Binomial distribution

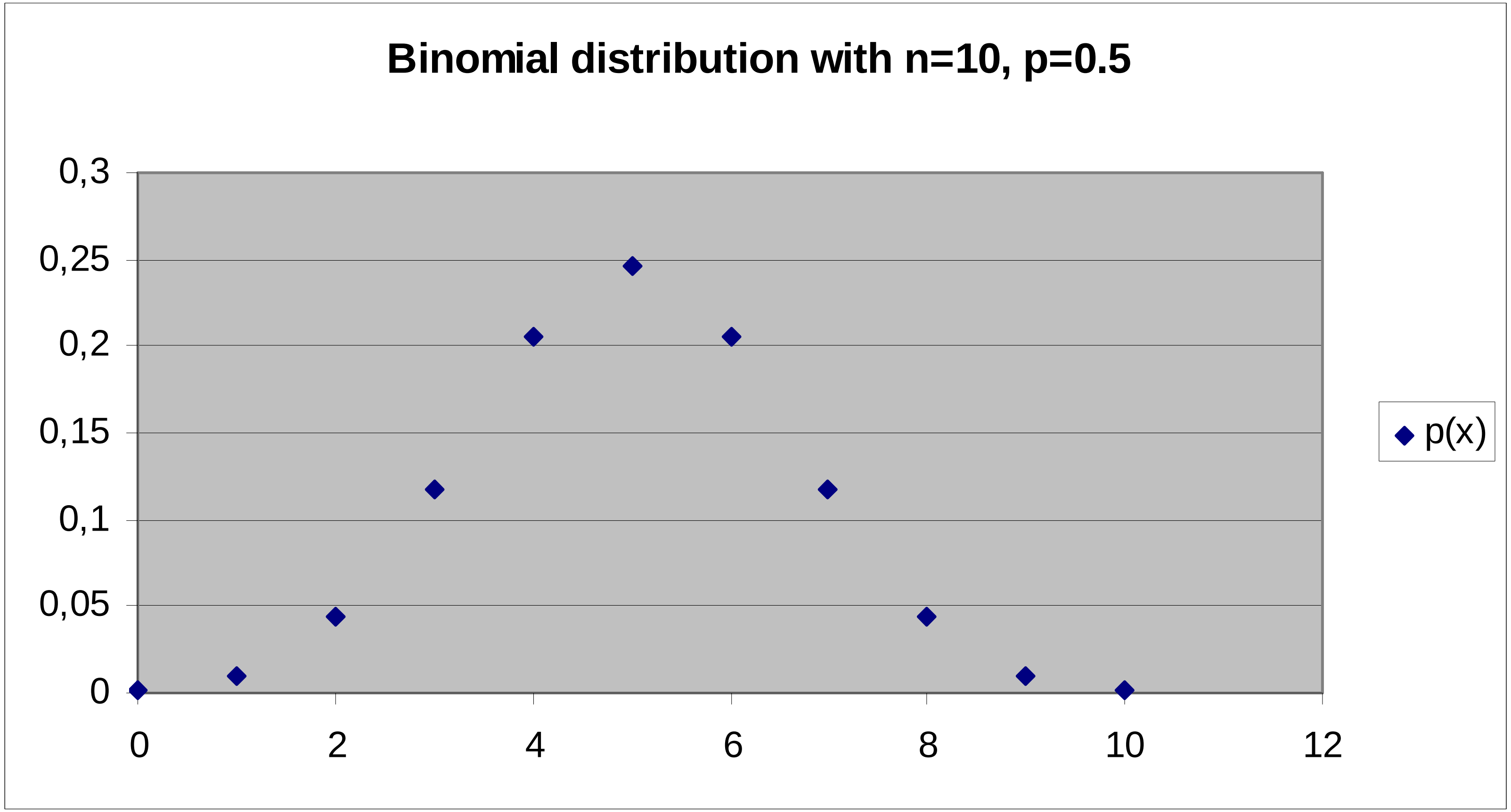
A **binomial** experiment has four properties:

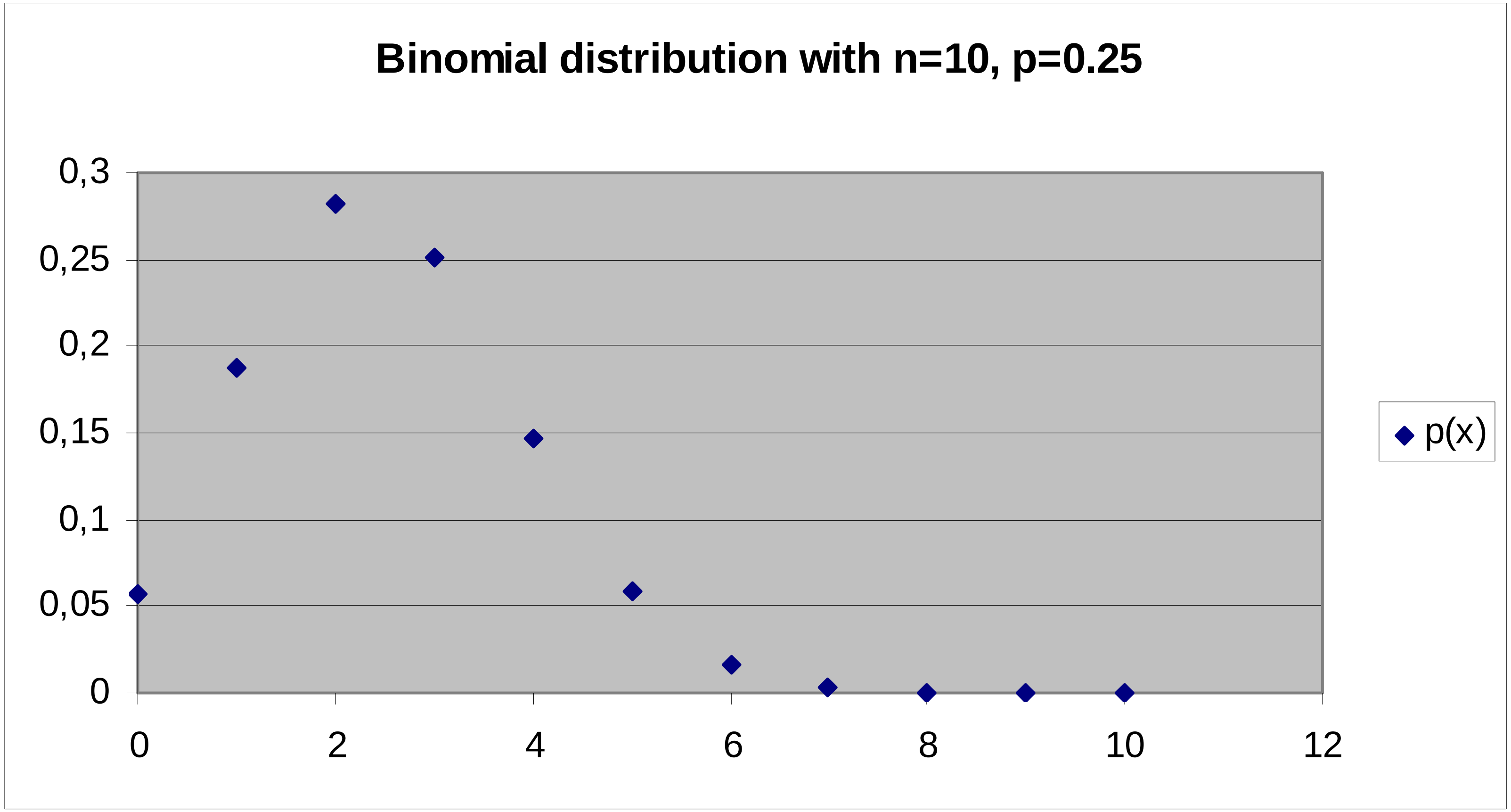
1) it consists of a sequence of n identical trials;
2) two outcomes, success or failure, are possible on each trial;
3) the probability of success on any trial, denoted p, does not change from trial to trial;
4) the trials are independent.

Let us conduct a binomial experiment with n trials, a probability p of success and let X be a random variable giving the number of successes in the binomial experiment. Then X is a discrete random variable with the range $\{0, 1, 2, ..., n\}$ and the probability function

$$p(i) = C_n^i p^i (1 - p)^{n-i}$$

We say that X has a ***binomial distribution***.


Binomial distribution with n=10, p=0.5
0,3
0,25
0,2
0,15
0,1
0,05
0
0
2
4
6
8
10
12
p(x)


Binomial distribution with n=10, p=0.25
0,3
0,25
0,2
0,15
0,1
0,05
0
0
2
4
6
8
10
12
p(x)

Expectancy and variance of a binomial distribution

For a random variable X with a binomial distribution Bi(n,p) we have

$$E(X) = pn$$

$$D(X) = np(1 - p)$$

Using the identity $$iC_n^i = \frac{i\,n!}{i!(n-i)!} = \frac{n(n-1)!}{(i-1)!(n-i)!} = nC_{n-1}^{i-1}$$

we can calculate

$$E(X) = \sum_{i=0}^{n} iC_n^i p^i q^{n-i} = \sum_{i=1}^{n} iC_n^i p^i q^{n-i} = \sum_{i=1}^{n} nC_{n-1}^{i-1} p^i q^{n-i} =$$

$$= p\sum_{i=1}^{n} nC_{n-1}^{i-1} p^{i-1} q^{n-i} = pn\sum_{i=0}^{n-1} C_{n-1}^{i} p^i q^{n-i} = pn(p+q)^{n-1} = pn$$

The same identity can be used to calculate $D(X)$

$$E\left(X^2\right)=\sum_{i=0}^{n} i^2 C_n^i p^i q^{n-i} = np\sum_{i=1}^{n} i C_{n-1}^{i-1} p^{i-1} q^{n-i} =$$

$$= np\left\{\sum_{i=1}^{n} C_{n-1}^{i-1} p^{i-1} q^{n-i} + \sum_{i=1}^{n} (i-1) C_{n-1}^{i-1} p^{i-1} q^{n-i}\right\} =$$

$$= np\left\{\sum_{i=0}^{n-1} C_{n-1}^{i} p^{i} q^{n-i} + (n-1)p\sum_{i=2}^{n} C_{n-2}^{i-2} p^{i-2} q^{n-i}\right\} =$$

$$= np\left\{(p+q)^{n-1} + (n-1)p\sum_{i=0}^{n-2} C_{n-2}^{i} p^{i} q^{n-i-2}\right\} =$$

$$= np\left\{1+(n-1)p(p+q)^{n-2}\right\} = np\{1+(n-1)p\} =$$

$$= np(1-p+np) = npq + (np)^2$$

$$D(X) = E\left(X^2\right) - [E(X)]^2 \Rightarrow D(X) = npq$$

Poisson distribution

Poisson distribution is defined for a discrete random variable X with a range {0, 1, 2, 3, ... } denoting the number of occurrences of an event A during a time interval (T_1, T_2) if the following conditions are fulfilled:

1) given any two occurrences A_1 and A_2, we have

$$P(A_2 \mid A_1) = P(A_2)$$

2) the probability of A occurring during an interval (t_1, t_2) with $T_1 \le t_1 < t_2 \le T_2$ equals to $c(t_2 - t_1)$ where c is fixed

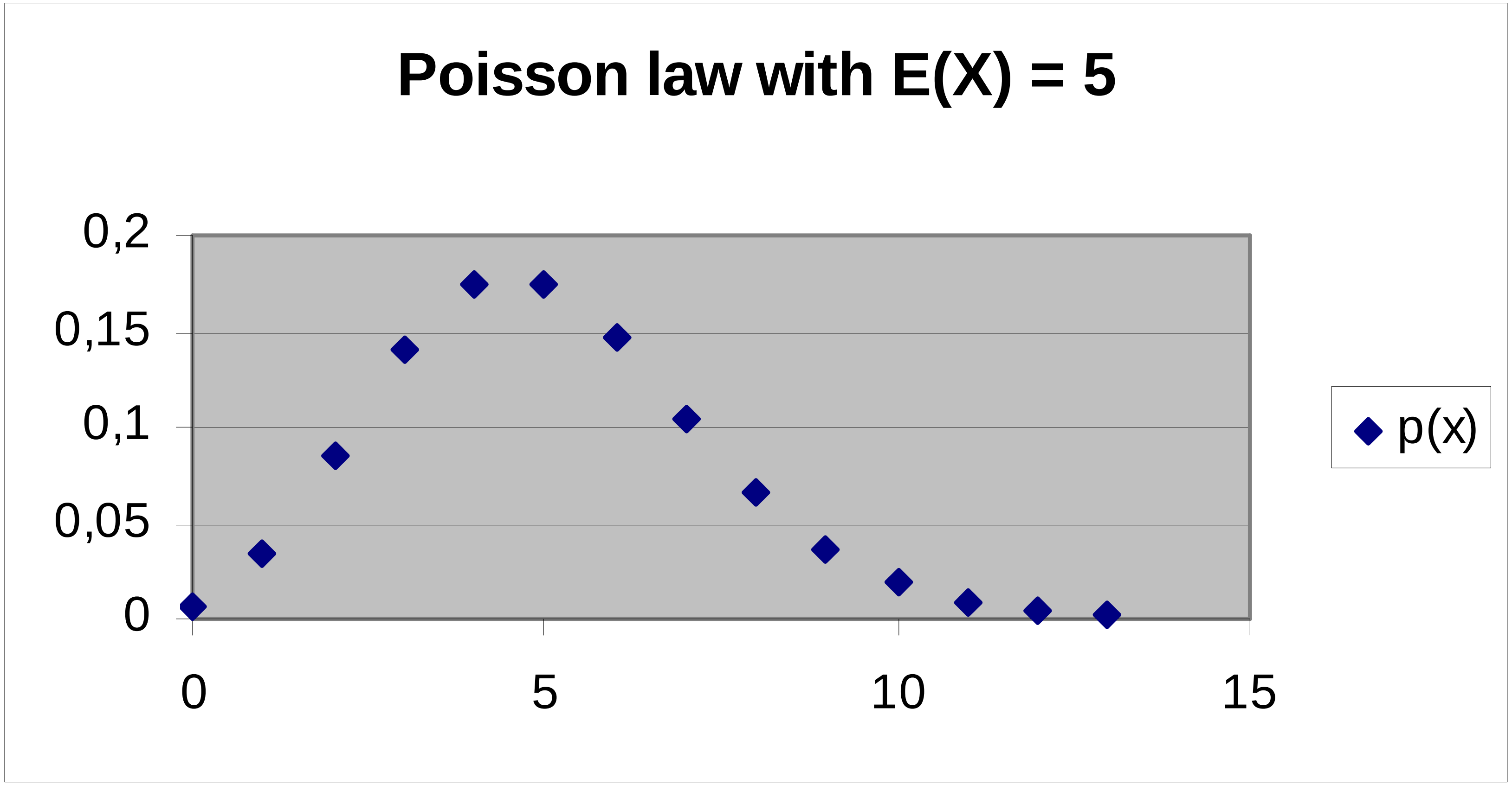
3) denoting by P_{12} the probability that E occurs at least twice between t_1 and t_2, we have

$$\lim_{t_1 \to t_2} \frac{P_{12}}{t_2 - t_1} = 0$$

The Poisson distribution Po(λ) is given by the probability function

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$\sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right) = e^{-\lambda} e^{\lambda} = 1$$



Poisson law with E(X) = 5
0,2
0,15
0,1
0,05
0
0
5
10
15
p(x)

The expectancy and variance of a random variable X with Poisson distribution are given by the following formulas

$$\boxed{E(X) = \lambda}$$

$$\boxed{D(X) = \lambda}$$

$$E(X) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} x = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} x = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda$$

$$E(X^2) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} x^2 = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} x^2 = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} x =$$

$$= e^{-\lambda} \lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} (x+1) = e^{-\lambda} \lambda \left(\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} x + \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \right) = e^{-\lambda} \lambda \left(e^{\lambda} \lambda + e^{\lambda} \right) =$$

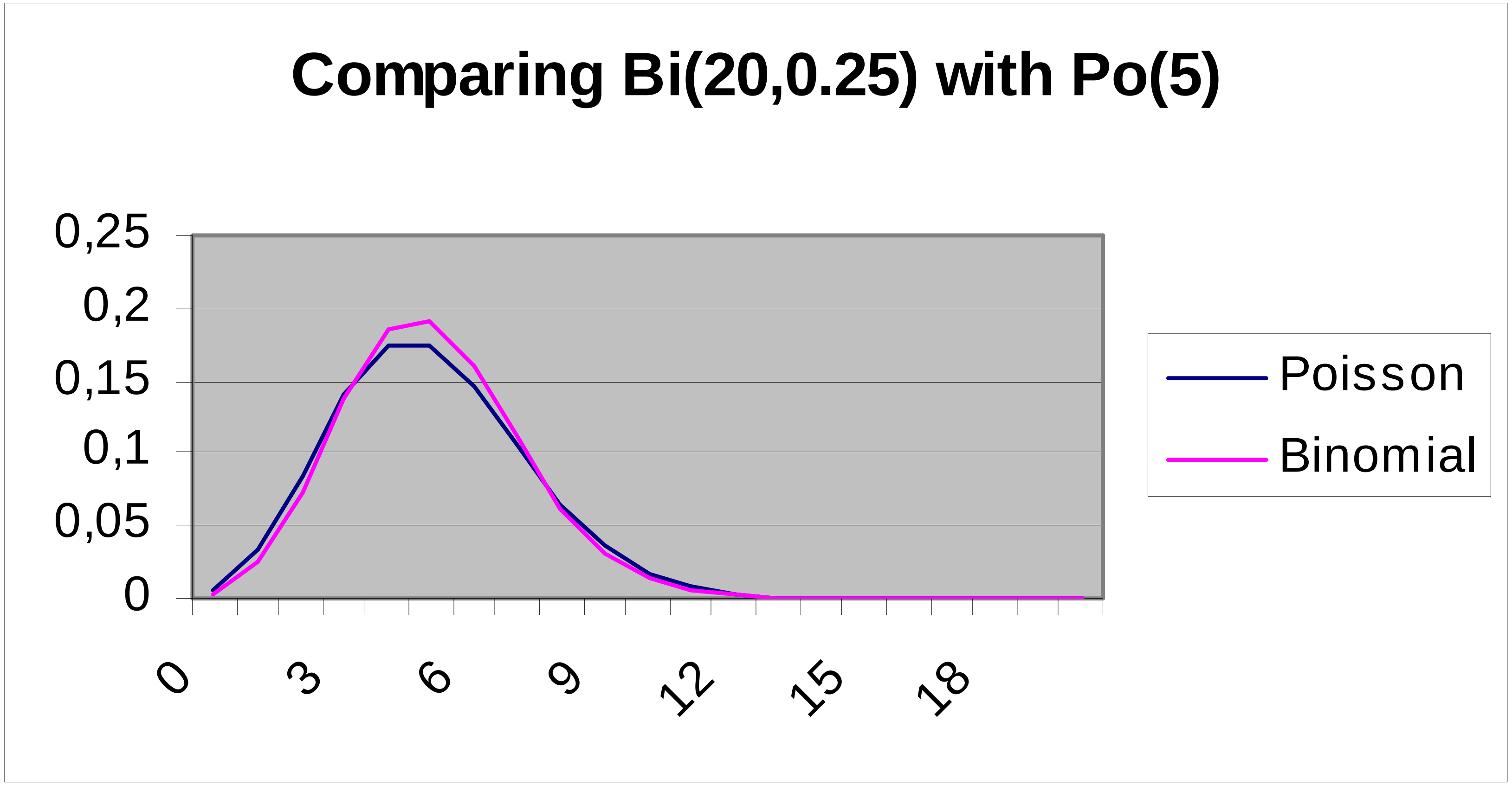
$$= \lambda^2 - \lambda$$

$$D(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Relationship between Bi(n,p) and Po(λ)

For large n, we can write

$$\mathbf{Bi}(n, p) \approx \mathbf{Po}(np)$$


Comparing Bi(20,0.25) with Po(5)
0,25
0,2
0,15
0,1
0,05
0
0
3
6
9
12
15
18
Poisson
Binomial

Normal distribution law

A continuous random variable *X* has a normal distribution law with *E*(*X*) = μ and *D*(*X*) = σ if its probability density is given by the formula

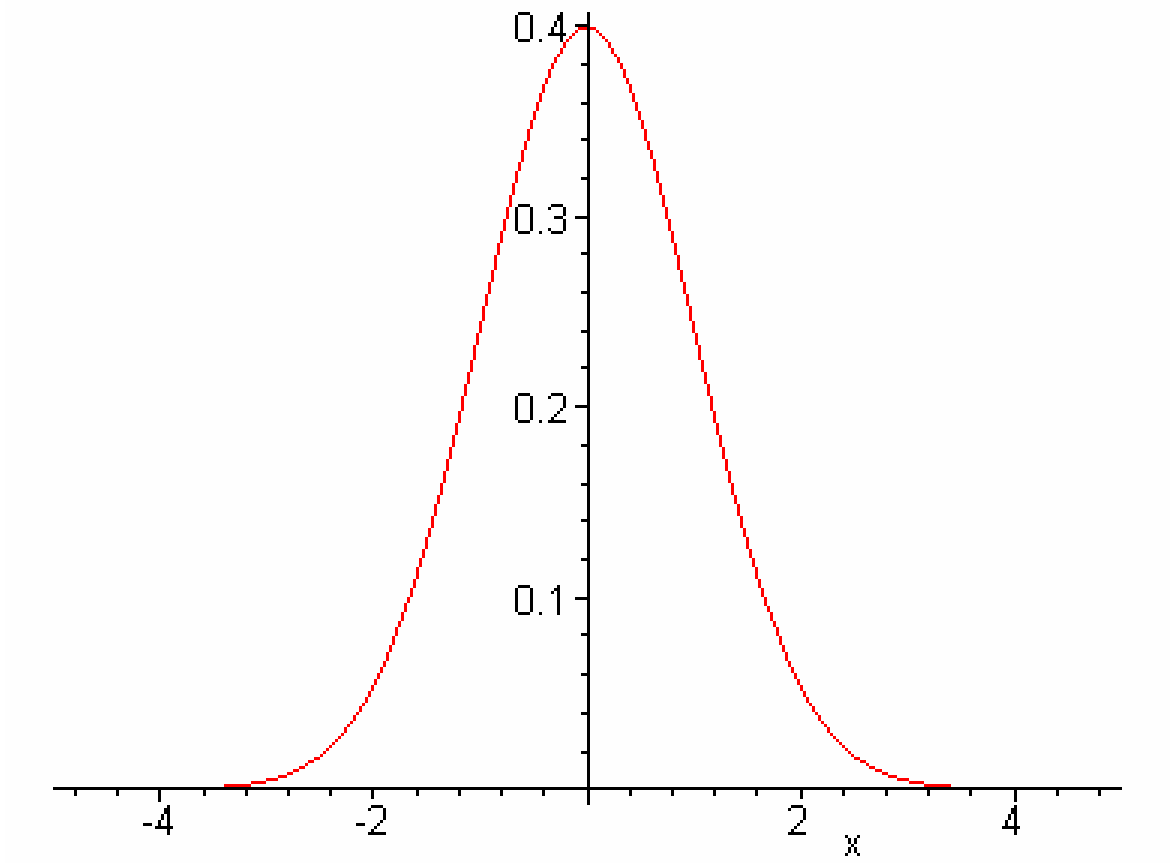
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Standardized normal distribution

A continuous random variable X has a standardized normal distribution if it has a normal distribution with $E(X) = 0$ and $D(X) = 1$ so that its probability density is given by the formula

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Standardized normal distribution law (Gaussian curve)



There are several different ways of obtaining a random variable X with a normal distribution law.

For large n, binomial and Poisson distributions asymptotically approximate normal distribution law

If an activity is undertaken with the aim to achieve a value μ, but a large number of independent factors are influencing the performance of the activity, the random variable describing the outcome of the activity will have a normal distribution with the expectancy μ.

Of all the random variables with a given expectancy μ and variance σ^2, a random variable with the normal distribution $N(\mu,\sigma^2)$ has the largest entropy.