Néhány alapvető eredmény a χ^2 statisztikáról

Tekintsük a következő statisztikai feladatot. Adott egy dobókocka, és el akarjuk dönteni, hogy az szabályos-e. Ennek érdekében feldobjuk a kockát n alkalommal, és megjegyezzük, hány alkalommal volt a dobás eredménye j, $1 \le j \le 6$. Jelölje $\nu_n(j)$, $1 \le j \le 6$, ezeket a (véletlen) számokat. Ha a $\nu_n(j) \sim \frac{n}{6}$ reláció teljesül minden $1 \le j \le 6$ számra akkor a kockát szabályosnak tekinthetjük, ellenkező esetben pedig szabálytalannak. Meg kell pontosabban fogalmazni, mit jelentenek a fenti aszimptotikus relációk. Ennek érdekében tekintsük a $\frac{1}{n}\sum\limits_{j=1}^{6}\left(\nu_n(j)-\frac{n}{6}\right)^2$ valószínűségi változókat. Be lehet látni, hogy amennyiben a kocka szabályos, azaz minden oldalára $\frac{1}{6}$ valószínűséggel esik, akkor ennek a kifejezésnek létezik határeloszlása $n \to \infty$ esetében, és ez a határeloszlás explicit módon megadható. Ez lehetővé teszi egy olyan természetes teszt megadását, amelyben egy szabályos dobókockát előírt p valószínűséggel tekintünk szabályosnak, és egy szabálytalan dobókockát nagy n kisérletszám esetén majdnem egy valószínűséggel fogunk szabálytalannak tekinteni.

A fenti problémát természetes módon általánosíthatjuk, és az új probléma megoldását megadhatjuk a fent említett határeloszlástétel egy általánosítása segítségével. Az általánosított feladat a következő: Legyen adva k urna, és ellenőrizni akarjuk azt a feltételezést, amely szerint ha egy golyót véletlenül bedobunk ezen urnák valamelyikébe, akkor az p_j, $p_j > 0$, valószínűséggel esik a j-ik urnába, $\sum\limits_{j=1}^{k} p_j = 1$. Ennek a feltételezésnek az ellenőrzése érdekében dobjunk egymástól függetlenül n golyót ezekbe az urnákba, és jelölje $\nu_n(j)$ a j-ik urnába eső golyók számát. Be lehet látni, hogy feltételezésünk teljesülése esetén a $\sum\limits_{j=1}^{k}\frac{(\nu_n(j)-np_j)^2}{np_j}$ valószínűségi változóknak létezik határeloszlása $n \to \infty$ esetén, és ezt a határeloszlást meg tudjuk adni. Ezt a határeloszlástételt írjuk le a 2. feladatban. E határeloszlástétel segítségével a kocka szabályosságának ellenőrzéséhez hasonlóan ellenőrizni tudjuk feltételezésünk helyességét.

E hatéreloszlástétel bizonyítása azon alapul, hogy egyrészt meg tudjuk adni a $\left\{\frac{\nu_n(j)-np_j}{\sqrt{np_j}},\ j=1,\dots,k\right\}$ véletlen vektorok határeloszlását a több-dimenziós centrális határeloszlástétel segítségével, másrészt, ha valamely $\left(\eta_1^{(n)},\dots,\eta_k^{(n)}\right)$ véletlen vektorok eloszlásban konvergálnak egy $(\eta_1,\dots,\eta_k)$ véletlen vektorhoz $n \to \infty$ esetében, és $f(x_1,\dots,x_k)$ folytonos k-változós függvény, akkor az $f\left(\eta_1^{(n)},\dots,\eta_k^{(n)}\right)$, $\quad n=1,2,\dots,$ valószínűségi változók eloszlásban konvergálnak az $f(\eta_1,\dots,\eta_k)$ valószínűségi változóhoz. Ez lehetővé teszi a minket érdeklő statisztikák határeloszlásának megadását. De ezt a határeloszlást egyszerűbb, jobban áttekinthető alakban akarjuk megadni. Ez lehetséges némi lineáris algebra segítségével. A számolás során felhasználjuk azt, hogy ha $(\eta_1,\dots,\eta_k)$ k-dimenziós nulla várható értékű D kovarianciamátrixú normális eloszlású véletlen vektor, A egy $k\times k$-as mátrix akkor $(\eta_1,\dots,\eta_k)A$ k-dimenziós nulla várható értékű A^*DA kovarianciamátrixú normális eloszlású véletlen vektor. A minket érdeklő határeloszlástétel bizonyítását megkönnyíti az alábbi önmagában is érdekes első feladat. Ezután fogalmazzuk meg a 2. feladatban a χ^2 próba alapját képező statisztikai állítást.

1.) Legyen $\eta = (\eta_1, \dots, \eta_k)$ k-dimenziós normális eloszlású valószínűségi változó nulla várható értékkel és D kovariancia mátrix-szal. Legyenek a $\lambda_1, \dots, \lambda_k$ számok a D mátrix sajátértékei (multiplicitással). Bizonyítsuk be (alapvető lineáris algebrai ismeretek felhasználásával), hogy a $\sum\limits_{j=1}^{k} \eta_j^2$ valószínűségi változó eloszlása megegyezik egy $\sum\limits_{j=1}^{k} \lambda_j \xi_j^2$ valószínűségi változó eloszlásával, ahol $\xi_1, \dots, \xi_k$ független standard normális eloszlású valószínűségi változók.

2.) Adott k darab urna, amelyekbe bedobunk egymástól függetlenül n golyót úgy, hogy mindegyik golyó p_j valószínűséggel esik a j-ik urnába, $1 \le j \le k$, $\sum\limits_{j=1}^{k} p_j = 1$. Jelölje $\nu_n(j)$ a j-ik urnába eső golyók számát. Lássuk be, hogy a $\sum\limits_{j=1}^{k} \dfrac{(\nu_n(j) - np_j)^2}{np_j}$ valószínűségi változók eloszlásban konvergálnak a $k-1$ szabadságfokú $\chi^2(k-1)$ eloszláshoz, ha $n \to \infty$, (a k szám rögzített), azaz egy olyan ζ_{k-1} valószínűségi változó eloszlásához, amelyet úgy kapunk, hogy tekintünk $k-1$ darab független standard normális eloszlású $\xi_1, \dots, \xi_{k-1}$ valószínűségi változót és elkészítjük segítségükkel a $\zeta_{k-1} = \sum\limits_{j=1}^{k-1} \xi_j^2$ valószínűségi változót.

A 2. feladatban szereplő határeloszlás csak az urnák k számától függ, de nem függ a p_j, $1 \le j \le k$, valószínűségektől. Ez jelzi azt, hogy természetes statisztikát vettünk, olyat amelyben a különböző urnákban levő golyók számának az eltérése annak várható értékétől egyforma fontos szerepet játszik. Az, hogy a határeloszlás a $\chi^2(k-1)$ eloszlás azzal függ össze, hogy bár k véletlen szám súlyozott négyzetösszegét tekintettük, (az egyes urnákba eső golyók számának eltérését tekintettük azok várható értékétől), de ezek között van egy determinisztikus összefüggés. Nevezetesen az, hogy az összes urnába eső golyók száma minusz azok várható értéke nullával egyenlő. A 2. feladat eredményét informálisan úgy szokták interpretálni, hogy $k-1$ szabadsági fokkal rendelkező véletlen vektorok koordinátáinak a négyzetösszegét tekintettük, illetve azok határeloszlását. Ilyen esetben a határeloszlást olyan véletlen összeg adja meg, amelyben mindegyik szabadsági foknak egy összeadandó felel meg, amelyik független a többi összeadandótól, és az egy standard normális eloszlású valószínűségi változó négyzete. A 4. illetve 4.$'$ feladatban a 2. feladat egy olyan általánosítását fogalmazzuk meg, mint amilyent a 2. feladat fenti heurisztikus interpretációja sugall. Ezelőtt azonban a 3. feladatban egy a 4. és 4.$'$ feladat bizonyításában hasznos technikai jellegű állítást tárgyalunk.

3.) Legyenek $X_1, \dots, X_k$ független standard normális eloszlású valószínűségi változók, $p_j \ge 0$, $1 \le j \le k$, tetszőleges nem negatív számok, amelyekre $\sum\limits_{j=1}^{k} p_j = 1$. Ekkor az $X_1^2 + \cdots + X_k^2 - \left(\sqrt{p_1}X_1 + \cdots + \sqrt{p_k}X_k\right)^2$ valószínűségi változó $\chi^2(k-1)$ eloszlású, és független a standard normális eloszlású $\sum\limits_{j=1}^{k} \sqrt{p_j} X_j$ valószínűségi változótól.

Tekintsük a következő modellt. Adva vannak urnák, amelyekbe egymástól függetlenül bedobunk n golyót, amelyek a j-ik urnába p_j valószínűséggel esnek. Tekintsük először azt az esetet, amikor bizonyos urnákat nem tudunk megkülönböztetni egymástöl, csak azt tudjuk megfigyelni, hogy rögzítve urnák k csoportját, a golyók az urnák melyik csoportjába esnek. Illetve annak valószínűségét is tudjuk, hogy az egyes golyók az urnák valamelyik csoportjába esnek. Ebben az esetben alkalmazhatjuk a 2. feladat eredményét, amely megadja az egyes urnacsoportokba eső golyók számának egy olyan függvényét, amelynek van határeloszlása $n \to \infty$ esetében, és az a $\chi^2(k-1)$ eloszlás. Ha meg tudjuk figyelni az egyes urnákba eső golyók számát is, és annak valószínűségét is ismerjük, hogy a golyók az egyes urnákba esnek, akkor ebben az esetben is alkalmazhatjuk a 2. feladat eredményét. Ezen eredmény szerint az egyes urnákba eső golyók számának egy alkalmas függvényét véve, annak van határeloszlása, és az χ^2 eloszlású. Ha ezt a két statisztikát egyidejűleg nézzük, akkor ezek együttesének is van határeloszlása az $n \to \infty$ esetben, és ez az együttes eloszlás olyan, mint amilyet a 2. feladat utáni heurisztikus érvelés sugall. Nevezetesen, a durvább megfigyeléshez tartozó statisztika határeloszlása megegyezik $k-1$ független standard normális eloszlású valószínűségi változó négyzetösszegének az eloszlásával. Ha az egyes urnákat is meg tudjuk különböztetni, és a durvább és pontosabb megfigyelésekhez tartozó statisztikák együttes határeloszlását akarjuk megadni, akkor a durvább megfigyeléseken alapuló statisztika határeloszlását leíró valószínűségi változóhoz hozzá kell adni annyi független standard normális eloszlású valószínűségi változónak a négyzetét, amennyivel a modell szabadságfoka nő a pontosabb megfigyelések által. E két valószínűségi változó együttes eloszlása adja meg a két statisztika együttes határeloszlását. Ezt az eredményt fogalmazzuk meg pontosabban a 4. feladatban, a 4.$'$ feladat pedig a 4. feladat eredményének egy természetes általánosítása.

4.) Legyen adva urnáknak k csoportja, és az urnák j-ik csoportjában legyen l_j, $l_j \ge 1$, urna, $1 \le j \le k$. Dobjunk le n golyót egymástól függetlenül az urnákba úgy, hogy mindegyik golyó a j-ik csoport s-ik urnájába $p_{j,s} > 0$ valószínűséggel esik, és jelölje $\nu_n(j,s)$ az ebbe az urnába eső golyók számát, $1 \le j \le k$, $1 \le s \le l_j$, $\sum\limits_{j=1}^{k} \sum\limits_{s=1}^{l_j} p_{j,s} = 1$. Legyen továbbá $p_j = \sum\limits_{s=1}^{l_j} p_{j,s}$, $\nu_n(j) = \sum\limits_{s=1}^{l_j} \nu_n(j,s)$, $1 \le j \le k$, azaz azon golyók száma, amelyek a j-ik csoportban levő urnák valamelyikébe esnek. Vezessük be továbbá az $L = \sum\limits_{j=1}^{k} l_j$ jelölést, és tekintsük az $U_n^{(0)} = \sum\limits_{j=1}^{k} \sum\limits_{s=1}^{l_j} \frac{(\nu_n(j,s)-np_{j,s})^2}{np_{j,s}}$ valamint az $U_n^{(1)} = \sum\limits_{j=1}^{k} \frac{(\nu_n(j)-np_j)^2}{np_j}$ valószínűségi változókat. Ekkor a 2. feladat eredménye alapján tudjuk, hogy az $U_n^{(0)}$ valószínűségi változók a $\chi^2(L-1)$, az $U_n^{(1)}$ valószínűségi változók pedig a $\chi^2(k-1)$ eloszláshoz konvergálnak, ha $n \to \infty$. Az ebből a két valószínűségi változóból álló $(U_n^{(0)}, U_n^{(1)})$ véletlen vektornak is van határeloszlása $n \to \infty$ esetén, és ez a következő módon jellemezhető:

Tekintsünk $L-1$ független $\xi_1, \dots, \xi_{L-1}$ standard normális eloszlású valószínűségi változót, és tekintsük azt az $U=(U_0,U_1)$ véletlen vektort, amelynek az első koordinátája az összes ξ_j valószínűségi változó négyzetösszege, a második koordináta pedig csak az első $k-1$ ξ_j valószínűségi változó négyzetösszege, azaz $U_0=\sum\limits_{j=1}^{L-1}\xi_j^2$, $U_1=\sum\limits_{j=1}^{k-1}\xi_j^2$. Ekkor az $(U_n^{(0)},U_n^{(1)})$ véletlen vektorok az $U=(U_0,U_1)$ véletlen vektorhoz konvergálnak eloszlásban $n\to\infty$ esetén.

4.′) Tekintsük az $\{1,\dots,k\}$ halmaz egymásba skatulyázott $\mathcal{A}_1\supset\mathcal{A}_2\supset\cdots\supset\mathcal{A}_t$ particióit, $\mathcal{A}_j=\{A_{j,1},\dots,A_{j,l_j}\}$, $1\le j\le t$, $\bigcup\limits_{s=1}^{l_j}A_{j,s}=\{1,\dots,k\}$, $A_{j,s}\cap A_{j,s'}=\emptyset$, ha $1\le s,s'\le l_j$, és $s\ne s'$. (Az $\mathcal{A}_j\supset\mathcal{A}_{j+1}$, $1\le j<t$, reláció azt jelenti, hogy minden $A_{j+1,s}$, $1\le s\le l_{j+1}$, halmazhoz létezik egy azt tartalmazó $A_{j,u}\supset A_{j+1,s}$, $1\le u\le l_j$ halmaz.) Rendeljünk hozzá minden $1\le r\le k$ ponthoz valamilyen $p_r>0$ súlyt úgy, hogy $\sum\limits_{r=1}^{k}p_r=1$, és rendeljük hozzá az $\mathcal{A}_j$ partició $A_{j,s}$ eleméhez a $p_s^{(j)}=\sum\limits_{r\in A_{j,s}}p_r$ súlyt, $1\le j\le t$, $1\le s\le l_j$. Dobjunk le n golyót k urnába egymástól függetlenül úgy, hogy mindegyik golyó p_r valószínűséggel esik az r-ik urnába, $1\le r\le k$, és jelölje $\nu_n(r)$ az r-ik urnába eső golyók számát. Az $\mathcal{A}_j$ partició segítségével definiáljuk a $\nu_n^{(j)}(s)=\sum\limits_{r\in A_{j,s}}\nu_n(r)$, $1\le s\le l_j$, számokat, azaz számoljuk meg azt, hogy hány golyó esik azon urnák valamelyikébe, amelyek indexei az $A_{j,s}$ halmazban vannak. Definiáljuk az $U_j^{(n)}=\sum\limits_{s=1}^{l_j}\frac{\left(\nu_n^{(j)}(s)-np_s^{(j)}\right)^2}{np_s^{(j)}}$ valószínűségi változókat, $1\le j\le t$. Ekkor az $(U_1^{(n)},\dots,U_t^{(n)})$ véletlen vektorok eloszlásban konvergálnak egy $(U_1,\dots,U_t)$ véletlen vektorhoz, ha $n\to\infty$, amelyet a következő módon definiálhatunk. Legyen $\xi_1,\dots,\xi_{k-1}$ $k-1$ darab független standard normális eloszlású valószínűségi változó, és $U_j=\sum\limits_{s=1}^{l_j-1}\xi_s^2$, $1\le j\le t$.

A következő feladat statisztikai jelentősége az, hogy segít megérteni azt, hogy a χ^2 statisztika miért jelenik meg természetes módon bizonyos statisztikai feladatokban, amikor konfidencia tartományokat akarunk konstruálni.

Legyen adva k urna, és dobjunk be ezekbe n golyót egymástól függetlenül úgy, hogy mindegyik golyó p_j, $p_j>0$, $\sum\limits_{j=1}^{k}p_j=1$, valószínűséggel esik a j-ik urnába, $1\le j\le k$. Ha az egyes urnákba m_j, $1\le j\le k$, golyó esik, akkor a természetes becslés a p_j valószínűségekre a $\hat p_j=\frac{m_j}{n}$ kifejezés, $1\le j\le k$. Ez egyben a p_j, $1\le j\le k$, paraméterek maximum likelihood becslése is. A matematikai statisztika egyik fontos feladata konfidencia-tartomány konstrukciója, amely jelen esetben a következő problémát jelenti. Rögzítsünk egy $0<u<1$ számot, és az $m_1,\dots,m_k$ megfigyelések segítségével adjunk egy viszonylag kis (véletlen) $U_n(m_1,\dots,m_k)$ tartományt úgy, hogy az igazi $(p_1,\dots,p_k)$ paraméterek u valószínűséggel vannak ebben az $U_n(m_1,\dots,m_k)$

tartományban, azaz formálisan,

$$P\left((p_1,\dots,p_k)\in U_n(m_1,\dots,m_k)\right)=u.$$

Az alább ismertetendő 5. feladat eredménye a következő (aszimptotikus) módszert sugallja ezen kérdés megoldására: Adva egy $0<u<1$ szám, definiáljuk azt az $x=x(u)$ számot, amelyre egy $\chi^2(k-1)$ eloszlású ζ valószínűségi változóra $P(\zeta<2x)=u$. Ezután definiáljuk a következő (véletlen) $\bar U_n(m_1,\dots,m_k)$ és $U_n(m_1,\dots,m_k)$ ellipszoidokat:

$$\bar U_n(m_1,\dots,m_k)=\left\{(u_1,\dots,u_k)\colon \sum_{j=1}^k\frac{(m_j-nu_j)^2}{n\hat p_j}\le x\right\},$$

(ebben a formulában $n\hat p_j=m_j$, $1\le j\le k$), és

$$U_n(m_1,\dots,m_k)=\bar U_n(m_1,\dots,m_k)\cap\left\{(u_1,\dots,u_k)\colon \sum_{j=1}^k u_j=1\right\}.$$

Ekkor a 2. feladat és alább megfogalmazandó 5. feladat eredménye alapján a

$$\lim_{n\to\infty}P\left((p_1,\dots,p_k)\in U_n(m_1,\dots,m_k)\right)=u,$$

reláció érvényes, és a maximum likelihood módszer ennek a relációnak az alkalmazását sugallja, ha a konfidenciatartományt akarjuk meghatározni a fenti problémában.

5.) Legyen adva k urna, és dobjunk ezekbe n golyót egymástól függetlenül úgy, hogy mindegyik golyó p_j valószínűséggel esik a j-ik urnába, $1\le j\le k$, $\sum\limits_{j=1}^k p_j=1$. Jelölje $m_1,\dots,m_k$ az egyes urnákba eső golyók számát, $\sum\limits_{j=1}^k m_j=n$. Tekintsük a likelihood függvény $L_n(p_1,\dots,p_k,m_1,\dots,m_k)$ logaritmusát, $\sum\limits_{j=1}^k p_j=1$, $\sum\limits_{j=1}^k m_j=n$, $p_j\ge0$, $m_j\ge0$, $1\le j\le k$, azaz tekintsük azon esemény valószínűségét, hogy az egyes urnákba $m_1,\dots,m_k$ golyó esik abban az esetben, ha annak valószínűsége, hogy a golyók a megfelelő urnába esnek az $p_1,\dots,p_k$ számokkal egyenlőek, majd vegyük ennek a valószínűségnek a logaritmusát. Ekkor

$$L_n(p_1,\dots,p_k,m_1,\dots,m_k)=\log\frac{n!}{m_1!\cdots m_k!}+\sum_{j=1}^k m_j\log p_j,$$

és az $L_n(p_1,\dots,p_k,m_1,\dots,m_k)$ függvény maximuma rögzített $m_1,\dots,m_k$ számokra a $p_j=\hat p_j=\frac{m_j}{n}$, $1\le j\le k$, pontban vétetik fel. Azaz ez a $p_1,\dots,p_k$ paraméterek maximum likelihood becslése. Tekintsük a maximum likelihood függvény logaritmusának különbségét a becsült $\hat{p}_j = \frac{m_j}{n}$, $1 \leq j \leq k$, és a valódi $p_1, \dots, p_k$ paraméterek között, azaz vegyük az

$$\begin{aligned} &B^{(n)}_{p_1,\dots,p_k}(m_1,\dots,m_k) \\ &\quad = L_n\left(\frac{m_1}{n},\dots,\frac{m_k}{n}, m_1,\dots,m_k\right) - L_n\left(p_1,\dots,p_k, m_1,\dots,m_k\right) \end{aligned}$$

különbséget. Ez a különbség a $\left(\frac{m_1}{n},\dots,\frac{m_k}{n}\right)$ és $(p_1,\dots,p_k)$ eloszlások közötti I-divergencia szorozva n-nel, azaz

$$B^{(n)}_{p_1,\dots,p_k}(m_1,\dots,m_k) = nI\left(\left(\frac{m_1}{n},\dots,\frac{m_k}{n}\right)\Big\|\,(p_1,\dots,p_k)\right) = n\sum_{j=1}^{k}\frac{m_j}{n}\log\frac{\frac{m_j}{n}}{p_j}.$$

A $B^{(n)}_{p_1,\dots,p_k}(m_1,\dots,m_k)$ valószínűségi változók aszimptotikusan megegyeznek a χ^2 próbában szereplő statisztikának felével. Ez azt jelenti, hogy ha a

$$C_n(m_1,\dots,m_k) = \sum_{j=1}^{k}\frac{(m_j - np_j)^2}{np_j}$$

statisztikát tekintjük, akkor $\frac{1}{2}C_n(m_1,\dots,m_k) - B^{(n)}_{p_1,\dots,p_k}(m_1,\dots,m_k) \Rightarrow 0$ az $n \to \infty$ esetben, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl.

Igaz ennek az állításnak a következő változata is: A fenti állítás érvényben marad, ha a $C_n((m_1,\dots,m_k)$ kifejezésben np_j-t az $n\hat{p}_j = m_j$ számmal helyettesítjük, azaz

$$\frac{1}{2}\bar{C}_n(m_1,\dots,m_k) - B^{(n)}_{p_1,\dots,p_k}(m_1,\dots,m_k) \Rightarrow 0, \quad \text{ha } n \to \infty,$$

ahol

$$\bar{C}_n(m_1,\dots,m_k) = \sum_{j=1}^{k}\frac{(m_j - np_j)^2}{m_j}.$$

Az 5.) feladatnak megfogalmazhatjuk az alábbi következményét. Definiáljuk minden $(m_1,\dots,m_k)$ paraméterre és x számra az alábbi ellipszoidot.

$$D((u_1,\dots,u_k)|m_1,\dots,m_k,x) = \left\{(u_1,\dots,u_k)\colon \sum_{j=1}^{k}\frac{(m_j - nu_j)^2}{m_j} < x\right\}$$

Ekkor a következőt mondhatjuk. Dobjunk n pontot egymástól függetlenül k urna valamelyikébe, és legyen p_j, $1 \leq j \leq k$, $\sum_{j=1}^{k} p_j = 1$, annak valószínűsége, hogy egy golyó a j-ik urnába esik. Ekkor

$$\lim_{n\to\infty} P((p_1,\dots,p_k) \in D((u_1,\dots,u_k)|m_1,\dots,m_k,x) = 1 - F_k(x), \tag{1}$$

ahol $F_k(x)$ a $k-1$ paraméterű χ^2 eloszlásfüggvény. Ez azt jelenti, hogy ha az n dobásszám nagy, és az egyes urnákba eső dobások száma $m_1,\dots,m_k$, akkor ezen pontok segítségével tudtunk definiáli egy olyan véletlen $D((u_1,\dots,u_k)|m_1,\dots,m_k,x)$ halmazt, amelyre teljesül az (1) reláció. Ez a tény hasznos, mind a statisztikai ellenőrzésben mind a konfidencia tartományok konstrukciójában.

Megoldások:

1.) A D mátrix felírható $D = U\Lambda U^*$ alakban, ahol U unitér, Λ pedig olyan diagonális mátrix, amelynek átlójában a D mátrix λ_j sajátértékei vannak. (Az U mátrix is felírható explicit módon a D nátrix sajátvektorainak segítségével, de erre a tényre itt nincs szükségünk.) Az $\eta = (\eta_1, \dots, \eta_k)$ véletlen vektor eloszlása megegyezik egy $\bar\eta = (\bar\eta_1, \dots, \bar\eta_k) = \xi\Lambda^{1/2}U^* = (\xi_1, \dots, \xi_k)\Lambda^{1/2}U^*$ véletlen vektor eloszlásával, ahol a $\xi = (\xi_1, \dots, \xi_k)$ véletlen vektor standard normális eloszlású. Valóban $\bar\eta$ normális eloszlású véletlen vektor, amelynek várható értéke nulla és kovariancia mátrixa a $(\Lambda^{1/2}U^*)^*\Lambda^{1/2}U^* = U\Lambda^{1/2}\Lambda^{1/2}U^* = U\Lambda U^* = D$ mátrix. Ezért a $\sum\limits_{j=1}^k \eta_j^2$ valószínűségi változó eloszlása megegyezik a $\sum\limits_{j=1}^k \bar\eta_j^2$ valószínűségi változó eloszlásával. Vegyük észre, hogy az $\tilde\eta = (\tilde\eta_1, \dots, \tilde\eta_k) = \bar\eta U = (\bar\eta_1, \dots, \bar\eta_k)U$ vektorra teljesül a $\sum\limits_{j=1}^k \bar\eta_j^2 = \sum\limits_{j=1}^k \tilde\eta_j^2$ azonosság, mert U unitér, tehát távolságtartó transzformáció. Viszont $\tilde\eta = \bar\eta U = \xi\Lambda^{1/2}U^*U = \xi\Lambda^{1/2}$. Ez azt jelenti, hogy a $\sum\limits_{j=1}^k \eta_j^2$ valószínűségi változó eloszlása megegyezik a $\sum\limits_{j=1}^k (\lambda_j^{1/2}\xi_j)^2 = \sum\limits_{j=1}^k \lambda_j\xi_j^2$ valószínűségi változó eloszlásával, és ez a feladat állítása.

2.) A többdimenziós centrális határeloszlástétel alapján a

$$\left(\frac{\nu_n(1) - np_1}{\sqrt{np_1}}, \dots, \frac{\nu_n(k) - np_k}{\sqrt{np_k}}\right)$$

véletlen vektorok eloszlásban konvergálnak $n \to \infty$ esetén egy olyan $(\eta_1, \dots, \eta_k)$ k-dimenziós normális eloszlású véletlen vektorhoz, amelyre $\operatorname{Var}\eta_j = (1-p_j)$, $E\eta_j = 0$, $1 \le j \le k$, és $\operatorname{Cov}(\eta_j, \eta_l) = -\sqrt{p_j p_l}$, ha $1 \le j, l \le k$, és $j \ne l$. Ezért elegendő belátni, hogy egy nulla várható értékű és a fenti kovarianciafüggvénnyel rendelkező normális eloszlású $(\eta_1, \dots, \eta_k)$ véletlen vektorra a $\sum\limits_{j=1}^l \eta_j^2$ kifejezés $\chi^2(k-1)$ eloszlású valószínűségi változó, mivel a $\sum\limits_{j=1}^k \frac{(\nu_n(j)-np_j)^2}{np_j}$ valószínűségi változók ehhez a valószínűségi változóhoz konvergálnak eloszlásban.

Ennek belátása érdekében tekintsük az $(\eta_1, \dots, \eta_k)$ véletlen vektor $D = (d_{j,l})$, $1 \le j, l \le k$, kovarianciamátrixát, amelyre $d_{j,j} = 1 - p_j$, $1 \le j \le k$, $d_{j,l} = -\sqrt{p_j p_l}$, ha $1 \le j, l \le k$, és $j \ne l$, és értsük meg a D kovarianciamátrixnak a szerkezetét. Azt állítjuk, hogy az $u = \left(\sqrt{p_1}, \dots, \sqrt{p_k}\right)$ vektor sajátvektora a D kovarianciamátrixnak nulla sajátértékkel és emellett, ha egy $x = (x_1, \dots, x_k)$ vektor merőleges erre az u vektorra, azaz $\sum\limits_{j=1}^k x_j\sqrt{p_j} = 0$, akkor az x vektor a D mátrixnak sajátvektora 1 sajátértékkel.

Valóban, tetszőleges $1 \le j \le k$ számra

$$\sum_{l:\, l\neq j} \sqrt{p_l} d_{j,l} = -\sqrt{p_j} \sum_{l:\, l\neq j} p_l = -\sqrt{p_j}(1-p_j) = -\sqrt{p_j} d_{j,j},$$

ahonnan $\sum\limits_{l=1}^{k} \sqrt{p_l} d_{j,l} = 0$, és ez azt jelenti, hogy az u vektor a D mátrix sajátvektora nulla sajátértékkel.

Ha $\sum\limits_{j=1}^{k} x_j \sqrt{p_j} = 0$, akkor minden $1 \le j \le k$ számra

$$\sum_{l:\, l\neq j} d_{j,l} x_l = -\sqrt{p_j} \sum_{l:\, l\neq j} \sqrt{p_l} x_l = \sqrt{p_j}\sqrt{p_j} x_j = p_j x_j,$$

ahonnan $\sum\limits_{l=1}^{k} d_{j,l} x_l = x_j p_j + x_j(1-p_j) = x_j$, és ez azt jelenti, hogy az x vektor a D mátrix sajátvektora egy sajátértékkel.

Ez azt jelenti, hogy a D kovariancia mátrixnak létezik egy nulla és $k-1$ 1 sajátértékkel rendelkező ortogonális sajátvektora. Valóban, az $u = \left(\sqrt{p_1}, \dots, \sqrt{p_k}\right)$ vektor és a rá merőleges altér tetszőleges $k-1$ vektorból álló ortogonális bázisa alkalmas választás. Ezért a bizonyítandó állítás következik az első feladat eredményéből. Ekkor ugyanis a sajátértékek rendszere az 1 számból áll $k-1$ multiplicitással és a nulla számból egy multiplicitással.

3.) Vezessük be a $Z_j = X_j - \sqrt{p_j} \sum\limits_{l=1}^{k} \sqrt{p_l} X_l$, $1 \le l \le k$, és $U = \sum\limits_{l=1}^{k} \sqrt{p_l} X_l$ valószínűségi változókat. Ekkor $(Z_1, \dots, Z_k, U)$ $k+1$ dimenziós normális eloszlású valószínűségi vektor. Továbbá, $EZ_jU = 0$ minden $1 \le j \le k$ számra, és $EU^2 = 1$. Innen, és a többdimenziós normális eloszlású valószínűségi változók tulajdonságaiból következik, hogy az $U = \sum\limits_{l=1}^{k} \sqrt{p_l} X_l$ standard normális eloszlású valószínűségi változó független a $(Z_1, \dots, Z_k)$ véletlen vektortól. Továbbá egyszerű algebrai számolás adja, hogy $\sum\limits_{j=1}^{k} Z_j^2 = X_1^2 + \cdots + X_k^2 - \left(\sqrt{p_1}X_1 + \cdots + \sqrt{p_k}X_k\right)^2$. Ezért a bizonyítás befejezéséhez elég megmutatni azt, hogy $\sum\limits_{j=1}^{k} Z_j^2$, $\chi^2(k-1)$ eloszlású valószínűségi változó. Viszont egyszerű számolás adja, hogy $EZ_j = 0$, $EZ_j^2 = 1-p_j$, $1 \le j \le k$, $EZ_jZ_l = -\sqrt{p_jp_l}$, ha $1 \le j, l \le k$, és $j \neq l$. Másrészt az előző feladatban bebizonyítottuk (ez volt a bizonyítás lényege), hogy egy ilyen kovariancia mátrixú, nulla várható értékű $(Z_1, \dots, Z_k)$ normális eloszlású véletlen vektorra a $\sum\limits_{j=1}^{k} Z_j^2$ valószínűségi változó $\chi^2(k-1)$ eloszlású.

4.) Tekintsük a $Z_{j,s,n} = \frac{\nu_n(j,s) - np_{j,s}}{\sqrt{np_{j,s}}}$, $1 \le j \le k$, $1 \le s \le l_j$, illetve a a $Z_{j,n} = \frac{\nu_n(j) - np_j}{\sqrt{np_j}}$, $1 \le j \le k$ valószínűségi változókat. A többdimenziós centrális határeloszlástétel alapján ezen valószínűségi változók együttes eloszlása konvergál egy nulla várható értékű és alkalmas kovarianciájú $Z_{j,s}$, $1 \le j \le k$, $1 \le s \le l_j$, és Z_j, $1 \le j \le k$, koordinátákból álló véletlen normális eloszlású vektorhoz. Innen következik, hogy az $(U_n^{(0)}, U_n^{(1)})$ vektoroknak van határeloszlása. Abból a célból, hogy ezt a határeloszlást megadjuk, először definiálunk olyan $Z_{j,s}$, és Z_j koordinátákból álló többdimenziós normális eloszlású vektort, amelynek eloszlása megegyezik a $Z_{j,s,n}$ és $Z_{j,n}$ koordinátájú véletlen vektorok határeloszlásával.

Legyenek $X_{j,s}$, $1 \le j \le k$, $1 \le s \le l_j$, független standard normális eloszlású valószínűségi változók, és definiáljuk segítségükkel az

$$X_j = \frac{1}{\sqrt{p_j}} \sum_{s=1}^{l_j} \sqrt{p_{j,s}} X_{j,s}, \qquad X = \sum_{j=1}^{k} \sqrt{p_j} X_j = \sum_{j=1}^{k} \sum_{s=1}^{l_j} \sqrt{p_{j,s}} X_{j,s},$$

$$Z_{j,s} = X_{j,s} - \sqrt{p_{j,s}} X,$$

továbbá a

$$Z_j = \frac{1}{\sqrt{p_j}} \sum_{s=1}^{l_j} \sqrt{p_{j,s}} Z_{j,s} = \frac{1}{\sqrt{p_j}} \sum_{j=1}^{l_j} \sqrt{p_{j,k}} X_{j,k} - \frac{p_j X}{\sqrt{p_j}} = X_j - \sqrt{p_j} X$$

valószínűségi változókat.

A $Z_{j,s}$, $1 \le j \le k$, $1 \le s \le l_j$ koordinátákból álló vektor eloszlása megegyezik a $Z_{j,s,n} = \frac{\nu_n(j,s) - np_{j,s}}{\sqrt{np_{j,s}}}$, $1 \le j \le k$, $1 \le s \le l_j$, koordinátákat tartalmazó véletlen vektor határeloszlásával $n \to \infty$ esetén, mert ez olyan normális eloszlású vektor, amelyre $EZ_{j,s} = 0$, $\operatorname{Var} Z_{j,s} = 1 - p_{j,s}$, $1 \le j \le k$, $1 \le s \le l_j$, és $\operatorname{Cov}(Z_{j,s}, Z_{j',s'}) = -\sqrt{p_{j,s} p_{j',s'}}$, ha $1 \le j, j' \le k$, $1 \le s \le l_j$, $1 \le s' \le l_{j'}$, és $(j,s) \ne (j',s')$. Továbbá, ha ezeket a vektorokat kiegészítjük a Z_j, $1 \le j \le k$ koordinátákkal akkor ezen vektorok eloszlása megegyezik azon vektorok határeloszlásával $n \to \infty$ esetében, amelyeket úgy kapunk, hogy a $Z_{j,s,n}$ koordinátákból álló vektort kiegészítjük a $Z_{j,n}$, $1 \le j \le k$ koordinátákkal. Ehhez elég észrevenni, hogy $Z_{j,n} = \frac{1}{\sqrt{p_j}} \sum\limits_{s=1}^{l_j} \sqrt{p_{j,s}} Z_{j,s,n}$, és hasonló reláció érvényes a Z_j és $Z_{j,s}$ vektorokra is.

A fentiek alapján az $(U_n^{(0)}, U_n^{(1)})$ vektorok határeloszlása megegyezik az $U^{(0)} = \sum\limits_{j=1}^{k} \sum\limits_{s=1}^{l_j} Z_{j,s}^2$ és $U^{(1)} = \sum\limits_{j=1}^{k} Z_j^2$ valószínűségi változókból álló $(U^{(0)}, U^{(1)})$ véletlen vektor eloszlásával.

Azt állítjuk, hogy $\sum\limits_{s=1}^{l_j} Z_{j,s}^2 - Z_j^2 = \sum\limits_{s=1}^{l_j} X_{j,s}^2 - X_j^2$ minden $1 \le j \le k$ számra. Valóban, $\sum\limits_{s=1}^{l_j} Z_{j,s}^2 = \sum\limits_{s=1}^{k_j} \left(X_{j,s} - \sqrt{p_{j,s}} X\right)^2 = \sum\limits_{s=1}^{l_j} X_{j,s}^2 + p_j X^2 - 2\sqrt{p_j} X X_j$, és $Z_j^2 = X_j^2 + p_j X^2 - 2\sqrt{p_j} X X_j$. E két azonosságot kivonva egymásból kapjuk a kívánt állítást.

Vezessük be a $V_j = \sum\limits_{s=1}^{l_j} Z_{j,s}^2 - Z_j^2$, $1 \le j \le k$ valószínűségi változókat. A fenti azonosságból és a 3. feladat eredményéből következik, hogy a (V_j, X_j) vektor koordinátái függetlenek egymástól, V_j $\chi^2(l_j - 1)$, X_j pedig standard normális eloszlású valószínűségi változó minden $1 \le j \le k$ indexre. Valóban, rögzítsünk egy $1 \le j \le k$ indexet, vezessük be a $q_s = \frac{p_{j,s}}{p_j}$ számokat, és használjuk fel a $\sum\limits_{s=1}^{l_j} Z_{j,s}^2 - Z_j^2 = \sum\limits_{s=1}^{l_j} X_{j,s}^2 - X_j^2$ azonosságot. Ekkor $X_j = \sum\limits_{s=1}^{l_j} \sqrt{q_s} X_{j,s}$, és bevezetve az $Y_{j,s} = X_{j,s} - \sqrt{q_s} X_j$ valószínűségi változókat, felírhatjuk, hogy $\sum\limits_{s=1}^{l_j} Y_{j,s}^2 = \sum\limits_{s=1}^{l_j} X_{j,s}^2 - X_j^2$. Ezért a 3.) feladat eredménye a benne szereplő változók alkalmas választásával megadja a kívánt eredeményt.

Továbbá a (V_j, X_j), $1 \le j \le k$, vektorok függetlenek egymástól, mert a (V_j, X_j) vektor az $X_{j,s}$, $1 \le s \le l_j$, valószínűségi változók függvénye. Vegyük továbbá észre, hogy a $V = \sum\limits_{j=1}^{k} X_j^2 - X^2$ vektorra $V = \sum\limits_{j=1}^{k} Z_j^2$, (mellesleg a V vektor független az X vektortól, de erre a tényre nem lesz szükségünk), és V $\chi^2(k-1)$ eloszlású vektor. Továbbá a konstrukció szerkezetéből és a V_j, X_j, $1 \le j \le k$, valószínűségi változók függetlenségéből következik, hogy a V_j, $1 \le j \le k$, és V véletlen vektorok függetlenek egymástól. Ezért a V és $\sum\limits_{j=1}^{k} V_j$ valószínűségi változók egymástól függetlenek $\chi^2(k-1)$ és $\chi^2(L-k)$ eloszlással. Ezenkívül az $U = (U_0, U_1)$ véletlen vektor megegyezik a $\left(V + \sum\limits_{j=1}^{k} V_j, V \right)$ vektorral. Innen következik a feladat állítása.

Megjegyzés: A bizonyítás módszere jobban érthető, ha felidézzük azt, hogy az egyes urnákba eső pontok számának alkalmas normalizáltjainak határeloszlása kifejezhető mint egy Brown bridge növekményeinek az eloszlása alkalmas intervallumokon. Továbbá egy Brown bridge-t elő lehet állítani, $B(t) = W(t) - tW(1)$ alakban, ahol $W(t)$, $0 \le t \le 1$, Wiener folyamat, és Wiener folyamattal egyszerűbb számolni mint Brown bridge-zsel, mert az független növekményű folyamat. A bizonyításban szereplő konstrukció hátterében ez az észrevétel áll. Jegyezzük meg, hogy ha a Brown bridge-t $B(t) = W(t) - tW(1)$, $0 \le t \le 1$, alakban reprezentáljuk egy Wiener folyamat segítségével, akkor ez a $B(\cdot)$ Brown bridge független a $W(1)$ valószínűségi változótól. Ez tény felel meg a 3. feladat állításának az $X_1^2 + \cdots + X_k^2 - \left(\sqrt{p_1} X_1 + \cdots + \sqrt{p_k} X_k\right)^2$ és $\sum\limits_{j=1}^{k} \sqrt{p_j} X_j$ valószínűségi változók függetlenségéről.

4.′) Először konstruálunk olyan $Z_s^{(j)}$, $1 \le j \le t$, $1 \le s \le l_j$ koordinátákból álló normális eloszlású véletlen vektort, amelynek eloszlása a $Z_{n,s}^{(j)} = \frac{\nu_n^{(j)}(s) - np_s^{(j)}}{\sqrt{np_s^{(j)}}}$, $1 \le j \le t$, $1 \le s \le l_j$ koordinátákból álló véletlen vektoroknak, határeloszlása $n \to \infty$ esetben, és amelyekkel viszonylag egyszerű számolni. Ennek érdekében tekintsünk X_r, $1 \le r \le k$, független standard normális eloszlású valószínűségi változókat, és definiáljuk segítségükkel az

$$
\begin{aligned}
X_s^{(j)} &= \frac{1}{\sqrt{p_s^{(j)}}} \sum_{r:\ r \in A_{j,s}} \sqrt{p_r} X_r, \qquad 1 \le j \le t, \quad 1 \le s \le l_j, \\
X &= \sum_{s=1}^{l_j} \sqrt{p_s^{(j)}} X_s^{(j)} = \sum_{r=1}^{k} \sqrt{p_r} X_r, \\
Z_s^{(j)} &= X_s^{(j)} - \sqrt{p_s^{(j)}} X, \qquad 1 \le j \le t, \quad 1 \le s \le l_j,
\end{aligned}
$$

valószínűségi változókat. Jegyezzük meg, hogy az X valószínűségi változó definíciója korrekt volt, azaz X valóban nem függ az első felírásában szereplő j indextől. Továbbá minden $1 \le j' \le j \le t$ és $1 \le s \le l_{j'}$ számpárra érvényes az alábbi azonosság:

$$
Z_s^{(j')} = \frac{1}{\sqrt{p_s^{(j')}}} \sum_{u:\ A_{j,u} \subset A_{j',s}} \sqrt{p_u^{(j)}} Z_u^{(j)}.
$$

Azt állítjuk, hogy a fent definiált $Z_s^{(j)}$, $1 \le j \le t$, $1 \le s \le l_j$, koordinátákból álló véletlen vektor eloszlása megegyezik a $Z_{n,s}^{(j)}$ $1 \le j \le t$, $1 \le s \le l_j$, koordinátákból álló véletlen vektorok határeloszlásával, ha $n \to \infty$. Ennek megmutatása érdekében tekintsük először a legfinomabb particiót, azaz a $j = t$ esetet és vegyük észre, hogy mivel $EZ_s^{(t)} = 0$, $\operatorname{Var} Z_s^{(t)} = 1 - p_s^{(t)}$, $1 \le s \le l_t$, $\operatorname{Cov}(Z_s^{(t)} Z_{s'}^{(t)}) = -\sqrt{p_s^{(t)} p_{s'}^{(t)}}$, ha $1 \le s, s' \le l_t$, és $s \ne s'$, ezért a a $Z_{n,s}^{(t)}$, $1 \le s \le l_t$, koordinátákból álló vektorok eloszlásban konvergálnak a $Z_s^{(t)}$, $1 \le s \le l_t$, vektorokból álló véletlen vektorhoz, ha $n \to \infty$. Mivel

$$
Z_{n,s}^{(j)} = \frac{1}{\sqrt{p_s^{(j)}}} \sum_{u:\ A_{t,u} \subset A_{j,s}} \sqrt{p_u^{(t)}} Z_{n,u}^{(j)},
$$

minden $1 \le j \le t$ és $1 \le s \le l_j$ számra, és hasonló reláció érvényes a $Z_s^{(j)}$ valószínűségi változókra, ezért az előző állításból következik a kívánt határeloszlás érvényessége.

A 4. feladatban végrehajtott számoláshoz hasonlóan kapjuk, hogy minden $1 < j \le t$, és $1 \le s \le l_{j-1}$ számokra érvényes a

$$
\begin{aligned}
V_{j-1,s} &= \sum_{r:\ A_{j,r} \subset A_{j-1,s}} Z_r^{(j)^2} - Z_s^{(j-1)^2} = \sum_{r:\ A_{j,r} \subset A_{j-1,s}} X_r^{(j)^2} - X_s^{(j-1)^2} \\
&= \sum_{r:\ A_{j,r} \subset A_{j-1,s}} \left(X_r^{(j)} - \frac{\sqrt{p_r^{(j)}}}{\sqrt{p_s^{(j-1)}}} X_s^{(j-1)} \right)^2
\end{aligned}
$$

azonosság. Továbbá, $EX_s^{(j-1)}\left(X_r^{(j)}-\frac{\sqrt{p_r^{(j)}}}{\sqrt{p_s^{(j-1)}}}X_s^{(j-1)}\right)=0$ minden $1\le s\le l_{j-1}$ számra és olyan r indexre, amelyre $A_{j,r}\subset A_{j-1,s}$. Ezért a többdimenziós normális eloszlású valószínűségi változók tulajdonságaiból következik, hogy az $X_s^{(j-1)}$ valószínűségi változó független az $\left(X_r^{(j)}-\frac{\sqrt{p_r^{(j)}}}{\sqrt{p_s^{(j-1)}}}X_s^{(j-1)},\ r\in A_{j-1,s}\right)$ véletlen vektortól. Innen, és a $V_{j-1,s}$ valószínűségi változók definiciójából következik, hogy az $X_s^{(j-1)}$ és $V_{j-1,s}$ valószínűségi változók függetlenek egymástól. Továbbá, mivel az $X_s^{(j-1)}$ és $V_{j-1,s}$ valószínűségi változók kifejezhetőek mint azon X_u valószínűségi változók függvényei, amelyekre $u\in A_{j-1,s}$, ezért az $(X_s^{(j-1)},V_{j-1},s)$, $1\le s\le l_{j-1}$, véletlen vektorok függetlenek. A fenti állításokból következik, hogy az $X_s^{(j-1)}$, $V_{j-1,s}$ valószínűségi változók, $1\le s\le l_{j-1}$, függetlenek egymástól.

Vezessük be az $U_j=\sum\limits_{s=1}^{l_j}{Z_s^{(j)}}^2$ valószínűségi változókat, $1\le j\le t$. A $V_{j-1,s}$ valószínűségi változókra bebizonyított azonosságokat összeadva kapjuk, hogy $U_j-U_{j-1}=\sum\limits_{s=1}^{l_{j-1}}V_{j-1,s}$. A $V_{j-1,s}$, $1\le s\le l_{j-1}$ valószínűségi változók függetlenek, továbbá a 3. feladat eredménye szerint $V_{j-1,s}$ $\chi^2(m(j-1,s)-1)$ eloszlású, ahol $m(j-1,s)$ az $A_{j-1,s}$ halmaz számosságát jelöli. Ezért U_j-U_{j-1} $\chi^2(l_j-l_{j-1})$ eloszlású. Továbbá az előző paragrafusban bizonyított függetlenségi relációból következik, hogy az $U_j-U_{j-1}=\sum\limits_{s=1}^{l_{j-1}}V_{j-1,s}$ valószínűségi változó minden $1<j\le t$ számra független az $\left\{X_s^{(j-1)},\ 1\le s\le l_{j-1}\right\}$ véletlen vektortól. Ezek az állítások a $j=1$ esetben is igazak, ha az $U_0=0$ és $l_0=1$ definiciót alkalmazzuk. (A $j=1$ esetben az X vektor játszhatja az U_1-U_0 valószínűségi változótól független vektorok rendszerét, de erre az észrevételre nincs szükségünk.)

Az U_m-U_{m-1}, $1\le m<j$, valószínűségi változók az $\left\{X_s^{(j-1)},\ 1\le s\le l_j\right\}$ véletlen vektor függvényeként is kifejezhetőek. Ezért az U_j-U_{j-1} valószínűségi változó független az $\{U_m-U_{m-1},\ 1\le m\le j-1\}$ véletlen vektortól. A fenti állításokból kapjuk, hogy az U_j-U_{j-1}, $1\le j\le t$, valószínűségi változók függetlenek egymástól, és $\chi^2(l_j-l_{j-1})$ eloszlásúak. Innen, és a U_j valószínűségi változók definiciójából következik a feladat állítása.

5.) Ha a golyók az egyes urnákba $p_1,\dots,p_k$ valószínűséggel esnek, akkor annak valószínűsége, hogy az első urnába m_1 a második urnába m_2, ... , a k-ik urnába m_k golyó esik $\frac{n!}{m_1!\cdots m_k!}p_1^{m_1}\cdots p_k^{m_k}$. Innen egyszerű számolással kapjuk az

$$L_n(p_1,\dots,p_k,m_1,\dots,m_k)$$

függvényre felírt kifejezést. Ennek szélsőértékét rögzített $m_1,\dots,m_k$ értékek esetén a Lagrange féle multiplikátor módszer segítségével (A $\sum\limits_{j=1}^{k}p_j=1$ feltétel miatt

érdemes alkalmazni a Lagrange féle multiplikátor módszert) a következő egyenletrendszer segítségével határozhatjuk meg: A Lagrange féle multiplikátor módszer szerint a $\sum\limits_{j=1}^{k} m_j \log p_j - \lambda(p_1 + \cdots + p_k)$ függvény parciális deriváltjai a $_j$ változó szerint nulla, és a $k+1$-ik egyenlet a $p_1 + \cdots + p_k = 1$ reláció.
Innen $\frac{m_j}{\hat{p}_j} - \lambda = 0$, $1 \le j \le k$. Az $m_j = \lambda p_j$ egyenleteket összeadva kapjuk hogy $\lambda = n$, ahonnan $\hat{p}_j = \frac{m_j}{n}$, $1 \le j \le k$, a maximum likelihood becslés a $p_1, \dots, p_k$ paraméterekre. Innen egyszerű számolással kapjuk a $B^{(n)}_{p_1,\dots,p_k}(m_1, \dots, m_k)$ kifejezésre megadott formulát. A $B^{(n)}_{p_1,\dots,p_k}(m_1, \dots, m_k)$ kifejezés aszimptotkájanak meghatározása érdekében írjuk fel a Taylor sorfejtés segítségével, hogy

$$\begin{aligned}
n\frac{m_j}{n} \log \frac{\frac{m_j}{n}}{p_j} &= -n\frac{m_j}{n} \log \left(1 + \frac{p_j - \frac{m_j}{n}}{\frac{m_j}{n}}\right) \\
&= -n\left(p_j - \frac{m_j}{n}\right) + \frac{1}{2}\frac{(m_j - np_j)^2}{n\frac{m_j}{n}} + \varepsilon_j(n) \\
&= -n\left(p_j - \frac{m_j}{n}\right) + \frac{1}{2}\frac{(m_j - np_j)^2}{np_j} + \bar{\varepsilon}_j(n)
\end{aligned}$$

minden $1 \le j \le k$ számra, ahol $\varepsilon_j(n) \Rightarrow 0$, $\bar{\varepsilon}_j(n) \Rightarrow 0$, ha $n \to \infty$, és $\Rightarrow$ sztochasztikus konvergenciát jelöl. Az $\varepsilon_j(n) \Rightarrow 0$ összefüggés azért teljesül, mert felhasználva azt, hogy az $\left(\frac{m_j}{n} - p_j\right)^3 n^{3/2}$ valószínűségi változó sztochasztikusan korlátos kapjuk, hogy a fenti Taylor sorban a harmadik tag hozadéka elhanyagolhatóan kicsi. Az itt használt érvelésekben felhasználjuk, hogy a centrális határeloszlástétel szerint $\frac{m_j - np_j}{\sqrt{np_j(1-p_j)}}$ eloszlásban konvergál a standard normális eloszláshoz. Ezután az $\frac{m_j}{np_j} \Rightarrow 1$ reláció és az $\frac{(m_j - np_j)^2}{n\frac{m_j}{n}}$ kifejezés sztochasztikus korlátossága miatt a fenti aszimptotikában a második sorozat helyettesíthető a harmadik sorral. A fenti relációkat összegezve minden $j = 1, \dots, k$-ra kapjuk, hogy, érvényes a

$$\begin{aligned}
B^{(n)}_{p_1,\dots,p_k}(m_1, \dots, m_k) &= -n\sum_{j=1}^{k}\left(p_j - \frac{m_j}{n}\right) + \frac{1}{2}\sum_{j=1}^{k}\frac{(m_j - np_j)^2}{np_j} + \varepsilon(n) \\
&= \frac{1}{2}\sum_{j=1}^{k}\frac{(m_j - np_j)^2}{np_j} + \varepsilon(n)
\end{aligned}$$

reláció, ahol $\varepsilon(n) \Rightarrow 0$, mivel $\sum\limits_{j=1}^{k}\left(p_j - \frac{m_j}{n}\right) = 0$. Ezért

$$\frac{1}{2}C_n(m_1, \dots, m_k) - B^{(n)}_{p_1,\dots,p_k}(m_1, \dots, m_k) \Rightarrow 0, \quad \text{ha } n \to \infty.$$

Továbbá, mivel minden $1 \le j \le k$ számra $\frac{m_j}{n} \Rightarrow p_j$, ha $n \to \infty$, és az $\frac{(m_j - np_j)^2}{np_j}$ valószínűségi változók sztochasztikusan korlátosak ezért

$$C_n(m_1, \dots, m_k) - \bar{C}_n(m_1, \dots, m_k) \Rightarrow 0, \quad \text{ha } n \to \infty.$$

Innen következik, hogy az utolsó relációban a C_n valószínűségi változót helyettesíthetjük a $\bar{C}_n$ valószínűségi változóval. Ezzel beláttuk a 5. feladat összes állítását.