Valószínűségszámítás és matematikai statisztika

Témakörök, elméleti összefoglaló

Mérnökinformatikus BSc, 2021-22 tavasz

Alábbi tételsor a 2021-22 tavaszi féléves valszám vizsgára készült, így ha későbbi vizsgára készülne belőle valaki, érdemes ellenőrizni, hogy a tételsor megegyezik-e. Ebben a tételsorban a paraméteres próbáktól kezdve egyik tétel sincs túl komplexen kidolgozva, részint azért, mert ezen témaköröket csak halványan érintettük a félév folyamán, illetve azért, mert személy szerint kétlem, hogy ezekből (nekünk legalábbis) adnának kérdést a vizsgán.

- Kacsa, Róka, Tsunami

Tartalom

1. tétel: Események

Kísérlet, kimenetel, elemi esemény, eseménytér, esemény. Lehetetlen esemény, biztos esemény. Műveletek eseményekkel. Teljes eseményrendszer.

- **Kísérlet**: véletlen, akárhányszor ismételhető tömegjelenség megfigyelése
- **Elemi esemény**: a kísérlet egy lehetséges kimenetele
- **Eseménytér** (Ω): az elemi események összesége
- **Esemény**: az eseménytér egy részhalmaza
- **Esemény bekövetkezik**, ha az adott esemény egy elemi esemény, és a kísérlet során adódik

Példa: kockadobás

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- A esemény: 7-esnél kisebbet dobunk => $A = \{1, 2, 3, 4, 5, 6\}$
    - Biztos esemény: a kísérlet minden kimenetele esetén bekövetkezik (összes elemi esemény tartalmazza)
- B esemény: 7-esnél nagyobbat dobunk => $B = 0$
    - Lehetetlen esemény: olyan esemény, mely sohasem következik be (egyik elemi esemény sem tartalmazza)
- C esemény: 2-est dobunk => $C = \{2\}$
- D esemény: páratlant dobunk => $D = \{1, 3, 5\}$
    - C és D egymást kizáró események: a két esemény egyszerre sosem következik be, azaz nincs közös halmazuk => $A \cap B = \emptyset$

- Események **összege**: $A + B$, legalább az egyik esemény bekövetkezik (unió)
- Események **szorzata**: $A * B$, mindkettő esemény bekövetkezik (metszet)
- Ellentett / **komplementer** esemény: egy A esemény komplementere $\bar{A}$, mely akkor következik be, amikor A esemény nem. Azaz: $\bar{A} = \Omega - A$
- Az események **kommutatívak, asszociatívak, disztributívak** (mivel halmazként vannak értelmezve)

- **Teljes eseményrendszer**: Az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, ha egymást páronként kizárják és összegük a biztos esemény => $A_1 + A_2 + \ldots + A_n = \Omega$.
    - Pl: A_1: a dobott szám négynél kisebb, A_2: a dobott szám a 4-es, A_3: a dobott szám négynél nagyobb
    - A három esemény páronként kizárja egymást, így teljes eseményrendszert alkot

2. tétel: Valószínűség

Véletlen tömegjelenség. Valószínűség, a valószínűségi mező.
A klasszikus valószínűségi mező. A geometriai valószínűségi mező.

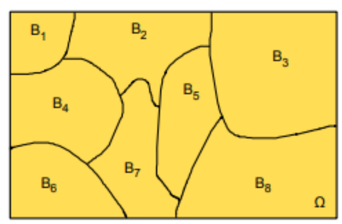
A valószínűség

Ha egy kísérletet azonos körülmények között *n*-szer lefuttatunk, és az *A* esemény (mint véletlen tömegjelenség) k_a alkalommal következik be, akkor azt mondjuk, hogy ez a k_a szám az *A* esemény gyakorisága, $\frac{k_a}{n}$ pedig az esemény relatív gyakorisága. A bekövetkezések és az összes kísérletek számának aránya egy meghatározott számérték körül ingadozik => ez az adott esemény valószínűsége, *P(A)*, melynek értéke 0 és 1 közötti szám. Biztos esemény valószínűsége 1, lehetetlen eseményé 0.

A valószínűségszámítás 3 axiómája

1. Az adott Ω eseménytér minden egyes A eseményéhez tartozik egy 0 és 1 közé eső P(A) szám, amelyet az A esemény valószínűségének (valószínűségi mértékének) nevezünk.
2. A biztos esemény valószínűsége 1, azaz P(Ω) = 1
3. Az egymást páronként kizáró események összegének a valószínűsége megegyezik az egyes események valószínűségeinek egyenként való összeadott összegével:
    a. P(A1 + A2 + . . .) = P(A1) + P(A2) + ...

Az axiómák egyszerűbb következményei



- A esemény komplementerének valószínűsége: 1 – P(A)
- Teljes eseményrendszer: Az $A_1, A_2, \ldots, A_n$ események teljes eseményrendszert alkotnak, ha egymást páronként kizárják és összegük a biztos esemény => $A_1 + A_2 + \ldots + A_n = \Omega$.
    - Pl: A_1: a dobott szám négynél kisebb, A_2: a dobott szám a 4-es, A_3: a dobott szám négynél nagyobb
    - A három esemény páronként kizárja egymást, így teljes eseményrendszert alkot
- Teljes eseményrendszert alkotó események valószínűségeinek összege mindig 1.
- A – B valószínűsége: P(A − B) = P(A) − P(A · B)
- Ha A esemény része a B-nek (A⊂B), akkor P(B − A) = P(B) − P(A)
- A és B események összegének valószínűsége: P(A + B) = P(A) + P(B) − P(A · B).

A klasszikus valósznűségi mező

Ha egy kísérlettel kapcsolatban az elemi események száma véges (*n*), és minden esemény valószínűsége egyenlő (1/*n*), akkor *k*-féleképpen bekövetkező A esemény valószínűsége: $P(A) = \frac{kedvező}{összes} = \frac{k}{n}$

Visszatevéses és visszatevés nélküli mintavétel

Véletlenszerűen kiválasztunk néhány mintát és a bennük található hibásak számából (*k*) *következtetünk* az összesben található hibások számára.

Visszatevéses: a minta elemeit egyesével választjuk ki, majd a vizsgálat után visszatesszük őket, és ezután vesszük a következőt, stb.

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Visszatevés nélküli: az elemeket kiválaszthatjuk egyszerre, mert nem számít a sorrend, vagy egyesével, ügyelve hogy ne tegyük vissza (számít a sorrend).

$$\frac{\binom{s}{k} \cdot \binom{N-s}{n-k}}{\binom{N}{n}},$$

A geometriai valósznűségi mező

Ha egy kísérlet leírható geometriai alakzatként (hosszúság, terület, térfogat) és az eseményeinek a valószínűsége felrajzolható geometriai alakzatként, ami része a kísérlet alakzatának, akkor az események és valószínűségeik geometriai valószínűségi mezőt alkotnak.

$$P(A) = \frac{A\ esemény\ részalakzata}{teljes\ alakzat} = \frac{m}{M}$$

Másképpen: ha a teljes eseménytérnek egy geometriai alakzat feleltethető meg; továbbá minden eseménynek ennek az alakzatnak egy részhalmaza feleltethető meg oly módon, hogy az esemény valószínűsége csak az alakzat mértékétől (hossz/terület/térfogat) függ, míg az alakzat elhelyezkedésétől és alakjától független.

3. tétel: Feltételes valószínűség

Feltételes valószínűség, események függetlensége.
A teljes valószínűség tétele, Bayes-tétel.

Feltételes valószínűség

Adott esemény valószínűségének kiszámításához valamilyen plusz információval rendelkezkünk, mely egy másik, ismert esemény bekövetkezését jelenti. Ekkor az A esemény B-re vonatkozó feltételes valószínűsége: $P(A|B) = \frac{együttes\ bekövetkezés}{feltétel\ bekövetkezés} = \frac{P(A*B)}{P(B)}$

Ebből következik, hogy: $P(A*B) = P(A|B) * P(B)$

A teljes valószínűség tétele és a Bayes-tétel

Teljes valószínűség tétele: Ha a B_1, B_2, . . . , B_n események teljes eseményrendszert alkotnak, és mindegyik B esemény valószínűsége nagyobb, mint 0, illetve *A* egy tetszőleges esemény, akkor:

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i) \,.$$

(azaz: az *A* esemény valószínűsége meghatározható minden adott B_i eseményre vonatkozó feltételes valószínűség és az adott B_i esemény valószínűségének szorzatösszegével)

Bayes-tétel: Ha a B_1, B_2, . . . , B_n események teljes eseményrendszert alkotnak, és mindegyik *B* esemény valószínűsége nagyobb, mint 0, illetve *A* egy tetszőleges pozitív valószínűségű esemény (P(A) > 0), akkor:

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} \,.$$

(ahol a nevező megegyezik a teljes valószínűség tételével, azaz P(A)-val)

Események függetlensége

Két esemény **független**, ha együttes bekövetkezésük valószínűsége egyenlő a valószínűségeik szorzatával (ergó nincs közös metszetük). => $P(A*B) = P(A) * P(B)$

Egymást kizáró események nem lehetnek függetlenek, mert akkor $P(A*B) = 0$

Teljes függetlenség: Ha A_1, A_2, …, A_n eseményekből bárhogyan kiválasztva *k* darabot, ezeknek együttes bekövetkezésének valószínűsége megegyezik valószínűségeik szorzatával. Minden ilyennek, azaz

$$P(A_i \cdot A_j) = P(A_i) \cdot P(A_j)$$
$$P(A_i \cdot A_j \cdot k) = P(A_i) \cdot P(A_j) \cdot P(A_k)$$
$$\vdots$$
$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$$

$\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$ db feltételnek teljesülnie kell.

Ha *A* és *B* esemény független, akkor egyik bekövetkezése sem változtatja meg a másik bekövetkezésének valószínűségét, azaz: $P(A|B) = P(A)$ **és** $P(B|A) = P(B)$.

Ha *A* és *B* független események, akkor $A \text{ és } \bar{B}, ill. \bar{A} \text{ és } B$ is független események.

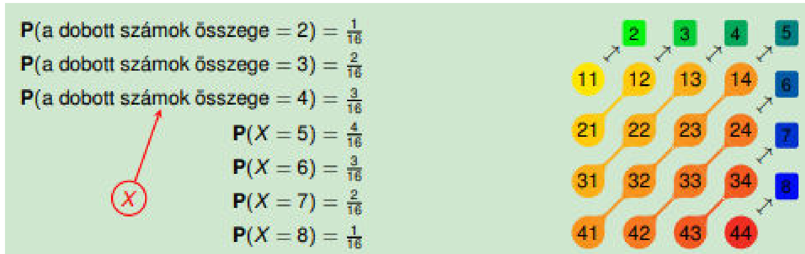
$$P(A \cdot \overline{B}) = P(A) \cdot P(\overline{B}),$$

4. tétel: Diszkrét valószínűségi változó

Valószínűségi változó. A valószínűségi változó eloszlása. Eloszlásfüggvény.
Módusz, medián, várható érték, momentum, szórás. A várható érték és a szórás tulajdonságai.

Valószínűségi változó

Azok a függvények, ahol az elemi események mindegyikéhez hozzárendelünk 1-1 valós számot (X : Ω → R)



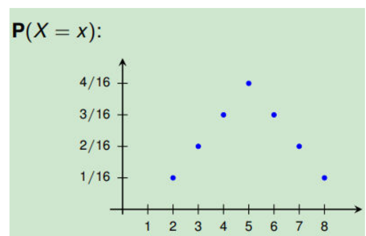
Ha egy kísérlet során során az a_i elemi esemény következik be, melyhez a valvál. megadásakor az x_i értéket rendeltük, akkor azt mondjuk, hogy az X valvál. az x_i értéket veszi fel, az x_i pedig az X egy lehetséges értéke.

Diszkrét valószínűségi változó: Olyan valvál., aminek véges sok, vagy megszámlálhatóan végtelen lehetséges értéke lehet. Megadjuk, hogy a valószínűségi változó milyen értékeket milyen valószínűséggel vesz fel (= eloszlás), melyet 4-féleképpen lehet ábrázolni:

- Felsorolással (első kép)
- Táblázattal
- Grafikonnal
- Képlettel

A táblázat első sorában a v.v. lehetséges értékei (x_i), alattuk a hozzájuk tartozó valószínűségek ($p_i = \mathbf{P}(X = x_i)$):

X:	2	3	4	5	6	7	8
	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$



Valószínűségfüggvény

Diszkrét valószínűségi változó “valószínűségfüggvénye”: $f: \mathbb{R} \to \mathbb{R}, f(x) = P(X = x)$

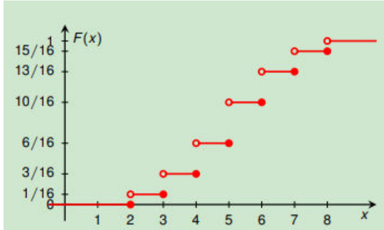
A „valószínűségfüggvény” tulajdonságai

1. $R_f \subseteq \mathbb{R}_0^+$.
2. $\sum_{x \in R_X} f(x) = 1$, ahol R_X az X értékkészlete ($|R_X| \leqslant \aleph_0$).

Eloszlásfüggvény

Diszkrét valószínűségi változó eloszlásfüggvénye: $F: \mathbb{R} \to \mathbb{R}, F(x) = P(X < x)$

Az X valvál. eloszlásfüggvénye az az F függvény, amely minden valós x értékhez hozzárendeli annak a valószínűségét, hogy az X valvál. annál az x-nél kisebb értéket vesz fel.



$$F(x) = \begin{cases} 0, & \text{ha } x \leqslant 2; \\ \frac{1}{16}, & \text{ha } 2 < x \leqslant 3; \\ \frac{3}{16}, & \text{ha } 3 < x \leqslant 4; \\ \frac{6}{16}, & \text{ha } 4 < x \leqslant 5; \\ \frac{10}{16}, & \text{ha } 5 < x \leqslant 6; \\ \frac{13}{16}, & \text{ha } 6 < x \leqslant 7; \\ \frac{15}{16}, & \text{ha } 7 < x \leqslant 8; \\ 1, & \text{ha } 8 < x. \end{cases}$$

Tulajdonságai:

- $D_F = \mathbb{R}$; $R_F \subseteq [0; 1]$.
- Monoton nő.
- Balról folytonos minden pontban.
- $\lim_{x \to -\infty} F(x) = 0; \qquad \lim_{x \to \infty} F(x) = 1.$

Középértékek

Módusz

X diszkrét valvál. leggyakoribb/legvalószínűbb értéke, azaz azon x_i, melyre p_i maximális. Jelölése: mod(X)

Medián

X diszkrét valvál. azon értéke, ahol "félbevágja" X eloszlását, azaz azon érték, melyre $P(X \leq x)$ és $P(X \geq x)$ egyaránt $\frac{1}{2}$. (ahol "átlépi" a 0.5-öt)

Jelölése: med(X)

Várható érték

Az X (diszkrét) valószínűségi változó várható értéke

$$\mathbf{E}(X) = \sum_i p_i x_i = \sum_{x \in R_X} x f(x),$$

ahol x_i a v.v. lehetséges értékei és p_i a megfelelő valószínűségek (ha a sor abszolút konvergens).

- A várható érték tulajdonságai:
  1. $\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$
  2. $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$
  3. $\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$

=> ha X és Y független valváltozók várható értéke létezik, akkor szorzatuknak is létezik várható értéke

A várható érték minden esetben egyértelmű. Sokkal jobban jellemzi az eloszlást, mert annak minden értéke és valószínűsége létezik. Jelentése: **A valvál. átlagos értékénél hol van az eloszlás súlypontja.**

Szórás

Terjedelem: az eloszlás lehetséges legnagyobb és legkisebb értékének különbsége.

Szórásnégyzet és szórás:

A X valószínűségi változó szórásnégyzete, illetve szórása: (ahol $\mathbf{E}(X^2) = \sum_i p_i x_i^2$).

$$\mathbf{D}^2(X) = \mathbf{E}\Big((X - \mathbf{E}(X))^2\Big) \qquad \text{és} \qquad \mathbf{D}(X) = \sqrt{\mathbf{E}\Big((X - \mathbf{E}(X))^2\Big)}.$$

A szórás megmutatja, hogy mennyire ingadoznak a valvál. értékei a várható érték körül.

Tulajdonságai:

- A szórás tulajdonságai:
  1. $\mathbf{D}(aX + b) = |a|\,\mathbf{D}(X)$
  2. $\mathbf{D}^2(X + Y) = \mathbf{D}^2(X) + \mathbf{D}^2(Y)$, ha X és Y függetlenek

5. tétel: Nevezetes diszkrét eloszlások

Indikátor (karakterisztikus) változó, binomiális eloszlás, hipergeometriai eloszlás, geometriai eloszlás, Poisson-eloszlás.

Indikátor (karakterisztikus) változó eloszlása

Egyetlen kísérletet végzünk, és arra vagyunk kíváncsiak, hogy egy adott valószínűségű *A* esemény bekövetkezik-e, vagy sem. Ha bekövetkezik, akkor X valvál. értéke legyen 1, különben 0. Ekkor az X az *A* esemény indikátor változója:

$$X : \left\{ \begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right.$$

- Várható értéke: $E(X) = 0 * (1-p) + 1 * p = p$
- Ugyanakkor: $E(X^2) = E(X) = p$
- Tehát: $D(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{p - p^2} = \sqrt{p(1-p)}$

Binomiális eloszlás

Egy adott számú (*n*) független kísérlet elvégzése esetén a megfigyelt *p* valószínűségű *A* esemény összesen hányszor következik be.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad k = 0,1,2,\ldots,n\,.$$

$$E(X) = n \cdot p \qquad D(X) = \sqrt{np(1-p)}\,.$$

Hipergeometriai eloszlás

Adott *m* elemből, melyből *s* db elem megkülönböztethető a maradék *m* – *s* darabtól (selejtes – nem selejtes) veszünk *n* darabot visszatevés nélkül. Ekkor az X valvál. értéke az *n* kiválasztott elemből a *k* darab megkülönböztetett elemek száma. (=> konkrétan megegyezik a visszatevés nélküli mintavétellel)

$$P(X = k) = \frac{\binom{s}{k} \cdot \binom{m-s}{n-k}}{\binom{m}{n}} \qquad k = 0,1,2,\ldots,n\,.$$

$$E(X) = n \cdot \frac{s}{m} = n \cdot p \qquad D(X) = \sqrt{np(1-p)\left(1 - \frac{n-1}{m-1}\right)}\,.$$

Geometriai eloszlás

Egy kísérlet ismételt végrehajtása addig, amíg egy adott esemény be nem következik. A szükséges kísérletek száma *k = 1, 2, …*, ahol felső korlát nincsen.

$$P(X = k) = (1-p)^{k-1} \cdot p \qquad k = 1,2,.$$

$$E(X) = \frac{1}{p} \qquad D(X) = \sqrt{\frac{1-p}{p^2}}.$$

Poisson-eloszlás

Annak meghatározására használjuk, hogy sok, független, kis valószínűségű esemény közül hány következik be, még akkor is, ha az egyes események valószínűségei különbözőek.

Például:

- Adott időintervallumon belül a bekövetkezések száma (1 óra alatt telefonálók száma)
- Adott térrészben történő bekövetkezések száma (sajtóhibák száma egy oldalon)
- Adott számú berendezés esetén adott idő alatti meghibásodások száma

X valvál. Poisson-eloszlású a $\lambda > 0$ paraméterrel, ha lehetséges értékei a nemnegatív egész számok (0,1, . . .), az eloszlása pedig:

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$$E(X) = \lambda \qquad D(X) = \sqrt{\lambda}.$$

6. tétel: Folytonos valószínűségi változó

A folytonos valószínűségi változó eloszlása. Eloszlásfüggvény, sűrűségfüggvény.
Módusz, medián, várható érték, momentum, szórás. A várható érték és a szórás tulajdonságai.

Eloszlásfüggvény

Az X valvál. eloszlásfüggvénye az az F függvény, amely minden valós x értékhez hozzárendeli annak a valószínűségét, hogy az X valvál. annál az x-nél kisebb értéket vesz fel.

Folytonos valvál.: Egy valószínűségi változót folytonosnak nevezünk, ha eloszlásfüggvénye abszolút folytonos függvény. (=> nem tudjuk az összes lehetséges értéket felsorolni)

Az eloszlásfüggvény tulajdonságai

1. $D_F = \mathbb{R}$, $R_F \subseteq [0; 1]$.
2. F monoton növő, azaz ha $x_1 < x_2$, akkor $F(x_1) \leqslant F(x_2)$.
3. F legfeljebb $\aleph_0$ sok[a] hely kivételével mindenütt folytonos; de ezeken a helyeken is balról folytonos.
4. $\lim\limits_{x \to -\infty} F(x) = 0$ és $\lim\limits_{x \to \infty} F(x) = 1$.

Valószínűségek számítása az eloszlásfüggvénnyel:

- $P(X < a) = F(a)$
- $P(X \geq a) = 1 - F(a)$
- $P(a \leq X \leq b) = F(b) - F(a)$

Sűrűségfüggvény

A folytonos eloszlású valószínűségi változó sűrűségfüggvénye az eloszlásfüggvényének deriváltja

Azaz: $f(x) = F'(x)$ $\quad F(x) = \int_{-\infty}^{x} f(t)dt$

A sűrűségfüggvény tulajdonságai

1. $R_f \subseteq \mathbb{R}_0^+$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$

(A 2. tulajdonságból következően: $\lim\limits_{x \to -\infty} f(x) = \lim\limits_{x \to \infty} f(x) = 0$, ha a határértékek léteznek.)

3. f(x) >= 0

Valószínűségek számítása a sűrűségfüggvénnyel:

- $P(X < a) = \int_{-\infty}^{a} f(x)\,dx$
- $P(X \geq a) = \int_{a}^{\infty} f(x)\,dx$
- $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$

Középértékek

Módusz

Az eloszlás módusza a sűrűségfüggvény maximumhelye. ($mod(X) = f'(x) = 0$)

Medián

Az eloszlás mediánja az a hely, ahol az eloszlásfüggvény értéke ½, ha ez egyértelmű. Különben az a hely, ahol ezt az ½-et “átugorja”. Ha több ilyen van, akkor azok átlaga adja meg a mediánt.

Várható érték

A folytonos eloszlású X valószínűségi változó várható értéke:

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx,$$

(ha az integrál abszolút konvergens).

K-adik momentum

Az X valvál k-adik momentuma X^k várható értéke, azaz:

$$m_k = E\left(X^k\right) = \int_{-\infty}^{\infty} x^k * f(x)\,dx$$

Szórás

X valvál. szórásnégyzete a várható értéktől való négyzetes eltérés várható értéke, azaz:

$\mathbf{D}^2(X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right)$ Ennek gyöke a szórás.

Egyszerűbb alakra hozva: $\mathbf{D}^2(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$

	Diszkrét	Folytonos
valószínűség	p_i	$f(x)$
$\mathbf{P}(a < X < b) =$	$\sum_{a<x_i<b} p_i$	$\int_a^b f(x)\,dx$
$1 =$	$\sum_i p_i$	$\int_{-\infty}^{\infty} f(x)\,dx$
$F(x) =$	$\sum_{x_i<x} p_i$	$\int_{-\infty}^{x} f(t)\,dt$
$\mathbf{E}(X) =$	$\sum_i p_i x_i$	$\int_{-\infty}^{\infty} x f(x)\,dx$
$m_k = \mathbf{E}(X^k) =$	$\sum_i p_i x_i^k$	$\int_{-\infty}^{\infty} x^k f(x)\,dx$

7. tétel: Nevezetes folytonos eloszlások

Egyenletes eloszlás, exponenciális eloszlás, normális eloszlás.

Egyenletes eloszlás

Tegyük fel, hogy egy valvál. az (a;b) intervallumon veheti fel értékeit, és ha ezen belül veszünk egy (c;d) részintervallumot, akkor annak a valószínűsége, hogy a valvál. ezen részintervallumba essen, a (c;d) hosszával arányos, annak (a;b)-beli elhelyezkedésétől független. Ebben az esetben mondjuk, hogy a valvál. egyenletes eloszlású az (a;b) intervallumon…

… ha, sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{ha } a < x < b; \\ 0 & \text{különben.} \end{cases}$$

… eloszlásfüggvénye pedig:

$$F(x) = \begin{cases} 0, & \text{ha } x \le a; \\ \dfrac{x-a}{b-a}, & \text{ha } a < x \le b; \\ 1, & \text{ha } x > b. \end{cases}$$

$$E(X) = \frac{a+b}{2}, \qquad D(X) = \frac{b-a}{\sqrt{12}}.$$

Exponenciális eloszlás

Exponenciális eloszlással modellezhető például a várakozási/sorbaállási idővel, bizonyos alkatrészek élettartalmával is.

Az X valvál. $\lambda > 0$ paraméterű exponenciális eloszlású, ha sűrűségfüggvénye:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, & \text{ha } x > 0; \\ 0 & \text{különben.} \end{cases}$$

… ezáltal eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - e^{-\lambda \cdot x}, & \text{ha } x > 0; \\ 0, & \text{ha } x \le 0. \end{cases}$$

$$E(X) = D(X) = \frac{1}{\lambda}.$$

Az exponenciális eloszlás örökifjú tulajdonságú. Azaz, ha pl. egy alkatrész élettartama exponenciális eloszlású, akkor az alkatrész bizonyos értelemben véve nem öregszik. Ennek matematikai jelentése az, hogy ha az alkatrész már működött *t* ideig, akkor annak valószínűsége, hogy még *s* ideig működik, nem függ a megelőző *t* időtartamtól, valószínűsége ugyan akkora, mint annak valószínűsége hogy a kezdettől számítva *s* ideig működni fog.

Tehát az exponenciális eloszlás örökifjú tulajdonságú, azaz ha X exp. elol.-ú valvál., akkor:

$$P(X \geq t + s | X \geq t) = P(X \geq s)\,.$$

Valószínűségek egyenletes és exponenciális esetén:

$$P(X < a) = F(a)$$

$$P(X \geq a) = 1 - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a)$$

Normális eloszlás

Általában, ha egy jelenséget nagy számú, egymástól független, vagy kevéssé függő, véletlenszerű tényező határoz meg, melyek egyenként csak kis mértékben járulnak hozzá a véletlentől függő ingadozáshoz, hatásaik pedig összegződnek, akkor a jelenség jól közelíthető normális eloszlással.

X folytonos eloszlású valvál., *m*, σ ($\sigma > 0$) paraméterű normális eloszlású, ha sűrűségfüggvénye:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} .$$

… eloszlásfüggvénye pedig:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} \, \mathrm{d}t .$$

$$E(X) = m, \qquad D(X) = \sigma .$$

(ezen fentebbi rész nem kell)

Ami ebből lényeges, hogy azon normális eloszlások között kitüntetett szerep jut annak, amelynek várható értéke 0, szórása pedig 1, azaz *m* = 0, σ = 1.Ebből jön a standard normális eloszlás, mely esetében:

Legyen $\Phi(x)$ a standard normális eloszlás eloszlásfüggvénye. Ekkor:

$$\Phi(-x) = 1 - \Phi(x) .$$

10.5. tétel: Legyen X normális eloszlású valószínűségi változó, a várható értéke m, a szórása σ. Ekkor az

$$X^* = \frac{X - m}{\sigma}$$

valószínűségi változó standard normális eloszlású.

8. tétel: Egyenlőtlenségek

Markov-egyenlőtlenség. Csebisev-egyenlőtlenség.
A nagy számok törvényeinek egyenlőtlenség alakja.

Markov-egyenlőtlenség

Ismeretlen eloszlású, nem-negatív valvál. esetében használjuk akkor, ha felső becslést adunk arra, hogy a valvál. valószínűsége nagyobb-egyenlő egy adott számnál. Két feltétele, hogy nem-negatív legyen a valvál, illetve, hogy ismerjük a várható értéket.

A valószínűségi változó csak viszonylag kis valószínűséggel vehet fel a várható értékhez képest nagy értékeket.

Ha X egy olyan nem-negatív értékeket felvevő valvál, melynek van várható értéke, és *a* egy tetszőleges, pozitív valós szám, akkor:

$$P\left(X \geq a\right) \leq \frac{E(X)}{a}.$$

A becslés használható az $X > a$ szigorú egyenlőtlenségre is, hiszen:

$$P(X > a) \leq P(X \geq a) \leq \frac{E(X)}{a}.$$

A Markov-egyenlőtlenség sajátossága, hogy nem alkalmazható olyan esetben, amikor arra akarunk becslést adni, hogy X értéke az adott értéknél kisebb – csak arra tudunk, hogy nagyobb-egyenlő. Ekkor kiszámoljuk az egyenlőtlenséget az $X > a$ módszerrel, majd az eredményt 100%-ból / 1-ből kivonva megkapjuk a megoldást az $X < a$ kérdésre is.

$$P\left(X < a\right) \geq 1 - \frac{E(X)}{a}.$$

Csebisev-egyenlőtlenség

Ismeretlen eloszlású valvál. esetén használjuk, ahol a Csebisev-egyenlőtlenség azt mondja ki, hogy a várható értéktől való eltérés nem lehet túl nagy. Ez már használható negatív valvál. esetén is, és ezt akkor használjuk alapból, ha a valvál.-nak ismert nem csak a várható értéke, de a szórása is.

Legyen X valvál., melynek van várható értéke és szórása, illetve legyen λ egy tetszőleges pozitív szám, ekkor:

$$P\left(|X - E(X)| \geq \lambda \cdot D(X)\right) \leq \frac{1}{\lambda^2}.$$

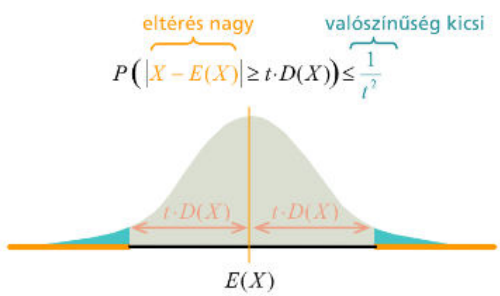
nem túl valószínű, hogy a szórás λ-szorosánál nagyobb legyen a várható értéktől való eltérés

A Csebisev-egyenlőtlenséget gyakran alkalmazzuk annak becslésére, hogy az X valvál. értéke milyen valószínűséggel esik egy adott, várható érték körüli szimmetrikus intervallumba. Ez konkrétan az alap Csebisev érték komplementere, azaz:

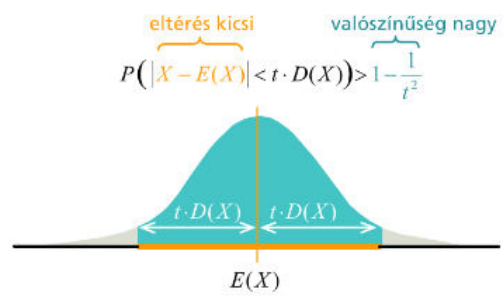
$$P\left(|X - E(X)| < \lambda \cdot D(X)\right) \geq 1 - \frac{1}{\lambda^2}.$$

várható értéktől való eltérés

sokkal valószínűbb, hogy ez az eltérés kisebb, mint a szórás λ-szorosa



alap Csebisev



fordított Csebisev

Példák:

1. Újságárus óránként átlag 64 db újságot ad el, a szórás pedig 8 db. Adjunk becslést annak valószínűségére, hogy az általa ladott lapok száma 50 db és 78 db közé esik.

Mivel a várható értékhez közelebb vagyunk, így a második Csebisev-egyenlőtlenséget használjuk fel, azaz:

$$E(X) = 64 \qquad D(X) = 8$$

$$Az\ eltérés\ 78 - 54 = 64 - 50 = 14$$

$$P\big(\,|X - E(X)| < \lambda\, * D(X)\big) > 1 - \frac{1}{\lambda^2}$$

$$\lambda * D(X) = 14 \to \lambda = \frac{14}{8}$$

$$\ldots \to 1 - \frac{1}{\lambda^2} = \mathbf{0,67}$$

2. Üzemben 150 mm hosszú csavarokat gyártanak, 2 mm szórással. Egy csavar selejtes, ha 146 mm-nél rövidebb, vagy 154 mm-nél hosszabb. Adjunk becslést a selejtarányra.

Mivel most a várható értéktől távolabb vizsgálódunk, így az első Csebisev-egyenlőtlenség fog kelleni, azaz:

$$E(X) = 64 \qquad D(X) = 8$$

$$Az\ eltérés\ 150 - 146 = 154 - 150 = 4$$

$$P\big(\,|X - E(X)| \geq \lambda\, * D(X)\big) \leq \frac{1}{\lambda^2}$$

$$\lambda * D(X) = 4 \to \lambda = 2$$

$$\ldots \to \frac{1}{\lambda^2} = \mathbf{0,25}$$

Plusz: Ha adódna olyan feladat (kétlem, hogy előfordulna), ahol az intervallum nem szimmetrikus, a várható érték körüli két irányba két különböző értékkel tér el, akkor kérdés, mit írunk a $\lambda * D(X)$ helyére. Ekkor először meghatározzuk, melyik egyenlőtlenséget kell használni, hogy várható értékhez közelebbi értéket számolunk, vagy selejtarányosat. Ezelőbbi esetben nyilván a második egyenlőtlenséget használjuk, ekkor **alsó** becslést adunk, így a két különböző mértékű eltérés közül a **kisebbel** számolunk, ha pedig selejtarányosat számolunk, akkor **felső** becslést adunk, tehát a **nagyobbal** számolunk.

A nagy számok törvényeinek egyenlőtlenség alakja

A NSzT segítségével meg tudjuk mondani, hogy hányszor kell elvégeznünk egy kísérletet ahhoz, hogy valamely esemény relatív gyakorisága annak valószínűségéhez nagyon közel kerüljön.

Például, ha egy dobókockával a hatos dobás valószínűsége 1/6, akkor nézzük meg, hányszor kell dobnunk a kockával ahhoz, hogy a hatos dobás valószínűségét a relatív gyakoriság 0,1-nél jobban megközelítse az esetek 95%-ban. Ez elég kusza, szóval vegyük szépen sorba..

A képletek, melyek közül a szöveg alapján könnyen meghatározható, melyik kell (kisebb – nagyobb szó):

Nagy Számok Törvénye

$$P\left(\left|\frac{X}{n}-p\right|<\varepsilon\right)\geq 1-\frac{p(1-p)}{n\varepsilon^2}$$

$$P\left(\left|\frac{X}{n}-p\right|>\varepsilon\right)<\frac{p(1-p)}{n\varepsilon^2}$$

A képletekben az $\frac{X}{n}$ a relatív gyakoriság, a *p* az elméleti valószínűség, *n* a kísérletek száma, ε pedig a "hibahatár", melyen belül (első -), vagy kívül kell vizsgálódnunk (második egyenlőtlenség). Így a feladatot újra megvizsgálva:

$$p=\frac{1}{6}\qquad \varepsilon=0{,}1$$

$$ún.\, megbízhatósági\ szint=0{,}95$$

$$P\left(\left|\frac{X}{n}-\frac{1}{6}\right|<0{,}1\right)\geq \underbrace{1-\frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n\cdot 0{,}1^2}}$$

$$0{,}95=1-\frac{\frac{5}{36}}{n\cdot 0{,}1^2}$$

$$n=\mathbf{277{,}78}$$

… azaz legalább 278-szor kell dobnunk a kockával ahhoz, hogy a hatos dobás valószínűségét a relatív gyakoriság 0,1-nél jobban megközelítse az esetek 95%-ban.

(kétlem, hogy adnának ilyen feladatot, de elméleten akár lehet, ha már a tételsorban benne van)

9. tétel: Határeloszlás tételek

A hipergeometriai eloszlás közelítése binomiális eloszlással. A binomiális eloszlás közelítése Poisson-eloszlással. A nagy számok törvénye (NSzT) az átlagra és a relatív gyakoriságra. A centrális határeloszlás tétel (CHT).

Előfordulhat, hogy egy véletlen jelenséggel kapcsolatos valószínűség meghatározásakor vagy nem ismerjük pontosan a jelenséget leíró valószínűségi eloszlást, vagy ismerjük ugyan, de még ennek a segítségével is csak nehezen tudjuk kiszámolni a valószínűség értékét. Ilyenkor megpróbálhatunk olyan egyszerűbb (kevesebb paramétert tartalmazó, vagy könnyebben számolható) eloszlást keresni, mellyel a kérdéses valószínűség egy jó közelítését kapjuk.

Hipergeometriai eloszlás közelítése binomiális eloszlással

Legyen *p* egy 0 és 1 közé eső szám, legyen továbbá s_m olyan egész értékű sorozat, melyre:

$$\log_{m\to\infty} \frac{s_m}{m} = p$$

Legyen *n* olyan *m*-től és s_m-től független egész állandó, melyre $n \leq s_m$és $n \leq m - s_m$. Ekkor tetszőleges rögzített *k* esetén $(0 \leq k \leq n)$:

$$\log_{m\to\infty} \frac{\binom{s_m}{k} * \binom{m-s_m}{n-k}}{\binom{m}{n}} = \binom{n}{k} * p^k * (1-p)^{n-k}$$

Tehát az *m, s_m* és *n* paraméterű hipergeometriai eloszlások sorozata az *n* és *p* paraméterű binomiális eloszláshoz közelít.

Azaz van *m* termékünk, melyből *s*-et megkülönböztetünk, kiválasztunk belőlük (visszatevés nélkül) *n* darabot, melyből majd *k* a kiválasztott selejtesek száma. Ha a kiválasztott *n* száma mind az *s* (selejtes) termékek számához, mind az *m-s* (hibátlan) termékek számához képest "kicsi", **akkor** jól közelíthető a hipergeom.-i eloszlás binomiálissal.

Binomiális eloszlás közelítése Poisson eloszlással

Legyen λ egy pozitív szám, illetve p_n (*n* = 1, 2, …) 0 és 1 közötti számok olyan sorozata, melynél

$$\lim_{n\to\infty} p_n * n = \lambda$$

Ekkor tetszőleges *k* esetén az *n* és p_n paraméterű binomiális eloszlások sorozata λ paraméterű Poisson eloszláshoz közelít.

$$\lim_{n\to\infty} \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Azaz akkor közelíthető binomiális Poissonnal, ha a *p* elég kicsi és *n* elég nagy. "Nagy elemszámú minta a kis selejtarányú sokaságból."

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \frac{(n\cdot p)^k}{k!} \cdot e^{-n\cdot p}.$$

NSzT az átlagra

A Csebisev-egyenlőtlenség segítségével becslést adunk arra, hogy független, azonos várható értékű és szórású valószínűségi változók átlaga a közös várható értéktől milyen mértékben térhet el. Az átlag túlságosan nem térhet el a várható értéktől, ezen nagy eltérés valószínűsége kicsi.

Legyenek $X_1, X_2, \ldots, X_i$ ($i = 1, 2, \ldots, n$) független valváltozók, azonos várható értékkel és szórással, azaz $E(X_i) = m$ és $D(X_i) = \sigma$. Ekkor:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n}.$$

Természetesen létezik ennek is komplementere, melyet akkor használunk, ha azt szeretnénk megvizsgálni, hogy a valváltozók milyen valószínűséggel közelítik meg a várható érték bizonyos környezetét:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - m\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2 \cdot n}.$$

(ε az éppen vizsált eltérés mértéke)

NSzT a relatív gyakoriságra (Bernoulli-féle törvény)

Adott számú független kísérlet esetén a relatív gyakoriságnak a valószínűségtől való adott eltérésének valószínűségére ad becslést. Kicsit pontatlanul megfogalmazva azt mondja ki a tétel, hogy a relatív gyakoriság nagyon nem térhet el az elméleti valószínűség értékétől, olyan értelemben, hogy a nagy eltérés valószínűsége kicsi.

Végezzünk el n darab független kísérletet a p valószínűségű A esemény megfigyelésére. Tegyük fel, hogy a kísérletek során az A esemény k-szor következett be. Ekkor tetszőleges $\varepsilon > 0$ esetén:

$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n}.$$

Komplementere esetén annak becslésére használjuk, hogy a relatív gyakoriság milyen valószínűséggel közelíti meg az adott A esemény p valószínűségét:

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n}.$$

(ε az éppen vizsált eltérés mértéke)

Vagyis ismeretlen p esetén az alábbi alakokban használhatjuk a tételt:

$$P\left(\left|\frac{k}{n}-p\right| \geq \varepsilon\right) \leq \frac{1}{4 \cdot \varepsilon^2 \cdot n},$$
$$P\left(\left|\frac{k}{n}-p\right| < \varepsilon\right) \geq 1-\frac{1}{4 \cdot \varepsilon^2 \cdot n}.$$

A centrális határeloszlás tétel (CHT)

Legyenek $X_1, X_2, \ldots, X_i$ ($i = 1, 2, \ldots, n$) független, azonos eloszlású valváltozók, létező és azonos várható értékkel és szórással, azaz $E(X_i) = m$ és $D(X_i) = \sigma$ mindegyik X_i -re. Ebben az esetben az X_i valváltozók összegének standardizáltja határesetben (azaz ha $n \to \infty$) standard normális eloszlású. Ez azt jelenti, hogy:

$$\lim_{n\to\infty} P\left(\frac{X_1+X_2+\ldots+X_n-n\cdot m}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, \mathrm{d}t = \Phi(x)\,.$$

Használatának legfontosabb feltételei a függetlenség, és a szórások végessége.

Érthetőbben: A centrális határeloszlás-tétel (CHT) azt mondja ki, hogy adott feltételek mellett, elegendően nagy számú és független valószínűségi változó várható értéke jó közelítéssel normális eloszlású, ha a független valószínűségi változók jól meghatározott várható érték és szórásnégyzettel rendelkeznek.

CHT az összegre:

$$X = X_1 + X_2 + \ldots + X_n \quad \Rightarrow \quad X \approx \mathrm{N}(\,nm\,;\,\sigma\sqrt{n}\,)$$

CHT az átlagra:

$$\overline{X} = \frac{X_1+X_2+\ldots+X_n}{n} \quad \Rightarrow \quad \overline{X} \approx \mathrm{N}\left(m\,;\,\frac{\sigma}{\sqrt{n}}\right)$$

10. tétel: Statisztika

Statisztikai minta, statisztika(i függvény).
Tapasztalati paraméterek: átlag, módusz, medián, szórásnégyzet(ek), szórás(ok).
A statisztikában használatos további eloszlások: a χ2-eloszlás, a Student-féle t-eloszlás

Statisztikai minta

A megfigyelt X valószínűségi változóval azonos eloszlású, független, $X_1, X_2, \ldots, X_n$ valószínűségi változók (mintavételi változók vagy mintaelemek) összességét nevezzük statisztikai mintának.

Ha az $X_1, X_2, \ldots, X_n$ valószínűségi változók nagyság szerint növekvő értékei közül az *i*-ediket jelöljük X_i^*-vel ($i = 1, 2, \ldots, n$), akkor az $X_1^*, X_2^*, \ldots, X_n^*$ statisztikai mintát rendezett mintának nevezzük. Ekkor

$$X_1^* \leq X_2^* \leq \cdots \leq X_n^*$$

Statisztikai függvény

Az $X_1, X_2, \ldots, X_n$ valószínűségi változókat a valós számok halmazára képező

$$\hat{\alpha}_n : \mathbb{R}^n \to \mathbb{R}$$

függvényt statisztikai függvénynek (statisztikának) nevezzük.

Tapasztalati paraméterek

Átlag

Az $X_1, X_2, \ldots, X_n$ mintaelemek mintaátlaga (empirikus közepe) az

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

valószínűségi változó.

Ezzel becsülhető a megfigyelt *X* valvál. $E(X) = m$ várható értéke.

Módusz

A módusz a statisztikai minta leggyakrabban előforduló mintaeleme.

Medián

Az $X_1^*, X_2^*, \ldots, X_n^*$ rendezett minta mediánja

$$\begin{cases} \dfrac{X_k^* + X_{k+1}^*}{2} & \text{ha } n = 2 \cdot k, \\ X_{k+1}^* & \text{ha } n = 2 \cdot k + 1. \end{cases}$$

középső elem és jobboldali szomszédjána átlaga, ha páros

középső elem jobboldali szomszédja, ha páratlan

Szórásnégyzet(ek)

Az $X_1, X_2, \ldots, X_n$ mintaelemek tapasztalati (empirikus) szórásnégyzetén a mintaelemek mintaátlagától való eltéréseinek négyzetes közepét értjük, azaz:

$$\hat{\sigma}_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \hat{m}_n)^2 = \frac{(X_1 - \hat{m}_n)^2 + (X_2 - \hat{m}_n)^2 + \ldots + (X_n - \hat{m}_n)^2}{n}$$

(képlet szinte fix, hogy nem kell lmao)

Mivel a fentebbi képlet torzított becslést ad, létezik korrigált szórásnégyzet, mely pontosabb becslést képet adni, ahol *n* helyett *n-1*-el osztunk le.

Szórás(ok)

?

χ2-eloszlás

Legyenek $X_1, X_2, \ldots, X_n$ **független**, standard normális eloszlású valószínűségi változók. Ekkor a belőlük képzett

$$\chi_n^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

$$E(\chi_n^2) = n \qquad \text{és} \qquad D^2(\chi_n^2) = 2n\,.$$

valószínűségi változó eloszlását *n* szabadsági fokú χ^2(khí négyzet)-eloszlásnak nevezzük. Lényegében *n* független, SN eloszlású valvál négyzetösszege.

Itt általában nem arra vagyunnk kíváncsiak, hogy az eloszlásfüggvény adott helyen milyen értéket vesz fel, hanem arra, hogy egy adott értéket hol vesz fel (az eloszlásfüggvény inverzének értéke).

Használhatjuk normális eloszlás szórására adott konfidencia-intervallumnál, illetve nem paraméteres próbáknál.

Student-féle t-eloszlás

Legyenek Y és Az X_1, X_2, …, X_n **független**, SN eloszlású valváltozók. Ekkor a belőlük képzett

$$t_n = \frac{Y}{\sqrt{\dfrac{X_1^2 + X_2^2 + \ldots + X_n^2}{n}}}$$

$$E(t_n) = \begin{cases} 0 & \text{ha } n \geq 2, \\ \text{nem létezik} & \text{ha } n = 1. \end{cases}$$

$$D^2(t_n) = \begin{cases} \dfrac{n}{n-2} & \text{ha } n \geq 3, \\ \text{nem létezik} & \text{ha } n = 1, 2. \end{cases}$$

valószínűségi változó eloszlását n szabadsági fokú Student-féle (t-eloszlásnak) nevezzük.

A statisztikában használt t-próbát adja. Szintén az érdekel minket, hogy egy adott értéket **hol** vesz fel az eloszlásfüggvény (azaz az eloszlásfüggvény inverzének értéke).

11. tétel: Becslések

Pontbecslés. A jó becslés kritériumai: torzítatlan, konzisztens, efficiens. Intervallumbecslés.
Konfidenciaintervallum normális eloszlás várható értékére ismert szórás esetén.
Konfidenciaintervallum normális eloszlás várható értékére ismeretlen szórás esetén.
Konfidenciaintervallum normális eloszlás szórására.

Pontbecslés

Ebben az esetben a valvál. értékét a mintából számított értékkel becsüljük. Ilyen módszer például a legnagyobb valószínűség elve, amikor azt keressük, hogy az eloszlás mely paraméterei mellett kapható meg a kísérlet során a legnagyobb valószínűséggel az adott minta.

A jó becslés kritériumai

<u>**Torzítatlan:**</u> Az $\alpha_n = \alpha_n(X_1; X_2; \ldots ; X_n)$ statisztika az *a* paraméter torzítatlan becslése, amennyiben várható értéke megegyezik a becsült paraméter értékével, azaz ha:

$$\mathbf{E}(\alpha_n) = a.$$

Ellenkező esetben a becslés torzított.

Konzisztens: Az $\alpha_n = \alpha_n(X_1; X_2; \ldots ; X_n)$ statisztika az *a* paraméter konzisztens becslése, ha a valváltozók becsléssorozata sztochatikusan konvergál az *a* paraméterhez, azaz ha:

$$\lim_{n\to\infty} \mathbf{P}(|\alpha_n - a| > \varepsilon) = 0.$$

Efficiens: Ha az *a* paraméternek van olyan α_n^* torzítatlan becslése, amelynek az α_n torzítatlan becslések közül minimális a szórásnégyzete, azaz:

$$\mathbf{D}^2(\alpha_n^*) \leqslant \mathbf{D}^2(\alpha_n),$$

akkor az α_n^* statisztika az *a* paraméter efficiens (hatásos) becslése.

180	176	164	167	170	174	172	178	186	181

- Adjunk becslést a 180 cm-nél magasabb diákok arányára: $\hat{p} = \frac{\#\{x_i > 180\}}{n} = \frac{2}{10} = 0{,}2.$
- Adjunk becslést a testmagasság várható értékére: $\hat{m} = \bar{x} = \frac{\sum x_i}{n} = \frac{1748}{10} = 174{,}8.$
- Adjunk becslést a testmagasság szórására: $S^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{305\,962}{10} - 174{,}8^2 = 41{,}16$
  $S^{*2} = \frac{n}{n-1} S^2 = \frac{10}{9} \cdot 41{,}16 \approx 45{,}733 \quad \Rightarrow \quad \hat{\sigma} = S^* \approx 6{,}763.$

Intervallumbecslés

Olyan intervallumot keresünk, amelybe előre adott, nagy valószínűséggel beleesik a becsült paraméter valódi értéke. A becsült paraméter értékére tehát egy intervallumot adunk meg, ezzel az adott mintából származó bizonytalanságot csökkentjük.

Konfidenciaintervallum: Ha a becsült α paraméter 1-ε valószínűséggel beleesik a (c_1; c_2) intervallumba.

23.12. definíció: A (c_1, c_2) intervallumot az X valószínűségi változó α paraméterére vonatkozó $(1-\varepsilon)\cdot 100\%$ megbízhatósági szintű konfidencia-intervalumnak nevezzük, ha

$$P(c_1 < \alpha < c_2) = 1 - \varepsilon$$

teljesül.

Ebből következi annak a valószínűsége, hogy a becsült paraméter valódi értéke az intervallumon kívül essen:

$$\mathbf{P}(a < c_1) = \mathbf{P}(a > c_2) = \frac{\varepsilon}{2}.$$

Konfidenciaintervallum normális eloszlás várható értékére ismert szórás esetén

$$\mathbf{P}\left(-u_\varepsilon < \frac{\bar{X} - m}{\sigma}\sqrt{n} < u_\varepsilon\right) = 1 - \varepsilon$$

Konfidenciaintervallum normális eloszlás várható értékére ismeretlen szórás esetén

$$\mathbf{P}\left(-t_\varepsilon < \frac{\bar{X} - m}{S^*}\sqrt{n} < t_\varepsilon\right) = 1 - \varepsilon$$

Konfidenciaintervallum normális eloszlás szórására

$$\mathbf{P}\left(\chi^2_{1-\frac{\varepsilon}{2}} < \frac{(n-1)\,S^{*2}}{\sigma^2} < \chi^2_{\frac{\varepsilon}{2}}\right) = 1 - \varepsilon$$

(komolyan mondom, ebből bármit számonkérnek vizsgán én megeszem a kalapom)

12. tétel: Hipotézisvizsgálat

Statisztikai hipotézisek. A hipotézisvizsgálat menete.
Szignifikanciaszint. Egy- és kétoldali próba. Első- és másodfajú hiba.

Statisztikai hipotézis

Statisztikai hipotézisnek egy vagy több valószínűségeloszlásra vonatkozó felvetést nevezünk. A statisztikai próba pedig maga az eljárás, mellyel a hipotézist elfogadjuk, vagy elutasítjuk.

Statisztikai próba / hipotézisvizsgálat menete

A statisztikai próbák menete minden hipotézivizsgálat során ugyan azok, az egyes próbák csak technikai részletekben térnek el egymástól.

1. H_0 (null-) és H_1 (ellenhipotézis) felállítása (a **nullhipotézis** tartalmazza az **egyenlőséget**, az **ellenhipotézis** pedig az **egyenlőtlenséget**, így ezek egymást kizárják). Általában úgy fogalmazzuk meg ezeket, hogy a nullhipotézis elvetése legyen fontos.
2. Próbafüggvény kiválasztása és értékének meghatározása.
3. A kritikus vagy elutasítási tartomány kijelölése a próbafüggvény eloszlása és szignifikanciaszint ismeretében.
4. Döntés a H_0 elfogadásáról vagy elutasításáról.

Hibalehetőségek:

- Elsőfajú hiba: a H_0 hipotézis igaz, de elutasítjuk
    - ε valószínűséggel lehet ilyen hiba, szóval minél nagyobb az (1 – ε), azaz a szignifikanciaszint, annál kisebb az elsőfajú hiba esélye
        - szignifikanciaszint szemléletes jelentése: valószínűséget ad arra, hogy a mért eredményt nem a véletlen okozta
- Másodfajú hiba: H_0 hipotézis hamis, de elfogadjuk
    - minél több mintával dolgozunk, annál kisebb az esélye
    - ha csökkentjük az elsőfajú hibát, akkor nő a másodfajú hiba esélye

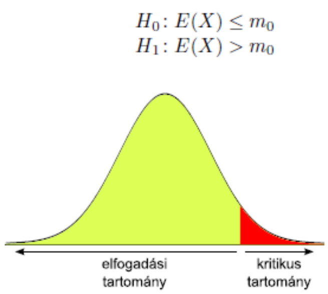
	H_0 igaz	H_0 hamis
H_0-t elfogadjuk	Jó döntés	Másodfajú hiba
H_0-t elutasítjuk	Elsőfajú hiba	Jó döntés

Egy- és kétoldali próba

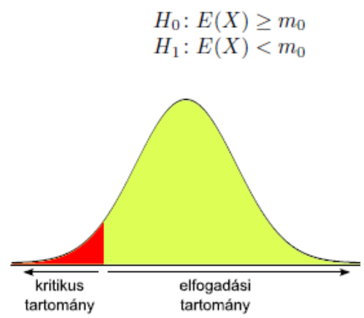
Az ellenhipotézistől függően a kritikus tartomány kijelölésében ez a két eset különböztethető meg.

Egyoldali próba: Az ellenhipotézis a feltételezett értéktől csak egyirányú eltérést tartalmaz. Vegyük észre, hogy ekkor a kritikus tartomány kijelölése az ellenhipotézis egyenlőtlenségének irányától függ.

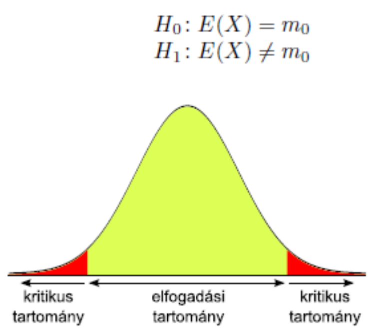
Jobb oldali próba:



Bal oldali próba:



Kétoldali próba: Az ellenhipotézis a feltételezett értéktől bármilyen irányú eltérést tartalmaz.



13. tétel: Paraméteres próbák

A várható értékre vonatkozó próbák:
egymintás u- és t-próba, kétmintás u- és t-próba.

Paraméteres próbák esetében a H_0 hipotézis és a H_1 ellenhipotézis a megfigyelt valószínűségi változó valamely paraméterére vonatkozik.

Várható értékre vonatkozü egymintás próbák

Ismeretlen várható értékű, normális eloszlású valószínűségi változó várható értékére vonatkozó $H_0 : E(X) = m_0$ hipotézis helyességét akarjuk ellenőrizni.

Egymintás u-próba

1. Amikor ismert a szórás: a próbastatisztika standard normális eloszlású a CHT miatt.
2. Amikor ismeretlen a szórás és nagy a minta elemszáma (n > 30): a próbastatisztika közel standard normális eloszlású a CHT miatt.

Egymintás t-próba

1. Ha a szórás nem ismert és a minta elemszáma kicsi: Próbastatisztika (n-1) szabadságfokú Student eloszlású.

Kétmintás u-próba

1. Amikor ismert a szórás: a próbastatisztika standard normális eloszlású a CHT miatt.
2. Azonos szórás, de a közös szórás nem ismert, és mintkét minta elemszáma nagy (n > 30): a próbastatisztika standard normális eloszlású

Kétmintés t-próba

1. A próbastatisztika (n1 + n2 − 2) szabadsági fokú Student-eloszlású.

	σ ismert	σ nem ismert
$n < 30$	u-próba: $u_p = \dfrac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$	t-próba: $t_p = \dfrac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$
$n \geq 30$	u-próba: $u_p = \dfrac{\hat{m}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$	u-próba: $u_p = \dfrac{\hat{m}_n - m_0}{\frac{\hat{s}_n}{\sqrt{n}}}$

$$\boxed{t = \frac{\bar{x} - m}{S^*}\sqrt{n}}$$

14. tétel: Nem-paraméteres próbák

Illeszkedésvizsgálat, homogenitásvizsgálat, függetlenségvizsgálat.

Nem paraméteres próbák esetében a H_0 hipotézis a megfigyelt valószínűségi változü ismeretlen eloszlására vonatkozik.

A χ^2 próba

Csak nagy elemszámú mintánál jó.

Illeszkedésvizsgálat

- Mire használjuk: **A minta származhat-e egy általunk kijelölt eloszlásból / Adott eloszlásra illeszkedik-e a minta.**
- Azaz: Azt vizsgáljuk, hogy a mintából származó gyakoriságok mennyire illeszkednek az eloszlásból származó becsült gyakoriságokra.
- Próbastatisztika eloszlára (r – 1) χ^2 eloszlás, ha H_0 fennáll.
- A próba mindig egyoldali, jobb oldali kritikus tartománnyal.

$$\chi_p^2 = \sum_{i=1}^{r} \frac{(\mu_i - n \cdot p_i)^2}{n \cdot p_i}$$

Homogenitásvizsgálat

- Mire használjuk: Ha azt akarjuk eldönteni, hogy **2 független minta azonos eloszlásból származik-e.**
- Próbastatisztika eloszlára (r – 1) χ^2 eloszlás, ha H_0 fennáll.
- A próba mindig egyoldali, jobb oldali kritikus tartománnyal.

$$\chi_p^2 = \frac{1}{m \cdot n} \sum_{i=1}^{r} \frac{(n \cdot \mu_i - m \cdot \nu_i)^2}{\mu_i + \nu_i}$$

Függetlenségvizsgálat

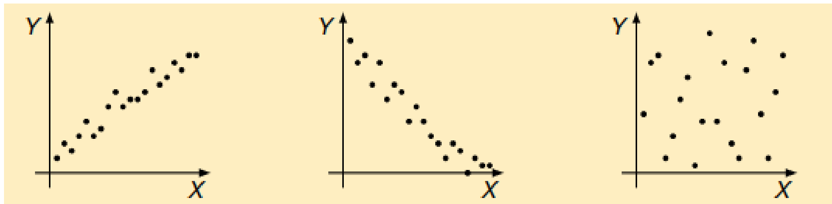
- Mire használjuk: **Két valószínűségi változó független-e egymástól** (X és Y)
- Próbastatisztika eloszlára (r – 1) χ^2 eloszlás, ha H_0 fennáll.
- A próba mindig egyoldali, jobb oldali kritikus tartománnyal.

$$\chi_p^2 = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n \cdot \mu_{ij} - \mu_{i*} \cdot \nu_{*j})^2}{\mu_{i*} \cdot \nu_{*j}}$$

15. tétel: Korreláció- és regresszióelemzés

Valószínűségi változók közötti sztochasztikus kapcsolat jellemzése, a korrelációs együttható.
A regressziós egyenes paramétereinek becslése, a legkisebb négyzetek módszere.

Valószínűségi változók közötti sztochasztikus kapcsolat jellemzése, a korrelációs együttható



1. Pozitív korrelációs együttható
2. Negatív korrelációs együttható
3. Korrelálatlanok

Sztochatikus kapcsolat: A valószínűségi változók között nincs függvényszerű kapcsolat, nem pontosan, de még így is elég jól meghatározzák egymást.

A korrelációs együttható: Azt írja le, mennyire függ össze a két valószínűségi változó.

1. Ha pozitív, akkor pozitív korreláció áll fenn
2. Ha negatív, akkor negatív korreláció áll fenn
3. Ha 0, akkor korrelálatlan

Két valószínűségi változó sztochasztikus kapcsolatát illetően legteljesebb információt a két változó együttes eloszlásából nyerünk. Ebből számítható a (lineáris) korrelációs együttható:

$$\varrho = \frac{\mathbf{E}\left[(X - \mathbf{E}(X))\,(Y - \mathbf{E}(Y))\right]}{\mathbf{D}(X)\,\mathbf{D}(Y)} = \frac{\mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y)}{\mathbf{D}(X)\,\mathbf{D}(Y)}.$$

Ha $0 < |\varrho| < 1$, akkor ez a sztochasztikus kapcsolat erősségét jellemző szám

A regressziós egyenes paramétereinek becslése, a legkisebb négyzetek módszere

Regresszió: A valószínűségi változók közti sztochatikus kapcsolat.

Regresziós egyenes: A gyakorlatban az (x,y) változópár együttes eloszlását nem ismerjük, így a rájuk vonatkozó kétdimenziós eloszlását sem ismerjük. Ezáltal saját vonatkoztatásból indulunk ki => Keressük azt a Y = ax + b egyenest, amely a legkisebb négyzetek elve értelmében a legjobban illeszkedik a pontfelhőhöz.

Röviden: A regressziós egyenes egy átfogó képet ad arról, hogy *Y* várhatóan hogyan változik *X* változásának hatására.

A legkisebb négyzetek módszere: A tényleges és becsült érték közötti négyzetes eltéréseket minimalizáljuk, és azt a paramétert választjuk, amelyik a legkisebb.