Probabil

Zestaw 8
Semestr zimowy 2023/2024
Kraków, 10 stycznia 2024

THEORETICAL COMPUTER SCIENCE
Jagiellonian University

Rozkład wykładniczy i proces Poissona

Zacznijmy od prostej obserwacji na temat splatania rozkładów. Niech X, Y będą niezależnymi zmiennymi losowymi ciągłymi. Mamy

$$F_{X+Y}(z) = \mathrm{P}(X + Y < z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy dx.$$

Korzystając z niezależności zmiennych X, Y mamy $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. Wyłączając stałą przed wewnętrzną całkę mamy

$$F_{X+Y}(z) = \cdots = \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{z-x} f_Y(y) dy dx.$$

Zauważmy, że wewnętrzna całka jest dystrybuantą Y obliczoną w punkcie $z - x$. Czyli

$$F_{X+Y}(z) = \cdots = \int_{-\infty}^{\infty} f_X(x) F_Y(z - x) dx$$

Weźmy teraz pochodną po z stronami. Wiemy, że z pochodną można wejść pod całkę, więc

$$f_{X+Y}(z) = F'_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

Ostatnią całkę nazywamy *splotem* funkcji f_X i f_Y. Jest to przydatny wzór, który ma bardzo naturaną interpretację graficzną Polecamy materiał na kanale 3Blue1Brown: https://youtu.be/IaSGqQa5O-M?si=1XCmcn-DFmZCEsPU.

Zadanie 1. Policz funkcje tworzące momenty dla rozkładu wykładniczego i jednostajnego.

Zadanie 2. Niech X oraz Y będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego z parametrem 1. Znajdź gęstość i dystrybuantę zmiennej $Z = X + Y$.

Zadanie 3. Niech X, Y będą zmiennymi losowymi o rozkładzie wykładniczym z parametrem 1. Oblicz

$$\lim_{R \to \infty} \mathrm{P}(\min(X, Y) > R \mid X + Y > 2R).$$

Zadanie 4. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego z parametrem 1. Znajdź gęstość i dystrybuantę zmiennej $Z = X_1 + \cdots + X_n$.

Zadanie 5. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego z parametrem 1. Wyznacz wartość oczekiwaną k-tej wartości w posortowanym ciągu tych zmiennych.

Zadanie 6. Rzucamy monetą, gdzie orzeł wypada z prawdopodobieństwem p. Każdy rzut trwa czas będący zmienną losową z rozkładu wykładniczego z parametrem 1. Pokaż, że łączny czas oczekiwania na orła jest zmienną z rozkładu wykładniczego o parametrze p.

Zadanie 7. Niech $(X_n)_{n=1}^{\infty}$ będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego z parametrem 1. Dla danego k niech $N(k) = \min\{n : \sum_{i=1}^{n} X_i > k\}$. Wyznacz $\mathrm{E}(N(k))$ na dwa sposoby:

(i) bezpośrednio korzystając z gęstości wyliczonych w jednym z poprzednich zadań i wzoru $\mathrm{E}(N(k)) = \sum_{j=1}^{\infty} \mathrm{P}(N(k) \geqslant j)$;

(ii) nie licząc ani jednej całki i korzystając ze wzoru $\mathrm{E}(\sum_{i=1}^{N(k)} X_i) = \mathrm{E}(X_1)\,\mathrm{E}(N(k))$ (wcześniej go dowodząc).

Zadanie 8. Niech X będzie zmienną losową z rozkładu jednostajnego na przedziale $[0, 1]$. Wykaż, że zmienna losowa $-\ln X$ ma rozkład wykładniczy z parametrem 1.

Zadanie 9. Niech $(X_n)_{n=1}^{\infty}$ będą niezależnymi zmiennymi losowymi z rozkładu jednostajnego na przedziale $[0, 1]$. Dla danego $k \in (0, 1)$ niech $N(k) = \min\{n : \Pi_{i=1}^{n} X_i < k\}$. Wyznacz $\mathrm{E}(N(k))$.

Mając dwie zmienne losowe X, Y o rozkładach ciągłych definiujemy wartość oczekiwaną warunkową w następujący sposób

$$\mathrm{E}(X \mid Y)(y) = \mathrm{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx dy.$$

Zadanie 10. Niech T_1, T_2 będą zmiennymi o rozkładzie wykładniczym z parametrem 1 i niech $T = T_1 + T_2$. Oblicz $\mathrm{E}(T_1 \mid T)$ i $\mathrm{E}(T \mid T_1)$.

Zadanie 11. Wykaż, że jeśli $\{N(t) \mid t \geqslant 0\}$ jest stochastycznym procesem takim, że

(i) $N(0) = 0$,

(ii) czasy pomiędzy zliczanymi zdarzeniami są niezależne i mają rozkład wykładniczy z parametrem λ,

to $\{N(t) \mid t \geqslant 0\}$ jest procesem Poissona z parametrem λ. (To jest twierdzenie 8.11 z książki.)

Zadanie 12. Niech $\{N(t) \mid t \geqslant 0\}$ będzie procesem Poissona z parametrem λ. Niech $Y_1, Y_2, \ldots$ będą czasami przyjścia kolejnych zdarzeń (pierwszego, drugiego,...). Oblicz:

(i) $P(N(3) = 5)$

(ii) $P(N(3) - N(1) = 4)$

(iii) $P(N(1) \geqslant 1 \,\cap\, N(3) \leqslant 2)$

(iv) $P(Y_1 > \frac{1}{2})$

(v) $P(Y_1 > \frac{1}{2} \,\cap\, Y_3 \leqslant \frac{3}{2})$

(vi) $P(Y_1 > 1 \,\cap\, Y_2 - Y_1 \leqslant \frac{1}{2})$

Zadanie 13. Rozważ proces Poissona $\{N(t) \mid t \geqslant 0\}$ z parametrem λ.

(i) Czy dla ustalonego $c > 0$ rodzina $\{c \cdot N(t) \mid t \geqslant 0\}$ jest procesem Poissona?

(ii) Oblicz $P(N(2021) = 31 \mid N(2022) = 31)$.

Zadanie 14. Niech $\{N(t)\}$ będzie procesem Poissona z parametrem 2023.

1. Jeśli $N(2023) = 2023$, to z jakim prawdopodobieństwem $N(2022) = 2022$?
2. Z jakim prawdopodobieństwem nie zajdzie żadne zdarzenie w przedziale czasowym $[2023, 2024]$? Z jakim prawdopodobieństwem nie zajdzie żadne zdarzenie w przedziale czasowym $[2023, 2024]$ jeśli wiemy, że $N(2023) = 2023$?
3. Dla $n > 1$, niech X_n będzie czasem pomiędzy $n-1$-wszym i n-tym zdarzeniem w procesie i niech X_1 to po prostu czas pierwszego zdarzenia. Oblicz $P(X_1 > X_2)$ i $P(X_1 + X_2 > X_3)$.

Zadanie bonusowe do spisania

Część 1 (1 punkt)

Rozważmy problem Set Cover. Instancją tego poblemu jest $(U, \mathcal{S}, \{c_S\}_{S \in \mathcal{S}})$: skończone uniwersum U, rodzina jego podzbiorów $\mathcal{S}$, oraz nieujemna waga c_S dla każdego $S \in \mathcal{S}$. Szukamy podrodziny $\mathcal{S}' \subseteq \mathcal{S}$ o minimalnej wartości $\sum_{S \in \mathcal{S}'} c_S$ takiej, że dla każdego $u \in U$ istnieje $S \in \mathcal{S}'$ taki, że $u \in S$.

Problem Set Cover jest obliczeniowo trudny. Przedstawimy jeden z klasycznych schematów przybliżania rozwiązania trudnych problemów. Najpierw zamieniamy problem na program liniowy całkowitoliczbowy. W naszym przypadku program wygląda tak:

$$\min \sum_{S \in \mathcal{S}} c_S x_S \quad \text{pod warunkiem, że}$$

$$\begin{aligned} x_S &\geqslant 0 && \text{dla każdego } S \in \mathcal{S} \\ \sum_{u \in S} x_S &\geqslant 1 && \text{dla każdego } u \in U. \end{aligned}$$

Następnie rozwiązujemy ten program w liczbach rzeczywistych korzystając np. z metody Simplex, to znaczy znajdujemy najlepsze wartości dla zmiennych x_S. Na koniec zaokrąglamy rozwiązanie w przemyślany sposób. Teraz przedstawimy pomysł na zaokrąglanie dla problemu Set Cover.

Ustalmy instancję $(U, \mathcal{S}, \{c_S\}_{S \in \mathcal{S}})$ problemu Set Cover i ustalmy rozwiązanie powyższego programu liniowego $(x_S)_{S \in \mathcal{S}}$.

Teraz wykonujemy następujący randomizowany algorytm zaokrąglania.

1. Dla każdego $S \in \mathcal{S}$ ustalamy $y_S = 1$.
2. Dla każdego $S \in \mathcal{S}$ wybieramy wartość Z_S z rozkładu wykładniczego z parametrem 1. Wszystkie wybory są niezależne.
3. Dla każdego $u \in U$ niech $S \in \mathcal{S}$ będzie jednym ze zbiorów, który świadczy wartość $\min\left\{\frac{Z_S}{x_S} \mid u \in S\right\}$. Ustalamy $y_S = 1$.

Problem 1. Wykaż, że wartość oczekiwana $\mathrm{E}(\sum_{S \in \mathcal{S}} c_S y_S) \leqslant (\ln s + 1)\mathrm{OPT}$, gdzie s jest maksymalnym rozmiarem zbioru w $\mathcal{S}$, a OPT jest wartością optymalnego rozwiązania.

Innymi słowy otrzymaliśmy algorytm aproksymacyjny dla problemu Set Cover ze współczynnikiem aproksymacji $\ln s + 1$. Co ciekawe jest on optymalny w następującym sensie. Wiadomo, że nie istnieje ogólny algorytm aproksymacyjny ze współczynnikiem $(1 - o(1)) \ln |U|$ o ile $P \neq NP$.

Część 2 (1 punkt)

Procesem Markowa nazywamy proces stochastyczny czasu ciągłego $\{X(t)\,|t \geqslant 0\}$, który spełnia tzw. Własność Markowa

$$\forall_{s,t} : \mathrm{P}\left(X\left(s+t\right) = x | X\left(u\right), 0 \leqslant u \leqslant t\right) = \mathrm{P}\left(X\left(s+t\right) = x | X\left(t\right)\right)$$

Na proces ten można patrzeć jak na złożenie dwóch procesów:

1. Łańcucha Markowa danego macierzą przejścia $P_{i,j}$ (proces dyskretnego czasu)
2. Wektora parametrów $(\theta_1, \theta_2, \ldots)$, który określa ile czasu proces spędza w stanie i nim go zmieni. Konkretnie czas postoju w stanie i określony jest rozkładem wykładniczym z parametrem θ_i.

Rozkład stacjonarny dla Procesu Markowa nie jest tym samym co rozkład stacjonarny dla Łańcucha Markowa. Podobnie jak w przypadku dyskretnym rozkład ten określa asymptotyczne prawdopodobieństwo tego, że proces znajduje się w konkretnym stanie. Rozkład stacjonarny spełnia następującą zależność:

$$\pi_i \theta_i = \sum_k \pi_k \theta_k p_{k,i}$$

gdzie $p_{k,i}$ są elementami macierzy przejścia Łańcucha Markowa.

Standardową notacją dla kolejek jest $Y/Z/n$, gdzie Y określa dystrybucję klientów przychodzących, Z dystrybucję czasu obsługi klienta, zaś n liczbę dostępnych serwerów. Literą M oznacza się standardowy proces Markowa, więc napis $M/M/n$ oznacza kolejkę obsługiwaną przez n serwerów, gdzie klienci przychodzą zgodnie z procesem Poissona oraz czas obsługi jest z rozkładu wykładniczego. Zwykłe kolejki określają także kolejność przetwarzania, jeśli takowa nie jest wyspecyfikowana, standardowym podejściem jest FIFO.

Problem 2. Przeczytaj i opanuj materiał z sekcji 8.5 (Continuous Time Markov Processes) oraz 8.6.1 i 8.6.2 o kolejkach Markowa.

Problem 3 (Model Ehrenfesta)**.** Pojemnik zawierający w sumie n cząstek, jest przedzielony na dwie części błoną półprzepuszczalną. Każda z cząstek, niezależnie od pozostałych, przechodzi do przeciwnej części po czasie będącym zmienną losową z rozkładu wykładniczego z parametrem 1.

(i) Wyznacz rozkład stacjonarny tego procesu.
(ii) Który stan w rozkładzie stacjonarnym ma największe prawdopodobieństwo, dlaczego?

Problem 4. Maszyna, nim ulegnie awarii, pracuje nieprzerwanie przez liczbę godzin będącą zmienną losową X z rozkładu wykładniczego z parametrem 0,1. W momencie wystąpienia awarii natychmiastowo podejmowane są działania naprawcze, które zajmują Y godzin, gdzie Y jest zmienną losową z rozkładu wykładniczego z parametrem 0,9 oraz niezależną od X. Po dokonaniu naprawy cały cykl się powtarza.

(i) Pokaż, że ten proces jest Procesem Markowa.
(ii) Wyznacz macierz przejścia Łańcucha Markowa stowarzyszonego z tym procesem.
(iii) Wyznacz rozkład stacjonarny Procesu Markowa oraz rozkład stacjonarny Łańcucha Markowa stowarzyszonego z tym procesem.
(iv) Asymptotycznie jaki jest procent czasu spędzony w naprawie?
(v) Podczas pracy maszyna przynosi dochód w wysokości \$60 na godzinę pracy. Naprawa kosztuje \$30 za godzinę. Jaki jest długoterminowy średni dochód netto?

Problem w części pierwszej jest wart 1 punkt. Problemy w części drugiej są warte $\frac{1}{3}$ punktu każdy. Przypominamy, że rozwiązania spisane **na komputerze** należy wysyłać na adres podany na stronie internetowej kursu w terminie do wtorku 9 stycznia o godzinie 23:59.