

4.2: Continuous Conditional Probability

In situations where the sample space is continuous we will follow the same procedure as in the previous section. Thus, for example, if X is a continuous random variable with density function $f(x)$, and if E is an event with positive probability, we define a conditional density function by the formula

$$f(x|E) = \begin{cases} f(x)/P(E), & \text{if } x \in E, \\ 0, & \text{if } x \notin E. \end{cases} \quad (4.2.1)$$

Then for any event F , we have

$$P(F|E) = \int_F f(x|E) dx. \quad (4.2.2)$$

The expression $P(F|E)$ is called the conditional probability of F given E . As in the previous section, it is easy to obtain an alternative expression for this probability:

$$P(F|E) = \int_F f(x|E) dx = \int_{E \cap F} \frac{f(x)}{P(E)} dx = \frac{P(E \cap F)}{P(E)}. \quad (4.2.3)$$

We can think of the conditional density function as being 0 except on E , and normalized to have integral 1 over E . Note that if the original density is a uniform density corresponding to an experiment in which all events of equal size are then the same will be true for the conditional density.

✓ Example 4.2.1:

In the spinner experiment (cf. Example 2.1.1), suppose we know that the spinner has stopped with head in the upper half of the circle, $0 \leq x \leq 1/2$. What is the probability that $1/6 \leq x \leq 1/3$?

Solution

Here $E = [0, 1/2]$, $F = [1/6, 1/3]$, and $F \cap E = F$. Hence

$$\begin{aligned} P(F|E) &= \frac{P(F \cap E)}{P(E)} \\ &= \frac{1/6}{1/2} \\ &= \frac{1}{3}, \end{aligned}$$

which is reasonable, since F is $1/3$ the size of E . The conditional density function here is given by

$$f(x|E) = \begin{cases} 2, & \text{if } 0 \leq x < 1/2, \\ 0, & \text{if } 1/2 \leq x < 1. \end{cases} \quad (4.2.4)$$

Thus the conditional density function is nonzero only on $[0, 1/2]$ and is uniform there.

✓ Example 4.2.2:

In the dart game (cf. Example 2.2.2), suppose we know that the dart lands in the upper half of the target. What is the probability that its distance from the center is less than $1/2$?

Solution

Here $E = \{(x, y) : y \geq 0\}$, and $F = \{(x, y) : x^2 + y^2 < (1/2)^2\}$. Hence,

$$\begin{aligned} P(F|E) &= \frac{P(F \cap E)}{P(E)} = \frac{(1/\pi)[(1/2)(\pi/4)]}{(1/\pi)(\pi/2)} \\ &= \frac{1}{4}. \end{aligned}$$

Here again, the size of $F \cap E$ is $1/4$ the size of E . The conditional density function is

$$f((x, y)|E) = \begin{cases} f(x, y)/P(E) = 2/\pi, & \text{if } (x, y) \in E, \\ 0, & \text{if } (x, y) \notin E. \end{cases} \quad (4.2.5)$$

✓ Example 4.2.3:

We return to the exponential density (cf. Example 2.2.7). We suppose that we are observing a lump of plutonium-239. Our experiment consists of waiting for an emission, then starting a clock, and recording the length of time X that passes until the next emission. Experience has shown that X has an exponential density with some parameter λ , which depends upon the size of the lump. Suppose that when we perform this experiment, we notice that the clock reads r seconds, and is still running. What is the probability that there is no emission in a further s seconds?

Solution

Let $G(t)$ be the probability that the next particle is emitted after time t . Then

$$\begin{aligned} G(t) &= \int_t^{\infty} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_t^{\infty} = e^{-\lambda t}. \end{aligned}$$

Let E be the event “the next particle is emitted after time r ” and F the event “the next particle is emitted after time $r + s$.” Then

$$\begin{aligned} P(F|E) &= \frac{P(F \cap E)}{P(E)} \\ &= \frac{G(r+s)}{G(r)} \\ &= \frac{e^{-\lambda(r+s)}}{e^{-\lambda r}} \\ &= e^{-\lambda s}. \end{aligned}$$

This tells us the rather surprising fact that the probability that we have to wait s seconds more for an emission, given that there has been no emission in r seconds, is of the time r . This property (called the *memoryless* property) was introduced in Example 2.17. When trying to model various phenomena, this property is helpful in deciding whether the exponential density is appropriate.

The fact that the exponential density is memoryless means that it is reasonable to assume if one comes upon a lump of a radioactive isotope at some random time, then the amount of time until the next emission has an exponential density with the same parameter as the time between emissions. A well-known example, known as the “bus paradox,” replaces the emissions by buses. The apparent paradox arises from the following two facts: 1) If you know that, on the average, the buses come by every 30 minutes, then if you come to the bus stop at a random time, you should only have to wait, on the average, for 15 minutes for a bus, and 2) Since the buses arrival times are being modelled by the exponential density, then no matter when you arrive, you will have to wait, on the average, for 30 minutes for a bus.

The reader can now see that in Exercises 2.2.9, 2.2.10, and 2.2.11, we were asking for simulations of conditional probabilities, under various assumptions on the distribution of the interarrival times. If one makes a reasonable assumption about this distribution, such as the one in Exercise 2.2.10, then the average waiting time is more nearly one-half the average interarrival time.

Independent Events

If E and F are two events with positive probability in a continuous sample space, then, as in the case of discrete sample spaces, we define E and F to be *independent* if $P(E|F) = P(E)$ and $P(F|E) = P(F)$. As before, each of the above equations imply the other, so that to see whether two events are independent, only one of these equations must be checked. It is also the case that, if E and F are independent, then $P(E \cap F) = P(E)P(F)$.

✓ Example 4.2.1:

In the dart game (see Example 4.1.12, let E be the event that the dart lands in the *upper* half of the target ($y \geq 0$) and F the event that the dart lands in the *right* half of the target ($x \geq 0$). Then $P(E \cap F)$ is the probability that the dart lies in the first quadrant of the target, and

$$\begin{aligned} P(E \cap F) &= \frac{1}{\pi} \int_{E \cap F} 1 \, dx \, dy \\ &= \frac{\text{Area}(E \cap F)}{\text{Area}(E) \text{Area}(F)} \\ &= \left(\frac{1}{\pi} \int_E 1 \, dx \, dy \right) \left(\frac{1}{\pi} \int_F 1 \, dx \, dy \right) \\ &= P(E)P(F) \end{aligned}$$

so that E and F are independent. What makes this work is that the events E and F are described by restricting different coordinates. This idea is made more precise below.

Joint Density and Cumulative Distribution Functions

In a manner analogous with discrete random variables, we can define joint density functions and cumulative distribution functions for multi-dimensional continuous random variables.

Definition 4.2.5

Let X_1, X_2, \dots, X_n be continuous random variables associated with an experiment, and let $\bar{X} = (X_1, X_2, \dots, X_n)$. Then the joint cumulative distribution function of \bar{X} is defined by

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n). \quad (4.2.6)$$

The joint density function of \bar{X} satisfies the following equation:

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) \, dt_n \, dt_{n-1} \cdots dt_1. \quad (4.2.7)$$

It is straightforward to show that, in the above notation,

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} \quad (4.2.8)$$

Independent Random Variables

As with discrete random variables, we can define mutual independence of continuous random variables.

Definition 4.2.6

Let X_1, X_2, \dots, X_n be continuous random variables with cumulative distribution functions $F_1(x), F_2(x), \dots, F_n(x)$. Then these random variables are if

$$F(x_1, x_2, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n) \quad (4.2.9)$$

for any choice of x_1, x_2, \dots, x_n .

Thus, if X_1, X_2, \dots, X_n are mutually independent, then the joint cumulative distribution function of the random variable $\bar{X} = (X_1, X_2, \dots, X_n)$ is just the product of the individual cumulative distribution functions. When two random variables are mutually independent, we shall say more briefly that they are

Using Equation 4.4, the following theorem can easily be shown to hold for mutually independent continuous random variables.

Theorem 4.2.2

Let X_1, X_2, \dots, X_n be continuous random variables with density functions $f_1(x), f_2(x), \dots, f_n(x)$. Then these random variables are *mutually independent* if and only if

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n) \quad (4.2.10)$$

for any choice of x_1, x_2, \dots, x_n

Let's look at some examples.

Example 4.2.5:

In this example, we define three random variables, X_1 , X_2 , and X_3 . We will show that X_1 and X_2 are independent, and that X_1 and X_3 are not independent. Choose a point $\omega = (\omega_1, \omega_2)$ at random from the unit square. Set $X_1 = \omega_1^2$, $X_2 = \omega_2^2$, and $X_3 = \omega_1 + \omega_2$. Find the joint distributions $F_{12}(r_1, r_2)$ and $F_{23}(r_2, r_3)$.

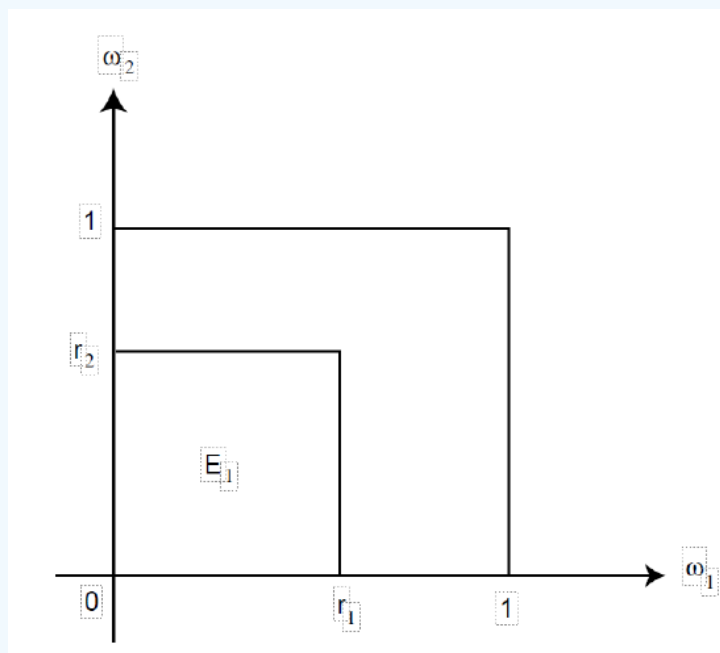


Figure 4.2.1: X_1 and X_2 are independent.

We have already seen (see Example 2.13 that

$$\begin{aligned} F_1(r_1) &= P(-\infty < X_1 \leq r_1) \\ &= \sqrt{r_1}, \quad \text{if } 0 \leq r_1 \leq 1, \end{aligned}$$

and similarly,

$$F_2(r_2) = \sqrt{r_2}, \quad (4.2.11)$$

if $0 \leq r_2 \leq 1$. Now we have (see Figure 4.2.1)

$$\begin{aligned} F_{12}(r_1, r_2) &= P(X_1 \leq r_1 \text{ and } X_2 \leq r_2) \\ &= P(\omega_1 \leq \sqrt{r_1} \text{ and } \omega_2 \leq \sqrt{r_2}) \\ &= \text{Area}(E_1) \\ &= \sqrt{r_1}\sqrt{r_2} \\ &= F_1(r_1)F_2(r_2). \end{aligned}$$

In this case $F_{12}(r_1, r_2) = F_1(r_1)F_2(r_2)$ so that X_1 and X_2 are independent. On the other hand, if $r_1 = 1/4$ and $r_3 = 1$, then (see Figure 4.2.2)

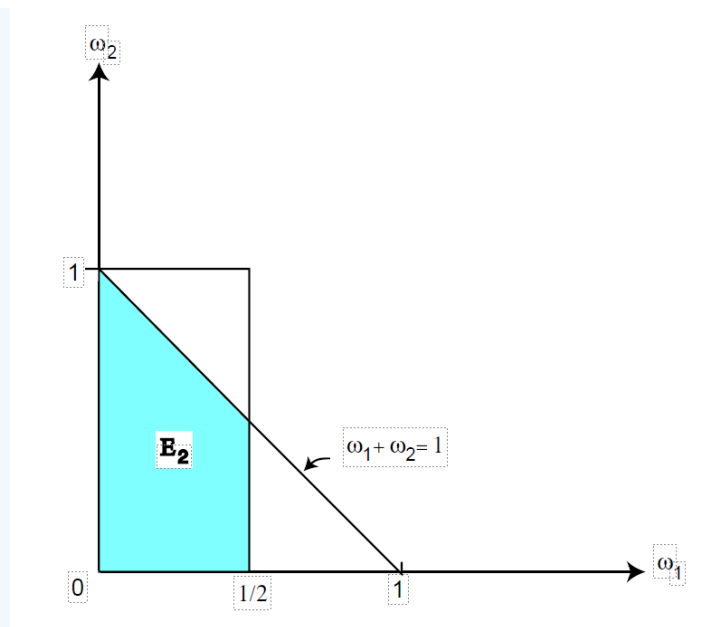


Figure 4.2.2: X_1 and X_3 are not independent

$$\begin{aligned}
 F_{13}(1/4, 1) &= P(X_1 \leq 1/4, X_3 \leq 1) \\
 &= P(\omega_1 \leq 1/2, \omega_1 + \omega_2 \leq 1) \\
 &= \text{Area}(E_2) \\
 &= \frac{1}{2} - \frac{1}{8} = \frac{3}{8}.
 \end{aligned}$$

Now recalling that

$$F_3(r_3) = \begin{cases} 0, & \text{if } r_3 < 0, \\ (1/2)r_3^2, & \text{if } 0 \leq r_3 \leq 1, \\ 1 - (1/2)(2 - r_3)^2, & \text{if } 1 \leq r_3 \leq 2, \\ 1, & \text{if } 2 < r_3, \end{cases} \quad (4.2.12)$$

(see Example 2.14, we have $F_1(1/4)F_3(1) = (1/2)(1/2) = 1/4$. Hence, X_1 and X_3 are not independent random variables. A similar calculation shows that X_2 and X_3 are not independent either.

Although we shall not prove it here, the following theorem is a useful one. The statement also holds for mutually independent discrete random variables. A proof may be found in Rényi.¹⁷

Theorem 4.2.1

Let X_1, X_2, \dots, X_n be mutually independent continuous random variables and let $\phi_1(x), \phi_2(x), \dots, \phi_n(x)$ be continuous functions. Then $\phi_1(X_1), \phi_2(X_2), \dots, \phi_n(X_n)$ are mutually independent.

Independent Trials

Using the notion of independence, we can now formulate for continuous sample spaces the notion of independent trials (see Definition 4.5).

Definition

A sequence X_1, X_2, \dots, X_n of random variables X_i that are mutually independent and have the same density is called an *independent trials process*.

As in the case of discrete random variables, these independent trials processes arise naturally in situations where an experiment described by a single random variable is repeated n times.

Beta Density

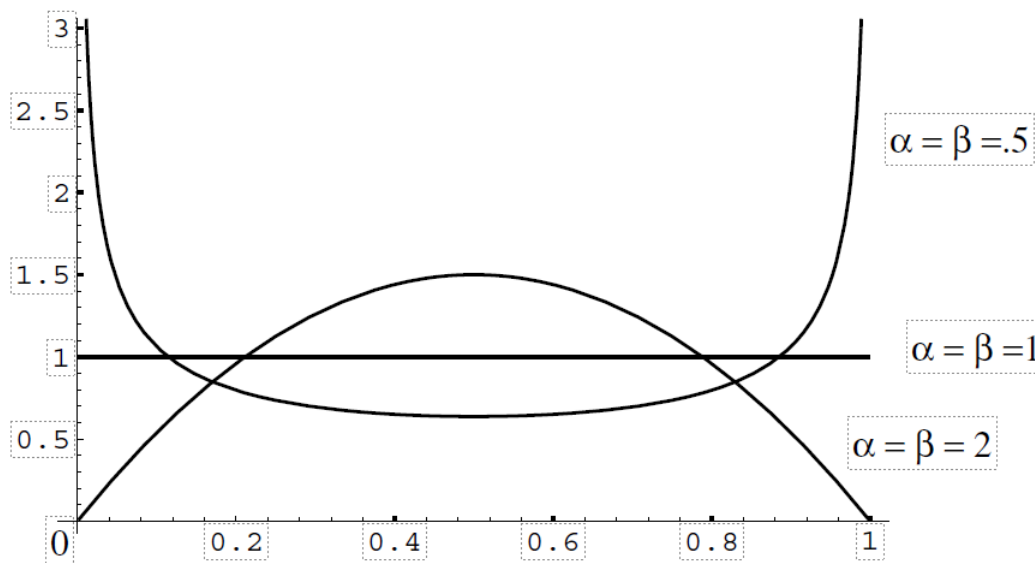


Figure 4.2.3: Beta density for $\alpha = \beta = .5, 1, 2$

We consider next an example which involves a sample space with both discrete and continuous coordinates. For this example we shall need a new density function called the *beta density*. This density has two parameters α , β and is defined by

$$B(\alpha, \beta, x) = \begin{cases} (1/B(\alpha, \beta))x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (4.2.13)$$

Here α and β are any positive numbers, and the beta function $B(\alpha, \beta)$ is given by the area under the graph of $x^{\alpha-1}(1-x)^{\beta-1}$ between 0 and 1:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx. \quad (4.2.14)$$

Note that when $\alpha = \beta = 1$ the beta density is the uniform density. When α and β are greater than 1 the density is bell-shaped, but when they are less than 1 it is U-shaped as suggested by the examples in Figure 4.9.

We shall need the values of the beta function only for integer values of α and β , and in this case

$$B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}. \quad (4.2.15)$$

Example 4.2.23

In medical problems it is often assumed that a drug is effective with a probability x each time it is used and the various trials are independent, so that one is, in effect, tossing a biased coin with probability x for heads. Before further experimentation, you do not know the value x but past experience might give some information about its possible values. It is natural to represent this information by sketching a density function to determine a distribution for x . Thus, we are considering x to be a continuous random variable, which takes on values between 0 and 1. If you have no knowledge at all, you would sketch the uniform density. If past experience suggests that x is very likely to be near $2/3$ you would sketch a density with maximum at $2/3$ and a spread reflecting your uncertainty in the estimate of $2/3$. You would then want to find a density function that reasonably fits your sketch. The beta densities provide a class of densities that can be fit to most sketches you might make. For example, for $\alpha > 1$ and $\beta > 1$ it is bell-shaped with the parameters α and β determining its peak and its spread.

Assume that the experimenter has chosen a beta density to describe the state of his knowledge about x before the experiment. Then he gives the drug to n subjects and records the number i of successes. The number i is a discrete random variable, so we may

conveniently describe the set of possible outcomes of this experiment by referring to the ordered pair (x, i) .

We let $m(i|x)$ denote the probability that we observe i successes given the value of x . By our assumptions, $m(i|x)$ is the binomial distribution with probability x for success:

$$m(i|x) = b(n, x, i) = \binom{n}{i} x^i (1-x)^j, \quad (4.2.16)$$

where $j = n - i$.

If x is chosen at random from $[0, 1]$ with a beta density $B(\alpha, \beta, x)$, then the density function for the outcome of the pair (x, i) is

$$\begin{aligned} f(x, i) &= m(i|x) B(\alpha, \beta, x) \\ &= \binom{n}{i} x^i (1-x)^j \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \\ &= \binom{n}{i} \frac{1}{B(\alpha, \beta)} x^{\alpha+i-1} (1-x)^{\beta+j-1}. \end{aligned}$$

Now let $m(i)$ be the probability that we observe i successes knowing the value of x . Then

$$\begin{aligned} m(i) &= \int_0^1 m(i|x) B(\alpha, \beta, x) dx \\ &= \binom{n}{i} \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+i-1} (1-x)^{\beta+j-1} dx \\ &= \binom{n}{i} \frac{B(\alpha+i, \beta+j)}{B(\alpha, \beta)}. \end{aligned}$$

Hence, the probability density $f(x|i)$ for x , given that i successes were observed, is

$$f(x|i) = \frac{f(x, i)}{m(i)} \quad (4.2.17)$$

$$= \frac{x^{\alpha+i-1} (1-x)^{\beta+j-1}}{B(\alpha+i, \beta+j)}, \quad (4.2.18)$$

that is, $f(x|i)$ is another beta density. This says that if we observe i successes and j failures in n subjects, then the new density for the probability that the drug is effective is again a beta density but with parameters $\alpha + i, \beta + j$.

Now we assume that before the experiment we choose a beta density with parameters α and β , and that in the experiment we obtain i successes in n trials. We have just seen that in this case, the new density for x is a beta density with parameters $\alpha + i$ and $\beta + j$.

Now we wish to calculate the probability that the drug is effective on the next subject. For any particular real number t between 0 and 1, the probability that x has the value t is given by the expression in Equation 4.5. Given that x has the value t , the probability that the drug is effective on the next subject is just t . Thus, to obtain the probability that the drug is effective on the next subject, we integrate the product of the expression in Equation 4.5 and t over all possible values of t . We obtain:

$$\frac{1}{B(\alpha+i, \beta+j)} \int_0^1 t \cdot x^{\alpha+i-1} (1-t)^{\beta+j-1} dt \quad (4.2.19)$$

$$= \frac{B(\alpha+i+1, \beta+j)}{B(\alpha+i, \beta+j)} \quad (4.2.20)$$

$$= \frac{(\alpha+i)!(\beta+j-1)!}{(\alpha+\beta+i+j)!} \cdot \frac{(\alpha+\beta+i+j-1)!}{(\alpha+i-1)!(\beta+j-1)!} \quad (4.2.21)$$

$$= \frac{\alpha+i}{\alpha+\beta+n} \quad (4.2.22)$$

If n is large, then our estimate for the probability of success after the experiment is approximately the proportion of successes observed in the experiment, which is certainly a reasonable conclusion.

The next example is another in which the true probabilities are unknown and must be estimated based upon experimental data.

✓ Example 4.2.7: (Two-armed bandit problem)

You are in a casino and confronted by two slot machines. Each machine pays off either 1 dollar or nothing. The probability that the first machine pays off a dollar is x and that the second machine pays off a dollar is y . We assume that x and y are random numbers chosen independently from the interval $[0, 1]$ and unknown to you. You are permitted to make a series of ten plays, each time choosing one machine or the other. How should you choose to maximize the number of times that you win?

Machine	Result
1	W
1	L
2	L
1	L
1	W
1	L
1	L
1	L
2	W
2	L

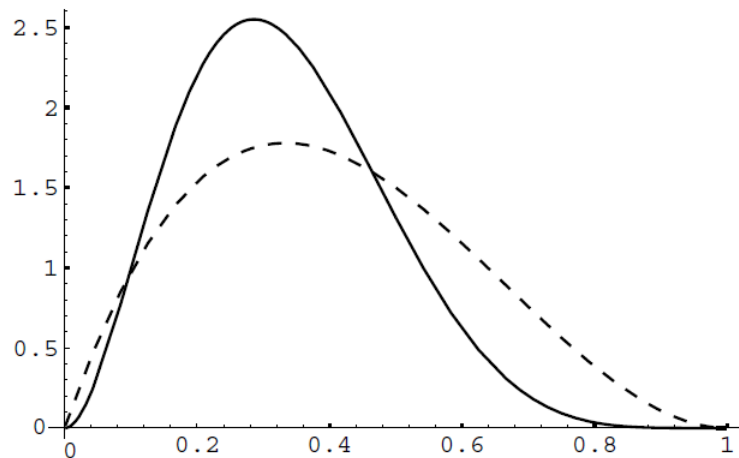


Figure 4.2.4: Play the best machine.

One strategy that sounds reasonable is to calculate, at every stage, the probability that each machine will pay off and choose the machine with the higher probability. Let $\text{win}(i)$, for $i = 1$ or 2 , be the number of times that you have won on the i th machine. Similarly, let $\text{lose}(i)$ be the number of times you have lost on the i th machine. Then, from Example 4.16 the probability $p(i)$ that you win if you choose the i th machine is

$$p(i) = \frac{\text{win}(i) + 1}{\text{win}(i) + \text{lose}(i) + 2} . \quad (4.2.23)$$

Thus, if $p(1) > p(2)$ you would play machine 1 and otherwise you would play machine 2. We have written a program **TwoArm** to simulate this experiment. In the program, the user specifies the initial values for x and y (but these are unknown to the experimenter). The program calculates at each stage the two conditional densities for x and y , given the outcomes of the previous trials, and then computes $p(i)$, for $i = 1, 2$. It then chooses the machine with the highest value for the probability of winning for the next play. The program prints the machine chosen on each play and the outcome of this play. It also plots the new densities for x (solid line) and y (dotted line), showing only the current densities. We have run the program for ten plays for the case $x = .6$ and $y = .7$. The result is shown in Figure 4.7

The run of the program shows the weakness of this strategy. Our initial probability for winning on the better of the two machines is $.7$. We start with the poorer machine and our outcomes are such that we always have a probability greater than $.6$ of winning and so we just keep playing this machine even though the other machine is better. If we had lost on the first play we would have switched machines. Our final density for y is the same as our initial density, namely, the uniform density. Our final density for x is different and reflects a much more accurate knowledge about x . The computer did pretty well with this strategy, winning seven out of the ten trials, but ten trials are not enough to judge whether this is a good strategy in the long run.

Another popular strategy is the *play-the-winner-strategy*. As the name suggests, for this strategy we choose the same machine when we win and switch machines when we lose. The program **TwoArm** will simulate this strategy as well. In Figure 4.11, we show the results of running this program with the play-the-winner strategy and the same true probabilities of $.6$ and $.7$ for the two machines. After ten plays our densities for the unknown probabilities of winning suggest to us that the second machine is indeed the better of the two. We again won seven out of the ten trials.

Machine	Result
1	W
1	W
1	L
2	L
1	W
1	W
1	L
2	L
1	L
2	W

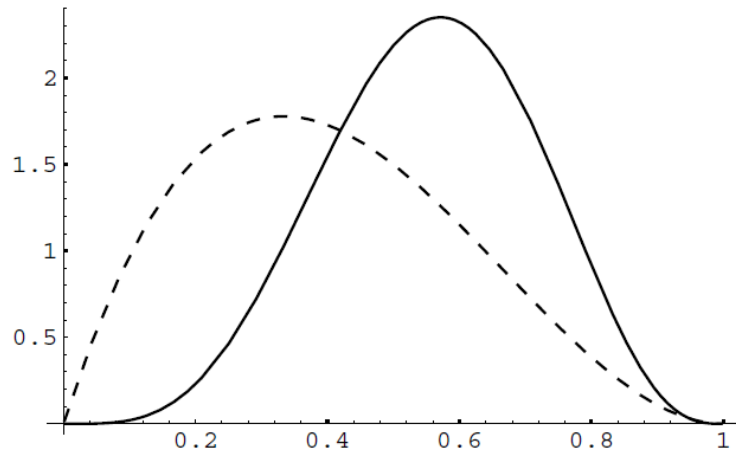


Figure 4.2.5: Play the winner

Neither of the strategies that we simulated is the best one in terms of maximizing our average winnings. This best strategy is very complicated but is reasonably approximated by the play-the-winner strategy. Variations on this example have played an important role in the problem of clinical tests of drugs where experimenters face a similar situation.

Exercises

Exercise 4.2.1

Pick a point x at random (with uniform density) in the interval $[0, 1]$. Find the probability that $x > 1/2$, given that

- $x > 1/4$.
- $x < 3/4$.
- $|x - 1/2| < 1/4$.
- $x^2 - x + 2/9 < 0$.

Exercise 4.2.2

A radioactive material emits α -particles at a rate described by the density function

$$f(t) = .1e^{-.1t} . \quad (4.2.24)$$

Find the probability that a particle is emitted in the first 10 seconds, given that

- no particle is emitted in the first second.
- no particle is emitted in the first 5 seconds.
- a particle is emitted in the first 3 seconds.
- a particle is emitted in the first 20 seconds.

Exercise 4.2.3

The Acme Super light bulb is known to have a useful life described by the density function

$$f(t) = .01e^{-.01t} , \quad (4.2.25)$$

where time t is measured in hours.

- Find the *failure rate* of this bulb (see Exercise 2.2.6)
- Find the *reliability* of this bulb after 20 hours.
- Given that it lasts 20 hours, find the probability that the bulb lasts another 20 hours.
- Find the probability that the bulb burns out in the forty-first hour, given that it lasts 40 hours.

Exercise 4.2.4

Suppose you toss a dart at a circular target of radius 10 inches. Given that the dart lands in the upper half of the target, find the probability that

- it lands in the right half of the target.
- its distance from the center is less than 5 inches.
- its distance from the center is greater than 5 inches.
- it lands within 5 inches of the point $(0, 5)$.

Exercise 4.2.5

Suppose you choose two numbers x and y , independently at random from the interval $[0, 1]$. Given that their sum lies in the interval $[0, 1]$, find the probability that

- $|x - y| < 1$.
- $xy < 1/2$.
- $\max\{x, y\} < 1/2$
- $x^2 + y^2 < 1/4$.
- $x > y$.

Exercise 4.2.6

Find the conditional density functions for the following experiments.

- A number x is chosen at random in the interval $[0, 1]$, given that $x > 1/4$.
- A number t is chosen at random in the interval $[0, \infty)$ with exponential density e^{-t} , given that $1 < t < 10$.
- A dart is thrown at a circular target of radius 10 inches, given that it falls in the upper half of the target.
- Two numbers x and y are chosen at random in the interval $[0, 1]$, given that $x > y$.

Exercise 4.2.7

Let x and y be chosen at random from the interval $[0, 1]$. Show that the events $x > 1/3$ and $y > 2/3$ are independent events.

Exercise 4.2.8

Let x and y be chosen at random from the interval $[0, 1]$. Which pairs of the following events are independent?

- $x > 1/3$.
- $y > 2/3$.
- $x > y$.
- $x + y < 1$.

Exercise 4.2.9

Suppose that X and Y are continuous random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively. Let $f(x, y)$ denote the joint density function of (X, Y) . Show that

$$\int_{-\infty}^{\infty} f(x, y) dy = f_X(x), \quad (4.2.26)$$

and

$$\int_{-\infty}^{\infty} f(x, y) dx = f_Y(y). \quad (4.2.27)$$

Exercise *4.2.10

In Exercise 2.2.12 you proved the following: If you take a stick of unit length and break it into three pieces, choosing the breaks at random (i.e., choosing two real numbers independently and uniformly from $[0, 1]$), then the probability that the three pieces form a triangle is $1/4$. Consider now a similar experiment: First break the stick at random, then break the longer piece at random. Show that the two experiments are actually quite different, as follows:

- Write a program which simulates both cases for a run of 1000 trials, prints out the proportion of successes for each run, and repeats this process ten times. (Call a trial a success if the three pieces do form a triangle.) Have your program pick (x, y) at

random in the unit square, and in each case use x and y to find the two breaks. For each experiment, have it plot (x, y) if (x, y) gives a success.

b. Show that in the second experiment the theoretical probability of success is actually $2 \log 2 - 1$.

Exercise 4.2.11

A coin has an unknown bias p that is assumed to be uniformly distributed between 0 and 1. The coin is tossed n times and heads turns up j times and tails turns up k times. We have seen that the probability that heads turns up next time is

$$\frac{j+1}{n+2}. \quad (4.2.28)$$

Show that this is the same as the probability that the next ball is black for the Polya urn model of Exercise 4.1.20. Use this result to explain why, in the Polya urn model, the proportion of black balls does not tend to 0 or 1 as one might expect but rather to a uniform distribution on the interval $[0, 1]$.

Exercise 4.2.12

Previous experience with a drug suggests that the probability p that the drug is effective is a random quantity having a beta density with parameters $\alpha = 2$ and $\beta = 3$. The drug is used on ten subjects and found to be successful in four out of the ten patients. What density should we now assign to the probability p ? What is the probability that the drug will be successful the next time it is used?

Exercise 4.2.13

Write a program to allow you to compare the strategies play-the-winner and play-the-best-machine for the two-armed bandit problem of Example 4.17. Have your program determine the initial payoff probabilities for each machine by choosing a pair of random numbers between 0 and 1. Have your program carry out 20 plays and keep track of the number of wins for each of the two strategies. Finally, have your program make 1000 repetitions of the 20 plays and compute the average winning per 20 plays. Which strategy seems to be the best? Repeat these simulations with 20 replaced by 100. Does your answer to the above question change?

Exercise 4.2.14

Consider the two-armed bandit problem of Example 4.24. Bruce Barnes proposed the following strategy, which is a variation on the play-the-best-machine strategy. The machine with the greatest probability of winning is played the following two conditions hold: (a) the difference in the probabilities for winning is less than .08, and (b) the ratio of the number of times played on the more often played machine to the number of times played on the less often played machine is greater than 1.4. If the above two conditions hold, then the machine with the smaller probability of winning is played. Write a program to simulate this strategy. Have your program choose the initial payoff probabilities at random from the unit interval $[0, 1]$, make 20 plays, and keep track of the number of wins. Repeat this experiment 1000 times and obtain the average number of wins per 20 plays. Implement a second strategy—for example, play-the-best-machine or one of your own choice, and see how this second strategy compares with Bruce's on average wins.

This page titled [4.2: Continuous Conditional Probability](#) is shared under a [GNU Free Documentation License 1.3](#) license and was authored, remixed, and/or curated by [Charles M. Grinstead & J. Laurie Snell \(American Mathematical Society\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.