Biostatisztika és informatika alapjai

4. előadás:

A valószínűségszámítás elemei

2014. szeptember 29.

Veres Dániel

Ebben az előadásban szó lesz valószínűségekről, illetve eloszlásokról.
Az előadás első felében néhány kísérletet mutatok elméletben, illetve gyakorlatban egy kis ízelítőül a valószínűség furcsa világáról. Ezeken keresztül megpróbálom megmutatni a minta nagyságának szerepét, hogy eljuthassunk a valószínűség egy definíciójához. Ez a definíció a nagy számok relatív gyakoriságokra vonatkozó törvényén alapszik.
Ezt követően az események valószínűségének alapfogalmaival ismerkedhetünk meg (jelölésrendszer; és/vagy kapcsolat események között; egymást kölcsönösen kizáró események és független események fogalma, használta; Kolmogorov axiómák; feltételes valószínűség és használata).
Ezt követően (röviden visszaidézve középiskolai vélhető ismereteiteket is) konkrét példákon keresztül megmutatom, hogy hogyan számíthatók valószínűségek az orvosi gyakorlatban is a valószínűségszámítás alapelemeivel, illetve elméleti eloszlások segítségével.
Definiáljuk a várható értéket és az elméleti varianciát is.
Majd rátérünk az egyes speciális eloszlásokra: az egyenletes, binomiális, poisson, normál, lognormál és exponenciális eloszlásokra.
Az utolsó néhány dián pedig megmutatom, hogy hogyan működik a „mindennapi gondolkodásunk”, illetve hogyan kellene működnie, ha a valószínűségszámítás megtanult elemeire is gondolunk...
Az előadáshoz képest néhány további diát is beletettem a jobb értelmezhetőség miatt.

Péter és Pál

(pétervári paradoxon)

Pénzfeldobásos játék, eredménye szerint

- a kezdeti nyeremény 2 dukát és mindig duplázódik, ha írást dobunk
- ha fej, akkor vége a játéknak, írásnál folytatódik
- a nyeremények:
  - ha az 1. dobás fej: Pál ad 2 dukátot Péternek
  - ha az 1. írás, 2. dobás fej: Pál ad 4 dukátot Péternek
  - ha csak a 3. lesz fej: Pál ad 8 dukátot Péternek...

Mennyit fizessen Péter a játékélményért egy játékra átlagosan?
(Úgy hogy ne járjon se Pál, se Péter rosszul)

A valószínűségszámítás egyik kezdeti, a fejlődést megindító, meghatározó problémája az 1715-ben megjelent *pétervári paradoxon* volt. Ez egy pénzfeldobásos játék két elméleti játékos Péter és Pál között. A játék szabályai a következőek:

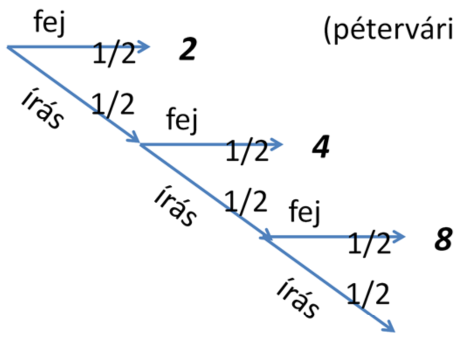
Pénzfeldobásos játék, eredménye szerint

- a kezdeti nyeremény 2 dukát és mindig *duplázódik*, ha írást dobunk
- ha fej, akkor vége a játéknak, írásnál folytatódik
- a nyeremények:
  - ha az *1. dobás fej*: Pál ad 2 dukátot Péternek
  - ha az 1. írás, *2. dobás fej*: Pál ad 4 dukátot Péternek
  - ha csak a *3. lesz fej*: Pál ad 8 dukátot Péternek...

Mennyit fizessen Péter a játékélményért egy játékra átlagosan? (Úgy hogy ne járjon se Pál, se Péter rosszul – mindketten a pénzüknél maradjanak)

Péter és Pál

(pétervári paradoxon)



Az egy játékra jutó átlagos nyeremény matematikailag („elméletileg”): végtelen!

$$\frac{1}{2}\cdot 2+\frac{1}{2^2}\cdot 4+\ldots+\frac{1}{2^n}\cdot 2^n$$

Tapasztalat: Buffon 2084 játékból átlagosan 9,82 dukátot nyert egy játékra nézve

Hányszor játszunk?

Tehát mi is lesz egy játékban Pál „várható átlagos” adománya?
Az esetek felében az első dobás fej – ez *½*2* dukátot jelent (az egyenlet 1. tagja) Péter számára egy átlagos játékban. Ha az első írás (ez ugye az esetek fele), akkor ezen esetek felében fej lesz a következő dobás eredménye, tehát átlagosan egy játékra nézve ez *½*½* esetben történik meg, azaz *¼*4* dukátot jelent Péternek. Ezt a gondolatmenetet folytatva figyelembe véve az összes lehetséges esetet Péter ezeknek az összegét nyeri egy játékban átlagosan. Ha megvizsgáljuk ezt az összeget, akkor láthatjuk, hogy *elméletileg* Péter egy játékban *átlagosan 1+1+1+1.... azaz végtelen dukátot kap*!
Na de *a gyakorlatban* azt tapasztaljuk, hogy egyetlen játékban sem kap végtelent Péter, illetve az átlagos nyereménye *egy játékra soha nem végtelen*!
A paradoxon elnevezés tehát erre a tapasztalati-elméleti (nem végtelen-végtelen) ellentétre utal.
Példának okáért Buffon (egy híres matematikus) 2048 játékot játszott és 9,82 dukátos átlagnyereményt produkált (közel sem végtelent...) Egymillió játékból pedig 10,94 dukátos átlagot kaptak egy számítógépes modellezésben.
A tapasztalat azt mutatta, hogy bár Péter átlagos nyereménye egy játékra nézve sohasem végtelen, de a *játékok számának növelésével mindig egyre több, közelítve az elméleti értékhez*!

Egy másik kísérlet....

Ebben a kísérletben egy járványügyi „eljárást” modellezünk.
Az adott betegség kimutatására rendelkezünk egy gyorsteszttel, amelynek az eredménye:

fehér: ha a vizsgált páciens egészséges
zöld: ha az illető beteg

Szeretnénk kideríteni a gyorsteszt segítségével, hogy egy kérdéses területen van-e járvány. A következőket tudjuk:

- A betegséggel nem sújtott („egészséges”) területeken:
    1 zöld eredményt kapunk 10 megvizsgált egyénre.
- A betegséggel sújtott („beteg”) területeken:
    9 zöld eredményt kapunk 10 megvizsgált egyénre

Felütötte-e a fejét a járvány az adott területen?

Ebben a kísérletben egy járványügyi „eljárást” modellezünk.
Az adott betegség kimutatására rendelkezünk egy gyorsteszttel, amelynek az eredménye:

fehér: ha a vizsgált páciens egészséges
zöld: ha az illető beteg

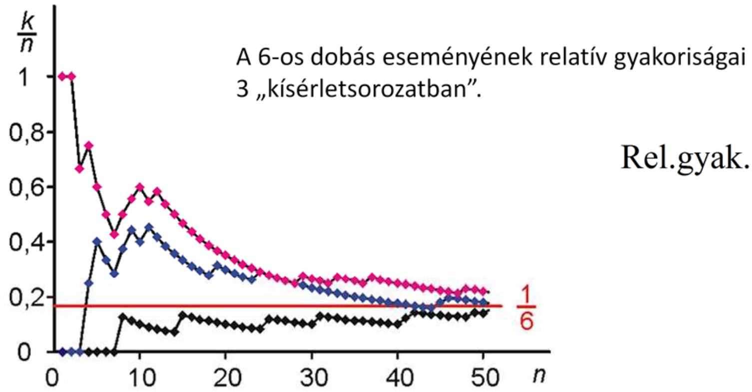
Szeretnénk kideríteni a gyorsteszt segítségével, hogy egy kérdéses területen van-e járvány. A következőket tudjuk:

- A betegséggel nem sújtott („egészséges”) területeken:
    1 zöld eredményt kapunk 10 megvizsgált egyénre.
- A betegséggel sújtott („beteg”) területeken:
    9 zöld eredményt kapunk 10 megvizsgált egyénre

Felütötte-e a fejét a járvány az adott területen?

Sajnálatos módon limitált forrás és idő áll rendelkezésünkre döntésünk meghozatalában, de döntenünk kell.
Az elvégzett kísérletből azt a tanulságot vonhattuk le, hogy az *elvégzett gyorstesztek (megvizsgált emberek) számának növelésével nő jó döntésünk „biztonsága”*.

Egy másik szemszögből...

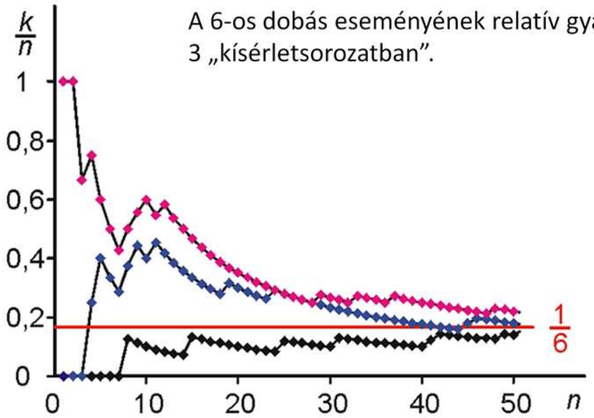


$$\text{Rel.gyak.} = \frac{N_{\text{"kedvező"}}}{N_{\text{összes}}}$$

Tapasztalati tény, hogy a **relatív gyakoriságok** ilyen sorozatai – bár ingadozásokat mindig mutatnak – a „kísérletsorozat” hosszának növekedtével egyre inkább **stabilizálódnak valamilyen érték körül**, továbbá ez az érték az aktuális „kísérletsorozattól” függetlenül lényegében ugyanakkora.

Ebben a kísérletben egy dobókockát dobunk 50-szer és közben rögzítjük a 6-os dobások relatív gyakoriságát. Ezt a kísérletet 3-szor végezzük el. Ezeknek az eredményei láthatók a dián. Azt tapasztalhatjuk, hogy a *relatív gyakoriság* („kedvező”/összes dobások száma) – bár folyamatosan ingadozva –, de *egy adott értékhez tart*, továbbá ez a stabilizálódó érték az aktuális *„kísérletsorozattól” függetlenül* lényegében ugyanakkora.

Valószínűség, mint mennyiség?



$$P = \text{Rel.gyak.} = \frac{N_{\text{"kedvező"}}}{N_{\text{összes}}}$$

$$N_{\text{összes}} \to \infty$$

Tapasztalati tény, hogy a relatív gyakoriságok ilyen sorozatai – bár ingadozásokat mindig mutatnak – a „kísérletsorozat” hosszának növekedtével egyre inkább stabilizálódnak valamilyen érték körül, továbbá ez az érték az aktuális „kísérlet sorozattól” függetlenül lényegében ugyanakkora. Legyen ez az érték a ***valószínűség***.

A ***nagy számok (relatív gyakoriságokra vonatkozó) tapasztalati törvényé***nek nevezzük ezt, amelyet **logikai úton bizonyítani nem lehet**.

Ezt a megfigyelést nevezzük a *nagy számok (relatív gyakoriságokra vonatkozó) tapasztalati törvényének*. Rendeljük hozzá a „kedvező” eseményünkhöz a „stabilizálódó” értéket, mint mennyiséget: jelen esetben a 6-os dobáshoz az 1/6-ot. Ezt az értéket nevezzük az *esemény valószínűségének*.
Tehát azt is mondhatjuk, hogy egy esemény relatív gyakorisága megegyezik az esemény valószínűségével, ha az ismétlések száma (kísérletsorozat hossza) végtelen.
Ez a törvény *tapasztalati törvény*, tehát logikai úton nem bizonyítható.

Események valószínűségei I.

Jelölések:

Esemény: **A**

(a páciensnek láza van)

Az *A* esemény bekövetkezésének valószínűsége**: P(A)**

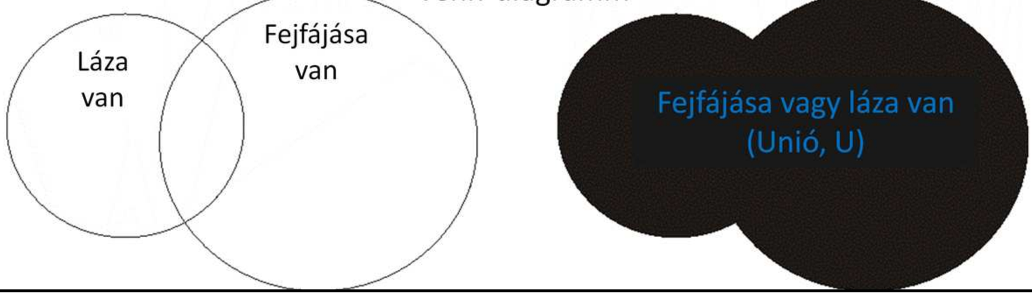
(annak a valószínűsége, hogy a páciensnek láza van)

Annak a valószínűsége, hogy *A* **vagy** *B* esemény bekövetkezik:

P(AvagyB), **P(A+B)**, **P(AUB)**

(annak a valószínűsége, hogy a páciensnek láza vagy fejfájása van)

Venn-diagramm



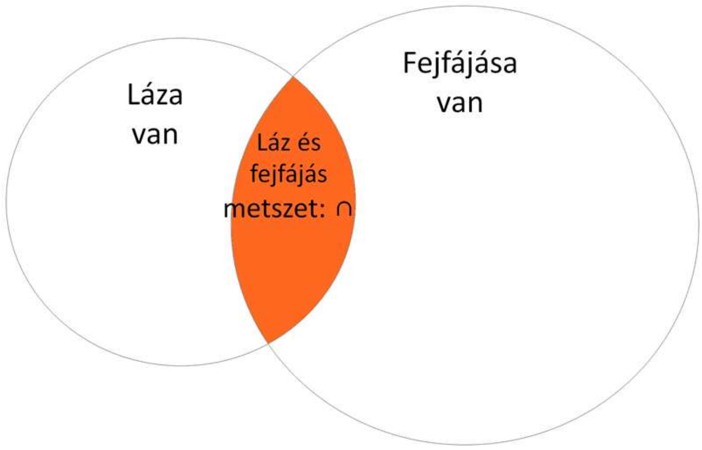
A következőkben az események valószínűségeiről beszélünk.
Először ismerkedjünk meg a jelölésrendszerrel. (a következőkben zárójelben, dőlt betűkkel találjátok a példákat)
Az egyes eseményeket nagy betűvel jelöljük – például *A = a betegnek láza van*. Az *A* esemény bekövetkezésének valószínűségét pedig *P(A)*-val jelöljük. Például *P(A) = annak a valószínűsége, hogy a páciensnek láza van*.
Annak a valószínűségét, hogy az *A vagy* a *B* esemény bekövetkezik (*a páciensnek láza vagy fejfájása van*) több féle módon jelölhetjük: *P(AvagyB), P(A*+*B)*, *P(AUB)*. Ez utóbbi jelölés a halmazoknál szokásos jelölésnek felel meg: az *U a halmazok unióját* jelenti, azaz legalább az egyik halmazhoz való tartozást. Ennek a Venn-diagrammon való ábrázolása látható az ábrán.

Események valószínűségei II.

Annak a valószínűsége, hogy *A* **és** *B* esemény egyaránt bekövetkezik:

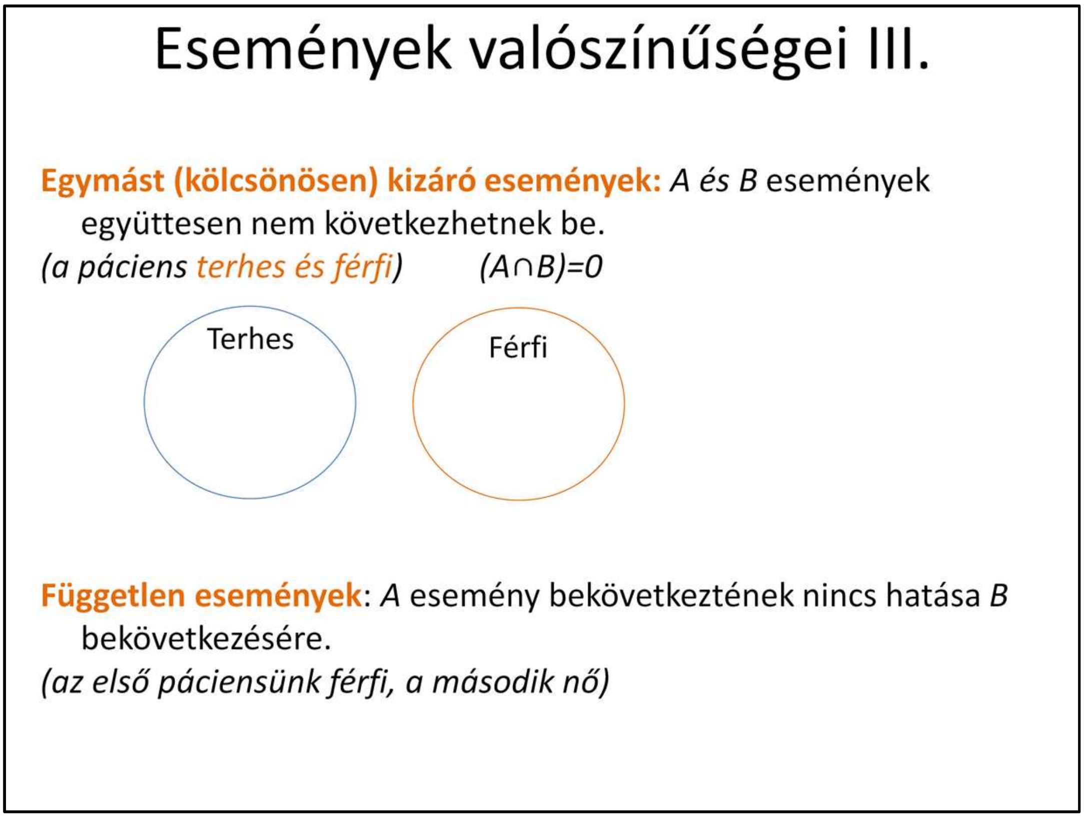
P(AésB), **P(A*B)**, **P(AB)**, **P(A∩B)**

(annak a valószínűsége, hogy a páciensnek láza és fejfájása van)



Annak a valószínűségét, hogy *A és B* esemény egyaránt bekövetkezik a következő módokon jelölhetjük: P(AésB), P(A*B), P(AB), P(A∩B). (*Annak a valószínűsége, hogy a páciensnek láza és fejfájása is van.)*

A ∩ jel a halmazok metszetét jelenti. A Venn-diagrammon ez a két halmaz közös része.



Ezen a dián kétféle esemény-kapcsolatra hívnám fel a figyelmet.
Az *egymást kölcsönösen kizáró események* azok, amelyek együttesen nem következhetnek be *(a páciens férfi és terhes)*. Halmazokkal ábrázolva őket látszik, hogy ez azt jelenti, hogy a két halmaz nem metszi egymást (metszetük üreshalmaz, azaz $(A \cap B)=0$).
Független eseményekre igaz, hogy az egyik esemény bekövetkezésének nincs hatása a másik esemény bekövetkeztére *(az első páciens férfi, a második pedig nő)*.

Események valószínűségei IV.

Feltételes valószínűség

A esemény bekövetkezésének valószínűsége B esemény bekövetkezését tudva, hogy B esemény bekövetkezett: **P(A|B).** *(annak a valószínűsége, hogy a páciensünknek influenza fertőzése van (ami egy vírusfertőzés) tudva azt, hogy vírusos eredetű a fertőzése)*

Mielőtt tovább haladnánk, definiáljuk a *feltételes valószínűséget*.
A „feltételesség” a feltételes gyakorisághoz hasonlóan itt is arra vonatkozik, hogy csak egy alcsoporton belül – azaz ha már egy esemény bekövetkezett – számítjuk a valószínűséget. Másként megfogalmazva *A* eseménynek *B*-re vonatkoztatott feltételes valószínűsége az *A* esemény valószínűsége, ha *B* bekövetkezett. Jelölése: P(A|B). Például *annak a valószínűsége, hogy a páciensünknek influenza fertőzése van tudva azt, hogy vírusos eredetű a fertőzése*.

Események valószínűségei V.

Események valószínűségének alaptörvényei ***(Kolmogorov-axiómák)***:

1. **0 ≤ P(A) ≤ 1**

2. **P(*biztos*) = 1** *(a páciens előbb vagy utóbb meghal)*
P(*lehetetlen*) = 0 *(a páciens teljesen egészséges*)*

*3. Egymást kölcsönösen kizáró eseményekre: (*P(AésB)=0)
P(AvagyB)=P(A)+P(B)
*(annak a valószínűsége, hogy páciensünk terhes **vagy** férfi)*

Ezekből levezethető:

+4. *Független eseményekre*: **P(AésB)=P(A)*P(B)**
*(annak a valószínűsége, hogy az első páciensünk férfi **és** a második nő)*

Az események valószínűségeinek leírását a Kolmogorov axiómák segítségével tehetjük meg. Az alábbiakban ezeket az axiómákat írom le (kissé egyszerűsített formában).
1. Egy esemény valószínűsége 0 és 1 közötti érték.
2. a) A biztos esemény valószínűsége 1 (*a páciens előbb-utóbb meghal* – mint tudjuk az élet egy szexuális úton fertőző halálos betegség ☺).
2. b) A lehetetlen esemény valószínűsége 0 (*a páciens teljesen egészséges* – mint tudjuk nincs egészséges ember, csak rosszul kivizsgált beteg ☺).
3. Annak a valószínűsége, hogy *A vagy B* esemény bekövetkezik, ha *A* és *B* események *egymást kölcsönösen kizáróak* egyenlő az egyik esemény valószínűségének és a másik esemény valószínűségének az összegével *(annak a valószínűsége, hogy következő páciensünk egy férfi lesz vagy egy terhes = annak a valószínűsége, hogy férfi lesz a következő páciensünk + annak a valószínűsége, hogy egy terhes lesz a következő páciensünk).*
Ezekből az axiómákból számos egyéb összefüggés levezethető, amelyek közül most egyet emelnék ki.
+4. Annak a valószínűsége, hogy *A és B* esemény is bekövetkezik, ha *A* és *B* események egymástól független események egyenlő az *A* esemény valószínűségének és a *B* esemény valószínűségének szorzatával *(annak a valószínűsége, hogy az első páciensünk férfi és a rákövetkező nő lesz = annak a valószínűsége, hogy férfi lesz a következő páciensünk*annak a valószínűsége, hogy a rákövetkező páciensünk nő lesz)*.
Megjegyzendő, hogy ezek a megállapítások (2-4) „fordítva" is igazak. Például ha P(AésB)=P(A)*P(B), akkor *A* és *B* események egymástól függetlenek.

Ismétlés...

Feltételes valószínűség: **P(A|B)*P(B)=P(A)**

annak a valószínűsége, hogy a páciensünknek influenza fertőzése van tudva azt, hogy vírusos eredetű a fertőzése 14% = P(A|B)

annak a valószínűsége, hogy a praxisunkban levő betegnek vírusos fertőzése van 8% = P(B)

*annak a valószínűsége, hogy páciensünknek influenza fertőzése van: P(A) = 0,14*0,08 = 0,0112 = 1,12%*

Felhívnám a figyelmet arra, hogy a feltételes valószínűséggel hasonlóképpen lehet számolni, mint a feltételes gyakoriságokkal (ld. első előadás függeléke). Ezt a következő példán mutatnám be.
Annak a valószínűsége, hogy a páciensünknek influenza fertőzése van tudva azt, hogy vírusos eredetű a fertőzése 14% = P(A|B) – ez az A esemény B-re vonatkoztatott feltételes valószínűsége.
Annak a valószínűsége, hogy a praxisunkban levő betegnek vírusos fertőzése van 8% = P(B) – ez a feltétel valószínűsége.
*Annak a valószínűsége, hogy páciensünknek influenza fertőzése van: P(A) = 0,14*0,08 = 0,0112 = 1,12%* - ez annak a valószínűsége, hogy praxisunkban influenza fertőzése van valakinek.

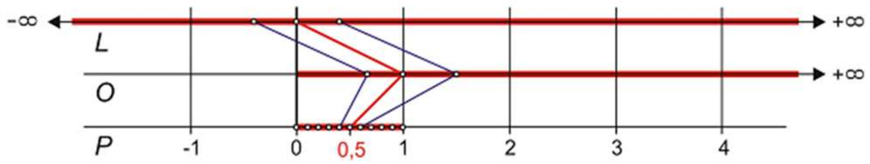
Esély

Esély (O - odds): „hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be”

$$O = \frac{P(A)}{1 - P(A)}$$

Logit (L): esély természetes alapú logaritmusa

Logit – Esély – Valószínűség



Egy további „valószínűség szerű” fogalomra hívnám fel a figyelmet, ez az esély. Ez az esemény bekövetkezésének valószínűségének és nem bekövetkezésének valószínűségének hányadosa, amely megmutatja, hogy hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be.

A logit nem más, mint az esély természetes alapú logaritmusa.

A logit-esély-valószínűség értékeinek kapcsolatát mutatja az ábra. Látható, hogy a valószínűség értéke 0 és 1 között, az esélyé 0 és végtelen között, amíg a logit értékei mínusz végtelen és plusz végtelen között vannak. Leolvashatjuk, hogy ha a valószínűség kisebb, mint 0,5, akkor az esély 1-nél kisebb, a logit pedig negatív. Természetesen további összefüggések is láthatóak...

Egy kis matek mindenkinek......

1.A. *Ismétlés nélküli permutáció.* (Hányféle sorrendben vizsgálhatjuk meg a kint várakozó 4 beteget?) $n!$

1.B. *Ismétléses permutáció.* (Hányféle sorrendben használhatjuk az oltóanyagokat a 3 influenza oltásra és 4 meningococcus oltásra váró gyermeknél, ha a megfelelő oltások ugyanolyanok?) $\frac{n!}{k_1!*k_2!*\ldots}$

2.A. *Ismétlés nélküli kombináció*. (Hányféleképpen választhatunk ki 4 diákot a kísérlethez a 10 jelentkező közül?) $\binom{n}{k}=\frac{n!}{(n-k)!*k!}$

2.B. *Ismétléses kombináció*. (Hányféleképpen választhatunk ki 4 diákot a kísérlethez a 10 jelentkező közül?- Egyet akár többször.) $\binom{n+k-1}{k}$

3.A. *Ismétlés nélküli variáció.* (Hányféleképpen és sorrendben választhatunk ki 3 pácienst a szívtranszplantációhoz a 10 alkalmas közül?) $\frac{n!}{(n-k)!}$

3.B. *Ismétléses variáció.* (Hányféleképpen és sorrendben választhatunk ki 30 pácienst véradásra a 100 alkalmas közül? – Többször is adhat.) n^k

Ezen a dián utalnék arra, amit elvileg mindenki tud, mert megtanulta a középiskolában.
A valószínűségszámítás alapját a permutációk, kombinációk és variációk jelentik.
(Néhány megjegyzés: ismétlés nélküli jelentése az, hogy nincs egyforma elem, kombinációnál a sorrend nem lényeges.)
A hallgatóság többségének szerencséjére azonban jelen kurzus során nem fogjuk ezeket ilyen formán alkalmazni.

Na az eddigiek meg mire vótak jók….

Influenzaszezont megelőzően a rendelőnkben az adott napra 4 oltóanyag áll rendelkezésre. Az előző években átlagosan 2989 páciensből 402 személyt kellett beoltanunk. Az előző év alapján mekkora a valószínűsége, hogy a rendelkezésre álló 4 oltóanyag elegendő lesz és el is fogy, ha 25 embert várunk aznapra?

$$P=\binom{n}{k}\cdot(p)^k\cdot(1-p)^{(n-k)}=\binom{25}{4}\cdot\left(\frac{402}{2989}\right)^4\cdot\left(1-\frac{402}{2989}\right)^{(25-4)}\approx 0{,}2$$

Mekkora a valószínűsége annak, hogy páciensünk 3.45 mmol/l-es K szintje még „egészséges”?

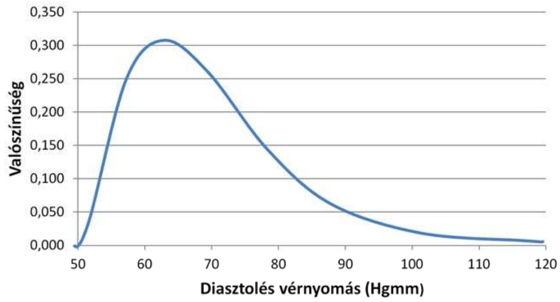
Hány szülés várható az esti ügyeletben , ha az éves statisztika 1000 szülést mutat éjfél és 8:00 között?

Az évfolyamból várhatóan hányan lesznek alkalmasak egy csípőprotézis elvégzésére (tömegük alapján)?

...... Hogyan számoljunk? Ismerjük a „képletet”?

Nade akkor ezek mire is jók nekünk? A következőkben olyan példát mutatok, ahol az ismétlés nélküli kombináció könnyen tetten érhető. Influenzaszezont megelőzően a rendelőnkben az adott napra 4 oltóanyag áll rendelkezésre. Az előző években átlagosan 2989 páciensből 402 személyt kellett beoltanunk. Az előző év alapján mekkora a valószínűsége, hogy a rendelkezésre álló 4 oltóanyag elegendő lesz és el is fogy, ha 25 embert várunk aznapra?
A kérdés megválaszolásához a Bernoulli eloszlást (lásd később) használjuk fel (vegyük észre, hogy a feladat során a 25 emberből választunk ki 4-et jelenti az ismétlés nélküli kombinációt).
Számos ehhez hasonló kérdést vetettem fel a dián. A kérdés az, hogy ezek megválaszolásához mit tegyünk, milyen „egyenletet” válasszunk, mikor melyiket?

Elméleti eloszlások

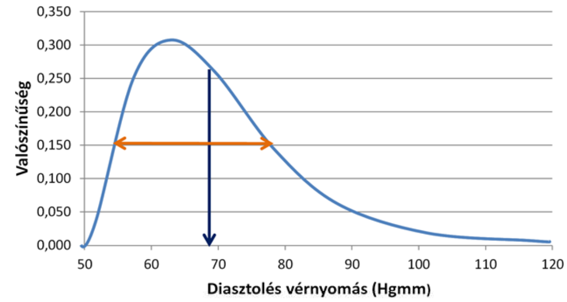


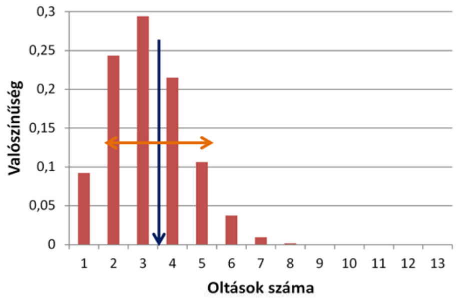
Ismerem az adott mennyiség valószínűségét. $N = \infty$

Ismerem, mert ki tudom számítani: *nevezetes eloszlások.*
(Vagy becsülni tudom)

A fent említett kérdések megválaszolásában segítenek minket az *elméleti eloszlások*. Ezek megmutatják egy adott érték előfordulásának valószínűségét. Ritka esetben igen-igen nagy számú (végtelen) mérés alapján ismert egy adott változó valószínűségi eloszlása – ld. például percentilis táblázatok, görbék. Az esetek nagyobb részében azonban néhány jellemző érték és úgynevezett *nevezetes eloszlások* alapján tudjuk meghatározni bármely értékhez tartozó, így az adott kérdéses értékünk (pl. oltóanyagszám) valószínűségét is. Ebben az esetben a kérdés csak annyi, hogy *melyek ezek a néhány jellemző értékek* – hogyan tudom ezeket megadni, illetve becsülni –, illetve melyik kérdésemre *melyik eloszlást* kell használnom.

Nevezetes eloszlások





- ***Várható érték (E, M, μ)***

$$E(\xi)=\int_{-\infty}^{\infty} p_i \cdot x_i \qquad\qquad E(\xi)=\sum_{i=1}^{m} p_i \cdot x_i$$

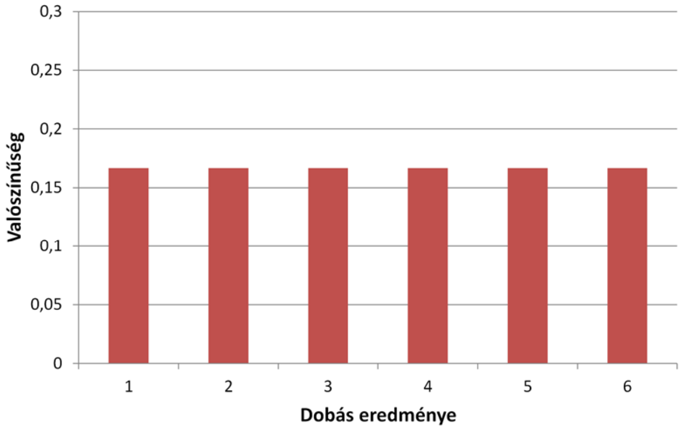
- Elméleti szórásnégyzet (Var, D^2, σ^2)

$$Var(\xi)=E\left[\left(\xi-E(\xi)\right)^2\right]$$

Medián, módusz, kvartilisek...

Vizsgáljuk meg először, hogy az elméleti eloszlásoknak melyek ezek a jellemzői. Két alapvető jellemzőt használunk: a *várható értéket* (E, M vagy μ jelöléssel) és az *elméleti szórást* (Var, D^2 vagy σ^2 jelöléssel). (Jelen esetben csak a számszerű változókról beszélek.)
A várható érték az eloszlás „közepét” mutatja (ahogyan a minta esetén a módusz, a medián és az átlag – tehát ezek használatosak a becslésére). Az elméleti variancia az elméleti eloszlás „szélességét” mutatja (ahogyan a minta varianciája, illetve kvantilistávolságai). Ez a két jellemző egyértelműen leírja az általunk használt speciális eloszlásokat – azaz ezek ismeretében bármely értékhez tartozó valószínűség meghatározható.
Az elméleti eloszlások is az adott változónak megfelelően lehetnek folytonosak, illetve diszkrétek.
Vizsgáljuk meg először a diszkrét esetben kapott várható érték definíciót. Vegyük észre, hogy p – a nagy számok törvényét használva – közelíthető a relatív gyakorisággal, amely pedig az abszolút gyakoriság osztva az elemszámmal. Tehát E felírható $E=(\Sigma(\text{abs.gyak.}_i * x_i))/n$, azaz az egyforma elemeket annyiszor adom össze, ahány van belőle (ez az $\text{abs.gyak.}_i * x_i$), és ezt minden elemnél megteszem – tehát összességében minden elemet összeadok, majd az így kapott értéket osztom az elemszámmal. Vegyük észre, hogy ez nem más, mint az átlag definíciója végtelen elemszám esetében.
Folytonos változó esetében ugyanezt az összegzést végzem el, csak végtelenül kicsi osztályszélességgel (ez az integrálás).
A következőkben megvizsgáljuk, hogy melyik speciális eloszlásnál hogyan adható meg a várható érték, illetve melyik eloszlás milyen kérdés megválaszolására alkalmas.
(Megjegyzés: a számunkra érdekes eloszlások várható értékei és szórásai szerepelnek a képlettárban.)

Egyenletes eloszlás



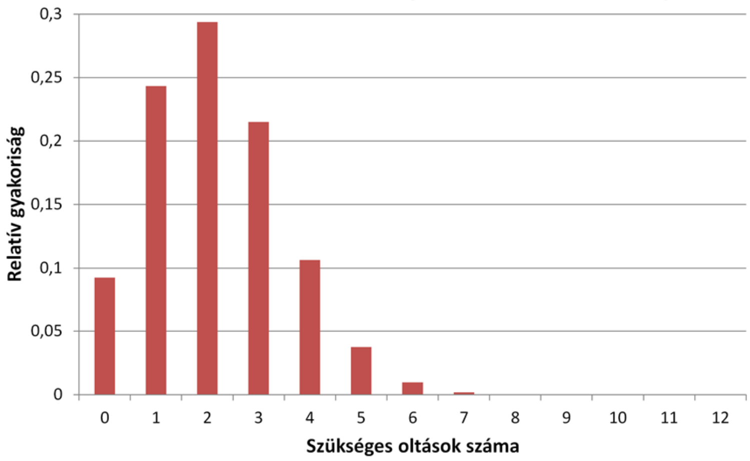
$$E(\xi) = \frac{1}{2}(a+b)$$

$$Var(\xi) = \frac{1}{12}(b-a)^2$$

Ideális kocka eredményeinek eloszlása
Ideális munkaterhelés eloszlása a nap folyamán
Hőmérséklet eloszlása egy üres terem különböző pontjain

Először tekintsük meg az *egyenletes eloszlást*.
Egyenletes eloszlást mutat például az ideális kocka dobáseredményei, vagy az ideális napi munkaterhelés (minden másodpercben ugyanannyire kell dolgoznunk), illetve a hőmérséklet eloszlása egy üres terem különböző pontjain. Ennek alapján például megadhatom annak a valószínűségét, hogy 4-est dobok. A várható érték és a variancia képletében az *a* és a *b* a legkisebb, illetve legnagyobb érték. Ezek alapján tehát a 6 oldalú dobókockával való dobás várható értéke: 0,5*(1+6)=3,5.

Binomiális (Bernoulli) eloszlás



$$E(\xi) = n \cdot p$$

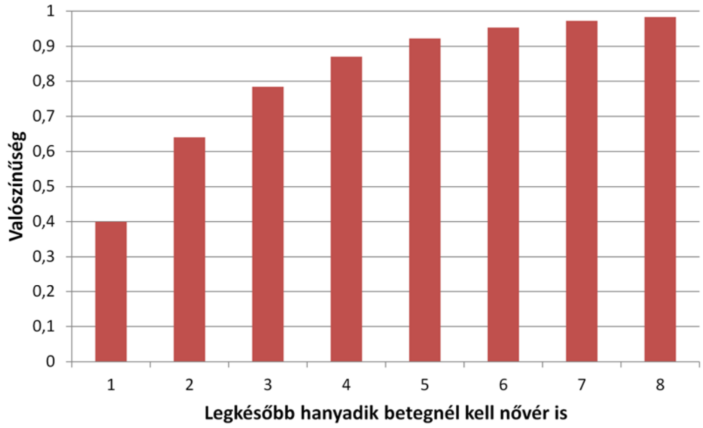
$$Var(\xi) = n \cdot p \cdot (1-p)$$

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

Szükséges oltóanyagszám eloszlása
Általánosan: egy n-szer megismételt jelenség x-szer következik be
Ha p „kicsi" Poisson eloszláshoz „közelít"
Ha n „nagy" és p 0,5-höz tart, akkor normál eloszláshoz „közelít"

A *binomiális (Bernoulli) eloszlást* használunk általánosságban, ha egy dologot n-szer ismételve egy esemény k-szor következik be ez alatt.
A korábban említett példában az n – az ismétlések száma – a megvizsgálandó paciensek számát jelenti, a k pedig a raktáron lévő oltóanyagok száma. Látható, hogy a várható érték, illetve a szórás (és így az adott értékhez tartozó egyenlet) meghatározásához szükséges tudnunk (vagy becsülnünk) a bekövetkezés valószínűségét – példánkban az oltóanyag szükségességének relatív gyakoriságát.
Megjegyzendő, hogyha az esemény bekövetkezésének valószínűsége (az ismétlésszámhoz képest) kicsi, akkor a binomiális eloszlás a Poisson eloszláshoz közelít. Ha az ismétlések száma nagy és p értéke 0,5-höz tart, akkor pedig a normál eloszláshoz válik hasonlóvá.

Geometriai eloszlás



$$E(\xi) = \frac{1}{p}$$

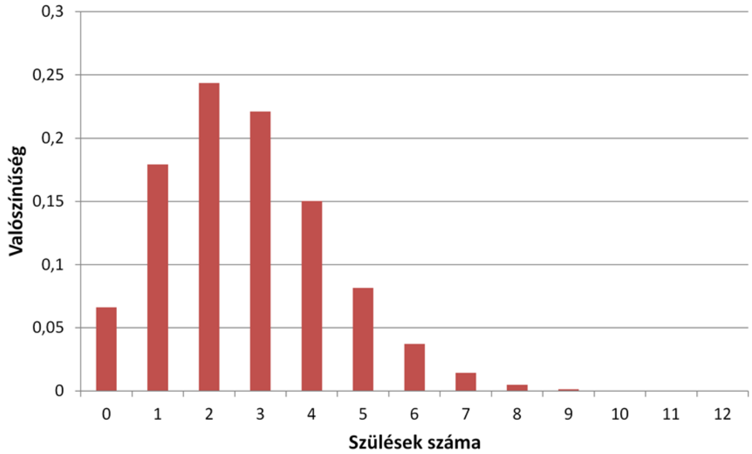
$$Var(\xi) = \sqrt{\frac{1-p}{p^2}}$$

$$P = p \cdot (1-p)^{(n-1)}$$

Független Bernoulli kísérletek egymásutánja
Leghamarabb hanyadik dobás lesz fej (Péter és Pál)
Legkésőbb hanyadik beteghez kell a nővér is

A *geometriai eloszlás* egy speciális Bernoulli eloszlás. Ezt az eloszlást kapjuk, ha egymástól független Bernoulli próbákat hajtunk végre egymás után (megtörténik az esemény, vagy nem történik meg).
Egy orvosi példa erre az eloszlásra: Mekkora a valószínűsége, hogy szükségünk van a nővér segítségére az első betegnél? Vagy csak a másodiknál kell hívnunk segítséget... Az ábrán azt tüntettem fel az első oszlopban, hogy mekkora a valószínűsége annak, hogy már az első betegnél kell segítség, a második oszlopban, hogy vagy az első vagy a második (vagy mindkettő) esetben kell segítség... tehát itt egy kumulált valószínűséget tüntettem fel, amit úgy fogalmazhatunk meg, hogy annak a valószínűsége, hogy legkésőbb hanyadik betegnél kell nővér segítsége is.
Természetesen ugyanilyen eloszlást mutat pétervári paradoxon játéka is.

Poisson eloszlás



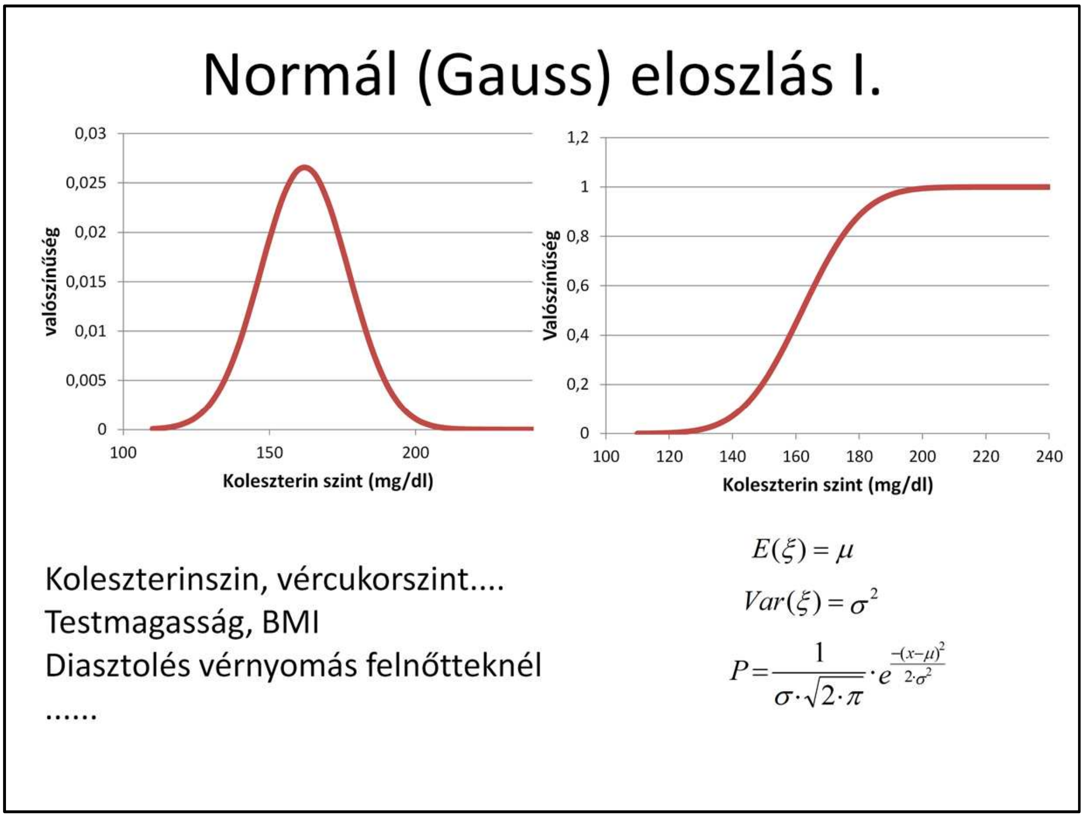
$$E(\xi) = \lambda$$

$$Var(\xi) = \lambda$$

$$P = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Szülések száma az ügyeletben
Új diagnosztizált karcinómák száma
Hálóban lévő halak száma
Radioaktív preparátumban adott idő alatt elbomló atomok száma
Normál eloszláshoz „közelíthető”

A *Poisson eloszlásnak* van egy különlegessége: *várható értéke és szórása megegyezik*, valamint *egyetlen paraméterrel megadható*.
Például ennek az eloszlásnak a segítségével megadhatjuk, hogy mekkora annak a valószínűsége, hogy 3 szülés történik az éjszakai ügyeletben. További példák az új diagnosztizált karcinómák számának, illetve a hálóban lévő halak számának eloszlása.
Általánosságban Poisson eloszlást követ a változónk, ha a kérdés egy adott elemszám egy adott intervallum alatt, illetve adott térfogatban, területen..., *ha az esemény bekövetkezésének valószínűsége kicsi*. (Példánkban a szülés valószínűsége kicsi – az emberek kis részével történik meg egy adott időben.)
A várható érték – a λ – az ismétlések számának és az adott esemény bekövetkeztének valószínűségének szorzatából adódik (n*p).
Megjegyzendő, hogy megfelelő körülmények esetében ez az eloszlás is normál eloszlással közelíthető.



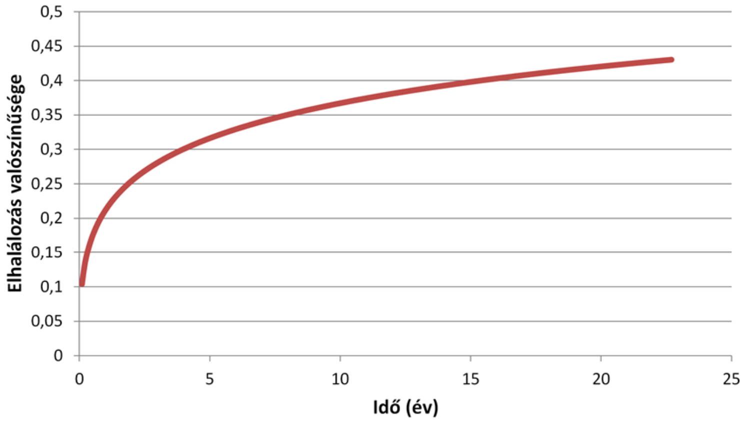
A *normál (Gauss) eloszlásnak* a fentebb már említetteken kívül további különlegessége van: *ez az orvosi gyakorlatban a leggyakoribb eloszlás*.
Az ábrán az eloszlás sűrűségfüggvényét és kumulatív eloszlásgörbéjét egyaránt feltüntettem, ugyanis ez nagyon fontos eloszlás számunkra. Mint látható, az előzőekkel ellentétben a normál eloszlás egy szimmetrikus eloszlás.
Látható, hogy valóban sok a változó, amely normál eloszlást követ: a koleszterinszint, a vércukorszint, a legtöbb enzimszint, a testmagasság, a BMI, a disztolés vérnyomás.... De vajon miért van ez?

Normál eloszlás II.

Centrális határeloszlás tétele: ha sok független valószínűségi változót összegzünk, akkor elég általános feltételek teljesülése esetén az összeg normális eloszlású valószínűségi változó lesz.

A normál eloszlás gyakoriságának okára a *centrális határeloszlás tétele* utal. Ez kimondja, hogy ha sok független valószínűségi változót összegzünk, akkor elég általános feltételek teljesülése esetén az összeg normális eloszlású valószínűségi változó lesz.
Az emberi test különböző jellemzői (mérhető értékei, változói) általában nagy sok más változó együtteséből alakulnak ki. Például az emberi testmagasság függ az apai és anyai génektől, a táplálkozástól, az életviteltől...

Lognormál (Galton) eloszlás



$$E(\xi) = e^{\mu+\sigma^2/2}$$

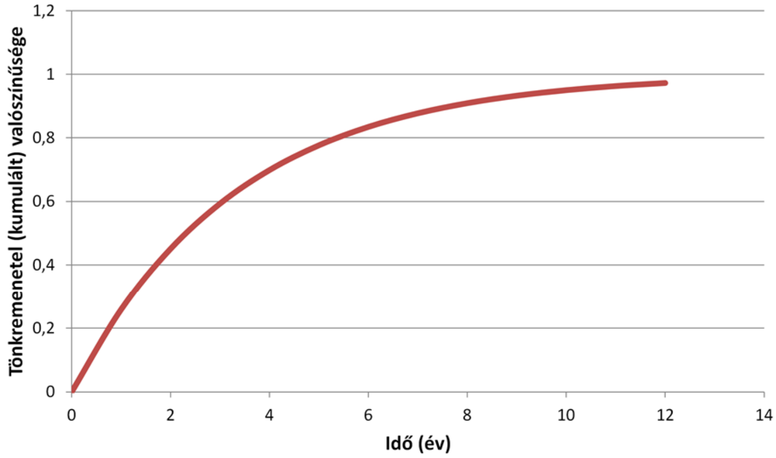
$$Var(\xi) = (e^{\sigma^2} - 1) \cdot e^{2\mu+\sigma^2}$$

$$P = \frac{1}{\sigma \cdot x \cdot \sqrt{2 \cdot \pi}} \cdot e^{\frac{-(\ln x - \mu)^2}{2 \cdot \sigma^2}}$$

Testtömeg, testmagasság gyermekkorban
Túlélési idő

A l*ognormál* eloszlásnak is nagy jelentősége van az orvosi gyakorlatban. Ilyen eloszlást követ például a testtömeg, testmagasság gyermekkorban, illetve a túlélési idő rosszindulatú daganatoknál.
Általánosságban azt mondhatjuk, hogy akkor lesz az eloszlásunk lognormál, ha a változónk értékei kicsik, de nem lehetnek 0-nál kisebbek.
(Az eloszlást azért nevezzük lognormálnak, mert a változó értékeinek logaritmizálásával normális eloszláshoz jutunk.)

Exponenciális eloszlás



$$E(\xi) = \frac{1}{\lambda}$$

$$Var(\xi) = \frac{1}{\lambda^2}$$

$$P = \lambda \cdot e^{-\lambda \cdot x}$$

Altatóberendezés működési ideje (az első hibáig).
Radioaktív bomlás során az egyes atomok élettartama.

Az *exponenciális eloszlás* igen gyakori a biofizikában és néhol van szerepe az orvosi gyakorlatban is. Ehhez két példát említenék: altatóberendezés működési ideje (az első hibáig eltelt idő), illetve radioaktív bomlás során az egyes atomok élettartama.

Az emberi gondolkodás...

Tomi csendes, visszahúzódó, szerény, szorgalmas fiú, aki másoknak szívesen segít. Melyiket tartod valószínűbbnek:

a) Tomi könyvtáros
b) Tomi kétkezi munkás

A mindennapi gondolkodásunk sok esetben eltér attól, ahogyan annak a valószínűségszámítás ismeretében működnie kellene.
A dián látható két állítás közül sokan közületek az elsőt (Tomi könyvtáros) tartotta valószínűbbnek. Pedig belegondolva abba, hogy mennyire gyakori a könyvtárosi állás, illetve a kétkezi munkás, egyértelmű, hogy annak nagyobb a valószínűsége, hogy Tomi kétkezi munkás.

Az emberi gondolkodás...

Linda tehetséges, független, filozófia szakot végzett 31 éves nő. Nagyon érzékeny a társadalmi igazságtalanságokra. Diákként részt vett az antinukleáris demonstrációkban. Sorszámozza meg az alábbi állításokat aszerint, hogy mennyire tartja valószínűnek (1-es sorszám a legvalószínűbb):

a) Linda tanító egy általános iskolában,
b) Linda könyvesboltban dolgozik, és jóga tanfolyamra jár,
c) Linda a nőszavazók ligájának tagja,
d) Linda bankpénztáros,
e) Linda biztosítási ügynök,
f) Linda bankpénztáros és feminista.

A következő példában két állításra hívnám fel a figyelmet: Linda bankpénztáros, illetve Linda bankpénztáros és feminista. Az előadás során többetek előrébb sorolta azt, hogy Linda bankpénztáros és feminista, mint azt, hogy Linda bankpénztáros. Pedig ismerve a 11. dián is leírt független események valószínűségének együttes előfordulási valószínűségét, beláthatjuk, hogy egy esemény valószínűsége mindig legalább ugyanakkora, mint ugyanezen esemény és egy másik esemény együttes előfordulásának valószínűsége...

Valószínűség másképp...

Az előadás végén kísérletet tettünk a pí értékének kísérleti meghatározására. Az elvégzett kísérletet hívják Buffon-féle tűproblémának és ehhez kapcsolódik a geometriai valószínűség fogalma.

Ellenőrző kérdések #1

- Definiáld a valószínűséget a nagy számok törvénye alapján.
- Ismertesd a nagy számok törvényét.
- Hogyan bizonyítható a nagy számok törvénye?
- Hogyan jelölhető az A vagy B esemény bekövetkezésének valószínűsége?
- Hogyan jelölhetjük azt a valószínűséget, hogy A és B esemény egyaránt bekövetkezik?
- Mit jelent két esemény metszete, illetve uniója?
- Definiált az egymást kölcsönösen kizáró eseményeket.
- Mondj példát az egymást kölcsönösen kizáró eseményekre.
- Mit tudsz az egymást kölcsönösen kizáró események metszetéről?
- Definiáld az egymástól független események fogalmát.
- Mondj példát független eseményekre.
- Mit jelent a feltételes valószínűség?
- Mondj példát a feltételes valószínűségre.
- Hogyan jelöljük a feltételes valószínűséget?
- Hogyan számítható P(A) ha P(A|B) és P(B) adott?
- Mik a Kolmogorov axiómák?
- Mit tudsz A és B esemény viszonyáról, ha P(AvagyB)=P(A)+P(B) igaz?
- Mit tudsz A és B esemény viszonyáról , ha P(AB)=P(A)*P(B) igaz?
- Mekkora a biztos esemény valószínűsége?
- Mekkora a lehetetlen esemény valószínűsége?
- Adj példát biztos és lehetetlen eseményekre.
- Mekkora lehet egy esemény valószínűségének értéke?
- Mit jelent az esély?
- Definiáld a logitot.
- Add meg az esemény logit értékét, ha az esemény valószínűsége 0,12.
- Mekkora az esély értéke, ha a valószínűség 0,4.
- Add meg a valószínűséget, ha az esély 3.
- Add meg a valószínűséget, ha a logit – 32.

A kérdések megválaszolhatók az előadáson elhangzottak, a gyakorlatvezetővel folytatott konzultációk, illetve saját utánaolvasás segítségével. Az ellenőrző kérdések egyben példák arra, hogy milyen tesztkérdések (feleletválasztós formában) fordulhatnak elő.

Ellenőrző kérdések #2

- Mondj egy példát ismétlés nélküli kombinációra.
- Hogyan számítható egy folytonos eloszlás várható értéke?
- Hogyan számítható egy diszkrét eloszlás várható értéke?
- Melyik középérték egyezik meg a várható értékkel egy populáció esetében?
- Mivel becsülhető egy elméleti eloszlás várható értéke?
- Mivel becsülhető egy elméleti eloszlás szórása?
- Definiáld a z elméleti varianciát.
- Melyik két mutató határoz meg egyértelműen egy speciális eloszlást?
- Ábrázold az egyenletes eloszlás gyakoriságfüggvényét.
- Ábrázold a Poisson eloszlás gyakoriságfüggvényét.
- Ábrázold a Bernoulli eloszlás gyakoriságfüggvényét.
- Ábrázold a geometriai eloszlás gyakoriságfüggvényét.
- Ábrázold a normál eloszlás gyakoriságfüggvényét.
- Ábrázold a Gauss eloszlás kumulált eloszlásfüggvényét.
- Ábrázold az exponenciális eloszlás gyakoriságfüggvényét.
- Ábrázold a lognormál eloszlás gyakoriságfüggvényét.
- Írj 2 példát az egyenletes eloszlásra.
- Írj 2 példát binomiális eloszlásra.
- Írj 2 példát a geometriai eloszlásra.
- Írj 3 példát a normál eloszlásra.
- Írj 2 példát a lognormál eloszlásra.
- Írj 2 példát a Poisson eloszlásra.
- Írj 2 példát az exponenciális eloszlásra.
- Hogyan számítható az egyenletes eloszlás várható értéke?
- Hogyan számítható a binomiális eloszlás várható értéke?
- Hogyan számítható a lognormál eloszlás várható értéke?
- Hogyan számítható az exponenciális eloszlás várható értéke?
- Hogyan számítható a Poisson eloszlás várható értéke?
- Hogyan számítható normál eloszlás várható értéke?
- Miről szól a centrális határeloszlás tétele?
- Miért követ a legtöbb orvosi gyakorlatban használt változó normál eloszlást?

Ellenőrző kérdések #3

- Add meg általánosan, hogy mikor kapunk általában binomiális eloszlást.
- Add meg általánosan, hogy mikor kapunk általában geometriai eloszlást.
- Add meg általánosan, hogy mikor kapunk általában Poisson eloszlást.
- Add meg általánosan, hogy mikor kapunk általában lognormál eloszlást.
- Lehet két esemény együttes bekövetkezésének valószínűsége nagyobb az egyes események bekövetkezésének valószínűségénél?