Lecture 17. Probability

DR. YARASLAU ZADVORNY

Continuous random variables

Let us start the conversation about random variables which assume not necessarily integer but arbitrary real values.

Example. You and your friend have agreed to meet between noon and 1 p.m. The moment your friend appears is a random variable ξ which assumes the values in [0, 1].

Suppose my friend can come at any moment from the noon to the 1 p.m. with the same probabilities. Then we say that the random variable ξ is uniformly distributed in [0, 1].

Continuous random variables

We say that the random variable ξ is **uniformly distributed** in $[a, b]$ if, **intuitively**, it may assume any value from a to b with the same probabilities.

By the way, then the probability $P(\xi = x) = 0$ for any $x \in [a, b]$.

Indeed, suppose $P(\xi = x) > 0$. Then for some N, $P(\xi = x) > 1/N$.

Choose some numbers $x_1, x_2, \ldots, x_N \in [a, b]$.

Then $P(\xi \in \{x_1, x_2, \ldots, x_N\}) = P(\xi = x_1) + P(\xi = x_2) + \ldots + P(\xi = x_N) > 1$!!!

Continuous random variables: How to work with it?

The is a problem.

When we work with a discrete random variable, we can list all its possible values with their probabilities.

Say, if ξ is the result of a throw of a dice then all the possible results are {1, 2, 3, 4, 5, 6} and

$$P(\xi = 1) = P(\xi = 2) = \ldots = P(\xi = 6) = \frac{1}{6}.$$

The same is impossible if there are infinitely many possible values, say, if the set of values is the segment [0, 1].

Continuous random variables: How to work with it?

Suppose ξ and η are respectively the results of the first and the second throw of the dice, respectively.

I am interested in P(ξ + η = 5).

$$P(\xi+\eta=5)=P(\xi=1)\cdot P(\eta=4)+P(\xi=2)\cdot P(\eta=3)+$$
$$+P(\xi=3)\cdot P(\eta=2)+P(\xi=4)\cdot P(\eta=1)=$$
$$=\frac{1}{6}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{6}=\frac{4}{36}=\frac{1}{9}$$

Continuous random variables: How to work with it?

Now suppose the random variables ξ and η assume the values in the segment [0, 1] and we are interested in P(ξ + η =0.5). Then

$$P(\xi+\eta=0.5)=P(\xi=0)\cdot P(\eta=0.5)+P(\xi=0.1)\cdot P(\eta=0.4)+$$
$$+P(\xi=0.17)\cdot P(\eta=0.33)+P(\xi=0.06)\cdot P(\eta=0.44)+\ldots$$

So, there are infinitely many summands!

How to define the distribution?

Let random variables ξ and η be uniformly distributed in the segments [0, 1] and [0, 2] respectively.

Let us try to define the distribution as $P_\xi(x) = P(\xi = x)$.

For the both random variables $P_\xi(x) = P_\eta(x) = 0$!!!

How to define the distribution?

We define **probability density function** P_ξ on sets: $P_\xi(B) = P(\xi \in B)$.

Often, we can also define **cumulative distribution function** F_ξ(x) : R → [0, 1] as

$$F_\xi(x) = P(\xi < x).$$

(the function is defined on the set of numbers, not on the set of sets!!!)

Example. You and your friend have agreed to meet between noon and 1 p.m. The moment your friend appears is a random variable ξ which assumes the values in [0, 1].

$$P_\xi([c,d]) = d - c; \qquad F_\xi(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

A random variable (simplified a bit!!!)

Example. You and your friend have agreed to meet between noon and 1 p.m. The moment your friend appears is a random variable ξ which assumes the values in [0, 1].

$$P_\xi([c,d]) = d - c; \quad F_\xi(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

Unexpectable detail: if the probability of an event *A* equals to 0, it doesn't mean that this event will not happen!

Indeed, there probability of any point equals to 0.
But there is a point to be chosen!

(Continuous) uniform distribution

A random variable ξ is said to be uniformly distributed on a segment [a, b] if for any $[c,d]\subseteq[a,b]$

$$P(\xi\in[c,d])=P_{\xi}([c,d])=\frac{d-c}{b-a}$$

(Continuous) uniform distribution

A random variable ξ is said to be uniformly distributed on a segment [a, b] if for any $[c,d] \subseteq [a,b]$

$$P_{\xi}([c,d]) = \frac{d-c}{b-a}$$

If $x \in [a, b]$ then $F_{\xi}(x) = P(\xi < x) = P_{\xi}([a,x]) = \frac{x-a}{b-a}$

Thus,

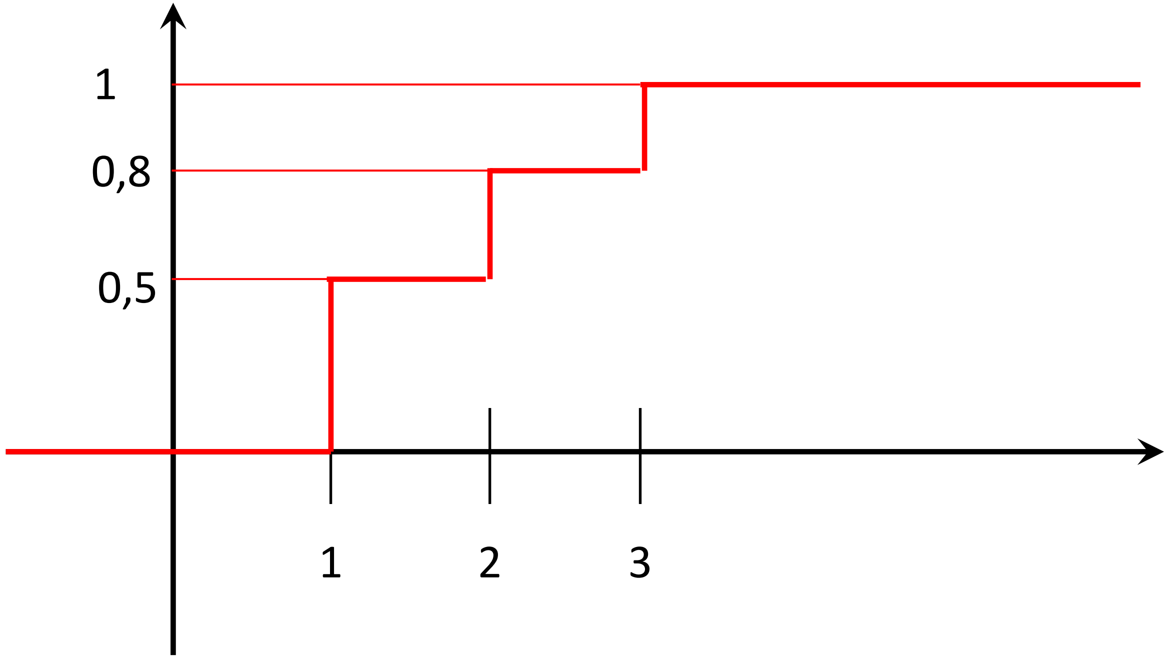
$$F_{\xi}(x) = \begin{cases} 0, x \le a \\ \frac{x-a}{b-a}, a < x < b, \\ 1, x \ge b \end{cases}$$

Example

Let P(ξ = 1) = 0.5, P(ξ = 2) = 0.3, P(ξ = 3) = 0.2.

Then

$$F_\xi(x) = \begin{cases} 0, x < 1 \\ 0.5, 1 \le x < 2 \\ 0.8, 2 \le x < 3 \\ 1, 3 \le x \end{cases}$$



Probability density function

We define **probability density function** P_ξ on sets: $P_\xi(B) = P(\xi \in B)$.

There is also a reason to define the probability density function as a function of *x*, not as a function of sets. It is defined as

$$P_\xi(x) = \left(F_\xi(x)\right)'$$

For a random variable ξ is said to be uniformly distributed on a segment [a, b] we have

$$F_\xi(x) = \begin{cases} 0, x \le a \\ \dfrac{x-a}{b-a}, a < x < b, \\ 1, x \ge b \end{cases} \qquad P_\xi(x) = \begin{cases} 0, x \notin [a,b] \\ \dfrac{1}{b-a}, x \in [a,b] \end{cases}$$

Probability density function

We define **probability density function** P_ξ on sets: $P_\xi(B) = P(\xi \in B)$.

There is also a reason to define the probability density function as a function of *x*, not as a function of sets. It is defined as

$$P_\xi(x) = \left(F_\xi(x)\right)'$$

Then, $P(\xi \in [a,b]) = \int_a^b P_\xi(x)dx$. Indeed,

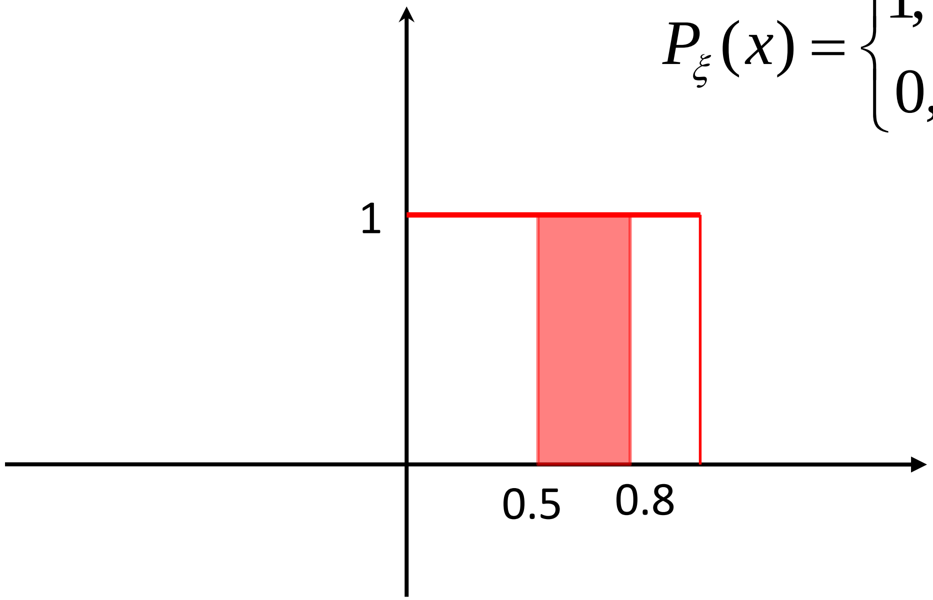
$$P(\xi \in [a,b]) = P(\xi < b) - P(\xi < a) = F_\xi(b) - F_\xi(a) =$$

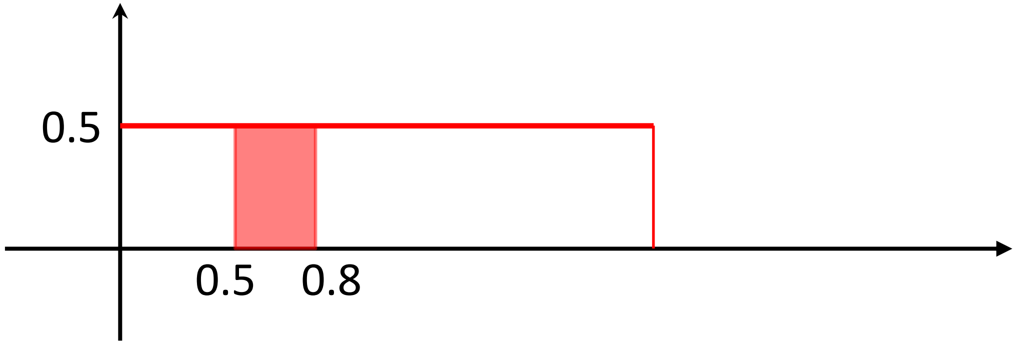
$$= \int_a^b \left(F_\xi(x)\right)' dx = \int_a^b P_\xi(x)dx$$

How to define the distribution?

Say, for $\Omega = [0, 1]$, $P([c, d]) = d - c$, $\xi(x) = x$ and $\eta(x) = 2x$ these functions are:

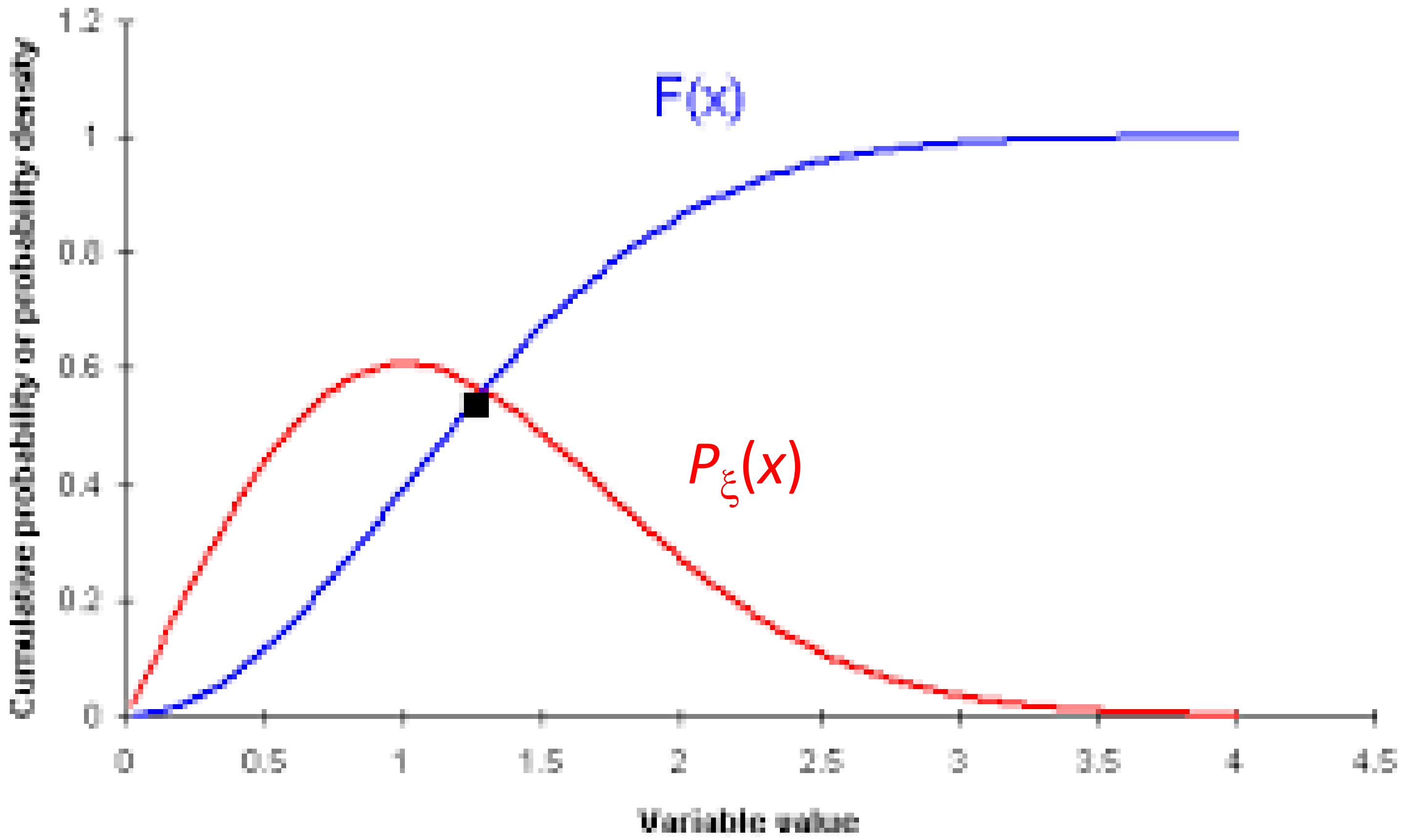
$$P_\xi(x) = \begin{cases} 1, x \in [0,1], \\ 0, x \notin [0,1] \end{cases}; P_\eta(x) = \begin{cases} 1/2, x \in [0,2], \\ 0, x \notin [0,2] \end{cases}$$





The area of the shadowed region is the probability of the set; but $P(x)$ still is not the same as $P(\xi = x)$!!!

How to define the distribution?



Mathematical Expectation

In this case, the mathematical expectation can be defined as

$$E\xi = \int_{-\infty}^{+\infty} xF_{\xi}'(x)dx$$

Mathematical Expectation

In this case, the mathematical expectation can be defined as

$$E\xi = \int_{-\infty}^{+\infty} xF_{\xi}'(x)dx = \int_{-\infty}^{+\infty} xP_{\xi}(x)dx$$

Compare to the discrete case:

$$E\xi = \sum_{k=1}^{n} x_k P(\xi = x_k)$$

$$E\xi = \int_{-\infty}^{+\infty} xF_{\xi}'(x)dx \approx \sum_{k=1}^{n} (x_k - x_{k-1}) x_k F_{\xi}'(x_k) \approx \sum_{k=1}^{n} (x_k - x_{k-1}) x_k \frac{F_{\xi}(x_k) - F_{\xi}(x_{k-1})}{x_k - x_{k-1}} =$$

$$= \sum_{k=1}^{n} x_k \left(F_{\xi}(x_k) - F_{\xi}(x_{k-1})\right) = \sum_{k=1}^{n} x_k P(x_{k-1} < x < x_k)$$

Mathematical Expectation

In this case, the mathematical expectation can be defined as

$$E\xi = \int_{-\infty}^{+\infty} xF_{\xi}'(x)dx$$

In particular, let ξ is uniformly distributed on [3. 5]. Recall that

$$F_{\xi}(x) = \begin{cases} 0, x \le 3 \\ \dfrac{x-3}{2}, 3 < x < 5, \\ 1, x \ge 5 \end{cases} \qquad F_{\xi}'(x) = \begin{cases} 0, x \le 3 \\ \dfrac{1}{2}, 3 < x < 5, \\ 0, x \ge 5 \end{cases}$$

$$E\xi = \int_{-\infty}^{+\infty} xF_{\xi}'(x)dx = \int_{3}^{5} x\frac{1}{2}dx = \left(\frac{x^2}{4}\right)_3^5 = \frac{25-9}{4} = 4.$$

Variance

Variance $D\xi$ is the measure of possible difference between ξ and $E\xi$.

For discrete random variable ξ which assumes the values from $\{x_1, x_2, \ldots, x_n\}$,

$$D\xi = \sum_{i=1}^{n} (x_i - E\xi)^2 P(\xi = x_i):$$

Variance

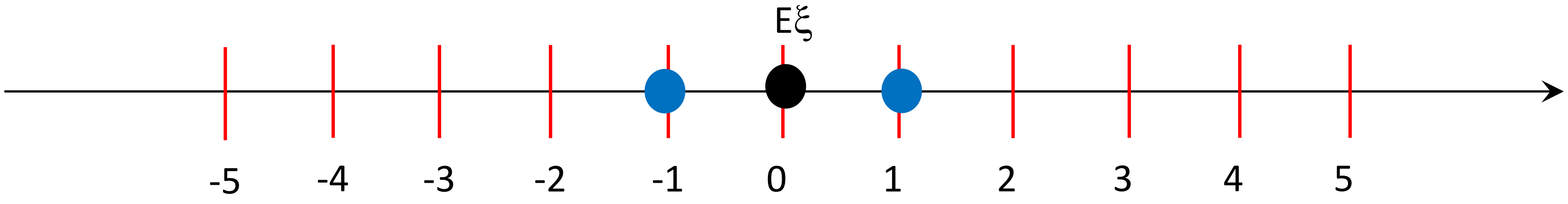
Variance Dξ is the measure of possible difference between ξ and Eξ.

For discrete random variable ξ which assumes the values from $\{x_1, x_2, \ldots, x_n\}$,

$$D\xi = \sum_{i=1}^{n} (x_i - E\xi)^2 P(\xi = x_i):$$

Let $P(\xi = 1) = 0.5$ and $P(\xi = -1) = 0.5$. Then $E\xi = 0$ and

$$D\xi = \frac{1}{2} \cdot (1-0)^2 + \frac{1}{2} \cdot (-1-0)^2 = 1$$



Variance

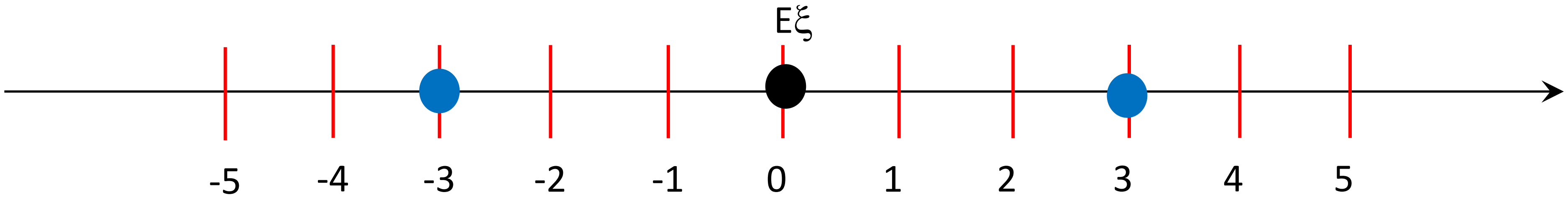
Variance Dξ is the measure of possible difference between ξ and Eξ.

For discrete random variable ξ which assumes the values from $\{x_1, x_2, \ldots, x_n\}$,

$$D\xi = \sum_{i=1}^{n} (x_i - E\xi)^2 P(\xi = x_i):$$

Let $P(\xi = 3) = 0.5$ and $P(\xi = -3) = 0.5$. Then $E\xi = 0$ and

$$D\xi = \frac{1}{2} \cdot (3-0)^2 + \frac{1}{2} \cdot (-3-0)^2 = 9$$



Variance

Variance Dξ is the measure of possible difference between ξ and Eξ.

For discrete random variable ξ which assumes the values from $\{x_1, x_2, \ldots, x_n\}$,

$$D\xi = \sum_{i=1}^{n} (x_i - E\xi)^2 P(\xi = x_i):$$

$$P(\xi = 1) = P(\xi = 2) = \ldots = P(\xi = 6) = \frac{1}{6}, E\xi = 3.5$$

$$D\xi = \frac{1}{6}\left((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\right) =$$

$$= \frac{1}{6}(6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25) = \frac{1}{6}(17.5) = \frac{35}{12}$$

Variance (continuous random variables)

For a continuous random variable,

$$D\xi = \int_{-\infty}^{+\infty} \left(x - E\xi\right)^2 F_\xi'(x)dx$$

Variance (continuous random variables)

For a continuous random variable, $$D\xi = \int_{-\infty}^{+\infty} \left(x - E\xi\right)^2 F_{\xi}'(x)dx$$

Suppose ξ is uniformly distributed on [*a*, *b*]:

$$D\xi = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}dx = \int_a^b \left(x^2 - x(a+b) + \left(\frac{a+b}{2}\right)^2\right)\frac{1}{b-a}dx =$$

$$= \int_a^b x^2 \frac{1}{b-a}dx - \int_a^b x(a+b)\frac{1}{b-a}dx + \int_a^b \left(\frac{a+b}{2}\right)^2 \frac{1}{b-a}dx =$$

$$= \frac{(b^3 - a^3)}{3(b-a)} - \frac{(a+b)(b^2-a^2)}{2(b-a)} + \frac{(a+b)^2(b-a)}{4(b-a)} =$$

$$= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{2} + \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

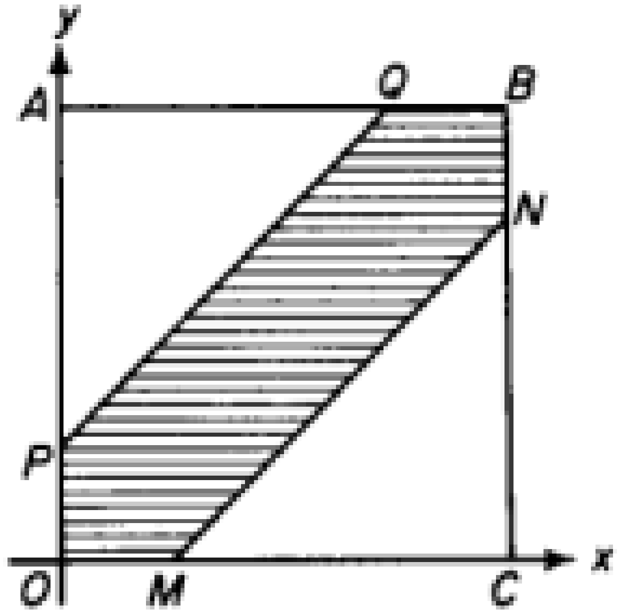
Geometric probability

Two people agree to meet at a given place between noon and 1 pm. By agreement, the first to arrive will wait 15 minutes for the second, after which he will leave. What is the probability that the meeting actually takes place if each of them selects his moment of arrival at random during the interval from 12 to 1 pm?

Geometric probability

Two people agree to meet at a given place between noon and 1 pm. By agreement, the first to arrive will wait 15 minutes for the second, after which he will leave. What is the probability that the meeting actually takes place if each of them selects his moment of arrival at random during the interval from 12 to 1 pm?

All the outcomes are defined by pairs of numbers (x, y), where $0 \le x \le 1$ and $0 \le y \le 1$. The favorable outcomes in our problem are those which correspond to points (x, y) for which $|x - y| \le 1/4$. These points make up the shaded region in the following figure:



The area of the shadowed region equals to 7/16. Thus, the desired probability is 7/16.