Generalization of probability density of random variables

Marcin Ziółkowski

Institute of Mathematics and Computer Science
Jan Dlugosz University of Częstochowa
al. Armii Krajowej 13/15, 42–201 Częstochowa, Poland
(e-mail: m.ziolkowski@ajd.czest.pl)

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 define analogous concept also for variables of the other types. It can be easily shown that we are able to generalize concept of density using distributions especially Dirac delta function.

1. Introduction

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

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

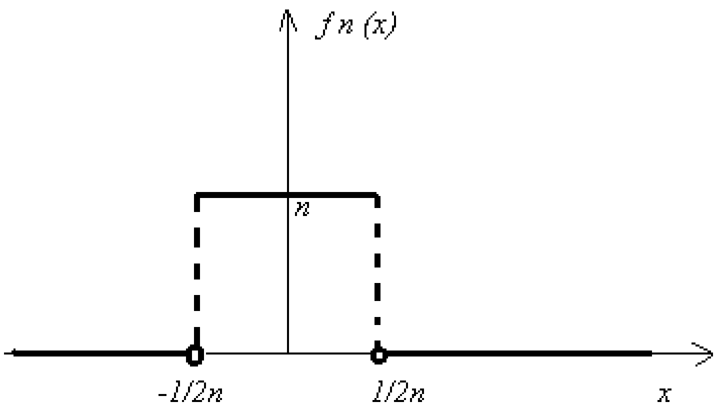

Fig.1.

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

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

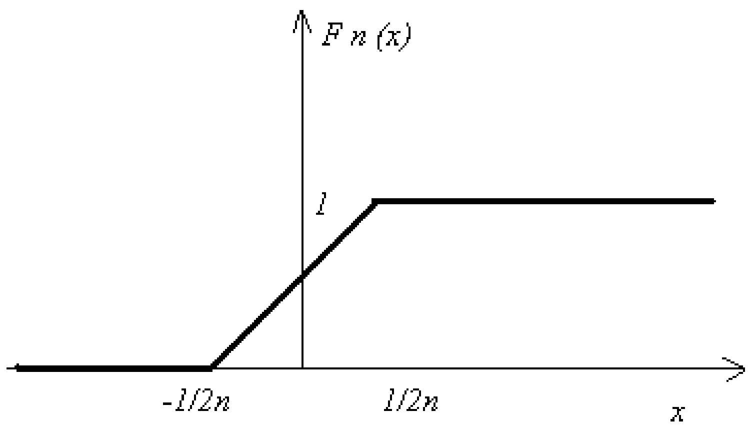

Fig.2.

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

Now we define a sequence a_n $(n = 1, 2, \ldots)$ 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 side the limit of sequence $f_n(x)$ is not a real function but distribution. This limit is called Dirac delta function – $\delta(x)$.

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

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

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

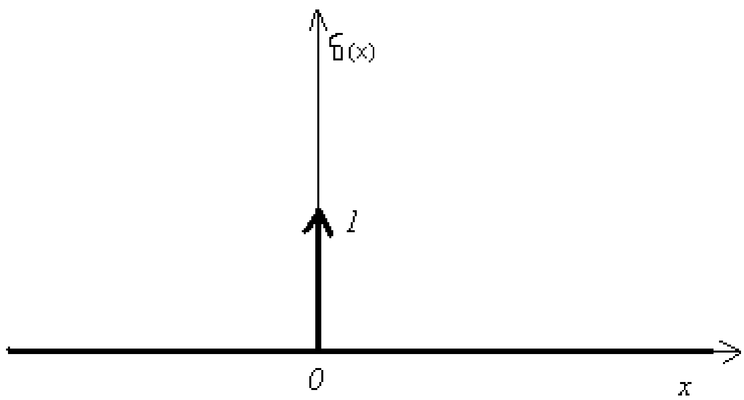

Fig.3.

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

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

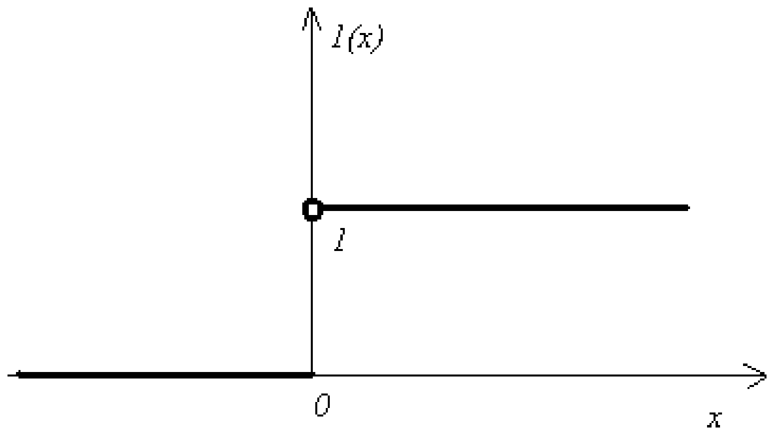

Fig.4.

2. Probability density for discrete random variables

Let us consider discrete random variable i.e 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 specified random variable we can define probability density by the following formula:

$$f(x) = \sum_k p_k \cdot \delta(x - x_k). \tag{1}$$

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

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

Diagram of such specified density is shown on Fig. 5.

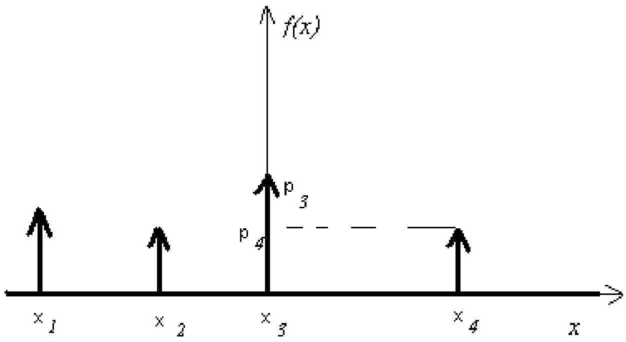

Fig.5.

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

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

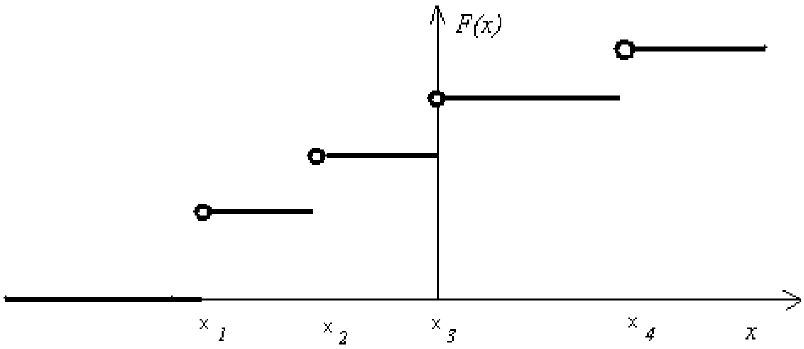

Fig.6.

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

3. Generalized probability density for the other random variables

It is clear that generalized probability density (using Dirac delta function) can be definited for the other types of random variables. Let us assume that random variable has finite or countable number of jump points (jump point is the point of discontinuity x of distribution function $F(x)$ for which we have for every $\varepsilon > 0$ inequality $F(x+\varepsilon) - F(x) > 0$).

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

Example.

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

Let us find formulas for distribution function and probability density for random variable which specify time of TV advertising.

Distribution function is defined by the following formula (Fig. 7):

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

Formula can be simplier if we use Heaviside function:

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

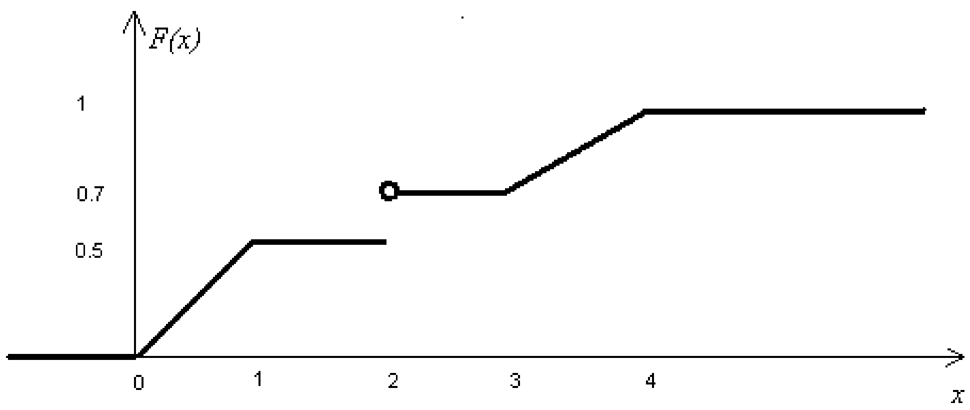

Fig.7.

Probability density can be defined by the formula:

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

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 following formula:

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

Diagram of such specified density is shown on Fig. 8.

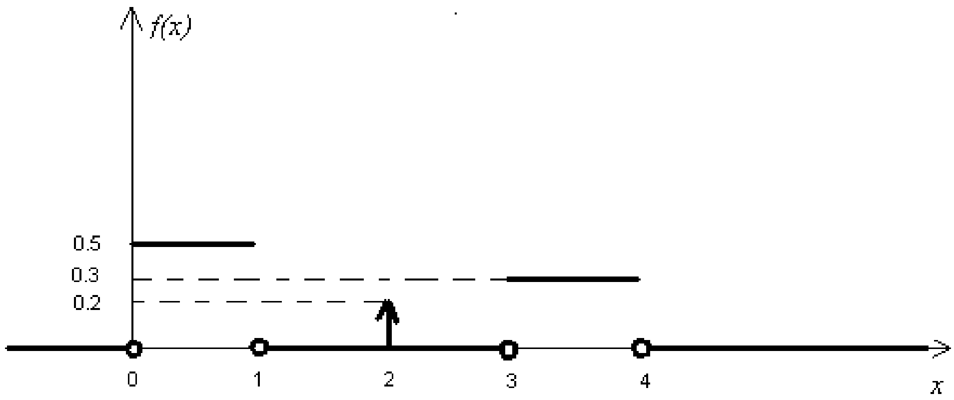

Fig.8.

4. Applications of generalized probability density

Generalized density of random variables 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 following formula:

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

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

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

$$E\xi = \int_{-\infty}^{\infty} x \; f(x) \; dx = \int_0^1 0.5x \; dx + \int_3^4 0.3x \; dx + \int_1^3 0.2x \cdot \delta(x-2) \; dx =$$

$$= 0.25x^2|_{x=0}^{x=1} + 0.15x^2|_{x=3}^{x=4} + 0.2x \cdot \mathbf{1}(x-2)|_{x=1}^{x=3} - 0.2 \cdot \int_1^3 \mathbf{1}(x-2) \; dx =$$

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

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

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

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

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

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

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

References

[1] Praca zbiorowa. 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.