2. Continuous Distributions

Basic Theory

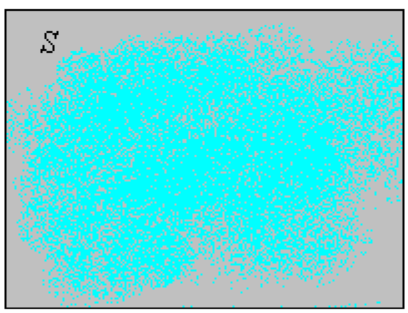
As usual, suppose that we have a random experiment with probability measure $\mathbb{P}$ on an underlying sample space Ω. A random variable X taking values in $S \subseteq \mathbb{R}^n$ is said to have a **continuous distribution** if

$$\mathbb{P}(X = x) = 0 \text{ for each } \quad x \in S$$

The fact that X takes any particular value with probability 0 might seem paradoxical at first, but conceptually it is 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.

1. Show that $\mathbb{P}(X \in C) = 0$ for any countable $C \subseteq S$.

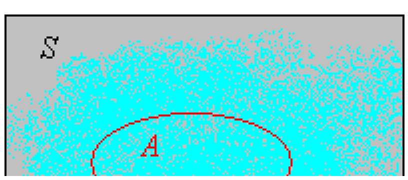
Thus, continuous distributions are in complete contrast with discrete distributions, for which all of the probability mass is concentrated on a discrete set. For a continuous distribution, the probability mass is *continuously* spread over S. Note also that S itself cannot be countable. In the picture below, the light blue shading is intended to suggest a continuous distribution of probability.

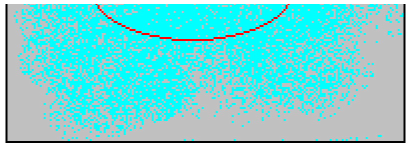


Probability Density Functions

Suppose again that X has a continuous distribution on $S \subseteq \mathbb{R}^n$. A real-valued function f defined on S is said to be a **probability density function** for X if f satisfies the following properties:

a. $f(x) \ge 0, \quad x \in S$
b. $\int_S f(x) dx = 1$
c. $\int_A f(x) dx = \mathbb{P}(X \in A), \quad A \subseteq S$





Property (c) in the definition is particularly important since it implies that the probability distribution of X is completely determined by the density function. Conversely, any function that satisfies properties (a) and (b) is a probability density function, and then property (c) can be used to define a continuous distribution on S.

If $n > 1$, the integrals in properties (b) and (c) are multiple integrals over subsets of $\mathbb{R}^n$ with $x = (x_1, x_2, \ldots, x_n)$ and $dx = dx_1\, dx_2 \cdots dx_n$. In fact, technically, f is a probability density function **relative to** the standard **n-dimensional measure**, which we recall is given by

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

Note that $\lambda_n(S)$ must be positive (perhaps infinite). In particular,

1. if $n = 1$, S must be a subset of $\mathbb{R}$ with positive length;
2. if $n = 2$, S must be a subset of $\mathbb{R}^2$ with positive area;
3. if $n = 3$, S must be a subset of $\mathbb{R}^3$ with positive volume.

However, we recall that except for exposition, the low dimensional cases ($n \in \{1, 2, 3\}$) play no special role in probability. Interesting random experiments often involve *several* random variables (that is, a random vector). Finally, note that we can always extend f to a probability density function on all of $\mathbb{R}^n$ by defining $f(x) = 0, \; x \notin S$. This extension sometimes simplifies notation.

An element $x \in S$ that maximizes the density f is called a **mode** of the distribution. If there is only one mode, it is sometimes used as a measure of the *center* of the distribution.

Probability density functions of continuous distributions differ from their discrete counterparts in several important ways:

- A probability density function f for a continuous distribution is not unique. Note that the values of f on a finite (or even countable) set of points could be changed to other nonnegative values, and properties (a), (b), and (c) would still hold. The critical fact is that only *integrals* of f are important.
- Note that it is possible to have $f(x) > 1$ for some (or even all) $x \in S$. Indeed, f can be unbounded on S.
- Finally, note that $f(x)$ is not a probability; it is the probability *density* at x. That is, $f(x)\,dx$ is approximately the probability that X is in an n-dimensional box at $x = (x_1, x_2, \ldots, x_n)$ with side lengths $(dx_1, dx_2, \ldots, dx_n)$, if these side lengths are small.

Constructing Densities

2. Suppose that g is a nonnegative function on $S \subseteq \mathbb{R}^n$. Let

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

Show that if $0 < c < \infty$ then $f(x) = \frac{1}{c} g(x), \; x \in S$ defines a probability density function on S.

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

Conditional Densities

Suppose that X is a random variable taking values in $S \subseteq \mathbb{R}^n$ with a continuous distribution that has density function f. The probability density function of X, of course, is based on the underlying probability measure $\mathbb{P}$ on the sample space Ω. This measure could be a conditional probability measure, conditioned on a given event $E \subseteq \Omega$ (with $\mathbb{P}(E) > 0$ of course). The usual notation is

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

Note, however, that except for notation, no new concepts are involved. The function above is a probability density function for a continuous distribution. That is, it satisfies properties (a) and (b) while property (c) becomes

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

.

All results that hold for probability density functions in general have analogies for conditional probability density functions.

3. Suppose that $B \subseteq S$ with $\mathbb{P}(X \in B) = \int_B f(x)dx > 0$. Show that the conditional density of X given $X \in B$ is

$$f(x|X \in B) = \begin{cases} \dfrac{f(x)}{\mathbb{P}(X \in B)}, & x \in B \\ 0, & x \in B^c \end{cases}$$

Examples and Applications

The Exponential Distribution

4. Let $f(t) = r\,e^{-r\,t}, \ t \ge 0$ where $r > 0$ is a parameter.

a. Show that f is a probability density function.
b. Sketch the graph of f and identify the mode.

The distribution defined by the 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. The exponential distribution is studied in detail in the chapter on Poisson Processes.

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

6. In the exponential experiment, set $r = \frac{1}{2}$ Run the simulation 1000 times, updating every 10 runs, and note the apparent convergence of the empirical density function to the true probability density function.

A Random Angle

7. In Bertrand's problem, a certain random angle Θ has probability density function $f(\theta) = \sin(\theta), \ 0 \le \theta \le \frac{\pi}{2}$.

a. Show that f is a probability density function.
b. Graph f and identify the mode.
c. Find $\mathbb{P}\left(\Theta < \frac{\pi}{4}\right)$.

8. In Bertrand's experiment, select the model with uniform distance. Run the simulation 200 times, updating every run, and compute the empirical probability of the event $\left\{\Theta < \frac{\pi}{4}\right\}$. Compare with the true probability in the previous exercise.

The Gamma Distribution

9. Let $g_n(t) = e^{-t} \frac{t^n}{n!}, \ t \ge 0$ where $n \in \mathbb{N}$ is a parameter.

a. Show that g_n is a probability density function for each n. *Hint:* Use induction on n.
b. Show that $g_n(t)$ is strictly increasing for $t < n$ and strictly decreasing for $t > n$, so that the mode occurs at $t = n$.
c. Graph g_n.

Remarkably, we showed in the last section on discrete distributions, that $f_t(n) = g_n(t)$ is a density function on $\mathbb{N}$ for each $t > 0$. The distribution defined by the density g_n is the **gamma distribution**; $n + 1$ is called the **shape parameter**. The gamma distribution is studied in detail in the chapter on Poisson Processes. Note that the special case $n = 0$ gives the exponential distribution with rate parameter 1.

10. Suppose that the lifetime of a device T (in 1000 hour units) has the gamma distribution with $n = 2$. Find $\mathbb{P}(T > 3)$.

11. In the gamma experiment, set $r = 1$ and $k = 3$. Run the experiment 200 times, updating every run. Compute the empirical probability of the event $\{T > 3\}$ and compare with the theoretical probability in the previous exercise.

Beta Distributions

12. Let $g(x) = x^2 (1 - x), \ 0 \le x \le 1$

a. Sketch the graph of g.
b. Find the probability density function f proportional to g.
c. Find the mode.

d. Find $\mathbb{P}\left(\frac{1}{2} < X < 1\right)$ where X is a random variable with the density in (b).

13. Let $g(x) = \frac{1}{\sqrt{x(1-x)}}$, $0 < x < 1$

a. Sketch the graph of g.
b. Find the probability density function f proportional to g.
c. Find $\mathbb{P}\left(0 < X < \frac{1}{4}\right)$ where X is a random variable with the density in (b).

Hint: In the integrals, first use the simple substitution $u = \sqrt{x}$ and then recognize the new integral as an arcsine integral.

The distributions defined in the last two exercises are examples of **beta distributions**. The particular beta distribution in Exercise 13 is also known as the **arcsine** distribution. Beta distributions are studied in detail in the chapter on Special Distributions.

14. In the random variable experiment, select the beta distribution. For the following parameter values, note the shape and location of the probability density function. Run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function to the true probability density function.

a. $a = 3$, $b = 2$. This gives the beta distribution in Exercise 12.
b. $a = \frac{1}{2}$, $b = \frac{1}{2}$. This gives the arcsine distribution in Exercise 13.

The Pareto Distribution

15. Let $g(x) = \frac{1}{x^b}$, $x > 1$ where $b > 0$ is a parameter.

a. Sketch the graph of g.
b. Show that for $0 < b \le 1$, there is no probability density function proportional to g.
c. Show that for $b > 1$, the normalizing constant is $\frac{1}{b-1}$.

The distribution defined in the last exercise 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**. The Pareto distribution is studied in detail in the chapter on Special Distributions.

16. Suppose that the income X (in appropriate units) of a person randomly selected from a population has the Pareto distribution with shape parameter 3. Find $\mathbb{P}(X > 2)$.

17. In the random variable experiment, select the Pareto with $a = 3$. Run the simulation 1000 times, updating every

10 runs. Compute the empirical probability of the event $\{X > 2\}$ and compare with the theoretical probability in the last exercise.

The Cauchy Distribution

18. Let $g(x) = \frac{1}{x^2 + 1}$, $x \in \mathbb{R}$.

a. Sketch the graph of g.
b. Show that the normalizing constant is π.
c. Find $\mathbb{P}(-1 < X < 1)$ where X has the density function proportional to g.

The distribution constructed in the previous exercise is known as the **Cauchy distribution**, named after Augustin Cauchy (it might also be called the **arctangent distribution**). It is a member of the **Student** t family of distributions; the t distributions are studied in detail in the chapter on Special Distributions. The graph of g is known as the **witch of Agnesi**, in honor of Maria Agnesi.

19. In the random variable experiment, select the student t distribution. Set $n = 1$ to get the Cauchy distribution. Run the simulation 1000 times, updating every 10 runs. Note how well the empirical density function fits the true probability density function.

The Standard Normal Distribution

20. Let $g(z) = e^{-\frac{1}{2} z^2}$, $z \in \mathbb{R}$.

a. Sketch the graph of g.
b. Show that the normalizing constant is $\sqrt{2\pi}$. *Hint*: If c denotes the normalizing constant, express c^2 as a double integral and convert to polar coordinates.

The distribution defined in the last exercise is the **standard normal distribution**, perhaps the most important distribution in probability. Normal distributions are widely used to model physical measurements that are subject to small, random errors. The family of normal distributions is studied in detail in the chapter on Special Distributions.

21. In the random variable experiment, select the normal distribution (the default parameters give the standard normal distribution). Run the simulation 1000 times, updating every 10 runs. Note how well the empirical density function fits the true probability density function.

The Extreme Value Distribution

22. Let $f(x) = e^{-x} e^{-e^{-x}}$ for $x \in \mathbb{R}$.

a. Show that f is a probability density function.
b. Sketch the graph of f and identify the mode. Note in particular the asymmetry of the graph.

c. Find $\mathbb{P}(X > 0)$, where X has probability density function f.

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

23. In the random variable experiment, select the extreme value distribution. Note the shape and location of the density function. Run the simulation 1000 times, updating every 10 runs. Note how well the empirical density function fits the true probability density function.

Additional Examples

24. Let $f(x) = -\ln(x),\ 0 < x < 1$

a. Sketch the graph of f
b. Show that f is a probability density function.
c. Find $\mathbb{P}\left(\frac{1}{3} \le X \le \frac{1}{2}\right)$ where X has the probability density function in (b).

25. Let $g(x) = e^{-x}(1 - e^{-x}),\ \ x \ge 0$.

a. Sketch the graph of g
b. Find the probability density function f proportional to g and identify the mode.
c. Find $\mathbb{P}(X \ge 1)$ where X has the probability density function in (b).

26. Let $f(x, y) = x + y,\ 0 \le x \le 1,\ 0 \le y \le 1$.

a. Show that f is a probability density function
b. Find $\mathbb{P}(Y \ge X)$ where (X, Y) has the probability density function in (a).
c. Find the conditional density of (X, Y) given $\left\{X < \frac{1}{2},\ Y < \frac{1}{2}\right\}$.

27. Let $g(x, y) = x + y,\ 0 \le x \le y \le 1$.

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

28. Let $g(x, y) = x^2 y,\ 0 \le x \le 1,\ 0 \le y \le 1$.

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

29. Let $g(x, y) = x^2 y, \; 0 \le x \le y \le 1$.

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

30. Let $g(x, y, z) = x + 2y + 3y, \; 0 \le x \le 1, \; 0 \le y \le 1, \; 0 \le z \le 1$.

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

31. Let $g(x, y) = e^{-x} e^{-y}, \; 0 < x < y < \infty$.

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

Continuous Uniform Distributions

The following exercises describe an important class of continuous distributions.

32. Suppose that $S \subseteq \mathbb{R}^n$ with $0 < \lambda_n(S) < \infty$. Show that

a. $f(x) = \frac{1}{\lambda_n(S)}, \; x \in S$ defines a probability density function on S.
b. If X has the density in (a) then $\mathbb{P}(X \in A) = \frac{\lambda_n(A)}{\lambda_n(S)}, \; A \subseteq S$.

A random variable X with the density function in Exercise 32 is said to have the **continuous uniform distribution** on S. From part (b), note that the probability assigned to a subset A of S is proportional to the standard measure of A. Note also that in both the discrete and continuous cases, a random variable X is uniformly distributed on a set S if and only if the density function is constant on S. Uniform distributions on rectangles in the plane play a fundamental role in Geometric Models.

33. Find $\mathbb{P}(X > 0, Y > 0)$ in each of the following cases:

a. (X, Y) is uniformly distributed on the square $S = [-6, 6]^2$.
b. (X, Y) is uniformly distributed on the triangle $S = \{(x, y) : -6 \le y \le x \le 6\}$.
c. (X, Y) is uniformly distributed on the circle $S = \{(x, y) : x^2 + y^2 \le 36\}$.

34. In the bivariate uniform experiment, select each of the following domains and then run the simulation 100 times,

updating every run. Watch the points in the scatter plot. Compute the empirical probability of the event $\{X > 0, Y > 0\}$ and compare with the true probability in the previous exercise.

a. Square
b. Triangle
c. Circle

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

a. Compute the probability using the probability density function.
b. Compute the probability using a combinatorial argument. *Hint*: Argue that each of the 6 orderings of (X, Y, Z) should be equally likely.

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

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

37. Suppose that $S \subseteq \mathbb{R}^n$ and that $0 < \lambda_n(S) < \infty$ and that $B \subseteq S$ with $\lambda_n(B) > 0$. Show that if X is uniformly distributed on S, then the conditional distribution of X given $X \in B$ is uniformly distributed on B.

Data Analysis Exercises

If $D \subseteq \mathbb{R}^n$ 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 density of points in each subset. Empirical density functions are studied in more detail in the chapter on Random Samples.

38. For the cicada data, W_B denotes body weight, L_B body length, and G gender. Construct an empirical density function for each of the following and display each as a bar graph:

a. W_B
b. L_B
c. W_B given $G =$ female

39. For the cicada data, L_W denotes wing length and W_W wing width. Construct an empirical density function for (L_W, W_W).

Degenerate Continuous Distributions

Unlike the discrete case, the existence of a density function for a continuous distribution is an *assumption* that we are making. A random variable can have a continuous distribution on a subset $S \subseteq \mathbb{R}^n$ but with no probability density function; the distribution is sometimes said to be **degenerate**. In this subsection, we explore the common ways in which such distributions can occur.

Reducing the Dimension

First, suppose that X is a random variable taking values in $S \subseteq \mathbb{R}^n$ with $\lambda_n(S) = 0$ It is possible for X to have a continuous distribution, but X could not have a probability density function relative to λ_n. In particular, property (c) in the definition could not hold, since $\int_A f(x)dx = 0$ for any function f and any $A \subseteq S$. However, in many cases, X may be defined in terms of continuous random variables on lower dimensional spaces that *do* have probability density functions.

For example, suppose that U is a random variable with a continuous distribution on $T \subseteq \mathbb{R}^k$ where $k < n$, and that $X = h(U)$ for some continuous function h from T into $\mathbb{R}^n$ Any event defined in terms of X can be changed into an event defined in terms of U. The following exercise illustrates this situation

40. Suppose that Θ is uniformly distributed on the interval $[0, 2\pi)$. Let $X = \cos(\Theta)$, $Y = \sin(\Theta)$.

a. Show that (X, Y) has a continuous distribution on the circle $C = \{(x, y) : x^2 + y^2 = 1\}$
b. Show that (X, Y) does not have a density function on C (with respect to the area measure λ_2 on $\mathbb{R}^2$).
c. Find $\mathbb{P}(Y > X)$.

Mixed Components

Another situation occurs when a random vector X in $\mathbb{R}^n$ (with $n > 1$) has some components with discrete distributions and others with continuous distributions. Such distributions with **mixed components** are studied in more detail in the section on mixed distributions; however, the following exercise gives an illustration.

41. Suppose that X is uniformly distributed on the set $\{0, 1, 2\}$, Y is uniformly distributed on the interval $[0, 2]$, and that X and Y are independent.

a. Show that (X, Y) has a continuous distribution on the product set $S = \{0, 1, 2\} \times [0, 2]$
b. Show that (X, Y) does not have a density function on S (with respect to the area measure λ_2 on $\mathbb{R}^2$).
c. Find $\mathbb{P}(Y > X)$.

Singular Continuous Distributions

Finally, it is also possible to have a continuous distribution on $S \subseteq \mathbb{R}^n$ with $\lambda_n(S) > 0$, yet still with no density function. Such distributions are said to be **singular**, and are rare in applied probability. However, it is not difficult to construct such a distribution. Let $(X_1, X_2, \ldots)$ be a sequence of Bernoulli trials with success parameter p. We will indicate the dependence of the probability measure $\mathbb{P}$ on the parameter p with a subscript. Thus, we have a sequence of independent indicator variables with

$$\mathbb{P}_p(X_i = 1) = p, \quad \mathbb{P}_p(X_i = 0) = 1 - p$$

We interpret X_i as the i^{th} binary digit (**bit**) of a random variable X taking values in $[0, 1]$ That is,

$$X = \sum_{i=1}^\infty \frac{X_i}{2^i}$$

Conversely, recall that every number $x \in [0, 1]$ can be written in binary form:

$$x = \sum_{i=1}^\infty \frac{x_i}{2^i} \quad \text{where} \quad x_i \in \{0, 1\} \quad \text{for each} \quad i \in \{1, 2, ...\}$$

This representation is unique except when x is a **binary rational** of the form $x = \frac{k}{2^n}$ for some $k \in \{1, 3, ..., 2^n - 1\}$. In this case, there are two representations, one in which the bits are eventually 0 and one in which the bits are eventually 1. By convention, we will use the first representation. Note, however, that the set of binary rationals is countable.

42. Show that X has a continuous distribution. That is, $\mathbb{P}_p(X = x) = 0$ for $x \in [0, 1]$.

Next, we define the set of numbers for which the limiting relative frequency of 1's is p. Let

.

$$C_p = \left\{ x \in [0, 1] : \frac{1}{n} \sum_{i=1}^n x_i \to p \text{ as } n \to \infty \right\}$$

43. Show that

a. $C_p \cap C_q = \emptyset$ for $p \neq q$.
b. $\mathbb{P}_p(X \in C_p) = 1$, *Hint:* Use the strong law of large numbers.

It follows that the distributions of X, as p varies from 0 to 1, are **mutually singular**.

44. Suppose that $p = \frac{1}{2}$. Show that X is uniformly distributed on $[0, 1]$. That is, the distribution of X in this case is the standard (Lebesgue) measure on $[0, 1]$.

From the last two exercises, it follows that when $p \neq \frac{1}{2}$, the distribution of X does not have a density relative to the standard measure on $[0, 1]$. For an applied example, see Bold Play in the game of Red and Black.