Lasse Berglund

Grattis på födelsedagarna!

Födelsedagsproblemet är en klassiker från 1930-talet som fortfarande kan förvåna och engagera elever. Idag står också datatekniska möjligheter till buds för elever – och andra – som vill utforska problemet.

Vi ska på följande sidor titta lite närmare på en fråga som då och då – felaktigt – anses vara en paradox. Det är det så kallade *födelsedagsproblemet* som vi ska ta oss an, vilket säkert är bekant för många läsare. För den som inte känner till födelsedagsproblemet så handlar det om sannolikheten att minst två personer kommer att dela samma födelsedag i en uppsättning av *n* slumpmässigt utvalda personer.

Man kan ha otur helt enkelt

Om man i en skolklass med 25 elever påstår att sannolikheten är så hög som 57 % att några av dem har samma födelsedag, får man nog räkna med en del motreaktioner då påståendet verkar vara så helt uppåt väggarna. Låter man därför eleverna i tur och ordning uppge sin födelsedag tills någon ropar ”samma här”, har man fått en bra start på sin lektion. Problemet med sannolikhetslära är emellertid att även mindre troliga händelser kan inträffa, och kanske ingen ropar ”samma här”. Och skulle man hälsa på i klassen intill; det kan bli samma resultat där. Hur skapar man en situation som bringar reda i detta?

Några rader modern matematikhistoria

I Skolverkets text *Matematik för grundskolan* kan man läsa att ”undervisningen ska ge eleverna förutsättningar att utveckla kunskaper om historiska sammanhang där viktiga begrepp och metoder i matematiken har utvecklats”.

En metod som kanske kan väcka nyfikenhet hos eleverna är *Monte Carlo-metoden*. Den växte fram under första halvan av 1900-talet och tog sen fart i och med utvecklingen av datorer. Idén bakom metoden är att *simulera* det förlopp man vill studera och där ett viktigt inslag är användningen av *slumpen*. Monte Carlo-metoden har visat sig vara framgångsrik på ett flertal skilda områden: exempelvis inom biologi och för finansmarknadsprognoser.

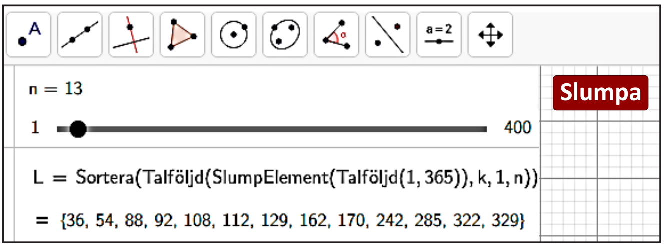
I en studie av födelsedagar behövs många observationer. Men då vi inte har möjlighet att springa runt på skolor och ute på allmänna platser i avsikt att bunta ihop folk i lämpliga grupper, återstår bara en sak: att göra simuleringar.

Två listor och en knapp

I övningarna nedan kommer vi att anta att årets alla dagar är lika troliga som födelsedag, och vi kommer även att bortse från skottår. Detta blir den modell vi ska arbeta med.

Vi väljer Geogebra för de simuleringar vi ska göra. Det är praktiskt att först skapa en glidare *n* med heltalsvärden från 1 till, säg, 400, och en knapp för att med den slumpa fram de listor vi ska studera; se appendix för att se hur knappen skapas.

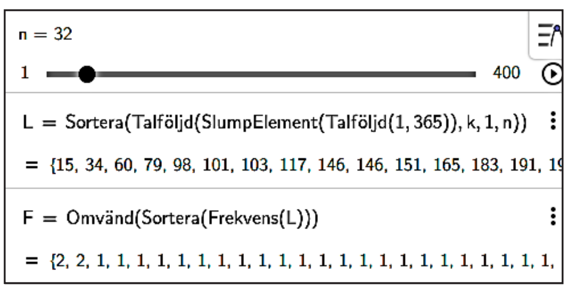
Vad vi behöver är en lista med heltal framslumpade av Geogebra från intervallet 1–365, och där en parameter *n* ska ingå för att med glidaren enkelt kunna ändra gruppstorleken. Trevligt är också om listan är sorterad. Figur 1 visar vår första inskrivning.



Figur 1.

Om eleverna provar att klicka några gånger på Slumpaknappen, ser de att värdena i lista L hela tiden ändras. Med $n = 13$ som i figur 1 ovan, visar listan tretton nya datum framslumpade för varje klick. Några dubbletter bland talen syns inte till, vilket innebär att det i denna grupp inte finns några personer som är födda samma datum.

En frekvenslista är ett lämpligt komplement. Vi sorterar även denna lista, och väljer samtidigt att vända på den, se figur 2.

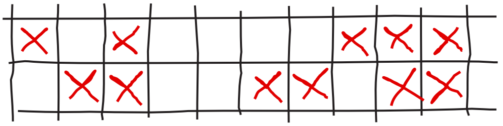


Figur 2.

Figur 2 visar en framslumpad grupp på 32 personer, och i frekvenslistan F framgår att i denna grupp finns två stycken par, där det ena paret syns i rutan: två personer är födda på årets 146:e dag.

Insamling och sammanställning av data

Nu vill eleverna så klart stilla sin nyfikenhet(!) och se om det verkligen stämmer, att chansen är så hög som 57 % för att några av personerna i en grupp på 25, är födda samma datum. Om en elev klickar kan någon annan med papper och penna sätta kryss var gång en 2:a eller större siffra visar sig i lista F, se figur 3.



Var ruta är en ny framslumpad grupp. Gruppstorleken är 25, och de röda kryssen markerar när det har varit annat än enbart 1:or i lista F. 11/20 = 0,55

Figur 3

Klassens resultat skriver man upp på tavlan allteftersom. Tabellen nedan visar hur det kan se ut efter fyra inlämningar.

$$\frac{11}{20}=0{,}55 \qquad \frac{15}{25}=0{,}6 \qquad \frac{7}{20}=0{,}35 \qquad \frac{31}{50}=0{,}62$$

Vi noterar att redan dessa resultat framtagna under en lektion, motsvarar att hälsa på i 115 klassrum. När eleverna sen beräknar det samlade medelvärdet, är det inte otänkbart att de gör det på olika sätt, se nedan.

$$\bar{x}=\frac{0{,}55+0{,}6+0{,}35+0{,}62}{4}=0{,}53 \qquad \bar{x}=\frac{11+15+7+31}{20+25+20+50}=0{,}56$$

Diskussioner kan uppstå. Varför blir resultaten olika? Vardagliga resonemang kan leda fram till att alla har inte gjort lika många simuleringar. Intresserade elever kan med algebra bevisa att de två metoderna ger samma resultat endast då alla insamlade mätvärden är komna ur lika stora urval. Här är alltså den andra metoden den korrekta.

Detta resultat som faktiskt pekar mot 57 %, kommer säkert ändå att framstå som obegripligt för en del. Vi återkommer till detta obegripliga. Men simuleringarna talar ändå sitt tydliga språk: eleverna känner att de själva på allvar har studerat frågan. De äger frågan, ungefär på samma sätt som den forskare gör när det är hennes data man diskuterar. Fler funderingar i klassrummet:

- Vad innebär det om lista F ser ut som följande: {4, 3, 3, 2, 1, 1 … 1}?
- Vad krävs för att få en lista som denna?

En trevlig utvikning

Rätt snart infinner sig frågan: hur många ska ingå i gruppen om man vill vara 100 % säker på en dubblett? Synpunkter kan dyka upp och kanske ett av förslagen är 365 personer. Men strax inser nån att ”om vi har otur, kan alla dessa 365 ha blivit utsmetade jämnt över hela kalendern. Men har vi ytterligare en person, så måste denne klämmas in nånstans i alla fall.” Så är det: med 366 personer garanterar den så kallade *lådprincipen* att åtminstone ett datum besätts av mer än en person. Denna princip påtalar att om man har n stycken lådor och $n+1$ stycken föremål, måste minst en låda innehålla två eller flera föremål. Resonemang och praktiska övningar kring detta kan vara en trevlig utvikning.

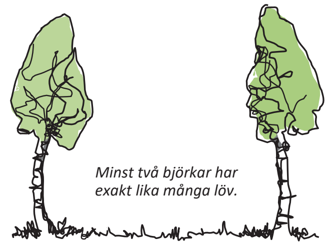
Mellanstadiet

Plocka fram fyra lådor och fem pennor och låt eleverna undersöka denna princip: *Minst en låda kommer att innehålla två eller flera pennor.*

Högstadiet

Fakta: Det finns fler björkar i Sverige än det finns löv på en björk. Låt löven vara lådorna i lådprincipen.

Vad kan man utifrån detta säga om björkarna i vårt land?



Vad säger teorin om en grupp på 25 personer?

Som så ofta inom sannolikhetsläran är det även denna gång lättare att börja i den motsatta änden: hur blir kalkylen om vi inte har någon dubblett alls?

Att den första personen är född ett visst datum är hundra procent säkert, det vill säga (365/365). Sannolikheten att nästa person inte fyller år samma dag blir då den betingade sannolikheten (364/365), då 364 stycken datum finns tillgängliga. Så här forsätter det att minska med 1. Tar vi med ytterligare en person (363/365), och multiplicerar dessa sannolikheter, får vi sannolikheten för att alla tre personerna har olika födelsedagar. Då vi söker *komplementhändelsen* till detta – att åtminstone någon dubblett finns – får vi här i denna artikel, men också i en Ma5-kurs:

Förlänger på lämpligt sätt.

$$P_3(\textit{någon dubblett}) = 1 - \frac{365}{365}\cdot\frac{364}{365}\cdot\frac{363}{365}\cdot\frac{362!}{362!} = 1 - \frac{365!}{365^3\cdot 362!}$$

$$= 1 - \frac{365!}{365^3\cdot(365-3)!} = 1 - \frac{\text{nPr}(365;3)}{365^3} = 0{,}008$$

där nPr(365;n) är Geogebras beteckning för de så kallade permutationerna, $365!/(365-n)!$. Här har vi nu ett samband som låter eleverna jämföra sina empiriska resultat med ett teoretiskt.

$$P_{25}(\textit{någon dubblett}) = 1 - \text{nPr}(365;25)/365^{25} = 0{,}568 \approx 57\,\%.$$

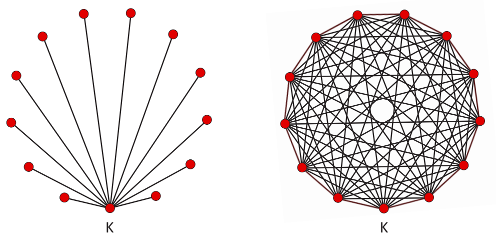
– Det stämmer tametusan!

– Jamen ändå ganska märkligt.

Paradoxen som försvann

Den jämförelse som vi här har studerat, förväxlas inte sällan med en annan jämförelse: alla i gruppen ska jämföras med *en själv*. Men det är inte *en* person mot alla de övriga, utan jämförelsen gäller alla mot alla.

Låt oss kika på en grupp med 13 personer och ställa frågan om någon i gruppen fyller år samma dag som, säg Kalle. Det blir nu enbart Kalle mot var och en av de andra: $1 \cdot 12$. Men som grundfrågan är ställd, så ska *var och en* av de 13 personerna jämföras med de övriga 12. Antalet jämförelser blir därmed: $13 \cdot 12/2 = 78$, där halveringen måste ingå därför att om Anna är jämförd med Lisa, så är Lisa därmed redan jämförd med Anna. Se figur 4, där vi har K som i Kalle.



13 personer, 12 stycken jämförelser

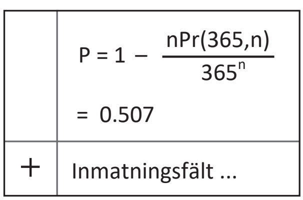
13 personer, 78 stycken jämförelser

Figur 4

Den klassiska frågan

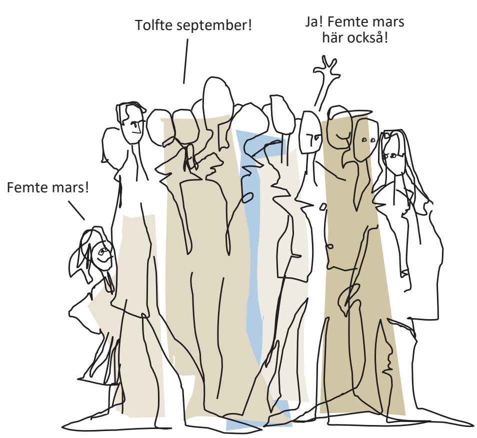
Nu kan eleverna också lägga in detta explicita samband i Geogebra med n som variabel. Här har eleverna nu möjlighet att få svar på den klassiska frågan: hur många personer måste en grupp bestå av, för att det ska vara mer sannolikt att några har samma födelsedag, än att inga har det?

Ekvationen som ska lösas för att få svar på frågan är $P_n = 0{,}5$. Men denna ekvation är utom räckhåll för både elever och lärare. Fast med tillgång till glidare n som eleverna kan dra i, kan de se hur värdena hela tiden ändras. I figur 5 syns ett värde: 0,507, det vill säga strax ovanför 50 %. Detta värde visas då $n = 23$. Svar: 23 personer räcker.



Figur 5

Det är detta resultat som då och då – felaktigt – betraktas som en paradox.



Inget exceptionellt här, än mindre något paradoxalt.

Två frågor i klassen där eleverna först kan simulera fram ett resultat och sedan jämföra detta med P_n:

- Sedan vår förste statsminister tillträdde år 1876 har Sverige haft 35 statsministrar och där två av dem, Nils Edén och Östen Undén, är födda den 25 augusti båda två. Är inte detta sammanträffande helt otroooligt!
- Man samlar 60 slumpvis valda tonåringar i en aula. Visa att detta antal räcker för att sannolikheten ska överstiga 50 % för att där finns personer med samma ÅÅMMDD.

Det kostar på att vara hundra

För att vara 100 % säker på en dubblett behövs en grupp på 366 personer. Att samla ihop så mycket folk kräver en del. Om vi därför väljer att skippa inte mindre än 300 av alla dessa för att minska detta jobb, får vi: $P_{66} = 99{,}8\,\%$. Kontenta: om vi avstår från kravet på absolut visshet, så sjunker kostnader och tidsåtgång avsevärt.

Någon slags omvänd analogi till detta är så kallade samlarbilder. Inledningsvis får samlaren bra utdelning; det blir nya fotbollsspelare hela tiden. Men efterhand kommer med nödvändighet dubbletterna. Utan att här göra några beräkningar, kan vi bara konstatera att har man fått sitt album fullt så när som på *en* bild, så kommer denna sista bild att bidra rätt rejält till den totala kostnaden. Det är den sista procenten som kostar.

Klassen igen

Om vi avslutningsvis går tillbaka till vår klass för att där ta fram ett värde som ligger i linje med vad folk vanligen förväntar sig, får vi välja att göra jämförelserna mot endast en av klassens 25 elever: Stina. Då är sannolikheten 364/365 för var och en av klassens övriga elever att de inte fyller år samma dag som Stina. Multiplikationsprincipen i samarbete med komplementet ger: S_{25}(samma födelsedag som Stina) $= 1 - (364/365)^{24} \approx 6{,}4\,\%$. Sannolikheten blev väsentligt lägre än värdet 57 % från tidigare.

Skolan och världen utanför

Med simulering som metod har vi i denna undersökning gjort ett försök att ansluta skolmatematiken till världen utanför, då det i verklig forskning kan gå till så här ibland. Man kan sen gå vidare och låta eleverna ersätta talet 365 i lista L med andra värden, som direkt överför frågan till händelser som att fylla år samma månad eller samma vecka, eller

- kanske frågan om samma namn. Vilka förenklingar av verkligheten behövs då?
- en simulering av en pokerhand kanske, med mod13 i en kompletterande lista
- något helt annat från tillvaron som eleverna själva hittar på.

Låt eleverna formulera hypoteser utifrån sina framslumpade värden och därefter använda P_n eller något annat passande samband för att jämföra sina resultat med vad teorin säger.

Appendix

För att skapa en knapp:

- Välj menyn där glidaren finns. Välj där *Skapa knapp*.
- Klicka i ett ritområde. En dialogruta dyker upp. I *Förklaring* skriver man Slumpa och i *GeoGebra script* skriver man UppdateraKonstruktion[ ]. Klicka OK och klicka sen en gång nånstans utanför ritområdet. Nu fungerar knappen.

PS: Om det är så, alla ni läsare av Nämnaren, att 50 stycken av er har läst denna artikel, så kan jag nästan ta gift på att några av er fyller år på samma dag. DS.

Jonas Hall