Univerza v Ljubljani
Fakulteta za matematiko in fiziko

Seminarska naloga pri predmetu SEMINAR

KRVNI TESTI

Ljubljana, 2009/2010 Diana Šinkovec

1. UVOD

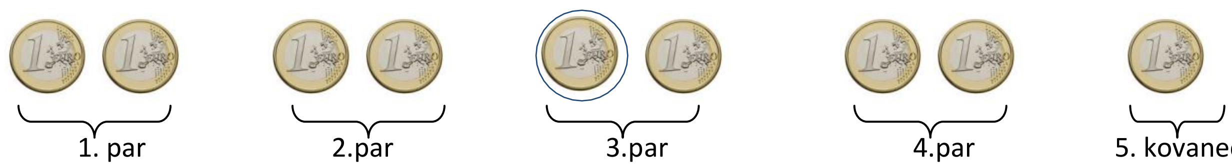
V tej seminarski nalogi sem se ukvarjala z raziskavami krvnih vzorcev. Zanimalo me je, kako lahko iz krvnih vzorcev dojenčkov razberemo njihovo bolezen (motnjo, virus, …). Seveda je najlažje pregledati kri vsakega dojenčka posebej (npr. če moramo pregledati 100 otrok, moramo narediti 100 testov). Ampak, ali obstaja način, da bi zmanjšali število testov (npr. da bi jih naredili manj kot 100)? Preden pa odgovorimo na to vprašanje, si poglejmo podoben primer.

2. PROBLEM TEHTANJA KOVANCEV

Za lažje razumevanje našega problema si poglejmo malo lažji in bolj predstavljiv problem – problem tehtanja kovancev. V tem poglavju bomo izmed vseh kovancev, ki nam bodo dani, poskušali na čim bolj učinkovit način (čim hitreje oz. s čim manj koraki) najti ponarejen kovanec.

Pa recimo, da imamo danih 9 kovancev: 8 je pravih, 1 pa je ponarejen in je malo težji kot navaden kovanec. Zanima nas, v koliko korakih (oz. tehtanjih) prepoznamo ponarejen kovanec?

Razporedimo kovance v 4 skupine po 2, eden pa naj ostane sam (ponarejen kovanec sem obkrožila z modro barvo):



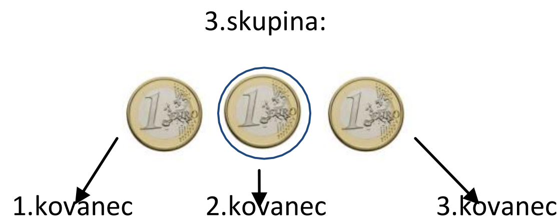
Sedaj moramo s pomočjo tehtnice (na podlagi teže kovancev) ugotoviti, kateri je ponarejen. Naša tehtnica nam pokaže le, katera stran je težja, ne pa točne teže kovancev.

Vzamemo 1. par in stehtamo en kovanec proti drugemu. Ugotovimo, da ni nobene razlike v teži, saj je tehtnica v ravnovesju. Torej sta oba kovanca prava. Nato vzamemo 2. par in stehtamo en kovanec proti drugemu. Tehtnica je zopet v ravnovesju. Vzamemo 3. par in damo oba kovanca na tehtnico. Vidimo, da je npr. leva stran težja od desne. Kovanec na levi strani je torej težji, zato ugotovimo, da je ponarejen. Četrtega para nam ni potrebno tehtati, saj smo ponarejeni kovanec že našli. Ugotovimo, da smo ponaredek našli v 3 tehtanjih. Če bi bil ponarejen kovanec v četrtem paru ali pa osamljen kovanec, pa bi potrebovali še eno tehtanje več, torej skupno 4 tehtanja.

A je to najbolj učinkovit način za iskanje ponarejenega kovanca?

Še vedno imamo 9 kovancev (8 pravih in 1 ponarejenega). Razporedimo jih v 3 skupine po 3 kovance:

Sedaj pa stehtajmo 1. skupino proti 2. skupini. Vidimo, da je tehtnica v ravnovesju. Sklepamo, da je ponaredek v tretji skupini. Sedaj moramo ugotoviti, kateri izmed teh treh kovancev ni pravi.



Sedaj pa 1. kovanec stehtamo proti tretjemu in opazimo, da ni razlike v teži. Takoj lahko torej sklepamo, da je ponaredek 2. kovanec (če bi stehtali 1. in 2. kovanec ali 2. in 3. kovanec, bi opazili, kateri del tehtnice je težji in bi tako našli ponaredek). V tem postopku smo za ugotovitev lažnega kovanca rabili le 2 tehtanji.

Kaj pa če bi imeli 27 kovancev (26 pravih in 1 ponaredek)?

Kovance najprej razdelimo v 3 skupine po 9:

Zopet se problema lotimo podobno kot v primeru z 9 kovanci. Stehtamo 1. skupino proti 2. skupini. Ugotovimo, da je 2. skupina težja. Ponarejen kovanec se torej skriva v njej. Vzamemo osumljeno deveterico. Sedaj iščemo ponaredek med devetimi kovanci. To pa že znamo. Razdelimo jih v tri skupine po tri. Najprej ugotovimo, v kateri skupini je ponaredek, nato pa, kateri je lažni kovanec. Pri 27 kovancih smo problem rešili s 3 tehtanji.

Kaj pa, če bi imeli 81 kovancev (80 pravih in 1 ponaredek)?

Tako kot doslej začnemo z razporeditvijo kovancev v 3 skupine. V vsaki imamo 27 kovancev. Stehtamo 2 skupini (eno proti drugi) in ugotovimo, kje je ponaredek. Nato vzamemo osumljeno skupino 27 kovancev in te razporedimo v tri deveterice. Spet stehtamo 2 skupini (eno proti drugi) in ugotovimo, v kateri je ponaredek. Nato to razdelimo v 3 skupine po 3 kovance. Stehtamo eno skupino proti drugi. Iz osumljene skupine ugotovimo še, kateri je ponarejen kovanec. Tu smo rabili 4 tehtanja.

V splošnem:

št. kovancev	št. tehtanj
$9 = 3^2$	2
$27 = 3^3$	3
$81 = 3^4$	4
$\vdots$	$\vdots$
3^m	m

Če imamo 3^m kovancev, pridemo do rešitve problema v m tehtanjih. Preveriti moramo še, če je m res najmanjša možna številka tehtanj.

Vedno ko stehtamo kovance, si postavimo vprašanje, ki ima 3 možne odgovore:

a) levi del tehtnice je težji,
b) tehtnica je v ravnovesju,
c) desni del tehtnice je težji.

Na začetku imamo 3^m kovancev (npr.: $3^4; m = 4$). Razporedimo jih v 3 skupine in s prvim tehtanjem določimo osumljeno skupino. Sedaj imamo 3^{m-1} kovancev (npr.: 3^3: $m - 1 = 3$). Nato zopet določimo skupino, v kateri je gotovo ponarejen kovanec. Imamo samo še 3^{m-2} kovancev (npr.: 3^2: $m - 2 = 2$). In nadaljujemo v tej smeri. Imamo samo še $3^0 = 1$ kovanec (ta je ponarejen). Ugotovimo, da ne moremo priti do ponaredka v manj kot m korakih. Lahko bi rekli tudi takole: če ima vsako vprašanje 3 možne odgovore, potem ne moremo med 3^m možnostmi razlikovati z manj kot m vprašanji. Lahko pa si postavimo vprašanje, ki ima dva možna odgovora:

a) da,
b) ne.

Če si postavimo tako vprašanje, potem lahko 2^m možnosti razlikujemo z m vprašanji. Privzemimo, da nas bodo od sedaj naprej zanimala vprašanja z dvema možnima odgovoroma.

$$n = 2^m \quad \Rightarrow \quad m = \log_2 n$$

Vidimo, da je število vprašanj veliko manjše od števila različnih možnosti, med katerimi razlikujemo ($m < 2^m$). To pa je zelo dobro za naš problem s krvnimi testi, kjer moramo delati poskuse na velikem številu dojenčkov, število testov pa bi radi, da bi bilo majhno.

3. KRVNI TESTI

Spomnimo se našega problema. S čim manj testi vzorcev krvi želimo preveriti, če ima kakšen dojenček virus ali kakšno motnjo. Lahko testiramo vsak vzorec posebej ali pa po nekaj vzorcev skupaj. Slednje zgleda bolj učinkovito. Pri takem testiranju lahko ugotovimo naslednje:

a) vsaj eden kaže virus oz. motnjo,
b) vsi so zdravi.

(Vsako vprašanje ima dva možna odgovora: a in b.)

Težave imamo lahko le pri določanju oz. ugotavljanju, kateri otrok je okužen. Recimo, da imamo n *otrok*. Vsak otrok ima dve možnosti: lahko je okužen (O) ali zdrav (Z). Vseh možnih stanj je torej 2^m. Naredimo primer s tremi otroki ($n = 3$). Imamo $2^3 = 8$ možnih situacij:

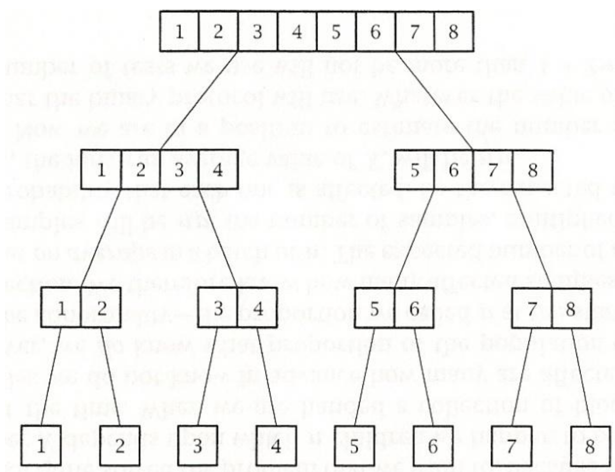
otrok 1	otrok 2	otrok 3
Z	Z	Z
Z	Z	O
Z	O	Z
O	Z	Z
Z	O	O
O	Z	O
O	O	Z
O	O	O

Če imamo torej n otrok, bomo za razlikovanje morali narediti n testov. Vendar to nam ni preveč všeč, saj to pomeni, da bomo morali narediti test pri vsakem otroku posebej. To je precej logično, saj je ena od možnosti tudi, da so okuženi vsi otroci in takrat bomo gotovo morali testirati prav vsakega (pri kovancih pa smo imeli vedno le en ponaredek).

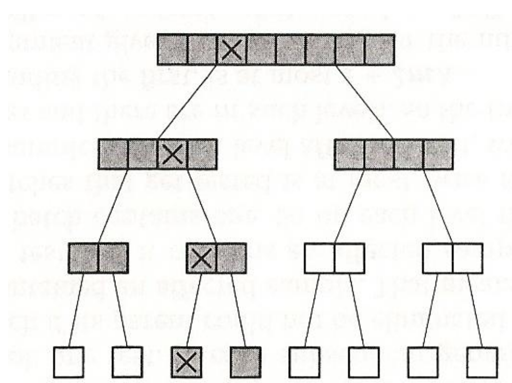
Recimo, da imamo 100 krvnih vzorcev in vemo, da je motnja prisotna le pri 1% populacije. Pričakovali bomo, da le eden od stotih ni zdrav ($1\%\ od\ 100$). Lažje bi torej bilo, če bi zagotovo vedeli, da je izmed vseh okužen le eden (to pa je podobno problemu tehtanja kovancev). V tem primeru ne bomo pričakovali, da moramo opraviti 100 testov, vendar le $\log_2 100 \quad oz.\ \ 6{,}6$. Kakšna je sedaj najučinkovitejša rešitev? Privzemimo tudi $n = 2^m \;\Rightarrow\; m = \log_2 n$. (Primer: Če imamo 100 vzorcev: $2^6 = 64\ in\ 2^7 = 128$, moramo narediti največ 7 testov, saj je $128 > 100$.) Postopek, pri katerem možnosti delimo v dve skupini, bomo poimenovali DVOJIŠKI POSTOPEK.

4. DVOJIŠKI POSTOPEK

Recimo, da je verjetnost za motnjo v populaciji enaka p. Imamo n testov in želimo izvesti dvojiški postopek. Lahko se zgodi, da ni okužen noben izmed n dojenčkov. Tako je naš test negativen in zaključimo postopek z enim samim testiranjem. Če je prvi test pozitiven, potem vzorce razdelimo v dve skupini: $\frac{n}{2} = 2^{m-1}$. Nato zopet testiramo ti dve skupini in tisto, pri kateri je test negativen (lahko tudi pri obeh), razdelimo zopet v dve skupini: $\frac{n}{4} = 2^{m-2}$. Nadaljujemo s tem postopkom. Za lažjo predstavo si to lahko ponazorimo s tako imenovanim *družinskim drevesom*:



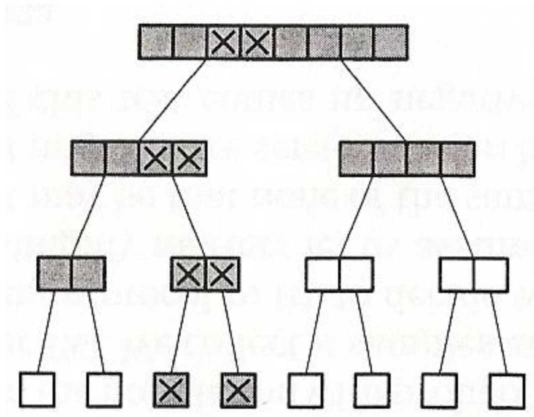
Naslednje drevo kaže, kaj se zgodi, če začnemo z $2^3 = 8$ vzorcev, vemo pa, da le eden kaže dojenčkovo motnjo (osenčena mesta v tabeli so tista, katera moramo še nadalje testirati):



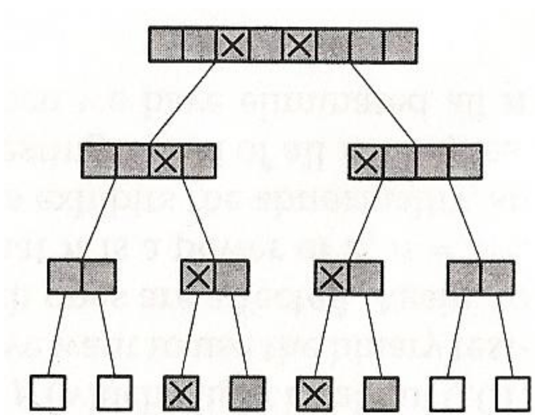
Kaj pa storimo, če imamo več okuženih vzorcev?

Naslednji dve družinski drevesi sta primera dveh različno porazdeljenih okuženih vzorcev:

1. primer: okužena vzorca sta na 3. in 4. mestu



2. primer: okužena vzorca sta na 3. in 5. mestu

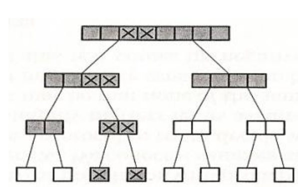


Sedaj bomo ocenili, koliko testov moramo narediti, da najdemo vse vzorce, ki kažejo zdravstvene motnje pri dojenčkih $(n = 2^m)$:

št. vzorcev, ki kažejo motnjo, označimo z X:

$$(*) \qquad \textbf{število testov} \le \mathbf{1 + 2mX}$$

Preverimo to formulo za $n = 8 = 2^3; m = 3;\ X = 1;\ 1 + 2mX = 1 + 2 \cdot 3 \cdot 1 = 7$. Torej ne bi smeli narediti več kot 7 testov (glej spodnjo sliko: naredili smo 7 testov, saj je osenčenih 7 mest).



Dokaz enačbe (*): Če je število vzorcev $\mathrm{n} = 2^{\mathrm{m}}$, ima družinsko drevo $(m + 1)$ generacij: zgornjo (1) in vse pod njo (m). Začnemo s testiranjem zgornje in naredimo na njej en test. V vsaki »veji« oz. stopnji spodaj testiramo le, če je starš (ena stopnja višje) okužen. Na vsaki stopnji je torej skupno število testiranj največ dvakrat večje od števila okuženih vzorcev (2X). Imamo m takih stopenj: $\underline{m \cdot 2X}$. Število testov je torej največ $1 + 2mX$. S tem smo enačbo dokazali.

Pišimo $X = np$, saj je p verjetnost za motnjo v populaciji: $1 + 2mX = 1 + 2mnp$.

Upoštevamo lahko še $m = \log_2 n$: $1 + 2mnp = 1 + 2(log_2 n)np$.

Predpostavimo, da je $p = \frac{1}{n}$ (kot v primeru s 100 vzorci in $p = 1\%$): $1 + 2(log_2 n)$.

Ta ocena je še bolj podobna $\log_2 n$.

SKLEP: Če vemo, da je okužen le en otrok, bo število testov seveda manjše. Če ne vemo, koliko dojenčkov je okuženih, pa bomo potrebovali le približno dvakrat toliko testov, kar pomeni, da je dvojiški postopek zelo dober.

5. DODATNE NALOGE

NALOGA: Na banko je prispelo 10 vreč s kovanci. Vsak kovanec je običajno težak 10 gramov. Vendar smo eno vrečo zamenjali z vrečo, v kateri so vsi kovanci težki 9 gramov. Kako s čim manj tehtanji ugotoviti, v kateri vreči so lažji kovanci? Dovoljeno je tudi odpiranje vreč. Tu je naša tehtnica taka, da nam pokaže točno težo in ne samo, katera stran je težja.
(http://forum.feri.uni-mb.si/Default.aspx?g=posts&t=3652)

To lahko naredimo s samo enih tehtanjem. Iz prve vreče vzamemo ven 1 kovanec, iz druge vzamemo ven 2, iz tretje 3, …, iz desete pa 10 kovancev. Vse te kovance damo na tehtnico. Če bi bili vsi kovance težki 10 gramov, bi tehtnica pokazala 550 gramov. Če je odstopanje za 1 gram (tehtnica pokaže 549 gramov), vemo, da so 9 gramski kovanci v prvi vreči. Če tehtnica pokaže 548 gramov, so lažji kovanci v 2. vreči. Itd. Kratka in zanimiva naloga.

NALOGA: Imamo 6 žog: 2 modri, 2 rdeči in 2 beli. Modri žogi nista enako težki (ena je težja od druge), ravno tako tudi nista obe beli žogi enako težki (ena je težja od druge) in enako velja za obe rdeči žogi (ena je težja od druge). Vemo pa še, da je ena od obeh modrih žog težka m, ena od obeh belih je tudi težka m in ena od 2 rdečih je težka m (vse tri so skupaj težke 3m – prva trojica žog). Ostale žoge (ena modra, ena rdeča in ena bela – druga trojica žog) pa so skupaj težke 3n (vsaka je težka n). Vemo še, da je ena trojica žog težja od druge trojice. Lahko z dvema tehtanjema razlikujemo žoge? Naša tehtnica nam pokaže le, katera stran je težja.
(http://www.cut-the-knot.org/blue/6Balls2Weighings.shtml)

6 različnih žog bomo označili tako: A1, A2, B1, B2, C1, C2.

<u>Sledi 1. tehtanje, pri katerem stehtamo (A1 in B1) proti (B2 in C1).</u> A1 in B1 damo na desno stran tehtnice, B2 in C1 pa na levo. Imamo tri možne izide tega tehtanja:

a) leva stran je težja,
b) ravnovesje,
c) desna stran je težja.

Poglejmo si najprej možnost a:

Kaj mora veljati, da sta C1 in B2 res težji od A1 in B1? Sklepamo lahko, da mora biti B1 lažji od B2, A1 in C1 pa mora biti ali oba težka ali pa oba lahka. Torej (znak < pomeni lažji, > pa težji):

$B1 < B2 \Rightarrow B1 = B_L in\ B2 = B_H$ (L-lažja žoga; H-težja žoga)

$$če\ A1 = A_H \Rightarrow C1 = C_H$$

$$če\ C1 = C_L \Rightarrow A1 = A_L$$

Sledi drugo tehtanje: (A1 in C1) proti (B1 in B2). Lahko se zgodi, da bo tehtnica v ravnovesju. Ker že vemo, da je $B1 = B_L in\ B2 = B_H$, lahko sklepamo naslednje: $\mathrm{A1} = \mathrm{A_L}$ in $\mathrm{C1} = \mathrm{C_H}$ in zato je $\mathrm{A2} = \mathrm{A_H}$in $\mathrm{C2} = \mathrm{C_L}$ (rešitev). Če tehtnica ni v ravnovesju, je lahko desna stran težja ali lažja od leve. Pa poglejmo najprej primer, ko je desna stran (B1 in B2) težja od leve (A1 in C1). Sklepamo lahko: $A1 = A_L\ in\ C1 = C_L$ in zato $A2 = A_H$ in $C2 = C_H$ (rešitev). Če pa je desna stran lažja, ugotovimo naslednje: $A1 = A_H\ in\ C1 = C_H$, posledično tudi $A2 = A_L\ in\ C2 = C_L$ (rešitev).

Poglejmo možnost b:

Leva in desna stran sta v ravnovesju. Sledi drugo tehtanje: B1 proti B2. Če je B1 lažji, velja: $B1 = B_L \Rightarrow A1 = A_H\ in\ C1 = C_L$, s tem smo določili tudi B2, A2 in C2 (rešitev). Če pa je B1 težji, ugotovimo naslednje: $B1 = B_H \Rightarrow A1 = A_L\ in\ C1 = C_H$ (rešitev).

Ostala nam je še možnost c:

Ta možnost je zelo podobna primeru a, zato ne bom pisala rešitve.

Nalogo smo uspeli rešiti z dvema tehtanjema.

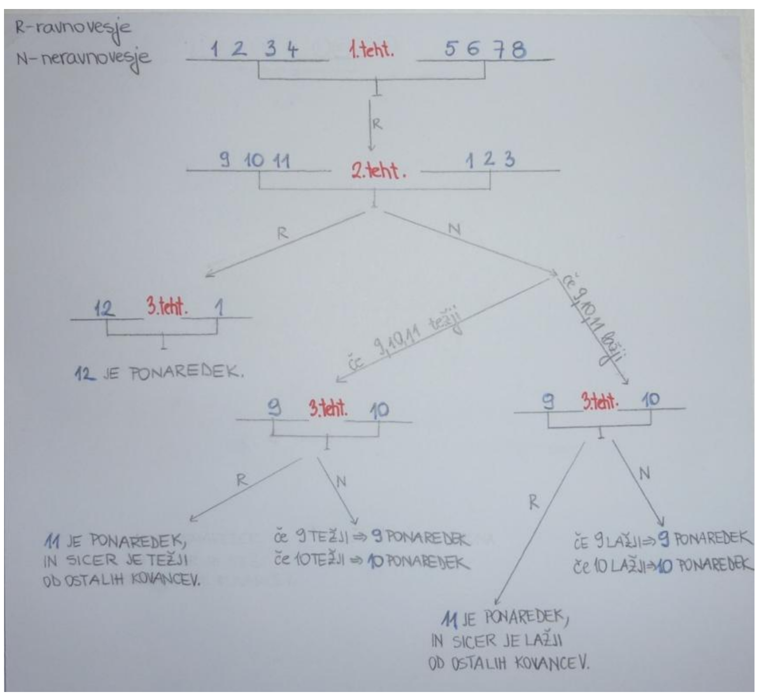
NALOGA: Danih imamo 12 kovancev, med katerimi je eden ponarejen, vendar ne vemo ali je lažji ali težji od ostalih. Ali lahko s tremi tehtanji ugotovimo, kateri je ponarejen in ali je težji ali lažji od pravih kovancev? Naša tehtnica nam pokaže le, katera stran je težja.

(http://mathforum.org/dr.math/)

Najprej označimo vse kovance: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ,11, 12.

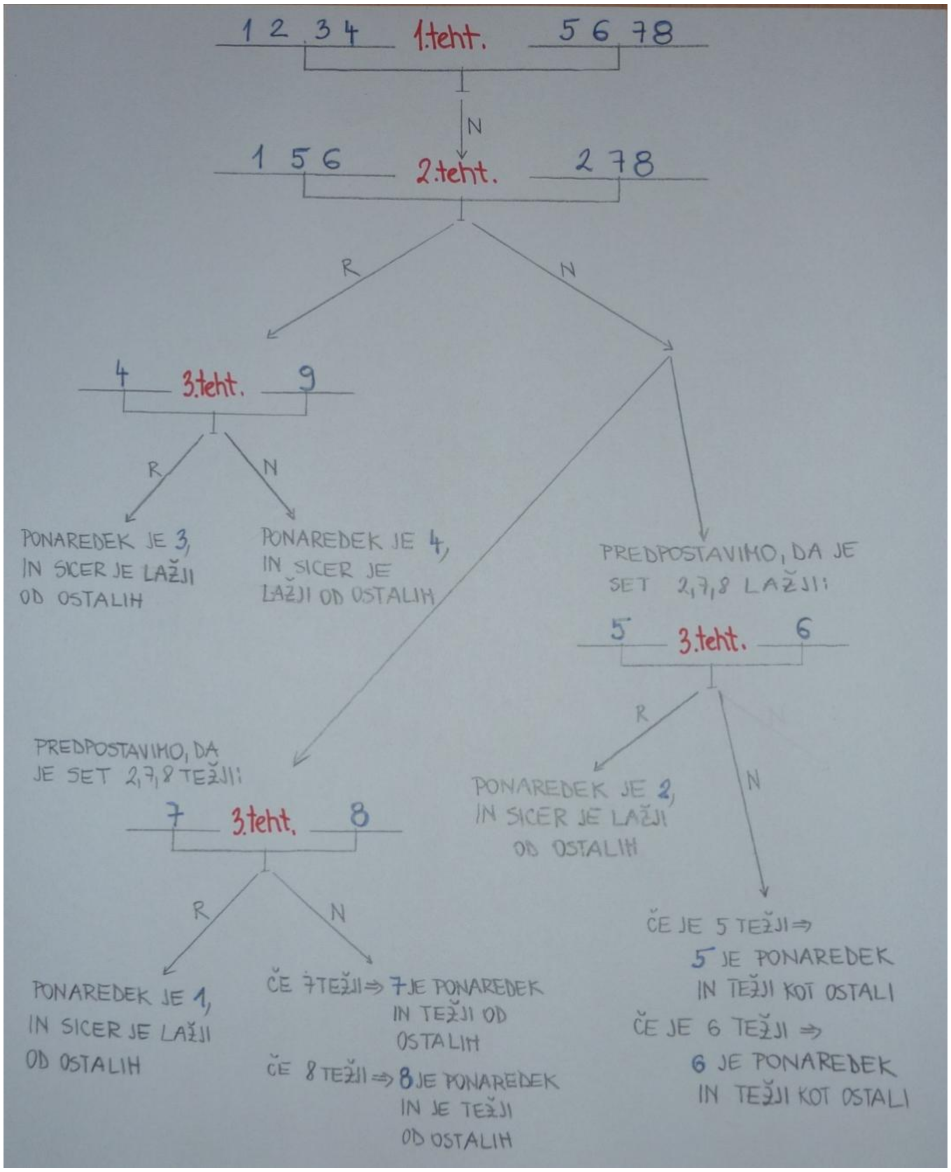
Za boljšo preglednost in lažje razumevanje sta spodaj sliki, ki nam pregledno kažeta potek tehtanja.

Prvo tehtanje: na levo stran tehtnice damo kovance 1, 2, 3 in 4, na desno pa 5, 6, 7 in 8. Poglejmo si najprej samo možnost, ko dobimo ravnovesje. (Kasneje si bomo pogledali še možnosti neravnovesja.) Sklepamo lahko, da je ponaredek gotovo v setu 9, 10, 11, 12. Sledi drugo tehtanje: 9, 10, 11 proti 1, 2, 3 (za te tri kovance smo s prejšnjim tehtanjem ugotovili, da so pravi). Če dobimo ravnovesje, vemo, da je ponaredek kovanec 12. Vendar hočemo izvedeti še ali je težji ali lažji od ostalih kovancev. Potrebno je torej še tretje tehtanje: kovanec 12 proti pravemu kovancu 1. Tako smo v treh tehtanjih dobili ponaredek (12) in ugotovili ali je lažji ali težji od pravih kovancev. Če pa tehtnica ni v ravnovesju, imamo dve možnosti: lahko so kovanci 9, 10 in 11 lažji od 1, 2, 3 ali pa težji. Če so težji, je ponaredek težji in seveda je med njimi. Če pa so lažji, je ponaredek med njimi lažji od ostalih kovancev. Pa poglejmo najprej možnost, ko so kovanci 9, 10 in 11 težji. Sledi 3. tehtanje: kovanec 9 proti kovancu 10. Pri uravnoteženi tehtnici lahko ugotovimo, da je ponaredek kovanec 11, in sicer je težji od ostalih kovancev. Če pa ni ravnovesja, imamo lahko dva možna izida: če je kovanec 9 težji, je ta ponaredek (težji od ostalih), če pa je težji 10, je ponaredek ta kovanec. Ogledati si moramo še možnost, ko so kovanci 9, 10 in 11 pri drugem tehtanju lažji od 1, 2 in 3. Ponaredek je gotovo lažji od ostalih kovancev. Če dobimo ravnovesje, je z enakim sklepanjem kot prej ponaredek gotovo kovanec 11. Če s tretjim tehtanjem ugotovimo, da je kovanec 9 lažji od kovanca 10, je ponaredek 9, sicer pa 10. Sedaj smo prišli s tremi tehtanji do rešitve, vendar smo si ogledali šele primer, ko po prvem tehtanju dobimo ravnotežje.



Kaj pa če imamo po prvem tehtanju neravnotežje?

Predpostavimo, da smo pri prvem tehtanju ugotovili, da je desna stran (5,6,7,8) težja (če je lažja, je postopek podoben). Sledi drugo tehtanje: 1,5 in 6 proti 2, 7 in 8. Če pri drugem tehtanju dobimo ravnovesje, vemo da je ponaredek kovanec 3 ali 4. Sledi 3. tehtanje: kovanec 4 proti 9. V primeru ravnovesja, je ponaredek kovanec 3 (lažji od ostalih), v primeru neravnovesja pa je ponaredek 4 (lažji). Če pri drugem tehtanju nimamo ravnovesja, lahko predpostavimo, da je desna stran (2, 7, 8) težja. Imamo dve možnosti: če je ponaredek 7 ali 8, je ta gotovo težji od ostalih, če pa je ponarejen kovanec 1, je ta lažji od ostalih. Sledi zadnje (tretje) tehtanje. Na levo stran damo kovanec 7, na desno pa 8. V primeru ravnovesja, je ponaredek 1. Če pa dobimo neravnovesje, imamo zopet dve možnosti: če je težji 7, je ta ponaredek, če pa je težji 8, je ponaredek ta kovanec. Obravnavamo lahko še možnost, da je skupina kovancev 2, 7, 8 lažja (v neravnovesju pri drugem tehtanju). Sledi tretje tehtanje: 5 proti 6. Če dobimo ravnovesje, lahko sklepamo, da je ponaredek kovanec 2, in sicer je ta lažji od ostalih. Če pa je kovanec 5 težji, je ponaredek in težji od pravih kovancev. Kovanec 6 je ponaredek, če je težji od kovanca 5.



Naloga je veliko lažja, če si pomagamo z risanjem.

6. LITERATURA

1. Pierce, J. R. 1980 An introduction to information theory: symbols, signals and noise, 2nd edn. New York: Dover
2. http://forum.feri.uni-mb.si/Default.aspx?g=posts&t=3652
3. http://mathforum.org/dr.math/
4. http://www.cut-the-knot.org/blue/6Balls2Weighings.shtml