Wprowadzenie do symulacji i metod Monte Carlo 22.10.2019

Lista nr 1, ver 2

Paweł Lorek

Szereg harmoniczny: $H_n = \sum_{k=1}^{n} \frac{1}{k}$.

Liczby Stirlinga drugiego rodzaju: $\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n$, jest to liczba sposobów podziału zbioru n elementów na k niepustych zbiorów podzbiorów.

1. Niech T będzie zmienną losową przyjmującą wartości całkowite nieujemne. Pokaż, że $ET = \sum_{k=0}^{\infty} P(T > k)$.

2. Udowodnij następujący

   Lemat 1 (Nierówność Markowa) *Niech X będzie nieujemną zmienną losową. Wtedy $\forall r > 0$*
   $$P(X \geq r) \leq \frac{EX}{r}$$

3. Udowodnij następujący

   Lemat 2 (Nierówność Czebyszewa) *Niech X będzie zmienną losową o średniej $\mu < \infty$ i wariancji $\sigma^2 < \infty$. Wtedy $\forall r > 0$*
   $$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

4. Pokaż, że dla każdego $x \in \mathbb{R}$ mamy $1 + x \leq e^x$

5. Udowodnij następujący

   Lemat 3 (Nierówność Chernoffa) *Niech X_i będą niezależnymi zmiennymi losowymi o rozkładzie $P(X_i = 1) = p_i = 1 - P(X_i = 0)$. Niech $S = \sum_{i=1}^{n} X_i$ oraz $\mu = ES = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p_i$. Wtedy dla każdego $\varepsilon > 0$ zachodzi*
   $$P(S > (1+\varepsilon)\mu) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}} \right)^{\mu}$$
   oraz
   $$P(S < (1-\varepsilon)\mu) \leq \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}} \right)^{\mu}$$

6. Pokaż, że poprzednie zadanie implikuje

$$P(S < (1-\varepsilon)\mu) \leq e^{\frac{-\mu\varepsilon^2}{2}}$$

7. Udowodnij:

$$\forall(1 \leq k \leq n): \qquad \left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$$

8. Niech X będzie zmienną losową o rozkładzie $Erl(n, \lambda)$ tj. mającą gęstość ($\lambda > 0$)

$$f(x) = \frac{1}{(n-1)!}\lambda^n x^{n-1} e^{-\lambda x}, \quad x \geq 0.$$

Pokaż, że dystrybuanta wyraża się wzorem

$$F(x) = P(X \leq x) = 1 - \sum_{r=0}^{n-1} \frac{(\lambda x)^r e^{-\lambda x}}{r!}.$$

9. Niech S_n będzie zbiorem wszystkich permutacji zbioru $\{1, 2, \ldots, n\}$. Niech Y oznacza zmienną losową o rozkładzie jednostajnym na S_n, tzn. $\forall \sigma \in S_n$ $Pr(Y = \sigma) = \frac{1}{n!}$. Dla $\sigma \in S_n$ zdefiniujmy

$$f(\sigma) = \#\{i : \sigma(i) = i\},$$

tzn. $f(\sigma)$ to liczba punktów stałych permutacji σ. Pokaż, że $Ef(Y) = 1$.

10. Niech Y będzie zmienną losową o rozkładzie jednostajnym na S_n (jak w poprzednim zadaniu). Dla $\sigma \in S_n$ zdefiniujmy

$$C_m(\sigma) = \#\{\text{ cykle długości } m\}$$

(liczba cykli długości m) oraz $C(\sigma) = \sum_{m=1}^{n} C_m(\sigma)$ (liczba cykli).

- Ile jest średnio cykli długości m w losowej permutacji n elementow? Tzn. ile wynosi $EC_m(Y)$?
- Ile jest średnio cykli w losowej permutacji? Tzn. ile wynosci $EC(Y)$?

11. Niech Y będzie zmienną losową o rozkładzie jednostajnym na S_{100}. Pokaż, że

$$Pr(Y \text{zawiera cykl długości } \geq 51) = H_{100} - H_{50} \approx .6881721793.$$

12. (zagadka) W pewnym pomieszczeniu jest 100 szuflad, w nich jest rozmieszczonych 100 kartek z numerami od 1 do 100 losowo (jednostajnie wybrana permutacja). Jest tam też 100 więźniów ponumerowanych od 1 do 100. Każdy z nich musi znaleźć kartkę ze swoim numerem. Zasady są takie, że każdy więzień może otworzyć maksymalnie 50 szuflad oraz, że więźniowie nie mogą się ze sobą komunikować (wchodzą oni do pomieszczenia pojedynczo w losowej kolejności). Wcześniej mogą oni ustalić wspólną strategię. Podaj jak najlepszą taką strategię i wylicz prawdopodobieństwo tego, że wszyscy odnajdą kartki ze swoimi numerami.

13. Niech $U_1, \dots, U_n \in [0,1]$. Niech $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_i \le t)$. Zdefiniujmy

$$D_n = \sup_{0\le t\le 1} |\hat{F}_n(t) - t|.$$

Niech $U_{(1)}, \dots, U_{(n)}$ będzie statystyką pozycyjną próby $U_1, \dots, U_n$. Pokaż, że $D_n = \max(D_n^+, D_n^-)$, gdzie

$$D_n^+ = \max_{1\le i\le n}\left(\frac{i}{n} - U_{(i)}\right), \qquad D_n^- = \max_{1\le i\le n}\left(U_{(i)} - \frac{i-1}{n}\right).$$

14. (test serii, średnia). Niech n, t, d będą całkowitymi liczbami dodatnimi. Załóżmy, że mamy obserwacje $U_1, \dots, U_{nt} \in [0,1]$. Dzielimy je grupy po t:

$$\mathbf{U}_1 = (U_1, \dots, U_t), \mathbf{U}_2 = (U_{t+1}, \dots, U_{2t}), \qquad \mathbf{U}_n = (U_{(n-1)t+1}, \dots, U_{nt}).$$

Rozpatrzmy kostkę $[0,1]^t$: dzielimy każdy z boków na d równych odcinków $[i/d, (i+1)/d], i = 0, \dots, d-1$ i otrzymujemy $k = d^t$ “mini” kostek. Załóżmy, że mamy jakąs numerację kostek $1, 2, \dots, k$ Niech

$$X_i = \#\{j : \mathbf{U}_j \text{ jest w minikostce o numerze } i\}.$$

Zdefiniujmy

$$T = \sum_{j=1}^{k} \frac{(X_j - \frac{n}{k})^2}{\frac{n}{k}}.$$

Zakładając, że $U_1, \dots, U_n$ są iid o rozkładzie $U(0,1)$ pokaż, że $ET = k - 1$

15. Rozmieszczamy losowo n kul w n urnach (tj. każde rozmieszczenie ma prawd. n^{-n}. Niech A_k oznacza zdarzenie, że jest dokładnie k urn pustych.

Pokaż, że

$$Pr(A_k) = \binom{n}{k}\sum_{i=0}^{n-k}(-1)^i\binom{n-k}{i}\left(1 - \frac{k+i}{n}\right)^n.$$

Wybrane zadania teoretyczne ze skryptu

http://www.math.uni.wroc.pl/~rolski/Zajecia/sym.pdf, Rozdział II.5

16. Niech $X_1, X_2, \dots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $(\bar{M})$, gdzie M jest liczbą naturalną oraz $\bar{M} = \{0, 1, \dots, M-1\}$. Niech $M = 2^k$ oraz $b(x)$ dla $x \in \bar{M}$ będzie rozwinięciem dwójkowym przedstawionym w reprezentacji k cyfr zero lub jeden. Pokazać, że ciąg zmiennych losowych zero-jedynkowych zdefiniowany przez

$$\xi_1, \xi_2, \dots = b(X_1), b(X_2)\dots$$

jest ciągiem prób Bernoulliego. Pokazać na przykładzie, ze tak nie musi być jeśli M nie jest potęgą dwójki.

17. Przez *funkcję błedu* $\mathrm{erf}(x)$ rozumie się

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt,$$

oraz przez *uzupełniającą funkcję błedu* $\mathrm{erfc}(x)$

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt.$$

Pokazać, że jeśli $\Phi(x)$ jest dystrybuantą standardowego rozkładu normalnego, to

$$\Phi(x) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}(x/\sqrt{2}) = \frac{1}{2}\mathrm{erfc}(-x/\sqrt{2}).$$

18. Obliczamy całkę $\bar{f} = \int_{[0,1)^d} f(\mathbf{u})\, d\mathbf{u}$ przy założeniu, że $\sigma_f^2 = Var f(\mathbf{U}) < \infty$. Korzystając z nierówności Czebyszewa pokaż, że z prawdopodobieństwem co najmniej $1-\delta$ mamy (gdzie $\mathbf{U}_1, \mathbf{U}_2, \ldots$ są iid o tym samym rozkładzie co $\mathbf{U}$).

$$\left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i) - \bar{f} \right| < \frac{\sigma_f}{\sqrt{\delta n}}.$$