Intelligens orvosi műszerek VIMIA023

Ha döntésünk költségeit is figyelembe vesszük

2018 ősz

http://www.mit.bme.hu/oktatas/targyak/vimia023

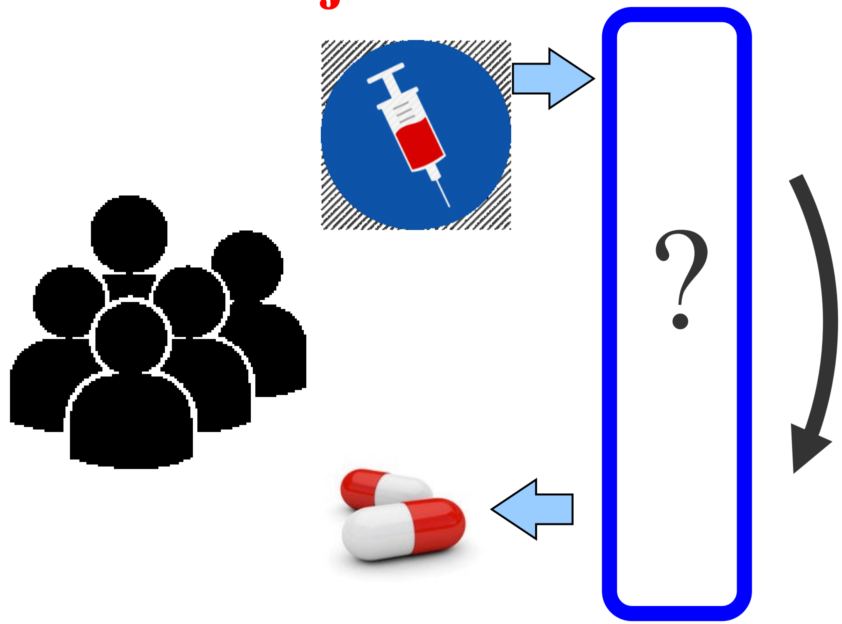
dr. Pataki Béla

pataki@mit.bme.hu

(463-)2679

Az észleléseink alapján térjünk át injekcióról tablettára!

Ez jó döntés?



Jó döntést szeretnénk hozni. De mi a jó döntés?

- <u>Általában nincs tökéletesen jó döntés</u>, csak az egyik jobb, a másik rosszabb!
- Ahhoz, hogy össze tudjunk hasonlítani két alternatívát, kell legyen <u>a döntés jóságát jellemző (skalár) mérőszámunk</u>
- Alapvetően bináris (jó/rossz, igaz/hamis, beteg/egészséges, ártatlan/bűnös) döntéseken mutatjuk be a módszereket

Mi a helyzet, ha a tévedéseknek különböző a költsége?

1. **típusú hiba hamis negatív (FN)**

   – azt hisszük egészséges (H0), pedig beteg (T1)

   - azt hisszük barát pedig ellenség

   költség: $\mathbf{C_{01}}$ (későn vesszük észre a bajt, bonyolult műtét vagy kezelés kell, azt hittük barát, pedig ellenség és lebombázza a hidat/gyárat)

2. **típusú hiba hamis pozitív (FP)**

   – azt hisszük beteg (H1), pedig egészséges (T0)

   - azt hisszük ellenség pedig barát

   költség: $\mathbf{C_{10}}$ (feleslegesen kezeljük vagy vizsgáljuk az egészségest, a baráti gépet lelőjük vagy vadászrepülőket ezt-azt küldünk rá)

legtöbbször a hamis negatív a rosszabb: $C_{01} >> C_{10}$

Ráadásul a helyes döntéseknek is van költsége!

valós negatív (TN)

– azt hisszük egészséges (H0), valóban az (T0)

- azt hisszük barát, és tényleg

költség: C_{00} vizsgálatot kell végeznünk

– orvosi szűrés, program teszt, katonai radar

valós pozitív (TP)

– azt hisszük beteg (H1), valóban az (T1)

- azt hisszük ellenség, és tényleg

költség: C_{11} vizsgálatot kell végeznünk + kezelnünk kell

– orvosi szűrés, program teszt, katonai radar

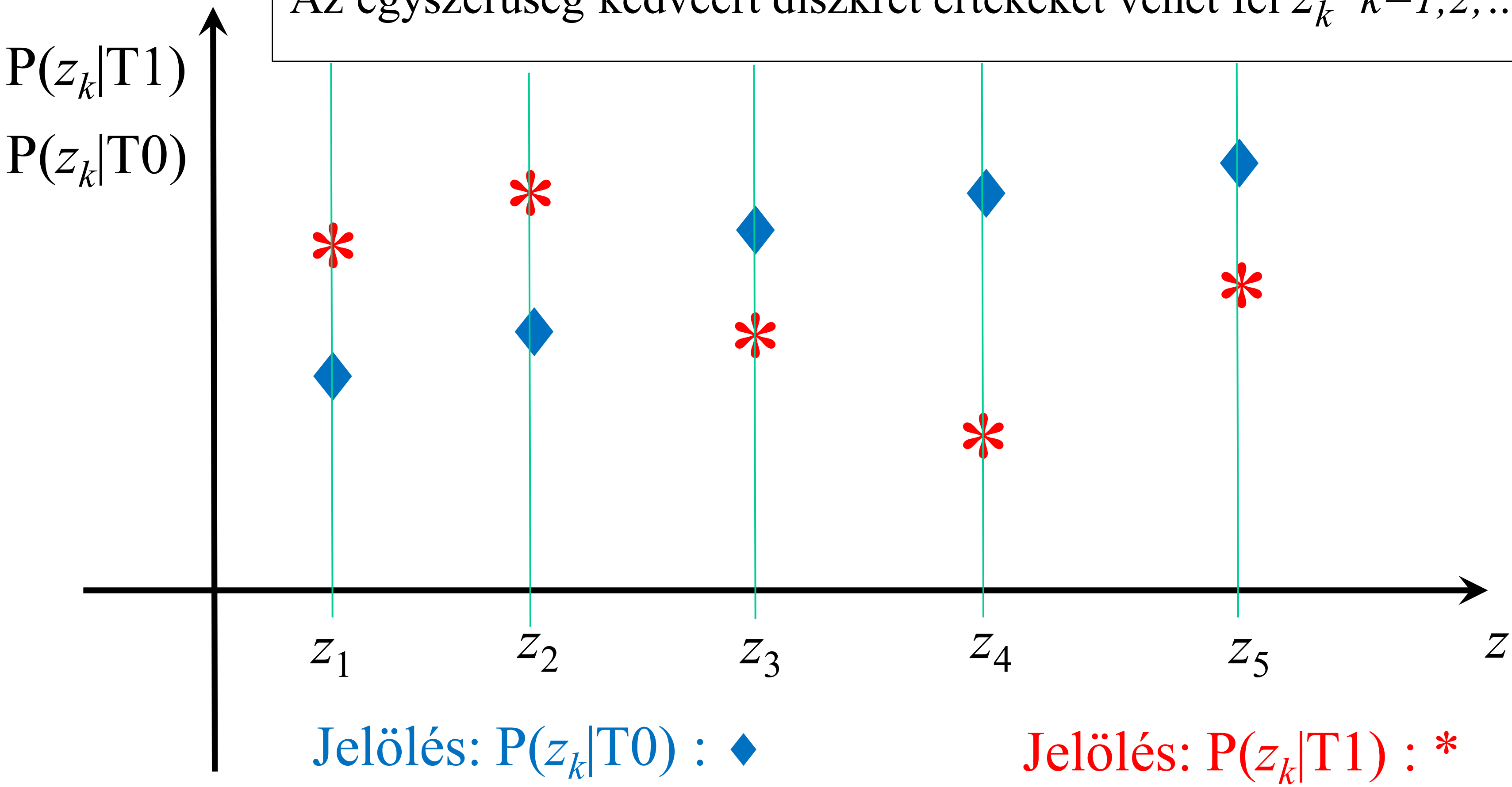
+ kezelés, légvédelmi rakéta, hibajavítás

tehát négyféle költség befolyásolja a döntést:

C_{00}, C_{11}, C_{01}, C_{01}

Skalár mért vagy származtatott paraméter, ami alapján döntünk : z

Az egyszerűség kedvéért diszkrét értékeket vehet fel z_k $k=1,2,\ldots,K$



Hogyan becsülhetjük meg ezeket az arányokat?

Egy tesztet végzünk (vagy az eddigi adatokból összegyűjtünk egy megfelelően nagy mintát).

Tudnunk kell, hogy milyen arányban produkálják a negatív és pozitív esetek az egyes lehetséges *z* értékeket.

Ezért tesztelünk N_T esetet, ezek közül N_N tartozik a negatív, N_P a pozitív csoportba ($N_N + N_P = N_T$)

k	**1**	**2**	**3**	**4**	**5**	**Σ**
z_k	z_1	z_2	z_3	z_4	z_5	
z_k-t produkál az negatív esetek közül	N_{N1} (pl. 1200)	N_{N2} (pl. 3100)	N_{N3} (pl. 4500)	N_{N4} (pl. 2700)	N_{N5} (pl. 1820)	$N_{N1}+N_{N2}+\ldots$ $\ldots+N_{N5} =$ $= N_N$ (N_N=13320)
$P(z_k\|T0)$	$\frac{N_{N1}}{N_N}$ $\left(=\frac{1200}{13320}\right)$	$\frac{N_{N2}}{N_N}$	$\frac{N_{N3}}{N_N}$	$\frac{N_{N4}}{N_N}$	$\frac{N_{N5}}{N_N}$	1
z_k-t produkál a pozitív esetek közül	N_{P1}	N_{P2}	N_{P3}	N_{P4}	N_{P5}	$N_{P1}+N_{P2}+\ldots$ $\ldots+N_{P5} =$ $= N_P$
$P(z_k\|T1)$	$\frac{N_{P1}}{N_P}$	$\frac{N_{P2}}{N_P}$	$\frac{N_{P3}}{N_1}$	$\frac{N_{P4}}{N_P}$	$\frac{N_{P5}}{N_P}$	1

Ha a populáció N elemű, és ebből a negatívok aránya $P_0 = \frac{N_0}{N}$, akkor a negatív esetek száma $N_0 = P_0 \cdot N$.
Az N_0 negatív esetből z_k értéket produkál $N_0 \cdot \mathrm{P}(z_k|\mathrm{T0})$ eset.

A pozitív esetek aránya $P_1 = \frac{N_1}{N}$, akkor a pozitív esetek száma $N_1 = P_1 \cdot N$.
Az N_1 pozitív esetből z_k értéket produkál $N_1 \cdot \mathrm{P}(z_k|\mathrm{T1})$ eset.

Ha úgy döntünk, hogy z_k esetén pozitívnak véleményezzük az ismeretlen esetet, akkor a költségünk:

$$\mathrm{K1} = \mathrm{C}_{10} \cdot \mathrm{N}_0 \cdot \mathrm{P}(z_k|\mathrm{T0}) + \mathrm{C}_{11} \cdot N_1 \cdot \mathrm{P}(z_k|\mathrm{T1}) =$$
$$= \mathrm{C}_{10} \cdot P_0 \cdot N \cdot \mathrm{P}(z_k|\mathrm{T0}) + \mathrm{C}_{11} \cdot P_1 \cdot N \cdot \mathrm{P}(z_k|\mathrm{T1})$$

Ha úgy döntünk, hogy z_k esetén negatívnak véleményezzük az ismeretlen esetet, akkor a költségünk:

$$\mathrm{K2} = \mathrm{C}_{00} \cdot \mathrm{N}_0 \cdot \mathrm{P}(z_k|\mathrm{T0}) + \mathrm{C}_{01} \cdot N_1 \cdot \mathrm{P}(z_k|\mathrm{T1}) =$$
$$= \mathrm{C}_{00} \cdot P_0 \cdot N \cdot \mathrm{P}(z_k|\mathrm{T0}) + \mathrm{C}_{01} \cdot P_1 \cdot N \cdot \mathrm{P}(z_k|\mathrm{T1})$$

A kérdés tehát az, hogy melyik esetben kisebb a költség?

Ha K1 < K2 azaz

$$\mathbf{C_{10} \cdot \boldsymbol{P_0} \cdot \boldsymbol{N} \cdot P(z_k|T0) + C_{11} \cdot \boldsymbol{P_1} \cdot \boldsymbol{N} \cdot P(z_k|T1) < C_{00} \cdot \boldsymbol{P_0} \cdot \boldsymbol{N} \cdot P(z_k|T0) + C_{01} \cdot \boldsymbol{P_1} \cdot \boldsymbol{N} \cdot P(z_k|T1)}$$

átrendezve:

$$(C_{10} - C_{00}) \cdot P_0 \cdot N \cdot P(z_k \mid T0) < (C_{01} - C_{11}) \cdot P_1 \cdot N \cdot P(z_k \mid T1)$$

egyszerűsítve:

$$(C_{10} - C_{00}) \cdot P_0 \cdot P(z_k \mid T0) < (C_{01} - C_{11}) \cdot P_1 \cdot P(z_k \mid T1)$$

akkor minden ilyen esetet (z_k-t produkál) **pozitívnak** veszünk.

****

Ha K1 > K2, akkor minden ilyen esetet (z_k-t produkál) **negatívnak** érdemes tekinteni.

****

A döntésünk nagy mértékben azon múlik, hogy jól becsüljük-e meg az egyes költségeket!!!
Ha C_{01} nagy, ettől félünk, akkor egész más alakul ki, mint ha C_{10}-tól félünk, azt vesszük nagynak.

Például a jog: attól félünk, hogy egy ártatlant elítélnek, vagy attól, hogy egy bűnös megússza?

Például a terrorista kérdés: attól félünk, hogy valahogy besurran egy terrorista, vagy attól, hogy jogtalanul vádolunk valakit?

Például az egészségügy: attól félünk-e jobban, hogy valakinek nem vesszük észre a betegségét, vagy attól, hogy feleslegesen ijesztgetjük, visszarendeljük vizsgálatra?

Példák lehetséges zh- vagy vizsgafeladatra

1. Jelöljük T0-val az egészségeseket, T1-gyel a betegeket. Egy z paramétert akarunk felhasználni diagnosztikai céllal, és vizsgálatot végeztünk 7670 emberen, hogy a lehetséges 6 érték eloszlását megismerjük.

k	**1**	**2**	**3**	**4**	**5**	**6**
z_k	**$z_1=100$**	**$z_2=120$**	**$z_3=140$**	**$z_4=160$**	**$z_5=180$**	**$z_6=2000$**
z_k-t produkált az egészségesek közül	1200	930	450	270	80	20
z_k-t produkált a betegek közül	200	300	450	700	1250	1820

Adja meg a fenti adatok alapján a $P(z_k|T0)$ és $P(z_k|T1)$ becslését!

2. Diagnosztikai eljárásunk egy *z* paraméter mérésén alapul, ami csak a táblázatban látható hétféle értéket veheti fel. Azt tekintjük *egészséges*nek, akinek a mért *z* értéke egy adott limitnél *nagyobb*, azokat pedig *beteg*nek tekintjük, akiknek a limit *alatti* a mért *z* értéke.

Eljárásunkat olyan mintán teszteltük, akik közül 23.710 egészséges, és 18.161 a beteg. A táblázat azt mutatja, hogy az egészségesek, illetve a betegek közül hányan produkálták az egyes *z* értékeket.

A következő nyolc *L* küszöbértéknél vizsgáltuk meg az eljárást L_1= -10, L_2= -6, L_3=-3, L_4= -1, L_5= 1, L_6= 3 , L_7= 6 , L_8= 10. Rajzolja fel az eljárást jellemző ROC görbét! Jelölje meg a görbén (legalább jellegre helyesen, tehát a 0,3 pont ne a 0,9-nél legyen!) az egyes kiszámolt pontokat, és azt, hogy melyik *L* értéknél kaptuk azokat!

z	**-7**	**-5**	**-2**	**0**	**2**	**5**	**7**
Az adott z értéket ennyi egészségesnél mértük	**345**	**1.532**	**2.421**	**3.211**	**4.587**	**5.239**	**6.375**
Az adott z értéket ennyi betegnél mértük	**5.462**	**4.388**	**3.632**	**2.222**	**1.543**	**539**	**375**

3. Egy diagnosztikai eljárást vizsgáltunk meg 8.006 személy segítségével, akik közül 2.629 beteg volt (bár tudjuk, hogy a népességben ez a betegség csak 1,5%-ban fordul elő). A diagnosztika vizsgálata során egy *z* paramétert mértünk (a szemléletesség kedvéért legyen a testhőmérséklet), ennek különböző értékeinél tekintjük betegnek, illetve egészségesnek a pácienst. A diagnosztikai eljárás 2.300 Ft-ba kerül személyenként, a tévesen egészségesnek diagnosztizált betegek kezelése 213.000 Ft-ba kerül, az időben diagnosztizált betegeké viszont csak 53.400 Ft-ba. A tévesen betegnek tekintetteknél átlagosan 9.000 Ft-ba kerül, mire kiderül, hogy valójában egészségesek.

Azt tapasztaltuk, hogy a következő tartományokba esnek a *z* értékek a táblázatban megadott számú betegnél és egészségesnél.

A. Ezen adatok alapján olcsóbb-e, ha a [37,38) tartományba eső értékek esetén betegnek tekintjük az embereket, vagy az az olcsóbb, ha egészségesnek?

B. Önmagában a szűrés költsége mennyiben befolyásolja, hogy az A. pontban vizsgált csoportot egészségesnek vagy betegnek célszerű-e tekinteni?

$[z_{jmin}, z_{jmax})$	[35,36)	[36,37)	[37,38)	[38,39)	[39,40)	[40,41)
Betegek, akiknél $[z_{jmin}, z_{jmax})$-ba esik a mért paraméter	**1874**	**538**	**124**	**54**	**32**	**7**
Egészségesek, akiknél $[z_{jmin}, z_{jmax})$-ba esik a mért paraméter	**52**	**168**	**469**	**725**	**1542**	**2421**

4. Testtömegindex (BMI) mérést végzünk: célunk, hogy ennek alapján szűrjük a kóros soványságot. A szűrési eljárás jellemzése céljából összesen 23.417 embert vizsgáltunk meg, akikről tudjuk, hogy kik – ebből a szempontból – egészségesek, és kik (kórosan sovány) betegek. Összesen 16.917 egészséges van a vizsgált mintában. Megmértük a vizsgálatban résztvevők testtömeg-indexét, és azt tapasztaltuk, hogy a táblázatban megadott B értékek alatti értéknél az alábbiak szerint alakult a betegek, ezen B értékek feletti értéknél az egészségesek száma. (Persze ez a valóságos szűrés erősen leegyszerűsített, és ezáltal torzított képe… ráadásul légbőlkapott adatokkal, szóval csak ROC görbe stb. gyakorlásra jó.) Rajzolja fel ezen vizsgálat alapján a – *B*-nél kisebb érték esetén beteg, a fölötti értéknél egészséges – eljárást jellemző ROC görbét!

B	**8**	**16**	**18**	**20**	**25**
Egészségesek, akiknek B fölé esik a mért BMI-je	16.917	16.917	15.863	12.132	0
Kórosan sovány betegek, akiknek B alá esik a mért BMI-je	0	3.879	5.803	6.194	6.500