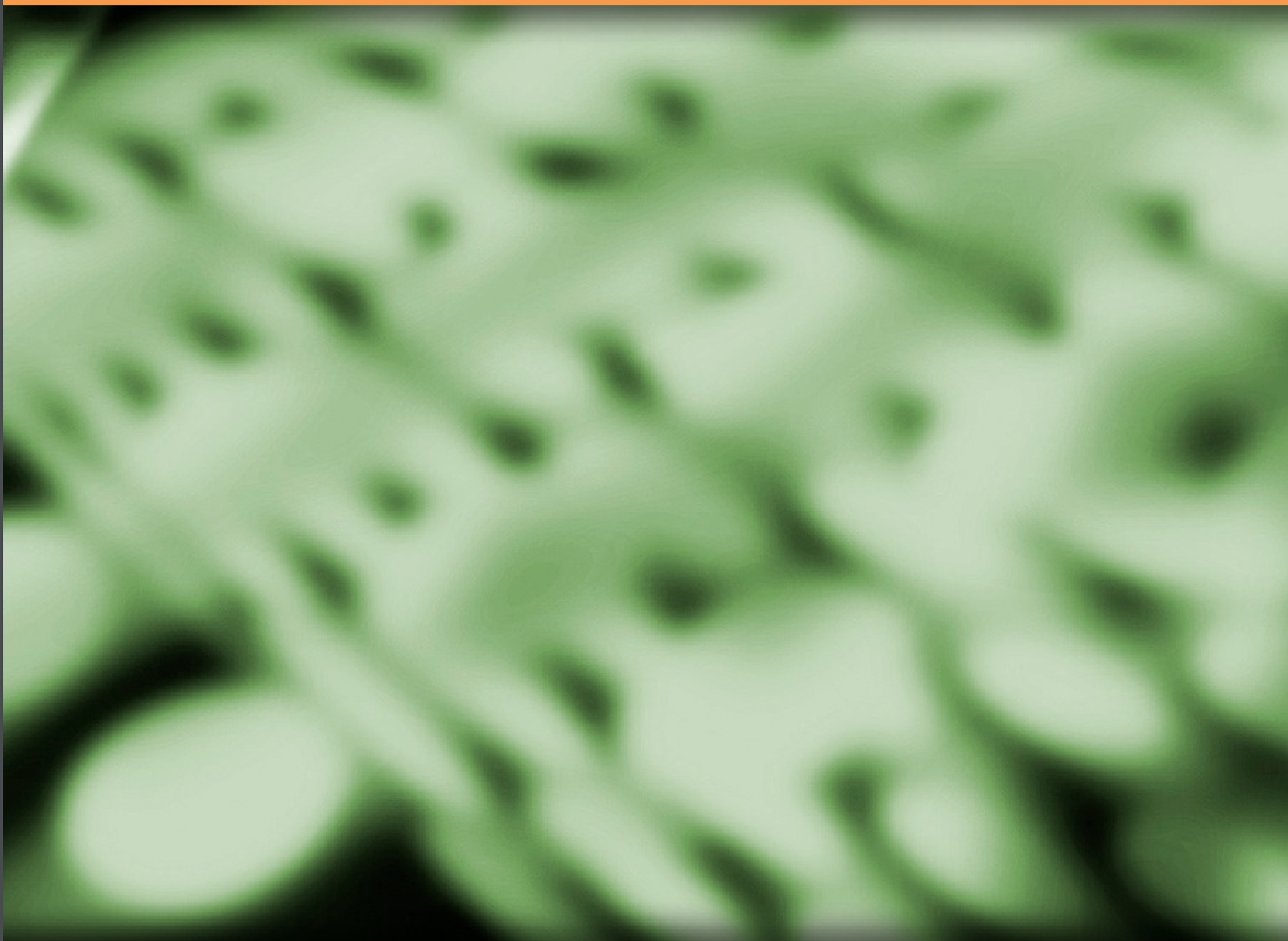
Random variables III

Probability Examples c-4

Leif Mejlbro



Leif Mejlbro

Probability Examples c-4
Random variables III

Probability Examples c-4 – Random variables III

ISBN 978-87-7681-519-6

Contents

Introduction

This is the fourth book of examples from the *Theory of Probability*. This topic is not my favourite, however, thanks to my former colleague, Ole Jørsboe, I somehow managed to get an idea of what it is all about. The way I have treated the topic will often diverge from the more professional treatment. On the other hand, it will probably also be closer to the way of thinking which is more common among many readers, because I also had to start from scratch.

The topic itself, *Random Variables*, is so big that I have felt it necessary to divide it into three books, of which this is the third one.

The prerequisites for the topics can e.g. be found in the *Ventus: Calculus 2* series, so I shall refer the reader to these books, concerning e.g. plane integrals.

Unfortunately errors cannot be avoided in a first edition of a work of this type. However, the author has tried to put them on a minimum, hoping that the reader will meet with sympathy the errors which do occur in the text.

Leif Mejlbro
26th October 2009

1 Some theoretical results

The abstract (and precise) definition of a *random variable* X is that X is a real function on Ω, where the triple $(\Omega, \mathcal{F}, P)$ is a probability field, such that

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F} \qquad \text{for every } x \in \mathbb{R}.$$

This definition leads to the concept of a *distribution function* for the random variable X, which is the function $F : \mathbb{R} \to \mathbb{R}$, which is defined by

$$F(x) = P\{X \leq x\} \qquad (= P\{\omega \in \Omega \mid X(\omega) \leq x\}),$$

where the latter expression is the mathematically precise definition which, however, for obvious reasons everywhere in the following will be replaced by the former expression.

A distribution function for a random variable X has the following properties:

$0 \leq F(x) \leq 1 \qquad$ for every $x \in \mathbb{R}$.

The function F is weakly increasing, i.e. $F(x) \leq F(y)$ for $x \leq y$.

$\lim_{x \to -\infty} F(x) = 0 \qquad$ and $\qquad \lim_{x \to +\infty} F(x) = 1$.

The function F is continuous from the right, i.e. $\lim_{h \to 0+} F(x+h) = F(x)$ for every $x \in \mathbb{R}$.

One may in some cases be interested in giving a crude description of the behaviour of the distribution function. We define a *median* of a random variable X with the distribution function $F(x)$ as a real number $a = (X) \in \mathbb{R}$, for which

$$P\{X \leq a\} \geq \frac{1}{2} \qquad \text{and} \qquad P\{X \geq a\} \geq \frac{1}{2}.$$

Expressed by means of the distribution function it follows that $a \in \mathbb{R}$ is a median, if

$$F(a) \geq \frac{1}{2} \qquad \text{and} \qquad F(a-) = \lim_{h \to 0-} F(x+h) \leq \frac{1}{2}.$$

In general we define a *p-quantile*, $p \in]0,1[$, of the random variable as a number $a_p \in \mathbb{R}$, for which

$$P\left\{X \leq a_p\right\} \geq p \qquad \text{and} \qquad P\left\{X \geq a_p\right\} \geq 1-p,$$

which can also be expressed by

$$F\left(a_p\right) \geq p \qquad \text{and} \qquad F\left(a_p-\right) \leq p.$$

If the random variable X only has a finite or a countable number of values, x_1, x_2, ..., we call it *discrete*, and we say that X has a *discrete distribution*.

A very special case occurs when X only has one value. In this case we say that X is *causally distributed*, or that X is constant.

The random variable X is called *continuous*, if its distribution function $F(x)$ can be written as an integral of the form

$$F(x) = \int_{-\infty}^{x} f(u)\,du, \qquad x \in \mathbb{R},$$

where f is a nonnegative integrable *function*. In this case we also say that X has a continuous *distribution*, and the integrand $f : \mathbb{R} \to \mathbb{R}$ is called a *frequency* of the random variable X.

Let again $(\Omega, \mathcal{F}, P)$ be a given probability field. Let us consider *two* random variables X and Y, which are both defined on Ω. We may consider the pair (X, Y) as a *2-dimensional random variable*, which implies that we then shall make precise the extensions of the previous concepts for a single random variable.

We say that the *simultaneous distribution*, or just the *distribution*, of (X, Y) is known, if we know

$$P\{(X, Y) \in A\} \qquad \text{for every Borel set } A \subseteq \mathbb{R}^2.$$

When the simultaneous distribution of (X, Y) is known, we define the *marginal distributions* of X and Y by

$$P_X(B) = P\{X \in B\} := P\{(X, Y) \in B \times \mathbb{R}\}, \qquad \text{where } B \subseteq \mathbb{R} \text{ is a Borel set,}$$

$$P_Y(B) = P\{Y \in B\} := P\{(X, Y) \in \mathbb{R} \times B\}, \qquad \text{where } B \subseteq \mathbb{R} \text{ is a Borel set.}$$

Notice that we can always find the *marginal distributions* from the *simultaneous distribution*, while it is far from always possible to find the *simultaneous distribution* from the *marginal distributions*. We now introduce

The *simultaneous distribution function* of the 2-dimensional random variable (X,Y) is defined as the function $F:\mathbb{R}^2\to\mathbb{R}$, given by

$$F(x,y) := P\{X \le x \wedge Y \le y\}.$$

We have

- If $(x,y)\in\mathbb{R}^2$, then $0 \le F(x,y) \le 1$.
- If $x\in\mathbb{R}$ is kept fixed, then $F(x,y)$ is a weakly increasing function in y, which is continuous from the right and which satisfies the condition $\lim_{y\to-\infty} F(x,y)=0$.
- If $y\in\mathbb{R}$ is kept fixed, then $F(x,y)$ is a weakly increasing function in x, which is continuous from the right and which satisfies the condition $\lim_{x\to-\infty} F(x,y)=0$.
- When both x and y tend towards infinity, then

  $$\lim_{x,\,y\to+\infty} F(x,y) = 1.$$

- If x_1, x_2, y_1, $y_2 \in\mathbb{R}$ satisfy $x_1\le x_2$ and $y_1 \le y_2$, then

  $$F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_2) \ge 0.$$

Given the simultaneous distribution function $F(x,y)$ of (X,Y) we can find the distribution functions of X and Y by the formulæ

$$F_X(x) = F(x,+\infty) = \lim_{y\to+\infty} F(x,y), \qquad \text{for } x\in\mathbb{R},$$

$$F_y(x) = F(+\infty,y) = \lim_{x\to+\infty} F(x,y), \qquad \text{for } y\in\mathbb{R}.$$

The 2-dimensional random variable (X,Y) is called *discrete*, or that it has a *discrete distribution*, if both X and Y are discrete.

The 2-dimensional random variable (X,Y) is called *continuous*, or we say that it has a *continuous distribution*, if there exists a nonnegative integrable function (a *frequency*) $f:\mathbb{R}^2\to\mathbb{R}$, such that the distribution function $F(x,y)$ can be written in the form

$$F(x,y) = \int_{-\infty}^{x}\left\{\int_{-\infty}^{y} f(t,u)\,du\right\}dt, \qquad \text{for } (x,y)\in\mathbb{R}^2.$$

In this case we can find the function $f(x,y)$ at the differentiability points of $F(x,y)$ by the formula

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

It should now be obvious why one should know something about the theory of integration in more variables, cf. e.g. the *Ventus: Calculus 2* series.

We note that if $f(x,y)$ is a frequency of the continuous 2-dimensional random variable (X,Y), then X and Y are both continuous 1-dimensional random variables, and we get their (marginal) frequencies by

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy, \qquad \text{for } x\in\mathbb{R},$$

and

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx, \qquad \text{for } y \in \mathbb{R}.$$

It was mentioned above that one far from always can find the simultaneous distribution function from the marginal distribution function. It is, however, possible in the case when the two random variables X and Y are *independent*.

Let the two random variables X and Y be defined on the same probability field $(\Omega, \mathcal{F}, P)$. We say that X and Y are *independent*, if for all pairs of Borel sets A, $B \subseteq \mathbb{R}$,

$$P\{X \in A \wedge Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\},$$

which can also be put in the simpler form

$$F(x,y) = F_X(x) \cdot F_Y(y) \qquad \text{for every } (x,y) \in \mathbb{R}^2.$$

If X and Y are *not* independent, then we of course say that they are *dependent*.

In two special cases we can obtain more information of independent random variables:

If the 2-dimensional random variable (X,Y) is *discrete*, then X and Y are independent, if

$$h_{ij} = f_i \cdot g_j \qquad \text{for every } i \text{ and } j.$$

Here, f_i denotes the probabilities of X, and g_j the probabilities of Y.

If the 2-dimensional random variable (X,Y) is *continuous*, then X and Y are independent, if their frequencies satisfy

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \text{almost everywhere.}$$

The concept *"almost everywhere"* is rarely given a precise definition in books on applied mathematics. Roughly speaking it means that the relation above holds outside a set in $\mathbb{R}^2$ of area zero, a so-called *null set*. The common examples of null sets are either finite or countable sets. There exists, however, also non-countable null sets. Simple examples are graphs of any (piecewise) C^1-curve.

Concerning maps of random variables we have the following very important results,

Theorem 1.1 *Let X and Y be independent random variables. Let $\varphi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be given functions. Then $\varphi(X)$ and $\psi(Y)$ are again independent random variables.*

If X is a continuous random variable of the frequency I, then we have the following important theorem, where it should be pointed out that one always shall check *all* assumptions in order to be able to conclude that the result holds:

Theorem 1.2 *Given a continuous random variable X of frequency f.*

1) Let I be an open interval, such that $P\{X \in I\} = 1$.

2) Let $\tau : I \rightarrow J$ be a bijective *map of I onto an* open *interval J.*

3) Furthermore, assume that τ is differentiable *with a* continuous *derivative τ', which satisfies*

$$\tau'(x) \neq 0 \qquad \text{for alle } x \in I.$$

Under the assumptions above $Y := \tau(X)$ is also a continuous random variable, and its frequency $g(y)$ is given by

$$g(y) = \begin{cases} f\left(\tau^{-1}(y)\right) \cdot \left|\left(\tau^{-1}\right)'(y)\right|, & \text{for } y \in J, \\ 0, & \text{otherwise.} \end{cases}$$

We note that if just one of the assumptions above is *not* fulfilled, then we shall instead find the distribution function $G(y)$ of $Y := \tau(X)$ by the general formula

$$G(y) = P\{\tau(X) \in]-\infty\,,\, y]\} = P\left\{X \in \tau^{\circ -1}(]-\infty\,,\, y])\right\},$$

where $\tau^{\circ -1} = \tau^{-1}$ denotes the inverse set map.

Note also that if the assumptions of the theorem are all satisfied, then τ is necessarily monotone.

At a first glance it may be strange that we at this early stage introduce 2-dimensional random variables. The reason is that by applying the simultaneous distribution for (X, Y) it is fairly easy to define the elementary operations of calculus between X and Y. Thus we have the following general result for a *continuous* 2-dimensional random variable.

Theorem 1.3 *Let (X, Y) be a* continuous *random variable of the frequency $h(x, y)$.*

The frequency of the sum $X + Y$ is $\qquad k_1(z) = \int_{-\infty}^{+\infty} h(x, z - x)\, dx.$

The frequency of the difference $X - Y$ is $\qquad k_2(z) = \int_{-\infty}^{+\infty} h(x, x - z)\, dx.$

The frequency of the product $X \cdot Y$ is $\qquad k_3(z) = \int_{-\infty}^{+\infty} h\left(x\,,\, \frac{z}{x}\right) \cdot \frac{1}{|x|}\, dx.$

The frequency of the quotient X/Y is $\qquad k_4(z) = \int_{-\infty}^{+\infty} h(zx\,,\, x) \cdot |x|\, dx.$

Notice that one must be very careful by computing the product and the quotient, because the corresponding integrals are improper.

If we furthermore assume that X and Y are *independent*, and $f(x)$ is a frequency of X, and $g(y)$ is a frequency of Y, then we get an even better result:

Theorem 1.4 *Let X and Y be* continuous *and* independent *random variables with the frequencies $f(x)$ and $g(y)$, resp..*

The frequency of the sum $X+Y$ is $\quad k_1(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)\,dx.$

The frequency of the difference $X-Y$ is $\quad k_2(z) = \int_{-\infty}^{+\infty} f(x)g(x-z)\,dx.$

The frequency of the product $X \cdot Y$ is $\quad k_3(z) = \int_{-\infty}^{+\infty} f(x)\,g\left(\frac{z}{x}\right) \cdot \frac{1}{|x|}\,dx.$

The frequency of the quotient X/Y is $\quad k_4 = \int_{-\infty}^{+\infty} f(zx)g(x) \cdot |x|\,dx.$

Let X and Y be *independent* random variables with the distribution functions F_X and F_Y, resp.. We introduce two random variables by

$$U := \max\{X,Y\} \qquad \text{and} \qquad V := \min\{X,Y\},$$

the distribution functions of which are denoted by F_U and F_V, resp.. Then these are given by

$$F_U(u) = F_X(u) \cdot F_Y(u) \qquad \text{for } u \in \mathbb{R},$$

and

$$F_V(v) = 1 - (1 - F_X(v)) \cdot (1 - F_Y(v)) \qquad \text{for } v \in \mathbb{R}.$$

These formulæ are general, provided only that X and Y are independent.

If X and Y are *continuous and independent*, then the frequencies of U and V are given by

$$f_U(u) = F_X(u) \cdot f_Y(u) + f_X(u) \cdot F_Y(u), \qquad \text{for } u \in \mathbb{R},$$

and

$$f_V(v) = (1 - F_X(v)) \cdot f_Y(v) + f_X(v) \cdot (1 - F_y(v)), \qquad \text{for } v \in \mathbb{R},$$

where we note that we shall apply both the frequencies and the distribution functions of X and Y.

The results above can also be extended to bijective maps $\underline{\varphi} = (\varphi_1, \varphi_2) : \mathbb{R}^2 \to \mathbb{R}^2$, or subsets of $\mathbb{R}^2$. We shall need the *Jacobian* of $\underline{\varphi}$, introduced in e.g. the *Ventus: Calculus 2* series.

It is important here to define the notation and the variables in the most convenient way. We start by assuming that D is an open domain in the $(x_1 \, x_2)$ plane, and that $\tilde{D}$ is an open domain in the (y_1, y_2) plane. Then let $\underline{\varphi} = (\varphi_1, \varphi_2)$ be a bijective map *of* $\tilde{D}$ *onto* D with the inverse $\underline{\tau} = \underline{\varphi}^{-1}$, i.e. the opposite of what one probably would expect:

$$\underline{\varphi} = (\varphi_1, \varphi_2) : \tilde{D} \to D, \qquad \text{with } (x_1, x_2) = \underline{\varphi}(y_1, y_2).$$

The corresponding *Jacobian* is defined by

$$J_{\underline{\varphi}} = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \dfrac{\partial \varphi_1}{\partial y_1} & \dfrac{\partial \varphi_2}{\partial y_1} \\[2ex] \dfrac{\partial \varphi_1}{\partial y_1} & \dfrac{\partial \varphi_2}{\partial y_2} \end{vmatrix},$$

where the independent variables (y_1, y_2) are in the "denominators". Then recall the *Theorem of transform of plane integrals*, cf. e.g. the *Ventus: Calculus 2* series: If $h : D \to \mathbb{R}$ is an integrable function, where $D \subseteq \mathbb{R}^2$ is given as above, then for every (measurable) subset $A \subseteq D$,

$$\int_A h(x_1, x_2) \, dx_1 dx_2 = \int_{\underline{\varphi}^{-1}(A)} h(x_1, x_2) \cdot \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1 dy_2.$$

Of course, this formula is not mathematically correct; but it shows intuitively what is going on: Roughly speaking we "delete the y-s". The correct mathematical formula is of course the well-known

$$\int_A h(x_1, x_2) \, dx_1 dx_2 = \int_{\underline{\varphi}^{-1}(A)} (\varphi_1(y_1, y_2), \varphi_2(y_1, y_2)) \cdot \left| J_{\underline{\varphi}}(y_1, y_2) \right| dy_1 dy_2,$$

although experience shows that it in practice is more confusing then helping the reader.

Theorem 1.5 *Let (X_1, X_2) be a continuous 2-dimensional random variable with the frequency $h(x_1, x_2)$. Let $D \subseteq \mathbb{R}^2$ be an open domain, such that*

$$P\{(X_1, X_2) \in D\} = 1.$$

Let $\underline{\tau} : D \to \tilde{D}$ be a bijective map of D onto another open domain $\tilde{D}$, and let $\underline{\varphi} = (\varphi_1, \varphi_2) = \underline{\tau}^{-1}$, where we assume that φ_1 and φ_2 have continuous partial derivatives and that the corresponding Jacobian is different from 0 in all of $\tilde{D}$.
Then the 2-dimensional random variable

$$(Y_1, Y_2) = \underline{\tau}(X_1, X_2) = (\tau_1(X_1, X_2), \tau_2(X_1, X_2))$$

has the frequency $k(y_1, y_2)$, given by

$$k(y_1, y_2) = \begin{cases} h(\varphi_1(y_1, y_2), \varphi_2(y_1, y_2)) \cdot \left| \dfrac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right|, & \text{for } (y_1, y_2) \in \tilde{D}, \\[2ex] 0, & \text{otherwise} \end{cases}$$

We have previously introduced the concept *conditional probability*. We shall now introduce a similar concept, namely the *conditional distribution*.

If X and Y are discrete, we define the *conditional distribution of X for given $Y = y_j$* by

$$P\{X = x_i \mid Y = y_j\} = \frac{P\{X = x_i \wedge Y = y_j\}}{P\{Y = y_j\}} = \frac{h_{ij}}{g_j}.$$

It follows that for fixed j we have that $P\{X = x_i \mid Y = y_j\}$ indeed is a distribution. We note in particular that we have the *law of the total probability*

$$P\{X = x_i\} = \sum_j P\{X = x_i \mid Y = y_j\} \cdot P\{Y = y_j\}.$$

Analogously we define for two continuous random variables X and Y the *conditional distribution function of X for given $Y = y$* by

$$P\{X \le x \mid Y = y\} = \frac{\int_{-\infty}^{x} f(u, y)\, du}{f_Y(y)}, \qquad \text{forudsat, at } f_Y(y) > 0.$$

Note that the conditional distribution function is not defined at points in which $f_Y(y) = 0$.

The corresponding frequency is

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}, \qquad \text{provided that } f_Y(y) = 0.$$

We shall use the convention that “0 times undefined = 0”. Then we get the *Law of total probability*,

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

We now introduce the *mean*, or *expectation* of a random variable, provided that it exists.

1) Let X be a *discrete* random variable with the possible values $\{x_i\}$ and the corresponding probabilities $p_i = P\{X = x_i\}$. The *mean*, or *expectation*, of X is defined by

$$E\{X\} := \sum_i x_i\, p_i,$$

provided that the series is absolutely convergent. If this is not the case, the mean does not exists.

2) Let X be a *continuous* random variable with the frequency $f(x)$. We define the *mean*, or *expectation* of X by

$$E\{X\} = \int_{-\infty}^{+\infty} x\, f(x)\, dx,$$

provided that the integral is absolutely convergent. If this is not the case, the mean does not exist.

If the random variable X only has nonnegative values, i.e. the image of X is contained in $[0, +\infty[$, and the mean exists, then the mean is given by

$$E\{X\} = \int_0^{+\infty} P\{X \geq x\}\, dx.$$

Concerning maps of random variables, means are transformed according to the theorem below, provided that the given expressions are absolutely convergent.

Theorem 1.6 *Let the random variable $Y = \varphi(X)$ be a function of X.*

1) If X is a discrete *random variable with the possible values $\{x_i\}$ of corresponding probabilities $p_i = P\{X = x_i\}$, then the* mean *of $Y = \varphi(X)$ is given by*

$$E\{\varphi(X)\} = \sum_i \varphi(x_i)\, p_i,$$

provided that the series is absolutely convergent.

2) If X is a continuous *random variable with the frequency $f(x)$, then the* mean *of $Y = \varphi(X)$ is given by*

$$E\{\varphi(X)\} = \int_{-\infty}^{+\infty} \varphi(x)\, g(x)\, dx,$$

provided that the integral is absolutely convergent.

Assume that X is a random variable of mean μ. We add the following concepts, where $k \in \mathbb{N}$:

The k-th moment,	$E\left\{X^k\right\}$.
The k-th absolute moment,	$E\left\{\lvert X\rvert^k\right\}$.
The k-th central moment,	$E\left\{(X-\mu)^k\right\}$.
The k-th absolute central moment,	$E\left\{\lvert X-\mu\rvert^k\right\}$.
The variance, i.e. the second central moment,	$V\{X\} = E\left\{(X-\mu)^2\right\}$,

provided that the defining series or integrals are absolutely convergent. In particular, the *variance* is very important. We mention

Theorem 1.7 *Let X be a random variable of mean $E\{X\} = \mu$ and variance $V\{X\}$. Then*

$$E\left\{(X-c)^2\right\} = V\{X\} + (\mu - c)^2 \qquad \text{for every } c \in \mathbb{R},$$

$$V\{X\} = E\left\{X^2\right\} - (E\{X\})^2 \qquad \text{for } c = 0,$$

$$E\{aX + b\} = a\,E\{X\} + b \qquad \text{for every } a,\, b \in \mathbb{R},$$

$$V\{aX + b\} = a^2 V\{X\} \qquad \text{for every } a,\, b \in \mathbb{R}.$$

It is not always an easy task to compute the distribution function of a random variable. We have the following result which gives an estimate of the probability that a random variable X differs more than some given $a > 0$ from the mean $E\{X\}$.

Theorem 1.8 (Čebyšev's inequality). *If the random variable X has the mean μ and the variance σ^2, then we have for every $a > 0$,*

$$P\{|X - \mu| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

If we here put $a = k\sigma$, we get the equivalent statement

$$P\{\mu - k\sigma < X < \mu + k\sigma\} \geq 1 - \frac{1}{k^2}.$$

These concepts are then generalized to 2-dimensional random variables. Thus,

Theorem 1.9 *Let $Z = \varphi(X,Y)$ be a function of the 2-dimensional random variable (X,Y).*

1) If (X,Y) is discrete*, then the* mean *of $Z = \varphi(X,Y)$ is given by*

$$E\{\varphi(X,Y)\} = \sum_{i,j} \varphi(x_i\,,\, y_j) \cdot P\{X = x_i \wedge Y = y_j\}\,,$$

provided that the series is absolutely convergent.

2) If (X,Y) is continuous*, then the* mean *of $Z = \varphi(X,Y)$ is given by*

$$E\{\varphi(X,Y)\} = \int_{\mathbb{R}^2} \varphi(x,y)\, f(x,y)\, dxdy,$$

provided that the integral is absolutely convergent.

It is easily proved that if (X,Y) is a 2-dimensional random variable, and $\varphi(x,y) = \varphi_1(x) + \varphi_2(y)$, then

$$E\{\varphi_1(X) + \varphi_2(Y)\} = E\{\varphi_1(X)\} + E\{\varphi_2(Y)\}\,,$$

provided that $E\{\varphi_1(X)\}$ and $E\{\varphi_2(Y)\}$ exists. In particular,

$$E\{X+Y\} = E\{X\} + E\{Y\}.$$

If we furthermore assume that X and Y are independent and choose $\varphi(x,y) = \varphi_1(x) \cdot \varphi_2(y)$, then also

$$E\{\varphi_1(X) \cdot \varphi_2(Y)\} = E\{\varphi_1(X)\} \cdot E\{\varphi_2(Y)\}\,,$$

provided that $E\{\varphi_1(X)\}$ and $E\{\varphi_2(Y)\}$ exists. In particular we get under the assumptions above that

$$E\{X \cdot Y\} = E\{X\} \cdot E\{Y\},$$

and

$$E\{(X - E\{X\}) \cdot (Y - E\{Y\})\} = 0.$$

These formulæ are easily generalized to n random variables. We have e.g.

$$E\left\{\sum_{i=1}^n X_i\right\} = \sum_{i=1}^n E\{X_i\}\,,$$

provided that all means $E\{X_i\}$ exist.

If two random variables X and Y are not independent, we shall find a measure of how much they "depend" on each other. This measure is described by the *correlation*, which we now introduce.

Consider a 2-dimensional random variable (X,Y), where

$$E\{X\} = \mu_X, \qquad E\{Y\} = \mu_Y, \qquad V\{X\} = \sigma_X^2 > 0, \qquad V\{Y\} = \sigma_Y^2 > 0,$$

all exist. We define the *covariance* between X and Y, denoted by $\text{Cov}(X,Y)$, as

$$\text{Cov}(X,Y) := E\left\{(X-\mu_X)\cdot(Y-\mu_Y)\right\}.$$

We define the *correlation* between X and Y, denoted by $\varrho(X,Y)$, as

$$\varrho(X,Y) := \frac{\text{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}.$$

Theorem 1.10 *Let X and Y be two random variables, where*

$$E\{X\}=\mu_X, \qquad E\{Y\}=\mu_Y, \qquad V\{X\}=\sigma_X^2>0, \qquad V\{Y\}=\sigma_Y^2>0,$$

all exist. Then

$$\text{Cov}(X,Y)=0, \qquad \textit{if } X \textit{ and } Y \textit{ are independent},$$

$$\text{Cov}(X,Y)=E\{X\cdot Y\}-E\{X\}\cdot E\{Y\},$$

$$|\text{Cov}(X,Y)|\le\sigma_X\cdot\sigma_y,$$

$$\text{Cov}(X,Y)=\text{Cov}(Y,X),$$

$$V\{X+Y\}=V\{X\}+V\{Y\}+2\text{Cov}(X,Y),$$

$$V\{X+Y\}=V\{X\}+V\{Y\}, \qquad \textit{if } X \textit{ and } Y \textit{ are independent},$$

$$\varrho(X,Y)=0, \qquad \textit{if } X \textit{ and } Y \textit{ are independent},$$

$$\varrho(X,X)=1, \qquad \varrho(X,-X)=-1, \qquad |\varrho(X,Y)|\le 1.$$

Let Z be another random variable, for which the mean and the variance both exist- Then

$$\text{Cov}(aX+bY,Z)=a\,\text{Cov}(X,Z)+b\,\text{Cov}(Y,Z), \qquad \textit{for every } a,\ b\in\mathbb{R},$$

and if $U=aX+b$ and $V=cY+d$, where $a>0$ and $c>0$, then

$$\varrho(U,V)=\varrho(aX+b,cY+d)=\varrho(X,Y).$$

Two independent random variables are always non-correlated, while two non-correlated random variables are not necessarily independent.

By the obvious generalization,

$$V\left\{\sum_{i=1}^n X_i\right\}=\sum_{i=1}^n V\left\{X_i\right\}+2\sum_{j=2}^n\sum_{i=1}^{j-1}\text{Cov}\left(X_i,X_j\right).$$

If all X_1, X_2, $\dots$, X_n are independent of each other, this is of course reduced to

$$V\left\{\sum_{i=1}^n X_i\right\}=\sum_{i=1}^n V\left\{X_i\right\}.$$

Finally we mention the various types of convergence which are natural in connection with sequences of random variables. We consider a sequence X_n of random variables, defined on the same probability field $(\Omega,\mathcal{F},P)$.

1) We say that X_n *converges in probability* towards a random variable X on the probability field $(\Omega, \mathcal{F}, P)$, if

$$P\{|X_n - X| \geq \varepsilon\} \to 0 \qquad \text{for } n \to +\infty,$$

for every fixed $\varepsilon > 0$.

2) We say that X_n *converges in probability* towards a constant c, if every fixed $\varepsilon > 0$,

$$P\{|X_n - c| \geq \varepsilon\} \to 0 \qquad \text{for } n \to +\infty.$$

3) If each X_n has the distribution function F_n, and X has the distribution function F, we say that the sequence X_n of random variables *converges in distribution* towards X, if at every point of continuity x of $F(x)$,

$$\lim_{n \to +\infty} F_n(x) = F(x).$$

Finally, we mention the following theorems which are connected with these concepts of convergence. The first one resembles *Čebyšev's inequality.*

Theorem 1.11 (The weak law of large numbers). *Let X_n be a sequence of* independent *random variables, all defined on $(\Omega, \mathcal{F}, P)$, and assume that they all have the same mean and variance,*

$$E\{X_i\} = \mu \qquad \text{and} \qquad V\{X_i\} = \sigma^2.$$

Then for every fixed $\varepsilon > 0$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right\} \to 0 \qquad \text{for } n \to +\infty.$$

A slightly different version of the weak law of large numbers is the following

Theorem 1.12 *If X_n is a sequence of* independent identical distributed *random variables, defined on $(\Omega, \mathcal{F}, P)$ where $E\{X_i\} = \mu$, (notice that we do not assume the existence of the variance), then for every fixed $\varepsilon > 0$,*

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right\} \to 0 \qquad \text{for } n \to +\infty.$$

We have concerning convergence in distribution,

Theorem 1.13 (Helly-Bray's lemma). *Assume that the sequence X_n of random variables converges in distribution towards the random variable X, and assume that there are real constants a and b, such that*

$$P\{a \leq X_n \leq b\} = 1 \qquad \text{for every } n \in \mathbb{N}.$$

If φ is a continuous function on the interval $[a, b]$, then

$$\lim_{n \to +\infty} E\{\varphi(X_n)\} = E\{\varphi(X)\}.$$

In particular,

$$\lim_{n \to +\infty} E\{X_n\} \qquad \text{and} \qquad \lim_{n \to +\infty} V\{X_n\} = V\{X\}.$$

Finally, the following theorem gives us the relationship between the two concepts of convergence:

Theorem 1.14 *1) If X_n converges in probability towards X, then X_n also converges in distribution towards X.*

2) If X_n converges in distribution towards a constant c, then X_n also converges in probability towards the constant c.

2 Maximum and minimum of random variables

Example 2.1 *Lad X_1, X_2 and X_3 be independent random variables of the same distribution function $F(x)$ and frequency $f(x)$, $x \in \mathbb{R}$. The random variables X_1, X_2 and X_3 are ordered according to size, such that we get three new random variables $X_1^\star$, $X_2^\star$ and $X_3^\star$, satisfying $X_1^\star < X_2^\star < X_3^\star$, and defined by*

$X_1^\star =$ *the smallest of* X_1, X_2 *and* X_3 $(= \min\{X_1, X_2, X_3\})$,

$X_2^\star =$ *the second smallest of* X_1, X_2 *and* X_3,

$X_3^\star =$ *the largest of* X_1, X_2 *and* X_3 $(= \max\{X_1, X_2, X_3\})$.

1. *Find, expressed by $F(x)$ and $f(x)$, the distribution functions and the frequencies of the random variables $X_1^\star$ and $X_3^\star$.*

2. *Prove that $X_2^\star$ has the distribution function $F_2^\star(x)$ given by*

$$F_2^\star(x) = 3\{F(x)\}^2\{1 - F(x)\} + \{F(x)\}^3, \qquad x \in \mathbb{R},$$

and find the frequency $f_2^\star(x)$ of $X_2^\star$.

We assume in the following that X_1, X_2 and X_3 are independent and rectangularly distributed over the interval $]0, a[$ (where $a > 0$).

3. *Compute the frequencies of $X_1^\star$, $X_2^\star$ and $X_3^\star$.*

4. *Prove that the three random variables $X_2^\star$, $\frac{1}{3}(X_1 + X_2 + X_3)$ and $\frac{1}{2}(X_1^\star + X_3^\star)$ all have the same mean, and find this mean.*

5. *Which one of the two random variables $X_2^\star$ and $\frac{1}{3}(X_1 + X_2 + X_3)$ has the smallest variance?*

1) It is easily seen that

$$F_3^\star(x) = P\{X_1 \le x \wedge X_2 \le x \wedge X_3 \le x\} = \{F(x)\}^3.$$

Then by a differentiation,

$$f_3^\star = 3\{F(x)\}^2 f(x).$$

Analogously,

$$F_1^\star = 1 - \{1 - F(x)\}^3.$$

By a differentiation we get

$$f_1^\star(x) = 3\{1 - F(x)\}^2 f(x).$$

2) An identification of the various possibilities then gives

$$\begin{aligned} F_2^\star(x) &= P\{X_2^\star \le x\} \\ &= \left.\begin{array}{l} P\{X_1 > x \wedge X_2 \le x \wedge X_3 \le x\} \\ +P\{X_1 \le x \wedge X_2 > x \wedge X_3 \le x\} \\ +P\{X_1 \le x \wedge X_2 \le x \wedge X_3 > x\} \end{array}\right\} \begin{array}{l}\text{two of the variables are } \le x, \\ \text{and the remaining one is } > x,\end{array} \\ &\quad\ +P\{X_1 \le x \wedge X_2 \le x \wedge X_3 \le x\} \qquad \text{All variables are } \le x, \\ &= 3\,F(x)^2\{1-F(x)\} + \{F(x)\}^3 = 3\,F(x)^2 - 2\,F(x)^3. \end{aligned}$$

By a differentiation we obtain the frequency

$$f_2^\star = 6\left\{F(x) - F(x)^2\right\} f(x) = 6\,F(x)\,\{1-F(x)\}\,f(x).$$

3) When X_1, X_2 and X_3 are rectangularly distributed over $]0,a[$, then

$$f(x) = \begin{cases} \dfrac{1}{a} & \text{for } x \in\,]0,a[, \\ \\ 0 & \text{otherwise,} \end{cases}$$

and

$$F(x) = \begin{cases} 0 & \text{for } x \le 0, \\ \\ \dfrac{x}{a} & \text{for } x \in\,]0,a[, \\ \\ 1 & \text{for } x \ge a. \end{cases}$$

By insertion we get for $x \in\,]0,a[$,

$$\begin{aligned} f_1^\star(x) &= 3\{1-F(x)\}^2 f(x) = \frac{3}{a}\left\{1-\frac{x}{a}\right\}^2 = \frac{3}{a^3}\,(a-x)^2, \\ f_2^\star(x) &= \frac{6}{a}\cdot\frac{x}{a}\left\{1-\frac{x}{a}\right\} = \frac{6}{a^3}\,x(a-x) = \frac{6}{a^3}\,\left(ax-x^2\right), \\ f_3^\star(x) &= \frac{3}{a}\left\{\frac{x}{a}\right\}^2 = \frac{3x^2}{a^2}. \end{aligned}$$

All frequencies are 0 for $x \notin\,]0,a[$.

4) The mean of $X_2^\star$ is

$$E\{X_2^\star\} = \frac{6}{a^3}\int_0^a \left(ax^2 - x^3\right\}\,dx = \frac{6}{a^3}\left(\frac{a^4}{3} - \frac{a^4}{4}\right) = \frac{a}{2}.$$

The mean of $\dfrac{1}{3}\,(X_1+X_2+X_3)$ is

$$E\left\{\frac{1}{3}\,(X_1+X_2+X_3)\right) = \frac{1}{3}\cdot 3\,E\{X_1\} = \frac{a}{2}.$$

Since $X_1^\star + X_2^\star + X_3^\star = X_1 + X_2 + X_3$, we get

$$\frac{1}{2}\left(X_1^\star + X_3^\star\right) = \frac{3}{2}\left\{\frac{1}{3}\left(X_1 + X_2 + X_3\right)\right\} - \frac{1}{2}\,X_2^\star,$$

hence

$$E\left\{\frac{1}{2}\left(X_1^\star + X_3^\star\right)\right\} = \frac{3}{2}\,E\left\{\frac{1}{3}\left(X_1 + X_2 + X_3\right)\right\} - \frac{1}{2}\,E\left\{X_2^\star\right\} = \frac{3}{2}\cdot\frac{a}{2} - \frac{1}{2}\cdot\frac{a}{2} = \frac{a}{2},$$

and the three means are all equal to $\dfrac{a}{2}$.

5) It is well-known that

$$V\left\{\frac{1}{3}\left(X_1 + X_2 + X_3\right)\right\} = \frac{1}{9}\left(V\left\{X_1\right\} + V\left\{X_2\right\} + V\left\{X_3\right\}\right) = \frac{1}{3}\,V\left\{X_1\right\} = \frac{1}{3}\cdot\frac{a^2}{12} = \frac{a^2}{36}.$$

Since

$$E\left\{(X_2^\star)^2\right\} = \frac{6}{a^3}\int_0^a \left(ax^3 - x^4\right)\,dx = \frac{6}{a^3}\left(\frac{a^5}{4} - \frac{a^5}{5}\right) = \frac{6}{20}\,a^2,$$

we obtain

$$V\left\{X_2^\star\right\} = E\left\{(X_2^\star)^2\right\} - \left(E\left\{X_2^\star\right\}\right)^2 = \frac{6}{20}\,a^2 - \frac{1}{4}\,a^2 = \frac{a^2}{20}.$$

It follows that the mean $\frac{1}{3}\,(X_1 + X_2 + X_3)$ has the smallest variance.

Example 2.2 *Let X_1, X_2, X_3 and X_4 be independent random variables of the same distribution function $F(x)$ and frequency $f(x)$, $x \in \mathbb{R}$, and let the random variables Y and Z be defined by*

$$Y = \min\left\{X_1, X_2, X_3, X_4\right\}, \qquad Z = \max\left\{X_1, X_2, X_3, X_4\right\}.$$

1. *Find, expressed by $F(x)$ and $f(x)$, the distribution functions and the frequencies of the random variables Y and Z.*

2. *Prove that the simultaneous frequency of (Y, Z) is given by*

$$g(y,z) = \begin{cases} 12\,f(y)\cdot f(z)\cdot\{F(z) - F(y)\}^2, & y \le z, \\ 0, & y > z, \end{cases}$$

HINT: *Start by finding $P\{Y > y \,\wedge\, Z \le z\}$ for $y \le z$.*

We assume in the following that

$$f(x) = \begin{cases} 1, & x \in]0,1[, \\ 0, & \textit{otherwise}. \end{cases}$$

3. *Find the frequencies of Y and Z, and the simultaneous frequency of (Y, Z).*

4. *Find the means $E\{Y\}$ and $E\{Z\}$.*

5. *Find the variances $V\{Y\}$ and $V\{Z\}$.*

We now introduce the width of the variation U *by $U = Z - Y$.*

6. *Find the mean $E\{U\}$.*

7. *Find the variance $V\{U\}$.*

1) We see that

$$F_Z(z) = P\left\{X_1 \le z \,\wedge\, X_2 \le z \,\wedge\, X_3 \le z \,\wedge\, X_4 \le z\right\} = \{F(z)\}^4$$

and

$$F_Y(y) = 1 - \{1 - F(y)\}^4.$$

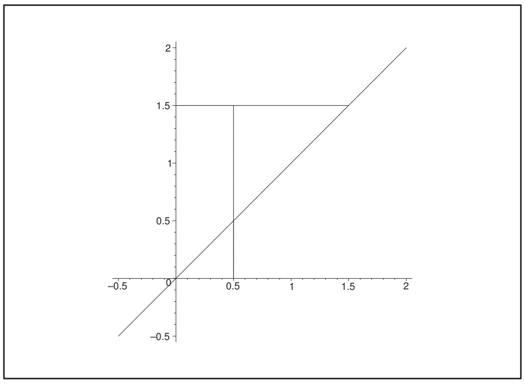


Figure 1: When $y < z$, the domain of integration is the triangle on the figure, where (y, z) are the coordinates of the rectangular corner.

By differentiation we get the frequencies

$$f_Y(y) = 4\{1 - F(y)\}^3 f(y)$$

and

$$f_Z(z) = 4\{F(z)\}^3 f(z).$$

2) By definition, $Y \le Z$, so clearly $g(y, z) = 0$ for $y > z$. If $y \le z$, then

$$\begin{aligned} P\{Y > y \wedge Z \le z\} &= P\{y < X_1 \le z \wedge y < X_2 \le z \wedge y < X_3 \le z \wedge y < X_4 \le z\} \\ &= P\{y < X_1 \le z\} \cdot P\{y < X_2 \le z\} \cdot P\{y < X_4 \le z\} \\ &= \{F(z) - F(y)\}^4, \end{aligned}$$

hence the distribution function of (Y, Z) is for $y \le z$ given by

$$F(y,z) = P\{Y \le y \wedge Z \le z\} = P\{Z \le z\} - P\{Y > y \wedge Z \le z\} = P\{Z \le z\} - \{F(z) - F(y)\}^4.$$

Then

$$g(y,z) = \frac{\partial^2 G}{\partial y \partial z} = 0 - \frac{\partial}{\partial z}\left\{-4(F(z) - F(y))^3 f(y)\right\} = 12\, f(y) \cdot f(z) \cdot \{F(z) - F(y)\}^2,$$

and the claim is proved.

3) Since $F(x) = x$ for $x \in]0,1[$, we get for $y,\ z \in]0,1[$ by insertion,

$$f_Y(y) = 4\,(1-y)^3 \qquad \text{and} \qquad f_Z(z) = 4z^3.$$

and $f_Y(y) = 0$ for $y \notin]0,1[$, and $f_Z(z) = 0$ for $z \notin]0,1[$.

When $0 < y < z < 1$, we get the simultaneous frequency

$$g(y,z) = 12 \cdot 1 \cdot 1 \cdot (z-y)^2 = 12\,(z-y)^2,$$

and $g(y, z) = 0$ otherwise.

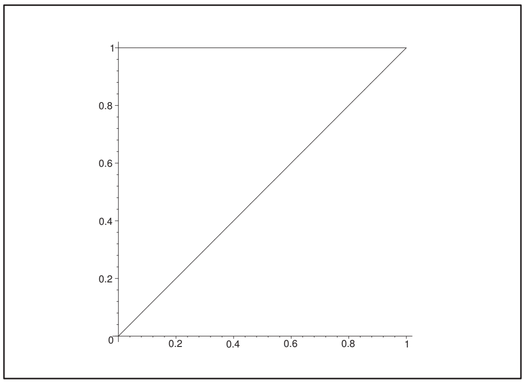


Figure 2: The domain D.

4) The means are given by

$$E\{Y\} = 4 \int_0^1 y(1-y)^3 \, dy = 4 \int_0^1 \left\{(1-y)^3 - (1-y)^4\right\} \, dy = 4 \left(\frac{1}{4} - \frac{1}{5} \right) = \frac{4}{20} = \frac{1}{5},$$

and

$$E\{Z\} = 4 \int_0^1 z^4 \, dz = \frac{4}{5}.$$

5) We first compute

$$\begin{aligned} E\left\{Y^2\right\} &= 4 \int_0^1 y^2(1-y)^3 <, dy = 4 \left[-\frac{1}{4} \, y^2 (1-y)^4 \right]_0^1 + 2 \int_0^1 y(1-y)^4 \, dy \\ &= 0 + 2 \left[-\frac{1}{5} \, y(1-y)^5 \right]_0^1 + \frac{2}{5} \int_0^1 (1-y)^5 \, dy = 0 + \frac{2}{5 \cdot 6} = \frac{1}{15}. \end{aligned}$$

The variance is

$$V\{Y\} = \frac{1}{15} - \left(\frac{1}{5} \right)^2 = \frac{1}{5} \left(\frac{1}{3} - \frac{1}{5} \right) = \frac{2}{75}.$$

From

$$E\left\{Z^2\right\} = 4 \int_0^1 z^5 \, dz = \frac{4}{6} = \frac{2}{3}.$$

follows that

$$V\{Z\} = \frac{2}{3} - \left(\frac{4}{5} \right)^2 = \frac{2}{3} - \frac{16}{25} = \frac{50-48}{75} = \frac{2}{75}.$$

6) The mean is of course

$$E\{U\} = E\{Z-Y\} = E\{Z\} - E\{Y\} = \frac{4}{5} - \frac{1}{5} = \frac{3}{5}.$$

7) Finally,

$$E\left\{U^2\right\} = E\left\{Z^2\right\} - 2E\{ZY\} + E\left\{Y^2\right\} = \frac{2}{3} + \frac{1}{15} - 2\,E\{ZY\},$$

where

$$\begin{aligned} E\{ZY\} &= \int\!\!\int_D yz\,g(y,z)\,dy\,dz = 12\int\!\!\int_D yz(z-y)^2\,dy\,dz = 12\int_0^1 z\left\{\int_0^z y(y-z)^2\,dy\right\}dz \\ &= 12\int_0^1 z\left\{\left[\frac{1}{3}\,y\cdot(y-z)^3\right]_0^z - \frac{1}{3}\int_0^z (y-z)^3\,dy\right\}dz \\ &= -4\int_0^1 z\left[\frac{1}{4}\,(y-z)^4\right]_0^z dz = \int_0^1 z^5\,dz = \frac{1}{6}, \end{aligned}$$

which gives by insertion

$$E\left\{U^2\right\} = \frac{2}{3} + \frac{1}{15} - \frac{1}{3} = \frac{1}{3} + \frac{1}{15} = \frac{6}{16} = \frac{2}{5}.$$

The variance is

$$V\{U\} = E\left\{U^2\right\} - (E\{U\})^2 = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{2}{5} - \frac{9}{25} = \frac{1}{25}.$$

Example 2.3 *Let X_1 and X_2 be independent, identically distributed random variables of frequency*

$$f(x) = \begin{cases} \dfrac{2x}{a^2}, & 0 < x < a, \\ 0, & \textit{otherwise}, \end{cases}$$

where a is a positive constant, and let the random variables Y and Z be given by

$$Y = \max\{X_1, X_2\}, \qquad Z = \min\{X_1, X_2\}.$$

1. *Compute the mean and the variance of X_1.*

2. *Find the frequency and the mean of Y.*

3. *Find the frequency and the mean of Z.*

4. *Prove that the simultaneous frequency of (Y, Z) is given by*

$$g(y, z) = \begin{cases} \dfrac{8yz}{a^4}, & 0 < z < y < a, \\ 0, & \textit{otherwise}. \end{cases}$$

HINT: *Start by computing $P\{Y \le y \wedge Z > z\}$ for $z < y$.*

We introduce the width of the variation U *by $U = Y - Z$.*

5. *Find the mean of U.*

6. *Find the frequency of U.*

1) By the usual computations,

$$E\{X_1\} = \int_0^a x \cdot \frac{2x}{a^2}\,dx = \frac{2}{3}\,a,$$

and

$$E\left\{X_1^2\right\} = \int_0^a x^2 \cdot \frac{2x}{a^2}\,dx = \frac{1}{2}\,a^2,$$

hence

$$V\{X_1\} = E\left\{X_1^2\right\} - (E\{X_1\})^2 = \left(\frac{1}{2} - \frac{4}{9}0\right) a^2 = \frac{1}{18}\,a^2.$$

2) Let $F(x)$ $\left[= \dfrac{x^2}{a^2} \text{ for } 0 < x < a\right]$ be the distribution function of X_1 and X_2. Then the distribution function of Y is in the interval $]0, a[$ given by

$$F_Y(y) = \{F(y)\}^2 = \frac{y^4}{a^4},$$

so the corresponding frequency is

$$f_Y(y) = \begin{cases} 4\,\dfrac{y^3}{a^4} & \text{for } 0 < y < a, \\ \\ 0 & \text{otherwise.} \end{cases}$$

The mean is

$$E\{Y\} = \int_0^a \frac{4y^4}{a^4}\,dy = \frac{4}{5}\,a.$$

3) Analogously, the distribution function of Z for $0 < z < a$ is given by

$$F_Z(z) = 1 - \{1 - F(z)\}^2 = 1 - \left(1 - \frac{z^2}{a^2}\right)^2 = \frac{1}{a^4}\left(2a^2z^2 - z^4\right).$$

We get the frequency by a differentiation,

$$f_Z(z) = \begin{cases} \dfrac{4}{a^4}\left\{a^2 z - z^3\right\} & \text{for } 0 < z < a, \\ \\ 0 & \text{otherwise.} \end{cases}$$

The mean is

$$E\{Z\} = \frac{4}{a^2}\int_0^a \left\{a^2z^2 - z^4\right\}\,dz = \frac{4}{a^4}\left(\frac{1}{3} - \frac{1}{5}\right)a^5 = \frac{8}{15}\,a.$$

4) It follows from the definitions of Y and Z that $g(y,z) = 0$, whenever we do *not* have $0 < z < y < a$. On the other hand, if these inequalities are fulfilled, then it follows, since X_1 and X_2 are independent that

$$\begin{aligned} P\{Y \le y \wedge Z > z\} &= P\{z < X_1 \le y \wedge z < X_2 \le y\} = P\{z < X_1 \le y\} \cdot P\{z < X_2 \le y\} \\ &= \{F(y) - F(z)\}^2 = \frac{1}{a^4}\left(y^2 - z^2\right)^2. \end{aligned}$$

Therefore, if $0 < z < y < a$, then the simultaneous distribution function is given by

$$G(y,z) = P\{Y \le y \wedge Z \le z\} = P\{Y \le y\} - P\{Y \le y \wedge Z > z\} = F_Y(y) - \frac{1}{a^4}\left(y^2 - z^2\right)^2,$$

hence

$$\frac{\partial G}{\partial z} = 0 - \frac{2}{a^4}\left(y^2 - z^2\right)\cdot(-2z) = \frac{4z}{a^4}\left(y^2 - z^2\right),$$

and

$$g(y,z) = \frac{\partial^2 G}{\partial y \partial z} = \frac{8yz}{a^4} \qquad 0 < z < y < a,$$

and $g(y,z) = 0$ otherwise.

5) The mean is of course

$$E\{U\} = E\{Y - Z\} = E\{Y\} - E\{Z\} = \frac{4}{5}\,a - \frac{8}{15}\,a = \frac{4}{15}\,a.$$

6) The frequency of $U = Y - Z$ is given by

$$f_U(u) = \int_{-\infty}^{\infty} g(y, y-u)\,dy.$$

The integrand is $\neq 0$, when $0 < y - u < y < a$, so we have the conditions

$$0 < y < a \qquad \text{and} \qquad 0 < u < y < a.$$

If $u \in]0,1[$, then the domain of integration is $u < y < a$, hence

$$\begin{aligned} f_U(u) &= \int_u^a \frac{8y}{a^4}\,(y-u)\,dy = \frac{8}{a^4}\int_u^a (yr - yu)\,dy = \frac{8}{a^4}\left[\frac{1}{3}\,y^3 - \frac{u}{2}\,y^2\right]_u^a \\ &= \frac{8}{a^4}\left\{\frac{a^3}{3} - \frac{a^2}{2}\,u - \frac{1}{3}\,u^3 + \frac{1}{2}\,u^3\right\} = \frac{8}{a^4}\left\{\frac{a^3}{3} - \frac{a^2}{2}\,u + \frac{1}{6}\,u^3\right\}, \end{aligned}$$

and $f_U(u) = 0$ otherwise.

A WEAK CHECK:

$$\int_0^a f_U(u)\,du = \frac{8}{a^4}\left\{\frac{a^3}{3}\cdot a - \frac{a^2}{4}\cdot a^2 + \frac{1}{24}\,a^4\right\} = 8\left(\frac{1}{3} - \frac{1}{4} + \frac{1}{24}\right) = \frac{8}{24}\,(8 - 6 + 1) = 1.$$

Example 2.4 *An instrument contains two components, the lifetimes of which T_1 and T_2 are independent random variables, both of the frequency*

$$f(t) = \begin{cases} a\,e^{-at}, & t > 0, \\ 0, & t \le 0, \end{cases}$$

where a is a positive constant.
We introduce the random variables X_1, X_2 and Y_2 by

$$X_1 = \min\{T_1, T_2\}, \qquad X_2 = \max\{T_1, T_2\}, \qquad Y_2 = X_2 - X_1.$$

Here, X_1 denotes the time until the first of the components fails, and X_2 the time, until the second component also fails, and Y_2 is the time from the first component fails to the second one fails.

1. *Find the frequency and the mean of X_1.*

2. *Find the frequency and the mean of X_2.*

3. *Find the mean of Y_2.*

The simultaneous frequency of (X_1, X_2) is given by

$$h(x_1, x_2) = \begin{cases} 2a^2 e^{-a(x_1+x_2)}, & 0 < x_1 < x_2, \\ 0, & \textit{otherwise.} \end{cases}$$

(One shall not prove this statement.)

4. *Find the simultaneous frequency of the 2-dimensional random variable (X_1, Y_2).*

5. *Find the frequency of Y_2.*

6. *Check if the random variables X_1 and Y_2 are independent.*

1) Concerning X_1,

$$P\{X_1 > x_1\} = P\{T_1 > x_1 \wedge T_2 > x_1\} = P\{T_1 > x_1\} \cdot P\{T_2 > x_2\} = e^{-2ax_1},$$

thus

$$P\{X_1 \le x_1\} = 1 - e^{-2ax_1}, \qquad x_1 > 0,$$

and X_1 is exponentially distributed of the frequency

$$f_{X_1} = \begin{cases} 2a\,e^{-2ax_1}, & x_1 > 0, \\ 0, & x_1 \le 0, \end{cases} \qquad \text{and mean } \frac{1}{2a}.$$

2) Concerning X_2,

$$\begin{aligned} P\{X_2 \le x_2\} &= P\{T_1 \le x_2 \wedge T_2 \le x_2\} = P\{T_1 \le x_2\} \cdot P\{T_2 \le x_2\} \\ &= \left(1 - e^{-ax_2}\right)^2, \qquad x_2 > 0, \end{aligned}$$

thus X_2 has the frequency

$$f_{X_2}(x_2) = 2a\,e^{-ax_2}\left(1 - e^{-ax_2}\right) = 2a\,e^{-ax_2} - 2a\,e^{-2ax_2} \qquad \text{for } x_2 > 0,$$

and

$$f_{X_2}(x_2) = 0 \qquad \text{for } x_2 \le 0.$$

The mean is

$$E\{X_2\} = \int_0^\infty x_2 f_{X_2}(x_2)\,dx_2 = \int_0^\infty \left\{2a\,x_2 e^{-ax_2} - 2a\,x_2 e^{-2ax_2}\right\} dx_2 = \frac{2}{a} - \frac{1}{2a} = \frac{3}{2a}.$$

Additional. The mean of X_2 is easily obtained from $X_1 + X_2 = T_1 + T_2$, i.e.

$$E\{X_2\} = E\{T_1\} + E\{T_2\} - E\{X_1\} = \frac{1}{a} + \frac{1}{a} - \frac{1}{2a} = \frac{3}{2a}.$$

3) This is trivial, because

$$E\{Y_2\} = E\{X_2\} - E\{X_1\} = \frac{3}{2a} - \frac{1}{2a} = \frac{1}{a}.$$

4) The simultaneous frequency $k(y_1, y_2)$ of

$$(Y_1, Y_2) = (X_1, X_2 - X_1)$$

can e.g. be obtained directly by using a formula, where $a = 1$, $b = 0$, $c = -1$ and $d = -1$,

$$\begin{aligned} k(y_1, y_2) &= h\left(\frac{dy_1 - by_2}{ad - bc}, \frac{-cy_1 + ay_2}{ad - bc}\right) \cdot \frac{1}{|ad - bc|} \\ &= h(y_1, y_1 + y_2) = 2a^2 e^{-a(2y_1 + y_2)} \qquad \text{for } y_1 > 0 \text{ and } y_2 > 0, \end{aligned}$$

and

$$k(y_1, y_2) = 0 \qquad \text{otherwise.}$$

This is also written

$$k(y_1, y_2) = \begin{cases} 2a\,e^{-2ay_1} \cdot a\,e^{-ay_2}, & \text{for } y_1 > 0 \text{ and } y_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

5) (and 6.) It follows immediately from **4.** that Y_1 $(= X_1)$ and Y_2 are independent, and that Y_2 has the frequency

$$k_{Y_2}(y_2) = \begin{cases} a\,e^{-ay_2}, & y_2 > 0, \\ 0, & y_2 \le 0. \end{cases}$$

Example 2.5 *An instrument A contains two components, the lifetimes of which X_1 and X_2 are independent random variables, both of the frequency*

$$f(x) = \begin{cases} a\,e^{-ax}, & x > 0, \\ \\ 0, & x \le 0, \end{cases}$$

where a is a positive constant.
The instrumentet A works as long as at least one of the two components is working, thus the lifetime X of A is

$$X = \max\{X_1, X_2\}\,.$$

Another instrument B has the lifetime Y of the frequency

$$g(y) = \begin{cases} a\,e^{-ay}, & y > 0, \\ \\ 0, & y \le 0. \end{cases}$$

1) Find the distribution function and the frequency of the random variable X.

2) Find the mean of X.

3) Find the simultaneous frequency of (X, Y), and find $P\{Y > X\}$.

4) Find the frequency of $X + Y$, and find the mean of $X + Y$.

1) Since X_1 and X_2 have the frequency

$$f(x) = a\,e^{-ax}, \qquad \text{for } x > 0,$$

the distribution function of each of them is

$$F(x) = 1 - e^{-ax}, \qquad \text{for } x > 0.$$

Then by a formula, $X = \max\{X_1, X_2\}$ has the frequency

$$F_X(x) = F_{X_1}(x) \cdot F_{X_2}(x) = \left\{1 - e^{-ax}\right\}^2 \qquad \text{for } x > 0,$$

hence the frequency for $x > 0$ is given by

$$f_X(x) = F_X'(x) = 2\left(1 - e^{-ax}\right) a\,e^{-ax} = 2a\,e^{-ax} - 2a\,e^{-2ax}.$$

2) The mean is

$$\begin{aligned} E\{X\} &= \int_0^\infty x\,f_X(x)\,dx = 2a\int_0^\infty x\,e^{-ax}\,dx - 2a\int_0^\infty x\,e^{-2ax}\,dx \\ &= 2a\left(\frac{1}{a^2} - \frac{1}{4a^2}\right) = 2a \cdot \frac{3}{4a^2} = \frac{3}{2a}. \end{aligned}$$

3) In the first quadrant the simultaneous frequency is given by

$$f_X(x)\,g_Y(y) = 2a\left(e^{-ax} - e^{-2ax}\right)\cdot a\,e^{-ay},$$

hence

$$\begin{aligned} P\{Y > X\} &= \int_{x=0}^{\infty} 2a\left(e^{-ax} - e^{-2ax}\right)\left\{\int_{y=x}^{\infty} a\,e^{-ay}\,dy\right\}dx = \int_0^{\infty} 2a\left(e^{-ax} - e^{-2ax}\right)e^{-ax}\,dx \\ &= \int_0^{\infty} 2a\left(e^{-2ax} - e^{-3ax}\right)\,dx = 2a\left(\frac{1}{2a} - \frac{1}{3a}\right) = \frac{1}{3}. \end{aligned}$$

4) The mean of $X + Y$ is of course

$$E\{X + Y\} = E\{X\} + E\{Y\} = \frac{3}{2a} + \frac{1}{a} = \frac{5}{2a}.$$

When $z > 0$, the frequency of $X + Y$ is given by

$$\begin{aligned} h(z) &= \int_0^z f_X(x)\,g_Y(z-x)\,dx \\ &= \int_0^z 2a\left(e^{-ax} - e^{-2ax}\right)a\,e^{-a(z-x)}\,dx = 2a^2\int_0^z \left(e^{-az} - e^{-ax}e^{-az}\right)\,dx \\ &= 2a^2e^{-az}\int_0^z \left(1 - e^{-ax}\right)\,dx = 2a^2e^{-az}\left\{z - \frac{1}{a}\left(1 - e^{-az}\right)\right\} \\ &= 2a^2z\,e^{-az} - 2a\,e^{-az} + 2a\,e^{-2az} = 2a\,e^{-az}\left(az - 1 + e^{-az}\right). \end{aligned}$$

3 The transformation formula and the Jacobian

Example 3.1 *Let* (X_1, X_2) *be a 2-dimensional random variable of the frequency*

$$h(x_1, x_2) = \begin{cases} \dfrac{1}{\pi}, & 0 < x_1^2 + x_2^2 < 1, \\ 0, & \textit{otherwise.} \end{cases}$$

1. *Find the frequencies of the random variables* X_1 *and* X_2.

2. *Find the means and the variances of the random variables* X_1 *and* X_2.

3. *Prove that* X_1 *and* X_2 *are non-correlated, but not independent.*

Let (Y_1, Y_2) *be given by*

$$X_1 = Y_1 \cos Y_2, \qquad X_2 = Y_1 \sin Y_2,$$

where $0 < Y_1 < 1$ *and* $0 \le Y_2 < 2\pi$.

4. *Find the frequency* $k(y_1, y_y)$ *for* (Y_1, Y_2).

Are Y_1 *and* Y_2 *independent?*

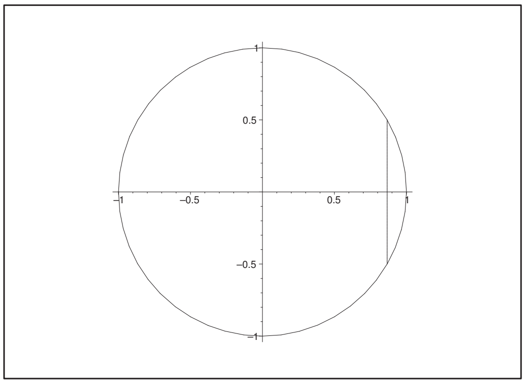


Figure 3: When $-1 < x_1 < 1$, then $-\sqrt{1-x_1^2} < x_2 < \sqrt{1-x_1^2}$.

1) It follows immediately that

$$f_{X_1}(x_1) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x_1^2}, & -1 < x_1 < 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_{X_2}(x_1) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x_2^2}, & -1 < x21 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

2) It follows from the above that

$$E\{X_1\} = E\{X_2\} = \frac{2}{\pi}\int_{-1}^{1} t\sqrt{1-t^2}\,dt = 0,$$

and

$$\begin{aligned} V\{X_1\} = V\{X_2\} &= E\{X_1^2\} = \frac{2}{\pi}\int_{-1}^{1} t^2\sqrt{1-t^2}\,dt = \frac{4}{\pi}\int_{0}^{1} t^2\sqrt{1-t^2}\,dt \\ &= \frac{4}{\pi}\int_0^{\frac{\pi}{2}} \sin^2 t\cdot\cos t\cdot\cos t\,dt = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \sin^2 2t\,dt = \frac{1}{4}. \end{aligned}$$

3) The support of the frequency is not a rectangle parallel to the axes. Hence, X_1 and X_2 cannot be independent.
It follows from the symmetry that $E\{X_1X_2\} = 0$. Hence

$$\operatorname{Cov}(X_1, X_2) = E\{X_1X_2\} - E\{X_1\}\,E\{X_2\} = 0,$$

and X_1 and X_2 are non-correlated.

4) The map

$$(x_1, x_2) = \varphi(y_1, y_2) = (y_1\cos y_2, y_1\sin y_2)$$

is bijective between the two given domains.
The Jacobian is

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} \dfrac{ddx_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \cos y_2 & -y_1\sin y_2 \\ \sin y_2 & y_1\cos y_2 \end{vmatrix} = y_1 \neq 0.$$

Then we get the frequency of (Y_1, Y_2),

$$k(y_1,y_2) = \begin{cases} \dfrac{1}{\pi}\,y_1, & \text{for } y_1 \in\,]0.1[\text{ and } y_2 \in [0.2\pi[, \\ 0 & \text{otherwise.} \end{cases}$$

5) It follows from

$$g_{Y_1}(y_1) = \begin{cases} 2y_1 & \text{for } y \in\,]0,1[, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$g_{Y_2}(y_2) = \begin{cases} \dfrac{1}{2\pi} & \text{for } y_2 \in [0.2\pi[, \\ 0 & \text{otherwise,} \end{cases}$$

that

$$k(y_1,y_2) = g_{Y_1}(y_1)\cdot g_{Y_2}(y_2),$$

hence Y_1 and Y_2 are independent.

Example 3.2 *Let* (X_1, X_2) *have the frequency*

$$h(x_1, x_2) = \begin{cases} e^{-x_1} \cdot \lambda\, e^{-\lambda x_2}, & x_1 > 0,\, x_2 > 0, \\ 0, & \text{otherwise}, \end{cases}$$

where λ *is a positive constant, and let* $(Y_1, Y_2') = \tau(X_1, X_2)$ *be given by*

$$Y_1 = X_1 + X_2, \qquad Y_2 = X_1 - X_2.$$

1) Prove that τ *maps* $]0,\infty[\times]0,\infty[$ *bijectively onto the domain*

$$D' = \left\{ (y_1, y_2) \in \mathbb{R}^2 \mid y_1 > 0,\, |y_2| < y_1 \right\}.$$

2) Find the frequency $k(y_1, y_2)$ *of* (Y_1, Y_2).

3) Prove that Y_1 *and* Y_2 *are non-correlated for precisely one value of* λ*, and find this value.*

4) Prove that Y_1 *and* Y_2 *are not independent for any choice of* λ.

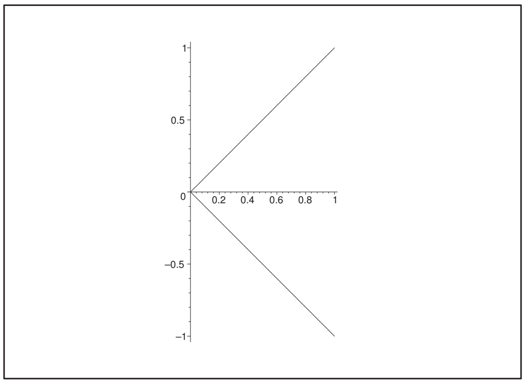


Figure 4: The domain D' is the angular space in the right half plane (and D is the first quadrant).

1) It follows from

$$y_1 = x_1 + x_2, \qquad y_2 = x_1 - x_2,$$

that

$$x_1 = \frac{1}{2}(t_1 + y_2), \qquad x_2 = \frac{1}{2}(y_1 - y_2).$$

Since (x_1, x_2) is uniquely determined (by an explicit expression as a function) from the given (y_1, y_2) and *vice versa*, the map is bijective.

In order to find the image D' of the first quadrant D by the map τ we start by determining the images of the boundary curves:

- The line $x_1 = 0$ is mapped into $y_1 + y_2 = 0$, i.e. into the line $y_2 = -y_1$.
- The line $x_2 = 0$ is mapped into $y_1 - y_2 = 0$, i.e. into the line $y_2 = y_1$.

Since τ is continuous and $y_1 > 0$, it follows from where the boundary curves are lying that the image is

$$D' = \left\{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 > 0,\ |y_2| < y_1\right\},$$

which has been indicated on the figure.

2) The Jacobian is

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Hence, if $(y_1, y_2) \in D'$, then the frequency of (Y_1, Y_2) is given by

$$\begin{aligned} k(y_1, y_2) &= \left|-\frac{1}{2}\right| \cdot h\left(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right) \\ &= \frac{\lambda}{2} \exp\left(-\frac{1}{2}(y_1 + y_2)\right) \cdot \exp\left(-\frac{\lambda}{2}(y_1 - y_2)\right) \\ &= \frac{\lambda}{2} \exp\left(-\frac{\lambda + 1}{2} y_1\right) \cdot \exp\left(\frac{\lambda - 1}{2} y_2\right), \end{aligned}$$

or more well-organized

$$k(y_1, y_2) = \begin{cases} \dfrac{\lambda}{2}\exp\left(-\dfrac{\lambda+1}{2}y_1\right)\cdot\exp\left(\dfrac{\lambda-1}{2}y_2\right), & y_1 > 0,\ |y_2| < y_1, \\ 0, & \text{otherwise.} \end{cases}$$

3) Since X_1 and X_2 are independent, it follows by a reduction that

$$\operatorname{Cov}(Y_1, Y_2) = \operatorname{Cov}(X_1 + X_2, X_1 - X_2) = V\{X_1\} - V\{X_2\}.$$

It follows from

$$V\{X_1\} = \int_0^\infty x_1^2\, e^{-x_1}\, dx_1 - \left\{\int_0^\infty x_1 e^{-x_1}\, dx_1\right\}^2 = 2! - (1!)^2 = 1,$$

and

$$V\{X_2\} = \int_0^\infty x_2^2\,\lambda\, e^{-\lambda x_2}\, dx_2 - \left\{\int_0^\infty x_2\cdot\lambda\, e^{-\lambda x_2}\, dx_2\right\}^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

that $\operatorname{Cov}(Y_1, Y_2) = 0$, precisely when $\lambda > 0$ is equal to $\lambda = 1$, hence Y_1 and Y_2 are non-correlated precisely when $\lambda = 1$.

4) Since D' is not a domain which is parallel to the axes, Y_1 and Y_2 cannot be independent for any choice of $\lambda > 0$.

Example 3.3 *A 2-dimensional random variable* (X, Y) *has the frequency*

$$h(x_1, x_2) = \begin{cases} 1, & 0 < x_1 < \infty,\ 0 < x_2 < e^{-x_1}, \\ 0, & \textit{otherwise.} \end{cases}$$

1. *Find the frequencies of the random variables* X_1 *and* X_2.

2. *Find the means* $E\{X_1\}$ *and* $E\{X_2\}$.

3. *Find the variances* $V\{X_1\}$ *and* $V\{X_2\}$.

4. *Find the correlation coefficient* $\varrho(X_1, X_2)$.

Let the 2-dimensional random variable $(Y_1, Y_2) = \tau(X_1, X_2)$ *be given by*

$$Y_1 = X_2\, e^{X_1}, \qquad Y_2 = e^{-X_1}.$$

5. *Find the frequency of* (Y_1, Y_2).

6. *Are* Y_1 *and* Y_2 *independent?*

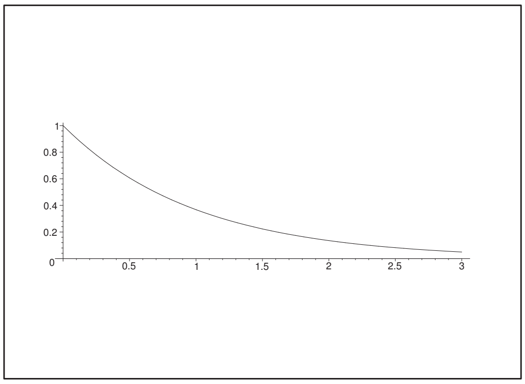


Figure 5: The domain D, where $h(x_1, x_2) > 0$.

1) We get for fixed $x_1 \in \mathbb{R}$ by a vertical integration,

$$f_{X_1}(x_1) = \begin{cases} e^{-x_1} & \text{for } x_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then by a horizontal integration for fixed x_2,

$$f_{X_2}(x_2) = \begin{cases} -\ln x_2 & \text{for } 0 < x_2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

2) The means are $E\{X_1\} = 1$, and

$$E\{X_2\} = -\int_0^1 x_2 \cdot \ln x_2 \, dx_2 = -\left[\frac{1}{2}\, x_2^2 \, \ln x_2\right]_0^1 + \int_0^1 \frac{1}{2}\, x_2 \, dx_2 = \frac{1}{4}.$$

3) The variance of X_1 can be found in a table, $V\{X_1\} = 1$.
Concerning X_2 we first compute

$$E\left\{X_2^2\right\} = -\int_0^1 x_2^2 \, \ln x_2 \, dx_2 = -\left[\frac{1}{3}\, x_2^3 \, \ln x_2\right]_0^1 + \int_0^1 \frac{1}{3}\, x_2^3 \, dx_2 = \frac{1}{9}.$$

The variance is

$$V\{X_2\} = E\left\{X_2^2\right\} - \left(E\{X_2\}\right)^2 = \frac{1}{9} - \frac{1}{16} = \frac{7}{144}.$$

4) It follows from

$$E\{X_1 X_2\} = \int_0^\infty x_1 \left\{ \int_0^{\exp(x_1)} x_2 \, dx_2 \right\} dx_1 = \frac{1}{2} \int_0^\infty x_1 \cdot e^{-2x_1} \, dx_1 = \frac{1}{8},$$

that

$$\operatorname{Cov}(X_1, X_2) = E\{X_1 X_2\} - E\{X_1\}\; E\{X_2\} = \frac{1}{8} - 1 \cdot \frac{1}{4} = -\frac{1}{8},$$

hence

$$\varrho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{V\{X_1\}\; V\{X_2\}}} = \frac{-\frac{1}{8}}{\sqrt{1 \cdot \dfrac{7}{144}}} = -\frac{12}{8\sqrt{7}} = -\frac{3\sqrt{7}}{14}.$$

5) It follows from

$$y_1 = x_2 \, e^{x_1}, \qquad y_2 = e^{-x_1},$$

that

$$x_1 = -\ln y_2 \quad \text{and} \quad x_2 = y_1 y_2.$$

Investigating the boundary we see that

- the curve $x_2 = 0$, $x_1 > 0$ is mapped into $y_1 = 0$ and $0 < y_2 < 1$,
- the curve $x_1 = 0$, $0 < x_2 < 1$, is mapped into $0 < y_1 < 1$ and $y_2 = 1$,
- the curve $x_2 = e^{-x_1}$, $x_1 > 0$ is mapped into $y_1 = 1$ and $0 < y_2 < 1$.

Finally, it follows from y_1, $y_2 > 0$ and $y_1 = x_2 \, e^{x_1} < 1$ that the image is $D' =]0,1[\times]0,1[$.
The Jacobian is

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} 0 & -\dfrac{1}{y_2} \\ y_2 & y_1 \end{vmatrix} = 1.$$

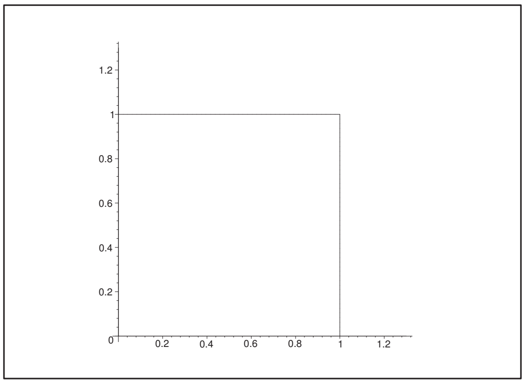


Figure 6: The image D'.

If $(y_1, y_2) \in D'$, then $k(y_1, y_2) = 1$, hence

$$k(y_1, y_2) = \begin{cases} 1 & \text{for } 0 < y_1 < 1,\ 0 < y_2 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

6) Obviously, Y_1 and Y_2 are independent. In fact,

$$k_1(y_1) = \begin{cases} 1 & \text{for } 0 < y_1 < 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$k_2(y_2) = \begin{cases} 1 & \text{for } 0 < y_2 < 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$k(y_1, y_2) = k_1(y_1) \cdot k_2(y_2).$$

Example 3.4 *A 2-dimensional random variable* (X_1, X_2) *has the frequency*

$$h(x_1, x_2) = \begin{cases} 4x_1^2 & i\ D, \\ 0 & otherwise, \end{cases}$$

where

$$D = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_2 < x_1 < 1\right\}.$$

1. *Find the marginal frequencies of* X_1 *and* X_2*.*

2. *Compute the means* $E\{X_1\}$ *and* $E\{X_2\}$*.*

3. *Compute the covariance* $\mathrm{Cov}(X_1, X_2)$*.*

We now define the random variables Y_1 *and* Y_2 *by*

$$(Y_1, Y_2) = \tau(X_1, X_2) = (X_1, X_1 - 2X_2).$$

4. *Prove that the vector function* τ *given by*

$$\tau(x_1, x_2) = (x_1, x_1 - 2x_2)$$

maps D *bijectively onto*

$$D' = \left\{(y_1, y_2) \in \mathbb{R}^2 \mid 0 < y_1 < 1,\ -y_1 < y_2 < y_1\right\}.$$

5. *Find the simultaneous frequency* $k(y_1, y_2)$ *of* (Y_1, Y_2)*.*

6. *Find the marginal frequencies of* Y_1 *and* Y_2*.*

7. *Compute the means* $E\{Y_1\}$ *and* $E\{Y_2\}$*.*

8. *Check if* Y_1 *and* Y_2 *are non-correlated.*

9. *are* Y_1 *and* Y_2 *independent?*

1) It follows by a vertical integration,

$$f_{X_1}(x_1) = \begin{cases} 4x_1^3 & \text{for } 0 < x_1 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then by a horizontal integration for $0 < x_2 < 1$,

$$f_{X_2}(x_2) = \int_{x_2}^{1} 4x_1^2\, dx_1 = \frac{4}{3}\left(1 - x_2^3\right),$$

hence

$$f_{X_2}(x_2) = \begin{cases} \frac{4}{3}\left(1 - x_2^3\right) & \text{for } 0 < x_2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

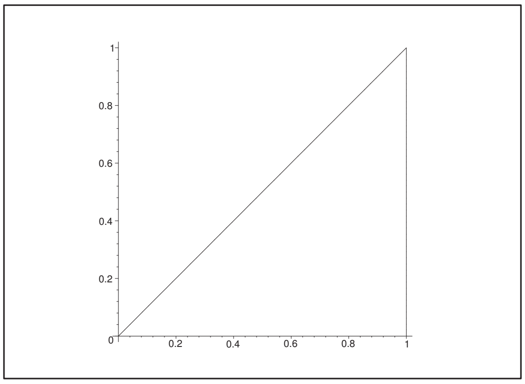


Figure 7: The domain D.

2) The means are

$$E\{X_1\} = \int_0^1 4x_1^4\,dx_1 = \frac{4}{5},$$

and

$$E\{X_2\} = \frac{4}{3}\int_0^1 \left(x_2 - x_2^4\right)\,dx_2 = \frac{4}{3}\left(\frac{1}{2} - \frac{1}{5}\right) = \frac{4}{3}\cdot\frac{3}{10} = \frac{2}{5}.$$

3) It follows from

$$E\{X_1X_2\} = \int_0^1 x_1 \left\{ \int_0^{x_1} x_2 \cdot 4x_1^2 \, dx_2 \right\} dx_1 = \int_0^1 4x_1^3 \cdot \frac{1}{2}\, x_1^2 \, dx_1 = \frac{2}{6} = \frac{1}{3},$$

that

$$\operatorname{Cov}(X_1, X_2) = E\{X_1X_2\} - E\{X_1\} \cdot E\{X_2\} = \frac{1}{3} - \frac{4}{5} \cdot \frac{2}{3} = \frac{1}{3} - \frac{8}{25} = \frac{1}{75}.$$

4) By solving the equations

$$y_1 = x_1 \qquad \text{and} \qquad y_2 = x_1 - 2x_2$$

with respect to (x_1, x_2) we get

$$x_1 = y_1 \qquad \text{and} x_2 = \frac{1}{2}\,(t_1 - y_2)\,,$$

proving that the map is bijective.

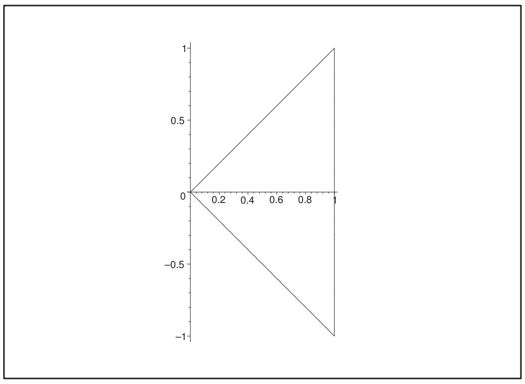


Figure 8: The image D'.

The images of the boundary curves are described by

- The line segment $0 < x_1 < 1$, $x_2 = 0$, is mapped into

$$(y_1, y_2) = (x_1, x_1)\,, \qquad 0 < x_1 < 1.$$

- The line segment $x_1 = 1$, $0 < x_2 < 1$, is mapped into

$$y_1 = 1 \qquad \text{and} \qquad y_2 = 1 - 2x_2, \quad 0 < x_2 < 1.$$

- The line segment $(x_1, x_2) = t(1, 1)$, $0 < t < 1$, is mapped into the line segment

$$(y_1, y_2) = (t, -t), \qquad 0 < t < 1.$$

Since a bounded set is mapped into a bounded set, it follows that D' is the triangle on the figure.

5) The Jacobian is

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Then by the transformation formula,

$$k(y_1,y_2) = \left|-\frac{1}{2}\right| \cdot 4y_1^2 = 2y_1^2 \qquad \text{i } D',$$

and

$$k(y_1,y_2) = 0 \qquad \text{for } (y_1,y_2) \notin D'.$$

6) By a vertical integration,

$$f_{Y_1}(y_1) = 2y_1 \cdot 2y_1^2 = 4y_1^3 \qquad \text{for } 0 < y_1 < 1,$$

and

$$f_{Y_1}(y_1) = 0 \qquad \text{otherwise.}$$

By a horizontal integration,

$$f_{Y_2}(y_2) = \int_{|y_2|}^{1} 2y_1^2\,dy_1 = \frac{2}{3}\left(1-|y_2|^3\right) \qquad \text{for } -1 < y_2 < 1,$$

and

$$f_{Y_2}(y_2) = 0 \qquad \text{otherwise.}$$

7) The means are

$$E\{Y_1\} = E\{X_1\} = \frac{4}{5}$$

and

$$E\{Y_2\} = E\{X_1 - 2X_2\} = \frac{4}{5} - 2\cdot\frac{2}{5} = 0.$$

Concerning $E\{Y_2\}$ one may alternatively apply that $f_{Y_2}(y_2)$ is an even function over a symmetric interval. The computations, however, are in this case far bigger.

8) Since $y_1 y_2 k(y_1,y_2)$ is an odd function in y_2, it follows by the symmetry with respect to the Y_1 axis that $E\{Y_1Y_2\} = 0$, hence

$$\text{Cov}(Y_1,Y_2) = E\{Y_1Y_2\} - E\{Y_1\}\cdot E\{Y_2\} = 0,$$

whence Y_1 and Y_2 are non-correlated.

9) The support D' of the frequency $k(y_1,y_2)$ is not rectangular. Hence Y_1 and Y_2 are not independent.

Example 3.5 *Let* (X_1, X_2) *be a 2-dimensional random variable of frequency*

$$h(x_1, x_2) = \begin{cases} \frac{3}{2}\, x_2, & (x_1, x_2) \in D, \\ \\ 0, & \textit{otherwise.} \end{cases}$$

where

$$D = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_2 < 1,\, -x_2 < x_1 < x_2 \right\}.$$

1. *Find the marginal frequencies of* X_1 *and* X_2.

2. *Compute the means* $E\{X_1\}$ *and* $E\{X_2\}$.

3. *Prove that* X_1 *and* X_2 *are non-correlated.*

4. *Are* X_1 *and* X_2 *independent?*

We now define the random variables Y_1 *and* Y_2 *by*

$$(Y_1, Y_2) = \tau(X_1, X_2) = (-X_1 + X_2, 2X_2).$$

Without proof we may use that the vector function τ *given by*

$$\tau(x_1, x_2) = (-x_1 + x_2, 2x_2)$$

maps D *bijectively onto*

$$D' = \left((y_1, y_2) \in \mathbb{R}^2 \mid 0 < y_1 < y_2 < 2 \right\}.$$

5. *Find the simultaneous frequency* $f(y_1, y_2)$ *of* (Y_1, Y_2).

6. *Find the marginal frequencies of* Y_1 *and* Y_2.

7. *Compute* $P\{Y_2 > 2Y_1\}$.

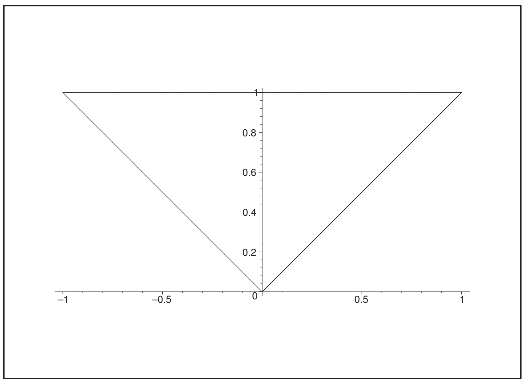


Figure 9: The domain D.

1) We get by a vertical integration,

$$f_{X_1}(x_1) = \int_{|x_1|}^{1} \frac{3}{2}\, x_2 \, dx_2 = \frac{3}{4}\left(1 - x_1^2\right) \qquad \text{for } -1 < x_1 < 1,$$

and

$$f_{X_1}(x_1) = 0 \qquad \text{otherwise.}$$

Then by a horizontal integration,

$$f_{X_2}(x_2) = \int_{-x_2}^{x_2} \frac{3}{2}\, x_2 \, dx_2 = 3x_2^2 \qquad \text{for } 0 < x_2 < 1,$$

and

$$f_{X_2}(x_2) = 0 \qquad \text{otherwise.}$$

2) The means are

$$E\{X_1\} = \int_{-1}^{1} x_1 \cdot \frac{3}{4}\left(1 - x_1^2\right) dx_1 = 0,$$

because the integrand is an odd function, and the interval of integration is symmetric with respect to 0, and

$$E\{X_2\} = \int_0^1 3x_2^3 \, dx_2 = \frac{3}{4}.$$

3) Now,

$$E\{X_1X_2\} = \int_0^1 \frac{3}{2}\,x_2^2 \left\{\int_{-x_1}^{x_2} x_1\,dx_1\right\} dx_2 = 0,$$

because the integrand is odd in x_1, and we integrate it over a symmetric interval with respect to 0 (the dependency of x_2 does not matter anything here)- Hence,

$$\text{Cov}\,(X_1, X_2) = E\{X_1X_2\} - E\{X_1\} \cdot E\{X_2\} = 0,$$

proving that X_1 and X_2 are non-correlated.

4) Since D is not a rectangular domain, X_1 and X_2 are not independent.

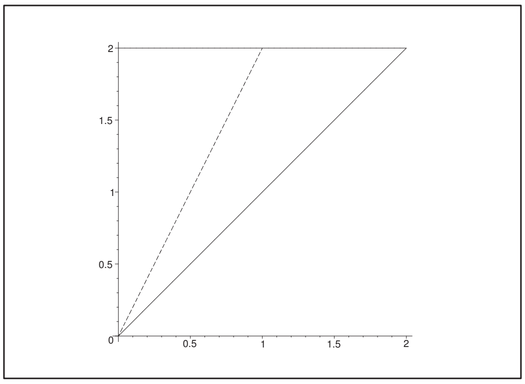


Figure 10: The domain D' with the cut by the line $y_2 = 2y_1$.

5) It follows from

$$y_1 = x_1 + x_2 \qquad \text{and} \qquad y_2 = 2x_2$$

that

$$x_2 = \frac{1}{2}\,y_2 \quad \text{and} \quad x_1 = -y_1 + x_2 = -y_1 + \frac{1}{2}\,y_2,$$

hence the Jacobian is

$$\frac{\partial\,(x_1, x_2)}{\partial\,(y_1, y_2)} = \begin{vmatrix} -1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

If $(y_1, y_2) \in D'$, i.e. $0 < y_1 < y_2 < 2$, then by the transformation formula,

$$k\,(y_1, y_2) = \left|-\frac{1}{2}\right| \cdot \frac{3}{2} \cdot \left(\frac{1}{2}\,y_2\right) = \frac{3}{8}\,y_2,$$

and

$$k\,(y_1, y_2) = 0 \qquad \text{otherwise.}$$

6) Then by a vertical integration,

$$f_{Y_1}(y_1) = \int_{y_1}^{2} \frac{3}{8}\, y_2\, dy_2 = \frac{3}{16}\left(4 - y_1^2\right) \qquad \text{for } 0 < y_1 < 2,$$

and

$$f_{Y_1}(y_1) = 0 \qquad \text{otherwise.}$$

Horizontal integrations then give

$$f_{Y_2}(y_2) = \frac{3}{8}\, y_2^2 \qquad \text{for } 0 < y_2 < 2,$$

and

$$f_{Y_2}(y_2) = 0 \qquad \text{otherwise.}$$

7) When we write the wanted probability as a planar integral, then

$$\begin{aligned} P\{Y_2 > 2Y_1\} &= \int_0^1 \left\{ \int_{2y_1}^{2} \frac{3}{8}\, y_2\, dy_2 \right\} dy_1 = \int_0^1 \frac{3}{8} \left[\frac{1}{2}\, y_2^2 \right]_{2y_1}^{2} dy_1 = \frac{3}{16} \int_0^1 \left\{4 - 4y_1^2\right\} dy_1 \\ &= \frac{3}{4} \int_0^1 \left(1 - y_1^2\right) dy_1 = \frac{3}{4}\left(1 - \frac{1}{3}\right) = \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}. \end{aligned}$$

ALTERNATIVELY and somewhat more sophisticated we notice that the line $y_2 = 2y_1$ intersects the triangle D' into two triangles of the same weight, because $k(y_1, y_2) = \frac{3}{8} y_2$ in D' only depends on y_2, and because the line $y_2 = 2y_2$ intersects every horizontal line segments through D' into two line segments of equal length.

Example 3.6 *Let* (X_1, X_2) *be a 2-dimensional random variable of frequency*

$$h(x_1, x_2) = \begin{cases} 4\, e^{-(x_1+2x_2)}, & (x_1, x_2) \in D, \\ 0, & \textit{otherwise}, \end{cases}$$

where

$$D = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_1 < 2x_2 < \infty \right\},$$

and let $(Y_1, Y_2) = \tau(X_1, X_2)$ *be given by*

$$Y_1 = X_1 + 2X_2, \qquad Y_2 = X_1 - 2X_2.$$

1) Prove that τ *maps* D *bijectively onto the domain*

$$D' = \left\{ (y_1, y_2) \in \mathbb{R}^2 \mid y_2 < 0,\ y_1 + y_2 > 0 \right\}.$$

2) Find the frequency $k(y_1, y_2)$ *of* (Y_1, Y_2).

3) Find the marginal frequencies of Y_1 *and* Y_2.

4) Check if Y_1 *and* Y_2 *are independent.*

5) Find the means of Y_1 *and* Y_2.

6) Find the variances of Y_1 *and* Y_2.

7) Compute the correlation coefficient $\varrho(Y_1, Y_2)$.

1) If

$$y_1 = x_1 + 2x_2, \qquad y_2 = x_1 - 2x_2,$$

then

$$x_1 = \frac{1}{2}(y_1 + y_2), \qquad x_2 = \frac{1}{4}(y_1 - y_2),$$

hence the x-s are uniquely determined by the y-s, which proves that the map is bijective.

We shall now describe the domain D'.
The half line $x_2 = \frac{1}{2} x_1$, $x_1 > 0$, is mapped into $y_2 = 0$, $y_1 + y_2 > 0$, i.e. into the positive y_1 axis.
The half line $x_1 = 0$, $x_2 > 0$, is mapped into $(y_1, y_2) = (2x_2, -2x_2)$, $x_2 > 0$, i.e. into $y_2 = -y_1$ and $y_1 > 0$.
We shall now decide which angular space is the right one. However, since also $y' > 0$, it follows that D' is uniquely determined as the angular space in the fourth quadrant between the line $y_2 = -y_1$ and the y_1 axis.

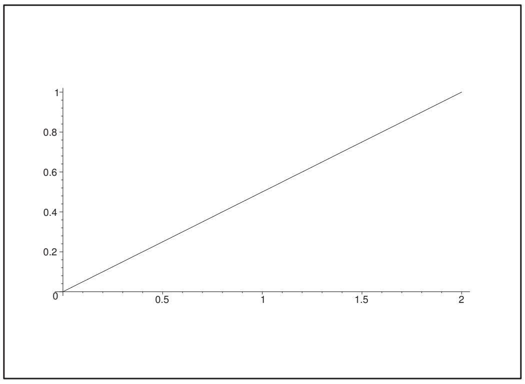


Figure 11: The domain D lies in the first quadrant above the line $x_2 = \frac{1}{2}\,x_1$.

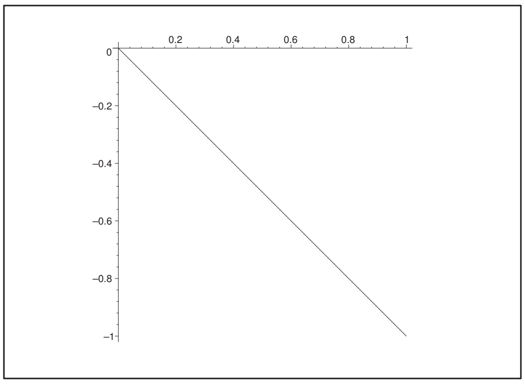


Figure 12: The domain D' lies in the fourth quadrant between the oblique line $y_2 = -y_1$ and the x axis.

2) The Jacobian is

$$\frac{\partial (x_1, x_2)}{\partial (y_1, y_2)} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} \end{vmatrix} = -\frac{1}{4}.$$

It follows from the transformation formula that

$$k(y_1, y_2) = \begin{cases} \left|-\dfrac{1}{4}\right| \cdot 4 \cdot \exp(-y_1) = e^{-y_1} & \text{for } (y_1, y_2) \in D', \\ 0, & \text{otherwise.} \end{cases}$$

3) By a vertical integration,

$$f_{Y_1}(y_1) = \begin{cases} y_1\, e^{-y_1} & \text{for } y_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

By a horizontal integration,

$$f_{Y_2}(y_2) = \begin{cases} \int_{-y_2}^{\infty} e^{-y_1}\,dy_1 = e^{y_2} & \text{for } y_2 < 0, \\ \\ 0 & \text{otherwise.} \end{cases}$$

4) Since D' is not a rectangle parallel to the axes, Y_1 and Y_2 ar not independent.

5) The means are

$$E\{Y_1\} = \int_0^{\infty} y_1^2\, e^{-y_1}\,dy_1 = 2,$$

and

$$E\{Y_2\} = \int_{-\infty}^{0} y_2\, e^{y_2}\,dy_2 = -1.$$

6) It follows from

$$E\left\{Y_1^2\right\} = \int_0^{\infty} y_1^3\, e^{-y_1}\,dy_1 = 3! = 6,$$

that

$$V\{Y_1\} = 6 - 2^2 = 2.$$

It follows from

$$E\left\{Y_2^2\right\} = \int_{-\infty}^{0} y_2^2\, e^{y_2}\,dy_2 = \int_0^{\infty} t^2 e^{-t}\,dt = 2,$$

that

$$V\{Y_2\} = 2 - (-1)^2 = 1.$$

7) We now compute

$$\begin{aligned} E\{Y_1Y_2\} &= \int_0^{\infty}\left\{\int_{-y_1}^{0} y_1y_2\, e^{-y_1}\,dy_2\right\}dy_1 = \int_0^{\infty} y_1\, e^{-y_1}\left[\frac{1}{2}\,y_2^2\right]_{-y_1}^{0} dy_1 \\ &= -\frac{1}{2}\int_0^{\infty} y_1^3\, e^{-y_1}\,dy_1 = -3. \end{aligned}$$

Then

$$\operatorname{Cov}(Y_1Y_2) = E\{Y_1\}\,E\{Y_2\} = -3 - 2(-1) = -1,$$

and hence

$$\varrho(Y_1, Y_2) = \frac{\operatorname{Cov}(Y_1, Y_2)}{\sqrt{V\{Y_1\}\,V\{Y_2\}}} = \frac{-1}{\sqrt{2\cdot 1}} = -\frac{\sqrt{2}}{2}.$$

Example 3.7 *Let* (X_1, X_2) *be a 2-dimensional random variable of the frequency*

$$h(x_1, x_2) = \begin{cases} 4\,e^{-(x_1+3x_2)}, & (x_1, x_2) \in D, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$D = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_2 < x_1 < \infty\right\}.$$

1. *Find the marginal frequencies of* X_1 *and* X_2.

2. *Find the means* $E\{X_1\}$ *and* $E\{X_2\}$.

We now define the random variables Y_1 *and* Y_2 *by*

$$(Y_1, Y_2) = \tau(X_1, X_2) = (-X_1 + X_2, X_1 + 3X_2).$$

Without proof we may use that the vector function τ *given by*

$$\tau(x_1, x_2) = (-x_1 + x_2, x_1 + 3x_2)$$

maps D *bijectively onto*

$$D' = \left\{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 + y_2 > 0,\ y_1 < 0\right\}.$$

3. *Find the simultaneous frequency* $k(y_1, y_2)$ *of* (Y_1, Y_2).

4. *Find the marginal frequencies of* Y_1 *and* Y_2.

5. *Compute the means* $E\{Y_1\}$ *and* $E\{Y_2\}$.

6. *Are* Y_1 *and* Y_2 *independent?*

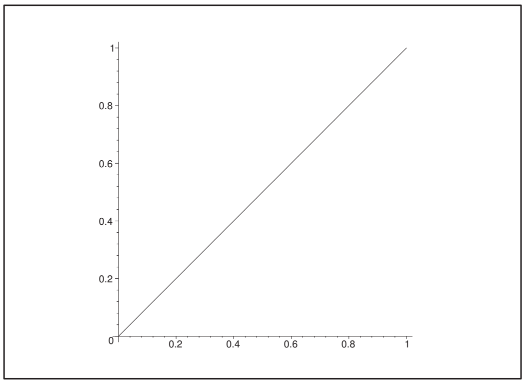


Figure 13: The domain D lies in the first quadrant between the oblique line $x_2 = x_1$ and the x_1 axis.

1) By a vertical integration for $x_1 > 0$,

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^{x_1} 4\, e^{-(x_1+3x_2)}\, dx_2 = 4\, e^{-x_1} \left[-\frac{1}{3}\, e^{-3x_2} \right]_0^{x_1} \\ &= \frac{4}{3}\, e^{-x_1} \left(1 - e^{-3x_1}\right) = \frac{4}{3} \left(e^{-x_1} - e4^{-4x_1}\right), \end{aligned}$$

and $f_{X_1}(x_1) = 0$ for $x_1 \le 0$.

By a horizontal integration for $x_2 > 0$,

$$f_{X_2}(x_2) = \int_{x_2}^{\infty} 4\, e^{-(x_1+3x_2)}\, dx_1 = 4\, e^{-3x_2} \left[-e^{-x_1} \right]_{x_2}^{\infty} = 4\, e^{-4x_2},$$

and $f_{X_2}(x_2) = 0$ otherwise.

2) The means are

$$E\{X_1\} = \frac{4}{3} \int_0^{\infty} \left\{ x_1\, e^{-x_1} - x_1\, e^{-4x_1} \right\} dx_1 = \frac{4}{3} \left\{ 1 - \frac{1}{4^2} \right\} = \frac{4}{3} \cdot \frac{15}{16} = \frac{5}{4},$$

and

$$E\{X_2\} = 4 \int_0^{\infty} x_2\, e^{-4x_2}\, dx_2 = \frac{1}{4}.$$

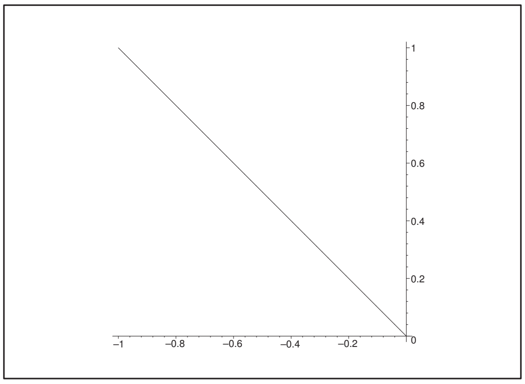


Figure 14: The domain D' lies in the second quadrant between the oblique line $y_2 = -y_1$ and the vertical y_2 axis.

3) It follows from

$$y_1 = -x_1 + x_2, \qquad y_2 = x_1 + 3x_2,$$

that

$$y_1 + y_2 = 4x_2, \qquad \text{i.e.} \qquad x_2 = \frac{1}{4}\,y_1 + \frac{1}{4}\,y_2,$$

and

$$x_1 = x_2 - y_1 = -\frac{3}{4}\,y_1 + \frac{1}{4}\,y_2,$$

i.e.

$$x_1 = -\frac{3}{4}\,y_1 + \frac{1}{4}\,y_2 \qquad \text{and} \qquad x_2 = \frac{1}{4}\,y_1 + \frac{1}{4}\,y_2.$$

Hence, we get the Jacobian

$$\frac{\partial\,(x_1, x_2)}{\partial\,(y_1, y_2)} = \begin{vmatrix} -\frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{vmatrix} = -\frac{1}{4}.$$

Then by the transformation formula,

$$k\,(y_1, y_2) = \begin{cases} \left|-\dfrac{1}{4}\right| \cdot 4 \cdot e^{-y_2} = e^{-y_2} & \text{for } (y_1, y_2) \in D', \\ 0 & \text{otherwise.} \end{cases}$$

4) Then by a vertical integration,

$$f_{Y_1}\,(y_1) = \begin{cases} \int_{-y_1}^{\infty} e^{-y_2}\,dy_2 = e^{y_1} & \text{for } y_1 < 0, \\ 0 & \text{for } y_1 \geq 0. \end{cases}$$

A horizontal integration gives

$$f_{Y_2}(y_2) = \begin{cases} \int_{-y_2}^{0} e^{-y_2}\,dy_1 = y_2 e^{-y_2} & \text{for } y_2 > 0, \\ 0 & \text{for } y_2 \le 0. \end{cases}$$

5) The means are

$$E\{Y_1\} = E\{-X_1 + X_2\} = -E\{X_1\} + E\{X_2\} = -\frac{5}{4} + \frac{1}{4} = -1,$$

and

$$E\{Y_2\} = E\{X_1 + 3X_2\} = E\{X_1\} + 3E\{X_2\} = \frac{5}{4} + \frac{3}{4} = 2.$$

6) Since D' is not a rectangle parallel to the axes, Y_1 and Y_2 are not independent.

Example 3.8 *A rectangle has its side lengths X_1 and X_2, where X_1 and X_2 are independent random variables, and where X_1 is rectangularly distributed over $]0,2[$, and X_2 is rectangularly distributed over $]0,1[$.*

1. *Find the mean of the circumference of the rectangle, $E\{2X_1+2X_2\}$.*

2. *Find the mean of the area of the rectangle, $E\{X_1X_2\}$.*

Let the 2-dimensional random variable $(Y_1,Y_2)=\tau(X_1,X_2)$ be given by

$$Y_1 = X_1X_2, \qquad Y_2 = \frac{X_1}{X_2}.$$

3. *Prove that τ maps $]0,2[\times]0,1[$ bijectively onto the domain*

$$D' = \left\{(y_1,y_2)\in\mathbb{R}^2 \mid 0<y_1<2,\, y_1<y_2<\frac{4}{y_1}\right\}.$$

4. *Find the frequency $k(y_1,y_2)$ of (Y_1,Y_2).*

5. *Find the marginal frequencies of Y_1 and Y_2.*

6. *Check if $Y_2 = X_1/X_2$ has a mean.*

7. *Find the probability*

$$P\left\{\frac{1}{3}X_1 < X_2 < 3X_1\right\}.$$

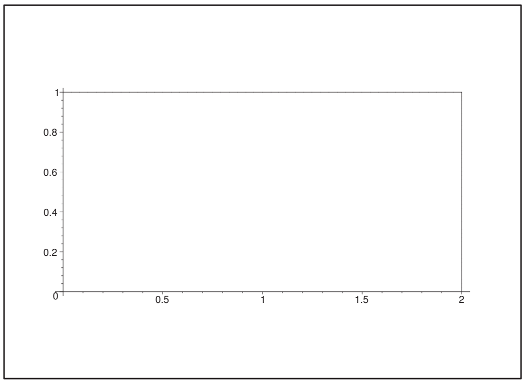


Figure 15: The domain D.

1) It follows from $E\{X_1\}=1$ and $E\{X_2\}=\dfrac{1}{2}$, that

$$E\{2X_1+2X_2\} = 2\left(E\{X_1\}+E\{X_2\}\right) = 3.$$

2) Since X_1 and X_2 are independent, we get

$$E\{X_1X_2\} = E\{X_1\}\cdot E\{X_2\} = 1\cdot\frac{1}{2} = \frac{1}{2}.$$

3) Then solve the equations

$$y_1 = x_1 x_2, \quad y_2 = \frac{x_1}{x_2}, \qquad 0 < x_1 < 2, \quad 0 < x_2 < 1,$$

with respect to x_1 and x_2. Clearly, $0 < y_1 < 2$ and $y_2 > 0$, so

$$x_1 = \sqrt{y_1 y_2} \qquad \text{and} \qquad x_2 = \sqrt{\frac{y_1}{y_2}}.$$

We conclude that the map is bijective.

Then we shall find the image $D' = \tau(D)$.

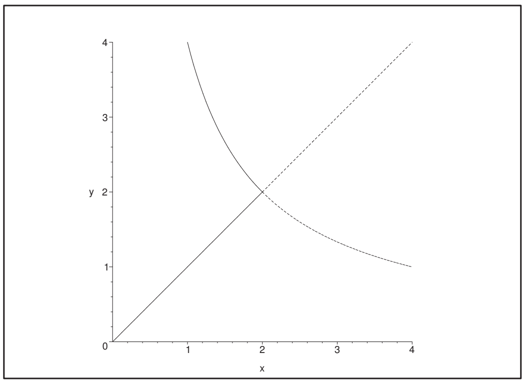


Figure 16: The domain D' lies between the hyperbolic arc and the line $y_2 = y_1$, and the vertical y_2 axis.

- When $x_1 = 0$ and $0 < x_2 < 1$, then s $y_1 = 0$ and $y_2 = 0$.
- When $x_1 = 2$ and $0 < x_2 < 1$, then $(y_1, y_2) = \left(2x_2, \dfrac{2}{x_2}\right)$, thus $y_2 = \dfrac{4}{y_1}$ and $0 < y_1 < 2$.
- When $x_2 = 1$ and $0 < x_1 < 2$, then $(y_1, y_2) = (x_1, x_1)$, i.e. $y_2 = y_1$.

We conclude from the continuity and the claim $0 < y_1 < 2$ that

$$D' = \left\{ (y_1, y_2) \in \mathbb{R}^2 \mid 0 < y_1 < 2,\, y_1 < y_2 < \frac{4}{y_1} \right\}.$$

4) Since $y_2 > 0$, the Jacobian becomes

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{1}{2}\sqrt{\frac{y_2}{y_1}} & \frac{1}{2}\sqrt{\frac{y_1}{y_2}} \\ \frac{1}{2}\sqrt{\frac{1}{y_1 y_2}} & -\frac{1}{2}\sqrt{\frac{y_1}{y_2}} \end{vmatrix} = -\frac{1}{4}\sqrt{\frac{y_2}{y_1} \cdot \frac{y_1}{y_2^3}} - \frac{1}{4}\sqrt{\frac{y_1}{y_2} \cdot \frac{1}{y_1 y_2}} = -\frac{1}{2y_2}.$$

From $h(x_1, x_2) = \dfrac{1}{2}$ for $(x_1, x_2) \in D$, follows that

$$k(y_1, y_2) = \begin{cases} \dfrac{1}{4y_2} & \text{for } (y_1, y_2) \in D', \\ 0 & \text{otherwise.} \end{cases}$$

5) When $0 < y_1 < 2$, we get by a vertical integration

$$f_{Y_1}(y_1) = \int_{y_1}^{4/y_1} \frac{1}{4y_2}\, dy_2)\frac{1}{4}\left[\ln y_2\right]_{y_1}^{4/y_1} = \frac{1}{4}\left(\ln\frac{4}{y_1} - \ln y_1\right) = \frac{1}{2}\ln\left(\frac{2}{y_1}\right),$$

hence

$$f_{Y_1}(y_1) = \begin{cases} \dfrac{1}{2}(\ln 2 - \ln y_1) & \text{for } 0 < y_1 < 2, \\ 0 & \text{otherwise.} \end{cases}$$

When $0 < y_2 \le 2$, we get by a horizontal integration,

$$f_{Y_2}(y_2) = \frac{y_2}{4y_2} = \frac{1}{4}.$$

If instead $y_2 > 2$, then

$$f_{Y_2}(y_2) = \frac{1}{4y_2} \cdot \frac{4}{y_2} = \frac{1}{y_2^2}.$$

Summing up,

$$f_{Y_2}(y_2) = \begin{cases} \dfrac{1}{4} & \text{for } 0 < y_2 \le 2, \\ \dfrac{1}{y_2^2} & \text{for } 2 < y < \infty, \\ 0 & \text{for } -\infty < y \le 0. \end{cases}$$

6) The improper integral

$$\int_2^{\infty} y_2\, f_{Y_2}(y_2)\, dy_2 = \int_2^{\infty} \frac{1}{y_2}\, dy_2 = \infty,$$

is clearly divergent, hence $E\{Y_2\}$ does not exist.

7) Since $X_2 > 0$, it follows by a small rewriting

$$\begin{aligned} P\left\{\frac{1}{3}X_1 < X_2 < 3X_1\right\} &= P\left\{\frac{1}{3}Y_2 < 1 < 3Y_2\right\} = P\left\{\frac{1}{3} < Y_2 < 3\right\} = \int_{\frac{1}{3}}^{3} f_{Y_2}(y_2)\, dy_2 \\ &= \int_{\frac{1}{3}}^{2} \frac{1}{4}\, dy_2 + \int_2^3 \frac{1}{y_2^2}\, dy_2 = \frac{1}{4}\left(2 - \frac{1}{3}\right) + \left[-\frac{1}{y_2}\right]_2^3 = \frac{5}{12} - \frac{1}{3} + \frac{1}{2} \\ &= \frac{5 - 4 + 6}{12} = \frac{7}{12}. \end{aligned}$$

4 Conditional distributions

Example 4.1 *Let (X, Y) be a 2-dimensional random variable of frequency $h(x, y)$ and marginal frequencies $f(x)$ and $g(y)$, and let $f(x \mid y)$ be the conditional frequency of X, given $Y = y$.*
Let φ be a function $: \mathbb{R} \to \mathbb{R}$, for which

$$\int_{-\infty}^{\infty} |\varphi(x)|\, f(x \mid y)\, dx < \infty \qquad \text{for alle } y \in \mathbb{R}.$$

In such a case we define the conditional mean of $\varphi(X)$, given $Y = y$, by

$$(1) \quad \int_{-\infty}^{\infty} \varphi(x)\, f(x \mid y)\, dx.$$

The conditional mean of $\varphi(X)$, given Y, *is the random variable, which for $Y = y$ has the value of (1). Hence, the conditional mean is a function in Y, and it is denoted by $E\{\varphi(X) \mid Y\}$.*
If $\varphi(x) = x$, we get in particular the conditional mean of X, given Y, *and for $\varphi(x) = (x - E\{X \mid Y\})^2$ we get the* conditional variance of X, given Y.

1) *Assuming that the random variable $E\{X \mid Y\}$ has a mean, prove that*

$$E\{X\} = E\{E\{X \mid Y\}\}.$$

2) *Find an analogous formula which expresses $V\{X\}$ by means of the conditional mean $E\{X \mid Y\}$ and the conditional variance $V\{X \mid Y\}$.*

3) *Let Ψ be a function $: \mathbb{R} \to \mathbb{R}$Prove that $E\{[X - \Psi(Y)]^2\}$ has its minimum for $\Psi(Y) = E\{X \mid Y\}$.*

1) We have

$$h(x, y) = f(x \mid y)\, g(y).$$

If we put $Z = \varphi(Y) = E\{X \mid T\}$, then Z has the values

$$\int_{-\infty}^{\infty} x\, f(x \mid y)\, dx, \qquad \text{if } g(y) \neq 0,$$

and 0 otherwise. Hence, the values of $Z = E\{X \mid Y\}$ are

$$z(y) = \begin{cases} \dfrac{1}{g(y)} \int_{-\infty}^{\infty} x\, h(x, y)\, dx & \text{for } g(y) \neq 0, \\[2ex] 0 & \text{for } g(y) = 0. \end{cases}$$

Since $g(y) = 0$ implies that $h(x, y) = 0$ almost everywhere, the mean of $Z = E\{X \mid Y\}$ is given by

$$\begin{aligned} E\{Z\} &= E\{E\{X \mid Y\}\} = \int_{-\infty}^{\infty} z(y)\, g(y)\, dy = \int_{g(y)\neq 0} \frac{1}{g(y)} \int_{-\infty}^{\infty} x\, h(x, y)\, dx \cdot g(y)\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, h(x, y)\, dx\, dy = E\{X\}. \end{aligned}$$

ALTERNATIVELY we may use that $E\{X \mid Y\}$ for $Y = y$ has the value

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

so

$$\begin{aligned} E\{E\{X \mid Y\}\} &= \int_{y=-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x\, f(x \mid y)\, dx \right\} g(y)\, dy = \int_{x=-\infty}^{\infty} x \left\{ \int_{y=-\infty}^{\infty} f(x \mid y) g(y)\, dy \right\} dx \\ &= \int_{x=-\infty}^{\infty} x \left\{ \int_{y=-\infty}^{\infty} f(x,y)\, dy \right\} dx = \int_{-\infty}^{\infty} x\, f(x)\, dx = E\{X\}. \end{aligned}$$

2) Then put $\varphi(x) = (x - E\{X \mid Y\})^2$. When $g(y) \neq 0$ it follows that $V\{X \mid Y\}$ has the values

$$\begin{aligned} &\int_{-\infty}^{\infty} (x - E\{X \mid Y = y\})^2 f(x,y)\, dx \\ &\quad = \frac{1}{g(y)} \int_{-\infty}^{\infty} \left[x^2 - 2x\, E\{X \mid y\} + (E\{X \mid y\})^2 \right] h(x,y)\, dx, \end{aligned}$$

thus

$$\begin{aligned}
E\{V\{X \mid Y\}\} &= \int_{-\infty}^{\infty} \frac{1}{g(y)} \int_{-\infty}^{\infty} \left[x^2 - 2x\, E\{X \mid y\} + (E\{X \mid y\})^2\right] h(x,y)\, dx \cdot g(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[x^2 - 2x\, E\{X \mid y\} + (E\{X \mid y\})^2\right]\, h(x,y)\, dx\, dy \\
&= E\left\{X^2\right\} - 2\int_{-\infty}^{\infty} g(y)\, E\{X \mid y\} \cdot \int_{-\infty}^{\infty} x\, f(x \mid y)\, dx\, dy \\
&\qquad + \int_{-\infty}^{\infty} g(y)\, (E\{X \mid y\})^2\, dy \\
&= E\left\{X^2\right\} - 2\int_{-\infty}^{\infty} g(y) \cdot E\{X \mid y\} \cdot E\{X \mid y\}\, dy \\
&\qquad + \int_{-\infty}^{\infty} g(y)\, (E\{X \mid y\})^2\, dy \\
&= E\left\{X^2\right\} - \int_{-\infty}^{\infty} (E\{X \mid y\})^2\, g(y)\, dy \\
&= V\{X\} + (E\{X\})^2 - \int_{-\infty}^{\infty} (E\{X \mid y\})^2\, g(y)\, dy \\
&= V\{X\} + (E\{E\{X \mid Y\}\})^2 - E\{(E\{X \mid Y\})^2\},
\end{aligned}$$

and hence

$$V\{X\} = E\{V\{X \mid Y\}\} - (E\{E\{X \mid Y\}\})^2 + E\left\{(E\{X \mid Y\})^2\right\}.$$

Alternatively and more sophisticated we first compute

$$\begin{aligned}
V\{X\} &= E\left\{(X - E\{X\})^2\right\} = E\left\{[(X - E\{X \mid Y\}) + E\{X \mid Y\} - E\{XY\})]^2\right\} \\
&= E\left\{(X - E\{X \mid Y\})^2\right\} + E\left\{(E\{X \mid Y\} - E\{X\})^2\right\} \\
&\qquad +2\, E\{(X - E\{X \mid Y\}) \cdot (E\{X \mid Y\} - E\{X\})\} \\
&= E\left\{(X - E\{X \mid Y\})^2\right\} + V\{E\{X \mid Y\}\} \\
&\qquad +2\, E\{(X - E(X \mid Y)\} \cdot E\{X \mid Y\}\} = 0.
\end{aligned}$$

Then the claim follows if we can prove that the third term above is 0.
We first compute the simpler expression

$$\begin{aligned}
E\{X \cdot E\{X \mid Y\}\} &= \int\!\!\int \{x \int x\, f(x \mid y)\, dx\} h(x,y)\, dx\, dy \\
&= \int\!\!\int \{x \int x\, f(x \mid y)\, dy\} f(x \mid y)\, g(y)\, dx\, dy \\
&= \int_y \left\{\left(\int_x x\, f(x \mid y)\, dx \cdot g(y)\right) \cdot \int_x x\, f(x \mid y)\, dx\right\} dy \\
&= \int_y g(y) \cdot \{x\, f(x \mid y)\, dx\}^2\, dy = E\left\{(E\{X \mid Y\})^2\right\}.
\end{aligned}$$

Then

$$0 = E\{X \cdot E\{X \mid Y\}\} - E\left\{(E\{X \mid Y\})^2\right\} = E\{(X - E\{X \mid Y\}) \cdot E\{X \mid Y\}\},$$

and we conclude that the third term is indeed 0 as claimed above, and it follows that

$$V\{X\} = V\{E\{X \mid Y\}\} + E\left\{[X - E\{X \mid Y\}]^2\right\}.$$

3) By a small computation,

$$\begin{aligned} E\left\{[X-\Psi(Y)]^2\right\} &= E\left\{[X - E\{X \mid Y\} + E\{X \mid Y\} + E\{X \mid Y\} - \Psi(Y)]^2\right\} \\ &= E\left\{[X - E\{X \mid Y\}]^2\right\} + 2\,E\{[X - E\{X \mid Y\}]\,[E\{X \mid Y\} - \Psi(Y)]\} \\ &\qquad + E\left\{[E(X \mid Y\} - \Psi(Y)]^2\right\}. \end{aligned}$$

Here

$$\begin{aligned} &2\,E\{[X - E\{X \mid Y\}]\,[E\{X \mid Y\} - \Psi(Y)]\} \\ &= 2\int_{-\infty}^{\infty} g(y)(E\{X \mid y\} - \Psi(Y)) \int_{-\infty}^{\infty} (x - E\{X \mid y\})\, f(x \mid y)\, dx\, dy \\ &= 2\int_{-\infty}^{\infty} g(y)\,[E\{X \mid y\} - \Psi(y)]\,[E\{X \mid y\} - E\{X \mid y\}]\, dy \\ &= 0. \end{aligned}$$

Hence

$$E\left\{[X-\Psi(Y)]^2\right\} = E\left\{[X - E\{X \mid Y\}]^2\right\} + E\left\{[E\{X \mid Y\} - \Psi(Y)]^2\right\}.$$

Since $E\left\{[E\{X \mid Y\} - \Psi(Y)]^2\right\} \geq 0$, and $E\left\{[E\{X \mid Y\} - \Psi(Y)]^2\right\} = 0$ imply that $\Psi(Y) = E\{X \mid Y\}$, the claim is proved.
ALTERNATIVELY,

$$E\left\{[X-\psi(Y)]^2\right\} = \int_y g(y)\left\{\int_x (x-\psi(y))^2 f(x \mid y)\, dx\right\} dy$$

is smallest, when

$$\int (x-\psi(y))^2 f(x \mid y)\, dx$$

is smallest. This is the case, if and only if

$$\psi(y) = \int_x x\, f(x \mid y)\, dx,$$

hence

$$\psi(Y) = E\{X \mid Y\}.$$

Example 4.2 *Let the 2-dimensional random variable (X,Y) have the frequency*

$$f(x,y)=\begin{cases}\dfrac{1}{2}\,x^3\,e^{-x(y+1)}, & x>0 \text{ and } y>0,\\[2mm] 0, & \text{otherwise.}\end{cases}$$

Find the conditional frequencies $f(x \mid y)$ and $f(y \mid x)$, and find the conditional means $E\{X \mid Y\}$ and $E\{Y \mid X\}$.

First find the marginal frequencies. When $x>0$, then

$$f_X(x)=\frac{1}{2}\int_0^{\infty} x^3 e^{-x(y+1)}\,dy=\frac{1}{2}\,x^2\,e^{-x}.$$

When $y>0$, then

$$f_Y(y)=\frac{1}{2}\int_0^{\infty} x^3\,e^{-x(y+1)}\,dy=\frac{1}{2}\,\frac{1}{(y+1)^4}\int_0^{\infty} t^3 e^{-t}\,dt=\frac{3}{(y+1)^4}.$$

Summing up,

$$f_X(x)=\begin{cases}\dfrac{1}{2}\,x^2\,e^{-x}, & x>0,\\[2mm] 0, & x\le 0,\end{cases}$$

and

$$f_Y(y)=\begin{cases}\dfrac{3}{(y+1)^4}, & y>0,\\[2mm] 0, & y\le 0.\end{cases}$$

Since

$$f(x,y)=f(x\mid y)\,f_Y(y)=f(y\mid x)\,f_X(x)$$

where $f(x \mid y)=0$ for $f_Y(y)=0$, and analogously, if follows for x, $y>0$, that

$$f(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{1}{2}\,x^3\,e^{-x(y+1)}\Big/\frac{3}{(y+1)^4}=\frac{1}{6}\,x^3\,(y+1)^4 e^{-x(y+1)},$$

and

$$f(y\mid x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{2}\,x^3 e^{-x(y+1)}\Big/\frac{1}{2}\,x^2 e^{-x}=x\,e^{-xy},$$

with the value 0 otherwise.

We get from Example 4.1 for given $Y=y>0$, that

$$\begin{aligned}E\{X\mid y\} &= \int_0^{\infty} x\,f(x\mid y)\,dx=\frac{1}{6}\,(y+1)^4\int_0^{\infty} x^4 e^{-x(y+1)}\,dx\\ &= \frac{1}{6}\,\frac{1}{y+1}\int_0^{\infty} t^4\,e^{-t}\,dt=\frac{24}{6}\cdot\frac{1}{y+1}=\frac{4}{y+1},\end{aligned}$$

hence

$$E\{X \mid Y\} = \frac{4}{Y+1}.$$

Analogously, for given $X = x > 0$,

$$E\{Y \mid x\} = \int_0^{\infty} y\, f(y \mid x)\, dy = \int_0^{\infty} y\, x\, e^{-xy}\, dy = \frac{1}{x} \int_0^{\infty} t\, e^{-t}\, dt = \frac{1}{x},$$

hence

$$E\{Y \mid X\} = \frac{1}{X}.$$

Example 4.3 *Let X_1 and X_2 be independent random variables of frequency*

$$f(x) = \begin{cases} a\,e^{-ax}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where a is a positive constant, and let the random variable Y be given by $Y = X_1 + X_2$.

1) Find the conditional frequency $f(x_1 \mid y)$ of X_1, for given $Y = y$.

2) Find the conditional mean $E\{X_1 \mid Y\}$.

1) First find the frequency $g(y)$ of Y. Obviously, $g(y) = 0$ for $y \leq 0$. When $y > 0$ we get

$$g(y) = \int_0^y a\,e^{-ax} \cdot a\,e^{-a(y-x)}\,dx = a^2 y\,e^{-ay}.$$

Let $Z = (X_1, Y) = (X_1, X_1 + X_2)$ have the frequency $h(x_1, y)$, and let $X = (X_1, X_2)$ have the frequency $k(x_1, x_2)$. Since X_1 and X_2 are independent, we get

$$k(x_1, x_2) = \begin{cases} a^2 e^{-a(x_1+x_2)} & \text{for } x_1 \geq 0 \text{ and } x_2 \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then we derive $h(x_1, y)$ from $k(x_1, x_2)$ in the following way. If we put

$$(y_1, y_2) = \psi(x_1, x_2) = (x_1, x_1 + x_2) \qquad [= (x_1, y)],$$

then the inverse map is given by

$$(x_1, x_2) = \varphi(y_1, y_2) = (y_1, y_2 - y_1) \qquad [= (x_1, y - x_1)].$$

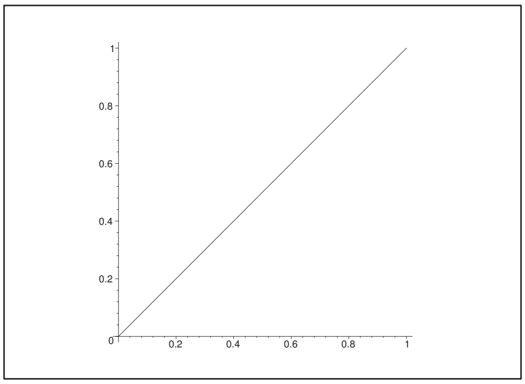


Figure 17: The domain D' is the angular space between the line $y_2 = y_1$ and the vertical y_2 axis.

The map ψ is bijective from $\mathbb{R}_+^2$ onto the domain

$$D' = \{(y_1, y_2) \mid 0 < y_1 < y_2\}.$$

The Jacobian is

$$\frac{\partial\left(x_1,x_2\right)}{\partial\left(y_1,y_2\right)}=\left|\begin{array}{cc}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2}\\ \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2}\end{array}\right|=\left|\begin{array}{cc}1 & 0\\ \\ -1 & 1\end{array}\right|=1,$$

so by the transformation formula,

$$h\left(y_1,y_2\right)=\begin{cases}f\left(x_1,y-x_1\right)\cdot 1=a^2\,e^{-a y_2} & \text{for } 0<y_1<y_2,\\ \\ 0 & \text{otherwise,}\end{cases}$$

thus

$$h\left(x_1,y\right)=\begin{cases}a^2\,e^{-ay} & \text{for } 0<x_1<y,\\ \\ \quad 0 & \text{otherwise.}\end{cases}$$

If $y \le 0$, then $f\left(x_1 \mid y\right)=0$, and if $y>0$, then

$$f\left(x_1 \mid y\right)=\frac{h\left(x_1,y\right)}{g(y)}=\frac{a^2 e^{-ay}}{a^2 y\, e^{-ay}}=\frac{1}{y} \qquad \text{for } 0<x_1<y,$$

and $f\left(x_1 \mid y\right)=0$ otherwise.

2) When $Y = y$ is given, we conclude from Example 4.1,

$$E\{X_1 \mid y\} = \frac{1}{y}\int_0^t x_1\,dx_1 = \frac{1}{y}\left[\frac{1}{2}x_1^2\right]_0^y = \frac{1}{2}\,y,$$

hence

$$E\{X_1 \mid Y\} = \frac{1}{2}\,Y.$$

Example 4.4 *Let X_1 and X_2 be independent random variables med frequency*

$$f(x) = \begin{cases} a\,e^{-ax}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where a is a positive constant. Let

$$(Y_1, Y_2) = \left(X_1^2, X_1 - X_2\right).$$

1) *Find the frequency of (Y_1, Y_2).*

2) *Find the conditional frequency of Y_1, given $Y_2 = y_2$.*

3) *Find the conditional mean of Y_1, given Y_2.*

4) *Find the correlation coefficient between Y_1 and Y_2.*

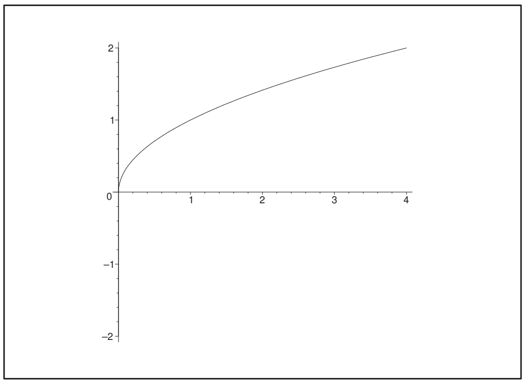


Figure 18: The domain Ω is that part of the right half plane, which lies below the parabolic arc $y_2 = \sqrt{y_1}$, $y_1 > 0$.

1) The function

$$(y_1, y_2) = \psi(x_1, x_2) = \left(x_1^2, x_1 - x_2\right)$$

maps the first quadrant $\mathbb{R}_+^2$ bijectively into the domain Ω of the figure, given by

$$\Omega = \{(y_1, y_2) \mid y_1 > 0,\, y_2 < \sqrt{y_1}\}\,.$$

The inverse map $\varphi : \Omega \to \mathbb{R}_+^2$ is given by

$$(x_1, x_2) = \varphi\,(y_1, y_2) = (\sqrt{y_1}, \sqrt{y_1} - y_2)\,.$$

The Jacobian is

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2\sqrt{y_1}} & 0 \\ \frac{1}{2\sqrt{y_1}} & -1 \end{vmatrix} = -\frac{1}{2\sqrt{y_1}} < 0.$$

If $(y_1, y_2) \in \Omega$, then the frequency of (Y_1, y_2) is given by

$$h\,(y_1, y_2) = f\,(\sqrt{y_1}) \cdot f\,(\sqrt{y_1} - y_2) \cdot \frac{1}{2\sqrt{y_1}} = a\,e^{-a\sqrt{y_1}} \cdot a\,e^{-a\sqrt{y_1} + a\,y_2} \cdot \frac{1}{2\sqrt{y_1}},$$

thus

$$h\,(y_1, y_2) = \begin{cases} \dfrac{a^2}{2\sqrt{y_1}}\,e^{-2a\sqrt{y_1} + a\,y_2} & \text{for } y_1 > 0 \text{ and } y_2 < \sqrt{y_1}, \\[2mm] 0 & \text{otherwise.} \end{cases}$$

2) First find the marginal frequency of Y_2.
If $y_2 \le 0$, then we get by a horizontal integration,

$$\begin{aligned} f_{Y_2}\,(y_2) &= \int_0^\infty h\,(y_1, y_2)\,dy_1 = a^2\,e^{a\,y_2} \int_0^\infty \frac{1}{2\sqrt{y_1}}\,e^{-2a\sqrt{y_1}}\,dy_1 \\ &= a^2 e^{a\,y_2} \int_0^\infty e^{-2at}\,dt = \frac{a}{2}\,e^{a\,y_2} = \frac{a}{2}\,e^{-a|y_2|}. \end{aligned}$$

If instead $y_2 > 0$, then by a horizontal integration,

$$\begin{aligned} f_{Y_2}\,(y_2) &= a^2 e^{a\,y_2} \int_{y_2^2}^\infty \frac{1}{2\sqrt{y_1}}\,e^{-2a\sqrt{y_1}}\,dy_1 = a^2 e^{a\,y_2} \int_{y_2}^\infty e^{-2at}\,dt \\ &= \frac{a}{2}\,e^{a\,y_2} \cdot e^{-2a\,y_2} = \frac{a}{2}\,e^{-a\,y_2} = \frac{a}{2}\,e^{-1|y_2|}. \end{aligned}$$

Summing up,

$$f_{Y_2}\,(y_2) = \frac{a}{2}\,e^{-a|y_2|}, \qquad y_2 \in \mathbb{R}.$$

If $(y_1, y_2) \in \Omega$, i.e. $y_1 > 0$ and $y_2 < \sqrt{y_1}$, then $f\,(y_1 \mid y_2)$ is given by

$$\begin{aligned} f\,(y_1 \mid y_2) &= \frac{h\,(y_1, y_2)}{f_{Y_2}\,(y_2)} = \frac{a^2}{2\sqrt{y_1}} \cdot e^{-2a\sqrt{y_1} + a\,y_2} \cdot \frac{2}{a}\,e^{a|y_2|} = \frac{a}{\sqrt{y_1}}\,e^{-2a\sqrt{y_1} + a(y_2 + |y_2|)} \\ &= \begin{cases} \dfrac{a}{\sqrt{y_1}}\,e^{-2a\sqrt{y_1} + 2a\,y_2} & \text{for } y_2 > 0, \\[2mm] \dfrac{a}{\sqrt{y_1}}\,e^{-2a\sqrt{y_1}} & \text{for } y_1 \le 0. \end{cases} \end{aligned}$$

3) If $y_2 > 0$, then we get for given $Y_2 = y_2$,

$$\begin{aligned}
E\{Y_1 \mid y_2\} &= \int_{y_2^2}^{\infty} y_1 \cdot \frac{a}{\sqrt{y_1}}\, e^{-2a\sqrt{y_1}+2ay_2}\, dy_1 = e^{2ay_2}\int_{y_2}^{\infty} t^2 \cdot 2a\, e^{-2at}\, dt \\
&= \frac{1}{4a^2}\, e^{2ay_2}\int_{2ay_2}^{\infty} u^2 e^{-u}\, du \\
&= \frac{1}{4a^2}\, e^{2ay_2}\left\{\left[-u^2e^{-u}\right]_{2ay_2}^{\infty} + 2\int_{2ay_2}^{\infty} u\, e^{-u}\, du\right\} \\
&= \frac{1}{4a^2}\, e^{2ay_2}\left\{4a^2y_2^2e^{-2ay_2} + 2\left[-u\, e^{-u}\right]_{2ay_2}^{\infty} + 2\left[-e^{-u}\right]_{2ay_2}^{\infty}\right\} \\
&= \frac{1}{4a^2}\left\{4a^2y_2^2 + 2\cdot 2ay_2 + 2\right\} = y_2^2 + \frac{1}{a}\, y_2 + \frac{1}{2a^2}.
\end{aligned}$$

On the other hand, if $y_2 \le 0$, then for given $Y_2 = y_2$,

$$E\{Y_1 \mid y_2\} = \int_0^{\infty} y_1 \cdot \frac{a}{\sqrt{y_1}}\, e^{-2a\sqrt{y_1}}\, dy_1 = 2\int_0^{\infty} a\, t^2 e^{-2at}\, dt = \frac{1}{4a^2}\int_0^{\infty} u^2 e^{-u}\, du = \frac{2!}{4a^2} = \frac{1}{2a^2}.$$

Summing up,

$$E\{Y_1 \mid Y_2\} = (\max\{Y_2, 0\})^2 + \frac{1}{a}\, \max\{Y_2, 0\} + \frac{1}{2a^2}.$$

4) Here the easiest method is to go back to the X-s. We get

$$E\{Y_1\} = E\left\{X_1^2\right\} = \int_0^{\infty} x_1^2 a\, e^{-ax_1}\, dx_1 = \frac{1}{a^2}\int_0^{\infty} t^2 e^{-t}\, dt = \frac{2}{a^2}$$

and

$$E\left\{Y_1^2\right\} = E\left\{X_1^4\right\} = \int_0^{\infty} x_1^4 a\, e^{-ax_1}\, dx_1 = \frac{1}{a^4}\int_0^{\infty} t^4 e^{-t}\, dt = \frac{24}{a^4},$$

hence

$$V\{Y_1\} = E\left\{Y_1^2\right\} - (E\{Y_1\})^2 = \frac{20}{a^4}.$$

Furthermore,

$$E\{Y_2\} = E\{X_1 - X_2\} = E\{X_1\} - E\{X_2\} = 0,$$

so

$$\begin{aligned}
V\{Y_2\} &= E\left\{Y_2^2\right\} = E\left\{(X_1 - X_2)^2\right\} = E\left\{X_1^2 - 2X_1X_2 + X_2^2\right\} \\
&= E\left\{X_1^2\right\} - 2\, E\{X_1\} \cdot E\{X_2\} + E\left\{X_2^2\right\} = 2\left(E\left\{X_1^2\right\} - (E\{X_1\})^2\right) = 2\, V\{X_1\} \\
&= 2\left\{\int_0^{\infty} x_1^2 a\, e^{-ax_1}\, dx_1 - \left(\int_0^{\infty} x_1 a\, e^{-ax_1}\, dx_1\right)^2\right\} = 2\left(\frac{2}{a^2} - \frac{1}{a^2}\right) = \frac{2}{a^2}.
\end{aligned}$$

Finally,

$$\begin{aligned} E\{Y_1Y_2\} &= E\left\{X_1^3 - X_1^2X_2\right\} \\ &= \int_0^{\infty} x_1^3 a\, e^{-ax_1}\, dx_1 - \int_0^{\infty} x_1^2 a\, e^{-ax_1}\, dx_1 \cdot \int_0^{\infty} x_2 a\, e^{-ax_2}\, dx_2 \\ &= \frac{3!}{a^3} - \frac{2!}{a^2} \cdot \frac{1}{a} = \frac{6-2}{a^3} = \frac{4}{a^3}, \end{aligned}$$

and we get

$$\text{Cov}\left(Y_1, Y_2\right) = E\{Y_1Y_2\} - E\{Y_1\} \cdot E\{Y_2\} = \frac{4}{a^3} - 0 = \frac{4}{a^3},$$

and

$$\varrho\left(Y_1, Y_2\right) = \frac{\text{Cov}\left(Y_1, Y_2\right)}{\sqrt{V\{Y_1\} \cdot V\{Y_2\}}} = \frac{\frac{4}{a^3}}{\sqrt{\frac{20}{a^4} \cdot \frac{2}{a^2}}} = \frac{4}{2\sqrt{10}} = \frac{2\sqrt{10}}{10} = \frac{\sqrt{10}}{5}.$$

5 Some theoretical results

Example 5.1 *Let X be a random variable, for which $P\{X>0\}=1$, and for which $E\{X\}$ and $E\left\{\frac{1}{X}\right\}$ exist.*
Prove that

$$1 \le E\{X\}\cdot E\left\{\frac{1}{X}\right\}.$$

HINT: *One may look at* $E\left\{\left(\sqrt{X}+t\cdot\frac{1}{\sqrt{X}}\right)^2\right\}$.

Remark 5.1 The proof is similar to the traditional proof of the Cauchy-Schwarz inequality. ◊

Since $P\{X>0\}=1$, it follows that $\sqrt{X}$ is defined.
Then by the rules of computation we get for every $t\in\mathbb{R}$ that

$$0 \le E\left\{\left(\sqrt{X}+t\cdot\frac{1}{\sqrt{X}}\right)^2\right\} = E\left\{X+2t+t^2\cdot\frac{1}{X}\right\} = t^2 E\left\{\frac{1}{X}\right\}+2t+E\{X\}.$$

The right hand side is a polynomial of second degree in t. Since it is ≥ 0 for every $t\in\mathbb{R}$, it is well-known from high school that the condition is

$$0 \ge \left(\frac{V}{2}\right)^2 - AC = 1-E\{X\}\cdot E\left\{\frac{1}{X}\right\},$$

hence by a rearrangement

$$1 \le E\{X\}\cdot E\left\{\frac{1}{X}\right\}.$$

Example 5.2 *Let X and Y be random variables where $E\left\{X^2\right\}<\infty$ and $E\left\{Y^2\right\}<\infty$.*
Prove that XY has a mean and that

$$E\{|XY|\} \le \sqrt{E\left\{X^2\right\}}\cdot\sqrt{E\left\{Y^2\right\}}.$$

We shall apply the same method as in Example 5.1.
For every $t\in\mathbb{R}$,

$$0 \le E\left\{(|X|+t|Y|)^2\right\} = E\left\{X^2+2t\,|XY|+t^2Y^2\right\} = t^2E\left\{Y^2\right\}+2t\,E\{|XY|\}+E\left\{X^2\right\},$$

where the right hand side is a non-negative polynomial of second degree in t. Then

$$|XY| \le \frac{1}{2}\,|X|^2+\frac{1}{2}\,|Y|^2,$$

exists $E\{|XY|\} < \infty$, hence $E\{XY\}$ also exists.
Finally, it follows from the condition of the discriminant that

$$(E\{|XY|\})^2 \le E\left\{X^2\right\} \cdot E\left\{Y^2\right\},$$

whence

$$E\{|XY|\} \le \sqrt{E\left\{X^2\right\}} \cdot \sqrt{E\left\{Y^2\right\}}.$$

Example 5.3 *Let (X,Y) have the frequency*

$$f(x,y) = \begin{cases} \dfrac{2}{\pi^2\left(1+x^2\right)\left(1+y^2\right)}, & x > 0, \\ \qquad\qquad 0, & x \le 0, \end{cases} \qquad y \in \mathbb{R}.$$

Prove that X and Y are independent, though not non-correlated.

If $x > 0$, then

$$f_X(x) = \int_{-\infty}^{\infty} \frac{2}{\pi^2\left(1+x^2\right)\left(1+y^2\right)}\,dy = \frac{2}{\pi} \cdot \frac{1}{1+x^2},$$

and $f_X(x) = 0$ for $x \le 0$.

Analogously we get for every $y \in \mathbb{R}$,

$$f_Y(y) = \int_0^{\infty} \frac{2}{\pi^2\left(1+x^2\right)\left(1+y^2\right)}\,dx = \frac{1}{\pi} \cdot \frac{1}{1+y^2}.$$

It follows from

$$f(x,y) = f_X(x) \cdot f_Y(y),$$

that X and Y are independent.

The phrase "*X and Y are non-correlated*" assumes that $\operatorname{Cov}(X,Y)$ exists and is $= 0$. The existence of $\operatorname{Cov}(X,Y)$ assumes again that $E\{XY\}$ exists. In the given situation this is *not* the case, because

$$\int_{-\infty}^{\infty}\left\{\int_0^{\infty} \frac{2|xy|\,dx}{\pi^2\left(1+x^2\right)\left(1+y^2\right)}\right\}dy = \frac{4}{\pi^2}\int_0^{\infty}\frac{x}{1+x^2}\,dx \cdot \int_0^{\infty}\frac{y}{1+y^2}\,dy = \infty.$$

6 The correlation coefficient

Example 6.1 *Let $X_1, X_2, \ldots, X_n$ be independent random variables for which*

$$E\{X_i\} = \mu, \quad V\{X_i\} = \sigma^2, \qquad i = 1, 2, \ldots, n.$$

Let $\overline{X}$ denote the random variable

$$\overline{X} = \frac{1}{n}\{X_1 + X_2 + \cdots + X_n\}.$$

Find the correlation coefficient $\varrho\left(\overline{X}, X_1\right)$.

Since the covariance is bilinear, and $X_1, X_2, \ldots, X_n$ are independent, it follows that

$$\operatorname{Cov}\left(\overline{X}, X_1\right) = \operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} X_i, X_1\right) = \frac{1}{n}\sum_{i=1}^{n}\operatorname{Cov}(X_i, X_1) = \frac{1}{n}\operatorname{Cov}(X_1, X_1) = \frac{1}{n}V\{X_1\} = \frac{1}{n}\sigma^2.$$

Furthermore,

$$V\{\overline{X}\} = V\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{X_i\} = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n},$$

hence

$$\varrho\left(\overline{X}, X_1\right) = \frac{\operatorname{Cov}\left(\overline{X}, X_1\right)}{\sqrt{V\{\overline{X}\}\, V\{X_1\}}} = \frac{\frac{1}{n}\sigma^2}{\frac{1}{\sqrt{n}}\sigma\cdot\sigma} = \frac{1}{\sqrt{n}}.$$

Example 6.2 *A random variable X is rectangularly distributed over $]-1, 1[$. Let $Y = X^2$ and $Z = X^3$. Find $\varrho(X, Y)$ and $\varrho(X, Z)$.*

It follows by the symmetry that

$$E\left\{X^{2n+1}\right\} = 0, \qquad n \in \mathbb{N}_0.$$

Furthermore,

$$E\left\{X^{2n}\right\} = \int_{-1}^{1}\frac{1}{2}x^{2n}\,dx = \int_{0}^{1}x^{2n}\,dx = \frac{1}{2n+1}, \qquad n \in \mathbb{N}.$$

Hence

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}\left(X, X^2\right) = E\left\{X^3\right\} - E\{X\}\cdot E\left\{X^2\right\} = 0,$$

and thus

$$\varrho(X, Y) = 0.$$

Furthermore,

$$\operatorname{Cov}(X,Z) = \operatorname{Cov}\left(X,X^3\right) = E\left\{X^4\right\} - E\{X\}\cdot E\left\{X^3\right\} = \frac{1}{5}.$$

Since

$$V\{X\} = E\left\{X^2\right\} - (E\{X\})^2 = E\left\{X^2\right\} = \frac{1}{3},$$

and

$$V\{Z\} = V\left\{X^3\right\} = E\left\{X^6\right\} - \left(E\left\{X^3\right\}\right)^2 = E\left\{X^6\right\} = \frac{1}{7},$$

we get

$$\varrho(X,Z) = \frac{\operatorname{Cov}(X,Z)}{\sqrt{V\{X\}\cdot V\{Z\}}} = \frac{\frac{1}{5}}{\sqrt{\frac{1}{3}\cdot\frac{1}{7}}} = \frac{\sqrt{21}}{5} \approx 0.917.$$

Example 6.3 *Let X and Y be random variables for which*

$$V\{X\}=1, \qquad V\{Y\}=9 \quad \text{and} \quad \varrho(X,Y)=\frac{1}{3}.$$

Let $U=X+aY$, $V=X+Y$, where a is a real constant.
Find a, such that U and V become non-correlated.

First we derive the condition,

$$\begin{aligned} 0 &= \operatorname{Cov}(U,Y)= \operatorname{Cov}(X+aY,X+Y) \\ &= \operatorname{Cov}(X,X)+a\operatorname{Cov}(Y,X)+ \operatorname{Cov}(X,Y)+a\operatorname{Cov}(Y,Y) \\ &= V\{X\}+(a+1)\operatorname{Cov}(X,Y)+a\,V\{Y\} \\ &= V\{X\}+(a+1)\varrho(X,Y)\sqrt{V\{X\}\cdot V\{Y\}}+a\,V\{Y\} \\ &= 1+(a+1)\cdot\frac{1}{3}\sqrt{1\cdot 9}+a\cdot 0=1+a+1+9a=2+10\,a. \end{aligned}$$

When this equation is solved with respect to a, we get $a=-\frac{1}{5}$.

Example 6.4 *Let X and Y be independent random variables of the frequency*

$$f(x)=\begin{cases} 1-|x|, & |x|<1, \\ 0, & |x|\geq 1. \end{cases}$$

Put $U=X^2+Y^2$ and $V=X^3+Y$. Find the correlation coefficients $\varrho(U,X)$, $\varrho(V,X)$ and $\varrho(U,V)$.

It follows from the symmetry that $E\{X\}=E\{Y\}=0$. Hence

$$V\{X\}=V\{Y\}=E\left\{X^2\right\}=\int_{-1}^{1}x^2(1-|x|)\,dx=2\int_0^1 x^2(1-x)\,dx=2\left(\frac{1}{3}-\frac{1}{4}\right)=\frac{1}{6}.$$

Analogously, $E\left\{X^{2n+1}\right\}=E\left\{Y^{2n+1}\right\}=0$, and

$$\begin{aligned} E\left\{X^{2n}\right\}=E\left\{Y^{2n}\right\} &= 2\int_0^1 x^{2n}(1-x)\,dx=\frac{2}{2n+1}-\frac{2}{2n+2} \\ &= \frac{2}{2n+1}-\frac{1}{n+1}=\frac{1}{(2n+1)(n+1)}. \end{aligned}$$

Since X and Y are independent, it follows from the above that

$$\begin{aligned} \operatorname{Cov}(U,V) &= \operatorname{Cov}\left(X^2+Y^2,X\right)= \operatorname{Cov}\left(X^2,X\right)+ \operatorname{Cov}\left(Y^2,X\right) \\ &= \operatorname{Cov}\left(X^2,X\right)=E\left\{X^3\right\}-E\left\{X^2\right\}\cdot E\{X\}=0, \end{aligned}$$

so $\varrho(U,X)=0$, and U and X are non-correlated.

Analogously;

$$\operatorname{Cov}(V,X)= \operatorname{Cov}\left(X^3+Y,X\right)= \operatorname{Cov}\left(X^3,x\right)=E\left\{X^4\right\}-E\left\{X^2\right\}\,E\{X\}=\frac{1}{5\cdot 3}=\frac{1}{15}.$$

Since

$$V\{V\} = V\left\{X^3\right\} + V\{Y\} = E\left\{X^6\right\} + E\left\{Y^2\right\} = \frac{1}{7 \cdot 4} + \frac{1}{3 \cdot 2} = \frac{3 + 2 \cdot 7}{7 \cdot 4 \cdot 3} = \frac{17}{84},$$

we get

$$\varrho(V, X) = \frac{\operatorname{Cov}(V, X)}{\sqrt{V\{V\} \cdot V\{X\}}} = \frac{\frac{1}{15}}{\sqrt{\frac{17}{84} \cdot \frac{1}{6}}} = \frac{6}{15}\sqrt{\frac{14}{17}} = \frac{2}{5}\sqrt{\frac{14}{17}} \approx 0,363.$$

Finally,

$$\begin{aligned} \operatorname{Cov}(U, V) &= \operatorname{Cov}\left(X^2 + Y^2, X^3 + Y\right) = \operatorname{Cov}\left(X^2, x^3\right) + \operatorname{Cov}\left(Y^2, Y\right) \\ &= E\left\{X^5\right\} - E\left\{X\right\} \cdot E\left\{X^3\right\} + E\left\{Y^3\right\} - E\left\{Y^2\right\} \cdot E\{Y\} = 0, \end{aligned}$$

hence $\varrho(U, V) = 0$, and U and V are non-correlated.

7 Maximum and minimum of linear combinations of random variables

Example 7.1 *1) Let X_1 and X_2 be two independent random variables, for which*

$$E\{X_1\} = E\{X_2\} = \mu \neq 0, \qquad V\{X_1\} = \sigma_1^2 > 0 \quad \textit{and} \quad V\{X_2\} = \sigma_2^2 > 0.$$

Find the constants a_1 and a_2, such that

$$E\{a_1X_1 + a_2X_2\} = \mu,$$

and such that

$$V\{a_1X_1 + a_2X_2\}$$

has its smallest value. Then find the corresponding minimum.
What is the minimum, when in particular $\sigma_1 = \sigma_2 = \sigma$?

2) Then let $X_1, X_2, \dots, X_n$ be independent random variables, for which

a) $E\{X_1\} = E\{X_2\} = \cdots = E\{X_n\} = \mu \quad (\neq 0)$,
b) $V\{X_1\} = V\{X_2\} = \cdots = V\{X_n\} = \sigma > 0$.

Find the constants $a_1, a_2, \dots, a_n$, such that

$$E\left\{\sum_{i=1}^n a_iX_i\right\} = \mu,$$

while

$$V\left\{\sum_{i=1}^n a_iX_i\right\}$$

takes its smallest value. Then find this smallest value.

1) It follows by the linearity that

$$E\{a_1X_1 + a_2X_2\} = a_1E\{X_1\} + a_2E\{X_2\} = (a_1 + a_2)\mu.$$

Since $\mu \neq 0$, this expression is $= \mu$, if and only if $a_1 + a_2 = 1$.

Put $a_1 = \lambda$. Then $a_2 = 1 - \lambda$, and

$$\varphi(\lambda) = V\{a_1X_1 + a_2X_2\} = \lambda^2V\{X_1\} + (1-\lambda)^2V\{X_2\} = \lambda^2\sigma_1^2 + (1-\lambda)^2\sigma_2^2$$

where

$$\varphi'(\lambda) = 2\lambda\sigma_1^2 + 2(\lambda - 1)\sigma_2^2 = 0$$

for

$$\lambda = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \qquad \text{thus} \qquad 1 - \lambda = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

On the other hand, we know that there *exists* a smallest value, and since the computations above give the coefficients of the only candidate, we must necessarily have

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} \qquad \text{and} \qquad a_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2},$$

corresponding to

$$V\left\{\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,X_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\,X_2\right\} = \frac{\sigma_2^4\sigma_1^2}{\left(\sigma_1^2+\sigma_2^2\right)^2} + \frac{\sigma_1^4\sigma_2^2}{\left(\sigma_1^2+\sigma_2^2\right)^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

Note that since $\sigma_1^2 > 0$ and $\sigma_2^2 > 0$, this variance is $< \min\left\{\sigma_1^2, \sigma_2^2\right\}$.

When $\sigma_1 = \sigma_2 = \sigma$, then the value of the smallest value is

$$\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma^4}{2\sigma^2} = \frac{1}{2}\,\sigma^2.$$

2) This is just a generalization. Since the equation

$$E\left\{\sum_{i=1}^n a_i X_i\right\} = \sum_{i=1}^n a_i E\left\{X_i\right\} = \sum_{i=1}^n a_i \mu = \mu \neq 0,$$

is only satisfied for

$$\sum_{i=1}^n a_i = 1,$$

we can eliminate one constant, e.g.

$$a_n = 1 - \sum_{i=1}^{n-1} a_i.$$

Then the task is reduced to minimize the function

$$\begin{aligned}\varphi\left(a_1,\ldots,a_{n-1}\right) &= V\left\{\sum_{i=1}^{n-1} a_i X_i + \left(1-\sum_{i=1}^{n-1} a_i\right) X_n\right\} = \sum_{i=1}^{n-1} a_i^2 V\left\{X_i\right\} + \left(1-\sum_{i=1}^{n-1} a_i\right)^2 V\left\{X_n\right\} \\ &= \left\{\sum_{i=1}^{n-1} a_i^2 + \left(\sum_{i=1}^{n-1} a_i - 1\right)^2\right\}\sigma^2.\end{aligned}$$

The equations of possible stationary points are

$$\frac{\partial\varphi}{\partial a_i} = \left\{2a_i + 2\left(\sum_{i=1}^{n-1} a_i - 1\right)\right\}\sigma^2 = 2\sigma^2\left\{a_i - a_n\right\} = 0,$$

for $i = 1, \ldots, n-1$, thus $a_i = a_n$ for all i. This implies that

$$\sum_{i=1}^n a_i = \sum_{i=1}^n a_n = n\,a_n = 1,$$

hence

$$a_n = \frac{1}{n} \qquad \text{and} \qquad a_i = \frac{1}{n}, \quad i = 1, \dots, n-1.$$

We have now proved that $\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ is the only stationary point.

Since $\varphi(a_1, \dots, a_{n-1})$ is of class C^∞ and is positive, and since $\varphi(a_1, \dots, a_{n-1}) \to \infty$ for $a_1^2 + \cdots + a_{n-1}^2 \to \infty$, a minimum exists. The only candidate is $\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, so this is indeed a minimum.

Finally, by insertion,

$$\varphi\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = V\left\{\sum_{i=1}^{n} \frac{1}{n} X_i\right\} = \frac{n}{n^2} V\{X_1\} = \frac{\sigma^2}{n}.$$

ALTERNATIVELY it is possible here to make some constructive guesses. We must again require that $\sum_{i=1}^{n} a_i = 1$, so getting an inspiration from the first question we guess that all $a_i = \frac{1}{n}$.

This can be proved in the following way:

Let the a_i be any such constants of $\sum_{i=1}^n a_i = 1$. Then

$$\begin{aligned} V\left\{\sum_{i=1}^n a_i X_i\right\} &= \sigma^2 \sum_{i=1}^n a_i^2 = \sigma^2 \sum_{i=1}^n \left\{\left(a_i - \frac{1}{n}\right) + \frac{1}{n}\right\}^2 = \sigma^2 \left\{\sum_{i=1}^n \left(a_i - \frac{1}{n}\right)^2 + \sum_{i=1}^n \frac{1}{n^2}\right\} \\ &= \sigma^2 \left\{\sum_{i=1}^n \left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n}\right\}. \end{aligned}$$

It follows that the minimum is obtained when the first term in the parenthesis is 0, i.e. when all $a_i = \frac{1}{n}$. With these choices we finally get the minimum $\frac{\sigma^2}{n}$.

Example 7.2 *Let $X_1, X_2, \dots, X_n$ be independent random variables, for which*

$$E\{X_i\} = \mu \quad (\neq 0), \qquad V\{X_i\} = \sigma_i^2 > 0, \qquad i = 1, 2, \dots, n.$$

Find constants $a_1, a_2, \dots, a_n$, such that

$$E\left\{\sum_{i=1}^n a_i X_i\right\} = \mu,$$

while

$$V\left\{\sum_{i=1}^n a_i X_i\right\}$$

takes on its minimum. Then find this minimum.

Remark 7.1 This example is of course a generalization of Example 7.1. ◊

1) First we compute

$$E\left\{\sum_{i=1}^n a_i X_i\right\} = \left(\sum_{i=1}^n a_i\right)\mu = \mu \neq 0 \quad \text{for} \quad \sum_{i=1}^n a_i = 1,$$

and

$$V\left\{\sum_{i=1}^n a_i X_i\right\} = \sum_{i=1}^n a_i^2 V\{X_i\} = \sum_{i=1}^n \sigma_i^2 a_i^2.$$

Since

$$a_n = 1 - \sum_{i=1}^{n-1} a_i \qquad \text{where} \qquad \frac{\partial a_n}{\partial a_i} = -1,$$

it follows that we shall minimize the function

$$\varphi(a_1, \dots, a_{n-1}) = \sum_{i=1}^{n-1} \sigma_i^2 a_i^2 + \sigma_n^2 \left(1 - \sum_{i=1}^{n-1} a_i\right)^2, \qquad a_n = 1 - \sum_{i=1}^{n-1} a_i.$$

2) The equations of possible stationary points are

$$\frac{\partial \varphi}{\partial a_i} = 2\sigma_i^2 a_i + 2\sigma_n^2 a_n \frac{\partial a_n}{\partial a_i} = 2\left(\sigma_i^2 a_i - \sigma_n^2 a_n\right) = 0, \qquad i = 1, \dots, n-1.$$

They imply that

$$a_i = \frac{\sigma_n^2}{\sigma_i^2}\, a_n, \qquad i = 1, \dots, n-1.$$

Then by insertion,

$$1 = \sum_{i=1}^n a_i = \left(\sum_{i=1}^n \frac{\sigma_n^2}{\sigma_i^2}\right) a_n = \sigma_n^2 \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right) \cdot a_n,$$

thus

$$a_n = \frac{1}{\sigma_n^2 \sum_{i=1}^n \left(\frac{1}{\sigma_i^2}\right)} \quad \text{and} \quad a_i = \frac{1}{\sigma_i^2 \sum_{i=1}^n \left(\frac{1}{\sigma_i^2}\right)}, \qquad i = 1, \dots, n-1,$$

giving us the coordinates of the only stationary point.

3) It follows from

$$\varphi\left(a_1, \dots, a_{n-1}\right) \to \infty \qquad \text{for} \quad a_1^2 + \cdots + a_{n-1}^2 \to \infty,$$

that we get a minimum at this stationary point. Hence, the minimum is given by

$$\left(a_1, \dots, a_n\right) = \frac{1}{\sum_{i=1}^n \left(\frac{1}{\sigma_i^2}\right)} \left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \dots, \frac{1}{\sigma_n^2}\right).$$

Here, the value is

$$V\left\{\sum_{i=1}^n a_i X_i\right\} = \frac{1}{\left\{\sum_{i=1}^n \frac{1}{\sigma_i^2}\right\}^2} \sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^4} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}.$$

ALTERNATIVELY we may pass straight ahead towards the task of finding the a_i, such that $\sum_{i=1}^n a_i = 1$, and $\sum_{i=1}^n a_i^2 \sigma_i^2$ is as small as possible.
If we put $x_i = a_i \sigma_i$, i.e. $a_i = \dfrac{x_i}{\sigma_i}$, we see that we shall find the x_i, such that

$$\sum_{i=1}^n \frac{1}{\sigma_i}\, x_i = 1 \text{ and } \sum_{i=1}^n x_i^2 \text{ is as small as possible.}$$

Here the condition

$$\sum_{i=1}^n \frac{1}{\sigma_i}\, x_i = 1$$

describes an hyperplane in $\mathbb{R}^n$ with the normed normal vector

$$\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_n}\right) \cdot \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}.$$

We obtain the smallest distance to the zero for

$$x_i = \frac{\frac{1}{\sigma_i}}{\sum_{j=1}^n \frac{1}{\sigma_j^2}}, \qquad \text{and the distance is} \qquad \frac{1}{\sum_{j=1}^n \frac{1}{\sigma_i^2}}.$$

The conclusion is that

$$a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^n \frac{1}{\sigma_i^2}}, \qquad i = 1, 2, \dots, n,$$

and that the minimum is

$$\frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}.$$

ALTERNATIVELY it was proved in Example 7.1, first question that the minimum is obtained for

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \qquad \text{and} \qquad a_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}.$$

Therefore, we *guess* that the minimum in the general case is obtained when

$$a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^n \frac{1}{\sigma_j^2}}, \qquad i = 1, \dots, n.$$

This can be proved in the following way:
Let the a_i be any numbers for which $\sum_{i=1}^n a_i = 1$. Then

$$\begin{aligned}
V\left\{\sum_{i=1}^n a_i X_i\right\} &= \sum_{i=1}^n a_i^2 \sigma_i^2 = \sum_{i=1}^n \left\{\left(a_i - \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}\right) + \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}\right\}^2 \sigma_i^2 \\
&= \sum_{i=1}^n \left\{a_i - \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}\right\}^2 \sigma_i^2 + \sum_{i=1}^n + \sum_{n=1}^n \frac{1/\sigma_i^2}{\left\{\sum_{j=1}^n 1/\sigma_j^2\right\}^2} \\
&\quad +2\sum_{i=1}^n \left\{a_i - \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}\right\} \cdot \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2} \cdot \sigma_i^2 \\
&= \sum_{i=1}^n \left\{a_i - \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}\right\}^2 \sigma_i^2 + \frac{1}{\sum_{j=1}^n 1/\sigma_j^2} + 0,
\end{aligned}$$

because it is easily seen that the last sum above is 0.
This implies that the minimum is obtained when all squares in the first sum are equal to 0, thus

$$a_i = \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}, \qquad i = 1, 2, \dots, n,$$

and the minimum is

$$\frac{1}{\sum_{j=1}^n 1/\sigma_j^2}.$$

Example 7.3 *Let $X_1, X_2, \dots, X_n$ be independent random variables, for which*

$$E\{X_1\} = E\{X_2\} = \cdots = E\{X_n\} = \mu \quad (\neq 0),$$

$$V\{X_1\} = V\{X_2\} = \cdots = V\{X_n\} = \sigma^2 > 0.$$

Find constants $a_1, a_2, \dots, a_n$, such that

$$a_i > 0, \qquad i = 1, 2, \dots, n,$$

$$E\left\{\sum_{i=1}^{n} a_i X_i\right\} = \mu,$$

while at the same time,

$$V\left\{\sum_{i=1}^{n} a_i X_i\right\}$$

takes its maximum, and find this maximum.

First note that taking the mean is a linear operation, so

$$\sum_{i=1}^{n} a_i \mu = \mu \neq 0, \qquad \text{thus } \sum_{i=1}^{n} a_i = 1.$$

Furthermore, all $a_i \geq 0$, $i = 1, 2, \dots, n$.

We shall maximize the function

$$\varphi(a_1, a_2, \dots, a_n) = V\left\{\sum_{i=1}^{n} a_i X_i\right\} = \sum_{i=1}^{n} a_i^2 \sigma^2,$$

under the conditions above.

Obviously,

$$1 = (a_1 + a_2 + \cdots + a_n)^2 \geq a_1^2 + \cdots + a_n^2 = \sum_{i=1}^{n} a_i^2,$$

so this maximum must be $\leq 1 \cdot \sigma^2$.

On the other hand, this value is obtained, when precisely one $a_i = 1$, and all others are $a_j = 0$, $j \neq i$. Thus, the maximum is

$$V\{X_1\} = V\{X_2\} = \cdots = V\{X_n\} = \sigma^2.$$

Example 7.4 *Let $X_1, X_2, \dots, X_n$ be independent Bernoulli distributed random variables of probabilities of success $p_1, p_2, \dots, p_n$, and let $Y = \sum_{i=1}^n X_i$. It is well-known that*

$$E\{Y\} = \sum_{i=1}^n p_i.$$

Prove that if $E\{Y\}$ is a fixed number s, then the variance $V\{Y\}$ is largest, if $p_1 = p_2 = \cdots = p_n$- Then find this maximum.

The Bernoulli distribution is given by

$$P\{X=1\} = p \qquad \text{and} \qquad P\{X=0\} = q,$$

where $p+q=1$, p, $q > 0+$. Then $E\{X\} = p$ and $E\left\{X^2\right\} = p$, hence

$$V\{X\} = E\left\{X^2\right\} - (E\{X\})^2 = p - p \qquad [= p(p-1) = pq].$$

If we assume that $0 < s < n$ is constant and that

$$\sum_{i=1}^n p_i = s, \qquad 0 < p_i < 1 \quad \text{for} \quad i = 1, \dots, n,$$

then we shall maximize

$$\begin{aligned} V\{Y\} &= \sum_{i=1}^n V\left\{X_i\right\} = \sum_{i=1}^n \left(p_i - p_i^2\right) = s - \sum_{i=1}^n p_1^2 = s - \sum_{i=1}^n \left\{\left(p_i - \frac{s}{n}\right) + \frac{s}{n}\right\}^2 \\ &= s - \sum_{i=1}^n \left(p_i - \frac{s}{n}\right)^2 - 2\,\frac{s}{n} \sum_{i=1}^n \left(p_i - \frac{s}{n}\right) - \sum_{i=1}^n \frac{s^2}{n^2} \\ &= s - \sum_{i=1}^n \left(p_i - \frac{s}{n}\right)^2 - 2\,\frac{s}{n} \cdot \left(s - n \cdot \frac{s}{n}\right) - \frac{s^2}{n} = s - \frac{s^2}{n} - \sum_{i=1}^n \left(p_i - \frac{s}{n}\right)^2. \end{aligned}$$

Clearly, this expression is largest, when $p_i = \dfrac{s}{n}$ for $i = 1, \dots, n$, and when this holds, then

$$V\{Y\} = s - \frac{s^2}{n} \qquad (> 0,\ \text{because } 0 < s < n).$$

Example 7.5 *1) Let X be a random variable of mean μ and variance σ^2. Prove that $E\left\{(X-a)^2\right\}$ has its minimum at $a=\mu$.*

2) Let X_1 and X_2 be random variables of means μ_1, μ_2, resp., variances σ_1^2, σ_2^2, resp., and correlation coefficient ϱ.
For which pairs of numbers (a,b) does

$$(2)\quad E\left\{[X_2-(aX_1+b)]^2\right\}$$

obtain its smallest value?
Then find this minimum.
HINT: *First keep a fixed and find the value of b, for which the expression (2) is as small as possible-*

1) A direct computation gives

$$\begin{aligned} E\left\{(X-a)^2\right\} &= E\left\{[(X-\mu)+(\mu-a)]^2\right\} \\ &= E\left\{(X-\mu)^2\right\}+E\{2(\mu-a)(X-\mu)\}+E\left\{(\mu-a)^2\right\} \\ &= E\left\{(X-\mu)^2\right\}+2(\mu-a)\,E\{X-\mu\}+(\mu-a)^2 \\ &= E\left\{(X-\mu)^2\right\}+(\mu-a)^2, \end{aligned}$$

from which immediately follows that $E\left\{(X-a)^2\right\}$ obtains its minimum for $a=\mu$.

2) Then by a simple reduction,

$$\begin{aligned} \varphi(a,b) &= E\left\{[X_2-a\,(X_1+b)]^2\right\} \\ &= E\left\{[(X_2-aX_1)-(\mu_2-a\mu_1)+(\mu_2-a\mu_1)+b]^2\right\} \\ &= E\left\{[(X_2-aX_1)-(\mu_2-a\mu_1)]^2\right\} \\ &\qquad +2\,(\mu_2-a\mu_1+b)\;E\left\{(X_2-aX_1)-(\mu_2-a\mu_1)\right\} \\ &\qquad +(\mu_2-a\mu_1-b)^2 \\ &= V\left\{X_2-aX_1\right\}+2\,(\mu_2-a\mu_1-b)\,[(\mu_2-a\mu_1)-(\mu_2-a\mu_1)] \\ &\qquad +(\mu_2-a\mu_1-b)^2 \\ &= V\left\{X_2\right\}-2\,a\,\mathrm{Cov}\,(X_1,X_2)+a^2V\left\{X_1\right\}+(a\mu_1+b-\mu_2)^2 \\ &= (a\mu_1+b-\mu_2)^2+a^2\sigma_1^2-2a\varrho\sigma_1\sigma_2+\sigma_2^2. \end{aligned}$$

We search possible stationary points of

$$\varphi(a,b)=(a\mu_1+b-\mu_2)^2+a^2\sigma_1^2-2a\varrho\sigma_1\sigma_2+\sigma_2^2.$$

The equations of the stationary points are

$$\begin{cases} \dfrac{\partial\varphi}{\partial a}=2\mu_1\,(a\mu_1+b-\mu_2)+2a\sigma_1^2-2\varrho\sigma_1\sigma_2=0, \\[2ex] \dfrac{\partial\varphi}{\partial b}=2\,(a\mu_1+b-\mu_2)=0. \end{cases}$$

By a subtraction,

$$2a\sigma_1^2 - 2\varrho\sigma_1\sigma_2 = 0,$$

hence

$$a = \frac{2\varrho\sigma_1\sigma_2}{2\sigma_1^2} = \frac{\varrho\sigma_2}{\sigma_1}.$$

We get by insertion into the latter equation,

$$b = \mu_2 - \frac{\varrho\sigma_2}{\sigma_1}\,\mu_1,$$

so the only stationary point is

$$(a,b) = \left(\frac{\varrho\sigma_2}{\sigma_1}, \mu_2 - \frac{\varrho\sigma_2}{\sigma_1}\,\mu_1\right).$$

Since $\varphi(a,b) \to \infty$ for $a^2 + b^2 \to \infty$, the stationary point must necessarily be a minimum.

Finally the minimum is found to be

$$E\left\{\left[X_2 - \frac{\varrho\sigma_2}{\sigma_1}\,X_1 - \mu_2 + \frac{\varrho\sigma_2}{\sigma_1}\,\mu_1\right]^2\right\} = \frac{\varrho^2\sigma_2^2}{\sigma_1^2}\,\sigma_1^2 - 2\,\frac{\varrho\sigma_2}{\sigma_1}\cdot\varrho\sigma_1\sigma_2 + \sigma_2^2 = \varrho^2\sigma_2^2 - 2\varrho^2\sigma_2^2 + \sigma_2^2 = \sigma_2^2\left(1-\varrho^2\right).$$

ALTERNATIVELY, if a is given,

$$E\left\{[X_2-(a\,X_1+b)]^2\right\}=E\left\{[(X_2-a\,X_1)-b]^2\right\}$$

obtains according to question 1 its minimum for

$$b=E\left\{X_2-a\,X_1\right\}=\mu_2-a\,\mu_1,$$

and it follows that the minimum is

$$V\left\{X_2-a\,X_1\right\}=\sigma_2^2+a^2\sigma_1^2-2a\varrho\sigma_1\sigma_2.$$

This function in a has its minimum for

$$a=\varrho\cdot\frac{\sigma_2}{\sigma_1},$$

which either follows from high school mathematics or by noticing that the graph is a parabola.
We conclude that we obtain the minimum for

$$a=\varrho\cdot\frac{\sigma_2}{\sigma_1}\qquad\text{and}\qquad b=\mu_2-\varrho\cdot\frac{\sigma_2}{\sigma_1}\cdot\mu_1,$$

and the minimum is

$$\sigma_2^2\left(1-\varrho^2\right).$$

Example 7.6 *Let $X_1, X_2, \dots, X_n$ be independent random variables, where*

$$E\{X_i\} = \mu, \quad V\{X_i\} = \sigma^2, \qquad i = 1, \dots, n,$$

and let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Prove that

$$E\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right\} = \sigma^2.$$

HINT: *Write*

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 = \sum_{i=1}^{n}\left\{(X_i-\mu)^2 + \left(\mu-\overline{X}\right)^2 + 2(X_i-\mu)\left(\mu-\overline{X}\right)\right\}.$$

We shall only compute and reduce:

$$\begin{aligned}
E\left\{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right\} &= \frac{1}{n-1}\,E\left\{\sum_{i=1}^{n}\left[(X_i-\mu)^2+\left(\mu-\overline{X}\right)^2+2(X_i-\mu)\left(\mu-\overline{X}\right)\right]\right\}\\
&= \frac{1}{n-1}\,E\left\{\sum_{i=1}^{n}(X_i-\mu)^2\right\} + \frac{1}{n-1}\,E\left\{n\left(\overline{X}-\mu\right)^2\right\} + \frac{2}{n-1}\,E\left\{\sum_{i=1}^{n}(X_i-\mu)\left(\mu-\overline{X}\right)\right\}\\
&= \frac{1}{n-1}\sum_{i=1}^{n}E\left\{(X_i-\mu)^2\right\} + \frac{n}{n-1}\,E\left\{\left(\overline{X}-\mu\right)^2\right\} + \frac{2}{n-1}\,E\left\{n\left(\overline{X}-\mu\right)\left(\mu-\overline{X}\right)\right\}\\
&= \frac{1}{n-1}\sum_{i=1}^{n}V\{X_i\} - \frac{n}{n-1}\,E\left\{\left(\overline{X}-\mu\right)^2\right\} = \frac{1}{n-1}\sum_{i=1}^{n}\sigma^2 - \frac{n}{n-1}\,V\left\{\overline{X}\right\}\\
&= \frac{n}{n-1}\,\sigma^2 - \frac{n}{n-1}\,V\left\{\frac{1}{n}\sum_{i=1}^{n}X_i\right\} = \frac{n}{n-1}\,\sigma^2 - \frac{n}{n-1}\cdot\frac{1}{n^2}\,V\left\{\sum_{i=1}^{n}X_i\right\}\\
&= \frac{n}{n-1}\,\sigma^2 - \frac{1}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}V\{X_i\} = \frac{n}{n-1}\,\sigma^2 - \frac{1}{n-1}\cdot\frac{1}{n}\cdot n\,\sigma^2\\
&= \frac{n}{n-1}\,\sigma^2 - \frac{1}{n-1}\,\sigma^2 = \sigma^2.
\end{aligned}$$

8 Convergence in probability and in distribution

Example 8.1 *In this example we use the notation $X_n \xrightarrow{P} X$, if (X_n) converges in probability towards X. Recall that $X_n \xrightarrow{P} X$, if for every $\varepsilon \in \mathbb{R}_+$,*

$$P\{|X_n - X| \ge \varepsilon\} \to 0 \qquad \text{for } n \to \infty.$$

This can also be written in the following way:

$X_n \xrightarrow{P} X$, *if the following condition is satisfied:*

$$\forall\, \varepsilon \in \mathbb{R}_+ \,\forall\, \eta \in \mathbb{R}_+ \,\exists\, n_0 \in \mathbb{N} \,\forall\, n \in \mathbb{N} : n > n_0 \Rightarrow P\{|X_n - X| \ge \varepsilon\} < \eta.$$

1) *Prove that if $X_n \xrightarrow{P} X$, and a is a real constant, then also $aX_n \xrightarrow{P} aX$.*

2) *Prove that if $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then also $X_n + Y_n \xrightarrow{P} X + Y$.*

3) *Prove that if $X_n \xrightarrow{P} X$, then also $|X_n| \xrightarrow{P} |X|$.*

4) *Prove that if $X_n \xrightarrow{P} 0$, then also $X_n^2 \xrightarrow{P} 0$.*

5) *Prove that if $X_n \xrightarrow{P} X$, and Y is a random variable, then $X_nY \xrightarrow{P} XY$.*
HINT: *To every $\delta \in \mathbb{R}_+$ there exists $c \in \mathbb{R}_+$, such that $P\{|Y| > c\} < \delta$.*

6) *Prove that if $X_n \xrightarrow{P} X$, then also $X_n^2 \xrightarrow{P} X^2$.*
HINT: *Write X_n in the form $X_n = (X_n - X) + X$, and apply some of the results of the previous questions.*

7) *Prove that if $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then also $X_nY_n \xrightarrow{P} XY$.*
HINT: *Apply the rewriting*

$$X_nY_n = \frac{1}{4}\left\{(X_n + Y_n)^2 - (X_n - Y_n)^2\right\}.$$

1) When $a = 0$, there is nothing to prove. When $a \ne 0$, there exists an $n_1 = n_1(\varepsilon, a, \eta)$, such that

$$P\{|aX_n - aX| \ge \varepsilon\} = P\left\{|X_n - X| \ge \frac{\varepsilon}{|a|}\right\} < \eta,$$

for every $n > n_1(\varepsilon, a, \eta)$.

2) It follows from

$$|(X_n + Y_n) - (X + Y)| \le |X_n - X| + |Y_n - Y|,$$

that if $|(X_n + Y_n) - (X + Y)| \ge \varepsilon$, then either

$$|X_n - X| \ge \frac{\varepsilon}{2} \qquad \text{or} \qquad |Y_n - Y| \ge \frac{\varepsilon}{2}.$$

Then

$$\{|(X_n + Y_n) - (X + Y)| \geq \varepsilon\} \subseteqq \left\{|X_n - X| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|Y_n - Y| \geq \frac{\varepsilon}{2}\right\},$$

hence

$$P\{|(X_n + Y_n) - (X + Y)| \geq \varepsilon\} \leq P\left\{|X_n - X| \geq \frac{\varepsilon}{2}\right\} + P\left\{|Y_n - Y| \geq \frac{\varepsilon}{2}\right\} < \eta$$

for $n > n_2\left(\varepsilon, \frac{\eta}{2}, (X_n), (Y_n)\right)$.

3) Analogously, we get from $||X_n| - |X|| \leq |X_n - X|$ that

$$P\{||X_n| - |X|| \geq \varepsilon\} \leq P\{|X_n - X| \geq \varepsilon\} < \eta,$$

and the claim is proved.

4) If $X = 0$, then $|X_n| \xrightarrow{P} 0$ by (3), and

$$P\left\{X_n^2 \geq \varepsilon\right\} = P\left\{|X_n| \geq \sqrt{\varepsilon}\right\} < \eta,$$

and the claim is proved.

5) First we use the hint to estimate in general,

$$\begin{aligned} P\{|X_n Y - XY| \geq \varepsilon\} &= P\{|Y| \cdot |X_n - X| \geq \varepsilon\} \\ &= P\{|Y| \cdot |X_n - X| \geq \varepsilon \wedge |Y| > c\} + P\{|Y| \cdot |X_n - X| \geq \varepsilon \wedge |Y| \leq c\} \\ &\leq P\{|Y| > c\} + P\{c \cdot |X_n - X| \geq \varepsilon\} < \delta + P\left\{|X_n - X| \geq \frac{\varepsilon}{c}\right\}. \end{aligned}$$

Choose $\delta = \frac{\eta}{2}$. In this way we fix the constant $c > 0$.
Nowchoose $n_0 \in \mathbb{N}$, such that

$$P\left\{|X_n - X| \geq \frac{\varepsilon}{c}\right\} < \frac{\eta}{2} \qquad \text{for every } n > n_0.$$

Then for $n > n_0$,

$$P\{|X_n Y - XY| \geq \varepsilon\} < \delta + P\left\{|X_n - X| \geq \frac{\varepsilon}{c}\right\} < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

6) Since $X_n = (X_n - X) + X$, we get

$$X_n^2 - X^2 = (X_n - X)^2 + 2X(X_n - X),$$

hence by putting $Y = X$,

$$\begin{aligned} P\left\{\left|X_n^2 - X^2\right| \geq \varepsilon\right\} &\leq P\left\{(X_n - X)^2 \geq \frac{\varepsilon}{2}\right\} + P\left\{2|XX_n - XX| \geq \frac{\varepsilon}{2}\right\} \\ &= P\left\{|X_n - X| \geq \sqrt{\frac{\varepsilon}{2}}\right\} + P\left\{|YX_n - YX| \geq \frac{\varepsilon}{4}\right\}. \end{aligned}$$

By assumption, $X_n \xrightarrow{P} X$, so

$$P\left\{|X_n - X| \geq \sqrt{\frac{\varepsilon}{2}}\right\} < \frac{\eta}{2} \qquad \text{for } n > n_1.$$

Since $Y X_n \xrightarrow{P} Y X$ and $Y = X$, we get

$$P\left\{2\,|X X_n - X X| \geq \frac{\varepsilon}{2}\right\} < \frac{\eta}{2} \qquad \text{for } n > n_2.$$

Then put $n_0 = \max\{n_1, n_2\}$, and we obtain for $n > n_0$ that

$$P\left\{\left|X_n^2 - X^2\right| > \varepsilon\right\} < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

7) It follows from

$$X_n Y_n = \frac{1}{4}\left\{(X_n + Y_n)^2 - (X_n - Y_n)^2\right\},$$

that

$$|X_n Y_n - XY| = \frac{1}{4}\left|\left\{(X_n + Y_n)^2 - (X + Y)^2\right\}\right| + \frac{1}{4}\left|\left\{(X - Y)^2 - (X_n - Y_n)^2\right\}\right|.$$

If $|X_n Y_b| \geq \varepsilon$, then at least one of the two terms on the right hand side is $\geq \dfrac{\varepsilon}{2}$, hence

$$\begin{aligned}
&P\{|X_nY_n - XY| \geq \varepsilon\} \\
&\quad \leq P\left\{\frac{1}{4}\left|(X_n+Y_n)^2-(X+Y)^2\right| \geq \frac{\varepsilon}{2}\right\} + P\left\{\frac{1}{4}\left|(X_n-Y_n)^2-(X-Y)^2\right| \geq \frac{\varepsilon}{2}\right\} \\
&\quad = P\left\{\left|(X_n+Y_n)^2-(X+Y)^2\right| \geq 2\varepsilon\right\} + P\left\{\left|(X_n-Y_n)^2-(X-Y)^2\right| \geq 2\varepsilon\right\}.
\end{aligned}$$

It follows from (2) that $X_n \pm Y_n \xrightarrow{P} X \pm Y$.
Applying (6) we get $(X_n \pm Y_n)^2 \xrightarrow{P} (X \pm Y)^2$.
In particular, we can find n_1 and n_2, such that

$$P\left\{\left|(X_n+Y_n)^2-(X+Y)^2\right| \geq 2\varepsilon\right\} < \frac{\eta}{2} \qquad \text{for } n > n_1,$$

and

$$P\left\{\left|(X_n-Y_n)^2-(X-Y)^2\right| \geq 2\varepsilon\right\} < \frac{\eta}{2} \qquad \text{for } n > n_2.$$

The claim follows, when $n > n_0 = \max\{n_1, n_2\}$.

Example 8.2 *Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables, such that (X_n) converges in distribution towards a constant a.*
Prove that (X_n) converges in probability towards the constant a.
Assume furthermore that every X_n has a mean. Is it possible to conclude that $E\{X_n\} \to a$ for $n \to \infty$?

If $X_n \xrightarrow{D} a$, then

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0 & \text{for } x < a, \\ 1 & \text{for } x \geq a. \end{cases}$$

We shall prove that

$$P\{|X_n - a| \geq \varepsilon\} \to 0 \qquad \text{for } n \to \infty.$$

We get

$$\begin{aligned}
P\{|X_n-a| \geq \varepsilon\} &= P\{X_n - a \geq \varepsilon\} + P\{X_n - a \leq -\varepsilon\} = P\{X_n \geq a+\varepsilon\} + P\{X_n \leq a-\varepsilon\} \\
&= 1 - P\{X_n < a+\varepsilon\} + P\{X_n \leq a-\varepsilon\} = 1 - F_(a+\varepsilon-) + F_n(a-\varepsilon) \\
&\to 1 - F(a+\varepsilon) + F(a-\varepsilon) = 1 - 1 + 0 = 0 \qquad \text{for } n \to \infty.
\end{aligned}$$

The latter claim is in general *not* true. Choose e.g.

$$F_n(x) = \begin{cases} 1 - \dfrac{n}{x^2+n^2} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Then clearly,

$$F_n(x) \to F(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ \\ 0 & \text{for } x < 0, \end{cases} \qquad \text{for } n \to \infty,$$

thus $a = 0$.
Here,

$$E\{X_n\} = \int_0^\infty \{1 - F_n(x)\}\, dx = \int_0^\infty \frac{n}{x^2+n^2}\, dx = \int_0^\infty \frac{1}{1+\left(\frac{x}{n}\right)^2}\, d\left(\frac{x}{n}\right) = \frac{\pi}{2} \neq a = 0.$$

Obviously one can modify such examples, so one can expect a lot of unpleasant anomalies.

Example 8.3 *A box contains $\frac{n(n+1)}{2}$ slips of paper, of which on slip has the number 1 written on it, two slips are provided with the number 2, etc. until finally n slips of paper are provided with the number n. Select at random one slip from the box. Let X_n denote the random variable, which indicates the number of the selected slip, and let another random variable Y_n be defined by*

$$Y_n = \frac{1}{n}\, X_n.$$

1) *Find the probabilities $P\{X_n = k\}$, $k = 1, 2, \dots, n$.*

2) *Find the mean $E\{X_n\}$.*

3) *Prove that the distribution function of Y_n on the interval $[0,1]$ is given by*

$$F_n(y) = \frac{[ny]([ny]+1)}{n(n+1)}.$$

(Here $[a]$ denotes the largest integer smaller than or equal to a).

4) *Prove that the sequence $\{Y_n\}$ converges in distribution towards a random variable Y, and find the distribution of Y.*

HINT: *It may be convenient to use the formula*

$$\sum_{k=1}^n k^2 = \frac{1}{6}\, n(n+1)(2n+1).$$

1) Clearly,

$$P\{X_n = k\} = \frac{k}{\frac{1}{2}\, n(n+1)} = \frac{2k}{n(n+1)}, \qquad k = 1, 2, \dots, n.$$

2) When we insert the result of (1), it follows by the definition,

$$E\{X_n\} = \sum_{k=1}^n k\, P\{X_n = k\} = \frac{2}{n(n+1)} \sum_{k=1}^n k^2 = \frac{2}{n(n+1)} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{2n+1}{3}.$$

3) First note that

$$P\left\{Y_n = \frac{k}{n}\right\} = P\left\{X_n = k\right\} = \frac{2k}{n(n+1)}.$$

Thus the distribution function for Y_n is

$$F_n(y) = P\left\{Y_n \le y\right\} = \sum_{k=1}^{[ny]} P\left\{Y_n = \frac{k}{n}\right\} = \sum_{k=1}^{[ny]} \frac{2k}{n(n+1)} = \frac{[ny]([ny]+1)}{n(n+1)},$$

because $\sum_{k=1}^{m} k = \frac{1}{2}\, m(m+1)$ for $m \in \mathbb{N}$.

4) It follows from

$$ny - 1 < [ny] \le ny,$$

that

$$y - \frac{1}{n} < \frac{[ny]}{n} \le y,$$

and we conclude that

$$\frac{[ny]}{n} \to y \quad \text{and} \quad \frac{[ny]+1}{n+1} \to y \qquad \text{for } n \to \infty, \quad y \in [0,1].$$

It follows that $F_n(y) \to y^2$ for $n \to \ infty$ and $y \in [0,1]$.
This means that (Y_n) converges in distribution towards a random variable Y, the distribution function of which is

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ y^2, & 0 \le y \le 1, \\ 1, & y \ge 1. \end{cases}$$

The corresponding frequency is

$$f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example 8.4 *Let X and Y be independent random variables, both rectangularly distributed over the interval $]0,1[$.*

1) Find the distribution function $F(v)$ and the frequency $f(v)$ of the random variable

$$V = \frac{Y}{X} + 1.$$

2) Check if the mean of V exists.

3) Prove that there exists a random variable U, such that

$$\lim_{n\to\infty} P\left\{ \sqrt[n]{V} \le v \right\} = P\{U \le v\} \qquad \text{for all } v \neq 1.$$

1) It is obvious that the values of V lie in $]1,\infty[$. When $v > 1$, then

$$F(v) = P\{V \le v\} = P\left\{\frac{Y}{X} + 1 \le v\right\} = P\left\{\frac{Y}{X} \le v-1\right\}.$$

The frequency of $\dfrac{Y}{X}$ is given by

$$\begin{aligned} k(s) &= \int_0^1 f_X(sx)\, f_Y(x)\, x\, dx \\ &= \begin{cases} \int_0^1 1\cdot 1\cdot x\, dx = \dfrac{1}{2} & \text{for } 0 < s < 1, \\ \int_0^{\frac{1}{s}} 1\cdot 1\cdot x\, dx = \dfrac{1}{2s^2} & \text{for } s > 1, \end{cases} \end{aligned}$$

hence

$$\begin{aligned} F(v) &= P\left\{\frac{Y}{X} \le v-1\right\} \\ &= \int_0^{v-1} k(s)\, ds = \begin{cases} \dfrac{1}{2}(v-1), & 1 < v \le 2, \\ \dfrac{1}{2} + \int_1^{v-1} \dfrac{ds}{2s^2} = \dfrac{1}{2} - \left[\dfrac{1}{2s}\right]_1^{v-1} = 1 - \dfrac{1}{2(v-1)}, & v > 2, \end{cases} \end{aligned}$$

and we get by a differentiation,

$$f_V(v) = k(v-1) = \begin{cases} \dfrac{1}{2}, & \text{for } 1 < v \le 2, \\ \dfrac{1}{2(v-1)^2}, & \text{for } v > 2. \end{cases}$$

2) The mean does not exist. In fact,

$$\int_1^\infty v\, f_V(v)\, dv = \int_1^2 \frac{v}{2}\, dv + \int_2^\infty \frac{v}{2(v-1)^2}\, dv = \infty.$$

3) To any $v > 1$ there exists an $N = N(v)$, such that $v^2 > 2$ for every $n > N$, and such that

$$P\left\{\sqrt[n]{V} \le v\right\} = P\left\{V \le v^n\right\} = 1 - \frac{1}{2\left(v^2 - 1\right)} \qquad \text{for } n > N.$$

Since $V > 1$, we have $P\left\{\sqrt[n]{V} \le v\right\} = 0$ for $v \le 1$. By taking the limit $n \to \infty$ we get

$$\lim_{n\to\infty} P\left\{\sqrt[n]{V} \le v\right\} = \begin{cases} 1 & \text{for } v > 1, \\ 0 & \text{for } v \le 1. \end{cases}$$

The right hand side is the distribution function of the causal random variable U, for which

$P\{U = 1\} = 1.$

Example 8.5 *A 2-dimensional random variable* (X, Y) *has the frequency*

$$h(x,y) = \begin{cases} x+y & \text{for } 0 \le x \le 1, \quad 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

1) Find the frequencies of the random variables X *and* Y*.*

2) Find the means and the variances of the random variables X *and* Y*.*

3) Find the frequency of the random variable $X + Y$*.*

4) Find for every $n \in \mathbb{N}$ *the distribution function* $F_n(x)$ *and the frequency* $f_n(x)$ *of the random variable* X^n *and prove that for every* $\varepsilon > 0$*,*

$$P\{X^n > \varepsilon\} \to 0 \qquad \text{for } n \to \infty.$$

1) If $x \in [0,1]$, then

$$f_X(x) = \begin{cases} \int_0^1 (x+y)\,dy = x + \dfrac{1}{2}, & x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

It follows by the symmetry,

$$f_Y(y) = \begin{cases} \int_0^1 (x+y)\,dx = y + \dfrac{1}{2}, & y \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

2) The means exist, and by the symmetry,

$$E\{X\} = E\{Y\} = \int_0^1 t\left(t + \frac{1}{2}\right) dt = \int_0^1 \left(t^2 + \frac{t}{2}\right) dt = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

3) Since the values of $X + Y$ lie in $[0,2]$, the frequency is for $s \in [0,2]$ given by

$$g(s) = \int_0^1 h(x, s-x)\,dx.$$

The integrand is $\neq 0$, when $0 \le s - x \le 1$, so the domain of integration is determined by $s - 1 \le x \le s$ and $0 \le 1$, hence

$$g(s) = \begin{cases} \int_0^s s\,dx = s^2 & \text{for } s \in [0,1], \\ \int_{s-1}^1 s\,dx = s(2-s) = 1 - (s-1)^2 & \text{for } s \in [1,2]. \end{cases}$$

Summing up,

$$g(s) = \begin{cases} s^2 & \text{for } s \in [0,1], \\ 1-(s-1)^2 & \text{for } s \in \,]1,2], \\ 0 & \text{otherwise.} \end{cases}$$

4) Since the values of X lie in $[0,1]$, we get for $x \in [0,1]$ that

$$F_n(x) = P\{X^n \le x\} = P\left\{X \le \sqrt[n]{x}\right\} = \int_0^{\sqrt[n]{x}} \left(t+\frac{1}{2}\right) dt = \frac{1}{2}\left(\sqrt[n]{x^2}+\sqrt[n]{x}\right) = \frac{1}{2}\left\{x^{\frac{2}{n}}+x^{\frac{1}{n}}\right\},$$

and

$$f_n(x) = \frac{1}{2}\left\{\frac{2}{n}\,x^{\frac{2}{n}-1}+\frac{1}{n}\,x^{\frac{1}{n}-1}\right\} = \begin{cases} \dfrac{1}{2nx}\left\{2\sqrt[n]{x^2}+\sqrt[n]{x}\right\} & \text{for } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Finally,

$$P\{X^n > \varepsilon\} = 1 - P\{X^n \le \varepsilon\} = 1 - \frac{1}{2}\left\{\varepsilon^{\frac{2}{n}}+\varepsilon^{\frac{1}{n}}\right\} \to 1 - \frac{1}{2}(1+1) = 0 \quad \text{for } n \to \infty.$$

Example 8.6 *Given a sequence of random variables $(X_n)_{n=1}^{\infty}$, where X_n has the frequency*

$$f_n(x) = \begin{cases} n(n+1)\,x^{n-1}(1-x), & x \in]0,1[, \\ 0, & \textit{otherwise.} \end{cases}$$

1. *Find the mean of X_n.*

For every fixed $n \in \mathbb{N}$ we define a random variable Y_n by

$$Y_n = (X_n)^n .$$

2. *Find the distribution function $G_n(y)$ and the frequency $g_n(y)$ of Y_n.*

3. *Prove that the sequence $(Y_n)_{n=1}^{\infty}$ converges in distribution towards a random variable Y.*

4. *Finally, find the frequency of Y.*

We start by noting that for $0 < x < 1$ the distribution function $F(x)$ of X is given by

$$F(x) = \int_0^x f_n(t)\,dt = (n+1)x^n - n\,x^{n+1}.$$

1) The mean of X_n is

$$E\{X_n\} = \int_0^1 x\,f_n(x)\,dx = n(n+1)\int_0^1 \left(x^n - x^{n+1}\right) dx = n(n+1)\left(\frac{1}{n+1}-\frac{1}{n+2}\right) = \frac{n}{n+2}.$$

2) The distribution function of $Y_n = X_n^n$ for $0 < y < 1$ is given by

$$G_n(y) = P\{Y_n = X_n^n \le y\} = P\left\{X_n \le y^{\frac{1}{n}}\right\} = (n+1)y - n\,y^{1+\frac{1}{n}},$$

thus

$$G_n(y) = \begin{cases} 0, & \text{for } y \le 0, \\ (n+1)y - n\,y^{1+\frac{1}{n}}, & \text{for } 0 < y < 1, \\ 1, & \text{for } y \ge 1, \end{cases}$$

and hence by differentiation,

$$g_n(y) = \begin{cases} (n+1)\left(1 - y^{\frac{1}{n}}\right) & \text{for } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

3) According to l'Hospital's theorem,

$$\lim_{x \to 0} \frac{1 - y^x}{x} = \lim_{x \to 0} \frac{-\ln y \cdot y^x}{1} = -\ln y.$$

Put $x = \dfrac{1}{n}$. Then by insertion and by taking the limit,

$$\lim_{n \to \infty} n\left(1 - y^{\frac{1}{n}}\right) = \lim_{n \to \infty} \frac{1 - y^{\frac{1}{n}}}{\frac{1}{n}} = -\ln y.$$

Then finally for $y \in]0,1[$,

$$G_n(y) = y + ny\left(1 - y^{\frac{1}{n}}\right) \to y - y\,\ln y \qquad \text{for } n \to .$$

Consequently, (Y_n) converges in distribution towards a random variable Y of the distribution function

$$G(y) = \begin{cases} 0, & \text{for } y \le 0, \\ y - y\,\ln y, & \text{for } 0 < y < 1, \\ 1, & \text{for } y \ge 1. \end{cases}$$

4) The frequency of Y is derived by differentiation, $g(y) = G'(y)$, thus

$$g(y) = \begin{cases} -\ln y, & \text{for } 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example 8.7 *We define a sequence of random variables* $(X_n)_{n=1}^{\infty}$ *by assuming that* X_n *has the distribution function*

$$F_n(x) = \begin{cases} 0, & x < 0, \\ x^n, & x \in [0,1], \\ 1, & x > 1. \end{cases}$$

1) Find the frequency $f_n(x)$ *of* X_n *and find the mean and the variance of* X_n*.*

2) Prove that the sequence (X_n) *converges in distribution towards a random variable* X*, and find the distribution of* X*.*

3) Prove that

$$E\{X_n\} \to E\{X\} \quad \text{and} \quad V\{X_n\} \to V\{X\} \qquad \text{for } n \to \infty.$$

4) Assuming that the variables X_2 *and* X_3 *above are independent, find the frequency of the random variable*

$$Z = X_2 + X_3.$$

1) The frequency of X_n is obtained from $F_n(x)$ by differentiation

$$f_n(x) = \begin{cases} n\,x^{n-1} & \text{for } x \in]0,1[, \\ 0 & \text{otherwise.} \end{cases}$$

The mean is

$$E\{X_n\} = \int_0^1 n\,x^n\,dx = \frac{n}{n+1}.$$

From

$$E\left\{X_n^2\right\} = \int_0^1 n\,x^{n+1}\,dx = \frac{n}{n+2},$$

we get the variance

$$\begin{aligned} V\{X_n\} &= E\left\{X_n^2\right\} - \left(E\{X_n\}\right)^2 = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 \\ &= \frac{n}{(n+2)(n+1)^2}\left\{(n+1)^2 - n(n+2)\right\} = \frac{n}{(n+2)(n+1)^2}. \end{aligned}$$

2) Trivially,

$$F(x) = \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \text{for } x < 1, \\ 1 & \text{for } x > 1, \end{cases}$$

and $F(x)$ is the distribution function of the causal random variable X, which is given by

$$P\{X = 1\} = 1.$$

3) We have for the causal distribution X that $E\{X\} = 1$ and $V\{X\} = 0$, and

$$\lim_{n\to\infty} E\{X_n\} = \lim_{n\to\infty} \frac{n}{n+1} = 1 = E\{X\},$$

and

$$\lim_{n\to\infty} V\{X_n\} = \lim_{n\to\infty} \frac{n}{(n+2)(n+1)^2} = 0 = V\{X\}.$$

4) The values of $Z = X_2 + X_3$ clearly lies in $]0,2[$. If $s \in]0,2[$, then the frequency of Z is given by the convolution integral

$$g(s) = \int_0^1 f_2(x)\, f_3(s-x)\, dx.$$

The integrand is $\neq 0$ for $0 < x < 1$ and $0 < s - x < 1$, thus $s - 1 < x < s$.
Then we must split the investigation into two cases.

a) If $s \in]0,1[$, then

$$\begin{aligned} g(s) &= \int_0^s 2x \cdot 32(s-x)^2\, dx = 6\int_0^s (s-t)t^2\, dt = 6\int_0^s \left(st^2 - t^3\right)\, dt = 6\left[\frac{1}{3}\,st^3 - \frac{1}{4}\,t^4\right]_0^s \\ &= 6\left\{\frac{1}{3} - \frac{1}{4}\right\} s^4 = \frac{1}{2}\, s^4. \end{aligned}$$

b) If $s \in]1,2[$, then we get instead

$$\begin{aligned}
g(s) &= \int_{s-1}^{1} 2x \cdot 3(s-x)^2 \, dx = 6 \int_{s-1}^{1} (s-t)t^2 \, dt = 6 \left[\frac{1}{3}\, st^3 - \frac{1}{4}\, t^4 \right]_{s-1}^{1} \\
&= 6\left(\frac{1}{3}\, s - \frac{1}{4} - \frac{1}{3}\, s(s-1)^3 + \frac{1}{4}\,(s-1)^4 \right) = 6\left(\frac{1}{3}\, s - \frac{1}{4} - (s-1)^3 \left(\frac{1}{3}\, s - \frac{1}{4}\,(s-1) \right) \right) \\
&= 6\left(\frac{1}{3}\, s - \frac{1}{4} - (s-1)^3 \left(\frac{1}{12}\, s + \frac{1}{4} \right) \right) = \frac{6}{12} \left(4s - 3 - (s+3)(s-1)^3 \right) \\
&= \frac{1}{2} \left(4s - 3 - \left\{ s^3 - 3s^2 + 3s - 1 \right\} \{s+3\} \right) \\
&= \frac{1}{2} \left(4s - 3 - \left(s^4 + 3s^3 - 3s^3 - 9s^2 + 3s^2 + 9s - s - 3 \right) \right) \\
&= \frac{1}{2} \left(4s - 3 - s^4 + 6s^2 - 8s + 3 \right) = -\frac{1}{2}\, s^4 + 3s^2 - 2s.
\end{aligned}$$

Summing up,

$$g(s) = \begin{cases} \dfrac{1}{2}\, s^4 & \text{for } s \in]0,1], \\[2ex] -\dfrac{1}{2}\, s^4 + 3s^2 - 2s & \text{for } s \in]1,2], \\[2ex] 0 & \text{otherwise.} \end{cases}$$

Example 8.8 *Three random variables X_1, X_2, X_3 are assumed to be independent, and the distribution function for each of them is given by*

$$(3)\quad F(x)=\begin{cases}0, & x<0,\\[2ex] 1-e^{-x}, & x\geq 0.\end{cases}$$

We define the random variable U by $U=\max\{X_1,X_2,X_3\}$.

1. *Find the distribution of U.*

2. *Find the mean of U.*

Let $(X_n)_{n=1}^{\infty}$ denote a sequence of independent random variables, each of them given the distribution function $F(x)$ as in (3).

3. *Let the random variables Y_n and Z_n for $n\in\mathbb{N}$ be given by*

$$Y_n=\max\left\{X_1,X_2,\dots,X_n\right\}\quad \textit{and}\quad Z_n=Y_n-\ln n.$$

Prove that the sequence (Z_n) converges in distribution towards a random variable Z of the distribution function

$$F_Z(z)=\exp\left(-e^{-z}\right),\qquad z\in\mathbb{R}.$$

1) Since X_1, X_2, X_3 are independent, the distribution function of $U=\max\left\{X_1,X_2,X_3\right\}$ is given by

$$G(u)=P\left\{X_1\leq u,\,X_2\leq u,\,X_3\leq u\right\}=P\left\{X_1\leq u\right\}\cdot P\left\{X_2\leq u\right\}\cdot P\left\{X_3\leq u\right\}=\{F(u)\}^3,$$

i.e.

$$G(u)=\begin{cases}0, & u\leq 0,\\[2ex] \left(1-e^{-u}\right)^3, & u>0.\end{cases}$$

The corresponding frequency is

$$g(u)=\begin{cases}0, & u\leq 0,\\[2ex] 3\left(1-e^{-u}\right)^2\cdot e^{-u}\quad \left[=3\left(e^{-3u}-2e^{-2u}+e^{-u}\right)\right], & u>0.\end{cases}$$

2) The *mean* is

$$\begin{aligned}E\{U\} &= \int_0^{\infty} u\,g(u)\,du=3\int_0^{\infty} u\left(e^{-3u}-2e^{-2u}+e^{-u}\right)\,du\\ &= 3\left\{\frac{1}{9}\int_0^{\infty} t\,e^{-t}\,dt-\frac{2}{4}\int_0^{\infty} t\,e^{-t}\,dt+\int_0^{\infty} t\,e^{-t}\,dt\right\}=3\left\{\frac{1}{9}-\frac{1}{2}+1\right\}=\frac{11}{6}.\end{aligned}$$

ALTERNATIVELY,

$$E\{U\}=\int_0^{\infty}\{1-G(u)\}\,du=\int_0^{\infty}\left\{e^{-3u}-3e^{-2u}+3e^{-u}\right\}du=\frac{1}{3}-\frac{3}{2}+3=\frac{11}{6}.$$

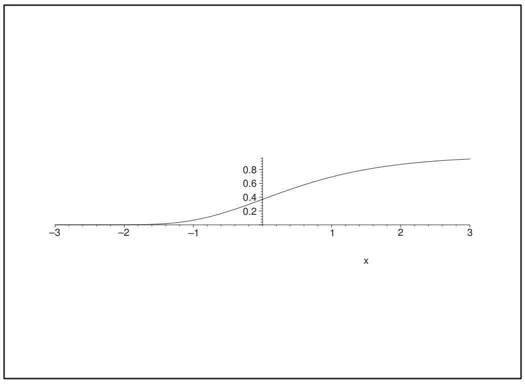


Figure 19: The graph of $F_Z(z) = \exp\left(-e^{-z}\right)$.

3) When (1) is generalized we get

$$P\{Y_n \le y\} = F(y)^n,$$

hence

$$P\{Z_n \le z\} = P\{Y_n \le z + \ln n\} = (F(z + \ln n))^n,$$

and whence

$$F_{Z_n}(z) = P\{Z_n \le z\} = \begin{cases} 0, & z \le -\ln n, \\ \left(1 - e^{-(z+\ln n)}\right)^n = \left(1 - \dfrac{1}{n}\, e^{-z}\right)^n & z > -\ln n. \end{cases}$$

Then for every fixed z,

$$\lim_{n\to\infty} P\{Z_n \le z\} = \lim_{n\to\infty} \left(1 - \frac{1}{n}\, e^{-z}\right)^n = \exp\left(-e^{-z}\right),$$

proving that the sequence (Z_n) converges in distribution towards a random variable Z of the distribution function

$$F_Z(z) = \exp\left(-e^{-z}\right), \qquad z \in \mathbb{R}.$$

Remark 8.1 We have above tacitly applied the well-known result

$$\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a \qquad \text{for } a \in \mathbb{R}, \qquad \Diamond$$

It is easily seen that $F_Z(z) = \exp\left(-e^{-z}\right)$ is increasing and continuous and

$$\lim_{z\to\infty} F_Z(z) = 0 \quad \text{and} \quad \lim_{z\to\infty} F_Z(z) = 1,$$

so $F_Z(z)$ is indeed a distribution function of a random variable Z. $\Diamond$

Example 8.9 *Let X_1, X_2, ... be independent random variables, all Cauchy distributed of the frequency*

$$f(x) = \frac{1}{\pi\left(1+x^2\right)}, \qquad x \in \mathbb{R}.$$

Let

$$Y_n = \max\left\{X_1, X_2, \dots, X_n\right\}, \quad Z_n = \frac{1}{n} Y_n, \qquad n \in \mathbb{N}.$$

1) Find the distribution function $G_n(z)$ of the random variable Z_n.

2) Prove that (Z_n) converges in distribution towards a random variable Z, and find the distribution function and the frequency of Z.

HINT: *It may be convenient to use the formula*

$$\text{Arctan } x + \text{Arctan } \frac{1}{x} = \frac{\pi}{2} \cdot \frac{x}{|x|}, \qquad x \neq 0.$$

1) The distribution function for each X_i is given by

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{dt}{1+t^2} = \frac{1}{\pi}\left[\text{Arctan } t\right]_{-\infty}^{x} = \frac{1}{\pi}\text{Arctan } x + \frac{1}{2}, \qquad x \in \mathbb{R}.$$

Thus

$$\begin{aligned} G_n(z) &= P\left\{\frac{1}{n}Y_n \le z\right\} = P\left\{Y_n \le nz\right\} = P\left\{\max\left\{X_1, \dots, X_n\right\} \le nz\right\} \\ &= \left(P\left\{X_1 \le nz\right\}\right)^n = \left(\frac{1}{2} + \frac{1}{\pi}\text{Arctan } nz\right)^n \quad (> 0). \end{aligned}$$

2) If $z \le 0$, then Arctan $nz \le 0$, hence

$$G_n(z) = \left(\frac{1}{2} + \frac{1}{\pi}\text{Arctan } nz\right) \le \frac{1}{2^n} \to 0 \qquad \text{for } n \to \infty.$$

If $z > 0$, then we use

$$\frac{1}{\pi}\text{Arctan}(nz) = \frac{1}{2} - \frac{1}{\pi}\text{Arctan } \frac{1}{nz},$$

to conclude that

$$G_n(z) = \left(1 - \frac{1}{\pi}\text{Arctan } \frac{1}{nz}\right)^n,$$

and

$$\begin{aligned} \ln G_n(z) &= n \ln\left\{1 - \frac{1}{\pi}\text{Arctan } \frac{1}{nz}\right\} = n\left\{-\frac{1}{\pi}\text{Arctan } \frac{1}{nz} - \frac{1}{nz}\,\varepsilon\left(\frac{1}{nz}\right)\right\} \\ &= -\frac{n}{\pi}\left\{\frac{1}{nz} + \frac{1}{nz}\,\varepsilon\left(\frac{1}{nz}\right)\right\} = -\frac{1}{\pi z} - \frac{1}{\pi z}\,\varepsilon\left(\frac{1}{nz}\right) \to -\frac{1}{\pi z} \qquad \text{for } n \to \infty. \end{aligned}$$

The distribution function is

$$G(z) = \begin{cases} \exp\left(-\dfrac{1}{\pi z}\right) & \text{for } z > 0, \\ 0 & \text{for } z \leq 0, \end{cases}$$

and the frequency is

$$g(z) = \begin{cases} \dfrac{1}{\pi z^2} \exp\left(-\dfrac{1}{\pi z}\right) & \text{for } z > 0, \\ 0 & \text{for } z \leq 0. \end{cases}$$

Example 8.10 *Let X and Y be independent random variables, where X is exponentially distributed of the frequency*

$$f_X(x) = \begin{cases} 2\,e^{-2x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

and Y is rectangularly distributed over the interval $]0,3[$.

1) *Find the mean and the variance for each of the three random variables X, Y and $Z = X + Y$.*

2) *Find the frequency of the random variable Z.*

3) *Now assume that X and Y_n are independent random variables, where X has the same distribution as above, while Y_n is rectangularly distributed over the interval $\left]0, \frac{1}{n}\right[$, $n \in \mathbb{N}$. Find for $z > \frac{1}{n}$, the distribution function $F_n(z)$ of the random variable $Z_n = X + Y_n$.*

4) *Find $\lim_{n\to\infty} F_n(z)$ for every $z \in \mathbb{R}$.*

1) Clearly,

$$E\{X\} = \int_0^\infty x \cdot 2e^{-2x}\,dx = \frac{1}{2}\int_0^\infty t\,e^{-t}\,dt = \frac{1}{2},$$

and since

$$E\left\{X^2\right\} = \int_0^\infty x^2 \cdot 2e^{-2x}\,dx = \frac{1}{4}\int_0^\infty t^2 e^{-t}\,dt = \frac{1}{4}\cdot 2! = \frac{1}{2},$$

it follows that

$$V\{X\} = E\left\{X^2\right\} - (E\{X\})^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

It follows from

$$f_Y(y) = \begin{cases} \frac{1}{3} & \text{for } x \in]0,3[, \\ 0 & \text{otherwise,} \end{cases}$$

that

$$E\{Y\} = \frac{1}{3}\int_0^3 y\,dy = \frac{1}{3}\left[\frac{y^2}{2}\right]_0^3 = \frac{1}{3}\cdot\frac{9}{2} = \frac{3}{2},$$

and

$$E\left\{Y^2\right\} = \frac{1}{3}\int_0^3 y^2\,dy = \frac{1}{3}\left[\frac{y^3}{3}\right]_0^3 = 3,$$

hence

$$V\{Y\} = E\left\{Y^2\right\} - (E\{Y\})^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

Remark 8.2 All results above are of course well-known, so the computations are strictly speaking not necessary. They are given here for completeness. $\Diamond$

Finally,

$$E\{Z\} = E\{X+Y\} = E\{X\} + E\{Y\} = \frac{1}{2} + \frac{3}{2} = 2,$$

and

$$V\{Z\} = V\{X\} + V\{Y\} = \frac{1}{4} + \frac{3}{4} = 1.$$

2) The frequency of Z is 0 for $z \le 0$. When $z > 0$, then

$$f_Z(z) = \int_0^{\infty} f_X(t)\, g_Y(z-t)\, dt.$$

The integrand is $\neq 0$, when $t > 0$ and $z - t \in]0,3[$, i.e. when $t \in]z-3, z[$.

a) If $z \in]0,3[$, then $z - 3 < 0$, hence

$$f_Z(z) = \int_0^z 2e^{-2t} \cdot \frac{1}{3}\, dt = \frac{1}{3} \left[-e^{-2t}\right]_0^z = \frac{1}{3} \left(1 - e^{-2z}\right).$$

b) If $z \ge 3$, then

$$f_Z(z) = \int_{z-3}^z 2e^{-2t} \cdot \frac{1}{3}\, dt = \frac{1}{3} \left[-e^{-1t}\right]_{z-3}^z = \frac{1}{3} \left(e^6 - 1\right) e^{-2z}.$$

Summing up,

$$f_Z(z) = \begin{cases} 0 & \text{for } z \le 0, \\ \frac{1}{3}\left(1 - e^{-2z}\right) & \text{for } 0 < z < 3, \\ \frac{1}{3}\left(e^6 - 1\right) e^{-2z} & \text{for } z \ge 3. \end{cases}$$

3) The frequency of Y_n is

$$f_{Y_n}(y) = \begin{cases} n & \text{for } y \in \left]0, \frac{1}{n}\right[, \\ 0 & \text{otherwise.} \end{cases}$$

If $z > \frac{1}{n}$, then the frequency of Z_n is given by

$$f_n(z) = \int_0^{\infty} f_X(t)\, f_{Y_n}(z-t)\, dt = \int_{z-\frac{1}{n}}^z 2e^{-2t}\, n\, dt = n \left[-e^{-2t}\right]_{z-\frac{1}{n}}^z = n\left\{e^{\frac{2}{n}} - 1\right\} e^{-2z}.$$

We conclude for $z > \dfrac{1}{n}$ that the distribution function is

$$\begin{aligned} F_n(z) &= \int_{-\infty}^{z} f_{Z_n}(t)\,dt = 1 - \int_{z}^{\infty} f_{Z_n}(t)\,dt = 1 - n\left\{e^{\frac{2}{n}} - 1\right\}\int_z^\infty e^{-2t}\,dt \\ &= 1 - n\left\{e^{\frac{2}{n}} - 1\right\}\left[-\frac{1}{2}\,e^{-2t}\right]_z^\infty = 1 - \frac{n}{2}\left\{e^{\frac{2}{n}} - 1\right\}e^{-2z}. \end{aligned}$$

4) If $z < 0$, then $F_n(z) = 0$, hence $\lim_{n\to\infty} F_n(z) = 0$.
If $z > 0$, then there exists an N, such that $z > \dfrac{1}{n}$ for every $n \geq N$, so

$$\begin{aligned} \lim_{n\to\infty} F_n(z) &= \lim_{n\to\infty}\left\{1 - \frac{n}{2}\left(e^{\frac{2}{n}} - 1\right)e^{-2z}\right\} = 1 - e^{-2z}\lim_{t\to\infty}\frac{n}{2}\left(e^{\frac{2}{n}} - 1\right) \\ &= 1 - e^{-2z}\lim_{n\to\infty}\left\{\frac{n}{2}\left(1 + \frac{2}{n} + \frac{2}{n}\,\varepsilon\left(\frac{2}{n}\right)\right) - 1\right\} = 1 - e^{-2z} = F_X(z). \end{aligned}$$

Example 8.11 *Let X_n, $n \in \mathbb{N}$, and X be random variables, and let a_n, $n \in \mathbb{N}$, and a be positive numbers. Prove that if the sequence (X_n) converges in distribution towards X, and the sequence (a_n) converges towards a, then the sequence (a_nX_n) converges in distribution towards aX.*

Let $F_n(x)$ be the distribution functions of X_n and $F(x)$ the distribution function of X. Let $G_n(y)$ be the distribution functions of $Y_n = a_nX_n$, and $G(y)$ the distribution function of $Y = aX$.

The assumptions are that $a_n > 0$ and $a > 0$, and

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{and} \quad \lim_{n\to\infty} a_n = a.$$

We prove that at any point of continuity y,

$$\lim_{n\to\infty} G_n(y) = G(y).$$

First rewrite in the following way,

$$\begin{aligned} G_n(y) &= P\{Y_n \leq y\} = p\{a_nX_n \leq y\} = P\left\{X_n \leq \frac{y}{a_n}\right\} = F_n\left(\frac{y}{a_n}\right) \\ &= F\left(\frac{y}{a}\right) + \left\{F_n\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a_n}\right)\right\} + \left\{F\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a}\right)\right\} \\ &= P\left\{X \leq \frac{y}{a}\right\} + \left\{F_n\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a_n}\right)\right\} + \left\{F\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a}\right)\right\} \\ &= P\{Y \leq y\} + \left\{F_n\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a_n}\right)\right\} + \left\{F\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a}\right)\right\}, \end{aligned}$$

thus

$$|G_n(y) - G(y)| \leq \left|F_n\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a_n}\right)\right| + \left|F\left(\frac{y}{a_n}\right) - F\left(\frac{y}{a}\right)\right|.$$

If $\dfrac{y}{a}$ is a point of continuity of F, then the right hand side will converge towards 0 for $n \to \infty$, and the claim is proved.

Alternatively we know that at the points of continuity $x \in \mathbb{R}$ of $F(x)$ we have the limit

$$\lim_{n\to\infty} P\left\{X_n \le x\right\} = P\{X \le x\} = F(x).$$

Let a_n and a be positive numbers, where $a_n \to a$, and let $\dfrac{x}{a}$ be a point of continuity of $F(x)$. Then

$$P\left\{a_n X_n \le x\right\} = P\left\{X_n \le \frac{x}{a_n}\right\}.$$

Choose any $\varepsilon > 0$. If $n \ge n(x,\varepsilon)$,, then

$$P\left\{X_n \le \frac{x-\varepsilon}{a}\right\} \le P\left\{X_n \le \frac{x}{a_n}\right\} \le P\left\{X_n \le \frac{x+\varepsilon}{a}\right\}.$$

Then restrict $\varepsilon > 0$, such that also $\dfrac{x-\varepsilon}{a}$ and $\dfrac{x+\varepsilon}{a}$ are points of continuity of F. (Here we exploit that since F is weakly monotonous, F has at most a countably many points of discontinuity, so this can always be obtained for ε "as small as we want it"). Letting $n \to \infty$, we get

$$P\left\{X \le \frac{x-\varepsilon}{a}\right\} \le \liminf_{n\to\infty} P\left\{X_n \le \frac{x}{a_n}\right\} \le \limsup_{n\to\infty} P\left\{X_n \le \frac{x}{a_n}\right\} \le P\left\{X \le \frac{x+\varepsilon}{a}\right\}.$$

If $\varepsilon \to 0$, then two of the terms will both tend towards

$$P\left\{X \le \frac{x}{a}\right\} = P\{a\,X \le x\},$$

and we have proved that

$$\lim_{n\to\infty} P\left\{a_n X_n \le x\right\} = \lim_{n\to\infty} P\left\{X_n \le \frac{x}{a_n}\right\} = P\{aX \le x\}.$$

Index