

A Valószínűségszámítás I. előadássorozat negyedik előadása.

2003. február 25.

Összefoglaló:

Rátértünk a Borel–Cantelli lemma bizonyítására. A bizonyításhoz szükségünk van a következő lemmára, amelynek állítását tartalmazza a 2. előadásban szereplő, ott a 6.–7. oldalon megfogalmazott Tétel A.

Lemma a valószínűségi mérték folytonosságáról. *Legyen adva $B_1 \subset B_2 \subset \dots$, $B_n \in \mathcal{A}$, $n = 1, 2, \dots$, eseményeknek növekvő sorozata egy (Ω, \mathcal{A}, P) valószínűségi mezőn. Legyen $B = \bigcup_{n=1}^{\infty} B_n$. Ekkor létezik a $\lim_{n \rightarrow \infty} P(B_n)$ határérték, és $\lim_{n \rightarrow \infty} P(B_n) = P(B)$.*

Legyen adva $C_1 \supset C_2 \supset \dots$, $C_n \in \mathcal{A}$, $n = 1, 2, \dots$, eseményeknek csökkenő sorozata egy (Ω, \mathcal{A}, P) valószínűségi mezőn. Legyen $C = \bigcap_{n=1}^{\infty} C_n$. Ekkor létezik a $\lim_{n \rightarrow \infty} P(C_n)$ határérték, és $\lim_{n \rightarrow \infty} P(C_n) = P(C)$.

A Borel–Cantelli lemma bizonyítása. Az a.) rész bizonyítása:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k) \quad \text{minden } n \text{ számra.}$$

Mivel $\sum_{k=1}^{\infty} P(A_k) < \infty$, minden $\varepsilon > 0$ számhoz létezik olyan $n = n(\varepsilon)$ szám, hogy $\sum_{k=n}^{\infty} P(A_k) < \varepsilon$, ahonnan az előző reláció szerint $P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \leq \varepsilon$. Innen következik az a.) rész állítása.

A b.) rész bizonyítása: Az előbbi lemma szerint a valószínűségi mérték folytonosságáról kapjuk, hogy az A_n halmazok függetlensége miatt

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) &= \lim_{N \rightarrow \infty} P\left(\bigcup_{k=N}^{\infty} A_k\right) = \lim_{N \rightarrow \infty} \left[\lim_{M \rightarrow \infty} P\left(\bigcup_{k=N}^M A_k\right) \right] \\ &= \lim_{N \rightarrow \infty} \left[\lim_{M \rightarrow \infty} \left(1 - P\left(\bigcap_{k=N}^M (\Omega \setminus A_k)\right) \right) \right] \\ &= \lim_{N \rightarrow \infty} \left[\lim_{M \rightarrow \infty} \left(1 - \prod_{k=N}^M (1 - P(A_k)) \right) \right]. \end{aligned}$$

Ezért elég belátni, hogy minden rögzített N számra $\lim_{M \rightarrow \infty} \prod_{k=N}^M (1 - P(A_k)) = 0$, ha

$\sum_{n=1}^{\infty} P(A_n) = \infty$. Viszont az $1 - x \leq e^{-x}$ egyenlőtlenség alapján $\prod_{k=N}^M (1 - P(A_k)) \leq \exp\left\{-\sum_{k=N}^M P(A_k)\right\}$, ahonnan $\lim_{M \rightarrow \infty} \prod_{k=N}^M (1 - P(A_k)) = 0$, ha $\sum_{n=1}^{\infty} P(A_n) = \infty$.

A felhasznált $1 - x \leq e^{-x}$ egyenlőtlenség például a következő módon látható: Az $f(x) = e^x$ függvény konvex, $f(0) = 1$, $f'(0) = 1$. Az $f(x) = e^x$ függvény konvexitása miatt a függvény $(0, f(0))$ ponton átmenő érintője minden valós x -re a az $f(x)$ függvény alatt van, azaz $e^x \geq x + 1$. Az x szám helyett $-x$ -et írva megkapjuk a kívánt állítást.

Megjegyzés: Az analízisben bizonyítják és használják a következő eredményt: Adva x_n valós számoknak olyan sorozata, amelyre $0 \leq x_n < 1$ minden $n = 1, 2, \dots$, számra $\prod_{k=1}^{\infty} (1 - x_k) > 0$, ha $\sum_{k=1}^{\infty} x_k < \infty$, és $\prod_{k=1}^{\infty} (1 - x_k) = 0$, ha $\sum_{k=1}^{\infty} x_k = \infty$. Bár nekünk erre nem lesz szükségünk, az előadás kiegészítésében megadom ennek a formulának előbb heurisztikus indoklását majd precíz bizonyítását, mivel ez tanulságos.

Korábbi feladatokban láttuk, hogy $A_1 \cup \dots \cup A_n$ alakú események valószínűségét egyszerűen egyszerűen ki lehet számítani a

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= 1 - P((\Omega \setminus A_1) \cap \dots \cap (\Omega \setminus A_n)) \\ &= 1 - \prod_{j=1}^n P((\Omega \setminus A_j)) = 1 - \prod_{j=1}^n (1 - P(A_j)) \end{aligned}$$

azonosság segítségével, ha az A_j , $1 \leq j \leq n$, események függetlenek. A fenti számolásban kihasználtuk az A_j események függetlenségét. A következőben egy példán keresztül megmutatjuk, hogy bizonyos érdekes esetekben, amikor olyan A_j események uniójának a valószínűségét akarjuk kiszámítani, melyek nem feltétlenül függetlenek, de ki tudjuk számolni az $A_{j_1} \cap \dots \cap A_{j_s}$ alakú események valószínűségét, akkor is meg tudjuk oldani feladatunkat egy szita formulának nevezett azonosság segítségével. A következő feladatot fogjuk tárgyalni:

Egy estélyen megjelenik n házaspár. Egy táncmester, aki nem tudja, hogy kik házasársak és kik nem véletlen módon párba rendezi a férfiakat és nőket a tánc előtt. Mi a valószínűsége annak, hogy egyetlen házaspár sem táncol együtt? Mi ennek a valószínűségnek a határértéke, ha $n \rightarrow \infty$?

Megoldás: Definiáljuk a következő A_j eseményeket:

$$A_j = \text{a } j\text{-ik házaspár együtt táncol, } 1 \leq j \leq n.$$

Ekkor minket a $P(\overline{A_1 \cup \dots \cup A_n}) = 1 - P(A_1 \cup \dots \cup A_n)$ valószínűség érdekel. Vegyük észre, hogy a $P(A_{j_1} \cap \dots \cap A_{j_k}) = \frac{(n-k)!}{n!}$ azonosság igaz. Ugyanis az összes lehetséges párbaállítások száma $n!$, míg az olyan párbaállítások száma, amelyben a j_1 -ik, j_2 -ik, \dots , j_k -ik házaspár egy párba kerül $(n-k)!$. Továbbá érvényes a következő, az irodalomban szita-formulának nevezett eredmény.

Szita formula. Legyenek A_1, \dots, A_n tetszőleges események egy (Ω, \mathcal{A}, P) valószínűségi mezőn. Ekkor

$$P(A_1 \cup \dots \cup A_n) = S_1 - S_2 + S_3 - \dots + (-1)^{n+1} S_n,$$

ahol

$$S_k = \sum_{1 \leq j_1 < \dots < j_k \leq n} P(A_{j_1} \cap \dots \cap A_{j_k}).$$

A szita-formula segítségével kapjuk, hogy n házaspár esetében

$$S_k = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Innen adódik, hogy a minket érdeklő valószínűség n házaspár esetén

$$\begin{aligned} P(\overline{A_1 \cup \dots \cup A_n}) &= 1 - P(A_1 \cup \dots \cup A_n) \\ &= 1 - S_1 + S_2 - \dots + (-1)^n S_n = \sum_{k=0}^n \frac{(-1)^k}{k!}, \end{aligned}$$

és ezért

$$P(\overline{A_1 \cup \dots \cup A_n}) = \sum_{k=0}^n \frac{(-1)^k}{k!} \rightarrow \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{1}{e} \quad \text{ha } n \rightarrow \infty.$$

A szita formula bizonyítását a kiegészítésben fogom tárgyalni. Viszont megfogalmazok egy érdekes ezzel a formulával kapcsolatos eredményt nem kötelező házi feladat formájában. E feladat megoldását megkaphatjuk a szita formula bizonyításának alkalmas adaptációjával.

Nem kötelező házi feladat:

Lássuk be, hogy a szitaformula jelöléseit használva

$$S_1 - S_2 = \sum_{j=1}^n P(A_j) - \sum_{1 \leq j < k \leq n} P(A_j \cap A_k) \leq P(A_1 \cup \dots \cup A_n) \leq S_1 = \sum_{j=1}^n P(A_j)$$

és általában

$$\sum_{k=1}^{2l} (-1)^{k+1} S_k \leq P(A_1 \cup \dots \cup A_n) \leq \sum_{k=1}^{2l-1} (-1)^{k+1} S_k$$

minden l indexre. (Legyen $S_k = 0$, ha $k > n$.) Ez az eredmény azt jelenti, hogy a $P(A_1 \cup \dots \cup A_n)$ valószínűséget a szita formulában kifejező $S_1 - S_2 + S_3 - \dots$ előjeles összeg páratlan számú tagot tartalmazó részletösszegei felülről és páros számú tagot tartalmazó részletösszegei alulról becsülik meg a vizsgált valószínűséget.

Diszkrét eloszlású valószínűségi változók és azok függetlensége

Először bevezettük a valószínűségi változó fogalmát:

Valószínűségi változó fogalma: Legyen adva egy (Ω, \mathcal{A}, P) valószínűségi mező. Egy ezen a téren értelmezett valós értékű (mérhető) $\xi(\omega)$ függvényt valószínűségi változónak fogunk nevezni. Azaz ξ az $\Omega \rightarrow R^1$ leképezés. Az a megkötés, hogy a függvény legyen mérhető azt jelenti, hogy minden x valós számra az $\{\omega: \xi(\omega) < x\} \in \mathcal{A}$ feltétel teljesül.

Időnként hasznos nemcsak valós értékű valószínűségi változóról beszélni, hanem kissé általánosabban olyan valószínűségi változókat definiálni, amelyek értékeiket egy általános téren veszik fel. (Például bizonyos esetekben érdemes olyan valószínűségi változókat tekinteni, amelyek értéke komplex szám, vagy nem egyetlen szám, hanem egy szám-n-es. Az ilyen valószínűségi változókat vektor értékűnek szokták hívni.) Annak érdekében, hogy ezt megtehesük tekintsünk valamilyen X halmazt, és e halmaz kitüntetett részhalmazainak \mathcal{X} osztályát, amelyek σ -algebrát alkotnak. Egy az (Ω, \mathcal{A}, P) valószínűségi mezőn definiált X térbeli értékeket felvevő (mérhető) függvényt X -tér értékű valószínűségi változónak nevezünk. Az, hogy ez a függvény mérhető azt jelenti, hogy minden $Y \in \mathcal{X}$ halmazra az $\{\omega: \xi(\omega) \in Y\} \in \mathcal{A}$ feltétel teljesül.

Megjegyzés: Az láttuk, hogy a valószínűségi változók (szép, azaz mérhető) függvényeket jelentenek. Az egyetlen különbség a korábban használt függvények és a most bevezetett valószínűségi változók között abban áll, hogy most egy általánosabb téren, egy (Ω, \mathcal{A}, P) valószínűségi mező Ω halmazán és nem feltétlenül a számegyenesen értelmezett függvényeket tekintünk. Ez egyben azt is jelenti, hogy valószínűségi változókkal ugyanúgy számolhatunk, mint ahogy azt eddig függvényekkel tettük.

Felmerülhet a kérdés, hogy miért vezettük be azt a mesterkéltnek tűnő feltételt, hogy a valószínűségi változó legyen mérhető függvény. Azért, mert természetes kívánság az, hogy legyen értelme annak valószínűségéről beszélni, hogy az $\{\omega: \xi(\omega) < x\}$ esemény bekövetkezett. Viszont csak \mathcal{A} -beli halmazoknak a valószínűségéről tudunk beszélni. Természetes ennél is többet kívánni, nevezetesen azt, hogy a számegyenes minden „szép” B halmazára az $\{\omega: \xi(\omega) \in B\}$ halmaz legyen a \mathcal{A} σ -algebrában, azaz legyen értelme ennek az eseménynek a valószínűségéről is beszélni. A mértékelméletben belátták, hogy amennyiben a számegyenes Borel mérhető részhalmazait tekintjük szép halmazoknak, akkor a mérhetőség általunk megkövetelt feltételeiből következik ez az erősebb tulajdonság is.

A fenti érvelés azt mutatja, hogy a mérhetőség megkövetelése természetes. Másrészt további eredmények azt is mutatják, hogy minden „definiálható” függvény egyben mérhető is, ezért ennek a tulajdonságnak a megkövetelése nem jelent igazi megszorítást, a továbbiakban megfeledezhetünk róla. Az általános terekben értelmezett mérhetőségről hasonlókat mondhatunk. Érdemes még megjegyezni, hogy bár formálisan a valós és általános térbeli függvény mérhetőségét másképp definiáltuk ez a különbség csak lát-szólagos. Vezessük be a számegyenesen a Borel mérhető halmazok \mathcal{B} σ -algebráját, és társítjuk a számegyenest ezzel a σ -algebrával. Ekkor a valós értékű mérhető függvényeknek az általános esetbeli definíciója az $(X, \mathcal{X}) = (R, \mathcal{B})$ választással megegyezik a valós értékű mérhető függvények eredeti definíciójával.

Diszkrét eloszlású valószínűségi változók és azok eloszlása. Egy ξ valószínűségi változót egy (Ω, \mathcal{A}, P) valószínűségi mezőn diszkrétnek vagy diszkrét eloszlásúnak hívunk, ha megadható egy véges vagy megszámlálhatóan végtelen számosságú x_1, x_2, \dots , halmaz úgy, hogy az $\{\omega, \xi(\omega) = x_n\}$, $n = 1, 2, \dots$, halmazok teljes eseményrendszert alkotnak. Egy diszkrét eloszlású valószínűségi változó eloszlását meghatározzák azok az x_1, x_2, \dots , értékek amelyeket az felvesz és a $p_n = P(\xi = x_n)$ valószínűségek. (Jegyezzük meg, hogy $\sum_n P(\xi = x_n) = 1$.)

Legyenek ξ_1, \dots, ξ_k diszkrét eloszlású valószínűségi változók ugyanazon az (Ω, \mathcal{A}, P) valószínűségi mezőn, amelyek valamilyen x_1, x_2, \dots értékeket vesznek fel. Ezek együttes eloszlását meghatározzák a $P(\xi_1 = x_{j_1}, \xi_2 = x_{j_2}, \dots, \xi_k = x_{j_k})$, $j_s = 1, 2, \dots, 1 \leq s \leq k$, valószínűségek.

Diszkrét eloszlású valószínűségi változók függetlensége. Legyenek ξ_1, \dots, ξ_k diszkrét eloszlású valószínűségi változók ugyanazon az (Ω, \mathcal{A}, P) valószínűségi mezőn, amelyek valamilyen x_1, x_2, \dots értékeket vesznek fel. Azt mondjuk, hogy ezek az $\xi_1, \xi_2, \dots, \xi_k$ valószínűségi változók függetlenek, ha

$$P(\xi_1 = x_{j_1}, \xi_2 = x_{j_2}, \dots, \xi_k = x_{j_k}) = P(\xi_1 = x_{j_1}) P(\xi_2 = x_{j_2}) \cdots P(\xi_k = x_{j_k})$$

minden $j_s = 1, 2, \dots, 1 \leq s \leq k$, indexre.

1. *Megjegyzés:* A fenti definícióban feltettük, hogy a diszkrét eloszlású valószínűségi változók amelyeknek az együttes eloszlását vagy függetlenségét definiáltuk ugyanazokat az értékeket veszik fel. Ez azonban nem jelent megszorítást. Ugyanis, ha egyesítjük ezen valószínűségi változók értékészletét, akkor újra legfeljebb megszámlálható számosságú halmazt kapunk, és mondhatjuk, hogy az egyes valószínűségi változók ezen halmazból veszik fel az értékeiket. Lehetséges, hogy ily módon sok nulla valószínűségi eseményt is bevezettünk, de ez nem okoz problémát.

2. *Megjegyzés:* Annak, hogy valószínűségi változóknak az eloszlását definiáltuk mélyebb oka van. A valószínűségszámításban olyan problémákkal foglalkozunk, amelyek csak attól függenek, hogy különböző eseményeknek mennyi a valószínűségük, különböző valószínűségi változók milyen értéket milyen valószínűséggel vesznek fel. Arra vagyunk kíváncsiak, hogy ezen ismereteknek milyen következményei vannak. Viszont a valószínűségszámításban vizsgált kérdésekre adott válasz nem függ attól, hogy milyen véletlen hatások váltották ki a megfigyelt véletlen eredményeket, azaz milyen az a valószínűségi mező, amelyben a vizsgált események és valószínűségi változók megjelentek. Ez az oka annak, hogy minket igazából nem maguk a valószínűségi változók, hanem azok (együttes) eloszlása érdekel.

Peladat:

Legyenek adva ξ_1, \dots, ξ_k független, diszkrét eloszlású valószínűségi változók egy (Ω, \mathcal{A}, P) valószínűségi mezőn, melyek valamilyen x_1, x_2, \dots értékeket vesznek fel. Legyenek $A_1 \subset \{x_1, x_2, \dots\}$, $A_2 \subset \{x_1, x_2, \dots\}$, \dots , $A_k \subset \{x_1, x_2, \dots\}$ tetszőleges halmazok. Ekkor

$$P(\xi_1 \in A_1, \xi_2 \in A_2, \dots, \xi_k \in A_k) = P(\xi_1 \in A_1) P(\xi_2 \in A_2) \cdots P(\xi_k \in A_k).$$

Speciálisan, ekkor tetszőleges $\{j_1, \dots, j_s\} \subset \{1, \dots, k\}$ indexhalmazra a $\xi_{j_1}, \dots, \xi_{j_s}$ valószínűségi változók függetlenek.

Halmaz indikátorfüggvényének a definíciója. Legyen adva egy $A \in \mathcal{A}$ esemény egy (Ω, \mathcal{A}, P) valószínűségi mezőn. Az A halmaz indikátorfüggvényén azt a $\chi_A(\omega)$ valószínűségi változót értjük, amelyre $\chi_A(\omega) = 1$, ha $\omega \in A$, és $\chi_A(\omega) = 0$, ha $\omega \notin A$.

Ezután megfogalmaztuk a következő gyakorlaton tárgyalandó feladatokat:

Legyenek A_1, \dots, A_k események egy (Ω, \mathcal{A}, P) valószínűségi mezőn. Az A_1, \dots, A_k események akkor és csak akkor függetlenek, ha azok $\chi_{A_1}(\omega), \dots, \chi_{A_k}(\omega)$ indikátorfüggvényei függetlenek.

Véletlenül meghívunk 30 embert. Tegyük fel, hogy az egyes embereknek egymástól függetlenül van születésnapjuk, és minden ember esetében $\frac{1}{365}$ annak a valószínűsége, hogy az év valamely napján született. Mi annak a valószínűsége, hogy van két ember a társaságban, akiknek ugyanaznap van a születésnapjuk?

Általánosabban, van n urna, amelyekbe bedobunk egymástól függetlenül k golyót úgy, hogy mindegyik golyó egyforma valószínűséggel esik az egyes urnákba. Mi annak a valószínűsége, hogy van olyan urna amelybe legalább két golyó esik? Érdekel minket továbbá ennek a valószínűségnek a viselkedése, ha mind az n mind a k szám nagy, és a $k = k(n)$ számnak megfelelő a nagyságrendje. Lássuk be, hogy a fenti valószínűségnek van határértéke, ha $n \rightarrow \infty$, $\frac{k}{\sqrt{n}} \rightarrow \alpha$ valamilyen $0 \leq \alpha < \infty$ számmal, és határozzuk meg ezt a határértéket.

Megoldás: Jelölje ξ_j azt a valószínűségi változót, hogy a j -ik embernek az év hanyadik napján van a születésnapja. Ekkor a ξ_j , $1 \leq j \leq 30$, valószínűségi változók függetlenek, $P(\xi_j = l) = \frac{1}{365}$, $1 \leq j \leq 30$, $1 \leq l \leq 365$, és $P(\xi_j \neq \xi_{j'} \text{ ha } j \neq j')$ annak a valószínűsége, hogy mindenkinek különböző nap van a születésnapja. Ez a valószínűség viszont

$$\frac{365 \cdot 364 \cdots (365 - 30 + 1)}{365^k} = \prod_{j=1}^{29} \left(1 - \frac{j}{365}\right),$$

mert annak valószínűsége, hogy az első ember születésnapja az l_1 -ik, a másodiké az l_2 -ik és így tovább a k -ik ember születésnapja az l_k -ik napon van $\frac{1}{365^k}$, tetszőleges

$1 \leq l_j \leq 365$, $1 \leq j \leq 30$ számok esetén, és ezeket a számokat $\prod_{j=0}^{k-1} (365 - j)$

módon választhatjuk úgy, hogy mindegyik l_j szám különböző legyen. Így annak a valószínűsége, hogy van két ember akinek ugyanazon a napon van a születésnapja

$$1 - \prod_{j=1}^{29} \left(1 - \frac{j}{365}\right).$$

Hasonlóan, annak valószínűsége, hogy ha k golyót dobunk n urnába az adott módon, akkor van olyan urna, amelyikbe legalább két golyó esik $1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right)$.

Adjunk jó közelítést a $\log \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) = \sum_{j=1}^{k-1} \log \left(1 - \frac{j}{n}\right)$ kifejezésre, ha $n \rightarrow \infty$,

$\frac{k(n)}{\sqrt{n}} \rightarrow \alpha$. Heurisztikus érvelés szerint mivel $\log \left(1 - \frac{j}{n}\right) \sim -\frac{j}{n}$ a $\log(1+x)$

függvény Taylor sorfejtése szerint, ezért $\sum_{j=1}^{k-1} \log \left(1 - \frac{j}{n}\right) \sim -\sum_{j=1}^{k-1} \frac{j}{n} = -\frac{(k-1)k}{2n}$,

ahonnan $\log \prod_{j=1}^{k(n)-1} \left(1 - \frac{j}{n}\right) \rightarrow -\frac{\alpha^2}{2}$, ha $n \rightarrow \infty$, és $\frac{k(n)}{\sqrt{n}} \rightarrow \alpha$. Ez a számolás precízre tehető, ha felhasználjuk például azt az egyenlőtlenséget, amely szerint

$$\left| \log \left(1 - \frac{j}{n}\right) + \frac{j}{n} \right| \leq \frac{2j^2}{n^2} \leq \frac{\text{const.}}{n}, \quad \text{ha } n \text{ elég nagy és } \frac{j}{\sqrt{n}} \leq \alpha + 1,$$

ami szintén következik a $\log(1+x)$ Taylor sorfejtéséből. Miért? Innen kapjuk, hogy $1 - \prod_{j=1}^{k(n)-1} \left(1 - \frac{j}{n}\right) \rightarrow 1 - e^{-\alpha^2/2}$, ha $n \rightarrow \infty$, és $\frac{k(n)}{\sqrt{n}} \rightarrow \alpha$.

A következőben a valószínűségszámítás egyik legfontosabb fogalmát, a várható értéket definiáltuk, és tárgyaltuk annak legfontosabb tulajdonságait. Egyelőre csak diszkrét eloszlású valószínűségi változók várható értékével foglalkozunk.

Diszkrét eloszlású valószínűségi változó várható értéke. Legyen ξ diszkrét eloszlású valószínűségi változó, amely x_1, x_2, \dots értékeket vesz fel $p_k = P(\xi = x_k)$ valószínűséggel, $k = 1, 2, \dots$, $\sum_{k=1}^{\infty} P(\xi = x_k) = 1$. A ξ valószínűségi változó várható értéke az

$$E\xi = \sum_{k=1}^{\infty} x_k P(\xi = x_k)$$

összeg, feltéve hogy ez az összeg abszolút konvergens. Ha ez az összeg nem abszolút konvergens, akkor nem definiáljuk az $E\xi$ várható értéket.

1. *Megjegyzés:* Most és a továbbiakban az egyszerűbb és egységesebb jelölés érdekében feltesszük, hogy a diszkrét eloszlású ξ valószínűségi változó (megszámlálhatóan) végtelen sok értéket vesz fel. Ez nem jelent megszorítást, mert ha ξ csak véges sok értéket vesz fel, akkor is feltehetjük, hogy ezenkívül még felvesz végtelen sok értéket nulla valószínűséggel.

2. *Megjegyzés:* A várható értéket az angol *expectation* szó kezdőbetűjével az E betűvel jelöljük. Manapság ez a legelterjedtebb jelölés, bár használják az M betűt is a német *mathematischer Mittelwert* szó alapján.

Emlékeztető: Azt mondjuk, hogy a $\sum_{n=1}^{\infty} y_n$ sor abszolút konvergens, ha nemcsak a $\sum_{n=1}^{\infty} y_n$, hanem a $\sum_{n=1}^{\infty} |y_n|$ sor is konvergens, azaz $\sum_{n=1}^{\infty} |y_n| < \infty$. Ha csak a $\sum_{n=1}^{\infty} y_n$ sor konvergens, és a $\sum_{n=1}^{\infty} |y_n|$ sor divergens, akkor feltételesen konvergens sorról beszélünk.

Idézzünk fel néhány eredményt, amelyek megmagyarázzák miért követeltük meg a várható érték definíciójában, hogy a $\sum_{k=1}^{\infty} x_k P(\xi = x_k)$ sor ne csak konvergens, hanem abszolút konvergens is legyen.

Ha a $\sum_{n=1}^{\infty} y_n$ sor abszolút konvergens, akkor ennek tetszőleges átrendezése azaz tetszőleges olyan $\sum_{n=1}^{\infty} y_{k(n)}$ összeg, amelyre a $k(n) = j$ egyenletnek pontosan egy megoldása van minden $1 \leq j < \infty$ egész számra szintén konvergens, és ugyanaz a határértéke mint az eredeti sorrendben. Ez a tulajdonság nem abszolút konvergens sorozatokra már nem érvényes.

Miért fontos ez a tulajdonság? Vegyünk például azt az esetet, ha a ξ valószínűségi változó az összes racionális számot veszi fel. Akkor ezek a racionális számok mivel megszámlálhatóan sokan vannak felsorolhatóak mint egy x_1, x_2, \dots , sorozat, de nincs a felsorolásban egy természetes sorrend. Ezért a várható értéket definiáló összeg csak akkor értelmes, ha az nem függ attól, hogy a ξ valószínűségi változó lehetséges x_k értékeit milyen sorrendben soroljuk fel.

Az abszolút konvergens sorok fenn említett tulajdonsága szerepelt az analízis tananyagban. A bizonyítás megismétlése helyett lássunk inkább egy példát, amely megmutatja, hogy a nem abszolút konvergens sorok másképpen viselkednek. Tekintsük az $S = 1 - 1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{3} - \frac{1}{3} + \frac{1}{4} - \frac{1}{4} + \dots$ sort. Ez a sor konvergens, határértéke zéró, ugyanakkor nem abszolút konvergens, mert $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. Tekintsük e sor következő átrendezését: $S = 1 + \sum_{k=1}^{\infty} \left(\frac{1}{2^k} + \frac{1}{2^k + 1} + \dots + \frac{1}{2^{k+1} - 1} - \frac{1}{k} \right)$. Ez az átrendezett sor divergens, mert a k -ik tagra $\frac{1}{2^k} + \frac{1}{2^k + 1} + \dots + \frac{1}{2^{k+1} - 1} - \frac{1}{k} \geq \frac{2^k}{2^{k+1}} - \frac{1}{k} = \frac{1}{2} - \frac{1}{k}$. Röviden, kissé nagyvonalúan, azt mondhatjuk, hogy abszolút konvergens sorozatokkal ugyanolyan szabadon számolhatunk (átrendezhetjük őket, abszolút konvergens sorok szorzataiban elvégezhetjük a tagonkénti szorzást) mint véges összegek esetében, míg nem abszolút konvergens sorok esetében ezt nem tehetjük meg.

Ezek után tekintsük a várható érték kiszámítását egy viszonylag egyszerű esetben.

Egy szabályos pénzdarabot feldobunk 100-szor egymás után. Számoljuk ki a fejdobások számának a várható értékét.

Megoldás: Annak valószínűsége, hogy pontosan k fejdobás történik 100 dobás ese-

tében $\binom{100}{k}2^{-100}$, $0 \leq k \leq 100$. Ezért $E\xi = \sum_{k=0}^{100} k \binom{100}{k} 2^{-100}$ a keresett várható érték. Ezt az összeget ki tudjuk számítani a következő észrevétel segítségével: $k \binom{n}{k} = n \binom{n-1}{k-1}$, ha $1 \leq k \leq n$. Innen $n = 100$ választással

$$\begin{aligned} E\xi &= \sum_{k=0}^{100} k \binom{100}{k} 2^{-100} = 100 \sum_{k=0}^{99} \binom{99}{k} 2^{-100} = 50 \sum_{k=0}^{99} \binom{99}{k} \left(\frac{1}{2}\right)^{99} \\ &= 50 \cdot \left(\frac{1}{2} + \frac{1}{2}\right)^{99} = 50. \end{aligned}$$

Heurisztikusan a következő módon érvelnénk: Egy pénzfeldobás esetén a fejek számának a várható értéke $\frac{1}{2}$, száz pénzfeldobás esetén a pénzdobások számának várható értéke $100 \cdot \frac{1}{2} = 50$.

A következő fontos eredmény, amelynek bizonyítása a következő előadás anyaga lesz megmagyarázza, hogy a fenti heurisztika miért adott helyes eredményt, és ennek segítségével hogyan lehet kiszámítani sok esetben egyszerűen valószínűségi változók várható értékét.

Tétel. *Legyenek ξ_1, ξ_2 (diszkrét eloszlású) valószínűségi változók, amelyeknek létezik várható értékük. Ekkor a $\xi_1 + \xi_2$ összegnek is létezik várható értéke, és*

$$E(\xi_1 + \xi_2) = E\xi_1 + E\xi_2.$$

Következmény. *Legyenek adva $\xi_1, \xi_2, \dots, \xi_k$ (diszkrét eloszlású) valószínűségi változók, amelyeknek létezik várható értékük, és legyenek c_1, \dots, c_k valós számok. Ekkor a $c_1\xi_1 + c_2\xi_2 + \dots + c_k\xi_k$ valószínűségi változónak is létezik várható értéke, és*

$$E(c_1\xi_1 + c_2\xi_2 + \dots + c_k\xi_k) = c_1E\xi_1 + c_2E\xi_2 + \dots + c_kE\xi_k.$$

Megjegyzés: Érdeemes hangsúlyozni, hogy a fenti eredményben nem tételeztük fel, hogy a tekintett valószínűségi változók függetlenek. Ez az észrevétel bizonyos alkalmazásokban hasznos. Később tanulni fogjuk, hogy a most megfogalmazott eredmény nemcsak diszkrét eloszlású valószínűségi változók összegeire érvényes.

Lássuk, hogy hogyan tudjuk kiszámítani az előző feladatban tekintett várható értéket a fenti eredmény segítségével.

Vezessük be a következő valószínűségi változókat: $\xi_j = 1$, ha a j -ik dobás eredménye fej, $\xi_j = 0$, ha a j -ik dobás eredménye írás. Azaz a ξ_j valószínűségi változó azt számolja, hogy mennyi a j -ik dobás eredményének hozadéka a fej-dobások számához.

Ekkor minket az $E\left(\sum_{j=1}^{100} \xi_j\right)$ várható érték érdekel. Viszont a fenti eredmény alapján

$E\left(\sum_{j=1}^{100} \xi_j\right) = \sum_{j=1}^{100} E\xi_j$, és $E\xi_j = 1 \cdot P(\xi_j = 1) + 0 \cdot P(\xi_j = 0) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}$ minden

$1 \leq j \leq 100$ indexre, ahonnan $E\left(\sum_{j=1}^{100} \xi_j\right) = 100 \cdot \frac{1}{2} = 50$.

Feladat:

Egy szabályos dobókockát feldobunk 100-szor egymás után. Számoljuk ki a dobás-eredmények összegének a várható értékét.

Kiegészítés:

Az előadásban megemlítettük a következő eredményt:

Állítás: Legyen x_n valós számoknak olyan sorozata, amelyre $0 \leq x_n < 1$ minden $n = 1, 2, \dots$, számra. Ekkor $\prod_{k=1}^{\infty} (1 - x_k) > 0$, ha $\sum_{k=1}^{\infty} x_k < \infty$, és $\prod_{k=1}^{\infty} (1 - x_k) = 0$, ha $\sum_{k=1}^{\infty} x_k = \infty$.

Indoklás: Heusztikusan a következő módon érvelhetünk: Logaritmust véve kapjuk, hogy $\prod_{k=1}^{\infty} (1 - x_k) > 0$ akkor és csak akkor, ha $\sum_{k=1}^{\infty} -\log(1 - x_k) < \infty$. Viszont azt, hogy az utóbbi összeg konvergens-e vagy divergens az dönti el, hogy kis x_k számokra a $-\log(1 - x_k)$ összeadandók milyen kicsik. Mivel $-\log(1 - x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots$, a $-\log(1 - x_k)$ mennyiségnek természetes jó közelítése az x_k kifejezés, (a Taylor sor első tagja), és ez azt sugallja, hogy a $\sum_{k=1}^{\infty} -\log(1 - x_k)$ illetve $\sum_{k=1}^{\infty} x_k$ végtelen sorok egyszerre konvergenssek vagy divergenssek.

Némi munkával a fenti érvelés precízzé tehető. Először azt kell meggondolni, hogy a feladat állításában a $\prod_{k=1}^{\infty} (1 - x_k) > 0$ tulajdonság a $\sum_{k=1}^{\infty} -\log(1 - x_k) < \infty$ egyenlőtlenséggel, a $\prod_{k=1}^{\infty} (1 - x_k) = 0$ tulajdonság pedig a $\sum_{k=1}^{\infty} -\log(1 - x_k) = \infty$ relációval ekvivalens. Ehhez azt kell felhasználni, hogy esetünkben $0 < 1 - x_k < 1$, és $-\log(1 - x_k) > 0$. Ezután vegyük észre, hogy a $\sum_{k=1}^{\infty} -\log(1 - x_k)$ és $\sum_{k=1}^{\infty} x_k$ sorok mindegyike divergens, ha rögzítve egy kis pozitív számot például az $\frac{1}{10}$ számot az x_k számsorozatnak végtelen sok olyan tagja van, amelyek nagyobbak mint ez az $\frac{1}{10}$ szám. Továbbá, egy végtelen összeg konvergenciáját vagy divergenciáját nem befolyásolja véges sok tagjának az értéke. Ezért az összegekben szereplő $x_k \geq \frac{1}{10}$ tagokat elhagyva elég csak azzal az esettel foglalkozni, amikor $x_k \leq \frac{1}{10}$ minden k -ra. Viszont ekkor $\frac{1}{2}x_k < -\log(1 - x_k) < 2x_k$ minden $k = 1, 2, \dots$ számra, ahonnan $\frac{1}{2} \sum_{k=1}^n x_k < \sum_{k=1}^n -\log(1 - x_k) < 2 \sum_{k=1}^n x_k$

minden n -re, és innen következik, hogy a $\sum_{k=1}^{\infty} x_k$ és $\sum_{k=1}^{\infty} -\log(1-x_k)$ sorok egyszerre konvergálnak vagy divergálnak.

A szita formula bizonyítása: Adva egy A esemény vezessük be az A^ε jelölést, ahol $\varepsilon = \pm 1$, $\varepsilon = 1$ estében $A^1 = A$, $\varepsilon = -1$ esetében $A^{-1} = \Omega \setminus A = \bar{A}$. Definiáljuk tetszőleges $r_s = 1$ vagy $r_s = -1$, $s = 1, \dots, n$ számokra az

$$A(r_1, \dots, r_n) = A^{r_1} \cap \dots \cap A^{r_n}$$

eseményeket. A bizonyítás gondolata abból áll, hogy a bizonyítandó azonosságnak mind a két oldalát kifejezzük az $A(r_1, \dots, r_n) = A^{r_1} \cap \dots \cap A^{r_n}$ események $P(A(r_1, \dots, r_n))$ valószínűségeinek lineáris kombinációjaként, és ellenőrizzük, hogy a két kifejezés megegyezik. Vegyük észre, hogy

$$A_1 \cup \dots \cup A_n = \bigcup_{(r_1, \dots, r_n): (r_1, \dots, r_n) \neq (-1, \dots, -1)} A(r_1, \dots, r_n),$$

és tetszőleges $A_{j_1} \cap \dots \cap A_{j_s}$, $1 \leq j_1 < j_2 < \dots < j_s \leq n$ alakú halmazra

$$A_{j_1} \cap \dots \cap A_{j_s} = \bigcup_{(r_1, \dots, r_n): r_{j_u}=1, 1 \leq u \leq s} A(r_1, \dots, r_n).$$

Továbbá az $A(r_1, \dots, r_n)$ események különböző (r_1, \dots, r_n) paraméterek esetében diszjunktak. Ezért

$$P(A_1 \cup \dots \cup A_n) = \sum_{(r_1, \dots, r_n): (r_1, \dots, r_n) \neq (-1, \dots, -1)} P(A(r_1, \dots, r_n)),$$

és tetszőleges $A_{j_1} \cap \dots \cap A_{j_s}$ alakú halmazra

$$P(A_{j_1} \cap \dots \cap A_{j_s}) = \sum_{(r_1, \dots, r_n): r_{j_u}=1, 1 \leq u \leq s} P(A(r_1, \dots, r_n)).$$

Ilyen módon a az $A(r_1, \dots, r_n)$ halmazok $P(A(r_1, \dots, r_n))$ valószínűségeinek a lineáris kombinációjaként fejezhető ki a szita-formula két oldalán szereplő kifejezés. Azt kell belátni, hogy a $P(A(r_1, \dots, r_n))$ szám a szitaformula két oldalán szereplő kifejezésben ugyanazzal az együtthatóval szerepel. Az azonosság baloldalán ez az együttható 1, ha $\{r_1, \dots, r_n\} \neq \{-1, \dots, -1\}$, és nulla, ha $\{r_1, \dots, r_n\} = \{-1, \dots, -1\}$. Ha az (r_1, \dots, r_s) halmaz l darab 1 és $n-l$ darab -1 jegyet tartalmaz, akkor ennek valószínűsége a szitaformula jobboldalán ezt $\binom{l}{1} - \binom{l}{2} + \dots \pm \binom{l}{l} = \sum_{k=1}^l (-1)^{k+1} \binom{l}{k}$ együtthatóval szerepel. Ugyanis az S_k összegben ez a kifejezés $\binom{l}{k}$ együtthatóval szerepel. Miért?

Innen a vizsgált együttható $\sum_{k=1}^l (-1)^{k+1} \binom{l}{k} = 1 - \sum_{k=0}^l (-1)^k \binom{l}{k} = 1 - (1-1)^l = 1$, ha az $A(r_1, \dots, r_n)$ halmaz r_1, \dots, r_n indexei között $l \neq 0$ 1-es van. A maradék esetben pedig az együttható nulla. Ez azt jelenti, hogy a szita formula két oldalán szereplő együtthatók megegyeznek.

Ismertetem a szita formula egy más bizonyítását, mely egy érdekes, és más esetben is alkalmazható gondolaton alapul. Nevezetesen az alább megfogalmazott lemmán, amelyet szintén be fogunk bizonyítani.

Lemma. *Legyenek adva valamely A_1, \dots, A_n események egy (Ω, \mathcal{A}, P) valószínűségi mezőn, és definiáljuk ezek felhasználásával véges sok unió, metszet és komplementer segítségével az $B_j = f_j(A_1, \dots, A_n)$. Rögzítsünk valamely c_1, \dots, c_k valós számokat. A*

$$\sum_{j=1}^k c_j P(f_j(A_1, \dots, A_n)) \geq 0$$

egyenlőtlenség teljesül tetszőleges valószínűségi mezőn definiált tetszőleges A_1, \dots, A_n halmazokra, ha speciálisan teljesül abban a (2^n számú) speciális esetben, amikor mindegyik A_j vagy az Ω biztos vagy az \emptyset üres esemény.

A lemma bizonyítása: Jelölje $A_j^{-1} = \Omega \setminus A_j$ az A_j halmaz komplementerét. Vezessük be az $A_j^{\varepsilon_j} = A_j$ jelölést és jelölje $(\varepsilon_1, \dots, \varepsilon_n)$, $\varepsilon_l = \pm 1$, $1 \leq l \leq n$, valamely n hosszúságú ± 1 sorozatot. Ekkor felírva mindegyik A_j eseményt konjunktív normálforma alakban. Felhasználva, hogy a konjunktív normálformában diszjunkt halmazok uniója jelenik meg, a vizsgálandó egyenlőtlenség

$$\sum_{\substack{(\varepsilon_{j_1}, \dots, \varepsilon_{j_n}) \\ E_{j_s} = \pm 1, s=1, \dots, n}} d_{j_1, \dots, j_n} P(A_1^{\varepsilon_{j_1}} \cap \dots \cap A_n^{\varepsilon_{j_n}}) \geq 0$$

alakba írható alkalmas d_{j_1, \dots, j_n} együtthatókkal. Ez az egyenlőtlenség akkor és csak akkor érvényes minden A_1, \dots, A_n halmaz esetén, ha az összes d_{j_1, \dots, j_n} együtthatóra. Adva egy $(\varepsilon_{j_1}, \dots, \varepsilon_{j_n})$ sorozat, tekintsük azt az A_1, \dots, A_n sorozatot, melyre $A_l = \Omega$, ha $\varepsilon_{j_l} = 1$, és $A_l = \emptyset$, ha $\varepsilon_{j_l} = -1$, $1 \leq l \leq j$. Ezzel a választással a lemma feltétele azt adja, hogy $d_{j_1, \dots, j_n} \geq 0$. Mivel ezt minden lehetséges d_{j_1, \dots, j_n} sorozatra elmondhatjuk, innen következik a lemma állítása.

A szita formula bizonyítása a lemma segítségével: A lemma alapján elegendő belátni a szita formulát abban a speciális esetben, ha mindegyik A_j esemény a biztos vagy az üres esemény. Ekkor ugyanis a szita formula bal és jobboldalán szereplő kifejezések különbségéről tudjuk, hogy egyrészt nagyobb vagy egyenlő mint nulla, másrészt kisebb egyenlő, mint nulla. Tekintsük azokat az eseteket, amikor az A_j események között k darab biztos esemény van, és $n - k$ üres esemény, $0 \leq k \leq n$. Ekkor $S_l = \binom{k}{l}$, ha $1 \leq l \leq k$, és $S_l = 0$, ha $l \geq k$. Ezért $1 \leq k \leq n$ esetben a szita formula jobboldalán álló kifejezés $\sum_{l=1}^k (-1)^{l+1} S_l = \sum_{l=1}^k (-1)^{l+1} \binom{k}{l} = 1 - \sum_{l=0}^k (-1)^l \binom{k}{l} = 1 - (1 - 1)^k = 1$, a baloldali kifejezés pedig szintén $P(A_1 \cup \dots \cup A_n) = P(\Omega) = 1$. A $k = 0$ esetben a szita formula mindkét oldala nullával egyenlő.