3. feladatsor

Egyszerű elágazó folyamat és nagy eltérések

2010. november 10. és 15.

1. Egy vállalkozó üzleteket nyit sorban. Idén nyitotta meg az elsőt, és innentől kezdve minden egyes üzlet minden évben
   - 0,2 valószínűséggel becsődöl, és be kell zárni;
   - 0,4 valószínűséggel nullszaldós lesz, ekkor folytatja a működést a következő évben;
   - 0,4 valószínűséggel akkora hasznot termel, amennyiből meg lehet nyitni még egy üzletet.

   (Úgy tekintjük, hogy minden üzlet teljesítménye a többitől és a saját korábbi teljesítményétől is független.)

   Modellezzük a feladatot elágazó folyamattal. Írjuk fel az utódeloszlás generátorfüggvényét.

   (a) Számítsuk ki, mennyi az üzletek számának várható értéke 5 év múlva.

   (b) Mekkora a valószínűsége, hogy a vállalkozó legfeljebb 2 éven belül becsődöl (azaz bezárja az összes üzletét)?

   (c) Mekkora a valószínűsége, hogy sohasem csődöl be teljesen a vállalkozó?

   Megoldás: Generátorfüggvény: $g(z) = 0,2 + 0,4z + 0,4z^2$.

   (a) Az órán tanultak miatt ez a várható érték

   $$\begin{aligned}(g(g(g(g(g(z))))))'\Big|_{z=0} &= g'(g(g(g(g(z)))))g'(g(g(g(z))))g'(g(g(z)))g'(g(z))g'(z)\Big|_{z=0} \\ &= (g'(1))^5 \\ &= 1,2^5.\end{aligned}$$

   (b) Ha Y_k a k-adik generáció mérete, akkor $\mathbf{P}(Y_k = 0) = \underbrace{g \circ \cdots \circ g}_{k}(0)$. Most $k = 2$-re $g(g(0)) =$ $0,2+0,4(0,2+0,4z+0,4z^2)+0,4(0,2+0,4z+0,4z^2)^2 = 0,2+0,4\cdot 0,2+0,4\cdot 0,2^2 = 0,296.$

   (c) A $g(z) = z$ egyenlet legkisebb $(0,1]$-beli megoldása adja az elágazó folyamatban a kihalás valószínűségét. Ez itt a

   $$0,4z^2 - 0,6z + 0,2 = 0$$

   másodfokú egyenletet adja, aminek a két megoldása 1/2 és 1. Tehát

   $$\mathbf{P}(\text{soha nincs csőd}) = \mathbf{P}(\text{túlélés}) = 1 - \mathbf{P}(\text{kihalás}) = 1 - \frac{1}{2} = \frac{1}{2}.$$

2. Tegyük fel, hogy egy végtelennek tekinthető számítógép populációban ha egy számítógép megfertőződik egy bizonyos típusú kártékony programmal, akkor a következő 1 nap során $p = \frac{1}{4}$ valószínűséggel kiirtják, így nem fertőz meg mást. $(1-p)p$ valószínűséggel nem irtják ki, de nem is fertőz. Továbbá $(1-p)^k p$ valószínűséggel k nem fertőzött gépet fertőz meg, ha beleszámoljuk a fertőzés forrásának tekinthető gépet, $k = 2, 3, 4, \ldots$. Tegyük fel, hogy minden megfertőzött számítógép az összes többi fertőzött számítógéptől függetlenül „szaporodik”.

   Modellezzük az egyes napokon a fertőzött számítógépek számát elágazó folyamattal. Milyen eloszlású az utódeloszlás?

   Tegyük fel, hogy az első napon egy fertőzött gép van és $p = \frac{1}{4}$. Írjuk fel az utódeloszlás generátorfüggvényét.

(a) Számítsuk ki, mennyi a fertőzött számítógépek várható száma 30 nap múlva.

(b) Mekkora a valószínűsége, hogy 3 nap múlva már nincs fertőzött gép?

(c) Mekkora a valószínűsége, hogy sohasem sikerül kiirtani a kártevőt?

Megoldás: Az utódeloszlás generátorfüggvénye $g(z) = \frac{p}{1-(1-p)z}$ (ld. 5. feladat megoldása).

(a) Utódeloszlás várható értéke 3, ezért az 1. feladatban levezetettek miatt a 30 nap múlva fertőzöttek számának várható értéke 3^{30}.

(b) $g(g(g(0))) = \frac{13}{40}$.

(c) $g(z) = z$ megoldása, ami az $(1-p)z^2 - z + p = 0$ egyenletre vezet.

$$z_{1,2} = \frac{1 \pm \sqrt{1-4(1-p)p}}{2(1-p)} = \frac{1 \pm \sqrt{(1-2p)^2}}{2(1-p)} = \frac{1 \pm (1-2p)}{2(1-p)},$$

ami $\frac{p}{1-p}$ vagy 1. A $\frac{p}{1-p} < 1$ egyenlőtlenség pontosan akkor teljesül, ha $p < \frac{1}{2}$, vagyis pontosan ekkor lesz $\frac{p}{1-p}$ a kihalás valószínűsége, egyébként 1. Ez annak is a valószínűsége, hogy sosem sikerül kiirtani a kártevőt.

3. Móricka frissen nyitott szoftver kereskedése nagyon jól megy. Egy rendelés kitöltésére fordított ideje 3 perc. Ezalatt j új vásárló áll be a várakozók közé p_j valószínűséggel, ahol $p_0 = 0,2$, $p_1 = 0,2$ és $p_2 = 0,6$. Tegyük fel, hogy a várakozó szoba kapacitása végtelen. Móricka csak akkor tud kávészünetet tartani, ha nincs várakozó igény. Mi a valószínűsége, hogy Mórickának valaha lesz kávészünete?

Ha lefixáljuk $p_1 = 0,2$-t és feltesszük, hogy $p_0 + p_1 + p_3 = 1$, akkor milyen érkezés eloszlás a feltétele a stabil kiszolgáló rendszernek, azaz 1 valószínűséggel elfogynak az igények véges időn belül?

Megoldás: Az általánosabb esetet oldjuk meg. Legyen $p_1 = 0,2$, p_0 paraméter, ekkor $p_2 = 0,8 - p_0$. A generátorfüggvény: $g(z) = p_0 + 0,2z + (0,8 - p_0)z^2$. A stabil kiszolgálórendszer feltétele, hogy az elágazó folyamat 1 valószínűséggel kihal. A kihalás valószínűsége a $g(z) = z$ egyenlet legkisebb $(0,1]$-beli megoldása. Ez a $(0,8 - p_0)z^2 - 0,8z + p_0 = 0$ egyenletre vezet.

$$z_{1,2} = \frac{0,8 \pm \sqrt{0,8 - 4(0,8-p)p}}{2(0,8-p)} = \frac{1 \pm \sqrt{(0,8-2p)^2}}{2(0,8-p)} = \frac{1 \pm (0,8-2p)}{2(0,8-p)},$$

aminek $\frac{p_0}{1-p_0}$ és 1 a két megoldása. Pontosan akkor 1 a kihalás valószínűsége, ha $\frac{p_0}{1-p_0} \geq 1$, azaz $p_0 \geq 0,4$.

4. Ottó gyermekeinek száma 0, 1, 2 vagy 3, mindegyik $1/4 - 1/4$ valószínűséggel. Ottó minden egyes leszármazottjának gyermekeinek száma is ugyanilyen eloszlású és a többiekétől független.

(a) Jelölje X Ottó ükunokáinak számát. Adjuk meg X generátorfüggvényét. Számítsuk ki $\mathbf{E}X$ és $\mathbf{D}X$ értékét.

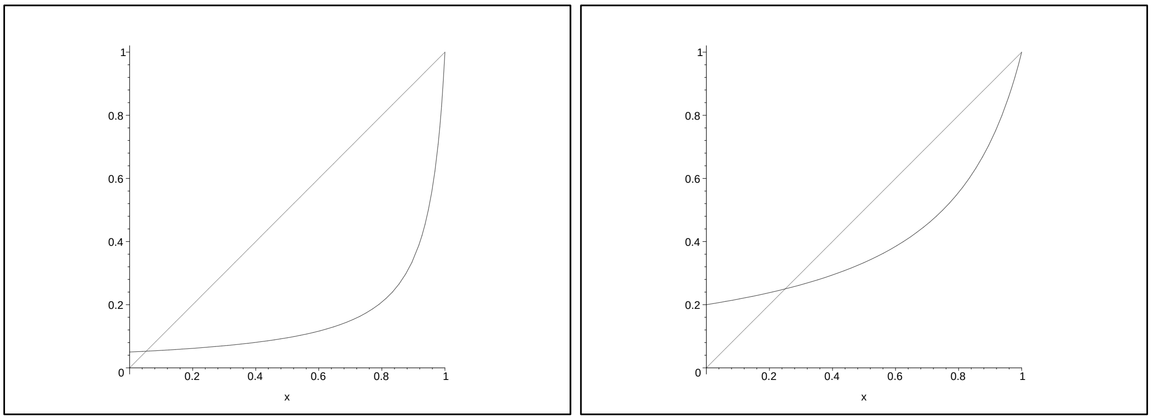
(b) Számítsuk ki annak az esélyét, hogy Ottó leszármazottai előbb-utóbb kihalnak.

5. Jelölje $\theta(p)$ annak a valószínűségét, hogy soha nem pusztul ki egy olyan elágazó folyamat, amelyben az utódok eloszlása geometriai eloszlású p paraméterrel. Rajzoljuk fel a $p \mapsto \theta(p)$ függvény grafikonját.

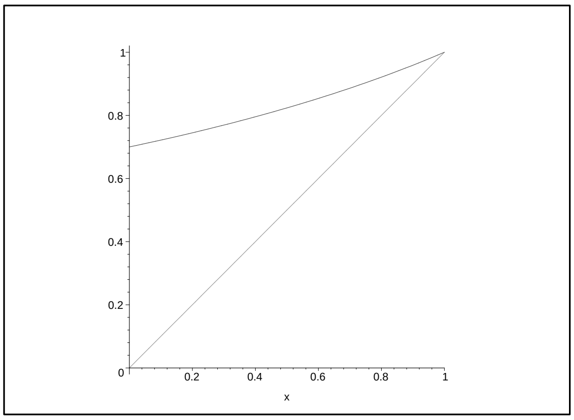
Megoldás: Először kiszámítjuk rögzített p érték esetén az utódeloszlás generátorfüggvényét.

$g(z) = \sum_{k=0}^{\infty} \mathbf{P}(X = k)z^k = \sum_{k=0}^{\infty} p(1-p)^k z^k = \frac{p}{1-(1-p)z}$, mert egy $(1-p)z$ hányadosú geometriai sor összege.

A kihalás valószínűsége a $g(z) = z$ egyenlet legkisebb nemnegatív gyöke. Hogy könnyebb legyen elképzelni, ábrázoljuk néhány p értékre a $g(z)$ függvényt. A bal oldali ábrán $p = 0,05$, a jobboldalin $p = 0,2$ (az ábrákra a z függvény is be van rajzolva).



Az alábbi ábrán a $p = 0,7$-hez tartozó $g(z)$ van berajzolva. $g(z)$ csak $z = 1$-ben metszi a z függvény grafikonját, ezért a kihalás valószínűsége 1.



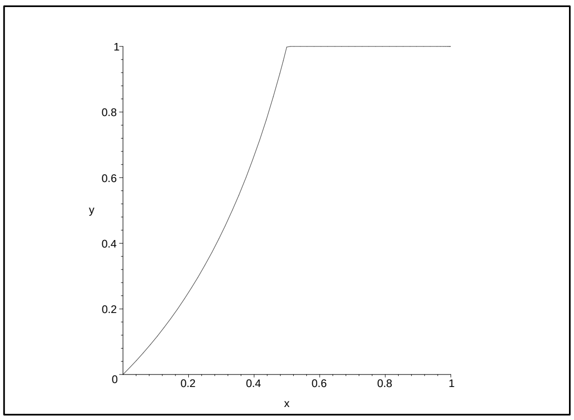
Általában igaz a következő: a kihalás valószínűsége 1 $\iff$ $G(z)$ végig z felett halad $\iff$ $g'(1) \leq 1$. ($g'(1)$ egyébként az utódeloszlás várható értéke.)

$g'(1) = \frac{1-p}{p}$, ami $p \geq 1/2$ esetén kisebb vagy egyenlő, mint 1, így $p \geq 1/2$ esetén a kihalás valószínűsége 1. A $p < 1/2$ esetben meg kell oldanunk a $g(z) = z$ egyenletet:

$\frac{p}{1-(1-p)z} = z \iff z = \frac{p}{1-p}$

Így a kihalás valószínűsége: $\theta(p) = \begin{cases} \frac{p}{1-p} & \text{ha } x < \frac{1}{2} \\ 1 & \text{ha } x \geq \frac{1}{2} \end{cases}$

A $\theta(p)$ függvény ábrázolva:



Hoeffding-egyenlőtlenség, centrális határeloszlás-tétel

6. Egy helyi internetszolgáltató 2015-ben 12000 előfizetőt szolgál ki. Csúcsidőben az előfizetők az előfizetésük, és internetezési szokásaik alapján a következő 3 kategóriába sorolhatók:

   - kezdő: az átlagos sávszélesség fogyasztás 100 Mbps, de semmiképpen nem több, mint 200 Mbps;
   - haladó: az átlagos sávszélesség fogyasztás 160 Mbps, de semmiképpen nem több, mint 280 Mbps;
   - mester: az átlagos sávszélesség fogyasztás 250 Mbps, de semmiképpen nem több, mint 400 Mbps.

   Az egyes kategóriákba tartozó felhasználók száma rendre 3500, 6500 és 2000. Mekkora legyen a kiszolgáló sávszélesség kapacitása, ha azt szeretnénk, hogy legfeljebb 10^{-6} eséllyel legyen kevés a csúcsidei szükséglet fedezésére? Mi lesz a kapacitás, ha a kapacitás túllépés valószínűsége legfeljebb 10^{-7}, illetve 10^{-8}?

7. Két egymáshoz közeli szélfarmon 400 A típusú és 200 B típusú szélerőművet telepítettek. Az A típusú termelése 0,5 MW és 1,6 MW között ingadozik 1 MW átlagos termeléssel. A B típusú termelése 1,2 MW és 2,8 MW között van, átlagosan 2 MW.

   (a) Számítsuk ki, hogy mekkora az a kapacitás, amit legalább $1 - 10^{-8}$ valószínűséggel nem lép túl a 600 erőmű össztermelése.

   (b) Tegyük fel, hogy adott egy névleges kapacitás, $C = 800$ MW. Mekkora valószínűséggel lépi túl C-t az aktuális kapacitás?

   (c) Oldjuk meg az előző feladatot, ha 360 A típusú és 200 B típusú szélerőművet kapcsolnak be. C ugyanaz.

   (d) Tegyük fel, hogy a B típusúak mind be vannak kapcsolva. Ha $C = 800$ MW-ot nem szeretnénk túllépni legalább $1 - 10^{-8}$ valószínűséggel, akkor hány A típusú erőművet kapcsoljunk be.

   Megoldás: Legyen $X_1, \dots, X_{400}$ az A típusú erőművek termelése, $Y_1, \dots, Y_{200}$ pedig a B típusúaké. Jelölje $S = \sum_{k=1}^{400} X_k + \sum_{k=1}^{200} Y_k$. Ekkor $\mathbf{E}(S) = 800$.

   (a) Olyan t számot keresünk, amelyre $\mathbf{P}(S - \mathbf{E}(S) \leq t) \geq 1 - 10^{-8}$, vagyis $\mathbf{P}(S - \mathbf{E}(S) \geq t) \leq 10^{-8}$. A legkisebb olyen t, amire a Höffding-egyenlőtlenségből adódó felső korlát sem nagyobb 10^{-8}-nál, az

   $$\exp\left(-\frac{2t^2}{400 \cdot 1,1^2 + 200 \cdot 1,6^2}\right) = 10^{-8}$$

   egyenlet megoldása. Ez

   $$t = \sqrt{\frac{8 \ln 10(400 \cdot 1,1^2 + 200 \cdot 1,6^2)}{2}}.$$

   (b) Itt a centrális határeloszlás-tételt kell alkalmazni. Az X_i-k összege és az Y_i-k összege is külön-külön normális eloszlással közelíthető, és a kiindulási változók függetlensége miatt ez a két normális eloszlású változó, amikkel közelítünk, szintén független egymástól. Független normálisok összege pedig normális, ami pontosan 1/2 súlyt ad a várható értéke (800 MW) feletti értékeknek, tehát 1/2 a kapacitás túllépésének valószínűsége is. (Ez speciális kérdés volt, a várható értéktől különböző kapacitás esetén a kérdezett valószínűséget nem lehet megmondani ennyi adatból, szükség van pl. a szórások ismeretére.)

   (c) Felső becslést tudunk adni. Legyen $S' = \sum_{k=1}^{360} X_k + \sum_{k=1}^{200} Y_k$. Ekkor $\mathbf{E}(S') = 760$.

   $$\mathbf{P}(S' > C) = \mathbf{P}(S' - \mathbf{E}(S') > 40) \leq \exp\left(-\frac{2 \cdot 40^2}{360 \cdot 1,1^2 + 200 \cdot 1,6^2}\right) \simeq 0,034.$$

   (d) Legyen $S^{(n)} = \sum_{k=1}^{n} X_k + \sum_{k=1}^{200} Y_k$, ahol $0 \leq n \leq 400$. Ekkor $\mathbf{E}(S^{(n)}) = 400 + n$. A

   $$\mathbf{P}(S^{(n)} > C) = \mathbf{P}(S^{(n)} - \mathbf{E}(S^{(n)}) > 400 - n) \leq \exp\left(-\frac{2 \cdot (400 - n)^2}{n \cdot 1,1^2 + 200 \cdot 1,6^2}\right)$$

becslés fennáll a Höffding-egyenlőtlenség miatt. Ha n-et úgy választjuk, hogy a fenti felső becslés jobb oldala 10^{-8} legyen, akkor a kapacitás túllépésének valószínűsége sem lehet ennél nagyobb. Vagyis

$$\exp\left(-\frac{2\cdot(400-n)^2}{n\cdot 1,1^2+200\cdot 1,6^2}\right)=10^{-8},$$

amiből az

$$n^2-(4\cdot 1,1^2\cdot\ln 10+2\cdot 400)n+400^2-4\cdot 200\cdot 1,6^2\cdot\ln 10=0$$

másodfokú egyenlet adódik. Ennek két megoldása n-re a $309,632$ és az $501,513$, melyek közül az utóbbi nyilván hamis gyök (hiszen akkor $S^{(n)}$ várható értéke is jóval 800 fölé esik, itt a Höffding-egyenlőtlenségben megjelenő t negatív, és valójában annak a valószínűsége kicsi, hogy az össztermelés a kapacitás alatt lesz). Tehát legfeljebb 309 erőművet kapcsolhatunk be biztosan.

8. Egy városban 40000 család él. Az egy család által egy nap alatt termelt szemét mennyisége semmiképpen nem több, mint 50 liter; a várható értéke 20 liter, szórása 10 liter.

   (a) Mekkora napi kapacitású szemétégető üzemet építsen az önkormányzat a háztartási szemétnek, ha azt szeretnék, hogy annak az esélye, hogy az üzem nem tudja feldolgozni az egy nap alatt termelődött szemetet, legfeljebb 1% legyen? Adjunk becslést a CHT alapján.

   (b) Miért nem alkalmazható a CHT, ha az önkormányzat 1% helyett 10^{-8}-os biztonságot szeretne? Ebben az esetben adjunk becslést a Hoeffding-korlát segítségével.

Megoldás: Legyen S az összes megtermelt szemét mennyisége. Ekkor $\mathbf{E}(S)=40000\cdot 20=800000$, $\mathbf{D}(S)=\sqrt{40000\cdot 10}$. A kérdéses kapacitást $\mathbf{E}(S)+t$ alakban adjuk meg.

(a)

$$\mathbf{P}(S>\mathbf{E}(S)+t)=\mathbf{P}\left(\frac{S-\mathbf{E}(S)}{\mathbf{D}(S)}>\frac{t}{\mathbf{D}(S)}\right)=0,01.$$

A CHT alapján $\frac{S-\mathbf{E}(S)}{\mathbf{D}(S)}$ standard normális eloszlással közelíthető, ezért fennáll a

$$\Phi\left(\frac{t}{\mathbf{D}(S)}\right)=1-0,01$$

egyenlet, amiből

$$t=\mathbf{D}(S)\Phi^{-1}(0,99)\simeq\sqrt{400000}\cdot 2,33\simeq 1473,6$$

(b) Höffding-korláttal az

$$\exp\left(-\frac{2t^2}{40000\cdot 50^2}\right)=10^{-8}$$

egyenletet kell megoldani, vagyis

$$t=\sqrt{\frac{8\ln 10\cdot 40000\cdot 50^2}{2}}\simeq 30348,5.$$

Megjegyzés: Az (a) részben a Höffding-egyenlőtlenségből

$$t=\sqrt{\frac{2\ln 10\cdot 40000\cdot 50^2}{2}}\simeq 15174,3$$

adódik, ami lényegesen pontatlanabb becslés, mint a CHT alapján kapott eredmény. A (b) részben pedig a nagyon kis valószínűségű ún. nagy eltérések becslésére a Höffding-korlát alkalmas.

9. A Sóder kft. az építőiparban tevékenykedik. Minden hónap elején 70% eséllyel kapnak munkát. A munka mindig ugyanaz, minden esetben egy hónapig tart, és a bevétel egy ilyen munkából 1 millió forint. Egy tétlenül töltött hónap bevétele 0.

   (a) Mennyi 30 hónap alatt a bevételük várható értéke?

(b) Becsüljük meg, mekkora annak az esélye, hogy 30 hónap alatt a bevételük kevesebb, mint 20 millió forint.

(c) Becsüljük meg, mekkora annak az esélye, hogy 30 hónap alatt a bevételük kevesebb, mint 15 millió forint.

10. Egy űrhajóban egy alkatrészre 200 különböző forrásból érkezik terhelés. A várható terhelés összesen 1000 Pa, a terhelés minden egyes forrásból legfeljebb 50 Pa. Mekkora legyen az alkatrész terhelhetősége, ha azt szeretnénk, hogy a túlterhelés esélye legfeljebb 10^{-10} legyen?