N.I. Lobachevsky State University
of Nizhni Novgorod

Probability theory and mathematical statistics:

Variance, covariance, correlation
Law of large numbers

Associate Professor
A.V. Zorine

Definition

The *variance* (also called the *dispersion*) of the random variable X is defined as

$$\operatorname{Var} X = \mathbf{E}(X - \mathbf{E}X)^2$$

Standard deviation is defined as

$$\sigma(X) = \sqrt{\operatorname{Var} X}$$

Properties:

1. $\operatorname{Var} X \geqslant 0$. If $X = C$ (a non-random constant), $\operatorname{Var} X = \mathbf{E}(C - \mathbf{E}C)^2 = 0$. If $\operatorname{Var} X = 0$ then X is constant (not random) almost sure: there exists C such that $\mathrm{P}(X = C) = 1$.

2. $\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2$

Proof. Direct computation:

$$\begin{aligned}\operatorname{Var} X &= \mathbf{E}(X - \mathbf{E}X)^2 \\ &= \mathbf{E}(X^2 - 2X\mathbf{E}X + (\mathbf{E}X)^2) \\ &= \mathbf{E}X^2 - 2(\mathbf{E}X)^2 + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.\end{aligned}$$

2. $\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2$

Proof. Direct computation:

$$\begin{aligned}
\operatorname{Var} X &= \mathbf{E}(X - \mathbf{E}X)^2 \\
&= \mathbf{E}(X^2 - 2X\mathbf{E}X + (\mathbf{E}X)^2) \\
&= \mathbf{E}X^2 - 2(\mathbf{E}X)^2 + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.
\end{aligned}$$

Example. Let X have the uniform distribution on the set $\{1, 2, \dots, n\}$, $\mathrm{P}(X = k) = \frac{1}{n}$, $k = 1, 2, \dots, n$. We've seen that $\mathbf{E}X = \frac{n+1}{2}$. Then

$$\begin{aligned}
\mathbf{E}X^2 &= \frac{1}{n}(1^2 + 2^2 + \ldots + n^2) \\
&= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6},
\end{aligned}$$

$$\operatorname{Var} X = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2 - 1}{12}.$$

3. $\operatorname{Var} X = \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2$.
Proof. Notice that $\mathbf{E}X^2 = \mathbf{E}[X(X-1+1)] = \mathbf{E}[X(X-1)] + \mathbf{E}X$, then use the previous formula.

3. $\operatorname{Var} X = \mathbf{E}[X(X-1)] + \mathbf{E}X - (\mathbf{E}X)^2$.
Proof. Notice that $\mathbf{E}X^2 = \mathbf{E}[X(X-1+1)] = \mathbf{E}[X(X-1)] + \mathbf{E}X$, then use the previous formula.

It is useful when X is an integer random variable.

Example. Let X have the binomial distribution. Then

$$\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} (k-1)nC_{n-1}^{k-1} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)C_{n-2}^{k-2} p^{(k-2)+2} (1-p)^{n-2-(k-2)} = n(n-1)p^2,
\end{aligned}$$

$$\operatorname{Var} X = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

Example. Let X have a geometric distribution, then

$$\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} \\
&= p(1-p)\sum_{k=2}^{\infty}((1-p)^k)'' \\
&= p(1-p)\Big(\frac{(1-p)^2}{1-(1-p)}\Big)'' \\
&= 2\frac{1-p}{p^2}, \\
\operatorname{Var} X &= \frac{1-p}{p^2}
\end{aligned}$$

Example. Let X have a Poisson distribution with parameter λ. Then

$$\mathbf{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}$$
$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} = \lambda^2,$$
$$\operatorname{Var} X = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

4. For a constant a, $\operatorname{Var}(aX) = a^2 \operatorname{Var} X$.

Proof. From the definition,

$$\begin{aligned} \operatorname{Var}(aX) &= \mathbf{E}(aX - \mathbf{E}(aX))^2 \\ &= \mathbf{E}(a(X - \mathbf{E}X))^2 = a^2 \mathbf{E}(X - \mathbf{E}X)^2 \end{aligned}$$

4. For a constant a, $\operatorname{Var}(aX) = a^2 \operatorname{Var} X$.

Proof. From the definition,

$$\begin{aligned}\operatorname{Var}(aX) &= \mathbf{E}(aX - \mathbf{E}(aX))^2 \\ &= \mathbf{E}(a(X - \mathbf{E}X))^2 = a^2\mathbf{E}(X - \mathbf{E}X)^2\end{aligned}$$

Definition

The *covariance* of X and Y is defined as

$$\operatorname{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)$$

If $\operatorname{Var} X > 0$, $\operatorname{Var} Y > 0$, then

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{Var} X \operatorname{Var} Y}}$$

is called the *correlation coefficient* of X and Y.

Another useful formula:

$$\operatorname{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y$$

Proof.

$$\operatorname{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = \mathbf{E}(XY - X\mathbf{E}Y - Y\mathbf{E}X + \mathbf{E}X\mathbf{E}Y)$$

$$= \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) - (\mathbf{E}X)(\mathbf{E}Y) + (\mathbf{E}X)(\mathbf{E}Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y).$$

It follows from the Cauchy-Bunyakowsky inequality:

$$\operatorname{cov}(X, Y) \leqslant \sigma(X)\sigma(Y)$$

More facts:

$$\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X),$$

$$\operatorname{cov}(aX_1 + bX_2, Y) = a \operatorname{cov}(X_1, Y) + b \operatorname{cov}(X_2, Y)$$

5. For random variables X and Y,

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbf{E}((X-\mathbf{E}X)+(Y-\mathbf{E}Y))^2 \\ &= \mathbf{E}(X-\mathbf{E}X)^2 + 2\mathbf{E}(X-\mathbf{E}X)(Y-\mathbf{E}Y) + \mathbf{E}(Y-\mathbf{E}Y)^2 \\ &= \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{cov}(X,Y).\end{aligned}$$

If X and Y are independent, so are $X-\mathbf{E}X$ and $Y-\mathbf{E}Y$. By a property of expectation,

$$\mathrm{cov}(X,Y) = \mathbf{E}(X-\mathbf{E}X)\mathbf{E}(Y-\mathbf{E}Y) = 0.$$

5′. For independent random variables X and Y

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y$$

Example. Given the joint distribution of random variables X and Y, compute $\operatorname{Var} X$, $\operatorname{Var} Y$, $\operatorname{cov}(X, Y)$.

$X \setminus Y$	-1	0	1
-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

Example. Given the joint distribution of random variables X and Y, compute $\operatorname{Var} X$, $\operatorname{Var} Y$, $\operatorname{cov}(X, Y)$.

$X \setminus Y$	-1	0	1
-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

$$\mathbf{E}X = (-1) \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) + 1 \cdot \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{4}\right) = 0,$$

$$\mathbf{E}X^2 = (-1)^2 \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) + 1^2 \cdot \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{4}\right) = 1,$$

$$\operatorname{Var} X = 1 - 0^2 = 1,$$

$$\mathbf{E}Y = (-1)\cdot(\frac{1}{6}+\frac{1}{8}) + 0\cdot(\frac{1}{6}+\frac{1}{8}) + 1\cdot(\frac{1}{6}+\frac{1}{4}) = \frac{1}{8},$$

$$\operatorname{Var} Y = (-1)^2\cdot(\frac{1}{6}+\frac{1}{8}) + 0^2\cdot(\frac{1}{6}+\frac{1}{8}) + 1^2\cdot(\frac{1}{6}+\frac{1}{4}) - \Big(\frac{1}{8}\Big)^2 = \frac{133}{192},$$

$$\mathbf{E}XY = (-1)\cdot(-1)\cdot\frac{1}{6} + (-1)\cdot 0\cdot\frac{1}{6} + (-1)\cdot 1\cdot\frac{1}{6} +$$

$$+1\cdot(-1)\cdot\frac{1}{8} + 1\cdot 0\cdot\frac{1}{8} + 1\cdot 1\cdot\frac{1}{4} = \frac{1}{8},$$

$$\operatorname{cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y = \frac{1}{8}.$$

Example. Let X and Y denote the number of Heads and Tails in n tosses. Find the covariance of X and Y and the correlation coefficient.

Since $X+Y=n$, $Y=n-X$, $\mathbf{E}Y=\mathbf{E}X=\frac{n}{2}$,

$$\operatorname{cov}(X,Y)=\mathbf{E}\left(X-\frac{n}{2}\right)\left(Y-\frac{n}{2}\right)$$

$$=\mathbf{E}(X-\frac{n}{2})(n-X-\frac{n}{2})=\mathbf{E}\left(nX-X^2-\frac{n^2}{4}\right),$$

$$\mathbf{E}X^2=\operatorname{Var}X+(\mathbf{E}X)^2=\frac{n}{4}+\frac{n^2}{4},$$

$$\operatorname{cov}(X,Y)=\frac{n^2}{2}-\frac{n}{4}-\frac{n^2}{4}-\frac{n^2}{4}=-\frac{n}{4},$$

$$\rho(X,Y)=-1.$$

Example. Let X and Y be independent random variables, then

$$\operatorname{Var}(X - Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

Indeed,

$$\operatorname{Var}(-Y) = (-1)^2 \operatorname{Var} Y = \operatorname{Var} Y,$$

$$\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

Example. Let X have the binomial distribution $b(n,p)$, I_k be the indicator variable of a success in the kth trial. Then I_1, I_2, $\ldots$, I_n are independent,

$$\mathbf{E}I_k = p, \quad \mathbf{E}I_k^2 = \mathbf{E}I_k = p, \quad \operatorname{Var} I_k = p - p^2 = p(1-p),$$

$$\operatorname{Var} X = \operatorname{Var} I_1 + \operatorname{Var} I_2 + \ldots + \operatorname{Var} I_n = np(1-p)$$

The result is re-established.

Notice that $\frac{X}{n}$ is the frequency of successes,

$$\mathbf{E}\left(\frac{X}{n}\right) = p, \quad \operatorname{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}.$$

Example. There are N balls in the urn, M blue balls and $N-M$ red balls. r balls are sampled without replacement ($r < M$, $r < N-M$). Let X denote the number of blue balls among r (i.e. in the sample). Find $\operatorname{Var} X$.

Let I_i equal 1 if the i ball is blue, equal 0 otherwise. It was proved in the previous lectures that $\mathrm{P}(I_i = 1) = \frac{M}{N}$ for each $i = 1, 2, \ldots, r$. Moreover it can be proved that

$$\mathrm{P}(I_i I_j = 1) = \frac{M(M-1)}{N(N-1)}, i \neq j.$$

Then $X = I_1 + I_2 + \ldots + I_r$,

$$\mathbf{E}X^2 = \mathbf{E}I_1^2 + \mathbf{E}I_2^2 + \ldots + \mathbf{E}I_r^2 + 2(\mathbf{E}I_1 I_2 + \ldots + \mathbf{E}I_{r-1}I_r) =$$

$$= r\frac{M}{N} + 2\frac{r(r-1)}{2}\frac{M(M-1)}{N(N-1)},$$

$$\operatorname{Var} X = r\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{M-r}{N-1}.$$

If $\operatorname{cov}(X,Y) = 0$ ($\rho(X,Y) = 0$) then X and Y are called *uncorrelated*. Consider two random variables X and Y. Suppose that only X can be observed. If X and Y are correlated we may expect that knowing the value of X allows us to make some inference about the values of the unobserved variable Y.

Any function $f(X)$ is called *an estimator* for X. We call an estimator $f^*(X)$ *the best mean-square estimator* if

$$\mathbf{E}(Y - f^*(X))^2 \leqslant \mathbf{E}(Y - f(X))^2$$

for any other estimator $f(X)$.

Let's find the best linear estimator. Consider the function $g(a,b) = \mathbf{E}(Y - (a + bX))^2$. Differentiating $g(a,b)$ with respect to x and b, we obtain

$$\frac{\partial g(a,b)}{\partial a} = -2\mathbf{E}(Y - (a + bX)),$$
$$\frac{\partial g(a,b)}{\partial b} = -2\mathbf{E}(Y - (a + bX))X,$$

and setting the derivatives equal to zero, we find

$$a = \mathbf{E}Y - b\mathbf{E}X, \quad b = \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}.$$

In other words, the best linear estimator

$$f^*(X) = \mathbf{E}Y + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var} X}(X - \mathbf{E}X).$$

The number $\mathbf{E}(Y - f^*(X))^2$ is called the *mean-square error of observation*. An easy calculation shows that it is equal to

$$(1 - \rho^2(X, Y)) \operatorname{Var} Y.$$

The larger (in absolute value) the correlation coefficient $\rho(X, Y)$ between X and Y, the smaller the mean-square error of observation. In particular, if $|\rho(X, Y)| = 1$, then this error is zero. On the other hand, if X and Y are uncorrelated, ($\rho(X, Y) = 0$), then $f^*(X) = \mathbf{E}Y$, i.e. in the absence of correlation the best estimate of Y in terms of X is simply $\mathbf{E}Y$.

Uncorrelated random variables are not necessary independent!

Example. Let $\Omega = \{0, \frac{\pi}{2}, \pi\}$, the outcomes are equally likely, $X(\omega) = \sin\omega$, $Y(\omega) = \cos\omega$. Then

$$\mathbf{E}X = \sin 0 \cdot \frac{1}{3} + \sin\frac{\pi}{2} \cdot \frac{1}{3} + \sin\pi \cdot \frac{1}{3} = \frac{1}{3},$$

$$\mathbf{E}Y = \cos 0 \cdot \frac{1}{3} + \cos\frac{\pi}{2} \cdot \frac{1}{3} + \cos\pi \cdot \frac{1}{3} = 0,$$

$$\operatorname{cov}(X, Y) = (\sin 0 - \frac{1}{3}) \cos 0 \cdot \frac{1}{3} + (\sin\frac{\pi}{2} - \frac{1}{3}) \cos\frac{\pi}{2} \cdot \frac{1}{3} + (\sin\pi - \frac{1}{3}) \cos\pi \cdot \frac{1}{3} = 0.$$

6. If $X \geqslant 0$, then

$$\mathrm{P}(X > a) \leqslant \frac{\mathbf{E}X}{a}.$$

Proof. Let $Y = 1$ when $X > a$, and $Y = 0$ whenever $X \leqslant a$. Then

$$\mathbf{E}X = \mathbf{E}(XY + X(1-Y)) = \mathbf{E}XY + \mathbf{E}X(1-Y) \geqslant \mathbf{E}XY \geqslant \mathbf{E}aY = a\mathrm{P}(X > a).$$

6′ (Chebyshev's inequality) For any $a > 0$

$$\mathrm{P}(|X - \mathbf{E}X| > a) \leqslant \frac{\operatorname{Var} X}{a^2}.$$

The standard deviation $\sigma(X)$ measures the average deviation of the random variable X from its expected value. This is called the *rule of 6 sigmas*:

$$\mathrm{P}(|X - \mathbf{E}X| > 3\sigma(X)) \leqslant \frac{\operatorname{Var} X}{9\sigma^2(X)} = \frac{1}{9}$$

The law of large numbers

We say that random variables $X_1, X_2, \ldots$ with finite expectations $\mathbf{E}X_1$, $\mathbf{E}X_2, \ldots$ follow the (weak) law of large numbers if for any $a > 0$

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}X_i\right| > a\right) = 0$$

Chebyshev's law of large numbers

Let random variables $X_1, X_2, \ldots$ be pairwise independent, have finite variances bounded by the same constant C ($\operatorname{Var} X_i \leqslant C$) then for any $a > 0$

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}X_i\right| > a\right) = 0$$

Proof. When the conditions of the theorem hold,

$$\operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) = \frac{1}{n^2}\sum_{k=1}^{n} \operatorname{Var} X_k,$$

hence

$$\operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) \leqslant \frac{C}{n}.$$

Now,

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}X_i\Big| > a\Big) \leqslant \frac{\operatorname{Var}\big(\frac{1}{n}\sum_{k=1}^{n} X_k\big)}{a} \leqslant \frac{C}{na^2} \to 0$$

as $n \to \infty$.

Now the Law of large numbers by Bernoulli follows from the previous theorem. Let $\mu(A;n)$ denote hte number of occurrences of the event A in n independent trials, then

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{\mu(A;n)}{n} - p\right| > a\right) = 0.$$

Poisson's law of large numbers. Let the probability of observing the event A in the kth trial equal p_k, then

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{\mu(A;n)}{n} - \frac{p_1 + p_2 + \ldots + p_n}{n}\right| > a\right) = 0.$$