Wstęp do rachunku prawdopodobieństwa
Wykład 13: Prawo Wielkich Liczb. Centralne Twierdzenie Graniczne.

Twierdzenie (Prawo Wielkich Liczb). *Niech $X_1, X_2, \ldots,$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie, dla których $\mathbb{E}X_1 = \mu$, $VarX_1 = \sigma^2 < \infty$. Wtedy dla zmiennej losowej $S_n = X_1 + \cdots + X_n$ i dowolnej stałej $\epsilon > 0$ zachodzi*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0.$$

Definicja. *Zmienna losowa X ma* **rozkład normalny**, *co zapisujemy $X \sim N(\mu, \sigma^2)$, jeśli jej gęstość dana jest wzorem*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \quad dla\ x \in (-\infty, \infty)\,.$$

Jeśli $X \sim N(\mu, \sigma^2)$, to $\mathbb{E}X = \mu$ i $VarX = \sigma^2$.

Jeżeli $X \sim N(0,1)$, to mówimy, że X ma **standaryzowany rozkład normalny**. *Dystrybuantę standaryzowanego rozkładu normalnego zwykle oznaczamy przez $\Phi(\cdot)$.*

Uwaga. Funkcji Φ nie możemy zapisać przy pomocy funkcji elementarnych. Jej wartości można odszukać w tablicach lub obliczyć korzystając z dostępnych programów, np.

`http://onlinestatbook.com/2/normal_distribution/standard_normal.html`

Najczęściej używane wartości Φ to: $\Phi(1.2816) \sim 0.9$, $\Phi(1.6449) \sim 0.95$, $\Phi(1.9600) \sim 0.975$, $\Phi(2.3263) \sim 0.99$ i $\Phi(2.5758) \sim 0.995$.

Twierdzenie (Centralne Twierdzenie Graniczne). *Niech $X_1, X_2, \ldots,$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie, dla których $\mathbb{E}X_1 = \mu$, $VarX_1 = \sigma^2 < \infty$ i $\mathbb{E}|X_1|^3 < \infty$. Wtedy dla zmiennej losowej $S_n = X_1 + \cdots + X_n$ i dowolnej stałej $x \in \mathbb{R}$ zachodzi*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx = \Phi(x)\,.$$

Centralne Twierdzenie Graniczne (w skrócie CTG) odnosi sie do sytuacji gdy n i $\mathbb{E}S_n$ są dostatecznie duże. Zwykle CTG dobrze przybliża rozkład S_n gdy $n \geqslant 30$ i $\mathbb{E}S_n \geqslant 10$. Jeśli stosujemy je do przybliżania zmiennej losowej S_n, która przyjmuje tylko wartości całkowite, należy pamiętać o uwzględnieniu "poprawki" wynoszącej $1/2$. Na przykład, chcąc użyć CTG do oszacowania prawdopodobieństwa $\mathbb{P}(S_n = k)$, należy je zastąpić przez prawdopodobieństwo $\mathbb{P}(k - 1/2 < S_n \leqslant k + 1/2)$, które z kolei szacujemy przez

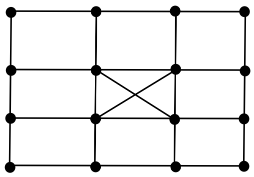
$$\begin{aligned}\mathbb{P}(k - 1/2 < S_n \leqslant k + 1/2) &= \mathbb{P}\left(\frac{k - n\mu - 1/2}{\sigma\sqrt{n}} < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant \frac{k - n\mu + 1/2}{\sigma\sqrt{n}}\right) \\ &\simeq \Phi\Big(\frac{k - n\mu + 1/2}{\sigma\sqrt{n}}\Big) - \Phi\Big(\frac{k - n\mu - 1/2}{\sigma\sqrt{n}}\Big)\,.\end{aligned}$$

1. Zadania na ćwiczenia

Zadanie 1. Dla łańcucha Markowa o podanej poniżej macierzy przejścia wyznacz przybliżone prawdopodobieństwo przejścia ze stanu 4 do stanu 2 w 2000 krokach. Błąd przybliżenia jest nieistotny.

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{6} & 0 & \frac{1}{2} & \frac{2}{6} \\ 0 & \frac{1}{6} & \frac{2}{6} & \frac{1}{2} \end{bmatrix}$$

Zadanie 2. Rozważmy błądzenie na poniższym grafie (proszę zauważyć, że jest to graf z zadania domowego). Załóżmy, że startujemy z lewego górnego rogu.



(a) Wyznacz przybliżone prawdopodobieństwo, że po 1234567 krokach znajdziemy się w jednym ze środkowych wierzchołków. Błąd przybliżenia jest nieistotny.
(b) Uzasadnij, że wartość oczekiwana czasu potrzebnego do odwiedzenia każdego wierzchołka jest mniejsza niż 2000.
(c) Niech T oznacza czas potrzebny do odwiedzenia każdego wierzchołka. Wykaż, że $\mathbb{P}(T \geqslant 6000) < \frac{1}{3}$.

Zadanie 3. Pchła spaceruje losowo po cyklu o 5 wierzchołkach. Uzasadnij, że z prawdopodobieństwem większym niż $\frac{2}{3}$ po 150 skokach wszystkie wierzchołki będą odwiedzone.

Zadanie 4. Mysz i kot błądzą losowo (i niezależnie) na tym samym spójnym grafie G. Gdy znajdą się w tym samym wierzchołku, kot zjada mysz. Uzasadnij, że oczekiwana liczba kroków, po których mysz zostanie zjedzona, jest mniejsza niż $8e(G)^2v(G)^2$.

Zadanie 5. (rezerwowe) Oceń poprawność poniższych zdań. Odpowiedź należy poprzeć albo uzasadnieniem ogólnym, albo przykładem, albo kontrprzykładem.

(a) Jeśli łańcuch Markowa posiada jednostajny rozkład stacjonarny, to $p_{ij} = p_{ji}$ dla dowolnych stanów i, j tego łańcucha.
(b) Jeśli łańcuch Markowa ma dokładnie jeden rozkład stacjonarny, to jest ergodyczny.
(c) Jeżeli $(\pi_1, \ldots, \pi_t)$ jest rozkładem stacjonarnym łańcucha Markowa, to dla każdego stanu j zachodzi $p_{1j}^n \to \pi_j$ przy $n \to \infty$.
(d) Jeśli bładzimy klasycznie na grafie G, który jest 4-regularny, to oczekiwana liczba kroków, po których wszystkie wierzchołki będą odwiedzone, jest mniejsza niż 5000.

---

2. Zadania domowe

Uwaga Do rozwiązania (ostatniej!) kartkówki 4 czerwca przyda się prosty kalkulator, np. taki jaki standardowo znajduje się w smartfonie.

Zadanie 1. Oznaczmy przez S_{100} liczbę szóstek wyrzuconych w 100 rzutach symetryczną kostką. Użyj kalkulatora i tablic statystycznych do:

(i) wyznaczenia wartości $\mathbb{P}(S_{100} = 15)$;
(ii) oszacowania $\mathbb{P}(S_{100} = 15)$ przy użyciu CTG.

Zadanie 2. Niech S_{400} oznacza liczbe orłów wyrzuconych w 400 rzutach monetą.

(i) Oszacuj, używając kalkulatora i tablic statystycznych, $\mathbb{P}\,(205 < S_{400} \leqslant 222)$.
(ii) Znajdź, przy pomocy CTG najmniejszą wartość k, dla której $\mathbb{P}\,(S_{400} \leqslant k) > 0.95$.
(iii) Rozwiąż zadanie (ii) używając nierówności Czebyszewa.

Zadanie 3. Ile co najmniej razy musimy rzucić symetryczna kostką, aby prawdopodobienstwo tego, że szóstka wypadnie więcej niż 20% było mniejsze niż 0.1? Rozwiąż to zadanie używając (i) nierówności Czebyszewa, (ii) CTG.

Zadanie 4. Mierzony w minutach czas T, jaki spędza codziennie pan January czekając na autobus, ma rozkład wykładniczy z parametrem 0.1, tzn. gęstość T wyraża się wzorem

$$f(x) = \begin{cases} 0.1e^{-x/10} & \text{dla } x \geqslant 0 \\ 0 & \text{dla } x \leqslant 0. \end{cases}$$

Używając CTG oszacuj prawdopodobieństwo, że w ciagu ostatnich 100 dni pan January czekał na autobus w sumie dłużej niż 1200 minut (tzn. średnio dłużej niż 12 minut).

Wskazówka Zarówno $\mathbb{E}T$ jak i $\mathrm{Var}T$ można odnaleźć w wypisach do jednego z poprzednich wykładów.