A valószínűségszámítás és a statisztika alapvonalai

Ferenci Tamás, tamas.ferenci@medstat.hu

2021. február 26.

Tartalom

Tartalom

Előszó

A valószínűségszámítást matematikai formalizmus nélkül fogjuk tárgyalni

A formalizálás kell a mélyebb tárgyaláshoz, azonban sok hallgató (orvosok, szociológusok stb.) számára ezek a részek úgysem fontosak

Viszont a matematikai formalizmus azt is megnehezíti, hogy az alapkérdéseket megértsék

Minden visszajelzést örömmel veszek a tamas.ferenci@medstat.hu email-címen

Tartalom

A valószínűségszámítás röviden

A valószínűségszámítás bizonyos értelemben a statisztika „alaptudománya”

Éppen ezért a statisztika megértéséhez is fontos érteni az alapjait (és önmagában is érdekes!)

Mi az, hogy véletlen?

A valószínűségszámítás alapfogalma a **véletlen kísérlet**: adott körülmények között akárhányszor megfigyelhető, de a kimenete nem mondható meg biztosan (csak az, hogy egyáltalán milyen kimenetek lehetségesek)

Valójában nagyon sokszor nem a szó szokásos értelemben vett „kísérletről” van szó és az sem nyilvánvaló, hogy milyen értelemben figyelhető meg akárhányszor (pl. a beteg meggyógyul-e)

Az is döntés kérdése, hogy mi véletlen, aminek „nem mondható meg biztosan” a kimenetet (pl. a kockadobás az? elvileg nem!) – közelítés kérdése

A valószínűségszámítás alapfogalmai 1.

A véletlen kísérlet lehetséges kimeneteleit **kimenetelnek** nevezzük (pl. kockadobásnál az, hogy ⚁)

Az összes lehetséges kimenetet egybefogjuk egy halmazba (kockadobásnál {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}), ennek neve **eseménytér**

Az eseménytér lehet véges (kockadobás) vagy végtelen, ez utóbbi esetben lehet megszámlálhatóan végtelen (hányszor járt egy véletlenszerűen választott magyar lakos külföldön) vagy nem megszámlálhatóan végtelen (mennyi a testtömege)

A véges vagy legfeljebb megszámlálhatóan végtelen eseménytereket együtt **diszkrét** eseménytérnek, a nem megszámlálhatóan végtelent **folytonos** eseménytérnek nevezzük

Valószínűségeket nem a kimenethez fogunk rendelni, mert szeretnénk bonyolultabb dolgokhoz (pl. „párosat dobtunk”) is valószínűségeket rendelni

A valószínűségszámítás alapfogalmai 2.

Kimenetekből egyet vagy többet bepakolunk egy halmazba (pl. {⚁, ⚃, ⚅}), ezt **eseménynek** hívjuk, és úgy definiáljuk, hogy bekövetkezik, ha olyan kimenetel következik be, ami benne van ebben a halmazban

(Lehet egyelemű halmaz is)

Biztos esemény: az egész eseménytér, lehetetlen esemény: üres halmaz (gondoljuk végig, teljesen logikus elnevezések)

Ilyen módon események „verbális" műveleteinek megfeleltethetünk halmaz-műveleteket

Pl. „4-nél nem nagyobb és páros számot dobunk" a „4-nél nem nagyobb számot dobunk" és a „páros számot dobunk" halmazainak a metszete: az „és"-nek megfelel a metszet

Hasonlóan a „vagy"-nak az unió

A valószínűségszámítás alapfogalmai 3.

Miért nem rendelünk egész egyszerűen az összes lehetséges eseményhez valószínűséget?

Diszkrét eseménytérnél működne, de folytonosnál nem, ott csúnya dolgok történnének, ha ezt megpróbálnánk

Éppen ezért bevezetünk egy halmazt, mely azokat az eseményeket tartalmazza, amelyekhez egyáltalán szeretnénk valószínűséget rendelni, ezt hívjuk a **megfigyelhető események** halmazának

A valószínűség bevezetése

A **valószínűség** egy függvény, amely a megfigyelhető események mindegyikéhez hozzárendel egy valós számot (pl. {⚁, ⚃, ⚅} $\mapsto$ 0,3, ekkor 0,3 valószínűséget rendeltünk ehhez az eseményhez, köznyelvileg ezt 30%-nak mondjuk)

Jele $\mathbb{P}$, előbbi példánkban $\mathbb{P}(A) = \mathbb{P}($ {⚁, ⚃, ⚅} $) = 0{,}3$

De nem akárhogy rendeljük hozzá: a valószínűségek nemnegatívak, a biztos esemény valószínűsége 1, és kizáró események uniójának a valószínűsége a valószínűségeik összege (ez utóbbi is jól érthető, ha a valószínűségre mint pacák területére gondolunk egy festővásznon)

Ezek axiómák (nem levezethetőek valamiből)

A valószínűség interpretációi 1.

Valószínűség interpretácója: a fenti matematikai konstrukciónak mi köze van valós jelenségekhez

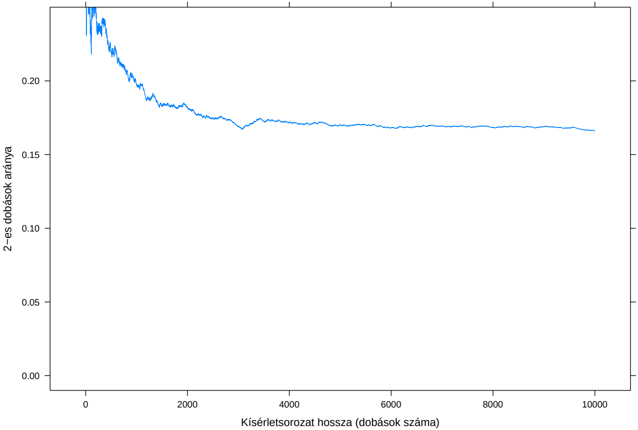
Legfontosabb interpretáció: elkezdünk kockát dobálgatni, és számoljuk a 2-es dobások relatív gyakoriságát

Azt fogjuk tapasztalni, hogy ez konvergál valamihez – ez a valami legyen a valószínűség!

Ez a valószínűség **frekventista interpretációja**

Azért bőven van további interpretációknak is hely (pl. ..30% a valószínűsége annak, hogy holnap esni fog az eső")

A valószínűség interpretációi 2.



A feltételes valószínűség 1.

Alapkérdés: hogy tudunk valószínűségbe valamilyen információt beépíteni

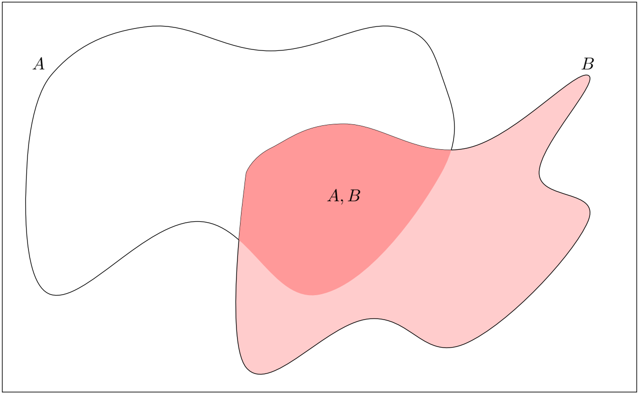
Ha tudjuk, hogy valami megtörtént, az hogyan módosítja a valószínűséget

Pl. *feltéve*, hogy legfeljebb 3-ast dobtunk, mennyi annak a valószínűsége, hogy párosat dobtunk?

Arra szűkítve a világunkat, ahol a „legfeljebb 3-at dobtunk” megtörtént, ezen a világon belül mekkora a páros dobás valószínűsége?

Nagyon jól megérezhető a megoldás, ha a pacás analógiára gondolunk: az egyik pacán *belül* mekkora a másik paca területe?

A feltételes valószínűség 2.



A feltételes valószínűség 3.

Így a **feltételes valószínűség** definíciója:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)},$$

Ismert információ (B) beépítésével „frissítettük" A valószínűségét

$\mathbb{P}(A)$ a valószínűség a plusz-információ megismerése előtt, ezért neve **prior valószínűség**, $\mathbb{P}(A \mid B)$ a valószínűség a plusz-információ megismerése után, ezért neve **poszterior valószínűség**

Átrendezve: $\mathbb{P}(A, B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B)$

„Együttes egyenlő feltételes, szorozva a feltétel valószínűségével"

Ha $\mathbb{P}(A \mid B) = \mathbb{P}(A)$: B ismerete nem változtatja meg A valószínűségét, ilyenkor azt mondjuk, hogy ezek **független** események

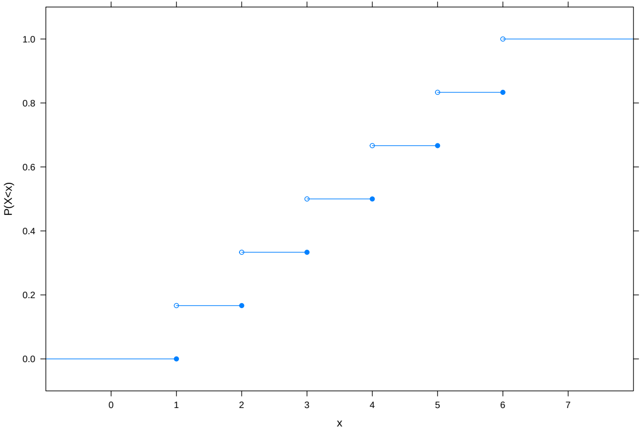
Szimmetrikus: ha ez fennáll, akkor $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ is

A feltételes valószínűség 4.

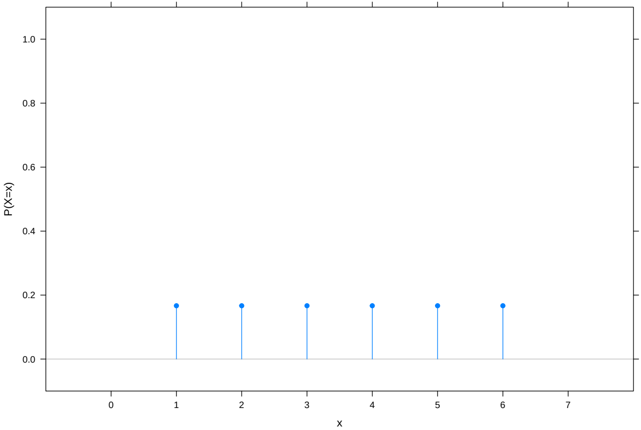
És mindezek egyenértékűek azzal, hogy $\mathbb{P}(A, B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

A Bayes-tétel

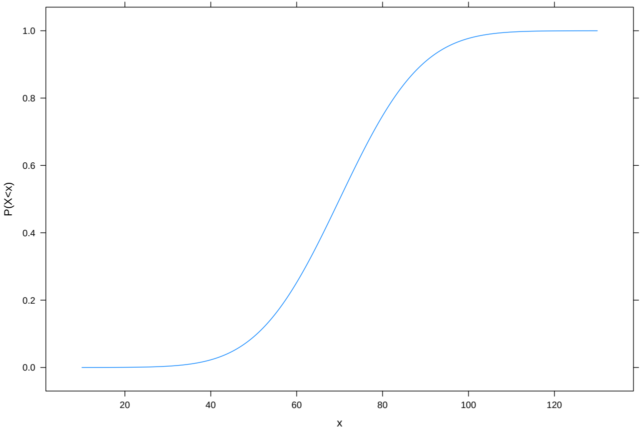
A valószínűségi változó 1.



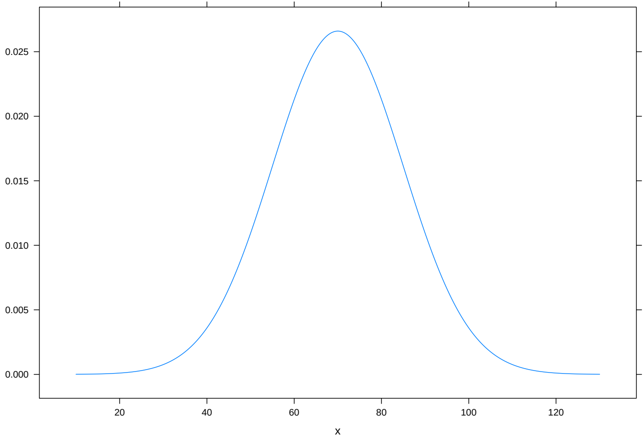
A valószínűségi változó 2.



A valószínűségi változó 3.

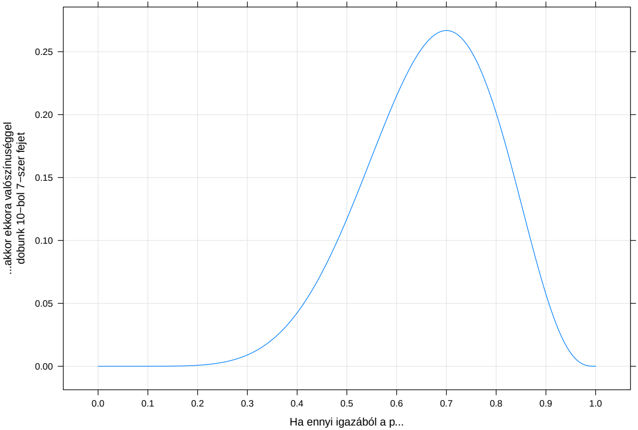


A valószínűségi változó 4.



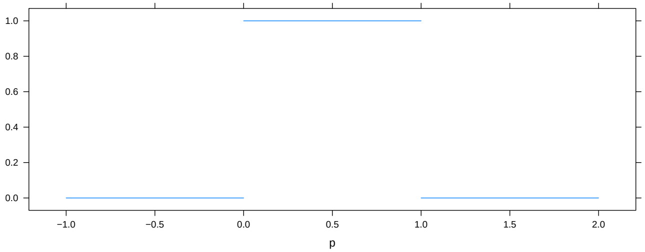
Tartalom

Egy rávezető gyakorlat

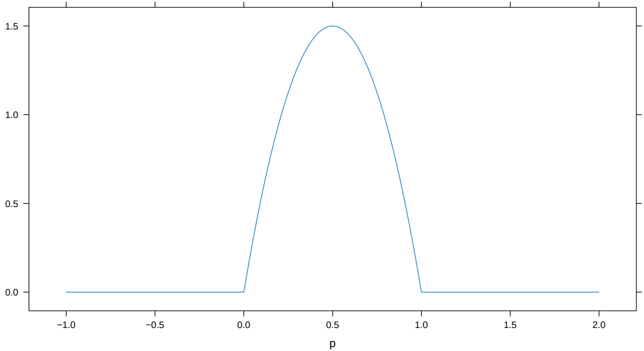


Egy becslőfüggvény felé

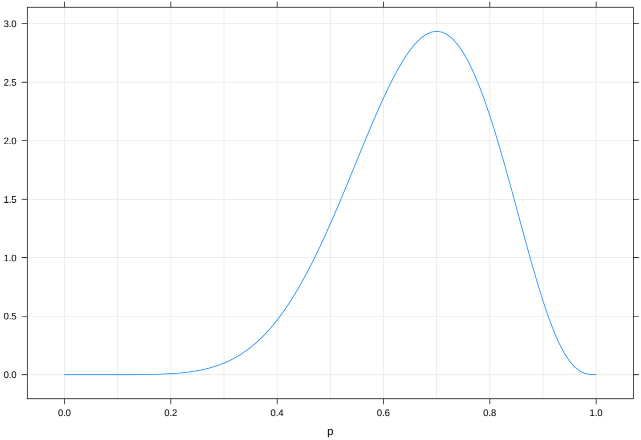
A bayes-i probléma 1.



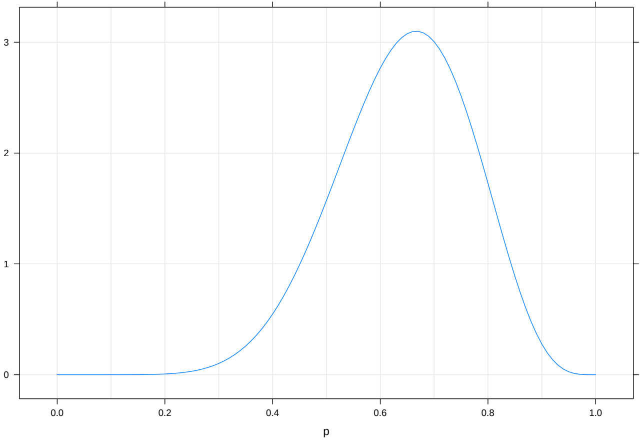
A bayes-i probléma 2.



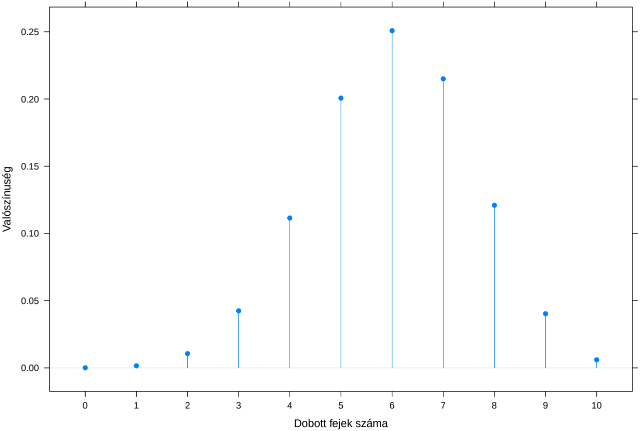
A bayes-i probléma 3.



A bayes-i probléma 4.



A frekventista statisztika építménye



Tartalom

Intervallumbecslés

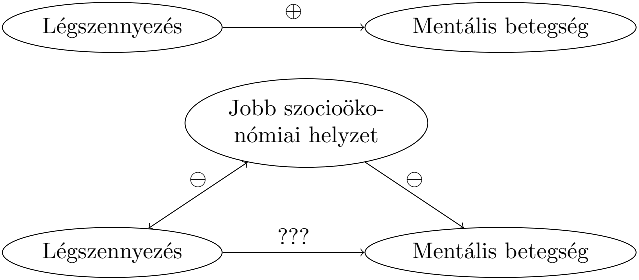
Hipotézisvizsgálat

Tartalom

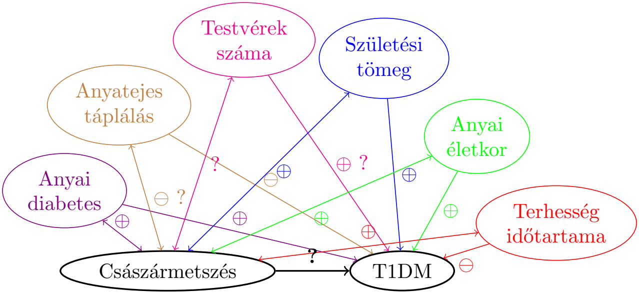
Az orvosi megismerés alapkérdései

Az okozatiság nyomában

A confounding problémája 1.



A confounding problémája 2.



Zavaró változóktól a megzavart olvasókig

Egy aranyérmes megoldás

Megfigyelés és kísérlet

A jó, a rossz, és a közepesnél némileg gyengébben jó