

3. Mixed Distributions

Basic Theory

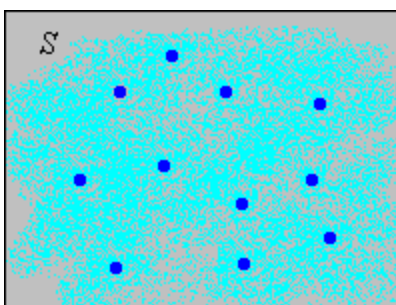
As usual, we start with a **random experiment** with **probability measure** \mathbb{P} on an underlying **sample space**. In this section, we will discuss two “mixed” cases for the distribution of a random variable: the case where the distribution is partly **discrete** and partly **continuous**, and the case where the variable has both discrete coordinates and continuous coordinates.

Distributions of Mixed Type

Suppose that X is a random variable for the experiment, taking values in $S \subseteq \mathbb{R}^n$. Then X has a **distribution of mixed type** if S can be **partitioned** into subsets D and C with the following properties:

1. D is countable and $0 < \mathbb{P}(X \in D) < 1$.
2. $\mathbb{P}(X = x) = 0$ for all $x \in C$.

Thus, part of the distribution of X is concentrated at points in a discrete set D ; the rest of the distribution is continuously spread over C . In the picture below, the light blue shading is intended to represent a continuous distribution of probability while the darker blue dots are intended to represent points of positive probability.



Let $p = \mathbb{P}(X \in D)$, so that $0 < p < 1$. We can define a function on D that is a **partial probability density function** for the discrete part of the distribution.

1. Let $g(x) = \mathbb{P}(X = x)$ for $x \in D$. Show that

- a. $g(x) \geq 0$ for $x \in D$
- b. $\sum_{x \in D} g(x) = p$
- c. $\mathbb{P}(X \in A) = \sum_{x \in A} g(x)$ for $A \subseteq D$

Usually, the continuous part of the distribution is also described by a **partial probability density function**.

Thus, suppose there is a nonnegative function h on C such that

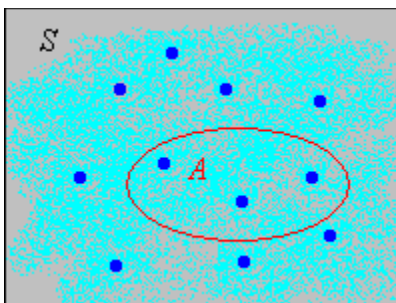
$$\mathbb{P}(X \in A) = \int_A h(x) dx \text{ for } A \subseteq C$$

2. Show that $\int_C h(x) dx = 1 - p$.

The distribution of X is completely determined by the partial probability density functions g and h . First, we extend the functions g and h to S in the usual way: $g(x) = 0$ for $x \in C$, and $h(x) = 0$ for $x \in D$.

3. Show that

$$\mathbb{P}(X \in A) = \sum_{x \in A} g(x) + \int_A h(x) dx, \quad A \subseteq S$$



The conditional distributions on D and on C are purely discrete and continuous, respectively.

4. Show that the conditional distribution of X given $X \in D$ is discrete, with probability density function

$$f(x|X \in D) = \frac{g(x)}{p}, \quad x \in D$$

5. Show that the conditional distribution of X given $X \in C$ is continuous, with probability density function

$$f(x|X \in C) = \frac{h(x)}{1 - p}, \quad x \in C$$

Thus, the distribution of X is a **mixture** of a discrete distribution and a continuous distribution. Mixtures are studied in more generality in the section on [conditional distributions](#).

Truncated Variables

Distributions of mixed type occur naturally when a random variable with a continuous distribution is *truncated* in a certain way. For example, suppose that $T \in [0, \infty)$ is the random lifetime of a device, and has a continuous distribution with probability density function f . In a test of the device, we can't wait forever, so we might select a positive constant a and record the random variable U , defined by **truncating** T at a as follows:

$$U = \begin{cases} T, & T < a \\ a, & T \geq a \end{cases}$$

6. Show that U has a mixed distribution. In particular, show that, in the notation above,

- $D = \{a\}$ and $g(a) = \int_a^\infty f(t)dt$
- $C = [0, a)$ and $h(t) = f(t)$ for $t \in [0, a)$

Suppose that random variable X has a continuous distribution on \mathbb{R} , with probability density function f . The variable is **truncated** at a and b ($a < b$) to create a new random variable Y as follows:

$$Y = \begin{cases} a, & X \leq a \\ X, & a < X < b \\ b, & X \geq b \end{cases}$$

7. Show that Y has a mixed distribution. In particular show that

- $D = \{a, b\}$, $g(a) = \int_{-\infty}^a f(x)dx$, $g(b) = \int_b^\infty f(x)dx$
- $C = (a, b)$ and $h(x) = f(x)$ for $x \in (a, b)$

Random Variable with Mixed Coordinates

Suppose X and Y are random variables for our experiment, and that X has a discrete distribution, taking values in a countable set S while Y has a continuous distribution on $T \subseteq \mathbb{R}^n$

8. Show that $\mathbb{P}((X, Y) = (x, y)) = 0$ for $(x, y) \in S \times T$. Thus (X, Y) has a continuous distribution on $S \times T$.

Usually, (X, Y) has a probability density function f on $S \times T$ in the following sense:

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

9. More generally, for $C \subseteq S \times T$ and $x \in S$, define the **cross section** at x by $C(x) = \{y \in T : (x, y) \in C\}$

Show that

$$\mathbb{P}((X, Y) \in C) = \sum_{x \in S} \int_{C(x)} f(x, y)dy, \quad C \subseteq S \times T$$

Technically, f is the probability density function of (X, Y) with respect to the product measure on $S \times T$ formed from counting measure on S and n -dimensional measure λ_n on T .

Random vectors with mixed coordinates arise naturally in applied problems. For example, the **cicada data set** has 4 continuous variables and 2 discrete variables. The **M&M data set** has 6 discrete variables and 1

continuous variable. Vectors with mixed coordinates also occur when a discrete parameter for a continuous distribution is randomized, or when a continuous parameter for a discrete distribution is randomized.

Examples and Applications

10. Suppose that X has probability $\frac{1}{2}$ uniformly distributed on the set $\{1, 2, \dots, 8\}$ and has probability $\frac{1}{2}$ uniformly distributed on the interval $[0, 10]$. Find $\mathbb{P}(X > 6)$.



11. Suppose that (X, Y) has probability $\frac{1}{3}$ uniformly distributed on $\{0, 1, 2\}^2$ and has probability $\frac{2}{3}$ uniformly distributed on $[0, 2]^2$. Find $\mathbb{P}(Y > X)$.



12. Suppose that the lifetime T of a device (in 1000 hour units) has the exponential distribution with probability density function $f(t) = e^{-t}$, $t \geq 0$. A test of the device is terminated after 2000 hours; the truncated lifetime U is recorded. Find each of the following:



- $\mathbb{P}(U < 1)$
- $\mathbb{P}(U = 2)$

13. Let

$$f(x, y) = \begin{cases} \frac{1}{3}, & x = 1, 0 \leq y \leq 1 \\ \frac{1}{6}, & x = 2, 0 \leq y \leq 2 \\ \frac{1}{9}, & x = 3, 0 \leq y \leq 3 \end{cases}$$

- Show that f is a mixed density in the sense defined above, with $S = \{1, 2, 3\}$ and $T = [0, 3]$
- Find $\mathbb{P}(X > 1, Y < 1)$.



14. Let $f(p, k) = 6 \binom{3}{k} p^{k+1} (1-p)^{4-k}$ for $k \in \{0, 1, 2, 3\}$ and $p \in [0, 1]$.

- Show that f is a mixed probability density function in the sense defined above.
- Find $\mathbb{P}(V < \frac{1}{2}, X = 2)$ where (V, X) is a random vector with probability density function f .



As we will see in the section on [conditional distributions](#), the distribution in the last exercise models the following experiment: a random probability V is selected, and then a coin with this probability of heads is tossed 3 times; X is the number of heads.

15. For the [M&M data](#), let N denote the total number of candies and W the net weight (in grams).
Construct an empirical density function for (N, W)

[Virtual Laboratories](#) > [3. Distributions](#) > [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#)

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Key words](#) | [Feedback](#) | ©