Chapter 11

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

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

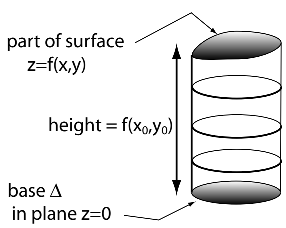
We have been working for some time with problems involving such pairs of random variables, but we have not needed to formalize the concept of a joint distribution. When both X and Y have continuous distributions, it becomes more important to have a systematic way to describe how one might calculate probabilities 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 $\mathbb{P}\{X + Y \leq 7\}$?

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 $\iint \{(x, y) \in B\} \ldots$ or $\iint \mathbb{I}\{(x, y) \in B\} \ldots$ instead of $\iint_B \ldots$.



The density function defines a surface, via the equation $z = f(x, y)$. The probability that the random point $(X(\omega), Y(\omega))$ 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).

Remark. For a rigorous treatment, Δ is not allowed to be too weirdly shaped. One can then show that the limit exists and equals $f(x_0, y_0)$ except for (x_0, y_0) in a region with zero area.

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 6.

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

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

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 <11.1>: (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 <11.2>: Joint densities for linear combinations

Read through the details of the following important special case, to make sure you understand the notation from Example <11.2>.

Example <11.3>: Linear combinations of independent normals

The method used in Example <11.2>, 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.

From Homework 9, you know that for independent random variables X and Y with $X \sim \text{gamma}(\alpha)$ and $Y \sim \text{gamma}(\beta)$, we have $X/(X+Y) \sim \text{beta}(\alpha, \beta)$ and $X + Y \sim \text{gamma}(\alpha + \beta)$. The next Example provides a slightly simpler way to derive these two results, plus a little more.

Example <11.4>: 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 $V \sim \text{gamma}(\alpha + \beta)$.

In general, if X and Y have a joint density function $f(x, y)$ then

$$\mathbb{P}\{X \in A\} = \iint \{x \in A, -\infty < y < \infty\} f(x, y)\,dx\,dy = \int \{x \in A\} f_X(x)\,dx,$$

where

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

That is, X has a continuous distribution with (marginal) density function f_X. Similarly, Y has a continuous distribution with (marginal) density function $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$. Remember that the word marginal is redundant; it serves merely to stress that a calculation refers only to one of the random variables.

©David Pollard

The conclusion about $X + Y$ from Example <11.4> 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 <11.5>: Sums of independent gamma random variables.

Examples for Chapter 11

<11.1> **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 $\delta > 0$ and $\epsilon > 0$,

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

By independence,

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

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

Remark. 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.

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,

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

In particular, if we take $C = D = \mathbb{R}$ then we get

$$\int_{-\infty}^{\infty} g(x)\, dx = K \qquad \text{and} \qquad \int_{-\infty}^{\infty} h(y)\, dy = 1/K$$

for some constant 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. Moreover, provided $\mathbb{P}\{Y \in D\} \neq 0$,

$$\mathbb{P}\{X \in C \mid Y \in D\} = \frac{\mathbb{P}\{X \in C, Y \in D\}}{\mathbb{P}\{Y \in D\}} = \frac{\mathbb{P}\{X \in C\}\mathbb{P}\{Y \in D\}}{\mathbb{P}\{Y \in D\}} = \mathbb{P}\{X \in C\}.$$

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 of a joint density (even if we do not know that the factors are the marginal densities) implies independence. □

<11.2> **Example.** Suppose X and Y have a jointly continuous distribution with joint density $f(x, y)$. For constants a, b, c, d, define

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

Find the joint density function $\psi(u,v)$ for (U,V), under the assumption that the quantity $\kappa = ad - bc$ is nonzero.

Think of the pair (U,V) as 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 $\det A = ad - bc = \kappa$. The assumption that $\kappa \neq 0$ ensures that the transformation is invertible:

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

That is,

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

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

It helps to distinguish between the two roles for $\mathbb{R}^2$, by referring to the domain of T as the (x,y)-plane and the range as the (u,v)-plane.

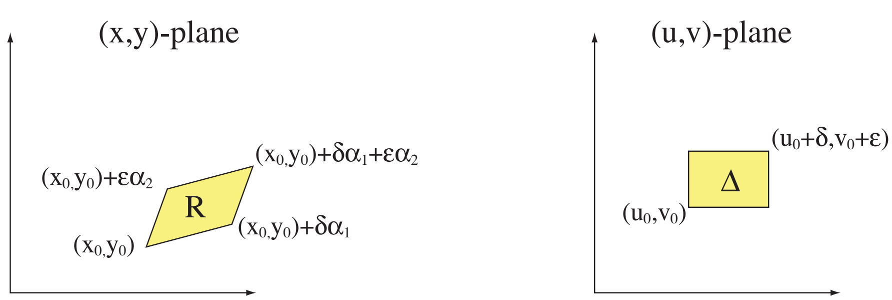
The joint density function $\psi(u,v)$ is characterized by the property that

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

for each (u_0, v_0) in the (u,v)-plane, and small, positive δ and ϵ. 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 $(u_0 + \delta, v_0 + \epsilon)$, 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. 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 R 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 R 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,

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

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

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

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.

<11.3> **Example.** Suppose X and Y are independent random variables, each distributed $N(0, 1)$. By Example <11.1>, the joint density for (X, Y) equals

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

By Example <11.2>, the joint distribution of the random variables

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

has the joint density

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

You'll learn more about joint normal distributions in Chapter 13. □

<11.4> **Example.** We are given independent random variables X and Y, with $X \sim \text{gamma}(\alpha)$ and $Y \sim \text{gamma}(\beta)$. That is, X has a continuous distribution with density

$$g(x) = x^{\alpha-1}e^{-x}\mathbb{I}\{x > 0\}/\Gamma(\alpha)$$

and Y has a continuous distribution with density

$$h(y) = y^{\beta-1}e^{-y}\mathbb{I}\{y > 0\}/\Gamma(\beta)$$

From Example <11.1>, the random variables have a jointly continuous distribution with joint density

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

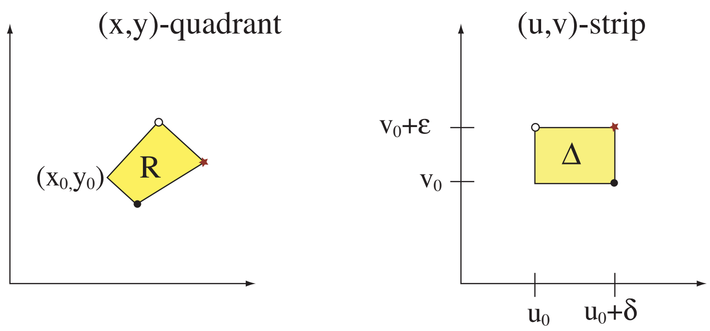
We need to find the joint density function $\psi(u, v)$ for the random variables $U = X/(X + Y)$ and $V = X + Y$.

The pair (U, V) takes values in the strip in the (u, v)-plane defined by $0 < u < 1$ and $0 < v < \infty$. 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:

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

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 $(x_0, y_0) = (u_0v_0, v_0(1 - u_0))$. More precisely, for small positive δ and ϵ, there is a small region R 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. First locate the points corresponding to the corners of Δ, under the maps $x = uv$ and $y = v(1 - u)$:

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

In matrix notation,

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

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

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

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

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

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

Substitute $x_0 = u_0v_0$ and $y_0 = (1 - u_0)v_0$ to get the joint density at (u_0, v_0):

$$\begin{aligned}
\psi(u_0, v_0) &= \frac{u_0^{\alpha-1}v_0^{\alpha-1}e^{-u_0v_0}}{\Gamma(\alpha)}\frac{(1 - u_0)^{\beta-1}v_0^{\beta-1}e^{-v_0+u_0v_0}}{\Gamma(\beta)}v_0\\
&= \frac{u_0^{\alpha-1}(1 - u_0)^{\beta-1}}{B(\alpha, \beta)}\frac{v_0^{\alpha+\beta-1}e^{-v_0}}{\Gamma(\alpha + \beta)}\frac{\Gamma(\alpha + \beta)B(\alpha, \beta)}{\Gamma(\alpha)\Gamma(\beta)}
\end{aligned}$$

That is,

$$\psi(u,v) = f_0(u) f_1(v) \frac{\Gamma(\alpha+\beta)B(\alpha,\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

where

$$f_0(u) = \frac{u^{\alpha-1}(1-u)^{\beta-1}\mathbb{I}\{0<u<1\}}{B(\alpha,\beta)} \qquad \text{the beta}(\alpha,\beta)\text{ density}$$

$$f_1(v) = \frac{v^{\alpha+\beta-1}e^{-v}\mathbb{I}\{0<v\}}{\Gamma(\alpha+\beta)} \qquad \text{the gamma}(\alpha+\beta)\text{ density.}$$

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

$$\begin{aligned} 1 &= \mathbb{P}\{0<U<1, 0<V<\infty\} \\ &= \iint \{0<u<1, 0<v<\infty\}\psi(u,v)\,du\,dv \\ &= \left(\int_0^1 f_0(u)\,du\right)\left(\int_0^\infty f_1(v)\,dv\right)\frac{B(\alpha,\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{B(\alpha,\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}. \end{aligned}$$

Once again we have derived the expression relating beta and gamma functions.

The joint density factorizes into a product of the marginal densities: the random variables U and V are independent. □

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

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

<11.5> **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 $t^{k/2-1}e^{-t}\mathbb{I}\{0<t\}/\Gamma(k/2)$. It follows that the sum $Z_1^2 + \ldots + Z_k^2$ has density

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

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. □

 ©David Pollard

Appendix: area of a parallelogram

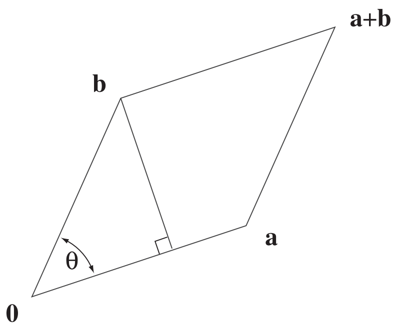
Let R be a parallelogram in the plane with corners at $\mathbf{0} = (0, 0)$, and $\mathbf{a} = (a_1, a_2)$, and $\mathbf{b} = (b_1, b_2)$, and $\mathbf{a} + \mathbf{b}$. The area of R is equal to the absolute value of the determinant of the matrix

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

That is, the area of R equals $|a_1 b_2 - a_2 b_1|$.

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

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



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

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

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

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

$\square$

©David Pollard