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

\[
P\{X = x_i, Y = y_j\} \quad \text{for } i = 1, 2, \ldots \text{ and } j = 1, 2, \ldots
\]

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

\[
P\{X < y\} \quad \text{or} \quad P\{X^2 + Y^2 \leq 9\} \quad \text{or} \quad P\{X + Y \leq 7\}?
\]

\(<10.1>\) **Definition.** Say that random variables \(X\) and \(Y\) have a jointly continuous distribution with joint density function \(f(x, y)\) if

\[
P\{(X, Y) \in B\} = \int \int \{ (x, y) \in B \} f(x, y) \, dx \, dy
\]

for each subset \(B\) of \(\mathbb{R}^2\). In particular, for a small region \(\Delta\) around a point \((x_0, y_0)\),

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

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

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

\[
f \geq 0 \quad \text{and} \quad \int \int \{ (x, y) \in \mathbb{R}^2 \} f(x, y) \, dx \, dy = 1.
\]

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 \(\Delta\) 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 \(\Delta\) 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,

\[
P\{(X, Y) \in \Delta\} = P\{x_0 \leq X \leq x_0 + \delta_x\} P\{y_0 \leq Y \leq y_0 + \delta_y\}
\]
Invoke the defining property of the densities \( g \) and \( h \) to approximate the last product by
\[
[g(x_0)\delta_x + \text{smaller order terms}] [h(y_0)\delta_y + \text{smaller order terms}] \approx \delta_x\delta_y g(x_0)h(y_0)
\]
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.

**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
\[
U = aX + bY \quad \text{and} \quad V = cX + dY
\]
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
\[
u = ax + by \quad \text{and} \quad v = cx + dy,
\]
or in matrix notation,
\[
(u, v) = (x, y)A \quad \text{where} \quad A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}
\]
Notice that \( \det A = ad - bc \). The assumption that \( ad - bc \neq 0 \) ensures that the transformation is invertible:
\[
(u, v)A^{-1} = (x, y) \quad \text{where} \quad A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}
\]
That is,
\[
\frac{du - bv}{ad - bc} = x \quad \text{and} \quad \frac{-cu + av}{ad - bc} = y
\]
Notice that \( A^{-1} = 1/(ad - bc) = 1/|\det A| \).

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
\[
P\{u_0 \leq U \leq u_0 + \delta_u, v_0 \leq V \leq v + \delta_v\} \approx \psi(u_0, v_0)\delta_u\delta_v
\]
for each \((u_0, v_0)\) in the \((U, V)\)-plane, and small \((\delta_u, \delta_v)\). 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 \( \Delta \), 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
\[
(x_0, y_0 + \delta_y) = (u_0, v_0 + \delta_v)A^{-1} \quad \text{and} \quad (x_0 + \delta_x, y_0 + \delta_y) = (u_0 + \delta_u, v_0 + \delta_v)A^{-1}
\]
\[
(x_0, y_0) = (u_0, v_0)A^{-1} \quad \text{and} \quad (x_0 + \delta_x, y_0) = (u_0 + \delta_u, v_0)A^{-1}
\]
More succinctly,
\[
(\delta_x, \delta_y) = (\delta_u, \delta_v)A^{-1} = \delta_u \alpha_u + \delta_v \alpha_v \quad \text{where} \quad A^{-1} \text{ has rows } \alpha_u \text{ and } \alpha_v.
\]
You’ll learn more about joint normal distributions in Chapter 12.

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 $U = aX + bY$ and $V = cX + dY$

has the joint density

$$\psi(u, v) = \frac{1}{2\pi(ad - bc)} \exp \left( -\frac{1}{2} \left( \frac{du - bv}{ad - bc} \right)^2 - \frac{1}{2} \left( \frac{cu + av}{ad - bc} \right)^2 \right)$$

$$= \frac{1}{2\pi(ad - bc)} \exp \left( -\frac{(c^2 + d^2)u^2 - 2(db + ac)uv + (a^2 + b^2)v^2}{2(ad - bc)^2} \right)$$

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

Exercise. Suppose $X$ and $Y$ are independent random variables, with $X$ having a gamma($\alpha$) distribution and $Y$ having a gamma($\beta$) distribution. Show that $X/(X + Y)$ has a beta($\alpha$, $\beta$) 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

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

\[
X = U V = u_0 v_0 \quad \text{and} \quad Y = V - U V = (1 - u_0)v_0
\]

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

\[
x_0 = u_0 v_0 \quad \text{and} \quad y_0 = (1 - u_0)v_0
\]

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 \(\delta_u\) and \(\delta_v\), determine the region \(R\) in the \((X, Y)\) quadrant corresponding to the small rectangle

\[
\Delta = \{(u, v) : u_0 \leq u \leq u_0 + \delta_u, v_0 \leq v \leq v_0 + \delta_v\}
\]

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

\[
\begin{align*}
(u_0 + \delta_u, v_0) &\mapsto (x_0, y_0) + (\delta_y v_0, -\delta_u v_0) \\
(u_0, v_0 + \delta_v) &\mapsto (x_0, y_0) + (\delta_v u_0, \delta_v (1 - u_0)) \\
(u_0 + \delta_u, v_0 + \delta_v) &\mapsto (x_0, y_0) + (\delta_y v_0 + \delta_v u_0, -\delta_u v_0 + \delta_v (1 - u_0)) + (\delta_u \delta_v, -\delta_u \delta_v)
\end{align*}
\]

In matrix notation,

\[
(u_0, v_0) + (\delta_u, 0) \mapsto (x_0, y_0) + (\delta_u, 0)J \quad \text{where} \quad J = \begin{pmatrix} v_0 & -v_0 \\ u_0 & 1 - u_0 \end{pmatrix}
\]

\[
(u_0, v_0) + (0, \delta_v) \mapsto (x_0, y_0) + (0, \delta_v)J
\]

\[
(u_0, v_0) + (\delta_u, \delta_v) \mapsto (x_0, y_0) + (\delta_u, \delta_v)J + \text{smaller order terms}
\]

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

\[
(0, 0) \quad \text{and} \quad a = (v_0, -v_0) \quad \text{and} \quad b = (u_0, 1 - u_0) \quad \text{and} \quad a + b
\]

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:

\[
\delta_u \delta_v \psi(u_0, v_0) \approx \mathbb{P}\{(U, V) \in \Delta\} = \mathbb{P}\{(X, Y) \in R\}
\]
\[
\approx f(x_0, y_0) (\text{area of } R)
\]
\[
\approx \frac{x_0^{a-1} e^{-x_0} y_0^{\beta-1} e^{-y_0}}{\Gamma(a) \Gamma(\beta)} \delta_u \delta_v v_0
\]

Substitute \( x_0 = u_0 v_0 \) and \( y_0 = (1 - u_0)v_0 \), then rearrange factors, to get the joint density
\[
\psi(u_0, v_0) = \frac{u_0^{a-1} v_0^{\beta-1} e^{-u_0 v_0}}{\Gamma(a) \Gamma(\beta)} (1 - u_0)^{\beta-1} v_0^{\beta-1} e^{-(1-u_0)v_0} v_0
\]

If we write
\[
g(u) = \frac{u^{a-1} (1-u)^{\beta-1}}{B(a, \beta)} \quad \text{the beta}(a, \beta) \text{ density}
\]
\[
h(v) = \frac{v^{\alpha-1} e^{-v}}{\Gamma(\alpha)} \quad \text{the gamma}(\alpha) \text{ density}
\]

then
\[
\psi(u, v) = g(u) h(v) \frac{B(a, \beta) \Gamma(\alpha + \beta)}{\Gamma(a) \Gamma(\beta)}
\]

for \( 0 < u < 1 \) 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
\[
1 = \mathbb{P}(0 < U < 1, 0 < V < \infty)
\]
\[
= \int \int \{0 < u < 1, 0 < v < \infty\} \psi(u, v) \, du \, dv
\]
\[
= \int_0^1 g(u) \, du \int_0^\infty h(v) \, dv \frac{B(a, \beta) \Gamma(\alpha + \beta)}{\Gamma(a) \Gamma(\beta)}
\]

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,

\[
B(a, \beta) = \frac{\Gamma(a) \Gamma(\beta)}{\Gamma(a + \beta)} \quad \text{for } \alpha > 0 \text{ and } \beta > 0
\]

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
\[
\mathbb{P}(U \in A) = \mathbb{P}(U \in A, 0 < V < \infty)
\]
\[
= \int \int \{u \in A, 0 < v < \infty\} \psi(u, v) \, du \, dv
\]
\[
= \int \{u \in A\} \psi_U(u) \, du
\]

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\((a, \beta)\) 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,
\[
P(U \in A, V \in B) = \int \int \{u \in A, v \in B\} \psi(u, v) \, du \, dv
\]
\[
= \int \{u \in A\} g(u) \, du \int \{v \in A\} h(v) \, dv
\]
\[
= P(U \in A) P(V \in B)
\]

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. \( \square \)

**Remarks:** In Chapter 9 we discovered that \( \Gamma(1/2) = \sqrt{\pi} \). This fact also follows from the equality
\[
\frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} = B(1/2, 1/2)
\]
\[
= \int_0^1 t^{-1/2}(1-t)^{-1/2} \, dt
\]
\[
= \int_{\pi/2}^{\pi} \frac{1}{\sin(\theta)\cos(\theta)} 2\sin(\theta)\cos(\theta) \, d\theta \quad \text{putting } t = \sin^2(\theta)
\]
\[
= \pi
\]

\( \square \)

**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. \( \square \)

**Example.** If \( X_1, X_2, \ldots, X_k \) are independent random variables, with \( X_i \) distributed \( \text{gamma}(\alpha_i) \) for \( i = 1, \ldots, k \), then
\[
X_1 + X_2 \sim \text{gamma}(\alpha_1 + \alpha_2),
\]
\[
X_1 + X_2 + X_3 = (X_1 + X_2) + X_3 \sim \text{gamma}(\alpha_1 + \alpha_2 + \alpha_3)
\]
\[
X_1 + X_2 + X_3 + X_4 = (X_1 + X_2 + X_3) + X_4 \sim \text{gamma}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)
\]
\[
\ldots
\]
\[
X_1 + X_2 + \ldots + X_k \sim \text{gamma}(\alpha_1 + \alpha_2 + \ldots + \alpha_k)
\]

A particular case has great significance for Statistics.

Suppose \( Z_1, \ldots, Z_k \) are independent random variables, each distributed \( \text{N}(0,1) \). From Chapter 9, the random variables \( Z_1^2/2, \ldots, Z_k^2/2 \) are independent \( \text{gamma}(1/2) \) distributed random variables. The sum
\[
(Z_1^2 + \ldots + Z_k^2)/2
\]
must have a \( \text{gamma}(k/2) \) distribution with density \( t^{k/2-1} e^{-t/2} / \Gamma(k/2) \) for \( t > 0 \). The sum \( Z_1^2 + \ldots + Z_k^2 \) has density
\[
\frac{(t/2)^{k/2-1} e^{-t/2}}{2^{k/2} \Gamma(k/2)} \quad \text{for } t > 0
\]
This distribution is called the CHI-SQUARED on \( k \) degrees of freedom, usually denoted by \( \chi_k^2 \). The letter \( \chi \) is a lowercase Greek chi.

\[ \text{APPENDIX: AREA OF A PARALLELOGRAM} \]

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

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

**Proof.** Let \( \theta \) denotes the angle between \( a \) and \( b \). Remember that

\[
\|a\| \times \|b\| \times \cos(\theta) = a \cdot b
\]

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

\[
\|a\|^2 \|b\|^2 \sin^2(\theta) = \|a\|^2 \|b\|^2 - \|a\|^2 \|b\|^2 \cos^2(\theta)
\]

\[
= (a \cdot a)(b \cdot b) - (a \cdot b)^2
\]

\[
= \det \begin{pmatrix} a \cdot a & a \cdot b \\ a \cdot b & b \cdot b \end{pmatrix}
\]

\[
= \det J' J
\]

\[
= (\det J)^2
\]

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.