Chapter 10

Joint densities

Consider the general problem of describing probabilities involving two random variables, X and Y. If both have discrete distributions, with X taking values $x_1, x_2, \ldots$ and Y taking values $y_1, y_2, \ldots$, then everything about the joint behavior of X and Y can be deduced from the set of probabilities

$$\mathbb{P}\{X = x_i, Y = y_j\} \qquad \text{for } i = 1, 2, \ldots \text{ and } j = 1, 2, \ldots$$

We have been working for some time with problems involving such pairs of random variables, but we have not needed to formalize the concept of a joint distribution. When both X and Y have continuous distributions, it becomes more important to have a systematic way to describe how one might calculate probabilities of the form $\mathbb{P}\{(X, Y) \in B\}$ for various subsets B of the plane. For example, how could one calculate $\mathbb{P}\{X < Y\}$ or $\mathbb{P}\{X^2 + Y^2 \leq 9\}$ or $\mathbb{P}\{X + Y \leq 7\}$?

<10.1> **Definition.** *Say that random variables X and Y have a jointly continuous distribution with* JOINT DENSITY *function $f(\cdot, \cdot)$ if*

•joint density

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

for each subset B of R^2. In particular, for a small region Δ around a point (x_0, y_0),

$$\mathbb{P}\{(x, y) \in \Delta\} \approx (\text{area of } \Delta)\, f(x_0, y_0),$$

at least if f is continuous at (x_0, y_0). □

To ensure that $\mathbb{P}\{(X, Y) \in B\}$ is nonnegative and that it equals one when B is the whole of $\mathbb{R}^2$, we must require

$$f \geq 0 \qquad \text{and} \qquad \iint \{(x, y) \in \mathbb{R}^2\} f(x, y)\, dx\, dy = 1.$$

Apart from the replacement of single integrals by double integrals, and the replacement of intervals of small length by regions of small area, the definition of a joint density is the same as the definition for densities on the real line in Chapter 6.

The small region Δ can be chosen in many ways—small rectangles, small disks, small blobs, small shapes that don't have any particular name—whatever suits the needs of a particular calculation.

<10.2> **Example.** When X has density $g(x)$ and Y has density $h(y)$, and X is independent of Y, the joint density is particularly easy to calculate. Let Δ be a small rectangle with one corner at (x_0, y_0) and small sides of length $\delta_x > 0$ and $\delta_y > 0$:

$$\Delta = \{(x, y) \in \mathbb{R}^2 : x_0 \leq x \leq x_0 + \delta_x, y_0 \leq y \leq y_0 + \delta_y\}$$

By independence,

$$\mathbb{P}\{(X, Y) \in \Delta\} = \mathbb{P}\{x_0 \leq X \leq x_0 + \delta_x\}\mathbb{P}\{y_0 \leq Y \leq y_0 + \delta_y\}$$

Invoke the defining property of the densities g and h to approximate the last product by

$$\big(g(x_0)\delta_x + \text{ smaller order terms}\big)\big(h(y_0)\delta_y + \text{ smaller order terms}\big) \approx \delta_x\delta_y g(x_0)h(y_0)$$

•marginal densities

Thus $f(x_0, y_0) = g(x_0)h(y_0)$. That is, the joint density f is the product of the MARGINAL DENSITIES g and h. The word *marginal* is used here to distinguish the joint density for (X, Y) from the individual densities g and h. □

When pairs of random variables are not independent it takes more work to find a joint density. The prototypical case, where new random variables are constructed as linear functions of random variables with a known joint density, illustrates a general method for deriving joint densities.

<10.3> **Exercise.** Suppose X and Y have a jointly continuous distribution with joint density $f(x, y)$. For constants a, b, c, d with $ad - bc \neq 0$ define

$$U = aX + bY \qquad \text{and} \qquad V = cX + dY$$

Find the joint density function $\psi(u, v)$ for (U, V).

SOLUTION: Think of the pair (U, V) are defining a new random point in $\mathbb{R}^2$. That is $(U, V) = T(X, Y)$, where T maps the point $(x, y) \in \mathbb{R}^2$ to the point $(u, v) \in \mathbb{R}^2$ with

$$u = ax + by \qquad \text{and} \qquad v = cx + dy,$$

or in matrix notation,

$$(u,\ v) = (x,\ y)A \qquad \text{where } A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Notice that $\det A = ad - bc$. The assumption that $ad - bc \neq 0$ ensures that the transformation is invertible:

$$(u,\ v)A^{-1} = (x,\ y) \qquad \text{where } A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

That is,

$$\frac{du - bv}{ad - bc} = x \qquad \text{and} \qquad \frac{-cu + av}{ad - bc} = y$$

Notice that $\det A^{-1} = 1/(ad - bc) = 1/(\det A)$

It helps to distinguish between the two roles for $\mathbb{R}^2$, referring to the domain of T as the (X, Y)-plane and the range as the (U, V)-plane.

The joint density function $\psi(u, v)$ is characterized by the property that

$$\mathbb{P}\{u_0 \leq U \leq u_0 + \delta_u, v_0 \leq V \leq v + \delta_v\} \approx \psi(u_0, v_0)\delta_u\delta_v$$

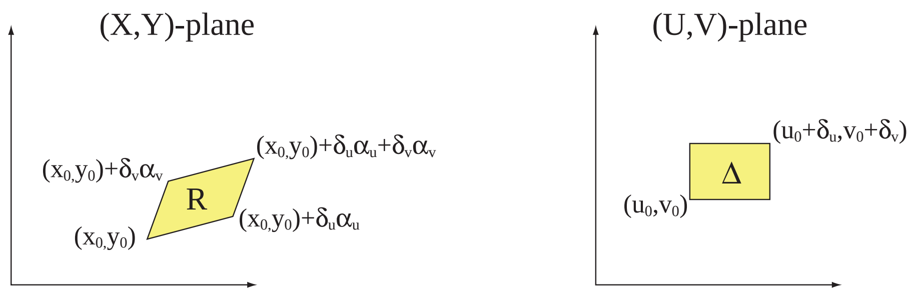
for each (u_0, v_0) in the (U, V)-plane, and small (δ_u, δ_v). To calculate the probability on the left-hand side we need to find the region R in the (X, Y)-plane corresponding to the small rectangle Δ, with corners at (u_0, v_0) and $(u_0 + \delta_u, v_0 + \delta_v)$, in the (U, V)-plane.

The linear transformation A^{-1} maps parallel straight lines in the (U, V)-plane into parallel straight lines in the (X, Y)-plane. The region R must be a parallelogram, with vertices

$$(x_0, y_0 + \delta_y) = (u_0, v_0 + \delta_v)A^{-1} \qquad \text{and} \qquad (x_0 + \delta_x, y_0 + \delta_y) = (u_0 + \delta_u, v_0 + \delta_v)A^{-1}$$
$$(x_0, y_0) = (u_0, v_0)A^{-1} \qquad \text{and} \qquad (x_0 + \delta_x, y_0) = (u_0 + \delta_u, v_0)A^{-1}$$

More succinctly,

$$\begin{aligned}(\delta_x, \delta_y) &= (\delta_u, \delta_v)A^{-1} \\ &= \delta_u\alpha_u + \delta_v\alpha_v \qquad \text{where } A^{-1} \text{ has rows } \alpha_u \text{ and } \alpha_v.\end{aligned}$$



From the formula in the Appendix, the parallelogram R has area

$$\left|\det\left(\delta_u\alpha_u', \delta_v\alpha_v'\right)\right| = \delta_u\delta_v |\det\left(A^{-1}\right)'| = \frac{\delta_u\delta_v}{|\det A|}$$

For small $\delta_u > 0$ and $\delta_v > 0$,

$$\begin{aligned}\psi(u_0, v_0)\delta_u\delta_v &\approx \mathbb{P}\{(U,V)\in\Delta\}\\ &= \mathbb{P}\{(X,Y)\in R\}\\ &\approx (\text{area of } R)\, f(x_0,y_0)\\ &\approx \delta_u\delta_v\, f(x_0,y_0)/|\det(A)|\end{aligned}$$

It follows that (U,V) have joint density

$$\psi(u,v) = \frac{1}{|\det A|} f(x,y) \qquad \text{where } (x,y) = (u,v)A^{-1}$$

In effect, we have calculated a Jacobian by first principles. □

<10.4> **Example.** Suppose X and Y are independent random variables, each distributed $N(0,1)$. By Example <10.2>, the joint density for (X,Y) equals

$$f(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)$$

By Exercise <10.3>, the joint distribution of the random variables

$$U = aX + bY \qquad \text{and} \qquad V = cX + dY$$

has the joint density

$$\begin{aligned}\psi(u,v) &= \frac{1}{2\pi(ad-bc)}\exp\left(-\frac{1}{2}\left(\frac{du-bv}{ad-bc}\right)^2 - \frac{1}{2}\left(\frac{-cu+av}{ad-bc}\right)^2\right)\\ &= \frac{1}{2\pi(ad-bc)}\exp\left(-\frac{(c^2+d^2)u^2 - 2(db+ac)uv + (a^2+b^2)v^2}{2(ad-bc)^2}\right)\end{aligned}$$

You'll learn more about joint normal distributions in Chapter 12. □

The calculations in Exercise <10.3> for linear transformations gives a good approximation for more general *smooth* transformations when applied to small regions. Densities describe the behaviour of distributions in small regions; in small regions smooth transformations are approximately linear; the density formula for linear transformations gives the density formula for smooth transformations in small regions.

<10.5> **Exercise.** Suppose X and Y are independent random variables, with X having a gamma(α) distribution and Y having a gamma(β) distribution. Show that $X/(X+Y)$ has a beta(α,β) distribution, independent of $X+Y$, which has a gamma$(\alpha+\beta)$ distribution.

SOLUTION: Write U for $X/(X+Y)$ and V for $X+Y$. The pair (X,Y) takes values ranging over the positive quadrant $(0,\infty)^2$, with joint density function

$$f(x,y) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \times \frac{y^{\beta-1}e^{-y}}{\Gamma(\beta)} \qquad \text{for } x>0, y>0.$$

The pair (U, V) takes values in the strip $(0,1) \otimes (0,\infty)$. That is, $0 < U < 1$ and $0 < V < \infty$. The joint density function, $\psi(u, v)$, for (U, V) remains to be determined.

Consider $\psi(\cdot,\cdot)$ near a point (u_0, v_0) in the strip. If $U = u_0$ and $V = v_0$ then

$$X = UV = u_0 v_0 \qquad \text{and} \qquad Y = V - UV = (1-u_0)v_0$$

Moreover, (U, V) lies near (u_0, v_0) when (X, Y) lies near (x_0, y_0), where

$$x_0 = u_0 v_0 \qquad \text{and} \qquad y_0 = (1-u_0)v_0$$

Notice how each (u_0, v_0) with $0 < u_0 < 1$ and $0 < v_0 < \infty$ corresponds to a unique (x_0, y_0) with $0 < x_0 < \infty$ and $0 < y_0 < \infty$.

For small positive δ_u and δ_v, determine the region R in the (X, Y) quadrant corresponding to the small rectangle

$$\Delta = \{(u, v) : u_0 \le u \le u_0 + \delta_u,\ v_0 \le v \le v_0 + \delta_v\}$$

in the (U, V) strip. First locate the points corresponding to the corners of Δ.

$$\begin{aligned}
(u_0 + \delta_u, v_0) &\mapsto (x_0, y_0) + (\delta_u v_0, -\delta_u v_0) \\
(u_0, v_0 + \delta_v) &\mapsto (x_0, y_0) + (\delta_v u_0, \delta_v(1-u_0)) \\
(u_0 + \delta_u, v_0 + \delta_v) &\mapsto (x_0, y_0) + (\delta_u v_0 + \delta_v u_0, -\delta_u v_0 + \delta_v(1-u_0)) + (\delta_u\delta_v, -\delta_u\delta_v)
\end{aligned}$$

In matrix notation,

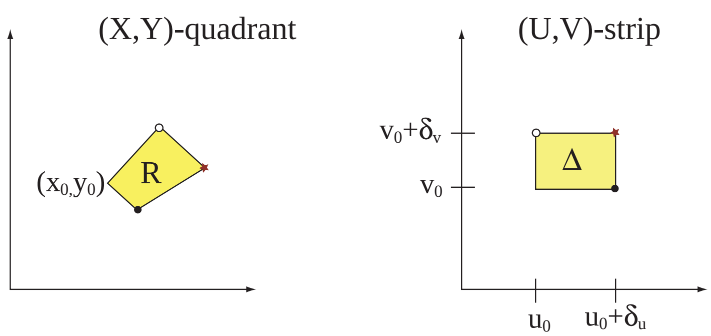
$$\begin{aligned}
(u_0, v_0) + (\delta_u, 0) &\mapsto (x_0, y_0) + (\delta_u, 0)J \qquad \text{where } J = \begin{pmatrix} v_0 & -v_0 \\ u_0 & 1-u_0 \end{pmatrix} \\
(u_0, v_0) + (0, \delta_v) &\mapsto (x_0, y_0) + (0, \delta_v)J \\
(u_0, v_0) + (\delta_u, \delta_v) &\mapsto (x_0, y_0) + (\delta_u, \delta_v)J + \text{ smaller order terms}
\end{aligned}$$

•Jacobian matrix

You might recognize J as the JACOBIAN MATRIX of partial derivatives

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix}$$

evaluated at (u_0, v_0). For small perturbations, the transformation from (u, v) to (x, y) is approximately linear.



The region R is approximately a rectangle, with the edges oblique to the coordinate axes. To a good approximation, the area of R is equal to $\delta_u\delta_v$ times the area of the rectangle with corners at

$$(0,0) \qquad \text{and} \qquad a = (v_0, -v_0) \qquad \text{and} \qquad b = (u_0, 1-u_0) \qquad \text{and} \qquad a + b$$

From the Appendix, the area of this rectangle equals $|\det(J)| = v_0$.

The rest of the calculation of the joint density $\psi(\cdot,\cdot)$ for (U, V) is easy:

$$\begin{aligned}
\delta_u\delta_v\psi(u_0, v_0) &\approx \mathbb{P}\{(U, V) \in \Delta\} \\
&= \mathbb{P}\{(X, Y) \in R\}
\end{aligned}$$

$$\approx f(x_0, y_0)(\text{area of } R)$$
$$\approx \frac{x_0^{\alpha-1}e^{-x_0}}{\Gamma(\alpha)} \frac{y_0^{\beta-1}e^{-y_0}}{\Gamma(\beta)} \delta_u \, \delta_v \, v_0$$

Substitute $x_0 = u_0 v_0$ and $y_0 = (1-u_0)v_0$, then rearrange factors, to get the joint density

$$\psi(u_0, v_0) = \frac{u_0^{\alpha-1} v_0^{\alpha-1} e^{-u_0 v_0}}{\Gamma(\alpha)} \frac{(1-u_0)^{\beta-1} v_0^{\beta-1} e^{-v_0+u_0 v_0}}{\Gamma(\beta)} v_0$$

If we write

$$g(u) = \frac{u^{\alpha-1}(1-u)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{the beta}(\alpha,\beta) \text{ density}$$
$$h(v) = \frac{v^{\alpha+\beta-1}e^{-v}}{\Gamma(\alpha+\beta)} \qquad \text{the gamma}(\alpha+\beta) \text{ density}$$

then

$$\psi(u,v) = g(u)h(v)\frac{B(\alpha,\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \qquad \text{for } 0 < u < 1 \text{ and } 0 < v < \infty$$

I have dropped the subscripting zeros because I no longer need to keep your attention fixed on a particular (u_0, v_0) in the (U, V) strip. The jumble of constants involving beta and gamma functions must reduce to the constant 1, because

$$\begin{aligned} 1 &= \mathbb{P}\{0 < U < 1, 0 < V < \infty\} \\ &= \iint \{0 < u < 1, 0 < v < \infty\}\psi(u,v)\,du\,dv \\ &= \int_0^1 g(u)\,du \int_0^\infty h(v)\,dv \frac{B(\alpha,\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

Notice how the double integral has split into a product of two single integrals because the joint density factorized into a product of a function of u and a function of v. Both the single integrals equal 1 because both g and h are density functions. We have earned a bonus,

beta vs. gamma

$$\boxed{B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \text{for } \alpha > 0 \text{ and } \beta > 0}$$

which is a useful expression relating beta and gamma functions.

The factorization of the joint density implies that the random variables U and V are independent. To see why, consider any pair of subsets A and B of the real line. The defining property of the joint density gives

$$\begin{aligned} \mathbb{P}\{U \in A\} &= \mathbb{P}\{U \in A,\ 0 < V < \infty\} \\ &= \iint \{u \in A,\ 0 < v < \infty\}\psi(u,v)\,du\,dv \\ &= \int \{u \in A\}\psi_U(u)\,du \end{aligned}$$

where $\psi_U(u) = \int_0^\infty \psi(u,v)\,dv$. That is, we get the MARGINAL DENSITY $\psi_U(u)$ for U by integrating the joint density with respect to V over its whole range. Specifically,

$$\psi_U(u) = \int_0^\infty g(u)h(v)dv = g(u)$$

That is,

U has a beta(α, β) distribution.

Similarly, V has a continuous distribution with density

$$\psi_V(v) = \int_0^1 \psi(u,v)\,du = \int_0^1 g(u)h(v)\,du = h(v)$$

That is, V has a gamma$(\alpha+\beta)$ distribution.

Finally,

$$\begin{aligned}\mathbb{P}\{U \in A,\ V \in B\} &= \iint \{u \in A, v \in B\}\psi(u,v)\,du\,dv \\ &= \int \{u \in A\}g(u)\,du \int \{v \in A\}h(v)\,dv \\ &= \mathbb{P}\{U \in A\}\mathbb{P}\{V \in B\}\end{aligned}$$

The events $\{U \in A\}$ and $\{V \in B\}$ are independent, for all choices of A and B. That is, the random variables U and V are independent. □

Remarks: In Chapter 9 we discovered that $\Gamma(1/2) = \sqrt{\pi}$. This fact also follow from the equality

$$\begin{aligned}\frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} &= B(1/2, 1/2) \\ &= \int_0^1 t^{-1/2}(1-t)^{-1/2}\,dt \\ &= \int_0^{\pi/2} \frac{1}{\sin(\theta)\cos(\theta)} 2\sin(\theta)\cos(\theta)\,d\theta \qquad \text{putting } t = \sin^2(\theta) \\ &= \pi\end{aligned}$$

□

marginal densities from joint density

Remarks: It is worthwhile to remember the method for deriving marginal densities from a joint density: In general, if X and Y have a jointly continuous distribution with density function f(x,y) then the (marginal) distribution of X is continuous, with (marginal) density

$$\int_{-\infty}^{\infty} f(x,y)\,dy,$$

and the (marginal) distribution of Y is continuous, with (marginal) density

$$\int_{-\infty}^{\infty} f(x,y)\,dx,$$

Remember that the word marginal is redundant; it serves merely to stress that a calculation refers only to one of the random variables. □

<10.6> **Example.** If $X_1, X_2, \ldots, X_k$ are independent random variables, with X_i distributed gamma(α_i) for $i = 1, \ldots, k$, then

$$\begin{aligned}&X_1 + X_2 \sim \text{gamma}(\alpha_1 + \alpha_2), \\ &X_1 + X_2 + X_3 = (X_1 + X_2) + X_3 \sim \text{gamma}(\alpha_1 + \alpha_2 + \alpha_3) \\ &X_1 + X_2 + X_3 + X_4 = (X_1 + X_2 + X_3) + X_4 \sim \text{gamma}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) \\ &\ldots \\ &X_1 + X_2 + \ldots + X_k \sim \text{gamma}(\alpha_1 + \alpha_2 + \ldots + \alpha_k)\end{aligned}$$

A particular case has great significance for Statistics.

Suppose $Z_1, \ldots Z_k$ are independent random variables, each distributed N(0,1). From Chapter 9, the random variables $Z_1^2/2, \ldots, Z_k^2/2$ are independent gamma$(1/2)$ distributed random variables. The sum

$$(Z_1^2 + \ldots + Z_k^2)/2$$

must have a gamma$(k/2)$ distribution with density $t^{k/2-1}e^{-t}/\Gamma(k/2)$ for $t > 0$. The sum $Z_1^2 + \ldots + Z_k^2$ has density

$$\frac{(t/2)^{k/2-1}e^{-t/2}}{2\Gamma(k/2)} \qquad \text{for } t > 0$$

•chi-squared

This distribution is called the CHI-SQUARED on k degrees of freedom, usually denoted by χ^2_k. The letter χ is a lowercase Greek chi. □

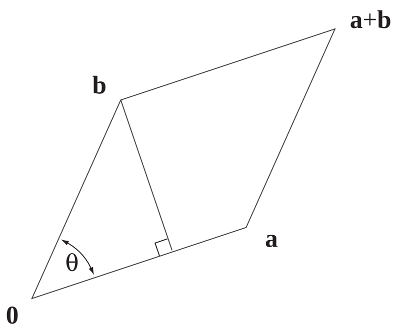
Appendix: area of a parallelogram

Let R be a rectangle in the plane $\mathbb{R}^2$ with corners at $\mathbf{0} = (0,0)$, $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, and $\mathbf{a} + \mathbf{b}$. The area of R is equal to the absolute value of the determinant of the matrix

$$J = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = (\mathbf{a}, \mathbf{b})$$

Proof. Let θ denotes the angle between $\mathbf{a}$ and $\mathbf{b}$. Remember that

$$\|\mathbf{a}\| \times \|\mathbf{b}\| \times \cos(\theta) = \mathbf{a} \cdot \mathbf{b}$$



With the side from $\mathbf{0}$ to $\mathbf{a}$, which has length $\|\mathbf{a}\|$, as the base, the vertical height is $\|\mathbf{b}\| \times |\sin\theta|$. The absolute value of the area equals $\|\mathbf{a}\| \times \|\mathbf{b}\| \times |\sin\theta|$. The square of the area equals

$$\begin{aligned} \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \sin^2(\theta) &= \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \cos^2(\theta) \\ &= (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2 \\ &= \det \begin{pmatrix} \mathbf{a} \cdot \mathbf{a} & \mathbf{a} \cdot \mathbf{b} \\ \mathbf{a} \cdot \mathbf{b} & \mathbf{b} \cdot \mathbf{b} \end{pmatrix} \\ &= \det J' J \\ &= (\det J)^2 \end{aligned}$$

If you are not sure about the properties of determinants used in the last two lines, you should rewrite the area as an explicit function of a_1, a_2, b_2, b_2 then grind out the algebra. □

If you know about Jacobians you should recognize what was going on in the proof.