THE ABSOLUTE FIRST CENTRAL MOMENT OF RANDOM VARIABLES

Gusev, Anatoly I., *Moscow State University of Railway Engineering (MIIT), Moscow, Russia.*

ABSTRACT

The geometric, Poisson, and binomial distribution laws are considered in the article. For each of them an analytic formula of the absolute first central moments is derived, which allows us to find the average distribution zone. The work is of a fundamental nature and can be used in studies on probability theory, in applied problems where these distribution laws are present.

Keywords: random variable, distribution laws, mathematical expectation, dispersion, root-mean-square deviation, moments of random variables.

Background. *Let a random variable X be taken (no matter which, continuous or discrete), for example, the worker's salary. The mathematical expectation MX (we take only quantities with a finite value MX) characterizes the average value of the random variable. When estimating the deviation of a random variable from the mean value in an extensive literature on probability theory (see, in part. [1–3]), the dispersion of the random variable* $D(X)=M(X-MX)^2$ *is considered, then the notion of the mean-square deviation* $\sigma(X)=\sqrt{D(X)}$ *is introduced and* $\sigma(X)$ *is called the deviation of the random variable from the mathematical expectation. What exactly characterizes the obtained deviation is not very clear, let's try to figure it out.*

Objective. *The objective of the author is to consider the absolute first central moment of random variables.*

Methods. *The author uses general scientific methods, mathematical apparatus, scientific description.*

Results.

Three zones of value

Let the salary have the distribution law shown in Table 1.

The average salary

$$MX = 10\cdot\frac{1}{5}+30\cdot\frac{1}{5}+90\cdot\frac{1}{5}+150\cdot\frac{1}{5}+170\cdot\frac{1}{5}=90.$$

The dispersion

$$D(X)=80^2\cdot\frac{1}{5}+60^2\cdot\frac{1}{5}+60^2\cdot\frac{1}{5}+80^2\cdot\frac{1}{5}=4000,$$

a $\sigma(X)=\sqrt{3560}=63,25.$

Now we calculate the true mean deviation of the random variable from the mathematical expectation, that is, we find the absolute first central moment for X (we denote it by $\Delta(X)$*):*

$$\Delta(X)=M|X-MX|=80\cdot\frac{1}{5}+60\cdot\frac{1}{5}+60\cdot\frac{1}{5}+80\cdot\frac{1}{5}=56.$$

As can be seen, the discrepancy between $\sigma(X)$ *and* $\Delta(X)$ *is large – 7,25 (thousand rubles). We take random values for which* $\Delta(X)$ *is a finite number (if* $\Delta(X) = 0$*, then the random variable X is a constant), while* $\sigma(X)$ *can be equal to* ∞ *(for example, X takes values* $\pm\frac{2^n}{n^{1.5}}$ *with probability* $\frac{1}{2^{n+1}}$*, n = 1; 2; …; …n).*

For an arbitrary random variable, we consider three zones of its values (Pic. 1).

The lower zone H: $x < MX-\Delta(X)$*.*

Middle zone C: $[MX-\Delta(X);\ MX+\Delta(X)]$*.*

The higher zone B: $x > MX+\Delta(X)$*.*

Accordingly, we have a zone of low earnings at a level of < 34, a zone of average earnings from 34 to 146, and a zone of the highest earnings at a level above 146. Earnings of the amount of 30 fall into the lower zone, but if we judge by the standard deviation of $\sigma(X)$*, then earnings fall into middle zone. It is possible that a random variable takes only two values of* $MX-\Delta(X)$ *and* $MX+\Delta(X)$*, then it is easy to assume that the probabilities of accepting these values = 1/2.*

At the same time, when a random variable takes more than two values, the interior of the middle zone is always not empty. Zone H can be empty, then zone B is not empty (that is, it means that only average and higher earnings are available, they can be called good, as it is favorable for society). And here MX = 16; $\Delta(X) = 8$*. An example of such a distribution is given in Table 2.*

If, on the contrary, zone B is empty, then zone H is not empty (this means that there are only medium and

Table 1

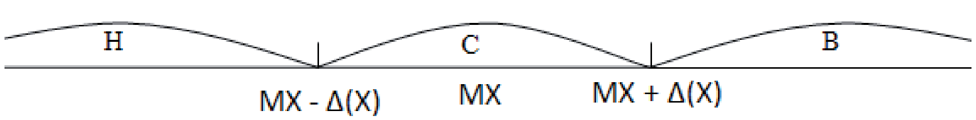
X (thous. rubles)	10	30	90	150	170
P	1/5	1/5	1/5	1/5	1/5

Table 2

X(thous. rubles)	10	22	34
P	4/6	1/6	1/6

Table 3

X (thous. rubles)	12	24	36
P	1/6	1/6	4/6



Pic. 1. Zones of values of an arbitrary random variable.

Table 4

X	1	2	3	...	n	...
P	p	pq	pq^2	...	pq^{n-1}	...

Table 5

$\left\|X-\frac{1}{p}\right\|$	$\frac{1}{p}-1$	$2-\frac{1}{p}$	$3-\frac{1}{p}$	...	$n-\frac{1}{p}$	...
P	p	pq	pq^2	...	pq^{n-1}	...

Table 6

$\left\|X-\frac{1}{p}\right\|$	$\frac{1}{p}-1$	$\frac{1}{p}-2$	...	$\frac{1}{p}-n$	$\frac{1}{p}-(n+1)$	$(n+2)-\frac{1}{p}$	...
P	p	pq	...	pq^{n-1}	pq^n	pq^{n+1}	...

Table 7

The values of σ(X), Δ(X) and σ(X) – Δ(X) for the geometric distribution law

p	б		Δ		б-Δ
1	0		0		0
0,9	0,351364		0,2		0,151364
0,8	0,559017		0,4		0,159017
0,7	0,782461		0,6		0,182461
0,6	1,054093		0,8		0,254093
0,5	1,414214		1		0,414214
1/3	2,44949		1,777778		0,671712
1/4	3,464102		2,53125		0,932852
1/5	4,472136		3,2768		1,195336
1/6	5,477226		4,018776		1,45845
1/7	6,480741		4,758833		1,721907
1/8	7,483315		5,497743		1,985572
1/9	8,485281		6,235909		2,249372
1/10	9,486833		6,973569		2,513264
1/20	19,49359		14,33944		5,154152
1/30	29,49576		21,69969		7,796072
1/40	39,49684		29,0586		10,43824
1/50	49,49747		36,41697		13,08051
1/60	59,4979		43,77508		15,72282
1/70	69,4982		51,13304		18,36516
1/80	79,49843		58,4909		21,00752
1/90	89,4986		65,84871		23,6499
1/100	99,49874		73,20647		26,29228

Table 8

X	0	1	2	...	n	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	...	$\frac{\lambda^n e^{-\lambda}}{n!}$	...

Table 9

$\lvert X-\lambda \rvert$	λ	$1-\lambda$	$2-\lambda$	...	$n-\lambda$	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	...	$\frac{\lambda^n e^{-\lambda}}{n!}$	...

Table 10

$\lvert X-\lambda \rvert$	λ		...	$\lambda-n$	$\lambda-(n+1)$	$(n+2)-\lambda$	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$		$\frac{\lambda^n e^{-\lambda}}{n!}$	$\frac{\lambda^{n+1} e^{-\lambda}}{(n+1)!}$	$\frac{\lambda^{n+2} e^{-\lambda}}{(n+2)!}$	...

Table 11

The values of σ(X), Δ(X) and σ(X) –Δ(X) for the Poisson distribution law

λ	σ	Δ	σ-Δ
1	1	0,735759	0,264241
2	1,414214	1,082682	0,331531
3	1,732051	1,344251	0,3878
4	2	1,562935	0,437065
5	2,236068	1,754674	0,481394
6	2,44949	1,927478	0,522012
7	2,645751	2,086039	0,559712
8	2,828427	2,233385	0,595043
9	3	2,371602	0,628398
10	3,162278	2,502201	0,660077
20	4,472136	3,553413	0,918723
30	5,477226	4,358072	1,119154
40	6,324555	5,035763	1,288792
50	7,071068	5,632501	1,438567
60	7,745967	6,171809	1,574157
70	8,3666	6,667639	1,698961
80	8,944272	7,129067	1,815205
90	9,486833	7,562392	1,924441
100	10	7,972199	2,027801

Table 12

X	0	1	...	k	...	n
P	q^n	npq^{n-1}	...	$C_n^k p^k q^{n-k}$	...	p^n

Table 13

$\lvert X-np \rvert$	np	1-np	...	k-np	...	n-np
P	q^n	npq^{n-1}	...	$C_n^k p^k q^{n-k}$	...	p^n

Table 14

$\lvert X-np \rvert$	np	np-1	...	np-k	k+1-np	...	n-np
P	q^n	npq^{n-1}	...	$C_n^k p^k q^{n-k}$	$C_n^{k+1} p^{k+1} q^{n-k-1}$	...	p^n

low earnings, and we are talking about unsatisfactory situation in society). Table 3 shows an example of such a distribution, where MX = 30; Δ (X) = 8.

From the general theory it follows that both for discrete and continuous random variables $\sigma(X) \geq \Delta(X)$, however, for prevailing random variables it is desirable to know their dependence shown in the form $\sigma(X) = K\Delta(X)$ (i.e. the explicit value of the coefficient K). The binomial, Poisson, and geometric distribution laws are widely used in various fields, so it is expedient to find the absolute first central moment $\Delta(X)$ for them.

Variant for the geometric law

The geometric distribution law (Table 4) has numerical characteristics: $\mathrm{MX} = \frac{1}{p}$; $D(\mathrm{X}) = \frac{q}{p^2}$;

$$\sigma(\mathrm{X}) = \frac{\sqrt{q}}{p}.$$

Theorem 1. Let $n \leq \frac{1}{p} < n+1$, *then* $\Delta(X) = 2nq^n$, *n = 1; 2;... The proof is by induction on n. For* $1 \leq \frac{1}{p} < 2$, *we have the variant of Table 5.*

$\frac{1}{p} - 1$ *is represented in the form*

$$\frac{1}{p} - 1 = 1 - \frac{1}{p} + 2\left(\frac{1}{p} - 1\right), \text{ then}$$

$$M\left|X - \frac{1}{p}\right| = \frac{1}{p} - \frac{1}{p} + 2\left(\frac{1}{p} - 1\right)\mathrm{p} = 2 - 2\mathrm{p} = 2\mathrm{q}$$

satisfies the formula to be proved.

Let at $n \leq \frac{1}{p} < n+1$ *we have* $\Delta(X) = 2nq^n$ *(see Table 6), we prove that at* $n+1 \leq \frac{1}{p} < n+2$ $\Delta(\mathrm{X}) = 2(\mathrm{n}+1)q^{\mathrm{n}+1}$.

$\frac{1}{p} - (\mathrm{n}+1)$ *is represented in the form*

$$(\mathrm{n}+1) - \frac{1}{p} + 2\left(\frac{1}{p} - (\mathrm{n}+1)\right), \text{ then}$$

$$M\left|X - \frac{1}{p}\right| = 2\mathrm{n}q^{\mathrm{n}} + 2\left(\frac{1}{p} - (\mathrm{n}+1)\right)\mathrm{p}q^{\mathrm{n}} =$$
$$= 2\mathrm{n}q^{\mathrm{n}} + 2q^{\mathrm{n}} - 2(\mathrm{n}+1)\mathrm{p}q^{\mathrm{n}} =$$
$$= 2(\mathrm{n}+1)q^{\mathrm{n}} - 2(\mathrm{n}+1)\mathrm{p}q^{\mathrm{n}} =$$
$$= 2(\mathrm{n}+1)q^{\mathrm{n}}(1-\mathrm{p}) = 2(\mathrm{n}+1)q^{\mathrm{n}+1},$$

which was to be proved.

Note. It is easy to verify that in the theorem we can replace $n \leq \frac{1}{p} < n+1$ *by* $n \leq \frac{1}{p} \leq n+1$, *i.e.* $\Delta(X)$ *– continuous function of p.*

To compare $\sigma(X)$ and $\Delta(X)$ Table 7 is compiled in the variant of the geometric distribution law.

Variant for the Poisson distribution law

Let's consider the Poisson distribution law (Table 8).

A random variable distributed according to the Poisson's law has numerical characteristics: MX = λ; D(X) = λ; $\sigma(\mathrm{X}) = \sqrt{\lambda}$.

Theorem 2. Let n ≤ λ < n + 1, then

$$\sigma(\mathrm{X}) = \frac{2\lambda^{n+1}e^{-\lambda}}{n!}, \ \mathrm{n} = 0;1;2;....$$

We carry out the proof by induction on n. For 0 ≤ λ < 1 we have the variant of Table 9.

λ is represented in the form λ = (0–λ)+2λ, then M|X–λ| = λ–λ+2λe$^{-\lambda}$ = 2λe$^{-\lambda}$ = λ–λ+2λe$^{-\lambda}$ – which satisfies the formula to be proved.

Let for n ≤ λ < n +1 we have $\Delta(\mathrm{X}) = \frac{2\lambda^{n+1}e^{-\lambda}}{n!}$ *(see Table 10), we prove that for*

$$\mathrm{n}+1 \leq \lambda < \mathrm{n}+2 \ \Delta(\mathrm{X}) = \frac{2\lambda^{n+2}e^{-\lambda}}{(n+1)!}.$$

λ–(n+1) is represented in the form (n+1)–λ+2(λ–(n+1)), then

$$\mathrm{M}|X-\lambda| = \frac{2\lambda^{n+1}e^{-\lambda}}{n!} + 2(\lambda - (\mathrm{n}+1))\frac{\lambda^{n+1}e^{-\lambda}}{(n+1)!} = \frac{2\lambda^{n+2}e^{-\lambda}}{(n+1)!},$$

which was to be proved.

Note. It is easy to verify that in the theorem we can replace n ≤ λ < n+1 by n ≤ λ ≤ n+1 i.e. Δ(X) is continious function of λ.

For comparison of σ(X) and Δ(X) in the Poisson distribution variant Table 11 is compiled.

Variant for the binominal distribution law

Let's consider the binominal distribution law (Table 12).

A random variable distributed according to the binomial law has numerical characteristics: MX = np; D(X) = npq; $\sigma(\mathrm{X}) = \sqrt{\mathrm{npq}}$.

Theorem 3. Let k–1 ≤ np ≤ k, k = 1;2;...n, then $\Delta(\mathrm{X}) = 2\mathrm{k}C_n^k p^k q^{n-k+}$. *We carry out the proof by induction on k. For k = 1, i.e. 0 ≤ np ≤1, we have the variant of Table 13.*

Here np is represented in the form 0–np+2np, then M|X–np| = np–np + 2npq^n = 2npq^n – which satisfies the formula to be proved.

Let for k–1 ≤ np ≤ k we have $\Delta(\mathrm{X}) = 2\mathrm{k}C_n^k p^k q^{n-k+1}$ *(see Table 14). Let us prove that for k ≤ np ≤ k+1* $\Delta(\mathrm{X}) = 2(\mathrm{k}+1)C_n^{k+1} p^{k+1} q^{n-k}$.

np–k is represented in the form np–k = k–np+2(np–k), then we have

$$M|X-np| = 2\mathrm{k}C_n^k p^k q^{n-k+1} + 2(\mathrm{np}-\mathrm{k})C_n^k p^k q^{n-k} =$$
$$= 2\mathrm{np}C_n^k p^k q^{n-k} + 2\mathrm{k}C_n^k p^k q^{n-k}(q-1)) =$$
$$= 2\mathrm{n}C_n^k p^{k+1} q^{n-k} - 2\mathrm{k}C_n^k p^{k+1} q^{n-k} = 2(\mathrm{n}-\mathrm{k})C_n^k p^{k+1} q^{n-k} =$$
$$= 2(\mathrm{k}+1)\frac{C_n^k(n-k)}{(k+1)} p^{k+1} q^{n-k} = 2(\mathrm{k}+1)C_n^{k+1} p^{k+1} q^{n-k},$$

which was to be proved.

For comparison of σ(X) and Δ(X) in the variant of the binominal distribution law two tables are compiled for n = 100 and for n = 1000.

Table 15

Values of σ(X), Δ(X) and σ(X) – Δ(X) for the binominal distribution law for n = 100

p	б	Δ	б-Δ
0	0	0	0
0,01	0,994987	0,732064683	0,262923
0,02	1,4	1,071782533	0,328217
0,03	1,705872	1,323899422	0,381973
0,04	1,959592	1,531303658	0,428288
0,05	2,179449	1,710169359	0,46928
0,06	2,374868	1,868659559	0,506209
0,07	2,55147	2,011576901	0,539893
0,08	2,712932	2,142031945	0,5709
0,09	2,861818	2,262177211	0,59964
0,1	3	2,373576243	0,626424
0.2	4	3.177606874	0.822393
0,3	4,582576	3,64492232	0,937653
0,4	4,898979	3,89851896	1,000461
0,5	5	3,979461869	1,020538
0,6	4,898979	3,89851896	1,000461
0,7	4,582576	3,64492232	0,937653
0,8	4	3,177606874	0,822393
0,9	3	2,373576243	0,626424
1	0	0	0

In view of the symmetry, it is clear that for symmetric values of p the results will be the same, for example, at p = 0,6 and p = 0,4 (see Table 15). Therefore, Table 16 is made up to p = 0,5.

Variant for classical continuous random variables

In this section we find absolute first central moments for uniform, exponential and normal distribution laws.

Uniform distribution law

A random variable X, uniformly distributed on the interval [a, b], has density p(x), equal to $\frac{1}{в-a}$ *, on the interval [a, b], and equal to 0 outside it. In this case* $\mathrm{MX}=\frac{a+в}{2},\ \mathrm{D}(\mathrm{X})=\frac{(в-a)^2}{12},\ \sigma(\mathrm{X})=\frac{в-a}{2\sqrt{3}}$.

Let's calculate Δ(X) (the result is, of course, obvious and equal to $\frac{в-a}{4}$ *):*

$$\Delta(\mathrm{X})=\int_a^{\frac{a+в}{2}}(\frac{a+в}{2}-x)\frac{1}{в-a}\mathrm{dx}+\int_{\frac{a+в}{2}}^{в}(x-\frac{a+в}{2})\frac{1}{в-a}\mathrm{dx}=$$

$$=\frac{1}{в-a}(\frac{a+в}{2}\mathrm{x}-\frac{x^2}{2})\Big|_a^{\frac{a+в}{2}}+\frac{1}{в-a}(\frac{x^2}{2}-\frac{a+в}{2}\mathrm{x})\Big|_{\frac{a+в}{2}}^{в}=$$

$$=\frac{1}{в-a}\left(\frac{(a+в)^2}{4}-\frac{(a+в)^2}{8}-\frac{a+в}{2}\mathrm{a}+\frac{a^2}{2}\right)-$$

$$-\frac{1}{в-a}\left(\frac{в^2}{2}-\frac{a+в}{2}\mathrm{в}-\frac{(a+в)^2}{8}+\frac{(a+в)^2}{4}\right)=$$

$$=\frac{1}{в-a}\left(\frac{(a+в)^2}{4}-\mathrm{aв}\right)=\frac{в-a}{4}\approx 0.866\ \sigma(\mathrm{X}).$$

Exponential distribution law

A random variable X, distributed in exponential order, has density p(x), equal to $\lambda e^{-\lambda x}$*, at* $x \geq 0$ *and zero density at* $x < 0$. $\mathrm{MX}=\frac{1}{\lambda}$; $\mathrm{D}(\mathrm{X})=\frac{1}{\lambda^2}$; $\sigma(\mathrm{X})=\frac{1}{\lambda}$.

Let's calculate Δ(X) (the result here is, of course, not obvious):

$$\Delta(\mathrm{X})=\int_0^{\frac{1}{\lambda}}(\frac{1}{\lambda}-x)e^{-\lambda x}\ \mathrm{dx}+\int_{\frac{1}{\lambda}}^{\infty}\left(\mathrm{x}-\frac{1}{\lambda}\right)e^{-\lambda x}\ \mathrm{dx}=$$

$$=\int_0^{\frac{1}{\lambda}}e^{-\lambda x}dx-\int_0^{\frac{1}{\lambda}}x\lambda e^{-\lambda x}dx+\int_{\frac{1}{\lambda}}^{\infty}x\lambda e^{-\lambda x}dx-\int_{\frac{1}{\lambda}}^{\infty}e^{-\lambda x}dx=$$

$$=2\int_0^{\frac{1}{\lambda}}e^{-\lambda x}dx-2\int_0^{\frac{1}{\lambda}}x\lambda e^{-\lambda x}\ \mathrm{dx}+\frac{1}{\lambda}-\frac{e^{-\lambda x}}{-\lambda}\Big|_0^{\infty}=$$

Table 16

The values of σ(X), Δ(X) and σ(X) – Δ(X) for the binominal distribution law for n = 1000

p	б	Δ	б-Δ
0	0	0	0
0,001	0,9995	0,735391	0,264109
0,002	1,412799	1,081599	0,3312
0,003	1,729451	1,342233	0,387218
0,004	1,995996	1,559805	0,436191
0,005	2,230471	1,750281	0,48019
0,006	2,44213	1,921686	0,520445
0,007	2,636475	2,078724	0,557751
0,008	2,817091	2,224432	0,592659
0,009	2,986469	2,360903	0,625566
0,01	3,146427	2,489656	0,65677
0,05	6,892024	5,489858	1,402166
0,1	9,486833	7,563022	1,923811
0,15	11,29159	9,004249	2,287341
0,2	12,64911	10,08812	2,560995
0,25	13,69306	10,92154	2,771524
0,3	14,49138	11,55882	2,932555
0,35	15,0831	12,03117	3,051933
0,4	15,49193	12,35751	3,13442
0,45	15,73213	12,54925	3,182887
0,5	15,81139	12,61251	3,198879

$$= 2\int_0^{\frac{1}{\lambda}} e^{-\lambda x}dx + 2\left(xe^{-\lambda x}\Big|_0^{\frac{1}{\lambda}} - \int_0^{\frac{1}{\lambda}} e^{-\lambda x}\,dx \right) =$$

$$= \frac{2}{e\lambda} \approx 0{,}736\sigma(X).$$

Normal distribution law

A random variable X, distributed according to the normal law, has density

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}},\ MX = a,$$

$$D(X) = \sigma^2,\ \sigma(X) = \sigma.$$

We can assume that a = 0 (otherwise we take a random variable X–a, which has the same Δ(X)).

Let's calculate Δ(X):

$$\Delta(X) = 2\int_0^{\infty} x\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_0^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,d\left(x^2\right) =$$

$$= \frac{-2\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\Bigg|_0^{\infty} = \sqrt{\frac{2}{\pi}}\sigma \approx 0.798\sigma.$$

Conclusion. *The analytical formulas for the absolute first central moments for the geometric, binomial, and Poisson distribution laws are derived. This allows us to find the average distribution zone, which is important for calculations in applied problems. The work can also be used in studies on probability theory.*

REFERENCES

1. Gnedenko, B. V. Course on probability theory [*Kurs teorii verojatnostej*]. 6th ed., rev. and enl. Moscow, Nauka publ., 1988. [Electronic resource]: http://www.booksshare.net/books/physics/gnedenko-bv/1988/files/kursteoriiveroyatnostey1988.pdf. Last accessed 15.11.2016.

2. Gmurman, V. E. Probability theory and mathematical statistics [*Teorija verojatnostej i matematicheskaja statistika*]. Moscow, Vysshaja shkola publ., 2003, 479 p.

3. Kremer, N. Sh. Probability theory and mathematical statistics: textbook [*Teorija verojatnostej i matematicheskaja statistika: uchebnik*]. 3rd ed., rev. and enl. Moscow, Unity publ., 2012, 552 p.

Information about the author:

Gusev, Anatoly I. – Ph.D. (Physics and Mathematics), associate professor of Moscow State University of Railway Engineering (MIIT), Moscow, Russia, +7 (495) 684–24–82.

Article received 09.11.2016, accepted 20.01.2017.