Jan Długosz University in Częstochowa
Scientific Issues, Mathematics XVI, Częstochowa 2011

CONVOLUTE AND GEOMETRICAL PROBABILITY SPACES

Maciej Major

Institute of Mathematics, Pedagogical University of Cracow
Podchorążych 2, 30-084 Cracow, Poland
e-mail: mmajor@up.krakow.pl

Abstract. Let X and Y be two independent random variables, either discrete or continuous. The question is "what is the probability distribution of $Z = X + Y$"? Clearly, the probability distribution of $Z = X + Y$ is some combination of f_X and f_Y which is called the convolution of f_X and f_Y. It is denoted by $*$. We have $f_Z(t) = f_{X+Y}(t) = f_X(t)*f_Y(t)$. In this paper it is shown how we can use geometrical probability spaces to find (without convolution) the distribution of random variable $Z = X + Y$.

1. Introduction

A random variable is one of important notions of the probability calculus. Of particular importance are continuous random variables because they have applications in mathematical statistics, economics, theory of insurance, and physics. Mathematical tools used for examining these random variables are rather complicated (the characteristic function, the Riemann–Stieltjes integral, the notion of the functional convolution). In the works [2], [3] and [4] it is presented how it is possible to use geometrical probability space for examining continuous random variables. In this work we suggest a method of finding the cumulative distribution function and the density function of the sum of independent continuous random variables, with the use of the geometrical probability space.

2. Basic definitions

To begin with, we recall definitions and theorems which are essential for subsequent part of this work.

Definition 1. *Let* $(\Omega, \mathcal{Z}, P)$ *be any probability space. A random variable in this probability space is defined as any function* X *from the set* Ω *in* $\mathbb{R}$ *that satisfies the condition:*

$$\{\omega \in \Omega : X(\omega) < x\} \in \mathcal{Z} \text{ for any } x \in \mathbb{R}. \tag{1}$$

Theorem 1. *If* X *is a random variable in the probability space* $(\Omega, \mathcal{Z}, P)$ *and* $\mathcal{B}$ *is the set of Borel subsets of a straight line and* P_X *is a function defined by formula:*

$$P_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}) \text{ for any } A \in \mathcal{B}, \tag{2}$$

then the triple $(\mathbb{R}, \mathcal{B}, P_X)$ *is also the probability space.*

Definition 2. *Let* X *be a random variable in the probability space* $(\Omega, \mathcal{Z}, P)$*. The function* P_X *defined by formula (2) on the set of Borel subsets of a straight line is called the probability generated on a straight line by the random variable* X *or the distribution of the random variable* X*, and the triple* $(\mathbb{R}, \mathcal{B}, P_X)$ *is called the probability space generated on the straight line by the random variable* X*.*

Definition 3. *A function* F_X *defined on* $\mathbb{R}$ *by the formula*

$$F_X(x) = P_X((-\infty, x)) \text{ for any } x \in \mathbb{R}$$

is called the cumulative distribution function (also the cumulative density function) or briefly the distribution function of a random variable X*.*

Definition 4. *A random variable* X*, for which there exists such a nonnegative and integrable function* f_X *defined on* $\mathbb{R}$ *that*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt,$$

is called continuous, and its distribution P_X *is called a continuous distribution. The function* f_X *is called the density of a random variable* X *or the density of a distribution* P_X*.*

Definition 5. *Random variables* $X_1, X_2, \ldots, X_n$ *from the same probability space* $(\Omega, \mathcal{Z}, P)$ *are called independent if for any Borel sets* $B_1, B_2, \ldots,$ B_n *on a straight line, the events* $A_1, A_2, \ldots, A_n$*, where* $A_j = \{\omega \in \Omega \colon X_j(\omega) \in B_j\}$ *for* $j = 1, 2, \ldots, n$*, satisfy the following condition:*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n).$$

Definition 6. *Let Ω be a subset of k-dimensional Euclidean space ($k = 1, 2, 3, \dots$) having the positive k-dimensional Lebesgue measure, let $\mathcal{Z}$ be a set of subsets of the Ω set having the Lebesgue measure and let P be a function defined on $\mathcal{Z}$ by the formula:*

$$P(A) = \frac{m_l(A)}{m_l(\Omega)}, \text{ where } m_l \text{ denotes the Lebesgue measure.} \tag{3}$$

The triple $(\Omega, \mathcal{Z}, P)$ is called the geometric probability space, and P is called the geometric probability.

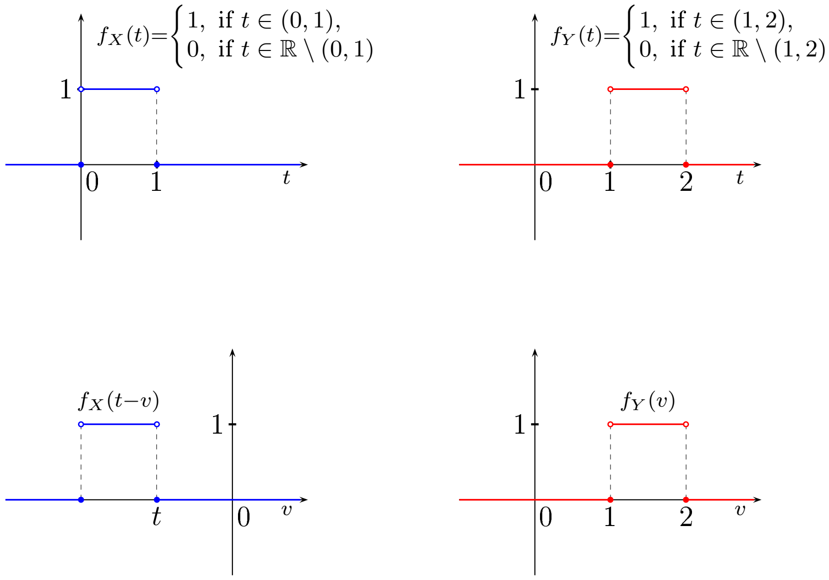
3. The sum of two independent uniform random variables – the classical method

Let X and Y be independent random variables with uniform distributions and let

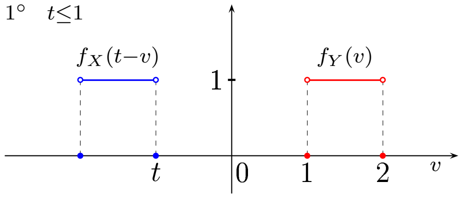
$$f_X(t) = \begin{cases} 1 \text{ if } 0 < t < 1, \\ 0 \text{ otherwise,} \end{cases} \qquad f_Y(t) = \begin{cases} 1 \text{ if } 1 < t < 2, \\ 0 \text{ otherwise.} \end{cases}$$

Let $Z = X + Y$.

The graphs below illustrate the method of determining the density function of the random variable Z.

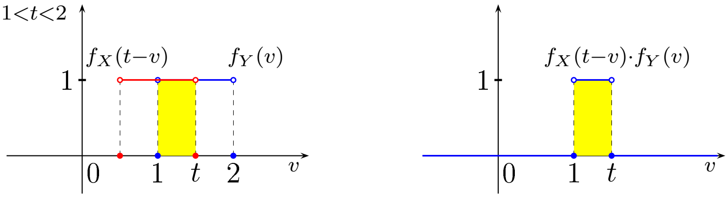


1° $t \leq 1$



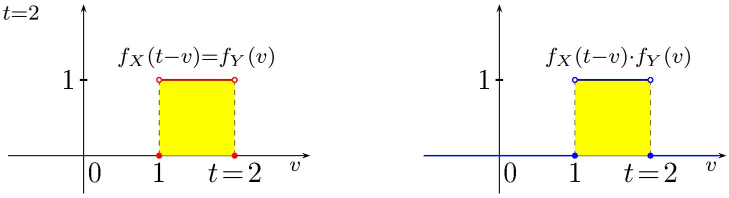
$$f_X(t-v)\cdot f_Y(v) = 0 \Rightarrow f_X * f_y(t) = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v)dv = 0$$

2° $1<t<2$



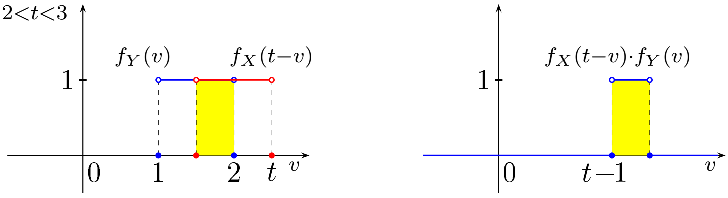
$$f_X * f_Y(t) = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v)dv = (t-1)\cdot 1 = t-1$$

3° $t=2$



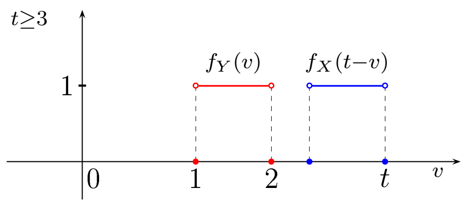
$$f_X * f_Y(t) = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v)dv = (2-1)\cdot 1 = 1$$

4° $2<t<3$



$$f_X * f_Y(t) = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v)dv = (2-(t-1))\cdot 1 = -t+3$$

5° $t \geq 3$



$$f_X(t-v)\cdot f_Y(v)=0 \Rightarrow f_X * f_Y(t) = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v)dv = 0$$

We have:

$$f_Z(t) = \begin{cases} 0 & \text{for } t \leq 1 \vee t \geq 3, \\ t-1 & \text{for } 1 < t < 2, \\ -t+3 & \text{for } 2 < t < 3. \end{cases}$$

4. The sum of two independent uniform random variables – the alternative method

Let us consider independent continous random variables X and Y with the density functions f_X and f_Y. Let

$$\Omega^{XY} := \{(x, y, z) \in \mathbb{R}^3 \colon 0 \leq z \leq f_X(x) f_Y(y)\}$$

and let $\mathcal{Z}$ be a family of subset of the set Ω^{XY} having the Lebesque measure. Let us notice that $m_l(\Omega) = 1$. The triple $(\Omega, \mathcal{Z}, P)$, where $P(A) = m_l(A)$, is a geometrical probability space. The probability space $(\Omega, \mathcal{Z}, P)$ will be called the *basic geometrical probability space of independent random variables X and Y*.

Let X and Y be independent random variables with uniform distributions and

$$f_X(t) = \begin{cases} 1 \text{ if } 0 < t < 1, \\ 0 \text{ otherwise}, \end{cases} \qquad f_Y(t) = \begin{cases} 1 \text{ if } 1 < t < 2, \\ 0 \text{ otherwise}. \end{cases}$$

Let $Z = X + Y$,

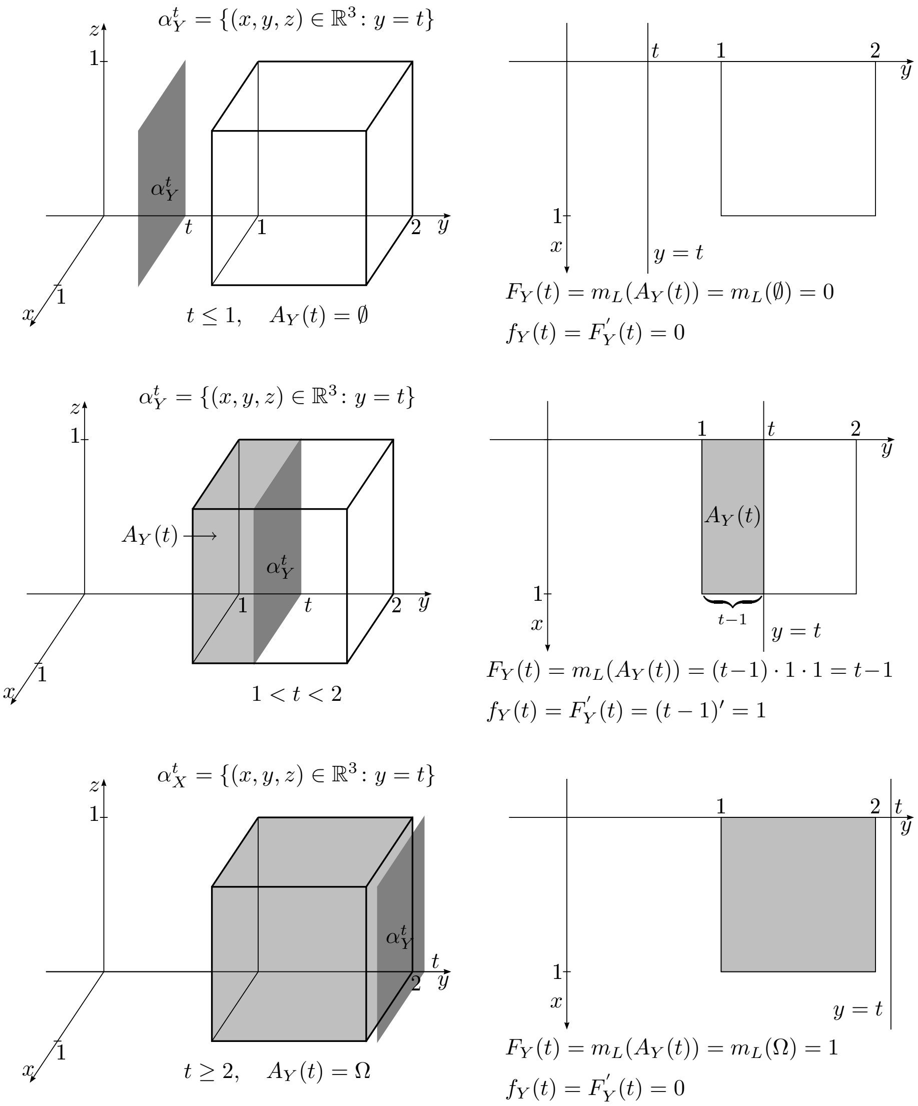
$$\Omega^{XY} = \{(x, y, z) \in \mathbb{R}^3 \colon 0 < x < 1,\ 1 < y < 2,\ 0 \leq z \leq 1\} = (0,1) \times (1,2) \times [0,1],$$

and $X \colon \Omega \to \mathbb{R}$ be given by the formula $X(x, y, z) = x$, whereas $Y \colon \Omega \to \mathbb{R}$ be given by the formula $Y(x, y, z) = y$. Let us define

$$A_X(t) = \{(x, y, z) \in \Omega^{XY} \colon X(\omega) < t\} = \{(x, y, z) \in \Omega^{XY} \colon x < t\}$$

for $t \in \mathbb{R}$.

The graphs below illustrate the method of determining the distribution function of the random variable X.



The cumulative distribution function of a random variable X is expressed by the formula:

$$F_X(t) = \begin{cases} 0 \text{ if } x \leq 0, \\ t \text{ if } 0 < t < 1, \\ 1 \text{ if } t \geq 1. \end{cases}$$

Hence it follows that the density function of a random variable X is given by the formula:

$$f_X(t) = \begin{cases} 1 \text{ if } 0 < t < 1, \\ 0 \text{ if } t \leq 0 \vee t \geq 1. \end{cases}$$

Reasoning in a similar way, one may state that Y is a continuous random variable for which

$$f_Y(t) = \begin{cases} 1 \text{ if } 1 < t < 2, \\ 0 \text{ if } t \leq 1 \vee t \geq 2. \end{cases}$$

It may be easily shown that random variables X and Y are independent random variables.

Let us now consider a random variable $Z = X + Y$. We have:

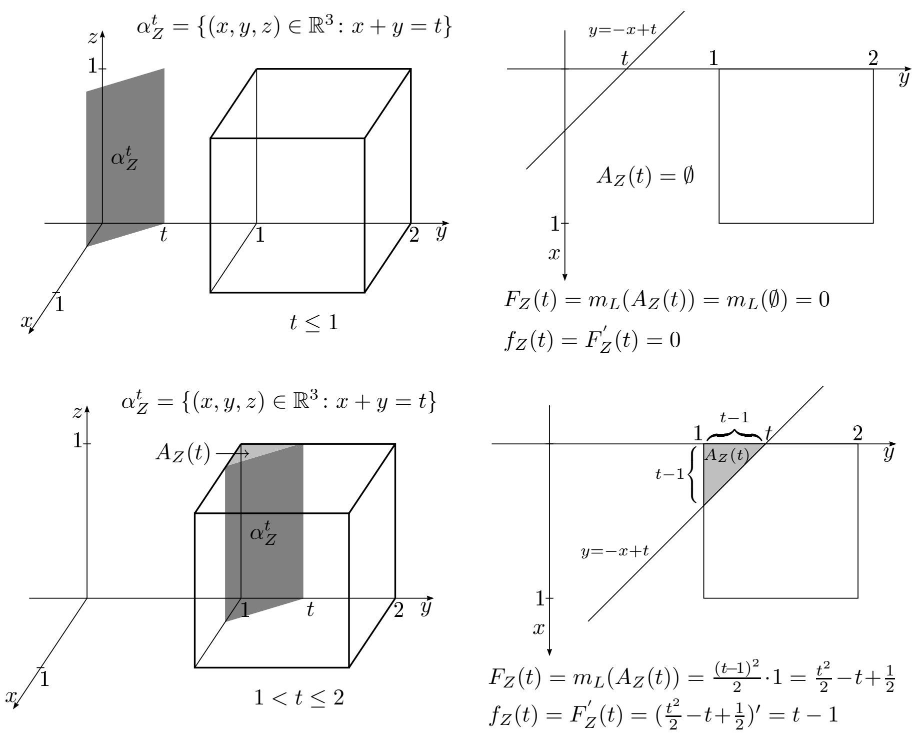
$$Z(x, y, z) = x + y \quad \text{for } (x, y, z) \in (0, 1) \times (1, 2) \times [0, 1].$$

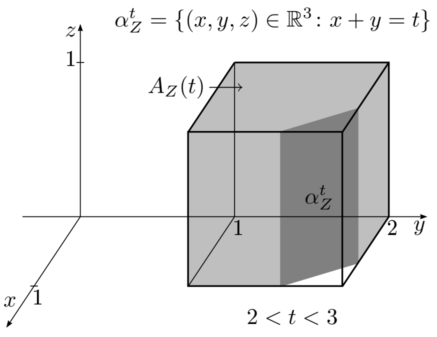
Now we determine the cumulative distribution function of the random variable Z. Let

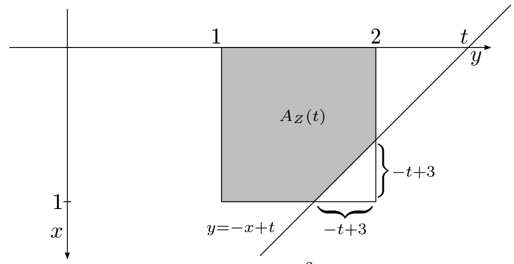
$$A_Z(t) = \{(x, y, z) \in (0, 1) \times (1, 2) \times [0, 1] \colon x + y < t\}$$

for $t \in \mathbb{R}$.

The graphs below illustrate the method of determining the distribution function of the random variable Z.

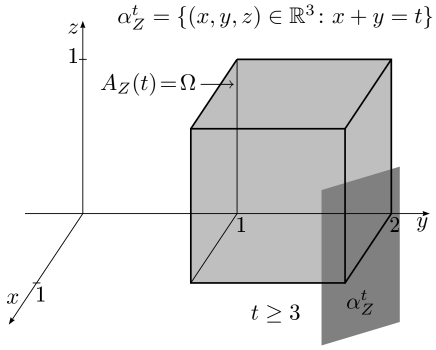


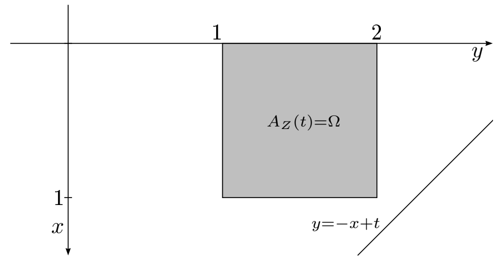




$F_Z(t)=m_L(A_Z(t))=(1-\frac{(-t+3)^2}{2})\cdot 1=\frac{-t^2}{2}+3t-\frac{7}{2}$

$f_Z(t) = F_Z'(t) = (\frac{-t^2}{2}+3t-\frac{7}{2})' = -t+3$





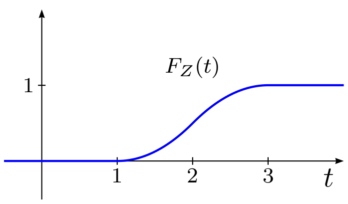
$F_Z(t) = m_L(A_Z(t)) = m_L(\Omega) = 1$

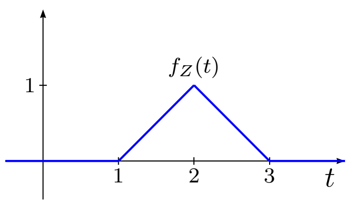
$f_Z(t) = F_Z'(t) = 1' = 0$

We have:

$$F_Z(x) = \begin{cases} 0 & \text{for } t \leq 1, \\ \frac{t^2}{2} - t + \frac{1}{2} & \text{for } 1 < t \leq 2, \\ \frac{-t^2}{2} + 3t - \frac{7}{2} & \text{for } 2 < t < 3, \\ 1 & \text{for } t \geq 3, \end{cases}$$

$$f_Z(t) = \begin{cases} 0 & \text{for } t \leq 1 \vee t \geq 3, \\ t-1 & \text{for } 1 < t < 2, \\ -t+3 & \text{for } 2 < t < 3. \end{cases}$$





5. Conclusion

It is worthwhile to solve the problems presented above with the students at mathematics teachers training majors. Proving the theorems with elementary methods with the use of mathematical analysis and geometrical methods allows us to consider the elements of probability calculus in a different (than traditional) way. Quite elementary tools make the presented problems simple to understand and to operative use.

References

[1] M. Krzyśko. *Wykłady z teorii prawdopodobieństwa*. Wydawnictwa Naukowo-Techniczne, Warszawa 2000.

[2] M. Major. Geometric probability space as the tool of examining continuous random variables. *Acta Mathematica, Nitra*, **11**, 147–152, 2008.

[3] M. Major. Continuous random variable. *Scientific Issues, Catholic University in Ružomberok, Mathematica*, **II**, 41–49, 2009.

[4] M. Major. Continuous random variables and geometrical probability space. *Acta Mathematica, Nitra*, **12**, 165–170, 2009.

[5] A. Płocki. *Stochastyka dla nauczyciela. Rachunek prawdopodobieństwa, kombinatoryka i statystyka matematyczna jako nauka "in statu nascendi"*. Wydawnictwo naukowe NOVUM, Płock 2002.