Chapter 11

Joint densities

11.1 Overview

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

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

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 $\mathbb{P}\{X + Y \leq 7\}$?

Definition. Say that random variables X and Y have a jointly continuous distribution with ***joint density*** function $f(\cdot, \cdot)$ if

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

for each subset B of $\mathbb{R}^2$.

Remark. To avoid messy expressions in subscripts, I will sometimes write $\iint \mathbf{1}\{(x, y) \in B\} \dots$ instead of $\iint_B \dots$.

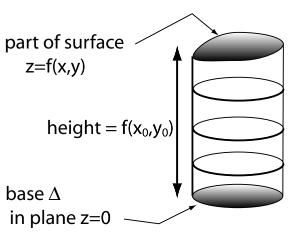
To ensure that $\mathbb{P}\{(X,Y) \in B\}$ is nonnegative and that it equals one when B is the whole of $\mathbb{R}^2$, we must require

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

The density function defines a surface, via the equation $z = f(x,y)$. The probability that the random point (X,Y) lands in B is equal to the volume of the "cylinder"

$$\{(x,y,z) \in \mathbb{R}^3 : 0 \leq z \leq f(x,y) \quad \text{and} \quad (x,y) \in B\}.$$

In particular, if Δ is small region in $\mathbb{R}^2$ around a point (x_0, y_0) at which f is continuous, the cylinder is close to a thin column with cross-section Δ and height $f(x_0, y_0)$, so that



$$\mathbb{P}\{(X,Y) \in \Delta\} = (\text{area of } \Delta) f(x_0,y_0) + \text{ smaller order terms}.$$

More formally,

$$\lim_{\Delta \downarrow \{x_0,y_0\}} \frac{\mathbb{P}\{(X,Y) \in \Delta\}}{\text{area of } \Delta} = f(x_0,y_0).$$

The limit is taken as Δ shrinks to the point (x_0, y_0).

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 essentially the same as the definition for densities on the real line in Chapter 7.

Example <11.1> Expectations of functions of random variable with jointly continuous distributions: $\mathbb{E}H(X,Y) = \iint_{\mathbb{R}^2} H(x,y) f(x,y)\,dx\,dy$.

The joint density for (X,Y) includes information about the ***marginal distributions*** of the random variables. To see why, write $A \times \mathbb{R}$ for the subset $\{(x,y) \in \mathbb{R}^2 : x \in A, y \in \mathbb{R}\}$ for a subset A of the real line. Then

$$\begin{aligned}
&\mathbb{P}\{X \in A\} \\
&\quad = \mathbb{P}\{(X,Y) \in A \times \mathbb{R}\} \\
&\quad = \iint \mathbf{1}\{x \in A, y \in \mathbb{R}\} f(x,y)\,dx\,dy \\
&\quad = \int_{-\infty}^{+\infty} \mathbf{1}\{x \in A\} \left(\int_{-\infty}^{+\infty} \mathbf{1}\{y \in \mathbb{R}\} f(x,y)\,dy \right) dx \\
&\quad = \int_{-\infty}^{+\infty} \mathbf{1}\{x \in A\} h(x)\,dx \qquad \text{where } h(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy.
\end{aligned}$$

It follows that X has a continuous distribution with ***(marginal) density*** h. Similarly, Y has a continuous distribution with (marginal) density $g(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$.

Remark. The word *marginal* is used here to distinguish the joint density for (X,Y) from the individual densities g and h.

When we wish to calculate a density, the small region Δ can be chosen in many ways—small rectangles, small disks, small blobs, and even small shapes that don't have any particular name—whatever suits the needs of a particular calculation.

Example <11.2> (Joint densities for independent random variables) Suppose X has a continuous distribution with density g and Y has a continuous distribution with density h. Then X and Y are independent if and only if they have a jointly continuous distribution with joint density $f(x,y) = g(x)h(y)$ for all $(x,y) \in \mathbb{R}^2$.

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.

Example <11.3> Suppose X and Y have a jointly continuous distribution with density function f. Define $S = X + Y$ and $T = X - Y$. Show that (S,T) has a jointly continuous distribution with density $\psi(s,t) = \frac{1}{2} f\left(\frac{s+t}{2}, \frac{s-t}{2}\right)$.

For instance, suppose the X and Y from Example <11.3> are independent and each is $N(0,1)$ distributed. From Example <11.2>, the joint density for (X,Y) is

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

The joint density for $S = X + Y$ and $T = X - Y$ is

$$\begin{aligned}\psi(s,t) &= \frac{1}{4\pi} \exp\left(\tfrac{1}{8}((s+t)^2 + (s-t)^2)\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \qquad \text{where } \sigma^2 = 2.\end{aligned}$$

It follows that S and T are independent, each with a $N(0, 2)$ distribution.

Example <11.3> also implies the convolution formula from Chapter 8. For if X and Y are independent, with densities g and h, then their joint density is $f(x, y) = g(x)h(y)$ and the joint density for $S = X + Y$ and $T = X - Y$ is

$$\psi(s, t) = \tfrac{1}{2} g\left(\frac{s+t}{2}\right) h\left(\frac{s-t}{2}\right)$$

Integrate over t to get the marginal density for S:

$$\begin{aligned}\int_{-\infty}^{+\infty} \psi(s, t)\, dt &= \int_{-\infty}^{+\infty} \tfrac{1}{2} g\left(\frac{s+t}{2}\right) h\left(\frac{s-t}{2}\right) dt \\ &= \int_{-\infty}^{+\infty} g(x)h(s-x)\, dx \qquad \text{putting } x = (s+t)/2.\end{aligned}$$

The argument for general linear combinations is slightly more complicated, unless you already know about Jacobians. You could skip the next Example if you don't know about matrices.

Example <11.4> Suppose X and Y have a jointly continuous distribution with joint density $f(x, y)$. For constants a, b, c, d, define $U = aX + bY$ and $V = cX + dY$. Find the joint density function $\psi(u, v)$ for (U, V), under the assumption that the quantity $\kappa = ad - bc$ is nonzero.

The method used in Example <11.4>, for linear transformations, extends to give 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 a good approximation to the density for smooth transformations in small regions.

Example <11.5> Suppose X and Y are independent random variables, with $X \sim \text{gamma}(\alpha)$ and $Y \sim \text{gamma}(\beta)$. Show that the random variables $U = X/(X + Y)$ and $V = X + Y$ are independent, with $U \sim \text{beta}(\alpha, \beta)$ and $V \sim \text{gamma}(\alpha + \beta)$.

The conclusion about $X + Y$ from Example <11.5> extends to sums of more than two independent random variables, each with a gamma distribution. The result has a particularly important special case, involving the sums of squares of independent standard normals.

Example <11.6> Sums of independent gamma random variables.

And finally, a polar coordinates way to generate independent normals:

Example <11.7> Building independent normals

11.2 Examples for Chapter 11

<11.1> **Example.** *Expectations of functions of a random variable with jointly continuous distributions*

Suppose X and Y have a jointly continuous distribution with joint density function $f(x,y)$. Let $Y = H(X,Y)$ be a new random variable, defined as a function of X and Y. An approximation argument similar to the one used in Chapter 7 will show that

$$\mathbb{E}H(X,Y) = \iint_{\mathbb{R}^2} H(x,y)f(x,y)\,dx\,dy.$$

For simplicity suppose H is nonnegative. (For the general case split H into positive and negtive parts.) For a small $\delta > 0$ define

$$A_n = \{(x,y) \in \mathbb{R}^2 : n\delta \le H(x,y) < (n+1)\delta\} \qquad \text{for } n = 0, 1, \dots$$

The function $H_\delta(x,y) = \sum_{n\ge 0} n\delta \mathbf{1}\{(x,y) \in A_n\}$ approximates H:

$$H_\delta(x,y) \le H(x,y) \le H_\delta(x,y) + \delta \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

In particular,

$$\mathbb{E}H_\delta(X,Y) \le \mathbb{E}H(X,Y) \le \delta + \mathbb{E}H_\delta(X,Y).$$

and

$$\iint_{\mathbb{R}^2} H_\delta(x,y)f(x,y)\,dx\,dy$$
$$\le \iint_{\mathbb{R}^2} H(x,y)f(x,y)\,dx\,dy \le \delta + \iint_{\mathbb{R}^2} H_\delta(x,y)f(x,y)\,dx\,dy$$

The random variable $H_\delta(X,Y)$ has a discrete distribution, with expected value

$$\begin{aligned}
\mathbb{E}H_\delta(X,Y) &= \mathbb{E}\sum\nolimits_{n\ge 0} n\delta\mathbf{1}\{(X,Y)\in A_n\}\\
&= \sum\nolimits_{n\ge 0} n\delta\,\mathbb{P}\{(X,Y)\in A_n\}\\
&= \sum\nolimits_{n} n\delta \iint_{\mathbb{R}^2} \mathbf{1}\{(x,y)\in A_n\} f(x,y)\,dx\,dy\\
&= \iint_{\mathbb{R}^2} \sum\nolimits_{n} n\delta\mathbf{1}\{(x,y)\in A_n\} f(x,y)\,dx\,dy\\
&= \iint_{\mathbb{R}^2} H_\delta(x,y) f(x,y)\,dx\,dy.
\end{aligned}$$

Deduce that

$$\begin{aligned}
&\iint_{\mathbb{R}^2} H(x,y)f(x,y)\,dx\,dy - \delta\\
&\qquad \le \mathbb{E}H(X,Y)\\
&\qquad \le \delta + \iint_{\mathbb{R}^2} H(x,y)f(x,y)\,dx\,dy
\end{aligned}$$

for every $\delta > 0$. □

<11.2> **Example.** *(Joint densities for independent random variables) Suppose X has a continuous distribution with density g and Y has a continuous distribution with density h. Then X and Y are independent if and only if they have a jointly continuous distribution with joint density $f(x,y) = g(x)h(y)$ for all $(x,y)\in\mathbb{R}^2$.*

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 $\delta > 0$ and $\epsilon > 0$,

$$\Delta = \{(x,y)\in\mathbb{R}^2 : x_0 \le x \le x_0+\delta, y_0 \le y \le y_0+\epsilon\}.$$

By independence,

$$\begin{aligned}
\mathbb{P}\{(X,Y)\in\Delta\} &= \mathbb{P}\{x_0 \le X \le x_0+\delta\}\mathbb{P}\{y_0 \le Y \le y_0+\epsilon\}\\
&\approx \delta g(x_0)\epsilon h(y_0) = \big(\text{area of }\Delta\big)\times g(x_0)h(y_0).
\end{aligned}$$

Thus X and Y have a jointly continuous distribution with joint density that takes the value $f(x_0,y_0) = g(x_0)h(y_0)$ at (x_0,y_0).

Conversely, if X and Y have a joint density f that factorizes, $f(x,y) = g(x)h(y)$, then for each pair of subsets C, D of the real line,

$$\begin{aligned}\mathbb{P}\{X \in C, Y \in D\} &= \iint \mathbf{1}\{x \in C, y \in D\} f(x,y)\,dx\,dy \\ &= \iint \mathbf{1}\{x \in C\}\mathbf{1}\{y \in D\} g(x)h(y) dx\,dy \\ &= \left(\int \mathbf{1}\{x \in C\} g(x)\,dx\right)\left(\int \mathbf{1}\{y \in D\} h(y)\,dy\right)\end{aligned}$$

Define $K := \int_{-\infty}^{+\infty} g(x)\,dx$. The choice $C = D = \mathbb{R}$ in the previous display then shows that $\int_{-\infty}^{+\infty} h(y)\,dy = 1/K$.

If we take only $D = \mathbb{R}$ we get

$$\mathbb{P}\{X \in C\} = \mathbb{P}\{X \in C, Y \in \mathbb{R}\} = \int_C g(x)/K\,dx$$

from which it follows that $g(x)/K$ is the marginal density for X. Similarly, $Kh(y)$ is the marginal density for Y, so that

$$\mathbb{P}\{X \in C, Y \in D\} = \int_C \frac{g(x)}{K}\,dx \times \int_D Kh(y)\,dy = \mathbb{P}\{X \in C\} \times \mathbb{P}\{Y \in D\}.$$

Put another way,

$$\mathbb{P}\{X \in C \mid Y \in D\} = \mathbb{P}\{X \in C\} \qquad \text{provided } \mathbb{P}\{Y \in D\} \neq 0.$$

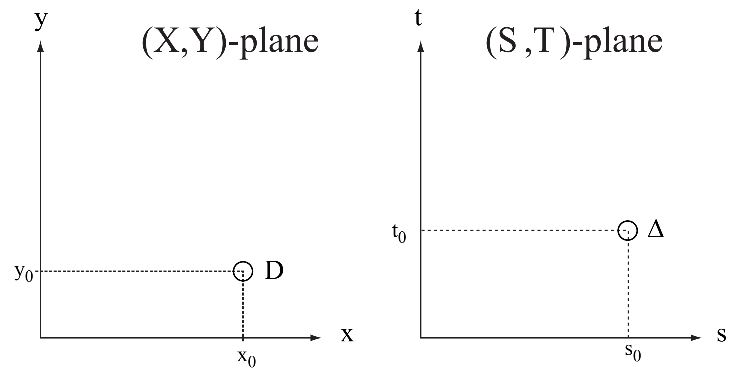
The random variables X and Y are independent.

Of course, if we know that g and h are the marginal densities then we have $K = 1$. The argument in the previous paragraph actually shows that any factorization $f(x,y) = g(x)h(y)$ of a joint density (even if we do not know that the factors are the marginal densities) implies independence. □

<11.3> **Example.** *Suppose X and Y have a jointly continuous distribution with density function f. Define $S = X+Y$ and $T = X-Y$. Show that (S,T) has a jointly continuous distribution with density $g(s,t) = \frac{1}{2}f\left(\frac{s+t}{2}, \frac{s-t}{2}\right)$.*

Consider a small ball Δ of radius ϵ centered at a point (s_0, t_0) in the plane. The area of Δ equals $\pi\epsilon^2$. The point (s_0, t_0) in the (S,T)-plane (the region where (S,T) takes its values) corresponds to the point (x_0, y_0) in the (X,Y)-plane for which $s_0 = x_0 + y + 0$ and $t_0 = x_0 - y_0$. That is, $x_0 = (s_0 + t_0)/2$ and $y_0 = (s_0 - t_0)/2$.

We need to identify $\{(S,T) \in \Delta\}$ with some set $\{(X,Y) \in D\}$.



By great luck (or by a clever choice for Δ) the region D in the (X,Y)-plane turns out to be another ball:

$$
\begin{aligned}
\{(S,T) \in \Delta\} &= \{(S-s_0)^2 + (T-t_0)^2 \le \epsilon^2\} \\
&= \{(X+Y-x_0-y_0)^2 + (X-Y-x_0+y_0)^2 \le \epsilon^2\} \\
&= \{2(X-x_0)^2 + 2(Y-y_0)^2 \le \epsilon^2\}.
\end{aligned}
$$

(Notice the cancellation of $(X-x_0)(Y-y_0)$ terms.) That is D is a ball of radius $\epsilon/\sqrt{2}$ centered at (x_0,y_0), with area $\pi\epsilon^2/2$, which is half the area of Δ. Now we can calculate.

$$
\begin{aligned}
\mathbb{P}\{(S,T) \in \Delta\} &= \mathbb{P}\{(X,Y) \in D\} \\
&\approx (\text{area of } D) \times f(x_0,y_0) \\
&= \tfrac{1}{2}(\text{area of } \Delta) \times f\left(\frac{s_0+t_0}{2}, \frac{s_0-t_0}{2}\right)
\end{aligned}
$$

It follows that (S,T) has joint density $g(s,t) = \frac{1}{2} f\left(\frac{s+t}{2}, \frac{s-t}{2}\right)$. □

<11.4> **Example.** *Suppose X and Y have a jointly continuous distribution with joint density $f(x,y)$. For constants a,b,c,d, define $U = aX + bY$ and $V = cX + dY$. Find the joint density function $\psi(u,v)$ for (U,V), under the assumption that the quantity $\kappa = ad - bc$ is nonzero.*

In matrix notation,

$$
(U,\, V) = (X,\, Y)A \qquad \text{where } A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.
$$

Notice that $\det A = ad - bc = \kappa$. The assumption that $\kappa \neq 0$ ensures that A has an inverse:

$$A^{-1} = \frac{1}{\kappa}\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

That is, if $(u, v) = (x, y)A$ then

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

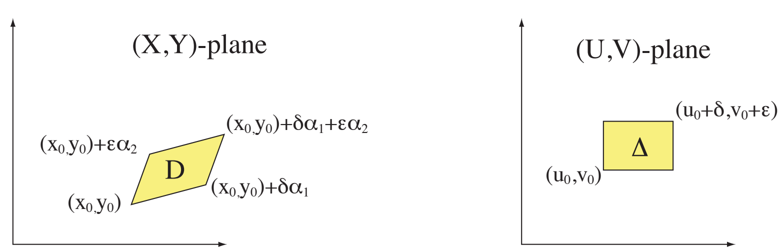
Notice that $\det\left(A^{-1}\right) = 1/\kappa = 1/(\det A)$.

Consider a small rectangle $\Delta = \{u_0 \le u \le u_0 + \delta,\ v_0 \le v \le v_0 + \epsilon\}$, for (u_0, v_0) in the (U, V)-plane and small, positive δ and ϵ. The joint density function $\psi(u, v)$ is characterized by the property that

$$\mathbb{P}\{(U, V) \in \Delta\} \approx \psi(u_0, v_0)\delta\epsilon$$

The event $\{(U, V) \in \Delta\}$ is equal to some event $\{(X, Y) \in D\}$. 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 D must be a parallelogram. We have only to determine its vertices, which correspond to the four vertices of the rectangle Δ. Define vectors $\alpha_1 = (d, -c)/\kappa$ and $\alpha_2 = (-b, a)/\kappa$, which correspond to the two rows of the matrix A^{-1}. Then D has vertices:

$$\begin{aligned}
(x_0, y_0) &= (u_0, v_0)A^{-1} = u_0\alpha_1 + v_0\alpha_2 \\
(x_0, y_0) + \delta\alpha_1 &= (u_0 + \delta, v_0)A^{-1} = (u_0 + \delta)\alpha_1 + v_0\alpha_2 \\
(x_0, y_0) + \epsilon\alpha_2 &= (u_0, v_0 + \epsilon)A^{-1} = u_0\alpha_1 + (v_0 + \epsilon)\alpha_2 \\
(x_0, y_0) + \delta\alpha_1 + \epsilon\alpha_2 &= (u_0 + \delta, v_0 + \epsilon)A^{-1} = (u_0 + \delta)\alpha_1 + (v_0 + \epsilon)\alpha_2
\end{aligned}$$



From the formula in the Appendix to this Chapter, the parallelogram D has area equal to $\delta\epsilon$ times the absolute value of the determinant of the matrix with rows α_1 and α_2. That is,

$$\text{area of } D = \delta\epsilon |\det(A^{-1})| = \frac{\delta\epsilon}{|\det A|}.$$

In summary: for small $\delta > 0$ and $\epsilon > 0$,

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

It follows that (U,V) have joint density

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

On the right-hand side you should substitute $(du - bv)/\kappa$ for x and $(-cu + av)/\kappa$ for y, in order to get an expresion involving only u and v. □

Remark. In effect, I have calculated a Jacobian by first principles.

<11.5> **Example.** *Suppose X and Y are independent random variables, with $X \sim \text{gamma}(\alpha)$ and $Y \sim \text{gamma}(\beta)$. Show that the random variables $U = X/(X+Y)$ and $V = X+Y$ are independent, with $U \sim \text{beta}(\alpha, \beta)$ and $V \sim \text{gamma}(\alpha+\beta)$.*

The random variables X and Y have marginal densities

$$g(x) = x^{\alpha-1}e^{-x}\mathbf{1}\{x > 0\}/\Gamma(\alpha) \qquad \text{and} \qquad h(y) = y^{\beta-1}e^{-y}\mathbf{1}\{y > 0\}/\Gamma(\beta)$$

From Example <11.2>, they have a jointly continuous distribution with joint density

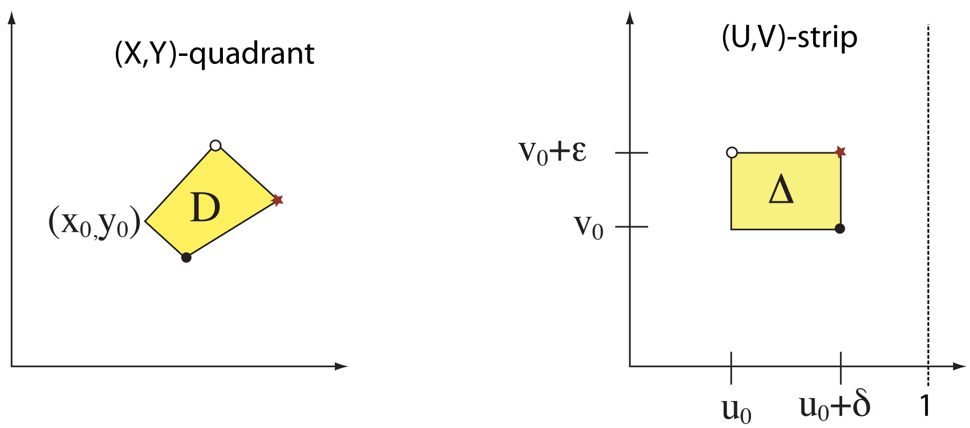
$$f(x,y) = g(x)h(y) = \frac{x^{\alpha-1}e^{-x}y^{\beta-1}e^{-y}}{\Gamma(\alpha)\Gamma(\beta)}\mathbf{1}\{x > 0, y > 0\}.$$

We need to find the joint density function $\psi(u,v)$ for the random variables $U = X/(X+Y)$ and $V = X+Y$. The pair (U,V) takes values in the strip defined by $\{(u,v) \in \mathbb{R}^2 : 0 < u < 1, 0 < v < \infty\}$. The joint density function ψ can be determined by considering corresponding points (x_0, y_0) in the (x,y)-quadrant and (u_0, v_0) in the (u,v)-strip for which

$$u_0 = x_0/(x_0 + y_0) \qquad \text{AND} \qquad v_0 = x_0 + y_0,$$

that is,

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



When (U,V) lies near (u_0,v_0) then (X,Y) lies near the point $(x_0,y_0) = (u_0v_0, v_0(1-u_0))$. More precisely, for small positive δ and ϵ, there is a small region D in the (X,Y)-quadrant corresponding to the small rectangle

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

in the (U,V)-strip. That is, $\{(U,V) \in \Delta\} = \{(X,Y) \in D\}$. The set D is not a parallelogram but it is well approximated by one. For small perturbations, the map from (u,v) to (x,y) is approximately linear. First locate the points corresponding to the corners of Δ, under the maps $x = uv$ and $y = v(1-u)$:

$$\begin{aligned}
(u_0, v_0) &\mapsto (x_0, y_0)\\
(u_0 + \delta, v_0) &\mapsto (x_0, y_0) + \delta(v_0, -v_0)\\
(u_0, v_0 + \epsilon) &\mapsto (x_0, y_0) + \epsilon(u_0, 1 - u_0).
\end{aligned}$$

The fourth vertex, $(u_0 + \delta, v_0 + \epsilon)$ corresponds to the point (x,y) with

$$\begin{aligned}
x &= (u_0 + \delta)(v_0 + \epsilon) = u_0v_0 + \delta v_0 + \epsilon u_0 + \delta\epsilon\\
y &= (v_0 + \epsilon)(1 - u_0 - \delta) = v_0u_0 + \epsilon(1 - u_0) - \delta v_0 - \delta\epsilon
\end{aligned}$$

Put another way,

$$\begin{aligned}(u_0, v_0) &\mapsto (x_0, y_0)\\(u_0, v_0) + (\delta, 0) &\mapsto (x_0, y_0) + (\delta, 0)J\\(u_0, v_0) + (0, \epsilon) &\mapsto (x_0, y_0) + (0, \epsilon)J\\(u_0, v_0) + (\delta, \epsilon) &\mapsto (x_0, y_0) + (\delta, \epsilon)J + \text{ smaller order terms}\end{aligned}$$

where

$$J = \begin{pmatrix} v_0 & -v_0 \\ u_0 & 1 - u_0 \end{pmatrix}.$$

You might recognize J as the ***Jacobian matrix*** of partial derivatives

$$\begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} \\[2ex] \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

evaluated at (u_0, v_0).

The region D is approximately a parallelogram, with the edges oblique to the coordinate axes. To a good approximation, the area of D is equal to $\delta\epsilon$ times the area of the parallelogram with corners at

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

which, from the Appendix to this Chapter, equals $|\det(J)| = v_0$.

The rest of the calculation of the joint density ψ for (U, V) is easy:

$$\begin{aligned}\delta\epsilon\psi(u_0, v_0) &\approx \mathbb{P}\{(U, V) \in \Delta\}\\&= \mathbb{P}\{(X, Y) \in R\}\\&\approx f(x_0, y_0)(\text{area of } D) \approx \frac{x_0^{\alpha-1}e^{-x_0}}{\Gamma(\alpha)} \frac{y_0^{\beta-1}e^{-y_0}}{\Gamma(\beta)}\, \delta\, \epsilon\, v_0\end{aligned}$$

Substitute $x_0 = u_0 v_0$ and $y_0 = (1 - u_0)v_0$ to get the joint density at (u_0, v_0):

$$\begin{aligned}\psi(u_0, v_0) &= \frac{u_0^{\alpha-1} v_0^{\alpha-1} e^{-u_0 v_0}}{\Gamma(\alpha)} \frac{(1 - u_0)^{\beta-1} v_0^{\beta-1} e^{-v_0 + u_0 v_0}}{\Gamma(\beta)} v_0\\&= \frac{u_0^{\alpha-1}(1 - u_0)^{\beta-1}}{B(\alpha, \beta)} \times \frac{v_0^{\alpha+\beta-1}e^{-v_0}}{\Gamma(\alpha + \beta)} \times \frac{\Gamma(\alpha + \beta)B(\alpha, \beta)}{\Gamma(\alpha)\Gamma(\beta)}.\end{aligned}$$

Once again the final constant must be equal to 1, which gives the identity

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The joint density factorizes; the random variables U and V are independent, with $U \sim \text{Beta}(\alpha, \beta)$ and $V \sim \text{gamma}(\alpha+\beta)$. □

Remark. The fact that $\Gamma(1/2) = \sqrt{\pi}$ also follows from the equality

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

<11.6> **Example.** If $X_1, X_2, \dots, X_k$ are independent random variables, with X_i distributed gamma(α_i) for $i = 1, \dots, k$, then

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

A particular case has great significance for Statistics. Suppose $Z_1, \dots Z_k$ are independent random variables, each distributed N(0,1). You know that the random variables $Z_1^2/2, \dots, Z_k^2/2$ are independent gamma$(1/2)$ distributed random variables. The sum

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

must have a gamma$(k/2)$ distribution with density $t^{k/2-1}e^{-t}\mathbf{1}\{0 < t\}/\Gamma(k/2)$. It follows that the sum $Z_1^2 + \dots + Z_k^2$ has density

$$\frac{(t/2)^{k/2-1}e^{-t/2}\mathbf{1}\{0 < t\}}{2\Gamma(k/2)}.$$

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

<11.7> **Example.** Here are the bare bones of the polar coordinates way of manufacturing two independent $N(0,1)$'s. Start with independent random variables $U \sim \text{Uniform}(0, 2\pi)$ and $W \sim \text{gamma}(1)$ (a.k.a. standard exponential). Define $R = \sqrt{2W}$ and $X = R\cos(U)$ and $Y = R\sin(U)$. Calculate the density for R as

$$g(r) = r\exp(-r^2/2)\mathbf{1}\{r > 0\}.$$

For $0 < \theta_0 < 1$ and $r_0 > 0$, and very small $\delta > 0$ and $\epsilon > 0$, check that the region

$$D = \{(u, r) \in (0,1) \times (0,\infty) : \theta_0 \le U \le \theta_0 + \delta,\ r_0 \le r \le r_0 + \epsilon\}$$

corresponds to the region Δ in the (X, Y)-plane that is bounded by circles of radius r_0 and $r_0+\epsilon$ and by radial lines from the origin at angles θ_0 and $\theta_0+\delta$ to the horizontal axis. The area of Δ is approximately $2\pi r_0\epsilon\delta$.

Deduce that the joint density f for (X, Y) satisfies

$$2\pi r_0\epsilon\delta f(x_0, y_0) \approx \epsilon g(r_0)\frac{\delta}{2\pi} \qquad \text{where } x_0 = r_0\cos(\theta_0), \quad y_0 = r_0\sin(\theta_0)$$

That is,

$$\begin{aligned} f(x, y) &= \frac{g(r)}{2\pi r} \qquad \text{where } x = r\cos(\theta), \quad y = r\sin(\theta) \\ &= \frac{1}{2\pi}\exp\left(-\frac{1}{2}(x^2 + y^2)\right). \end{aligned}$$

The random variables X and Y are independent, with each distributed $N(0,1)$. □

11.3 Appendix: area of a parallelogram

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

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

That is, the area of R equals $|a_1b_2 - a_2b_1|$.

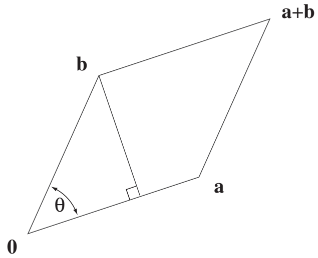
Proof Let θ denotes the angle between $\mathbf{a}$ and $\mathbf{b}$. Remember that

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

The area of R is twice the area of the triangle with vertices at $\mathbf{0}$, $\mathbf{a}$, and $\mathbf{b}$. The triangle has area

$$\tfrac{1}{2}(\text{base length}) \ \times (\text{height}) = \tfrac{1}{2} \|\mathbf{a}\| \times (\|\mathbf{b}\| \times |\sin\theta|)$$

The square of the area of R equals



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

If you are not sure about the properties of determinants used in the last two lines, you should check directly that

$$(a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2 = (a_1 b_2 - a_2 b_1)^2.$$

□