Losowe rozbicia odcinka, proces Poissona i paradoks czasu oczekiwania

Rafał SZTENCEL

Rozkład wykładniczy ma pewną godną uwagi własność. Wyobraźmy sobie, że opisuje on czas oczekiwania T na pewne zjawisko, zatem

$$P(T > t) = e^{-\lambda t}, \quad t \geqslant 0,\ \lambda > 0.$$

Podaliśmy ogon rozkładu, czyli szansę, że czas oczekiwania przekroczy t. Przypuśćmy, że czekamy już s jednostek czasu. Jaka jest szansa, że poczekamy jeszcze co najmniej t? Obliczamy za pomocą prawdopodobieństwa warunkowego:

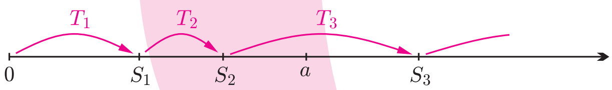
$$P(T > s+t | T > s) = \frac{P(T > s+t, T > s)}{P(T > s)} =$$

$$= \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} =$$

$$= e^{-\lambda t} = P(T > t), \quad s, t \geqslant 0.$$

Rezultat nie zależy od s, dlatego powyższą własność określa się jako brak pamięci dla rozkładu wykładniczego. Wynika z niej, że nawet gdy czekamy już bardzo długo, średni dalszy czas oczekiwania jest i tak równy $ET = 1/\lambda$.

Rozpatrzmy teraz ciąg niezależnych czasów oczekiwania $T_1, T_2, \ldots$, o tym samym rozkładzie co T. Mogą to być, na przykład, czasy oczekiwania na zgłoszenia do centrali telefonicznej. Proces startuje w chwili 0, a kolejne zgłoszenia nadchodzą w momentach $S_n = T_1 + \ldots + T_n$. W chwili a rozpoczynamy obserwację procesu. Jaki jest średni czas oczekiwania $m(a)$ na najbliższe zgłoszenie?



Możliwe są dwie odpowiedzi: (1) z własności braku pamięci wynika, że $m(a) = 1/\lambda$, bo nie jest istotne, kiedy miało miejsce ostatnie zgłoszenie poprzedzające a; (2) ale przecież średni odstęp między zgłoszeniami jest równy $1/\lambda$, i skoro punkt a znajduje się wewnątrz odcinka o średniej długości $1/\lambda$, to można się spodziewać, że $m(a)$ będzie z grubsza dwukrotnie mniejsze.

Zanim rozstrzygniemy ten paradoks, mała dygresja o losowym rozbiciu okręgu. Jeśli wybierzemy losowo (dokładniej: zgodnie z rozkładem jednostajnym) i niezależnie punkty A i B na okręgu o długości 1, to średnie długości łuków AB i BA będą równe $1/2$. A jeśli przed losowaniem ustalimy punkt P? Teraz łuk zawierający ten punkt będzie miał średnią długość $2/3$. Istotnie, trzy punkty wybrane losowo i niezależnie z okręgu dzielą go na łuki o średniej długości $1/3$. Wystarczy teraz obrócić okrąg tak, by np. pierwszy wylosowany punkt pokrył się z P.

Odpowiednio wykształcona intuicja probabilistyczna powinna podpowiedzieć, że dłuższy przedział ma większe szanse przykrycia punktu P niż krótszy. Można pokazać, że rozkład długości łuku AB jest jednostajny, ale rozkład długości łuku zawierającego punkt P ma gęstość $2x$ dla $x \in [0, 1]$.

Analogiczne zjawisko powoduje, że argumenty (1) i (2) nie są sprzeczne. Przedział zawierający punkt a jest średnio dłuższy niż $1/\lambda$ i odpowiedź (1) nie prowadzi do paradoksu, bowiem rzeczywiście średni czas oczekiwania na najbliższe zgłoszenie po chwili a jest równy $1/\lambda$.

Natomiast średni czas od ostatniego zgłoszenia poprzedzającego a do a jest równy $\frac{1}{\lambda}(1 - e^{-\lambda a})$, więc dla dużych a przedział nakrywający punkt a ma długość niewiele mniejszą niż $2/\lambda$. Oto szkic dowodu.

Po pierwsze, można wyliczyć gęstość g_n i dystrybuantę F_n zmiennej losowej S_n:

$$g_n(x) = \frac{\lambda^n x^{n-1}}{(n-1)!},$$

$$F_n(x) = 1 - e^{-\lambda x}\left(1 + \lambda x + \ldots + \frac{(\lambda x)^{n-1}}{(n-1)!}\right), \quad x \geqslant 0.$$

Stąd liczba N_a zgłoszeń w przedziale $[0, a]$ ma rozkład Poissona z parametrem λa:

$$P(N_a = k) = P(S_k \leqslant a, S_{k+1} > a) =$$

$$= P(\{S_k \leqslant a\} \setminus \{S_{k+1} \leqslant a\}) =$$

$$= F_k(a) - F_{k+1}(a) = \frac{(\lambda a)^k}{k!} e^{-\lambda a}.$$

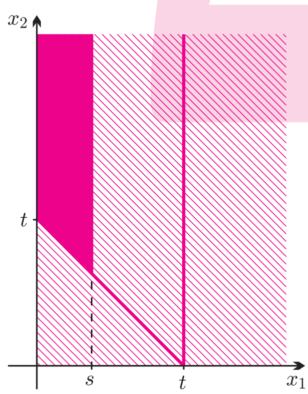
Po drugie, jeśli wiemy, że w przedziale $[0, a]$ było k zgłoszeń, to układają się one tak, jak k punktów rzuconych losowo i niezależnie[1]. Dlatego przedział $[0, a]$ zostanie podzielony na $(k+1)$ przedziałów o średniej długości $a/(k+1)$. Łatwo to zobaczyć, gdy odwołamy się do losowego rozbicia okręgu punktami $X_1, X_2, \ldots, X_{k+1}$. Jasne jest, że wszystkie łuki mają równe średnie długości. Po rozcięciu okręgu w punkcie X_{k+1} otrzymujemy losowe rozbicie odcinka.

Teraz za pomocą uogólnienia wzoru na prawdopodobieństwo całkowite obliczamy średni odstęp a od poprzedzającego zgłoszenia. Jest on równy

$$\sum_{k=0}^{\infty} \frac{a}{k+1} \frac{(\lambda a)^k}{k!} e^{-\lambda a} = \frac{1}{\lambda} \sum_{k=0}^{\infty} \frac{(\lambda a)^{k+1}}{(k+1)!} e^{-\lambda a} =$$

$$= \frac{1}{\lambda}\left(e^{\lambda a} - 1\right) e^{-\lambda a} = \frac{1}{\lambda}\left(1 - e^{-\lambda a}\right).$$

Podaliśmy tu jedną z możliwych konstrukcji procesu Poissona. Z formalnego punktu widzenia jest on rodziną zmiennych losowych $(N_t)_{t \geqslant 0}$, liczących ilość zgłoszeń do chwili t.



[1] Dla $k = 1$ można wykazać, że $P(X_1 < s | X_1 \leqslant t, X_1 + X_2 > t) = \frac{s}{t}$, gdy $0 \leqslant s \leqslant t$ na podstawie rysunku (patrz obok); wystarczy skorzystać z faktu, że gęstość łącznego rozkładu (X_1, X_2), wyrażająca się w pierwszej ćwiartce wzorem $\lambda^2 e^{-\lambda(x_1+x_2)}$, jest stała na odcinkach $x_1 + x_2 = c$, $x_1, x_2 \geqslant 0$. W wyższych wymiarach idea dowodu jest ta sama.