# Chapter 3: Joint Distributions

Donghyun Ko

May 27, 2026

A *joint distribution* describes how multiple random variables behave *together*. It extends one-variable probability to the multivariate setting and enables us to talk about dependence, conditioning, and transformations.

## Contents

# 1 Why this post? What you'll learn

In introductory probability, we often study a single random variable $X$ and its CDF $F_X(x)$, PMF $p_X(x)$, or PDF $f_X(x)$. But most real data come as *vectors*: $(X, Y)$, or even $(Y_1, \ldots, Y_K)$. Once we have more than one random variable, two new questions appear:

- **How do we compute probabilities of events involving both variables?** For example, $\mathbb{P}(X \le 2,\ Y > 0)$ or $\mathbb{P}(1 < X < 3,\ 0 < Y < 2)$.
- **How do we describe dependence?** Even if we know the distributions of $X$ and $Y$ individually, that does not tell us whether they tend to be large together, move in opposite directions, or are independent.

This note expands a concise lecture slide given by Prof. Ana-Maria Staicu in the Statistics department of NC state university, and aligns the main ideas with Rice's Chapter 3 topics. What you'll be able to do after reading this posting are:

- Define and interpret the joint CDF and compute rectangle probabilities;
- Obtain marginal and conditional distributions from a joint PMF/PDF;
- Test independence in discrete and continuous settings;
- Use the change-of-variables formula (Jacobian) to find distributions of transformations;
- Derive distributions of maxima/minima and the $k$th order statistic.

# 2 Joint cumulative distribution function (Joint CDF)

## 2.1 Definition

Let $X$ and $Y$ be random variables on the same probability space. The most general way to describe their joint behavior is the joint CDF:

$$F_{X,Y}(x, y) = \mathbb{P}(X \le x,\ Y \le y), \quad \text{where} \quad x, y \in \mathbb{R}.$$

Think of $(X, Y)$ as a random point in the plane. Then, $F_{X,Y}(x, y)$ is the probability that the point lands in the region to the *left* of the vertical line $x$ and *below* the horizontal line $y$. This is a direct multivariate analogue of the one-dimensional CDF. Once the joint distribution is specified, the distribution of each component is obtained by "ignoring" the other variable. For example, the (one-dimensional) distribution function of $X$ is

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(X \le x,\ Y \in \mathbb{R}).$$

We call $F_X$ the *marginal CDF* of $X$ to emphasize that it is derived from the joint distribution of $(X, Y)$ by marginalizing over $Y$. Similarly, the marginal CDF of $Y$ is

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X \in \mathbb{R},\ Y \le y).$$

Equivalently, using the defining property of joint CDFs, these marginals can be written as

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y), \qquad F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y).$$

When a joint density exists, this limit representation has a clear geometric meaning. The joint CDF $F_{X,Y}(x, y)$ is the probability accumulated over the rectangle $(-\infty, x] \times (-\infty, y]$. Letting $y \to \infty$ simply extends this rectangle to include all possible values of $Y$, so the probability reduces to that of the event $\{X \le x\}$ alone. Hence,

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y).$$

From the density viewpoint, the same idea is expressed by integrating out $Y$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

Thus, the limit in the CDF formulation and the integral in the density formulation represent the same operation—removing any restriction on $Y$—written in two different languages. The joint CDF is useful because many probabilities can be computed from it by inclusion–exclusion. In particular, probabilities of rectangles are easy.

### 2.2 A key formula: Rectangle probabilities

Let $a < b$ and $c < d$. Consider the event

$$\{a < X \le b,\ c < Y \le d\}.$$

This event is a rectangle in the plane. Using inclusion–exclusion in two dimensions,

$$\mathbb{P}(a < X \le b,\ c < Y \le d) = F(b, d) - F(a, d) - F(b, c) + F(a, c),$$

where $F$ denotes $F_{X,Y}$.

**Example**

Suppose you know the joint CDF $F$. Compute $\mathbb{P}(0 < X \le 1,\ 2 < Y \le 3)$ in terms of $F$.

**Solution.** Use the rectangle formula with $(a, b) = (0, 1)$ and $(c, d) = (2, 3)$:

$$\mathbb{P}(0 < X \le 1,\ 2 < Y \le 3) = F(1, 3) - F(0, 3) - F(1, 2) + F(0, 2).$$

### 2.3 Extension to $K$ variables

For random variables $Y_1, \ldots, Y_K$, define

$$F(y_1, \ldots, y_K) = \mathbb{P}(Y_1 \le y_1, \ldots, Y_K \le y_K).$$

The same ideas (marginalization, independence, conditioning) extend to higher dimensions, but drawings become harder, so notation must be handled carefully. Knowing $F_X$ and $F_Y$, for example in 2D probabilistic space, alone usually does *not* determine $F_{X,Y}$. Many different joint distributions can share the same marginals. What differs is dependence.

## 3 Independence and conditional random variables

We say that two random variables $X$ and $Y$ are *independent* if their joint CDF factorizes as

$$F_{X,Y}(x,y) = F_X(x)\,F_Y(y), \qquad \forall x,y \in \mathbb{R},$$

where $F_X(\cdot)$ and $F_Y(\cdot)$ are the marginal CDFs of $X$ and $Y$, respectively. This definition means that the probability of the event $\{X \le x,\ Y \le y\}$ can be computed by multiplying the probability of $\{X \le x\}$ and the probability of $\{Y \le y\}$. In other words, knowing information about $Y$ does not change the probability behavior of $X$, and vice versa.

The concept of independence extends directly to any finite collection of random variables $Y_1, \dots, Y_K$: they are independent if the joint CDF equals the product of their marginal CDFs. An important practical consequence is that when random variables are independent, the joint distribution is often much easier to compute than in the dependent case.

### 3.1 When variables are not independent

If $X$ and $Y$ are *not* independent, then the distribution of $X$ generally depends on the value taken by $Y$. In this case, the marginal CDF $F_X(x)$ no longer provides a complete description of $X$ once information about $Y$ is available.

To describe this situation, we introduce the *conditional random variable* $X \mid (Y = y)$, read as "$X$ given $Y = y$". This conditional random variable has its own distribution, called the conditional distribution of $X$ given $Y = y$. A key point is that when $X$ and $Y$ are dependent, the distribution of $X \mid (Y = y)$ *changes with* $y$.

### 3.2 Discrete and continuous cases

The form of a conditional distribution depends on whether the random variables are *discrete* or *continuous*. Although the formulas look different, the underlying idea is the same in both cases: we restrict attention to outcomes consistent with the condition and then rescale probabilities so that they form a valid distribution.

**Discrete case.** Suppose $X$ and $Y$ are discrete random variables with joint PMF $p(x,y)$. If $\mathbb{P}(Y = y) > 0$, the conditional PMF of $X$ given $Y = y$ is defined as

$$p_{X|Y}(x|y) = \mathbb{P}(X = x \mid Y = y) = \frac{p(x,y)}{p_Y(y)}, \quad \text{where}$$

$p_Y(y) = \sum_x p(x,y)$is the marginal PMF of $Y$. This formula follows directly from the definition of conditional probability.

**Continuous case.** Now suppose $(X,Y)$ has a joint PDF $f_{X,Y}(x,y)$ and marginal density $f_Y(y)$. Since events of the form $\{Y = y\}$ have probability zero, we cannot define conditional probabilities directly. Instead, we define the *conditional density* of $X$ given $Y = y$ by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad \text{for } f_Y(y) > 0,$$

where

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

In both settings, conditioning can be interpreted as first restricting attention to outcomes consistent with the condition ($Y = y$ in the discrete case, or $Y$ near $y$ in the continuous case), and then renormalizing so that the resulting probabilities sum (or integrate) to one.

Example

**Discrete example.** Suppose the joint PMF of $(X, Y)$ is given by

$$\mathbb{P}(X = 1, Y = 1) = \tfrac{1}{4}, \quad \mathbb{P}(X = 2, Y = 1) = \tfrac{1}{4}, \quad \mathbb{P}(X = 1, Y = 2) = \tfrac{1}{8}, \quad \mathbb{P}(X = 2, Y = 2) = \tfrac{3}{8}.$$

First compute the marginal PMF of $Y$:

$$p_Y(1) = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}, \qquad p_Y(2) = \tfrac{1}{8} + \tfrac{3}{8} = \tfrac{1}{2}.$$

Then, the conditional PMF of $X$ given $Y = 1$ is

$$p_{X|Y}(1|1) = \frac{1/4}{1/2} = \tfrac{1}{2}, \qquad p_{X|Y}(2|1) = \frac{1/4}{1/2} = \tfrac{1}{2}.$$

Example

**Continuous example.** Let $(X, Y)$ have joint PDF

$$f_{X,Y}(x, y) = 2, \qquad 0 < y < x < 1,$$

and $f_{X,Y}(x, y) = 0$ otherwise. First compute the marginal density of $Y$:

$$f_Y(y) = \int_y^1 2\,dx = 2(1 - y), \qquad 0 < y < 1.$$

The conditional density of $X$ given $Y = y$ is therefore

$$f_{X|Y}(x|y) = \frac{2}{2(1-y)} = \frac{1}{1-y}, \qquad y < x < 1.$$

Thus, given $Y = y$, the random variable $X$ is uniformly distributed on $(y, 1)$.

# 4 Joint distributions of discrete random variables

Assume $X$ and $Y$ take values in countable sets (finite or countably infinite). The joint PMF is

$$p(x, y) = \mathbb{P}(X = x,\ Y = y).$$

A valid joint PMF satisfies $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

## 4.1 From joint PMF to joint CDF

For discrete random variables, the joint CDF is obtained by accumulating the probability mass over all pairs $(u, v)$ that lie in the lower-left region $\{u \leq x,\ v \leq y\}$. Specifically,

$$F_{X,Y}(x, y) = \sum_{u \leq x} \sum_{v \leq y} p(u, v).$$

This formula simply adds up the probabilities of all outcomes $(u, v)$ for which $X$ does not exceed $x$ and $Y$ does not exceed $y$. Thus, the joint CDF is the discrete analogue of integrating a joint density over a region in the continuous case: sums replace integrals, and probability masses replace densities.

Example

**Example.** Suppose $(X, Y)$ takes values in $\{1, 2, 3\} \times \{1, 2, 3\}$ and has joint PMF given by

| $p(x, y)$ | $y = 1$ | $y = 2$ | $y = 3$ |
|---|---|---|---|
| $x = 1$ | 0.05 | 0.10 | 0.05 |
| $x = 2$ | 0.10 | 0.20 | 0.10 |
| $x = 3$ | 0.10 | 0.20 | 0.10 |

Compute the joint CDF at $(x, y) = (2.5, 2.5)$:

$$F_{X,Y}(2.5, 2.5) = \mathbb{P}(X \leq 2.5,\ Y \leq 2.5).$$

Since $X \leq 2.5$ corresponds to $X \in \{1, 2\}$ and $Y \leq 2.5$ corresponds to $Y \in \{1, 2\}$, we sum the joint PMF over all such pairs:

$$\begin{aligned} F_{X,Y}(2.5, 2.5) &= p(1, 1) + p(1, 2) + p(2, 1) + p(2, 2) \\ &= 0.05 + 0.10 + 0.10 + 0.20 \\ &= 0.45. \end{aligned}$$

## 4.2 Marginal PMFs

The marginal PMF of a random variable is obtained by summing the joint PMF over all possible values of the other variable. For $X$ and $Y$,

$$p_X(x) = \sum_y p(x, y), \qquad p_Y(y) = \sum_x p(x, y).$$

This procedure is called *marginalization*, because we remove one variable from the joint description while retaining the probability structure of the remaining variable. Marginalization plays a central role in multivariate probability: it allows us to recover the individual behavior of each variable from the joint distribution, even though information about their dependence is lost.

Example

**Example.**
Using the joint PMF from the previous example, we compute the marginal PMFs by applying the general marginalization formulas

$$p_X(x) = \sum_y p(x, y), \qquad p_Y(y) = \sum_x p(x, y).$$

These formulas state that a marginal PMF is obtained by summing the joint PMF over all possible values of the other variable.

**Marginal PMF of $X$.** For each fixed value of $x$, we sum over all values of $y \in \{1, 2, 3\}$:

$$p_X(x) = \sum_{y=1}^{3} p(x, y).$$

Applying this formula to each value of $x$, we obtain

$$\begin{aligned}
p_X(1) &= \sum_{y=1}^{3} p(1, y) = p(1, 1) + p(1, 2) + p(1, 3) \\
&= 0.05 + 0.10 + 0.05 = 0.20, \\
p_X(2) &= \sum_{y=1}^{3} p(2, y) = p(2, 1) + p(2, 2) + p(2, 3) \\
&= 0.10 + 0.20 + 0.10 = 0.40, \\
p_X(3) &= \sum_{y=1}^{3} p(3, y) = p(3, 1) + p(3, 2) + p(3, 3) \\
&= 0.10 + 0.20 + 0.10 = 0.40.
\end{aligned}$$

**Marginal PMF of $Y$.** Similarly, for each fixed value of $y$, we sum over all values of $x \in \{1, 2, 3\}$:

$$p_Y(y) = \sum_{x=1}^{3} p(x, y).$$

Thus,

$$\begin{aligned}
p_Y(1) &= \sum_{x=1}^{3} p(x, 1) = p(1, 1) + p(2, 1) + p(3, 1) \\
&= 0.05 + 0.10 + 0.10 = 0.25, \\
p_Y(2) &= \sum_{x=1}^{3} p(x, 2) = p(1, 2) + p(2, 2) + p(3, 2) \\
&= 0.10 + 0.20 + 0.20 = 0.50, \\
p_Y(3) &= \sum_{x=1}^{3} p(x, 3) = p(1, 3) + p(2, 3) + p(3, 3) \\
&= 0.05 + 0.10 + 0.10 = 0.25.
\end{aligned}$$

As a consistency check, note that

$$\sum_x p_X(x) = 0.20 + 0.40 + 0.40 = 1, \qquad \sum_y p_Y(y) = 0.25 + 0.50 + 0.25 = 1,$$

confirming that both marginal PMFs are valid probability distributions.

### 4.3 Conditional PMF

If $p_Y(y) > 0$, the conditional PMF of $X$ given $Y = y$ is

$$p_{X|Y}(x|y) = \mathbb{P}(X = x \mid Y = y) = \frac{p(x, y)}{p_Y(y)}.$$

This formula is just the definition of conditional probability.

Example

**Example.** Using the joint PMF from the previous example, we compute the conditional PMF of $X$ given a specific value of $Y$. Recall the definition of the conditional PMF:

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}, \qquad \text{for } p_Y(y) > 0.$$

**Step 1: Fix the value of $Y$.** Suppose we are interested in the conditional distribution of $X$ given $Y = 2$.

**Step 2: Compute the marginal probability $p_Y(2)$.** From the marginalization example,

$$p_Y(2) = p(1, 2) + p(2, 2) + p(3, 2) = 0.10 + 0.20 + 0.20 = 0.50.$$

**Step 3: Apply the conditional PMF formula.** For each possible value of $x \in \{1, 2, 3\}$,

$$p_{X|Y}(1|2) = \frac{p(1, 2)}{p_Y(2)} = \frac{0.10}{0.50} = 0.20,$$

$$p_{X|Y}(2|2) = \frac{p(2, 2)}{p_Y(2)} = \frac{0.20}{0.50} = 0.40,$$

$$p_{X|Y}(3|2) = \frac{p(3, 2)}{p_Y(2)} = \frac{0.20}{0.50} = 0.40.$$

**Step 4: Interpret and check.** The conditional PMF satisfies

$$p_{X|Y}(1|2) + p_{X|Y}(2|2) + p_{X|Y}(3|2) = 0.20 + 0.40 + 0.40 = 1,$$

so it is a valid probability distribution. This distribution describes how $X$ behaves when we know that $Y = 2$.

### 4.4 Multinomial counts

In the previous sections, we studied joint PMFs of a small number of discrete random variables such as $(X, Y)$, along with marginalization and conditioning. A very important extension of this idea arises when we consider *counts across categories*. Instead of tracking individual outcomes, we record how many times each category occurs. The resulting random vector of counts follows a *multinomial distribution*. Suppose we perform $n$ independent and identical trials, where each trial produces exactly one of $r$ categories with probabilities

$$p_1, \dots, p_r, \qquad \sum_{i=1}^{r} p_i = 1.$$

Let $N_i$ denote the number of trials that result in category $i$. Then, the random vector $(N_1, \ldots, N_r)$ is a joint discrete random variable satisfying

$$N_1 + \cdots + N_r = n.$$

The joint PMF of $(N_1, \ldots, N_r)$ is given by

$$\mathbb{P}(N_1 = n_1, \ldots, N_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}, \quad \sum_{i=1}^{r} n_i = n.$$

This is called the *multinomial distribution*. It can be viewed as a high-dimensional analogue of the binomial distribution and fits naturally into the framework of joint PMFs discussed earlier. From this perspective, a histogram is nothing more than a realization of a multinomial random vector: each bin count corresponds to one component $N_i$. Even when the underlying distribution is perfectly uniform, the counts $(N_1, \ldots, N_r)$ are random, which explains why histograms typically exhibit visible fluctuations.

Example

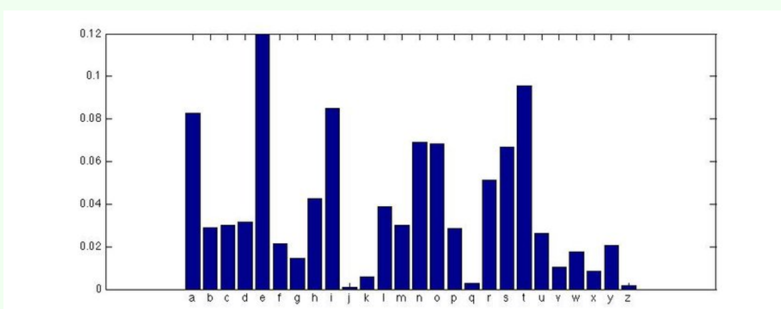
**Example.** Draw $n = 100$ i.i.d. samples from the uniform distribution $\mathrm{Unif}(0, 1)$ and divide the interval $[0, 1]$ into $r = 26$ ('a' to 'z') equal-width bins. Let $N_i$ be the number of observations falling into bin $i$. Then the vector of bin counts $(N_1, \ldots, N_{26})$ has a multinomial distribution with

$$p_i = \tfrac{1}{26}, \qquad i = 1, \ldots, 26.$$

For each bin,

$$\mathbb{E}[N_i] = np_i \approx 3.85, \qquad \mathrm{Var}(N_i) = np_i(1 - p_i) \approx 3.70.$$

Thus, although the expected count in each bin is 3–4, it is common to observe values such as 0, 1, 5 or 6 in practice. Figure 4.4 shows a histogram from one such random sample. The uneven bar heights are not evidence against uniformity; they are a direct consequence of the randomness captured by the multinomial model.




## 5 Joint distributions of continuous random variables

In this section, we extend the ideas of joint PMFs to the continuous setting. Instead of probability masses, continuous random variables are described by *joint probability density functions (PDFs)*, and probabilities are computed by integration over regions in the plane.

## 5.1 Joint PDF and probabilities of regions

Assume that $(X, Y)$ has a joint density $f(x, y)$. A function $f(x, y)$ is a valid joint PDF if

$$f(x, y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

The most important role of the joint PDF is that it allows us to compute probabilities of events by integrating over regions:

$$\mathbb{P}[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy,$$

where $A$ is any region in $\mathbb{R}^2$. This formula is the continuous analogue of summing a joint PMF over a set of points in the discrete case.

## 5.2 Support: the first thing to write down

In most applications, the joint PDF is nonzero only on a specific subset of the plane, called the *support* of $(X, Y)$. Typical supports include rectangles, triangles, or more general regions. Always identify the support before performing any calculation. Ignoring the support is one of the most common sources of errors in multivariate integration. The support determines:

- the correct limits of integration,
- when a probability is automatically zero,
- why a joint CDF often must be written in a piecewise form.

## 5.3 Relationship between the joint PDF and the joint CDF

The joint cumulative distribution function (CDF) of $(X, Y)$ is defined by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x,\ Y \leq y).$$

When a joint PDF exists, this probability is obtained by integrating the density over the rectangle $(-\infty, x] \times (-\infty, y]$:

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du.$$

Under suitable smoothness conditions, the joint PDF can be recovered from the joint CDF by differentiation.

**Theorem**

If $(X, Y)$ has joint density $f$ and joint CDF $F$, then

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du, \quad \text{and} \quad f(x, y) = \frac{\partial^2}{\partial x\, \partial y} F(x, y),$$

whenever the mixed partial derivative exists.

Example

Let $f(x,y) = \frac{12}{7}(x^2 + xy), \quad 0 \le x \le 1,\ 0 \le y \le 1$, and $f(x,y) = 0$ otherwise. Compute the joint CDF: $F(x,y) = \mathbb{P}(X \le x,\ Y \le y)$ for the interior case $0 < x < 1$ and $0 < y < 1$.

**Step 1: Identify the support and the integration region.** Since the density is nonzero only on the unit square $[0,1] \times [0,1]$, and we are assuming $0 < x < 1$ and $0 < y < 1$, the event $\{X \le x,\ Y \le y\}$ corresponds exactly to the rectangle $[0,x] \times [0,y]$ inside the support. Therefore,

$$F(x,y) = \int_0^x \int_0^y f(u,v)\,dv\,du = \int_0^x \int_0^y \frac{12}{7}\big(u^2 + uv\big)\,dv\,du.$$

**Step 2: Pull out constants and integrate with respect to $v$.** Because $\frac{12}{7}$ does not depend on $u$ or $v$, we can factor it out:

$$F(x,y) = \frac{12}{7} \int_0^x \left[\int_0^y \big(u^2 + uv\big)\,dv\right] du.$$

Now treat $u$ as a constant while integrating in $v$:

$$\int_0^y \big(u^2 + uv\big)\,dv = \int_0^y u^2\,dv + \int_0^y uv\,dv.$$

Compute each piece:

$$\int_0^y u^2\,dv = u^2 \int_0^y 1\,dv = u^2\,(y - 0) = u^2 y,$$

$$\int_0^y uv\,dv = u \int_0^y v\,dv = u\left[\frac{v^2}{2}\right]_0^y = u \cdot \frac{y^2}{2}.$$

Therefore, the inner integral equals

$$\int_0^y \big(u^2 + uv\big)\,dv = u^2 y + \frac{uy^2}{2}.$$

**Step 3: Integrate the resulting expression with respect to $u$.** Substitute the inner integral back:

$$F(x,y) = \frac{12}{7} \int_0^x \left(u^2 y + \frac{uy^2}{2}\right) du.$$

Since $y$ is constant with respect to $u$, split the integral:

$$F(x,y) = \frac{12}{7}\left[y \int_0^x u^2\,du + \frac{y^2}{2} \int_0^x u\,du\right].$$

Compute the two standard integrals:

$$\int_0^x u^2\,du = \left[\frac{u^3}{3}\right]_0^x = \frac{x^3}{3}, \qquad \int_0^x u\,du = \left[\frac{u^2}{2}\right]_0^x = \frac{x^2}{2}.$$

Hence,

$$F(x,y) = \frac{12}{7}\left[y \cdot \frac{x^3}{3} + \frac{y^2}{2} \cdot \frac{x^2}{2}\right] = \frac{12}{7}\left(\frac{yx^3}{3} + \frac{y^2x^2}{4}\right).$$

**(Optional check.)** Because $F(x,y)$ is a probability, it should be increasing in $x$ and $y$ and satisfy $0 \le F(x,y) \le 1$ on $0 < x < 1$, $0 < y < 1$. The expression above has these properties.

### 5.4 Piecewise form of the joint CDF

When the support of $f(x, y)$ is bounded, the joint CDF must be defined piecewise, because the rectangle $(-\infty, x] \times (-\infty, y]$ may:

- lie entirely outside the support,
- partially intersect the support,
- fully contain the support.

For example, if $f$ is supported on $(0, 1) \times (0, 1)$, then

$$F(x, y) = \begin{cases} 0, & x \le 0 \text{ or } y \le 0, \\ \int_0^x \int_0^y f(u, v)\, dv\, du, & 0 < x < 1,\ 0 < y < 1, \\ F_X(x), & 0 < x < 1,\ y \ge 1, \\ F_Y(y), & x \ge 1,\ 0 < y < 1, \\ 1, & x \ge 1,\ y \ge 1. \end{cases}$$

This piecewise structure reflects careful accounting of which parts of the support are included in the probability calculation.

### 5.5 Marginal densities

Marginal densities are obtained by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

When the support is bounded, the limits of integration must match the support. Marginalization allows us to recover the individual behavior of each variable from the joint distribution, although information about dependence is lost.

### 5.6 Independence and conditional densities

The random variables $X$ and $Y$ are independent if and only if

$$f(x, y) = f_X(x) f_Y(y) \quad \text{for all } x, y.$$

This criterion is often easier to check than the corresponding condition involving CDFs. If $X$ and $Y$ are not independent, the distribution of $X$ generally depends on the value of $Y$. This dependence is described by the conditional density.

If $f_Y(y) > 0$, the conditional density of $X$ given $Y = y$ is defined as

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Theorem

If $f$ is a joint density and $f_Y(y) > 0$, then $f_{X|Y}(x|y)$ is a valid density in $x$. Moreover,

$$X \perp Y \quad \Longleftrightarrow \quad f_{X|Y}(x|y) = f_X(x) \text{ for all } x, y.$$

Proof. Integrating the conditional density,

$$\int f_{X|Y}(x|y)\,dx = \frac{1}{f_Y(y)} \int f(x,y)\,dx = \frac{f_Y(y)}{f_Y(y)} = 1,$$

so it is properly normalized. If $f(x,y) = f_X(x)f_Y(y)$, then $f_{X|Y}(x|y) = f_X(x)$. Conversely, if the conditional density equals $f_X(x)$ for all $y$, then $f(x,y) = f_{X|Y}(x|y)f_Y(y) = f_X(x)f_Y(y)$, which implies independence.

# 6 Transformations of random variables: Jacobian method

Often we define new random variables as functions of old ones such that:

$$U = g_1(X,Y), \qquad V = g_2(X,Y).$$

For continuous variables, the main tool is the change-of-variables formula using a Jacobian. The key idea is: probability must be preserved when we re-express the same random point in different coordinates.

Theorem

**(Change of variables in two dimensions)**
Let $(X,Y)$ be a continuous random vector with joint density $f(x,y)$. Suppose that $(U,V) = g(X,Y)$ is a one-to-one transformation with continuous partial derivatives, and that the inverse transformation $(X,Y) = g^*(U,V)$ exists and is differentiable. Then the joint density of $(U,V)$ is given by

$$h(u,v) = f\big(g_1^*(u,v),\, g_2^*(u,v)\big)\, \left|\det \frac{\partial(x,y)}{\partial(u,v)}\right|,$$

where $\det \dfrac{\partial(x,y)}{\partial(u,v)}$ is the Jacobian determinant of the inverse transformation.

**Proof.** Let $A$ be any region in the $(u,v)$-plane, and let $B = g^*(A)$ be the corresponding region in the $(x,y)$-plane under the inverse transformation. By definition of $(U,V)$,

$$\mathbb{P}\big((U,V) \in A\big) = \mathbb{P}\big((X,Y) \in B\big).$$

Since $(X,Y)$ has joint density $f$, this probability can be written as

$$\mathbb{P}\big((X,Y) \in B\big) = \iint_B f(x,y)\,dx\,dy.$$

Using the multivariable change-of-variables formula from calculus, the integral over $B$ can be rewritten as an integral over $A$:

$$\iint_B f(x,y)\,dx\,dy = \iint_A f\big(g_1^*(u,v), g_2^*(u,v)\big) \left|\det \frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv.$$

Therefore,

$$\mathbb{P}\big((U,V)\in A\big) = \iint_A h(u,v)\,du\,dv,$$

where

$$h(u,v) = f\big(g_1^*(u,v), g_2^*(u,v)\big)\,\left|\det \frac{\partial(x,y)}{\partial(u,v)}\right|.$$

Since this holds for every region $A$, the function $h(u,v)$ is the joint density of $(U,V)$. □

Example

Let $X_1$ and $X_2$ be independent standard normal random variables, i.e.,

$$X_1 \sim N(0,1), \quad X_2 \sim N(0,1), \quad X_1 \perp X_2.$$

Define a transformation from $(X_1, X_2)$ to $(Y_1, Y_2)$ by

$$Y_1 = X_1, \qquad Y_2 = X_1 + X_2.$$

**Goal.** Find the joint density $f_{Y_1,Y_2}(y_1, y_2)$.

**Step 1: Start from the joint density of $(X_1, X_2)$.** Because $X_1$ and $X_2$ are independent and each has density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

their joint density factorizes:

$$f_{X_1,X_2}(x_1,x_2) = \phi(x_1)\phi(x_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\}.$$

**Step 2: Write the inverse transformation $(x_1, x_2) = g^*(y_1, y_2)$.** From $Y_1 = X_1$, we immediately get

$$x_1 = y_1.$$

From $Y_2 = X_1 + X_2$ and $x_1 = y_1$, we solve for $X_2$:

$$x_2 = y_2 - x_1 = y_2 - y_1.$$

So the inverse mapping is

$$(x_1, x_2) = g^*(y_1, y_2) = (y_1,\ y_2 - y_1).$$

**Step 3: Compute the Jacobian determinant of the inverse mapping.** We need the matrix of partial derivatives of $(x_1, x_2)$ with respect to $(y_1, y_2)$:

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Its determinant is

$$\det \frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = (1)(1) - (0)(-1) = 1,$$

so the absolute Jacobian determinant is $|1| = 1$.

**Step 4: Apply the change-of-variables formula.** By the 2D change-of-variables theorem,

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\big(x_1 = y_1,\ x_2 = y_2 - y_1\big) \left|\det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right|.$$

Substitute the joint density of $(X_1, X_2)$ and the Jacobian:

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left((y_1)^2 + (y_2 - y_1)^2\right)\right\} \cdot 1 \\ &= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(y_1^2 + (y_2 - y_1)^2\right)\right\}. \end{aligned}$$

This formula is valid for all $(y_1, y_2) \in \mathbb{R}^2$, since the transformation is one-to-one on $\mathbb{R}^2$.

**Step 5: Interpretation (why they are not independent).** Even though $X_1$ and $X_2$ are independent, the new variables are not: $Y_2$ contains $Y_1$ because $Y_2 = X_1 + X_2$ and $Y_1 = X_1$. Indeed,

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1,\ X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Cov}(X_1, X_2) = 1 + 0 = 1,$$

so the correlation is nonzero. Therefore $(Y_1, Y_2)$ is a bivariate normal vector with dependence between the components.

If $X$ and $Y$ are independent and you transform them separately, $U = g_1(X)$ and $V = g_2(Y)$, then $U$ and $V$ remain independent. Intuitively, you did not "mix" the two sources of randomness.

# 7 Extrema and order statistics

Order statistics describe the sorted sample:

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

They are essential in reliability (lifetimes), quality control, and nonparametric statistics.

## 7.1 Maximum and minimum

Let $X_1, \ldots, X_n$ be i.i.d. with CDF $F$ and density $f$. Define the maximum $U = \max_{1\le i\le n} X_i$ and the minimum $V = \min_{1\le i\le n} X_i$. The key observation is:

- $U \le u$ happens exactly when all $X_i \le u$,
- $V \ge v$ happens exactly when all $X_i \ge v$.

**Theorem**

Let $X_1, \ldots, X_n$ be i.i.d. continuous random variables with CDF $F$ and density $f$. Define

$$U = \max_{1\le i\le n} X_i, \qquad V = \min_{1\le i\le n} X_i.$$

Then, for all $u$ (where $F$ is defined),

$$F_U(u) = [F(u)]^n, \qquad f_U(u) = n\, f(u)\, [F(u)]^{n-1}.$$

Also, for all $v$,

$$F_V(v) = 1 - [1 - F(v)]^n, \qquad f_V(v) = n\, f(v)\, [1 - F(v)]^{n-1}.$$

**Proof.** We begin with the maximum $U$.

*Step 1 (CDF of the maximum).* By definition,

$$F_U(u) = \mathbb{P}(U \le u) = \mathbb{P}\big(\max_{1 \le i \le n} X_i \le u\big).$$

The event $\{\max_i X_i \le u\}$ occurs if and only if *every* observation is at most $u$:

$$\{\max_{1 \le i \le n} X_i \le u\} = \bigcap_{i=1}^{n} \{X_i \le u\}.$$

Therefore,

$$F_U(u) = \mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i \le u\}\Big).$$

Because $X_1, \dots, X_n$ are independent, the probability of the intersection factors:

$$F_U(u) = \prod_{i=1}^{n} \mathbb{P}(X_i \le u) = \prod_{i=1}^{n} F(u) = [F(u)]^n.$$

*Step 2 (PDF of the maximum).* Assuming $F$ is differentiable with derivative $f$, we differentiate:

$$f_U(u) = \frac{d}{du} F_U(u) = \frac{d}{du} \big[F(u)\big]^n.$$

By the chain rule,

$$\frac{d}{du} \big[F(u)\big]^n = n\, [F(u)]^{n-1}\, F'(u) = n\, [F(u)]^{n-1}\, f(u).$$

Hence,

$$f_U(u) = n\, f(u)\, [F(u)]^{n-1}.$$

Now consider the minimum $V$.

*Step 3 (CDF of the minimum).* By definition,

$$F_V(v) = \mathbb{P}(V \le v) = \mathbb{P}\big(\min_{1 \le i \le n} X_i \le v\big).$$

It is often easier to compute this via the complement event:

$$\mathbb{P}(V \le v) = 1 - \mathbb{P}(V > v).$$

The event $\{V > v\}$ means that *all* observations exceed $v$:

$$\{V > v\} = \bigcap_{i=1}^{n} \{X_i > v\}.$$

Thus,

$$\mathbb{P}(V > v) = \prod_{i=1}^{n} \mathbb{P}(X_i > v) = \prod_{i=1}^{n} \big(1 - F(v)\big) = \big[1 - F(v)\big]^n,$$

where we used $\mathbb{P}(X_i > v) = 1 - \mathbb{P}(X_i \le v) = 1 - F(v)$. Therefore,

$$F_V(v) = 1 - \big[1 - F(v)\big]^n.$$

*Step 4 (PDF of the minimum).* Differentiate $F_V(v)$:

$$f_V(v) = \frac{d}{dv} F_V(v) = \frac{d}{dv}\left(1 - \big[1 - F(v)\big]^n\right) = -\frac{d}{dv}\big[1 - F(v)\big]^n.$$

Apply the chain rule carefully. Let $g(v) = 1 - F(v)$, so $g'(v) = -f(v)$. Then

$$\frac{d}{dv}\big[g(v)\big]^n = n\,[g(v)]^{n-1} g'(v) = n\,[1 - F(v)]^{n-1}(-f(v)).$$

Hence,

$$f_V(v) = -\left(n\,[1 - F(v)]^{n-1}(-f(v))\right) = n\,f(v)\,[1 - F(v)]^{n-1}.$$

This completes the proof. □

Example

**(Series system lifetime: numerical example)** A series system fails when the *first* component fails, so the system lifetime is the minimum of component lifetimes:

$$T = \min\{X_1, \ldots, X_n\}.$$

Assume $X_1, \ldots, X_n$ are i.i.d. $\mathrm{Exp}(\lambda)$ with survival function $\mathbb{P}(X_i > t) = e^{-\lambda t}$ for $t \ge 0$.

**Step 1: Distribution of the minimum.** For $t \ge 0$,

$$\mathbb{P}(T > t) = \mathbb{P}(X_1 > t, \ldots, X_n > t) = \prod_{i=1}^{n} \mathbb{P}(X_i > t) = \left(e^{-\lambda t}\right)^n = e^{-n\lambda t}.$$

Thus $T \sim \mathrm{Exp}(n\lambda)$.

**Step 2: Plug in numbers.** Suppose each component has rate $\lambda = 0.2$ per hour (mean lifetime $1/\lambda = 5$ hours) and there are $n = 5$ components in series. Then

$$T \sim \mathrm{Exp}(n\lambda) = \mathrm{Exp}(1.0 \text{ per hour}),$$

so the system mean lifetime is

$$\mathbb{E}[T] = \frac{1}{n\lambda} = \frac{1}{1.0} = 1 \text{ hour}.$$

**Step 3: Compute a failure probability.** The probability that the system fails within 2 hours is

$$\mathbb{P}(T \le 2) = 1 - \mathbb{P}(T > 2) = 1 - e^{-n\lambda \cdot 2} = 1 - e^{-2} \approx 1 - 0.1353 = 0.8647.$$

So even though each component individually has mean lifetime 5 hours, a 5-component series system has mean lifetime 1 hour and has about an 86.5% chance to fail within 2 hours.

### 7.2 Density of the $k$th order statistic

Rice provides a standard closed form for the density of $X_{(k)}$ (continuous case). To understand the formula, think: for $X_{(k)}$ to be near $x$, we need

- $k-1$ observations below $x$,
- one observation near $x$,
- $n-k$ observations above $x$.

This produces powers of $F(x)$ and $1-F(x)$, and a combinatorial coefficient counts the number of ways such an arrangement can occur.

Theorem

**(Density of $X_{(k)}$; continuous i.i.d. case)** Let $X_1, \dots, X_n$ be i.i.d. continuous random variables with CDF $F$ and density $f$. Let $X_{(k)}$ denote the $k$th order statistic (the $k$th smallest observation). Then, $X_{(k)}$ has density

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}\, f(x)\, [F(x)]^{k-1}\, [1-F(x)]^{n-k}.$$

**Proof.** Fix a point $x$ and consider a small interval $(x, x+dx)$ with $dx > 0$ very small. The event $\{X_{(k)} \in (x, x+dx)\}$ occurs precisely when:

- exactly $k-1$ of the $n$ observations fall in $(-\infty, x]$,
- exactly one observation falls in $(x, x+dx)$,
- the remaining $n-k$ observations fall in $[x+dx, \infty)$.

Because the $X_i$ are i.i.d., for a *specified* choice of which observation plays each role, the probability of this configuration is approximately

$$[F(x)]^{k-1} \cdot \mathbb{P}(x < X \le x+dx) \cdot [1 - F(x+dx)]^{n-k}.$$

Since $\mathbb{P}(x < X \le x+dx) = F(x+dx) - F(x) \approx f(x)\,dx$ and $1 - F(x+dx) = 1 - F(x) + o(1)$ as $dx \to 0$, we may write

$$\mathbb{P}(\text{specified configuration}) \approx [F(x)]^{k-1} \cdot (f(x)\,dx) \cdot [1-F(x)]^{n-k}.$$

Now count how many ways we can choose which observations fall into each of the three categories. We must choose:

- which $k-1$ observations are below $x$,

- which 1 observation lands in $(x, x + dx)$,
- the remaining $n - k$ are then automatically above $x$.

The number of such assignments is the multinomial coefficient

$$\frac{n!}{(k-1)!\,1!\,(n-k)!}.$$

Therefore,

$$\mathbb{P}(X_{(k)} \in (x, x + dx)) \approx \frac{n!}{(k-1)!(n-k)!}\,[F(x)]^{k-1}\,(f(x)\,dx)\,[1 - F(x)]^{n-k}.$$

Finally, divide by $dx$ and take the limit as $dx \to 0$:

$$f_{X_{(k)}}(x) = \lim_{dx \to 0} \frac{\mathbb{P}(X_{(k)} \in (x, x + dx))}{dx} = \frac{n!}{(k-1)!(n-k)!}\,f(x)\,[F(x)]^{k-1}\,[1 - F(x)]^{n-k}.$$

□