**Valószínűségi feladatok**

Szerepelt a küvetkező feladat a generátorfüggvény módszer segítségével:

1. Adva $n$ házaspár, valaki véletlenszerűen párbaállítja őket. Mi annak a valószínűsége, hogy van olyan házaspár, mely egy párba kerül. Oldjuk meg ezt a feladatot a következő szita formula segítségével.
   Lássuk be a következő szita formulát: Ha $A_1, \dots, A_n$ tetszőleges események egy valószínűségi mezőn, akkor
   $$P(A_1 \cup \dots \cup A_n) = S_1 - S_2 + S_3 - \dots + (-1)^{n+1} S_n,$$
   ahol
   $$S_k = \sum_{1 \le j_1 < \dots < j_k \le n} P(A_{j_1} \cap \dots \cap A_{j_k}), \quad 1 \le k \le n.$$
2. Lássuk be a szitaformula következő általánosítását: Legyenek $A_1, \dots, A_n$ tetszőleges események egy valószínűségi mezőn,
   $$B_1 = B_1(A_1 \dots, A_n), \dots, B_m = B_m(A_1 \dots, A_n),$$
   ezen események függvényei, azaz ezek a $B_1, \dots, B_m$ események az $A_1, \dots, A_n$ események segítségével metszet, unió és komplementerképzés segítségével kifejezhetőek. Legyenek $c_1, \dots, c_m$ tetszőleges valós számok. A
   $$\sum_{j=1}^{m} c_j P(B_j) \ge 0$$
   egyenlőtlenség teljesül, ha teljesül abban a $2^n$ számú speciális esetben, ha a $B_j$ eseményeket meghatározó $A_1, \dots, A_n$ események mindegyike vagy a biztos vagy az üres esemény. Lássuk be a szita formulát e reláció segítségével.
3. Mi a valószínűsége annak, hogy egy 30 tagú véletlen társagágban van két ember akinek ugyanakkor van a születésnapja? Általánosabban, ha adott $k$ ember, mindegyikük kap egy számot (születésnap) 1 és $n$ között egymástól függetlenül, mindegyik számot egyforma valószínűséggel, akkor mi a valószínűsége annak, hogy van két ember aki ugyanazt a számot kapta.

Az előző feladatban az igazán érdekes probléma nem is az, hogy mi a pontos értéke a fenti valószínűségnek, hanem az, hogy nagy $n$ és $k$ paraméterekre milyen ennek a valószínűségnek az aszimptotikus viselkedése. Egyszerűbb és tanulságos ezt a kifejezést az úgynevezett Poisson approximáció segítségével vizsgálni. Ezt tesszük a következő feladatban.

4. Legyen adott $n$ urna, és $\xi$ véletlen számú golyó, melyek eloszlása $k$ paraméterű Poisson eloszlású valószínűségi változó, azaz $P(\xi = k) = \dfrac{1}{k!} e^{-k}$. Dobjuk be a golyókat egymástól (és a $\xi$ valószínűségi valtozó értékétől) függetlenül úgy hogy minden urnába egyforma valószínűséggel esnek a golyók. Lássuk be, hogy az egyes

urnákba eső golyók száma $n$ darab egymástól független $\frac{k}{n}$ paraméterű Poisson eloszlású valószínűségi változó. Lássuk be, hogy a nagy számok törvénye szerint $\xi \sim k$ majdnem egy valószínűséggel, ha $k$ nagy. E tények segítségével mutassuk meg, hogy az előző feladatban $k \sim c\sqrt{n}$ esetén a valószínűség értéke tart $e^{-c}$ számhoz $n \to \infty$ esetén.

5. Legyen adva $k$ urna, amelyikbe bedobunk $n$ golyót egymástól függetlenül, az egyes unákba $p_1, \dots, p_k$ valószínűséggel, $\sum\limits_{j=1}^{k} p_j = 1$. Ekkor a nagy számok törvénye szerint az egyes urnákba eső golyók számának relativ gyakorisága tart a $p_1, \dots, p_k$ számokhoz, ha $n \to \infty$. Annak valószínűsége, hogy a relatív gyakorisága $q_1, \dots, q_k$ valamilyen más $q_1, \dots, q_k$ exponenciálisan kicsi. Célunk ebben a feladatban ennek a valószínűségnek a pontosabb meghatározása.

   Legyen $q_1, \dots, q_k$ olyan, hogy $\sum\limits_{j=1}^{k} q_j = 1$, $\varepsilon > 0$, $n_1, \dots, n_k$, $\sum\limits_{j=1}^{k} n_j = n$ az egyes urnákba eső golyók száma, $\kappa_j = \frac{n_j}{n}$, $A_n = A_n(\varepsilon) = \{(1-\varepsilon)q_j \le \kappa_j \le (1+\varepsilon)q_j\}$. Ekkor

$$\lim_{\varepsilon \to 0} \limsup \frac{1}{n} \log P(A_n(\varepsilon)) = I(\{q_1, \dots, q_k\} \| \{p_1, \dots, p_k\})$$
$$\lim_{\varepsilon \to 0} \limsup \frac{1}{n} \log P(A_n(\varepsilon)) = I(\{q_1, \dots, q_k\} \| \{p_1, \dots, p_k\}),$$

   ahol $I(\{q_1, \dots, q_k\} \| \{p_1, \dots, p_k\}) = -\sum\limits_{j=1}^{k} q_j \log \frac{p_j}{q_j}$.

Az előző feladat egy olyan relációt fejez ki, amelyik hasznos a következő szemináriumon tárgyalandó entrópiának és az információelméletnek a megértéséhez. Végül egy népszerű kérdéskör, az úgynevezett tiszta fejek probémáját fogjuk vizsgálni. A kérdés az, hogy egyszabályos pénzdarab feldobása során keletkezett $n$ hosszúságú fej-írás sorozat milyen hosszúságú csak fej jeleket tartalmazó blokkot tartalmaz. Durván szólva $\log n$-nél valamivel rövidebb sorozat majdnem biztos van, $\log n$-nél valamivel hosszabb sorozat majdnem biztos nincsen.

6. Adjunk jó becslést arra, hogy egy $n = A2^k$ hosszúságú szabályos fej-írás sorozat tartalmaz $k$ hosszúságú fej blokkot.Tekintsük az első $m$ darab egymás után kövekező tiszta fej blokkot. (Ha két írás következik egymás után, akkor azt tekintsük úgy, hogy közöttük nulla hosszúságú fej-blokk van.) Milyen becslést tudunk adni arra, hogy az első $m$ blokk tartalmaz vagy nem tartalmaz $k$ hosszúságú sorozatot, hogy a hosszaik összege kisebb vagy nagyobb mint $n$? Milyen becsléseket kapunk így az eredeti feladatra?
7. Legyen $T_n$ annak a valószínűsége, hogy a szabályos fej-írás sorozt első $n$ tagja nem tartalmaz $k$ hosszúságú tiszta fej sorozatot. Lássuk be, hogy $T_n = \sum\limits_{j=0}^{k-1} \frac{T_{n-j}}{2^j}$, $T_1 = \dots = T_{k-1} = 1$. Lássuk be, hogy a $2^n T_n$ sorozat egy általánosított Fibonacci sorozatnak tekinthető. Milyen következtetéseket tudunk ennek alapján levonni ennek viselkedéséről? Mi e sorozat generátor függvénye?