Kłopoty z wartością oczekiwaną

Rafał SZTENCEL *

W poprzednim odcinku była mowa o medianie zmiennej losowej X. Tak nazywamy każdą liczbę a, dla której

$$P(X \leqslant a) \geqslant \frac{1}{2}, \quad P(X \geqslant a) \geqslant \frac{1}{2}.$$

Niestety, mediana na ogół nie jest wyznaczona jednoznacznie. Zajmiemy się teraz innym parametrem rozkładu: wartością oczekiwaną. Ta z kolei jest wyznaczona jednoznacznie, co nie oznacza, że nie ma z nią innych kłopotów. Po pierwsze, nie zawsze istnieje. Po drugie, problem stanowi podanie takiej definicji, która obejmowałaby wszystkie zmienne losowe (albo lepiej: wszystkie rozkłady prawdopodobieństwa).

Jeśli zmienna losowa X ma rozkład dyskretny, czyli przyjmuje wartości $x_1, x_2, \ldots, x_n, \ldots$ z prawdopodobieństwami odpowiednio $p_1, p_2, \ldots, p_n, \ldots$, to wartością oczekiwaną nazwiemy liczbę

$$EX = \sum_{i=1}^{\infty} x_i p_i, \tag{1}$$

o ile szereg po prawej stronie jest zbieżny bezwzględnie. Zwykła zbieżność nie wystarczy, jeśli bowiem szereg nie jest zbieżny bezwzględnie, to przestawiając jego wyrazy, można otrzymać dowolną sumę, a nawet szereg rozbieżny. Niewielki byłby pożytek z parametru rozkładu, który zależałby od tego, w jakim porządku wypiszemy wartości zmiennej losowej.

Być może lepiej nazywać EX wartością średnią, mamy bowiem po prostu do czynienia ze średnią ważoną liczb x_i, gdzie wagami są prawdopodobieństwa p_i. Jeśli X jest liczbą oczek wyrzuconych na uczciwej kostce, to $EX = 3\frac{1}{2}$, co – przyznajmy – jest wartością raczej nieoczekiwaną w tym prostym doświadczeniu.

Jeśli rozkład zmiennej losowej X jest ciągły i ma gęstość g, to na mocy definicji

$$EX = \int_{-\infty}^{\infty} x g(x)\,\mathrm{d}x, \tag{2}$$

gdzie żąda się zbieżności całki $\int_{-\infty}^{\infty} |x| g(x)\,\mathrm{d}x$.

Przypuśćmy, że w odległości 30 stóp od płotu, do którego przytwierdzono asa pik, stoi pan Andrzej Kmicic i popisuje się pewnością ręki. Jednak z rozmaitych przyczyn tor kuli tworzy z prostą przechodzącą przez pana Andrzeja i asa pik (ta prosta jest – jak można się domyślić – prostopadła do płotu) kąt θ, który jest zmienną losową o rozkładzie jednostajnym na przedziale $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$.

Jeżeli uznamy asa pik za punkt zerowy, to jako współrzędną punktu trafienia możemy przyjąć $\operatorname{tg}\theta$. Jak się wydaje, pan Andrzej średnio trafia w asa pik. Czy na pewno?

Nie ulega wątpliwości, że ze względu na symetrię medianą zmiennej losowej $X = \operatorname{tg}\theta$ jest 0. A jaka jest wartość średnia?

Niestety, definicja zmusza nas do wyznaczenia gęstości. Wyznaczymy dystrybuantę X:

$$F_X(t) = P(X \leqslant t) = P(\operatorname{tg}\theta \leqslant t) = P(\theta \leqslant \operatorname{arc\,tg} t) =$$
$$= \frac{\operatorname{arc\,tg} t + \frac{1}{2}\pi}{\pi} = \frac{1}{\pi}\operatorname{arc\,tg} t + \frac{1}{2} = \int_{-\infty}^{t} \frac{\mathrm{d}x}{\pi(1+x^2)}.$$

Otrzymaliśmy dystrybuantę w postaci całki z gęstości

$$g(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}.$$

Jednak całka

$$\int_{-\infty}^{\infty} \frac{x\,\mathrm{d}x}{\pi(1+x^2)}$$

nie jest zbieżna, jako że funkcja pierwotna, $\frac{\ln(1+x^2)}{2\pi}$, ma w plus i minus nieskończoności granicę nieskończoną. Wobec tego wartość średnia zmiennej losowej X nie istnieje.

Przypuśćmy teraz, że jeśli w n-tym rzucie nieskończonej serii rzutów symetryczną monetą wypadnie orzeł, to wygramy $2 \cdot 3^{-n}$ złotych. Niech X oznacza łączną wygraną. Jak obliczyć EX?

Teraz $0 \leqslant X \leqslant 1$, więc wartość średnia powinna istnieć, ale próba wyznaczenia rozkładu zmiennej losowej X i skorzystania z definicji prowadzi na manowce. Istotnie, można zauważyć, że X nie przyjmuje wartości z przedziału $(\frac{1}{3}, \frac{2}{3})$: jeśli w pierwszym rzucie wypadł orzeł, wygramy co najmniej $\frac{2}{3}$, w przeciwnym razie – co najwyżej $\frac{1}{3}$. Rozpatrując drugi, trzeci i dalsze rzuty, widzimy, że zabronione są przedziały $(\frac{1}{9}, \frac{2}{9})$, $(\frac{7}{9}, \frac{8}{9})$, etc. Suma długości zabronionych przedziałów da się obliczyć. Jest to

$$\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \ldots = 1.$$

Ale dystrybuanta musi być stała na przedziałach zabronionych, zatem gęstość – gdyby istniała – musiałaby tam być równa zeru. Wtedy jednak całka z gęstości na przedziale $[0, 1]$ byłaby równa zeru, a nie jedności.

Czytelnik zapewne zorientował się, że X przyjmuje wartości ze zbioru Cantora. Idąc tym tropem można pokazać, że żadna wartość zmiennej losowej X nie jest przyjmowana z dodatnim prawdopodobieństwem. Wzór (1) jest więc bezużyteczny.

Można argumentować, że X i $1 - X$ mają ten sam rozkład, ponieważ $1 - X$ jest wygraną w grze, w której zamieniono rolami orła i reszkę. Mielibyśmy zatem

$$EX = E(1-X) = 1 - EX = \frac{1}{2}.$$

Od porządnej teorii oczekujemy jednak porządnej definicji, dającej się zastosować do każdej zmiennej losowej; propozycja wymyślania za każdym razem nowych sztuczek nie jest poważna. Ale o tym za miesiąc.

*Instytut Matematyki Uniwersytetu Warszawskiego