A valószínűségszámítás és a statisztika alapvonalai

Ferenci Tamás, tamas.ferenci@medstat.hu

2019. december 7.

Tartalom

Tartalom

Előszó

- A valószínűségszámítást matematikai formalizmus nélkül fogjuk tárgyalni

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

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

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”

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)

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)

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

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

- A véletlen kísérlet lehetséges kimeneteleit **kimenetelnek** nevezzük (pl. kockadobásnál az, hogy ⚁)
- Az összes lehetséges kimenetet egybefogjuk egy halmazba (kockadobásnál {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}), ennek neve **eseménytér**

A valószínűségszámítás alapfogalmai

- 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 valószínűségszámítás alapfogalmai

- 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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

A valószínűségszámítás alapfogalmai

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

- Miért nem rendelünk egész egyszerűen az összes lehetséges eseményhez valószínűséget?

A valószínűségszámítás alapfogalmai

- Miért nem rendelünk egész egyszerűen az összes lehetséges eseményhez valószínűséget?
- Diszkrét eseménytérnél működne, de folytonosnál nem, ott csúnya dolgok történnének, ha ezt megpróbálnánk

A valószínűségszámítás alapfogalmai

- 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)

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

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)

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

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

A valószínűség interpretációi

- Valószínűség interpretácója: a fenti matematikai konstrukciónak mi köze van valós jelenségekhez
- Legfontosabb interpretáció: elkezdünk kockát dobálgatni, és számoljuk a 2-es dobások relatív gyakoriságát

A valószínűség interpretációi

- 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!

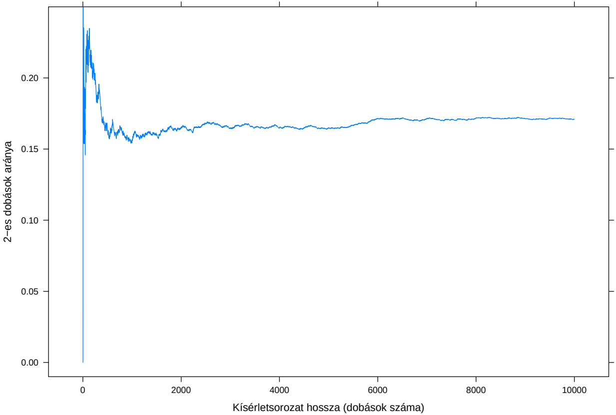
A valószínűség interpretációi

- 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**

A valószínűség interpretációi

- 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



A feltételes valószínűség

- Alapkérdés: hogy tudunk valószínűségbe valamilyen információt beépíteni

A feltételes valószínűség

- Alapkérdés: hogy tudunk valószínűségbe valamilyen információt beépíteni
- *Ha* tudjuk, hogy valami megtörtént, az hogyan módosítja a valószínűséget

A feltételes valószínűség

- 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?

A feltételes valószínűség

- 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?

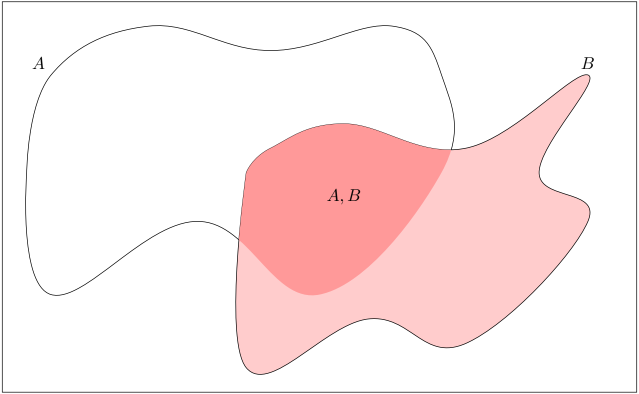
- Arra szűkítve a világunkat, ahol a „legfeljebb 3-at dobtunk” megtörtént, ezen a világon belül mekkora a páros dobás valószínűsége?

A feltételes valószínűség

- 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?

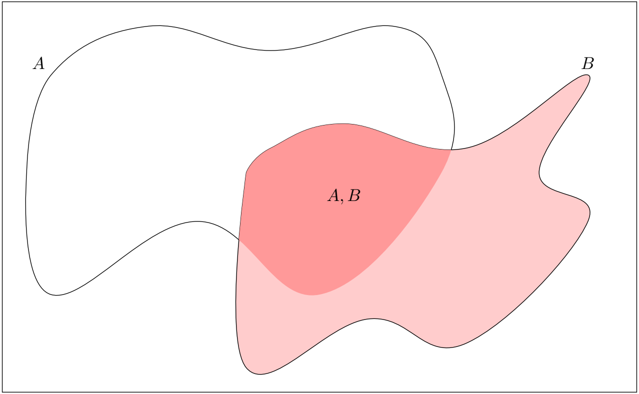
A feltételes valószínűség



- Így a **feltételes valószínűség** definíciója:

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

A feltételes valószínűség

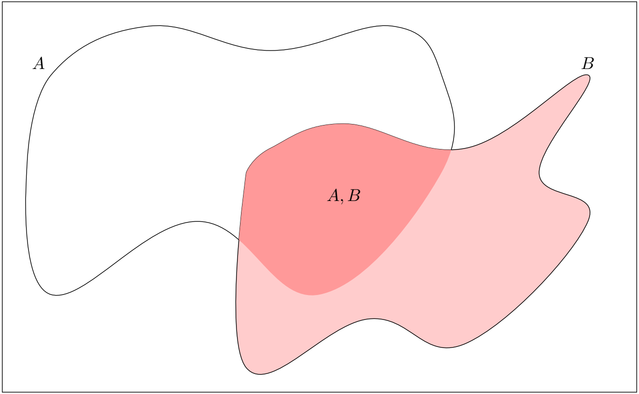


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A feltételes valószínűség

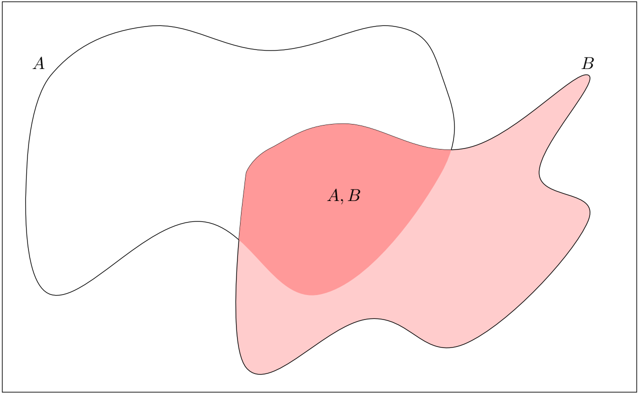


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük" A valószínűségét

A feltételes valószínűség

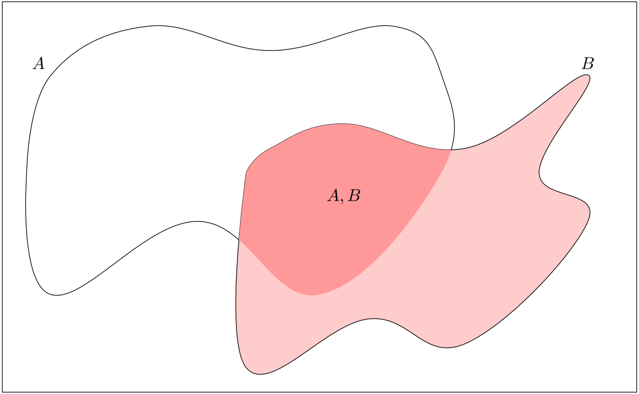


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A feltételes valószínűség

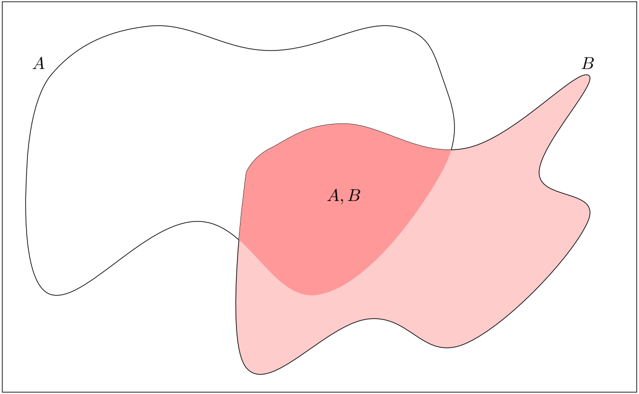


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A feltételes valószínűség

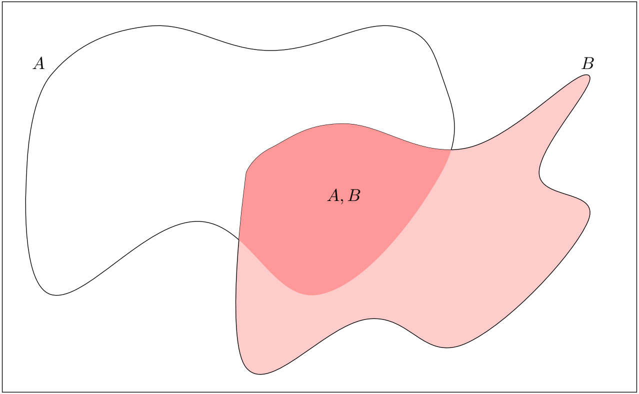


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A feltételes valószínűség

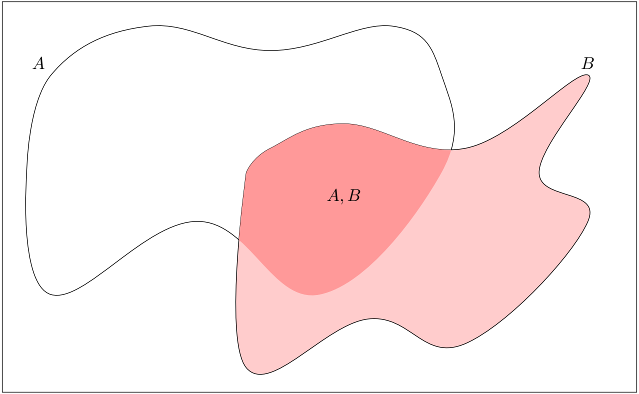


- Így a **feltételes valószínűség** definíciója:

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

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A feltételes valószínűség



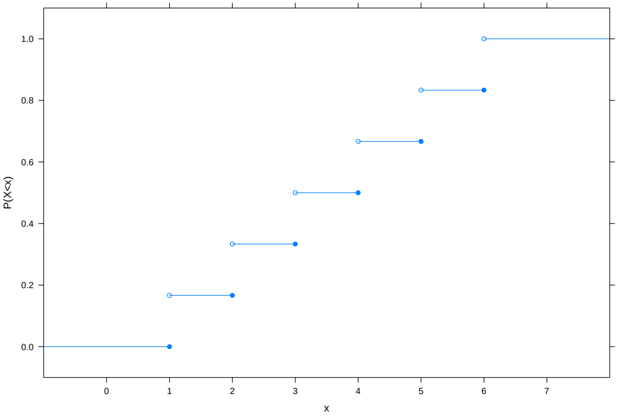
- Így a **feltételes valószínűség** definíciója:

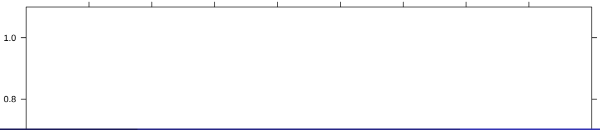
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)},$$

- Ismert információ (B) beépítésével „frissítettük” A valószínűségét

A Bayes-tétel

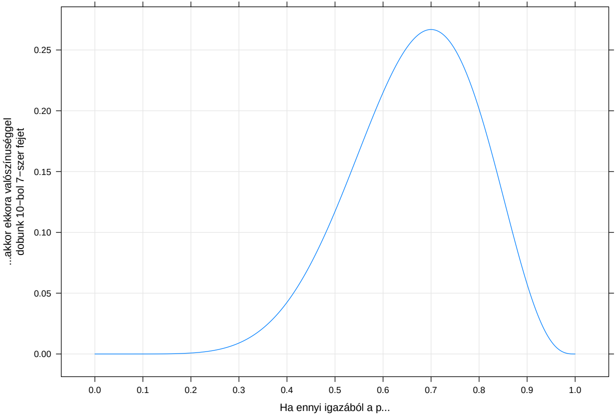
A valószínűségi változó





Tartalom

Egy rávezető gyakorlat



Egy becslőfüggvény felé

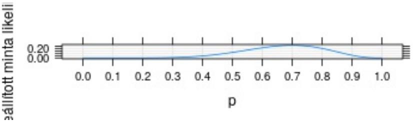
Likelihood különböző minták mellett

Dobások száma:

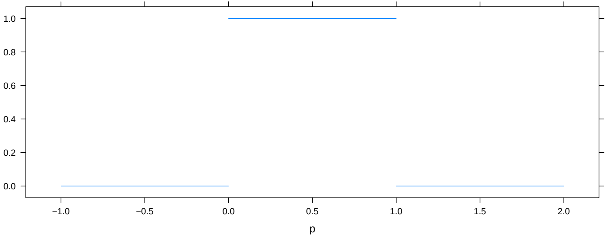
1 10 100

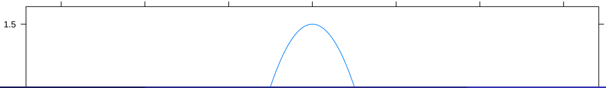
Fejek száma:

0 7 10

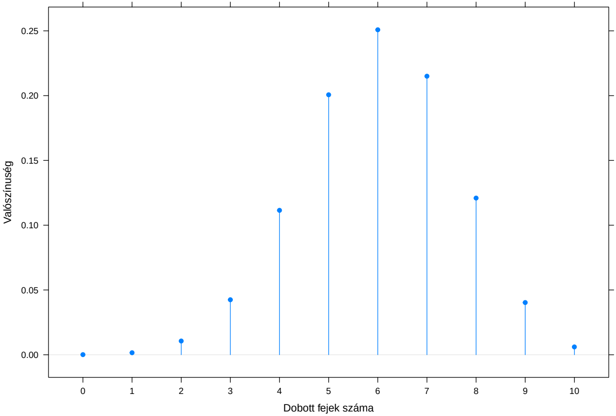


A bayes-i probléma





A frekventista statisztika építménye



Becslés frekventista megközelítése

Dobások száma:

1 10 100

Tartalom

Intervallumbecslés

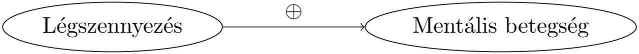
Hipotézisvizsgálat

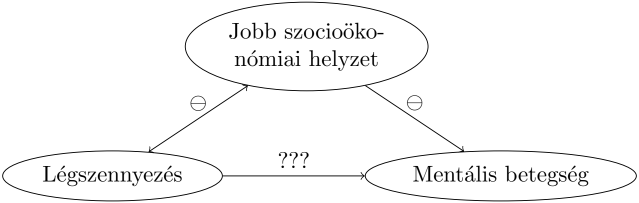
Tartalom

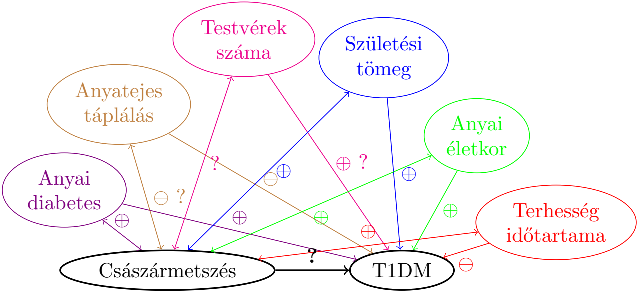
Az orvosi megismerés alapkérdései

Az okozatiság nyomában

A confounding problémája







Zavaró változóktól a megzavart olvasókig

Egy aranyérmes megoldás

Megfigyelés és kísérlet

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