Anna Janicka

Probability Calculus 2017/2018
Lecture 6

1. Expected Value

Is it worthwhile to gamble? To take part in lotteries? What does the answer to these questions depend on? Obviously, the willingness to engage in different lotteries depends on the psychological characteristics of the potential participant, whether he is risk-seeking or risk averse, etc. There exists, however, an objective measure, with which we can assess the essence of the game, if the game is more of a "loss" than a "win", i.e. the (average) profitability. This measure is the *expected value.*

Let us assume we are presented with the possibility to participate in the following lottery: we roll a die; if a six appears, we win \$100, if the result is different, we pay \$30. What should we expect of this game? If we were to play this game n times, then a six appears in on average $\frac{n}{6}$ cases, so our win after n games would intuitively be on average

$$\frac{n}{6}\cdot 100-\frac{5n}{6}\cdot 30=-\frac{50n}{6}<0,$$

so the result is not favorable. On average, we lose. In a single game, we expect that this loss would amount, on average, to

$$\frac{1}{6}\cdot 100+\frac{5}{6}\cdot(-30)=-\frac{50}{6}<0.$$

In other words, when thinking about the average, we weigh the possible results with the probabilities of obtaining specific results. This leads to the following definition:

Definition 1. *Let X be a random variable with a discrete distribution, concentrated on $S\subset\mathbb{R}$, and let $p_x=\mathbb{P}(X=x)$ for $x\in S$. We will say that the expected value of X is finite if the sum*

$$\sum_{x\in S}|x|p_x<\infty.$$

Then we can define this **expected value** *of X as*

$$\mathbb{E}X=\sum_{x\in S}xp_x.$$

The expected value is often referred to as the **mean value**. Note that we need the condition that the series of products of values and probabilities converges absolutely – otherwise it would be possible to obtain different values of $\mathbb{E}X$ depending on the order of summation. The expected value depends only on the distribution of the random variable.

Note that the expected value of a random variable which takes on values from a finite set always exists (the series converges, because it is finite).

Examples:

(1) For the most simple random variable, the Dirac delta: if $\mathbb{P}(X=a)=1$ for a given $a\in\mathbb{R}$, then $\mathbb{E}X=a\cdot 1=a$.

(2) We roll a die. Let X denote the number obtained. $\mathbb{P}(X=k)=\frac{1}{6}$ for $k=1,2,\ldots,6$, so

$$\mathbb{E}X=\sum_{i=1}^{6}i\cdot\frac{1}{6}=3.5.$$

(3) Let X be a random variable from a binomial distribution with parameters n and p. We have:

$$\mathbb{E}X=\sum_{k=0}^{n}k\mathbb{P}(X=k)=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}$$

$$\overset{l=k-1}{=}np\sum_{l=0}^{n-1}\binom{n-1}{l}p^l(1-p)^{n-1-l}=np.$$

(4) Let X be a random variable over $\{1,2,\ldots\}$ such that

$$\mathbb{P}(X=k)=\frac{1}{k(k+1)}=\frac{1}{k}-\frac{1}{k+1},\quad k=1,2,\ldots.$$

Then, the expected value of X does not exist:

$$\sum_{k=1}^{\infty}k\mathbb{P}(X=k)=\sum_{k=1}^{\infty}\frac{1}{k+1}=\infty.$$

(5) Let X be a random variable such that

$$\mathbb{P}\left(X=\frac{(-2)^k}{k}\right)=\frac{1}{2^k},\quad k=1,2,\ldots.$$

Then, the expected value of X does not exist:

$$\sum_{k=1}^{\infty}|x_k|\mathbb{P}(X=x_k)=\sum_{k=1}^{\infty}\frac{|-2|^k}{k}\cdot\frac{1}{2^k}=\sum_{k=1}^{\infty}\frac{1}{k}=\infty,$$

although the series from the definition of the expected value itself does converge:

$$\sum_{k=1}^{\infty}x_k\mathbb{P}(X=x_k)=\sum_{k=1}^{\infty}\frac{(-1)^k}{k}=-\ln 2.$$

For continuous random variables, the definition is similar (we change sums into integrals and probability into density):

Definition 2. *Let X be a random variable with density g. If*

$$\int_{\mathbb{R}}|x|g(x)dx<\infty,$$

then we will say the expected value of X exists. We define this **expected value** *of X as*

$$\mathbb{E}X=\int_{\mathbb{R}}xg(x)dx.$$

Again, as in the discrete case, the expected value depends only on the distribution of the random variable. At this point we can only note that both the discrete and the continuous definitions are special cases of a more general definition which "works" for all types of random variables, but is too complicated for this course. However, we will see below how the above definitions may be extended to "work" also in case of random variables which are neither discrete nor continuous.

Note that if a continuous random variable is limited (say, takes on only values from a limited range (a,b) with probability 1), then the expected value of the variable exists:

$$\int_{\mathbb{R}}|x|g(x)dx\leqslant\int_{\mathbb{R}}\max\{|a|,|b|\}g(x)dx=\max\{|a|,|b|\}.$$

Examples:

(1) Let X be uniformly distributed over (a,b). Then X is limited and thus the expected value exists:

$$\mathbb{E}X=\int_{\mathbb{R}}xg(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{a+b}{2}.$$

(2) Let X be a random variable from the standard normal distribution $\mathcal{N}(0,1)$. Then

$$\int_R|x|\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)dx=\frac{2}{\sqrt{2\pi}}\int_0^{\infty}x\exp(-x^2/2)dx=\frac{2}{\sqrt{2\pi}}\left(-e^{-x^2/2}\right)\Big|_0^{\infty}=\frac{2}{\sqrt{2\pi}},$$

so the expected value exists and since the variable is symmetric around 0, is equal to

$$\int_{\mathbb{R}}x\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)dx=0.$$

(3) Let X be a random variable from a Cauchy distribution, i.e. with density $g(x) = \frac{1}{\pi(1+x^2)}$. Although this variable too is symmetric around 0, the expected value does not exist, since

$$\int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2} dx = \frac{2}{\pi} \left(\ln(1+x^2)\right)\Big|_0^\infty = \infty.$$

Due to the specificity of the definition of the expected value, the properties of this operator are "inherited" from the operators used for definition, i.e. sums and integers – which means that the expected value is a linear operator and has the following properties, which simplify many calculations:

Theorem 1. *Let X and Y be random variables with expected values.*
(i) If $X \geqslant 0$, then $\mathbb{E}X \geqslant 0$.
(ii) If $X \leqslant Y$, then $\mathbb{E}X \leqslant \mathbb{E}Y$.
(iii) $\mathbb{E}X \leqslant \mathbb{E}|X|$.
(iv) If $a, b \in \mathbb{R}$, then $aX + bY$ has an expected value and $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$.
(v) If $X = 1_A$, then $\mathbb{E}X = \mathbb{P}(A)$.

Note that property *(iv)*, by induction, may be further generalized: if $X_1, X_2, \ldots, X_n$ are random variables with expected values and $a_1, a_2, \ldots, a_n \in \mathbb{R}$, then the variable $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ has an expected value, and

$$\mathbb{E}(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1\mathbb{E}X_1 + a_2\mathbb{E}X_2 + \ldots + a_n\mathbb{E}X_n.$$

In particular, we have

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_n.$$

This last property is extremely useful, as shown by the following examples:

(1) We roll a die 100 times and let X denote the sum of numbers obtained. If we wanted to calculate the expected value of X from the definition, we would first have to find the distribution of X, which is complicated and would lead to horrible calculations. We can, however, decompose X into a sum of the results of rolls $1, 2, \ldots, 100$: $X = X_1 + X_2 + \ldots X_{100}$, where X_i is the number obtained in the i-th roll, which permits us to write

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_{100} = 100 \cdot 3.5 = 350.$$

(2) There are 20 addressed envelopes and 20 letters. A secretary randomly puts letters into envelopes. Let X denote the number of letters which were placed in the correct envelopes. What is the expected value of X? Again, if we wanted to calculate the expected value from definition, we would need to go through complicated calculations to find the distribution of X (and then the expected value). We can, however, apply the same method and decompose X into a sum of variables $X_1 + X_2 + \ldots X_{20}$, where

$$X_i = 1_{\{i\text{–th letter is in the correct envelope}\}}, \\ = \begin{cases} 1 & \text{if the } i\text{-th letter is in the correct envelope,} \\ 0 & \text{otherwise.} \end{cases}$$

We then have

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_{20} = 20 \cdot \mathbb{P}(\text{a specific letter is in the correct envelope}) = \frac{20}{20} = 1.$$

We may wish to calculate not only the expected value of a random variable, but also of a function of this random variable. The following theorem justifies a simple procedure:

Theorem 2. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a Borel function.*

(i) If X is discrete, concentrated on the set S, and $p_x = \mathbb{P}(X = x)$ for $x \in S$, then the random variable $\phi(X)$ has an expected value if and only if

$$\sum_{x \in S} |\phi(x)| p_x < \infty,$$

and the expected value is equal to

$$\mathbb{E}\phi(X) = \sum_{x \in S} \phi(x) p_x.$$

(ii) If X is continuous with density g, then the random variable $\phi(X)$ has an expected value if and only if

$$\int_R |\phi(x)| g(x) dx < \infty,$$

and the expected value is equal to

$$\mathbb{E}\phi(X) = \int_R \phi(x) g(x) dx.$$

Examples:

(1) Let X denote the number obtained in a single die roll. The expected value of any transformation of X exists, since X is concentrated on a finite set. We have

$$\mathbb{E}X^2 = \sum_{k=1}^{6} k^2 \frac{1}{6} = \frac{91}{6}.$$

(2) Let X be uniformly distributed over $[0, \frac{\pi}{2}]$. Then the expected value of the sine of this angle is

$$\mathbb{E}\sin X = \int_0^{\frac{\pi}{2}} \sin x \cdot \frac{2}{\pi} dx = \frac{2}{\pi}.$$

(We do not need to check the existence of the expected value, since sine is a limited function and thus the result is obvious).

(3) The above theorem also permits to calculate the expected values of random variables which are neither discrete nor continuous, if only they can be presented as a function of such variables. For example, let X have a uniform distribution over $[0, 2]$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be given by the formula $\phi(x) = \min\{x, 1\}$. $\phi(X)$ is neither continuous nor discrete. However, it suffices that X is. Due to the fact that both X and $\phi(x)$ are limited, the expected value of $\phi(X)$ exists; it is equal to

$$\begin{aligned}\mathbb{E}\min\{X, 1\} = \mathbb{E}\phi(X) &= \int_{\mathbb{R}} \phi(x) g(x) dx = \int_0^2 \min\{x, 1\} \cdot \frac{1}{2} dx \\ &= \int_0^1 x \cdot \frac{1}{2} dx + \int_1^2 1 \cdot \frac{1}{2} dx = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.\end{aligned}$$

We will conclude our considerations of the methods of calculation of expected values by considering an alternative way of calculating $\mathbb{E}X$ for non-negative random variables – with the use of the cumulative distribution function (or rather $1 - \text{CDF}$).

Let us first consider a non-negative integer-valued random variable X. If we were to calculate $\mathbb{E}X$ from the definition, we would have:

$$\mathbb{E}X = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k).$$

We can decompose this sum in the following way:

$$\begin{array}{rllll}\mathbb{E}X = & \mathbb{P}(X=1)+ & & & \\ & \mathbb{P}(X=2)+ & \mathbb{P}(X=2)+ & & \\ & \mathbb{P}(X=3)+ & \mathbb{P}(X=3)+ & \mathbb{P}(X=3)+ & \\ & \mathbb{P}(X=4)+ & \mathbb{P}(X=4)+ & \mathbb{P}(X=4)+ & \mathbb{P}(X=4)+ \\ & \dots & & & \end{array}$$

The initial formula includes summation first within rows, and then taking the sum of the sums of rows. All the elements in the series are non-negative, therefore we may change the order of summation without a change in the outcome; we may first sum the elements within columns, and then take the sum of all columns.

The sum of elements in the first column is equal to $\mathbb{P}(X \geqslant 1)$ or $\mathbb{P}(X > 0)$, the sum of elements in the second column is equal to $\mathbb{P}(X \geqslant 2)$ or $\mathbb{P}(X > 1)$, etc. Therefore, the initial sum in the expected value may be written as

$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geqslant k) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

This result may be generalized to any non-negative valued random variable.

Theorem 3. *Let X be a non-negative random variable.*
(i) If $\int_0^\infty \mathbb{P}(X > t)dt < \infty$, then X has an expected value and

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)dt.$$

(ii) If $p \in (0, \infty)$ and $\int_0^\infty pt^{p-1}\mathbb{P}(X > t)dt < \infty$, then X^p has an expected value and

$$\mathbb{E}X^p = \int_0^\infty pt^{p-1}\mathbb{P}(X > t)dt.$$

In many cases, the above theorem allows to simplify calculations (avoid integration by parts) or allows the calculation of an expected value which otherwise could not be calculated given the definitions used until now (in particular, calculation of the expected value of random variables which are neither discrete nor continuous).

(1) Let X be a random variable from a geometric distribution. We have $\mathbb{P}(X = k) = p(1-p)^{k-1}$ for $k = 1, 2, \ldots$, and $P(X > k) = (1-p)^k$ for such k. We then have

$$\mathbb{E}X = \underline{\int_0^\infty \mathbb{P}(X > t)dt = \sum_{k=0}^{\infty} \int_k^{k+1} \mathbb{P}(X > t)dt = \sum_{k=0}^{\infty} \int_k^{k+1} (1-p)^k dt} = \sum_{k=0}^{\infty}(1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Note that we could have skipped the underlined part in the calculations and used the special case for non-negative integer-valued random variables.

(2) Let X be a random variable from an exponential distribution with parameter λ, i.e. with a CDF $F_X(t) = (1 - e^{-\lambda t})\mathbf{1}_{(0,\infty)}(t)$. We have

$$\mathbb{E}X = \int_0^\infty e^{-\lambda t}dt = \left(-\frac{1}{\lambda}e^{-\lambda t}\right)\Big|_0^\infty = \frac{1}{\lambda}.$$

(3) Let X be a random variable with density $g(x) = \frac{2}{x^2}1_{[2,\infty)}(x)$. Let $p \in (0, \infty)$. Does $\mathbb{E}X^p$ exist? We have

$$\mathbb{P}(X > t) = \begin{cases} 1 & \text{if } t < 2, \\ \frac{2}{t} & \text{if } t \geqslant 2. \end{cases}$$

On the basis of the above theorem, to determine the existence of $\mathbb{E}X^p$ it suffices to check the integral

$$\int_0^\infty pt^{p-1}\mathbb{P}(X > t)dt = \int_0^2 pt^{p-1}dt + \int_2^\infty pt^{p-1}\mathbb{P}(X > t)dt = 2^p + 2p\int_2^\infty t^{p-2}dt.$$

The last integral is finite if $p < 1$, and is equal to $2^p p/(1-p)$. Therefore, the expected value of X^p exists if and only if $p < 1$ and is equal to $2^p/(1-p)$.

(4) Let X be an exponentially distributed random variable with parameter 1, and let $Y = \max\{X, 2\}$. We have $F_Y(t) = (1 - e^{-t})\mathbf{1}_{[2,\infty)}(t)$. We can calculate $\mathbb{E}Y$ as

$$\int_0^\infty 1 - F_Y(t)dt = \int_0^2 1dt + \int_2^\infty e^{-t}dt = 2 + (-e^{-t})\Big|_2^\infty = 2 + e^{-2}.$$