5. Conditional Distributions

Basic Theory

As usual, we start with a random experiment with probability measure $\mathbb{P}$ on an underlying sample space Ω. Suppose that X is a random variable for the experiment, taking values in a set S. The purpose of this section is to study the conditional probability measure given $X = x$ for $x \in S$. Thus, if $E \subseteq \Omega$ is an event for the experiment, we would like to define and study

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

We will see that if X has a discrete distribution, no new concepts are involved, and the simple definition of conditional probability suffices. When X has a continuous distribution, however, 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 $g(x) > 0$ for $x \in S$

1. Show that if E is an event in the experiment then
$$\mathbb{P}(E|X = x) = \frac{\mathbb{P}(X = x,\ E)}{g(x)}, \ \ x \in S$$

2. Prove that if E is an event in the experiment and A is a subset of S then
$$\mathbb{P}(X \in A,\ E) = \sum\nolimits_{x \in A} \mathbb{P}(E|X = x)\, g(x)$$

Conversely, this result completely characterizes the conditional distribution given $\{X = x\}$,

3. Suppose that the function $Q(x,\ E)$ defined for $x \in S$ and for events E satisfies
$$\mathbb{P}(E,\ X \in A) = \sum\nolimits_{x \in A} Q(x,\ E)\, g(x), \quad A \subseteq S$$

   Show that $Q(x,\ E) = \mathbb{P}(E|X = x)$ for all $x \in S$ and all events E.

The Continuous Case

Suppose now that X has a continuous distribution on $S \subseteq \mathbb{R}^n$, 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 the conditional probability of an event E given $\{X = x\}$, because the conditioning event has probability 0 for every x. 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. Thus, based on the characterization above, we *define* the conditional probability

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

by the requirement that for any (measurable) subset A of S.

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

For now, we will accept the fact that $\mathbb{P}(E|X = x)$ can be defined by this condition. However, we will return to this point in the section on Conditional Expectation in the chapter on Expected Value.

Conditioning and Bayes' Theorem

Again, suppose that X is a random variable and that E is an event. From our discussion in the last two sections, we have the basic formulas for computing the probability of E by conditioning on X, in the discrete and continuous cases, respectively:

$$\mathbb{P}(E) = \sum_{x \in S} g(x)\,\mathbb{P}(E|X = x)$$
$$\mathbb{P}(E) = \int_S g(x)\,\mathbb{P}(E|X = x)dx$$

Bayes' Theorem, named after Thomas Bayes, gives a formula for the conditional probability density function of X given E, in terms of the probability density function of X and the conditional probability of E given $X = x$

⊞ 4. Suppose that X has probability density function g and that E is an event with $\mathbb{P}(E) > 0$. Show that the conditional probability density function of X given E is as follows, in the discrete and continuous cases, respectively.

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

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

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 $g(x)\mathbb{P}(E|X = x)$, the sum or integral is simply the normalizing constant.

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. Thus, suppose that Y is a random variable taking values in a set T. Then (X, Y) is a random variable taking values in the product set $S \times T$. We will assume that (X, Y) has (joint) probability density function f. (In particular, we are assuming one of the standard distribution types: jointly discrete, jointly continuous with a probability density function, or mixed components with a probability density function.) As before, g denotes the probability density function of X and we assume $g(x) > 0$ for $x \in S$.

⊞ 5. Show that the function defined below is a probability density function in $y \in T$ for each $x \in S$:

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

The next exercise shows that $y \mapsto h(y|x)$ is the conditional probability density function of Y given $X = x$

◈ 6. Prove the following results, in the case that Y has a discrete or continuous distribution, respectively: if $B \subseteq T$ then
$$\mathbb{P}(Y \in B | X = x) = \sum_{y \in B} h(y|x), \quad x \in S$$
$$\mathbb{P}(Y \in B | X = x) = \int_B h(y|x) dy, \quad x \in S$$

The following theorem gives Bayes' theorem for probability density functions. We use the notation established above, and additionally, let $g(x|y)$ denote the conditional probability density function of X at $x \in S$ given $Y = y$, for $y \in T$.

◈ 7. Prove the following results, in case that X has a discrete or continuous distribution, respectively:
$$g(x|y) = \frac{g(x) h(y|x)}{\sum_{s \in S} g(s) h(y|s)}, \quad x \in S, \ y \in T$$
$$g(x|y) = \frac{g(x) h(y|x)}{\int_S g(s) h(y|s) ds}, \quad x \in S, \ y \in T$$

In the context of Bayes' theorem, g is the **prior probability density function** of X and $x \mapsto g(x|y)$, is the **posterior probability density function** of X given $Y = y$. Note that $x \mapsto g(x|y)$ is proportional to $x \mapsto g(x) h(y|x)$. The sum or integral 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.

◈ 8. Show that the following conditions are equivalent:

  a. X and Y are independent.
  b. $h(y|x) = h(y)$ for $x \in S, \ y \in T$
  c. $g(x|y) = g(x)$ for $x \in S, \ y \in T$

In many cases, a conditional distribution arises when a parameter of a given distribution is **randomized**. Note this situation in some of the exercises that follow.

Examples and Applications

Coins and Dice

◈ 9. 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.

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

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

10. 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.

  a. Find the joint density of (N, Y)
  b. Find the density of Y.
  c. Find the conditional density of N given $Y = k$ for each $k \in \{0, 1, 2, 3, 4, 5, 6\}$

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

  a. Run the simulation of the1000 times, updating every 10 runs. Compare the empirical density function of Y with the true density function in the previous exercise
  b. Run the simulation 200 times, updating after each run. Compute the empirical conditional density function of N given $Y = k$ for each k, and compare with the density function in the previous exercise

12. 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.

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

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

  a. Run the simulation 1000 times, updating every 10 runs. Compare the empirical density function of Y with the true density in the previous exercise.
  b. Run the simulation 200 times, updating after each run. Compute the empirical conditional probability density function of X given $Y = 2$, and compare with the density function in the previous exercise.

14. 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 V denote the probability of heads of the selected coin, and Y the number of heads.

  a. Find the joint probability density function of (V, Y).
  b. Find the probability density function of Y.
  c. Find the conditional probability density function of V given $Y = k$ for $k \in \{0, 1, 2\}$.

15. Suppose that V has probability density function $g(p) = 6\,p\,(1-p)$, $0 \le p \le 1$. This is a member of the **beta** family of probability density functions; beta distributions are studied in more detail in the chapter on Special Distributions. Given $V = p$, a coin with probability of heads p is tossed 3 times. Let Y denote the number of heads.

   a. Find the joint density of (V, Y).
   b. Find the probability density of function of Y.
   c. Find the conditional probability density of V given $Y = k$ for $k \in \{0, 1, 2, 3\}$. Graph these on the same axes. Each conditional distribution is also a member of the beta family.

Compare Exercise 15 with Exercise 14. In the latter exercise, we effectively choose a coin from a box with a continuous infinity of coin types.

_____

16. 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.

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

17. Suppose that N has the Poisson distribution with parameter 1, and given $N = n$, Y has the binomial distribution with parameters n and p.

   a. Find the joint probability density function of (N, Y).
   b. Find the probability density function of Y.
   c. Find the conditional probability density function of N given $Y = k$.

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

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

19. Suppose that (X, Y) has probability density function $f(x, y) = x + y$ for $0 \le x \le 1$, $0 \le y \le 1$.

   a. Find the conditional density of X given $Y = y$.
   b. Find the conditional density of Y given $X = x$.
   c. Are X and Y independent?

20. Suppose that (X, Y) has probability density function $f(x, y) = 2(x + y)$ for $0 \leq x \leq y \leq 1$.

   a. Find the conditional density of X given $Y = y$.
   b. Find the conditional density of Y given $X = x$.
   c. Are X and Y independent?

21. Suppose that (X, Y) has probability density function $f(x, y) = 15\, x^2\, y$ for $0 \leq x \leq y \leq 1$.

   a. Find the conditional density of X given $Y = y$.
   b. Find the conditional density of Y given $X = x$.
   c. Are X and Y independent?

22. Suppose that (X, Y) has probability density function $f(x, y) = 6\, x^2\, y$ for $0 \leq x \leq 1, \; 0 \leq y \leq 1$.

   a. Find the conditional density of X given $Y = y$.
   b. Find the conditional density of Y given $X = x$.
   c. Are X and Y independent?

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

   a. Find the conditional probability density function of X given $Y = y$.
   b. Find the conditional probability density function of Y given $X = x$.
   c. Are X and Y independent?

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

   a. Find the joint density of (X, Y).
   b. Find the probability density function of Y.
   c. Find the conditional probability density function of X given $Y = y$ for $y \in (0, 1)$.

Multivariate Uniform Distributions

Multivariate uniform distributions give a geometric interpretation of some of the concepts in this section. Recall first that the standard measure λ_n on $\mathbb{R}^n$ is

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

In particular, λ_1 is the length measure on $\mathbb{R}$, λ_2 is the area measure on $\mathbb{R}^2$, and λ_3 is the volume measure on $\mathbb{R}^3$.

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 $R \subseteq \mathbb{R}^{j+k}$. Thus, by definition, the joint density function of (X, Y) is

$$f(x, y) = \frac{1}{\lambda_{j+k}(R)}, \quad (x, y) \in R$$

Let S and T be the **projections** of R onto $\mathbb{R}^j$ and $\mathbb{R}^k$ respectively, defined as follows:
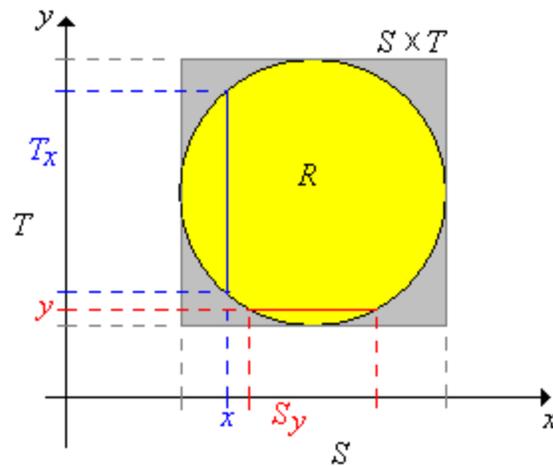
$$S = \left\{ x \in \mathbb{R}^j : (x, y) \in R \text{ for some } y \in \mathbb{R}^k \right\}$$
$$T = \left\{ y \in \mathbb{R}^k : (x, y) \in R \text{ for some } x \in \mathbb{R}^j \right\}$$

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

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

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 exercise shows, the conditional distributions are always uniform.

⊞ 25. Show that

     a. The conditional distribution of Y given $X = x$ is uniformly distributed on T_x.
     b. The conditional distribution of X given $Y = y$ is uniformly distributed on S_y.

⊞ 26. 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:

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

27. In the bivariate uniform experiment, run the simulation 5000 times, updating every 10 runs, in each of the following cases. Watch the points in the scatter plot and the graphs of the marginal distributions. Interpret what you see in the context of the discussion above.

  a. square
  b. triangle
  c. circle

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

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

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. The parameters a, b, and c are nonnegative integers with $a + b + c \le m$. We sample n objects from the population at random, and without replacement. Denote the number of type 1, 2, and 3 objects in the sample by U, V, and W, respectively. Hence, the number of type 0 objects in the sample is $n - U - V - W$. In the problems below, the variables i, j, and k are nonnegative integers.

29. Use both a combinatorial argument and an analytic argument to show that the conditional distribution of (U, V) given $W = k$ is hypergeometric and has the density function given below. The essence of the combinatorial argument is that we are selecting a random sample of size $n - k$ 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.

$$\mathbb{P}(U = i, V = j | W = k) = \frac{\binom{a}{i}\binom{b}{j}\binom{m - a - b - c}{n - i - j - k}}{\binom{m - c}{n - k}}, \quad i + j + k \le n$$

30. Use both a combinatorial argument and an analytic argument to show that the conditional distribution of U given $V = j$ and $W = k$ is hypergeometric, and has the density function given below. The essence of the combinatorial argument is that we are selecting a random sample of size $n - j - k$ from a population of size $m - b - c$, with a objects of type 1, $m - a - b - c$ objects type 0.

$$\mathbb{P}(U = i | V = j, W = k) = \frac{\binom{a}{i}\binom{m - a - b - c}{n - i - j - k}}{\binom{m - b - c}{n - j - k}}, \quad i + j + k \le n$$

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. The hypergeometric distribution and the multivariate hypergeometric distribution are studied in detail in the chapter on Finite Sampling Models.

⊞ 31. Recall that a **bridge hand** consists of 13 cards selected at random and without replacement from a standard deck of 52 cards. Let U, V, and W denote the number of spades, hearts, and diamonds, respectively, in the hand. Find the density function of each of the following:

  a. (U, V) given $W = 3$
  b. U given $V = 3$ and $W = 2$

  [?]

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 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, and r are nonnegative numbers satisfying $p + q + r \leq 1$. Denote the number of times that outcome 1, outcome 2, and outcome 3 occurred in the n trials by U, V, and W respectively. Of course, the number of times that outcome 0 occurs is $n - U - V - W$. In the problems below, variables i, j, and k are nonnegative integers.

⊞ 32. Use a probability argument and an analytic argument to show that the conditional distribution of (U, V) given $W = k$ is also multinomial, and has the density function given below. The essence of the probability argument is that effectively, we have $n - k$ independent trials, and on each trial, outcome 1 occurs with probability $\dfrac{p}{1-r}$ and outcome 2 with probability $\dfrac{q}{1-r}$.

$$\mathbb{P}(U = i, V = j | W = k) = \binom{n - k}{i, \, j} \left(\frac{p}{1 - r} \right)^i \left(\frac{q}{1 - r} \right)^j \left(1 - \frac{p}{1 - r} - \frac{q}{1 - r} \right)^{n - i - j - k}, \quad i + j + k \leq n$$

⊞ 33. Use a probability argument and an analytic argument to show that the conditional distribution of U given $V = j$ and $W = k$ is binomial, with the density function given below. The essence of the probability argument is that effectively, we have $n - j - k$ independent trials, and on each trial, outcome 1 occurs with probability $\dfrac{p}{1-q-r}$.

$$\mathbb{P}(U = i | V = j, W = k) = \binom{n - j - k}{i} \left(\frac{p}{1 - q - r} \right)^i \left(1 - \frac{p}{1 - q - r} \right)^{n - i - j - k}, \quad i + j + k \leq n$$

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 has a multinomial. The binomial distribution and the multinomial distribution are studied in detail in the chapter on Bernoulli Trials.

⊞ 34. Recall that an **ace-six flat die** is a standard 6-sided die in which faces 1 and 6 have probability $\frac{1}{4}$ each, while faces 2, 3, 4, and 5 have probability $\frac{1}{8}$ each. Suppose that an ace-six flat die is thrown 50 times; let X_i denote the number of times that score i occurred for $i \in \{1, 2, 3, 4, 5, 6\}$. Find the density function of each of the following:

a. (X_1, X_2, X_4, X_5) given $X_3 = 8$
b. (X_1, X_2, X_5) given $X_3 = 5$ and $X_4 = 7$
c. (X_1, X_5) given $X_2 = 4$, $X_3 = 7$, and $X_4 = 5$
d. X_3 given $X_1 = 8$, $X_2 = 4$, $X_4 = 6$, $X_5 = 7$

Bivariate Normal Distributions

35. Suppose that (X, Y) has probability density function

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

a. Find the conditional density function of X given $Y = y$.
b. Find the conditional density function of Y given $X = x$.
c. Are X and Y independent?

36. Suppose that (X, Y) has probability density function

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

a. Find the conditional density function of X given $Y = y$.
b. Find the conditional density function of Y given $X = x$.
c. Are X and Y independent?

The joint distributions in the last two exercises are examples of **bivariate normal distributions**. The conditional distributions are also normal. 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.

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.

37. First suppose that S is countable, and that g is the probability density function of a discrete distribution on S. Show that $\mathbb{P}$ defined below is a probability measure on T:
$$\mathbb{P}(B) = \sum_{x \in S} g(x) P_x(B), \quad B \subseteq T$$

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

given by

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

⊞ 39. Suppose now that $S \subseteq \mathbb{R}^n$ and that g is a probability density function of a continuous distribution on S. Show that $\mathbb{P}$ defined below is a probability measure on T:

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

⊞ 40. In the setting of the previous exercise, suppose that P_x is a discrete (respectively continuous) distribution with probability density function h_x for each $x \in S$. Show that $\mathbb{P}$ is also discrete (respectively continuous) with density function h given by

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

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. Suppose that X either has a discrete or continuous distribution, with probability density function g. The following exercise is simply a restatement of the law of total probability.

⊞ 41. Show that the distribution of Y is a mixture of the conditional distributions of Y given $X = x$, over $x \in S$, with mixing density g.

⊞ 42. Suppose that X is a random variable taking values in $S \subseteq \mathbb{R}^n$, with a mixed discrete and continuous distribution. Show that the distribution of X is a mixture of a discrete distribution and a continuous distribution, in the sense defined here.