C} \mathbb{P}(\{x\}) = 0 \quad (3.2.1)$$

Thus, continuous distributions are in complete contrast with discrete distributions, for which all of the probability mass is concentrated on the points in a discrete set. For a continuous distribution, the probability mass is *continuously* spread over S in some sense. In the picture below, the light blue shading is intended to suggest a continuous distribution of probability.

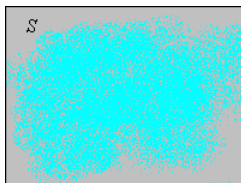


Figure 3.2.1: A continuous probability distribution on S

Typically, S is a region of \mathbb{R}^n defined by inequalities involving elementary functions, for example an interval in \mathbb{R} , a circular region in \mathbb{R}^2 , and a conical region in \mathbb{R}^3 . Suppose that \mathbb{P} is a continuous probability measure on S . The fact that each point in S has probability 0 is conceptually the same as the fact that an interval of \mathbb{R} can have positive length even though it is composed of points each of which has 0 length. Similarly, a region of \mathbb{R}^2 can have positive area even though it is composed of points (or curves) each of which has area 0. In the one-dimensional case, continuous distributions are used to model random variables that take values in intervals of \mathbb{R} , variables that can, in principle, be measured with any degree of accuracy. Such variables abound in applications and include

- length, area, volume, and distance

- time
- mass and weight
- charge, voltage, and current
- resistance, capacitance, and inductance
- velocity and acceleration
- energy, force, and work

Usually a continuous distribution can usually be described by certain type of function.

Suppose again that \mathbb{P} is a continuous distribution on S . A function $f : S \rightarrow [0, \infty)$ is a *probability density function* for \mathbb{P} if

$$\mathbb{P}(A) = \int_A f(x) dx, \quad A \in \mathcal{S} \quad (3.2.2)$$

Details

Technically, f must be measurable and is a probability density function of \mathbb{P} *with respect* to Lebesgue measure, the standard measure on \mathbb{R}^n . Moreover, the integral is the Lebesgue integral, but the ordinary Riemann integral of calculus suffices for the sets that occur in typical applications.

So the probability distribution \mathbb{P} is completely determined by the probability density function f . As a special case, note that $\int_S f(x) dx = \mathbb{P}(S) = 1$. Conversely, a nonnegative function on S with this property defines a probability measure.

A function $f : S \rightarrow [0, \infty)$ that satisfies $\int_S f(x) dx = 1$ is a *probability density function* on S and then \mathbb{P} defined as follows is a continuous probability measure on S :

$$\mathbb{P}(A) = \int_A f(x) dx, \quad A \in \mathcal{S} \quad (3.2.3)$$

Proof

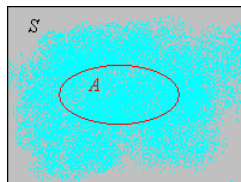


Figure 3.2.2: A continuous distribution is completely determined by its probability density function

Note that we can always extend f to a probability density function on a subset of \mathbb{R}^n that contains S , or to all of \mathbb{R}^n , by defining $f(x) = 0$ for $x \notin S$. This extension sometimes simplifies notation. Put another way, we can be a bit sloppy about the “set of values” of the random variable. So for example if $a, b \in \mathbb{R}$ with $a < b$ and X has a continuous distribution on the interval $[a, b]$, then we could also say that X has a continuous distribution on (a, b) or $[a, b)$, or $(a, b]$.

The points $x \in S$ that maximize the probability density function f are important, just as in the discrete case.

Suppose that \mathbb{P} is a continuous distribution on S with probability density function f . An element $x \in S$ that maximizes f is a *mode* of the distribution.

If there is only one mode, it is sometimes used as a measure of the *center* of the distribution.

You have probably noticed that probability density functions for continuous distributions are analogous to probability density functions for discrete distributions, with integrals replacing sums. However, there are essential differences. First, every discrete distribution has a unique probability density function f given by $f(x) = \mathbb{P}(\{x\})$ for $x \in S$. For a continuous distribution, the existence of a probability density function is not guaranteed. The advanced section on absolute continuity and density functions has several examples of continuous distribution that do not have density functions, and gives conditions that are necessary and sufficient for the existence of a probability density function. Even if a probability density function f exists, it is *never* unique. Note that the values of f on a finite (or even countably infinite) set of points could be changed to other nonnegative values and the new function would still be a probability density function for the same distribution. The critical fact is that only *integrals* of f are important. Second, the values of the PDF f for a discrete distribution are probabilities, and in particular $f(x) \leq 1$ for $x \in S$. For a

continuous distribution the values are not probabilities and in fact it's possible that $f(x) > 1$ for some or even all $x \in S$. Further, f can be unbounded on S . In the typical calculus interpretation, $f(x)$ really is probability *density* at x . That is, $f(x) dx$ is approximately the probability of a "small" region of size dx about x .

Constructing Probability Density Functions

Just as in the discrete case, a nonnegative function on S can often be scaled to produce a probability density function.

Suppose that $g: S \rightarrow [0, \infty)$ and let

$$c = \int_S g(x) dx \quad (3.2.4)$$

If $0 < c < \infty$ then f defined by $f(x) = \frac{1}{c}g(x)$ for $x \in S$ defines a probability density function for a continuous distribution on S .

Proof

Technically, the function g is measurable. Technicalities aside, the proof is trivial. Clearly $f(x) \geq 0$ for $x \in S$ and

$$\int_S f(x) dx = \frac{1}{c} \int_S g(x) dx = \frac{c}{c} = 1 \quad (3.2.5)$$

Note again that f is just a scaled version of g . So this result can be used to construct probability density functions with desired properties (domain, shape, symmetry, and so on). The constant c is sometimes called the *normalizing constant* of g .

Conditional Densities

Suppose now that X is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that X has a continuous distribution on S . A probability density function for X is based on the underlying probability measure on the sample space (Ω, \mathcal{F}) . This measure could be a conditional probability measure, conditioned on a given event $E \in \mathcal{F}$ with $\mathbb{P}(E) > 0$. Assuming that the conditional probability density function exists, the usual notation is

$$f(x | E), \quad x \in S \quad (3.2.6)$$

Note, however, that except for notation, no new concepts are involved. The defining property is

$$\int_A f(x | E) dx = \mathbb{P}(X \in A | E), \quad A \in \mathcal{S} \quad (3.2.7)$$

and all results that hold for probability density functions in general hold for conditional probability density functions. The event E could be an event described in terms of the random variable X itself:

Suppose that X has a continuous distribution on S with probability density function f and that $B \in \mathcal{S}$ with $\mathbb{P}(X \in B) > 0$. The conditional probability density function of X given $X \in B$ is the function on B given by

$$f(x | X \in B) = \frac{f(x)}{\mathbb{P}(X \in B)}, \quad x \in B \quad (3.2.8)$$

Proof

For $A \in \mathcal{S}$ with $A \subseteq B$,

$$\int_A \frac{f(x)}{\mathbb{P}(X \in B)} dx = \frac{1}{\mathbb{P}(X \in B)} \int_A f(x) dx = \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in B)} = \mathbb{P}(X \in A | X \in B) \quad (3.2.9)$$

Of course, $\mathbb{P}(X \in B) = \int_B f(x) dx$ and hence is the normalizing constant for the restriction of f to B , as in (8)

Examples and Applications

As always, try the problems yourself before looking at the answers.

The Exponential Distribution

Let f be the function defined by $f(t) = re^{-rt}$ for $t \in [0, \infty)$, where $r \in (0, \infty)$ is a parameter.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and state the important qualitative features.

Proof

1. Note that $f(t) > 0$ for $t \geq 0$. Also $\int_0^\infty e^{-rt} dt = \frac{1}{r}$ so f is a PDF.
2. f is decreasing and concave upward so the mode is 0. $f(x) \rightarrow 0$ as $x \rightarrow \infty$.

The distribution defined by the probability density function in the previous exercise is called the *exponential distribution* with rate parameter r . This distribution is frequently used to model random times, under certain assumptions. Specifically, in the Poisson model of random points in time, the times between successive arrivals have independent exponential distributions, and the parameter r is the average rate of arrivals. The exponential distribution is studied in detail in the chapter on Poisson Processes.

The lifetime T of a certain device (in 1000 hour units) has the exponential distribution with parameter $r = \frac{1}{2}$. Find

1. $\mathbb{P}(T > 2)$
2. $\mathbb{P}(T > 3 \mid T > 1)$

Answer

1. $e^{-1} \approx 0.3679$
2. $e^{-1} \approx 0.3679$

In the gamma experiment, set $n = 1$ to get the exponential distribution. Vary the rate parameter r and note the shape of the probability density function. For various values of r , run the simulation 1000 times and compare the empirical density function with the probability density function.

A Random Angle

In *Bertrand's problem*, a certain random angle Θ has probability density function f given by $f(\theta) = \sin \theta$ for $\theta \in [0, \frac{\pi}{2}]$.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph f , and state the important qualitative features.
3. Find $\mathbb{P}(\Theta < \frac{\pi}{4})$.

Answer

1. Note that $\sin \theta \geq 0$ for $0 \leq \theta \leq \frac{\pi}{2}$ and $\int_0^{\pi/2} \sin \theta d\theta = 1$.
2. f is increasing and concave downward so the mode is $\frac{\pi}{2}$.
3. $1 - \frac{1}{\sqrt{2}} \approx 0.2929$

Bertrand's problem is named for Joseph Louis Bertrand and is studied in more detail in the chapter on Geometric Models.

In Bertrand's experiment, select the model with uniform distance. Run the simulation 1000 times and compute the empirical probability of the event $\{\Theta < \frac{\pi}{4}\}$. Compare with the true probability in the previous exercise.

Gamma Distributions

Let g_n be the function defined by $g_n(t) = e^{-t} \frac{t^n}{n!}$ for $t \in [0, \infty)$ where $n \in \mathbb{N}$ is a parameter.

1. Show that g_n is a probability density function for each $n \in \mathbb{N}$.
2. Draw a careful sketch of the graph of g_n , and state the important qualitative features.

Proof

1. Note that $g_n(t) \geq 0$ for $t \geq 0$. Also, g_0 is the probability density function of the exponential distribution with parameter 1. For $n \in \mathbb{N}_+$, integration by parts with $u = t^n/n!$ and $dv = e^{-t} dt$ gives $\int_0^\infty g_n(t) dt = \int_0^\infty g_{n-1}(t) dt$. Hence it follows by induction that g_n is a PDF for each $n \in \mathbb{N}_+$.

2. g_0 is decreasing and concave downward, with mode $t = 0$. For $n > 0$, g_n increases and then decreases, with mode $t = n$. g_1 is concave downward and then upward, with inflection point at $t = 2$. For $n > 1$, g_n is concave upward, then downward, then upward again, with inflection points at $n \pm \sqrt{n}$. For all $n \in \mathbb{N}$, $g_n(t) \rightarrow 0$ as $t \rightarrow \infty$.

Interestingly, we showed in the last section on discrete distributions, that $f_t(n) = g_n(t)$ is a probability density function on \mathbb{N} for each $t \geq 0$ (it's the Poisson distribution with parameter t). The distribution defined by the probability density function g_n belongs to the family of *Erlang distributions*, named for Agner Erlang; $n + 1$ is known as the *shape parameter*. The Erlang distribution is studied in more detail in the chapter on the Poisson Process. In turn the Erlang distribution belongs to the more general family of *gamma distributions*. The gamma distribution is studied in more detail in the chapter on Special Distributions.

In the gamma experiment, keep the default rate parameter $r = 1$. Vary the shape parameter and note the shape and location of the probability density function. For various values of the shape parameter, run the simulation 1000 times and compare the empirical density function with the probability density function.

Suppose that the lifetime of a device T (in 1000 hour units) has the gamma distribution [above](#) with $n = 2$. Find each of the following:

1. $\mathbb{P}(T > 3)$.
2. $\mathbb{P}(T \leq 2)$
3. $\mathbb{P}(1 \leq T \leq 4)$

Answer

1. $\frac{17}{2}e^{-3} \approx 0.4232$
2. $1 - 5e^{-2} \approx 0.3233$
3. $\frac{5}{2}e^{-1} - 13e^{-4} \approx 0.6816$

Beta Distributions

Let f be the function defined by $f(x) = 6x(1 - x)$ for $x \in [0, 1]$.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and state the important qualitative features.

Answer

1. Note that $f(x) \geq 0$ for $x \in [0, 1]$. Also, $\int_0^1 x(1 - x) dx = \frac{1}{6}$, so f is a PDF
2. f increases and then decreases, with mode at $x = \frac{1}{2}$. f is concave downward. f is symmetric about $x = \frac{1}{2}$ (in fact, the graph is a parabola).

Let f be the function defined by $f(x) = 12x^2(1 - x)$ for $x \in [0, 1]$.

1. Show that f is a probability density function.
2. Draw a careful sketch the graph of f , and state the important qualitative features.

Answer

1. Note that $f(x) \geq 0$ for $0 \leq x \leq 1$. Also $\int_0^1 x^2(1 - x) dx = \frac{1}{12}$, so f is a PDF.
2. f increases and then decreases, with mode at $x = \frac{2}{3}$. f is concave upward and then downward, with inflection point at $x = \frac{1}{3}$.

The distributions defined in the last two exercises are examples of *beta distributions*. These distributions are widely used to model random proportions and probabilities, and physical quantities that take values in bounded intervals (which, after a change of units, can be taken to be $[0, 1]$). Beta distributions are studied in detail in the chapter on Special Distributions.

In the special distribution simulator, select the beta distribution. For the following parameter values, note the shape of the probability density function. Run the simulation 1000 times and compare the empirical density function with the probability density function.

1. $a = 2, b = 2$. This gives the first beta distribution [above](#).
2. $a = 3, b = 2$. This gives the second beta distribution [above](#).

Suppose that P is a random proportion. Find $\mathbb{P}\left(\frac{1}{4} \leq P \leq \frac{3}{4}\right)$ in each of the following cases:

1. P has the first beta distribution [above](#).
2. P has the second beta distribution [above](#).

Answer

1. $\frac{11}{16}$
2. $\frac{11}{16}$

Let f be the function defined by

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}}, \quad x \in (0, 1) \quad (3.2.10)$$

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and state the important qualitative features.

Answer

1. Note that $f(x) > 0$ for $0 < x < 1$. Using the substitution $u = \sqrt{x}$ gives

$$\int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = \int_0^1 \frac{2}{\sqrt{1-u^2}} du = 2 \arcsin u \Big|_0^1 = \pi \quad (3.2.11)$$

Thus f is a PDF.

2. f is symmetric about $x = \frac{1}{2}$. f decreases and then increases, with minimum at $x = \frac{1}{2}$. $f(x) \rightarrow \infty$ as $x \downarrow 0$ and as $x \uparrow 1$ so the distribution has no mode. f is concave upward.

The distribution defined in the last exercise is also a member of the beta family of distributions. But it is also known as the (standard) *arcsine distribution*, because of the arcsine function that arises in the proof that f is a probability density function. The arcsine distribution has applications to a very important random process known as Brownian motion, named for the Scottish botanist Robert Brown. Arcsine distributions are studied in more generality in the chapter on Special Distributions.

In the special distribution simulator, select the (continuous) arcsine distribution and keep the default parameter values. Run the simulation 1000 times and compare the empirical density function with the probability density function.

Suppose that X_t represents the change in the price of a stock at time t , relative to the value at an initial reference time 0. We treat t as a continuous variable measured in weeks. Let $T = \max\{t \in [0, 1] : X_t = 0\}$, the last time during the first week that the stock price was unchanged over its initial value. Under certain ideal conditions, T will have the arcsine distribution. Find each of the following:

1. $\mathbb{P}\left(T < \frac{1}{4}\right)$
2. $\mathbb{P}\left(T \geq \frac{1}{2}\right)$
3. $\mathbb{P}\left(T \leq \frac{3}{4}\right)$

Answer

1. $\frac{1}{3}$
2. $\frac{1}{2}$
3. $\frac{2}{3}$

Open the Brownian motion experiment and select the *last zero* variable. Run the experiment in single step mode a few times. The random process that you observe models the price of the stock in the previous exercise. Now run the experiment 1000 times and compute the empirical probability of each event in the previous exercise.

The Pareto Distribution

Let g be the function defined by $g(x) = 1/x^b$ for $x \in [1, \infty)$, where $b \in (0, \infty)$ is a parameter.

1. Draw a careful sketch the graph of g , and state the important qualitative features.
2. Find the values of b for which there exists a probability density function f (8)proportional to g . Identify the mode.

Answer

1. g is decreasing and concave upward, with $g(x) \rightarrow 0$ as $x \rightarrow \infty$.
2. Note that if $b \neq 1$

$$\int_1^\infty x^{-b} dx = \frac{x^{1-b}}{1-b} \Big|_1^\infty = \begin{cases} \infty, & 0 < b < 1 \\ \frac{1}{b-1}, & 1 < b < \infty \end{cases} \quad (3.2.12)$$

When $b = 1$ we have $\int_1^\infty x^{-1} dx = \ln x \Big|_1^\infty = \infty$. Thus, when $0 < b \leq 1$, there is no PDF proportional to g . When $b > 1$, the PDF proportional to g is $f(x) = \frac{b-1}{x^b}$ for $x \in [1, \infty)$. The mode is 1.

Note that the qualitative features of g are the same, regardless of the value of the parameter $b > 0$, but only when $b > 1$ can g be normalized into a probability density function. In this case, the distribution is known as the *Pareto distribution*, named for Vilfredo Pareto. The parameter $a = b - 1$, so that $a > 0$, is known as the *shape parameter*. Thus, the Pareto distribution with shape parameter a has probability density function

$$f(x) = \frac{a}{x^{a+1}}, \quad x \in [1, \infty) \quad (3.2.13)$$

The Pareto distribution is widely used to model certain economic variables and is studied in detail in the chapter on Special Distributions.

In the special distribution simulator, select the Pareto distribution. Leave the scale parameter fixed, but vary the shape parameter, and note the shape of the probability density function. For various values of the shape parameter, run the simulation 1000 times and compare the empirical density function with the probability density function.

Suppose that the income X (in appropriate units) of a person randomly selected from a population has the Pareto distribution with shape parameter $a = 2$. Find each of the following:

1. $\mathbb{P}(X > 2)$
2. $\mathbb{P}(X \leq 4)$
3. $\mathbb{P}(3 \leq X \leq 5)$

Answer

1. $\frac{1}{4}$
2. $\frac{15}{16}$
3. $\frac{16}{225}$

The Cauchy Distribution

Let f be the function defined by

$$f(x) = \frac{1}{\pi(x^2 + 1)}, \quad x \in \mathbb{R} \quad (3.2.14)$$

1. Show that f is a probability density function.
2. Draw a careful sketch the graph of f , and state the important qualitative features.

Answer

1. Note that $f(x) > 0$ for $x \in \mathbb{R}$. Also

$$\int_{-\infty}^\infty \frac{1}{1+x^2} dx = \arctan x \Big|_{-\infty}^\infty = \pi \quad (3.2.15)$$

and hence f is a PDF.

2. f increases and then decreases, with mode $x = 0$. f is concave upward, then downward, then upward again, with inflection points at $x = \pm \frac{1}{\sqrt{3}}$. f is symmetric about $x = 0$.

The distribution constructed in the previous exercise is known as the (standard) *Cauchy distribution*, named after Augustin Cauchy. It might also be called the *arctangent distribution*, because of the appearance of the arctangent function in the proof that f is a probability density function. In this regard, note the similarity to the arcsine distribution [above](#). The Cauchy distribution is studied in more generality in the chapter on Special Distributions. Note also that the Cauchy distribution is obtained by normalizing the function $x \mapsto \frac{1}{1+x^2}$; the graph of this function is known as the *witch of Agnesi*, in honor of Maria Agnesi.

In the special distribution simulator, select the Cauchy distribution with the default parameter values. Run the simulation 1000 times and compare the empirical density function with the probability density function.

A light source is 1 meter away from position 0 on an infinite, straight wall. The angle Θ that the light beam makes with the perpendicular to the wall is randomly chosen from the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. The position $X = \tan(\Theta)$ of the light beam on the wall has the standard Cauchy distribution. Find each of the following:

1. $\mathbb{P}(-1 < X < 1)$.
2. $\mathbb{P}\left(X \geq \frac{1}{\sqrt{3}}\right)$
3. $\mathbb{P}(X \leq \sqrt{3})$

Answer

1. $\frac{1}{2}$
2. $\frac{1}{3}$
3. $\frac{2}{3}$

The Cauchy experiment (with the default parameter values) is a simulation of the experiment in the last exercise.

1. Run the experiment a few times in single step mode.
2. Run the experiment 1000 times and compare the empirical density function with the probability density function.
3. Using the data from (b), compute the relative frequency of each event in the previous exercise, and compare with the true probability.

The Standard Normal Distribution

Let ϕ be the function defined by $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ for $z \in \mathbb{R}$.

1. Show that ϕ is a probability density function.
2. Draw a careful sketch the graph of ϕ , and state the important qualitative features.

Proof

2. Note that $\phi(z) > 0$ for $z \in \mathbb{R}$. Let $c = \int_{-\infty}^{\infty} e^{-z^2/2} dz$. Then

$$c^2 = \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy \quad (3.2.16)$$

Change to polar coordinates: $x = r \cos \theta$, $y = r \sin \theta$ where $r \in [0, \infty)$ and $\theta \in [0, 2\pi)$. Then $x^2 + y^2 = r^2$ and $dx dy = r dr d\theta$. Hence

$$c^2 = \int_0^{2\pi} \int_0^{\infty} r e^{-r^2/2} dr d\theta \quad (3.2.17)$$

Using the simple substitution $u = r^2$, the inner integral is $\int_0^{\infty} e^{-u} du = 1$. Then the outer integral is $\int_0^{2\pi} 1 d\theta = 2\pi$. Hence $c = \sqrt{2\pi}$ and so ϕ is a PDF.

3. Note that ϕ is symmetric about 0. ϕ increases and then decreases, with mode $z = 0$. ϕ is concave upward, then downward, then upward again, with inflection points at $z = \pm 1$. $\phi(z) \rightarrow 0$ as $z \rightarrow \infty$ and as $z \rightarrow -\infty$.

The distribution defined in the last exercise is the *standard normal distribution*, perhaps the most important distribution in probability and statistics. Its importance stems largely from the central limit theorem, one of the fundamental theorems in probability. In particular, normal distributions are widely used to model physical measurements that are subject to small, random errors. The family of normal distributions is studied in more generality in the chapter on Special Distributions.

In the special distribution simulator, select the normal distribution and keep the default parameter values. Run the simulation 1000 times and compare the empirical density function and the probability density function.

The function $z \mapsto e^{-z^2/2}$ is a notorious example of an integrable function that does not have an antiderivative that can be expressed in closed form in terms of other elementary functions. (That's why we had to resort to the polar coordinate trick to show that ϕ is a probability density function.) So probabilities involving the normal distribution are usually computed using mathematical or statistical software.

Suppose that the error Z in the length of a certain machined part (in millimeters) has the standard normal distribution. Use mathematical software to approximate each of the following:

1. $\mathbb{P}(-1 \leq Z \leq 1)$
2. $\mathbb{P}(Z > 2)$
3. $\mathbb{P}(Z < -3)$

Answer

1. 0.6827
2. 0.0228
3. 0.0013

The Extreme Value Distribution

Let f be the function defined by $f(x) = e^{-x}e^{-e^{-x}}$ for $x \in \mathbb{R}$.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and state the important qualitative features.
3. Find $\mathbb{P}(X > 0)$, where X has probability density function f .

Answer

1. Note that $f(x) > 0$ for $x \in \mathbb{R}$. Using the substitution $u = e^{-x}$,

$$\int_{-\infty}^{\infty} e^{-x} e^{-e^{-x}} dx = \int_0^{\infty} e^{-u} du = 1 \quad (3.2.18)$$

(note that the integrand in the last integral is the exponential PDF with parameter 1.

2. f increases and then decreases, with mode $x = 0$. f is concave upward, then downward, then upward again, with inflection points at $x = \pm \ln[(3 + \sqrt{5})/2]$. Note however that f is not symmetric about 0. $f(x) \rightarrow 0$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$.
3. $1 - e^{-1} \approx 0.6321$

The distribution in the last exercise is the (standard) *type 1 extreme value distribution*, also known as the *Gumbel distribution* in honor of Emil Gumbel. Extreme value distributions are studied in more generality in the chapter on Special Distributions.

In the special distribution simulator, select the extreme value distribution. Keep the default parameter values and note the shape and location of the probability density function. Run the simulation 1000 times and compare the empirical density function with the probability density function.

The Logistic Distribution

Let f be the function defined by

$$f(x) = \frac{e^x}{(1 + e^x)^2}, \quad x \in \mathbb{R} \quad (3.2.19)$$

1. Show that f is a probability density function.
2. Draw a careful sketch the graph of f , and state the important qualitative features.
3. Find $\mathbb{P}(X > 1)$, where X has probability density function f .

Answer

1. Note that $f(x) > 0$ for $x \in \mathbb{R}$. The substitution $u = e^x$ gives

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \frac{1}{(1+u)^2} du = 1 \quad (3.2.20)$$

2. f is symmetric about 0. f increases and then decreases with mode $x = 0$. f is concave upward, then downward, then upward again, with inflection points at $x = \pm \ln(2 + \sqrt{3})$. $f(x) \rightarrow 0$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$.
3. $\frac{1}{1+e} \approx 0.2689$

The distribution in the last exercise is the (standard) *logistic distribution*. Logistic distributions are studied in more generality in the chapter on Special Distributions.

In the special distribution simulator, select the logistic distribution. Keep the default parameter values and note the shape and location of the probability density function. Run the simulation 1000 times and compare the empirical density function with the probability density function.

Weibull Distributions

Let f be the function defined by $f(t) = 2te^{-t^2}$ for $t \in [0, \infty)$.

1. Show that f is a probability density function.
2. Draw a careful sketch the graph of f , and state the important qualitative features.

Answer

1. Note that $f(t) \geq 0$ for $t \geq 0$. The substitution $u = t^2$ gives $\int_0^{\infty} f(t) dt = \int_0^{\infty} e^{-u} du = 1$.
2. f increases and then decreases, with mode $t = 1/\sqrt{2}$. f is concave downward and then upward, with inflection point at $t = \sqrt{3/2}$. $f(t) \rightarrow 0$ as $t \rightarrow \infty$.

Let f be the function defined by $f(t) = 3t^2e^{-t^3}$ for $t \geq 0$.

1. Show that f is a probability density function.
2. Draw a careful sketch the graph of f , and state the important qualitative features.

Answer

1. Note that $f(t) \geq 0$ for $t \geq 0$. The substitution $u = t^3$ gives

$$\int_0^{\infty} f(t) dt = \int_0^{\infty} e^{-u} du = 1 \quad (3.2.21)$$

2. f increases and then decreases, with mode $t = (\frac{2}{3})^{1/3}$. f is concave upward, then downward, then upward again, with inflection points at $t = (1 \pm \frac{1}{3}\sqrt{7})^{1/3}$. $f(t) \rightarrow 0$ as $t \rightarrow \infty$.

The distributions in the last two exercises are examples of *Weibull distributions*, name for Waloddi Weibull. Weibull distributions are studied in more generality in the chapter on Special Distributions. They are often used to model random failure times of devices (in appropriately scaled units).

In the special distribution simulator, select the Weibull distribution. For each of the following values of the shape parameter k , note the shape and location of the probability density function. Run the simulation 1000 times and compare the empirical density function with the probability density function.

1. $k = 2$. This gives the first Weibull distribution [above](#).
2. $k = 3$. This gives the second Weibull distribution [above](#).

Suppose that T is the failure time of a device (in 1000 hour units). Find $\mathbb{P}(T > \frac{1}{2})$ in each of the following cases:

1. T has the first Weibull distribution [above](#).
2. T has the second Weibull distribution [above](#).

Answer

1. $e^{-1/4} \approx 0.7788$
2. $e^{-1/8} \approx 0.8825$

Additional Examples

Let f be the function defined by $f(x) = -\ln x$ for $x \in (0, 1]$.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and state the important qualitative features.
3. Find $\mathbb{P}(\frac{1}{3} \leq X \leq \frac{1}{2})$ where X has the probability density function in (a).

Answer

1. Note that $-\ln x \geq 0$ for $0 < x \leq 1$. Integration by parts with $u = -\ln x$ and $dv = dx$ gives

$$\int_0^1 -\ln x \, dx = -x \ln x \Big|_0^1 + \int_0^1 1 \, dx = 1 \quad (3.2.22)$$

2. f is decreasing and concave upward, with $f(x) \rightarrow \infty$ as $x \downarrow 0$, so there is no mode.
3. $\frac{1}{2} \ln 2 - \frac{1}{3} \ln 3 + \frac{1}{6} \approx 0.147$

Let f be the function defined by $f(x) = 2e^{-x}(1 - e^{-x})$ for $x \in [0, \infty)$.

1. Show that f is a probability density function.
2. Draw a careful sketch of the graph of f , and give the important qualitative features.
3. Find $\mathbb{P}(X \geq 1)$ where X has the probability density function in (a).

Answer

1. Note that $f(x) > 0$ for $0 < x < \infty$. Also, $\int_0^\infty (e^{-x} - e^{-2x}) \, dx = \frac{1}{2}$, so f is a PDF.
2. f increases and then decreases, with mode $x = \ln(2)$. f is concave downward and then upward, with an inflection point at $x = \ln(4)$. $f(x) \rightarrow 0$ as $x \rightarrow \infty$.
3. $2e^{-1} - e^{-2} \approx 0.6004$

The following problems deal with two and three dimensional random vectors having continuous distributions. The idea of [normalizing](#) a function to form a probability density function is important for some of the problems. The relationship between the distribution of a vector and the distribution of its components will be discussed later, in the section on joint distributions.

Let f be the function defined by $f(x, y) = x + y$ for $0 \leq x \leq 1, 0 \leq y \leq 1$.

1. Show that f is a probability density function, and identify the mode.
2. Find $\mathbb{P}(Y \geq X)$ where (X, Y) has the probability density function in (a).
3. Find the conditional density of (X, Y) given $\{X < \frac{1}{2}, Y < \frac{1}{2}\}$.

Answer

1. mode $(1, 1)$
2. $\frac{1}{2}$
3. $f(x, y \mid X < \frac{1}{2}, Y < \frac{1}{2}) = 8(x + y)$ for $0 < x < \frac{1}{2}, 0 < y < \frac{1}{2}$

Let g be the function defined by $g(x, y) = x + y$ for $0 \leq x \leq y \leq 1$.

1. Find the probability density function f that is [proportional](#) to g .
2. Find $\mathbb{P}(Y \geq 2X)$ where (X, Y) has the probability density function in (a).

Answer

1. $f(x, y) = 2(x + y), 0 \leq x \leq y \leq 1$
2. $\frac{5}{12}$

Let g be the function defined by $g(x, y) = x^2y$ for $0 \leq x \leq 1, 0 \leq y \leq 1$.

1. Find the probability density function f that is **proportional** to g .
2. Find $\mathbb{P}(Y \geq X)$ where (X, Y) has the probability density function in (a).

Answer

1. $f(x, y) = 6x^2y$ for $0 \leq x \leq 1, 0 \leq y \leq 1$
2. $\frac{2}{5}$

Let g be the function defined by $g(x, y) = x^2y$ for $0 \leq x \leq y \leq 1$.

1. Find the probability density function f that is **proportional** to g .
2. Find $P(Y \geq 2X)$ where (X, Y) has the probability density function in (a).

Answer

1. $f(x, y) = 15x^2y$ for $0 \leq x \leq y \leq 1$
2. $\frac{1}{8}$

Let g be the function defined by $g(x, y, z) = x + 2y + 3z$ for $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$.

1. Find the probability density function f that is **proportional** to g .
2. Find $\mathbb{P}(X \leq Y \leq Z)$ where (X, Y, Z) has the probability density function in (a).

Answer

1. $f(x, y, z) = \frac{1}{3}(x + 2y + 3z)$ for $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$
2. $\frac{7}{36}$

Let g be the function defined by $g(x, y) = e^{-x}e^{-y}$ for $0 \leq x \leq y < \infty$.

1. Find the probability density function f that is **proportional** to g .
2. Find $\mathbb{P}(X + Y < 1)$ where (X, Y) has the probability density function in (a).

Answer

1. $f(x, y) = 2e^{-x}e^{-y}, 0 < x < y < \infty$
2. $1 - 2e^{-1} \approx 0.2642$

Continuous Uniform Distributions

Our next discussion will focus on an important class of continuous distributions that are defined purely in terms of geometry. We need a preliminary definition.

For $n \in \mathbb{N}_+$, the standard measure λ_n on \mathbb{R}^n is given by

$$\lambda_n(A) = \int_A 1 \, dx, \quad A \subseteq \mathbb{R}^n \quad (3.2.23)$$

In particular, $\lambda_1(A)$ is the *length* of $A \subseteq \mathbb{R}$, $\lambda_2(A)$ is the *area* of $A \subseteq \mathbb{R}^2$, and $\lambda_3(A)$ is the *volume* of $A \subseteq \mathbb{R}^3$.

Details

Technically, λ_n is Lebesgue measure on the σ -algebra of measurable subsets of \mathbb{R}^n . The name is in honor of Henri Lebesgue. The representation above in terms of the standard Riemann integral of calculus works for the sets that occur in typical applications. For the remainder of this discussion, we assume that all subsets of \mathbb{R}^n that are mentioned are measurable

Note that if $n > 1$, the integral above is a multiple integral. Generally, $\lambda_n(A)$ is referred to as the *n-dimensional volume* of $A \subseteq \mathbb{R}^n$.

Suppose that $S \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$ with $0 < \lambda_n(S) < \infty$.

1. the function f defined by $f(x) = 1/\lambda_n(S)$ for $x \in S$ is a probability density function on S .
2. The probability measure associated with f is given by $\mathbb{P}(A) = \lambda_n(A)/\lambda_n(S)$ for $A \subseteq S$, and is known as the *uniform distribution* on S .

Proof

The proof is simple: Clearly $f(x) > 0$ for $x \in S$ and

$$\int_A f(x) dx = \frac{1}{\lambda_n(S)} \int_A 1 dx = \frac{\lambda_n(A)}{\lambda_n(S)}, \quad A \subseteq S \quad (3.2.24)$$

In particular, when $A = S$ we have $\int_S f(x) dx = 1$.

Note that the probability assigned to a set $A \subseteq \mathbb{R}^n$ is proportional to the size of A , as measured by λ_n . Note also that in both the discrete and continuous cases, the uniform distribution on a set S has constant probability density function on S . The uniform distribution on a set S governs a point X chosen “at random” from S , and in the continuous case, such distributions play a fundamental role in various Geometric Models. Uniform distributions are studied in more generality in the chapter on Special Distributions.

The most important special case is the uniform distribution on an interval $[a, b]$ where $a, b \in \mathbb{R}$ and $a < b$. In this case, the probability density function is

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b \quad (3.2.25)$$

This distribution models a point chosen “at random” from the interval. In particular, the uniform distribution on $[0, 1]$ is known as the *standard uniform distribution*, and is very important because of its simplicity and the fact that it can be transformed into a variety of other probability distributions on \mathbb{R} . Almost all computer languages have procedures for simulating independent, standard uniform variables, which are called *random numbers* in this context.

Conditional distributions corresponding to a uniform distribution are also uniform.

Suppose that $R \subseteq S \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$, and that $\lambda_n(R) > 0$ and $\lambda_n(S) < \infty$. If \mathbb{P} is the uniform distribution on S , then the conditional distribution given R is uniform on R .

Proof

The proof is very simple: For $A \subseteq R$,

$$\mathbb{P}(A | R) = \frac{\mathbb{P}(A \cap R)}{\mathbb{P}(R)} = \frac{\mathbb{P}(A)}{\mathbb{P}(R)} = \frac{\lambda_n(A)/\lambda_n(S)}{\lambda_n(R)/\lambda_n(S)} = \frac{\lambda_n(A)}{\lambda_n(R)} \quad (3.2.26)$$

The last theorem has important implications for simulations. If we can simulate a random variable that is uniformly distributed on a set, we can simulate a random variable that is uniformly distributed on a subset.

Suppose again that $R \subseteq S \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$, and that $\lambda_n(R) > 0$ and $\lambda_n(S) < \infty$. Suppose further that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of independent random variables, each uniformly distributed on S . Let $N = \min\{k \in \mathbb{N}_+ : X_k \in R\}$. Then

1. N has the geometric distribution on \mathbb{N}_+ with success parameter $p = \lambda_n(R)/\lambda_n(S)$.
2. X_N is uniformly distributed on R .

Proof

1. Since the variables are uniformly distributed on S , $\mathbb{P}(X_k \in R) = \lambda_n(R)/\lambda_n(S)$ for each $k \in \mathbb{N}_+$. Since the variables are independent, each point is in R or not independently. Hence N , the index of the first point to fall in R , has the geometric distribution on \mathbb{N}_+ with success probability $p = \lambda_n(R)/\lambda_n(S)$. That is, $\mathbb{P}(N = k) = (1-p)^{k-1}p$ for $k \in \mathbb{N}_+$.
2. Note that $p \in (0, 1]$, so $\mathbb{P}(N \in \mathbb{N}_+) = 1$ and hence X_N is well defined. We know from our work on independence and conditional probability that the distribution of X_N is the same as the conditional distribution of X given $X \in R$, which by the previous theorem, is uniformly distributed on R .

Suppose in particular that S is a Cartesian product of n bounded intervals. It turns out to be quite easy to simulate a sequence of independent random variables $\mathbf{X} = (X_1, X_2, \dots)$ each of which is uniformly distributed on S . Thus, the last theorem give an algorithm for simulating a random variable that is uniformly distributed on an irregularly shaped region $R \subseteq S$ (assuming that we have an algorithm for recognizing when a point $x \in \mathbb{R}^n$ falls in R). This method of simulation is known as the *rejection method*, and as we will see in subsequent sections, is more important that might first appear.

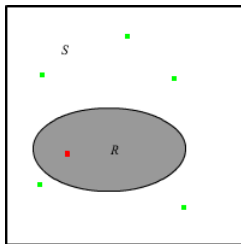


Figure 3.2.3: With a sequence of independent, uniformly distributed points in S , the first one to fall in R is uniformly distributed on R .

In the simple probability experiment, random points are uniformly distributed on the rectangular region S . Move and resize the events A and B and note how the probabilities of the 16 events that can be constructed from A and B change. Run the experiment 1000 times and note the agreement between the relative frequencies of the events and the probabilities of the events.

Suppose that (X, Y) is uniformly distributed on the circular region of radius 5, centered at the origin. We can think of (X, Y) as the position of a dart thrown “randomly” at a target. Let $R = \sqrt{X^2 + Y^2}$, the distance from the center to (X, Y) .

1. Give the probability density function of (X, Y) .
2. Find $\mathbb{P}(n \leq R \leq n+1 \text{ for } n \in \{0, 1, 2, 3, 4\})$.

Answer

1. $f(x, y) = \frac{1}{25\pi}$ for $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 25\}$
2. $\mathbb{P}(n \leq R \leq n+1) = \frac{2n+1}{25}$ for $n \in \{0, 1, 2, 3, 4\}$

Suppose that (X, Y, Z) is uniformly distributed on the cube $S = [0, 1]^3$. Find $\mathbb{P}(X < Y < Z)$ in two ways:

1. Using the probability density function.
2. Using a combinatorial argument.

Answer

1. $\mathbb{P}(X < Y < Z) = \int_0^1 \int_0^z \int_0^y 1 \, dx \, dy \, dz = \frac{1}{6}$
2. Each of the 6 strict orderings of (X, Y, Z) are equally likely, so $\mathbb{P}(X < Y < Z) = \frac{1}{6}$

The time T (in minutes) required to perform a certain job is uniformly distributed over the interval $[15, 60]$

1. Find the probability that the job requires more than 30 minutes
2. Given that the job is not finished after 30 minutes, find the probability that the job will require more than 15 additional minutes.

Answer

1. $\frac{2}{3}$
2. $\frac{1}{6}$

Data Analysis Exercises

If D is a data set from a variable X with a continuous distribution, then an *empirical density function* can be computed by partitioning the data range into subsets of small size, and then computing the probability density of points in each subset. Empirical probability density functions are studied in more detail in the chapter on Random Samples.

For the cicada data, BW denotes body weight (in grams), BL body length (in millimeters), and G gender (0 for female and 1 for male). Construct an empirical density function for each of the following and display each as a bar graph:

1. BW
2. BL
3. BW given $G = 0$

Answer

1.	BW	(0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]
	Density	0.8654	5.8654	3.0769	0.1923
2.	BL	(15, 29]	(20, 25]	(25, 30]	(30, 35]
	Density	0.0058	0.1577	0.0346	0.0019
3.	BW	(0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]
	Density given $G = 0$	0.3390	4.4068	5.0847	0.1695

This page titled [3.2: Continuous Distributions](#) 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.