UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Danica Petric

Skupni rojstni dnevi in normalna porazdelitev

Seminar II

Ljubljana, 2010

Kazalo

1. Uvod

Predpostavimo, da imamo sobo v kateri je 23 oseb. Kolikšna je verjetnost, da se v tej sobi nahajata dve osebi, ki praznujeta rojstni dan na isti dan v letu? Problem lahko povemo tudi nekoliko drugače. Imamo 23 imen, ki jih naključno porazdelimo med 365 škatel. Vsaka škatla predstavlja en dan v letu. Kolikšna je verjetnost, da bo med temi škatlami takšna škatla, ki bo vsebovala več kot eno ime? Jasno je, da 23 imen porazdeljenih med 365 škatlami predstavlja slabo distribucijo, saj bo velika količina škatel praznih.

Kaj pa, če imamo v sobi 100 oseb? Sedaj lahko rečemo, da imamo vsaj razumno možnost zadetka. Koliko zadetkov lahko dobimo? Nekaj že; če je sreča, vsaj tri.

Resnica, ki se zdi na prvi pogled za lase privlečena, je naslednja:

- Če imamo 23 oseb je verjetnost 50:50, da nek par teh oseb deli rojstni dan.[1]
- Če imamo 100 oseb pričakujemo od 13 do 14 parov. Zato je v skupini stotih oseb približno četrtina takih, ki delijo rojstni dan z nekom drugim iz skupine.

Cilja tega poglavja sta:

a) Premisliti, zakaj je resnica takšna, kot je.
b) Z uporabo cilja a) spoznati pomembno načelo v verjetnostni teoriji, *centralni limitni izrek*.

2. Kolikšna je verjetnost zadetka?

Naključno porazdelimo 23 imen med 365 škatel. Kako izračunamo verjetnost, da bo med škatlami taka škatla, ki bo vsebovala več kot eno ime? Da se izognemo uporabi števil (npr. 365), lahko rečemo, da imamo k oseb in n škatel (ali datumov). Najenostavnejši način za izračun verjetnosti zadetka je ta, da izračunamo verjetnost ne zadetka in nato odštejemo to verjetnost od 1. (Če je verjetnost ne zadetka 40%, je potem verjetnost enega ali več zadetkov 60%.)

Kolikšna je verjetnost, da ni zadetka? Predstavljajmo si metanje imen v škatle, eno za drugo. Vržemo prvo ime "Tina" v katerokoli škatlo, saj so še vse škatle prazne. Sedaj vzamemo drugo ime "Matija". Če se želimo izogniti zadetku, potem Matija ne smemo dati v isto škatlo kot Tino. Tako imamo za Matija na izbiro $n-1$ škatel od vseh n-tih. Torej je verjetnost, da se izognemo zadetku, enaka

$$\frac{n-1}{n}=1-\frac{1}{n}.$$

Kaj pa Mojca? Kakšna je njena verjetnost, da se izogne zadetku? Ona ima na izbiro $n-2$ škatel od vseh n-tih, če imena shranjujemo ločeno. Torej je verjetnost, da se Mojca izogne zadetkom, enaka

$$1-\frac{1}{n}.$$

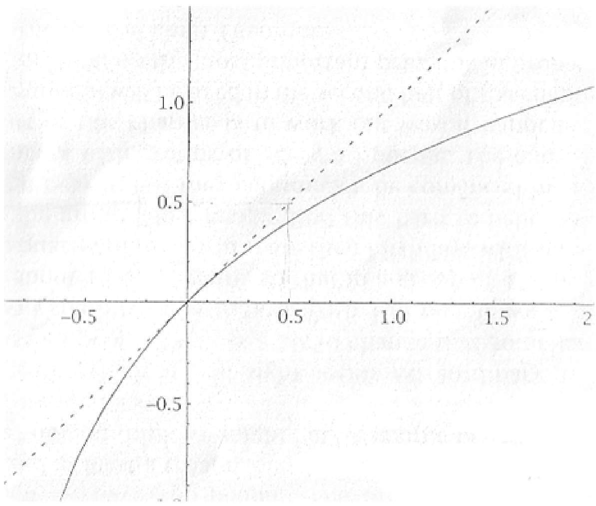
[1] Ta trditev je v prilogi preverjena z dejanskim preizkusom na skupini 40 oseb.

Če nadaljujemo z razmišljanjem v tej smeri, ugotovimo, da je verjetnost ne zadetka potem, ko smo vrgli vseh k imen v n škatel, enaka

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right). \tag{1}$$

Za velik n, npr. 365 in k samo 23, je vsak člen v tem produktu dokaj blizu 1. Najmanjši je zadnji člen, ki je približno enak 0,9497. Začetni členi so mnogo bližje 1 (npr. prvi člen je približno enak 0,9972, drugi člen 0,9945, itd.). Zato dobimo občutek, da je tudi produkt dokaj blizu 1. Zapomniti pa si moramo, da govorimo o ne zadetku. Če je verjetnost le-tega blizu 1, potem je verjetnost zadetka blizu 0.

Za posebne vrednosti k in n lahko uporabimo kalkulator oziroma računalnik. Z njuno pomočjo lahko poiščemo natančno numerično vrednost za produkt (1) . Mi navedenih pripomočkov ne bomo uporabili, temveč bomo preizkusili oceniti velikost dobljenega produkta.



Slika 1: Dotik tangentne črte s logaritmom.

Dajmo produkt označiti s P:

$$\mathrm{P}=\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right).$$

Poizkusimo oceniti velikost tega produkta. Iz izkušenj vemo, da je ocenjevati vsoto veliko lažje kot ocenjevati produkt. Produkt P lahko spremenimo v vsoto s pomočjo logaritma:

$$\ln\mathrm{P}=\ln\left(1-\frac{1}{n}\right)+\ln\left(1-\frac{2}{n}\right)+\cdots+\ln\left(1-\frac{k-1}{n}\right).$$

Sedaj, ko smo spremenili produkt v vsoto, poizkusimo oceniti vsak člen posebej. Ker je členov veliko in so si ti med seboj zelo podobni, je nesmiselno delati oceno za vsak člen posebej. Naredimo raje oceno za poljubni člen

$$\ln\left(1-\frac{j}{n}\right).$$

Za lažje določanje ocene si pomagajmo s sliko 1. Ta prikazuje graf funkcije $y = \ln(1+x)$ in tangento $y = x$ v točki $(0,0)$. Tangenta zagotavlja dober približek krivulji, če je x blizu 0. Za majhne vrednosti x je število $\ln(1+x)$ približno enako x. To pomeni, da toliko časa, dokler je j znatno manjši od n, lahko uporabimo tangentni približek z $x = -j/n$ in dobimo

$$\ln\left(1-\frac{j}{n}\right) \approx -\frac{j}{n}.$$

Vrnimo se nazaj na $\ln \mathrm{P}$:

$$\ln \mathrm{P} = \ln\left(1-\frac{1}{n}\right) + \ln\left(1-\frac{2}{n}\right) + \cdots + \ln\left(1-\frac{k-1}{n}\right).$$

Toliko časa, dokler je k znatno manjši kot n, pričakujemo, da lahko aproksimiramo vsak logaritem z uporabo tangentnega približka. Torej

$$\ln \mathrm{P} \approx -\frac{1}{n} - \frac{2}{n} - \cdots - \frac{k-1}{n}.$$

Na desni strani zapisa lahko izpostavimo $-1/n$. Ostane nam vsota

$$1 + 2 + 3 + 4 + \cdots + (k-1).$$

To vsoto poznamo kot vsoto aritmetičnega zaporedja in jo zato lahko nadomestimo s $k(k-1)/2$. Torej pričakujemo, da je

$$\ln \mathrm{P} \approx -\frac{k(k-1)}{2n}. \qquad (2)$$

Pri takšni aproksimaciji moramo biti nekoliko previdni, saj smo ocenili vsak člen vsote $\ln(1-j/n)$ in nato združili še ocene skupaj. Tako smo lahko naredili kar veliko napak.

Pomožen problem:
Pokažimo, da velja

$$\ln \mathrm{P} \le -\frac{k(k-1)}{2n}.$$

Rešitev pomožnega problema:
Verjetnost P je enaka

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$$

in njegov logaritem je

$$\ln \mathrm{P} = \ln\left(1-\frac{1}{n}\right) + \ln\left(1-\frac{2}{n}\right) + \cdots + \ln\left(1-\frac{k-1}{n}\right).$$

Graf na sliki 1 prikazuje, da ni samo $\ln(1+x) \approx x$, če je x majhno število, ampak je tudi $\ln(1+x) \le x$ za katerokoli vrednost x. Torej je

$$\ln \mathrm{P} \le -\frac{1}{n} - \frac{2}{n} - \cdots - \frac{k-1}{n} = -\frac{k(k-1)}{2n}. \qquad \blacksquare$$

Vzemimo za trenutek, da je ocena (2) dobra. Tedaj z jemanjem eksponentov dobimo oceno za verjetnost samega P:

$$\ln \mathrm{P} \approx e^{-k(k-1)/(2n)}. \tag{3}$$

Kaj dobimo, ko je $k = 23$ in $n = 365$? V tem primeru je

$$\frac{k(k-1)}{2n} = \frac{23 \cdot 22}{730} \approx 0{,}69.$$

Če to uporabimo v naši oceni je verjetnost ne zadetka za 23 oseb približno $e^{-0{,}69}$. Vemo pa, da je $\ln 2 \approx 0{,}69$. Tako je verjetnost ne zadetka $e^{-\ln 2} = 1/2$. Potem je verjetnost za zadetek tudi približno $1/2$. Če imamo v sobi 23 oseb je tu grobih 50:50 možnosti, da imamo ujemajoč rojstni dan med temi osebami.

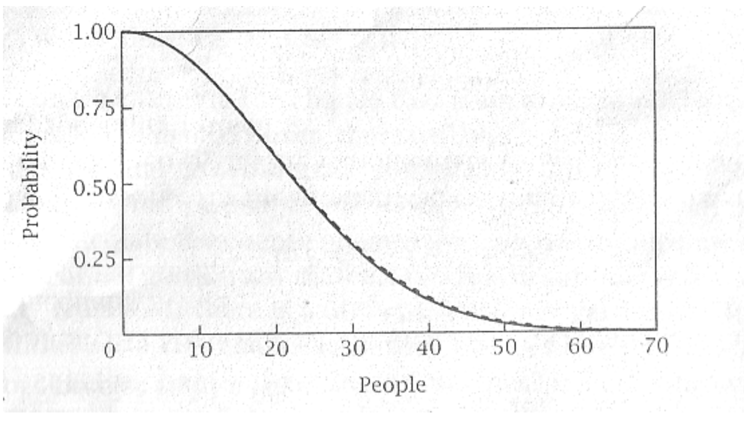
Kaj pa, če imamo $k = 100$ in $n = 365$? Opazimo, da k/n ni več tako majhen, kar pomeni, da naša ocena mogoče ne bo tako dobra. Po drugi strani pa lahko s pomočjo pomožnega problema zaključimo, da verjetnost ne zadetkov ne presega $e^{-k(k-1)/(2n)}$. Za $k = 100$ in $n = 365$ je

$$\frac{k(k-1)}{2n} = \frac{100 \cdot 99}{730} \approx 13.56.$$

To da oceno

$$e^{-13{,}56} \approx 1{,}3 \cdot 10^{-6}.$$

Verjetnost za ne zadetek je skoraj ena proti miljon. To nam pove, da je verjetnost zadetka zelo velika. Dejanska verjetnost, da ima 100 naključno izbranih oseb na isti dan rojstni daj, je približno $3 \cdot 10^{-7}$, kar je četrtina verjetnosti dane iz naše ocene. Naša ocena ni tako dobra za $k = 100$, narobe je za faktor 4. Ker pa je razlika med dejansko in ocenjeno verjetnostjo tako majhna, nas ta napaka ne skrbi.



Slika 2: Dejanska in ocenjena verjetnost.

Če pogledamo graf (slika 2), ki prikazuje verjetnost ne zadetka za vsako možno število oseb in oceno $e^{-k(k-1)/(2n)}$, opazimo, da se obeh narisanih krivulj skoraj ne da ločiti. Dejanska verjetnost, uporabljena na grafu, je bila izračunana z računalnikom.

Na grafu vidimo, da bo s samo majhnim številom oseb verjetnost, da bo obstajal nek par rojen na isti dan, dobra. To pa se ne ujema z našo intuicijo. Zakaj nas je intuicija pustila na cedilu?

Če pogledamo nazaj na izraz za verjetnost izognitve vseh zadetkov,

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$$

je res, da je vsak faktor kar blizu 1. Njihov produkt se konča majhen. Ko smo sešteli eksponente

$$\frac{1}{n}+\frac{2}{n}+\cdots+\frac{k-1}{n},$$

smo dobili

$$\frac{k(k-1)}{2n}.$$

Bistvo je, da smo dobili na vrhu nekaj podobnega k^2 in če naj bilo to število razmeroma veliko v primerjavi z n, bi morali za k jemati števila okoli drugega korena n. Torej je veliko manjši od n sam k. To bo postalo bolj jasno v naslednjem razdelku, v katerem bomo izračunali možno število zadetkov.

3. Koliko ujemanj dobimo?

Poizkusimo odgovoriti na drugo vprašanje iz predstavitve. V sobi imamo 100 oseb; koliko število zadetkov pričakujemo? Na voljo imamo več načinov za izračun pričakovanega števila zadetkov. Pokazali pa bomo samo enega.

Za razumevanje vprašanja začnimo s pomožnim primerom. Če vržemo pošten kovanec 40 krat, lahko izračunamo verjetnost vsake možne cifre in glave z uporabo binomskega koeficienta. Verjetnost k glav je

$$\binom{40}{k}\frac{1}{2^{40}},$$

za vsak k med 0 in 40. (V razdelku 5 je to bolj podrobno razloženo.) Za izračun pričakovanega števila glav lahko dodamo vse te verjetnost pomnožene z ustreznimi utežmi. Toda pričakovano število glav ni težko izračunati, saj je to očitno 20. V povprečju pričakujemo 20 glav v 40 metih, saj vsak met (v povprečju) prispeva polovico glave. Pričakovano številko glav je potem enako verjetnosti dobitka glave ob vsakem metu, pomnoženo s številom poizkusov.

Vrnimo se sedaj nazaj k rojstnim dnevom in poizkusimo igrati isto igro. Kaj pomeni met v tem primeru? Met pomeni samo par oseb, katere bi lahko ali pa tudi ne delile rojstne dneve. Zato moramo izračunati verjetnost, da določen par deli rojstni dan in število možnih parov.

a) Kakšna je verjetnost, da določen par oseb deli rojstni dan? To je $1/n$. Razlog zato je, kadarkoli je moj rojstni dan, imaš ti na voljo le $1/n$ možnosti, da izbereš ta isti dan.
b) Koliko možnih parov oseb je tam? Število parov oseb, katere je možno izbrati med k osebami, je

$$\binom{k}{2}=\frac{k(k-1)}{2}.$$

Torej tu je $k(k-1)/2$ možnih zadetkov in vsak od teh zadetkov ima $1/n$ verjetnosti, da se zgodi. Pričakovano število zadetkov je tako

$$\frac{k(k-1)}{2} \cdot \frac{1}{n} = \frac{k(k-1)}{2n}.$$

Uporabimo ta izračun za primer, ko imamo 100 oseb ($k = 100$). Število parov, ki jih je možno izbrati med stotimi osebami je $(100 \cdot 99)/2 = 4950$. Vsak možen par ima $1/365$ verjetnosti, da se pojavi. Pričakovano število, ki pride je $4950/365$ in to je približno 13,5 parov. Torej pričakujemo 13 ali 14 parov[2].

V naslednjem razdelku bomo izračunali pričakovano število oseb, ki so vpletene v pare.

4. Koliko oseb si deli rojstni dan?

Če je k imen naključno porazdeljenih med n škatel, potem pričakujemo, da bo med njimi v povprečju $k(k-1)/(2n)$ takih, ki si delijo škatle. Cilj tega razdelka bo ugotoviti, koliko oseb v skupini k oseb bo delilo rojstni dan z nekom drugim iz te skupine?

Navidezno logičen odgovor je, da si bo v skupini delilo rojstni dan dvakrat več oseb kot je parov. Saj sta v vsak par vpleteni dve osebi. To pomeni, da pričakujemo

$$\frac{k(k-1)}{n} \tag{4}$$

oseb, ki so "delivci rojstnega dneva". Za 100 oseb je pričakovano število $9900/365 \approx 27{,}1$.

Ali bi lahko to izračunali tudi na kakšen drugačen način? V skupini imamo ponovno k oseb. Če bi lahko izračunali verjetnost p, da vsak deli svoj rojstni dan še s kom, potem lahko izračunamo pričakovano število delivcev rojstnega dne kot kp. Kakšna je verjetnost, da določena oseba, Tina, deli svoj rojstni dan? Kot že prej je lažje izračunati verjetnost, da Tina ne deli rojstnega dne in nato to odšteti od števila 1. Da Tina ne deli svojega rojstnega dneva, je potrebno, da vsa druga $k-1$ imena vržemo v preostale škatle in v ne škatlo, kjer se nahaja njeno ime. Preostalih škatel imamo $n-1$. Tako je za vsa različna $k-1$ imena verjetnost

$$\frac{n-1}{n} = 1 - \frac{1}{n},$$

da končajo v drugi škatli, kot je Tina. Verjetnost, da vsa imena končajo v različnih škatlah, je potem

$$\left(1-\frac{1}{n}\right)^{k-1}.$$

Takšna je verjetnost, da Tina ne deli škatle oziroma da praznuje sama. Verjetnost, da deli ta dan z drugimi, je

$$p = 1 - \left(1-\frac{1}{n}\right)^{k-1}.$$

---

[2] S stotimi osebami je tu enaka možnost, da dobiš deset parov, ki si delijo rojstni dan in enega trojčka, ki prispeva tri delitve.

Ker pa je v skupini k oseb, vsaka z verjetnostjo p deljenja rojstnega dne, je pričakovano število oseb, ki si delijo rojstni dan, enako

$$kp = k\left(1-\left(1-\frac{1}{n}\right)^{k-1}\right).$$

Za primer 100 oseb imamo približno 23,8 oseb, ki delijo rojstni dan. Videli pa smo, da je pričakovano število parov za 100 oseb 13,56 in na podlagi tega tudi ugotovili, da je število delivcev rojstnega dne dvakrat večje, se pravi grobih 27,1. Kje smo naredili napako?
Napako smo naredili pri prvem razmisleku, kjer smo privzeli, da sta v vsak par vpleteni le dve osebi. Ta privzetek je dober za zmerno majhen k (npr. $k = 23$). Saj je za tak k verjetnost, da je vpletenih več oseb, zamerljivo majhna, zato si ta privzetek potem lahko dovolimo. Za $k = 100$ pa ta privzetek ni več dober. Vzrok temu je naslednji: če imajo tri osebe isti rojstni dan, potem predstavljajo tri različne pare oseb, ki si delijo rojstni dan. Na primer, če so Ana, Bob in Miha rojeni prvega oktobra, so AB, AM in BM šteti kot pari, ki delijo rojstni dan. Za 100 oseb je velika verjetnost, da bodo tri osebe imele isti rojstni dan. Torej le tri osebe bodo prispevale tri pare, namesto morda pričakovanih šestih oseb. V drugem razmisleku te napake ni.

5. Krivulja v obliki zvona

Predstavljamo si, da vržemo pošten kovanec n krat v zrak. Izid metanja bo zaporedje n glav in cifer. Za primer, ko je $n = 3$, je možen izid katerokoli od naslednjih osmih zaporedij:

C C C,	G C C,
C C G,	G C G,
C G G,	G G C,
C G C,	G G G.

Vsako od teh zaporedij je enako verjetno. Tako ima vsako zaporedje verjetnost $1/2^3$ (ali v splošnem $1/2^n$). Skupna verjetnost vseh verjetnostnih izidov je vedno 1.

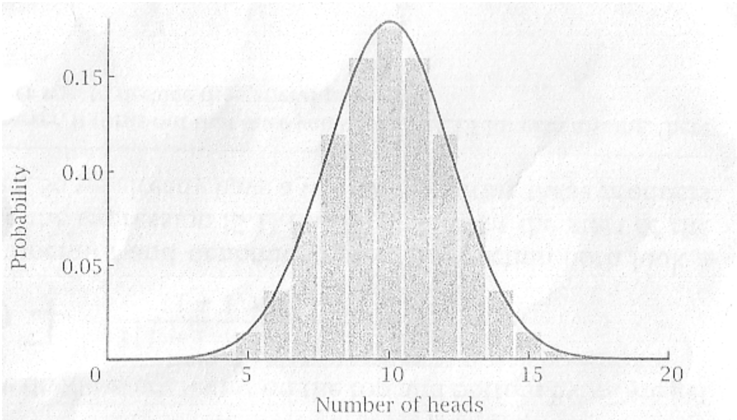
Sedaj bomo šteli celotno število glav (oziroma grbov), ki se pojavijo v n metih. To bo vsako celo število j med 0 in n. V vsakem zaporedju predstavlja j glav število načinov izbire j lokacij za glave znotraj zaporedja n lokacij. To je binomski koeficient

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}.$$

Verjetnost, da dobiš j glav v n metih, je

$$\binom{n}{j}\frac{1}{2^n}.$$

Če to narišmo s grafikonom, dobimo nekaj podobnega, kot je narisano na sliki 3. Ta slika prikazuje rezultate za $n = 20$ metov. Smiselno je, da so verjetnosti proti sredini grafikona ogromne, saj že pričakujemo, da bo število glav polovica števila metov. Ko se odmikamo od sredine, pa se verjetnosti hitro zmanjšujejo.



Slika 3: 20 metov poštenega kovanca.

Stolpičast grafikon je pokrit s krivuljo, ki da dober približek za višino stolpcev. Ta krivulja ima enačbo

$$y = \sqrt{\frac{2}{n\pi}}\, e^{-\frac{2}{n}\left(x-\frac{n}{2}\right)^2}. \tag{5}$$

To je primer normalne oziroma zvončaste oblike krivulje. Lahko tudi rečemo, da je ta krivulja kopija standardne normalne krivulje, katere enačba se glasi

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Dejstvo, da lahko verjetnost metanja kovanca v zrak aproksimiramo s to krivuljo, ni samo slučaj. Eden od pomembnih načel v matematiki je *centralni limitni izrek*, ki pravi, da kadarkoli seštejemo veliko število majhnih neodvisnih učinkov, je rezultat približno podan z normalno krivuljo. Ta izrek se uporablja predvsem v statistični analizi.

V nadaljevanju ne bomo naredili nobenega poizkusa za demonstracijo *centralno limitnega izreka*. Temveč bomo pokazali, kako nastane v posebnem primeru metanja poštenega kovanca v zrak. Privzemimo, da je število metov n celoštevilsko in enako $2m$. Verjetnost, da vržemo v $2m$ metih j glav, je enaka

$$\binom{2m}{j}\frac{1}{2^{2m}}.$$

Cilj tega razdelka bo preučiti, kako se spreminjajo verjetnosti, ko spreminjamo število j. To naredimo s primerjanjem verjetnosti različnih j-jev s verjetnostjo na sredini,

$$\binom{2m}{m}\frac{1}{2^{2m}}.$$

Da se osredotočimo na naš namen, zapišimo *j* kot $m+k$, pri čemer je *k* premik *j*-ja iz sredinske vrednosti *m*. Razmerje med verjetnostjo pridobitve $m+k$ glav in verjetnostjo pridobitve *m* glav je

$$\binom{2m}{m+k}\binom{2m}{m}^{-1} = \frac{(2m)!}{(m+k)!(m-k)!}\left(\frac{(2m)!}{m!m!}\right)^{-1}$$
$$= \frac{m!m!}{(m+k)!(m-k)!}$$
$$= \frac{m(m-1)\cdots(m-k+1)}{(m+k)(m+k+1)\cdots(m+1)}.$$

Če delimo vsak faktor v števcu in imenovalcu z *m*, potem dobimo

$$\frac{\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\cdots\left(1-\frac{k-1}{m}\right)}{\left(1+\frac{1}{m}\right)\left(1+\frac{2}{m}\right)\cdots\left(1+\frac{k}{m}\right)}.$$

Števec in imenovalec tega zapisa sta podobna zapisu (1), ki smo ga izračunali na začetku poglavja. Torej že poznamo način za ocenjevanje teh produktov, če je *k* znatno manjši od *m*. Potem je števec enak verjetnosti *P*, ki se pojavi v (1) (le da *n* zamenjamo z *m*). Torej je števec približno enak

$$e^{-k(k-1)/(2m)},$$

kot smo videli v zapisu (3). Za imenovalec lahko naredimo približek tudi z jemanjem logaritmov in z uporabo približka $\ln(1+x)\approx x$, potem je

$$\left(1+\frac{1}{m}\right)\left(1+\frac{2}{m}\right)\cdots\left(1+\frac{k}{m}\right) = e^{\frac{k(k+1)}{2m}},$$

če je *k* znatno manjši od *m*.

Sedaj lahko združimo naši oceni skupaj in dobimo

$$\binom{2m}{m+k}\binom{2m}{m}^{-1} \approx \frac{e^{-k(k-1)/(2m)}}{e^{k(k+1)/(2m)}}$$
$$= e^{-\frac{k^2}{m}}.$$

Tu opazimo obrise formule za normalno krivuljo. Če se vrnemo nazaj in spremenimo *k* v $j-m$ in *m* v $n/2$, dobimo

$$\binom{n}{j}\binom{n}{n/2}^{-1} \approx e^{-\frac{2}{n}\left(j-\frac{n}{2}\right)^2}.$$

Tako je verjetnost dobitka vseh *j* glav v *n* metih

$$\binom{n}{j}\frac{1}{2^n} \approx \binom{n}{n/2}\frac{1}{2^n}e^{-\frac{2}{n}\left(j-\frac{n}{2}\right)^2}.$$

Opazimo, da to zgleda zelo podobno kot enačba normalne krivulje (5). S samo to razliko, da je začetni faktor $\sqrt{2/(n\pi)}$ zamenjan z vrednostjo dobitka $n/2$ glav

$$\binom{n}{n/2}\frac{1}{2^n}.$$

Dokazali smo, da lahko približamo stolpičast grafikon s krivuljo, ki ima isto "obliko" kot normalna krivulja, ki je bila razrezana vertikalno. Najboljši približek je na sredini grafikona, kar je tudi vidno na sliki 3.

Seveda, ko smo našli pravilno obliko krivulje, ni presenečenje, da jo lahko z raztegom naredimo ustreznejšo, torej da dobimo pravo vrednost na sredini. Skrivnost je, zakaj je sredinska velikost tako blizu številu $\sqrt{2/(n\pi)}$. Navedimo razloge, zakaj je boljše imeti formulo s $\sqrt{2/(n\pi)}$ kot pa z

$$\binom{n}{n/2}\frac{1}{2^n}.$$

En razlog je, da je $\sqrt{2/(n\pi)}$ lažje izračunati kot pa verjetnost

$$\binom{n}{n/2}\frac{1}{2^n}.$$

Toda tu je še en pomembnejši razlog, ki pa tudi razjasni videz števila $\sqrt{2/(n\pi)}$.

Stolpičast grafikon je bil izdelan zato, da je višina vsakega stolpička predstavljala verjetnost dobitka števila grbov v n metih kovanca. (Vsaka od verjetnosti števila glav je predstavljena s stolpičem.) To pomeni, da je vsota višin stolpičev enaka 1, celotni verjetnosti. Predpostavimo, da je vsak stolpič širine 1. Potem lahko tudi rečemo, da je vsota ploščin stolpičev enaka 1. Ko predstavimo distribucijo v tej smeri, vedno dobimo celotno območje 1. Sedaj se želimo prepričati, da je krivulja, ki smo jo uporabili za približevanje verjetnostne distribucije, sama "verjetnostna krivulja" – taka, da je ploščina celotnega območja pod njo enaka 1. Kot bomo videli, je število $\sqrt{2/(n\pi)}$ dobro izbrano z namenom, da je ploščina območja pod krivuljo enaka 1.

Situacija je sedaj takšna. Ko smo enkrat našli pravilno obliko krivulje, smo jo raztegnili vertikalno tako, da je ploščina območja pod krivuljo enaka 1 namesto, da bi jo izmerili s sredinsko vrednostjo. Obe metodi nam dasta dober približek. Toda metoda, ki naredi ploščino celotnega območja enako 1, najbolje ustreza teoriji verjetnosti. Vsaka metoda, ki da pravilno ploščino, da tudi dober približek za sredinsko vrednost. Če lahko preverimo, da je $\sqrt{2/(n\pi)}$ pravilni faktor, ki da ploščino enako 1, bomo lahko sklepali, da je ta številka blizu faktorja

$$\binom{n}{n/2}\frac{1}{2^n}.$$

Zadnji del tega poglavja je namenjen dokazovanju, da je ploščina območja pod krivuljo

$$y=\sqrt{\frac{2}{n\pi}}\,e^{-\frac{2}{n}\left(x-\frac{n}{2}\right)^2},$$

res enaka 1. S tem bomo zaključili dokaz, da ta normalna krivulja aproksimira verjetnosti metanja kovanca.

6. Območje pod normalno krivuljo

Krivulja

$$y=\sqrt{\frac{2}{n\pi}}\,e^{-\frac{2}{n}\left(x-\frac{n}{2}\right)^2},$$

je bila prikazana na sliki 3, za primer n = 20. Ploščino območja pod to krivuljo lahko izračunamo z integralom

$$\int_{-\infty}^{\infty}\sqrt{\frac{2}{n\pi}}\,e^{-\frac{2}{n}\left(x-\frac{n}{2}\right)^2}dx\,.$$

Ta ploščina se ne spremeni, če premaknemo funkcijo na levo za razdaljo $n/2$ oziroma, če zamenjamo $x-n/2$ z x. Pokazali bi radi, da je

$$\int_{-\infty}^{\infty}\sqrt{\frac{2}{n\pi}}\,e^{-\frac{2}{n}\left(x-\frac{n}{2}\right)^2}dx=1\,.$$

Za integral na levi se zdi, da je odvisen od *n,* medtem ko se za število 1 to ne zdi. Zato je smiselno, da poizkusimo integral spremeniti tako, da bo število *n* izginilo. To lahko dosežemo tako, da zamenjamo *x* z $t\sqrt{n/2}$. Integral se potem spremeni v

$$\int_{-\infty}^{\infty}\sqrt{\frac{1}{\pi}}\,e^{-t^2}dt\,.$$

Torej pokazati želimo, da velja

$$\int_{-\infty}^{\infty}\sqrt{\frac{1}{\pi}}\,e^{-t^2}dt=1$$

oziroma

$$\int_{-\infty}^{\infty}e^{-t^2}dt=\sqrt{\pi}\,. \tag{6}$$

Za lažje nadaljevanje vpeljimo oznako I :

$$I=\int_{-\infty}^{\infty}e^{-t^2}dt\,.$$

<u>*Zanimivost:*</u>

Neka stara zgodba pravi, da je nekoč matematik in fizik Lord Kelvin rekel: "*Matematik je nekdo, kateremu je očitno, da je*

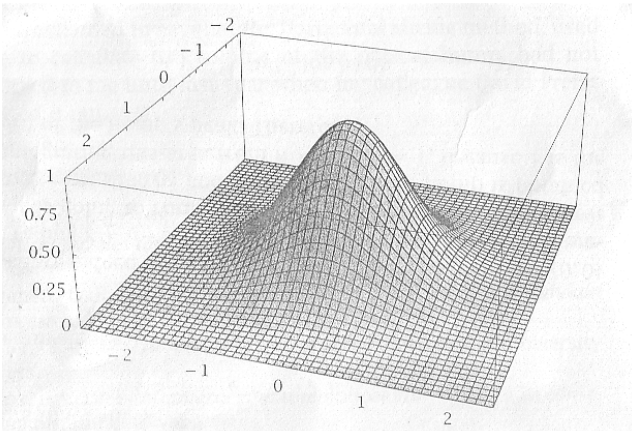
$$\int_{-\infty}^{\infty}e^{-t^2}dt=\sqrt{\pi}\text{ "}.$$

Metoda, katero bomo uporabili za dokaz formule (6), je zelo dober trik. Namesto proučevanja funkcije $t \to e^{-t^2}$ bomo preučili funkcijo, ki je v točki (*s*,*t*) definirana s predpisom

$$g(s,t) = e^{-(s^2+t^2)}.$$

V vsaki točki ima ta funkcija realno vrednost. Na primer, v točki (1,2) je vrednost funkcije $e^{-(1+4)} = e^{-5}$. Lahko narišemo graf *g*, kjer vrednost *g* predstavlja višino nad ustreznimi točkami. Graf narišemo kot površino in ne kot krivuljo. Narisan graf je prikazan na sliki 4. Mogoče je najenostavnejša pot za razmišljanje ta, da funkcijo *g* opišemo kot kraj, v katerem je *g*(*s,t*) nadmorska višina za zemljepisno širino *t* in dolžino *s*.

Funkcija *g* je očitno povezana na nek način z normalno krivuljo, saj zgleda $e^{-(s^2+t^2)}$ kot e^{-t^2}. Bistven korak pri izračunu površine pod normalno krivuljo bo ta, da izračunamo volumen območja pod površino grbe (slika 4).



Slika 4: Normalna grba.

Če najdemo tudi povezavo med volumnom grbe in ploščino, ki jo želimo izračunati, bomo lahko izrazili ploščino z znanim volumnom. Začnimo z iskanjem potrebnih povezav in nato še z izračunom samega volumna.

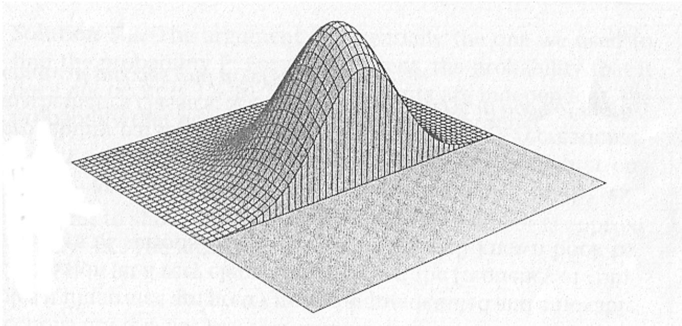
Višina našega površja *g*(*s,t*) v točki (*s,t*) je odvisna od koordinat *s* in *t*, na poseben način. Zaradi lastnosti eksponenta lahko višino zapišemo kot produkt

$$e^{-(s^2+t^2)} = e^{-s^2}e^{-t^2},$$

v katerem je vsak faktor odvisen samo od ene koordinate. Privzemimo za trenutek, da smo določili posebno vrednost *s*, recimo $s = 1$. Poglejmo, kaj se dogaja z *g*, ko se *t* spreminja. Vrednost g(1,t) je

$$e^{-1}e^{-t^2}.$$

Ker je funkcija *t*-ja, je potem enaka normalni krivulji, ki je pomnožena s številom e^{-1}.



Slika 5: Vertikalen prerez grbe.

Na sliki 5 lahko vidimo, da je prerez površine krivulja, ki zgleda kot normalna krivulja. Ploščina prereza je kot ploščina pod normalno krivuljo, le da je ta pomnožena z e^{-1}. Torej je ploščina enaka integralu

$$\int_{-\infty}^{\infty} e^{-1} e^{-t^2} dt = e^{-1} I \; .$$

Podobno je tudi za preostale vrednosti *s*. Zato lahko za splošen *s* rečemo, da je

$$\int_{-\infty}^{\infty} e^{-s^2} e^{-t^2} dt = e^{-s^2} I \; .$$

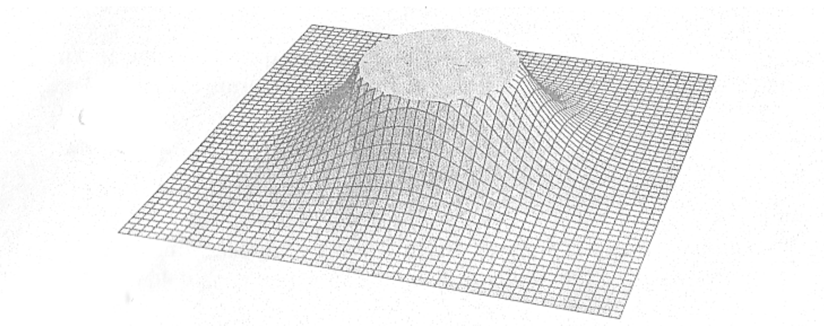
Sedaj lahko izračunamo volumen grbe z združitvijo vseh *s*-jev. Torej je volumen grbe enak

$$V = \int_{-\infty}^{\infty} e^{-s^2} I \, ds \quad .$$

Ta novi integral zgleda natanko tako kot integral normalne krivulje, sedaj pomnožene z *I*. Tako je integral enak I^2. Potem je volumen grbe

$$V = I^2 .$$

To je prvi korak v dokazu, da je ploščina *I* pod normalno krivuljo enaka $\sqrt{\pi}$. Pokazati moramo le, da je $V = \pi$. To zgleda lažje, saj smo se znebili kvadratnega korena. Ugotovili smo, da je $V = I^2$, če režemo grbo na vertikalne trakove. Da dokažemo, da je $V = \pi$, pa narežemo grbo horizontalno.



Slika 6: Horizontalni prerez grbe.

Slika 6 prikazuje grbo, ki ima odrezan vrh na višini $1/2$. Iz te slike ni težko ugotoviti, da je horizontalni prerez grbe disk. Bolj splošno, prerez na višini h je sestavljen iz vseh točk (s,t), kjer je višina grbe vsaj h, torej točke, ki zadoščajo neenakosti

$$e^{-s^2-t^2} \geq h$$

ali ekvivalentnemu pogoju

$$s^2 + t^2 \leq -\ln h \, .$$

To so točke diska, ki ima center v točki (0,0) in radij $r = \sqrt{-\ln h}$. Postaja razvidno, od kje bomo dobili število π. Površina diska pri višini h je $\pi r^2 = \pi(-\ln h)$. Volumen grbe dobimo z integriranjem te površine v mejah od $h = 0$ do $h = 1$ (to število je vrh grbe). Torej

$$V = \int_0^1 \pi(-\ln h)\, dh \, .$$

Integral rešimo z integracijo per partes in dobimo, da je $V = \pi$. S tem smo dosegli cilj tega razdelka, saj smo pokazali, da je

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \, .$$

S tem pa smo tudi pojasnili enačbo normalne krivulje, s katero aproksimiramo verjetnosti pri metanju kovancev.

7. Literatura

- Keith Ball, *Strange Curves*, *Counting Rabbits and Other Mathematical Explorations*, Princeton University Press 2003.

8. Priloga

št. oseb	rojstni dan	št. oseb	rojstni dan
1	20.7.	**21**	9.10.
2	12.4.	**22**	19.2.
3	20.6.	**23**	9.7.
4	30.7.	**24**	11.12.
5	10.12	**25**	12.10.
6	10.10.	**26**	27.12.
7	14.9.	**27**	4.5.
8	16.5.	**28**	11.12.
9	17.2.	**29**	2.10.
10	28.10.	**30**	17.11.
11	10.12.	**31**	20.9.
12	31.13.	**32**	29.6.
13	22.5.	**33**	1.7.
14	22.2.	**34**	15.3.
15	12.8.	**35**	29.4.
16	21.8.	**36**	3.6.
17	15.4.	**37**	13.9.
18	6.6.	**38**	7.2.
19	19.1.	**39**	18.9.
20	30.10.	**40**	18.12.

Rojstni dan delita 5. in 11. oseba.