6. Gyakorlat

Döntési fa, Bayes-tétel

Elméleti összefoglaló

A Bayes-tétel

Gyakorlati feladatok megoldása során gyakran elfordul, hogy ismert egy $B_1, B_2, ..., B_n$ teljes eseményrendszer a $P(B_j)$ $(j = 1, 2, ..., n)$ valószínségekkel együtt, és a vizsgált A eseménynek erre a teljes eseményrendszerre vonatkoztatott $P(A|B_j)$ $(j = 1, 2, ..., n)$ feltételes valószínségeit is ismerjük. Ha feltételezzük, hogy az A eseményt valamelyik B_j esemény okozta, akkor felmerülhet a kérdés, hogy a B_j-k mekkora mértékben járulnak hozzá A bekövetkezéséhez. Ez a $P(B_j|A)$
$(j = 1, 2, ..., n)$ ún. *Bayes-valószínség*ek meghatározását jelenti a fenti adatokból. Erre szolgál *Bayes tétel*e, mely Thomas Bayes (1701 – 1761) angol statisztikus és filozófustól származik.

A Bayes-tétel alkalmazása különböz adatokkal

Tegyük fel, hogy $P(A) \neq 0$ és $P(B_j) \neq 0$ $(j = 1, 2, ..., n)$. Ekkor a Bayes-tétel a következ alakokban írható fel:

$$\textbf{(1.)}\quad P(B_j|A) = \frac{P(A \cdot B_j)}{P(A)}$$

$$\textbf{(2.)}\quad P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{P(A)}$$

(valószínségek szorzástétele a számlálóra)

$$\textbf{(3.)}\quad P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{k=1}^{n} P(A|B_k) \cdot P(B_k)}$$

(teljes valószínség tétel a nevezre)

A feti három alak mindegyikét hívhatjuk Bayes-tételnek (habár az els egy szimpla feltételes valószínség definíció), és az általunk ismert/meghatározott adatok döntik el, hogy melyiket alkalmazzuk. Például, ha egy korábbi feladatrészben már meghatároztuk a $P(A)$ és $P(A \cdot B_j)$ valószínségeket, akkor elég az els alakot használni. A tankönyvek gyakran csak a legösszetettebb, **(3.)** alakot közlik, mely egyesíti a feltételes valószínség definícióját, a valószínségek szorzástételét és a teljes valószínség tételt.

Döntési fa diagram, inverz fa diagram

A Bayes-formulát sokszor kétfázisú kísérletekben alkalmazzuk, vagyis amikor az $A = A_i$ esemény egy második $A_1, A_2, ..., A_m$ teljes eseményrendszernek az eleme.

A kétfázisú döntési folyamatot szokás *döntési fa diagram*mal ábrázolni: a Start csúcsból minden B_j eseményhez vezet él (elágazás n felé), majd a második szinten minden B_j eseménytl

minden A_i eseményhez vezet él (elágazás m felé). Az els szint élein a $P(B_j)$ valószínségek, a második szint élein pedig a $P(A_i|B_j)$ feltételes valószínségek szerepelnek. Minden ág végén feltüntetjük a hozzá tartozó $P(A_i \cdot B_j)$ szorzat-valószínséget. A szorzatesemények teljes eseményrendszert alkotnak, így a szorzat-valószínségek összege 1.
Ha a döntési fa két logikai szintjét felcseréljük, akkor jutunk az *inverz fa diagram*hoz. Itt az els szint élein az $P(A_i)$ valószínségek, a második szint élein pedig a $P(B_j|A_i)$ Bayes-valószínségek szerepelnek. Elbbieket a teljes valószínség tétellel, utóbbiakat pedig a Bayes-tétellel számíthatjuk ki. A szorzat-valószínségek változatlanok maradnak, de a sorrendjük megváltozik(!) az eredeti döntési fához képest.

Kidolgozott példafeladatok

1. Feladat (Színes golyók húzása két dobozból / 2.)

Sárga és kék golyókat helyezünk el vegyesen két dobozba az alábbi ábra szerint:

1. doboz:	2. doboz:
41 sárga 49 kék	9 sárga 1 kék

Eztuán egy golyót sorsolunk ki a következ eljárással: elbb a dobozok közül véletlenszeren kiválasztjuk az egyiket, majd abból véletlenszeren kihúzunk egy golyót.
a) Adott szín kisorsolt golyó milyen valószínséggel származik egyik vagy másik dobozból?
b) Rajzoljuk fel a feladat inverz döntési fa diagramját a megfelel valószínségekkel!
c) Mi lenne a válasz az elz kérdésre, ha az 1. ill. 2. dobozt a bennük lév golyók számával arányos valószínséggel választanánk ki?

Megoldás

```
> restart;
```

A feladatban megadott mennyiségek:

```
> s1 := 41; k1 := 49; s2 := 9; k2 := 1;
> pd1 := 1/2; # 1. doboz választásának valószínsége
> pd2 := 1 - pd1; # 2. doboz választásának valószínsége
```

Az összes golyók száma, a sárga ill. kék golyók száma és az 1. ill. 2. dobozban lév golyók száma:

```
> g := s1 + k1 + s2 + k2; s := s1 + s2; k := k1 + k2; d1
  := s1 + k1; d2 := s2 + k2;
```

Annak valószínségei, hogy az 1. ill. 2. dobozból sárga ill. kék golyót húzunk:

```
> `P(k|d1)` := k1/d1; `P(s|d1)` := s1/d1; `P(k|d2)` :=
  k2/d2; `P(s|d2)` := s2/d2;
```

a) Adott szín kisorsolt golyó milyen valószínséggel származik egyik vagy másik dobozból?

Egy ilyen, több ágon végigfuttatható kísérlet során a különböz ágak összesített valószínsége mellett a másik jellemz kérdés az lehet, hogy ha egy adott kimenet bekövetkezett, akkor amögött milyen valószínséggel történhetett adott ágon való lefutás. Például meg lehet kérdezni, hogy ha a feladatbeli kísérlet végeredményeként sárga golyó van a kezünkben, akkor az milyen valószínséggel származik egyik vagy másik dobozból. Utóbbi kérdés megválaszolási módjára vonatkozik a *Bayes-tétel*. Lényegét tekintve ez arról

szól, hogy egy adott (már bekövetkezett) végkifejlet mellett annak valószínsége, hogy ez éppen egy kiszemelt ágon történt, ezen ág valószínségének az összes, ide vezet ág valószínségeinek összegéhez viszonyított aránya.
Pl. sárga golyó kisorsolásához mindkét dobozon át el lehet jutni. Az egyiken át a valószínség:

```
> `P(s|d1)`*pd1;
```

A másikon át:

```
> `P(s|d2)`*pd2;
```

Ezek összege a *teljes valószínség tétel* szerint a sárga golyó húzásának valószínsége:

```
> ps := `P(s|d1)`*pd1 + `P(s|d2)`*pd2; evalf(%);
```

Annak valószínsége, hogy egy kisorsolt sárga golyó az els dobozból érkezett (az els dobozon át sárga sorsolódás valószínsége arányítva a sárga sorsolódás teljes valószínségéhez):

```
> `P(d1|s)` := `P(s|d1)`*pd1/ps; evalf(%);
```

Annak valószínsége, hogy egy kisorsolt sárga golyó a második dobozból érkezett (a második dobozon át sárga sorsolódás valószínsége arányítva a sárga sorsolódás teljes valószínségéhez):

```
> `P(d2|s)` := `P(s|d2)`*pd2/ps; evalf(%);
```

Mivel sárga golyó kiválasztására összesen ez a két lehetség van, a két valószínség összege természetesen 1:

```
> `P(d1|s)` + `P(d2|s)`;
```

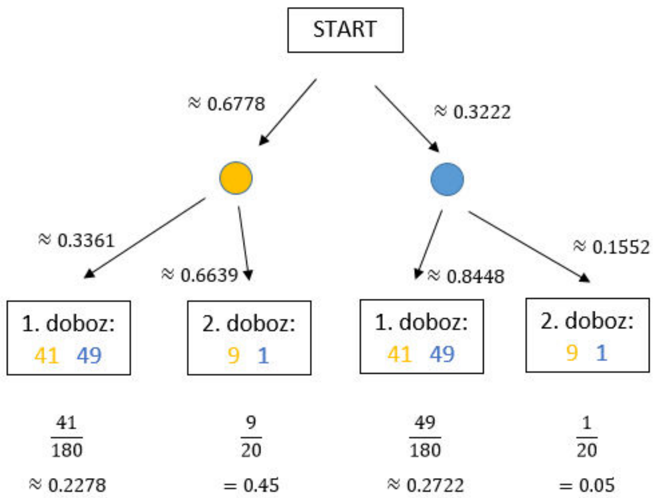
Látszik, hogy a sárga golyó kisorsolásával végzd kísérletek nagyobb részében a húzott sárga golyó a második dobozból származik (kb. kétszer annyi esetben). Ez összhangban van azzal, hogy míg onnan szinte majdnem mindig sárga érkezik, az els dobozból csak az esetek nem egészen felében.
Hasonlóan számítható kék golyó egyik vagy másik dobozból való származásának valószínsége:

```
> pk := `P(k|d1)`*pd1 + `P(k|d2)`*pd2; evalf(%);
> `P(d1|k)` := `P(k|d1)`*pd1/pk; evalf(%);
> `P(d2|k)` := `P(k|d2)`*pd2/pk; evalf(%);
```

Látható, hogy a kék golyó húzásával végzd kísérletek nagy részében a húzott golyó az els (arányaiban több kéket tartalmazó) dobozból érkezik. A két valószínség összege természetesen itt is 1:

```
> `P(d1|k)` + `P(d2|k)`;
```

b) Rajzoljuk fel a feladat inverz döntési fa diagramját a megfelel valószínségekkel!
Ebben a szakaszban az elz feladatsorban szerepl "Színes golyók húzása két dobozból / 1." feladathoz képest fordított módon közelítettük meg a kérdést. Ehhez a fordított megközelítéshez illeszkedik az inverz (fordított) fa diagram:



Ez a fa úgy születik, hogy az eredetihez képest felcseréljük az elágazások sorrendjét: ezúttal elször szín, majd utána a doboz szerint ágaztatjuk el. Az els elágazás valószínségei a fentebb kiszámított teljes valószínségek lesznek. Az egyes ágak valószínségei az eredeti fáéval azonosak (lásd az elz feladatsort). Ezekbl akár visszaosztással is számíthatók a második elágazási szint (feltételes) valószínségei, melyek egyébként épp a fent kiszámított Bayes-valószínségekkel azonosak.

c) Mi lenne a válasz az elz kérdésre, ha az 1. ill. 2. dobozt a bennük lév golyók számával arányos valószínséggel választanánk ki?

Bár a kék és sárga golyók száma egyenl, az adott kísérleti konstrukcióban túlnyomóan sárga golyók sorsolódnak ki. Ez azzal magyarázható, hogy az arányában sok sárgát, viszont darabszám szerint kevés golyót tartalmazó második doboz szerepe túlértékeldik azáltal, hogy kevés golyó mellett is ugyanazzal az 50 %-os kiválasztási valószínséggel rendelkezik, mint a 9-szer annyi golyót tartalmazó els doboz. A múlt heti feladatsorban láttuk, hogy a számarányok jobban tükrözdnének, ha az els lépésben, a dobozok kiválasztásánál a valószínségek arányosak lennének a tartalmazott golyók számával: a több golyót tartalmazó doboz arányosan többször kerülne a kisorsolásba.

```
> pd1 := d1/(d1 + d2);
  pd2 := 1 - pd1;
```

```
> pk := `P(k|d1)`*pd1 + `P(k|d2)`*pd2;
  ps := `P(s|d1)`*pd1 + `P(s|d2)`*pd2;
```

A Bayes-valószínségek ebben az esetben:

```
> `P(d1|k)` := `P(k|d1)`*pd1/pk;
```

```
> `P(d2|k)` := `P(k|d2)`*pd2/pk;
```

```
> `P(d1|s)` := `P(s|d1)`*pd1/ps;
```

```
> `P(d2|s)` := `P(s|d2)`*pd2/ps;
```

Azt kaptuk, hogy adott szín golyó húzása esetén az egyes dobozok választásának Bayes-valószínsége tükrözi az adott szín golyók számarányát az egyes dobozokban.

2. Feladat (Orvosi diagnosztikai teszt)

Tegyük fel, hogy egy adott betegségre létezik egy diagnosztikai teszt, amivel az orvos egy páciensrl megkísérli eldönteni, hogy beteg-e. A teszt kifejleszti a megbízhatósággal kapcsolatban azt állítják, hogy 99 %-os biztonsággal mutat betegnek egy valóban beteg embert, és 95 %-os biztonsággal mutat egészségesnek egy valóban egészséges embert (de ezek szerint néha téved a teszt: átlagosan minden századik betegnél elmulasztja kimutatni a betegséget, noha jelen van, viszont átlagosan minden huszadik egészséges embernél tévesen riaszt). A tapasztalatok szerint a páciensek 10 %-a szenved az adott betegségben. Végezzük el az alábbi valószínűségelméleti vizsgálatokat, hogy kiderítsük, mennyire használható jól ez a teszt!
a) A páciensek hány százalékát jelzi betegnek, és hány százalékát egészségesnek a teszt?
b) Tegyük fel, hogy egy páciens tesztje pozitív lett. Mekkora az esélye, hogy tényleg beteg?
c) Tegyük fel, hogy egy páciens tesztje negatív lett. Mekkora az esélye, hogy tényleg egészséges?
d) Mennyire jó ez a teszt?
e) Mi a helyzet, ha egy ugyanilyen megbízhatósággal rendelkez tesztet egy olyan járványszer betegség diagnosztizálására használunk, mely (ers járvány esetén) a páciensek 50 %-át érinti? Hogy alakulnak a fenti valószínségek? Mennyire használható a teszt?
f) Mi a helyzet, ha egy ugyanilyen megbízhatósággal rendelkez tesztet egy olyan nagyon ritka betegség diagnosztizálására használunk, mely mindössze a páciensek 0.1 %-át érinti? Hogyan alakulnak a fenti valószínségek? Mennyire használható a teszt?

Megoldás

```
> restart;
```

a) A páciensek hány százalékát jelzi betegnek, és hány százalékát egészségesnek a teszt?

A kísérlet maga a teszt elvégzése a (beteg vagy egészséges) páciensen. A kísérlet eseménytere a páciens állapotából {beteg, egészséges} ill. a teszt eredményébl {pozitív, negatív} álló rendezett párokból áll: {(beteg, pozitív), (beteg, negatív), (egészséges, pozitív), (egészséges, negatív)}, ezekhez szeretnénk meghatározni a valószínségeket. Vegyük fel megfelel jelölésekkel az adatokat (b=beteg, e=egészséges, `+`=pozitív, `-`= negatív)!

```
> `P(b)` := 0.1; `P(e)` := 1 - `P(b)`; `P(+|b)` := 0.99;
  `P(-|b)` := 1 - `P(+|b)`; `P(-|e)` := 0.95; `P(+|e)` :=
  1 - `P(-|e)`;
```

Milyen valószínséggel jelez betegnek egy pácienst a teszt (vagy: a páciensek hányad részét jelzi betegnek)?

A teljes valószínség tétele (TVT) szerint a megfelel ágak valószínségeinek kiszámítása majd ezek összeadása után:

```
> `P(+)` := `P(+|b)`*`P(b)` + `P(+|e)`*`P(e)`;
> `P(-)` := `P(-|b)`*`P(b)` + `P(-|e)`*`P(e)`;
```

Azt látjuk, hogy a valóságban 10 % : 90 % beteg-egészséges arányt a teszt 14.4 % : 85.6 % arányra torzította. Ez természetesen a teszt tévedéseibl adódik. A torzulás tendenciája láthatóan az, hogy több beteget mutat, mint amennyi valóban van. Ez azzal magyarázható, hogy a teszt a kétféle tévedési lehetsége közül a nagyobbat (5 %) egészségesek beteggé minsítésével követi el, ráadásul az egészségesek vannak többen (90 %). Tehát a teszt során nagyobb mintából nagyobb arányban kerülnek át páciensek egészségesbl beteg státuszba, míg fordítva ez jóval kisebb: kisebb mintából kisebb arányban minsít betegeket egészségesnek a teszt.

b) Tegyük fel, hogy egy páciens tesztje pozitív lett. Mekkora az esélye, hogy tényleg beteg?

Ez a fordított (*Bayes*) feladat: egy bekövetkezett végeredmény mellé keressük azt a valószínséget, amellyel a több lehetséges ág közül éppen azon jutottunk idáig.

```
> `P(b|+)` := `P(+|b)`*`P(b)`/`P(+)`;
```

c) Tegyük fel, hogy egy páciens tesztje negatív lett. Mekkora az esélye, hogy tényleg egészséges?

Szintén Bayes-tétellel a másik esetre:

```
> `P(e|-)` := `P(-|e)`*`P(e)`/`P(-)`;
```

d) Mennyire jó ez a teszt?

Látjuk, hogy a negatív teszteredmény mellett valóban szinte biztos (99.9 %), hogy egészséges a páciens. Érdemes megfigyelni, hogy ez a biztonság még a teszt jobbik irányú (99 %) megbízhatóságánál is nagyobb, ugyanis a negatív teszteredmény páciensek legnagyobb része a valóban egészségesek közül jön; közéjük csak az amúgy is csak 10 százaléknyi beteg közül érkezik az az 1 %, akit a teszt eltéveszt. Azért ilyen jó tehát a kizárás hatásfoka, mert egyrészt a betegek amúgy is kevesebben vannak, másrészt a teszt még közülük is csak nagyon keveseknél téved – elhanyagolható számban kerülnek betegek a negatív teszteredményesek közé.

Ugyanakkor a teszt által betegnek jelzett pácienssrl csak 69 %-os bizonyossággal állítható, hogy tényleg beteg. Ez az arány nem sokkal jobb a pénzfeldobással való eldöntésnél, ezért javasolható az orvosnak, hogy ha lehet, próbáljon meg más módszert is a diagnosztizálásra. A bizonytalanság egyébként onnan fakad, hogy mivel az egészségesek aránya elég magas, ráadásul pedig a teszt is nagyobb arányban hajlamos közülük is tévesen riasztani, a teszt során a valódi betegek közé nem elhanyagolható számban kerülnek egészségesek is.

Ez a teszt jó arra, hogy az egészségesek nagy részét jó hatásfokkal kiszrje a rendszerbl: a páciensek 85.6 %-ával nem kell tovább foglalkozni, miközben ezek 99.9 %-a valóban egészséges (csak 0.1 % fel nem ismert beteg marad köztük). A pozitív eredményekre ugyan meghagy egy akkora bizonytalanságot, ami további vizsgálatot tehet indokolttá, mégis hasznos volt, amennyiben az esetlegesen szükséges következ (bonyolultabb, költségesebb stb.) vizsgálatot már csak szkebb körre kell elvégezni.

A teszt által helyesen kezelt páciensek aránya:

```
> `P(+|b)`*`P(b)` + `P(-|e)`*`P(e)`;
```

Ugyanerre vezet az alternatív (inverz fa alapján való) számolás is:

```
> `P(b|+)`*`P(+)` + `P(e|-)`*`P(-)`;
```

Ez önmagában nem tnik rossznak, de (mint láttuk és késbb is látni fogjuk) a helyzet összetettebb annál, hogy ez az egy érték elég legyen a minség megítéléséhez.

e) Mi a helyzet, ha egy ugyanilyen megbízhatósággal rendelkez tesztet egy olyan járványszer betegség diagnosztizálására használunk, mely (ers járvány esetén) a páciensek 50 %-át érinti? Hogy alakulnak a fenti valószínségek? Mennyire használható a teszt?

Ezúttal már nem részletezve, csupán a fenti számítás megismétlésével, a betegek arányának beállítása után:

```
> `P(b)` := 0.5; `P(e)` := 1 - `P(b)`;
> `P(+)` := `P(+|b)`*`P(b)` + `P(+|e)`*`P(e)`;
> `P(-)` := `P(-|b)`*`P(b)` + `P(-|e)`*`P(e)`;
> `P(b|+)` := `P(+|b)`*`P(b)`/`P(+)`;
> `P(e|-)` := `P(-|e)`*`P(e)`/`P(-)`;
```

Azt látjuk, hogy arányaiban a teszt torzítása kisebb, és a megbízhatóságok is kiegyenlítdtek: mindkét döntés (beteg, egészséges) megbízhatósága közel került 100 %-

hoz. Ez esetben a teszt tehát önmagában (további orvosi vizsgálatok nélkül) is viszonylag megbízhatóan alkalmas betegek és egészségesek szétválasztására. Igaz, hogy eszerint kb. minden huszadik valóban beteg embert nem ismernek fel (és mondjuk nem kap gyógyszert), de ha ez nem egy nagyon súlyos betegség, és inkább csak kicsit több kellemetlenséggel, vagy egy-két nappal hosszabb felépüléssel jár, akkor a teszt hasznos volt, hiszen mégis elég jól beazonosította az embereknek azt a felét, akiknek gyógyszert kell kapniuk. (Emellett látjuk, hogy kb. minden századik egészséges ember tévedésbl, feleslegesen kap gyógyszert, de ha nincs különösebben káros hatása, akkor ez is "belefér").
Összességében a teszt által helyesen megítélt páciensek aránya:

```
> `P(+|b)`*`P(b)` + `P(-|e)`*`P(e)`;
```

Önmagában ez a szám nem sokkal jobb az elz esetnél (10 % beteg), viszont kiegyenlítettebb a kétféle állapot megítélése. Ahhoz képest, hogy a páciensek fele beteg, így minden vizsgálattal két azonos valószínség eset között kell dönteni, kimondottan jó eredmény, hogy ez az esetek 97 %-ában jól sikerül.

f) Mi a helyzet, ha egy ugyanilyen megbízhatósággal rendelkez tesztet egy olyan nagyon ritka betegség diagnosztizálására használunk, mely mindössze a páciensek 0.1 %-át érinti? Hogyan alakulnak a fenti valószínségek? Mennyire használható a teszt?

A számítás megismétlésével, a betegek arányának beállítása után:

```
> `P(b)` := 0.001;
  `P(e)` := 1 - `P(b)`;
> `P(+)` := `P(+|b)`*`P(b)` + `P(+|e)`*`P(e)`;
> `P(-)` := `P(-|b)`*`P(b)` + `P(-|e)`*`P(e)`;
> `P(b|+)` := `P(+|b)`*`P(b)`/`P(+)`;
> `P(e|-)` := `P(-|e)`*`P(e)`/`P(-)`;
```

Itt már komoly aránytalanság figyelhet meg. A betegség ugyan átlagosan 1000-bl csak egy pácienst érint, mégis kb. 51-nél pozitív lesz a teszt, tehát elég sok a téves riasztás. Ez egyrészt azzal jár, hogy akinek ennek ellenére negatív a tesztje, az elég biztos lehet benne (99.999 %), hogy tényleg egészséges. Ugyanakkor az egészségesek magas arányszáma folytán a pozitív tesztek köre annyira felhígul egészségesekkel (1 valódi betegre kb. 50 téves riasztás jut), hogy a pozitív teszt gyakorlatilag semmit sem jelent. Ez a teszt alkalmas egy pácienskör biztonságos kizárására, de a valódi betegek megtalálásához más tesztet kell alkalmazni. Bár tesztünk 1000 páciensbl kb. 949-et megbízhatóan kizár, a megmaradó 51 között nem lehet megtalálni vele az 1 valódi beteget. Mindenképp kell másik teszt is. Ezek után viszont a tesztek ára, bonyolultsága, fájdalmassága stb. dönti el, hogy érdemes-e egyáltalán elszrként alkalmazni a tárgyalt tesztet (hogy a következt már csak a páciensek huszadára kelljen), vagy eleve találunk olyat, ami egymagában is pontosabban és költséghatékonyabban végzi el a szrést.
Összességében a teszt által helyesen megítélt páciensek aránya:

```
> `P(+|b)`*`P(b)` + `P(-|e)`*`P(e)`;
```

Ez az eredmény valójában nagyon rossz, már csak ahhoz képest is, hogy ha teszt nélkül mindenkit automatikusan egészségesnek minsítenénk, akkor ugyanez az eredményességi mérszám 99.9 % lenne! Ugyanakkor fentebb azt is láttuk, hogy amilyen gyenge teljesítmény a teszt a beteg azonosításában, annyira jó a páciensek nagy részénél (95 %) a betegség kizárásában. Mindezeket az árnyalatokat nem képes visszaadni önmagában a helyesen jelzett páciensek aránya, ezért önmagában jósági mérszámként nem alkalmazható.

3. Feladat (Tesztkérdések tippelése)

Egy diák $\frac{2}{3}$ valószínséggel tudja egy tesztkérdésre a helyes választ. Ha nem tudja,

akkor tippel és a helyes tipp esélye $\frac{1}{4}$.

a) Mekkora a helyes válasz esélye?

b) Tudjuk, hogy a diák válasza helyes. Mekkora az esélye annak, hogy nem tudta a helyes választ?

c) Tudjuk, hogy a diák válasza helyes. Mekkora az esélye annak, hogy tudta a helyes választ?

d) Tudjuk, hogy a diák válasza helytelen. Mekkora az esélye annak, hogy nem tudta a helyes választ?

e) Rajzoljuk fel a feladat döntési fa diagramját és az inverz fa diagramját a megfelel valószínségekkel!

f) Adjuk meg az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Mekkora a helyes válasz esélye?

Rövidítések:

$T =$ tudja a helyes választ, $Tc =$ nem tudja a helyes választ

$J =$ a válasz jó, $N =$ a válasz nem jó.

```
> pT, `P(J|Tc)` := 2/3, 1/4;
```

A tagadás események valószínségeit megkapjuk 1-bl való kivonással!

```
> pTc, `P(N|Tc)` := 1 - 2/3, 1 - 1/4;
```

Ha a diák tudja a választ, a helyes eredmény valószínsége 1 és a rossz válasz esélye 0.

```
> `P(J|T)`, `P(N|T)` := 1, 0;
```

Helyes választ kétféleképpen kaphatunk: vagy tudás útján ($P(J \cdot T)$) vagy tippelés útján ($P(J \cdot \overline{T})$). Ezeket összegezve kapjuk a helyes válasz valószínségét. Ez egy teljes valószínség tétel.

```
> pJT := `P(J|T)`*pT;
  pJTc := `P(J|Tc)`*pTc;
  pJ := pJT + pJTc;
```

A rossz válasz (*N*) ennek a tagadása.

```
> pN := 1 - pJ;
```

A két utóbbi valószínség fog szerepelni az inverz fa els döntési elágazásán!

b) Tudjuk, hogy a diák válasza helyes. Mekkora az esélye annak, hogy nem tudta a helyes választ?

A kérdés a $P(\overline{T}|J)$ valószínségre vonatkozik, melyet egy Bayes-tétel alkalmazásával (feltételes valószínség definíciójával) számolhatunk:

```
> `P(Tc|J)` := pJTc/pJ;
```

c) Tudjuk, hogy a diák válasza helyes. Mekkora az esélye annak, hogy tudta a helyes választ?

A kérdés a $P(T|J)$ valószínségre vonatkozik, melyet a b) ponthoz hasonlóan számolhatunk:

```
> `P(T|J)` := pJT/pJ;
```

d) Tudjuk, hogy a diák válasza helytelen. Mekkora az esélye annak, hogy nem tudta a helyes választ?

A kérdés a $P(\overline{T}|N)$ valószínségre vonatkozik, melyet ismét a b) ponthoz hasonlóan számolhatunk, de elbb meg kell határoznunk a $P(N \cdot T)$ szorzat-valószínséget egy szorzástétellel:

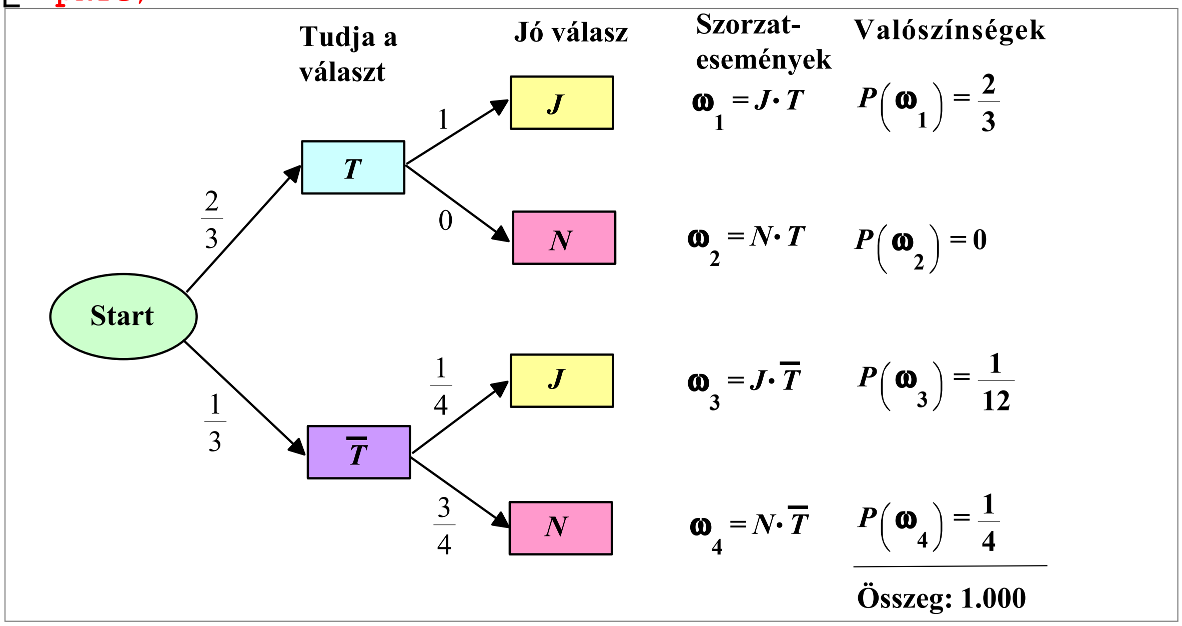
```
> pNTc := `P(N|Tc)`*pTc;
  `P(Tc|N)` := pNTc/pN;
```

Ez logikus, hiszen ha valaki tudja a választ, az mindig jól válaszol, így az összes rossz válasznak azoktól kell jönnie, akik eleve nem tudták a választ.

e) Rajzoljuk fel a feladat döntési fa diagramját és az inverz fa diagramját a megfelel valószínségekkel!

A döntési fa diagram els oszlopa tartalmazza a szorzateseményeket, a második oszlop pedig ezek valószínségeit! A szorzatesemények közül már hármat meghatároztunk, csak egy hiányzik:

```
> pJT;
  pJTc;
  pNT := `P(N|T)`*pT;
  pNTc;
```



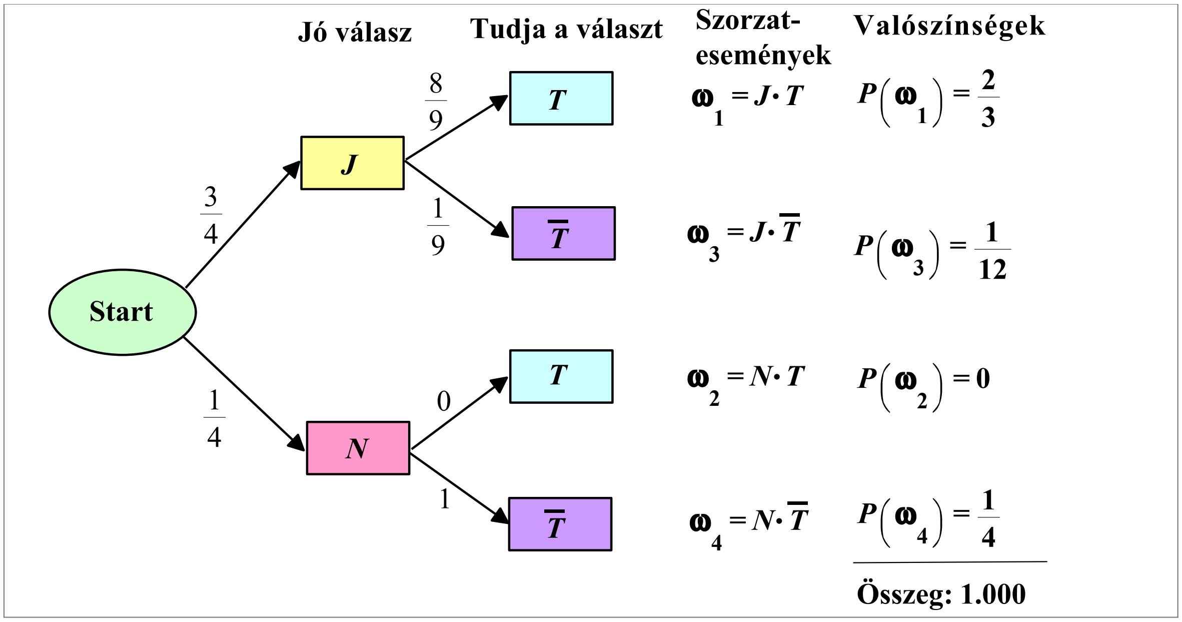
Az inverz fa felcseréli a két döntési fázist! A szorzatesemények az eredeti és az inverz fánál ugyanazok, kivéve a sorrendet! Az inverz fa második döntési fázisához tartozó feltételes valószínségek a Bayes-valószínségek. Ezek közül már hármat meghatároztunk, még egy hiányzik:

```
> `P(T|J)`;
  `P(Tc|J)`;
  `P(T|N)` := pNT/pN;
  `P(Tc|N)`;
```

Megjegyzés: Ezeket a feltételes valószínségeket közvetlenül, az alapadatokból is meghatározhatjuk a Bayes-tétel segítségével. Például:

```
> `P(T|J)` = `P(J|T)`*pT/(`P(J|T)`*pT + `P(J|Tc)`*pTc);
> `P(Tc|N)` = `P(N|Tc)`*pTc/(`P(N|T)`*pT + `P(N|Tc)`*pTc);
> # stb...
```

Jelen esetben azonban egyszerbb volt a már kiszámolt adatokat felhasználni (tulajdonképpen a Bayes-tétel fenti, **(3.)** alakjában szerepl számlálót és nevezt határoztuk meg korábban, és azokkal számoltunk).



f) Adjuk meg az együttes események valószínségeinek táblázatát!

Szorzatesemények 1

	A	B	C	D	E	F
1		*A válasz jó*	*A válasz nem jó*			
2	*Tudja a választ*	$\frac{2}{3}$	0	$\frac{2}{3}$		
3	*Nem tudja a válszt*	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{3}$		
4		$\frac{3}{4}$	$\frac{1}{4}$	1		
5						
6						
7						

4. Feladat (Repülgépek késése)

A Liszt Ferenc repülterrl naponta Amszterdamba induló 3 repülgépjáratot figyeltek utasterheltség és késési id szerint. Statisztikai megfigyelések alapján az összes utas 30 %-a utazott az 1. járattal, 50 %-a a 2. járattal és 20 %-a a 3. járattal! Az egyes járatok gépei késve érkeznek Amszterdamba, rendre 0.25, 0.2 és 0.15 valószínséggel. (Jelölje *K* a késés eseményét!)

a) Szemléltessük a járatok utasterheltségi - késési megoszlását döntési fa diagramon!

b) Az utasok hányadrésze érkezik késve Amszterdamba?

c) Készítsük el az inverz fa diagramot!

d) A késve érkez utasok között melyik járaton utazók vannak legkisebb arányban? Mennyi ez az arány?

e) A pontosan érkez utasok között melyik járaton utazók vannak legnagyobb arányban?

f) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Szemléltessük a járatok utasterheltségi - késési megoszlását döntési fa diagramon!

Rögzítsük az adatokat! Az 1., 2. és 3. járattal utazók aránya:

```
> pJ1, pJ2, pJ3 := 0.3, 0.5, 0.2;
```

A késés valószínsége, feltéve, hogy az 1., 2. ill. 3. járattal utazik valaki:

```
> `P(K|J1)`, `P(K|J2)`, `P(K|J3)` := 0.25, 0.2, 0.15;
```

A feltételre szorítkozó valószínségi mezn számolva mindhárom járat esetén megkapjuk a pontos érkezés, mint ellentett esemény valószínségét:

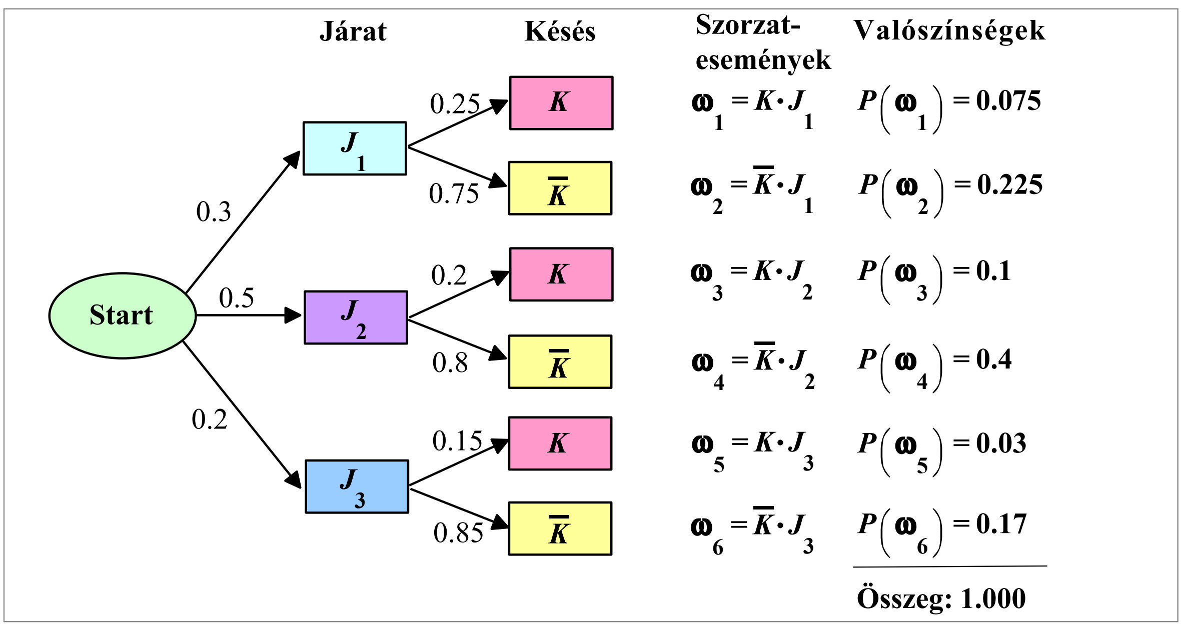
```
> `P(Kc|J1)` := 1 - `P(K|J1)`;
  `P(Kc|J2)` := 1 - `P(K|J2)`;
  `P(Kc|J3)` := 1 - `P(K|J3)`;
```

Most használjuk a valószínségek szorzástételét a szorzatesemények valószínségeinek meghatározásához!

```
> pKJ1 := `P(K|J1)`*pJ1;
  pKJ2 := `P(K|J2)`*pJ2;
  pKJ3 := `P(K|J3)`*pJ3;
```

```
> pKcJ1 := `P(Kc|J1)`*pJ1;
  pKcJ2 := `P(Kc|J2)`*pJ2;
  pKcJ3 := `P(Kc|J3)`*pJ3;
```

A fenti valószínségek ismeretében felrajzolhatjuk a döntési fát. A diagram els szintjén a járat szerint, a 2. szintjén pedig a késés ténye szerint bontjuk fel az eseteket.



b) Az utasok hányadrésze érkezik késve Amszterdamba?

A teljes valószínség tételét alkalmazzuk a $P(K)$ valószínség meghatározására: összeadjuk a döntési fa késéssel végzd ágain számolt szorzat-valószínségeket.

```
> pK := pKJ1 + pKJ2 + pKJ3;
```

Az utasok 20.5 %-a késve érkezik!

c) Készítsük el az inverz fa diagramot!

Az inverz fában az eredeti döntési fa szintjei fel vannak cserélve. Az els szint éleire a késés és a pontos érkezés valószínsége kerül, a második szint éleire pedig az ún. Bayes-valószínségek kerülnek, melyek az adott járattal érkezés valószínségét adják meg a késés/pontos érkezés feltétele mellett.

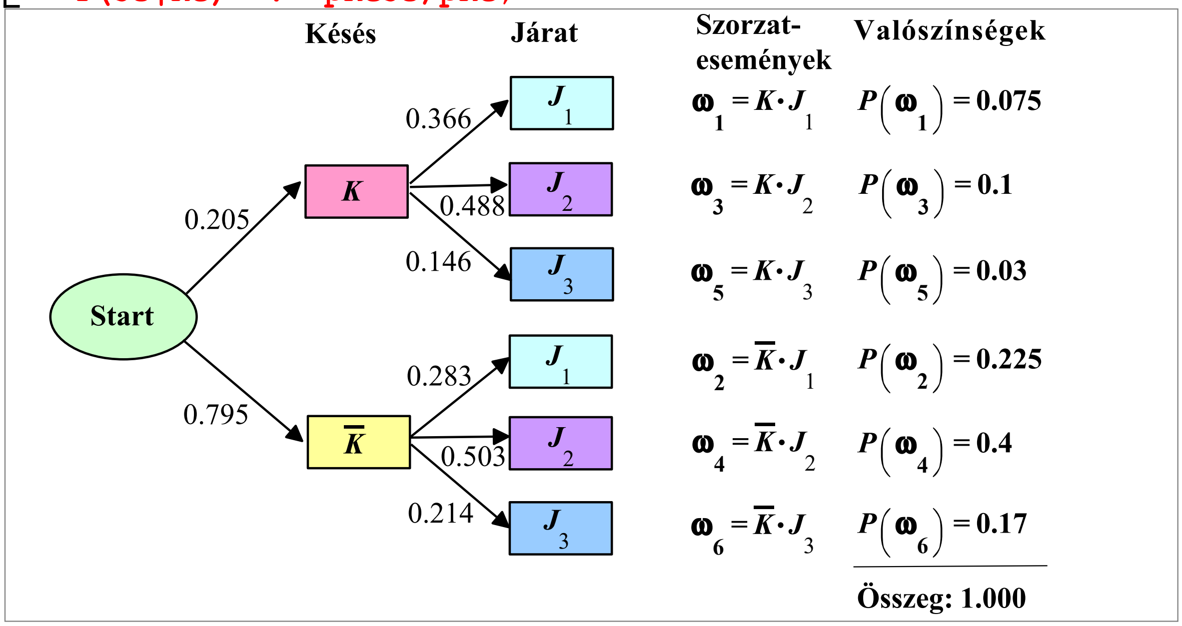
A pontos érkezés valószínségét a $P(K) + P(\overline{K}) = 1$ azonosság segítségével (vagy a b) ponthoz hasonlóan teljes valószínség tétellel) határozhatjuk meg

```
> pKc := 1 - pK;
```

A Bayes-valószínségek a Bayes-tétel segítségével számolhatók (itt felhasználjuk a korábban kiszámolt $P(K)$ és $P(\overline{K})$ valószínségeket):

```
> `P(J1|K)` := pKJ1/pK;
  `P(J2|K)` := pKJ2/pK;
  `P(J3|K)` := pKJ3/pK;
```

```
> `P(J1|Kc)` := pKcJ1/pKc;
  `P(J2|Kc)` := pKcJ2/pKc;
  `P(J3|Kc)` := pKcJ3/pKc;
```



d) A késve érkez utasok között melyik járaton utazók vannak legkisebb arányban? Mennyi ez az arány?

A kérdés az alábbi feltételes valószínségek közül kérdezi a legkisebbet:

```
> <`P(J1|K)`, `P(J2|K)`, `P(J3|K)`>;
```

Ez a 0.146, amely a J_3 járathoz tartozik.

e) A pontosan érkez utasok között melyik járaton utazók vannak legnagyobb arányban?

A kérdés az alábbi feltételes valószínségek közül kérdezi a legnagyobbat:

```
> <`P(J1|Kc)`, `P(J2|Kc)`, `P(J3|Kc)`>;
```

A pontosan érkez utasok között a J_2 járaton utazók vannak legnagyobb arányban (0.503).

f) Készítsük el az együttes események valószínségeinek táblázatát!

Együttes valószínségek 1								
	A	B	C	D	E	F	G	H
1		*Késik*	*Pontosan érkezik*					
2	*1. Járat*	0.075	0.225	0.3				
3	*2. Járat*	0.10	0.40	0.5				
4	*3. Járat*	0.030	0.170	0.2				
5		0.205	0.795	1.0				
6								
7								

5. Feladat (Ivóvíz nitrát-tartalma, ZH, 2016, A csoport)

Az EU területén az ivóvíz nitrát-tartalmának küszöbértéke 50 $\frac{mg}{l}$. Ivóvíz mintákat vettek az *A*, *B* és *C* területekrl. Az összes vízminta 30 %-a az *A* területrl, 35 %-a *B* területrl és a többi a *C* területrl való.
Az *A* területrl vett vízminták 90 %-ának nitrát-tartalma nem haladta meg az egészségügyi küszöbértéket, a *B* területrl vett minták 5 %-a meghaladta a küszöbértéket, míg a *C* területrl vett minták 85 %-a nem haladta meg az egészségügyi küszöbértéket (*K*)
.
a) Szemléltessük a vízminták területi és egészségügyi küszöbérték-meghaladás szerinti döntéseit fa diagram felrajzolásával és a hozzá tartozó valószínségek kiszámításával!
b) Az összes vízminta hányadrésze haladta meg az elírt egészségügyi küszöbértéket?
c) Készítsük el az inverz fa diagramot a hozzá tartozó valószínségek számításával!
d) Az elírt küszöbértéken belül maradók között melyik területrl vett minták aránya volt a legkisebb?
e) Az elírt küszöbértéket meghaladók között melyik területrl vett minták aránya volt a legnagyobb?
f) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Szemléltessük a vízminták területi és egészségügyi küszöbérték-meghaladás szerinti döntéseit fa diagram felrajzolásával és a hozzá tartozó valószínségek kiszámításával!

Rögzítsük az adatokat! Az *A*, *B* és *C* területrl származó vízminták aránya:

```
> pA, pB := 0.3, 0.35;
  pC := 1 - pA - pB;
```

Annak valószínsége, hogy az egyes területekrl vett vízminták nitrát-tartalma nem haladta meg az egészségügyi küszöbértéket:

```
> `P(K|A)`, `P(K|B)`, `P(K|C)` := 0.9, 0.95, 0.85;
```

A feltételre szorítkozó valószínségi mezn számolva mindhárom terület esetén megkapjuk az egészségügyi küszöbérték túllépésének, mint ellentett eseménynek a valószínségét:

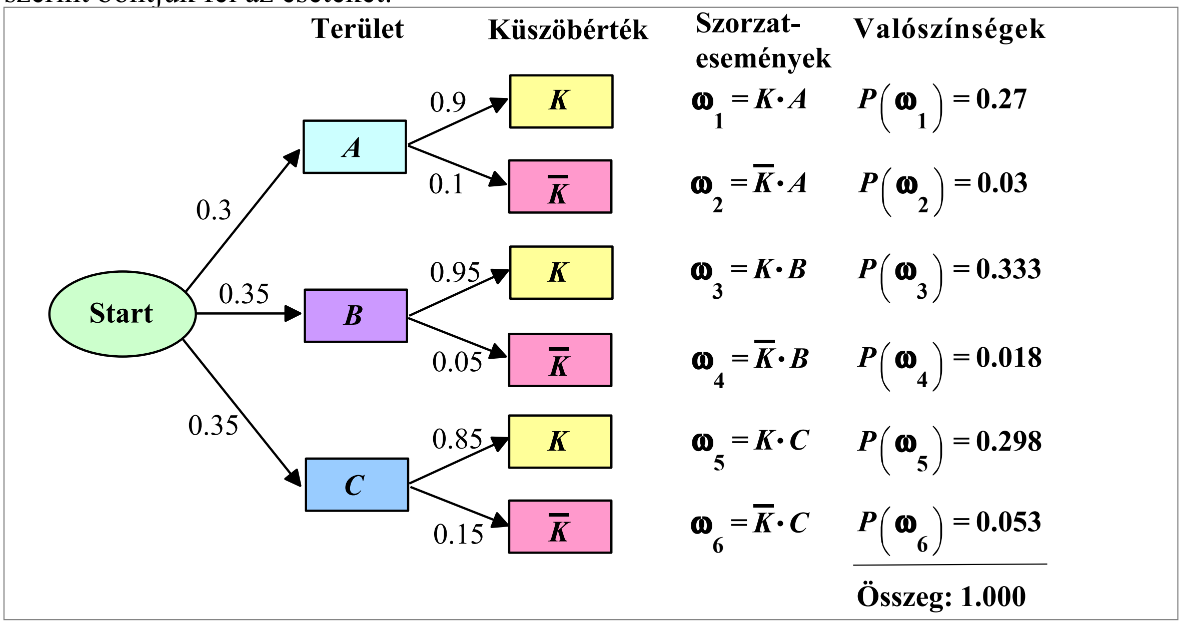
```
> `P(Kc|A)` := 1 - `P(K|A)`;
  `P(Kc|B)` := 1 - `P(K|B)`;
  `P(Kc|C)` := 1 - `P(K|C)`;
```

Most használjuk a valószínségek szorzástételét a szorzatesemények valószínségeinek meghatározásához!

```
> pKA := `P(K|A)`*pA;
  pKB := `P(K|B)`*pB;
  pKC := `P(K|C)`*pC;
```

```
> pKcA := `P(Kc|A)`*pA;
  pKcB := `P(Kc|B)`*pB;
  pKcC := `P(Kc|C)`*pC;
```

Ezeknek a valószínségeknek az ismeretében felrajzolhatjuk a döntési fát. A diagram els szintjén a terület szerint, a 2. szintjén pedig az egészségügyi küszöbérték alatt maradás szerint bontjuk fel az eseteket.



b) Az összes vízminta hányadrésze haladta meg az elírt egészségügyi küszöbértéket?
A teljes valószínség tételét alkalmazzuk a $P(\overline{K})$ valószínség meghatározására: összeadjuk a döntési fa azon ágain számolt szorzat-valószínségeket, melyek az egészségügyi küszöbérték túllépéséhez tartoznak.

```
> pKc := pKcA + pKcB + pKcC;
```

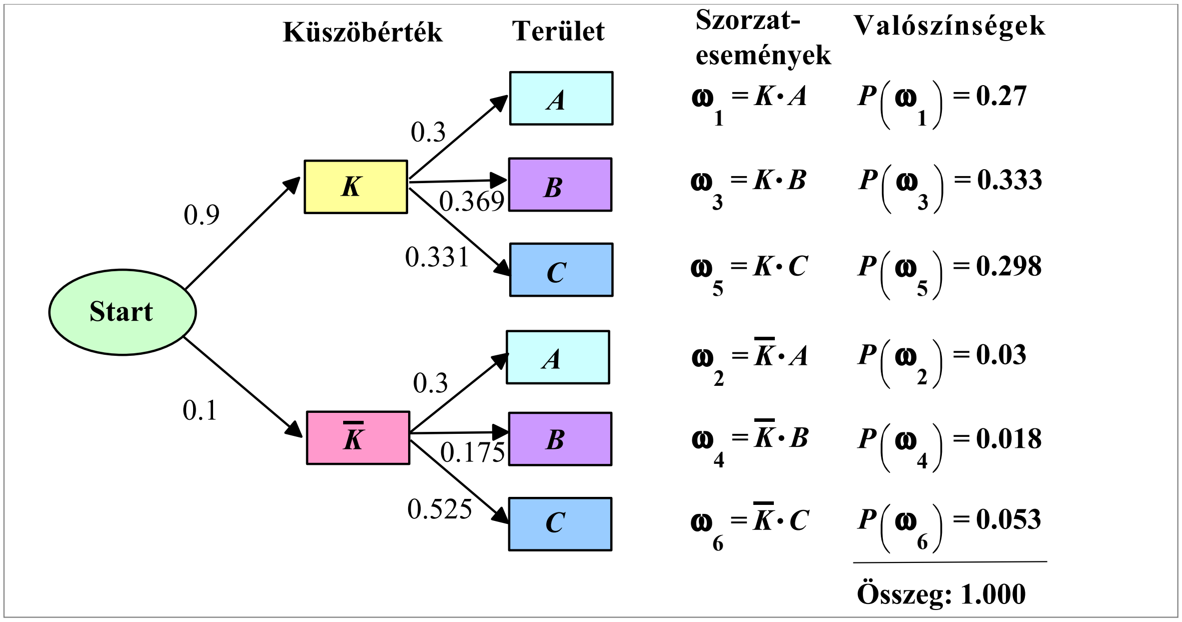
Tehát a vízminták 10 %-a haladja meg az egészségügyi küszöbértéket!
c) Készítse el az inverz fa diagramot a hozzá tartozó valószínségek számításával!
Az inverz fában az eredeti döntési fa szintjei fel vannak cserélve. Az els szint éleire az egészségügyi küszöbérték meghaladásának/nem meghaladásának valószínsége, a második szint éleire pedig az ún. Bayes-valószínségek kerülnek, melyek annak valószínségét adják meg, hogy a vízminta egy adott területrl érkezett az egészségügyi küszöbérték meghaladásának/nem meghaladásának feltétele mellett.
Annak valószínségét, hogy egy vízminta nitrát-tartalma az egészségügyi határérték alatt marad, a $P(K) + P(\overline{K}) = 1$ azonosság segítségével (vagy a b) ponthoz hasonlóan teljes valószínség tétellel) határozhatjuk meg:

```
> pK := 1 - pKc;
```

A Bayes-valószínségek a Bayes-tétel segítségével számolhatók (itt felhasználjuk a korábban kiszámolt $P(K)$ és $P(\overline{K})$ valószínségeket):

```
> `P(A|K)` := pKA/pK;
  `P(B|K)` := pKB/pK;
  `P(C|K)` := pKC/pK;
```

```
> `P(A|Kc)` := pKcA/pKc;
  `P(B|Kc)` := pKcB/pKc;
  `P(C|Kc)` := pKcC/pKc;
```



d) Az elírt küszöbértéken belül maradók között melyik területrl vett minták aránya volt a legkisebb?

A kérdés az alábbi feltételes valószínűségek közül kérdezi a legkisebbet:

```
> <`P(A|K)`, `P(B|K)`, `P(C|K)`>;
```

Tehát az *A* területrl vett minták között volt legkisebb (0.3) az elírt küszöbértéken belül maradók aránya.

e) Az elírt küszöbértéket meghaladók között melyik területrl vett minták aránya volt a legnagyobb?

A kérdés az alábbi feltételes valószínűségek közül kérdezi a legnagyobbat:

```
> <`P(A|Kc)`, `P(B|Kc)`, `P(C|Kc)`>;
```

Tehát a *C* területrl vett minták között volt legnagyobb (0.52) az elírt küszöbértéket meghaladók aránya.

f) Készítse el az együttes események valószínűségeinek táblázatát!

Együttes valószínűségek 2

	A	B	C	D	E	F
1		*küszöbértéken belül marad*	*meghaladja a küszöbértéket*			
2	*A terület*	0.27	0.03	0.3		
3	*B terület*	0.3325	0.0175	0.35		
4	*C terület*	0.2975	0.0525	0.35		
5		0.9000	0.1000	1.00		
6						
7						

6. Feladat (Grapefruit, ZH, 2016, B csoport)

Dél-Spanyolországban bizonyos fajta grapefruit gyümölcs shonos az *A*, *B* és *C* területeken. Az összes grapefruit termés 25 %-a az *A* területen, 30 %-a a *B* területen és a többi a *C* területen terem. A grapefruit termést idnként parazitafertzés (*F*) károsítja. Felmérések szerint az *A* területen a termés 5 %-a fertzött, *B* területen a termés 90 %-a ép

maradt és a *C* területen a termés 8 %-a fertzött.
a) Szemléltessük döntési fa diagramon a grapefruit termés területek szerinti és fertzöttség szerinti kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!
b) Az összes termés hány százaléka fertzött (*F*)?
c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!
d) Adjuk meg az összes fertzött termés százalékos megoszlását területek szerint!
e) Melyik terület hozzájárulása volt a legkisebb az összes ép terméshez? Hány százalékot jelent ez?
f) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Szemléltessük döntési fa diagramon a grapefruit termés területek szerinti és fertzöttség szerinti kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!

Rögzítsük az adatokat! A grapefruit termés *A*, *B* és *C* területre es hányada:

```
> pA, pB, pC := 0.25, 0.30, 0.45;
```

Annak a valószínsége, hogy az egyes területeken a termést parazitafertzés károsítja:

```
> `P(F|A)`, `P(F|B)`, `P(F|C)` := 0.05, 0.1, 0.08;
```

A feltételre szorítkozó valószínségi mezn számolva mindhárom terület esetén megkapjuk az ép termés részarányát:

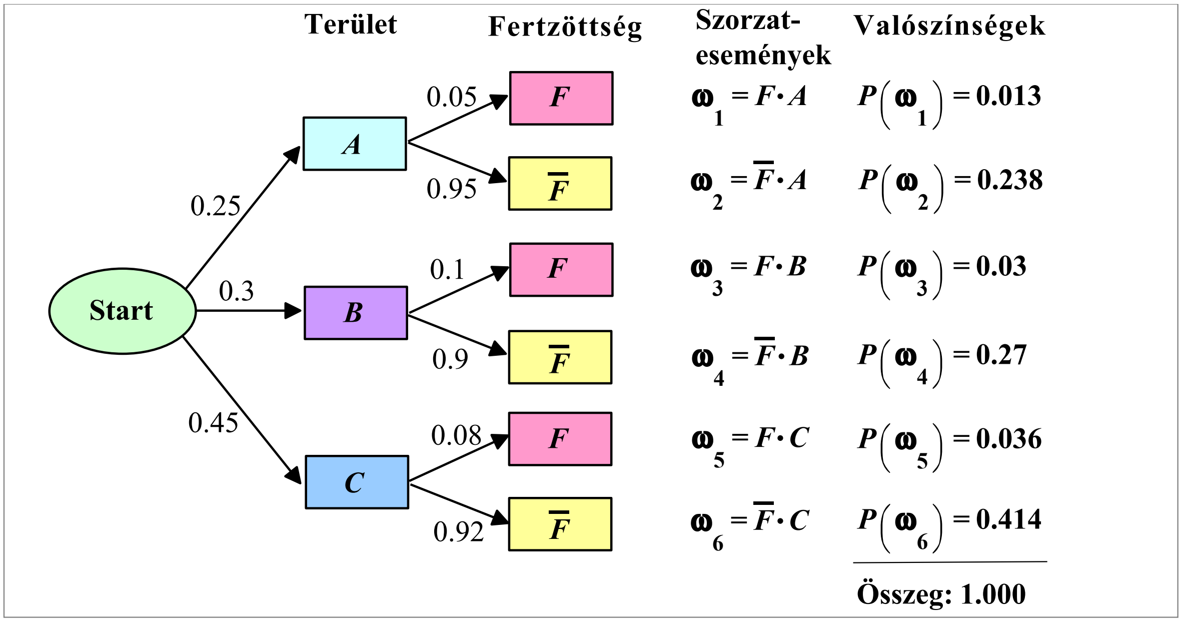
```
> `P(Fc|A)` := 1 - `P(F|A)`;
  `P(Fc|B)` := 1 - `P(F|B)`;
  `P(Fc|C)` := 1 - `P(F|C)`;
```

Most használjuk a valószínségek szorzástételét a szorzatesemények valószínségeinek meghatározásához!

```
> pFA := `P(F|A)`*pA;
  pFB := `P(F|B)`*pB;
  pFC := `P(F|C)`*pC;
```

```
> pFcA := `P(Fc|A)`*pA;
  pFcB := `P(Fc|B)`*pB;
  pFcC := `P(Fc|C)`*pC;
```

Ezeknek a valószínségeknek az ismeretében felrajzolhatjuk a döntési fát. A diagram els szintjén a terület szerint, a 2. szintjén pedig a fertzöttség szerint bontjuk fel az eseteket.



b) Az összes termés hány százaléka fertzött (*F*)?

A teljes valószínség tételét alkalmazzuk a $P(F)$ valószínség meghatározására: összeadjuk a döntési fa azon ágain számolt szorzat-valószínségeket, melyek a fertzött terméshez tartoznak.

```
> pF := pFA + pFB + pFC;
```

Tehát a grapefruit termés 7.85 %-a fertzött!

c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!

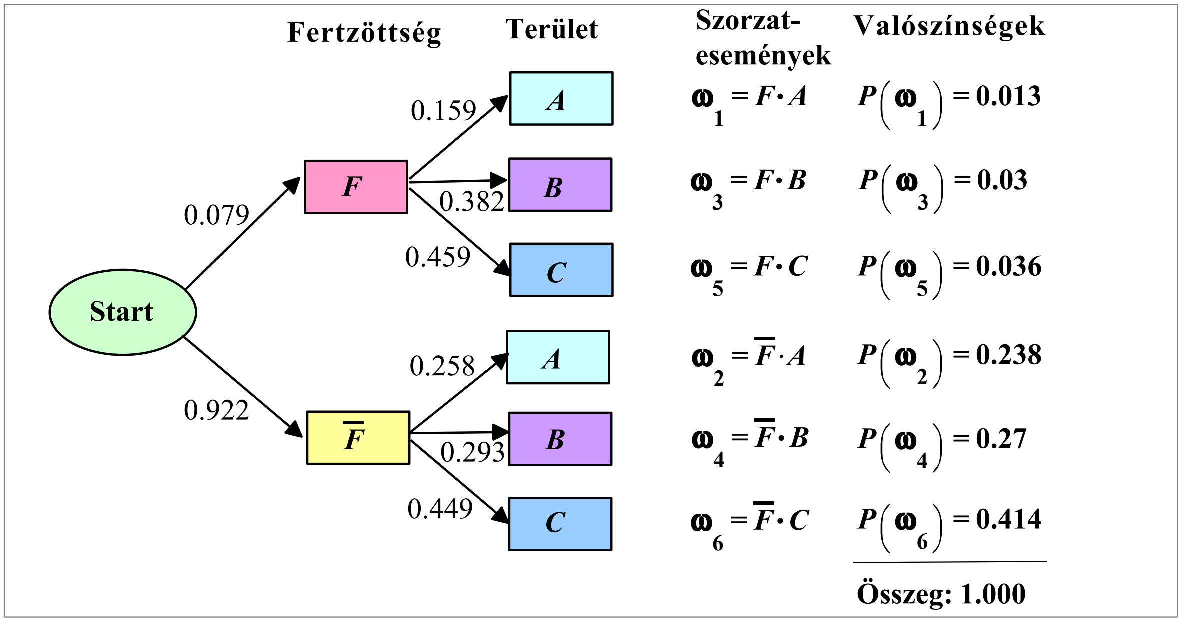
Az inverz fában az eredeti döntési fa szintjei fel vannak cserélve. Az els szint éleire a fertzöttség/épség valószínsége, a második szint éleire pedig az ún. Bayes-valószínségek kerülnek, melyek annak valószínségét adják meg, hogy egy grapefruit egy adott területrl származik a fertzöttség/épség feltétele mellett.

Annak valószínségét, hogy egy grapefruit ép maradt, a $P(F) + P(\overline{F}) = 1$ azonosság segítségével (vagy a b) ponthoz hasonlóan teljes valószínség tétellel) határozhatjuk meg:

```
> pFc := 1 - pF;
```

A Bayes-valószínségek a Bayes-tétel segítségével számolhatók (itt felhasználjuk a korábban kiszámolt $P(F)$ és $P(\overline{F})$ valószínségeket):

```
> `P(A|F)` := pFA/pF; `P(B|F)` := pFB/pF; `P(C|F)` :=
  pFC/pF;
> `P(A|Fc)` := pFcA/pFc; `P(B|Fc)` := pFcB/pFc; `P(C|Fc)`
  := pFcC/pFc;
```



d) Adjuk meg az összes fertzött termés százalékos megoszlását területek szerint!
A kérdés az alábbi feltételes valószínségek felsorolását jelenti:

```
> <`P(A|F)`, `P(B|F)`, `P(C|F)`>;
```

A százalékként való megjelenítéshez kattintsunk jobb egérgombbal az outputra és használjuk a *Numeric Formatting > Percent* (*Numerikus formázás > Százalék*) opciót.

e) Melyik terület hozzájárulása volt a legkisebb az összes ép terméshez? Hány százalékot jelent ez?
A kérdés az alábbi feltételes valószínségek közül kérdezi a legkisebbet:

```
> <`P(A|Fc)`, `P(B|Fc)`, `P(C|Fc)`>;
```

Tehát az *A* terület hozzájárulása volt a legkisebb az összes ép terméshez, mely 25.8 %.

f) Készítsük el az együttes események valószínségeinek táblázatát!

Szorzatesemények 3

	A	B	C	D	E	F	G	H
1		*Fertőzött*	*Nem fertőzött*					
2	*A terület*	0.0125	0.2375	0.25				
3	*B terület*	0.030	0.270	0.30				
4	*C terület*	0.0360	0.4140	0.45				
5		0.0785	0.9215	1.00				
6								
7								

7. Feladat (Gyümölcsös cukorkák)

Egy gyárban két gyártósoron háromféle gyümölcsös cukorkát gyártanak: citromosat (*C*), narancsosat (*N*) és málnásat (*M*). Az *A* gyártósor az össztermelés 30 %-áért felels, míg a *B* gyártósor az összes többiért. Az *A* gyártósoron készült cukorkák 40 %-a citromos, 35 %-a narancsos és a maradék málnás, míg a *B* gyártósoron készült cukorkák 20 %-a citromos, 30 %-a narancsos és a maradék málnás.
a) Szemléltessük döntési fa diagramon a cukorkák gyártósorok ill. ízesítés szerinti

kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!
b) Mekkora annak a valószínsége, hogy egy véletlenszeren kiválasztott cukorka citromos, narancsos, ill. málnás?
c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!
d) Melyik gyártósorról származik a narancsos cukorkák nagyobbik hányada. Hány százalákot jelent ez?
e) Melyik gyártósoron készül kevesebb málnás cukorka?
f) Ha egy cukorkáról tudjuk, hogy nem citromos, akkor mekkora az esélye, hogy az *A* gyártósoron készült?
g) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Szemléltessük döntési fa diagramon a cukorkák gyártósorok ill. ízesítés szerinti kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!

Rögzítsük az adatokat! Az A részesedése a termelésbl:

```
> pA := 0.3;
```

Ha egy cukorka nem az A gyártósoron készült, akkor a B-n, így $P(B)$ az ellentét-esemény valószínsége:

```
> pB := 1 - pA;
```

Annak valószínsége, hogy az A gyártósorról véletlenszeren választott cukorka citromos ill. narancsos:

```
> `P(C|A)`, `P(N|A)` := 0.4, 0.35;
```

Mivel minden cukorka citromos, narancsos vagy málnás az A gyártósoron, ezért $P(C|A) + P(N|A) + P(M|A) = 1$, amibl megkaphatjuk $P(M|A)$-t:

```
> `P(M|A)` := 1 - `P(C|A)` - `P(N|A)`;
```

Hasonlóan számolhatunk a B gyártósor esetében:

```
> `P(C|B)`, `P(N|B)` := 0.2, 0.3;
  `P(M|B)` := 1 - `P(C|B)` - `P(N|B)`;
```

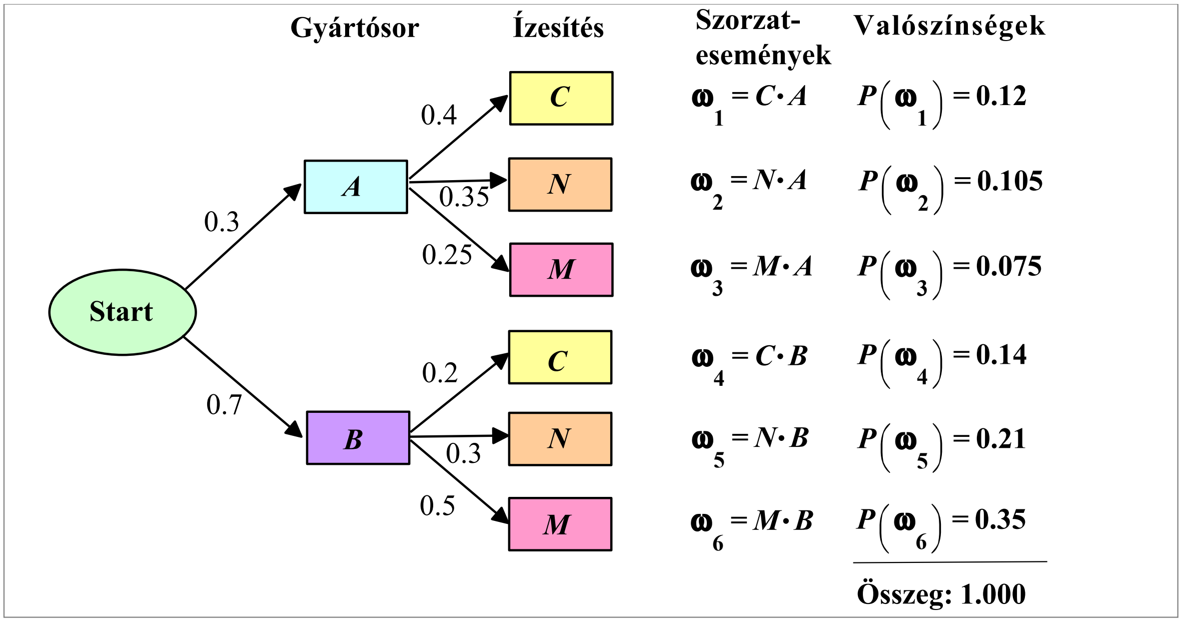
Ezután határozzuk meg a szorzatesemények valószínségeit a valószínségek szorzástételének segítségével!

```
> pCA := `P(C|A)`*pA;
  pNA := `P(N|A)`*pA;
  pMA := `P(M|A)`*pA;
> pCB := `P(C|B)`*pB;
  pNB := `P(N|B)`*pB;
  pMB := `P(M|B)`*pB;
```

A szorzatesemények valószínségeinek összege 1-et tesz ki:

```
> pCA + pNA + pMA + pCB + pNB + pMB;
```

Ezeknek a valószínségeknek az ismeretében felrajzolhatjuk a döntési fát. A diagram els szintjén a gyártósor szerint (kétfelé ágazás), a 2. szintjén pedig az ízesítés szerint (háromfelé ágazás) bontjuk fel az eseteket. Vegyük észre, hogy ez nem más, mint az $\{A, B\}$ és a $\{C, N, M\}$ teljes eseményrendszerek szerinti felbontás egymásutánja.



b) Mekkora annak a valószínsége, hogy egy véletlenszeren kiválasztott cukorka citromos, narancsos, ill. málnás?

A teljes valószínség tételét alkalmazzuk a $P(C)$, $P(N)$ és $P(M)$ valószínségek meghatározására: összeadjuk a döntési fa megfelel leveleire írt szorzat-valószínségeket:

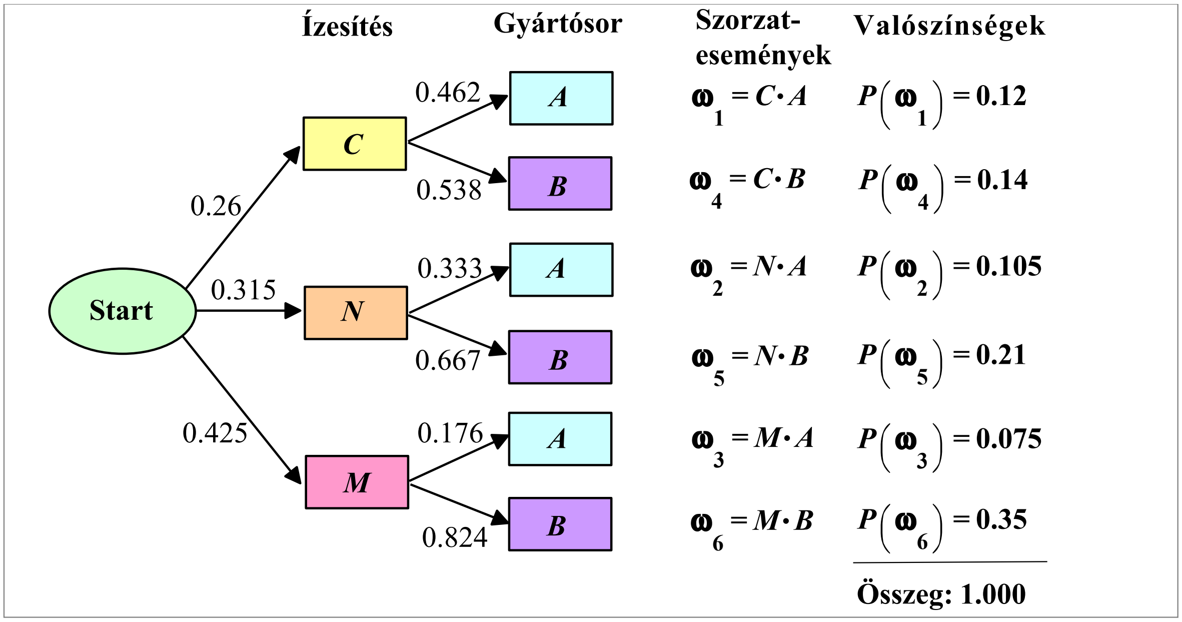
```
> pC  := pCA + pCB;
  pN  := pNA + pNB;
  pM  := pMA + pMB;
```

c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!

Az inverz fában az eredeti döntési fa szintjei fel vannak cserélve. Az els szint éleire az egyes ízesítések valószínsége, a második szint éleire pedig az ún. Bayes-valószínségek kerülnek, melyek az adott gyártósorról származás valószínségét adják meg, adott ízesítés feltétele mellett.

A Bayes-valószínségek a Bayes-tétel segítségével számolhatók (itt felhasználjuk az a) feladatrészbl a szorzatesemények valószínségeit, ill. a b) feladatrészben kiszámolt $P(C)$, $P(N)$ és $P(M)$ valószínségeket):

```
> `P(A|C)` := pCA/pC;
  `P(A|N)` := pNA/pN;
  `P(A|M)` := pMA/pM;
> `P(B|C)` := pCB/pC;
  `P(B|N)` := pNB/pN;
  `P(B|M)` := pMB/pM;
```



d) Melyik gyártósorról származik a narancsos cukorkák nagyobbik hányada. Hány százalákot jelent ez?
A kérdés az alábbi két feltételes valószínség közül kérdezi a kisebbiket:

```
> <`P(A|N)`, `P(B|N)`>;
```

Tehát a B gyártósorról származik a narancsos cukorkák többsége, 66.67 %-a (azaz $\frac{2}{3}$-a).

e) Melyik gyártósoron készül kevesebb málnás cukorka?
A kérdés az alábbi két feltételes valószínség közül kérdezi a kisebbiket:

```
> <`P(A|M)`, `P(B|M)`>;
```

Tehát az A gyártósoron készül kevesebb málnás cukorka.

f) Ha egy cukorkáról tudjuk, hogy nem citromos, akkor mekkora az esélye, hogy az A gyártósoron készült?
Ha egy cukorka nem citromos, akkor narancsos vagy málnás. A feladatunk tehát a $P(A|N+M)$ feltételes valószínség meghatározása. Definíció szerint

$$P(A|N+M) = \frac{P(A\cdot(N+M))}{P(N+M)} = \frac{P(A\cdot N + A\cdot M)}{P(N+M)}.$$

Mivel N és M egymást kölcsönösen kizáró események (egy cukorka nem lehet egyszerre narancsos és málnás íz is), ezért kihasználhatjuk a valószínségi mérték additivitását:

$$\frac{P(A\cdot N + A\cdot M)}{P(N+M)} = \frac{P(A\cdot N) + P(A\cdot M)}{P(N) + P(M)}.$$

A megoldás tehát:

```
> `P(A|Cc)` := (pNA + pMA)/(pN + pM);
```

Azaz nagyjából 24.32 % annak az esélye, hogy egy nem citromos cukorka az A gyártósoron készült.

g) Készítsük el az együttes események valószínségeinek táblázatát!

	Spreadsheet(1)							
	A	B	C	D	E	F	G	H
1		*Citromos*	*Narancsos*	*Málnás*				
2	*A gyártósor*	0.12	0.105	0.075	0.3			
3	*B gyártósor*	0.14	0.21	0.35	0.7			
4		0.26	0.315	0.425	1.0			
5								
6								
7								

8. Feladat (Cipkereskedés)

Egy cip-nagykeresked az *X*, *Y* és *Z* gyártól vásárol cipket 20, 35, ill. 45 %-os arányban. Az *X*, *Y* ill. *Z* gyár cipi 5, 4, ill. 2 %-ban hibásaknak (*H*) bizonyultak.
a) Szemléltessük döntési fa diagramon a cipk gyártók szerinti ill. hibásság szerinti kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!
b) Mekkora annak a valószínsége, hogy egy véletlenszeren kiválasztott cip hibás ill. hibátlan?
c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!
d) Adjuk meg a hibás cipk százalékos megoszlását gyártók szerint!
e) A hibátlan cipk között melyik gyártó termékei vannak legnagyobb arányban? Mekkora ez az arány?
f) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

a) Szemléltessük döntési fa diagramon a cipk gyártók szerinti ill. hibásság szerinti kétszint megoszlását! Számoljuk ki a szorzatesemények valószínségeit is!

Rögzítsük az adatokat! Az *X*, *Y* és *Z* gyártótól vásárolt cipk aránya:

```
> pX, pY, pZ := 0.2, 0.35, 0.45;
```

Annak a valószínsége, hogy az egyes gyártóktól vásárolt cip hibásnak bizonyul:

```
> `P(H|X)`, `P(H|Y)`, `P(H|Z)` := 0.05, 0.04, 0.02;
```

A feltételre szorítkozó valószínségi mezn számolva mindhárom gyártó esetén megkapjuk a hibátlan cipk részarányát:

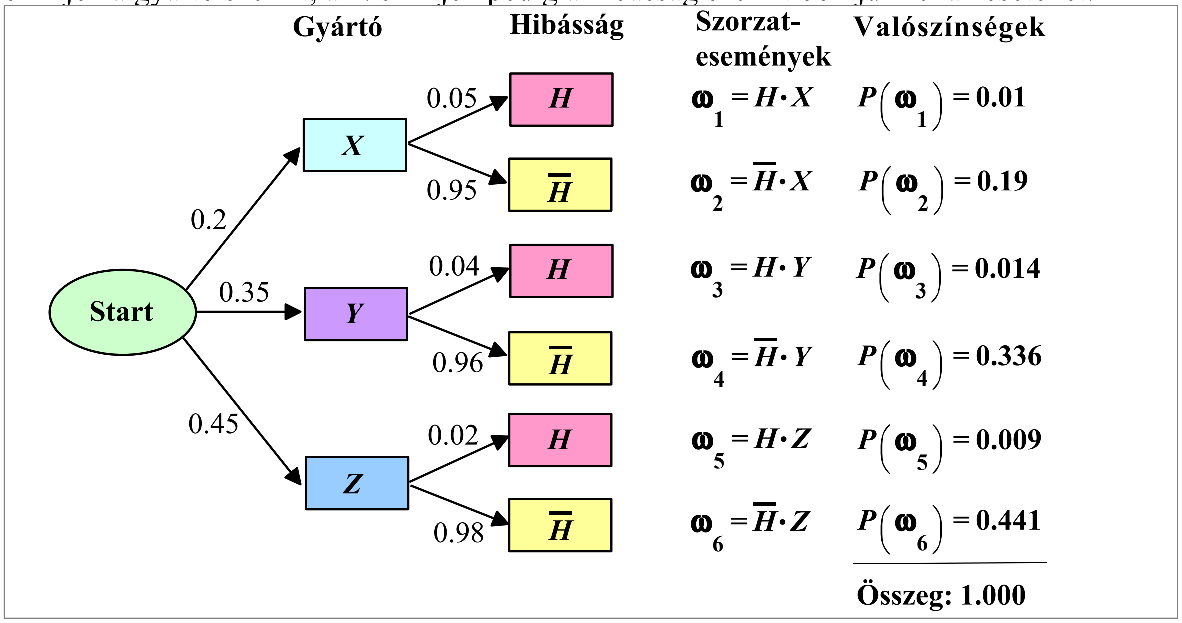
```
> `P(Hc|X)` := 1 - `P(H|X)`;
  `P(Hc|Y)` := 1 - `P(H|Y)`;
  `P(Hc|Z)` := 1 - `P(H|Z)`;
```

Most használjuk a valószínségek szorzástételét a szorzatesemények valószínségeinek meghatározásához!

```
> pHX := `P(H|X)`*pX;
  pHY := `P(H|Y)`*pY;
  pHZ := `P(H|Z)`*pZ;
```

```
> pHcX := `P(Hc|X)`*pX;
  pHcY := `P(Hc|Y)`*pY;
  pHcZ := `P(Hc|Z)`*pZ;
```

Ezeknek a valószínségeknek az ismeretében felrajzolhatjuk a döntési fát. A diagram els szintjén a gyártó szerint, a 2. szintjén pedig a hibásság szerint bontjuk fel az eseteket.



b) Mekkora annak a valószínsége, hogy egy véletlenszeren kiválasztott cip hibás ill. hibátlan?

A teljes valószínség tételét alkalmazzuk a $P(H)$ valószínség meghatározására: összeadjuk a döntési fa azon ágain számolt szorzat-valószínségeket, melyek a hibás cipkhöz tartoznak.

```
> pH := pHX + pHY + pHZ;
```

Tehát a cipk 3.3 %-a hibás.

A hibátlan cipk arányát megkapjuk, ha 1-bl kivonjuk a hibás cipk arányát (H és $\overline{H}$ ellentét események). (Vagy teljes valószínség tételt is alkalmazhatunk, $P(H)$ kiszámításához hasonlóan.)

```
> pHc := 1 - pH;
```

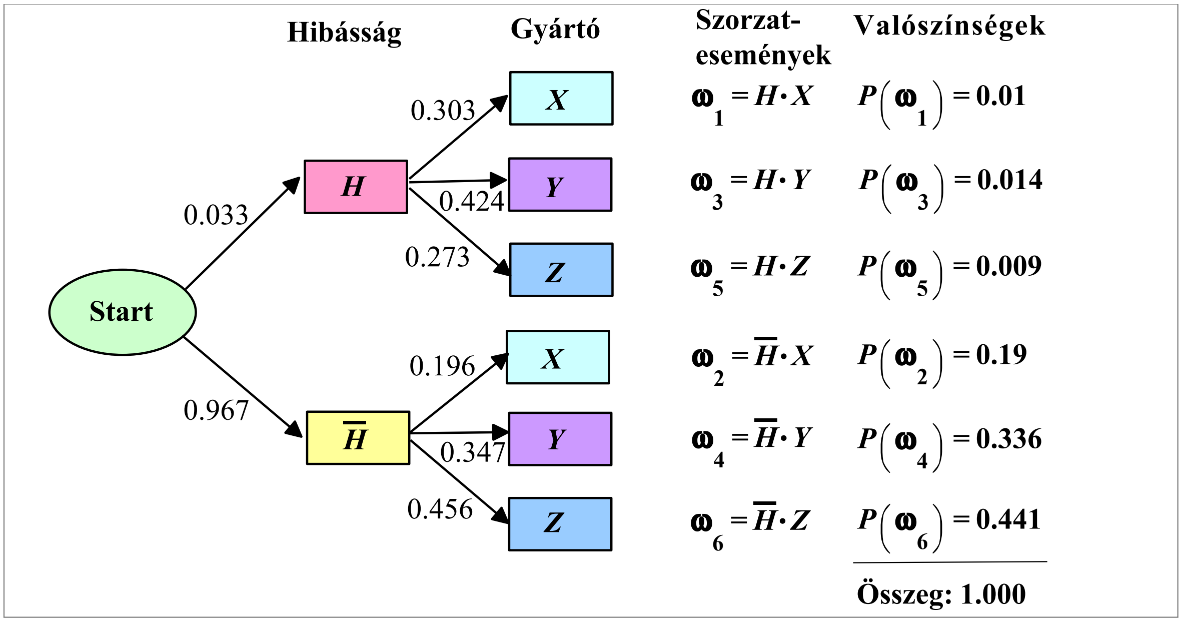
Tehát a cipk 96.7 %-a hibátlan.

c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!

Az inverz fában az eredeti döntési fa szintjei fel vannak cserélve. Az els szint éleire a hibásság/hibátlanság valószínsége, a második szint éleire pedig az ún. Bayes-valószínségek kerülnek, melyek annak valószínségét adják meg, hogy egy cip egy adott gyártótól származik a hibásság/hibátlanság feltétele mellett.

A Bayes-valószínségek a Bayes-tétel segítségével számolhatók (itt felhasználjuk a korábban kiszámolt $P(H)$ és $P(\overline{H})$ valószínségeket):

```
> `P(X|H)` := pHX/pH; `P(Y|H)` := pHY/pH; `P(Z|H)` :=
  pHZ/pH;
> `P(X|Hc)` := pHcX/pHc; `P(Y|Hc)` := pHcY/pHc; `P(Z|Hc)`
  := pHcZ/pHc;
```



d) Adjuk meg a hibás cipk százalékos megoszlását gyártók szerint!

A kérdés az alábbi feltételes valószínségek felsorolását jelenti:

```
> <`P(X|H)`, `P(Y|H)`, `P(Z|H)`>;
```

A százalékként való megjelenítéshez kattintsunk jobb egérgombbal az outputra és használjuk a *Numeric Formatting > Percent* (*Numerikus formázás > Százalék*) opciót.

e) A hibátlan cipk között melyik gyártó termékei vannak legnagyobb arányban? Mekkora ez az arány?

A kérdés az alábbi feltételes valószínségek közül kérdezi a legnagyobbat:

```
> <`P(X|Hc)`, `P(Y|Hc)`, `P(Z|Hc)`>;
```

Tehát a hibátlan cipk között az *Z* gyártó termékei vannak a legnagyobb arányban. Ez az arány 45.6 *%*.

f) Készítsük el az együttes események valószínségeinek táblázatát!

Szorzatesemények 3

	A	B	C	D	E	F	G	H
1		*Hibás*	*Hibátlan*					
2	*X gyártó*	0.010	0.190	0.2				
3	*Y gyártó*	0.0140	0.3360	0.35				
4	*Z gyártó*	0.0090	0.4410	0.45				
5		0.0330	0.9670	1.00				
6								
7								

Gyakorló feladatok

Gy/1. Feladat (Áruház beszállítói / 2.)

Egy áruház három különböz beszállítótól veszi a banánt.

Az "A" beszállító termékeinek 20 %-a minségileg enyhén kifogásolható (másodosztályú), de mégis ettl veszik a teljes mennyiség 50 %-át, mert olcsó.
A "B" beszállító 95 %-ban els osztályú terméket szállít, de a többiekhez képest drágábban, ezért a teljes készlet 20 %-át veszik csak tlük.
A maradék a "C" beszállítótól származik, melynek termékei 10 %-ban másodosztályúak.
a) Készítsünk az adatokból fa diagramot. Számoljuk ki a szorzatesemények valószínségeit is!
b) A bolt teljes kínálatában mennyi az els osztályú termékek aránya?
c) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!
d) Melyik beszállítótól érkezik a legtöbb másodosztályú banán?
e) Készítsük el az együttes események valószínségeinek táblázatát!

Megoldás

```
> restart;
```

Gy/2. Feladat (Hamis dobókocka)

Van 4 dobókockánk, 3 szabályos és 1 hamis. A hamis kockán a 4-es szám helyén is 6-os van (így összesen két oldalán van 6-os). Véletlenszeren kiválasztunk egy kockát és dobunk vele.
a) Mekkora annak a valószínsége, hogy 6-ost (*H*) illetve hogy 4-est (*N*) dobtunk?
b) Mekkora annak a valószínsége, hogy nem dobtunk sem 4-est, sem 6-ost (*S*)?
c) Feltéve, hogy 6-ost dobtunk, mennyi a valószínsége, hogy a cinkelt kockával dobtunk?
d) Feltéve, hogy 4-est dobtunk, mennyi a valószínsége, hogy egy szabályos kockával dobtunk?
e) Ha tudjuk, hogy nem dobtunk sem 4-est, sem 6-ost, akkor melyiknek van nagyobb esélye: annak, hogy egy szabályos vagy annak, hogy a hamis kockával dobtunk?
f) Készítsük el az inverz fa diagramot a megfelel valószínségekkel!

Megoldás

```
> restart;
```