Big Data
Lista zadań

Jacek Cichoń, WiT, PWr, 2022/23

1 Wstęp

Zadanie 1 — Pobierz plik z kilkoma dramatami Szekspira ze strony wykładu. Wybierz jeden z dramatów.

1. Oczyść wybrany plik. Podziel go na słowa.
2. Usuń z niego "Stop Words" i usuń z niego słowa o długości mniejszej lub równej 2.
3. Zbuduj chmurę wyrazów (word cloud) z otrzymanej listy. Możesz skorzystać np. z serwisu `http://www.wordclouds.com/`

Celem tego zadania jest wygenerowanie mniej więcej takiego obrazka (dla poematu "Pan Tadeusz"):

Zadanie 2 — To jest kontynuacja poprzedniego zadania.

1. Zastosuj część funkcji które napisałeś do realizacji poprzedniego zadania do wyznaczenia indeksów TF.IDF dla wszystkich wyrazów z dokumentów w dramatów Szekspira znajdujących się w pliku ze strony wykładu.
2. Zbuduj chmury wyrazów oparte o TF.IDF dla wszystkich rozważanych dramatów.

Zadanie 3 — Pokaż, że jeśli chcesz jednoznacznie wyreprezentować każdą z liczb ze zbioru $\{0, 1, \ldots, n\}$ za pomocą b bitów to $b \geqslant \lceil \log_2(n+1) \rceil$.

Zadanie 4 — Pokaż, że jeśli $x = \sum_{k=0}^{s} a_k 2^k$, gdzie $a_i \in \{0, 1\}$ oraz $a_s = 1$ to $s = \lceil \log_2(x+1) \rceil$

Zadanie 5 — Rozważmy następującą modyfikację licznika Morrisa: ustalamy liczbę $\alpha > 0$ oraz rozważamy tak oprogramowany licznik:

```
init :: C =0
onInc :: if (random() < (1/(1+α))^C) then C = C+1
onGet :: return (?????)
```

Niech C_n oznacza wartość zmiennej C po n wywołaniach metody onInc.

1. Wyznacz $\mathrm{E}\left[(1+\alpha)^{C_n}\right]$

2. Uzupełnij funkcję onGet tak aby otrzymać nieobciążony estymator liczby użyć metody onInc.

Zadanie 6 — Niech C_n będzie wartością klasycznego licznika Morris'a po n krotnym wywołaniu funkcji `onInc()`.

1. Pokaż, że $\mathrm{E}\left[4^{C_n}\right] = 1 + \frac{3}{2}n(n+1)$.
2. Pokaż, że $\mathrm{var}\left[2^{C_n}\right] = \frac{1}{2}n(n-1)$.
3. Skorzystaj z nierówności Jensena dla wartości oczekiwanej zmiennej losowej do pokazania, że $\mathrm{E}\left[C_n\right] \leqslant \log_2(n+1)$.

Zadanie 7 — Załóżmy, że $X_1, \ldots, X_m$ są niezależnymi zmiennymi losowymi o wartości oczekiwanej μ oraz wariancji σ^2. Niech

$$L = \frac{X_1 + \ldots + X_m}{m} \; .$$

1. Pokaż/sprawdź, że $\mathrm{E}\left[L\right] = \mu$ oraz $\mathrm{var}\left[L\right] = \frac{1}{m}\sigma^2$.
2. Pokaż, że $\Pr[|L - \mu| \geqslant \epsilon\mu] \leqslant \frac{\sigma^2}{\epsilon^2}$.

Zadanie 8 — Rozważamy ciąg $B_1, \ldots, B_n$ niezależnych zdarzeń, takich, że $\Pr[B_1] = \ldots = \Pr[B_n] = \frac{3}{4}$. Niech X oznacza liczbę sukcesów, czyli $X = \sum_{i=1}^{n} X_i$, gdzie $X_i = 1$ jeśli zaszło zdarzenie B_i oraz $X_i = 0$ w przeciwnym przypadku.

1. Korzystając z nierówności Czernoffa dla rozkładu dwumianowego pokaż, że

$$\Pr[X \leqslant \frac{1}{2}n] \leqslant \exp\left(-\frac{n}{24}\right)$$

2. Niech $\delta > 0$. Pokaż, że jeśli $n \geqslant 24 \ln \frac{1}{\delta}$, to $\Pr[X \leqslant \frac{n}{2}] \leqslant \delta$.
3. Skorzystaj z następującej wersji nierówności Czernoffa

$$\Pr[X \leqslant \mu - \lambda], \Pr[X \geqslant \mu + \lambda] \leqslant \exp\left(-\frac{2\lambda^2}{n}\right)$$

dla zmiennej losowej X o rozkładzie dwumianowym $\mathrm{Binom}(n, \mu)$ do wzmocnienia wyników z poprzednich dwóch punktów.

Zadanie 9 — Niech $x_1, \ldots, x_n$ będzie ciągiem liczb rzeczywistych. Rozważamy dwie funkcje $f(x) = \sum_{i-1}^{n} |x_i - x|$ oraz $g(x) = \sum_{i=1}^{n}(x_i - x)^2$

1. Pokaż, że funkcja g osiąga minimum w średniej arytmetycznej liczb $x_1, \ldots, x_n$
2. Pokaż, że funkcja h osiąga minimum w medianie ciągu $x_1, \ldots, x_n$.
   Wskazówka: Możesz założyć, że $x_1 \leqslant x_x \leqslant \cdots \leqslant x_n$. Przyjrzy się najpierw pomocniczej funkcji $\phi(x) = |x_1 - x| + |x_n - x|$.

Zadanie 10 — Niech $x_1, \ldots, x_n$ będzie ciągiem liczb rzeczywistych oraz niech $a < b$ będą dowolnymi liczbami rzeczywistymi. Załóżmy, że

$$|\{i \in \{1, \ldots, n\} : x_i \in (a, b)\}| > \frac{n}{2} \; .$$

Pokaż, że wtedy mediana ciągu $x_1, \ldots, x_n$ należy do odcinka (a, b).
Wskazówka: Rozważ oddzielnie przypadek parzystego i nienarzystego n.

2 Hashinig

Zadanie 11 — **("Rolling hash")** Rozważamy metodę haszowania opartą na wzorze

$$h_{r,p}([x_0, \ldots, x_k]) = \sum_{i=0}^{k} x_i \cdot r^i \mod p$$

1. Zastosuj metodę Hornera do implementacji tej metody haszowania i oszacuj złożoność obliczeniową tej metody.
2. Załóżmy, że p jest liczbą pierwszą. Rozważamy ciąg $[x_0, \ldots, x_m]$. Niech $0 \leqslant a < b < m$. Pokaż, że można wyznaczyć $h_{r,p}[x_{a+1}, \ldots, x_{b+1}])$ można wyznaczyć z $h_{r,p}[x_a, \ldots, x_{a+b}])$ w stałym czasie.
3. Załóżmy, że p jest liczbą pierwszą. Niech $\vec{x}$ i $\vec{y}$ będą ciągami długości r. Losujemy z jednakowym prawdopodobieństwem liczbę r ze zbioru $\{0, \ldots, p-1\}$. Pokaż, że
$$\Pr[h_{r,p}(\vec{x}) = h_{r,p}(\vec{y})] \leqslant \frac{r-1}{p} \ .$$
4. Zapoznaj się z algorytmem Rabina-Karpa wyszukiwania wzorca w w ciągu t. Pokaż, że jeśli to tego algorytmu zastosujemy funkcję haszującą $h_{r,p}$ z p będącym liczbą pierwszą taką, że $p > |t|^2$ zaś r jest losową liczbą ze zbioru $\{0, \ldots, p-1\}$, to algorytm ten działa w średnim czasie $O(|s|+|t|)$ ($|x|$ oznacza tu długość ciągu x).

Zadanie 12 — Do n urn wkładamy niezależnie k kul (rozważamy rozkład jednostajny). Niech $L_{n,k}$ oznacza wartość oczekiwaną liczby pustych urn. Oblicz

1. $\lim_{n\to\infty} \mathrm{E}\left[\frac{L_{n,n}}{n}\right]$
2. $\lim_{n\to\infty} \mathrm{E}\left[L_{n,n\ln n}\right]$
3. $\lim_{n\to\infty} \mathrm{E}\left[\frac{1}{\sqrt{n}}(n - L_{n,\sqrt{n}})\right]$

Zadanie 13 — Rozważamy dwie zmienne losowe X, Y o wartościach w zbiorze $\{1, \ldots, n\}$. Niech $\Pr[X = i] = \Pr[Y = i] = p_i$ dla $i \in \{1, \ldots, n\}$.

1. Pokaż, że $\Pr[X = Y] = \sum_{i=1}^{n} p_i^2$
2. Pokaż, że $\Pr[X = Y]$ przyjmuje wartość minimalną dla rozkładu jednostajnego na $\{1, \ldots, n\}$.

Zadanie 14 — Niech $f(x) = \ln(x)\ln(1-x)$ dla $x \in (0,1)$.

1. Pokaż, że $f(x) = f(\frac{1}{2} - x)$.
2. Pokaż, że $\lim_{x\to 0} f(x) = 0$.
3. Pokaż, że f osiąga maksimum w punkcie $x = \frac{1}{2}$.
4. Naszkicuj wykres funkcji f.

Zadanie 15 — Oprogramuj w języku Python Filtr Blooma. Skorzystaj z funkcji MurMurHash z biblioteki mmh3 (instalacja: pip install mmh3). Do implementacji tablicy wykorzystaj tablicę bitów (skorzystaj z bibliteki bitarray). Filtr zrealizuj jako obiekt. Przetestuj działanie filtru Blooma na słowach z pliku Hamlet.txt (użyj 8 funkcji haszujących, ustaw rozmiar tablicy na liczę słów w Hamlet.txt).

3 Sampling

Zadanie 16 — Pokaż, że jeśli $\mathcal{H}$ jest $(k+1)$-niezależną rodziną haszującą, to jest również k-niezależną rodziną haszującą.

Zadanie 17 — Pokaż, że 2-niezależna rodzina funkcji haszujących jest rodziną uniwersalną.

Zadanie 18 — Pokaż, że jeśli $\mathcal{H}$ jest 2- niezależną rodziną funkcji haszujących z U do V, to dla dowolnych $x \in U$ oraz $v \in V$ mamy
$$\Pr_{h \leftarrow \mathcal{H}}[h(x) = y] = \frac{1}{|V|} \ .$$

Zadanie 19 — Załóżmy, że $1 < n < p$. Niech X będzie zmienną losową o rozkładzie jednostajnym na zbiorze $\{0, \ldots, p-1\}$. Niech $\phi(x) = x \mod n$. Wyznacza rozkład zmiennej losowej $\phi \circ X$.

Zadanie 20 — Załóżmy, że $h : U \to R$ jest funkcją różnowartościową. Pokaż, że jednoelementowa rodzina $\mathcal{H} = \{h\}$ jest rodziną uniwersalną ale nie jest 2-niezależna.

Zadanie 21 — Jak z liczb $S = \sum_{i=1}^n x_i$, $SS = \sum_{i=1}^n x_i^2$ oraz n możesz wyznaczyć wariancję $\frac{1}{n}\sum_{i=1}^n (x-\mu)^2$, gdzie μ oznacza średnią $\mu = \frac{1}{n}\sum_{i=1}^n x_i$?

Zadanie 22 — Zaimplementuj prostą (z użyciem generatora liczb pseudolosowych po wczytaniu każdego elementu) wersję algorytmu **R** Vittera. Pobierz z sieci notowania dzienne bitcoina z ostatnich 5 lat (możesz posłużyć się biblioteką Pandas języka Python), wydobądź z danych notowania otwarcia, wygeneruj losową próbkę 40 elementów i wygeneruj wykresy notowań i losowej próbki.

Zadanie 23 — Ustalmy liczby naturalne $1 \leqslant k \leqslant n$. Rozważamy przestrzeń probabilistyczną $[n]^k = \{X \subseteq \{1,\ldots n\} : |X| = k\}$ z prawdopodobieństwem jednostajnym ($\Pr[X] = \binom{n}{k}^{-1}$). Ustalmy zbiór $A \subseteq \{1,\ldots,n\}$ taki, że $|A| = k-1$. Rozważmy następujący proces: (1) losujemy $B' \in [n]^k$; (2) z wylosowanego B' usuwamy losowo wybrany element (każdy z prawdopodobieństwem $\frac{1}{k}$) i otrzymujemy zbiór B.

1. Sprecyzuj powyższe rozumowanie korzystając z przestrzeni probabilistycznej $[n]^k \times \{1,\ldots,k\}$
2. Wyznacz prawdopodobieństwo otrzymania zbioru A.

Zadanie 24 — **(Własności dystrybuanty)** Celem tego zadania jest przypomnienie sobie podstawowych własności dystrubuant zmiennych losowych o wartościach w liczbach rzeczywistych.

1. Niech F_X będzie dystrybuantą zmiennej losowej X (czyli $F_X(x) = \Pr[X \leqslant x]$). Pokaż, że $\lim_{x\to-\infty} F(x) = 0$ oraz $\lim_{x\to\infty} F(x) = 1$
2. Niech F_X będzie dystrybuantą zmiennej losowej X. Pokaż, że F jest prawostronnie ciągła w każdym punkcie x, czyli, że dla każdego $a \in \mathbb{R}$ mamy $\lim_{x\to a+} F(x) = F(a)$.
3. Załóżmy, że F_X jest ostro rosnąca oraz, że $\mathrm{rgn}(F) = (0,1)$. Niech U będzie zmienną losową o rozkładzie jednostajnym w odcinku $(0,1)$. Pokaż, że zmienna losowa $F^{-1} \circ U$ ma taki sam rozkład, co zmienna X.
4. Załóżmy, że F jest dystrybuantą zmiennej losowej X. Uogólnioną odwrotnością dystrybuanty F nazywamy funkcję zdefiniowaną wzorem

   $$F^{\leftarrow}(p) = \inf\{x : F(x) \geqslant p\}\ .$$

   Zbadaj podstawowe własności tej funkcji (np. $F^{\leftarrow}$ jest niemalejąca, $F^{\leftarrow}(F(x)) \leqslant x$, $F(F^{\leftarrow}(p)) \geqslant p$) oraz pokaż, że $F^{\leftarrow} \circ U$ ma taki sam rozkład co zmienna X, gdzie U, podobnie jak w poprzednim punkcie, ma rozkład jednostajny na odcinku $(0,1)$.

Zadanie 25 — Zaimplementuj podstawową wersję algorytmu Bravermana, Ostrovsky'iego, Zaniolo z pracy *Optimal sampling from sliding windows.*

1. Sprawdź poprawność działania implementacji generując odpowiedni histogram (możesz użyć polecenia `plt.hist(sample, density=True, bins=N)` języka Python) ze wskazywanych przez ten algorytm elementów.
2. Przetestuj swoją implementację dla okna długości 5 i po zaobserwowaniu 10000 elementów. Po wczytaniu każdego elementu zapamiętaj w jakiejś strukturze pozycję wskazywanego elementu. Zastosuj test χ^2 p-wartością $p = 0.01$ dla hipotezy zerowej

   $$H_0 = \text{próbka pochodzi z rozkładu jednostajnego}$$

   (wartość krytyczna dla tej wartości p oraz 4 stopni swobody wynosi 11.345). Możesz też posłużyć się biblioteką `scipy.stats` do przeprowadzenia tego testu.

Zadanie 26 — **(Paradoks urodzinowy)** Niech $(X_k)_{k\geqslant 1}$ będzie rodziną niezależnych zmiennych losowych o wartościach w zbiorze $\{1,\ldots,n\}$. Niech $G_{n,k}$ oznacza zdarzenie "$(\forall i,j \leqslant k)(i \neq j \to X_i \neq X_j)$". Wiemy, że

$$\Pr[G_{n,k}] = \prod_{i=1}^{k}\left(1 - \frac{i}{n}\right)\ .$$

1. Naszkicuj wykres ciągu $\Pr[G_{365,k}]$ dla $k \in \{1,\ldots,365\}$. Wskazówka: Skorzystaj z dowolnego pakietu
2. Pokaż, że $1 - x > \exp(-x - x^2)$ dla $x \in (0, 0.5)$. Wskazówka: Wszystkie chwyty są dozwolone

3. Korzystając z nierówności z punktu (1) znajdź oszacowanie dolne na $\Pr[G_{n,k}]$ dla $k < \frac{n}{2}$.
4. Pokaż, że

$$\lim_{n\to\infty} \Pr[G_{n,\sqrt{n/\ln n}}] = 1 \ .$$

4 Locality sensitive hashing

Zadanie 27 — Załóżmy, że $a, b > 0$. Pokaż, że $\lim_{p->\infty}(a^p + b^p)^{\frac{1}{p}} = \max(a, b)$.

Zadanie 28 — Pokaż, że funkcja $d(A, B) = |A \triangle B|$ jest metryką na przestrzeni niepustych skończonych podzbiorów ustalonego zbioru X.

Zadanie 29 — Niech $f : [0, \infty) \to [0, \infty)$ będzie funkcją rosnącą i wklęsłą.

1. Pokaż, że dla $a, b \geqslant 0$ mamy $f(a + b) \leqslant f(a) + f(b)$.
   Wskazówka: Zacznij od pokazania, że jeśli $\beta \in [0, 1]$ i x$\geqslant$ 0, to $f(\beta x) \geqslant \beta f(x)$.
   Zauważ, że możemy założyć, że $a+b > 0$; następnie zauważ, że $a = (a+b)\frac{a}{a+b}$ oraz $b = (a+b)\frac{b}{a+b}$ i zastosuj nierówność Jensena dla funkcji wklęsłych
2. Załóżmy dodatkowo, że $f(0) = 0$. Niech d będzie metryką na zbiorze X. Pokaż, że funkcja $\rho(x, y) = f(d(x, y))$ jest również metryką na zbiorze X.
3. Pokaż, że jeśli $\epsilon \in (0, 1)$ oraz d jest metryką na zbiorze X, to funkcja $\rho(x, y) = d(x, y)^\epsilon$ jest metryką na zbiorze X.
4. Pokaż, że jeśli d jest metryką na zbiorze X, to funkcja $\rho(x, y) = \frac{d(x,y)}{1+d(x,y)}$ jest metryką na zbiorze X.

Zadanie 30 — **(Twierdzenie Steinhausa)** Niech d będzie metryką na zbiorze X. Ustalmy element $a \in X$ i zdefiniujmy funkcję

$$\rho(x, y) = \frac{2d(x, y)}{d(x, a) + d(y, a) + d(x, y)}$$

Celem tego zadania jest pokazanie, że ρ jest metryką na zbiorze X.

1. Pokaż najpierw, że jeśli $0 < p \leqslant q$ oraz $r \geqslant 0$ to $\frac{p}{q} \leqslant \frac{p+r}{q+r}$.
2. Wprowadź oznaczenia $p = d(x, y)$, $q = d(x, y) + d(x, a) + d(y, a)$ oraz $r = d(x, z) + d(y, z) - d(x, y)$ i zastosuj obserwację z poprzedniego punktu do pokazania nierówności trójkąta dla funkcji ρ.

Zadanie 31 — Zastosuj twierdzenie Steinhausa do przestrzeni metrycznej $\mathbb{R}$ ze standardową metryką $d(x, y) = |x - y|$ oraz do punktu $a = 1$.

1. Naszkicuj wykres funkcji $f(x) = \rho(x, 1)$
2. Wyjaśnij zachowanie tej funkcji dla $x \leqslant 0$.

Zadanie 32 — Mamy ustalony zbiór Ω. Przez V oznaczamy zbiór wszystkich skończonych podzbiorów zbioru Ω. Zbiór krawędzi definiujemy następująco:

$$E = \{\{A, B\} \in [V]^2 : (\exists c \in A)(B = A \setminus \{c\}) \vee (\exists c \notin A)B = A \cup \{c\}\} \ .$$

Wyznacz odległość grafową w grafie (V, E).

Zadanie 33 — Jak można zdefiniować odległość edycyjną za pomocą odległości grafowej ?

Zadanie 34 — Załóżmy, że S jest takim podobieństwem obiektów przestrzeni Ω, że istnieje rodzina funkcji haszujących $\mathcal{H}$ oraz prawdopodobieństwo na rodzinie $\mathcal{H}$ takie, że dla dowolnych dwóch obiektów $A, B \in \Omega$ mamy

$$P_{h\in\mathcal{H}}[h(A) = h(B)] = S(A, B)$$

Pokaż, że wtedy funkcja $d(A, B) = 1 - S(A, B)$ jest metryką na zbiorze Ω.

Zadanie 35 — Oprogramuj funkcję `minHash`, która dla łańcucha L oraz ciągu funkcji haszującej $[h_1, \ldots, h_k]$ o wartościach w liczbach całkowitych zwraca wektor

$$[\min\{h_1(x) : x \in L\}, \ldots, \min\{h_k(x) : x \in L\}] \ .$$

Zadanie 36 — (Porównywanie stylu) Niech $k \geqslant 1$ i niech $X = [x_1, \ldots, x_n]$ będzie dowolnym ciągiem. k-gramem ciągu X nazywamy dowolny podciąg X postaci $[x_i, x_{i+1}, \ldots, x_{i+k-1}]$, gdzie $1 \leqslant i \leqslant n-k+1$.

1. Napisz funkcję, która dla danego ciągu X, liczby k oraz funkcji haszującej h wyznacza $\min\{h(y) : y \in X^{(k)}\}$, gdzie $X^{(k)}$ oznacza zbiór wszystkich k - gramów ciągu X.
2. Rozszerz kolekcję dramatów Szekspira o tekst książki *Ulysses* James Joyce'a (możesz go pobrać ze strony `https://archive.org/stream/ulysses04300gut/ulyss12.txt`. Zastosuj metodę minhash do wyznaczenia podobieństwa Jaccarda między 7-gramami powyższej kolekcji plików. Przetestuj ten algorytm dla liczby funkcji haszujących $h \in \{64, 128, 256\}$.
3. Porównaj otrzymaną aproksymację podobieństwa Jaccarda 7-gramów z jej dokładnymi wartościami.
4. Zastosuj metodę klasteryzacji k-means do przeanalizowanych dokumentów.

Pamiętaj o wygenerowaniu wspólnej rodziny funkcji haszujących dla wszystkich analizowanych tekstów. Pamiętaj również o wstępnym oczyszczeniu analizowanych dokumentów (minimum: usuń zbędne spacje i znaki specjalne).

Zadanie 37 — Napisz procedurę służącą do wyznaczania sygnatur kosinusowych plików tekstowych korzystających z 1024 losowych wektorów z $\mathbb{R}^n$ (n tutaj oznacza moc wspólnego zbioru słów występujących w badanych dokumentach). Dokumenty reprezentowane mają być przez wektor częstotliwości słów. Zastosuj tę metodę do plików z Zadanie 2.

4.1 Klątwa wymiarowości

Zadanie 38 — Narysuj wykres objętości kul jednostkowych w przestrzeni $\mathbb{R}^n$ dla $n = 1 \ldots 20$ oraz objętości kul jednostkowych w $\mathbb{R}^n$ podzielonych przez 2^n.

Zadanie 39 — Losujemy zgodnie z jednostajnym rozkładem punkt X z kuli jednostkowej $B_n = \{x \in \mathbb{R}^n : \|x\|_2 <= 1\}$. Jaka jest wartość oczekiwania $\|X\|_2$?

Zadanie 40 — Rozważmy następującą procedurę generowania punktu z kuli B_2: (1) generujemy niezależnie dwie liczby losowe x, y z odcinka $[0, 1]$ zgodnie z rozkładem jednostajnym (2) zwracamy punkt $(\sqrt{x}\cos(2\pi y), \sqrt{x}\sin(2\pi y))$.

1. Pokaż, że metoda ta generuje losowy punkt z B_2 z rozkładem jednostajnym.
2. Jaki rozkład otrzymamy gdy "zapomnimy" o pierwiastku?

Zadanie 41 — Przypomnij sobie dowód równości

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{2\pi}\ .$$

Zadanie 42 — Funkcja Gamma Eulera zdefiniowana jest wzorem

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

dla $z > 0$.

1. Oblicz $\Gamma(1)$.
2. Pokaż, że $\Gamma[z+1] = z\Gamma(z)$.
   Wskazówka: Zauważ, że $(e^{-t})' = -e^{-t}$; skorzystaj z całkowania przez części.
3. Pokaż, że $\Gamma(n+1) = n!$ dla wszystkich liczb naturalnych n.

Zadanie 43 — Wiedząc, że $V_n(r) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} r^n$, i $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

1. sprawdź, że $V_{2n}(r) = \frac{\pi^n}{n!} r^n$
2. uprość wzór na $V_{2n+1}(r)$.

Zadanie 44 — Ustalmy parametr $n \geqslant 1$ oraz $k \geqslant 1$. Napisz procedurę, która generuje zbiór X złożony z k losowych punktów z przestrzeni $[0,1]^n$ zgodnie z rozkładem jednostajnym a następnie wyznacza zbiór wszystkich możliwych odległości podzielonych przez $\sqrt{n}$ wszystkich par różnych punktów ze zbioru X i w końcu wyświetla histogram otrzymanego zbioru odległości. Przeanalizuj zbudowaną procedurę dla $k = 100$ oraz $n = 1, 10, 100, 1000, 10000$.

Zadanie 45 — Niech $S(s,k,m) = 1 - (1 - s^k)^m$. Znajdź, stosując dowolny pakiet obliczeń numerycznych, $k \in [0,100]$, $m \in [0,1000]$ takie, że $S(\frac{1}{3},k,m) \approx \frac{1}{10}$ oraz $S(\frac{1}{2},k,m) \approx \frac{9}{10}$.

c.d.n.
Powodzenia,
Jacek Cichoń