

5. Valószínűesszámitási és statisztikai alapok

Alapfogalmak

Véletlen kísérlet: Egy olyan történés / jelenség, aminek az eredményeinek halmazát ismerjük előre, de a pontos kimenet ezek között véletlen alakul. Azonos feltételek mellett megismételhető.

A kockadobásunk esetén maga a dobás a kísérlet, az eredmények halmaza 1, 2, 3, 4, 5, 6.

A (0, 1)-beli véletlen szám választása esetén az eredmények halmaza a (0, 1) intervallum.

Esemény: Egy olyan állítás, aminek az igazságtartalma a kísérlet elvégzése után kiértékelhető. Az eseményeket nagybetűvel jelöljük. Két esemény kizáró, ha egyszerre nem következhet be mindkettő.

Egy A esemény *része* egy B eseménynek (A maga után vonja B-t), ha A teljesülése esetén B mindenképp teljesül. Jelölése: $A \subseteq B$. Két esemény ekvivalens, ha mindkettő része a másiknak. Jelölése: $A = B$. A biztos esemény jele: Ω , a lehetetlené \emptyset .

Kockadobás esetén az alábbiak mind események:

$A = \{4\text{-et dobunk}\}$, $B = \{4\text{-et vagy } 6\text{-ot dobunk}\}$, $C = \{\text{Összetett számot dobunk}\}$,

$D = \{\text{Páratlant dobunk}\}$, $E = \{100\text{-nál kisebbet dobunk}\}$, $F = \{3, 5\text{-öt dobunk}\}$.

A és D *kizáróak*, A *része* B-nek, B és C ekvivalensek, E a *biztos esemény*, F pedig a *lehetetlen*.

Teljes eseményrendszer: $A_1, A_2, \dots, A_n, \dots$ páronként kizáró események rendszere teljes eseményrendszer, ha a kísérlet bármely kimenetele esetén valamelyik A_i bekövetkezik (az alább definiált összeadást felhasználva átfogalmazható: az összegük Ω).

A kockadobás esetén például

$T1 = \{\{\text{párosat dobunk}\}, \{\text{páratlant dobunk}\}\}$, vagy

$T2 = \{\{1\text{-et dobunk}\}, \{1\text{-et vagy } 2\text{-et dobunk}\}, \{4\text{-et vagy } 6\text{-ot dobunk}\}\}$

is *teljes eseményrendszert* alkotnak.

Műveletek eseményekkel

Egy A esemény *ellentettjén* (vagy *komplementerén*) azt az eseményt értjük, ami pontosan akkor következik be, amikor A nem. Jele \bar{A} . (Pl: $\bar{\Omega} = \emptyset$).

Két esemény összege (vagy uniója) az az esemény, ami pontosan akkor következik be, ha a két esemény közül legalább egy teljesül. Jele $A + B$ vagy $A \cup B$.

Két esemény szorzatán (vagy metszetén) azt az eseményt értjük, ami pontosan akkor következik be, amikor mindkét esemény teljesül. Jele $A \cdot B$, AB vagy $A \cap B$.

Egy A és egy B esemény különbségén azt az eseményt értjük, amikor az A bekövetkezik, de a B nem. Jele $A - B$ vagy $A \setminus B$. Valamint $A - B = A \cdot \bar{B}$.

σ algebra

Legyen X egy halmaz. Egy $\mathcal{R} \subseteq 2^X$ halmaz pontosan akkor **σ algebra** X felett, ha teljesíti az alábbi három feltételt:

- $X \in \mathcal{R}$
- \mathcal{R} zárt a komplementerképzésre: $A \in \mathcal{R} \Rightarrow \bar{A} = X - A \in \mathcal{R}$
- \mathcal{R} zárt a megszámlálható összegre: $A_1, A_2, \dots, A_n, \dots \in \mathcal{R} \Rightarrow \sum_i A_i \in \mathcal{R}$

Egy kísérlet kimeneteleinek a halmazának nem feltétlen minden eleme érdekes számunkra. A vizsgálatainkhoz fontos a kísérleten definiálható események közül a számunkra megfigyelhetőket mindig megadni előre.

Elemi esemény: Azokat a nem lehetetlen, megfigyelhető eseményeket, amiknek nincs olyan nem lehetetlen, megfigyelhető részeseménye, ami nem ekvivalens vele, elemi eseményeknek nevezzük. Azaz $A \in \mathcal{R}$ pontosan akkor **elemi esemény**, ha:

$$A \neq \emptyset \text{ és } \forall B \neq \emptyset, B \in \mathcal{R}, B \subseteq A : A \subseteq B$$

Az elemi eseményeket ω -val jelöljük.

A kockadobásnál a megfelelő elemi események a következők:

$$\omega_1 = \{1\text{-et dobunk}\}, \omega_2 = \{2\text{-t dobunk}\}, \omega_3 = \{3\text{-at dobunk}\}, \omega_4 = \{4\text{-et dobunk}\}, \\ \omega_5 = \{5\text{-öt dobunk}\}, \omega_6 = \{6\text{-ot dobunk}\}.$$

Lehetne elemi esemény az, hogy "x lett a dobott érték, amit y oldalról átfordulva értünk el", de persze ez nem célszerű választás.

Eseménytér: Az elemi események halmaza. Mivel az összes elemi esemény összege Ω , ezért az eseményteret is szokásosan Ω -val jelöljük. Fontos megjegyezni, hogy abban az esetben, ha Ω nem megszámlálható, akkor az eseménytér önmagában nem határozza meg a megfigyelhető események halmazát.

A kockadobás esetén az előző példában látott elemi események mellett az eseménytér

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$$

Valószínűség

Definíció. Megfigyelhető események egy \mathcal{R} σ algebráján definiálhatjuk a **valószínűséget**, mint a következő tulajdonságokat teljesítő $P : \mathcal{R} \rightarrow [0, 1]$ valós függvényt:

- A valószínűség nem negatív: $0 \leq P(A)$
- A biztos esemény valószínűsége 1: $P(\Omega) = 1$
- $\forall A \in \mathcal{R}$ -ra, páronként kizáró ($A_i \cdot A_j = \emptyset, i \neq j$)
- Kizáró események megszámlálható összegének valószínűsége a valószínűségek összege:

$$A_1, A_2, \dots, A_n, \dots \in \mathcal{R}, \forall i \neq j : A_i A_j = \emptyset \Rightarrow P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

Definíció. Az eseménytér, a megfigyelhető események halmaza és a rajtuk definiált valószínűség együtt úgynevezett (**Kolmogorov-féle**) valószínűségi mezőt alkot: (Ω, \mathcal{R}, P) hármas.

A kockadobás esetén, ha szabályos kockáról beszélünk (és minden értékű kimenetel megfigyelhető), akkor vehetjük $\frac{1}{6}$ -nak minden lehetséges érték valószínűségét, és így rögtön megkapjuk a 3. tulajdonságból minden esemény valószínűségét (az eseménybe foglalt lehetséges kimenetek száma szorozva $\frac{1}{6}$ -al). Ha az eseménytér megszámlálható elemszámú, akkor működik is ez a módszer, azaz az elemi események valószínűségeiből minden esemény valószínűségét megkaphatjuk (nem megszámlálható esetben nem működik, például az egységintervallumon egyenletesen választott pont esetén minden egyes pont valószínűsége nulla, az intervallumok azonban már a hosszuknak megfelelő valószínűséget képviselnek).

Diszkrét és folytonos valószínűségi változók

Valószínűségi változónak nevezünk egy eseménytér (kísérlet) elemeihez valós számokat rendelő $X : \Omega \rightarrow \mathbb{R}$ függvényt, ha

$$\forall x \in \mathbb{R} : \left\{ \omega \in \Omega : X(\omega) < x \right\} \in \mathcal{R} \text{ (halmaz eseményt alkot)}$$

A kockadobás esetén legegyszerűbb módon úgy definiálhatunk egy X valószínűségi változót, hogy az értékének a dobás kimenetelét adjuk.

A X valószínűségi változó **diszkrét**, ha az értékkészlete véges vagy megszámlálhatóan végtelen.

Ha a X diszkrét valószínűségi változó értékkészlete $\{x_1, x_2, \dots\}$, akkor a $P(X = x_i)$ számokat X **eloszlásának** nevezzük.

Megjegyzés: Véges vagy végtelen sok szám, akkor és csak akkor alkot diszkrét eloszlást, ha nem negatívak és az összegük **1**.

A X valószínűségi változó **eloszlásfüggvénye**: $F_X : \mathbb{R} \rightarrow \mathbb{R}$

$$F_X(x) = P(X < x)$$

Tétel. Az $F_X : \mathbb{R} \rightarrow \mathbb{R}$ függvény pontosan akkor **eloszlásfüggvénye** valamely valószínűségi változónak, ha

1. monoton növekvő: $\forall x_1 < x_2 : F_X(x_1) \leq F_X(x_2)$
2. balról folytonos: $\forall x_0 \in \mathbb{R} : \lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0)$
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
4. $\lim_{x \rightarrow +\infty} F_X(x) = 1$

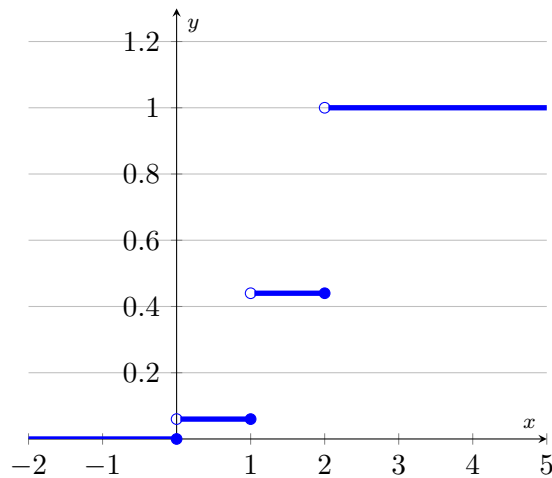
Egy diszkrét valószínűségi változó eloszlásfüggvénye olyan lépcsős függvény, amely a lehetséges értékeknél ugrik, és az ugrás nagysága az adott érték valószínűsége.

Ketten lőnek céltáblára. Az A találati esélye 0,7. és a B találati esélye 0,8.

Mindketten egy lövést adnak le egymástól függetlenül.

Jelentse X a találatok számát és adjuk meg az eloszlásfüggvényt!

találatok száma	valószínűség
0	$(1 - 0.7) * (1 - 0.8) = 0.3 * 0.2 = \mathbf{0.06}$
1	$0.7 * (1 - 0.8) + (1 - 0.7) * 0.8 = 0.7 * 0.2 + 0.3 * 0.8 = \mathbf{0.38}$
2	$0.7 * 0.8 = \mathbf{0.56}$



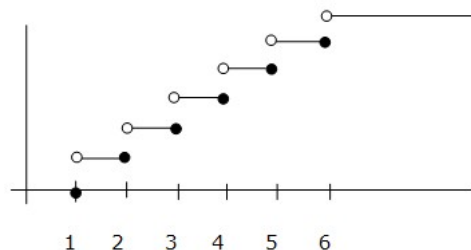
Ha X diszkrét valószínűségi változó értékkészlete $\{x_1, x_2, \dots\}$, akkor

$$F_X(x) = \sum_{i: x_i < x} P(X = x_i)$$

A szabályos kockadobás esetén az előző példában definiált X valószínűségi változónak adjuk meg az eloszlásfüggvényét. Látható, hogy az 1, 2, 3, 4, 5 és 6 értékeknél lesz csak változás, ezeken a pontokon kívül nem.

Az egyes intervallumokra a megfelelő valószínűségeket kiszámolva megkapjuk, hogy

- $F_X(x) = 0$, ha $x \leq 1$,
- $F_X(x) = \frac{i}{6}$, ha $i < x \leq i + 1$ és $1 < x \leq 6$,
- $F_X(x) = 1$, ha $x > 6$.



1. ábra

Legyen a X diszkrét valószínűségi változó értékkészlete $\{x_1, x_2, \dots\}$.

Ekkor X **várható értéke**

$$E(X) = \sum_k x_k \cdot P(X = x_k)$$

(feltéve, hogy ez a sor abszolút konvergens, azaz $\sum_k |x_k| \cdot P(X = x_k) < \infty$).

Példa: 3 darab 10 dollárossal befektetési terveink vannak, egy rulett segítségével.
A terv a következő: felteszünk 10 dollárt a pirosra.

Ha nyer, akkor megdupláztuk a 10 dollárt és abbahagyjuk a játékot.

Ha veszít, akkor újabb 10 dollárt teszünk a pirosra, és ha ezúttal nyerünk, akkor szintén abbahagyjuk a játékot.

Ha másodszorra sem nyerünk, akkor az utolsó 10 dollárt is felrakjuk a pirosra.

A kérdés, hogy várhatóan mennyi pénzünk lesz a tranzakció végén.

A ruletten 18 piros, 18 fekete és egy zöld mező található.

NY: Nyert, V: Veszített.

1.	2.	3.	X_i	$P(X_i)$
NY			40	$\frac{18}{37} = 0.0487$
V	NY		30	$\frac{19}{37} \cdot \frac{18}{37} = 0.25$
V	V	NY	20	$\frac{19}{37} \cdot \frac{19}{37} \cdot \frac{18}{37} = 0.128$
V	V	V	0	$\frac{19}{37} \cdot \frac{19}{37} \cdot \frac{19}{37} = 0.135$

$$EX = \sum_k x_k \cdot P(X = x_k) = 40 \cdot 0,487 + 30 \cdot 0.25 + 20 \cdot 0.128 = 29,54$$

Folytonos valószínűségi változók

A X valószínűségi változó abszolút folytonos, ha létezik olyan $f_X : \mathbb{R} \rightarrow \mathbb{R}$ függvény, hogy $F_X = \int_{-\infty}^x f_X(t) dt$ minden $x \in \mathbb{R}$ esetén. Ekkor az f_X függvényt a X valószínűségi változó **sűrűségfüggvényének** nevezzük.

Tétel. Egy $f : \mathbb{R} \rightarrow \mathbb{R}$ függvény, „akkor és csak akkor” **sűrűségfüggvénye** valamilyen abszolút folytonos valószínűségi változónak, ha

- f nem negatív (azaz $f(x) \geq 0, \forall x \in \mathbb{R}$)

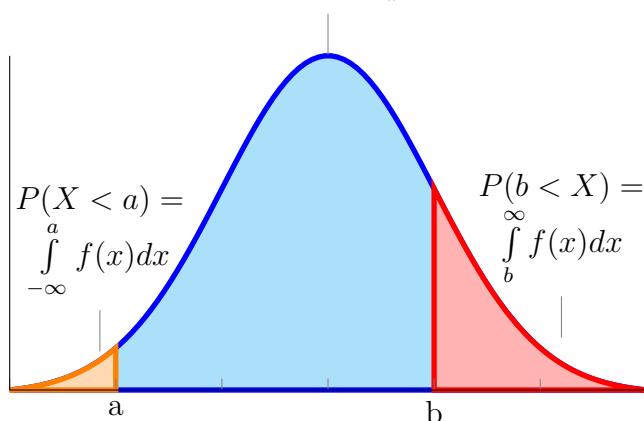
- $\int_{-\infty}^{\infty} f(x) dx = 1$

Megjegyzés:

- $P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (a, b \in \mathbb{R}, a < b \text{ esetén})$

- Ha X abszolút folytonos valószínűségi változó, akkor $F_X(x)$ mindig folytonos függvény, és $f_X(x)$ folytonos az $x \in \mathbb{R}$ pontban, akkor F_X differenciálható x -ben és $F'_X(x) = f_X(x)$.

$$P(a < X < b) = \int_a^b f(x) dx$$



$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^a f(x) dx + \int_a^b f(x) dx + \int_b^{\infty} f(x) dx = 1$$

Legyen a X abszolút folytonos valószínűségi változó sűrűségfüggvénye $f_X(x)$.
Ekkor X **várható értéke**

$$EX = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

(feltéve, hogy ez az integrál abszolút konvergens, azaz $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty$)

A X valószínűségi változó **szórásnégyzete** (vairanciája):

$$D^2 X = E(X - EX)^2 = EX^2 - E^2 X$$

(feltéve, hogy ez létezik.)

Megjegyzés: A X valószínűségi változó szórása

$$DX = +\sqrt{D^2 X}$$

- Ha X diszkrét valószínűségi változó:

$$D^2 X = \sum_k x_k^2 \cdot P(X = x_k) - \left(\sum_k x_k \cdot P(X = x_k) \right)^2$$

Roulettes példa: $D^2 X = 40^2 * 0,487 + 30^2 * 0.25 + 20^2 * 0.128 = 1055.2$

- Ha X abszolút folytonos valószínűségi változó:

$$D^2 X = \int_{-\infty}^{\infty} x^2 \cdot f_X dx - \left(\int_{-\infty}^{\infty} x \cdot f_X(x) dx \right)^2$$

Diszkrét valószínűségi változók

Értékkészlete legfeljebb megszámlálhatóan végtelen, azaz $\{x_1, \dots, x_n, \dots\}$ elemekből áll.

Ekkor eloszlása: $p_k := P(X = x_k)$.

Név	Értelmezés	Eloszlás	EX	D^2X
indikátor $Ind(p)$	Egy p valószínűségű esemény bekövetkezik-e vagy sem.	$P(X = 1) = p$ $P(X = 0) = 1 - p$	p	$p(1 - p)$
geometriai (Pascal) $Geo(p)$	Hányadikra követ- kezik be először egy p valószínűségű esemény.	$P(X = k) =$ $p(1 - p)^{k-1}$ $k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$

Ha a kísérletünk a kockadobás az első 6-os dobásig és az
X valószínűségi változó a dobások száma, akkor 6 geometriai eloszlású.

hipergeometriai $Hipgeo(N, M, n)$	Visszatevés nélküli mintavétel.	$P(X = k) =$ $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ $k = 0, 1, \dots, n$	$n \frac{M}{N}$	$n \frac{M}{N} (1 - \frac{M}{N}) (1 - \frac{n-1}{N-1})$
--------------------------------------	------------------------------------	--	-----------------	---

Ha N és M a végtelenbe tart úgy, hogy M / N egy 0 és 1 közötti p konstanshoz tart, akkor
a hipergeometrikus eloszlások sorozata a binomiális eloszláshoz tart.

binomiális $Bin(n, p)$	Visszatevéses min- tavétel.	$P(X = k) =$ $\binom{n}{k} p^k (1 - p)^{n-k}$ $k = 0, 1, \dots, n$	np	$np(1 - p)$
negatív bino- miális $Negbin(n, p)$	Hányadikra követke- zik be n . alkalommal egy p valószínűségű esemény.	$P(X = k) =$ $\binom{k-1}{n-1} p^n (1 - p)^{k-n}$ $k = n, n + 1, \dots$	$\frac{n}{p}$	$\frac{n(1 - p)}{p^2}$
Poisson $Poi(\lambda)$	Ritka esemény.	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ

Példa:

Hipergeometrikus:

Egy úton 30 nap alatt 12 napon történt baleset. Ebből a 30 napból kiválasztunk egy hetet, mi a valószínűsége, hogy ezen a héten 2 balesetes nap van?

Hipergeometriainál ismert, hogy mennyi az összes elem és az összes selejt N, K, és a minta n.
Az összes elem $N = 30$ nap, ebből (selejt) a balesetes nap, $M = 12$. A minta egy hét, vagyis
 $n = 7$, és itt $X = 2$ balesetes napot szeretnénk.

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\binom{12}{2} \binom{30-12}{7-2}}{\binom{30}{7}} = \frac{\binom{12}{2} \binom{18}{5}}{\binom{30}{7}} = \frac{565488}{2035800} = 0.27777188$$

$$EX = n * \frac{M}{N} = 7 * \frac{12}{30} = 0.23$$

A másik két feladatban csak valamilyen százalékos érték, a várható, az átlag, az arány vagy valószínűség. Ez esetben nem tudjuk, hogy mennyi baleset történik a 30 nap alatt, csak azt tudjuk, hogy várhatóan mennyi.

Binomiális:

Egy úton 30 nap alatt átlag 12 balesetes nap van. Mi a valószínűsége, hogy egy adott héten 2 balesetes nap van?

A binomiálisnál vett példában nem lehet több a balesetes napok száma, mint 7, így X korlátos.
 X = balesetes napok száma.

$$n = 7, p = \frac{M}{N} = \frac{12}{30} = 0.4.$$

$$P(X = k = 2) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{7}{2} * 0.4^2 * (1-0.4)^{7-2} = 21 * 0.16 * 0.07776 = 0.2612736$$

$$EX = n * p = 7 * 0.4 = 2.8$$

Poisson:

Egy úton 30 nap alatt átlag 12 baleset történik. Mi a valószínűsége, hogy egy adott héten 2 baleset van?

Az első két esetben X a balesetes napok száma, a harmadikban pedig a balesetek száma.

A harmadik feladatnál a Poisson eloszlás esetében baleset tetszőleges számú lehet, átlagban 12 30 naponta, de lehet akár 1000 is, tehát itt X nem korlátos.

X = balesetek száma.

λ = várhatóan hány baleset van egy héten 30 naponta 12 baleset szokott lenni, tehát naponta $\frac{12}{30} = 0.4$ és így hetente hétszer annyi $7 * 0.4 = 2.8$.

$$P(X = k = 2) = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{2.8^2}{2!} e^{-2.8} = \frac{7.84}{2} * 0.0608101 = 0.238375592$$

$$E(X) = \lambda = 2.8$$

Név	Eloszlásfüggvény	Sűrűségfüggvény	EX	D^2X
egyenletes $E(a, b)$	$\begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & b < x \end{cases}$	$\begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{különben} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
exponenciális $Exp(\lambda)$	$\begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{különben} \end{cases}$	$\begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & \text{különben} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
normális $N(m, \sigma^2)$...	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in \mathbb{R}$	m	σ^2
standard normális $N(0, 1^2)$	$\Phi(x) = \dots$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \in \mathbb{R}$	0	1
gamma $\Gamma(\alpha, \lambda)$...	$\begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{különben} \end{cases}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

Fogalmak

- *Konvolúció:* X, Y független valószínűségi változók, konvolúciójuk az $X + Y$ v. v.
- *Függetlenség:* $P(X_1 < x_1, \dots, X_n < x_n) = \prod_{i=1}^n P(X_i < x_i)$ vagy diszkrét esetben:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

- *Várható érték:* (Ω, \mathcal{R}, P) valószínűségi mező, $X : \Omega \rightarrow \mathbb{R}$ valószínűségi változó, $EX = \int_{\Omega} X dP$, ha ez létezik. Diszkrét esetben $EX = \sum_k x_k \cdot p_k$, ha abszolút konvergens. Abszolút folytonos esetben $EX = \int_{-\infty}^{\infty} x \cdot f(x) dx$, ha abszolút folytonos.
- *Szórásnégyzet:* $D^2X = E((X - EX)^2) = EX^2 - E^2X$
- *l. momentum:* $EX^l = \int_{\Omega} x^l dP$, ha létezik.
- *Szórás:* $DX = \sqrt{D^2X}$
- *Kovariancia:* $cov(X, Y) = E((X - EX)(Y - EY)) = E(XY) - EX \cdot EY$. Ha $cov(X, Y) = 0$, akkor X és Y korrelálatlan. (Megjegyzés: ha két v.v. független, akkor $cov(X, Y) = 0$, vagyis korrelálatlanok; illetve $cov(X, X) = D^2X$.)
- *Korreláció:* $R(X, Y) = \frac{cov(X, Y)}{DX \cdot DY}$, két v.v. lineáris kapcsolatát méri.
 - $R > 0 \rightarrow$ pozitív
 - $R < 0 \rightarrow$ negatív
 - $R^2 \sim 0 \rightarrow$ gyenge
 - $R^2 \sim 0.5 \rightarrow$ közepes
 - $R^2 \sim 1 \rightarrow$ erős

Nagy számok törvénye

Gyenge törvény

X_1, X_2, \dots függetlenek, azonos eloszlásúak, $EX_i = m < \infty$, $D^2X_i = \sigma^2 < \infty$.

$$P\left(\frac{X_1 + \dots + X_n}{n} - m \geq \varepsilon\right) \rightarrow 0 \quad (n \rightarrow \infty)$$

$\forall \varepsilon > 0$ -ra (sztochasztikus konvergencia).

Erős törvény

X_1, X_2, \dots függetlenek, azonos eloszlásúak, $EX_1 = m < \infty$, $D^2X_1 = \sigma^2 < \infty$.

$$\frac{X_1 + \dots + X_n}{n} \rightarrow m \quad (n \rightarrow \infty)$$

1 valószínűséggel.

Megjegyzés: Csebisev-egyenlőtlenséggel bizonyítjuk. $\left(\frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad (n \rightarrow \infty)\right)$

Csebisev-egyenlőtlenség

EX véges.

$$\text{Ekkor } P(|X - EX| \geq \lambda) \leq \frac{D^2X}{\lambda^2}$$

Megjegyzés: Bizonyítás Markov-egyenlőtlenséggel.

Markov-egyenlőtlenség

$X \geq 0, c > 0$.

$$\text{Ekkor } P(X \geq c) \leq \frac{EX}{c}$$

Konvergenciafajták

$X_n \rightarrow X$, vagyis X konvergens.

- *sztochasztikusan*: ha $\forall \varepsilon > 0$ -ra $P(|X_n - X| \geq \varepsilon) \rightarrow 0 \quad (n \rightarrow \infty)$.
- *1 valószínűséggel (majdnem mindenütt)*: ha $P(\omega : X_n(\omega) \rightarrow X(\omega)) = 1$.
- *L^p -ben*: ha $E(|X_n - X|^p) \rightarrow 0 \quad (n \rightarrow \infty)$ ($p > 0$ rögzített).
- *eloszlásban*: ha $F_{X_n}(x) \rightarrow F_X(x) \quad (n \rightarrow \infty)$ az utóbbi minden folytonossági pontjában.

Kapcsolataik: 1 valószínűségű és L^p -beli a legerősebb, ezekből következik a sztochasztikus, ebből pedig az eloszlásbeli.

Centrális határeloszlás tétel

X_1, X_2, \dots függetlenek, azonos eloszlásúak, $EX_1 = m < \infty$, $D^2X_1 = \sigma^2 < \infty$.

Ekkor $\frac{X_1 + \dots + X_n - nm}{\sqrt{n}\sigma} \rightarrow N(0, 1) \quad (n \rightarrow \infty)$ eloszlásban, azaz

$$P\left(\frac{X_1 + \dots + X_n - nm}{\sqrt{n}\sigma} < x\right) \rightarrow \Phi(x) \quad (n \rightarrow \infty)$$

Statisztikai mező

$(\Omega, \mathcal{R}, \mathcal{P})$ hármas, ha $\mathcal{P} = \{P_\vartheta\}_{\vartheta \in \Theta}$ és $(\Omega, \mathcal{R}, P_\vartheta)$ Kolmogorov-féle valószínűségi mező $\forall \vartheta \in \Theta$ -ra.

Fogalmak

- *Minta*: $\underline{X} = (X_1, \dots, X_n) : \Omega \rightarrow X \in \mathbb{R}^n$. (X_i valószínűségi változó)
- *Mintatér*: X , minta lehetséges értékeinek halmaza, gyakran $\mathbb{R}^n, \mathbb{Z}^n$.
- *Minta [realizációja]*: $\underline{x} = (x_1, \dots, x_n)$, konkrét megfigyelés.
- *Statisztika*: $T : X \rightarrow \mathbb{R}^k$.
- *Statisztika alaptétele*: (Glivenko–Cantelli-tétel) X_1, X_2, \dots független, azonos eloszlású F eloszlásfüggvénnyel. Ekkor az F_n tapasztalati eloszlásfüggvényre teljesül, hogy

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \rightarrow 0 \quad (n \rightarrow \infty)$$

1 valószínűséggel.

Statisztikai becslések

$(\Omega, \mathcal{R}, \mathcal{P})$ statisztikai mező, $\vartheta \in \Theta$, $P_\vartheta(X_1 < x_1, \dots, X_n < x_n) = F_\vartheta(\underline{x})$

$T(\underline{X})$ a ϑ becslése, ha $T : \mathbb{R}^n \rightarrow \Theta$.

$T(\underline{X})$ a $h(\vartheta)$ becslése, ha $T : \mathbb{R}^n \rightarrow h(\Theta)$.

Torzítatlanság: $T(\underline{X})$ torzítatlan becslése $h(\vartheta)$ -nak, ha $E_\vartheta T(\underline{X}) = h(\vartheta) \forall \vartheta \in \Theta$.

Aszimptotikusan torzítatlan: $T(\underline{X})$ aszimptotikusan torzítatlan a $h(\vartheta)$ -ra, ha $E_\vartheta T(\underline{X}) \rightarrow h(\vartheta) (n \rightarrow \infty) \forall \vartheta \in \Theta$.

A T_1 torzítatlan becslés *hatásosabb* T_2 torzítatlan becslésnél, ha $D_\vartheta^2 T_1 \leq D_\vartheta^2 T_2 \forall \vartheta \in \Theta$.

Hatásos, ha minden más torzítatlan becslésnél hatásosabb. Ha van hatásos becslés, akkor az egyértelmű.

- *Maximum-likelihood becslés*: Likelihood függvény: $L(\vartheta, \underline{x}) = \begin{cases} P_\vartheta(\underline{X} = \underline{x}) & \text{diszkr.} \\ f_{\vartheta, \underline{X}}(\underline{x}) & \text{absz.folyt.} \end{cases}$

Független esetben: $L(\vartheta, \underline{x}) = \begin{cases} \prod_{i=1}^n P_\vartheta(X_i = x_i) & \text{diszkr.} \\ \prod_{i=1}^n f_{\vartheta, X_i}(x_i) & \text{absz.folyt.} \end{cases}$

$\hat{\vartheta}$ a ϑ ismeretlen paraméter maximum-likelihood becslése, ha $L(\hat{\vartheta}, \underline{X}) = \max_{\vartheta \in \Theta} L(\vartheta, \underline{X})$.

- *Momentum-módszer becslés*: $\vartheta = (\vartheta_1, \dots, \vartheta_k)$, X_1, \dots, X_n , l . momentum: $M_l(\underline{\vartheta}) = E_{\underline{\vartheta}} X_i^l$,

$$\sum_{i=1}^n X_i^l$$

tapasztalati l . momentum: $\hat{M}_l = \frac{\sum_{i=1}^n X_i^l}{n}$.

$\hat{\underline{\vartheta}}$ a $\underline{\vartheta}$ momentum módszer szerinti becslése, ha megoldása az $M_l(\underline{\vartheta}) = \hat{M}_l$, $l = 1..k$ egyenletrendszernek.

Hipotézisvizsgálat

Felteszünk egy hipotézist, és vizsgáljuk, hogy igaz-e. Elfogadjuk vagy elutasítjuk. Lehet paraméteres vagy nem paraméteres, vizsgálhatjuk várható értékek, szórások egyezőségét, értékét, teljes eseményrendszerek függetlenségét. Illeszkedésvizsgálattal megállapíthatjuk, hogy a valószínűségi változók adott eloszlásfüggvényűek-e, homogenitásvizsgálattal pedig azt, hogy ugyanolyan eloszlású-e két minta.

H_0 : nullhipotézis, $\vartheta \in \Theta_0$; H_1 : ellenhipotézis, $\vartheta \in \Theta_1$; $\Theta = \Theta_0 \cup \Theta_1$.

Egy- és kétoldali vizsgálat: Kétoldali ellenhipotézisnél a nem egyezőséget tesszük fel, egyoldalinál valamilyen relációt. Kétoldalinál a próba értékének abszolút értékét vizsgáljuk, hogy az elfogadási tartományon belül van-e, ekkor például az u-próbánál az adott hibaszázalékot meg kell felelni a számításhoz, hiszen a Φ függvény szimmetrikus az y tengelyre.

- *Statisztikai próba*: $X = X_e \cup X_k$ (diszjunkt halmazok) elfogadási és kritikus tartomány. Ez a felbontás a statisztikai próba. Ha a megfigyelés eleme a kritikus tartománynak, akkor elutasítjuk a nullhipotézist, ha nem eleme, akkor elfogadjuk. $T(x) = \begin{cases} 1 & x \in X_k \\ 0 & \text{otherwise} \end{cases}$
- *Elsőfajú hiba*: H_0 igaz, de elutasítjuk. Valószínűsége: $P_\vartheta(\underline{X} \in X_k)$, $\vartheta \in \Theta_0$.
- *Másodfajú hiba*: H_0 hamis, de elfogadjuk. Valószínűsége: $P_\vartheta(\underline{X} \notin X_k)$, $\vartheta \in \Theta_1$. Az a cél, hogy ezek a hibák minél kisebbek legyenek. Egymás kárára javítható a két valószínűség, ha a megfigyelések száma rögzített.
- *Próba terjedelme*: α a próba terjedelme, ha $P_\vartheta(\underline{X} \in X_k) \leq \alpha$, $\vartheta \in \Theta_0$.
 α a próba pontos terjedelme, ha $\sup_{\vartheta \in \Theta_0} P_\vartheta(\underline{X} \in X_k) = \alpha$.

Klasszikus statisztikai próbák

Alapfogalmak

Alaphipotézisnek vagy **nullhipotézis**nek (H_0) nevezzük az alapfelvetésünket, amit igazolni szeretnénk. Ezzel szemben áll az ellenhipotézis, ami akkor teljesül, ha a nullhipotézis nem igaz.

Elfogadási tartománynak nevezzük egy olyan halmazt, amelyben a nullhipotézis teljesülése esetén a vizsgált statisztika értéke nagy valószínűséggel ($1 - p$) elhelyezkedik. Ezzel szemben a **kritikus tartomány**ba várhatóan akkor esik a h statisztika értéke, ha az ellenhipotézis teljesül.

Hibák		H_0 igaz	H_0 hamis
	H_0 elfogadása	helyes döntés	másodfajú hiba
	H_0 elvetése	elsőfajú hiba	helyes döntés

Egy **próba terjedelme** az elsőfajú hiba valószínűségének felső határa.

Ha β a másodfajú hiba valószínűsége, akkor a **próba erőfüggvénye**: $1 - \beta$.

Azonos terjedelmű próbák közül **erősebbnek** nevezzük azt, amelyiknek az erőfüggvénye minden ponton nem kisebb a másikénál.

Továbbiakban:

- X normális eloszlású, ismert szórású változó n elemű mintája:

$$X_i \sim N(m, \sigma^2), \quad i = 1, \dots, n$$

- X, Y egymástól független normális eloszlású, ismert szórású változó n, m elemű mintája:

$$\begin{aligned} X_i &\sim N(m, \sigma^2), & i &= 1, \dots, n, \\ \eta_j &\sim N(m_2, \sigma_2^2), & j &= 1, \dots, m \end{aligned}$$

- \bar{X} a vizsgált valószínűségi változó átlaga a mintában,
- m_i az előre megadott érték, amihez az átlagot viszonyítjuk,
- σ a vizsgált valószínűségi változó ismert szórása.

u-próbák

Egymintás eset

A próba azt ellenőrzi, hogy egy adott statisztikai ismerv esetén a mintabeli átlag szignifikánsan eltér-e a populációs átlagtól. Más szavakkal, hogy egy valószínűségi változó átlaga szignifikánsan különbözik-e egy adott m értéktől.

Adott egy X normális eloszlású, ismert szórású változó n elemű mintája. Nullhipotézisünk az, hogy a várható érték megegyezik-e egy konkrét értékkel ($H_0 : m = m_0$), másképpen fogalmazva azt vizsgáljuk, hogy a mintabeli átlag nem tér-e el szignifikánsan m_0 -tól.

Legyen a próbastatisztika:

$$u = \sqrt{n} \frac{\bar{X} - m_0}{\sigma}$$

Ekkor a p valószínűség mellett elfogadjuk a nullhipotézist, ha $|u| \leq u_{\frac{p}{2}}$, ahol $u_{\frac{p}{2}}$ a standard normális eloszlás $1 - \frac{p}{2}$ kvantilise. Ha igaz a nullhipotézis, akkor ez közel standard normális eloszlású.

Ha az ellenhipotézis egyoldalú, akkor az elfogadási tartomány $u \leq u_p$ -re, illetve $u \geq -u_p$ -re módosul.

Kétmintás eset

A próba azt vizsgálja, hogy egy valószínűségi változó átlaga két külön mintában szignifikánsan különböző-e.

Két független (a fenti feltételeket teljesítő) mintára (X, Y) vonatkozó nullhipotézisünk, hogy várható értékük egyenlő ($H_0 : m_1 = m_2$). Erre a próbastatisztika:

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Tétel. Az u -próbák konzisztens, torzítatlan és legerősebb próbák.

- *u-próba*: Feltételezzük, hogy a minta normális eloszlású ($X_i \sim N(m, \sigma^2)$), $i = 1..n$, és hogy a szórás ismert.

- *Egymintás*: A nullhipotézis az, hogy a várható érték megegyezik-e egy konkrét értékkel (m_0), másképpen fogalmazva azt vizsgáljuk, hogy a mintabeli átlag nem tér-e el szignifikánsan m_0 -tól. Tehát $H_0 : m = m_0$, és kétoldali esetben $H_1 : m \neq m_0$, egyoldali pedig például $H_1 : m \geq m_0$ vagy $H_1 : m < m_0$.

Az u-próba értéke: $u = \sqrt{n} \frac{\bar{X} - m_0}{\sigma}$. Ha igaz a nullhipotézis, akkor ez közel standard normális eloszlású.

ε hibavalószínűséggel vizsgáljuk a hipotézist, ehhez szükségünk van a $\Phi(u_{1-\varepsilon}) = 1 - \varepsilon$ értékre.

Kétoldali esetben H_0 -t elutasítjuk, ha $|u| > u_{1-\frac{\varepsilon}{2}}$, és elfogadjuk, ha $|u| \leq u_{1-\frac{\varepsilon}{2}}$.

Egyoldali esetben $u > u_{1-\varepsilon}$ (jobb) és $u < u_{1-\varepsilon}$ (bal) esetét vizsgáljuk, ezen esetekben utasítjuk el H_0 -t.

- *Kétmintás*: Itt a feltételek a következők: $X_i \sim N(m_1, \sigma_1^2)$, $i = 1..n$ és $\eta_j \sim N(m_2, \sigma_2^2)$, $j = 1..m$. A szórások szintén ismertek. $H_0 : m_1 = m_2$, és $u = \frac{\bar{X} - \bar{\eta}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$. $H_1 : m_1 > m_2$, ez a felső (jobb?) oldali, $H_1 : m_1 < m_2$ pedig az alsó (bal?) ellenhipotézis.

- *t-próba*: Ennél a próbánál nem ismert a szórás, viszont ugyanúgy normális eloszlást feltételezünk, mint az u-próbánál. $X_i \sim N(m, \sigma^2)$, $i = 1..n$.

- *Egymintás*: $H_0 : m = m_0$. Ellenhipotézis az u-próbához hasonlóan. $t = \sqrt{n} \frac{\bar{X} - m_0}{\sqrt{\sigma_*^2}}$, ahol σ_*^2 a korrigált tapasztalati szórásnégyzet, amit a mintából számíthatunk ki. (Megjegyzés: n helyett $n - 1$ -gyel osztunk a képletben.) Ez az érték t -eloszlású H_0 esetén, ami $n - 1$ szabadságfokú. Más néven szokás ezt a próbát Student-próbának is nevezni.

- *Kétmintás*: $X_i \sim N(m_1, \sigma_1^2)$, $i = 1..n$ és $\eta_j \sim N(m_2, \sigma_2^2)$, $j = 1..m$. Ez esetben sem ismert a szórás, viszont feltételezzük, hogy a két minta szórása megegyezik. Ekkor $t_{n+m-2} = \frac{\sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{X} - \bar{\eta}}{\sqrt{\sum (X_i - \bar{X})^2 + \sum (\eta_j - \bar{\eta})^2}}}{\sqrt{\sum (X_i - \bar{X})^2 + \sum (\eta_j - \bar{\eta})^2}}$. $n + m - 2$ a próba szabadságfoka.

- *f-próba*: Két minta esetén használható. Ez a próba szórások egyezőségének vizsgálatára alkalmas, tehát itt $H_0 : \sigma_1 = \sigma_2$. Ha a két minta szórásnégyzete megegyezik, akkor a hányadosuk 1-hez tart. $f_{n-1, m-1} = \max(\frac{\sigma_1^2}{\sigma_2^2}, \frac{\sigma_2^2}{\sigma_1^2})$. A két szabadsági fok közül az első az f számlálójához tartozó minta elemszáma -1 , a második a nevezőjéhez.

- *Welch-próba*: Más néven d-próba. Hasonló, mint a kétmintás t-próba, de itt a szórások egyezőségét nem kell feltenni. Szabadsági foka bonyolult képlettel számítható.

- *szekvenciális próbák*: $V_n = \frac{\prod f_1(x_i)}{\prod f_0(x_i)} = \frac{L_1(\underline{x})}{L_0(\underline{x})}$. f_0 a nullhipotézis szerinti sűrűségfüggvény, f_1 az ellenhipotézis szerinti. Adott egy A és egy B érték, $A < B$. Ha $V_n \geq B$, akkor elutasítjuk H_0 -t, ha $V_n \leq A$, akkor elfogadjuk, és ha $A < V_n < B$, akkor új mintaelemet

veszünk.

Stein tétele szerint N 1 valószínűséggel véges. $N = \min\{n : V_n \leq A \vee V_n \geq B\}$.

- *Minőség-ellenőrzés:* n_1 elemet nézünk, $c_1 < c_2$ és c_3 határértékek. Ha $X_1 \leq c_1$, akkor elfogadjuk H_0 -t, ha $X_1 \geq c_2$, akkor elutasítjuk. Ha $c_1 < X_1 < c_2$, akkor megnézzük n_2 elemet, és ha $X_1 + X_2 \leq c_3$, akkor szintén elfogadjuk H_0 -t. A várható mintaelemszám méri a hatékonyságát.
- χ^2 -próba: $H_0 : A_1, \dots, A_n$ teljes eseményrendszer. $P(A_i) = p_i, i = 1..n$, ν_i a gyakoriság. Ha teljesül a nullhipotézis, akkor $\frac{\nu_i}{n} \sim p_i$.
 $\chi^2 = \sum \frac{(\nu_i - np_i)^2}{np_i}$. Ez χ^2 eloszlású, aminek $r - 1$ szabadságfoka van. r az összeadott csoportok száma. (Megjegyzés: ha túl kicsi lenne 1-1 csoportban a gyakoriság, akkor azokat összevonjuk.)
 χ^2 -próbát használhatunk illeszkedés-, homogenitás- és függetlenségvizsgálatra is. (Megjegyzés: más képlet van mindhez.)
- *Egyéb próbák:*
 - *Kolmogorov-Szmirnov-próba:* 2 tapasztalati eloszlásfüggvény megegyezik-e (homogenitásvizsgálat), vagy 1 minta esetén megegyezik-e valamilyen eloszlásfüggvénnyel. $D_{m,n} = \max_x |F_n(x) - G_m(x)|$. X_i F eloszlásfüggvénnyel, Y_j G -vel. $H_0 : F \equiv G$.
 - *Előjel-próba:* Hányszor teljesül, hogy valami pozitív.
 - *Wilcoxon-próba:* (rangstatisztika), $P(X > Y) = \frac{1}{2}$ tesztelésére összeszámoljuk, hogy hány párra teljesül, hogy $X_i > Y_j$.