1. **[Karetní hra (10 bodů)]** V balíku je n černých karet a n červených, balík je rovnoměrně náhodně zamíchaný. Ve hře máte následující akce:
   - vzít další kartu
   - ukončit hru

   Začínáte s 0 body. Za každou vzatou červenou kartu dostanete jeden bod, za každou černou ztratíte jeden bod.
   - Jaká je nejvyšší střední hodnota výhry pro $n = 1, 2, 3, 4$?
   - Popište efektivní algoritmus, který na vstupu dostane n a určí střední hodnotu výhry a optimální strategii.

2. **[Vlastní čísla bipartitní kliky (10 bodů)]** Určete vlastní čísla matice sousednosti úplného bipartitního grafu $K_{m,n}$ (jedna partita má n vrcholů, druhá m vrcholů).

3. **[Nejdelší orientovaná procházka (10 bodů)]**
   - Popište silně souvislý orientovaný graf na n vrcholech, který bude mít co největší hitting time mezi nějakými dvěma vrcholy (existují v něm vrcholy u, v, že střední doba náhodné procházky z u do v je veliká).

     Nápověda: zkuste například orientovanou cestu rozdělit graf na tři množiny, kde v prvních dvou půjde hrana do počátečního vrcholu, ve třetí půjdou do všech vrcholů první množiny. Za jiné lepší konstrukce můžete dostat bonusové body.
   - Zkuste udělat co nejlepší horní odhad na hitting time mezi dvěma vrcholy silně souvislého orientovaného grafu na n vrcholech. I za triviální (konečný) horní odhad s důkazem budou body.

4. **[Streamovací algoritmus na k-tý moment (30 bodů)]**

   Připomenutí definic: Uvažujeme algoritmy, které počítají na vstupu $\sigma = a_1, a_2, \ldots, a_m$ kde jednotlivé tokeny jsou čísla $a_i \in \{1, 2, \ldots, n\}$. Stream dostáváme postupně (online) a chceme spočítat nějakou funkci ze streamu v co nejmenším prostoru vzhledem k n, m a případně dalším parametrům. $\mathcal{A}(\sigma)$ nechť je výstup randomizovaného streamovacího algoritmu na vstupním streamu σ, který má počítat funkci $\varphi(\sigma)$. Řekneme, že $\mathcal{A}$ (ε, δ)-aproximuje φ, pokud:

   $$\Pr\left[\left|\frac{\mathcal{A}(\sigma)}{\varphi(\sigma)} - 1\right| \geq \varepsilon\right] \leq \delta$$

   Někdy nás zajímá nějaká statistika σ jako multimnožiny, pak definujeme frekvenční vektor $\vec{f} = f_1, f_2, \ldots, f_n$ takto: $f_j = |\{i \mid a_i = j\}|$ (počet výskytů j v σ).

   Pro $k \in \mathbb{N}$ definujeme k-tý moment $F_k = \sum_{j=1}^{n} f_j^k = \|f\|_k^k$. Uvažme algoritmus 11, který má aproximovat F_k, a jeho několik vylepšení:

   - Ukažte, že algoritmus je ekvivalentní tomuto:
     - Vyberte jednu hodnotu tokenu $a \in [n]$ s pravděpodobností $\Pr[a = j] = f_j/m$ pro každé $j \in [n]$.
     - Vyberte jeden výskyt a ze všech f_a uniformně náhodně.
   - Nechť X je náhodná proměnná – výstup algoritmu, A a R nechť jsou náhodné proměnné určující a, r z algoritmu po zpracování σ. Ukažte (první pomocí teleskopické sumy):

     $$\mathbb{E}[X \mid A = j] = \frac{m}{f_j}(f_j^k - 0^k)$$

     $$\mathbb{E}[X] = F_k$$
   - Ukažte, že $\mathrm{Var}[X] \leq \mathbb{E}[X^2] = m \sum_{j=1}^{n} \sum_{i=1}^{f_j} (i^k - (i-1)^k)^2$

Vstup : stream σ
Výstup : odhad F_k
Inicializace:
$m = 0$
$r = 0$
$a = 0$
Zpracuj : zpracování tokenu j
$m = m + 1$
$\beta =$ random bit with $\Pr[\beta = 1] = \frac{1}{m}$
if $\beta == 1$ **then**
  $a = j$
  $r = 0$
if $a == j$ **then**
  $r = r + 1$
return $m(r^k - (r-1)^k)$

- Pomocí $x^k - (x-1)^k \leq kx^{k-1}$ (pro $x \geq 1$) ukažte:

$$\mathrm{Var}[X] \leq kF_1F_{2k-1}$$

- Nechť $n > 0$, $x_1, \cdots, x_n \geq 0$, $k \geq 1$ jsou reálná čísla. Ukažte, že pro $x_* = \max\{x_i\}$:

$$x_*^{k-1} \leq \left(\sum x_i^k\right)^{\frac{k-1}{k}}$$

$$\frac{1}{n}\sum x_i \leq \left(\frac{1}{n}\sum x_i^k\right)^{1/k}$$

  a z toho hlavní mezivýsledek

$$\left(\sum x_i\right)\left(\sum x_i^{2k-1}\right) \leq n^{1-\frac{1}{k}}\left(\sum x_i^k\right)^2$$

  Tedy: $\mathrm{Var}[X] \leq kn^{1-\frac{1}{k}}F_k^2$

- Nemůžeme použít Čebyševovu nerovnost rovnou, ale můžeme napřed průměrovat a pak použít trik s mediánem.
  Formálně: nechť X je náhodná proměnná taková, že $\mathbb{E}[X] = Q$. Nechť $\{X_{i,j}\}_{i\in[t],j\in[k]}$ jsou nezávislé náhodné proměnné distribuované jako X, kde

$$t = c\log\frac{1}{\delta}$$

$$k = \frac{3\,\mathrm{Var}[X]}{\varepsilon^2\mathbb{E}[X]^2}$$

  Nechť $Z = \mathrm{median}_{i\in[t]}\left(\frac{1}{k}\sum_{j=1}^{k} X_{i,j}\right)$ pak máme $\Pr[|Z - Q| \geq \varepsilon Q] \leq \delta$ (použijte Čebyševovu a Černovovu nerovnost).

- Kolik paměti používá tento výsledný algoritmus, který (ε, δ)-aproximuje F_k?