A szeptember 11-i gyakorlat témája
Rövid összefoglaló

A gyakorlaton a következő az irodalomban coupon problémának nevezett problémával illetve néhány hozzá kapcsolódó kérdéssel foglalkoztunk:

Tegyük fel, hogy egy termék megvásárlásakor a termékkel valamilyen cimkét (coupont) is kapunk. Összesen n különböző tipusú coupon van, és ha mindegyiket összegyüjtjük, akkor értékes nyereményt kapunk. Minden vásárlásnál a lehetséges n tipusú coupon valamelyikét egyforma, $\frac{1}{n}$ valószínűséggel kapjuk meg, és az egyes vásárlások esetén szerzett couponok tipusa egymástól független. Hány vásárlást kell tennünk annak érdekében, hogy az összes lehetséges coupont megszerezzük?

Természetesen a szükséges vásárlások száma a véletlentől is függ. De értékes információt ad a szükséges vásárlások számáról, annak várható értéke és szórásnégyzete. Ugyancsak érdekel minket az, hogy a szükséges couponok p-ed részének megszerzéséhez hány vásárlás szükséges. Mi ennek a várható értéke és szórásnégyzete? E kérdés hátterében a következő probléma áll. Szeretnénk tudni, hogy ha a szükséges couponoknak például a 95%-át már megszereztük, akkor a szükséges vásárlások nagyrészét várhatólag már megtettük vagy esetleg a még szükséges 5% megszerzése még több vásárlást igényel. E kérdés eldöntése érdekében érdemes kiszámolni a fenti várható értékeket és szórásnégyzeteket.

Vezessük be a következő $\eta_1, \eta_2, \dots,$ valószínűségi változókat: $\eta_j = k$, $j = 1, 2, \dots,$ $1 \le k \le n$, ha a j-ik vásárlás során a kapott coupon tipusa k. Akkor feltételezésünk formálisan úgy fogalmazható meg, hogy az η_j, $j = 1, 2, \dots,$ valószínűségi változók függetlenek, egyforma eloszlásúak, és $P(\eta_j = k) = \frac{1}{n}$, $j = 1, 2, \dots,$ $1 \le k \le n$. Kérdésünk pedig úgy fogalmazható meg, hogy mi az a legkisebb véletlentől függő N szám, melyre az $\eta_1, \dots, \eta_N$ sorozat az $1, 2, \dots, n$ számok mindegyikét felveszi. Mi ennek a véletlen N számnak a várható értéke és szórásnégyzete? Illetve ennek a kérdésnek azt a megfelelőjét tekintettük, amelyben azt követeltük meg, hogy az $\eta_1, \dots, \eta_N$ sorozat legalább pn értéket vegyen fel. Mind a két probléma vizsgálatában érdemes bevezetni a következő ζ_j, $j = 1, 2, \dots, n$ valószínűségi változókat:

Legyen $\zeta_1 = 1$. Vegyenek fel a $\zeta_1, \zeta_2, \dots$ valószínűségi változók pozitív egész értékeket, és legyen $1 = \zeta_1 < \zeta_2 < \cdots$. Ha a ζ_s, $1 \le s \le j$, valószínűségi változókat már definiáltuk, akkor legyen ζ_{j+1} az a legkisebb m index, melyre $\eta_m \ne \eta_{\zeta_s}$, minden $1 \le s \le j$ számra. Szemléletesen a ζ_j, $j = 1, 2, \dots,$ számokat úgy definiáltuk mint azokat a vásárlási időpontokat, melyekben a j-ik alkalommal szereztünk egy új tipusú coupont. A feladat megoldása érdekében a ζ_j valószínűségi változók viselkedését kell megértenünk.

1. Lássuk be, hogy a $\zeta_1, \zeta_2, \dots,$ valószínűségi változók függetlenek, és $P(\zeta_j = l) = p_j(l) = \left(\dfrac{j-1}{n}\right)^{l-1} \dfrac{n-j+1}{n}$, $l = 1, 2, \dots$.

1a.) Minden $0 < p < 1$ számra létezik $p_n = (1-p)^{n-1}(1-p)$, $n = 1, 2, \dots$ alakú eloszlás, azaz létezik olyan ζ valószínűségi változó, melyre $P(\zeta = n) = (1-p)^{n-1}p$. Mi az ilyen alakú eloszlások szemléletes tartalma?

1b.) Legyenek $\zeta_1, \zeta_2, \dots,$ olyan pozitív egész értékeket felvevő valószínűségi változók, melyekre $P(\zeta_n = j_n | \zeta_1 = j_1, \dots, \zeta_{n-1} = j_{n-1}) = p_n(j_n)$ minden $j_1, \dots, j_n$ pozitív egészekből álló szám n-esekre. Lássuk be, hogy ekkor $\zeta_1, \zeta_2, \dots,$ független valószínűségi változók, és $P(\zeta_n = j) = p_n(j)$, minden $n = 1, 2, \dots$ számra.

1a.) *Megoldás:* Mivel $\sum\limits_{n=1}^{\infty} p_n = \sum\limits_{n=1}^{\infty} (1-p)p^{n-1} = (1-p)\dfrac{1}{1-p} = 1$, és $p_n \geq 0$ minden $n = 1, 2, \dots$ számra az adott p_n számok valóban eloszlást határoznak meg (a pozitív egész számokon). Tekintsük egy olyan pénzdarab egymástól független dobásait, melyek p valószínűséggel esnek a fej, és $1-p$ valószínűséggel az írás oldalra. Akkor p_n annak valószínűsége, hogy az n-ik dobásban jelenik meg az első fej-dobás. Ezt az eloszlást az irodalomban (p paraméterű) negatív binomiális eloszlásnak nevezik. Az 1. feladat állítása szerint ζ_j $p_j = \dfrac{j-1}{n}$ paraméterű negatív binomiális eloszlású valószínűségi változó.

1b.) *Megoldás:* Azt kell belátni, hogy

$$P(\zeta_n = j_n, \zeta_{n-1} = j_{n-1}, \dots, \zeta_1 = j_1) = \prod_{k=1}^{n} p_k(j_k)$$

Viszont

$$\begin{aligned} &P(\zeta_n = j_n, \zeta_{n-1} = j_{n-1}, \dots, \zeta_1 = j_1) \\ &\quad = \prod_{k=2}^{n} P(\zeta_k = j_k | \zeta_{k-1} = j_{k-1}, \dots, \zeta_1 = j_1) P(\zeta_1 = j_1) = \prod_{k=1}^{n} p_k(j_k) \end{aligned}$$

a feladat feltételei szerint.

1. *Megoldás:* Az 1b.) feladat eredménye alapján elég belátni, hogy

$$P(\zeta_k = j_k | \zeta_{k-1} = j_{k-1}, \dots, \zeta_1 = j_1) = \left(\frac{k-1}{n}\right)^{j_k - 1} \frac{n-k+1}{n}$$

minden $k = 1, 2, \dots,$ $j_k = 1, 2, \dots$ számokra. Ehhez elegendő belátni, hogy az $M = \sum\limits_{l=1} k_l$ pozitív egész számra és minden olyan $\eta_1, \dots, \eta_M$ sorozatra, melyekre az $\eta_1, \dots, \eta_{M-1}$ sorozat pontosan $k-2$ különböző értéket vesz fel, az η_M valószínűségi változó pedig egy új $k-1$-ik értéket vesz fel, teljesül az alábbi azonosság: (Az előbb megfogalmazott feltételek azt fejezik ki, hogy egy olyan vásárlás sorozat történt, amelyikben az M-ik vásárláskor szereztük meg a $k-1$-ik coupont.) Megmutatjuk, hogy

$$P(\zeta_k = j_k | \eta_1 = l_1, \dots, \eta_M = l_M) = \left(\frac{k-1}{n}\right)^{j_k - 1} \frac{n-k+1}{n}$$

pozitív egész számok tetszőleges olyan $l_1, \dots, l_M$ sorozatára, amelyik pontosan $k-1$ különböző értéket vesz fel. Annak belátásához, hogy az utolsó összefüggésből következik a kívánt állítás azt kell észrevenni, hogy az $\{\eta_1 = l_1, \dots \eta_M = l_M\}$ események diszjunktak, a $\{\zeta_{k-1} = j_{k-1}, \dots, \zeta_1 = j_1\}$ esemény ilyen alakú (diszjunkt) események uniója, és az utolsó azonosság jobboldalán szereplő valószínűség nem függ az $\{\eta_1 = l_1, \dots \eta_M = l_M\}$ feltételtől.

A bizonyítandó állítás viszont egyszerűen ellenőrizhető. Azt kell észrevenni, hogy egymástól független kisérleteket kell vizsgálni, melyek $p = \dfrac{n-k+1}{n}$ valószínűséggel járnak sikerrel és $1-p$ valószínűséggel sikertelenek. Azt kell nézni, hogy mi annak a valószínűsége, hogy pontosan a j_k-ik kisérlet lesz sikeres. A felírt azonosság ezt a valószínűséget adja meg. (Kisérletnek az M-ik vásárlás után következő új vásárlásokat nevezzük, és e kisérleteket akkor nevezzük sikeresnek, ha a vásárlás során új coupont szereztünk.)

Ezután megbeszéltük azt, hogyan kell kiszámítani valószínűségi változük összegének a várható értékét és szórásnégyzetét, mi a jelentősége a függetlenségnek, hogyan lehet számolni abban az esetben is, ha a függetlenség feltéte nem teljesül. Ezt a kérdést későbbi gyakorlatokon részletesen tárgyaltuk, ezért itt nem írom le. Az vizsgált problémában független valószínűségi változók összegének a várható értékét kell vizsgálnunk. A feladat megoldása érdekében érdemes külön tárgyalni a bevezető valószínűség előadáson szerepelt negatív binomiális eloszlás várható értékét és szörásnégyzetét.

2. Számoljuk ki egy $(1,p)$ paraméterű negativ binomiális eloszlású valószínűségi változó azaz egy olyan ξ valószínűségi változó várható értékét és szórásnégyzetét, melyre $P(\xi = k) = p^{k-1}(1-p)$, $k = 1, 2, \dots$.

   Megoldás: Ha ξ $(1,p)$ paraméterű negativ binomiális eloszlású valószínűségi változó, akkor
$$E\xi = \sum_{k=1}^{\infty} k(1-p)p^{k-1}, \qquad E\xi^2 = \sum_{k=1}^{\infty} k^2(1-p)p^{k-1}.$$

   Ezen összegek egyik lehetséges kiszámolása: $\sum\limits_{k=0}^{\infty} x^k = \dfrac{1}{1-x}$, ha $|x| < 1$. Két egymás utáni deriválással kapjuk, hogy $\sum\limits_{k=1}^{\infty} kx^{k-1} = \dfrac{1}{(1-x)^2}$ és $\sum\limits_{k=2}^{\infty} k(k-1)x^{k-2} = \dfrac{2}{(1-x)^3}$. Innen $\sum\limits_{k=1}^{\infty} kp^{k-1} = \dfrac{1}{(1-p)^2}$, $\sum\limits_{k=1}^{\infty} k^2p^{k-1} = \sum\limits_{k=1}^{\infty} kp^{k-1} + p\sum\limits_{k=2}^{\infty} k(k-1)p^{k-2} = \dfrac{1}{(1-p)^2} + \dfrac{2p}{(1-p)^3}$, és $E\xi = \dfrac{1}{1-p}$, $E\xi^2 = \dfrac{1}{1-p} + \dfrac{2p}{(1-p)^2}$. Ezért $\operatorname{Var}\xi = E\xi^2 - (E\xi)^2 = \dfrac{p}{(1-p)^2}$.

3. Oldjuk meg a bevezetésben megfogalmazott feladatot. Lássuk be, hogy n coupon esetén az összes coupon megszerzéséhez szükséges vásárlások ζ számának várható értéke $E\zeta = \sum\limits_{k=1}^{n} \dfrac{n}{n-k+1}$, szórásnégyzete pedig $\operatorname{Var}\zeta = \sum\limits_{k=1}^{n} \dfrac{n(k-1)}{(n-k+1)^2}$. Számoljuk ki ezen mennyiségek aszimptotikus viselkedését nagy n számokra.

   Megoldás: Láttuk, hogy $\zeta = \sum\limits_{k=1}^{n} \zeta_k$, ahol ζ_k független valószínűségi változók negatív binomiális eloszlással $p_k = \frac{k-1}{n}$ paraméterrel, $k = 1, \dots, n$. Ezért $E\zeta = \sum_{k=1}^{n} E\zeta_k$, $\operatorname{Var}\zeta = \sum_{k=1}^{n} \operatorname{Var}\zeta_k$. Továbbá, az előző feladat alapján $E\zeta_k = \frac{n}{n-k+1}$, $\operatorname{Var}\zeta_k = \frac{n(k-1)}{(n-k+1)^2}$ minden $k = 1, \dots, n$ számra. Innen következnek a felírt azonosságok. Továbbá nagy n-re $E\zeta = \sum_{k=1}^{n} \frac{n}{n-k+1} \sim n \int_1^n \frac{1}{n-u+1}\, du = n \log n$. Hasonlóan $\operatorname{Var}\zeta \sim n \int_1^n \frac{(u-1)}{(n-u+1)^2}\, du$. Továbbá

$$\int_1^n \frac{(u-1)}{(n-u+1)^2}\, du = -\int_1^n \frac{1}{n-u+1}\, du + n \int_1^n \frac{1}{(n-u+1)^2}\, du \sim \int_1^\infty \frac{n}{u^2}\, du.$$

Ezért $\operatorname{Var}\zeta \sim \frac{n^2}{3}$.

4. Számoljuk ki annak várható értékét és szórásnégyzetét, hogy hány vásárlás szükséges ahhoz, hogy a szükséges couponok p-ed részét megszerezzük. Számoljuk ki a várható érték aszimptotikáját nagy n és rögzített p számra.

   Megoldás: Ezt a feladatot az elöző feladathoz hasonlóan oldhatjuk meg. Az egyetlen különbség, hogy most nem egytől n-ig, hanem 1-től pn-ig kell összegezni. Ezért a várható érték $\sum_{k=1}^{np} \frac{n}{n-k+1}$, a szórásnégyzet pedig $\sum_{k=1}^{np} \frac{n(k-1)}{(n-k+1)^2}$. Hasonlóan az előző feladat számolásához a várható érték nagy n-re közelítőleg

$$\int_1^{pn} \frac{n}{n-u+1}\, du = n \int_{n-np+1}^{n} \frac{du}{u} = n \log \frac{n}{n(1-p)+1} \sim n \log \frac{1}{1-p}.$$

   (Hasonló számolás mutatja, hogy a szórásnégyzet is n-nel lesz arányos.)