4. Joint Distributions

Basic Theory

As usual, we start with a random experiment with probability measure $\mathbb{P}$ on an underlying sample space. Suppose now that X and Y are random variables for the experiment, and that X takes values in S while Y takes values in T. We can think of (X, Y) as a random variable taking values in the product set $S \times T$. The purpose of this section is to study how the distribution of (X, Y) is related to the distributions of X and Y individually. In this context, the distribution of (X, Y) is called the **joint distribution**, while the distributions of X and of Y are referred to as **marginal distributions**. Note that X and Y themselves may be vector valued.

The first simple, but very important point, is that the marginal distributions can be obtained from the joint distribution, but not conversely.

1. Show that

a. $\mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times T)$ for $A \subseteq S$
b. $\mathbb{P}(Y \in B) = \mathbb{P}((X, Y) \in S \times B)$ for $B \subseteq T$

If X and Y are independent, then by definition,

$$\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B), \quad A \subseteq S,\ B \subseteq T$$

and as we have noted before, this completely determines the distribution (X, Y) on $S \times T$. However, if X and Y are dependent, the joint distribution cannot be determined from the marginal distributions. Thus in general, the joint distribution contains much more information than the marginal distributions individually.

Joint and Marginal Densities

In the discrete case, note that $S \times T$ is countable if and only if S and T are countable.

2. Suppose that (X, Y) has a discrete distribution with probability density function f on a countable set $S \times T$. Show that X and Y have density functions g and h, respectively, given by

a. $g(x) = \sum_{y \in T} f(x, y), \quad x \in S$
b. $h(y) = \sum_{x \in S} f(x, y), \quad y \in T$

For the continuous case, suppose that $S \subseteq \mathbb{R}^j$ and $T \subseteq \mathbb{R}^k$ so that $S \times T \subseteq \mathbb{R}^{j+k}$

3. Suppose that (X, Y) has a continuous distribution on $S \times T$ with probability density function f. Show that X and Y have continuous distributions with probability density functions g and h, respectively, given by

a. $g(x) = \int_T f(x, y) d y, \quad x \in S$
b. $h(y) = \int_S f(x, y) d x, \quad y \in T$

In the context of Exercises 2 and 3, f is called the **joint probability density function** of (X, Y), while g and h are called the **marginal density functions** of X and of Y, respectively.

Independence

When the variables are independent, the joint density is the product of the marginal densities.

4. Suppose that X and Y are independent, either both with discrete distributions or both with continuous distributions. Let g and h denote the probability density functions of X and Y respectively. Show that (X, Y) has probability density function f given by

$$f(x, y) = g(x) h(y), \quad (x, y) \in S \times T$$

The following exercise gives a converse to Exercise 4. If the joint probability density factors into a function of x only and a function of y only, then X and Y are independent.

5. Suppose that (X, Y) has either a discrete or continuous distribution, with probability density function f. Suppose that

$$f(x, y) = u(x) v(y), \quad (x, y) \in S \times T$$

where $u : S \to [0, \infty)$ and $v : T \to [0, \infty)$. Show that X and Y are independent and that there exists a nonzero constant c such that $g(x) = c\, u(x), \quad x \in S$ is a probability density function for X and $h(y) = \frac{1}{c} v(y), \quad y \in T$ is a probability density function for Y.

Random Variables with Mixed Coordinates

The results of this section have natural analogies in the case that (X, Y) has coordinates with different distribution types, as discussed in the section on mixed distributions. For example, suppose X has a discrete distribution, Y has a continuous distribution, and (X, Y) has joint probability density function f on $S \times T$. Then the results in exercises 2(a), 3(b), 4, and 5 hold.

Examples and Applications

Dice

6. Suppose that two standard, fair dice are rolled and the sequence of scores (X_1, X_2) recorded. Let $Y = X_1 + X_2$ and $Z = X_1 - X_2$ denote the sum and difference of the scores, respectively.

a. Find the probability density function of (Y, Z).
b. Find the probability density function of Y
c. Find the probability density function of Z.

d. Are Y and Z independent?

7. Suppose that two standard, fair dice are rolled and the sequence of scores (X_1, X_2) 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 probability density function of (U, V).
b. Find the probability density function of U
c. Find the probability density function of V.
d. Are U and V independent?

---

8. 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 probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

12. Suppose that (X, Y, Z) has probability density function $f(x, y, z) = 2(x + y)z$ for $0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le z \le 1$.

a. Find the probability density function of each pair of variables.
b. Find the probability density function of each variable.
c. Determine the dependency relationships between the variables.

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

Multivariate Uniform Distributions

Multivariate uniform distributions give a geometric interpretation of some of the concepts in this section. Recall first that the standard Lebesgue measure 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 probability density function of (X, Y) is

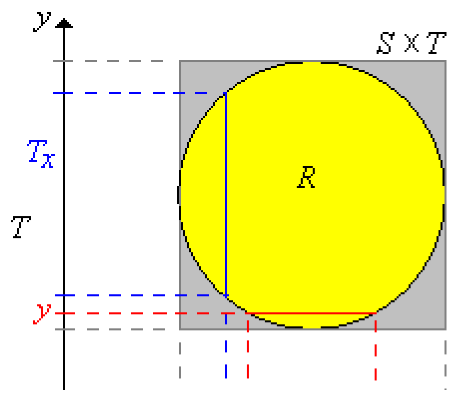
$$f(x, y) = \frac{1}{\lambda_{j+k}(R)}, \quad (x, y) \in R$$

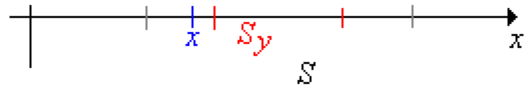
Let S and T be the **projections** of R onto $\mathbb{R}^j$ and $\mathbb{R}^k$ respectively, defined as follows:

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

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

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





14. Show that X takes values in S and that the probability density function g of X is proportional to the cross-sectional measure:

$$g(x) = \frac{\lambda_k(T_x)}{\lambda_{j+k}(R)}, \quad x \in S$$

15. Show that Y takes values in T and that the probability density function h of Y is proportional to the cross-sectional measure:

$$h(y) = \frac{\lambda_j(S_y)}{\lambda_{j+k}(R)}, \quad y \in T$$

In particular, note from previous exercises that X and Y are not in general either independent nor uniformly distributed.

16. Suppose that $R = S \times T$. Show that

a. X is uniformly distributed on S.
b. Y is uniformly distributed on T.
c. X and Y are independent.

17. In each of the following cases, find the joint and marginal densities, and determine if X and Y are independent.

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

18. In the bivariate uniform experiment, run the simulation 5000 times, updating every 10 runs for 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

19. Suppose that (X, Y, Z) is uniformly distributed on the cube $[0, 1]^3$.

a. Give the joint probability density function of (X, Y, Z).
b. Find the probability density function of each pair of variables.
c. Find the probability density function of each variable
d. Determine the dependency relationships between the variables.

20. Suppose that (X, Y, Z) is uniformly distributed on $\{(x, y, z) : 0 \le x \le y \le z \le 1\}$.

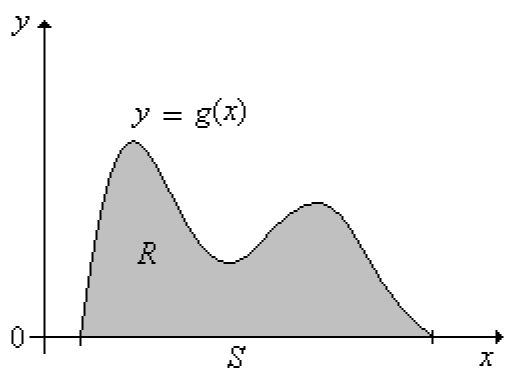
a. Give the joint density function of (X, Y, Z).
b. Find the probability density function of each pair of variables.
c. Find the probability density function of each variable
d. Determine the dependency relationships between the variables.

The following exercise shows how an arbitrary continuous distribution can be obtained from a uniform distribution. This result is useful for **simulating** certain continuous distributions.

21. Suppose that g is a probability density function for a continuous distribution on $S \subseteq \mathbb{R}^n$. Let

$$R = \{(x, y) : (x \in S) \text{ and } (0 \le y \le g(x))\} \subseteq \mathbb{R}^{n+1}$$

Show that if (X, Y) is uniformly distributed on R, then X has probability density function g. A picture in the case $n = 1$ is given below:



The Multivariate Hypergeometric Distribution

Suppose that a population consists of m objects, and that each object is one of four types. There are a type 1 objects, b type 2 objects, c type 3 objects and $m - a - b - c$ type 0 objects. 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 take values in the set $\{0, 1, \ldots, n\}$.

22. Use a combinatorial argument to show that (U, V, W) has a **(multivariate) hypergeometric distribution**, with probability density function:

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

23. Use both a combinatorial argument and an analytic argument to show that (U, V) also has a (multivariate) hypergeometric distribution, with the probability density function given below. The essence of the combinatorial

argument is that we are selecting a random sample of size n from a population of m objects, with a objects of type 1, b objects of type 2, and $m - a - b$ objects of other types.

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

24. Use both a combinatorial argument and an analytic argument to show that U has an ordinary hypergeometric distribution, with the probability density function given below. The essence of the combinatorial argument is that we are selecting a random sample of size n from a population of size m, with a objects of type 1 and $m - a$ objects of other types.

$$\mathbb{P}(U = i) = \frac{\binom{a}{i}\binom{m-a}{n-i}}{\binom{m}{n}}, \quad i \in \{0,\ 1,\ \ldots,\ n\}$$

These results generalize in a straightforward way to a population with any number of types. In brief, if a random vector has a hypergeometric distribution, then any sub-vector also has a hypergeometric distribution. In other terms, all of the marginal distributions of a hypergeometric distribution are themselves hypergeometric. The hypergeometric distribution and the multivariate hypergeometric distribution are studied in detail in the chapter on Finite Sampling Models.

25. 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. $(U,\ V,\ W)$
b. $(U,\ V)$
c. U

Multinomial Trials

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 occurs with probability $1 - p - q - r$. The parameters p, q, and r are nonnegative numbers with $p + q + r \le 1$. 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, the variables i, j, and k take values in the set $\{0,\ 1,\ \ldots,\ n\}$.

26. Use a probability argument (based on combinatorics and independence) to show that $(U,\ V,\ W)$ has a **multinomial distribution**, with probability density function given by

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

27. Use a probability argument and an analytic argument to show that $(U,\ V)$ also has a multinomial distribution, with the probability density function given below. The essence of the probability argument is that we have n independent trials, and on each trial, outcome 1 occurs with probability p, outcome 2 with probability q, and some

other outcome with probability $1 - p - q$.

$$\mathbb{P}(U = i, V = j) = \binom{n}{i,\ j} p^i q^j (1 - p - q)^{n-i-j}, \quad i + j \le n$$

28. Use a probability argument and an analytic argument to show that U has a **binomial distribution**, with the probability density function given below. The essence of the probability argument is that we have n independent trials, and on each trial, outcome 1 occurs with probability p and some other outcome with probability $1 - p$

$$\mathbb{P}(U = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i \le n$$

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 any sub-vector also has a multinomial distribution. In other terms, all of the marginal distributions of a multinomial distribution are themselves multinomial. The binomial distribution and the multinomial distribution are studied in detail in the chapter on Bernoulli Trials.

29. 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 10 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. $(X_1, X_2, X_3, X_4, X_5)$
b. (X_1, X_3, X_4, X_5)
c. (X_3, X_5, X_6)
d. (X_1, X_3)
e. X_3

Bivariate Normal Distributions

30. Suppose that (X, Y) has probability the density function given below:

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

31. Suppose that (X, Y) has the probability density function given below:

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

a. Find the density function of X.
b. Find the density function of Y.

c. Are X and Y independent?

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

The Exponential Distribution

Recall that the **exponential distribution** has probability density function

$$f(x) = r\, e^{-r\, x}, \quad x \geq 0$$

where $r > 0$ is the **rate parameter**. The exponential distribution is widely used to model random times, and is studied in more detail in the chapter on the Poisson Process.

32. Suppose X and Y have exponential distributions with parameters a and b, respectively, and are independent. Show that $\mathbb{P}(X < Y) = \frac{a}{a+b}$

33. Suppose X, Y, and Z have exponential distributions with parameters a, b, and c, respectively, and are independent. Show that

a. $\mathbb{P}(X < Y < Z) = \frac{a}{a+b+c}\,\frac{b}{b+c}$
b. $\mathbb{P}(X < Y,\ X < Z) = \frac{a}{a+b+c}$

If X, Y, and Z are the lifetimes of devices that act independently, then the results in the previous two exercises give probabilities of various failure orders. Results of this type are also very important in the study of continuous-time Markov processes. We will continue this discussion in the section on transformations of random variables.

---

34. Suppose X takes values in the finite set $\{1, 2, 3\}$, Y takes values in the interval $[0, 3]$, and that the joint density function f is given by

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

a. Find the probability density function of X.
b. Find the probability density function of Y.
c. Are X and Y independent?

35. Suppose that V takes values in the interval $[0, 1]$, X takes values in the finite set $\{0, 1, 2, 3\}$, and that (V, X) has joint probability density function f given by

$$f(p, k) = 6 \binom{3}{k} p^{k+1} (1 - p)^{4-k}, \quad (p, k) \in [0, 1] \times \{0, 1, 2, 3\}$$

.

a. Find the probability density function of V.
b. Find the probability density function of X.
c. Are V and X independent?

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.

Data Analysis Exercises

36. For the cicada data, G denotes gender and S denotes species type.

a. Find the empirical density of (G, S).
b. Find the empirical density of G.
c. Find the empirical density of S.
d. Do you believe that S and G are independent?

37. For the cicada data, W_B denotes body weight and L_B denotes body length (in mm).

a. Construct an empirical density for (W_B, L_B).
b. Find the corresponding empirical density for W_B .
c. Find the corresponding empirical density for L_B .
d. Do you believe that W_B and L_B are independent?

38. For the cicada data, G denotes gender and W_B denotes body weight (in grams).

a. Construct and empirical density for (G, W_B).
b. Find the empirical density for G.
c. Find the empirical density for W_B .
d. Do you believe that G and W_B are independent?