Metody Probabilistyczne i Statystyka

LISTA 10 do wykładu 11

Informatyka algorytmiczna (I st.) WIT – 2021/2022

Zadanie 1. Na wykładzie omawialiśmy metodę szacowania wielkości $|p - \hat{p}|$, gdzie p to rzeczywiste prawdopodobieństwo zdarzenia A, zaś $\hat{p}$ to częstość wystąpienia A w N niezależnych próbach. Zakładaliśmy, że N jest stosunkowo duże, więc można było użyć przybliżenia przez rozkład normalny.
Opracuj alternatywne podejście dla małych N oparte na nierówności Czebyszewa.

Zadanie 2. Samorząd gminny musi przyjąć lub odrzucić uchwałę X. Sprawa jest trudna, nie jesteśmy w Szwajcarii (gdzie w takich sprawach robi się referenda lokalne), więc postanowiono spytać się N losowych osób i postąpić zgodnie z wolą większości z nich.
Jednak zanim podejmie się decyzję, to trzeba mieć gwarancję, że decyzja odzwierciedla wolę całej lokalnej społeczności z prawdopodobieństwem co najmniej 0.9. Jak dobrać liczbę N? Liczba ta powinna być jak najmniejsza (dotarcie do losowych mieszkańców kosztuje dużo pracy).
Warunki:

- nie znamy frakcji p zwolenników X w całej populacji,
- jeśli p jest odległe od $\frac{1}{2}$ to wiadomo, że bezcelowe jest pytanie dużej liczby mieszkańców. Z drugiej strony, jak $p \approx \frac{1}{2}$ to należy się liczyć z koniecznością spytania dużej liczby osób.

Zadanie 3. Rozważamy sumę N niezależnych zmiennych losowych X_i , każda o wartości oczekiwanej μ i wariancji σ. Niech $X = \frac{1}{N}\sum_{i=1}^{N} X_i$. Wariancja X to $\frac{\sigma^2}{N}$.
Na podstawie eksperymentu chcielibyśmy wyznaczyć σ. Jak słyszeliście, estymator nieobciążony dla wariancji σ dany jest wzorem:

$$\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

gdzie $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ jest zaobserwowaną średnią pomiarów.
Zbadać dla małych N jaka jest wartość tego estymatora dla sumy N zmiennych X_i z rozkładem $\Pr(X_i = 0) = \Pr(X_i = 1)$. Dla porównania policzyć to samo dla wzoru

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

Zadanie 4. (Zadanie programistyczne) Zakładamy, że błąd w systemie występuje zgodnie z rozkładem Poissona dla współczynnika λ. Jednak λ nie jest stałe: jeśli między i-tym a $i + 1$-szym błędem, czas był krótszy niż pomiędzy wystąpieniem $i-1$-szego a i-tego błędu, to $\lambda_{i+1} = 2 \cdot \lambda_i$. Jeśli był dłuższy, to $\lambda_{i+1} = 0.5 \cdot \lambda_i$. Początkowo wartość λ to $\lambda_0 = 1$.
Zadanie: oszacować oczekiwany czas na wystąpienie 20 błędów.