Jan Długosz University in Częstochowa
Scientific Issues, Mathematics XIV, Częstochowa 2009

GENERALIZATION OF PROBABILITY DENSITY OF RANDOM VARIABLES

Marcin Ziółkowski

Institute of Mathematics and Computer Science
Jan Długosz University in Częstochowa
al. Armii Krajowej 13/15, 42-200 Częstochowa, Poland
e-mail: m.ziolkowski@ajd.czest.pl

Abstract
In this paper we present generalization of probability density of random variables. It is obvious that probability density is definite only for absolute continuous variables. However, in many practical applications we need to define the analogous concept also for variables of other types. It can be easily shown that we are able to generalize the concept of density using distributions, especially Dirac's delta function.

1. Introduction

Let us consider a functional sequence $f_n : R \to [0, +\infty)$ given by the formula (Fig. 1):

$$f_n(x) = \begin{cases} n \quad \text{for } x \in [-\frac{1}{2n}, \frac{1}{2n}] \\ 0 \quad \text{for } x \notin [-\frac{1}{2n}, \frac{1}{2n}] \end{cases}, \quad n = 1, 2, \ldots .$$

It is clear that every term of this sequence can be treated as a probability density of some absolute continuous random variable. Suitable distribution functions are given by the following formula (Fig. 2):

$$F_n(x) = \begin{cases} 0 \quad \text{for } x < -\frac{1}{2n} \\ nx + \frac{1}{2} \quad \text{for } x \in [-\frac{1}{2n}, \frac{1}{2n}] \\ 1 \quad \text{for } x > \frac{1}{2n} \end{cases}, \quad n = 1, 2, \ldots .$$

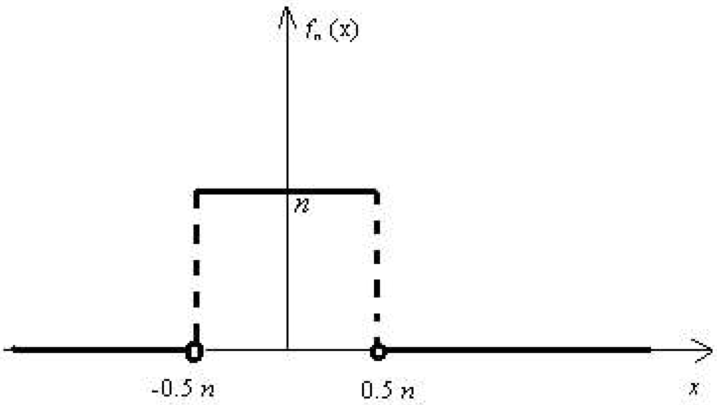

Fig. 1.

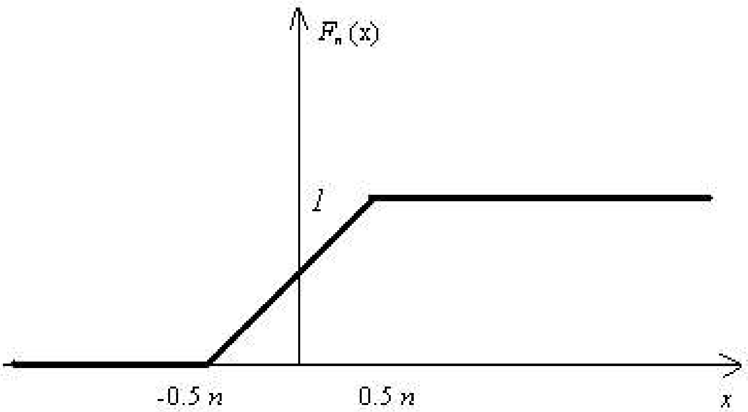

Fig. 2.

Distribution given by functions $f_n(x)$ or $F_n(x)$ $(n = 1, 2, \dots)$ is uniform on the interval $[-\frac{1}{2n}, \frac{1}{2n}]$.

Now we define a sequence a_n $(n = 1, 2, \dots)$ by the formula:

$$a_n = \int_{-\infty}^{\infty} f_n(x)\ dx.$$

The sequence a_n is constant ($a_n = 1$ for all $n \in N_+$), so $a_n \to 1$ when $n \to \infty$. On the other hand, the limit of the sequence $f_n(x)$ is not a real function but distribution. This limit is called Dirac's delta function $\delta(x)$.

$$\delta(x) := \lim_{n\to\infty} f_n(x).$$

By intuition, we understand Dirac delta function as below (Fig. 3):

$$\delta(x) = \begin{cases} 0 & \text{for } x \neq 0, \\ +\infty & \text{for } x = 0. \end{cases}$$

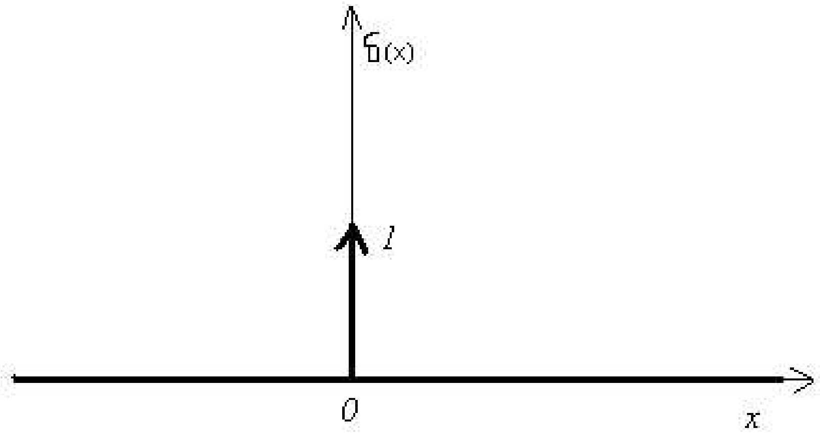

Fig. 3.

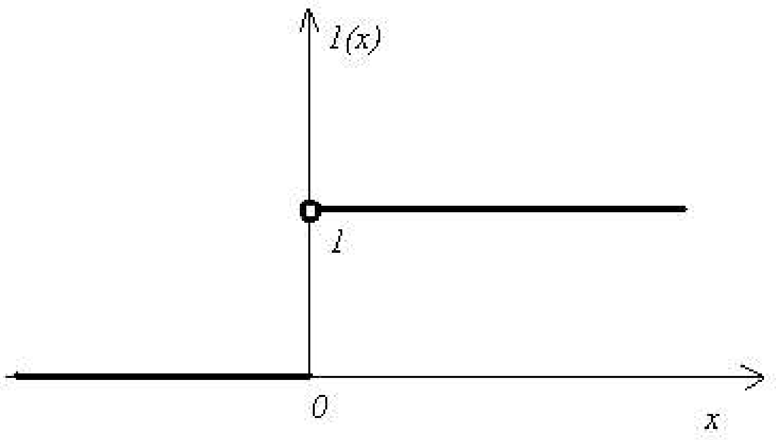

Fig. 4.

To show the properties of the sequences a_n and $f_n(x)$ let us treat Dirac's delta function as a density of a degenerate random variable (a discrete random variable which has only one value).

Let us note additionally that Dirac's delta function can be treated as a distributive derivative of Heaviside function $\mathbf{1}(x)$ (Fig. 4), which is an obvious distribution function of degenerate random variable.

2. Probability density for discrete random variables

Consider a discrete random variable, i.e a random variable which has finite or countable number of values x_k with probability p_k. Assume that $p_k > 0$ for all k and $\sum_k p_k = 1$. For such a specified random variable we can define the probability density by the following formula:

$$f(x) = \sum_k p_k \cdot \delta(x - x_k). \quad (1)$$

It is obvious that $f(x) \geq 0$ for all $x \in R$. After calculation, because of properties of a sequence a_n, we also obtain

$$\int_{-\infty}^{\infty} f(u)\, du = 1.$$

A diagram of such a specified density is shown in Fig. 5.

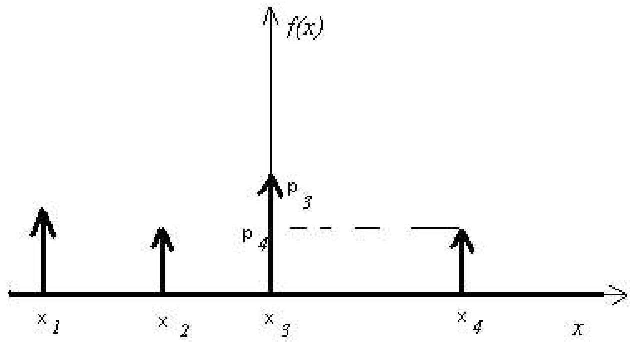

Fig. 5.

The distribution function of a discrete random variable can be defined using the Heaviside function by the formula (Fig. 6):

$$F(x) = \sum_k p_k \cdot \mathbf{1}(x - x_k). \quad (2)$$

We can now notice that $f(x) = \frac{\partial F(x)}{\partial x}$ in the sense of distributive derivative.

3. Generalized probability density for other random variables

It is clear that generalized probability density (using Dirac's delta function) can be defined for other types of random variables. Let us assume that a random variable has finite or countable number of jump points (the jump point

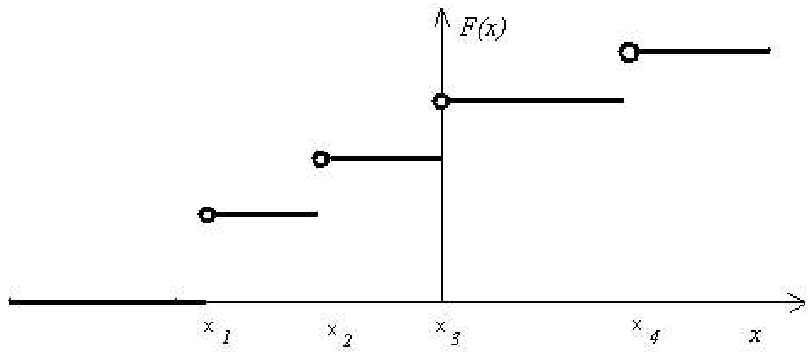

Fig. 6.

is the point of discontinuity x of distribution function $F(x)$ for which we have the inequality $F(x+\varepsilon) - F(x) > 0)$ for every $\varepsilon > 0$.

By this assumption, we can specify the probability density as a distributive derivative of a distribution function ($f(x) = F'(x)$).

Example.

TV advertising can be divided into three types. Half of them lasts from 0 to 1 minute (let us assume that this is a uniform distribution), 20% exactly 2 minutes (political sets) and the rest lasts from 3 to 4 minutes (also a uniform distribution).

Let us find formulas for distribution function and probability density for random variable which specify time of TV advertising. The distribution function is defined by the following formula (Fig. 7):

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 0.5x & \text{for } x \in [0,1] \\ 0.5 & \text{for } x \in (1,2] \\ 0.7 & \text{for } x \in (2,3) \\ 0.3x - 0.2 & \text{for } x \in [3,4] \\ 1 & \text{for } x > 4. \end{cases}$$

This formula can be simplied if we use the Heaviside function:

$$F(x) = \begin{cases} 0.5x & \text{for } x \in [0,1] \\ 0.3x - 0.2 & \text{for } x \in [3,4] \\ 0.5 \cdot \mathbf{1}(x-1) + 0.2 \cdot \mathbf{1}(x-2) + 0.3 \cdot \mathbf{1}(x-4) & \text{otherwise.} \end{cases}$$

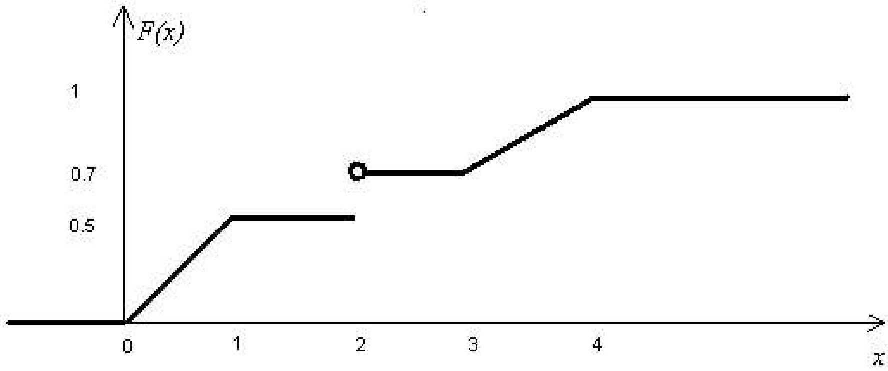

Fig. 7.

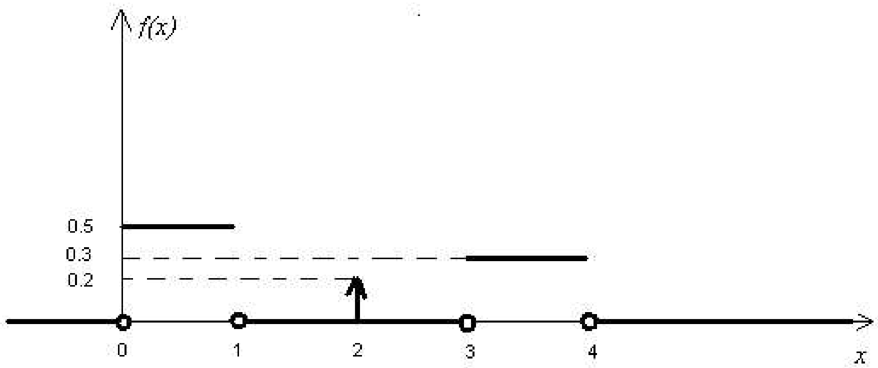

Fig. 8.

Probability density can be defined by the following formula:

$$f(x) = F'(x) = \begin{cases} 0.5 & \text{for } x \in [0,1] \\ 0.3 & \text{for } x \in [3,4] \\ 0.5 \cdot \delta(x-1) + 0.2 \cdot \delta(x-2) + 0.3 \cdot \delta(x-4) & \text{otherwise.} \end{cases}$$

But if $x \in (-\infty, 0) \cup (1,3) \cup (4, +\infty)$, then we have

$$0.5 \cdot \delta(x-1) + 0.2 \cdot \delta(x-2) + 0.3 \cdot \delta(x-4) = 0.2 \cdot \delta(x-2).$$

So finally we have:

$$f(x) = F'(x) = \begin{cases} 0.5 & \text{for } x \in [0,1] \\ 0.3 & \text{for } x \in [3,4] \\ 0.2 \cdot \delta(x-2) & \text{otherwise.} \end{cases}$$

A diagram of such a specified density is shown in Fig. 8.

4. Applications of generalized probability density

A generalized density of random variable is very useful in practical calculations. It can be easily shown that formulas, in which density of absolute continuous random variable is present, can be generalized for many other types of random variables. For example, we can use only one formula to calculate moments of random variables independently on their types. We can use the following formula:

$$E\xi^k = \int_{-\infty}^{\infty} x^k \ f(x) \ dx, \tag{3}$$

where $f(x)$ is a generalized density of random variable. In many cases it is a distribution, not a function (then we use Dirac's delta function to define it like in an example).

Let us calculate the first moment of the random variable presented in the example. Because of definition of generalized density $f(x)$, using integration by parts, we obtain

$$E\xi = \int_{-\infty}^{\infty} x \ f(x) \ dx = \int_{0}^{1} 0.5x \ dx + \int_{3}^{4} 0.3x \ dx + \int_{1}^{3} 0.2x \cdot \delta(x-2) \ dx =$$

$$= 0.25x^2\big|_{x=0}^{x=1} + 0.15x^2\big|_{x=3}^{x=4} + 0.2x \cdot \mathbf{1}(x-2)\big|_{x=1}^{x=3} - 0.2 \cdot \int_{1}^{3} \mathbf{1}(x-2) \ dx =$$

$$= 0.25 + 1.05 + 0.6 - 0.2 = 1.7.$$

We can note that the first moment can be obtained without using the generalized density but the result is exactly the same.

In queueing theory we often assume that a density of random variable exists to define the service intensity. In this case, service intensity is defined by the formula:

$$\mu(x) = \frac{f(x)}{1 - F(x)}, \tag{4}$$

where $f(x)$ and $F(x)$ are a density and a distribution function of service time, respectively.

The function $\mu(x)$ reduces calculations in many real models. Now we can use service intensity almost always, when we assume that $f(x)$ is a generalized density.

As it was shown, the concept of generalization of random variable probability density is very useful. It helps to reduce calculations in many practical problems.

References

[1] *Leksykon matematyczny.* Wiedza Powszechna, Warszawa 1993.

[2] H. Marcinkowska. *Dystrybucje, przestrzenie Sobolewa, równania różniczkowe.* PWN, Warszawa 1993.

[3] O. Tikhonenko. *Metody probabilistyczne analizy systemów informacyjnych.* EXIT, Warszawa 2006.