# Probability Theory with Simulations
-
# Part-IV
# Two-dimensional continuous distributions
-

Andras Vetier

2013 09 11

## Contents

# 1 Two-dimensional random variables and distributions

In this chapter, we start to work with two-dimensional continuous random variables and distributions. When two random variables, say $X$ and $Y$ are considered, then we may put them together to get a pair of random numbers, that is, a random point $(X,Y)$ in the two-dimensional space. Some examples:

1. Let us choose a Hungarian man, and let

   $X =$ his height

   $Y =$ his weight

   Then $(X,Y) =$ (height,weight) is a two-dimensional random variable.

2. Let us generate three independent random numbers, and let

   $X =$ the smallest of them

   $Y =$ the largest of them

   Then $(X,Y) =$ (smallest,biggest) is a two-dimensional random variable.

**Density function.** Two-dimensional continuous random variables are described mainly by their density function $f(x,y)$, which integrated on a set $A$ gives the probability of the event that the value of $(X,Y)$ is in the set $A$:

$$\mathbf{P}(A) = \mathbf{P}\left((X,Y) \in A\right) = \iint\limits_{A} f(x,y)\,dx\;dy$$

The characteristic properties of two-dimensional density functions are:

$$f(x,y) \geq 0$$

$$\iint\limits_{R^2} f(x,y)\,dx\;dy = 1$$

These two properties are characteristic for two-dimensional density functions, because, on one side, they are true for two-dimensional density functions of any continuous random variables, and on the other side, if a function $f(x,y)$ is given which has these two properties, then it is possible to define a two-dimensional random variable $(X,Y)$ so that its density function is the given function $f(x,y)$.

**The density function as a constant of approximate proportionality.** If $A$ is a small set around a point $(x,y)$, then

$$\mathbf{P}\left((X,Y) \in A\right) = \iint\limits_{A} f(x,y)\,dx\;dy \approx f(x,y) \;\times\; \text{area of } A$$

and

$$f(x,y) \approx \frac{\mathbf{P}\left((X,Y) \in A\right)}{\text{area of } A}$$

We emphasize that the value $f(x,y)$ of the density function does not represent any probability value. If $(x,y)$ is a fixed point, then $f(x,y)$ may be interpreted as a constant of approximate proportionality: if $A$ is a small set around a point $(x,y)$, then the probability that the point $(X,Y)$ is in $A$ is approximately equal to $f(x,y) \times$ area of $A$:

$$\mathbf{P}\left((X,Y) \in A\right) \approx f(x,y) \ \times \ \text{area of } A$$

**Approximating the density function.** If $A$ is a small rectangle with sides of lengths $\Delta x$ and $\Delta y$, then we get that

$$f(x,y) \approx \frac{\mathbf{P}\left(x < X < x + \Delta x \ \text{ and } \ y < Y < y + \Delta y\right)}{\Delta x \ \Delta y}$$

This formula is useful to determine the density function in some problems.

**Conditional probability.** If $A$ and $B$ are subsets of the plane, then both $(X,Y) \in A$ and $(X,Y) \in B$ defines an event. The conditional probability of the event $(X,Y) \in B$ on condition that the event $(X,Y) \in A$ occurs is denoted by $\mathbf{P}\left((X,Y) \in B | (X,Y) \in A\right)$ or $\mathbf{P}(B|A)$ for short. This conditional probability can be be calculated obviously as tho ratio of two integrals:

$$\mathbf{P}(B|A) = \frac{\iint\limits_{A \cap B} f(x,y)\, dx \ dy}{\iint\limits_{A} f(x,y)\, dx \ dy}$$

If $B \subseteq A$, then $A \cap B = B$, and we get that

$$\mathbf{P}(B|A) = \frac{\iint\limits_{B} f(x,y)\, dx \ dy}{\iint\limits_{A} f(x,y)\, dx \ dy}$$

**Conditional density function.** If $A$ is a subset of the plane, which has a positive probability, and we know that the condition $(X,Y) \in A$ is fulfilled, then the density function of $(X,Y)$ under this condition is, obviously

$$f(x,y|A) = \frac{f(x,y)}{\mathbf{P}(A)} = \frac{f(x,y)}{\iint\limits_{A} f(x,y)\, dx \ dy} \quad \text{if } \ (x,y) \in A$$

**Multiplication rule for independent random variables.** If $X$ and $Y$ are independent, and their density functions are $f_1(x)$ and $f_2(y)$ respectively, then the density function $f(x,y)$ of $(X,Y)$ is the direct product of the density functions $f_1(x)$ and $f_2(y)$:

$$f(x,y) = f_1(x) \ f_2(y)$$

**Proof.**

$$f(x,y) \approx \frac{\mathbf{P}\left(x < X < x+\Delta x \ \text{ and } \ y < Y < y+\Delta y\right)}{\Delta x \ \Delta y} \approx$$

$$\frac{\mathbf{P}\left(x < X < x+\Delta x\right)}{\Delta x} \ \frac{\mathbf{P}\left(y < Y < y+\Delta y\right)}{\Delta y} \approx$$

$$f_1(x) \ f_2(y)$$

**General multiplication rule.** If the density function of $X$ is $f_1(x)$ and the density function of $Y$ under the condition that $X = x$ is $f_{2|1}(y|x)$, then the density function of $(X,Y)$ is

$$f(x,y) = f_1(x) \ f_{2|1}(y|x)$$

Similarly:

$$f(x,y) = f_2(y) \ f_{1|2}(x|y)$$

**Proof.** We give the proof of the first formula:

$$f(x,y) \approx \frac{\mathbf{P}\left(x < X < x+\Delta x \ \text{ and } \ y < Y < y+\Delta y\right)}{\Delta x \ \Delta y} =$$

$$\frac{\mathbf{P}\left(x < X < x+\Delta x\right)}{\Delta x} \ \frac{\mathbf{P}\left(y < Y < y+\Delta y \mid x < X < x+\Delta x\right)}{\Delta y} \approx$$

$$\frac{\mathbf{P}\left(x < X < x+\Delta x\right)}{\Delta x} \ \frac{\mathbf{P}\left(y < Y < y+\Delta y \mid X \approx x\right)}{\Delta y} \approx$$

$$f_1(x) \ f_{2|1}(y|x)$$

**Distribution function.** The distribution function of a two-dimensional random variable is defined by

$$F(x,y) = \mathbf{P}(X < x, Y < y)$$

The distribution function can be calculated from the density function by integration:

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

The density function can be calculated from the distribution function by differentiation:

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

The probability of a rectangle can be calculated from the distribution function like this:

$$P(x_1 < X < x_2 \quad \text{and} \quad y_1 < Y < y_2) =$$

$$= P(X < x_2 \quad \text{and} \quad Y < y_2) - P(X < x_1 \quad \text{and} \quad Y < y_2) -$$

$$-P(X < x_2 \quad \text{and} \quad Y < y_1) + P(X < x_1 \text{ and } Y < y_1) =$$

$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

**Expected value of a function of $(X, Y)$.** If we make $N$ experiments for a two-dimensional random variable $(X, Y)$, and we substitute the experimental results

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$$

into the function $y = t(x, y)$, and we consider the values

$$t(X_1, Y_1), t(X_2, Y_2), \ldots, t(X_N, Y_N)$$

then their average is close to the expected value of $t(X, Y)$:

$$\frac{t(X_1, Y_1) + t(X_2, Y_2) + \ldots + t(X_N, Y_N)}{N} \approx \mathbf{E}\left(t(X, Y)\right)$$

where $\mathbf{E}\left(t(X, Y)\right)$ is the expected value of $t(X, Y)$, which is calculated by a double integral:

$$\mathbf{E}\left(t(X, Y)\right) = \iint\limits_{R^2} t(x, y) f(x, y)\, dx\ dy$$

**Expected value of the product.** As an example, and because of its importance, we mention here that the expected value of the product $XY$ of the random variables $X$ and $Y$ is calculated by a double integral:

$$\mathbf{E}\left(XY\right) = \iint\limits_{R^2} x\ y\ \ f(x, y)\, dx\ dy$$

which means that if $N$ is large, then

$$\frac{X_1 Y_1 + X_2 Y_2 + \ldots + X_N Y_N}{N} \approx \mathbf{E}\left(XY\right) = \iint\limits_{R^2} x\ y\ \ f(x, y)\, dx\ dy$$

**Covariance and covariance matrix.** The notion of the covariance is an auxiliary notion in two-dimensions. For a two-dimensional data-set, it is

$$\frac{(X_1 - \overline{X})(Y_2 - \overline{Y}) + (X_2 - \overline{X})(Y_2 - \overline{Y}) + \ldots + (X_N - \overline{X})(Y_N - \overline{Y})}{N} =$$

$$\frac{X_1Y_2 + X_2Y_2 + \ldots + X_NY_N}{N} - \overline{X}\,\overline{Y}$$

For a two-dimensional random variable and distribution the covariance $\mathbf{COV}(X,Y)$ is defined by

$$\mathbf{COV}(X,Y) = \mathbf{E}\left((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))\right) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) =$$

$$\iint\limits_{R^2} (x-\mu_1)\;(y-\mu_2)\;f(x,y)\,dx\;dy = \iint\limits_{R^2} x\;y\;f(x,y)\,dx\;dy - \mu_1\mu_2$$

The covariance and the variances of the coordinates can be arranged into a matrix. This matrix is called the **covariance matrix**:

$$C = \begin{pmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{pmatrix}$$

This matrix can be considered as a two-dimensional generalization or the notion of the variance.

**Correlation coefficient.** The notion of the correlation coefficient plays an important role in describing the relation between the coordinates of a two-dimensional data-set of random variable. Its definition is:

$$\mathbf{CORR}(X,Y) = \frac{\mathbf{COV}(X,Y)}{\sigma_1\sigma_2}$$

Its value is always between $-1$ and 1: $-1 \leq r \leq 1$. If $r > 0$, then larger $x$-coordinates mostly imply larger $y$-coordinates, if $r < 0$, then larger $x$-coordinates mostly imply smaller $y$-coordinates. In the first case, we say that the coordinates have a **positive correlation**, in the second case, we say that the coordinates have a **negative correlation**. If $|r|$ is close to 1, then the coordinates are in a **strong correlation**, if $|r|$ is close to 0, then the coordinates are in a **loose correlation**.

**Using a calculator.** More sophisticated calculators have a key to determine the covariance and the correlation coefficient of a two-dimensional data-set, as well.

# 2 Uniform distribution on a two-dimensional set

If $S$ is a set in the two-dimensional plane, and $S$ has a finite area, then we may consider the density function equal to the reciprocal of the area of $S$ inside $S$, and equal to 0 otherwise:

$$f(x,y) = \frac{1}{\text{area of } S} \quad \text{if } (x,y) \in S$$

The distribution associated to this density function is called **uniform distribution on the set** $S$. Since the integral of a constant on a set $A$ is equal to the area of $A$ multiplied by that constant, we get that

$$\mathbf{P}(A) = \iint\limits_{A} f(x,y)\,dx\;dy = \iint\limits_{A} \frac{1}{\text{area of } S}\,dx\;dy = \frac{\text{area of } A}{\text{area of } S}$$

for any subset $A$ of $S$. Thus, uniform distribution on $S$ means that, for any subset $A$ of $S$, the probability of $A$ is proportional to the area of $A$.

The reader probably remembers that in Chapter 6 of Part I, under the title "Geometrical problems, uniform distributions", we worked with uniform distributions. Now it should become clear that the uniform distribution on the set $S$ is a special continuous distribution whose density function is equal to a constant on the set $S$.

# 3 *** Beta distributions in two-dimensions

Assume that $n$ people arrive between noon and 1pm independently of each other according to uniform distribution, and let $X$ be the $i$th, and let $Y$ be the $j$th arrival time. This real-life problem can be simulated like this: we generate $n$ uniformly distributed independent random points between 0 and 1, and $X$ = $i$th smallest, and $Y$ = $j$th smallest among them. We calculate here the density function of the two-dimensional random variable $(X,Y)$. Let $0 < x < y < 1$, let $[x_1, x_2]$ be a small interval around $x$, and let $[y_1, y_2]$ be a small interval around $y$. We assume that $x_2 < y_1$. By the meaning of the density function:

$$f(x,y) \approx \frac{\mathbf{P}(X \in \Delta x, Y \in \Delta y)}{(x_2 - x_1)\,(y_2 - y_1)}$$

The event $X \in \Delta x$ , $Y \in \Delta y$, which stands in the numerator, means that the $i$th smallest point is in $[x_1, x_2)$, and the $j$th smallest point is in $[y_1, y_2)$, which means that

| | | | | | |
|---|---|---|---|---|---|
| there is | at least one | point $X$ | in | $[x_1, x_2)$, | and |
| there is | at least one | point $Y$ | in | $[y_1, y_2)$, | and |
| there are | $i-1$ | points | in | $[0, X)$, | and |
| there are | $j-i-1$ | points | in | $[X, Y)$, | and |
| there are | $n-j$ | points | in | $[Y, 1]$. | |

This, with a very good approximation, means that

| | | | | | |
|---|---|---|---|---|---|
| there are | $i-1$ | points | in | $[0, x_1)$, | and |
| there is | $1$ | point | in | $[x_1, x_2)$, | and |
| there are | $j-i-1$ | points | in | $[x_2, y_1)$, | and |
| there is | $1$ | point | in | $[y_1, y_2)$, | and |
| there are | $n-j$ | points | in | $[y_2, 1]$. | |

Using the formula of the poly-hyper-geometrical distribution, we get that the probability of the event $X \in \Delta x, Y \in \Delta y$ is approximately equal to

$$\frac{n!}{(i-1)!\,1!\,(j-i-1)!\,1!\,(n-j)!}\;x_1^{i-1}(x_2-x_1)^1(y_1-x_2)^{j-i-1}(y_2-y_1)^1(1-y_2)^{n-j}$$

Since $1! = 1$, we may omit some unnecessary factors and exponents, and the formula simplifies to

$$\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\;x^{i-1}\,(x_2-x_1)\,(y-x)^{j-i-1}\,(y_2-y_1)\,(1-y)^{n-j}$$

Dividing by $(x_2-x_1)\,(y_2-y_1)$, we get that the density function, for $0 < x < y < 1$, is

$$f(x,y) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\;x^{i-1}\,(y-x)^{j-i-1}\,(1-y)^{n-j}$$

**Special cases:**

1. **Three independent random numbers** (uniformly distributed between 0 and 1) are generated.

   (a) $X =$ **the smallest,** $Y =$ **the biggest** of them.
   Replacing $n = 3, i = 1, j = 3$, we get:

   $$f(x,y) = 6(y-x) \quad \text{if } 0 < x < y < 1$$

   (b) $X =$ **the smallest,** $Y =$ **the second smallest** of them.
   Replacing $n = 3, i = 1, j = 2$, we get:

   $$f(x,y) = 6(1-y) \quad \text{if } 0 < x < y < 1$$

   (c) $X =$ **the second smallest,** $Y =$ **the biggest** of them.
   Replacing $n = 3, i = 2, j = 3$, we get:

   $$f(x,y) = 6x \quad \text{if } 0 < x < y < 1$$

2. **Four independent random numbers** (uniformly distributed between 0 and 1) are generated.

   (a) $X =$ **the smallest,** $Y =$ **the biggest** of them.
   Replacing $n = 4, i = 1, j = 4$, we get:

   $$f(x,y) = 12\;(y-x)^2 \quad \text{if } 0 < x < y < 1$$

   (b) $X =$ **the smallest,** $Y =$ **the second smallest** of them.
   Replacing $n = 4, i = 1, j = 2$, we get:

   $$f(x,y) = 12(1-y)^2 \quad \text{if } 0 < x < y < 1$$

(c) $X=$ **the second smallest,** $Y=$ **the third smallest** of them.

$$f(x,y) = 24x(1-y)x \quad \text{if } \ 0<x<y<1$$

3. **Ten independent random numbers** (uniformly distributed between 0 and 1) are generated.

   $X=$ **the 3rd smallest,** $Y=$ **the 7th smallest** of them.

   Replacing $n=10, i=3, j=7$, we get:

$$f(x,y) = \frac{10!}{2!\ 3!\ 3!} \ x^2\ (y-x)^3\ (1-y)^3 \quad \text{if } \ 0<x<y<1$$

**More general two-dimensional beta distributions.** If the people arrive between $A$ and $B$ instead of 0 and 1, that is, the $n$ independent, uniformly distributed random numbers are generated between $A$ and $B$, and

$X=$ the $i$th smallest of them

$Y=$ the $j$th smallest of them

of them, then for $A<x<B$, the density function of $(X,Y)$ is

$$f(x,y) =$$

$$\frac{1}{(B-A)^2} \frac{n!}{(i-1)!\ (j-i-1)!\ (n-j)!} \left(\frac{x-A}{B-A}\right)^{i-1} \left(\frac{y-x}{B-A}\right)^{j-i-1} \left(\frac{B-y}{B-A}\right)^{n-j}$$

Files to study two-dimensional beta point-clouds:

*Demonstration file: Two-dimensional beta point-cloud related to size* $2$ *and ranks* $1$ *and* $2$
*ef-200-69-00*

*Demonstration file: Two-dimensional beta point-cloud related to size* $3$ *and ranks* $1$ *and* $2$
*ef-200-70-00*

*Demonstration file: Two-dimensional beta point-cloud related to size* $3$ *and ranks* $1$ *and* $3$
*ef-200-71-00*

*Demonstration file: Two-dimensional beta point-cloud related to size* $3$ *and ranks* $2$ *and* $3$
*ef-200-72-00*

*Demonstration file: Two-dimensional beta point-cloud related to size* $5$ *and ranks* $k_1$ *and* $k_2$
*ef-200-73-00*

*Demonstration file: Two-dimensional beta point-cloud related to size 10 and ranks $k_1$ and $k_2$*
*ef-200-74-00*

File to study two-dimensional point-clouds for arrival times:

*Demonstration file: Two-dimensional gamma distribution*
*ef-200-68-00*

# 4 Projections and conditional distributions

**Projections.** If the density function of the two-dimensional random variable $(X, Y)$ is f(x,y), then the density function $f_1(x)$ of the random variable $X$ can be calculated by integration:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\ dy$$

**Sketch of proof.** The interval $[x, x + \Delta x]$ on the horizontal line defines a vertical strip in the plane:

$$S_{[x,x+\Delta x]} = \{(x, y) : x \in [x, x + \Delta x]\}$$

so that the event $X \in [x, x + \Delta x]$ is equivalent to $(X, Y) \in S_{[x,x+\Delta x]}$. Using this fact we get that

$$f_1(x) \approx \frac{\mathbf{P}\left(X \in [x, x + \Delta x]\right)}{\Delta x} = \frac{\mathbf{P}\left((X, Y) \in S_{[x,x+\Delta x]}\right)}{\Delta x} =$$

$$\frac{\iint\limits_{S_{[x,x+\Delta x]}} f(x, y)\, dx\ dy}{\Delta x} = \frac{\int\limits_{x}^{x+\Delta x} \left( \int\limits_{-\infty}^{\infty} f(x, y)\ dy \right)\ dx}{\Delta x} \approx \int\limits_{-\infty}^{\infty} f(x, y)\ dy$$

Similarly, the density function $f_2(x)$ of the random variable $Y$ is

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\ dx$$

**Conditional distributions.** If a two-dimensional random variable $(X, Y)$ is considered, and somehow the actual value $x$ of $X$ is known, but the value of $Y$ is unknown, then we may need to know the conditional distribution of $Y$ under the condition that $X = x$. The conditional density function can be calculated by division:

$$f_{2|1}(y|x) = \frac{f(x, y)}{f_1(x)}$$

Similarly, the conditional density function of $X$ under the condition that $Y = y$ is

$$f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)}$$

**Sketch of proof.** We give the proof of the first formula.

$$f_{2|1}(y|x) \approx \frac{1}{\Delta y} \ \mathbf{P}\left(Y \in [y, y+\Delta y] \ \mid \ X = x\right) \approx$$

$$\frac{1}{\Delta y} \ \mathbf{P}\left(Y \in [y, y+\Delta y] \ \mid \ X \in [x, x+\Delta x]\right) =$$

$$\frac{1}{\Delta y} \ \frac{\mathbf{P}\left(X \in [x, x+\Delta x] \text{ and } Y \in [y, y+\Delta y]\right)}{\mathbf{P}\left(X \in [x, x+\Delta x]\right)} =$$

$$\frac{1}{\Delta y} \ \frac{\frac{\mathbf{P}(X \in [x,x+\Delta x] \text{ and } Y \in [y,y+\Delta y])}{\Delta x}}{\frac{\mathbf{P}(X \in [x,x+\Delta x])}{\Delta x}} =$$

$$\frac{\frac{\mathbf{P}(X \in [x,x+\Delta x] \text{ and } Y \in [y,y+\Delta y])}{\Delta x \ \Delta y}}{\frac{\mathbf{P}(X \in [x,x+\Delta x])}{\Delta x}} \approx \frac{f(x,y)}{f_1(x)}$$

**Product rules.** It often happens that the density function of $(X, Y)$ is calculated from one of the product rules:

$$f(x,y) = f_1(x) \ f_{2|1}(y|x)$$

$$f(x,y) = f_2(y) \ f_{1|2}(x|y)$$

**Conditional distribution function.** The distribution function of the conditional distribution is calculated from the conditional density function by integration:

$$F_{2|1}(y|x) = \int_{\infty}^{y} f_{2|1}(y|x) \ dy = P\left(Y < y \mid X = x\right)$$

Similarly,

$$F_{1|2}(x|y) = \int_{\infty}^{x} f_{1|2}(x|y) \ dx = P\left(X < x \mid Y = y\right)$$

On the contrary, the conditional density function is a partial derivative of the conditional distribution function:

$$f_{2|1}(y|x) = \frac{\partial F_{2|1}(y|x)}{\partial y}$$

Similarly,

$$f_{1|2}(x|y) = \frac{\partial F_{1|2}(x|y)}{\partial x}$$

**Conditional probability.** The conditional probability of an interval for $Y$, under the condition that $X = x$, can be calculated from the conditional density by integration:

$$\mathbf{P}(\ y_1 < Y < y_2 \mid X = x\ ) = \int_{y_1}^{y_2} f_{2|1}(y|x)\ dy$$

Similarly, the conditional probability of an interval for $X$, under the condition that $Y = y$, can be calculated from the other conditional density by integration:

$$\mathbf{P}(\ x_1 < X < x_2 \mid Y = y\ ) = \int_{x_1}^{x_2} f_{1|2}(x|y)\ dx$$

The conditional probability of an interval for $Y$, under the condition that $X = x$, can be calculated from the conditional distribution function as a difference:

$$\mathbf{P}(\ y_1 < Y < y_2 \mid X = x\ ) = F_{2|1}(y_2|x) - F_{2|1}(y_1|x)$$

Similarly, the conditional probability of an interval for $X$, under the condition that $Y = y$, can be calculated from the other conditional distribution function as a difference:

$$\mathbf{P}(\ x_1 < X < x_2 \mid Y = y\ ) = F_{1|2}(x_1|y) - F_{1|2}(x_1|y)$$

**Remark.** Notice that in the conditional probability

$$\mathbf{P}(\ y_1 < Y < y_2 \mid X = x\ )$$

the probability of the condition is zero:

$$\mathbf{P}(X = x\ ) = 0$$

Thus, the definition

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$$

would not be applicable to define $\mathbf{P}(\ y_1 < Y < y_2 \mid X = x\ )$.

**Conditional median.** Solving the equation

$$F_{2|1}(y|x) = \frac{1}{2}$$

for $y$, that is, expressing $y$ in terms of $x$, we get the conditional median of $Y$ on condition that $X = x$.

Similarly, solving the equation

$$F_{1|2}(x|y) = \frac{1}{2}$$

for $x$, that is, expressing $x$ in terms of $y$, we get the conditional median of $X$ on condition that $Y = y$.

**Conditional expected value.** The conditional expected value is the expected value of the conditional distribution:

$$\mathbf{E}(Y|X = x) = \mu_{2|1}(|x) = \int_{-\infty}^{\infty} y \ f_{2|1}(y|x) \ dy$$

$$\mathbf{E}(X|Y = y) = \mu_{1|2}(|y) = \int_{-\infty}^{\infty} x \ f_{1|2}(x|y) \ dx$$

**Conditional variance.** The variance of the conditional distribution is the conditional variance:

$$\mathbf{VAR}(Y|X = x) = \sigma^2_{2|1}(|x) =$$

$$\int_{-\infty}^{\infty} \left(y - \mu_{2|1}(|x)\right)^2 \ f_{2|1}(y|x) \ dy = \int_{-\infty}^{\infty} y^2 \ f_{2|1}(y|x) \ dy \quad - \left( \ \mu_{2|1}(|x) \ \right)^2$$

$$\mathbf{VAR}(X|Y = y) = \sigma^2_{1|2}(|y) =$$

$$\int_{-\infty}^{\infty} \left(x - \mu_{1|2}(|y)\right)^2 \ f_{1|2}(x|y) \ dx = \int_{-\infty}^{\infty} x^2 \ f_{1|2}(x|y) \ dx \quad - \left( \ \mu_{1|2}(|y) \ \right)^2$$

**Conditional standard deviation.** The standard deviation of the conditional distribution is the conditional standard deviation:

$$\mathbf{SD}(Y|X = x) = \sigma_{2|1}(|x) =$$

$$\sqrt{\int_{-\infty}^{\infty} \left(y - \mu_{2|1}(|x)\right)^2 \ f_{2|1}(y|x) \ dy} = \sqrt{\int_{-\infty}^{\infty} y^2 \ f_{2|1}(y|x) \ dy - \left(\mu_{2|1}(|x)\right)^2}$$

$$\mathbf{SD}(X|Y = y) = \sigma_{1|2}(|y) =$$

$$\sqrt{\int_{-\infty}^{\infty} \left(x - \mu_{1|2}(|y)\right)^2 \ f_{1|2}(x|y) \ dx} = \sqrt{\int_{-\infty}^{\infty} x^2 \ f_{1|2}(x|y) \ dx - \left(\mu_{1|2}(|y)\right)^2}$$

**Remark.** The notion of the conditional variance and the conditional standard deviation can obviously be introduced and calculated for discrete distributions as well: in the above formulas, instead of integration summation is taken.

**Example.** We choose a random number between $0$ and $1$ according to uniform distribution, let it be $X$. If $X = x$, then we choose another random number between $0$ and $x$ according to uniform distribution, let it be $Y$. We shall calculate now all the density functions. By the definition of $X$ and $Y$, we may write:

$$f(x) = 1 \quad \text{if} \quad 0 < x < 1$$

$$f_{2|1}(y|x) = \frac{1}{x} \quad \text{if} \quad 0 < y < x < 1$$

By the product rule:

$$f(x,y) = f_1(x)\ f_{2|1}(y|x) = 1\ \frac{1}{x} = \frac{1}{x} \quad \text{if} \quad 0 < y < x < 1$$

By the integration rule:

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\ dx = \int_y^y \frac{1}{x}\ dx = -\ln(y) \quad \text{if} \quad 0 < y < 1$$

By the division rule:

$$f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{\frac{1}{x}}{-\ln(y)} = -\frac{1}{x\ \ln(y)} \quad \text{if} \quad 0 < y < x < 1$$

The conditional distribution function is

$$F_{1|2}(x|y) = \int_{\infty}^{x} f_{1|2}(x|y)\ dx =$$

$$\int_y^x \left( -\frac{1}{x\ \ln(y)} \right)\ dx = \frac{1}{-\ln(y)} \int_y^x \frac{1}{x}\ dx =$$

$$\frac{1}{-\ln(y)}\ (\ln(x) - \ln(y)) = 1 - \frac{\ln(x)}{\ln(y)} \quad \text{if} \quad 0 < y < 1$$

The conditional median is the solution for $x$ of the equation

$$F_{1|2}(x|y) = 1 - \frac{\ln(x)}{\ln(y)} = \frac{1}{2}$$

that is

$$\frac{\ln(x)}{\ln(y)} = \frac{1}{2}$$

$$\ln(x) = \frac{1}{2} \ln(y)$$

$$\ln(x) = \ln(\sqrt{y})$$

$$x = \sqrt{y}$$

The conditional expected value is

$$\mathbf{E}(X|Y=y) = \mu_{1|2}(|y) = \int_{-\infty}^{\infty} x \; f_{1|2}(x|y) \; dx =$$

$$\int_{y}^{1} x \; \left(-\frac{1}{x \; \ln(y)}\right) \; dx = \int_{y}^{1} \; \left(-\frac{1}{\ln(y)}\right) \; dx =$$

$$\left(-\frac{1}{\ln(y)}\right) \; (1-y) = \frac{y-1}{\ln(y)} \quad \text{if} \quad 0 < y < 1$$

*Files to visualize projections and conditional distributions:*

*Demonstration file: $X = RND_1$ , $Y = X \; RND_2$ , projections and conditional distributions*
*ef-200-79-00*

*Demonstration file: Two-dim beta distributions, $n = 10$ , projections and conditional distributions*
*ef-200-80-00*

*Demonstration file: Two-dim beta distributions, $n \leq 10$ , projections and conditional distributions*
*ef-200-81-00*

*Files to study construction of a two-dimensional continuous distribution using conditional distributions:*

*Demonstration file: Conditional distributions, uniform on parallelogram*
*ef-200-84-00*

*Demonstration file: Conditional distributions, $(RND_1, RND_1 RND_2)$*
*ef-200-85-00*

*Demonstration file: Conditional distributions, uniform on triangle*
*ef-200-86-00*

*Demonstration file: Conditional distributions, Bergengoc bulbs*
*ef-200-87-00*

*Demonstration file: Conditional distributions, standard normal*
*ef-200-88-00*

*Demonstration file: Conditional distributions, normal*
*ef-200-89-00*

# 5 Normal distributions in two-dimensions

**Standard normal distribution.** The two-dimensional standard normal distribution is defined on the whole plane by its density function:

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

Since the value of the density function depends on $x$ and $y$ only through $x^2 + y^2$, the distribution is circular symmetrical. Actually, the surface defined by the density function resembles a hat or a bell.

It is easy to check that the projections of the standard normal distribution onto both axes are one-dimensional standard normal distributions.

The conditional distributions both on the vertical and the horizontal lines are one-dimensional standard normal distributions, as well.

**General two-dimensional normal distributions.** General two-dimensional normal distributions are defined by their density function:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\, e^{-\frac{1}{2}\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2r\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2}}{1-r^2}}$$

where the parameters $\mu_1$, $\mu_2$ are real numbers, $\sigma_1$, $\sigma_2$ are positive numbers, and $r$ is a number between $-1$ and $1$, equality permitted. The surface defined by the density function resembles a hat which is compressed in one direction so that looking at it from above it has an elliptical shape.

It can be shown that the level curves of the density function are ellipses, whose center is at the point $(\mu_1, \mu_2)$, the directions of the axes are determined by the directions of the eigen-vectors of the covariance matrix, and the sizes of the axes are proportional to the square roots of the eigen-values.

**Projections.** It is easy to check that the projections of a two-dimensional normal distribution onto both axes are normal distributions. The projection onto the horizontal axis is a normal distribution with parameters $\mu_1$ and $\sigma_1$. The projection onto the vertical axis is a normal distribution with parameters $\mu_2$ and $\sigma_2$. The density function of the projection onto the horizontal axis is:

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\, e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

The density function of the projection onto the vertical axis is:

$$f_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2}\, e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

**Correlation coefficient .** The value of the correlation coefficient can be calculated according to its definition. It turns out that its value is equal to the value of the parameter $r$.

**Conditional distributions.** The conditional distributions are normal distributions. The conditional density function along the vertical line, when $X = x$, is:

$$f_{2|1}(y|x)\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}}\ e^{-\frac{1}{2}\left(\frac{y-\left(\mu_2+r\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)}{\sigma_2\sqrt{1-r^2}}\right)^2}$$

The conditional density function along the horizontal line, when $Y = y$, is:

$$f_{1|2}(x|y)\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-r^2}}\ e^{-\frac{1}{2}\left(\frac{x-\left(\mu_1+r\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)}{\sigma_1\sqrt{1-r^2}}\right)^2}$$

**Conditional median and expected value.** Both conditional medians and expected values depend on the condition linearly. Namely, the conditional median and expected value of $Y$, when $X = x$, is:

$$\mu_2 + r\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

The straight line defined by this formula is the so called **first regression line**.

Similarly, the conditional median and expected value of $X$, when $Y = y$, is:

$$\mu_1 + r\frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

The straight line defined by this formula is the so called **second regression line**. The regression lines mark the place of the conditional medians and expected values.

**Conditional standard deviation.** Both conditional standard deviations do not depend on the condition, they are a constants. Namely, the conditional standard deviation of $Y$, when $X = x$, is:

$$\sigma_2\sqrt{1-r^2}$$

If we move the first regression line up and down by an amount equal to the value of the conditional standard deviation we get the so called conditional **standard deviation lines** associated to the first regression line.

Similarly, the conditional standard deviation of $X$, when $Y = y$, is:

$$\sigma_1\sqrt{1-r^2}$$

If we move the second regression line to the right and to the left by an amount equal to the value of the conditional standard deviation we get the so called conditional **standard deviation lines** associated to the second regression line.

**Standard deviation rectangle.** In order to get a better view of a normal distribution, let us consider the rectangle defined by the direct product of the intervals

$$(\mu_1 - \sigma_1; \mu_1 + \sigma_1) \text{ and } (\mu_2 - \sigma_2; \mu_2 + \sigma_2)$$

We may take also the rectangles defined by the direct product of the intervals

$$(\mu_1 - s\sigma_1; \mu_1 + s\sigma_1) \text{ and } (\mu_2 - s\sigma_2; \mu_2 + s\sigma_2)$$

where $s$ is a positive number. This rectangle may be called the **standard deviation rectangle** of size $s$. Let us put a scale from $-1$ to $1$ on each of the sides of the standard deviation rectangle so that the $-1$ is at the left end, and the is $1$ at the right end on the horizontal sides, and the $-1$ is at the lower end, and the is $1$ at the upper end on the vertical sides. The points on the sides which correspond to $r$, the value of the correlation coefficient, play an interesting role, since the following facts are true:

1. The sides of the standard deviation rectangles are *tangents to the ellipses* which arise as level curves, and the common points of the ellipses and the standard deviation rectangles correspond to $r$, the value of the *correlation coefficient*, on the scales on the sides of the standard deviation rectangles.

2. The *regression lines* intersect the standard deviation rectangles at points which correspond to $r$ on the scales on the sides of the standard deviation rectangles.

3. If we draw the standard deviation rectangle of size 1, and consider the ellipse which arises as a level curve, touching the standard deviation rectangle at the point having a position $r$ on the on the scales on the sides of the standard deviation rectangle, then we may draw the tangent lines to the ellipse, parallel to the regression lines. These lines that we get are the *standard deviation lines* associated to the regression lines.

*Files to study the height and weight of men as a two-dimensional normal random variable:*

*Demonstration file: Height and weight*
*ef-200-65-00*

*Demonstration file: Height and weight, ellipse, eigen-vectors (projections and conditional distributions are also studied)*
*ef-200-82-00*

*Demonstration file: Two-dim normal distributions normal distributions, projections and conditional distributions*
*ef-200-83-00*

*File to study voltages as a two-dimensional normal random variable:*

*Demonstration file: Measuring voltage*
*ef-200-66-00*

# 6 Independence of random variables

The discrete random variables $X$ and $Y$ are **independent**, if any (and then all) of the following relations hold:

$$p_{2|1}(y|x) = p_2(y) \quad \text{for all } x \text{ and } y$$

$$p_{1|2}(x|y) = p_1(x) \quad \text{for all } x \text{ and } y$$

$$p(x, y) = p_1(x)p_2(y) \quad \text{for all } x \text{ and } y$$

The continuous random variables $X$ and $Y$ are **independent**, if any (and then all) of the following relations hold:

$$f_{2|1}(y|x) = f_2(y) \quad \text{for all } x \text{ and } y$$

$$f_{1|2}(x|y) = f_1(x) \quad \text{for all } x \text{ and } y$$

$$f(x, y) = f_1(x)f_2(y) \quad \text{for all } x \text{ and } y$$

If some random variables are not independent, then we call them **dependent**.

*File to study the notion of dependence and independence:*

*Demonstration file: Lengths dependent or independent*
*ef-200-91-00*

# 7 Generating a two-dimensional random variable

It is important that for a given two-dimensional distribution that a two-dimensional random variable can be generated by a calculator or a computer so that its distribution is the given two-dimensional continuous distribution. If the distribution is continuous, then the method described below defines such a two-dimensional random variable. (The discrete case is left for the reader as an exercise.)

In order to find the desired distribution in the continuous case, let the distribution function of its projection onto the horizontal axis be denoted by $F_1(x)$. Let $F^{-1}(u)$ be its inverse. (If $F_1(x)$ not strictly increasing on the the whole real-line, but only on an interval $(A, B)$, then $F^{-1}(u)$ should be the inverse of restriction of $F(x)$ onto that interval.) The way how we technically find a formula for $F_1^{-1}(u)$ is that we solve the equation

$$u = F_1(x)$$

for x, that is, we express $x$ from the equation in terms of $u$:

$$x = F_1^{-1}(u)$$

In a similar way, let the distribution function of the conditional distributions on the vertical axes be denoted by $F_{2|1}(y|x)$, and their inverse be denoted by $F_{2|1}^{-1}(v|x)$. The way how we find technically a formula for $F_{2|1}^{-1}(v|x)$ is that we solve the equation

$$v = F_{2|1}(y|x)$$

for y, that is, we express $y$ from the equation in terms of $v$:

$$y = F_{2|1}^{-1}(v|x)$$

In this calculation $x$ plays the role of a parameter.

Now we define the random value $X$ by

$$X = F_1^{-1}(\mathrm{RND}_1)$$

It is easy to be convinced that the distribution function of the random variable $X$ is the function $F_1(x)$. Then we define the random variable $Y$ by

$$Y = F_{2|1}^{-1}(\mathrm{RND}_2|X)$$

It is easy to be convinced that the conditional distribution function of the random variable $Y$ on condition that $X = x$ is the function $F_{2|1}(y|x)$. The two facts that

1. the distribution function of the random variable $X$ is the function $F_1(x)$
2. the conditional distribution function of the random variable $Y$ on condition that $X = x$ is the function $F_{2|1}(y|x)$

mean that the random variable $(X, Y)$ has the given two-dimensional continuous distribution.

# 8 Properties of the expected value, variance and standard deviation

In this section, we present the most important properties of the expected value, variance and standard deviation. In the formulas bellow $X$, $Y$, $X_1$, $X_2$, ..., $X_n$ represent random variables, and $a$, $b$, $c$, $n$, $a_1$, $a_2$, ..., $a_n$, represent constants.

1. **Addition rule for the expected value:**

   (a) For two terms:

   $$\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$$

   (b) For more terms:

   $$\mathbf{E}(X_1 + X_2 + \ldots + X_n) = \mathbf{E}(X_1) + \mathbf{E}(X_2) + \ldots + \mathbf{E}(X_n)$$

2. **Expected value of constant times a random variable:**

   $$\mathbf{E}(cX) = c\mathbf{E}(X)$$

3. **Linearity of the expected value:**

   (a) For two terms:

   $$\mathbf{E}(a\ X + b\ Y) = a\ \mathbf{E}(X) + b\ \mathbf{E}(Y)$$

   (b) For more terms:

   $$\mathbf{E}(a_1X_1 + a_2X_2 + \ldots + a_nX_n) =$$

   $$a_1\mathbf{E}(X_1) + a_2\mathbf{E}(X_2) + \ldots + a_n\mathbf{E}(X_n)$$

4. **Expected value of the sum with identical expected values:**
   If $X_1, X_2, \ldots, X_n$ have an identical expected value $\mu$, then

   $$\mathbf{E}(X_1 + X_2 + \ldots + X_n) = n\mu$$

5. **Expected value of the average with identical expected values:**
   If $X_1, X_2, \ldots, X_n$ have an identical expected value $\mu$, then

   $$\mathbf{E}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \mu$$

6. **Expected value of the product for independent random variables:**
   If $X$ and $Y$ are independent, then

   $$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$$

7. **Variance of a sum:**

   $$\mathbf{VAR}(X+Y) = \mathbf{VAR}(X) + \mathbf{VAR}(Y) + 2\mathbf{COV}(X,Y)$$

   where $\mathbf{COV}(X,Y)$ is the covariance between $X$ and $Y$ (see the definition of the covariance later).

8. **Variance of the sum of independent random variables:**
   If $X$ and $Y$ are independent, then

   $$\textbf{VAR}(X+Y) = \textbf{VAR}(X) + \textbf{VAR}(Y)$$

9. **Variance of constant times a random variable:**

   $$\textbf{VAR}(cX) = c^2\ \textbf{VAR}(X)$$

10. **Variance of the sum of independent random variables:**
    If $X_1, X_2, \ldots, X_n$ are independent and have a common variance $\sigma^2$, then

    $$\textbf{VAR}(X_1 + X_2 + \ldots + X_n) = n\,\sigma^2$$

11. **Variance of the average of independent random variables:**
    If $X_1, X_2, \ldots, X_n$ are independent and have a common variance $\sigma^2$, then

    $$\textbf{VAR}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

12. **Standard deviation of constant times a random variable:**

    $$\textbf{SD}(cX) = |c|\ \textbf{SD}(X)$$

13. **Square root law for the standard deviation of the sum:**
    If $X_1, X_2, \ldots, X_n$ are independent and have a common standard deviation $\sigma$, then

    $$\textbf{SD}(X_1 + X_2 + \ldots + X_n) = \sqrt{n}\,\sigma$$

14. **Square root law for the standard deviation of the average**:
    If $X_1, X_2, \ldots, X_n$ are independent and have a common standard deviation $\sigma$, then

    $$\textbf{SD}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}$$

*File to study the "average"-property of the standard deviation:*

*Demonstration file: Standard deviation of the average*
*ef-200-61-00*

*Demonstration file: Standard deviation of the average*
*ef-200-61-01*

## 9 Transformation from plane to line

When the two-dimensional random continuous variable $(X, Y)$ has a density function $f(x, y)$, and $z = t(x, y)$ is a given function, then the distribution function $R(z)$ of the random variable $Z = t(X, Y)$ is

$$R(z) = \iint\limits_{A_z} f(x,y)\,dx\;dy$$

where the set $A_z$ is the inverse image of the interval $(-\infty, z)$ at the transformation $z = t(x, y)$:

$$A_z = \{(x,y):\; t(x,y) < z)\}$$

**Sketch of proof.** The event $Z < z$ is equivalent to the event $(X, Y) \in A_z$, so

$$R(z) = \mathbf{P}(Z < z) = \mathbf{P}(\ (X,Y) \in A_z) = \iint\limits_{A_z} f(x,y)\,dx\;dy$$

Taking the derivative with respect to $z$ on both sides, we get the density function:

$$r(z) = R'(z)$$

*Files to study transformations from plane to line:*

*Demonstration file: Transformation from square to line by product*
*ef-300-02-00*

*Demonstration file: Transformation from square to line by ratio*
*ef-300-03-00*

*Demonstration file: Transformation from plane into chi distribution*
*ef-300-04-00*

*Demonstration file: Transformation from plane into chi-square distribution*
*ef-300-05-00*

**Projections from plane onto the axes.** If $t(x, y) = x$, then the transformation means the projection onto the x-axis. Recall that the density function $f_1(x)$ can be calculated from f(x,y) by integration with respect to $y$:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\;dy$$

Similarly, if $t(x, y) = y$, then the transformation means the projection onto the y-axis. The density function $f_2(x)$ can be calculated from f(x,y) by integration with respect to $x$:

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

*Files to study projections from plane to axes:*

*Demonstration file: Projection from triangle onto axes:* $(\max(\mathit{RND}_1, \mathit{RND}_2); \min(\mathit{RND}_1, \mathit{RND}_2))$
*ef-300-06-00*

*Demonstration file: Projection from triangle onto axes:* $(\mathit{RND}_1; \mathit{RND}_1 \mathit{RND}_2)$
*ef-300-07-00*

*Demonstration file: Projection from sail onto axes:* $(\mathit{RND}_1 \mathit{RND}_2; \mathit{RND}_1/\mathit{RND}_2)$
*ef-300-08-00*

*Demonstration file: Projection from sale onto axes:* $(\sqrt{\mathit{RND}_1 \mathit{RND}_2}; \sqrt{\mathit{RND}_1/\mathit{RND}_2})$
*ef-300-09-00*

## 10 *** Transformation from plane to plane

**General case.** Assume that a the density function of a distribution on the plane is $f(x, y)$. Consider a one-to-one smooth transformation $t$ from the $(x, y)$-plane onto the $(u, v)$-plane given by a pair of functions:

$$u = u(x, y)$$

$$v = v(x, y)$$

Let the inverse of the transformation be given by the pair of functions

$$x = x(u, v)$$

$$y = y(u, v)$$

The Jacobian matrix of the inverse transformation plays an important role in the formula we will state. This is why we remind the reader that the Jacobian matrix of the inverse transformation is a two by two matrix consisting of partial derivatives:

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{\partial x(u,v)}{\partial u} & \frac{\partial x(u,v)}{\partial v} \\ \frac{\partial y(u,v)}{\partial u} & \frac{\partial y(u,v)}{\partial v} \end{pmatrix}$$

As the $(x, y)$-plane is transformed into the $(u, v)$-plane, the distribution on the $(x, y)$-plane is also transformed into a distribution on the $(u, v)$-plane. Let the density function of the arising distribution on the $(u, v)$-plane denoted by $s(u, v)$. Then the value of the new density function is equal to the value of the old density function multiplied by the absolute value of the determinant of the Jacobian matrix of the inverse transformation:

$$s(u, v) = f\left(x(u, v), y(u, v)\right) \quad \left| \det\left(\frac{\partial(x, y)}{\partial(u, v)}\right)\right|$$

**Sketch of proof.** Let us consider a small rectangle $B$ at the point $(u, v)$. Its inverse image on the $(x, y)$ plane is approximately a small parallelogram-like set $A$ at the point $(x, y)$ so that the ratio of their areas is approximately equal to the absolute value of the Jacobian matrix of the inverse transformation:

$$\frac{\text{area of } A}{\text{area of } B} \approx \left| \det\left(\frac{\partial(x, y)}{\partial(u, v)}\right)\right|$$

Using this fact we get that

$$s(u, v) \approx \frac{\mathbf{P}\left((U, V) \in B\right)}{\text{area of } B} = \frac{\mathbf{P}\left((X, Y) \in A\right)}{\text{area of } A} \; \frac{\text{area of } A}{\text{area of } B} \approx f(x, y) \quad \left| \det\left(\frac{\partial(x, y)}{\partial(u, v)}\right)\right|$$

*Files to study a transformation from plane to plane:*

*Demonstration file: Transformation from square onto a "sail"*
*ef-300-09-50*

**Special case: Multiplying by a matrix.** Let us consider the special case, when the transformation is a linear transformation

$$u = a_{11}x + a_{12}y$$
$$v = a_{21}x + a_{22}y$$

that is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Introducing the matrix notations

$$\mathbf{u} = \begin{pmatrix} u \\ v \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

the linear transformation can be written briefly as

$$\mathbf{u} = \mathbf{A}\mathbf{x}$$

Then the inverse transformation is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{u}$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$ and the Jacobian matrix of the inverse transformation is the inverse matrix $\mathbf{A}^{-1}$ itself.

So the new density function $s(\mathbf{u})$ expressed in terms of the old density $f(\mathbf{x})$ looks like this:

$$s(\mathbf{u}) = f(\mathbf{A}^{-1}\mathbf{u}) \ \left|\det\left(\mathbf{A}^{-1}\right)\right|$$

**Special case: Multiplying by a matrix and adding a vector.** Let us consider the special case, when the transformation is a linear transformation

$$\begin{aligned} u &= a_{11}x + a_{12}y + b_1 \\ v &= a_{21}x + a_{22}y + b_2 \end{aligned}$$

Introducing the notation

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

the linear transformation can be written briefly as

$$\mathbf{u} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

Then the inverse transformation is

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{u} - \mathbf{b})$$

So the new density function $s(\mathbf{u})$ expressed in terms of the old density $f(\mathbf{x})$ looks like as this:

$$s(\mathbf{u}) = f(\ \mathbf{A}^{-1}(\mathbf{u} - \mathbf{b})\ ) \ \left|\det\left(\mathbf{A}^{-1}\right)\right|$$

**Linear transformation of two-dimensional normal distributions.** If a two-dimensional normal distribution is transformed by a linear transformation

$$\mathbf{u} = \mathbf{A}\ \mathbf{x} \ + \ \mathbf{b}$$

then the new distribution is a normal distribution, too. If the expected value of the old distribution is $\mathbf{m}_{\text{old}}$, then the expected value of the new distribution is

$$\mathbf{m}_{\text{new}} = \mathbf{A}\ \mathbf{m}_{\text{old}} \ + \ \mathbf{b}$$

If the covariance matrix of the old distribution is $\mathbf{C}_{\text{old}}$, then the covariance matrix of the new distribution is

$$\mathbf{C}_{\text{new}} = \mathbf{A}\ \mathbf{C}_{\text{old}}\ \mathbf{A}^T$$

*Demonstration file: Linear transformation of the standard normal point-cloud*
*ef-300-10-00*

*Demonstration file: Linear transformation of normal distributions*
*ef-300-11-00*

# 11 *** Sums of random variables. Convolution

**Discrete random variables, general case.** Assume that the two-dimensional random variable $(X, Y)$ has a distribution $p(x, y)$. Let $Z$ denote the sum of $X$ and $Y$, that is, $Z = X + Y$. Then the distribution $r(z)$ of the sum is:

$$r(z) = \sum_{(x,y):\, x+y=z} p(x, y) = \sum_x p(x, z - x) = \sum_y p(z - y, y)$$

Notice that
- in the first summation, for a given value of $z$, the summation takes place for all possible values of $(x, y)$ for which $x + y = z$,
- in the second summation, for a given value of $z$, the summation takes place for all possible values of $x$.
- in the third summation, for a given value of $z$, the summation takes place for all possible values of $y$.

**Remark.** If $p(x, y)$ is zero outside a region $S$, then, in the first summation, for a given value of $z$, the summation can be restricted to the set $\{(x, y) : (x, y) \in S\}$:

$$r(z) = \sum_{(x,y)\in S:\, x+y=z} p(x, y)$$

In the second summation it can be restricted to the set $A_z = \{x : (x, z - x) \in S\}$:

$$r(z) = \sum_{x\in A_z} p(x, z - x)$$

In the third summation it can be restricted to the set $B_z = \{y : (z - y, y) \in S\}$:

$$r(z) = \sum_{y\in B_z} p(z - y, y)$$

**Discrete, independent random variables.** Assume now that the discrete random variables $X$ and $Y$ are independent. Since $p(x, y) = p_1(x)p_2(y)$, the above formulas reduce to:

$$r(z) = \sum_{(x,y):\ x+y=z} p_1(x)\ p_2(y) = \sum_x p_1(x)\ p_2(z-x) = \sum_y p_1(z-y)\ p_2(y)$$

We recognize that $r(z)$ is the **convolution** of the distributions $p_1(x)$ and $p_2(y)$.

**Example 1. Convolving binomial distributions.** If we convolve a binomial distribution with parameters $n_1$ and $p$ with a binomial distribution with parameters $n_2$ and $p$ (the parameter $p$ is the same for both distributions), then we get a binomial distribution with parameters $n_1 + n_2$ and $p$.

**Example 2. Convolving Poisson distributions.** If we convolve a Poisson distribution with parameter $\lambda_1$ with a Poisson distribution with parameter $\lambda_2$, then we get a Poisson distribution with parameters $\lambda_1 + \lambda_2$.

**Example 3. Convolving geometrical distributions.** If we convolve a geometrical distribution with parameter $p$ with a geometrical distribution with parameter $p$ (the parameter $p$ is the same for both distributions), then we get a second order negative binomial distribution with parameter $p$.

**Example 4. Convolving negative binomial distributions.** If we convolve a negative binomial distribution with parameters $r_1$ and $p$ with a negative binomial distribution with parameters $r_2$ and $p$ (the parameter $p$ is the same for both distributions), then we get a negative binomial distribution with parameters $r_1 + r_2$ and $p$.

*Files to study how the distribution of the sum can be calculated:*

*Demonstration file: Summation of independent random variables, fair dice*
*ef-300-12-00*

*Demonstration file: Summation of independent random variables, unfair dice*
*ef-300-13-00*

**Continuous random variables, general case.** Assume that the two-dimensional random variable $(X, Y)$ has a density function $f(x, y)$. Let us consider the sum of $X$ and $Y$: $Z = X + Y$. Then the density function $r(z)$ of the sum is:

$$r(z) = \int_{-\infty}^{\infty} f(x, z-x)\ dx = \int_{-\infty}^{\infty} f(z-y, y)\ dy$$

**Sketch of proof.** We shall perform the transformation in two steps. First we transform the distribution onto the $(u, v)$-plane by the linear transformation given by the equations

$$\begin{aligned} u &= x + y \\ v &= \phantom{x + {}} y \end{aligned}$$

and then we project onto the horizontal axis. Since the first coordinate of the above transformation is $u = x + y$, after the projection, we get what we need. The inverse transformation is

$$\begin{array}{ccccc} x & = & u & - & v \\ y & = & & & v \end{array}$$

The Jacobian matrix is

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ \\ & 1 \end{pmatrix}$$

so the Jacobian determinant is equal to 1. Thus

$$s(u,v) = f\ (\ u - v\ , v\ ) \quad 1 = f\ (\ u - v\ , v\ )$$

Now projecting onto the horizontal axis, the value of the density function at $u$ turns out to be

$$\int_{-\infty}^{\infty} s(u,v)\ dv = \int_{-\infty}^{\infty} f\ (\ u - v\ , v\ )\ dv$$

Replacing formally the letter $u$ by the letter $z$, and the integration variable $v$ by $y$, we get that the value of the density function at $z$ is

$$r(z) = \int_{-\infty}^{\infty} f\ (\ z - y\ , y\ )\ dy$$

**Remark.** If $f(x,y)$ is zero outside a region $S$, then, for a given $z$ value, the interval $(-\infty, \infty)$ in the first integral can be replaced by the set $A_z = \{x : (x, z - x) \in S\}$:

$$r(z) = \int_{A_z} f(x, z - x)\ dx$$

Similarly, the interval $(-\infty, \infty)$ in the second integral can be replaced by the set $B_z = \{y : (z - y, y) \in S\}$:

$$r(z) = \int_{B_z} f(z - y, y)\ dy$$

**Continuous, independent random variables.** Assume now that the continuous random variables $X$ and $Y$ are independent. Since $f(x,y) = f_1(x)f_2(y)$, the above formulas reduce to:

$$r(z) = \int_{-\infty}^{\infty} f_1(x)\ f_2(z - x)\ dx = \int_{-\infty}^{\infty} f_1(z - y)\ f_2(y)\ dy$$

We recognize that $r(z)$ is the **convolution** of the density functions $f_1(x)$ and $f_2(y)$.

**Example 1. Convolving exponential distributions.** If we convolve an exponential distribution with parameter $\lambda$ with an exponential distribution with parameter $\lambda$ (the parameter $\lambda$ is the same for both distributions), then we get a second order gamma distribution with parameter $\lambda$.

**Example 2. Convolving gamma distributions.** If we convolve a gamma distribution with parameters $n_1$ and $\lambda$ with a gamma distribution with parameters $n_2$ and $\lambda$ (the parameter $\lambda$ is the same for both distributions), then we get a gamma distribution with parameters $n_1 + n_2$ and $\lambda$.

**Example 3. Convolving normal distributions.** If we convolve a normal distribution with parameters $\mu_1$ and $\sigma_1$ with a normal distribution with parameters $\mu_2$ and $\sigma_2$ (the parameter $p$ is the same for both distributions), then we get a normal distribution with parameters $\mu_1 + \mu_2$ and $\sqrt{\sigma_1^2 + \sigma_2^2}$.

## 12 Limit theorems to normal distributions

**Moivre-Laplace theorem.** If, for a fixed $p$ value, we consider a binomial distribution with parameters $n$ and $p$ so that $n$ is large, then the binomial distribution can be well approximated by a normal distribution. The parameters, that is, the expected value and the standard deviation of the normal distribution should be taken to be equal to the expected value and the standard deviation of the binomial distribution, that is, $np$ and $\sqrt{np(1-p)}$.

If we standardize the binomial distribution, then the arising standardized binomial distribution will be close to the standard normal distribution.

Here is a file to study binomial approximation of normal distribution:

*Demonstration file: Binomial approximation of normal distribution*
*ef-300-14-00*

**Central limit theorem.** If we add many independent random variables, then the distribution of the sum can be calculated from the distributions of the random variables by convolution. It can be shown that, under very general conditions (which we do not give here), the distribution of the sum will approximate a normal distribution. The parameters, that is, the expected value and the standard deviation of the normal distribution should be taken to be equal to the expected value and the standard deviation of the sum.

If we standardize the sum, the arising standardized value will be distributed approximately according to the standard normal distribution.

**Central limit theorem in two-dimensions.** If we add many independent two-dimensional random variables (random vectors), then the distribution of the sum, under very general conditions (which we do not give here), the distribution of the sum will approximate a two-dimensional normal distribution.

Files to study how convolutions approximate normal distributions:

*Demonstration file: Convolution with uniform distribution*
*ef-300-15-00*

*Demonstration file: Convolution with asymmetrical distribution*
*ef-300-16-00*

*Demonstration file: Convolution with U-shaped distribution*
*ef-300-17-00*

*Demonstration file: Convolution with randomly chosen distribution*
*ef-300-18-00*

*File to study how gamma distributions approximate normal distributions:*

*Demonstration file: Gamma distribution approximates normal distribution*
*ef-300-19-00*

*Files to study the two-dimensional central limit theorem:*

*Demonstration file: Two-dimensional central-limit theorem, rectangle*
*ef-300-20-00*

*Demonstration file: Two-dimensional central-limit theorem, parallelogram*
*ef-300-21-00*

*Demonstration file: Two-dimensional central-limit theorem, curve*
*ef-300-22-00*