

3.5: Conditional Distributions

In this section, we study how a probability distribution changes when a given random variable has a known, specified value. So this is an essential topic that deals with how probability measures should be updated in light of new information. As usual, if you are a new student or probability, you may want to skip the technical details.

Basic Theory

Our starting point is a random experiment modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So to review, Ω is the set of outcomes, \mathcal{F} is the collection of events, and \mathbb{P} is the probability measure on the underlying sample space (Ω, \mathcal{F}) .

Suppose that X is a random variable defined on the sample space (that is, defined for the experiment), taking values in a set S .

Details

Technically, the collection of events \mathcal{F} is a σ -algebra, so that the sample space (Ω, \mathcal{F}) is a measurable space. Similarly, S will have a σ -algebra of admissible subsets, so that (S, \mathcal{S}) is also a measurable space. Random variable X is measurable, so that $\{X \in A\} \in \mathcal{F}$ for every $A \in \mathcal{S}$. The *distribution* of X is the probability measure $A \mapsto \mathbb{P}(X \in A)$ for $A \in \mathcal{S}$.

The purpose of this section is to study the conditional probability measure given $X = x$ for $x \in S$. That is, if E is an event, we would like to define and study the probability of E given $X = x$, denoted $\mathbb{P}(E | X = x)$. If X has a discrete distribution, the conditioning event has positive probability, so no new concepts are involved, and the simple definition of conditional probability suffices. When X has a continuous distribution, however, the conditioning event has probability 0, so a fundamentally new approach is needed.

The Discrete Case

Suppose first that X has a discrete distribution with probability density function g . Thus S is countable and we can assume that $g(x) = \mathbb{P}(X = x) > 0$ for $x \in S$.

If E is an event and $x \in S$ then

$$\mathbb{P}(E | X = x) = \frac{\mathbb{P}(E, X = x)}{g(x)} \quad (3.5.1)$$

Proof

The meaning of *discrete distribution* is that S is countable and $\mathcal{S} = \mathcal{P}(S)$ is the collection of all subsets of S . Technically, g is the probability density function of X with respect to counting measure $\#$ on S , the standard measure for discrete spaces. In the displayed equation above, the comma separating the events in the numerator of the fraction means *and*, and thus functions just like the intersection symbol. This result follows immediately from the definition of conditional probability:

$$\mathbb{P}(E | X = x) = \frac{\mathbb{P}(E, X = x)}{\mathbb{P}(X = x)} = \frac{\mathbb{P}(E, X = x)}{g(x)} \quad (3.5.2)$$

The next result is a special case of the law of total probability, and will be the key to the definition when X has a continuous distribution.

If E is an event then

$$\mathbb{P}(E, X \in A) = \sum_{x \in A} g(x) \mathbb{P}(E | X = x), \quad A \subseteq S \quad (3.5.3)$$

Conversely, this condition uniquely determines $\mathbb{P}(E | X = x)$.

Proof

As noted, the displayed equation is just a special case of the law of total probability. For $A \subseteq S$, the countable collection of events $\{\{X = x\} : x \in A\}$ partitions $\{X \in A\}$ so

$$\mathbb{P}(E, X \in A) = \sum_{x \in A} \mathbb{P}(E, X = x) = \sum_{x \in A} \mathbb{P}(E | X = x) \mathbb{P}(X = x) = \sum_{x \in A} \mathbb{P}(E | X = x) g(x) \quad (3.5.4)$$

Conversely, suppose that the function $Q(x, E)$, defined for $x \in S$ and $E \in \mathcal{F}$, satisfies

$$\mathbb{P}(E, X \in A) = \sum_{x \in A} g(x) Q(x, E), \quad A \subseteq S \quad (3.5.5)$$

Letting $A = \{x\}$ for $x \in S$ gives $\mathbb{P}(E, X = x) = g(x)Q(x, E)$, so $Q(x, E) = \mathbb{P}(E, X = x)/g(x) = \mathbb{P}(E | X = x)$.

The Continuous Case

Suppose now that X has a continuous distribution on $S \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$, with probability density function g . We assume that $g(x) > 0$ for $x \in S$. Unlike the discrete case, we cannot use simple conditional probability to define $\mathbb{P}(E | X = x)$ for an event E and $x \in S$ because the conditioning event has probability 0. Nonetheless, the concept *should* make sense. If we actually run the experiment, X will take on some value x (even though *a priori*, this event occurs with probability 0), and surely the information that $X = x$ should in general alter the probabilities that we assign to other events. A natural approach is to use the results obtained in the discrete case as *definitions* in the continuous case.

If E is an event and $x \in S$, the *conditional probability* $\mathbb{P}(E | X = x)$ is defined by the requirement that

$$\mathbb{P}(E, X \in A) = \int_A g(x)\mathbb{P}(E | X = x) dx, \quad A \subseteq S \quad (3.5.6)$$

Details

Technically, S is a measurable subset of \mathbb{R}^n and the σ -algebra \mathcal{S} consists of the subsets of S that are also measurable as subsets of \mathbb{R}^n . The function g is also required to be measurable, and is the density function of X with respect to Lebesgue measure λ_n . Lebesgue measure is named for Henri Lebesgue and is the standard measure on \mathbb{R}^n .

We will accept the fact that $\mathbb{P}(E | X = x)$ can be defined uniquely, up to a set of measure 0, by the condition above, but we will return to this point in the section on Conditional Expectation in the chapter on Expected Value. Essentially the condition means that $\mathbb{P}(E | X = x)$ is defined so that $x \mapsto g(x)\mathbb{P}(E | X = x)$ is a density function for the finite measure $A \mapsto \mathbb{P}(E, X \in A)$.

Conditioning and Bayes' Theorem

Suppose again that X is a random variable with values in S and probability density function g , as described above. Our discussions above in the discrete and continuous cases lead to basic formulas for computing the probability of an event by *conditioning on X* .

The law of total probability. If E is an event, then $\mathbb{P}(E)$ can be computed as follows:

1. If X has a discrete distribution then

$$\mathbb{P}(E) = \sum_{x \in S} g(x)\mathbb{P}(E | X = x) \quad (3.5.7)$$

2. If X has a continuous distribution then

$$\mathbb{P}(E) = \int_S g(x)\mathbb{P}(E | X = x) dx \quad (3.5.8)$$

Proof

1. This follows from the [discrete theorem](#) with $A = S$.
2. This follows from the [fundamental definition](#) with $A = S$.

Naturally, the law of total probability is useful when $\mathbb{P}(E | X = x)$ and $g(x)$ are known for $x \in S$. Our next result is, *Bayes' Theorem*, named after Thomas Bayes.

Bayes' Theorem. Suppose that E is an event with $\mathbb{P}(E) > 0$. The conditional probability density function $x \mapsto g(x | E)$ of X given E can be computed as follows:

1. If X has a discrete distribution then

$$g(x | E) = \frac{g(x)\mathbb{P}(E | X = x)}{\sum_{s \in S} g(s)\mathbb{P}(E | X = s)}, \quad x \in S \quad (3.5.9)$$

2. If X has a continuous distribution then

$$g(x | E) = \frac{g(x)\mathbb{P}(E | X = x)}{\int_S g(s)\mathbb{P}(E | X = s) ds}, \quad x \in S \quad (3.5.10)$$

Proof

1. In the discrete case, as usual, the ordinary simple definition of conditional probability suffices. The numerator in the displayed equation is $\mathbb{P}(X = x)\mathbb{P}(E | X = x) = \mathbb{P}(E, X = x)$. The denominator is $\mathbb{P}(E)$ by part (a) of the [law of total probability](#). Hence the fraction is $\mathbb{P}(E, X = x)/\mathbb{P}(E) = \mathbb{P}(X = x | E)$.
2. In the continuous case, as usual, the argument is more subtle. We need to show that the expression in the displayed equation satisfies the defining property of a PDF for the conditional distribution. Once again, the denominator is $\mathbb{P}(E)$ by part (b) of the [law of total probability](#). If $A \subseteq S$ then using the [fundamental definition](#),

$$\int_A g(x | E) dx = \frac{1}{\mathbb{P}(E)} \int_A g(x)\mathbb{P}(E | X = x) dx = \frac{\mathbb{P}(E, X \in A)}{\mathbb{P}(E)} = \mathbb{P}(X \in A | E) \quad (3.5.11)$$

By the meaning of the term, $x \mapsto g(x | E)$ is the conditional probability density function of X given E .

In the context of Bayes' theorem, g is called the *prior probability density function* of X and $x \mapsto g(x | E)$ is the *posterior probability density function* of X given E . Note also that the conditional probability density function of X given E is proportional to the function $x \mapsto g(x)\mathbb{P}(E | X = x)$, the sum or integral of this function that occurs in the denominator is simply the normalizing constant. As with the law of total probability, Bayes' theorem is useful when $\mathbb{P}(E | X = x)$ and $g(x)$ are known for $x \in S$.

Conditional Probability Density Functions

The definitions and results above apply, of course, if E is an event defined in terms of another random variable for our experiment. Here is the setup:

Suppose that X and Y are random variables on the probability space, with values in sets S and T , respectively, so that (X, Y) is a random variable with values in $S \times T$. We assume that (X, Y) has probability density function f , as discussed in the section on Joint Distributions. Recall that X has probability density function g defined as follows:

1. If Y has a discrete distribution on the countable set T then

$$g(x) = \sum_{y \in T} f(x, y), \quad x \in S \quad (3.5.12)$$

2. If Y has a continuous distribution on $T \subseteq \mathbb{R}^k$ then

$$g(x) = \int_T f(x, y) dy, \quad x \in S \quad (3.5.13)$$

Similarly, the probability density function h of Y can be obtained by summing f over $x \in S$ if X has a discrete distribution or integrating f over S if X has a continuous distribution.

Suppose that $x \in S$ and that $g(x) > 0$. The function $y \mapsto h(y | x)$ defined below is a probability density function on T :

$$h(y | x) = \frac{f(x, y)}{g(x)}, \quad y \in T \quad (3.5.14)$$

Proof

The result is simple, since $g(x)$ is the normalizing constant for $y \mapsto h(y | x)$. Specifically, fix $x \in S$. Then $h(y | x) \geq 0$. If Y has a discrete distribution then

$$\sum_{y \in T} h(y | x) = \frac{1}{g(x)} \sum_{y \in T} f(x, y) = \frac{g(x)}{g(x)} = 1 \quad (3.5.15)$$

Similarly, if Y has a continuous distribution then

$$\int_T h(y | x) dy = \frac{1}{g(x)} \int_T f(x, y) dy = \frac{g(x)}{g(x)} = 1 \quad (3.5.16)$$

The distribution that corresponds to this probability density function is what you would expect:

For $x \in S$, the function $y \mapsto h(y | x)$ is the conditional probability density function of Y given $X = x$. That is,

1. If Y has a discrete distribution then

$$\mathbb{P}(Y \in B \mid X = x) = \sum_{y \in B} h(y \mid x), \quad B \subseteq T \quad (3.5.17)$$

2. If Y has a continuous distribution then

$$\mathbb{P}(Y \in B \mid X = x) = \int_B h(y \mid x) dy, \quad B \subseteq T \quad (3.5.18)$$

Proof

There are four cases, depending on the type of distribution of X and Y , but the computations are identical, except for sums in the discrete case and integrals in the continuous case. The main tool is the [basic theorem](#) when X has a discrete distribution and the [fundamental definition](#) when X has a continuous distribution, with the event E replaced by $\{Y \in B\}$ for $B \subseteq T$. The other main element is the fact that f is the PDF of the (joint) distribution of (X, Y) .

1. Suppose that Y has a discrete distribution on the countable set T . If X also has a discrete distribution on the countable set S then

$$\sum_{x \in A} g(x) \sum_{y \in B} h(y \mid x) = \sum_{x \in A} \sum_{y \in B} g(x) h(y \mid x) = \sum_{x \in A} \sum_{y \in B} f(x, y) = \mathbb{P}(X \in A, Y \in B), \quad A \subseteq S \quad (3.5.19)$$

In this jointly discrete case, there is a simpler argument of course:

$$h(y \mid x) = \frac{f(x, y)}{g(x)} = \frac{\mathbb{P}(X = x, Y = y)}{P(X = x)} = \mathbb{P}(Y = y \mid X = x), \quad y \in T \quad (3.5.20)$$

If X has a continuous distribution on $S \subseteq \mathbb{R}^j$ then

$$\int_A g(x) \sum_{y \in B} h(y \mid x) dx = \int_A \sum_{y \in B} g(x) h(y \mid x) dx = \int_A \sum_{y \in B} f(x, y) dx = \mathbb{P}(X \in A, Y \in B), \quad A \subseteq S \quad (3.5.21)$$

2. Suppose that Y has continuous distributions on $T \subseteq \mathbb{R}^k$. If X has a discrete distribution on the countable set S then

$$\sum_{x \in A} g(x) \int_B h(y \mid x) dy = \sum_{x \in A} \int_B g(x) h(y \mid x) dy = \sum_{x \in A} \int_B f(x, y) dy = \mathbb{P}(X \in A, Y \in B), \quad A \subseteq S \quad (3.5.22)$$

If X has a continuous distribution $S \subseteq \mathbb{R}^j$ then

$$\int_A g(x) \int_B h(y \mid x) dy dx = \int_A \int_B g(x) h(y \mid x) dy dx = \int_A \int_B f(x, y) dy dx = \mathbb{P}(X \in A, Y \in B), \quad A \subseteq S \quad (3.5.23)$$

The following theorem gives Bayes' theorem for probability density functions. We use the notation established above.

Bayes' Theorem. For $y \in T$, the conditional probability density function $x \mapsto g(x \mid y)$ of X given $Y = y$ can be computed as follows:

1. If X has a discrete distribution then

$$g(x \mid y) = \frac{g(x)h(y \mid x)}{\sum_{s \in S} g(s)h(y \mid s)}, \quad x \in S \quad (3.5.24)$$

2. If X has a continuous distribution then

$$g(x \mid y) = \frac{g(x)h(y \mid x)}{\int_S g(s)h(y \mid s)ds}, \quad x \in S \quad (3.5.25)$$

Proof

In both cases the numerator is $f(x, y)$ while the denominator is $h(y)$.

In the context of Bayes' theorem, g is the *prior probability density function* of X and $x \mapsto g(x \mid y)$ is the *posterior probability density function* of X given $Y = y$ for $y \in T$. Note that the posterior probability density function $x \mapsto g(x \mid y)$ is proportional to the function $x \mapsto g(x)h(y \mid x)$. The sum or integral in the denominator is the normalizing constant.

Independence

Intuitively, X and Y should be independent if and only if the conditional distributions are the same as the corresponding unconditional distributions.

The following conditions are equivalent:

1. X and Y are independent.
2. $f(x, y) = g(x)h(y)$ for $x \in S, y \in T$
3. $h(y | x) = h(y)$ for $x \in S, y \in T$
4. $g(x | y) = g(x)$ for $x \in S, y \in T$

Proof

The equivalence of (a) and (b) was established in the section on joint distributions. Parts (c) and (d) are equivalent to (b). For a continuous distribution as described in the details in (4), a probability density function is not unique. The values of a PDF can be changed to other nonnegative values on a set of measure 0 and the resulting function is still a PDF. So if X or Y has a continuous distribution, the equations above have to be interpreted as holding for x or y , respectively, except on a set of measure 0.

Examples and Applications

In the exercises that follow, look for special models and distributions that we have studied. A special distribution may be embedded in a larger problem, as a conditional distribution, for example. In particular, a conditional distribution sometimes arises when a parameter of a standard distribution is *randomized*.

A couple of special distributions will occur frequently in the exercises. First, recall that the discrete uniform distribution on a finite, nonempty set S has probability density function f given by $f(x) = 1/\#(S)$ for $x \in S$. This distribution governs an element selected at random from S .

Recall also that *Bernoulli trials* (named for Jacob Bernoulli) are independent trials, each with two possible outcomes generically called *success* and *failure*. The probability of success $p \in [0, 1]$ is the same for each trial, and is the basic parameter of the random process. The number of successes in $n \in \mathbb{N}_+$ Bernoulli trials has the *binomial distribution* with parameters n and p . This distribution has probability density function f given by $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x \in \{0, 1, \dots, n\}$. The binomial distribution is studied in more detail in the chapter on Bernoulli trials

Coins and Dice

Suppose that two standard, fair dice are rolled and the sequence of scores (X_1, X_2) is recorded. Let $U = \min\{X_1, X_2\}$ and $V = \max\{X_1, X_2\}$ denote the minimum and maximum scores, respectively.

1. Find the conditional probability density function of U given $V = v$ for each $v \in \{1, 2, 3, 4, 5, 6\}$.
2. Find the conditional probability density function of V given $U = u$ for each $u \in \{1, 2, 3, 4, 5, 6\}$

Answer

1.	$g(u v)$	$u = 1$	2	3	4	5	6
	$v = 1$	1	0	0	0	0	0
	2	$\frac{2}{3}$	$\frac{1}{3}$	0	0	0	0
	3	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	0	0	0
	4	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	0	0
	5	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	0
	6	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{1}{11}$

2.	$h(v u)$	$u = 1$	2	3	4	5	6
	$v = 1$	$\frac{1}{11}$	0	0	0	0	0
	2	$\frac{2}{11}$	$\frac{1}{9}$	0	0	0	0
	3	$\frac{2}{11}$	$\frac{2}{9}$	$\frac{1}{7}$	0	0	0
	4	$\frac{2}{11}$	$\frac{2}{9}$	$\frac{2}{7}$	$\frac{1}{5}$	0	0
	5	$\frac{2}{11}$	$\frac{2}{9}$	$\frac{2}{7}$	$\frac{2}{5}$	$\frac{1}{3}$	0

$h(v u)$	$u = 1$	2	3	4	5	6
6	$\frac{2}{11}$	$\frac{2}{9}$	$\frac{2}{7}$	$\frac{2}{5}$	$\frac{2}{3}$	1

In the die-coin experiment, a standard, fair die is rolled and then a fair coin is tossed the number of times showing on the die. Let N denote the die score and Y the number of heads.

- Find the joint probability density function of (N, Y) .
- Find the probability density function of Y .
- Find the conditional probability density function of N given $Y = y$ for each $y \in \{0, 1, 2, 3, 4, 5, 6\}$

Answer

- and b.

$f(n, y)$	$n = 1$	2	3	4	5	6	$h(y)$
$y = 0$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{48}$	$\frac{1}{96}$	$\frac{1}{102}$	$\frac{1}{384}$	$\frac{63}{384}$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{16}$	$\frac{1}{24}$	$\frac{5}{192}$	$\frac{1}{64}$	$\frac{120}{384}$
2	0	$\frac{1}{24}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{96}$	$\frac{5}{128}$	$\frac{99}{384}$
3	0	0	$\frac{1}{48}$	$\frac{1}{24}$	$\frac{5}{96}$	$\frac{5}{96}$	$\frac{64}{384}$
4	0	0	0	$\frac{1}{96}$	$\frac{5}{192}$	$\frac{5}{128}$	$\frac{29}{384}$
5	0	0	0	0	$\frac{1}{192}$	$\frac{1}{64}$	$\frac{8}{384}$
6	0	0	0	0	0	$\frac{1}{384}$	$\frac{1}{384}$
$g(n)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$g(n y)$	$n = 1$	2	3	4	5	6
$y = 0$	$\frac{32}{63}$	$\frac{16}{63}$	$\frac{8}{63}$	$\frac{4}{63}$	$\frac{2}{63}$	$\frac{1}{63}$
1	$\frac{16}{60}$	$\frac{16}{60}$	$\frac{12}{60}$	$\frac{8}{60}$	$\frac{5}{60}$	$\frac{3}{60}$
2	0	$\frac{16}{99}$	$\frac{24}{99}$	$\frac{24}{99}$	$\frac{20}{99}$	$\frac{15}{99}$
3	0	0	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{5}{16}$
4	0	0	0	$\frac{4}{29}$	$\frac{10}{29}$	$\frac{15}{29}$
5	0	0	0	0	$\frac{1}{4}$	$\frac{3}{4}$
6	0	0	0	0	0	1

In the die-coin experiment, select the fair die and coin.

- Run the simulation of 1000 times and compare the empirical density function of Y with the true probability density function in the previous exercise
- Run the simulation 1000 times and compute the empirical conditional density function of N given $Y = 3$. Compare with the conditional probability density functions in the previous exercise.

In the coin-die experiment, a fair coin is tossed. If the coin is tails, a standard, fair die is rolled. If the coin is heads, a standard, ace-six flat die is rolled (faces 1 and 6 have probability $\frac{1}{4}$ each and faces 2, 3, 4, 5 have probability $\frac{1}{8}$ each). Let X denote the coin score (0 for tails and 1 for heads) and Y the die score.

- Find the joint probability density function of (X, Y) .
- Find the probability density function of Y .
- Find the conditional probability density function of X given $Y = y$ for each $y \in \{1, 2, 3, 4, 5, 6\}$

Answer

1. and b.

$f(x, y)$	$y = 1$	2	3	4	5	6	$g(x)$
$x = 0$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{2}$
$h(y)$	$\frac{5}{24}$	$\frac{7}{24}$	$\frac{7}{48}$	$\frac{7}{48}$	$\frac{7}{48}$	$\frac{5}{24}$	1

3.

$g(x y)$	$y = 1$	2	3	4	5	6
$x = 0$	$\frac{2}{5}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{4}{7}$	$\frac{2}{5}$
1	$\frac{3}{5}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{3}{5}$

In the coin-die experiment, select the settings of the previous exercise.

1. Run the simulation 1000 times and compare the empirical density function of Y with the true probability density function in the previous exercise.
2. Run the simulation 100 times and compute the empirical conditional probability density function of X given $Y = 2$. Compare with the conditional probability density function in the previous exercise.

Suppose that a box contains 12 coins: 5 are fair, 4 are biased so that heads comes up with probability $\frac{1}{3}$, and 3 are two-headed. A coin is chosen at random and tossed 2 times. Let P denote the probability of heads of the selected coin, and X the number of heads.

1. Find the joint probability density function of (P, X) .
2. Find the probability density function of X .
3. Find the conditional probability density function of P given $X = x$ for $x \in \{0, 1, 2\}$.

Answer

1. and b.

$f(p, x)$	$x = 0$	1	2	$g(p)$
$p = \frac{1}{2}$	$\frac{5}{48}$	$\frac{10}{48}$	$\frac{5}{48}$	$\frac{5}{12}$
$\frac{1}{3}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{1}{27}$	$\frac{4}{12}$
1	0	0	$\frac{1}{4}$	$\frac{3}{12}$
$h(x)$	$\frac{109}{432}$	$\frac{154}{432}$	$\frac{169}{432}$	1

3.

$g(p x)$	$x = 0$	1	2
$p = \frac{1}{2}$	$\frac{45}{109}$	$\frac{45}{77}$	$\frac{45}{169}$
$\frac{1}{3}$	$\frac{64}{109}$	$\frac{32}{77}$	$\frac{16}{169}$
1	0	0	$\frac{108}{169}$

Compare the [die-coin experiment](#) with the [box of coins experiment](#). In the first experiment, we toss a coin with a *fixed* probability of heads a *random* number of times. In the second experiment, we effectively toss a coin with a *random* probability of heads a *fixed* number of times.

Suppose that P has probability density function $g(p) = 6p(1 - p)$ for $p \in [0, 1]$. Given $P = p$, a coin with probability of heads p is tossed 3 times. Let X denote the number of heads.

1. Find the joint probability density function of (P, X) .
2. Find the probability density function of X .
3. Find the conditional probability density of P given $X = x$ for $x \in \{0, 1, 2, 3\}$. Graph these on the same axes.

Answer

1. $f(p, x) = 6 \binom{3}{x} p^{x+1} (1-p)^{4-x}$ for $p \in [0, 1]$ and $x \in \{0, 1, 2, 3, 4\}$
2. $h(0) = h(3) = \frac{1}{5}$, $h(1) = h(2) = \frac{3}{10}$.
3. $g(p | 0) = 30p(1-p)^4$, $g(p | 1) = 60p^2(1-p)^3$, $g(p | 2) = 60p^3(1-p)^2$, $g(p | 3) = 30p^4(1-p)$, in each case for $p \in [0, 1]$

Compare the [box of coins experiment](#) with the [last experiment](#). In the second experiment, we effectively choose a coin from a box with a continuous infinity of coin types. The prior distribution of P and each of the posterior distributions of P in part (c) are members of the family of *beta distributions*, one of the reasons for the importance of the beta family. Beta distributions are studied in more detail in the chapter on Special Distributions.

In the simulation of the beta coin experiment, set $a = b = 2$ and $n = 3$ to get the experiment studied in the previous exercise. For various “true values” of p , run the experiment in single step mode a few times and observe the posterior probability density function on each run.

Simple Mixed Distributions

Recall that the *exponential distribution* with rate parameter $r \in (0, \infty)$ has probability density function f given by $f(t) = re^{-rt}$ for $t \in [0, \infty)$. The exponential distribution is often used to model random times, under certain assumptions. The exponential distribution is studied in more detail in the chapter on the Poisson Process. Recall also that for $a, b \in \mathbb{R}$ with $a < b$, the continuous uniform distribution on the interval $[a, b]$ has probability density function f given by $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$. This distribution governs a point selected at random from the interval.

Suppose that there are 5 light bulbs in a box, labeled 1 to 5. The lifetime of bulb n (in months) has the exponential distribution with rate parameter n . A bulb is selected at random from the box and tested.

1. Find the probability that the selected bulb will last more than one month.
2. Given that the bulb lasts more than one month, find the conditional probability density function of the bulb number.

Answer

Let N denote the bulb number and T the lifetime.

1. $\mathbb{P}(T > 1) = 0.1156$

2. n	1	2	3	4	5
$g(n T > 1)$	0.6364	0.2341	0.0861	0.0317	0.0117

Suppose that X is uniformly distributed on $\{1, 2, 3\}$, and given $X = x \in \{1, 2, 3\}$, random variable Y is uniformly distributed on the interval $[0, x]$.

1. Find the joint probability density function of (X, Y) .
2. Find the probability density function of Y .
3. Find the conditional probability density function of X given $Y = y$ for $y \in [0, 3]$.

Answer

1. $f(x, y) = \frac{1}{3x}$ for $y \in [0, x]$ and $x \in \{1, 2, 3\}$.
2. $h(y) = \begin{cases} \frac{11}{18}, & 0 \leq y \leq 1 \\ \frac{5}{18}, & 1 < y \leq 2 \\ \frac{2}{18}, & 2 < y \leq 3 \end{cases}$
3. For $y \in [0, 1]$, $g(1 | y) = \frac{6}{11}$, $g(2 | y) = \frac{3}{11}$, $g(3 | y) = \frac{2}{11}$
 For $y \in (1, 2]$, $g(1 | y) = 0$, $g(2 | y) = \frac{3}{5}$, $g(3 | y) = \frac{2}{5}$
 For $y \in (2, 3]$, $g(1 | y) = g(2 | y) = 0$, $g(3 | y) = 1$.

The Poisson Distribution

Recall that the Poisson distribution with parameter $a \in (0, \infty)$ has probability density function $g(n) = e^{-a} \frac{a^n}{n!}$ for $n \in \mathbb{N}$. This distribution is widely used to model the number of “random points” in a region of time or space; the parameter a is proportional to the size of the region. The Poisson distribution is named for Simeon Poisson, and is studied in more detail in the chapter on the Poisson Process.

Suppose that N is the number of elementary particles emitted by a sample of radioactive material in a specified period of time, and has the Poisson distribution with parameter a . Each particle emitted, independently of the others, is detected by a counter with probability $p \in (0, 1)$ and missed with probability $1 - p$. Let Y denote the number of particles detected by the counter.

1. For $n \in \mathbb{N}$, argue that the conditional distribution of Y given $N = n$ is binomial with parameters n and p .
2. Find the joint probability density function of (N, Y) .
3. Find the probability density function of Y .
4. For $y \in \mathbb{N}$, find the conditional probability density function of N given $Y = y$.

Answer

1. Each particle, independently, is detected (success) with probability p . This is the very definition of Bernoulli trials, so given $N = n$, the number of detected particles has the binomial distribution with parameters n and p .
2. The PDF f of (N, Y) is defined by

$$f(n, y) = e^{-a} a^n \frac{p^y (1-p)^{n-y}}{y! (n-y)!}, \quad n \in \mathbb{N}, y \in \{0, 1, \dots, n\} \quad (3.5.26)$$

3. The PDF h of Y is defined by

$$h(y) = e^{-pa} \frac{(pa)^y}{y!}, \quad y \in \mathbb{N} \quad (3.5.27)$$

This is the Poisson distribution with parameter pa .

4. The conditional PDF of N given $Y = y$ is defined by

$$g(n | y) = e^{-(1-p)a} \frac{[(1-p)a]^{n-y}}{(n-y)!}, \quad n \in \{y, y+1, \dots\} \quad (3.5.28)$$

This is the Poisson distribution with parameter $(1-p)a$, shifted to start at y .

The fact that Y also has a Poisson distribution is an interesting and characteristic property of the distribution. This property is explored in more depth in the section on thinning the Poisson process.

Simple Continuous Distributions

Suppose that (X, Y) has probability density function f defined by $f(x, y) = x + y$ for $(x, y) \in (0, 1)^2$.

1. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$
2. Find the conditional probability density function of Y given $X = x$ for $x \in (0, 1)$
3. Find $\mathbb{P}\left(\frac{1}{4} \leq Y \leq \frac{3}{4} \mid X = \frac{1}{3}\right)$.
4. Are X and Y independent?

Answer

1. For $y \in (0, 1)$, $g(x | y) = \frac{x+y}{y+1/2}$ for $x \in (0, 1)$
2. For $x \in (0, 1)$, $h(y | x) = \frac{x+y}{x+1/2}$ for $y \in (0, 1)$
3. $\frac{1}{2}$
4. X and Y are dependent.

Suppose that (X, Y) has probability density function f defined by $f(x, y) = 2(x + y)$ for $0 < x < y < 1$.

1. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in (0, 1)$.
3. Find $\mathbb{P}\left(Y \geq \frac{3}{4} \mid X = \frac{1}{2}\right)$.
4. Are X and Y independent?

Answer

1. For $y \in (0, 1)$, $g(x | y) = \frac{x+y}{3y^2}$ for $x \in (0, y)$.
2. For $x \in (0, 1)$, $h(y | x) = \frac{x+y}{(1+3x)(1-x)}$ for $y \in (x, 1)$.
3. $\frac{3}{10}$
4. X and Y are dependent.

Suppose that (X, Y) has probability density function f defined by $f(x, y) = 15x^2y$ for $0 < x < y < 1$.

1. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in (0, 1)$.
3. Find $\mathbb{P}\left(X \leq \frac{1}{4} \mid Y = \frac{1}{3}\right)$.
4. Are X and Y independent?

Answer

1. For $y \in (0, 1)$, $g(x \mid y) = \frac{3x^2}{y^3}$ for $x \in (0, y)$.
2. For $x \in (0, 1)$, $h(y \mid x) = \frac{2y}{1-x^2}$ for $y \in (x, 1)$.
3. $\frac{27}{64}$
4. X and Y are dependent.

Suppose that (X, Y) has probability density function f defined by $f(x, y) = 6x^2y$ for $0 < x < 1$ and $0 < y < 1$.

1. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in (0, 1)$.
3. Are X and Y independent?

Answer

1. For $y \in (0, 1)$, $g(x \mid y) = 3x^2$ for $y \in (0, 1)$.
2. For $x \in (0, 1)$, $h(y \mid x) = 2y$ for $y \in (0, 1)$.
3. X and Y are independent.

Suppose that (X, Y) has probability density function f defined by $f(x, y) = 2e^{-x}e^{-y}$ for $0 < x < y < \infty$.

1. Find the conditional probability density function of X given $Y = y$ for $y \in (0, \infty)$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in (0, \infty)$.
3. Are X and Y independent?

Answer

1. For $y \in (0, \infty)$, $g(x \mid y) = \frac{e^{-x}}{1-e^{-y}}$ for $x \in (0, y)$.
2. For $x \in (0, \infty)$, $h(y \mid x) = e^{x-y}$ for $y \in (x, \infty)$.
3. X and Y are dependent.

Suppose that X is uniformly distributed on the interval $(0, 1)$, and that given $X = x$, Y is uniformly distributed on the interval $(0, x)$.

1. Find the joint probability density function of (X, Y) .
2. Find the probability density function of Y .
3. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$.
4. Are X and Y independent?

Answer

1. $f(x, y) = \frac{1}{x}$ for $0 < y < x < 1$
2. $h(y) = -\ln y$ for $y \in (0, 1)$
3. For $y \in (0, 1)$, $g(x \mid y) = -\frac{1}{x \ln y}$ for $x \in (y, 1)$.
4. X and Y are dependent.

Suppose that X has probability density function g defined by $g(x) = 3x^2$ for $x \in (0, 1)$. The conditional probability density function of Y given $X = x$ is $h(y \mid x) = \frac{3y^2}{x^3}$ for $y \in (0, x)$.

1. Find the joint probability density function of (X, Y) .
2. Find the probability density function of Y .
3. Find the conditional probability density function of X given $Y = y$.
4. Are X and Y independent?

Answer

1. $f(x, y) = \frac{9y^2}{x}$ for $0 < y < x < 1$.
2. $h(y) = -9y^2 \ln y$ for $y \in (0, 1)$.

3. For $y \in (0, 1)$, $g(x | y) = -\frac{1}{x \ln y}$ for $x \in (y, 1)$.
4. X and Y are dependent.

Multivariate Uniform Distributions

Multivariate uniform distributions give a geometric interpretation of some of the concepts in this section.

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

$$\lambda_n(A) = \int_A 1 \, dx, \quad A \subseteq \mathbb{R}^n \quad (3.5.29)$$

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 defined on the σ -algebra of measurable subsets of \mathbb{R}^n . In the discussion below, we assume that all sets are measurable. The integral representation is valid for the sets that occur in typical applications.

Suppose now that X takes values in \mathbb{R}^j , Y takes values in \mathbb{R}^k , and that (X, Y) is uniformly distributed on a set $R \subseteq \mathbb{R}^{j+k}$. So $0 < \lambda_{j+k}(R) < \infty$ and then the joint probability density function f of (X, Y) is given by $f(x, y) = 1/\lambda_{j+k}(R)$ for $(x, y) \in R$. Now let S and T be the projections of R onto \mathbb{R}^j and \mathbb{R}^k respectively, defined as follows:

$$S = \{x \in \mathbb{R}^j : (x, y) \in R \text{ for some } y \in \mathbb{R}^k\}, \quad T = \{y \in \mathbb{R}^k : (x, y) \in R \text{ for some } x \in \mathbb{R}^j\} \quad (3.5.30)$$

Note that $R \subseteq S \times T$. Next we denote the cross sections at $x \in S$ and at $y \in T$, respectively by

$$T_x = \{t \in T : (x, t) \in R\}, \quad S_y = \{s \in S : (s, y) \in R\} \quad (3.5.31)$$

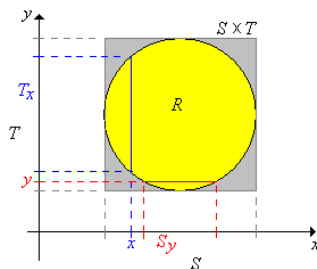


Figure 3.5.1: The projections S and T , and the cross sections at x and y

In the last section on Joint Distributions, we saw that even though (X, Y) is uniformly distributed, the marginal distributions of X and Y are not uniform in general. However, as the next theorem shows, the conditional distributions are always uniform.

Suppose that (X, Y) is uniformly distributed on R . Then

1. The conditional distribution of Y given $X = x$ is uniformly on T_x for each $x \in S$.
2. The conditional distribution of X given $Y = y$ is uniformly on S_y for each $y \in T$.

Proof

The results are symmetric, so we will prove (a). Recall that X has PDF g given by

$$g(x) = \int_{T_x} f(x, y) \, dy = \int_{T_x} \frac{1}{\lambda_{j+k}(R)} \, dy = \frac{\lambda_k(T_x)}{\lambda_{j+k}(R)}, \quad x \in S \quad (3.5.32)$$

Hence for $x \in S$, the conditional PDF of Y given $X = x$ is

$$h(y | x) = \frac{f(x, y)}{g(x)} = \frac{1}{\lambda_k(T_x)}, \quad y \in T_x \quad (3.5.33)$$

and this is the PDF of the uniform distribution on T_x .

Find the conditional density of each variable given a value of the other, and determine if the variables are independent, in each of the following cases:

1. (X, Y) is uniformly distributed on the square $R = (-6, 6)^2$.

2. (X, Y) is uniformly distributed on the triangle $R = \{(x, y) \in \mathbb{R}^2 : -6 < y < x < 6\}$.
3. (X, Y) is uniformly distributed on the circle $R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 36\}$.

Answer

The conditional PDF of X given $Y = y$ is denoted $x \mapsto g(x | y)$. The conditional PDF of Y given $X = x$ is denoted $y \mapsto h(y | x)$.

1.
 - For $y \in (-6, 6)$, $g(x | y) = \frac{1}{12}$ for $x \in (-6, 6)$.
 - For $x \in (-6, 6)$, $h(y | x) = \frac{1}{12}$ for $y \in (-6, 6)$.
 - X, Y are independent.
2.
 - For $y \in (-6, 6)$, $g(x | y) = \frac{1}{6-y}$ for $x \in (y, 6)$
 - For $x \in (-6, 6)$, $h(y | x) = \frac{1}{x+6}$ for $y \in (-6, x)$
 - X, Y are dependent.
3.
 - For $y \in (-6, 6)$, $g(x | y) = \frac{1}{2\sqrt{36-y^2}}$ for $x \in (-\sqrt{36-y^2}, \sqrt{36-y^2})$
 - For $x \in (-6, 6)$, $g(x | y) = \frac{1}{2\sqrt{36-x^2}}$ for $y \in (-\sqrt{36-x^2}, \sqrt{36-x^2})$
 - X, Y are dependent.

In the bivariate uniform experiment, run the simulation 1000 times in each of the following cases. Watch the points in the scatter plot and the graphs of the marginal distributions.

1. square
2. triangle
3. circle

Suppose that (X, Y, Z) is uniformly distributed on $R = \{(x, y, z) \in \mathbb{R}^3 : 0 < x < y < z < 1\}$.

1. Find the conditional density of each pair of variables given a value of the third variable.
2. Find the conditional density of each variable given values of the other two.

Answer

The subscripts 1, 2, and 3 correspond to the variables X, Y , and Z , respectively. Note that the conditions on (x, y, z) in each case are those in the definition of the domain R . They are stated differently to emphasize the domain of the conditional PDF as opposed to the given values, which function as parameters. Note also that each distribution is uniform on the appropriate region.

1. For $0 < z < 1$, $f_{1,2|3}(x, y | z) = \frac{2}{z^2}$ for $0 < x < y < z$
2. For $0 < y < 1$, $f_{1,3|2}(x, z | y) = \frac{1}{y(1-y)}$ for $0 < x < y$ and $y < z < 1$
3. For $0 < x < 1$, $f_{2,3|1}(y, z | x) = \frac{2}{(1-x)^2}$ for $x < y < z < 1$
4. For $0 < y < z < 1$, $f_{1|2,3}(x | y, z) = \frac{1}{y}$ for $0 < x < y$
5. For $0 < x < z < 1$, $f_{2|1,3}(y | x, z) = \frac{1}{z-x}$ for $x < y < z$
6. For $0 < x < y < 1$, $f_{3|1,2}(z | x, y) = \frac{1}{1-y}$ for $y < z < 1$

The Multivariate Hypergeometric Distribution

Recall the discussion of the (multivariate) hypergeometric distribution given in the last section on joint distributions. As in that discussion, suppose that a population consists of m objects, and that each object is one of four types. There are a objects of type 1, b objects of type 2, and c objects of type 3, and $m - a - b - c$ objects of type 0. We sample n objects from the population at random, and without replacement. The parameters a, b, c , and n are nonnegative integers with $a + b + c \leq m$ and $n \leq m$. Denote the number of type 1, 2, and 3 objects in the sample by X, Y , and Z , respectively. Hence, the number of type 0 objects in the sample is $n - X - Y - Z$. In the following exercises, $x, y, z \in \mathbb{N}$.

Suppose that $z \leq c$ and $n - m + c \leq z \leq n$. Then the conditional distribution of (X, Y) given $Z = z$ is hypergeometric, and has the probability density function defined by

$$g(x, y | z) = \frac{\binom{a}{x} \binom{b}{y} \binom{m-a-b-c}{n-x-y-z}}{\binom{m-c}{n-z}}, \quad x + y \leq n - z \quad (3.5.34)$$

Proof

This result can be proved analytically but a combinatorial argument is better. The essence of the argument is that we are selecting a random sample of size $n - z$ without replacement from a population of size $m - c$, with a objects of type 1, b objects of type 2, and $m - a - b - c$ objects of type 0. The conditions on z ensure that $\mathbb{P}(Z = z) > 0$, or equivalently, that the new parameters make sense.

Suppose that $y \leq b$, $z \leq c$, and $n - m + b \leq y + z \leq n$. Then the conditional distribution of X given $Y = y$ and $Z = z$ is hypergeometric, and has the probability density function defined by

$$g(x | y, z) = \frac{\binom{a}{x} \binom{m-a-b-c}{n-x-y-z}}{\binom{m-b-c}{n-y-z}}, \quad x \leq n - y - z \quad (3.5.35)$$

Proof

Again, this result can be proved analytically, but a combinatorial argument is better. The essence of the argument is that we are selecting a random sample of size $n - y - z$ from a population of size $m - b - c$, with a objects of type 1 and $m - a - b - c$ objects type 0. The conditions on y and z ensure that $\mathbb{P}(Y = y, Z = z) > 0$, or equivalently that the new parameters make sense.

These results generalize in a completely straightforward way to a population with any number of types. In brief, if a random vector has a hypergeometric distribution, then the conditional distribution of some of the variables, given values of the other variables, is also hypergeometric. Moreover, it is clearly not necessary to remember the hideous formulas in the previous two theorems. You just need to recognize the problem as sampling without replacement from a multi-type population, and then identify the number of objects of each type and the sample size. The hypergeometric distribution and the multivariate hypergeometric distribution are studied in more detail in the chapter on Finite Sampling Models.

In a population of 150 voters, 60 are democrats and 50 are republicans and 40 are independents. A sample of 15 voters is selected at random, without replacement. Let X denote the number of democrats in the sample and Y the number of republicans in the sample. Give the probability density function of each of the following:

1. (X, Y)
2. X
3. Y given $X = 5$

Answer

1. $f(x, y) = \frac{1}{\binom{150}{15}} \binom{60}{x} \binom{50}{y} \binom{40}{15-x-y}$ for $x + y \leq 15$
2. $g(x) = \frac{1}{\binom{150}{15}} \binom{60}{x} \binom{90}{15-x}$ for $x \leq 15$
3. $h(y | 5) = \frac{1}{\binom{90}{10}} \binom{50}{y} \binom{40}{10-y}$ for $y \leq 10$

Recall that a *bridge hand* consists of 13 cards selected at random and without replacement from a standard deck of 52 cards. Let X , Y , and Z denote the number of spades, hearts, and diamonds, respectively, in the hand. Find the probability density function of each of the following:

1. (X, Y, Z)
2. (X, Y)
3. X
4. (X, Y) given $Z = 3$
5. X given $Y = 3$ and $Z = 2$

Answer

1. $f(x, y, z) = \frac{1}{\binom{52}{13}} \binom{13}{x} \binom{13}{y} \binom{13}{z} \binom{13}{13-x-y-z}$ for $x + y + z \leq 13$.
2. $g(x, y) = \frac{1}{\binom{52}{13}} \binom{13}{x} \binom{13}{y} \binom{26}{13-x-y}$ for $x + y \leq 13$
3. $h(x) = \frac{1}{\binom{52}{13}} \binom{13}{x} \binom{39}{13-x}$ for $x \leq 13$
4. $g(x, y | 3) = \frac{1}{\binom{39}{10}} \binom{13}{x} \binom{13}{y} \binom{13}{10-x-y}$ for $x + y \leq 10$
5. $h(x | 3, 2) = \frac{1}{\binom{26}{8}} \binom{13}{x} \binom{13}{8-x}$ for $x \leq 8$

Multinomial Trials

Recall the discussion of *multinomial trials* in the last section on joint distributions. As in that discussion, suppose that we have a sequence of n independent trials, each with 4 possible outcomes. On each trial, outcome 1 occurs with probability p , outcome 2 with probability q , outcome 3 with probability r , and outcome 0 with probability $1 - p - q - r$. The parameters $p, q, r \in (0, 1)$, with $p + q + r < 1$, and $n \in \mathbb{N}_+$. Denote the number of times that outcome 1, outcome 2, and outcome 3 occurs in the n trials by X, Y , and Z respectively. Of course, the number of times that outcome 0 occurs is $n - X - Y - Z$. In the following exercises, $x, y, z \in \mathbb{N}$.

For $z \leq n$, the conditional distribution of (X, Y) given $Z = z$ is also multinomial, and has the probability density function.

$$g(x, y | z) = \binom{n-z}{x, y} \left(\frac{p}{1-r} \right)^x \left(\frac{q}{1-r} \right)^y \left(1 - \frac{p}{1-r} - \frac{q}{1-r} \right)^{n-x-y-z}, \quad x + y \leq n - z \quad (3.5.36)$$

Proof

This result can be proved analytically, but a probability argument is better. First, let I denote the outcome of a generic trial. Then $\mathbb{P}(I = 1 | I \neq 3) = \mathbb{P}(I = 1) / \mathbb{P}(I \neq 3) = p / (1 - r)$. Similarly, $\mathbb{P}(I = 2 | I \neq 3) = q / (1 - r)$ and $\mathbb{P}(I = 0 | I \neq 3) = (1 - p - q - r) / (1 - r)$. Now, the essence of the argument is that effectively, we have $n - z$ independent trials, and on each trial, outcome 1 occurs with probability $p / (1 - r)$ and outcome 2 with probability $q / (1 - r)$.

For $y + z \leq n$, the conditional distribution of X given $Y = y$ and $Z = z$ is binomial, with the probability density function

$$h(x | y, z) = \binom{n-y-z}{x} \left(\frac{p}{1-q-r} \right)^x \left(1 - \frac{p}{1-q-r} \right)^{n-x-y-z}, \quad x \leq n - y - z \quad (3.5.37)$$

Proof

Again, this result can be proved analytically, but a probability argument is better. As before, let I denote the outcome of a generic trial. Then $\mathbb{P}(I = 1 | I \notin \{2, 3\}) = p / (1 - q - r)$ and $\mathbb{P}(I = 0 | I \notin \{2, 3\}) = (1 - p - q - r) / (1 - q - r)$. Thus, the essence of the argument is that effectively, we have $n - y - z$ independent trials, and on each trial, outcome 1 occurs with probability $p / (1 - q - r)$.

These results generalize in a completely straightforward way to multinomial trials with any number of trial outcomes. In brief, if a random vector has a multinomial distribution, then the conditional distribution of some of the variables, given values of the other variables, is also multinomial. Moreover, it is clearly not necessary to remember the specific formulas in the previous two exercises. You just need to recognize a problem as one involving independent trials, and then identify the probability of each outcome and the number of trials. The binomial distribution and the multinomial distribution are studied in more detail in the chapter on Bernoulli Trials.

Suppose that peaches from an orchard are classified as *small*, *medium*, or *large*. Each peach, independently of the others is small with probability $\frac{3}{10}$, medium with probability $\frac{1}{2}$, and large with probability $\frac{1}{5}$. In a sample of 20 peaches from the orchard, let X denote the number of small peaches and Y the number of medium peaches. Give the probability density function of each of the following:

1. (X, Y)
2. X
3. Y given $X = 5$

Answer

1. $f(x, y) = \binom{20}{x, y} \left(\frac{3}{10} \right)^x \left(\frac{1}{2} \right)^y \left(\frac{1}{5} \right)^{20-x-y}$ for $x + y \leq 20$
2. $g(x) = \binom{20}{x} \left(\frac{3}{10} \right)^x \left(\frac{7}{10} \right)^{20-x}$ for $x \leq 20$
3. $h(y | 5) = \binom{15}{y} \left(\frac{5}{7} \right)^y \left(\frac{2}{7} \right)^{15-y}$ for $y \leq 15$

For a certain crooked, 4-sided die, face 1 has probability $\frac{2}{5}$, face 2 has probability $\frac{3}{10}$, face 3 has probability $\frac{1}{5}$, and face 4 has probability $\frac{1}{10}$. Suppose that the die is thrown 50 times. Let X, Y , and Z denote the number of times that scores 1, 2, and 3 occur, respectively. Find the probability density function of each of the following:

1. (X, Y, Z)
2. (X, Y)
3. X
4. (X, Y) given $Z = 5$

5. X given $Y = 10$ and $Z = 5$

Answer

1. $f(x, y, z) = \binom{50}{x, y, z} \left(\frac{2}{5}\right)^x \left(\frac{3}{10}\right)^y \left(\frac{1}{5}\right)^z \left(\frac{1}{10}\right)^{50-x-y-z}$ for $x + y + z \leq 50$
2. $g(x, y) = \binom{50}{x, y} \left(\frac{2}{5}\right)^x \left(\frac{3}{10}\right)^y \left(\frac{3}{10}\right)^{50-x-y}$ for $x + y \leq 50$
3. $h(x) = \binom{50}{x} \left(\frac{2}{5}\right)^x \left(\frac{3}{5}\right)^{50-x}$ for $x \leq 50$
4. $g(x, y | 5) = \binom{45}{x, y} \left(\frac{1}{2}\right)^x \left(\frac{3}{8}\right)^y \left(\frac{1}{8}\right)^{45-x-y}$ for $x + y \leq 45$
5. $h(x | 10, 5) = \binom{35}{x} \left(\frac{4}{5}\right)^x \left(\frac{1}{4}\right)^{10-x}$ for $x \leq 35$

Bivariate Normal Distributions

The joint distributions in the next two exercises are examples of *bivariate normal distributions*. The conditional distributions are also normal, an important property of the bivariate normal distribution. In general, normal distributions are widely used to model physical measurements subject to small, random errors. The bivariate normal distribution is studied in more detail in the chapter on Special Distributions.

Suppose that (X, Y) has the bivariate normal distribution with probability density function f defined by

$$f(x, y) = \frac{1}{12\pi} \exp\left[-\left(\frac{x^2}{8} + \frac{y^2}{18}\right)\right], \quad (x, y) \in \mathbb{R}^2 \quad (3.5.38)$$

1. Find the conditional probability density function of X given $Y = y$ for $y \in \mathbb{R}$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in \mathbb{R}$.
3. Are X and Y independent?

Answer

1. For $y \in \mathbb{R}$, $g(x | y) = \frac{1}{2\sqrt{2\pi}} e^{-x^2/8}$ for $x \in \mathbb{R}$. This is the PDF of the normal distribution with mean 0 and variance 4.
2. For $x \in \mathbb{R}$, $h(y | x) = \frac{1}{3\sqrt{2\pi}} e^{-y^2/18}$ for $y \in \mathbb{R}$. This is the PDF of the normal distribution with mean 0 and variance 9.
3. X and Y are independent.

Suppose that (X, Y) has the bivariate normal distribution with probability density function f defined by

$$f(x, y) = \frac{1}{\sqrt{3\pi}} \exp\left[-\frac{2}{3}(x^2 - xy + y^2)\right], \quad (x, y) \in \mathbb{R}^2 \quad (3.5.39)$$

1. Find the conditional probability density function of X given $Y = y$ for $y \in \mathbb{R}$.
2. Find the conditional probability density function of Y given $X = x$ for $x \in \mathbb{R}$.
3. Are X and Y independent?

Answer

1. For $y \in \mathbb{R}$, $g(x | y) = \sqrt{\frac{2}{3\pi}} e^{-\frac{2}{3}(x-y/2)^2}$ for $x \in \mathbb{R}$. This is the PDF of the normal distribution with mean $y/2$ and variance $3/4$.
2. For $x \in \mathbb{R}$, $h(y | x) = \sqrt{\frac{2}{3\pi}} e^{-\frac{2}{3}(y-x/2)^2}$ for $y \in \mathbb{R}$. This is the PDF of the normal distribution with mean $x/2$ and variance $3/4$.
3. X and Y are dependent.

Mixtures of Distributions

With our usual sets S and T , as above, suppose that P_x is a probability measure on T for each $x \in S$. Suppose also that g is a probability density function on S . We can obtain a new probability measure on T by averaging (or *mixing*) the given distributions according to g .

First suppose that g is the probability density function of a discrete distribution on the countable set S . Then the function \mathbb{P} defined below is a probability measure on T :

$$\mathbb{P}(B) = \sum_{x \in S} g(x) P_x(B), \quad B \subseteq T \quad (3.5.40)$$

Proof

Clearly $\mathbb{P}(B) \geq 0$ for $B \subseteq T$ and $\mathbb{P}(T) = \sum_{x \in S} g(x) 1 = 1$. Suppose that $\{B_i : i \in I\}$ is a countable, disjoint collection of subsets of T . Then

$$\mathbb{P}\left(\bigcup_{i \in I} B_i\right) = \sum_{x \in S} g(x) P_x\left(\bigcup_{i \in I} B_i\right) = \sum_{x \in S} g(x) \sum_{i \in I} P_x(B_i) = \sum_{i \in I} \sum_{x \in S} g(x) P_x(B_i) = \sum_{i \in I} \mathbb{P}(B_i) \quad (3.5.41)$$

Reversing the order of summation is justified since the terms are nonnegative.

In the setting of the previous theorem, suppose that P_x has probability density function h_x for each $x \in S$. Then \mathbb{P} has probability density function h given by

$$h(y) = \sum_{x \in S} g(x) h_x(y), \quad y \in T \quad (3.5.42)$$

Proof

As usual, we will consider the discrete and continuous cases for the distributions on T separately.

1. Suppose that T is countable so that P_x is a discrete probability measure for each $x \in S$. By definition, for each $x \in S$, $h_x(y) = \mathbb{P}_x(\{y\})$ for $y \in T$. So the probability density function h of P is given by

$$h(y) = P(\{y\}) = \sum_{x \in S} g(x) P_x(\{y\}) = \sum_{x \in S} g(x) h_x(y), \quad y \in T \quad (3.5.43)$$

2. Suppose now that P_x has a continuous distribution on $T \subseteq \mathbb{R}^k$, with PDF g_x for each $x \in S$. For $B \subseteq T$,

$$\mathbb{P}(B) = \sum_{x \in S} g(x) P_x(B) = \sum_{x \in S} g(x) \int_B h_x(y) dy = \int_B \sum_{x \in S} g(x) h_x(y) dy = \int_B h(y) dy \quad (3.5.44)$$

So by definition, h is the PDF of \mathbb{P} . Again, the interchange of sum and integral is justified because the functions are nonnegative. Technically, we also need $y \mapsto h_x(y)$ to be measurable for $x \in S$ so that the integral makes sense.

Conversely, given a probability density function g on S and a probability density function h_x on T for each $x \in S$, the function h defined in the previous theorem is a probability density function on T .

Suppose now that g is the probability density function of a continuous distribution on $S \subseteq \mathbb{R}^j$. Then the function \mathbb{P} defined below is a probability measure on T :

$$\mathbb{P}(B) = \int_S g(x) P_x(B) dx, \quad B \subseteq T \quad (3.5.45)$$

Proof

The proof is just like the proof of Theorem (45) with integrals over S replacing the sums over S . Clearly $\mathbb{P}(B) \geq 0$ for $B \subseteq T$ and $\mathbb{P}(T) = \int_S g(x) P_x(T) dx = \int_S g(x) dx = 1$. Suppose that $\{B_i : i \in I\}$ is a countable, disjoint collection of subsets of T . Then

$$\mathbb{P}\left(\bigcup_{i \in I} B_i\right) = \int_S g(x) P_x\left(\bigcup_{i \in I} B_i\right) dx = \int_S g(x) \sum_{i \in I} P_x(B_i) dx = \sum_{i \in I} \int_S g(x) P_x(B_i) dx = \sum_{i \in I} \mathbb{P}(B_i) \quad (3.5.46)$$

Reversing the integral and the sum is justified since the terms are nonnegative. Technically, we need the subsets of T and the mapping $x \mapsto P_x(B)$ to be measurable.

In the setting of the previous theorem, suppose that P_x is a discrete (respectively continuous) distribution with probability density function h_x for each $x \in S$. Then \mathbb{P} is also discrete (respectively continuous) with probability density function h given by

$$h(y) = \int_S g(x) h_x(y) dx, \quad y \in T \quad (3.5.47)$$

Proof

The proof is just like the proof of Theorem (46) with integrals over S replacing the sums over S .

1. Suppose that T is countable so that P_x is a discrete probability measure for each $x \in S$. By definition, for each $x \in S$, $h_x(y) = \mathbb{P}_x(\{y\})$ for $y \in T$. So the probability density function h of P is given by

$$h(y) = P(\{y\}) = \int_S g(x)P_x(\{y\}) dx = \int_S g(x)h_x(y) dx, \quad y \in T \quad (3.5.48)$$

Technically, we need $x \mapsto P_x(\{y\}) = h_x(y)$ to be measurable for $y \in T$.

2. Suppose now that P_x has a continuous distribution on $T \subseteq \mathbb{R}^k$, with PDF g_x for each $x \in S$. For $B \subseteq T$,

$$\mathbb{P}(B) = \int_S g(x)P_x(B) dx = \int_S g(x) \int_B h_x(y) dy dx = \int_B \int_S g(x)h_x(y) dx dy = \int_B h(y) dy \quad (3.5.49)$$

So by definition, h is the PDF of \mathbb{P} . Again, the interchange of sum and integral is justified because the functions are nonnegative.

Technically, we also need $(x, y) \mapsto h_x(y)$ to be measurable so that the integral makes sense.

In both cases, the distribution \mathbb{P} is said to be a *mixture* of the set of distributions $\{P_x : x \in S\}$, with mixing density g .

One can have a mixture of distributions, without having random variables defined on a common probability space. However, mixtures are intimately related to conditional distributions. Returning to our usual setup, suppose that X and Y are random variables for an experiment, taking values in S and T respectively and that X probability density function g . The following result is simply a restatement of the law of total probability.

The distribution of Y is a mixture of the conditional distributions of Y given $X = x$, over $x \in S$, with mixing density g .

Proof

Only the notation is different.

1. If X has a discrete distribuion on the countable set S then

$$\mathbb{P}(Y \in B) = \sum_{x \in S} g(x)\mathbb{P}(Y \in B \mid X = x), \quad B \subseteq T \quad (3.5.50)$$

2. If X has a continuous distribution $S \subseteq \mathbb{R}^j$ then

$$\mathbb{P}(Y \in B) = \int_S g(x)\mathbb{P}(Y \in B \mid X = x) dx, \quad B \subseteq T \quad (3.5.51)$$

Finally we note that a mixed distribution (with discrete and continuous parts) really is a mixture, in the sense of this discussion.

Suppose that \mathbb{P} is a mixed distribution on a set T . Then \mathbb{P} is a mixture of a discrete distribution and a continuous distribution.

Proof

Recall that *mixed distribution* means that T can be partitioned into a countable set D and a set $C \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}_+$ with the properties that $\mathbb{P}(\{x\}) > 0$ for $x \in D$, $\mathbb{P}(\{x\}) = 0$ for $x \in C$, and $p = \mathbb{P}(D) \in (0, 1)$. Let $S = \{d, c\}$ and define the PDF g on S by $g(d) = p$ and $g(c) = 1 - p$. Recall that the conditional distribution P_d defined by $P_d(A) = \mathbb{P}(A \cap D)/\mathbb{P}(D)$ for $A \subseteq T$ is a discrete distribution on T and similarly the conditional distribution P_c defined by $P_c(A) = \mathbb{P}(A \cap C)/\mathbb{P}(C)$ for $A \subseteq T$ is a continuous distribution on T . Clearly with this setup,

$$\mathbb{P}(A) = g(c)P_c(A) + g(d)P_d(A), \quad A \subseteq T \quad (3.5.52)$$

This page titled [3.5: Conditional 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.