Mathematical Statistics I

Chapter 3: Joint Distributions

Jesse Wheeler

Contents

1 Introduction

Introduction

- This material is based on the textbook by Rice (2007, Chapter 3).
- Our goal is to better understand the joint probability structure of more than one random variable, defined on the same sample space.
- One reason that studying joint probabilities is an important topic is that it enables us to use what we know about one variable to study another.

Joint cdf

- Just like the univariate case, the joint behavior of two random variables, X and Y, is determined by the cumulative distribution function

$$F(x, y) = P(X \leq x, Y \leq y).$$

- This is true for both discrete and continuous random variables.
- The any set $A \subset \mathbb{R}^2$, the joint cdf can give $P\big((X, Y) \in A\big)$.
- For example, let A be the rectangle defined by $x_1 < X < x_2$, and $y_1 < Y < y_2$. (It helps to draw a picture...)
- $F(x_2, y_2)$ gives $P(X < x_2, Y < y_2)$, an area that is too big, so we subtract off pieces
  - $F(x_2, y_1) = P(X < x_2, Y < y_1)$ (we already have the area $X < x_2$, but now subtract away the area $Y < y_1$).
  - $F(x_1, y_2) = P(X < x_1, Y < y_2)$ (Now subtracting the area $X < x_1$)
  - We have "double subtracted" the area $\{X < x_1, Y < y_1\}$, so we add it back.

	y			
x	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Table 1: Frequency table for X and Y, flipping a fair coin three times.

$$P\big((X,Y) \in A\big) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

- The definition also applies to more than two random variables.
- Let $X_1, \ldots, X_n$ be jointly distributed random variables defined on the same sample space. Then

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n).$$

- Like the univariate case, we can also define the pmf and pdf of jointly distributed random variables as well.

2 Discrete Random Variables

Discrete Random Variables

Definition: Joint pmf
Let X and Y be discrete random variables define on the same sample space, and take on values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, respectively. The *joint pmf* (or joint frequency function), is

$$p(x_i, y_j) = P(X = x_i, Y = y_j).$$

- For discrete RVs, it's often useful to describe the joint pmf as a frequency table.
- Suppose a fair coin is tossed 3 times. Let X denote the number of heads on the first toss, and Y the total number of heads.
- The sample space is

$$\Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}.$$

- The joint pmf can be expressed as the frequency table below (Table 1).
- Note that the probabilities in Table 1 sum to one.
- Using the probability laws we have already learned, we can calculate *marginal* probabilities.

$$\begin{aligned}
p_Y(0) &= P(Y = 0) \\
&= P(Y = 0, X = 0) + P(Y = 0, X = 1) \\
&= \frac{1}{8} + 0 = \frac{1}{8} \\
p_Y(1) &= P(Y = 1) \\
&= P(Y = 1, X = 0) + P(Y = 1, X = 1) \\
&= \frac{2}{8} + \frac{1}{8} = \frac{3}{8}.
\end{aligned}$$

- In general, to find the frequency function for Y and X, we just need to sum the appropriate columns or rows, respectively.
- $p_X(x) = \sum_i P(x, y_i)$ and $p_Y(y) = \sum_j P(x_j, y)$.
- The case with multiple random variables is similar:

$$p_{X_i}(x_i) = \sum_{x_j : j \neq i} p(x_1, x_2, \ldots, x_n).$$

- We can also get marginal frequencies for more than one variable:

$$p_{X_i X_j}(x_i, x_j) = \sum_{x_k : k \notin \{i,j\}} p(x_1, x_2, \ldots, x_n).$$

Example: Multinomial Distribution

- The *multinomial* distribution is a generalization of the binomial distribution.
- Suppose there are n independent trials, each with r possible outcomes, with probabilities $p_1, p_2, \ldots, p_r$, respectively.
- Let N_i be the total number of outcomes of type i in the n trials, with $i \in \{1, 2, \ldots, r\}$.
- The probability of any particular sequence $(N_1, N_2, \ldots, N_r) = (n_1, n_2, \ldots, n_r)$ is

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

- The total number of ways to do this was an identity from Chapter 1 (Proposition 1.3):

$$\binom{n}{n_1 \cdots n_r}.$$

- Combining this gives us the pmf of the multinomial distribution:

Multinomial Distribution
Let $N_1, N_2, \ldots, N_r$ be random variables that follow a multinomial distribution with parameters N and $(p_1, \ldots, p_r)$. The joint pmf is

$$p(n_1, n_2, \ldots, n_r) = \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

- The marginal distribution for any N_i can be found by summing the joint frequency function over the other n_j.
- While possible, this is a non-trivial algebraic exercise.
- The simple alternative is to reframe the problem: Let N_i be the number of successes in n trials, and $\tilde{N}_i = \sum_{j \neq i} N_j$ be the number of failures. The probability of success is still p_i, leaving the probability of failure to be $1 - p_i$.
- Thus, we see that the marginal distribution for N_i must follow a binomial distribution:

$$\begin{aligned} p_{N_i}(n_i) &= \sum_{n_j : j \neq i} \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r} \\ &= \binom{n}{n_i} p_i^{n_i} (1 - p_i)^{n - n_i} \end{aligned}$$

3 Continuous Random Variables

Continuous Random Variables

- Let X, Y be continuous random variables with joint cdf $F(x, y)$.
- Their *joint density function* is a piecewise continuous function of two variables, $f(x, y)$.
- A few properties:
  - $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}$ (or the support).
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.
  - For any "measureable set" $A \subset \mathbb{R}^2$, $P\big((X, Y) \in A\big) = \int \int_A f(x, y) dx dy$
  - In particular, $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv$.
- From the fundamental theorem of multivariable calculus, it follows that

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y),$$

  wherever the derivative is defined.

Finding joint probabilities
Let X, Y be jointly defined RVs with pdf

$$f(x, y) = \frac{12}{7}(x^2 + xy), \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1.$$

Find $P(X > y)$.

$$\begin{aligned} P(X > Y) &= \frac{12}{7} \int_0^1 \int_0^x (x^2 + xy) dy dx \\ &= \frac{9}{14}. \end{aligned}$$

Marginal cdf
The *marginal cdf* of X, denoted F_X, is

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(X \leq x \cap Y \in \mathbb{R}) = P(X \leq x \cap Y < \infty) \\ &= \lim_{y \to \infty} F(x, y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y) dy du. \end{aligned}$$

By taking the derivative of both sides of the equation, we get the *marginal density* of X:

$$f_X(x) = F_X'(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Calculating Marginal Densities

Using the same joint distribution as the previous example, find the marginal density of X:

$$\begin{aligned} f_X(x) &= \int_Y f(x,y)dy \\ &= \frac{12}{7}\int_0^1 (x^2+xy)dy \\ &= \frac{12}{7}\Big(x^2y+\frac{x}{2}y^2\Big)\Big|_0^1 \\ &= \frac{12}{7}\Big(x^2+\frac{x}{2}\Big) \end{aligned}$$

More than two random variables

- For several jointly continuous random variables, we can make the obvious generalizations.
- That is, to find the *marginal* densities, we need to "marginalize-" or "integrate-" out the *nusaince* variables.
- This means integrating out any combination of variables that we want.
- Example: Let X, Y, and Z be jointly continuous RVs with pdf $f(x,y,z)$. Then the two-dimensional marginal distribution of X and Z is:

$$f_{XZ}(x,z) = \int_{-\infty}^{\infty} f(x,y,z)dy.$$

Example: constructing bivariate cdfs

- Suppose that $F(x)$ and $G(y)$ are cdfs for random variables X and Y, resp.
- It can be shown that the following function, $H(x,y)$, is always a bivariate cdf for all $-1 \le \alpha \le 1$:

$$H(x,y) = F(x)G(y)\Big(1+\alpha\big(1-F(x)\big)\big(1-G(y)\big)\Big).$$

- Because $\lim_{x\to\infty} F(x) = \lim_{y\to\infty} G(x) = 1$, the marginal distributions are:

$$\begin{aligned} \lim_{y\to\infty} H(x,y) &= F(x) \\ \lim_{x\to\infty} H(x,y) &= G(y) \end{aligned}$$

- Thus, we can use this approach to build an infinite number of biviariate distributions that have a particular marginal distribution.
- One important example is when the marginal distributions are uniformly distributed.
- Let $F(x) = x, 0 \le x \le 1$, and $G(y) = y, 0 \le y \le 1$.
- By selecting $\alpha = -1$, we have

$$\begin{aligned} H(x,y) &= xy[1-(1-x)(1-y)] \\ &= x^2y + y^2x - x^2y^2, \quad 0 \le x, y \le 1. \end{aligned}$$

- The density is

$$\begin{aligned} h(x,y) &= \frac{\partial^2}{\partial x \partial y} H(x,y) \\ &= 2x + 2y - 4xy, \quad 0 \le x, y \le 1. \end{aligned}$$

- Here is a link to a 3D rendering of this function.
- Now, let's select $\alpha = 1/2$:

$$\begin{aligned} H(x,y) &= xy\Big(1 + \frac{1}{2}\big(1 - F(x)\big)\big(1 - G(y)\big)\Big) \\ &= \frac{1}{2}x^2y^2 - \frac{1}{2}x^2y - \frac{1}{2}xy^2 + \frac{3}{2}xy. \end{aligned}$$

- Taking the derivative, we get:

$$\begin{aligned} h(x,y) &= \frac{\partial^2}{\partial x \partial y} H(x,y) \\ &= 2xy - x - y + \frac{3}{2}, \quad 0 \le x, y \le 1. \end{aligned}$$

- Here is a link to a 3D rendering of this function.
- The last two joint cdfs were examples of a *copula*.

  Definition: Copulas

  A copula is a joint cdf that has uniform marginal distributions.

- Let $C(u,v)$ be a copula. One immediate consequence of the definition is that if U and V are uniform random variables, then $P(U \le u) = C(u,1) = u$, and $P(V \le v) = C(1,v) = v$.
- Let $C(u,v)$ be a copula, we will restrict ourselves to the case where it is twice differentiable, such that $c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v) \ge 0$.
- let F_X and F_Y be the cdfs of X and Y, resp.
- Now define $U = F_X(X)$, and $V = F_Y(Y)$. From Proposition 2.2, U and V are uniformly distributed.
- Now consider the function $H(x,y) = C(u,v) = C\big((F_X(x), F_Y(y)\big)$.
- Thus, by the property that $C(u,1) = u$ and $C(1,v) = v$, we have

$$\begin{aligned} C\big(F_X(x), 1\big) &= F_X(x) \\ C\big(1, F_Y(y)\big) &= F_Y(y). \end{aligned}$$

  Therefore by definition, $F_{XY}(x,y) = H(x,y) = C\big((F_X(x), F_Y(y)\big)$.

- Using the chain rule, we can differentiate to obtain

$$f_{XY}(x,y) = c\big(F_X(x), F_Y(y)\big) f_X(x) f_Y(y).$$

- *Takeaway:* We took arbitrary marginal distributions F_X and F_Y, and created a family of joint density functions, defined by *any* copula. Thus: the marginal distributions do not determine the joint distribution.

- There is a Theorem known as Sklar's Theorem (Wikipedia contributors, 2025) that generalizes this statement: All joint distributions can be expressed using a copula and marginal distributions, *and* the representation is unique.
- That is, the copula can be thought of as a way to describe the dependence between the variables in any joint distribution.

Uniform on specific region

- So far when we have talked about *uniform distributions*, we think about being uniform over $[0, 1]$, or a higher dimensional box: $[a, b]^d$.
- It's often useful to have a uniform distribution for other regions of space.
- Let $R \subset \mathbb{R}^2$ be any region of interest. The two-dimensional uniform distribution over R is defined by the probability

$$P\big((X, Y) \in A\big) = \frac{|A|}{|R|},$$

where $|\,|$ denotes the measure of the area.

- Example: Suppose a point is chosen randomly in a disk of radius 1.
- The area of the disk is $\pi r^2 = \pi$, and therefore the joint pdf for the location (X, Y) is

$$f(x, y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Now let R be the random variable denoting the distance of the point from the origin.
- Note that $R \leq r$ if and only if the point lies in a disk of radius r. This disk has area πr^2, and therefore the joint probability is

$$P(R \leq r) = \frac{\pi r^2}{\pi} = r^2, \quad 0 \leq r \leq 1.$$

- Taking a derivative, the corresponding density function is

$$f_R(r) = 2r, \quad 0 \leq r \leq 1.$$

- Now let us compute the marginal density of the x coordinate:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\pi} \times 1[x^2 + y^2 \leq 1] dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy \\ &= \frac{2}{\pi}\sqrt{1 - x^2}, \quad -1 \leq x \leq 1. \end{aligned}$$

Acknowledgments

- Compiled on August 18, 2025 using R version 4.5.1.

- We acknowledge students and instructors for previous versions of this course / slides.

References

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA. 1

Wikipedia contributors (2025). "Copula (statistics)#Sklar's Theorem: Wikipedia, The Free Encyclopedia." [Online; accessed 18-August-2025], URL https://en.wikipedia.org/w/index.php?title=Copula_(statistics)&oldid=1303484991. 8