Kłopoty z wartością oczekiwaną (II)

Rafał SZTENCEL [*]

W tym odcinku pokażemy, jak można obliczać wartość oczekiwaną zmiennej losowej X, mając do dyspozycji jej dystrybuantę F. Najpierw założymy dla wygody, że X jest nieujemna. Niech

$$G(t) = P(X > t) = 1 - F(t).$$

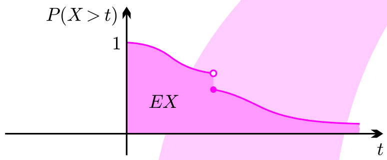
Funkcję G nazywamy ogonem zmiennej losowej.

Przypuśćmy, że rozkład X jest niezwykle regularny i istnieje ciągła, ograniczona gęstość g. Wtedy

$$G(t) = P(X > t) = \int_t^\infty g(s)\mathrm{d}s,$$

zatem na mocy podstawowego twierdzenia rachunku całkowego $G'(t) = -g(t)$. Do standardowego wzoru na wartość oczekiwaną zastosujemy całkowanie przez części:

$$EX = \int_0^\infty xg(x)\mathrm{d}x = -xG(x)\Big|_0^\infty + \int_0^\infty G(x)\mathrm{d}x.$$



Ale $0 \cdot G(0) = 0$, ponadto

$$xG(x) = x\int_x^\infty g(s)\mathrm{d}s \leqslant \int_x^\infty sg(s)\mathrm{d}s,$$

Jeśli zatem EX istnieje, to całka $\int_0^\infty sg(s)\mathrm{d}s$ jest zbieżna, więc

$$\lim_{x\to\infty} \int_x^\infty sg(s)\mathrm{d}s = 0,$$

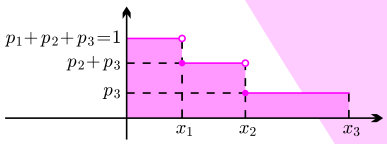
a stąd

$$\lim_{x\to\infty} xG(x) = 0.$$

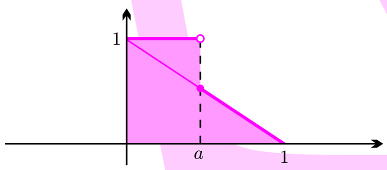
Ostatecznie

$$(*) \qquad EX = \int_0^\infty G(x)\mathrm{d}x = \int_0^\infty P(X > t)\mathrm{d}t.$$

Funkcja podcałkowa jest nierosnąca i prawostronnie ciągła, wobec tego ma co najwyżej przeliczalnie wiele punktów nieciągłości, a to oznacza, że na każdym przedziale ograniczonym jest całkowalna w sensie Riemanna. Można więc ostatecznie obliczyć wartość oczekiwaną zmiennej losowej, unikając całki Lebesgue'a, uważanej w pewnych kręgach za trudną, o czym świadczyłyby ponawiane co pewien czas próby eksmisji tej procedury całkowania na wyższe lata uniwersyteckich studiów matematycznych.



Warto w tym miejscu podkreślić, że bynajmniej nie udowodniliśmy wzoru $(*)$. Nie mamy przecież porządnej definicji wartości oczekiwanej. Pokazaliśmy tylko, że arbitralnie wprowadzony wzór z gęstością po przekształceniu nagle zaczyna mieć sens dla wszystkich rozkładów prawdopodobieństwa. Mogliśmy równie dobrze zacząć od rozkładów dyskretnych: niech $P(X = x_i) = p_i$, $i = 1, 2, \ldots, n$, $0 \leqslant x_1 < x_2 < \ldots < x_n$. Teraz



$$EX = x_1p_1 + x_2p_2 + \ldots + x_np_n =$$

$$= x_1(p_1 + \ldots + p_n) + (x_2 - x_1)(p_2 + \ldots + p_n) + \ldots + (x_n - x_{n-1})p_n =$$

$$= \int_0^\infty P(X > t)\mathrm{d}t.$$

Oto typowe zastosowanie wzoru $(*)$: niech X ma rozkład jednostajny na przedziale $[0, 1]$ i niech $a \in (0, 1)$. Obliczymy $E\max(X, a)$. Rozkład tej zmiennej losowej nie jest ani dyskretny, ani ciągły.

Jak widać, $E\max(X, a) = a + \frac{1}{2}(1 - a)^2$.

Wzór $(*)$ jest używany nie tylko w rachunku prawdopodobieństwa. W hydrologii stosuje się go do obliczania średniej głębokości jeziora, a tym samym – jego objętości. Wystarczy bowiem zinterpretować $G(t)$ jako stosunek powierzchni zamkniętej przez poziomicę $-t$ metrów do powierzchni jeziora.

W następnym odcinku zobaczymy, jak przenieść wzór $(*)$ na dowolne zmienne losowe.

[*] Instytut Matematyki Uniwersytetu Warszawskiego