TÉMAVÁZLAT 8-11. ÓRA

Kémiai Számítástechnika Gyakorlat (1), Kémia BSc I. évf. 2016/2017. tanév I. félév

Valószínűségszámítási és statisztikai alapismeretek kémiai alkalmazásokkal

(összeállította: Tóth Gergely)

Ajánlott irodalom: Solt György: Valószínűségszámítás (Bolyai-sorozat, Műszaki Könyvkiadó)

Ami már volt: kombinatorikai előzetes

Alapok

Kísérlet: jelenség kb. azonos körülmények között, tetszőlegesen sokszor ismételhető → többféle kimenet. Kimenetelét eseménynek nevezzük (pontosabban elemi eseménynek), nagybetűvel jelöljük. Ha B mindig bekövetkezik, ha A bekövetkezik: $A \subset B$. Ha $A \subset B$ és $B \subset A$ elég csak az egyikről beszélni.

T = eseménytér = összes elemi esemény　　O = lehetetlen esemény

I = biztos esemény　　$\overline{A}$ = ellentet esemény

Példa: dobozban fekete és fehér golyók, 2 golyót húzunk. A=(fehér, fehér), B=(fehér, fekete),

C=(fekete, fehér), D=(fekete,fekete), $\overline{A}$ ={B, C, D}

Példák kísérletekre és elemi eseményekre:

- Pénzérme feldobása 1 alkalommal: {fej, írás}
- Egyszerre 4 kockával dobunk: {(1,1,1,1); (1,1,1,2); ……}
  előre rögzítsük, hogy számít-e a sorrend, tehát (1,1,1,2) azonos-e (2,1,1,1)-vel
- 3 egymást követő évben fagy-e (igen, nem, igen)
- Magasságmérés végtelenül pontosan: végtelen sok elemi esemény
- Magasságmérés cm-es pontossággal: 100-220 cm, 1 cm- es beosztással
- Radioaktív részecske, elbomlott-e a megfigyelési idő alatt {igen, nem} Adott időhöz rendelt!

Feladat: Osztálylétszám 40 fő, egy tárgyból az átlag 3,7. A*: az osztályban van 5-ös tanuló,* B*: pontosan 5 tanuló bukott meg. Igaz-e, hogy* $B \subset A$*?*

Műveletek eseményekkel

A+B　　legalább A vagy B bekövetkezik, A+B=B+A (kommutatív), A+(B+C)=(A+B)+C (asszociatív)

B-A　　B bekövetkezik, de A nem

AB　　mind a kettő bekövetkezik, AB=BA (kommutatív), A(BC)=(AB)C (asszociatív)

AB=O　　ha A és B egymást kizáró események

Igazak-e a következő egyenlőségek? A+A=A, AA=A, A+O=A, AO=O, AI=A, A+I=I, $A+\overline{A}$ =I, $A\overline{A}$ =O

Lássuk be Venn-diagram segítségével: disztributivitás A(B+C)=AB+AC *és* A+(BC)=(A+B)(A+C)

Összetett esemény: legalább két különböző elemi esemény összege

*Feladat: Egy telephelyre vasúton (*A*) és közúton(*B*) is szállíthatnak az adott napon. Mondja el, mit jelentenek a következő események:* A+B, AB, B-A, $\overline{A}$, $\overline{A}$+B, A$\overline{B}$, $\overline{A+B}$, $\overline{AB}$, $\overline{A}\ \overline{B}$, $A\overline{B}+\overline{A}B$, $\overline{A}+\overline{B}$, AB+$\overline{A}\ \overline{B}$, A+$\overline{A}$B

Esemény valószínűsége

ha n kísérletből k alkalommal következik be A $$P(A)=\lim_{n\to\infty}\frac{k}{n}$$

1) 0≤P(A)≤1
2) P(O)=0, P(I)=1
3) ha AB=O, akkor P(A+B)=P(A)+P(B)

Igazak-e?

ha A⊂B, P(A)≤P(B)

P(A+B)=P(A)+P(B)-P(AB)

P(A)+P($\overline{A}$)=1

Ezek alapján elemi események valószínűségéből összetett események valószínűsége számítható.

Geometriai analógiák, ha pl. n és k nem megszámolható:

Feladat: mekkora a valószínűsége, hogyha egyenletesen dobálunk egy négyzetbe, akkor a legnagyobb beleírható körön belülre dobunk?

Feladat: Ha nem találom a kulcsomat, fél-fél a valószínűsége annak, hogy otthon vagy a munkahelyemen hagytam. Ha a munkahelyemen, akkor 9 fiók valamelyikében lehet egyenlő valószínűséggel. Egy konkrét esetben a munkahelyemen már 8 fiókban megnéztem, de egyikben sem volt. Mekkora a valószínűsége, hogy a 9. fiókban van? (Vigyázat, a 8 fiók átnézésével csökken az eseménytér!)

Feltételes valószínűség

bekövetkezett B esetén mekkora A bekövetkezésének a valószínűsége

ha P(B)≠0, P(A|B)=P(AB)/P(B)

Feladat: 32 lapos magyar kártya esetén mi a mi a valószínűsége, hogy először egy 7-est, utána 9-est, majd utána megint egy 7-est kapunk?

Megoldás: $P(A_1)=4/32$, $P(A_2|A_1)=4/31$, $P(A_3|A_1A_2)=3/30$, $P(A_1A_2A_3)=P(A_3|A_1A_2)P(A_2|A_1)P(A_1)$

Teljes valószínűség tétele

Ha $B_1 \ldots B_n$ teljes eseményrendszer és egymást kizáró események,

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

Feladat: Egy termék 40%-át az első, 30-30%-át a második és a harmadik napon készítik el. Az első napon gyártottak 5%-a, a második napon gyártottak 7%-a, a harmadik napon gyártottak 10%-a hibás. Az összes termék hány %-a hibás?

Feladat: A hallgatók 80%-a az alap matematikát hallgatja, 20%-a a haladó kurzusra jár. Az alap matematikán a hallgatók 60%-a végzi el sikeresen a gyakorlatot, a haladón 80%-a. Összességében hány % végzi el sikeresen a gyakorlatot?

Bayes tétele

Hogyan számítható ki B_i A-ra vonatkoztatott feltételes valószínűsége A B_i-re vonatkoztatottjából.

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}$$

Függetlenség

A és B események függetlenek, ha P(A)P(B)=P(AB) (ez a definíciója!)

Feladat: Független-e a borsók alakja és színe, ha az alábbi az előfordulási valószínűségük?(igen)

alak\szín	*zöld*	*sárga*
kerek	*9/16*	*3/16*
szögletes	*3/16*	*1/16*

Feladat: Független esemény-e egy kocka háromszori feldobásakor, hogy az első dobás 6-os és a három dobás összege 10?(nem)

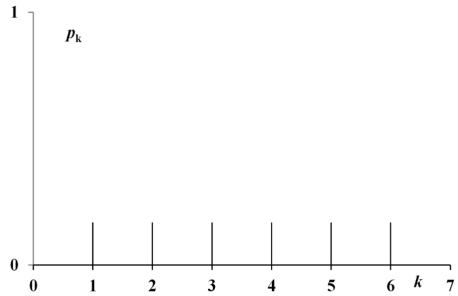
Valószínűségi változó

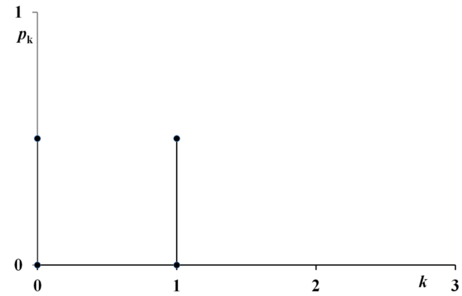
T eseménytér elemei → egy-egy számértéket rendelünk hozzá, ezt a számértéket valószínűségi változónak nevezzük, jele ξ. Vigyázat! Bár változónak nevezzük, igazából egy az események terén értelmezett függvény, aminek az értékkészletét $\{x_1,x_2,x_3\ldots\}$nevezzük valószínűségi változónak. Ha az esemény számértéket ad, többnyire azt a számot rendeljük az eseményhez. Más esetben pl. egész számokat.

Diszkrét valószínűségi változó esetén:

$p_k=P(\xi=x_k)=P(A_k)$

p_k - x függvényében - diszkrét valószínűségi változó eloszlása

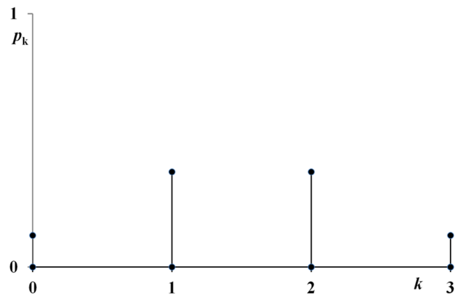




kockadobás eloszlása

pénzfeldobás eloszlása

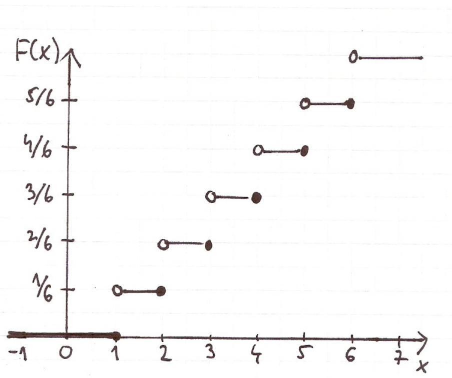
Feladat: Mi a valószínűsége, hogy 0, 1, 2, vagy 3 piros lámpánál kell megállnia, ha három lámpán halad át és mindegyiknél 50%-os valószínűséggel kap pirosat. Rajzolja le az eloszlást! Megoldás:



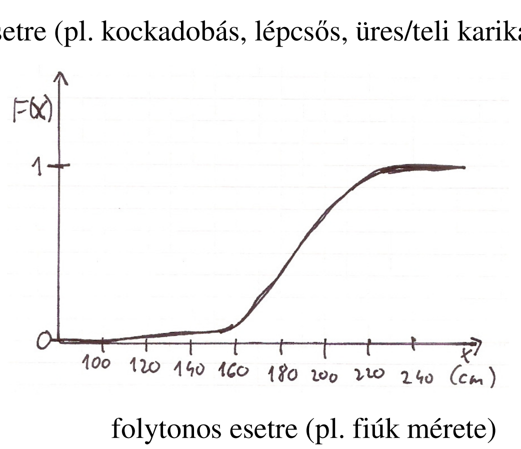
<u>Eloszlásfüggvény</u>

$F(x)=P(\xi<x)$, ahol $x \in R$ (az értelmezési tartomány a teljes számegyenes, nem csak ott, ahol ξ van!)

Tulajdonságai: monoton nő, $\lim_{x\to-\infty}F(x)=0$, $\lim_{x\to\infty}F(x)=1$, mindenütt folytonos balról



diszkrét esetre (pl. kockadobás, lépcsős, üres/teli karikák helye!)



folytonos esetre (pl. fiúk mérete)

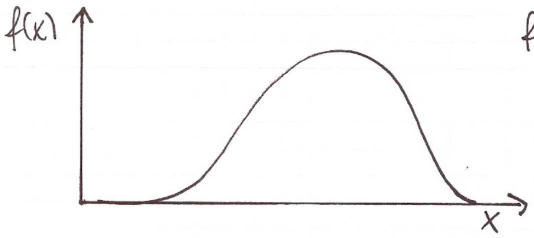
Sűrűségfüggvény

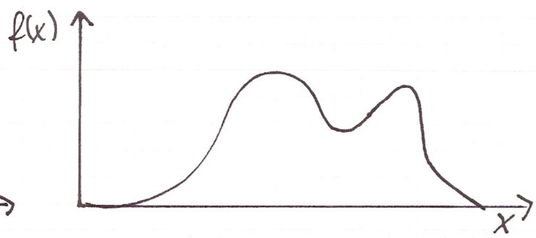
folytonos ξ esetén, ha létezik $f(x)$, úgyhogy

$$F(x)=\int_{-\infty}^{x} f(t)dt$$

Tulajdonságai: nem negatív, $\int_{-\infty}^{\infty} f(t)dt=1$,

Mit jelent? $p=\int_{x_0}^{x_0+\Delta x} f(t)dt$ az adott $[x;x+\Delta x]$ intervallumba esés valószínűsége. De ha $\Delta x \to 0$, akkor $p \to 0$, tehát $f(x)$ adott értéke nem azonos x valószínűségével!





unimodális (pl. férfiak magassága) bimodális (pl. nők+férfiak egyszerre mérve)

Kérdések

x lehet-e negatív?(igen)

f(x), F(x) lehet-e negatív? (nem,nem)

Diszkrét valószínűségi változónál lehet-e sűrűségfüggvény?(nem)

Diszkrét valószínűségi változónál van-e eloszlásfüggvény?(igen)

Várható érték – eloszlás helyét jellemzi a számegyenesen

diszkrét ξ-re $E(\xi)=\sum p_k x_k$

folytonos ξ-re $E(\xi)=\int_{-\infty}^{\infty} xf(x)dx$ (másik neve eloszlás első momentuma)

Feladatok

Mennyi a kockadobás várható értéke?

Két dobókockával dobunk, a kapott nyeremény a dobások összege, kivéve, ha van 6-os a dobások között, akkor mi fizetjük be a dobások összegét. Mekkora a nyeremény várható értéke?

Az 5-ös Lottón az öt találatra 10^8, a négyesre 10^6, a hármasra 10^4, a kettesre 10^3 Ft-t fizetnek. Mekkora a nyeremény várható értéke?

Szórás – szórásnégyzet (variancia) – eloszlás szélességét jellemzi

szórásnégyzet, variancia $\sigma^2=D^2(\xi)=E\left[(\xi-E(\xi))^2\right]$

szórás $\sigma = \sqrt{\sigma^2}$

sokszor így is számolják: $\sigma^2 = E(\xi^2) - [E(\xi)]^2$

diszkrét ξ-re $\sigma^2 = \sum p_i (x_i - E(\xi))^2 = \sum p_i x_i^2 - \left(\sum p_i x_i\right)^2$

folytonos ξ-re $\sigma^2 = \int_{-\infty}^{\infty} (x - E(\xi))^2 f(x) dx$ (másodrendű centrális momentum)

Mire jó a várható érték és a szórás? Majd konkrét eloszlásoknál és a statisztikánál látunk rá példákat.

Csebisev-egyenlőtlenség

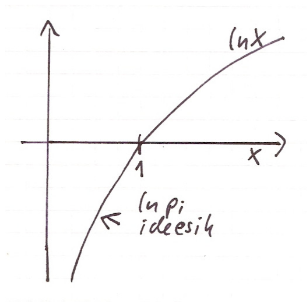
felső korlát a várhatóérték körüli szimmetrikus intervallumokon kívülre esés valószínűségére

$P(\varepsilon \leq |\xi - E(\xi)|) \leq \sigma^2/\varepsilon^2$, ahol $\varepsilon \in R^+$

Mennyi a valószínűség $\varepsilon = k\sigma$, $k=1,2,3$ esetén?

Valószínűségi változó entrópiája

$S = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx$ vagy $S = -\sum p_i \ln p_i$



az ln$x$ függvény (ln$x$ negatív, ha $x < 1$)

Feladat: A szabályos vagy a cinkelt dobókockával való dobások entrópiája nagyobb, ha a cinkelés következtében a 6-os kétszer gyakrabban jön ki, mint a többi szám külön-külön?

Feladat: Melyiknek nagyobb a (valószínűségi) entrópiája, ha egy ideális gáz a rendelkezésre álló teret egészében egyenletesen tölti ki, vagy csak a tér felében található meg, ott egyenletes eloszlással?

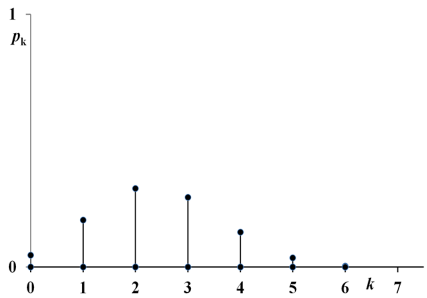
Fontosabb eloszlások

Ha a természetben lejátszódó és az általunk kitalált folyamatokat valószínűségi alapon vizsgáljuk, a folyamatok nagy része pár alap eloszlással leírható.

Binomiális eloszlás

n alkalommal végrehajtunk egy kísérletet. Ebből *k* alkalommal következik be az A esemény. Az A esemény valószínűségét *p*=P(A) jelölve, annak a valószínűsége, hogy pontosan *k* alkalommal következik be:

$p_k = \binom{n}{k} p^k (1-p)^{n-k}$, ahol 0≤*k*≤*n* és *k*∈N.



binomiális eloszlás, *n*=6, *p*=0,4

ξ diszkrét valószínűségi változó és értéke azonos A bekövetkezésének számával. Az eloszlás alakja két paramétertől függ: *n* és *p*. (1-*p*)-t szokás *q*-val külön jelölni.

E(ξ)=np $\sigma^2=np(1-p)$

Feladat: Ellenőrizzük, hogy logikai alapon a binomiális eredmény képletével azonoshoz jutunk-e a következő példában! Egy készletben (pl. kémcső) minden századik hibás. Ha 20 kémcsövet vizsgálunk, mi a valószínűsége, hogy mind jó? Hogy pontosan egy hibás?

Megoldás: p=0,01, n=20 *mind jó:* $(1-p)^{20}$ *, egy hibás:* $20*p*(1-p)^{19}$ *(az elsőnek hibásat választunk, a többinek jót, és ezt 20-szal szorozzuk, mert annyi helyre választhatnánk a hibásat)*

Feladat: Magyarországon a 0-s vércsoport gyakorisága p=0,32. Mi a valószínűsége, hogy véletlenszerűen kiválasztott n=3 emberből egy se 0-s vércsoportú? Pont 1 ember 0-ás vércsoportú? Kevesebb, mint 2 ember 0-s? Legalább 2 ember 0-s?

Megoldás EXCEL-lel: BINOM.ELOSZLÁS(k,n,p,0/1) Az utolsó helyre 0-t írunk, ha pont az adott k valószínűségét számoljuk. 1-t, ha $\sum_{i=0}^{k} p_i$ *-t szeretnénk kiszámolni. (Az EXCEL hibásan tartalmazza az elméletet: sűrűségfüggvényt rendel diszkrét valószínűségi változóhoz, valamint F(x)=P(ξ≤x)-t használ az elvi F(x)=P(ξ<x) helyett.)*

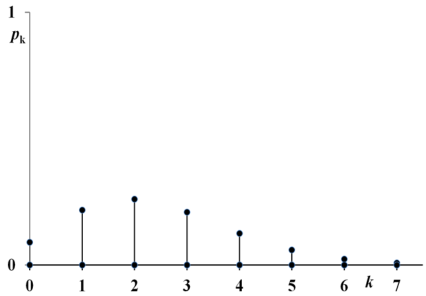
Feladat: A kémcsöves feladatnál mi a valószínűsége, hogy 15-nél több hibás? Használja ki, hogy $P(B)+P(\overline{B})=1$*, vagyis néha a komplementert érdemes kiszámolni és annak valószínűségét 1-ből kivonni.*

Poisson-eloszlás

Ha binomiálisnál *n*→∞ és *p*→0, vagyis ritka események eloszlása, ha az esemény bekövetkezése arányos a mérettel és/vagy időintervallummal, valamint az egymás utáni események függetlenek egymástól.

$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, ahol 0<*λ* és *λ*∈R és *k*∈N (*k*=0-t is beleértve)

egy paraméteres (*λ*), *ξ* diszkrét, *E(ξ)=λ*, illetve $\sigma^2=\lambda$



Poisson-eloszlás λ=2,4 paraméterrel

Kémiai példa: adott nagy anyagmennyiségnél időegység alatt várható radioaktív bomlások száma.

Feladat: GM számlálóval radioaktív háttérsugárzást mérünk. Egy óra alatt 2700-t jelzett a gép. Mi a valószínűsége, hogy 1 másodperc alatt egyet se mérünk? Többet, mint kettőt mérünk? 2 másodperc alatt egyet se mérünk?

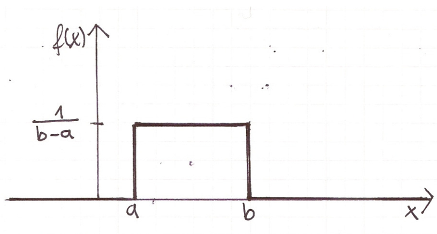
Megoldás EXCEL-lel: POISSON(k,λ,0/1) Először a kérdezett időegységre vonatkozó λ-t kell kiszámolni. A többet, mint kettőnél a komplementert kell 1-ből kivonni: $p=1-(p_0+p_1+p_2)$.

Feladat: Egy nagyvárosban naponta átlagosan 12 traumatológiai ellátást igénylő súlyos baleset történik. Tegyük fel, hogy 4 órára foglal le egy műtőt egy sérült. Hány műtő kell, hogy az 95%-ban legyen üres műtő? (4 órás intervallumra vonatkozó Poisson eloszlással dolgozzon!)

*Megoldás: POISSON(k,12*4/24,1) használatával próba-szerencse alapon megkeresni azt a k-t, ahol a valószínűség meghaladja a 0,95-t.*

<u>Egyenletes eloszlás</u>

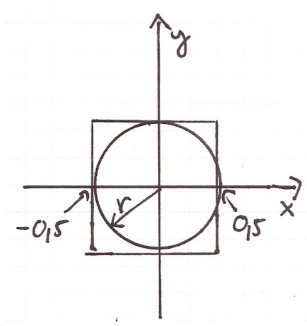
$$f(x)=\begin{cases} 0,\text{ha}(x\le a) \\ \dfrac{1}{b-a},\text{ha}(a<x<b) \\ 0,\text{ha}(b\le x)\end{cases} \quad F(x)=\begin{cases} 0,\text{ha}(x\le a) \\ \dfrac{x-a}{b-a},\text{ha}(a<x<b) \\ 1,\text{ha}(b\le x)\end{cases}$$



egyenletes eloszlás sűrűségfüggvénye

(*a;b*) intervallumban folytonos valószínűségi változó. Könnyű ilyen véletlen számot generálni számítógéppel (pszeudo véletlen számot), pl. VÉL(), illetve *<u>Eszközök/</u> Adatelemzés / Véletlenszám-generálás*

Feladat: A korábban említett geometriai analógia alapján becsüljük meg π értékét. Egyenletesen generáljunk pontokat egy origó központú egységnyi élhosszú kockába (pl. 100 x és y koordináta), és annak alapján becsüljünk, hogy a pontok hányadrésze kerül bele az origó középpontú egységnyi átmérőjű körbe.

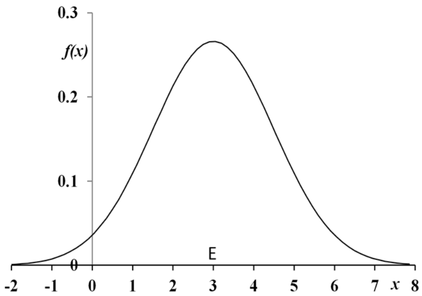


A négyzet és kör javasolt elrendezése

Megoldás: [-0,5;0,5] intervallumba eső egyenletes eloszlású véletlenszámokat generálunk az A és B oszlopokba. A C oszlopban kiszámoljuk a pontok (x-y számpárok) távolságát az origótól. A DARABTELI függvénnyel tudja megszámolni, hány távolság esik 0,5 alá.

<u>Normális eloszlás</u> (Gauss-eloszlás, haranggörbe)

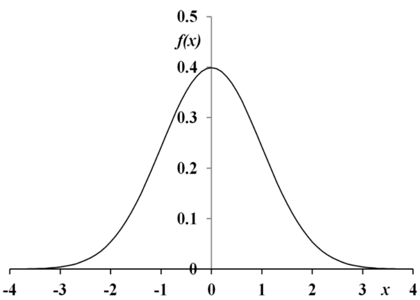
$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-E)^2}{2\sigma^2}}$, ahol $\sigma, E \in R$ és $0<\sigma$



normális eloszlás sűrűségfüggvénye, E=3, σ=1,5

ξ folytonos valószínűségi változó, $\xi \in R$,az eloszlás két paramétere az eloszlás várható értéke és szórása, *f(x)* szimmetrikus „haranggörbe” vagy „Gauss-görbe” *E*-re

$F(x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-E)^2}{2\sigma^2}}dt$, nem adható meg analitikus alakban elemi függvényekkel. Ma számítógéppel, számológéppel számoljuk, régen táblázatokban kerestük ki. Táblázatban csak egy volt megadva: <u>standard normális eloszlás</u> (E=0, σ=1). Erre átvihető mindegyik másik a változó standardizálásával: $x_{st}=\frac{x-E}{\sigma}$



a standard normális eloszlás sűrűségfüggvénye

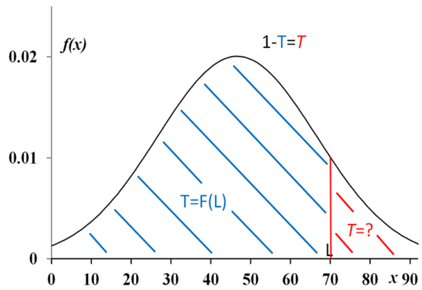
Egy számot célzó természettudományos mérések eredményei folytonos esetben többnyire ilyen eloszlásúak, mert „központi határeloszlás tétele”: ξ sok dologtól függ és sokat mérünk sok mintán → az eredmény normális eloszlású.

Értelmezzük az alábbi táblázat adatait! Hasonlítsuk össze a Csebisev-egyenlőtlenség értékeivel!

x_{st}	$f(x)$	$F(x)$
0	0,39	0,50
1	0,34	0,84
2	0,054	0,977
3	0,005	0,999

Feladat: Budaörsön 1998-ban NO_2 koncentrációjára $E=46{,}6\ \mu g/m^3$ és $\sigma=19{,}9\ \mu g/m^3$ értékeket határoztak meg a napi átlagokra. A napok hány %-ban lépték túl az akkori $L=70\ \mu g/m^3$ egészségügyi határértéket, ha az adatokra normális eloszlást feltételezünk? Mivel indokolja, hogy egy kisvárosban ilyen magas NO_2 értékeket mérnek?

Megoldás menete: NORM.ELOSZL($x,E,\sigma,1$) használatával az ábrának megfelelően.

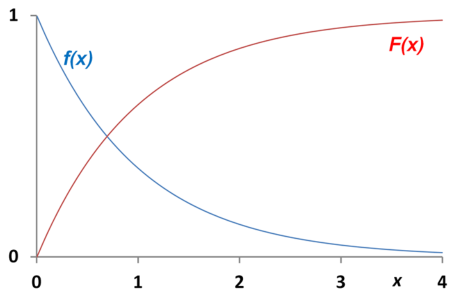


megoldás komplementer számításával

<u>Exponenciális eloszlás</u>

$$f(x)=\begin{cases} 0,\text{ha}(x\le 0) \\ \lambda e^{-\lambda x},\text{ha}(0<x)\end{cases}\text{, ahol } 0<\lambda,\ \lambda\in R \text{ és } F(x)=P(\xi<x)=\begin{cases} 0,\text{ha}(x\le 0) \\ 1-e^{-\lambda x},\text{ha}(0<x)\end{cases}$$

ξ folytonos $(0,\infty)$-ben és $\xi\in R^+$, egyparaméteres (λ), $E(\xi)=1/\lambda$, $\sigma=1/\lambda$, $\lim_{x\to+0} f(x)=\lambda$



exponenciális eloszlás sűrűség- és eloszlásfüggvénye $\lambda=1$ paraméterrel

Általában olyan folyamatokra, ahol két esemény bekövetkezése közötti idő a meghatározó. Reakciókinetikában az ütközések a meghatározóak, x=idő (pl. bomlási folyamat). Nem öregedő élettartammal kapcsolatos eloszlás (mindegy honnan kezdjük, a görbe alakja ugyanaz lesz). Felezési idővel való kapcsolata: $t_{1/2}=\ln 2/\lambda$ (ennyi idő alatt csökken az adott anyag mennyisége a felére pl. radioaktív bomlásnál).

Kapcsolata a bomlások differenciál-egyenletével:

$\frac{dN}{dt}=-kN$ megoldása: $N(t)=N_0e^{-kt}=N_0(1-F(t))$, ahol t az időt, N az anyagmennyiséget, N_0 a kezdeti anyagmennyiséget, k a reakciósebességi együtthatót (reakciósebességi állandót) jelenti.

Feladat: Egy bomlási folyamat $\lambda=0{,}0001$ év^{-1} paraméterű exponenciális eloszlással írható le. Az anyag hányadrésze bomlik el 1000 év alatt? Mennyi marad 20000 év után? Mennyi a felezési idő? Hányadrésze marad meg három felezési idő után? EXP.ELOSZLÁS(x,λ,1)

Hipergeometriai eloszlás

Már találkoztunk vele a Lottónál.

$$p(m,n,M,N)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$, ahol $N,M,n,m \in \mathbb{N}$, $0<N$, $n\leq N$, $M\leq N$, $m\leq n$, $m\leq M$

Feladat: Mi a valószínűsége annak, hogy egy 10 fős mintában 4 lány van, ha az iskolába 1000-n járnak és ebből 527 lány?

Megoldás: HIPERGEOM.ELOSZLÁS(4,10,527,1000)

Feladat: Mi az ötös Lottón a 2, 3, 4 és 5 találat valószínűsége?

Gyakorló feladatok eloszlásokhoz

1) Egy mérés során átlagosan minden 50-ik adatot hibásan jegyeznek fel. Mekkora a valószínűsége, hogy 50 kiválasztott adat mind helyes? Mekkora a valószínűsége, hogy 100 adatból pontosan 2 hibás.

2) Egy szennyezett vízben literenként átlagosan 10000 baktérium van. Mennyi annak a valószínűsége, hogy 0,5 ml-es mintában legalább 4 baktérium van? Mekkora mintát kell venni, hogy 95%-os valószínűséggel legalább egy baktérium legyen benne?

3) Egy monitoron átlagosan minden tízezredik pixel hibás. Mekkora a valószínűsége, hogy egy 150x150-szeres darabon minden pixel jó?

4) Azt tapasztalták, hogy 10 nap alatt egy anyag tömegének a fele elbomlott. Mennyi ideig kell várnunk, hogy már csak az eredeti mennyiség tizede maradjon? Mekkora része marad 20 nap után? 40 nap után?

5) Mekkora a valószínűsége, hogy egy standard normális eloszlással rendelkező valószínűségi változó értéke 0,2 és 0,8 közé esik? Milyen valószínűséggel negatív az értéke? Mennyi a valószínűsége, hogy –1 és 1 közé esik? Mennyi a valószínűsége., hogy 2-nél nagyobb?

6) Számítsuk ki két szabályos pénzérme feldobása esetén a kapott „fej” eredmények számának várható értékét és szórását!

7) Egy szerves kémiai reakciót átlagosan 5 kísérletből 4-szer sikerül reprodukálni. Mekkora a valószínűsége, hogy 10 próbálkozásból pontosan 8 sikerül?

8) Egy hallgató 30 tételből 20-t tanult meg. mekkora az esélye, hogy a húzott 3 tételből legalább kettőt megtanult?

9) Egy test többször megismételt tömegmérése során 5 mg várható értéket és 0,1 mg szórást állapítottak meg. Legalább mekkora tömeget mérünk 95%-os valószínűséggel? Melyik intervallumba esik az átlaghoz legközelebbi 80%-a méréseknek?

10) GM csöves méréskor átlagosan 4 beütést rögzítettek percenként a háttérsugárzásra. Mi a valószínűsége, hogy 12-nél több beütést rögzítenek egy percben? Mi a valószínűsége, hogy másfél óra alatt 1000 beütésnél többet észlelnek?

11) Egy palackozó gép átlagosan 0,998 dm^3 anyagot 0,006 dm^3 szórással tölt az üvegekbe. Mekkora palackot kell rendelni, hogy csak az esetek 2,3%-ban csorduljon túl a palack?

12) A csúcsidőben egy liftbe mindig a maximálisan megengedett 8 fő száll be, férfiak és nők 50-50%-os valószínűséggel. Ha az összeterhelés 660 kg feletti, a lift nem indul, valakinek ki kell szállnia. Mekkora a valószínűsége, hogy valakinek ki kell szállnia, ha a nőknél 60 kg, a férfiaknál 90 kg testsúllyal számolunk?

13) Ha a buszon a büntetés a jegy árának 20-szorosa, megéri-e bliccelni, ha az ellenőrzés gyakorisága λ=0,0667 paraméterű exponenciális eloszlással írható le? A példa természetesen csak züllött társadalomban értelmes!

Megoldások: binomiális: 1, 3, 6, 7, 12;Poisson: 2, 10; exponenciális: 4, 13; normális: 5, 9, 11; hipergeom.: 8.

Statisztika

Célja: egy halmazból, sokaságból kiválasztott minta alapján az egész halmazra vonatkozó következtetéseket vonjunk le. Az események eloszlását a véges minta miatt nem ismerjük tökéletesen, miként alkalmazzuk a valószínűségszámítás fogalmait ilyen esetben.

Várható érték becslése

várható érték N elemű mintára (y_i elemek): $E(\xi) \approx \overline{y} = \frac{\sum_{i=1}^{N} y_i}{N}$ (EXCEL-ben: ÁTLAG)

medián ($y_{medián}$): középső érték, vagy két középső átlaga. Kevésbé érzékeny a kilógó (elszúrt) adatra. MEDIÁN

módusz: leggyakoribb adat (diszkrét eloszlásnál értelmes) MÓDUSZ

további EXCEL függvények: MIN, MAX, KICSI, NAGY

Szórás és szórásnégyzet (variancia) becslése mintából

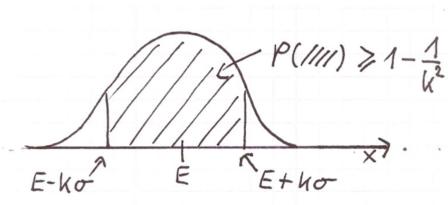
$$\sigma^2 \approx s^2 = \frac{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}{N-1} \qquad \sigma \approx s = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}{N-1}}$$

neve: becsült szórás(négyzet), korrigált tapasztalati szórás(négyzet), VAR, SZÓRÁS

σ^2 becslése csak N-es osztás esetén „torzított" lenne, (N-1) osztás esetén torzítatlan (Bessel-féle korrekció). Sajnos σ becslése így is torzított, tehát az igazi statisztikus varianciákkal és nem szórásokkal dolgozik!

Kilógó mérési adat kiválasztása

Háttér a Csebisev-egyenlőtlenség,



Csebisev-egyenlőtlenség szemléltetése tetszőleges eloszlásra

Szimmetrikus intervallumba esési valószínűségek a Csebisev-egyenlőtlenség alapján, illetve a normális eloszlásra

intervallum	tetszőleges eloszlásra P	normális eloszlásra P
$E\pm1\sigma$		P=0,682
$E\pm2\sigma$	0,75≤P	P=0,954
$E\pm3\sigma$	0,88≤P	P=0,997

Normális eloszlásnál P=0,95 esetén a szorzó 1,96.

Ha a mintában lehet hiba (pl. félremért kilógó adat), az átlagot célszerű a mediánnal becsülni, a terjedelmet az ún. kvartilisok segítségével.

alsó kvartilis ($y_{1/4}$): az elem, aminél az adatok negyede kisebb, háromnegyede nagyobb

felső kvartilis ($y_{3/4}$): az elem, aminél az adatok háromnegyede kisebb, negyede nagyobb

interkvartilis távolság: $y_{3/4}$-$y_{1/4}$

gyanúsak - eldobhatóak azok a kilógó adatok, amik kívül vannak a

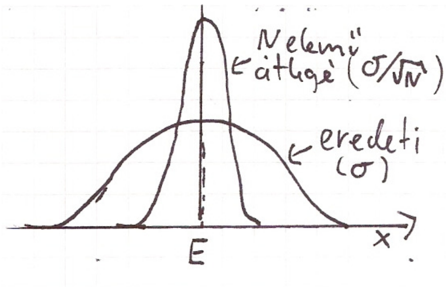
$y_{\text{medián}}\pm1{,}5*(y_{3/4}-y_{1/4})$, esetleg a $y_{\text{medián}}\pm2{,}25*(y_{3/4}-y_{1/4})$ intervallumon

KVARTILIS, PERCENTILIS

Várható érték szórása

Vegyünk N elemű mintát egy E várható értékű és σ szórású eloszlásból, számoljuk ki $\bar{y}$-t. Ismételjük ezt meg sokszor. Mi lesz a számolt $\bar{y}$-k szórása (σ_N-nel jelölve)?

Bemutatható, hogy $\sigma_N = \sigma / \sqrt{N}$. Ugyanez érvényes a becsült szórásokra is. Tehát: $\lim_{N\to\infty}\sigma_N=0$



Egy eredeti sokaság és az abból képzett N elemű átlagok sűrűségfüggvényei

Várható érték megbízhatósági intervalluma

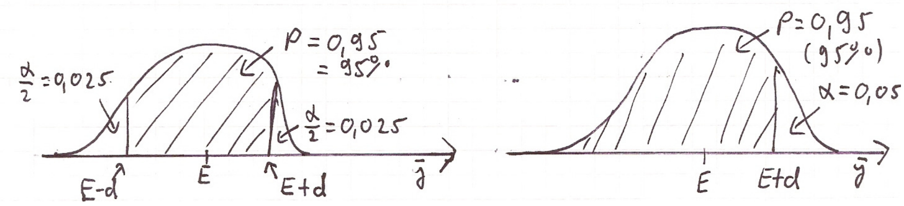
N mérés → $\bar{y}$ Mit írjunk le? $\bar{y}$ ±valamit, úgy, hogy tükrözze a várható érték pontosságát!

Ugyanaz az átlaga a két mérési sornak, de ugyanazt írnánk le?

a) 10,001; 10,002; 10,000; 9,999; 9,998
b) 10,000; 7,000; 13,000; 9,000; 11,000

Megbízhatósági (konfidencia) intervallumokat adjunk meg a várható értékre:

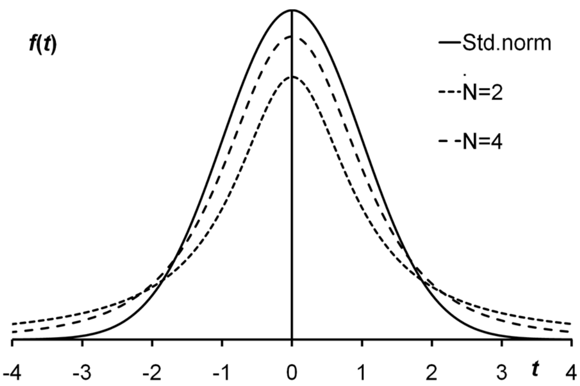
Olyant, ahol P(($\bar{y}$ -d)<E<($\bar{y}$ +d)) valószínűsége nagyobb legyen, mint mondjuk 90%, vagy 95%, vagy 99%. Többnyire kétoldali intervallumot adunk meg, de lehet csak egyoldali is! Sokszor nem a minimális P-t hanem α=1-P szignifikancia szintet adják meg.



kétoldali és egyoldali megbízhatósági intervallumok szemléltetése

A ma elfogadott megoldás (Gosset=”Student” 1908, Fisher 1925): t-eloszlás

$t = \frac{\bar{y} - E}{s / \sqrt{N}}$, ahol $\bar{y}$ és s az N elemű normális eloszlású mintából számolt várható érték és becsült szórás, E a sokaság (elméleti) várható értéke. t eloszlása kis N-re nem standard normális eloszlást ad, hanem ún. (N-1) szabadsági fokú t-eloszlást (más néven student-eloszlást). Szabadsági fok ≈ független adatok száma. Ha megjelenik egy az adatokat összekötő egyenlet (pl. várható érték számolása miatt), az csökkenti a szabadsági fokok számát.



t-eloszlás (*N*=2-re és *N*=4-re) és a standard normális eloszlás

Tehát várható érték megadása konfidencia intervallumával együtt:

kétoldali konfidencia intervallummal: $\bar{y} \pm t_{inverz}(\alpha/2; N-1)\frac{s}{\sqrt{N}}$, ahol $t_{\mathrm{inverz}}(\alpha/2;N\text{-}1)$ azt az értéket szolgáltatja, hogy a valószínűségi változó milyen értékénél lesz az *N*-1 szabadsági fokú *t*-eloszlás eloszlás függvényének értéke 1-α/2

egyoldali konfidencia intervallumnál pl. csak a felső érték: $\bar{y} + t_{inverz}(\alpha; N-1)\frac{s}{\sqrt{N}}$

EXCEL-ben $t_{\mathrm{inverz}}(\alpha/2;N\text{-}1)$ számolása: INVERZ.T(α;*N*-1) (mert automatikusan felezi α-t)

EXCEL-ben $t_{\mathrm{inverz}}(\alpha;N\text{-}1)$ számolása: INVERZ.T(2*α;*N*-1) (mert automatikusan felezi α-t)

EXCEL-ben *s* számolása: SZÓRÁS(adattartomány)

EXCEL-ben $\bar{y}$ számolása: ÁTLAG(adattartomány)

±-t nem értelmezi az EXCEL, tehát külön-külön cellába kerüljön $\bar{y}$ és a ± utáni rész!

30<*N* esetén a *t*-eloszlás közelíthető a standard normális eloszlással (INVERZ.NORM, MEGBÍZHATÓSÁG), tehát kétoldali 95%-os intervallumnál $t_{\mathrm{inverz}}(\alpha/2;N\text{-}1)$ helyett számolhatunk 1,96-tal.

Feladatok:

Egy termék esetén a következő tömegeket mértük g-ban: 7,7; 14,0; 4,2; 7,8; 4,7; 9,1; 9,9; 7,9; 13,2; 4,6; 6,5; 6,6. Számoljuk ki a 95%-os kétoldali konfidencia intervallumot az átlagos tömegre!

A következő koncentrációkat mértük mol/dm^3 *egységben: 0,120; 0,130; 0,110; 0,105; 0,125; 0,162; 0,135. Számoljuk ki a 95%-os kétoldali konfidencia intervallumot az átlagos tömegre! Határozzuk meg a minimális illetve a maximális értékét az átlagos tömegnek 95%-os egyoldali konfidencia sávval!*