N.I. Lobachevsky State University
of Nizhni Novgorod

Probability theory and mathematical statistics:

Mathematical expectation of random variable

Associate Professor
A.V. Zorine

Suppose there were the following outcomes in 10 dice rolls:

⚀, ⚀, ⚃, ⚅, ⚁, ⚀, ⚅, ⚂, ⚀, ⚄

The average number of points is

$$\frac{1+1+4+6+2+1+6+3+1+5}{10} =$$
$$= 1 \cdot \frac{4}{10} + 2 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} + 4 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 6 \cdot \frac{2}{10} =$$
$$= 3$$

Let numbers $x_1, x_2, \ldots, x_k$ be the possible values of a discrete random variable X, $p_i = \mathrm{P}(X = x_i)$. Denote by n_i the number of occurrences of x_i in n independent trials. Then the sample average value of X is

$$x_1 \cdot \frac{n_1}{n} + x_2 \cdot \frac{n_2}{n} + \ldots + x_k \cdot \frac{n_k}{n}$$

On the other hand,

$$\frac{n_i}{n} \approx p_i \qquad \text{for } i = 1, 2, \ldots, k.$$

So, sample average value should be close to

$$x_1 p_1 + x_2 p_2 + \ldots + x_k p_k$$

Definition.

Let X be a discrete random variable with values $x_1, x_2, \ldots, x_k$, $p_i = \mathrm{P}(X = x_i)$. The number

$$\mathbf{E}X = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k$$

is called *the mathematical expectation* of the random variable X, or *the expected value* of X.

They say that the mathematical expectation is a characteristics of location of a random variable, its *mean value* in some sense. (Some other definitions of a *mean value* are possible).

Example. Let $I_A = I_A(\omega)$ denote the indicator of an event A, i.e.

$$I_A(\omega) = \begin{cases} 1 & \text{if } A \text{ occured } (\omega \in A) \\ 0 & \text{otherwise } (\omega \in \bar{A}) \end{cases}$$

Then

$$\mathbf{E} I_A = 0 \cdot \mathrm{P}(I_A = 0) + 1 \cdot \mathrm{P}(I_A = 1) = \mathrm{P}(A)$$

Example. Let X have uniform distribution on the set $\{1, 2, \ldots, n\}$. Then

$$\begin{aligned}
\mathrm{P}(X = i) &= \frac{1}{n}, \qquad i = 1, 2, \ldots, n; \\
\mathbf{E}X &= 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \ldots + n \cdot \frac{1}{n} \\
&= \frac{1}{n}(1 + 2 + \ldots + n) \\
&= \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}
\end{aligned}$$

For the number of points shown on a dice, X has uniform distribution on the set $\{1, 2, 3, 4, 5, 6\}$,

$$\mathbf{E}X = \frac{7}{2}.$$

Example. Let random variable X have the binomial distribution with parameters n, p,

$$\mathrm{P}(X = k) = C_n^k p^k (1-p)^{n-k}$$

Observe that

$$k \cdot C_n^k = \frac{k \cdot n!}{k!(n-k)!} = \frac{n \cdot (n-1)!}{(k-1)!(n-1-(k-1))!} = n \cdot C_{n-1}^{k-1}$$

Example. Let random variable X have the binomial distribution with parameters n, p,

$$\mathrm{P}(X = k) = C_n^k p^k (1-p)^{n-k}$$

Observe that

$$k \cdot C_n^k = \frac{k \cdot n!}{k!(n-k)!} = \frac{n \cdot (n-1)!}{(k-1)!(n-1-(k-1))!} = n \cdot C_{n-1}^{k-1}$$

$$\begin{aligned} \mathbf{E}X &= \sum_{k=0}^{n} k \cdot C_n^k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} n \cdot C_{n-1}^{k-1} p^{k-1+1} (1-p)^{n-1-(k-1)} = \\ &= np \sum_{m=0}^{n-1} C_{n-1}^{m} p^m (1-p)^{n-1-m} = np \end{aligned}$$

Example. Let X have the hyper-geometric distribution:

$$\mathrm{P}(X = k) = \frac{C_M^k C_{N-M}^{r-k}}{C_N^r}, \quad k = 0, 1, \dots, r$$

where $0 < M < N$, $0 < r < M$, $0 < r < N - M$. Then

$$\mathbf{E}X = \sum_{k=0}^{r} k \cdot \frac{C_M^k C_{N-M}^{r-k}}{C_N^r} = \sum_{k=1}^{r} \frac{M C_{M-1}^{k-1} C_{N-M}^{r-k}}{C_{N-1}^{r-1} \cdot \frac{N}{r}} =$$

$$= r\frac{M}{N} \sum_{m=1}^{r} \frac{C_{M-1}^{k-1} C_{N-1-(M-1)}^{r-1-(k-1)}}{C_{N-1}^{r-1}} = r\frac{M}{N}$$

Suppose there are events $A_1, A_2, \ldots, A_s$ such that $X(\omega)$ is constant on each of them:

$$X(\omega) = x'_i, \quad \text{for } \omega \in A_i, i = 1, 2, \ldots, s$$

(some of the numbers $x'_1, x'_2, \ldots, x'_s$ can be equal to each other). Then $\mathbf{E}X$ can be computed as

$$x'_1 \mathrm{P}(A_1) + x'_2 \mathrm{P}(A_2) + \ldots + x'_s \mathrm{P}(A_s).$$

Example. Two dice are rolled. Let random variable X count the number of ⚅s.
An outcome $\omega = (a_1, a_2)$ indicates the number a_1 of points on the first dice, and the number a_2 of points on the second dice. There are $6^2 = 36$ elementary outcomes in total.
Let $A_1 = \{(a_1, a_2) \colon a_1 \neq 6, a_2 \neq 6\}$. Then $X(\omega) = 0$ for each $\omega \in A_1$, $N(A_1) = 5^2 = 25$, $\mathrm{P}(A_1) = \frac{25}{36}$.
Let $A_2 = \{(a_1, a_2) \colon a_1 = 6, a_2 \neq 6\}$. Then $X(\omega) = 1$ for each $\omega \in A_2$, $N(A_2) = 5$, $\mathrm{P}(A_2) = \frac{5}{36}$.
Let $A_3 = \{(a_1, a_2) \colon a_1 \neq 6, a_2 = 6\}$. Then $X(\omega) = 1$ for each $\omega \in A_3$, $N(A_3) = 5$, $\mathrm{P}(A_3) = \frac{5}{36}$.
Let $A_4 = \{(a_1, a_2) \colon a_1 = 6, a_2 = 6\}$. Then $X(\omega) = 2$ for each $\omega \in A_4$, $N(A_4) = 1$, $\mathrm{P}(A_2) = \frac{1}{36}$.

Then $\mathrm{P}(X=0)=\mathrm{P}(A_1)=\frac{25}{36}$, $\mathrm{P}(X=1)=\mathrm{P}(A_2\cup A_3)=\frac{10}{36}$,
$\mathrm{P}(X=2)=\mathrm{P}(A_4)=\frac{1}{36}$,

$$\begin{aligned}\mathbf{E}X &= 0\cdot\mathrm{P}(X=0)+1\cdot\mathrm{P}(X=1)+2\cdot\mathrm{P}(X=2)=\frac{12}{36}\\ &= 0\cdot\mathrm{P}(A_1)+1\cdot\mathrm{P}(A_2\cup A_3)+2\cdot\mathrm{P}(A_4)\\ &= 0\cdot\mathrm{P}(A_1)+1\cdot\mathrm{P}(A_2)+1\cdot\mathrm{P}(A_3)+2\cdot\mathrm{P}(A_4)\end{aligned}$$

1. Let C be a constant. Then

$$\mathbf{E}C = C.$$

Proof. A constant takes single value wth the probability 1. Thus, $\mathbf{E}C = C \cdot 1$.

1. Let C be a constant. Then

$$\mathbf{E}C = C.$$

Proof. A constant takes single value wth the probability 1. Thus, $\mathbf{E}C = C \cdot 1$.

2. If $X \geqslant 0$, then

$$\mathbf{E}X \geqslant 0.$$

Proof. Obviously $X \geqslant 0$ means that $x_1 \geqslant 0, x_2 \geqslant 0, \ldots x_k \geqslant 0$, then

$$\mathbf{E}X = x_1p_1 + x_2p_2 + \ldots + x_kp_k \geqslant 0$$

3. Let X and Y be random variables, a, b constants. Then

$$\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y.$$

Proof. Let $x_1, x_2, \ldots, x_k$ be the possible values of random variable X, $y_1, y_2, \ldots, y_m$ be the possible values of Y. Then the possible values of $aX + bY$ are of the form $ax_i + by_j$.

3. Let X and Y be random variables, a, b constants. Then

$$\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y.$$

Proof. Let $x_1, x_2, \ldots, x_k$ be the possible values of random variable X, $y_1, y_2, \ldots, y_m$ be the possible values of Y. Then the possible values of $aX + bY$ are of the form $ax_i + by_j$. Then

$$\mathbf{E}(aX + bY) = \sum_{i=1}^{k} \sum_{j=1}^{m} (ax_i + by_j) \mathrm{P}(X = x_i, Y = y_j)$$

$$
\begin{aligned}
&= a\sum_{i=1}^{k} x_i \sum_{j=1}^{m} \mathrm{P}(X = x_i, Y = y_j) \\
&\quad + b\sum_{j=1}^{m} y_j \sum_{i=1}^{k} \mathrm{P}(X = x_i, Y = y_j) \\
&= a\sum_{i=1}^{k} x_i \mathrm{P}(X = x_i) + b\sum_{j=1}^{m} y_j \mathrm{P}(Y = y_j) \\
&= a\mathbf{E}X + b\mathbf{E}Y.
\end{aligned}
$$

Example. Let X have the binomial distribution with parameters n, p. Suppose X can be represented as

$$X = I_1 + I_2 + \ldots + I_n$$

where I_i is an indicator of a success in the ith trial. Then

$$\mathrm{P}(I_i = 1) = 1 - \mathrm{P}(I_i = 0) = p,$$

$$\mathbf{E}I_i = p,$$

$$\mathbf{E}X = \mathbf{E}I_1 + \mathbf{E}I_2 + \ldots + \mathbf{E}I_n = p + p + \ldots + p = np$$

Example. There are N balls in the urn, M blue balls and $N-M$ red balls. r balles are sampled without replacement ($r<M$, $r<N-M$). Let X denote the number of blue balls among r. Find $\mathbf{E}X$.
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 $i=1,2,\ldots,r$. Then

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

$$\mathbf{E}X=\mathbf{E}I_1+\mathbf{E}I_2+\ldots+\mathbf{E}I_r=r\frac{M}{N}.$$

Note that x has hyper-geometric distribution, so the known result is re-established.

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

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

Proof. Using notation from the previous proof, observe that the possible values of $X \cdot Y$ has the form $x_i y_j$. Independence means that

$$\mathrm{P}(X = x_i, Y = y_j) = \mathrm{P}(X = x_i)\mathrm{P}(Y = y_j).$$

Then

$$
\begin{aligned}
\mathbf{E}(XY) &= \sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j \mathrm{P}(X = x_i, Y = y_j) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j \mathrm{P}(X = x_i)\mathrm{P}(Y = y_j) \\
&= \sum_{i=1}^{k} x_i \mathrm{P}(X = x_i) \sum_{j=1}^{m} y_j \mathrm{P}(Y = y_j) \\
&= (\mathbf{E}X) \cdot (\mathbf{E}Y).
\end{aligned}
$$

5. (Cauchy-Bunyakowsky inequation)

$$(\mathbf{E}(XY))^2 \leqslant \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2)$$

Proof. For any real number t random variable $(X + tY)^2$ is nonegative,

$$\mathbf{E}(X + tY)^2 = \mathbf{E}(X^2) + 2t\,\mathbf{E}(XY) + t^2\,\mathbf{E}(Y^2) \geqslant 0.$$

This means that

$$D = (2\,\mathbf{E}(XY))^2 - 4\mathbf{E}(X^2)\mathbf{E}(Y^2) \leqslant 0.$$

Thus

$$(\mathbf{E}(XY))^2 \leqslant \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2)$$

6. Let X be a discrete random variable, $g(u)$ a function of real argument u, then

$$\mathbf{E}g(X) = \sum_{i=1}^{k} g(x_i)p_i.$$

Definiton

The number

$$m_k = \mathbf{E}X^k, \quad m_k^\circ = \mathbf{E}(X - \mathbf{E}X)^k$$

is called the kth moment of random variable X and the kth centered moment correspondingly.

Centered moments measure the average deviation of random variable X from its expected value.

Theorem

The best mean-square predictor of X is $\mathbf{E}X$.

Proof. We shall prove that for any real a

$$\mathbf{E}(X-\mathbf{E}X)^2 \leqslant \mathbf{E}(X-a)^2$$

In fact,

$$\begin{aligned}\mathbf{E}(X-a)^2 &= \mathbf{E}(X-\mathbf{E}X+\mathbf{E}X-a)^2 \\ &= \mathbf{E}(X-\mathbf{E}X)^2 - 2(\mathbf{E}X-a)\underbrace{\mathbf{E}(X-\mathbf{E}X)}_{\text{equals } 0} + \mathbf{E}\underbrace{(\mathbf{E}X-a)^2}_{\text{constant}} \\ &= \mathbf{E}(X-\mathbf{E}X)^2 + (\mathbf{E}X-a)^2 \geqslant \mathbf{E}(X-\mathbf{E}X)^2.\end{aligned}$$

For event A, $\mathrm{P}(A) > 0$, denote by

$$\mathbf{E}(X|A) = x_1\mathrm{P}(X = x_1|A) + x_2\mathrm{P}(X = x_2|A) + \ldots + x_k\mathrm{P}(X = x_k|A)$$

the *conditional expectation of random variable X given event A occurred.*

7. Let $H_1, H_2, \ldots, H_s$ be mutually exclusive events, $\mathrm{P}(H_r) > 0$, for $r = 1, 2, \ldots, r$ and $H_1 \cup H_2 \cup \ldots \cup H_s = \Omega$. Then

$$\mathbf{E}X = \mathrm{P}(H_1)\mathbf{E}(X|H_1) + \mathrm{P}(H_2)\mathbf{E}(X|H_2) + \ldots + \mathrm{P}(H_s)\mathbf{E}(X|H_s)$$

This is called the *repeated expectation formula*.

Proof.

$$
\begin{aligned}
\mathbf{E}X &= \sum_{i=1}^{k} x_i \mathrm{P}(X = x_i) \\
&= \sum_{i=1}^{k} x_i \sum_{r=1}^{s} \mathrm{P}(H_r)\mathrm{P}(X = x_i | H_r) \\
&= \sum_{r=1}^{s} \mathrm{P}(H_r) \sum_{i=1}^{k} x_i \mathrm{P}(X = x_i | H_r) \\
&= \sum_{r=1}^{s} \mathrm{P}(H_r)\mathbf{E}(X | H_r).
\end{aligned}
$$

Example. A worker manages n similar machines placed at equal distance a from each other. Find the mean transition of the worker between machines.

Enumerate machines from 1 to n. Introduce hypotheses $H_1, H_2, \ldots, H_n$, H_i stating that the worker is at the ith machine. Since all machines are similar, the probability that ith machine will demand worker's attention next equals $\frac{1}{n}$. Given that the worker is at the kth machine, his transition equal

$$\lambda = \begin{cases} (k-i)a & \text{for } k \geqslant i, \\ (i-k)a & \text{for } k < i. \end{cases}$$

Then

$$\begin{aligned}\mathbf{E}(\lambda|H_k) &= \frac{1}{n}\Big(\sum_{i=1}^{k}(k-i)a + \sum_{i=k+1}^{n}(i-k)a\Big)\\ &= \frac{a}{n}\Big(\frac{k(k-1)}{2} + \frac{(n-k)(n-k+1)}{2}\Big)\\ &= \frac{a}{2n}(2k^2 - 2(n+1)k + n(n+1))\end{aligned}$$

Since $\mathrm{P}(H_k) = \frac{1}{n}$,

$$\begin{aligned}\mathbf{E}\lambda &= \sum_{k=1}^{n}\frac{a}{2n^2}(2k^2 - 2(n+1)k + n(n+1))\\ &= \frac{a(n^2-1)}{3n}.\end{aligned}$$

For discrete random variable X with denumerable number of possible values assume that

$$\sum_{i=1}^{\infty} |x_i| p_i < \infty.$$

The by definition

$$\mathbf{E}X = \sum_{i=1}^{\infty} x_i p_i.$$

All properties formulated above hold for this kind of random variables.

Let X have Poisson distribution with parameter λ. Then

$$\begin{aligned}
\mathbf{E}X &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\
&= \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = \lambda.
\end{aligned}$$

Let X have geometric distribution with parameter p. Then

$$\begin{aligned}
\mathbf{E}X &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p \\
&= p \sum_{k=1}^{\infty} (-(1-p)^k)' \\
&= -p \Big(\sum_{k=1}^{\infty} (1-p)^k \Big)' \\
&= -p \Big(\frac{1-p}{1-(1-p)} \Big)' \\
&= -p \frac{-p-(1-p)}{p^2} = \frac{1}{p}.
\end{aligned}$$