

Det här är underlaget till den första tryckningen av boken "Är tärningen kastad? Sannolikhetslära för vem som helst", som jag skrev våren 2008. Men istället för att de få böcker som trycktes ska samla damm på diverse bibliotek, så har jag valt att tillgängliggöra materialet även här på webben, ifall något skulle vara användbart. Materialet innehåller vissa felaktigheter, men någon errata-lista finns f.n. ej sammanställd.

Andreas Svensson
Uppsala, 3 oktober 2013

Andreas Svensson

Är tärningen kastad?

Sannolikhetslära för vem som helst



Tillgängliggjort via webben 3 oktober 2013
<http://user.it.uu.se/~andsv164/ArTarningenKastad.pdf>

e-post till författaren: andreas.svensson@it.uu.se



FÖRORD

”Någon talade om för mig att varje ekvation som jag tog med i boken skulle halvera försäljningen” har Stephen Hawking skrivit i förordet till sin bok Kosmos – en kort historik. Istället för att följa det rådet, vill jag (med risk för att låta naiv) hellre övertyga denne ”någon” och alla dessa läsare som eventuellt skulle utebli om att formler och ekvationer inte alls är farliga! Man behöver inte alls vara någon matematiker för att ha glädje och nytta av matematikkunskaper. Och alla kan förstå mer av matematiken än vad man först tror! Det gäller bara att släppa fördomarna om att det är ”svårt”, ”obegripligt” och att man ”har svårt att lära”, för såväl den som lär sig som för den som lär ut.

Min förhoppning är att den här boken ska kunna läsas både av dem som vill bilda sig utan att ha så mycket på fötterna sedan tidigare, och av drivna högstadie- och gymnasieelever som vill läsa utöver kursen i skolan på egen hand. Flera av kapitlen innehåller sådana moment som inte ingår i ens de högsta matematikkurserna på gymnasiet, så det finns något för de flesta att hämta här.

De läsare som känner att de har grundläggande kunskaper inom området sedan tidigare kan med gott samvete nöja sig med att ögna igenom vissa kapitel, men för den läsare som börjar från grunden rekommenderar jag att man läser boken från pärm till pärm. Vissa delar går snabbt att läsa och sätta sig in i, medan vissa delar kanske behövs läsas ett par gånger innan förståelsen kommer. Det är inga konstigheter med det, och det är den oslagbara fördelen med en sådan här typ av bok – du väljer tempot alldeles själv!

INNEHÅLLSFÖRTECKNING

Den här boken har kommit till under många timmars skrivande, läsande och funderande. Att det skulle vara så här roligt att skriva en bok kunde jag inte ana när jag gav mig in på projektet! Det har varit väldigt lärorikt och utvecklande för min egen del, och skrivandet har inneburit intressanta diskussioner, funderingar och möten med andra personer – som jag annars inte skulle ha träffat – som har givit mig idéer och stöd. Jag hoppas nu verkligen att den här boken kan få vara lärorik även för alla läsare.

Det är långt ifrån alla gånger självklart vilket matematiskt resonemang man ska föra för att nå fram till rätt svar, men jag hoppas att den här boken ska vara en inspirationskälla och bilda er läsare inom området, men också stimulera er till egna matematiska funderingar (och kanske till och med studier?). Eftersom jag själv inte är någon riktig matematiker, utan bara en intresserad gymnasieelev, kan jag inte garantera att det inte har smugit sig in någon matematisk felaktighet någonstans, men jag har naturligtvis försökt att undvika alla sådana tvivelaktigheter.

Det finns många personer som har läst och kommenterat hela eller delar av manuskriptet och jag vill naturligtvis tacka alla, men de fem personer som enskilt har betytt mest för bokens innehåll och utformning och som jag vill rikta ett extra stort tack till är mina korrekturläsare Niklas Järvstråt, Martin Gellerstedt och Bo Svensson – alla tre på Högskolan Väst – samt Lars Holmblad på Nils Ericsonsgymnasiet, och min förläggare Marianne Rugård Järvstråt.

Andreas Svensson
Trollhättan den 26 maj 2008

Ps. Som författare är det alltid roligt att få kommentarer från sina läsare. Min e-postadress finns på sidan fyra. Ds.

Förord	7
Innehållsförteckning	9
Inledning	11
Kapitel 1 Första kapitlet	15
Kapitel 2 Sannolikhetskalkylens grundvalar – intressantare än du anar!	23
Kapitel 3 Det klassiska getproblemet	37
Kapitel 4 En slump?	45
Kapitel 5 Det långa loppet	53
Kapitel 6 När man talar om trollen, och några andra talesätt	63
Kapitel 7 Kombinera!	77
Kapitel 8 Öka din lottovinst	87

INLEDNING

Kapitel 9	
Beroende händelser	103
Kapitel 10	
Kuvertproblemet	111
Kapitel 11	
Svälta räv	121
Kapitel 12	
Normalfördelning	129
Kapitel 13	
Tristan Miller: Varför jag aldrig kommer att ha en flickvän	139
Kapitel 14	
Benfords begynnelsesiffror	145
Kapitel 15	
Osannolika sannolikheter	155
Efterord	165
För den intresserade läsaren	167
Svar och lösningar till övningsuppgifterna	169
Källförteckning	175
Register	181

Spel är en stor del i den här boken. Dels för att det är ett område där kunskaperna snabbt kommer till användning, dels för att reglerna ofta brukar vara tydliga så att man har full kontroll på omständigheterna. Men framförallt beror det på att sannolikhetsläran började studeras från en vetenskaplig vinkel tack vare just spel; Fransmännen Blaise Pascal och Pierre de Fermat började på 1600-talet på det som skulle komma att bli läran om sannolikheter, i och med en brevväxling om hur man skulle dela upp vinster mellan spelare i ett visst tärningsspel när det inte fanns möjlighet att slutföra omgången. Innan dess hade det som skulle kunna motsvaras av sannolikhetsläran endast bestått av spelares känsla och erfarenhet. Själva starten för den moderna sannolikhetsläran anses dock ligga ett par hundra år senare, i och med Pierre-Simon Laplaces verk från 1812, *Théorie analytique des Probabilités*.

Trots att läran om sannolikheter är förhållandevis ung, så talas det redan i Gamla Testamentet i Bibeln om lotter: ”Sedan skall han ta de båda bockarna och ställa dem inför Herren vid ingången till uppenbarelsetältet och kasta lott om dem: en lott för Herren och en lott för Asasel.” (3 Mos 16:7-8)

Att försöka bemästra slumpen kan verka lockande när lotterier lockar med högvinster, det finns tips i horoskoperna om vad och hur man ska spela, och Svenska Spel publicerar statistik över hur ofta de olika siffrorna i Lotto har blivit dragna. Men kom ihåg – det finns bara en garanterad vinnare i ett lotteri, och det är arrangören! Och för arrangören har den moraliska aspekten i konflikt med

vinsten alltid funnits där; 1725 förbjöd påven Benedictus XIII lotteriet i Genua (som anses vara ursprunget till dagens lotto), men bara sex år senare startade hans efterträdare Clement XII ett eget lotteri i Rom för att öka Vatikanens inkomster. I Sverige går årligen mer än en miljard av vinsten från Svenska Spel till barn- och ungdomsidrott samt föreningslivet. Den cyniske kan säkert hävda att det är ett sätt att rentvå sitt samvete för de miljarder man tjänar på det svenska folkets spelande.

Vi kan aldrig vet vad slumpens nästa drag är. Vi kan göra gissningar utifrån vad vi tidigare sett och utifrån teoretiska beräkningar. Men vi kan aldrig veta säkert. Matematikern Bertrand Russel uttryckte det som ”Hur kan vi tala om slumpens lagar? Är då inte slumpen lagbundenhetens fullständiga motsats?”

Kapitlen i boken kan i de flesta fall läsas fristående, och man kan ofta hoppa över något kapitel och ändå förstå det nästkommande. För läsaren som vill få en helhet rekommenderar jag dock att man läser boken från pärm till pärm, något som för de flesta torde vara överkomligt.

Vi börjar i kapitel ett och två med lite grundläggande principer, och tittar sedan i kapitel tre på ett välkänt och klurigt problem, det så kallade getproblemet. I kapitel fyra och fem kommer det mer grundläggande teori, som innehåller färre formler och mera funderingar än de tidigare kapitlen. Kapitel sex är ett väldigt roligt kapitel, där vi tittar på sannolikheter i vardagen, följt av det användbara kapitel sju som handlar om teorin bakom kombinationer. Kapitel åtta handlar om lotto, och innehåller inte så stora delar ren matematik, det handlar istället mycket om beteenden hos spelare. Kapitel nio är sedan ett teorikapitel om händelser som är beroende av varandra, något vi får användning för i de två följande kapitlen. Kapitel tio handlar om det gäckande

kuvertproblemet, och kapitel elva om det klassiska kortspelet svälta räv. Kapitel tolv och fjorton är orienterande inom statistikens område, med den intressanta normalfördelningen och den kanske ännu mer fascinerande Benfords lag. Kapitel tretton är en lite annorlunda tillämpning av normalfördelningen. Slutligen knyter kapitel femton ihop säcken med en diskussion om sådant man – skenbart – aldrig trodde skulle inträffa, som till exempel olyckor i kärnkraftsverk, och lite samlarbilder.

Välkommen till en bok om slumpen, där vi aldrig kan veta, men i det långa loppet kan göra väldigt bra gissningar! Det hela handlar om det som kallas för sannolikheter.

1.

FÖRSTA KAPITLET



Sannolikhetslära handlar om chanser och risker. Men hur vet man om det är en chans eller risk? Ja, det är en bra fråga. Chans brukar man säga om det är något positivt som kan hända, och risk om det är något negativt. För att undvika de problemen säger vi istället sannolikhet, så blir det upp till läsaren att avgöra om det är fråga om en chans eller risk.

Man kan tala om sannolikheten för att något ska hända, och ange det som ett tal mellan 0 och 1, eller 0% respektive 100%. Sannolikheten för att få en klave när man singlar slant är hälften, alltså 0,5 eller 50%. Det är viktigt att när man lägger ihop alla sannolikheter vid en händelse, exempelvis ett tärningskast, så måste det bli just 1 eller 100%. Sannolikheten att få klave är som

sagt 0,5 eller 50% och sannolikheten att få krona är ju också hälften, alltså 0,5 eller 50%. Och $0,5 + 0,5$ är 1, och $50\% + 50\%$ är 100%. Sannolikheten 1, eller 100%, innebär nämligen att det säkert kommer att hända.

Det är klart att något måste hända, när man singlar slant. Vad skulle hända annars? Att den aldrig skulle landa, eller aldrig sluta rulla runt på bordet? *Något* av alla möjliga tänkbara utfall (att det exempelvis blir klave kallas för ett utfall) måste ju hända, alltså måste sannolikheten för alla utfall sammanlagt bli 1. Det kan vara bra att ha i bakhuvudet.

Man kan mäta och räkna sannolikhet i procent, det går jättebra. Och procent är säkert för många lätt att tänka i, exempelvis journalister verkar älska procent, märker man när man läser en tidning. Men av lite olika anledningar föredrar matematiker att använda sig av bråk eller decimalformer istället, så därför gör jag det också, till största del. En av anledningarna är att en tredjedel blir 33,3333...% om man använder procent, och då är det enklare och mer exakt att skriva $1/3$. Om du tycker det känns främmande att sannolikheten för något är 0,4, tänk då 40% istället när du läser det, så ska du se att du snart vänjer dig.

Sju kronor i rad kan väl inte vara normalt?

Singla slant en gång. Vad fick du? Fick du det du trodde att du skulle få? Hade du ens en anledning att tro att du skulle få just krona och inte klave (eller tvärtom)?



Eftersom myntet är lika tjockt på alla ställen och ser i princip likadant ut på båda sidorna (förutom att det är olika bilder, men det påverkar knappast hur myntet landar) så finns det ingen anledning att tro att det ska bli just krona eller klave. Sannolikheten är alltså 0,5 för krona och 0,5 för klave. Det vill säga, singlar vi slant många gånger kan vi förvänta oss att *ungefär* hälften av gångerna blir det krona, och andra hälften klave.

S singlar vi slant tvåhundra gånger, så är det rimligt att tro att ungefär hälften, hundra gånger, av dem blir krona. Men efter en krona blir det inte alltid klave, det kan ju bli flera kronor i rad emellanåt, testa gärna själv! Någonstans runt hälften av de hundra gångerna det blir krona, kommer troligen även nästa mynt i rad att bli en krona, för ett mynt har inget minne, och kommer inte ihåg hur det landade förra gången. Det betyder att ungefär femtio kronor kommer att följas av ytterligare en krona, det vill säga det kommer att vara minst två kronor i rad ungefär femtio gånger.

På samma sätt, så lär hälften av de gångerna det blir minst två kronor i rad, nu är vi nere i tjugofem, följas av ytterligare en krona, alltså minst tre kronor i rad ungefär tjugofem gånger. Vidare, så kommer ungefär hälften av de tjugofem gångerna följas av ännu en krona, det vill säga att vi kan misstänka att ungefär tolv gånger kommer det att vara minst fyra kronor i rad. Om man fortsätter, så kommer det att runt sex (hälften av tolv) gånger vara minst fem kronor i rad, tre gånger minst sex kronor i rad och ungefär en gång sju kronor i rad! Lite överraskande kanske?

Det är alltså inte så oväntat som det först kan verka med sju kronor i rad om man singlar slant ett par hundra gånger. Även åtta kronor i rad skulle inte vara förvånande om man singlar slant så många gånger. Och singlar man slant tusen gånger skulle inte ens tio kronor i rad vara särskilt otroligt.

‘Tycker du att det känns osannolikt med sju kronor i rad och tycker att “så betar sig ju inte slumpen”? Du är inte ensam; det sägs att Theodore Hill, nuförtiden professor emeritus i matematik i Atlanta, gav sina nya studenter i läxa att singla slant tvåhundra gånger och skriva upp utfallen i en lista. Några av studenterna tyckte att det var en onödig, tidsödande uppgift och fuskade genom att skriva ner vad de tyckte verkade rimligt istället för att verkligen singla slant så många gånger. Dem kunde Hill ofta lätt avslöja genom att ögna igenom deras listor – de fuskande studenterna skrev sällan fler än tre lika i rad, för att det intuitivt kändes osannolikt. Men som vi precis sett, så är det inte alls omöjligt – utan snarare troligt – att vi har sex eller till och med sju lika i rad! De med maximalt tre kronor i rad kunde Hill med stor sannolikhet alltså anklaga för fusk. “Sanningen är”, säger Hill, “att folk inte känner till vad som egentligen är sannolikt. Därför kan de sällan fejka ihop trovärdiga data utan att det märks.”

Lägg märke till alla mina uttryck som förmodligen, ungefär, kanske, cirka, lär och så vidare. Om du så singlar slant två tusen gånger och aldrig får mer än fem kronor i rad så *behöver* inte det betyda att det är fel på myntet eller slumpen, trots att det är ett relativt osannolikt resultat – det är tvärtom ett tecken på att det är just slumpen! Efter att ha läst den här boken kommer du förmodligen att ha fått en känsla för hur slumpen fungerar eller inte fungerar och vad den är eller inte är.



Ett enkelt och fiffigt tärningsknep, en liten rivstart

Jag vill inte uppmana till spel om pengar med den här boken, de allra flesta som spelar om pengar går back (hur skulle Svenska Spel kunna göra sina årliga miljardvinster om alla de som spelade också gick med vinst?). Det finns inga lottorader som har större chans att vinna än andra. Däremot skadar det ju inte att vinna lite prestige av vännerna och familjen runt soffbordet en lördagskväll, och det är den formen av spel jag själv föredrar.

Om man kastar en tärning en gång, så finns det ingen anledning att tro att någon viss sida ska komma upp (om det inte är något lurt med tärningen, som att en viss kant eller sida väger extra mycket, eller så). Men om man kastar två tärningar en gång var och räknar ihop summan av ögonen på dem, finns det någonting att tänka på då? Ja, till att börja med är det dumt att satsa sitt lördagsgodis på att summan blir ett, eftersom det inte går. Om båda tärningarna visar det lägsta möjliga, en etta var, så blir det två sammanlagt. Och summan två på tärningarna går bara att få på *ett* sätt (båda tärningarna måste visa ett), medan exempelvis summan fyra går att få på *tre* sätt – ena tärningen visar ett, den andra tre; båda tärningarna visar två; ena tärningen visar tre och andra ett. (Värt att notera är alltså att först en etta och sedan är trea inte är samma sak som först en trea och sedan en etta, medan två tvåor bara är en möjlighet.) Vi kan alltså ana att vissa summor är bättre än andra att satsa på. Men blir det bättre desto större summa? Det kanske är störst sannolikhet att få tolv? Det är ju den största summan. Fast det går å andra sidan bara att få på ett enda sätt; båda tärningarna måste visa sexor. På nästa sida kommer en tabell för att få ordning på det här.

Första kastet							
Andra kastet	1	2	3	4	5	6	
	1	1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
	2	2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
	3	3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
	4	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
	5	5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
	6	6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

Allt som allt blir det 36 olika möjliga tärningspar, eller utfall som vi ska kalla dem, och eftersom sannolikheten är lika stor för vilket utfall som helst på vardera tärningen, och de inte påverkar varandra, så blir sannolikheten för varje tärningspar lika stort, det vill säga att sannolikheten för utfallet “tre på första kastet och fem på andra kastet” är precis lika stor som sannolikheten för utfallet “två på första kastet och fyra på andra kastet”. Och eftersom det finns flest par med summan sju (närmare bestämt sex sådana par), blir sannolikheten störst för summan sju, när man kastar två tärningar.



Det finns fem par där summan är sex, och lika många med summan åtta, vilket betyder att sannolikheten att få summan sex på två tärningskast är lika stor som att få summan åtta.

Vill man räkna på *hur stora* sannolikheterna är, (och man inte bara nöjer sig med att veta vilken som är störst), kan man göra det. Det finns som sagt 36 möjliga utfall, i form av 36 tärningspar. Och som vi har konstaterat så är sannolikheten lika stor för varje utfall om tärningarna är korrekta. Och sannolikheten för att det blir *något* av de trettiosex möjliga utfallen måste vara 1, eftersom det inte finns några andra möjligheter. Alltså blir sannolikheten för ett speciellt utfall, ett särskilt tärningspar, $1/36$. (Det kommer mer om hur man räknar ut sådana här sannolikheter i nästa kapitel.)

Men hur stor är då sannolikheten att man får summan sju? Jo, sex olika utfall ger summan sju, och sannolikheten för varje utfall är som vi nyss sagt $1/36$, så det är bara att lägga ihop sannolikheterna för varje utfall: $1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36$ vilket är $6/36$ eller $1/6$. Sannolikheten att få summan sju av två tärningskast är därmed lika stor som sannolikheten att få, exempelvis, en femma vid *ett* tärningskast, $1/6$.

Sannolikheten att få summan tre är lika stor som att få summan tio vid två tärningskast ($3/36$ eller $1/12$). Så visst finns det taktik i vilken summa man ska satsa på att få, om man slår med två tärningar. Samma principer gäller för tre tärningar, fast där blir det lite mer avancerat. Sannolikheten att få tre sexor, summan 18, är något så litet som $1/216$, och här är det summorna 10 och 11 som är de vanligaste, båda med sannolikheten $1/8$. När du har kommit en bit in i boken kommer du förmodligen kunna räkna på sådant på egen hand.

Övningsuppgifter

För att du ska kunna testa dig själv, finns det i slutet på flera av kapitlen lite övningsuppgifter du själv kan räkna på, och handlar om sådant som förekommit i kapitlet. Kör du fast, så finns det lösningar och facit längst bak i boken.

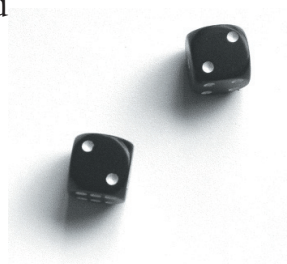
1.

Du passerar ett trafikljus på morgonen på väg till jobbet/skolan varje dag (vi struntar i helgdagarna). I genomsnitt var fjärde gång är det rött. Försök att rangordna dessa alternativ, i ordning efter vilket som är mest sannolikt. Kontrollera sedan med facit och se om din uppfattning stämde!

- a) grönt fem morgnar i rad
- b) rött fem morgnar i rad
- c) rött två morgnar i rad
- d) grönt två morgnar i rad

2.

Du kastar två tärningar, först en blå och sedan en röd. Ungefär hur stor är sannolikheten att du får samma eller ett större antal ögon på den blå tärningen, det första kastet, än den röda? Mer eller mindre än 0,5? Tips: tag hjälp av tabellen med tärningspar tidigare i kapitlet.



2.

SANNOLIKHETSKALKYLENS GRUNDVALAR

– intressantare än du anar!



Det finns en tanke bakom den lite smått jobbiga rubriken – att visa att ett svårt namn inte behöver vara så jobbigt som det låter. Du ska få se. Hela det här kapitlet handlar om sannolikhetskalkyl, vi kommer att gå igenom de grundläggande principerna som behövs för att räkna på sannolikheter. Det öppnar fler dörrar än vad man först kan tro.

Gynnsamma och möjliga fall

Det kanske mest grundläggande för förståelsen inom sannolikhetsläran är att sannolikheten, P , för att något särskilt ska inträffa beräknas som

$$P = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$$

Gynnsamma utfall är de utfallen man vill "räkna på". Ska man exempelvis räkna på sannolikheten att få en trea på ett tärningskast, så är utfallet {tre} det gynnsamma, och utfallen {etta}, {två}, {tre}, {fyra}, {fem} och {sex} de möjliga. (Klammerparenteserna används för att beteckna utfall eller utfallsserier. Här i boken används de inte konsekvent, utan bara där de behövs för att förtydliga att något är ett utfall.)

Sannolikheten att få en trea på ett tärningskast blir då enligt följande:

$$\frac{\text{tre}}{\text{etta, två, tre, fyra, fem, sex}} = \frac{1 \text{ fall}}{6 \text{ fall}} = \frac{1}{6}$$

eftersom det bara är trean som är ett gynnsamt utfall.

Vill man räkna ut sannolikheten för att få ett udda antal ögon när man kastar en tärning en gång, så blir det $3/6$ (som är samma sak som $1/2$ eller $0,5$), eftersom det kan bli ett udda antal ögon på tre sätt (etta, tre eller fem) och det finns sex tänkbara fall (etta, två, tre, fyra, fem och sex).

En förutsättning för att $P = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$ ska gälla är att sannolikheten är lika stor för varje fall. Med en tärning utgör det inget problem, men om man kastar en plastmugg så är inte sannolikheten lika stor att den hamnar ståendes eller upp-och-ner

som att den hamnar på sidan. Vi ska titta mera på det i kapitel fem.

Det hade varit enkelt om man hade kunnat använda $P = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$ i en tävling (med en vinnare och många förlorare) och resonera att det finns två fall – antingen vinner jag eller så vinner jag inte. Det skulle innebära att chansen att vinna skulle vara $1/2$ eller $0,5$. Men riktigt så enkelt är det inte, och det kan man ju förstå med sitt förnuft. Eftersom sannolikheterna att du vinner respektive förlorar förmodligen inte är lika stora går det inte att använda den formeln.

Säg att vi vet lite mer om tävlingen. Vi antar att det är ett lotteri där vi vet att det finns 50 lotter, varav en kommer att vinna och 49 stycken inte kommer att vinna, och ingen särskild lott har större chans att vinna än någon annan. Sannolikheten för att vinna blir då $1/50$ eller $0,02$; antingen vinner man, det kan man bara göra på ett sätt, eller förlorar man, och det kan man göra på 49 sätt, eftersom det finns 49 lotter utan vinst.



$P(A)$, ett användbart skrivsätt

Ett vanligt skrivsätt för sannolikheter är $P(A)$, som utläses "sannolikheten för A". A betyder här en särskild händelse. P:et har att göra med latinets *probabilitas*, engelskans *probability* och franskans *probabilité*, som alla betyder sannolikhet. Det är alltså inga nyheter för oss, mer än att skrivsättet $P(A)$ blir mycket

kortare och tydligare än att skriva ut det i ord varje gång. För visst minskar risken för missförstånd om man skriver ” $P(\text{fyra på ett tärningskast}) = 1/6$ ” istället för ”sannolikheten för att få en fyra på ett tärningskast är en sjättedel”?

OCH

En annan grundläggande princip inom sannolikhetsläran är att om man ska räkna ut sannolikheten för att både händelse A *och* händelse B ska inträffa, så multiplicerar man deras sannolikheter. Detta kallar vi för multiplikationssatsen. Alltså

$$P(A \text{ och } B) = P(A) \cdot P(B)$$

Om vi vill räkna ut sannolikheten för att få två femmor när vi kastar två tärningar multiplicerar vi sannolikheten för att få en femma med sannolikheten för att få en femma, alltså $1/6 \cdot 1/6$, vilket blir $1/36$. Det verkar stämma, om vi jämför med tabellen med tärningsparen i förra kapitlet där sannolikheten för ett tärningspar är $1/36$, och {femma, femma} är just ett sådant par.

Dockskadetläggasinettlitenförbehållhär—multiplikationssatsen gäller bara om händelserna är *oberoende*, som exempelvis tärningar. Om däremot utfallet av händelse A påverkar sannolikheten för utfallen i händelse B fungerar inte multiplikationssatsen. Detta tas upp längre fram i boken, i kapitel nio.

ELLER

Om man vill veta sannolikheten för att antingen händelse A *eller* händelse B inträffar, så adderar man sannolikheterna. Alltså

$$P(\text{antingen A eller B}) = P(A) + P(B)$$

Vill man räkna ut sannolikheten för att få antingen en tvåa *eller* en trea på ett tärningskast, så får vi $P(\text{antingen tvåa eller trea}) = P(\text{tvåa}) + P(\text{tre})$, och med siffervärden $1/6 + 1/6 = 2/6 = 1/3$.

Det man matematiskt brukar avse med ”eller” är dock ”antingen så händer A *eller* så händer B *eller* så händer A *och* B”. I dagligt tal brukar man normalt inte precisera sig så noga vad man menar med ”eller”, men i matematiken finns det behovet. Använder vi den matematiska definitionen av ”eller” och frågar om det har regnat igår *eller* idag, så får vi alltså svaret ”ja” om det dels har regnat igår, dels idag, men även om det har regnat båda dagarna.

Därför är definitionen av den så kallade additionssatsen

$$P(A \text{ eller } B) = P(A) + P(B) + P(A \text{ och } B)$$

När det till exempel handlar om att få en trea *eller* en fyra på ett tärningskast behöver vi inte göra någon skillnad på ”antingen *eller*” eller enbart ”eller”, eftersom man helt enkelt inte kan få en trea *och* en fyra på samma kast.

Men den uppmärksamme läsaren undrar säkert vad det här ska vara bra för. Man kan ju lika gärna använda formeln $P = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$ för att räkna ut sannolikheten för en tvåa *eller* trea; det blir ju två gynnsamma fall och sex möjliga, alltså $2/6$ eller $1/3$. Så varför ska vi ha en formel till för att räkna ut samma sak?

Det är en helt riktig iakttagelse, men ibland är det lättare att använda additionssatsen. Och ibland fungerar helt enkelt inte $p = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$ för vårt ändamål; ibland vet vi inte antalet möjliga fall, och vi vet ju också att den inte kan användas när sannolikheterna för de olika fallen är olika stora, som när man till exempel kastar en plastmugg.

OCH och ELLER

Vad är sannolikheten att få *en* trea *och* *en* fyra när man kastar en tärning två gånger (det spelar alltså ingen roll i vilken ordning man får dem)? Ja, det är lite lurigare att räkna ut än man först kan tro. Vi kan istället börja med frågan ”vad är sannolikheten att *först* få en trea och *sedan* en fyra när man kastar en tärning två gånger?” Då blir svaret $P(\text{tre} \text{ på första kastet}) \cdot P(\text{fyra på andra kastet})$, alltså $1/6 \cdot 1/6 = 1/36$.

En trea och *en* fyra på två kast kan man ju antingen få genom att först få en trea och sedan en fyra *eller* först få en fyra och sedan en trea. Alltså blir svaret på den första frågan: $P(\text{tre} \text{ på första kastet}) \cdot P(\text{fyra på andra kastet}) + P(\text{fyra på första kastet}) \cdot P(\text{tre} \text{ på andra kastet}) = 1/6 \cdot 1/6 + 1/6 \cdot 1/6 = 1/36 + 1/36 = 2/36 = 1/18$.

Återigen kan vi kolla svaret i tabellen med tärningspar i förra kapitlet (att kasta en tärning två gånger eller två tärningar en gång var fungerar precis likadant) – där finns det två par med en trea och en fyra, alltså borde sannolikheten vara $1/36 + 1/36 = 1/18$, så det verkar stämma. Det här med att man kan räkna ut sannolikheten

för en händelse på flera olika sätt (här fungerar exempelvis både tabellen och formlerna) kommer du att träffa på flera gånger. Ibland är det uppenbart vilken metod som är smidigast, men inte alltid. Då kan man kanske välja den metod man känner sig säkrast på, eller den metod man känner sig mest osäker på för att lära sig den, och eventuellt kontrollera med den andra metoden.

Man kan också fråga sig vad sannolikheten att bara få treor och/eller fyror på två kast är (alltså inga ettor, tvåor, femmor eller sexor). Då får vi tänka lite annorlunda; i första kastet kan vi få en trea *eller* fyra för att det ska stämma *och* i andra kastet kan vi också få en trea *eller* fyra. Det blir alltså $(P(\text{tre} \text{ på första kastet}) + P(\text{fyra på första kastet})) \cdot (P(\text{fyra på andra kastet}) + P(\text{tre} \text{ på andra kastet})) = (1/6 + 1/6) \cdot (1/6 + 1/6) = 2/6 \cdot 2/6 = 4/36 = 1/9$.

Svaret på den senaste frågan, om endast treor och fyror, går även att få fram på ett annat sätt: det finns fyra olika möjligheter att bara få treor och/eller fyror på två kast. Antingen enbart treor *eller* enbart fyror *eller* en trea på första kastet och sedan en fyra *eller* först en fyra och sedan en trea. Då får vi en lite längre beräkning, nämligen $(P(\text{tre} \text{ på första kastet}) \cdot P(\text{tre} \text{ på andra kastet})) + (P(\text{fyra på första kastet}) \cdot P(\text{fyra på andra kastet})) + (P(\text{tre} \text{ på första kastet}) \cdot P(\text{fyra på andra kastet})) + (P(\text{fyra på första kastet}) \cdot P(\text{tre} \text{ på andra kastet})) = (1/6 \cdot 1/6) + (1/6 \cdot 1/6) + (1/6 \cdot 1/6) + (1/6 \cdot 1/6) = 1/36 + 1/36 + 1/36 + 1/36 = 4/36 = 1/9$. Samma svar! Och det ska det bli också, eftersom det är samma sak vi har räknat ut.

Den sistnämnda metoden påminner en del om så kallade ”trädd”, som vi ska titta på nu.

Träd

När man singlar slant två gånger, för att ta ett enkelt exempel, så får man krona eller klave på första kastet, och krona eller klave på andra kastet. Antingen får man alltså {krona, krona} *eller* {krona, klave} *eller* {klave, krona} *eller* {klave, klave}. Man kan rita upp det som en sorts träd, där varje gren är ett utfall (i det här fallet krona eller klave).

För den som inte är så bevandrad inom numismatiken (vetenskapen om mynt), så är krona sidan med kungen på och klave sidan med en krona.

Skulle man skriva ut sannolikheten för varje gren vid alla förgreningar (i vårt krona-klave träd blir det 0,5 vid varje gren), så blir det tydligt hur man på ett enkelt sätt kan räkna ut sannolikheten för en speciell serie av krona och klave. Om vi med trädets hjälp vill räkna ut sannolikheten för att få krona, krona så är det bara att följa de grenar som innebär krona, krona och multiplicera alla sannolikheter man ”stöter på” på vägen, alltså $0,5 \cdot 0,5 = 0,25$.



Man kan även utveckla trädets grenar, som med fyra slantsinglingar, och räkna på det. Som en (i det här fallet kanske lite onödigt omständlig) kontroll att det verkar stämma kan man räkna ut sannolikheten för alla serier av utfall och addera sannolikheterna för alla utfall och kontrollera att det blir 1, så som det ska vara.

Ett ”sannolikhets-träd” behöver

inte beskriva just slantsinglingar,

det kan vara vad som helst,

bara man känner till

alla möjliga utfall och

sannolikheten för dessa.

Men finns det många

olika utfall och många

steg, så blir trädets

lätt väldigt rörigt och

svårt att både rita

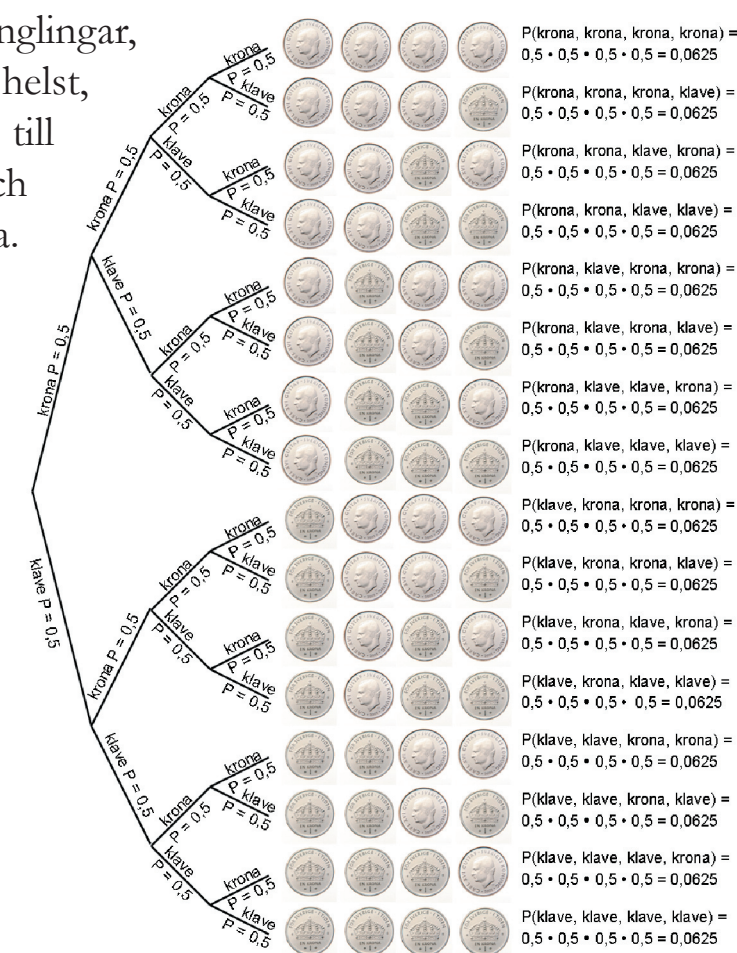
upp och hitta i. Men

även om man inte

ritar upp trädets, så

är det ett bra sätt att

tänka på.



Komplementhändelser

Vill vi räkna ut sannolikheten för att få *minst en* klave på fyra slantsinglingar, kan vi ta hjälp av det senaste trädet. I trädet motsvaras det av alla utfall utom det översta, som bara är kronor. Vi kan då räkna ut sannolikheten för att få antingen bara klavar (understa utfallsserien i figuren) *eller* tre klavar och en krona (näst understa) *eller* två klavar, en krona och en klave (nästnäst understa) och så vidare till och med den näst översta utfallsserien och summera dem. Det är en lite omständlig metod, men den fungerar och känns förmodligen intuitivt rätt.

Om vi vill göra det hela lite enklare kan vi använda oss av ett litet knep, som bygger på att summan av sannolikheterna för alla möjliga utfall alltid är 1. Man räknar helt enkelt ut sannolikheten för att något *inte ska* hända, och sen subtraherar vi det från 1 så får vi sannolikheten att det *ska* inträffa. I det här fallet räknar vi ut sannolikheten för att det inte är med någon klave i utfallsserien och sen subtraherar det från 1, så får vi alltså sannolikheten att det är med en klave. Enda utfallsserien som saknar klave är den med fyra kronor, och sannolikheten för den är $0,5 \cdot 0,5 \cdot 0,5 \cdot 0,5 = 0,0625$, alltså måste sannolikheten att få *minst en* klave på fyra slantsinglingar vara $1 - 0,0625 = 0,9375$.

Sannolikheten att något *inte ska* hända kallas för komplementhändelsen till att det ska hända. Att beräkna sannolikheten med hjälp av komplementhändelser är ett knep man kan ta till för att få enklare uträkningar, om utfallen där det efterfrågade händer är fler än de utfall där de efterfrågade *inte* händer. Det hela bygger alltså på att

$$P(\text{det efterfrågade händer}) + P(\text{det efterfrågade händer inte}) = 1.$$

Sannolikhetskalkyl var det

Det här är alltså det grundläggande inom sannolikhetsläran. Du kommer antagligen inte ihåg allting efter första gången, och man kan behöva läsa det igen och fundera över det. Men så småningom är jag övertygad om att det kommer att visa sig att det inte var så jobbigt det här, det handlar mest om sättet att tänka på. Nu är det dags för lite övningsuppgifter, så att du får fundera på egen hand.

Övningsuppgifter

1.

I ett lotteri med 100 lotter, där det finns ett förstapris och ett tröstpris, köper du 11 lotter. Samma lott kan inte vinna både förstapriset och tröstpriset, och ingen lott har större chans att vinna än någon annan.

a) Vad är sannolikheten att du vinner förstapris?

b) Du vann inte första pris, men efter att vinnarlotten är dragen drar de också tröstprislotten. Hur stor är chansen att du ska vinna det?

2.

a) Skriv med ord: " $P(\text{Färjestad vinner över Frölunda i nästa match}) = 0,6$ "

b) Skriv med formen $P(A)$: "Det är fifty-fifty om det är jag eller min kollega som kommer på besöket imorgon."



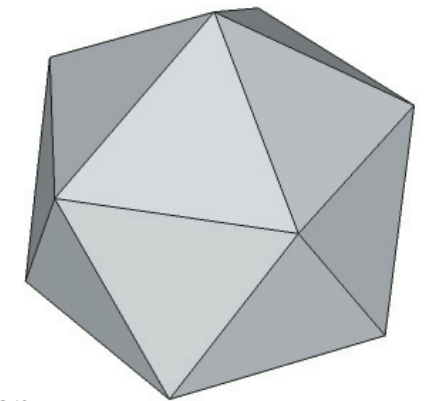
3.

a) Vad är sannolikheten att man ska få just serien {krona, klave, krona} om man singlar slant tre gånger?

b) Vad är sannolikheten att man ska få antingen serien {krona, klave, krona} eller {klave, krona, klave} om man singlar slant tre gånger?

4.

Vår vanliga sexsidiga tärning är en av de fem platonska kropparna, som är de enda kropparna som består av plana, identiska sidor. Att en kropp består av identiska sidor innebär att sannolikheten är jämnt fördelad mellan sidorna om man skulle kasta den som en tärning. Den vanliga tärningen utgörs av sex kvadrater, och är en kub. Av liksidiga trianglar går det att göra tre andra av de platonska kropparna: En tetraeder (en fyrsidig kropp, portionsmjölkförpackningar till kaffe, så kallade tetrapack, brukar ha den formen), en åttasidig oktaeder och en tjugosidig ikosaeder. Av pentagoner, femhörningar, kan man göra en tolvsidig dodekaeder, vilket är den femte platonska kroppen.



a) Vad är sannolikheten att få ett udda tal om man kastar en oktaeder, som är numrerad 1-8?

b) Det finns ingen femsidig platonsk kropp, men hur kan man göra för att skaffa sig en tärning som bygger på en platonsk kropp och där sannolikheten är jämnt fördelad mellan utfallen {etta}, {tvåa}, {trea}, {fyra} och {femma}?

5.

Vi vet att sannolikheten att hästen Thunderbolt vinner i första loppet är 0,4, och om han vinner där är sannolikheten att han vinner i även andra 0,5. Men skulle han förlora i första loppet, så är risken att han förlorar i det andra loppet 0,7.

- a) Rita upp ett sannolikhetsträd för Thunderbolt i de två loppen, och fyll i sannolikheterna för varje gren.
- b) Beräkna sannolikheten att Thunderbolt förlorar första loppet och vinner det andra.
- c) Beräkna sannolikheten att Thunderbolt vinner minst ett lopp.



3.

DET KLASSISKA GETPROBLEMET



Det finns ett klassiskt problem, som finns med i princip alla böcker och skrifter om sannolikhetslära. Här får det vara ett exempel på att det gäller att ta tillvara all information som finns tillgänglig, för att kunna göra ett så bra val som möjligt. Det kommer ursprungligen från en amerikansk TV-show, och kallas populärt för "getproblemet". Anledningen till att det är så populärt är förmodligen för att det strider mot intuitionen. Det är ett väldigt bra sätt att visa att det ibland går att säga mer om sannolikheterna än man först kan tro!

Så går det till

En tävlande tas upp på en scen, där det finns tre dörrar att välja mellan. Bakom två av dörrarna finns det en get (därav namnet), och bakom den tredje dörren finns det en bil. Detta vet den tävlande om, men hon har ingen aning om bakom vilken av dörrarna bilen finns, och inte då heller bakom vilka av dörrarna getterna finns. Lyckas den tävlande välja dörren med bilen bakom, så får hon den. Programledaren ber sedan den tävlande att välja den dörren där hon tror att bilen finns, och eftersom hon inte har någon aning om var den finns, blir detta val helt slumpmässigt.

Så långt var det inga konstigheter, och sannolikheten att den tävlande väljer rätt dörren (vi förutsätter att den tävlande hellre vill ha en bil än en get) är alltså $1/3$.

Nu kommer vi till klurigheten i det hela: Programledaren öppnar då en *annan* dörren än den som den tävlande har valt. Programledaren vet bakom vilken av dörrarna bilen finns, och öppnar aldrig den, utan öppnar istället en dörren med en get bakom. Efter det får den tävlande välja att byta dörren till den dörren som programledaren inte har öppnat, eller fortfarande hålla kvar vid sitt första val. Hur bör hon göra för att ha så stor sannolikhet som möjligt att få bilen?



Klurigheter

Efter att programledaren har öppnat en dörren som den tävlande inte har valt och som bilen inte finns bakom, så är det alltså en situation med två dörrar. Många tycker då det är självklart att sannolikheten blir en halv för vardera dörren, och att det alltså varken är bättre eller sämre att ändra sitt val av dörren. Det skulle förvisso vara sant, om vi skulle hoppa in i tävlingen med bara två dörrar kvar, utan att veta om vad som har hänt innan. Men nu är situationen inte sådan, programledarens taktik med dörröppningen ger faktiskt ledtrådar om bakom vilken av dörrarna bilen finns.

Det är inte alltid saker och ting ligger till som man kan förvänta sig, och emellanåt gäller det att tänka till en gång extra. Vi hoppar in i ett annat litet problem nu, innan vi ger oss på getproblemet igen: Du har en påse med tre mynt. Ett av dem är ett vanligt mynt, med en krona på ena sidan och en klave på den andra, men de andra två är falska på det sättet att det ena har kronor på båda sidorna och det andra klavar på båda sidorna. Du blundar, tar upp ett mynt och lägger det på bordet. När du öppnar ögonen ser du att det är en klave på det. Hur stor är sannolikheten att det är en krona på den andra sidan?

Gör en gissning själv – men som du säkert kan misstänka, så är svaret inte en halv, eftersom jag knappast hade tagit upp problemet nu i så fall.

När du tar upp ett av de tre mynten och lägger upp det på bordet så finns det sex möjliga utfall:

- A: Det äkta myntet hamnar med krona uppåt
- B: Det äkta myntet hamnar med klave uppåt
- C: Det ena falska myntet, det med bara kronor, hamnar med den ena kronan uppåt
- D: Det ena falska myntet, det med bara kronor, hamnar med den andra kronan uppåt
- E: Det andra falska myntet, det med bara klavar, hamnar med den ena klaven uppåt
- F: Det andra falska myntet, det med bara klavar, hamnar med den andra klaven uppåt



Sannolikheten måste rimligen vara lika stor för varje utfall, eftersom sannolikheten är lika stor för att välja respektive mynt ur påsen och jämnt fördelad mellan vilken av sidorna som hamnar uppåt när vi lägger upp myntet på bordet. Eftersom det är en klave som ligger uppåt måste det vara antingen B, E eller F. Av B, E och F är det bara i fallet B som det är en krona på andra sidan, i de andra fallen är det ju klavar på baksidorna. Alltså måste sannolikheten för att det ska vara en krona på andra sidan myntet vara just $1/3$.

Förhoppningsvis har du nu fått upp ögonen för att man kan behöva resonera på andra sätt än vad man från början tänkte för att få rätt svar. Nu återvänder vi till getproblemet.

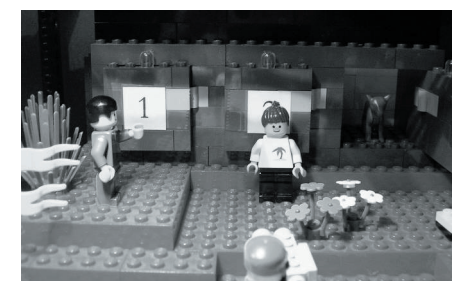
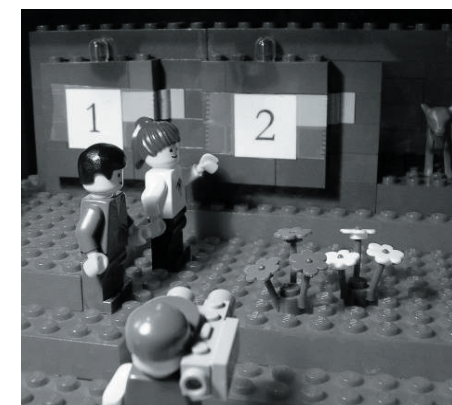
Tusen dörrar istället för tre

Vi tänker oss en lite modifierad version: Istället för tre dörrar har den tävlande tusen dörrar att välja mellan. Bakom 999 av dessa står det getter, och bakom en står det en bil. Den tävlande väljer en dörr, och programledaren öppnar 998 dörrar med getter bakom. Ska den tävlande byta dörr?

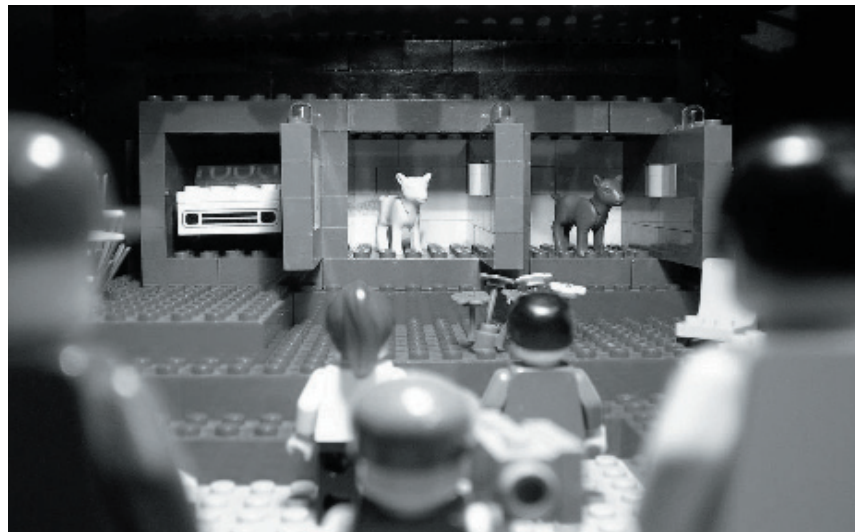
Sannolikheten att den tävlande valde rätt från början är försvinnande liten – bara $1/1000$. Sannolikheten att bilen finns bakom *någon* av de andra dörrarna är alltså $999/1000$. När programledaren senare öppnar de 998 dörrarna, skulle han kunna säga ”*om* bilen finns bakom *någon* av dessa 999 dörrar du inte har valt, så är det bakom den dörr jag *inte* öppnar nu”. Sannolikheten att den tävlande valde rätt dörr från början är fortfarande $1/1000$ (bilen har inte flyttat på sig!), och eftersom det nu bara finns en annan dörr att välja på så är sannolikheten att hon hittar bilen bakom den $999/1000$.

Rådet man ska ge henne blir i det här fallet tydligt: byt!

Samma princip gäller för tre dörrar, men med sannolikheten att hon väljer rätt från början $1/3$. Det programledaren gör är alltså att visa att ”*om* bilen finns bakom *någon* av dessa två dörrar du inte har valt, så är det bakom den jag *inte* öppnar nu”. Sannolikheten att bilen finns bakom den dörren programledaren inte öppnar är därmed $2/3$.



Känns det fel? Du är inte ensam, det sägs till och med att en och annan professor till en början hade svårt att acceptera det korrekta svaret. Men svaret är likväl att den tävlande bör byta dörr! Sannolikheten att bilen finns bakom den andra dörren är större än att den finns bakom den dörren hon valde från början.



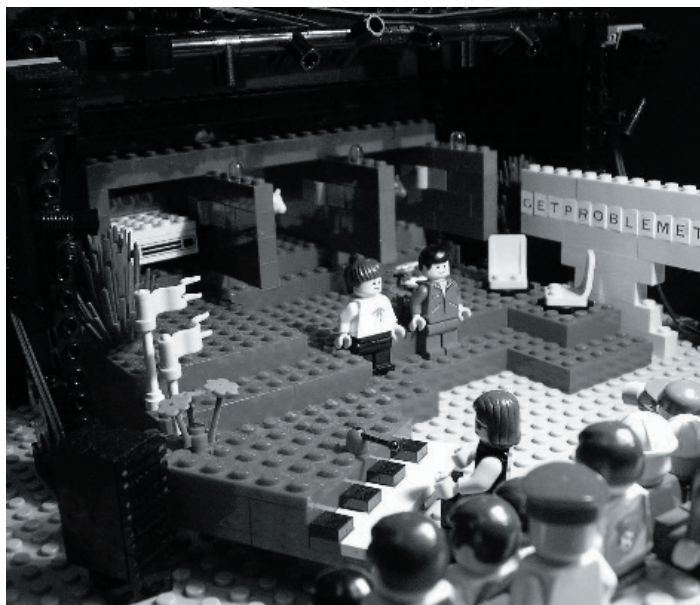
Övningsuppgifter

1.

Roa dig med att göra en statistisk undersökning på egen hand! Låt vänner välja mellan tre olika burkar, två tomma och en med en godisbit i. Ge dem samma förutsättningar som i TV-showen: De väljer en, du öppnar en tom som de inte har valt och låter dem byta till den andra burken om de vill. Se till att föra statistik över hur ofta de byter och får något godis, respektive hur många gånger de inte byter och får något godis. Du kommer förmodligen rätt snart märka att de får godiset oftare om de byter, precis som teorin förutsäger!

2.

Vi förflyttar oss till en gammal klassisk monarki, där tronföljden ärvs av den äldste sonen. Vi är på besök, och får höra att kungen har ett, och endast ett, syskon. Nyfikna som vi är funderar vi på sannolikheten att det är en kvinna, och kommer fram till – ja, hur stor är sannolikheten att det är en kvinna som är kungens enda syskon i denna monarki?



4.

EN SLUMP?



Så här i början, innan vi räknar alltför mycket på sannolikheter, ska vi fundera på begreppen slump och sannolikhet. Vad innebär egentligen begreppet sannolikhet och vad är det vi i själva verket kallar för slumpen?

Vaddå $P = 0,5$?

Vad betyder det egentligen att sannolikheten för att någonting ska inträffa är 0,5 eller 50%?

Man kan tänka sig att vi har en platta med en väl balanserad visare, som det går att snurra på (ungefär som en kompass, men

utan det magnetiska). Hälften av plattan är gul, och andra hälften är blå, och gränsen mellan färgerna går på en linje genom den punkt som visaren roterar kring. Vi sätter fart på visaren.

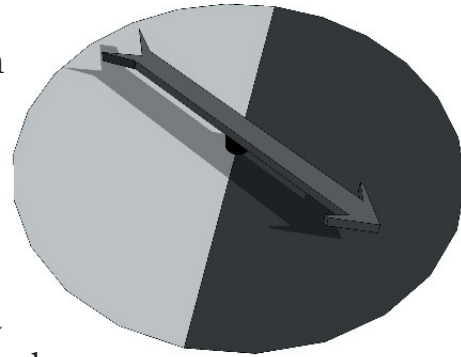
Sannolikheten att visaren ska stanna på det blå området är nu en halv, eller 0,5, och

detsamma för att den ska stanna på det gula området. Det är vad sannolikheten 0,5 innebär. (Risken finns att visaren stannar på linjen precis mitt emellan det gula och blåa, men den risken är försvinnande liten – även om den stannar väldigt nära, så pekar den antagligen pyttelite mer åt ena färgen än den andra.)

Ibland utförs någonting endast *en* gång. Då inträffar antingen det ena *eller* det andra, och inte hälften av båda. Kan sannolikheten verkligen vara något annat än 1 eller 0 om någonting bara utförs en enda gång?

Jodå, för även om snurran bara snurras *en enda* gång, och den gången stannar på det blåa, så kan vi fortfarande se att sannolikheten att den stannade där var just 0,5. Den kunde faktiskt lika gärna ha stannat på det gula, även om den nu inte gjorde det just denna enda gång.

Det är inte ofta man i verkligheten kan se sannolikheterna så tydligt som i den här modellen, men det är en bra modell som illustrerar vad sannolikheten 0,5 innebär.



Subjektiva sannolikheter

Ibland använder man sannolikheter för att beskriva exempelvis chanserna för att en affär går igenom eller risken för att få ett nej. Då handlar det oftast inte om den typen av matematiskt mätbara sannolikheter som tärningar och snurror, utan om en sorts subjektiva sannolikheter, någons sorts upplevelse av förutsättningarna inför något, men det går alldeles utmärkt att räkna med dem på samma sätt som med tärningar.

Om man upplever sannolikheten att en affär går igenom hos part A är 0,60 och hos part B 0,30, och att utfallet hos part A inte påverkar part B och tvärtom, så blir sannolikheten att affären går igenom hos part A *och* part B $0,60 \cdot 0,30 = 0,18$. Precis som med alla andra sannolikheter, bara att man här har kommit fram till 0,60 och 0,30 genom subjektiva bedömningar. Och för att 0,18 ska stämma, så måste förstås 0,30 och 0,60 vara riktiga.



När är det slumpen?

Så här säger Nationalencyklopedin om ”Slump”: ”Inom sannolikhetsteorin benämning på det oförutsägbara. Man säger att ett försök är ett slumpförsök om det vid upprepningar ger olika resultat. (...)”

Det handlar alltså om oförutsägbarhet. Men vad är oförutsägbart och vad är inte det? Om det ligger en tärning på bordet med en fyra uppåt, och du puttar lite grann på den i sidled så kommer det att fortsätta vara en fyra. Men om du lyfter tärningen en centimeter över bordet och släpper den försiktigt? Då kommer den *förmodligen* att landa med fyran uppåt, men det beror på vilken vinkel den träffade bordet med. Om du däremot tar och kastar tärningen genom halva rummet så har du faktiskt ingen aning om vilken sida den kommer att ha uppåt när den landat. Men är det slumpen i det fallet och inte de andra? Och var går i så fall gränsen?

Att kasta en tärning är ingen *riktig* slump, det går alltid att räkna teoretiskt med fysikaliska lagar och komma fram till hur tärningen kommer att landa, det blir bara svårare och komplexare desto större avstånd och fler krafter som är inblandade. För att kunna förutsäga hur tärningen ska landa vid ett vanligt tärningskast behöver man ta hänsyn till en stor mängd parametrar, till exempel exakt vilken position tärningen har när den lämnar handen och vilken hastighet och rotationshastighet den har. Man skulle även behöva ta hänsyn till om det kommer någon vindpust som kan påverka tärningen, exakt hur hörnen ser ut på tärningen om den skulle studsas och exakt vilken smuts som ligger på bordet och kan påverka tärningen i en studs, och så vidare och så vidare.

Så nej – ett tärningskast kan egentligen inte betecknas som ren slump, eftersom det teoretiskt skulle kunna gå att räkna på hur den kommer att landa. Och går det att räkna ut, så skulle man i princip kunna göra sådana kast att det alltid blir det önskade utfallet, om man känner till precis hur man ska kasta med de gällande förutsättningarna. Och en viss tärningskastare kanske alltid har samma sida upp på tärningen varje gång han kastar den

och gör ungefär samma rörelse med handen, och då skulle det kunna innebära att det oftare blir ett särskilt utfall. Men eftersom det är så oerhört komplext att räkna på det (och man förmodligen ändå skulle missa någon parameter så att man inte får rätt svar ändå) så kan vi, när allt kommer omkring, ändå använda tärningen med gott samvete som slump och anse att det trots allt är ett slumpmässigt utfall vi får. Till våra sätt att använda tärningskasten i den här boken duger det jättebra. Vi skapar oss helt enkelt en modell – vi bestämmer att utfallet på ett tärningskast är (tillräckligt) oförutsägbart, och vi kallar det för slumpen, trots att man (om man vill vara petig) med rätta kan hävda att det inte är någon riktig slump.

Finns det någon riktig slump?

Nu kommer vi in på ett ämne som inte är särskilt väsentligt för resten av boken, men det kan vara kul att känna till ändå. Eftersom tärningskast alltså inte är en riktig slump, då de teoretiskt går att förutsäga, diskvalificeras också andra liknande tillvägagångssätt som till exempel myntsinglingar. Även rouletthjul och andra roterande föremål, som modellen i början av kapitlet, blir diskvalificerade eftersom man *skulle* kunna räkna ut precis hur stor kraft man behöver ta i med för att det ska stanna på ett särskilt ställe. Det är oerhört svårt att verkligen lyckas med det, men tanken på att det *skulle* kunna gå är tillräcklig i det här fallet för att inte kunna anse det som en äkta slump.

Det finns sätt att skaffa sig riktiga slumpmässiga utfall på, som fysiker anser vara riktig slump och som fysikaliskt sett inte går att

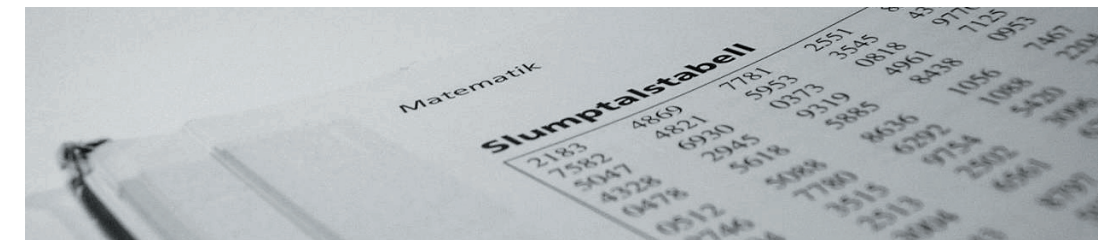
räkna på. Radioaktivt sönderfall är ett sådant fenomen som anses slumpmässigt. Det handlar om instabila atomer som sönderfaller till andra atomer, och ger ifrån sig joniserande strålning. En atom kan nämligen sönderfalla om tre sekunder eller tre miljoner år – det går inte att förutsäga, och därför går den att använda som en slumpalsgenerator. Mer om radioaktivitet kommer i nästa kapitel.

En annan typ av slump finns även i Werner Heisenbergs obestämbarhetsrelation, som går ut på att man inte kan bestämma både en elektrons riktning *och* hastighet, oavsett precisionen på mätinstrumentet, eftersom man så fort man observerar en enskild elektron också påverkar den. Vi kan alltså inte veta hastigheten *och* läget hos en enskild elektron (eller foton eller motsvarande partikel), vi kan bara prata om olika sannolikheter för var den befinner sig. Så i dessa elementarpartiklar finns det också en slump man inte kan räkna på. ”Gud spelar inte tärning” lär fysikern Albert Einstein ha fällt som kommentar om detta, då det inte alls stämde med hans bild av ett förutsägbart Universum. Men idag är denna centrala del av kvantmekaniken allmänt accepterad av fysiker.

Slumptalsgeneratorer i räknare och datorer ger däremot inga riktiga slumpal på samma sätt som det radioaktiva sönderfallet. I en dator finns det inget sådant fysikaliskt fenomen som skulle behövas för det, utan slumptalsgenereringen får skötas av någon sorts program, och de går inte att göra *riktigt* slumpmässiga. Men de duger – på samma sätt som tärningarna – att till praktiska ändamål använda som slump.

Om man inte känner för att ödsla tid på att slå med tärningar men ändå vill ha slumpal att utgå från för att göra några

beräkningar, eller något annat man behöver många slumpal till, kan man titta i en slumpalstabell (som man ofta kan hitta i matematiska formelsamlingar). Där kan man hitta långa serier med slumpal, som ofta är uppställda efter sådana fysikaliskt riktiga slumpalsprinciper, som exempelvis radioaktivt sönderfall.



Matematik									
Slumptalstabell									
2183	4869	7181	2551	4	971	7125			
7382	4821	5953	3545	0818	4901	0953			
5047	6930	0573	8438	1056	1089	5420			
4328	2345	5885	8036	6592	9754	2503			
0478	5618	5088	7780	3515	2513	6681			
0512	7744	3515	2513	6681	4799				

Definitionen av slump är inte lätt, och det finns flera olika synsätt på vad som egentligen bör anses som slump. För oss duger i den här boken dock en definition som bygger på att vi inte har någon rimlig förutsättning att kunna förutsäga utfallet. Det skulle då innefatta tärningar som man kastar så att de snurrar runt i luften (och inte bara släpps rakt ner på bordet) och andra liknande metoder, som myntsinglingar, rouletthjul och väl blandade kortlekar.

Övningsuppgifter

1.

I vilket/vilka av följande fall kan vi till vardags anse att vi har med slumpen att göra, det vill säga att vi inte har någon praktiskt rimlig förutsättning att kunna förutsäga utfallet?

- a) Vädret imorgon i stora drag
- b) Vädret om exakt ett år



- c) Vilken färg man får på en seg råtta om man sticker ner handen i en godispåse med bara sega råttor av olika färger och tar upp en utan att titta
- d) Om man vinner på bingo
- e) Om GAIS vinner sin nästa fotbollsmatch
- f) Om det är något stvafel i den här boken
- g) Vilket kort som ligger nederst i en väl blandad kortlek

5.

DET LÅNGA LOPPET



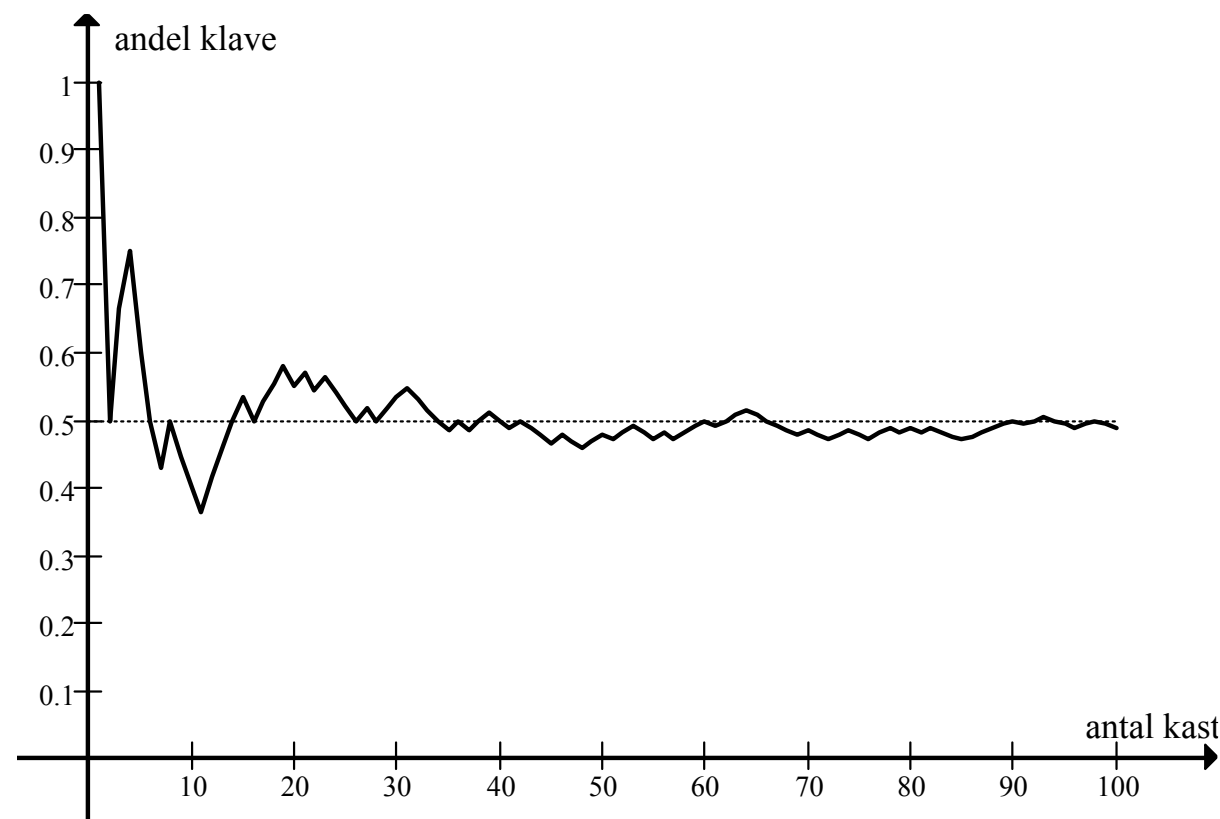
Det här kapitlet handlar om Stora talens lag, som innebär samma sak som det mer vardagliga uttrycket ”det jämnar ut sig i det långa loppet”. I slutet av kapitlet tar vi även upp lite grann om vad modeller är för något.

Slår man upp ”Stora talens lag” i Nationalencyklopedin står det så här: ”Sats i sannolikheteorin, som säger att om man gör ett stort antal oberoende observationer av en slumpvariabel så kommer dess aritmetiska medelvärde med stor sannolikhet att ligga nära variabelns väntevärde (om detta existerar).” Det Nationalencyklopedin säger är alltså att om man singlar slant 1000 gånger så kommer *ungefär* hälften av singlarerna att visa klavar och hälften kronor, alltså någonstans runt 500 klavar och 500 kronor, om sannolikheten att det ska bli krona respektive klave är 0,5.

Väntevärde

Variabelns väntevärde är ungefär vad det låter som – det är vad vi kan förvänta oss, vilket också är samma sak som sannolikheten för någonting. För en tärning kan man teoretiskt beräkna väntevärdet genom att konstatera att den är helt regelbunden och homogen och att det inte finns något i fysiken kring den som talar för att den skulle landa på en speciell sida. Alltså är det teoretiska väntevärdet för en speciell sida på en tärning $1/6$, och det är precis det vi kallat för sannolikheten, eller $P(\text{en särskild sida på en 6-sidig tärning}) = 1/6$.

Vi kan illustrera att man singlar slant 100 gånger med en graf. Efter varje kast räknar vi ut hur stor andel klave vi har fått av det totala antalet kast som är gjorda. Enligt Stora talens lag ska vi så



småningom närma oss den horisontella streckade linjen på 0,5 (vårt väntevärde, eller sannolikhet, för klave vid myntsingling).

Som synes så kan man konstatera att uttrycket ”det jämnar ut sig i det långa loppet” verkar stämma bra; vi kan se hur kurvan för andelen av de erhållna klavarna vid slantsinglingen stabiliserar sig kring linjen för den förväntade andelen. Efter de första tio kasten är det $4/10$ andel klave, det vill säga fyra klavar och sex kronor, en skillnad på två. Vid 85 kast skiljer det fem, mäter man noggrant i grafen kan man nämligen räkna ut att det är 40 klavar jämfört med 45 kronor, alltså $40/85$ andelar klave. Poängen här är att det skiljer ett *större antal* mellan klavarna och kronorna, men skillnaden mätt i *andelar* är mindre, för $4/10$ är längre ifrån det förväntade 0,5 än vad $40/85$ är.

Det är alltså viktigt att tänka på att det är *andelen* klavar som närmar sig väntevärdet, *inte antalet*!

Radioaktivitet

I förra kapitlet talades det om att radioaktivt sönderfall var slumpmässigt. Samtidigt så finns det väl bestämda halveringstider för så kallade radioaktiva isotoper, vilket kan tyckas vara en motsägelse.

Vissa isotoper (atomvarianter, samma sorts atom kan finnas som olika sorts isotoper) är nämligen instabila, och faller förr eller senare sönder till andra sorters isotoper eller andra sorters atomer och sänder då ut joniserande strålning, som slarvigt kallas för radioaktiv strålning. De flesta atomer faller sönder i början, och

processen avtar allt efter hand som det blir färre och färre atomer av just det här slaget kvar. Ett vanligt sätt att mäta hastigheten på sönderfallet är att ange hur lång tid det tar tills hälften av en viss mängd har sönderfallit. Det fiffiga med det hela är att först sönderfaller *hälften* av atomerna på en viss tid, och sedan tar det lika lång tid för *hälften* av det som kvarstår att sönderfalla:



I grafen visas sönderfallet för en tänkt isotop med en halveringstid på sex år: Efter sex år är bara hälften av alla ursprungliga atomer kvar, och efter 12 år är hälften av hälften, alltså en fjärdedel, av alla ursprungliga atomer kvar, och så vidare. Kontrollera själv!

Det finns en isotop som kallas för kol-14, eller ^{14}C som kemisterna föredrar att skriva den. Den brukar användas för att åldersbestämma arkeologiska fynd. Kol-14 nybildas i atmosfären hela tiden, och då det hela tiden finns kol-14 i luften hamnar en viss del så småningom i växter, som bygger upp sig själva av

koldioxiden i luften, och sedan i djur, som äter växter. Men kol-14-isotopen är instabil, och sönderfaller med en halveringstid på 5730 år till en stabil kväve-isotop. När ett djur har dött så äter det inte längre några nya växter, och kol-14-halten sjunker i djuret så småningom, allt eftersom tiden går och allt fler kol-14-atomer har sönderfallit.

Genom att mäta andelen kol-14 och jämföra det med andelen kol-12 i ett arkeologiskt prov kan man bestämma provets ålder. Tack vare att man kan gissa sig till den ursprungliga mängden kol-14 (då man, bland annat, vet hur stor andel som är kol-12 respektive kol-14 i nu levande växter och djur), kan man tala om hur länge provet har fått ligga för att rätt mängd kol-14 ska ha hunnit sönderfalla.

Radioaktivt sönderfall sker slumpmässigt. Hur kan man då veta hur lång tid det tar tills hälften av allt kol-14 i ett prov har sönderfallit? Man kan aldrig säkert *veta*, men tack vare Stora talens lag så kan man ändå göra bra beräkningar som med stor sannolikhet stämmer. Det är nämligen så att alla isotoper inte är lika instabila, det går att hitta *medelvärden* för hur snabbt en viss sorts isotop sönderfaller, och för kol-14 visar sig detta vara just 5730 år. Metoden bygger alltså på att atomer är så otroligt små; i några hundra gram trä från idag kan det röra sig om några biljarder kol-14-atomer. Men vi kan aldrig tala om när en enskild atom kommer att falla sönder, det kan vara om en sekund eller tre miljoner år. Vi har ingen aning, vi känner bara till ett medelvärde, som inte alls går att använda på en enskild atom!

Att tänka på...

Det går inte att bestämma några exakta antal med hjälp av Stora talens lag. Formuleringen ”med stor sannolikhet att ligga nära” från Nationalencyklopedin betyder att om vi singlar slant 1000 gånger så kan vi inte säga att det bestämt ska bli 500 stycken kronor och lika många klavar. Det är inte ens särskilt troligt att det ska bli precis 500 stycken. Men däremot är det rimligt att tro att det blir *i närheten* av 500 kronor.

Något man lätt kan glömma bort är att ett mynt eller en tärning inte har något minne. Om man har fått fem stycken klavar i rad, så är sannolikheten för att vi ska få en sjätte 0,5. Det är inte så stor sannolikhet att få sex klavar i rad, men sannolikheten att däremot få den sjätte klaven när man redan har fem är som vanligt 0,5. Det är alltså skillnad på att få *sex klavar i rad* och att få den *sjätte klaven* i rad när man redan har fått fem.

Det finns en ytterliggare aspekt på det. Tänk dig att det är en tävling och man singlar slant med varandra och det gäller att få så många klavar som möjligt. Om en av spelarna leder, så finns det ingenting som säger att den andre har större sannolikhet att komma i kapp bara för att hans mynt ”ligger efter” med klavar. Stora talens lag säger att de två tävlandena i det långa loppet kommer att få *ungefär* lika många klavar efter många kast (om båda spelar med schysta mynt). Men om en av de tävlande leder med tio klavar, så kan man ju lika gärna ”börja tänka på” Stora talens lag precis då, och tänka att ”från och med *nu* kommer vi att få ungefär lika stor andel klavar som



kronor, och jag har ju redan tio klavar”. På så sätt kan man komma fram till att spelaren som ledde mycket möjligt kan fortsätta leda hela tiden och vinna. Stora talens lag handlar om att det blir hyfsat jämnt i det långa loppet, men akta dig för att använda lagen fel!

Om man inte har ett väntevärde då?

Men det är inte alltid vi kan förutsäga sannolikheten, och därmed väntevärdet, på allting (därav Nationalencyklopedins klausul ”om detta existerar”). Om man kastar en plastmugg många gånger så har vi ingen aning om vad vi kan förvänta oss för fördelning mellan de tre alternativen som finns; den landar ståendes upprätt, upp-och-ner eller på sidan (såvida ingen har gjort någon undersökning tidigare och talat om resultatet för oss). Vi kan göra en rätt bra intuitiv gissning att den mestadels kommer att landa på sidan, men hur ofta? Hälften av gångerna? Nio av tio gånger?

För att kunna svara på det får vi ta hjälp av Stora talens lag, och utnyttja den ”baktänkes”, det vill säga vi kastar en plastmugg ett någorlunda stort antal gånger, exempelvis 200, och noterar vad vi får för resultat. I min egna undersökning kastade jag en plastmugg 200 gånger och fick följande resultat:

	efter 200 kast	andel av alla kast
upprätt	29	0,145
upp-och-ner	11	0,055
på sidan	160	0,80

Naturligtvis gäller de här siffrorna bara för just den sortens mugg jag kastade, hade det varit en lägre och bredare modell hade muggen antagligen oftare landat upprätt och upp-och-ner, och vice versa om den hade varit högre och smalare. Vi *vet* inte säkert att sannolikheten för att en mugg av den här typen ska hamna på sidan är 0,80, men vi har, enligt Stora talens lag, anledning att tro att det ligger i närheten av det. Vill vi ha ett bättre värde får vi kasta flera gånger, och desto fler kast vi gör desto troligare är det att vi närmar oss den teoretiska sannolikheten. Men vi kan aldrig bli riktigt säkra – det går alltid att göra ett kast till och få ett lite bättre värde.

Är 200 • 1 samma sak som 20 • 10?

Det ska erkännas att jag i min undersökning med plastmuggen inte kastade en mugg 200 gånger. Jag hade istället tio likadana muggar och kastade dem tjugo gånger. Men blir det samma sak?

Om alla muggar var identiska, så skulle vi få samma resultat. Förmodligen är inte plastmuggarna helt identiska, men tillräckligt för att det ska duga för oss här och nu. På samma sätt kan man göra om man ska göra en stor mängd tärningskast – är det inte för någon spets-forskning om tärningar eller liknande, så duger det gott att använda flera likadana tärningar och kasta samtidigt, istället för att göra alla kast med samma tärning.

Att man faktiskt kan vara bekväm av sig och använda flera tärningar (eller plastmuggar) om man vill göra många kast är tack vare att de inte har något minne. Hade det däremot varit frågan

om en nonstop-påse, där man äter upp nonstopet när man tagit upp det, så hade det inte fungerat. Nonstop-påsar har minne – om vi tänker oss att det finns 8 gröna bitar, av sammanlagt 25, så är sannolikheten att ta upp en grön 8/25. Men har man tagit upp en grön och ätit upp den, så är sannolikheten för nästa gröna 7/24, alltså mindre. Här påverkar våra tidigare utfall kommande utfall, de är vad som brukar kallas för beroende. Vi kommer att titta mer på sådana situationer längre fram, i kapitel nio.

Modeller

Nu är det inte alltid så intressant att veta hur muggar hamnar. Nej, ärligt talat, den enda gången man kan vilja veta det i praktiska livet är nog om man tappar en mugg med dricka i, fast då lär allt rinna ur på ett eller annat sätt ändå, och då spelar det trots allt inte så stor roll hur muggen till slut landar.

Att jag använde mig just av muggar i det ovanstående exemplet var för att det är en bra *modell* att illustrera hur man kan räkna ut en ungefärlig sannolikhet med hjälp av försök. Modeller är något man ofta använder sig av inom matematiken (och många andra vetenskaper) för att förklara verkligheten. I förra kapitlet använde jag mig också av en modell, en snurra, för att visa vad sannolikheten 0,5 innebär klart och tydligt.

Det är inte alltid det är så lätt att ta en vardagssituation och börja räkna på den – de kan vara väldigt komplexa. För att till exempel räkna ut hur stor sannolikheten är att det inte ska vara någon kö i mataffären vid ett visst klockslag får man ta hänsyn till många olika faktorer, bland annat hur dags alla slutar jobbet, hur

ofta folk handlar, hur välbesökt mataffären är över huvud taget, om det finns något särskilt som lockar/skrämmer kunder just denna dag och så vidare och så vidare. Du kan säkert komma på fler och bättre exempel.

För att få förståelsen för de grundläggande principerna får vi applicera dem på enkla situationer, modeller, för att sedan plocka ihop fler modeller och göra en mer komplex modell av verkligheten om det är det vi vill ha. Ibland kanske en enkel modell stämmer överens med verkligheten nästan lika bra som en komplex, och då kan man föredra den enklare, dels för att den är lättare att räkna med, men framförallt lättare att förstå.

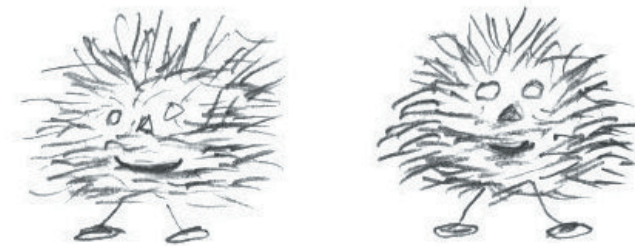
En enkel modell kan också vara en bra början om man vill undersöka något nytt. Beroende på vilket resultat man får av den enkla modellen (om det exempelvis var så intressant som man hade hoppats på) kan man välja om man vill gå vidare och lägga tid och kraft på att utveckla en mer avancerad modell, som ger ett noggrannare resultat, eller nöja sig med resultatet man har och försöka hitta något annat intressant istället.

Det långa loppet

Sammanfattningsvis kan vi alltså konstatera att det jämnar ut sig i det långa loppet, men inte förr! Det spelar ingen roll hur exakt vi har definierat vårt väntevärde, i det enskilda utfallet har vi ingen möjlighet att veta vad vi kan förvänta oss.

6.

NÄR MAN TALAR OM TROLLEN och några andra talesätt



“När man talar om trollen så står de i farstun” är ett gammalt talesätt. Men finns det något vetenskapligt som ligger bakom det, och andra liknande talesätt och uppfattningar? Vi ska titta på lite olika situationer och fundera på hur det egentligen ligger till.

”Telefonen ringer alltid när jag står i duschen”

Du kanske tycker att (det åtminstone känns som om) telefonen alltid ringer så fort du är i duschen? Eller att din kompis alltid brukar ringa så fort du är på väg att ringa honom? Eller att

du alltid möter grannen när du går i trapphuset? Att det alltid passerar ett tåg på järnvägsbron när du åker under den? Eller att trollen alltid kommer in i farstun när man talar om dem? Idén är att man upplever att något ”alltid” inträffar samtidigt som något annat, trots att det inte verkar finnas den minsta anledning till det. Vi ska titta och räkna överskådligt på just ”Telefonen ringer alltid när jag är i duschen”-syndromet, för att få en uppfattning om hur det egentligen ligger till, sett från en lite mer objektiv nivå.

För att räkna behöver vi göra lite antaganden om antal och tider. Låt oss säga att telefonen ringer i genomsnitt två gånger per dygn. Vi antar också att du har så snälla vänner att de inte ringer varken när du sover eller jobbar eller är i skolan, utan endast mellan klockan 16 och 22 på dygnet. Det innebär att de ringer någon gång under loppet av sex timmar. Varje gång telefonen ringer har du en minut på dig att svara från första signalen, vilket betyder att det i genomsnitt finns två minuter på sex timmar som telefonen ringer, alltså blir sannolikheten att telefonen ringer en särskild minut eller $P(\text{telefonen ringer denna minut}) = 2/(6 \cdot 60) = 1/180$.

Vi antar att du duschar en kvart per dygn, och detta gör du också någon gång mellan 16 och 22 (alltså samma tider som telefonen kan ringa). Då finns det femton minuter du duschar, av sex timmar. Sannolikheten att du ska duscha en särskild minut mellan 16 och 22 blir $15/(6 \cdot 60)$ eller $P(\text{du står i duschen denna minut}) = 1/24$.

Nu får vi ta till multiplikations-satsen för att räkna ut sannolikheten att telefonen ska ringa *och* att du står i duschen på en särskild minut. $P(\text{du står i duschen denna minut och$



telefonen ringer denna minut) = $P(\text{Du står i duschen denna minut}) \cdot P(\text{Telefonen ringer denna minut}) = 1/180 \cdot 1/24 = 1/4320$. Det är inte någon särskilt stor sannolikhet, men vi är inte färdiga än.

Än så länge har det bara handlat om en enda minut, men på ett dygn finns det ju sex timmar det skulle kunna inträffa på.

För att göra det så enkelt som möjligt för oss, tar vi till knepet med komplementhändelser och räknar ut sannolikheten att telefonen *inte* ringer samtidigt som du duschar någon gång under kvällen. Anledningen till att vi gör det är att vi annars först skulle behöva räkna på sannolikheten att telefonen ringer och du är i duschen den första minuten, och aldrig mer under kvällen, sedan sannolikheten att telefonen ringer och du är i duschen samtidigt den andra minuten och aldrig mer, och så vidare. Dessutom skulle vi också behöva räkna på sannolikheten att telefonen ringer två gånger när du duschar, och kanske till och med tre. Det är mycket enklare att räkna ut sannolikheten för att telefonen inte ringer alls när du är i duschen, och sen subtrahera den från 1, eftersom vi då får sannolikheten för att telefonen ska ringa minst en gång då du är i duschen under kvällen.

Först får vi ta fram sannolikheten för $P(\text{telefonen ringer inte någon gång på hela kvällen samtidigt som du duschar})$. Men vad är då $P(\text{telefonen ringer inte någon gång på hela kvällen samtidigt som du duschar})$? Hela kvällen innebär 360 minuter, och sannolikheten att telefonen *inte* ringer när du duschar en särskild minut blir $1 - 1/4320 = 4319/4320$ (komplementhändelsen igen!).

$P(\text{telefonen ringer inte någon gång på hela kvällen samtidigt som du duschar}) = P(\text{telefonen ringer inte samtidigt som du duschar, första minuten efter klockan 16}) \cdot P(\text{telefonen ringer inte samtidigt som du duschar, andra minuten efter klockan 16}) \cdot \dots \cdot$

$P(\text{telefonen ringer inte samtidigt som du duschar, minuten innan klockan 22}) = \left(\frac{4319}{4320}\right)^{360} \approx 0,92$, enligt multiplikationssatsen.

Då är det bara att ta $1 - P(\text{telefonen ringer inte någon gång på hela kvällen samtidigt som du duschar}) \approx 1 - 0,92 = 0,08 \approx 1/12$. Sannolikheten att du någon gång under de sex aktuella timmarna på dygnet ska stå i duschen samtidigt som telefonen ringer är alltså $1/12$, vilket innebär att i genomsnitt var tolfte kväll inträffar detta, om nu mina antaganden om tider och telefonsamtal stämmer.

Var tolfte kväll är knappast varenda, och inte heller varannan kväll. Finns det då någon anledning till att tanken kan ha uppkommit om att det händer ”varje kväll”? Jodå. Vi skulle kunna tänka oss följande situation: Du är extra stressad en dag, och detta är en dag då telefonen ringer när du är i duschen. Du tänker förmodligen extra mycket på det och irriterar dig just därför. Det är ingen omöjlighet att telefonen ringer igen när du duschar dagen efter, och då lär du notera det och tycka att det händer ”alltid nu för tiden”; det hände ju igår också. Men i bakhuvudet är du förmodligen medveten om att du med ”alltid” inte menar precis varenda dag, utan bara ofta.

Varken duschen eller telefonen har något ”minne”, bara för att telefonen ringde när du duschade igår betyder det inte att sannolikheten är mindre att den gör det idag. Så trots att det spontant kan verka otroligt, så är det trots allt inget oväntat om telefonen åter igen ringer när du duschar, bara ett par dagar efter de två dagarna i rad ovan. ”Tre gånger på fem dagar” känns förmodligen väldigt nära ”alltid”. Om det sedan går ett par veckor innan det inträffar nästa gång, så lär det inte minska din övertygelse så mycket – eftersom du duschar dagligen tänker du antagligen inte ”idag ringde inte telefonen”, det är bara när telefonen ringer

som du reagerar över det, även om det dröjer ett par veckor till nästa gång: ”Nu hände det *igen*!”.

Räknar du gångerna du duschar och hur många gånger av dem telefonen ringer under en lång tid, i storleksordningen ett år eller mer, kommer du med stor sannolikhet att hamna vid $1/12$ av dagarna. Det är nämligen så att det här med telefonen och duschen (och trollen i farstun) till syvende och sist i de allra flesta fall handlar om någon sorts psykologi (och regelbundna vanor). När du talar om trollen, men de *inte* står i farstun tänker du inte på det – det är bara när du talar om trollen och de verkligen står i farstun som du reflekterar över det!

Katter och smörgåsar

Hur är det då med att katter alltid landar med benen nedåt, och ostsmörgåsar alltid med osten mot golvet? Ligger det någon sanning i något av de påståendena då, eller är det samma fenomen som med telefonen och duschen, att det bara är i vissa fall man reflekterar över det?

Att katter oftast landar på fötterna är känt, redan 1894 efterlyste den franska vetenskapsakademien ”en fysikalisk förklaring till hur en katt vid fall från inte alltför låg höjd lyckas att alltid hamna på fötterna”. Forskningen har senare visat att de med hjälp av sin rörlighet i kroppen vänder sig i fallet om de råkar hamna upp-och-ner. De utnyttjar det så kallade rörelsemängdsmomentet, samma



fysikaliska princip som en konståkare på skridskor utnyttjar då de tar fart in i en piruett med armarna utsträcka och snurrar runt relativt sakta, och sedan drar in armarna till kroppen och får en betydligt högre rotationshastighet. Katten å sin sida drar in frambenen och vrider runt dem ett halvt varv så att de pekar nedåt (om den nu är upp-och-ner, alltså), medan den har bakbenen utsträckta för att hålla sig stabil. Sedan faller den ut frambenen så att de stabiliserar, medan katten drar in bakbenen och vrider dem nedåt. Och vips, så har katten vänt sig i fallet!

För den intresserade finns det universitetsuppsatser som behandlar detta ingående. Enligt Franny Syufy, som skriver på den amerikanska kattsajten cats.about.com, har en katt fallit 18 våningar och överlevt, vilket är det inofficiella rekordet i New York. Men det är inte alltid de klarar sig, det beror naturligtvis på olika omständigheter, men de har i vilket fall en vetenskapligt bekräftad förmåga att klara av att vända på sig när de faller, så att de landar med tassarna nedåt.

Så med katterna stämmer talesättet alltså, men hur är det med smörgåsar? Landar de också alltid upp-och-ner, det vill säga med smöret nedåt?

År 2000 lät man ungefär 1000 brittiska skolbarn göra en undersökning i ”Tumbling Toast Test”. Varje elev hade var sin rostad smörgås, med smör på ovansidan, som låg på en skärbräda (eller motsvarande), som i sin tur låg på bordskanten. De lutade skärbrädan sakta mer och mer tills smörgåsen gled över kanten, för att försöka efterlikna det sätt man oftast tappar smörgåsar. (Trots att det ser festligt ut, så är det väl inte så ofta du har snubblat på ett bananskal och smörgåsen har flugit iväg genom hela rummet i loopar? De allra flesta gånger man tappar smörgåsar i golvet är allt när de glider ut över bords- eller bänkkanten.)

Alla elever noterade åt vilket håll smörgåsarna landade, och varje elev gjorde om försöket i genomsnitt tio gånger. Testet utfördes även lika många gånger utan smör (de hade skrivit ett ”B”, som i engelskans ”butter”, på den sidan som var uppåt innan fallet istället), och dessutom gjordes ungefär två tusen ”fall” från två och en halv meters höjd, för att se om dessa faktorer – höjden och smöret – påverkade resultatet.

Av de tio tusen försöken med smörade brödsivor landade hela 62% med smöret nedåt. Eftersom tio tusen är ett så stort tal, så går det inte bara att blunda för de 12% som skiljer mot det antagligen förväntade 50%. Men hur förklarar man det då? Varför i hela fridens namn skulle smörgåsar med smör ha större tendens att landa nedåt? Beror det kanske på smöret?

Undersökningen innehöll som bekant även en del där eleverna lät smörgåsar utan smör falla på golvet från bordet, och resultatet man fick där var att 58% av smörgåsarna landade så att den sidan som från början låg uppåt nu låg nedåt. Det var alltså bara 4% skillnad mot testen med smör, fortfarande en klar majoritet för att den sidan som låg uppåt från början landade nedåt, så det går inte bara att skylla på smöret.

Vad beror då det hela på? I den tredje delen av testet, som var något mindre med ”bara” två tusen försök, lät man smörgåsen falla från två och en halv meters höjd, jämfört med den tidigare mer normala bordshöjden. Resultatet här blev att endast 47% landade med smörsidan upp! Det är antagligen höjden mellan bordet och golvet som är förklaringen – på en vanlig standardbordshöjd, i storleksordningen en meter, hinner brödskivan lagom rotera ett halvt varv, om den tippas av från exempelvis en skärbräda eller puttats över kanten. På högre höjder däremot hinner brödet med fler rotationer.

Påståendet att ostsmörgåsar alltid skulle landa med osten mot golvet ligger det alltså något i! Men till allra största del får det nog ändå tillskrivas samma fenomen som trollen i farstun; det är bara när den landar med osten ned på det smutsiga golvet som man riktigt reagerar över det och ilsknar till, men när den landar åt andra hållet tänker man inte så mycket på det och plockar lydigt upp den. Även om 60% är mer än hälften är det ju långt ifrån alltid. Och om sannolikheten för att smöret ska landa nedåt är $6/10$, så lär du inte märka någon skillnad jämfört med om det hade varit $1/2$ om du tappar en smörgås endast en handfull gånger, du behöver nog tappa smörgåsar åtminstone ett par hundra gånger för att kunna upptäcka skillnaden ...

Slutsatsen blir att det ibland faktiskt kan ligga någonting bakom vissa talesätt, även om talesättet om ostsmörgåsen i vardagen mer beror på ”trollen i farstun-effekten”.

Klasskamrater

Hur stor är sannolikheten att det finns två personer i en klass som fyller år samma dag?

Gör först en gissning. Ungefär hur stor skulle du tro sannolikheten är att (minst) två personer fyller år samma dag i en klass med 25 elever? $1/10$ eller $1/3$ kanske? Det finns som bekant 365 dagar på ett år, så sannolikheten borde väl vara rätt liten att två ska fylla år samma dag, kan man kanske tycka.

Här kan vi utnyttja knepet med komplementhändelser från kapitel två. Vi räknar först ut sannolikheten att *ingen* av eleverna

fyller år på samma dag, och sen subtraherar det från 1 för att få sannolikheten att *inte ingen* av eleverna fyller år samma dag, det vill säga att minst två elever har en gemensam födelsedag.

Först får vi skapa oss en modell, där vi säger att det är 365 dagar på ett år (det är ju inte riktigt sant i och med skottdagen, men den skulle ge så små skillnader i resultatet att vi kan bortse från den). Vi väljer att anse elevernas födelsedagar för oberoende, det vill säga att en elevs födelsedag inte säger något om någon annan elevs födelsedag. Vi beslutar oss också för att anse födelsedagarna jämnt fördelade över årets alla dagar. Det är inte riktigt sant, under exempelvis 2007 var toppen- respektive bottennoteringarna maj, då det föddes det 9 739 personer i Sverige, samt december med 7 720 födda, alltså en skillnad på fler än 2000 personer, enligt Statistiska Centralbyrån. Men för vår modell får det alltså duga att anse fördelningen jämn över året, det ger en tillräcklig fingervisning om hur stor sannolikheten kan tänkas vara.

Vi tänker oss att vi går igenom klasslistan ovanifrån. Vi bestämmer i tur och ordning sannolikheten för respektive elev att han eller hon inte fyller år på samma dag som någon *ovanför* på klasslistan, på det sättet missar vi inget. Sedan multiplicerar vi alla elevers sannolikheter: För den första eleven är sannolikheten 1 att han eller hon inte fyller år på samma dag som någon annan – det finns ingen ovanför på klasslistan att dela födelsedag med, alltså spelar det ingen roll vilken dag hon fyller år eftersom alla dagar är ”lediga”. För den andra eleven finns det $365 - 1 = 364$ dagar som inte eleven ovanför fyller år på, av 365 möjliga. Sannolikheten att elev nummer två är ensam om sin födelsedag blir alltså $364/365$.

Elev nummer tre på klasslistan har $365 - 2 = 363$ ”lediga” dagar att fylla år på, av 365 möjliga. Elev nummer fyra har

sannolikheten $(365 - 3)/365$ att inte fylla år på samma dag som någon ovanstående, elev nummer fem $(365 - 4)/365$ och så vidare till och med elev nummer tjugofem med sannolikheten $(365 - 24)/365$.

Sannolikheten för att ingen ska fylla år samma dag blir alltså $P(\text{första och andra och tredje och ... och tjugofemte eleven är ensamma om sina respektive födelsedagar}) = P(\text{första eleven är ensam om sin födelsedag}) \cdot P(\text{andra eleven är ensam om sin födelsedag}) \cdot P(\text{tredje eleven är ensam om sin födelsedag}) \cdot \dots \cdot P(\text{tjugofemte eleven är ensam om sin födelsedag})$ enligt multiplikationssatsen. I siffror blir detta $P(\text{inga elever har gemensamma födelsedagar}) = 365/365 \cdot 364/365 \cdot 363/365 \cdot \dots \cdot 341/365 \approx 0,431$. Sannolikheten för motsatsen till detta, att minst två personer fyller år på samma dag, blir då $P(\text{minst två personer har en gemensam födelsedag}) = 1 - P(\text{inga elever har gemensamma födelsedagar}) = 1 - 0,431 = 0,569$.

0,569 är mer än 0,5. Sannolikheten att det i en klass med 25 elever ska finnas minst två elever med en gemensam födelsedag är alltså mer än hälften!

Det var kanske lite oväntat? Men så är det faktiskt. Tänk efter själv – nog har det funnits två elever med samma födelsedag i någon skolklass du gått i? Eller i parallellklassen? Så värst ovanligt är det inte, och här är det alltså inte ett ”trollen i farstun-fenomen”, det är faktiskt vanligt med gemensamma födelsedagar i klasser.

En viktig faktor i sammanhanget är hur många elever det går i klassen. På nästa sida följer en liten tabell över hur stor sannolikheten är för att minst två elever ska ha en gemensam födelsedag i några olika stora klasser.

Antal elever i klassen	P(Minst två elever har en gemensam födelsedag)
15	0,253
20	0,411
25	0,569
30	0,706

En apa vid en skrivmaskin

Ibland kan man höra att ”om man skulle sätta en apa vid en skrivmaskin (eller dator får det väl bli i en modern tappning) och lät den trycka slumpvis på tangenterna skulle den förr eller senare skriva William Shakespeares samlade verk”. Finns det någon rimlighet i det påståendet?



Själva idén är det inget fel på, om man trycker ner slumpvisa tangenter på ett tangentbord så blir det förr eller senare ett ord någonstans, och rent hypotetiskt skulle det även kunna bli meningar och hela böcker. Vi kan tänka oss att det på ett tangentbord finns fyrtio olika tangenter som ger ett tecken på skärmen (vi bortser från stora och små bokstäver och plockar bort alla funktionstangenter samt enter och escape, et cetera). Vi antar också att apan trycker ner alla tangenter helt slumpmässigt, för att skapa oss en enkel modell.

Hur stor är sannolikheten att apan lyckas skriva ordet ”bok”? Sannolikheten att en bokstav ska bli ett ”b” är $1/40$, ett ”o” $1/40$ och ”k” också $1/40$. $P(\text{Apan skriver först ett b och sedan ett o och efter det ett k}) = 1/40 \cdot 1/40 \cdot 1/40 = 1/64\,000$. Det innebär att det inte kommer att inträffa ofta alls. Och förlänger vi

förväntningarna till en hel mening blir sannolikheten otroligt liten, och då inser du snabbt hur försvinnande liten sannolikheten är att en apa skulle kunna få ihop Shakespeares samlade verk genom att sitta vid tangentbordet! Sannolikheten att apan lyckas slumpa ihop en artikel ur morgondagens tidning blir ju faktiskt större än att det blir ens Romeo och Julia, eftersom tidningsartikeln är kortare!

I just det här stycket är det 357 tecken (exklusive mellanslag), så sannolikheten att apan lyckas skriva ens det här lilla stycket genom sina slumpmässiga tryckningar på tangentbordet är inte större än $(1/40)^{354} \approx 1/(1 \text{ med } 500 \text{ nollor efter sig})$. Det vill säga, praktiskt taget är det omöjligt, apan kommer hinna dö innan. Och det känns ju rätt betryggande ändå, det skulle ju vara lite snopet om jag skrev likadant som en apa.

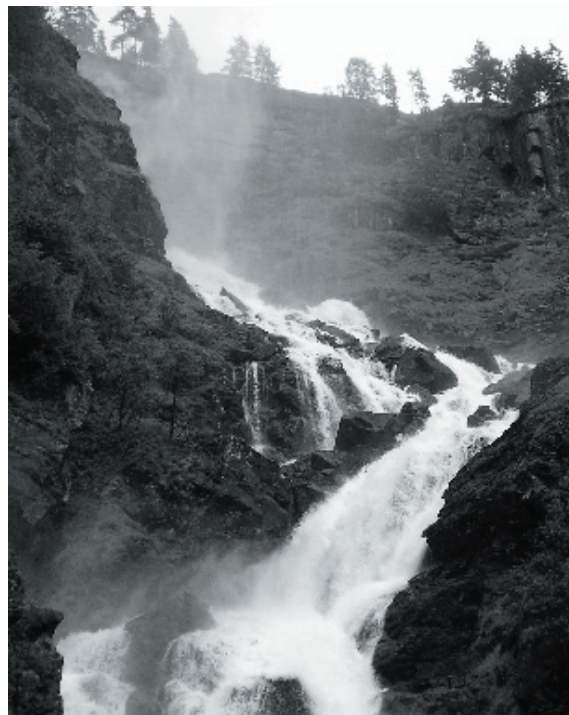


Övningsuppgifter

1.

Du ska hålla tal för 500 personer. För att väcka lite extra uppmärksamhet skulle du vilja börja med att sjunga ”Har den äran”, och hoppas att någon verkligen har födelsedag idag. Fast du vill inte ta för stora risker, och kommer fram till att du är beredd att ta en risk på 0,20 att göra bort dig. Ska du sjunga sången, det vill säga: Är sannolikheten att någon av de 500 personerna *inte* har sin födelsedag i dag större än 0,20?

(Till den som inte har tillgång till en räknare med de funktioner som krävs, vill jag rekommendera att söka efter ”online calculator” på webben.)



7.

KOMBINERA!



Kombinatorik handlar om på hur många sätt man kan kombinera allt möjligt. För att kunna beräkna sannolikheten för ett särskilt utfall, måste man veta hur många utfall det finns sammanlagt, och det är här kombinatoriken kommer till användning i sannolikhetsläran.

Upphöjt till

Hur många heltal finns det från och med 0 till och med 99? En självklar fråga kanske, men hur *räknar* man ut det på något lättare sätt än att skriva upp alla talen på ett papper och se hur många det blev?

Den första siffran i ett tal från 0 till och med 99 kan vara 0 (till exempel 04), 1 (exempelvis 14), 2 (exempelvis 24), 3, 4, 5, 6, 7, 8, eller 9. Och samma gäller för den andra siffran, den kan vara 0 (som i bland annat 30), 1 (exempelvis 31), 2, 3, 4, 5, 6, 7, 8 eller 9. Till varje *position* i talet finns det alltså tio olika möjliga siffror, och det finns två positioner.

Till varje möjlig siffra (det finns tio stycken) på första positionen finns det tio möjliga till den andra. Det går att sammanfatta som $10 \cdot 10$, vilket kan skrivas som 10^2 (uttalas ”tio upphöjt till två” och innebär att man multiplicerar 10 med sig själv). 10^2 är 100, så det finns 100 heltal från och med 0 till och med 99.

Om man kastar en tärning fyra gånger, hur många olika möjliga utfallsserier går det att få då? (En serie skulle kunna vara {etta, trea, femma, tvåa}.) Det första kastet kan vara en etta, tvåa, trea, fyra, femma eller sexa, allt som allt sex möjligheter. Detsamma gäller för var och en av de andra tre kasten, alltså blir det $6 \cdot 6 \cdot 6 \cdot 6 = 6^4 = 1296$ möjligheter.

En generell formel för det här blir alltså a^b där a är antalet möjliga utfall per gång (a är i exemplen ovan antalet möjliga siffror per position respektive antalet möjliga utfall per tärningskast), och b är antalet gånger vi ska utföra a (b är antalet positioner i talet respektive antalet tärningskast i exemplen ovan).

BIL 001

Vad är sannolikheten att få ”BIL 001” som registreringsskylt när man köper en bil? Vi får förutsätta att sannolikheten för

varje kombination är lika stor, så då gäller det gamla vanliga: $p = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$. Antalet gynnsamma utfall är lätta att räkna – det är ett; det är ju bara BIL 001 som vi vill ha. Frågan är hur många möjliga utfall det finns.

Vi börjar med siffrorna. De är inte så svåra, det finns 999 kombinationer (001 till och med 999, kombinationen 000 används inte). Så för att få fram antalet möjliga skyltar är det bara att multiplicera 999 med antalet möjliga bokstavskombinationer, eftersom det finns lika många möjliga bokstavskombinationer *för varje* sifferkombination. Men hur många bokstavskombinationer finns det då?

Det finns 23 möjliga bokstäver (A-Z utan Q, V och I, men med W) som får användas, och ungefär 150 kombinationer är inte tillåtna för att de ansetts olämpliga, som exempelvis FEM, HAT och WAR. Den första bokstaven kan väljas på 23 sätt (eftersom det finns 23 bokstäver att välja mellan), den andra kan också väljas på 23 sätt och likaså den tredje. Med det resonemanget lyckas vi täcka in alla möjliga kombinationer, från AAA till ZZZ. Det här kan vi skriva som $23 \cdot 23 \cdot 23$ eller 23^3 vilket blir 12 167. Men då är även våra förbjudna kombinationer med, så vi får subtrahera 150, och vi landar på 12 017.

För att få fram det totala antalet möjliga skyltar får vi multiplicera 12 017 (de möjliga bokstavskombinationerna) med 999 (de möjliga sifferkombinationerna), och vi får då 12 154 833 möjliga skyltar, varav *en* måste vara BIL 001. Så sannolikheten att få just den skylten på sin bil är $1/12\,154\,833$, eller ungefär en på tolv miljoner. Så det är nog inget att hoppas på till nästa gång du köper en bil.

Permutationer och fakultet

Vi har en klass på femton elever. I hur många olika ordningsföljder kan de gå in i ett klassrum? Hinner man prova alla olika följderna på en skoldag? En timme? Eller behövs det kanske en hel vecka?

Till att vara första elev att gå in finns det 15 kandidater (det vill säga vem som helst ur hela klassen); man kan alltså välja den första på 15 sätt. Till att vara nummer två finns det 14 möjliga elever kvar (en har redan gått in); man kan välja nummer två på 14 sätt. Den tredje kan väljas på 13 sätt, den fjärde på 12 sätt och så vidare, till och med den femtonde som bara kan väljas på ett sätt (eftersom det finns bara en elev kvar på slutet). Vi får alltså $15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ möjliga sätt att gå in i klassrummet. Här kan vi använda skrivsättet $15!$ (uttalas ”15-fakultet”), som betyder precis samma sak. Och $15!$ är 1 307 674 368 000, lite drygt 1 biljon, så det är nog ingen realistisk idé att prova alla ordningsföljderna på en dag eller ens en hel vecka, det kan nog räcka att läsa om det teoretiskt och använda tiden till något vettigare.

Om bara fyra av de femton eleverna ska gå in i klassrummet, hur många olika alternativ finns det då? Ja, det spelar fortfarande roll i vilken ordning de går in. Och dessutom spelar det roll vilka fyra det är som går in. Blir det fler eller färre sätt än när hela klassen skulle gå in?

Med samma resonemang som förut kan vi välja den första på 15 sätt, den andra på 14 sätt, den tredje på 13 sätt och den fjärde på 12 sätt, alltså $15 \cdot 14 \cdot 13 \cdot 12$. Har vi fått med alla tänkbara sätt nu? Jodå – den första måste vara en av de femton som finns, den andra en av de fjorton som är kvar, och så vidare.

$15 \cdot 14 \cdot 13 \cdot 12$ kan skrivas som $15!/11!$, eftersom det innebär $\frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$, så att elvan till och med ettan ”tar ut” varandra i närmaren och täljaren, och kvar blir $15 \cdot 14 \cdot 13 \cdot 12$. $15!/11!$ är 32 760. Det är många sätt det också, men det är betydligt färre än när hela klassen skulle gå in.

De båda tidigare exemplen har handlat om permutationer, och utifrån det kan man konstatera följande formel för permutationer:

Antalet permutationer när man ur en mängd med N stycken element väljer R stycken är $\frac{N!}{(N-R)!}$. Ett element är i föregående exempel en elev, och mängden är i det fallet klassen.

I det första av exemplen så var $R = N$ när alla i klassen skulle gå in i klassrummet, och nämnaren i formeln $\frac{N!}{(N-R)!}$ blev 1 ($0!$ är nämligen 1), så kvar blev bara $N!$, alltså $15!$.

Kombinationer

I permutationer spelade ordningen roll; Om eleven Alice går in först i klassrummet och sedan eleven Bertil, eller om Bertil går in först och sedan Alice, så blir det två olika alternativ. Men ibland är vi inte intresserade av att ha med någon ordning. Om man vill svara på frågan ”På hur många sätt kan man välja fyra elever ur en klass med femton elever?” så spelar det ingen roll i vilken *ordning* man väljer dem, det blir samma fyra elever oavsett om man väljer Alice före Bertil eller tvärtom.

Det smidigaste sättet att räkna ut hur många *kombinationer* det finns när man ur en mängd med N stycken element väljer R stycken är att först räkna ut antalet permutationer och sedan dela med R!, för att ”bli av” med de permutationerna som innehåller samma element, som bara ligger i olika ordning. Formeln för antalet möjliga kombinationer av N element ur en mängd med R element blir alltså $\frac{N!}{R!(N-R)!}$.

För att svara på frågan tidigare, ”På hur många sätt kan man välja fyra elever ur en klass med femton elever?”, så är det bara att sätta in värdena i formeln. Det finns femton elever i klassen, alltså är $N = 15$. Fyra elever ska väljas ut, alltså är $R = 4$. Det blir då $\frac{15!}{4!(15-4)!} = \frac{15 \cdot 14 \cdot 13 \cdot 12}{4 \cdot 3 \cdot 2 \cdot 1} = 1\,365$. Ska man välja en grupp om fyra elever av femton stycken, så finns det 1 365 möjliga grupper!

En liten användbar detalj när det gäller kombinationer är att det finns en symmetri i dem; har man en mängd med femton element och ska välja tre av dem, ger det lika många kombinationer som när man har femton element i mängden och ska välja tolv (femton minus tre). När man tänker lite närmare på det så är det rätt naturligt, eftersom man istället för att fråga ”på hur många sätt kan man välja tre element?” lika gärna kan fråga ”på hur många sätt kan man välja bort tolv element, så att det blir tre kvar?”.

Om man skulle få frågan ”på hur många sätt kan man välja ut 15 elever i en klass med 15 elever?” så fungerar formeln fint och ger svaret 1, om man nu inte skulle klara av att räkna ut detta i huvudet. Samma sak gäller med att välja ut 0 elever, svaret blir då också 1.

Även om du inte kommer ihåg formlerna (och för all del, det gör inte så mycket, de går lätt att hitta igen), så kan det vara bra

att ha en uppfattning om hur fort antalet möjliga kombinationer och permutationer växer med större mängder att välja från. Här nedan finns det en tabell med olika stora mängder, N, och olika antal element som ska väljas, R, samt hur många kombinationer respektive permutationer det då finns. Titta gärna lite noggrannare i tabellen, och konstatera att antalet permutationer växer betydligt fortare än antalet kombinationer med växande N, och att antalet permutationer beror till väldigt stor del på antalet som ska väljas, R, samt att kombinationerna har sitt maximum då R är hälften av N.

Antal element i mängden, N	Antal som därav ska väljas, R	Antal permutationer (ordningen spelar roll)	Antal kombinationer (ordningen spelar ingen roll)
1	1	1	1
2	1	2	2
5	2	20	10
5	3	60	10
10	2	90	45
10	5	30 240	252
10	8	1 814 400	45
15	5	360 360	3003
15	10	10 897 286 400	3003
15	14	1 307 674 368 000	15
20	2	380	190
20	5	1 860 480	15 504
20	10	670 442 572 800	184 756
20	15	20 274 183 400 000 000	15 504
50	10	37 276 043 020 000 000	10 272 278 170
50	25	2 sextiljarder	126 biljoner
100	50	3 quindeciljarder	101 kvadriljarder

Övningsuppgifter

1.

På hur många olika sätt kan man fylla i en tipsrad? (Stryktipset, där man tippar på fotbollsmatcher om de antingen slutar oavgjort, med vinst eller förlust för hemmalaget. Det finns 13 matcher att tippa på. Du kanske har hört uttrycket ”ha 13 rätt på tipset”?)

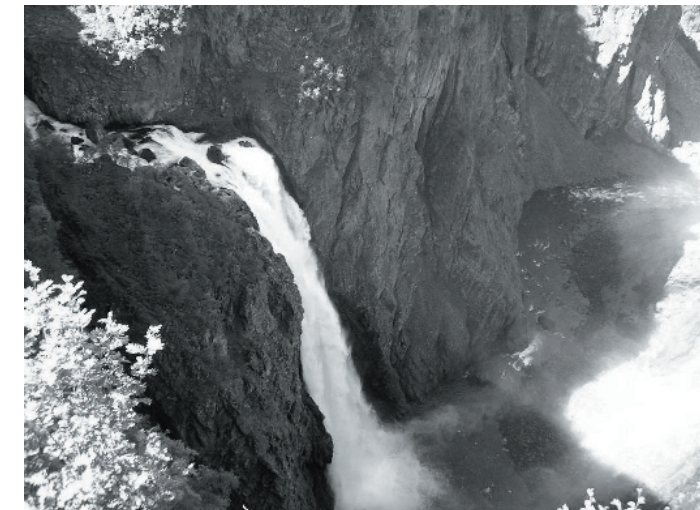


2.

Emellanåt kommer det, i varje fall till min e-postinkorg, meddelande som kortfattat säger att ”MSN har ett problem, användare registrerar flera e-postadresser per person, och snart är de slut! Vi har bara 578 lediga namn kvar. Därför kommer vi hädanefter att börja ta betalt för e-postkonton. Men om du skickar det här meddelandet vidare till 18 personer så slipper du betala.” Det behövdes inte lång tid för mig för att konstatera att det var någon som bluffade, och att några av mina vänner hade blivit lurade och vidarebefordrat meddelandet även till mig. Klarar du också, med ledning av det här kapitlet, att se att det så uppenbart är en bluff?

3.

Kalle har tre syltburkar, men alla har tappat etiketterna. I en burk finns det körsbärssylt, i en jordgubbssylt och i en lingonsylt. Eftersom alla sorterna är röda kan han inte se någon skillnad på dem. Kalle är inte så samvetsgrann av sig, så han tillverkar tre nya etiketter med rätt namn, och klistrar dem på burkarna, utan att ha en aning om vilken som är vilken. Hur stor är sannolikheten att alla tre blir rätt?





8.

ÖKA DIN LOTTOVINST



Det finns inga lottorader som har större chans att vinna än några andra. Men det innebär inte att det inte finns taktik i hur man ska välja sin lottorad – för om du skulle vinna, så vill du väl knappast dela vinsten med någon annan?

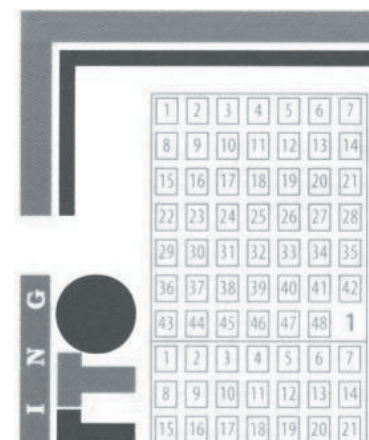
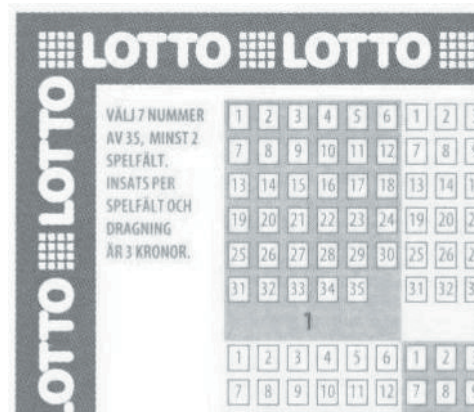
Glöm dock aldrig bort att spelbolaget går med vinst genom att arrangera lotto, vilket innebär att de allra, allra flesta spelare förlorar på att spela. Oavsett hur taktisk man är, så handlar lotto till största delen enbart om tur!

Lotto

Lotto är ett spel som i Sverige anordnas av Svenska Spel, men det är vanligt i många andra länder runt om i världen också. Spelet går ut på att man väljer en så kallad ”rad” genom att kryssa ett visst antal nummer på en kupong. Det är sedan dragning i TV, och beroende på hur många tal man själv har valt som överensstämmer med raden som dras i TV-dragningen hamnar man i olika vinstgrupper, där man tillsammans med alla andra som är i samma vinstgrupp får dela på en i förväg bestämd andel av alla prispengar. I storleksordningen hälften av spelbolagets inkomster på spelet går tillbaka som vinstpengar till spelarna, den andra hälften behåller bolaget. Är det ingen som hamnar i den högsta vinstkategorin, det vill säga att ingen har alla rätt, går de vinstpengarna vidare till nästkommande dragning, och blir en så kallad Jackpott så att man i nästa dragning kan vinna extra mycket.

I Svenska Spels Lotto ska man välja 7 nummer av 35, och på en vanlig vecka är det dragning inte mindre än fyra gånger, två gånger varje onsdag och två gånger varje lördag. En ordinarie rad på Lotto kostar 3 kronor, och den största ”rena” lottovinsten, utan Joker och Drömvinsten, som delats ut på svenska Lotto är knappt 49 miljoner kronor.

Viking Lotto är ett samarbete mellan spelbolagen i de nordiska länderna och Estland, där man ska välja 6 av 48 nummer och det är dragning varje onsdag. Att spela en rad kostar 4 kronor i Sverige, varav 0,04 euro går till den gemensamma högsta vinstgruppen.



Den största vinsten som delats ut på Viking Lotto är ungefär 18,5 miljoner kronor.

Svenska Spel har även ett spel kallat Keno som till en början kan påminna om ett lotto, eftersom man ska välja rader och det sedan är dragning. I Keno delar man däremot inte på prispengarna inom vinstgruppen, utan det finns en fastställd vinstplan som gäller oavsett om man är ensam eller en av hundra personer i sin vinstgrupp.

Framöver kommer det i första hand att syftas på Svenska Spels Lotto, om inget annat anges, men principen är densamma för alla lotto-spel runt om i världen. Den enda stora skillnaden mellan länderna är hur många nummer det finns att välja på i spelfälten, och hur många av dem man ska välja. Svenska Spels Lotto är ett 7/35-lotto, och Viking Lotto ett 6/48-lotto. Storbritannien har The National Lottery, ett 6/49-lotto, vilket också Tysklands ”Lotto 6 aus 49” är. I Norge är det (förutom Viking Lotto) Norsk Tippings 7/34-lotto som gäller.

Lotto, den svenska varianten

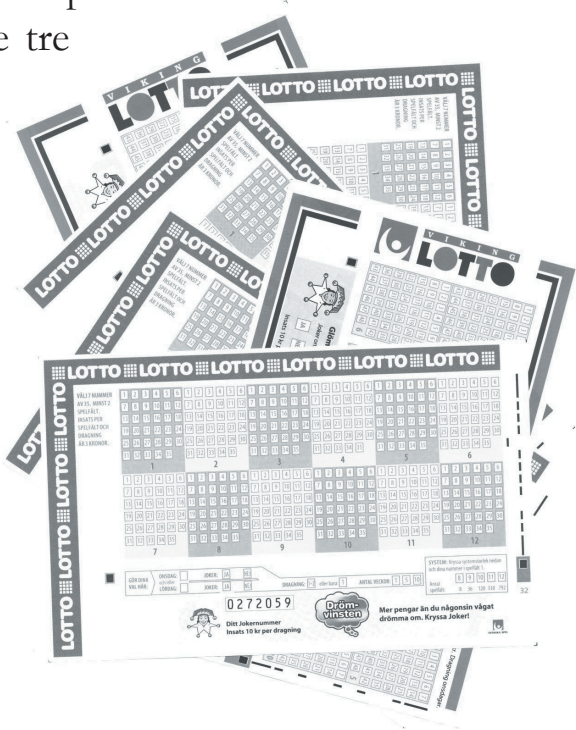
Det finns lite olika detaljer som skiljer sig åt mellan olika spelbolags lottospel, och så här fungerar just Svenska Spels Lotto:

Utöver de sju egentliga vinstnumren, dras det fyra tilläggsnummer. Har man sex rätt av de egentliga vinstnumren, och ett rätt

av tilläggsnumren, hamnar man i vinstgruppen som nedan kallas 6+1. Så här fördelas vinstpengarna mellan de olika grupperna:

7 rätt	34%
6+1 rätt	12%
6 rätt	7%
5 rätt	12%
4 rätt	35%

Det finns dessutom ett tillägg, Drömvinsten, som man kan vinna om man – utöver att ha alla rätt av de ordinarie 7 numren – spelar med ett så kallat Jokernummer. Jokernumret är ett 7-siffrigt tal, och det gäller att antingen de två första eller de två sista siffrorna i talet ska överensstämma med Jokernumret som Svenska Spel drar i samband med Lottodragningen. Drömvinsten börjar alltid på 50 miljoner kronor och växer för varje vecka tills någon vinner den, då den börjar om på 50 miljoner. Drömvinsten har på de tre åren mellan februari 2005 och februari 2008 endast delats ut fyra gånger. Den största Drömvinsten – tillika största lotterivinsten i Sverige – som delats ut var på närmare 123 miljoner. Att lägga till Joker till sin Lottorad, så att man tävlar om drömvinsten, kostar 10 kronor. En Lottorad med Joker i en dragning kostar alltså 13 kronor.



Sannolikheten för varje nummer är lika stort i dragningen (det kontrolleras noga), så det finns inga rader som vinner oftare än andra. Samtidigt publicerar Svenska Spel statistik på sin hemsida över hur ofta de respektive 35 numren blivit dragna de senaste 6 månaderna, tillsammans med texten ”Visst är det kul med statistik? Men kom ihåg att det ändå alltid är slumpen som avgör vilka nummer som dras.”

Finns det någon taktik i hur man ska välja en bra lottorad?

Lite sannolikheter

En rad består av 7 siffror, av 35 möjliga, och man kan inte ha samma siffra två gånger i samma rad. Från kapitel sju, om kombinatorik, vet vi då att antalet möjliga kombinationer ges av $\frac{N!}{R!(N-R)!}$ där N här motsvaras av antalet möjliga siffror och R av antalet vi ska välja, alltså 35 respektive 7. $\frac{35!}{7!(35-7)!} = 6\,724\,520$, så det finns en hel del möjliga rader att välja mellan. Sannolikheten att en slumpvis vald rad ska innehålla samma sju tal som dras i TV är alltså endast 1/6 724 520. Som en jämförelse kan det nämnas att det bor ungefär nio miljoner människor i Sverige. Sannolikheten att du ska vinna på Lotto på just en särskild rad i en särskild dragning är alltså endast något större än att just du skulle vinna i ett lotteri där alla svenskar (från norr till söder, från Södermalm i Stockholm till en enslig stuga i norra Norrland, från västkustens fiskelägen till Dalarnas fäbodrar!) deltog med varsin lott. Och det är ändå ”bara” att få alla rätt på Lotto, utan Drömvinsten.

Deltar du med två (olika) rader i samma dragning, eller med samma rad i två olika dragningar, så dubblas dina chanser. Och deltar du med 10 olika rader i en och samma dragning, alternativt samma rad i tio dragningar, har din chans till alla rätt vuxit till 1 på 674 252, alltså blivit tio gånger större.

Sannolikheten att vinna Drömvinsten är naturligtvis ännu mindre än att ”bara” få alla sju rätt. Vi låter bli att sätta oss in i hur Jokerspelet fungerar och undviker att begrava oss i uträkningar, och nöjer oss istället med att hämta sannolikheterna från Svenska Spels hemsida:

Sannolikheten att vinna Drömvinsten är $1/337\,915\,578$, alltså ungefär en femtiondel av sannolikheten att få sju rätt. Sannolikheten att vinna något över huvud taget är $1/55$, men av de vinsterna består mer än $9/10$ av vinstgruppen fyra rätt, där vinsten brukar landa på i storleksordningen 20 kronor per rad.

Sannolikheten att lyckas skrapa ihop alla rätt på Viking Lotto med en rad i en dragning är $1/12\,271\,512$, alltså ungefär hälften jämfört med Lotto.

Taktik

Alla lottorader har lika stor chans att vinna, såväl $\{1, 2, 3, 4, 5, 6, 7\}$ som $\{1, 2, 3, 5, 8, 13, 21\}$ och $\{3, 7, 9, 22, 30, 31, 33\}$. Men trots det så finns det faktiskt strategier i hur man ska välja sin rad.

Finessen som går att utnyttja med lotto är att vinsten delas mellan alla som är i samma vinstgrupp. Det gäller alltså att välja en rad som ingen annan har. För om din rad skulle lyckas att kamma hem storkovan vill du väl inte behöva dela din vinst med andra?

Argumentet att ”raden $\{1, 2, 3, 4, 5, 6, 7\}$ har lika stor sannolikhet som den inte lika regelbundna raden $\{3, 7, 10, 19, 22, 28, 33\}$ ” känner många till, och många väljer då hellre raden $\{1, 2, 3, 4, 5, 6, 7\}$ för att enkelt komma ihåg vilken rad man har valt, eller bara som en fix idé. Så den raden är med stor sannolikhet inget bra att satsa på – om du skulle vinna på den skulle du antagligen behöva dela din vinst med många andra. Men det finns även fler rader du bör undvika.

Det finns en utförlig bok på ämnet, *How to win more* av Norbert Henze och Hans Riedwyl. De har studerat den tillgängliga statistiken i olika länder och försökt notera hur spelare väljer sina rader, och utifrån det dragit slutsatser om hur man bör och inte bör välja sina rader. (Normalt publicerar aldrig spelbolag statistik över hur deras spelare väljer sina rader, men ibland har det skett undantag, och det är främst dessa undantag Henze och Riedwyl har tittat närmare på och dragit slutsatser utifrån.) Det handlar egentligen mer om psykologi, vad människor *uppfattar* som slumpmässigt och hur man tror på kombinationer som för tur med sig, än om matematisk sannolikhetslära. Men även det är intressant, och passar bra in i den här boken.

Deras sammanfattande uppmaning till spelare som vill vara ensamma om sina rader är att:

Välj aldrig dina nummer till din lottorad enligt någon regel som skulle kunna följas av andra spelare

Innebörden blir i praktiken att man aldrig ska välja sin rad enligt någon regel, för så fort du har en regel så finns risken att någon,

några, eller ännu värre – många, också har kommit på den regeln och väljer sina rader efter den. Henze och Riedwyl har brutit ner sin uppmaning till tolv mer specifika punkter, som jag presenterar här.

1. Välj inte aritmetiska serier

Den enklaste aritmetiska serien du kan välja som lottorad är {1, 2, 3, 4, 5, 6, 7}. Det var till och med den i särklass populäraste serien i det finska 7/39-lottot i dragningen den 27 augusti 1994, enligt Henze och Riedwyls statistik. Andra populära aritmetiska serier skulle kunna vara {2, 4, 6, 8, 10, 12, 14} och {5, 10, 15, 20, 25, 30, 35}. En aritmetisk serie är alltså en talserie där avståndet mellan talen är förutbestämd.

2. Inga tidigare vinnare

En enkel princip, men kan vara lite klurig att följa. Kombinationer som tidigare vunnit tenderar nämligen att hålla sig kvar i minnet hos vissa spelare upp till ett par år, enligt Henze och Riedwyl, så därför bör man alltså ha koll på de senaste årens vunna rader och undvika alla dem, eftersom andra spelare verkar spela på dem, kanske i tron att de för tur med sig. Den *enda* anledningen till att låta bli rader som nyss har vunnit är alltså att det är många spelare som spelar just dessa rader, sannolikheten att just förra veckans rad ska upprepas är varken större eller mindre än för någon annan rad.

3. Undvik de ”hetaste” och ”svalaste” numren...

Svenska Spel publicerar på sin hemsida en tabell med hur ofta respektive nummer har blivit dragna det senaste halvåret. Det här säger precis ingenting om hur nästa rad kommer att se ut. (Vi kan nämligen förutsätta att inga av de bollarna med nummer som används i dragningen är större/mindre/ rundare/plattare, eller vad de nu ska vara för att bli dragna oftare, än några andra. Bortsett från att Lotteriinspektionen gör kontroller, skulle spelbolaget riskera så otroligt mycket i anseende om det framkom att de fuskade, att de antagligen inte ens skulle fundera på det.)

Trots detta så finns det ändå spelare som spelar på ”heta” alternativt ”svala” rader, genom att ta de sju mest alternativt minst frekventa numren. I dragningen den 10 februari 1990 i det schweiziska 6/45-lotto var det i medeltal ungefär två spelare per rad. Inte mindre än 992 spelare spelade raden med de sex ”hetaste” numren (enligt Henze och Riedwyl)!

4. ... och de nummer som har ”vilat” längst

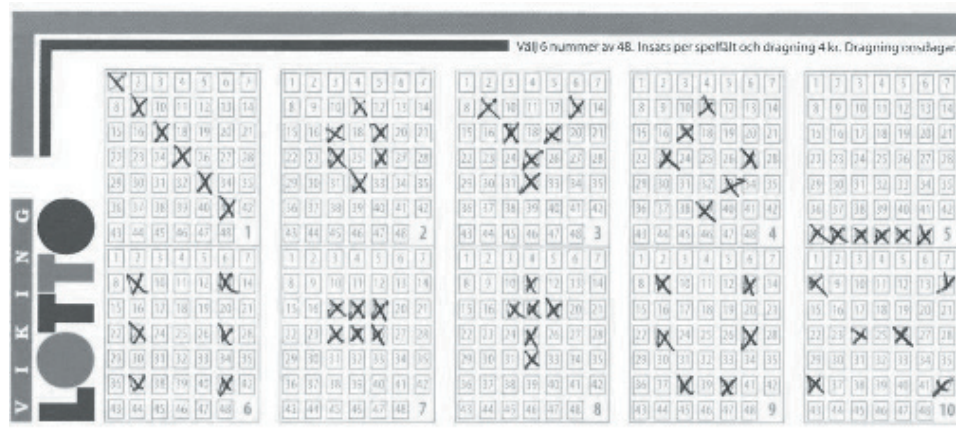
På Svenska Spels hemsida kan man relativt enkelt se vilka nummer som inte alls har blivit dragna på längst tid, det vill säga de numren som har ”vilat” längst. I det tyska 6/49-lottot i dragningen den 16 oktober 1993, där varje rad hade i medeltal färre än en halv spelare, spelade 460 på raden med de talen som hade ”väntat längst”. Notera skillnaden att ett ”svalt” nummer mycket väl kan ha blivit dragit föregående vecka, men inte särskilt ofta innan dess, medan det måste ha dröjt många dragningar för ett nummer kan kallas ”vilande”.

5. Välj aldrig lottorader som nyligen har vunnit i andra länder

En enkel princip; Henze och Riedwyls statistik visar att det förekommer att vissa spelare har koll på nyligen vinnande lottorader i grannländerna, och väljer att spela på dem själva, kanske i tron att ”de kommer hit snart”. I samma schweiziska lottodragning som nämndes tidigare konstateras att inte mindre än 1 283 spelare spelade på den raden som förra veckan drogs i Österrike!

6. Se upp för rader som ger ett snyggt mönster på spelfältet

Ett kryss eller kanske en våg? Eller kanske ett kors eller en lyckobringande hästsko? Många har säkerligen sina favoriter bland alla mönster som går att göra med sju kryss på spelfälten (eller sex för Viking Lotto och många andra länder), och det finns här bara ett råd att ge – undvik alla rader som ger visuella mönster på spelfältet, som symboler, symmetriska mönster, diagonaler och så vidare.



7. Gör inte en enkel aritmetisk eller geometrisk modifikation av en tidigare vinnande rad

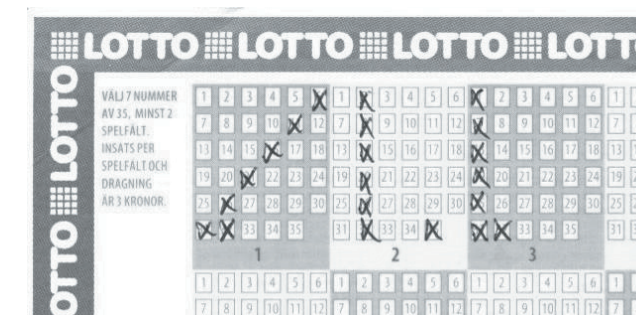
Utöver att undvika rader som vunnit nyligen är det också rekommenderat att undvika aritmetiska modifikationer av gamla vinnande rader, som att exempelvis lägga till ett till varje tal i raden, så att {3, 4} blir {4, 5}, eller dra bort eller lägga till två och så vidare. Geometrisk modifikation kan vara att flytta hela den vinnande raden ett steg ”nedåt” eller ”uppåt” på spelfältet, eller att till exempel spegelvända den.

8. Inga sifferlekar!

Man kan välja primtal som en rad, exempelvis {1, 2, 3, 5, 7, 11, 13} eller {5, 7, 11, 13, 17, 19, 23}. En annan variant är {3, 4, 6, 9, 13, 18, 24}, som bygger på principen ”lägg till en mer”; mellan tre och fyra skiljer det ett, mellan fyra och sex skiljer det två. Fibonaccis talföljd, där man lägger ihop de två föregående numren: {1, 2, 3, 5, 8, 13, 21}, finns det med stor sannolikhet också spelare som följer när de väljer lottorad. Och då gör du smartast i att låta bli, om du vill vara ensam om din eventuella vinst.

9. Försök inte approximera ett idealt geometrisk eller aritmetiskt mönster

Diagonalen är sex tal på de svenska spelfälten, och när man ska välja sju tal kan man välja en diagonal plus en annan ruta som ligger i närheten av var den sjunde



rutan ”borde” legat. Eller vill man följa en kant, får man sätta det sjunde krysset någon annanstans, och då är det inte ovanligt att sätta den så nära som möjligt där den ”borde” suttit.. Man kan också vilja välja talserier som 2, 8, 14, 20, 26, 32, 38. Men eftersom 38 inte finns, så skulle en sådan serie kunna resultera i {2, 8, 14, 20, 26, 32, 35}, eftersom 35 är så nära 38 som det går.

10. Summan av talen i raden bör vara minst 147 (gäller för 7/35-lotto som Lotto, för 6/48-lottot Viking Lotto rekommenderas minst summan 177)

Det verkar förekomma att spelare väljer rader med siffror som innehåller datum, och väljer därför aldrig rader med nummer högre än 31. Det blir inte så tydligt i det svenska Lotto, med 35 som högst, men i 6/49-lotton och liknande, som Henze och Riedwyl i första hand studerat, så är det påtagligt att summan av alla nummer är lägre än vad man först kan förvänta sig. Dessutom finns det många som har sina ”turnummer” bland de lägre numren, som 3, 7 och 13. Just summan 147 är inte huggen i sten, utan är en uppskattning.

11. 2, 3, 4, 5 eller 6 (men inte 7) av numren bör ha gemensamma kanter eller hörn på spelfältet (För Viking Lotto gäller då 2, 3, 4 eller 5 men inte 6)

Spelare verkar ha en tendens att välja bort rader med nummer som ”sitter ihop” i en klump på spelfältet, antingen över eller under varandra eller har kontakt med varandra via hörnen. Men sitter precis alla nummer ihop i en enda stor klump finns det dock en risk att de bildar någon sorts geometrisk symbol, och det vill vi också undvika.

12. 4, 5 eller 6 av numren bör ligga på kanten på spelfältet (För Viking Lotto blir motsvarande antal 3, 4 eller 5)

20 av 35 nummer ligger på spelplanen så att de har kontakt med en kant, vilket innebär att i genomsnitt fyra av sju nummer borde ligga där. Spelare verkar dock ha en tendens att inte kryssa så många nummer på kanterna som det ”borde” vara, om fördelningen mellan alla rutor vore jämn. Samtidigt bör man undvika att bara spela på nummer på kanterna, eftersom det inte är omöjligt alls att vissa spelare väljer sina rader efter principen ”bara nummer på kanterna”.

Slumpade rader

Om man spelar Lotto på Svenska Spels webbsida finns det en knapp som heter ”Hux Flux”. Klickar man på den så slumpas det en rad, och är man missnöjd är det bara att klicka på den igen så slumpas det en ny rad. Går man till spelhörnan i mataffären så finns det färdigfyllda kuponger, och det finns till och med spelalternativ där man aldrig ens ser vilka rader man spelar med, man uppger bara hur mycket pengar man ska satsa så sköter Svenska Spel resten, och rapporterar till dig om du vann något.

Om alla spelade med slumpade rader skulle inga av de tolv reglerna som nyss presenterades vara aktuella. Det är klart att

20	21	22	23	24
26	27	28	29	30
32	33	34	35	
xFlux				

19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	
HuxFlux					

19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	
HuxFlux					

2	3	4	5	6
8	9	10	11	12
14	15	16	17	18
20	21	22	23	24

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24

visa av raderna som spelare deltog med skulle vara just tjugiga geometriska mönster eller aritmetiska serier eller gamla vinnare, men de skulle inte förekomma oftare än några andra. Henze och Riedwyl har sammanställt väldigt intressant statistik från Kaliforniens 6/51-lotto under 1993 och 1994: Medelvinsten hos de spelare som spelade med slumpade rader och hade alla rätt var 10 526 198 amerikanska dollar, medan medelvinsten hos de med alla rätt som själva valde sina rader var lägre, nämligen 7 468 543 amerikanska dollar. Slutsatsen man kan dra utifrån det är att de som spelade med slumpade rader förmodligen inte behövde dela sin vinst lika ofta och med lika många, som de spelare som spelade med rader de valt själva.

För att minimera risken att behöva dela sin eventuella storvinst med andra, så bör man alltså välja slumpade rader. Antingen hämtar man färdiga kuponger, klickar på "Hux Flux" på webben eller gör sig en kortlek med alla 35 numren och drar sju kort ur den. Men en slumpad rad kan dock råka bli någon av de typer av rader vi nyss har tittat på, så för att undvika det bör du kontrollera att de uppfyller de tolv punkterna. Brister raden på någon av de tolv punkterna bör du kassera den och slumpa fram en ny. Allt för att minimera risken att behöva dela en eventuell storvinst med några andra!

Vad man dock inte ska glömma bort är att om alla lottospelare i Sverige läser den här boken, så kommer ingen längre att satsa på exempelvis raden {1, 2, 3, 4, 5, 6, 7}, och då är det helt plötsligt en bra taktik att spela på den! Hela strategin som föreslås i det här kapitlet bygger på att du vet mer än de andra spelarna. Och att du har sådan tur att du vinner också. Ty enda sättet att öka sannolikheten att vinna över huvud taget är förstås att spela *fler* rader, till det hjälper inga taktiska rader.

Övningsuppgifter

1.

Vilken/vilka av följande slumpade rader bör man spela med, enligt Henze och Riedwyls tolv punkter?

7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
10	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35



9.

BEROENDE HÄNDELSE



Matematiskt beroende handlar inte om det vi i dagligt tal kallar för beroende. Det handlar istället av hur tidigare utfall påverkar våra kommande. Det är egentligen något vi har berört lite grann tidigare, men vi ska i det här kapitlet titta närmare på det, tillsammans med något som kallas för betingade sannolikheter. Vid vissa tillfällen kan det nämligen vara extra bra att ha beroendesituationen klar för sig.

Kulpsåsar

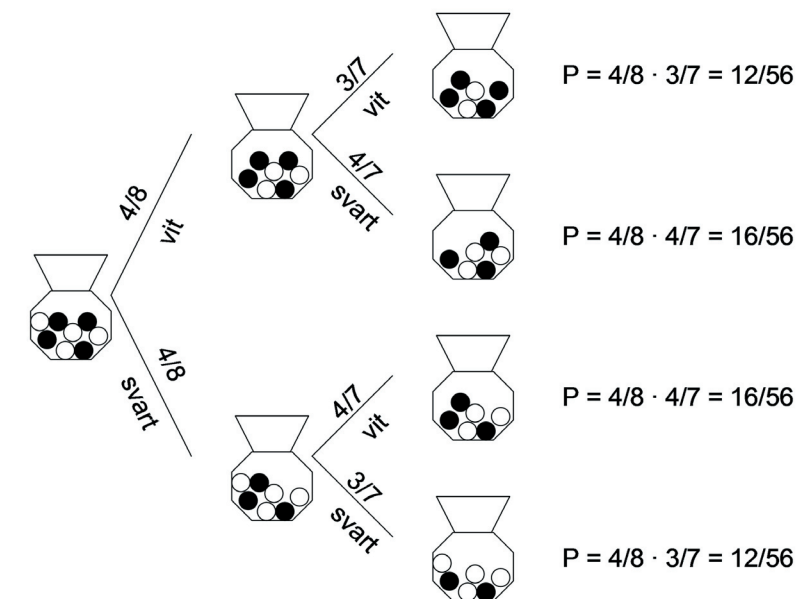
Det klassiska modellen av beroende är en påse med ett visst antal vita och svarta kulor. Man tar upp en vit kula utan att lägga tillbaka den, och det finns då färre vita kvar, och sannolikheten att få en svart i nästa drag är således större. I de tidigare kapitlen har inte ordningen som händelserna utförts i poängterats så mycket, eller vad förra händelsen var, men nu spelar det stor roll. Ett räkneexempel:

Du har en kulpåse med åtta kulor. Fyra stycken är svarta, fyra är vita. Du tar upp en kula på måfå, och lägger inte tillbaka den. Är den svart finns det tre svarta och fyra vita kvar i påsen. Sannolikheten för att ta en svart och sedan en vit kula blir då $4/8 \cdot 4/7 = 16/56 = 2/7$ eller ungefär 0,29. Sannolikheten att ta två vita är däremot $4/8 \cdot 3/7 = 12/56 = 3/14$, vilket är ungefär 0,21. Även om skillnaden 0,08 i sannolikheterna inte är jättestor, är det ändå en tydlig skillnad.

Den här typen av förfaranden brukar kallas för att man drar *utan återläggning*, det vill säga att man tar upp en kula utan att lägga tillbaka den. Sannolikheterna för olika utfall blir då *beroende* av utfallet i den tidigare händelsen (om det finns en sådan). Motsatsen är att man drar *med återläggning*, och händelserna blir då *inte beroende* av varandra.

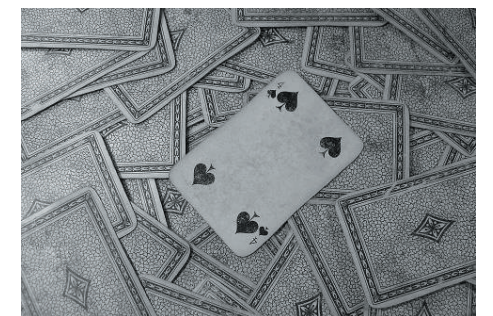
För att illustrera beroende kan träd vara lämpliga att använda, som du kan se på nästa sida.

Det luriga med beroendet ligger alltså i att alla förgreningar inte längre ser likadana ut, som de gjorde i kapitlet då vi ritade träd över slantsinglingar.



Kortlekar

Ett väldigt tydligt ställe man stöter på beroendet är i kortlekar. Du har sju kort på hand, varav fyra är hjärter och du har inte sett några andra kort alls. Sannolikheten för att ett kort du tar upp ur en hög eller motsvarande är en hjärter är mindre än att det exempelvis är en ruter, eftersom det finns fler ruter än hjärter bland de korten du inte känner till. Har du å andra sidan inga som helst hjärter på handen, men minst en av varje av de andra färgerna, så är sannolikheten att du ska ta upp en hjärter *större* än för var och en av de andra färgerna.



Samma resonemang gäller även för valörer – har du sju kort på handen, varav tre är tre ess och inga kungar, så är sannolikheten

att du ska få upp en kung $4/(52 - 7) = 4/45$, medan sannolikheten att du ska få upp ett ess endast $1/(52 - 7) = 1/45$, eftersom det bara finns ett ess kvar i högen, men fyra kungar.

Musikspelare

Beroende och oberoende slumpprocesser påverkar oss i vardagen på lite olika sätt. Bland annat när vi lyssnar på musik.

I CD- och mp3-spelare brukar det finnas en slumpfunktion, som slumpvis väljer låt att spela upp av de som finns på skivan, eller i mappen, eller i spellistan. Ofta brukar funktionen kallas för "shuffle" eller "random". Vissa spelare har funktionen med principen utan återläggning, och vissa spelare



principen med återläggning. Om funktionen väljer låtar utan återläggning innebär det att först väljs en helt slumpmässig låt, och nästa låt som väljs är någon av de resterande. Den tredje låten som väljs är också av de som hittills inte har spelats, och så håller det på tills alla låtar är spelade en gång var, då den börjar om

från början. Det betyder att om du sätter på spelaren, väljer slump-läget och får höra din absoluta favoritlåt så är det inte någon idé att hoppas på att få höra den snart igen – spelaren kommer att gå igenom *alla* låtar på skivan eller i mappen eller spellistan innan favoritlåten ens är med i urvalet igen. Och den gången kan du

ha sådan otur att din favoritlåt inte spelas upp förrän sist av alla låtarna.

Den andra principen, med återläggning, gör valet mellan *alla* låtar varje gång, oavsett om de har spelats precis innan eller aldrig tidigare. Så sätter du igång en mp3-/CD-spelare med en slumpfunktion som fungerar enligt principen med återläggning och får höra din favoritlåt, så är det inte alls omöjligt att den kommer en gång till direkt. Men å andra sidan kan det dröja i princip hur länge som helst innan den kommer igen, även om den förr eller senare kommer att spelas, enligt Stora talens lag.

Betingade sannolikheter

Herr X löper en viss risk att få en sjukdom, och därför gör han emellanåt besök hos en läkare. Det finns några olika möjliga tillstånd för honom: antingen är han frisk och läkaren har inte haft någon anledning att ställa en diagnos, eller så är han sjuk och har fått en diagnos. Men det kan också vara så att han är sjuk, fast läkaren inte har upptäckt det och därför inte ställt någon diagnos, alternativt att han fått en (felaktig) diagnos av läkaren trots att han egentligen är frisk.

Sannolikheten att herr X är sjuk över huvud taget är 0,045. Men vet vi någonting, exempelvis att herr X har fått diagnosen, så är sannolikheten att han verkligen är sjuk betydligt större. Här är lite olika sannolikheter för de olika fallen:

$P(\text{Herr X är frisk och läkaren har ingen anledning att ställa en diagnos}) = 0,95$

$P(\text{Herr X har fått en (felaktig) diagnos}) = 0,005$

$P(\text{Herr X är sjuk men läkaren har inte hittat några symptom och ställer ingen diagnos}) = 0,015$

$P(\text{Herr X är sjuk och har fått diagnosen}) = 0,03$

Om vi nu vet att herr X har fått diagnosen, hur stor är då sannolikheten att han faktiskt har sjukdomen, det vill säga vad är $P(\text{Herr X är sjuk givet att han har diagnosen})$? Vi får sortera ut andelen av den totala sannolikheten att Herr X har diagnosen, och det gör vi på följande sätt:

$P(\text{Herr X är sjuk, givet att han har diagnosen})$

$$= \frac{P(\text{Herr X har sjukdomen och diagnosen})}{P(\text{Herr X har diagnosen})}$$

$$= \frac{P(\text{Herr X har sjukdomen och diagnosen})}{P(\text{Herr X är frisk och har diagnosen}) + P(\text{Herr X har sjukdomen och diagnosen})}$$

$$= \frac{0,03}{0,005 + 0,03} = \frac{0,03}{0,035} = 0,86$$

Sannolikheten att Herr X verkligen har sjukdomen när han har fått diagnosen är alltså 0,86. Det här ger den mer generella formeln

$$P(A \text{ givet } B) = \frac{P(A \text{ och } B)}{P(B)}$$

och sannolikheten för A kallas här för en *betingad* sannolikhet eftersom den påverkas av att vi känner till B.

Definitioner

För att öva lite på vad man kan göra med matematiska formler kan vi utifrån de betingade sannolikheter göra lite definitioner på vad beroende och oberoende innebär. För Herr X var sannolikheten att han hade sjukdomen beroende av om han hade en diagnos eller ej, men däremot påverkade det knappast om det regnade eller inte. Att $P(\text{Herr X har sjukdomen, vi vet att det regnar}) = P(\text{Herr X har sjukdomen})$ är rätt självklart, och är beviset på att Herr Xs eventuella sjukdom och regn är oberoende. Mer generellt kan vi skriva att:

Om $P(A \text{ givet } B) = P(A)$ är sant, så är A och B *oberoende*.

Om $P(A \text{ givet } B) \neq P(A)$ är sant (som fallet med Herr Xs diagnos) så är A och B *beroende*.

Övningsuppgifter

1.

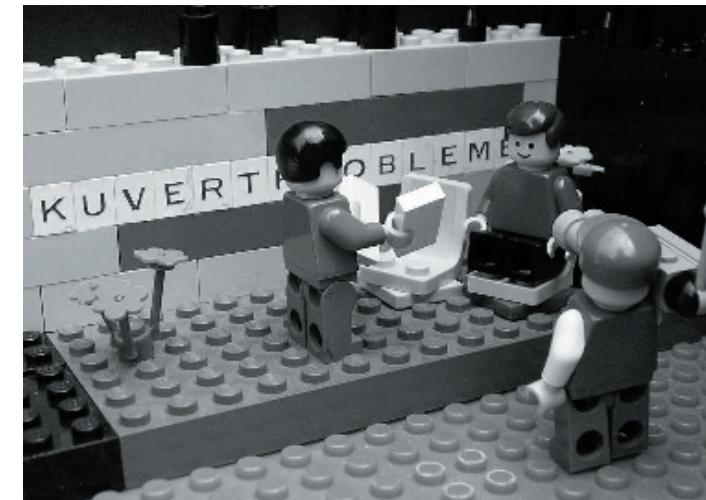
Använd formeln $P(A \text{ givet } B) = \frac{P(A \text{ och } B)}{P(B)}$ för att beräkna $P(\text{Tärningen visar en fyra, givet att den visar ett jämnt tal})$.

2.

Räkna nu med samma formel på följande uppgift: Det är en folkomröstning om studiemedlet ska höjas, på bekostnad av barnbidraget. 20% av de röstande är studenter och 90% av dem röstar för en höjning. Totalt röstar 40% av alla för en höjning. Hur stor är sannolikheten att en person är student, om man vet att hon röstat för en höjning? (Tips: börja med att definiera A och B tydligt)

10.

KUVERTPROBLEMET



Först vill jag poängtera att det här kapitlet är “formel-tätt”, och att du gott kan nöja dig med att ögna igenom det här kapitlet och sedan hoppa till nästa kapitel, som inte alls har samma karaktär som det här kapitlet. Om du är på det humöret, vill säga. Känner du dig å andra sidan beredd att ta itu med lite formler och matematiska tankegångar, så är det bara att fortsätta läsa!

Det här är, liksom getproblemet i kapitel tre, också ett klassiskt problem inom sannolikhetsläran. Men på grund av den något mer svårsmälta förklaringen så är det inte lika väl spritt. I den här boken är det antagligen det allra mest avancerade inslaget, och det är inga konstigheter om du inte ”köper” förklaringarna. Statistikern Magnus Arnér avslutar i sin bok På irrfärd i slumpens

värld sitt kapitel om kuvertproblemet med följande mening, efter att ha försökt utreda hur det ligger till: ”Detta strider mot mitt förnuft och jag måste medge att jag återigen är totalt förvirrad.”

Det matematiska resonemanget har jag tills största del hämtat ur Olle Häggströms bok Slumpens skördar, men skulle det ha smugit sig in brister här är naturligtvis felet mitt.

Problemet

Vi befinner oss återigen i den amerikanska TV-showen, som i kapitel tre, men den här gången är det inga dörrar, getter eller bilar som gäller, utan istället kuvert och pengar. Programledaren håller i två kuvert, och informerar dig om att det i det ena kuvertet finns dubbelt så mycket pengar som i det andra. Han blandar båda kuverten väl, och du får välja ett av dem och öppna det. Om du vill får du behålla de pengar som ligger i det, eller byta kuvert och ta pengarna i det andra kuvertet, men då utan att först få se hur mycket pengar som finns i det andra.

Så långt borde det inte vara några konstigheter. Nu ställer vi oss frågan: bör man byta kuvert till det andra efter att man öppnat det första? Vi kan tänka oss att vi har öppnat kuvert A och att det låg 100 kronor i det. I kuvert B finns det då antingen hälften, 50 kronor, eller dubbelt så mycket, 200 kronor, av vad som finns i A. Vi valde kuvert A med sannolikheten 0,5, så rimligen borde chansen att B innehåller dubbelt så mycket pengar vara lika stor som risken att B bara innehåller hälften. Genomsnittet på innehållet i kuvert B borde då vara $1/2 \cdot 50 + 1/2 \cdot 200 = 125$, det vill säga att vi i genomsnitt gör en vinst i att byta.

Hade vi valt kuvert B hade vi kunnat resonera likadant om kuvert A, och vi hade även då vunnit på att byta. Egentligen behöver vi inte öppna kuvertet ens, vi kan alltid resonera att det finns $1/2 \cdot 0,5 + 1/2 \cdot 2 = 1,25$ gånger pengarna i det andra kuvertet jämfört med det vi har valt. Och byter vi utan att först öppna kuvertet, så kan vi resonera så om att byta till det andra igen! Och igen, och igen, och för varje byte så gör vi en vinst på 25% av vad vi redan hade. Vad bra, vi har uppfunnit en finansiell evighetsmaskin! Står vi där och velar en stund har vi tjänat flera miljoner.

Någonstans blev det något fel i resonemanget. När vi kan dra en slutsats om att det lönar sig att byta kuvert utan att ens öppna det måste något vara galet, det finns väl ingen som helst grund för att tro att det finns mer pengar i det ena kuvertet än det andra...?

En brasklapp

Till att börja med ska jag lägga in ett litet förbehåll: en del anser nämligen inte detta problem höra hemma i sannolikhetsläran alls, för att det är alldeles för odefinierat. Som vi snart ska se så blir problemet väldigt märkligt när det inte finns någon gräns på hur mycket programledaren får dela ut.

Maximalt 1000 kronor

För att inse att det är något orimligt i frågeställningen kan vi resonera så här: Om vi vet att det aldrig i något kuvertet finns mer än, säg, 1000 kronor, och att sannolikheten för varje belopp 1-1000 kr är lika stor i kuvertet med den större summan, så skulle vi kunna tänka oss att vi satte upp följande lilla tumregel: I *genomsnitt* borde då summan i kuvertet med den högre summan vara ungefär 500 kronor (egentligen 500,5 men vi bortser från det nu), och på samma sätt borde den lägre summan i genomsnitt vara ungefär 250 kronor. Det totala medelvärdet blir då ungefär 375, det vill säga att om vi deltar i programmet väldigt många gånger, och alltid öppnar kuvertet vi först väljer och alltid behåller det, kommer vi i genomsnitt hitta 375 kronor.

Utifrån detta resonemang går det att formulera en taktik. Vet vi att det i genomsnitt finns ungefär 375 kronor i varje kuvert, kan vi bestämma oss att byta kuvert om vi hittar mindre än 375 kronor, eftersom det är under genomsnittet. Öppnar vi däremot kuvertet och hittar mer än 375 kronor så ska vi inte byta, eftersom det är över genomsnittet.

Detta är ingen optimal taktik, men det skulle kunna vara en taktik i vilket fall. Poängen är att den är formulerad utifrån det maximala beloppet som kan finnas i ett kuvert (och vi förutsätter att sannolikheten är jämnt fördelad mellan de olika summorna).

37,5% av oändligheten

Men nu känner vi inte till någon övre gräns. Antar vi att programledaren kan dela ut oändligt med pengar (fortfarande med en jämnt fördelad sannolikhet för alla möjliga summor) får vi strategin att byta kuvert om summan inte överstiger 37,5% av oändligheten (eftersom det borde bli medelvärdet om vi byter ut tusen mot oändligheten i resonemanget ovan). Hur mycket är 37,5% av oändligheten? Det är förvisso en intressant fråga, men just här informerar den oss mest bara om att det finns ett fundamentalt fel i problemformuleringen – vi saknar en övre gräns för hur mycket pengar det får finnas i kuvertet, och hur vanliga de olika möjliga summorna är (nyss antog vi att sannolikheten för de olika summorna var jämnt fördelad, men det behöver den ju inte vara. Programledaren är kanske snål och väljer oftare en lägre summa än en högre?). Uttryckt med matematiska termer saknar vi ett definierat utfallsrum. Låt oss titta närmare på det ur en lite annan synvinkel.

Betingad sannolikhet

För att göra det lite enklare för oss inför vi två storheter: X är det lägre av de två beloppen som tävlingsledaren har valt, och Y är beloppet i det kuvertet vi själva har valt. I exemplet ovan var alltså $Y = 100$, men frågan är om $X = 100$ eller 50, det vill säga om vi har valt kuvertet med den högre eller lägre summan pengar. Är $X = 100$, så finns det 200 kronor i kuvertet med mest pengar, och

vi har då valt kuvertet med den lägre summan pengar. Men är $X = 50$ så finns det bara 100 kronor i kuvertet med mest pengar, och vi har då valt kuvertet med den högre summan i.

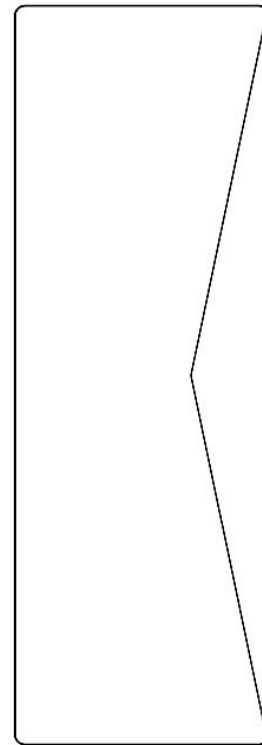
Felet kan ligga i resonemanget när vi kommer fram till att medelvärdet för det andra kuvertet är 1,25 gånger högre än det vi har valt, som i sin tur bygger på den betingade sannolikheten $P(X = 100 \text{ givet } Y = 100) = P(X = 50 \text{ givet } Y = 100) = 1/2$, som vi antog eftersom det verkade rimligt. Men vi får ta och se om vi kan definiera $P(X = 100 \text{ givet } Y = 100)$ utifrån andra förutsättningar vi känner till, och se om det kanske är här felet ligger.

Det här handlar alltså om betingade sannolikheter, som vi gick igenom teorin bakom i kapitel nio. Vad är egentligen den betingade sannolikheten att $X = 100$ kronor, om vi vet att $Y = 100$ kronor, det vill säga att vi valde kuvertet med den lägre summan pengar?

Eftersom kuverten är ordentligt blandade, kan vi konstatera att sannolikheten att, om den lägsta summan programledaren har lagt i ena kuvertet, X , är 100, så måste sannolikheten vara lika stor att den tävlande tar kuvertet med 100 kronor som att han tar kuvertet med 200 kronor eftersom de blandas väl. Vi definierar alltså

$$P(Y = 100 \text{ givet } X = 100) = 1/2 \quad (1)$$

vilket måste stämma så länge kuverten blandas ordentligt. (Notera att $P(X = 100 \text{ givet } Y = 100)$ *inte* är samma uttryck som $P(Y = 100 \text{ givet } X = 100)$). På samma sätt kan vi med all säkerhet definiera sannolikheten att den tävlande tar kuvertet med 100 kronor, om kuvertet med minst antal pengar i innehåller 50 kronor:



$$P(Y = 100 \text{ givet } X = 50) = 1/2 \quad (2)$$

eftersom kuverten blandas väl. Från förra kapitlet minns vi formeln $P(A \text{ givet } B) = \frac{P(A \text{ och } B)}{P(B)}$. Vi sätter in $P(Y = 100 \text{ givet } X = 100)$ som vänsterled, och kan sedan förlänga båda sidor med $P(X = 100)$, så att vi får

$$P(Y = 100 \text{ givet } X = 100) \cdot P(X = 100) = P(Y = 100 \text{ och } X = 100) \quad (3)$$

och på liknande sätt

$$P(Y = 100 \text{ givet } X = 50) \cdot P(X = 50) = P(Y = 100 \text{ och } X = 50) \quad (4)$$

som vi kommer att behöva senare. Vi ska också konstatera att om vi vet att $Y = 100$ kronor, så finns det endast två möjliga X , 50 respektive 100. Utifrån det kan vi fastslå att

$$P(Y = 100) = P(X = 100 \text{ och } Y = 100) + P(X = 50 \text{ och } Y = 100) \quad (5)$$

$P(X = 100 \text{ givet } Y = 100)$

Vi börjar nu om med formeln $P(A \text{ givet } B) = \frac{P(A \text{ och } B)}{P(B)}$ för att förhoppningsvis få någon ordning på om det är $P(X = 100 \text{ givet } Y = 100)$ som spökar. Vi antog som sagt i början av kapitlet att denna sannolikhet, som innebär att vi valt kuvertet med den lägre summan pengar om vi vet att det är 100 kronor i kuvertet, hade värdet $1/2$, men med hjälp av det vi nu har fastslagit, (1), (2), (3), (4) och (5), ska vi se om detta antagande var riktigt.

$$\begin{aligned}
 P(X = 100 \text{ givet } Y = 100) &= \frac{P(X = 100 \text{ och } Y = 100)}{P(Y = 100)} \\
 &\text{(med hjälp av (5) gör vi en substitution, "ersättning", i nämnaren)} \\
 &= \frac{P(X = 100 \text{ och } Y = 100)}{P(X = 100 \text{ och } Y = 100) + P(X = 50 \text{ och } Y = 100)} \\
 &\text{(ytterliggare substitutioner, "ersättningar", enligt (3) och (4) i nämnaren och täljaren)} \\
 &= \frac{P(X = 100) \cdot P(Y = 100 \text{ givet } X = 100)}{P(X = 100) \cdot P(Y = 100 \text{ givet } X = 100) + P(X = 50) \cdot P(Y = 100 \text{ givet } X = 50)} \\
 &\text{(substitution igen, enligt (1) och (2) i både nämnaren och täljaren)} \\
 &= \frac{1/2 \cdot P(X = 100)}{1/2 \cdot P(X = 100) + 1/2 \cdot P(X = 50)} \\
 &\text{(vi förlänger, multiplicerar både nämnare och täljare, med 2)} \\
 &= \frac{P(X = 100)}{P(X = 100) + P(X = 50)}
 \end{aligned}$$

Och vad kan man visa med detta? Jo, för att kunna beräkna $P(X = 100 \text{ givet } Y = 100)$ måste vi veta någonting om $P(X = 100)$ och $P(X = 50)$.

Låt oss återigen säga att programledaren som mest får lägga 1000 kronor i ett kuvert, det vill säga att $X_{\max} = 500$. Vi antar också att han bara lägger jämna 50-tal, och att sannolikheten är jämnt fördelad för alla tillåtna X , vilket innebär att $P(X = 50) = P(X = 100) = 1/10$. Vi sätter då in det så här: $P(X = 100 \text{ givet } Y = 100) = \frac{P(X = 100)}{P(X = 100) + P(X = 50)} = \frac{1/10}{1/10 + 1/10} = 1/2$. Det här var knappast någon revolution, om vi går tillbaka till början och använder detta $1/2$ med gott samvete, så får vi ju den här mystiska 25%-vinsten varje gång vi byter. Vad är skillnaden nu?

Jo, för att ha fått $P(X = 100 \text{ givet } Y = 100) = 1/2$ har vi antagit ett utfallsrum och sannolikheter för $P(X = 50)$ och $P(X = 100)$. Nu satte vi $P(X = 50) = P(X = 100)$, att det är lika troligt att programledaren väljer en stor som liten summa, men så behöver vi inte göra. Om programledaren gillar att dela ut små summor

pengar kanske $P(X = 50) = 1/5$ och $P(X = 100) = 1/10$, det vill säga att han oftare väljer $X = 50$ än $X = 100$. Med dessa värden blir $P(X = 100 \text{ givet } Y = 100) = 1/3$, och vi har inte längre någon 25% vinst vid ett byte ($2/3 \cdot 50 + 1/3 \cdot 200 = 100$), vi kan nu varken räkna med vinst eller förlust vid ett byte.

Sätter vi en ännu mindre sannolikhet på $X = 100$ i förhållande till $X = 50$, så kommer vi att få ett resultat att vi förväntas göra en förlust när vi byter kuvert. Med ett lite mer ledigt språk kan vi konstatera att om det är ovanligt att programledaren väljer stora summor, och vi har hittat en stor summa i kuvertet, så ska vi vara nöjda och inte byta till det andra kuvertet. Detta eftersom det är så pass ovanligt med så stora summor som vi hittade att vi bör förvänta oss att hitta en mindre i det andra kuvertet.

Men om nu $P(X = 50) = P(X = 100)$, det är ointressant om det är $1/10$ eller $1/1000$, blir $P(X = 100 \text{ givet } Y = 100)$ fortfarande $1/2$ (prova själv!), så då lönar det sig alltså att byta? Ja, men under ett egentligen ganska självklart villkor; att $Y \leq X_{\max}$. $X_{\max}$ innebär det maximala värdet som X kan anta. Låt mig förklara:

Om programledaren får dela ut 1000 kronor som maximalt, alltså $X_{\max} = 500$, och vi öppnar kuvertet och hittar, Y , 700 kronor ($Y > X_{\max}$) så vore det idiotiskt att byta, eftersom vi då vet att det andra kuvertet måste innehålla 350 kronor eftersom 1400 kronor är mer än vad programledaren får dela ut.

Hur ska man då välja när vi befinner oss framför TV-kameran? Jo, programledaren har inte oändligt med pengar att dela ut, det måste finnas en gräns någonstans. Troligen väljer han heller inte alla summor med lika stor sannolikhet, det kanske är vanligast med summor i något medelintervall, och det bara i undantagsfall förekommer riktigt stora och små summor? Det är upp till den

tävlande att först skaffa sig en modell över hur sannolikheterna för de olika möjliga summorna ser ut. Den tävlande får då i nästa steg placera in de antagna sannolikheterna i formeln $P(X = 100 \text{ givet } Y = 100) = \frac{P(X = 100)}{P(X = 100) + P(X = 50)}$ där $P(X = 100)$ ska tolkas som sannolikheten för att programledaren väljer den summa du nyss har konstaterat ligga i det öppnade kuvertet som lägsta summa, och $P(X = 50)$ innebär sannolikheten att programledaren väljer hälften av summan vi hittade som en lägsta summa. När man sedan får svaret får man då göra en bedömning av om man skulle kunna förväntas tjäna på ett byte eller ej. Och så måste man ta hänsyn till om $Y \leq X_{\max}$.



11.

SVÄLTA RÄV



I det här kapitlet ska vi titta lite närmare på kortspelet svälta räv. Svälta räv är kanske inte det mest rafflande kortspelet, men det är ett bra spel för att få lite känsla för hur kortspel fungerar. Det är också ett väldigt tydligt exempel på beroende, som kapitel nio handlade om.

Regler

Alla kort ur kortleken blandas väl, så att vi inte har en aning om vad som ligger var. De delas sedan ut, lika många till varje spelare. Spelarna får inte titta på dem, utan har dem liggandes på bordet

framför sig i var sin hög med baksidan upp. En omgång börjar med att en spelare tar det översta kortet ur sin hög och lägger upp det mitt på bordet, så att alla kan se vad det är. Så fortsätter alla spelare runt bordet i tur och ordning och lägger upp ett kort var på den högen som nu har skapats mitt på bordet. Den spelare som först vänder upp ett kort ur sin egen hög och lägger i mitten på bordet med samma färg (klöver, spader, hjärter eller ruter) som det kortet som ligger underst i högen mitt på bordet får hela den gemensamma högen, och lägger alla de korten underst i sin egen hög. Samma spelare lägger sedan ut ett nytt kort i mitten av bordet som början på nästa omgång.

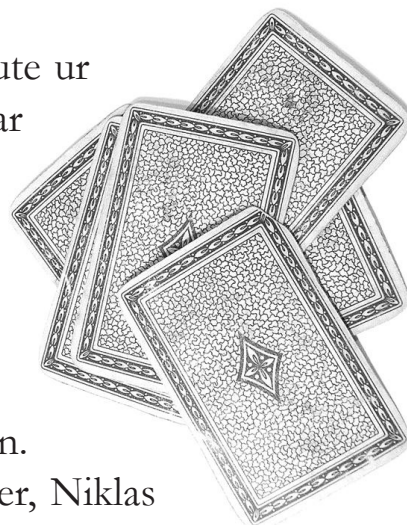
Blir man av med alla sina kort är man ute ur spelet (man blir ”svält”), och den som är kvar längst vinner.

Vi tar ett litet exempel: Kurt, Lisa, Moa och Niklas spelar svälta räv. Kurt börjar att lägga upp en ruter åtta på bordet. Den som nu först lägger en ruter får därmed ta upp ruteråttan och alla kort som lagts ovanpå den.

Lisa lägger en spader, Moa lägger en spader, Niklas lägger en hjärter, och det blir då Kurts tur igen. Kurt lägger en spader, och efter Kurt lägger Lisa en ruter dam.

Lisa får då ta upp alla sex korten, lägga dem underst i sin egen hög och lägga ut ett nytt kort. Det visade sig bli klöver fem. Det är då Moas tur, och hon får upp en klöveråtta ur sin hög och får då ta och lägga det kortet, tillsammans med klöverfemman, underst i sin hög och sedan lägga ut ett nytt kort i mitten på bordet.

För att det ska bli överskådligt att räkna på, får vi göra ett litet antagande, som man kan diskutera hur riktigt det är. När man tar



alla kort från högen på bordet och lägger underst i sin hög, så vet man ju vad som ligger där, så hela högen är inte ordentligt blandad längre. Men vi antar ändå att spelarna har så pass dåligt minne, att ingen kommer ihåg vad man lagt in i högen, och på så sätt anser vi att allas högar är ordentligt blandade hela tiden.

Vi återgår till Kurt, Lisa, Moa och Niklas spel. De har precis börjat en ny omgång. Moa börjar och lägger hjärter sju. Vad är sannolikheten att Niklas, som sitter i tur, också ska lägga en hjärter och på så sätt få ta upp korten?

Det finns 52 kort i leken, varav 13 är hjärter. Vi vet att en hjärter redan ligger på bordet, alltså finns det 51 kort kvar varav 12 är hjärter. Och vi har ingen aning om vilka kort som ligger i just Niklas hög, så vi kan inte precisera oss mer än att sannolikheten är $12/51 = 4/17 \approx 0,235$ att Niklas lägger en hjärter.

Nu lägger inte Niklas en hjärter, och inte Kurt heller. Hur stor är sannolikheten att Lisa ska lägga en hjärter? Det finns 49 okända kort kvar, och 12 av dem måste vara hjärter. Alltså är sannolikheten $12/49 \approx 0,246$ att Lisas kort ska vara en hjärter. Att sannolikheten är större i det senare fallet än det tidigare är inte så konstigt, eftersom andelen hjärter är större bland de kvarvarande korten när man har plockat ut två ”icke-hjärtrar”.

P(näste spelare får ta upp högen) = 1

Nu ska vi titta lite på en Extremsituation för att förhoppningsvis få ännu bättre förståelse för spelet. När är sannolikheten 1 att näste spelare ska lägga ut en hjärter från sin hög, om det var hjärter som

lades som första kort? Det vill säga, när kan man vara säker på att en spelare kommer att få ta upp högen, om det är ett hjärter som ligger först?

För att sannolikheten ska vara 1, måste antalet gynnsamma fall vara lika med antalet möjliga fall, för att formeln $P = \frac{\text{antalet gynnsamma fall}}{\text{antalet möjliga fall}}$ från kapitel två ska stämma, eller med andra ord: alla utfall ska vara gynnsamma, det vill säga alla möjliga utfall ska vara hjärter. Det betyder att det bara får finnas hjärter kvar i högen. För att omgången fortfarande ska kunna hålla på får inte någon mer hjärter än det första vara lagt, det vill säga att det måste finnas 12 hjärter kvar att lägga ut. Och inga andra kort, eftersom det bara skulle finnas hjärter kvar.

Svaret på frågan blir alltså att om man efter att ha lagt ut $52 - 12 = 40$ kort inte har lagt ut ett enda hjärter (förutom det kortet som ligger underst i högen), så är sannolikheten för att nästa kort är en hjärter lika med 1. Man *vet* alltså att nästa kort kommer att vara en hjärter i det fallet.

Vad är sannolikheten för att man ska hamna i det läget som beskrivs ovanför, att man lagt ut 40 kort varav bara det första var ett hjärter? Det kan väl inte vara så värst vanligt?

Först måste vi titta på sannolikheten att första kortet blir en hjärter. Det finns 52 kort i leken, varav 13 är hjärter. Alltså är sannolikheten för att det första kortet är hjärter $13/52 = 1/4$. Sedan får vi räkna ut sannolikheten för att det andra kortet som lades var ett icke-hjärter, att tredje kortet som lades var ett icke-hjärter, att fjärde kortet som lades också var ett icke-hjärter, och så vidare och så vidare.

För andra kortet finns det 51 kort kvar, varav 39 måste vara något annat än hjärter. För det tredje kortet finns det 50 kort kvar, och 38 kort av dem är inte hjärter, och så vidare. Det blir alltså $1/4 \cdot 39/51 \cdot 38/50 \cdot 37/49 \cdot 36/48 \cdot \dots \cdot 2/14 \cdot 1/13$. Med fakultet (som vi minns från kapitlet om kombinatorik) kan vi skriva det som $1/4 \cdot \frac{39!}{51!12!}$, och det är ungefär 0,000 000 000 001 574 eller $3/2\,000\,000\,000\,000$. Med andra ord är det 3 av 2 biljoner gånger man spelar svälta räv som man hamnar i den situationen!

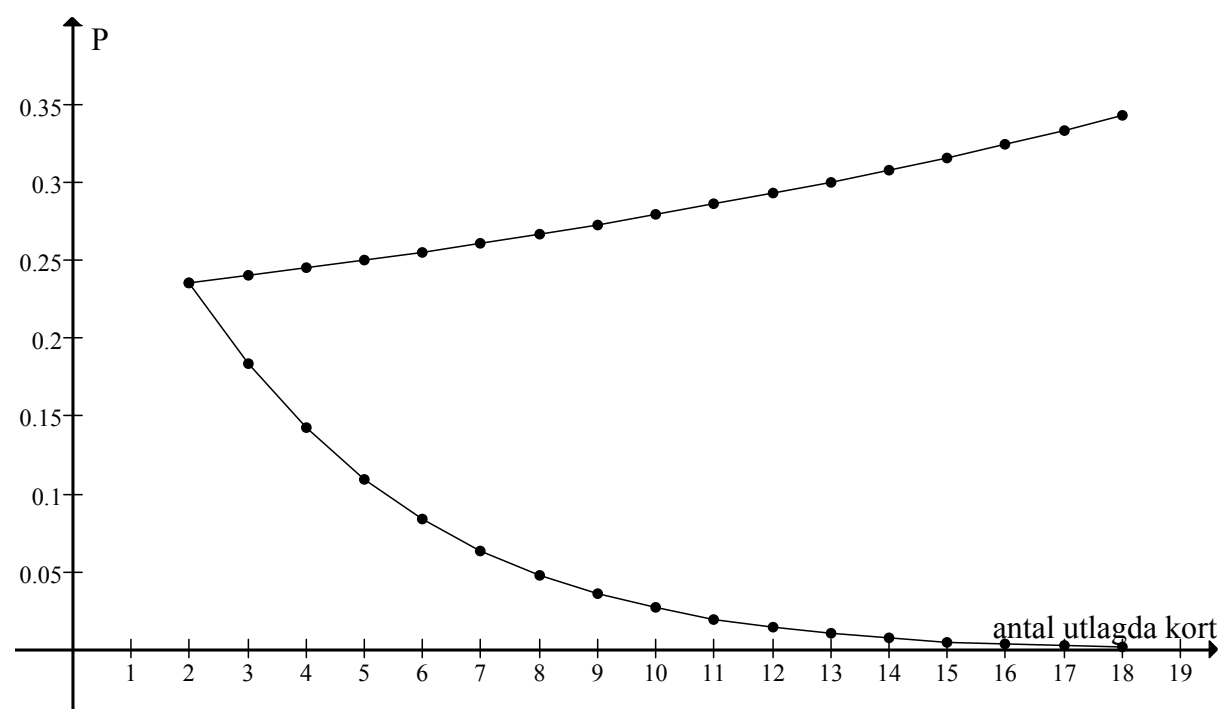
Sannolikheten att man hamnar i den situationen som vi beskrivit fast utan någon specifik färg är fyra gånger så stor, eftersom vi ska ta bort $1/4$ ur uträkningarna. $1/4$ kom från att första kortet skulle vara just en hjärter och inget annat, men när det inte spelar någon roll vilken färg det är, så tar vi bort det. Så spelar du och någon vän svälta räv, och det slutar med att *alla* de kort ni har kvar i era högar är i samma färg som det första kortet, oavsett vilken färg det är, så ring tidningen, det är något särdeles unikt! Om nu tidningen inser det unika och det går att göra något spännande reportage, men statistiskt sett är det anmärkningsvärt i alla fall.

Lite grafer

Vi ska titta på sannolikheten att man ska få ta upp högen med just det kortet man ska lägga, och se hur det beror på hur många kort som lagts innan. Det handlar alltså inte om sannolikheten att spelet ska ha pågått så länge, utan bara sannolikheten att man ska få ta upp högen i just det aktuella draget.

Efter att det första kortet är utlagt är sannolikheten $12/51$ att man ska få rätt färg på sitt kort så att man får ta upp högen. Om det allra första kortet plus ytterliggare ett är utlagt, så blir sannolikheten $12/50$ att ha rätt färg på sitt kort. 12 kommer alltså alltid att stå kvar i täljaren, eftersom det måste finnas tolv kort kvar av den färgen, annars skulle omgången redan vara slut. Nämnaren sjunker däremot med 1 för varje kort som är utlagt, vilket innebär att sannolikheten att få ta upp högen med nästa kort blir högre ju längre en omgång har hållit på.

Det kan också vara intressant att titta på en annan sannolikhet, nämligen hur sannolikheten för att en omgång ska hålla på till att ett visst antal kort lagts ser ut. Att man får ta upp högen med det andra kortet som läggs ut, det första möjliga kortet, och därmed avslutar omgången, är som vi tidigare konstaterat $12/51$. För att avsluta omgången med det tredje, andra möjliga, kortet gäller dels att spelaren innan inte tog upp det med sitt kort *och* att man själv tar upp det med sitt egna kort, alltså $39/51 \cdot 12/50$. Vi kan rita båda dessa sannolikheter i samma graf:



(I numreringen på den vågräta axeln räknas alltså det första kortet som läggs ut med, trots att man inte kan ta upp hela högen med det.)

Den övre kurvan är sannolikheten att det kortet som läggs upp kommer att vara av samma färg som det som ligger underst i högen, det vill säga att man får ta upp högen. Sannolikheten för det femtonde kortet som läggs ut är drygt 0,3. Det betyder att vi kan förvänta oss att ungefär en tredjedel av de gångerna det redan ligger fjorton kort på bordet kommer den spelare som vänder upp det femtonde kortet att få ta upp högen.

Den övre kurvan stiger med antal utlagda kort. Det betyder att sannolikheten att spelaren som ska lägga nästa kort får ta upp högen blir större och större desto fler kort det ligger i högen på bordet. Men inte nog med det – vi kan också se att den övre kurvan är lite böjd och inte en rät linje. Det betyder att sannolikheten att näste spelare ska få ta upp hela högen ökar med olika stora steg, där varje steg blir större än det föregående, för varje spelare som inte får ta upp högen: När kort nummer två, det första kortet man kan få ta upp högen med, läggs ut och det visar sig att det inte är av den färgen som ligger underst, så är sannolikheten att näste spelare ska lägga ett kort och få ta upp högen $12/50 - 12/51 \approx 0,005$ större. Om däremot det tolfte kortet visar sig att det inte heller är ett kort av "rätt" färg, så är sannolikheten att nästa kort har "rätt" färg $12/40 - 12/41 \approx 0,007$ större.

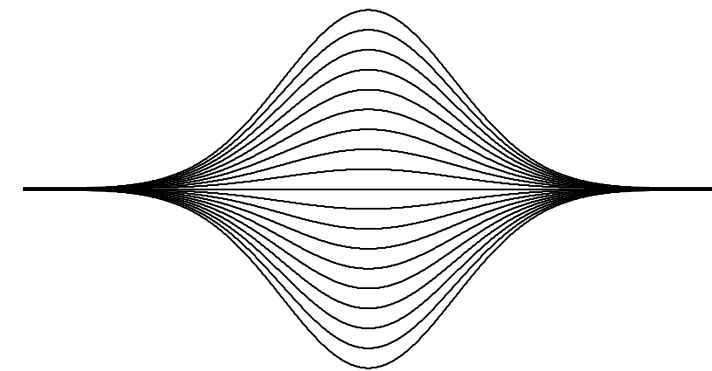
Den undre kurvan visar sannolikheten att en omgång håller på precis så länge att det hinner läggas ut så många kort som visas på den horisontella axeln. Sannolikheten att en omgång ska hålla på tills exakt det fjärde kortet läggs ut är ungefär 0,1. Det betyder alltså att ungefär var tionde omgång slutar i och med det fjärde kortet. En del omgångar slutar tidigare, andra senare.

Vad är sannolikheten att en omgång slutar med antingen två, tre eller fyra utlagda kort? Vi får addera sannolikheterna för de tre möjligheterna. Sannolikheten för att det slutar efter andra utlagda kortet ser vi i grafen är 0,24, för tredje kortet 0,19 och för det fjärde 0,15. Sannolikheten för att omgången ska sluta efter två, tre eller fyra utlagda kort (inräknat det första) är alltså $0,24 + 0,19 + 0,15 = 0,58$, alltså mer än hälften. Statistiskt sett slutar alltså mer än hälften av omgångarna på det första varvet om man är fyra spelare.

En annan poäng värd att notera är att det är väldigt viktigt att fundera på vad man egentligen frågar efter; ”sannolikheten att jag får ta upp högen med mitt kort efter just så här många kort” kan låta misstänkt likt ”sannolikheten att omgången håller på till ett visst antal kort är utlagda”, trots att det ger helt skilda värden! Och det är inte alla gånger alla lyckas med sina formuleringar så bra, utan frågar efter något annat än vad man egentligen vill veta. Så håll ögonen öppna, och koppla in det sunda förnuftet!

12.

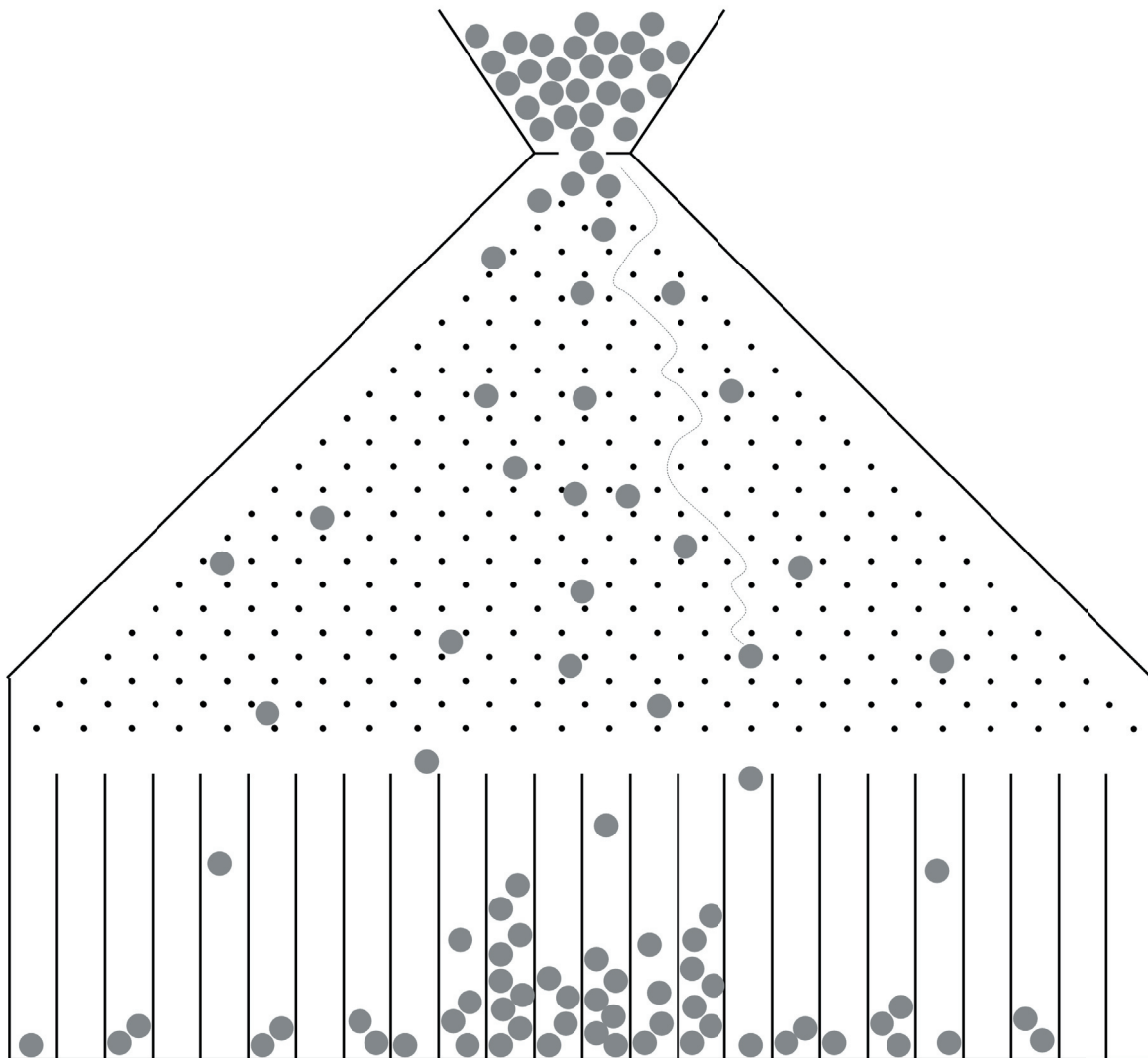
NORMALFÖRDELNING



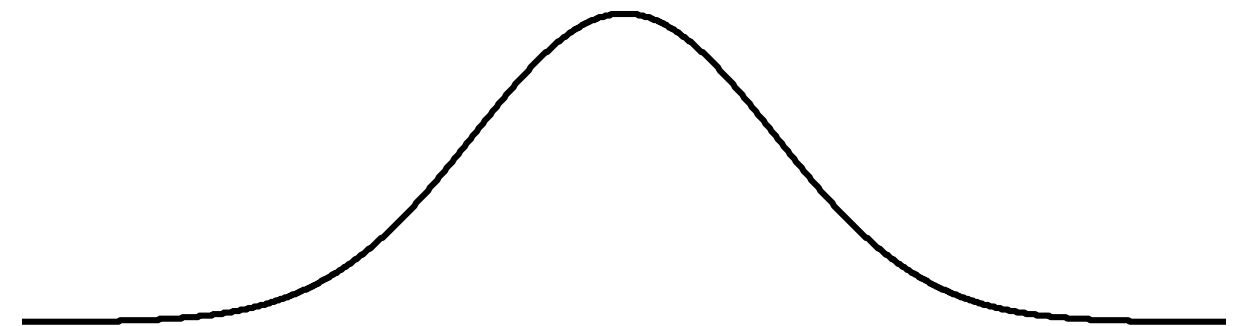
Normalfördelning är ett intressant område, det handlar om hur saker och ting fördelar sig i naturen, och ger oss möjligheter att förutsäga sådant vi kommer att möta. Det här kapitlet handlar mer om statistik än sannolikhetslära, och är bara orienterande inom området. Till den som blir intresserad och vill lära sig området ordentligt rekommenderar jag ett besök på biblioteket, där det går att hitta läroböcker i statistik.

Galtons Bönmaskin

Charles Darwins kusin Francis Galton var en mångsysslare inom vetenskapen. Han konstruerade bland annat en anordning för att kunna illustrera normalfördelningen. Egentligen ska den vara betydligt än mer omfattande än skissen nedan, men principen är densamma. Titta på skissen en stund och fundera på hur kulorna borde fördela sig på botten i det långa loppet. Var blir det flest? På kanterna eller i mitten? Eller blir det lika många överallt?

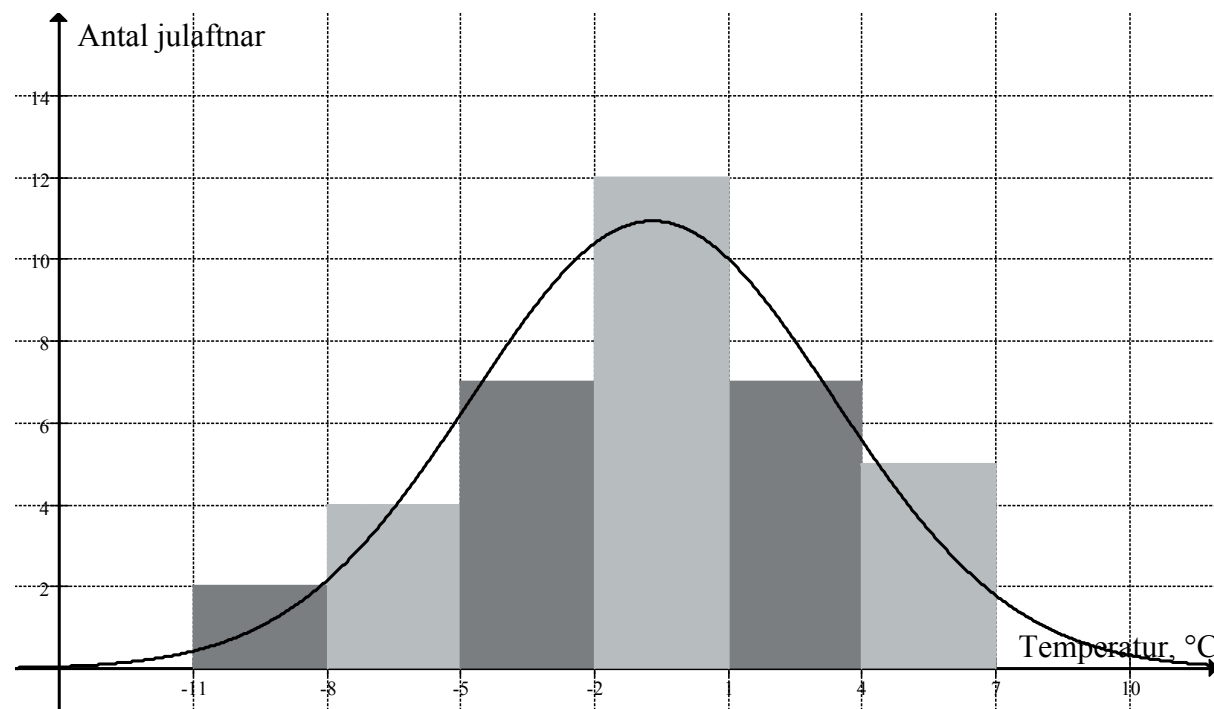


Som du nog har kommit fram till, så kommer fler kulor att hamna i mitten och färre på sidorna. Faktum är att det inte kommer att se ut som ett vanligt tippigt tak, utan kulstaplarna kommer i det långa loppet att få en väldigt karakteristisk form. Eftersom det är så pass få kulor och få fack de kan hamna i, så kommer det förmodligen inte att bli en helt perfekt kurva om man byggde modellen i verkligheten och använde den. Men skulle man lägga till betydligt fler kulor, eller bönor som det egentligen ska vara, och fler fack för dem att hamna i, skulle man få en kurva av stapelhöjderna som skulle se ut så här:



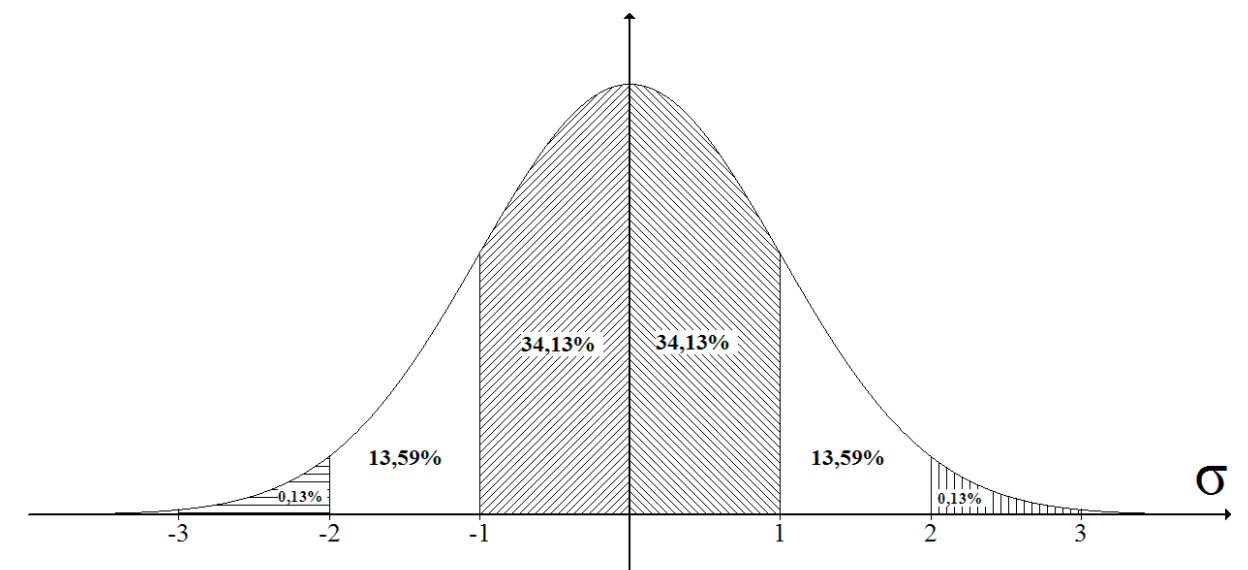
Kurvan kallas en normalfördelningskurva och går att använda på i princip alla naturligt förekommande storheter. Om man väger granbarr kommer man att få en kurva med den formen, när man har vikten på den vågräta axeln och antalet på den lodräta. Går du runt och mäter längden på personer du möter på stan kommer du också att få en sådan här normalfördelningskurva, eller Gauss-klocka som den också ibland kallas, efter matematikern Carl Friedrich Gauss. Mäter man temperaturer en viss dag på året flera år i rad, kan man också förvänta sig att få en normalfördelningskurva.

Här nedan har jag ritat in temperaturerna i Enköping på julafton åren 1970 – 2006, tillsammans med den motsvarande teoretiska normalfördelningskurvan, anpassad efter de här värdena:



Standardavvikelse

Ett viktigt mått i samband med normalfördelningen är den så kallade standardavvikelsen. Vi har ett medelvärde, som brukar betecknas med lilla μ . μ är vid toppen på kurvan, och sedan talar standardavvikelsen om hur mycket värdena avviker från medelvärdet μ . Standardavvikelsen brukar betecknas med den grekiska bokstaven lilla sigma, σ . I en normalfördelningskurva är 68,27% av arean under kurvan i intervallet $\mu - \sigma$ till $\mu + \sigma$, det vill säga inom en standardavvikelse från medelvärdet.



I temperaturstatistiken ovan är standardavvikelsen 4,05 från medelvärdet $-0,7$. Det innebär att vi kan förvänta oss att någonstans runt 68% av de julaftnarna har haft en temperatur i intervallet $-4,8$ till $3,4$ grader ($-0,7 - 4,05$ respektive $-0,7 + 4,05$). Faktum är att 65% av värdena är inom det intervallet, vilket är förhållandevis väldigt nära 68% med tanke på att det bara är värden från endast 37 tillfällen.

Inom området $\mu - 2\sigma$ till $\mu + 2\sigma$, det vill säga upp till och med två standardavvikelser från medelvärdet, återfinns 95,45% av värdena, och inom tre standardavvikelser från medelvärdet hittar man 99,37% av värdena. Utifrån medelvärdet och standardavvikelsen för temperaturen på julafton i Enköping som finns ovan, så kan vi dra slutsatsen att antalet julaftnar med en temperatur mer än tre standardavvikelser ifrån $-0,7$ grader, det vill säga antingen över 11,5 eller under $-12,9$ grader, statistiskt sett knappt förekommer en gång på 100 år (men det gäller i det långa loppet, två extremt varma julaftnar i rad är inget bevis för att modellen är falsk, det är bara en bekräftelse på att det finns variationer).

Man kan ofta, av praktiska skäl, behöva nöja sig med en uppskattad standardavvikelse utifrån ett stickprov, när vi inte har tillgång eller möjlighet att mäta hela den så kallade populationen. Man får då en skattad standardavvikelse, som man brukar beteckna med ett s, för att skilja den från σ .

För att räkna ut den skattade standardavvikelsen, alltså standardavvikelsen ur ett stickprov, använder man sig av följande formel

$$s = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n-1}}$$

där $x_1, x_2, \dots, x_n$ är de olika observationerna, och n antalet observationer. Man kvadrerar alltså differensen mellan utfallet och medelvärdet, lägger ihop alla dessa kvadrater och dividerar allt med antalet observationer minus ett, och drar slutligen roten ur detta.

Medelvärde μ
Standardavvikelse σ
Skattad standardavvikelse s

Midsommarafton i Enköping

På nästa sida återfinns en tabell med liknande data som exemplet med temperaturen på julafton byggde på, men det handlar nu istället om temperaturen den 21 juni (alltså ungefär midsommarafton) på samma ställe, Enköping.

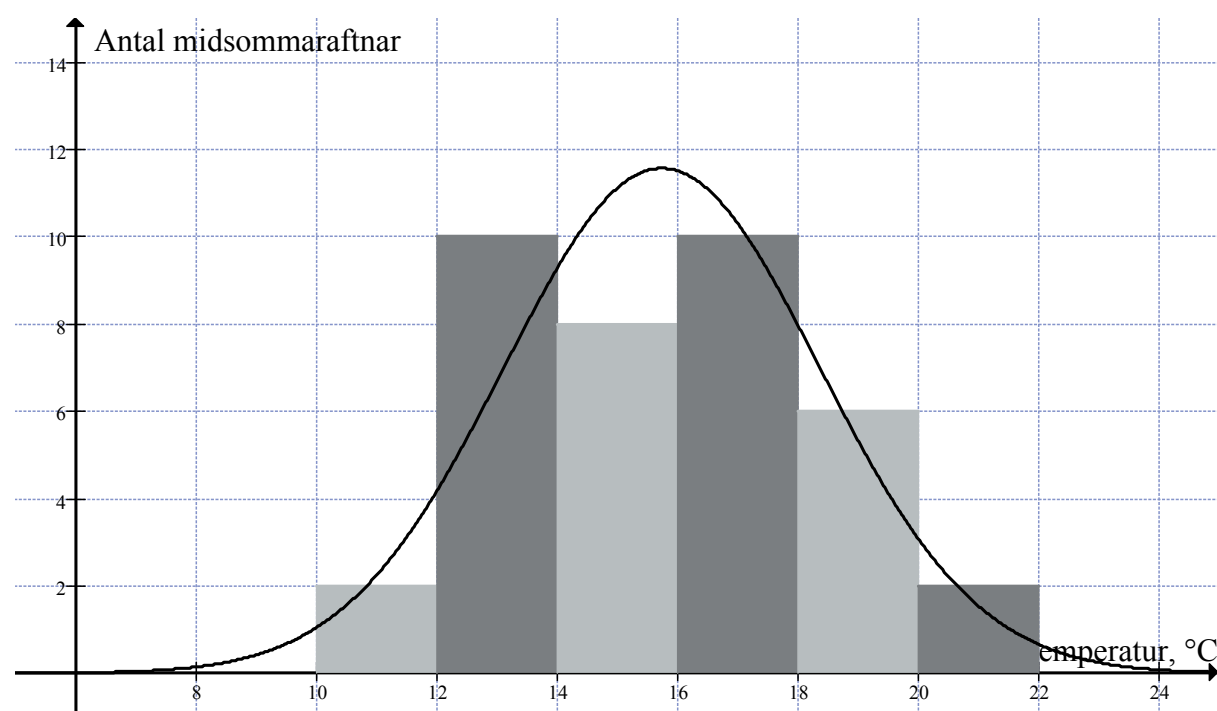
År	Temperatur, °C	År	Temperatur, °C
1970	21,5	1989	18,7
1971	16,3	1990	17,7
1972	14,1	1991	13,3
1973	18,8	1992	14,6
1974	17,6	1993	13,2
1975	21,2	1994	16,5
1976	13,3	1995	17,4
1977	14,3	1996	13,7
1978	18,5	1997	14,7
1979	19,3	1998	11,2
1980	14,5	1999	16,4
1981	13,3	2000	18,5
1982	10,1	2001	14,9
1983	13,4	2002	13,9
1984	17,6	2003	14,2
1985	18,1	2004	13,2
1986	12,8	2005	17,9
1987	13,6	2006	17,5
1988	17,4	2007	15,4

Vi ska nu titta noggrannare på dessa data, för att se hur kurvan ser ut för temperaturerna jämfört med motsvarande normalfördelningskurva. Först börjar vi med att beräkna vårt μ , alltså medelvärdet. Genom att addera alla värden och sedan dela på 38 (antalet värden) får vi fram att medelvärdet är 15,75. Sexton grader på midsommarafton verkar ju som ett rimligt medelvärde, bara som en kontroll att resultatet inte har blivit helt galet. Värt att tänka på här är dock att medelvärdet, och även standardavvikelsen, naturligtvis är beroende av hur exakt våra mätningar är utförda.

Vi vill också ha standardavvikelsen. Vi beräknar differensen mellan 15,75 och varje enskilt värde, kvadrerar den, delar allt med $38 - 1$, alltså 37, och drar slutligen roten ur det vi får.

Efter en liten stunds räknande kommer man fram till att standardavvikelsen, s , är 2,66.

Med $\mu = 15,75$ och $s = 2,66$ och med hjälp av något grafritande program (se mina rekommendationer till sådant längst bak i boken) kan vi rita upp vår graf. Men den blir inget intressant förrän vi kan jämföra den med hur temperaturerna verkligen är. För att kunna rita upp temperaturerna bra, så bör vi välja några lämpliga intervall att rita in temperaturerna i. Tvågradersintervall fungerar bra, med start på 10 grader. (För att få normalfördelningskurvan att täcka lika stor area som staplarna i grafen behöver vi lägga till koefficienten 76 framför normalfördelningsfunktionen, intervallstorleken multiplicerat med antal mätvärden.)



Mätvärdena och normalfördelningskurvan ser ut att ha ungefär samma form, särskilt med tanke på att det är så förhållandevis få värden, 38 stycken, den bygger på. För normalfördelningen lyder också under Stora talens lag – ju fler mätvärden du har, desto närmare en perfekt normalfördelning kommer du att hamna. Om vi lägger till fler värden, så gör vi klokast i att också beräkna μ och s på nytt, eftersom medelvärdet kanske blir högre eller lägre, vårt stickprov (i form av åren 1970 till 2007) kanske var ovanligt varma?



13.

Tristan Miller: VARFÖR JAG ALDRIG KOMMER ATT HA EN FLICKVÄN



Det här kapitlet är min fria översättning av Tristan Millers Why I Will Never Have a Girlfriend. Innehållet och resonemanget ska läsas med en nypa salt, jag tror knappast att Miller har skrivit uppsatsen på blodigaste allvar, det är snarare lite absurd studenthumor. Läs det som ett roande och kanske oväntat sätt att tillämpa normalfördelningen och sannolikheter i största allmänhet!

Beräkningarna och siffrorna han använder är helt korrekta, men man kanske kan fundera på hans krav på hur vacker han vill att sin eventuella flickvän ska vara, det kanske är ett lite väl högt krav han har ... Avgör själv, och ta inte slutsatsen på för stort allvar!

Varför har inte jag en flickvän?

Det här är en fråga som praktiskt taget varje kille ställer sig själv vid ett eller annat tillfälle i livet. Tyvärr finns det sällan ett klart och tydligt svar. Många män försöker trots det resonera sig igenom problemet, och kommer ofta fram till en rad löjliga förklaringar, den ena mer deprimerande än den andra: ”Är det för att jag är för blyg, och inte tillräckligt utmanande? Är det mina öppningsreplikar? Är jag en tråkig person? Är jag för tjock eller smal? Eller är jag helt enkelt ful och oattraktiv för kvinnor?” När alla rimliga förklaringar har räknats bort, så slutar det med att ”det måste vara något fel med mig” och man ger sedan upp sina drömmar. (Efter en kort period av grubblande kommer dock de flesta av dessa män att komma till den riktiga insikten att de hade för höga förväntningar, och hittar sedan någon medioker kvinna att gifta sig med.)

Inte jag, dock. Jag, för min del, har inte tänkt att spendera resten av livet grubblandes över min otur med kvinnor. Trots att jag är den första att erkänna att mina chanser att ingå ett meningsfullt äktenskap med någon speciell är lika med noll, så hävdar jag ständaktigt att det inte har med någon inneboende egenskap hos *mig* att göra. Istället är jag övertygad om att situationen kan beskrivas med endast vetenskapliga termer, utan att använda mer än befolkningsstatistik och lite grundläggande statistiska beräkningar.

För att ingen ska tro att mina förväntningar på kvinnor är för höga, så ska jag dämpa de farhågorna genom att precisera mina kriterier. Först och främst, så ska flickvännen vara i min ålder – låt oss säga 21 plus minus tre, fyra år. För det andra, så ska hon vara

vacker (med den termen avser jag både inre och yttre skönhet). För det tredje, så ska hon vara hyfsat smart – hon behöver inte vara medlem i Mensa, men gärna kunna klara av att komma med kvicka, genomtänkta argument. Där är de tre enkla kraven jag har, som jag tror alla tycker kan vara rimliga.

Nu ska jag presentera varför sannolikheten att finna en lämplig kandidat som uppfyller de ovan nämnda tre kraven praktiskt taget är obefintlig – med andra ord, varför jag aldrig kommer att ha en flickvän. Jag ska försöka göra beviset så noggrant som den tillgängliga informationen tillåter. Och jag ska också tillägga, att det inte är något statistiskt lurendrejeri här: All data och alla beräkningar är korrekta, på webbsidan en.nothingisreal.com/wiki/Why_I_Will_Never_Have_a_Girlfriend finns alla källor hänvisade, och alla nödvändiga beräkningar redovisade.

Antal invånare på jorden (1998): 5 592 830 000

Vi börjar med den största mängden jag skulle kunna vara intresserad av – alla människor på jorden. Det betyder inte att jag är emot idén om ett interstellärt förhållande, jag tror bara inte att det är särskilt sannolikt att jag skulle träffa någon söt flicka från Andromedagalaxen.

... varav kvinnor: 2 941 118 000

Enligt titeln på uppsatsen, så var det kvinnor det handlade om, om någon hade missat det. Alltså är det bara hälften av jordens befolkning som är intressant. Tyvärr, killar.

... i ”utvecklade” länder: 605 601 000

Hädanefter begränsar vi oss till I-länderna. Det är inte av förakt till de som inte har det lika bra ekonomiskt ställt, det är bara baserat på att sannolikheten att jag träffar en tjej från Bhutan eller Ghana, personligen eller via Internet, är väldigt låg. Jag kommer spendera nästan hela mitt liv i Nordamerika, Europa och Australien, så det är till dessa länder jag har begränsat siffrorna.

... i åldern 18 till 25: 65 399 083

Jag vill gärna ha en kvinna i min egen ålder.

... som är vackra: 1 487 838

Personlig attraktivitet, både fysisk och mental, är en viktig del i ett förhållande. Vad som är vackert är såklart subjektivt. Tyvärr finns det ingen statistik över hur vanligt förekommande vackra kvinnor är, men att antagande kan vara att det är normalfördelat. Det är baserat på att det finns gott om medelsnygga tjejer, och bara några enstaka riktigt snygga och riktigt fula. En kvinna som för min del är tillräckligt vacker (såväl invändigt som utvändigt) för att ha ett förhållande med är i så fall minst två standaravvikelser över medelvärdet, vilket innebär 2,3% av alla kvinnor.

... och intelligenta: 236 053

Vad som avses med intelligens kan också vara väldigt personligt, men återigen borde detta vara normalfördelat. På den här punkten nöjer jag mig med en standardavvikelse över medelvärdet, och vi kan då räkna bort 84,1% av de som hittills kvarstått. Jag utgår från att faktorerna ”vackra” och ”intelligenta” är oberoende.

... och inte redan upptagna: 118 027

Jag har inte kunnat hitta någon tydlig statistik över hur många kvinnor som är upptagna, men informella observationer gör att jag tror att andelen är runt 50%. Av moraliska skäl (och kanske för att skydda mig själv) har jag inte tänkt att försöka dejta de som redan har pojkvänner, är förlovade, gifta eller motsvarande.

... och som också skulle kunna gilla mig: 18 726

Bara att jag hittar en passande tjej behöver inte innebära att hon tycker om mig. Låt oss anta att även det här är normalfördelat, och att en tjej skulle ge en kille en chans om, och endast om, han befinner sig åtminstone en standardavvikelse över ”medelkillen”. Alltså skulle 15,9 % kunna tänka sig att ge mig en chans till ett förhållande.

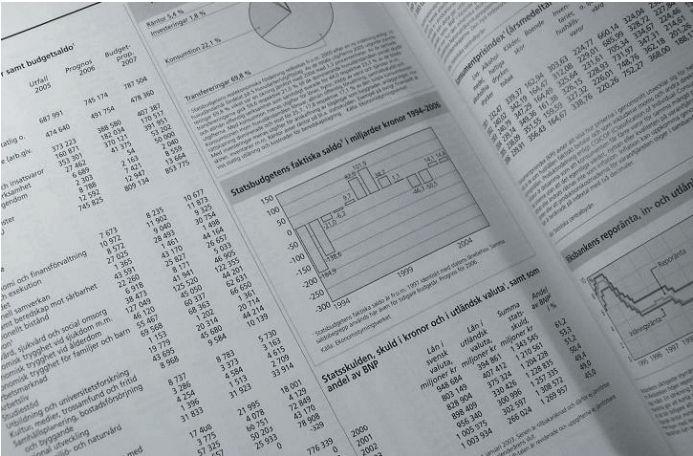
Slutsats

De finns där, 18 726 tänkbara kvinnor, där vi slutade den statistiska analysen. Till en början, så kan ett underlag på 18 726 kvinnor inte verka så tokigt. Men tänk så här: Anta att jag går på blind-dejt med en ny tjej varje vecka, som befinner sig i rätt ålder. Då skulle jag behöva 3493 veckor innan jag hittade *en* av de 18 726. Det är nästan 67 år! Som en nordamerikan född i slutet av 70-talet så är min förväntade livslängd lite mer än 70 år, så vi kan säkert säga att jag är död innan jag har hittat mina drömmars flickvän. Men å andra sidan, så kommer hon också att hinna dö innan dess.

Publicerad med tillstånd av Tristan Miller, 2008-04-06. Originalet återfinns på en.nothingisreal.com/wiki/Why_I_Will_Never_Have_a_Girlfriend



BENFORDS BEGYNNELSESIFFROR



Benfords lag är egentligen mer statistik än sannolikhetslära. Men som du kanske märkt, så kryper de ämnena närmare och närmare varandra desto längre man kommer inom sannolikhetsläran. Och Benfords lag är rolig att känna till, så den har fått komma med i boken.

Om du frågar någon du möter om vilket husnummer han eller hon bor på förväntar du dig förmodligen att sannolikheten är 1/9 att numret ska börja med en etta, 1/9 att numret ska börja med en tvåa och så vidare. Det finns ju nio möjliga siffror (eftersom husnummer inte börjar med nollor), och sannolikheten borde väl vara lika stor för varje begynnelsesiffra?

Benfords lag

Fysikern Frank Benford formulerade 1938 en lag om begynnelsesiffrors frekvenser i statistiska tabeller. (Egentligen var det inte Benford som var först med att upptäcka fenomenet, och det var först på 1980-talet som Theodore Hill bevisade den, men den bär ändå Benfords namn.) Lagen är lika enkel som den först låter osannolik:

Sannolikheterna, P, för respektive förstasiffra, D, i en lista, statistisk tabell eller motsvarande är fördelade enligt följande:

D	P	D	P
1	30,10%	6	6,69%
2	17,61%	7	5,80%
3	12,49%	8	5,11%
4	9,69%	9	4,58%
5	7,92%		

(För den matematiskt hugade, så svarar tabellen mot formeln $P = \lg(1 + 1/D)$)

Tabellen säger bland annat att sannolikheten för att ett tal, slumpmässigt valt ur en statistisk tabell (eller motsvarande), ska börja med en tvåa är ungefär 0,18 och att det börjar med en femma är ungefär 0,08. Lagen talar alltså tvärtemot det sunda förnuftet och det vi lärt oss hittills! Men den stämmer, och det finns det många exempel på.

Benfords lag går att använda på husnummer. Benford själv publicerade en undersökning där han kontrollerade att lagen stämde på gatuadresserna till det amerikanska fysikerförbundets medlemmar. Han publicerade dessutom undersökningar där han visade att lagen stämde överens med prislistor, sportresultat, tal i tidningsartiklar, ytor som floder avvattnar och elektricitetsräkningar.

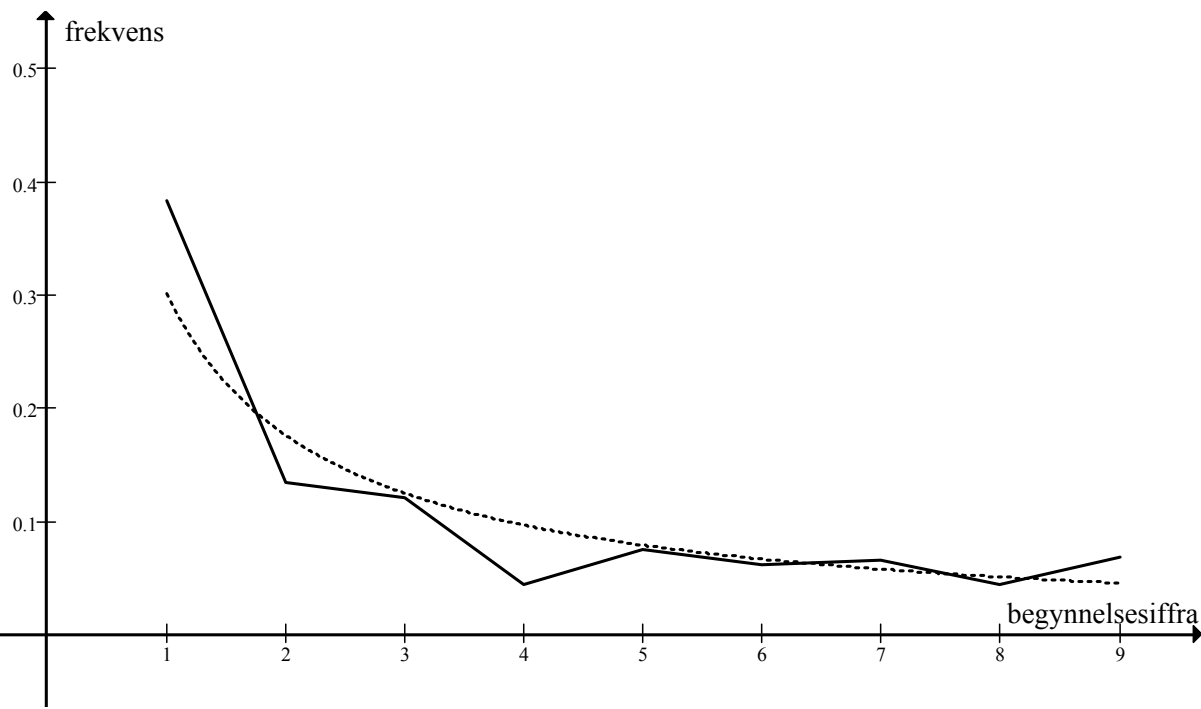


Kommunstatistik

För att titta på lite mer aktuella data, och testa dem mot Benfords lag, så grupperade jag Statistiska Centralbyråns statistik över antalet invånare per kommun i Sverige den sista december 2006 efter begynnelsesiffror. Den minsta kommunen hade 2 541 invånare (Bjurholm), och den största 782 885 invånare (Stockholm), och sammanlagt fanns det då 9 113 257 invånare i hela Sverige. Det finns 290 kommuner i Sverige, ett relativt stort tal, så man borde kunna vänta sig att Benfords lag stämmer något sänär på den här statistiken. På nästa sida finns resultatet.

Fördelningen mellan olika begynnelsesiffror, D, i antalet invånare per kommun den 31 december 2006

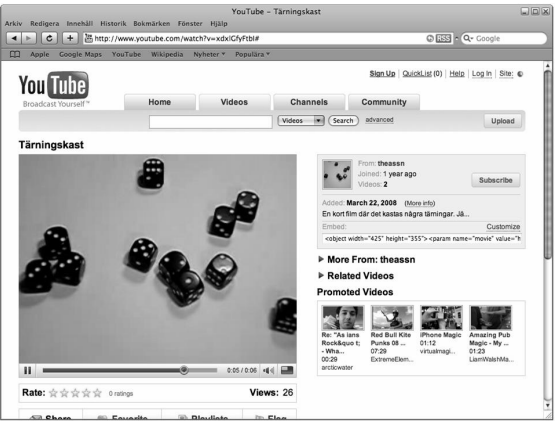
D	antal	frekvens	Benfords lag
1	111	38,28%	30,10%
2	39	13,45%	17,61%
3	35	12,07%	12,49%
4	13	4,48%	9,69%
5	22	7,59%	7,92%
6	18	6,21%	6,69%
7	19	6,55%	5,80%
8	13	4,48%	5,11%
9	20	6,90%	4,58%



Den prickade kurvan är Benfords lag, och den heldragna är den relativa frekvensen av respektive förstasiffra i den nyss nämnda kommunstatistiken. Det är lite avvikelser här och var mellan kurvorna, men det är ingen tvekan om att kurvorna påminner om varandra! Hade vi haft ännu fler kommuner i Sverige, så att vi hade fått ännu fler siffror i underlaget, hade vi enligt Stora talens lag förmodligen hamnat ännu närmare Benfords förutsägningar.

Youtube

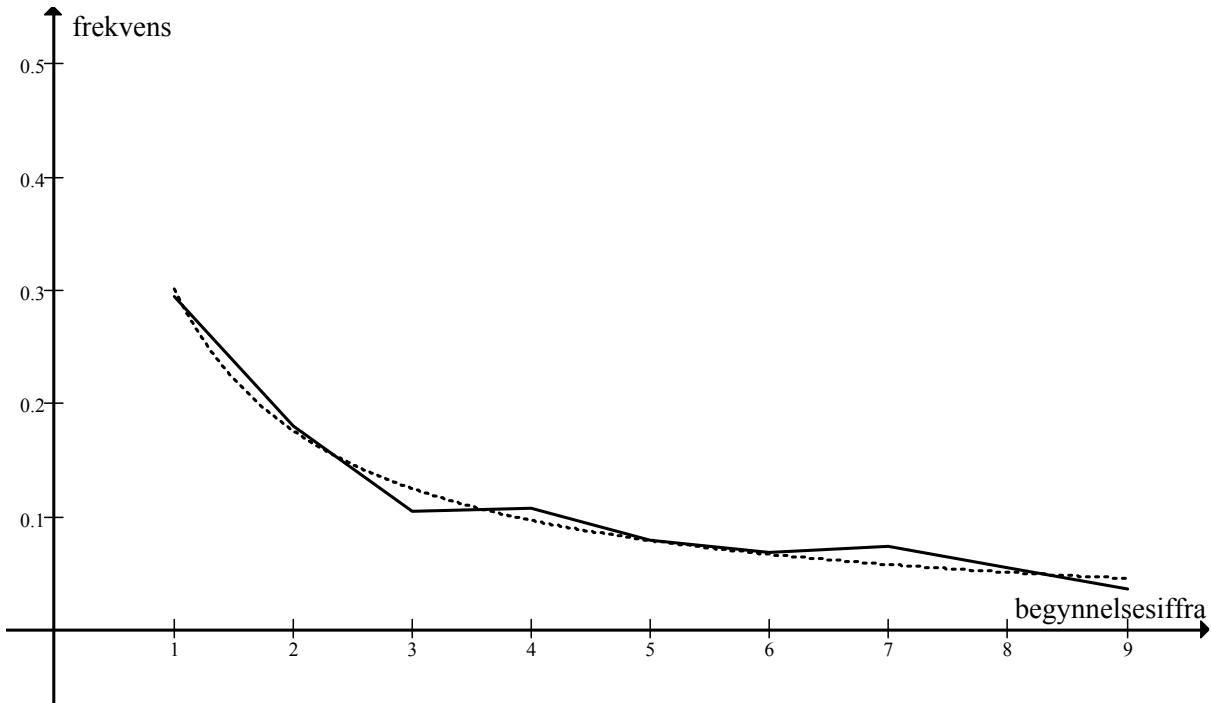
Kommunstatistik är kanske inte alla särskilt intresserade av. Men det finns – som du redan sett – även andra områden där Benfords lag gäller. Tittarsiffrorna på videosajten youtube.com, exempelvis. Där finns det, som många av er läsare säkert redan vet, en väldigt stor mängd korta videofilmer som användarna själva bidrar med, så att antalet filmer växer hela tiden. Användarna kan sedan ranka varandras filmer och kan bland annat se topplistor där de högst rankade filmerna hamnar, så att ännu fler kan hitta dem. Man kan även söka bland alla videoklipp med ord, exempelvis en särskild artist, och sökningen sker då i filmens titel och i en beskrivning, som användaren som bidrog med filmen har skrivit.



När man tittar på ett klipp kan man bland annat se hur många gånger klippet har visats, och det finns även länkar till andra liknande klipp, så att man kan ”klicka sig fram” till nya filmer hela tiden. Jag klickade mig på detta sätt genom 1000 filmer, och antecknade hur många gånger varje klipp hade visats, en siffra som varierar från 1 till tiotals miljoner gånger. Eftersom underlaget nu var mer än tre gånger större än med kommunerna, så kan vi förvänta oss att frekvensen för de olika begynnelsesiffrorna borde vara närmare vad Benfords lag förutsäger än för kommunerna, enligt Stora talens lag. Hur det gick framgår av följande tabell och graf:

Fördelningen mellan olika begynnelsesiffror, D , i antalet visningar per filmklipp, av 1000 klipp från youtube.com

D	antal	frekvens	Benfords lag
1	294	29,40%	30,10%
2	180	18,00%	17,61%
3	105	10,50%	12,49%
4	108	10,80%	9,69%
5	80	8,00%	7,92%
6	68	6,80%	6,69%
7	74	7,40%	5,80%
8	55	5,50%	5,11%
9	36	3,60%	4,58%



Den streckade linjen är Benfords lag, den heldragna är frekvensen för de olika begynnelsesiffrorna bland tittarsiffrorna på youtube.com. Inte alls tokigt, och här har vi drygt det trefaldiga underlaget jämfört med kommunerna – och visst stämmer Benfords lag ännu bättre här, helt enligt Stora talens lag.

Att denna lag stämmer så pass bra går att utnyttja. Det har utvecklats datorprogram som analyserar bokföring och bygger på Benfords lag för att se om kostnaderna och inkomsterna är fejkade eller inte. Om bokföringen är riktig, så ska alltså ungefär 30% av talen börja med en etta, och 18% med en tvåa och så vidare. Men en påhittad bokföring gör sällan det, eftersom människan ofta inte är särskilt medveten om hur verklighetens ”slumptalsgeneratorer” fungerar.

Hur gick det med normalfördelningen?

Säger inte Benfords lag emot normalfördelningsprincipen från kapitel tolv då? Den handlar också om statistiska värden, men säger att det ska finnas mest mätvärden kring medelvärdet och att de ska vara symmetrisk fördelade.

Det är en bra invändning, men det finns en väldigt viktig skillnad mellan de två principerna: Benfords lag handlar endast om begynnelseffran. Exempelvis 12 och 125 hamnar här i "samma kategori", medan de i en normalfördelning hamnar långt från varandra. De är alltså inte jämförbara.

Statistiken som Benfords lag går att tillämpa på kan vara normalfördelad, men behöver inte vara det. För husnumren är det förmodligen 1, 2 och 3 som är de vanligaste värdena, och sedan avtar de uppåt, eftersom alla gator inte är lika långa. Men nedåt så tar det bara abrupt slut (eftersom man inte kan ha negativa husnummer), så de är inte normalfördelade. Likaså är kommunernas storlekar inte naturligt fördelade – skulle man titta på hur många det bor i varje tätort, skulle det nog kunna tänkas vara normalfördelat, men kommunerna är sammanslagna för att få lagom stora administrativa enheter. Även om det skiljer väldigt mycket mellan största och minsta kommunen, så är storlekarna ändå tillrättalagda för att vara praktiska att hantera för samhället, och därför finns det väldigt få kommuner med färre än 5000 invånare, exempelvis, trots att det finns gott om tätorter med betydligt färre invånare.

Skonummer däremot är typiskt normalfördelad statistik, och här gäller inte Benfords lag, eftersom vi rör oss inom ett relativt begränsat intervall, kanske mellan 30 och 49, så att andra

begynnelseffrorna än tre och fyra nästan aldrig kommer att förekomma.

Benford själv visade att ytor på områden som avvattnas av olika floder omfattas av Benfords lag, och de är med största sannolikhet även normalfördelade. Så visst går det att kombinera Benfords lag och normalfördelningen. Men det beror som sagt på hur värdena ser ut.

Men varför blir det på detta viset?

Jag tänker inte försöka bevisa Benfords lag, men jag ska försöka motivera den så att den inte verkar så osannolik.

Det hela handlar om att vi inte har några specifika intervall, och att sannolikheten är inte samma för varje tänkbart utfall. När vi kastar en tärning så kommer resultatet alltid att hamna i intervallet 1-6, och sannolikheten är precis jämnt fördelad mellan de olika utfallen.

Men i en leksaksaffär är priserna spridda från en krona till några tusenlappar. De flesta priserna ligger kanske runt 50 till 150 kronor, och en hel del upp till 250 kronor och däröver bara några enstaka, vilket betyder att hela hundraintervallet, som börjar med en etta, blir väl representerade i statistiken, och tvåhundraintervallet får också en del av uppmärksamheten, men inte lika mycket. Femmorna till och med niorna som begynnelseffror i statistiken kommer från 50-100-intervallet, och treor och fyror samlas ihop med lite priser på såväl 30 och 40 som 300 och 400 kronor. Här har vi alltså inte alls samma klarhet i var intervallerna börjar och slutar, och sannolikheten att vi ska hitta något som exempelvis

kostar 105 kronor är större än att vi hittar något som kostar 210 kronor. (Var inte det här ett tillrättalat exempel, för att få Benfords lag att stämma? Vi skulle kunna tänka oss att det är flest priser kring 600 kronor istället, och då blir sexorna antagligen lite mer representerade i statistiken. Men å andra sidan, så finns det i så fall antagligen betydligt fler varor som kostar en tusenlapp och mer, där det samlas ihop ettor till statistiken igen.)

Om man tänker sig en pris-stegring på 25% per år i tio år, och priset börjar på 25 kronor, får man följande serie:

25, 31, 39, 49, 61, 76, 95, 119, 149, 186, 233.

Här får vi återigen en antydning om hur vanliga ettorna är, i förhållande till de andra begynnelsesiffrorna. Du får gärna räkna vidare själv, du ska se att ettorna kommer att fortsätta vara de mest förekommande oavsett var du slutar i serien!

Visst går det att hitta tillfällen då Benfords lag inte gäller, men då är det antagligen frågan om distinkta intervall, som med skonommer eller tärningskast.



15.

OSANNOLIKA SANNOLIKHETER



“Jag menar, före Harrisburg så var det ju ytterst osannolikt att det som hände i Harrisburg skulle hända, men så fort det hade hänt rakade ju sannolikheten upp till inte mindre än 100 procent så att det blev nästan sant att det hade hänt.”

Så sa Tage Danielsson i sin monolog i lunchrevyn ”Under dubbelgöken”, som gick 1979-80, apropå kärnkraftsomröstningen som snart skulle äga rum samt kärnkraftverksolyckan i Harrisburg den 28 mars 1979.

Det här sista kapitlet handlar om händelser som är så otroliga att man lätt kan tro att de aldrig kommer att inträffa, men ändå gör de det. På slutet finns det också några ord om ”kupongsamlarens dilemma”, som det kallas i facklitteraturen.

Försumbara sannolikheter?

Tage Danielsson beskriver en försumbar sannolikhet som att ”den finns inte, fast bara lite” i sin nyss nämnda monolog. Sannolikheten för att en olycka ska inträffa på ett kärnkraftverk kan kanske kallas för försumbar, eftersom den är så otroligt liten? Risken att träffas av blixten på väg till jobbet är väldigt liten och det är nog väldigt få som oroar sig för det, så den risken kan nog också klassas som försumbar. Chansen att vinna Drömvinsten på en rad på Lotto är 1 på ungefär 340 miljoner (se kapitel åtta), och kan om något anses försumbar.

Trots det, så har alla dessa händelser inträffat, flera gånger dessutom. Bara för att man anser att sannolikheten för en händelse är så otroligt liten att den är försumbar, betyder det inte att den aldrig kommer att inträffa någon gång, någonstans. Till en början kan det verka motsägelsefullt, men det handlar om att det finns två ”nivåer”: ”just här och nu” respektive ”någon gång någonstans”.

Rekord

Sannolikheten att det sätts ett världsrekord i just 10 000 meters löpning för herrar vid nästa sommar-OS är antagligen inte så stor. Men å andra sidan är sannolikheten att det sätts *något* nytt världsrekord vid nästa olympiska spel förmodligen rätt stor, och på samma sätt är gissningsvis sannolikheten att det *någon* gång inom de kommande 20 åren sätts ett nytt världsrekord i löpning 10 000 meter för herrar också relativt stor.

Det jag vill poängtera här är att även om sannolikheten att något särskilt ska inträffa just *här och nu* kan vara liten, så är sannolikheten att något ska inträffa *någon* gång *någonstans* ändå stor, och därför inträffar det förr eller senare.

Det här resonemanget gäller analogt med resonemanget om kärnkraftsolyckor. Om Harrisburg kan man säga som Tage Danielsson: ”En sån olycka inträffar ju enligt alla sannolikhetsberäkningar bara en gång på flera tusen år, och då är det ju i varje fall inte troligt att den har hänt redan nu.” Nej, det var inte särskilt troligt att just olyckan i Harrisburg skulle inträffa, men att det skulle inträffa någon kärnkraftverksolycka *någon* gång *någonstans* var (och är naturligtvis fortfarande) å andra sidan troligt.

Lite senare i monologen gör han dock ett matematiskt felsteg: ”Sen är det också det att om det som hände i Harrisburg verkligen hände, mot förmodan, så är ju sannolikheten för att det ska hända en gång till, den är ju så oerhört löjligt jätteliten så att på sätt och vis kan man säga att det var nästan bra att det som hände i Harrisburg hände, om det nu gjorde det. För jag menar, då kan man ju nästan säkert säga att det inte kommer hända igen.” Sannolikheten att något väldigt osannolikt ska hända två gånger i rad är just ”oerhört löjligt jätteliten”, men att det ska hända *en andra gång* när det redan har inträffat en första är lika stor som den var då händelsen inträffade första gången! (Vi bortser från att Harrisburg-olyckan förhoppningsvis medförde att säkerheten skärptes på andra kärnkraftverk för att undvika just denna typ av olycka, eftersom det finns så många andra sätt som olyckor kan inträffa på, som uppenbarligen inte åtgärdades: Tjernobylolyckan inträffade endast sju år senare!)

Skillnaden mellan sannolikheten att något ska inträffa två gånger i rad respektive en andra gång nämndes redan i det första kapitlet, men det handlade då om slantsinglingar – sannolikheten att få sex klavar i rad är liten, men sannolikheten att få den sjätte klaven när man redan har fem i rad är exakt 0,5.

Försäkringar

Hur stor är sannolikheten att du riskerar att bli utsatt för ett inbrott i hemmet? Fundera på hur många av dina bekanta som har råkat ut för det under det senaste året, så kommer du troligen fram till att det kanske handlar om i storleksordningen någon procents risk på ett år, inte mycket mer. Hur mycket har du och dina bekanta betalat under det senaste året i försäkringar, sett i proportion till hur mycket ni har fått tillbaka från försäkringsbolaget? Försäkringsbolaget har antagligen gjort en ganska stor vinst på er. Men samtidigt vill du inte stå där utan ersättning när inbrottstjuven gjort ett besök.

Tanken på att det *skulle* kunna hända duger tydligen för att motivera försäkringskostnaderna, även om risken att råka ut för det är så liten att vi nästan skulle kunna kalla den för försumbar. På samma sätt motiverar uppenbarligen tanken att man *skulle* kunna vinna det svenska folket att spela för ungefär 11 miljarder kronor per år hos Svenska Spel. Trots att det allra största flertalet av dessa spelare förlorar på spelandet.

Försäkringsbolag går med vinst, om det råder det nog ingen tvekan. Och för att ett bolag ska gå med vinst så måste man ju få in mer intäkter än vad man har utgifter, det är inga konstigheter.

Om alla kunder hos ett försäkringsbolag skulle råka ut för ett inbrott, så skulle bolaget förmodligen behöva betala tillbaka mer i ersättningar än vad de har fått in av kunderna när de betalat sina försäkringspremier, så för bolagen är Stora talens lag väldigt intressant. Som du kanske minns från kapitel fem så handlade Stora talens lag om ett ”stort antal”, och att man hamnar närmare och närmare väntevärdet ju fler observationer man gör – eller i försäkringsbolagets fall: ju fler kunder man har. Ett bolag som bara har tio kunder har svårt att förlita sig på Stora talens lag. Det skulle inte vara helt otroligt om alla tio kunder råkade ut för allvarliga olyckor under samma år, så att bolaget skulle gå minus det året. Ett bolag med tio tusen kunder kan däremot förlita sig på Stora talens lag, eftersom de har ett så stort antal kunder så att risken för att alla skulle råka ut för allvarliga olyckor under samma år är i praktiken obefintlig för dem.

En del myndigheter försäkrar inte sina datorer mot stöld, eftersom försäkringen skulle kosta mer än vad inköp av nya datorer kostar. Anledningen till det torde vara att försäkringsbolaget ser risken så pass stor att de kommer att bli stulna, att de väljer att sätta en premie som motsvarar att de behöver köpa in nya (plus lite till, för vinsten). För privatpersoner däremot kalkylerar försäkringsbolagen förmodligen med att i storleksordningen några procent av kunderna verkligen kommer att råka ut för olyckorna de försäkrar sig mot, och kan därför sätta en lägre premie.

I hemmen finns det många som försäkrar elektronik, som ljud- och filmanläggningar, utifall de skulle gå sönder av



exempelvis ett asknedslag. Men om man överväger vad det skulle kosta att köpa in ny utrustning, jämfört med hur liten risken faktiskt är att de går sönder på ett sådant sätt att de täcks av försäkringen, kommer man antagligen komma fram till att det kanske inte är så lönsamt. (Resonerar vi utifrån att försäkringsbolaget med all säkerhet går med vinst, kan vi ju vara säkra på att alla försäkringstagare tillsammans i det långa loppet går med förlust på att försäkra utrustningen, även om en särskilt olycksdrabbad försäkringstagare kanske kan tjäna på det.)

Poängen med att försäkra exempelvis ett hus är väl i första hand att man inte vill råka ut för någon obehaglig privatekonomisk överraskning, om något skulle inträffa och man inte har råd att ersätta vad det nu är. Men när det gäller TV-apparater, mobiltelefoner och stereoanläggningar har nog de flesta råd att bekosta en ny den dagen den går sönder, och som det nämndes nyss så gör man antagligen en förlust i längden med att försäkra dem, jämfört med om man inte hade försäkringen.

Avstår man försäkringar får man å andra sidan vara beredd att ta smällarna de gånger man annars skulle ha utnyttjat försäkringen. Vetskapen om att man är försäkrad att slippa stora oväntade engångskostnader är alltså något man i längden får betala för, även om man utnyttjar försäkringen emellanåt. Men å andra sidan är det kanske något som det är värt att betala för?



Samlarkort

För ungefär tio år sedan samlade nästan alla unga på Pokémon-kort. Nu för tiden är Pokémon inte alls lika stort, men det finns andra ersättare, som till exempelvis Digimon och Duel Masters. Bilder på hockeyspelare och artister har varit andra typer av samlarobjekt som har varit populära, och det finns och har funnits otaliga samlarserier i kex- och cornflakespaket och i chokladägg, så kallade Kinder-ägg.

De flesta av dessa har det gemensamma att man inte har någon aning om vilket av alla samlarobjekten som finns i dem när man köper dem. Pokémon-kort köptes i flerpäck i aluminiumfolieförpackningar, och i ett Kinder-ägg kan man inte se vad man har fått förrän man har ätit upp chokladen.

Låt oss som exempel tänka oss en samlarserie med 50 kort, som på något sätt säljs styckvis utan att man kan se vilket av alla korten det är. Vi antar att alla kort är lika vanliga. Hur många köp behöver man i medeltal göra för att få ihop en komplett serie, alltså alla 50 korten?

Till att börja med, så måste vi åtminstone ha 50 köp, för att ens ha den hypotetiska möjligheten att få ihop alla. Men det lär krävas betydligt fler – när vi har fått ihop 49 kort, så finns det i nästa köp 50 möjliga kort, varav endast ett är det vi efterfrågar. Vi bör alltså vara beredda på att behöva göra i medeltal ungefär 50 köp bara för att få det sista vi saknar!

På det här sättet kan vi med fördel resonera oss fram till hur många inköp vi behöver göra: För det första kortet räcker det med att göra ett inköp, vi får ju ett kort vi vill ha oavsett vilket det är (eftersom vi inte har något sedan tidigare).

För det andra kortet finns det 49 kort vi vill ha, och 50 möjliga. Det ger oss sannolikheten 49/50, vilket innebär att 49 av 50 köp ”lyckas”. Ett smidigt begrepp att införa här vore ”antalet köp per lyckade kort”, det vill säga hur många kort man i medeltal behöver köpa för att få rätt kort. Sannolikheten att få ett ”lyckat kort” innebär ju hur många kort som i medeltal blir ”lyckade” per köp. Antalet köp per lyckade kort är då inversen till sannolikheten, vilket betyder att det bara är att ta just inversen på sannolikheten, ”vända sannolikheten upp-och-ner”, för att få antalet köp per lyckade kort.

För kort nummer två i serien blir då antalet köp per lyckat kort 50/49 köp, alltså lite mer än 1, vilket ju är mycket rimligt med tanke på att det finns ett kort vi inte vill ha och 49 kort vi vill ha, och att det oftast borde bli något av dessa 49.

För det tredje kortet i serien blir antalet köp per lyckade kort 50/48, och så fortsätter det till det femtionde kortet, där vi redan har konstaterat att antalet köp per lyckat kort är 50/1, eller helt enkelt 50.

Summan av antalet köp per lyckade kort för alla kort, från det första till det femtionde, blir då enligt följande (vi skriver 50/50 istället för 1, för att få en mer lättanvänd talserie):

$$50/50 + 50/49 + 50/48 + 50/47 + \dots + 50/1 \approx 225$$

Vi behöver alltså göra i medeltal 225 köp för att få alla våra 50 samlarbilder. Som du säkert förstår så kan det både behövas fler och färre köp, men det ger en lämplig fingervisning om vad det är man ger sig in på om man exempelvis slår vad om att man ska hinna före någon annan att få ihop alla figurer från cornflakes-paketen.

Det kan vara värt att notera att det knappast är de första tio olika objekten som drar iväg i antal köp, utan det är att få ihop de sista tio som gör det så omständligt och (förmodligen) dyrt att få ihop hela serien. Det här fenomenet, som förövrigt kallas för ”kupongsamlarens dilemma”, med att det är ”de sista som kostar” för oss raskt vidare in i nästa lilla avdelning.

Skillnaden mellan sant och sannolikt

Du minns kanske födelsedagarna från slutet av kapitel sex? Där konstaterade vi att det räckte med endast 23 personer för att vi skulle få en sannolikhet över 0,5 för att minst två personer hade en gemensam födelsedag. Om man studerade tabellen kunde man också konstatera att redan vid 30 personer låg samma sannolikhet över 0,7. För att fortsätta där tabellen slutade, så är det vid 40 personer en sannolikhet kring 0,9 för minst två gemensamma födelsedagar, och för 50 personer ca 0,97. Men sannolikheten blir inte 1 förrän vid 366 personer, då vi säkert kan veta att det måste finnas en dubblett! Detta eftersom det bara finns 365 dagar på året (Vi bortser fortfarande från skottdagen. Den läsare som irriterar sig på detta kan naturligtvis lägga till skottdagen i resonamang).

Vid en folksamling på 50 personer kan vi alltså utropa att ”jag slår vad om att minst två personer har en gemensam födelsedag” och känna att vi har *sannolikheten* på vår sida med god marginal. Men, det är inte förrän med ytterliggare 316 personer som vi kan vara *säkra*!



EFTERORD

Hoppas att du, när du nu antingen har läst hela boken från pärm till pärm eller hoppat mellan kapitlen och till slut hamnat här, känner att du har lärt dig något. Oavsett om du tyckte att varenda sida har varit fylld med självklarheter, eller du har läst varenda kapitel tre gånger i hopp om att i varje fall förstå lite grann, så är min önskan att du ändå har lärt dig *någonting!*

Är du nu intresserad av att gå vidare, så har jag flera hänvisningar till olika populärvetenskapliga böcker inom området på kommande sidor. Men det allra roligaste vore förstås om du väljer att gå någon utbildning inom sannolikhetslära och statistik!

Känner du dig mätt på temat, så är det bara att ställa tillbaka boken i bokhyllan eller lämna tillbaka den till biblioteket. Förhoppningsvis så har lite av dina nyvunna kunskaper fastnat, så att du härnäst kan vandra runt i vardagen och känna dig lite klokare än de flesta andra. Du kan briljera med taktiker för hur man bör spela på Lotto, hur principerna för sannolikheter i kortspel fungerar, hur man kan kombinera olika element, vad slump är för något, lurigheterna med att försöka samla hela samlarserier, hur händelser jämnar ut sig i det långa loppet, vad normalfördelningen är, vad Benfords lag innebär, hur kuvert- och getproblemen fungerar, vad trollen gör i farstun och mycket, mycket mer!

Har du några funderingar eller kommentarer om boken och dess innehåll som du vill dela med dig av, så hittar du min e-postadress alldeles i början av boken.

Andreas

FÖR DEN INTRESSERADE LÄSAREN



Till den som vill utföra beräkningar på egen hand, men inte har tillgång till en räknare med de funktioner som ibland behövs, rekommenderar jag ”On-Line Calculator” på University of South Carolinas webbplats, www.math.sc.edu/cgi-bin/sumcgi/calculator.pl som har alla funktioner som behövs för uppgifterna i den här boken. Den kräver några minuter innan man förstår hur man ska använda den, men sedan är den mycket smidig och kraftfull!

Vill man göra slumptalsserier med exempelvis tärningar, mynt eller siffror så är random.org en smidig webbsida om man är lat och inte vill ägna tid åt att göra dem på egen hand.

Till er läsare som vill pröva att rita normalfördelningar, och grafer i allmänhet, på egen hand vill jag rekommendera det lilla gratisprogrammet Graph, som finns att hämta fritt på padowan.dk/graph/ för den som så önskar. Den finns dock endast för operativsystemet Windows.

Det finns många böcker att gå vidare med, som både sannolikhetslära och statistik, och matematik i allmänhet. När det gäller sannolikhetslära- och statistikböcker börjar de i allmänhet på en något högre matematisk nivå än den här boken, men efter att ha läst den här boken så borde du gott kunna gå vidare med någon av dessa böcker. Allan Guts böcker finns normalt på stadsbibliotek, och de flesta andra borde gå att hitta på något högskole- eller universitetsbibliotek (ditt lokala bibliotek kan säkert göra fjärrlån från dem), alternativt köpa via någon välsorterad bokhandel

eller internetbokhandel. De böcker jag föreslår här är mer eller mindre populärvetenskapliga, men utöver dessa så finns det också konventionella läroböcker.

- Sant eller sannolikt - Tankar kring matematik, statistik och sannolikheter. *En läsvärd bok skriven av Uppsala-professorn Allan Gut. Till stor del orienterande, utan matematiska övningar. Handlar främst om sannolikhetslära, men även en del statistik. Utgiven på Norstedts förlag.*
- Konsten att räkna - tankar om siffror och statistik. *Allan Gut, samma författare som Sant eller sannolikt, men i den här boken behandlar han mer matematik i allmänhet och statistik i synnerhet. Norstedts förlag.*
- Slumpens skördar. *Chalmersprofessorn Olle Häggström är författaren till den här boken, som främst riktar sig till läsare på gymnasie- eller högskolenivå, som inspirerande bredvidläsning. Studentlitteratur.*
- The Pleasures of Probability, av Richard Isaac. Påminner om den här boken, fast med högre tempo, fler formler och färre bilder. För den som på egen hand vill lära sig räkna ordentligt med sannolikhetslära. På engelska. Förlaget Springer.
- Taking Chances - Winning with Probability. Mig veterligt finns det ingen svensk översättning av denna mycket välskrivna bok av John Haigh på University of Sussex, som främst behandlar sannolikheten och strategier i spel av olika slag. Oxford University Press.
- The Jungles of Randomness - A Mathematical Safari. *En bok nästan helt utan matematiska formler som handlar mer om fenomenet slumpen och dess mönster runt omkring oss. Skriven av den prisbelönta Ivars Peterson. Även denna bok saknar svensk översättning. Utgiven på John Wiley & Sons.*

SVAR OCH LÖSNINGAR TILL ÖVNINGSUPPGIFTERNA

1. Första kapitlet

1.

Rätt ordning är d), a), c) och sist b).

Sannolikheterna för de respektive händelserna är ungefär

d) 0,56 (väldigt vanligt)

a) 0,24 (hyfsat vanligt)

c) 0,06 (inträffar ibland, men inte så ofta)

b) 0,001 (ovanligt)

2.

Mer än 0,5.

Det du behöver göra är att räkna i tabellen hur många par det finns där den vänstra tärningen har samma eller ett högre antal ögon än den högra. Det är 21 stycken, så sannolikheten blir $21/36 = 7/12$ eller ungefär 0,58, alltså mer än hälften.

2. Sannolikhetskalkylens grundvalar

1.

a) $11/100$

b) $11/99 = 1/9$. Det är en färre lott nu, eftersom den redan har gått till förstapriset.

2.

a) Chansen (eller risken, välj själv) att Färjestad vinner över Frölunda i nästa match är 60% (eller 6 på 10 eller 3 på 5)

b) $P(\text{Jag kommer imorgon}) = P(\text{Min kollega kommer imorgon}) = 0,5$

3.

a) $1/2 \cdot 1/2 \cdot 1/2 = 1/8 = 0,125$

b) $1/2 \cdot 1/2 \cdot 1/2 + 1/2 \cdot 1/2 \cdot 1/2 = 1/8 + 1/8 = 1/4 = 0,25$

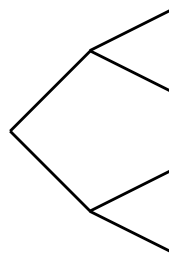
4.

a) $4/8 = 1/2$

b) Numrera en tjugosidig tärning (ikosaeder) med fyra ettor, fyra tvåor, fyra treor, fyra fyror och fyra femmor, så blir sannolikheten att få en femma $4/20 = 1/5$.

5.

a)



(V = Vinner F = Förlorar)

b) $P(\text{Thunderbolt förlorar det första och vinner det andra loppet}) = 0,6 \cdot 0,3 = 0,18$

c) $P(\text{Thunderbolt vinner minst ett lopp}) = 1 - 0,6 \cdot 0,7 = 0,58$

3. Getproblemet

2.

$2/3$

Det finns tre möjligheter till kungens enda syskon; antingen är det (a) en lillebror, (b) en storasyster eller (c) en lillasyster. Det finns inget som talar för att något av alternativen skulle vara mer troligt än något annat, alltså har varje alternativ sannolikheten $1/3$. I (b) och (c) är kungens syskon en kvinna, alltså blir sannolikheten $1/3 + 1/3 = 2/3$.

4. En slump?

1.

Alla svar på den här uppgiften är bara förslag, och man kan ha andra åsikter om hur man vill definiera slumpen och ha rätt ändå, men jag hoppas att du kan hålla med i stora drag.

a) Inte slump. Ibland kan vädret upplevas slumpartat, men vi har så pass god insyn i vädrets mekanismer att vi inte kan kalla morgondagens väder för slumpartat. Det finns förvisso många detaljer vi inte kan förutse, men i stora drag kan vi göra en träffsäker prognos för följande dag.

b) Slump. Så exakta prognoser kan vi inte göra, det är för komplext.

c) Slump. Vi vet kanske inte ens hur många det finns och inte heller hur fördelningen är mellan de olika färgerna.

d) Slump. Är bollarna, eller hur man nu drar siffrorna, väl blandade och lika stora så kan vi inte avgöra vilken som kommer att dras.

e) Tveksamt. Är det slumpen som avgör vilket lag som vinner en fotbollsmatch? Eller kan vi förutsäga det? Svaret lämnas över till den sportintresserade läsaren.

f) Ingen slump, bara slarv från min sida.

g) Slump, när den verkligen är välblandad.

6. När man talar om trollen

1.

Nej, du ska inte sjunga sången.

Sannolikheten att någon har sin födelsedag idag är $1/365$, men som du förmodligen märkt i kapitlet är komplementhändelserna oftast smidigare att använda. Sannolikheten att en person inte har sin födelsedag idag är $1 - 1/365 = 364/365$. Sannolikheten att inte person A och inte person B har sin födelsedag idag är $364/365 \cdot 364/365$. Alltså blir sannolikheten att ingen av de 500 personerna har sin födelsedag i dag $(364/365)^{500} = 0,254$, vilket ju är större än 0,2.

7. Kombinera!

1.

1 594 323

Det finns 3 möjligheter 13 gånger, det vill säga 3^{13} möjliga kombinationer, eller 1 594 323 möjliga sätt att fylla i en tipsrad. Det är alltså inte så värst stor chans att "råka" tippa rätt.

2.

Att e-postadresserna skulle vara slut är en absurd tanke, det räcker med att vi tittar på e-postadressen framför @-tecknet. Vi tänker oss att det får vara maximalt 15 tecken framför @ (i själva verket får det vara många, många fler, men med fler än 15 tecken börjar en e-postadress bli otymplig att använda) och att det finns 40 tillåtna tecken (a-z, 0-9 och lite specialtecken), så finns det till en e-postadress med 15 tecken 40^{15} möjliga kombinationer, eller lite drygt en kvadriljon! På jorden bor det ungefär sex miljarder människor, det skulle betyda att varenda människa på jorden hade registrerat 200 miljarder adresser var för att de ska ta slut, och det är ju helt orimligt. Om man dessutom tänker på att e-postadresser inte alls behöver vara just 15 tecken långa, och att en stor del av världens befolkning förmodligen inte alls använder MSN, så blir det ännu mer absurt.

3.

1/6

Det finns $3! = 6$ olika sätt att placera etiketterna, varav ett är rätt. Den första burken har han tre möjliga etiketter till, och till den andra finns det två möjliga. Till den tredje finns det kvar en etikett, så $3 \cdot 2 \cdot 1 = 3! = 6$.

8. Öka din lottovinst

1.

1 – Nej, för låg summa

2 – Ja

3 – Nej, för låg summa

4 – Ja

5 – Nej, för låg summa

6 – Nej, alla nummer är på kanterna och för låg summa

7 – Nej, alla nummer är på kanterna

8 – Nej, för låg summa

9 – Nej, för låg summa

10 – Nej, för låg summa

11 – Nej, för låg summa

12 – Nej, för låg summa

9. Beroende händelse

1.

1/3

$$P(\text{tärningen visar ett jämnt tal}) = \frac{P(\text{tärningen visar en fyra och ett jämnt tal})}{\frac{1/6}{1/2}} = \frac{1/6}{1/2} = 1/3$$

2.

0,45

A: en person är student

B: en person röstar för

$P(\text{en person är student givet att hon röstar för}) =$

$$\frac{P(\text{en person är student och röstar för})}{P(\text{en person är student})} = \frac{0,2 \cdot 0,9}{0,4} = 0,45$$

KÄLLFÖRTECKNING

Källor vars innehåll använts och/eller refererats ett flertal gånger

Arné, Magnus (2001), *På irrfärd i slumpens värld*, Lund: Studentlitteratur

Gut, Allan (2007), *Konsten att räkna*, Stockholm: Norstedts

Gut, Allan (2002), *Sant eller sannolikt: Tankar kring matematik, statistik och sannolikheter*, Stockholm: Norstedts

Haigh, John (1999, 2003), *Taking Chances*, Oxford: Oxford University Press

Isaac, Richard (1995), *The Pleasures of probability*, New York: Springer-Verlag, New York, Inc.

Henze, Norbert och Riedwyl, Hans (1998), *How to win more: Strategies for increasing a lottery win*, Natick MA: A K Peters, Ltd.

Häggström, Olle (2004), *Slumpens skördar: Strövtåg i sannolikhetsteorin*, Lund: Studentlitteratur

Klefsjö, Bengt och Hellmer, Stefan (1991), *Räkna med slumpen*, Lund: Studentlitteratur

Klefsjö, Bengt och Hellmer, Stefan (1991), *Räkna med slumpen – Problemsamling*, Lund: Studentlitteratur

Peterson, Ivars (1998), *The jungles of randomness: a Mathematical safari*, New York: John Wiley & Sons, Inc.

Schneider, Reto U. (2004), *De galna experimentens bok*, Stockholm: Fahrenheit

Singh, Simon (1997), *Fermats gåta*, Stockholm: Norstedts

Thompson, Jan m. fl. (1991), *Wahlströms och Widstrands matematiklexikon*, Stockholm: Wahlströms och Widstrands

Särskilda fakta- och sifferuppgifter, citat etc.

Förord

Hawking, S. W. (2001). *Kosmos: en kort historik*. Stockholm: Pan

Inledning

Bibeln : Bibelkommissionens översättning : noter, parallellbänvisningar, uppslagsdel. (2001). Stockholm: Verbum

Svenska Spel: Om SvS – Företagspresentation – Presentation.[Elektronisk]
Tillgänglig: <http://www.svenskaspel.se/pl.aspx?PageID=421&menuid=8&parentid=4520> [2008-03-20].

3. Getproblemet

Sempler, Kaianders (1999), Getproblemet upprör läsare, [Elektronisk] *Ny Teknik*, 11 september. Tillgänglig: http://www.nyteknik.se/efter_jobbet/kaianders/article21099.ece [2007-09-10]

4. Det långa loppet

Nationalencyklopedin (1989-1996). Höganäs: Bra böcker
Liljegren, Ronnie (1993). Kol-fjorton-datering. I *Nationalencyklopedin*. Bd 11, s. 170.

Nationalencyklopedin (1989-1996). Höganäs: Bra böcker
Nationalencyklopedin, (1995) Stora talens lag. I *Nationalencyklopedin*. Bd 17, s. 292.

5. En slump?

Bergström, Lars m fl (2005). *Heureka! Fysik för gymnasieskolan kurs B – Mer om kvantmekanik och relativitetsteori*, Stockholm: Bokförlaget Natur och Kultur (Elektronisk) Tillgänglig: <http://www.heureka.nu/laromedel/heureka/lararstod/index.asp?i=pdf/HeuBKap18.pdf> [2008-05-17]

Nationalencyklopedin (1989-1996). Höganäs: Bra böcker
Håstad, Johan. (1995). Slumptal. I *Nationalencyklopedin*. Bd 16, s. 630.

Nationalencyklopedin (1989-1996). Höganäs: Bra böcker
Laurent, Bertel. (1993). Kvantmekanik. I *Nationalencyklopedin*. Bd 11, s. 565-567.

Nationalencyklopedin (1989-1996). Höganäs: Bra böcker
Ryde, Hans. (1992). Heisenberg, Werner. I *Nationalencyklopedin*. Bd 8, s. 494.

6. När man talar om trollen

About.com: Does a Cat always land feet-side down - What Makes Cats Work Page 2 [Elektronisk] Tillgänglig: http://cats.about.com/cs/catmanagement101/a/how_cat_work_2.htm [2008-03-27]

Matthews, R.A.J. (2001) Testing Murphy's Law: urban myths as a source of school science projects, *School Science Review* 83 Tillgänglig: <http://ourworld.compuserve.com/homepages/rajm/ssrsum.htm> [2008-04-01]

8. Öka din lottovinst

Norsk tipping: Lotto – Norsk Tipping [Elektronisk] Tillgänglig: <http://www.norsk-tipping.no/page?id=17&tp=lsr> [2008-03-27]

Svenska Spel: Lotto – Spelpresentation. [Elektronisk] Tillgänglig: <http://www.svenskaspel.se/pl.aspx?PageID=226&menuid=4&parentid=4121> [2008-03-24].

Svenska Spel: Viking Lotto – Spelpresentation. [Elektronisk] Tillgänglig: <http://www.svenskaspel.se/pl.aspx?PageID=239&menuid=4&parentid=4124> [2008-03-24].

12. Normalfördelning

Björk, Lars-Eric och Brolin, Hans (2002), *Matematik 3000 Breddningshäfte Sannolikhetslära*, Stockholm: Bokförlaget Natur och Kultur

Wolfram Mathworld: Normal Distribution – from Wolfram Mathworld. [Elektronisk] Tillgänglig: <http://mathworld.wolfram.com/NormalDistribution.html> [2008-03-26]

Nordisk familjebok (1904-1926). Stockholm: Nordisk familjeboks förlags aktiebolag
A Wahlberg och S. A. von Schéele (1904-1926). Galton. I *Nordisk familjebok*. Bd. 9 s. 654.

13. Tristan Miller: Varför jag aldrig kommer att ha en flickvän

Tristan Miller: Why I Will Never Have A Girlfriend [Elektronisk] Tillgänglig: http://en.nothingisreal.com/wiki/Why_I_Will_Never_Have_a_Girlfriend [2008-04-06]

14. Benfords begynnelseiffror

Sempler, Kaianders (1999), Sifferanalys avslöjar kreativ bokföring, [Elektronisk] *Ny Teknik*, 22 augusti. Tillgänglig: http://www.nyteknik.se/efter_jobbet/kaianders/article3526.ece [2007-10-15]

Statistiska Centralbyrån: Folkmängd i riket, län och kommuner 31/12/2006 och befolkningsförändringar 2006 – Statistik från SCB. [Elektronisk] Tillgänglig: http://www.scb.se/templates/tableOrChart____193256.asp [2008-01-12].

Theodore P. Hill: Ted Hill's Home Page [Elektronisk] Tillgänglig: www.math.gatech.edu/~hill/ [2008-03-26]

Wolfram Mathworld: Benford's Law – from Wolfram Mathworld. [Elektronisk] Tillgänglig: <http://mathworld.wolfram.com/BenfordsLaw.html> [2008-03-26]

15. Osannolika sannolikheter

Svenska Spel: Pressrum – Statistik. [Elektronisk] Tillgänglig: <http://www.svenskaspel.se/pl.aspx?PageID=3487&menuid=8&parentid=4845> [2008-03-24].

Schöier, S. & Wermelin, S. (2005). *Hasse & Tage : Svenska ord & Co : saga & sanning*. Stockholm: Bonnier

Bildkällor

Alla bilder och all grafik är producerade av Andreas Svensson med ett undantag: Bilder på enskilda mynt (bland annat i grafik, kapitel 2): Riksbanken

Grafer är producerade i programmet Graph 4.3 som finns att hämta fritt på www.padowan.dk/graph

REGISTER

Additionssatsen	27
Apa	73
Benedictus XIII	11
Benford, Frank	146
Benfords lag	145
Beroende	26
Beroende	103
Betingad sannolikhet	107, 116
Bibeln	11
Chans och risk	15
Clement XII	12
Danielsson, Tage	155
Diskret	134
Drömvinsten	90, 92
Eller	27, 28
Fakultet	80
Fermat, Pierre de	11
Fibonaccis talföljd	97
Försäkringar	158
Galton, Francis	130
Getproblemet	37
Heisenbergs obestämbarhetsrelation	50
Henze, Norbert	93
Hill, Theodore	18, 146
Häggström, Olle	112
Katter	67
Klass, gemensamma födelsedagar	70, 163
Klass, gå in i ett klassrum	80
Kombinationer	81, 91

Kombinatorik	77
Kommuner	147
Komplementhändelser	32
Kontinuerlig	134
Kortlek	105, 121
Kulpåsar	104
Kupongsamlarens dilemma	161
Kuvertproblemet	111
Laplace, Pierre-Simon	11
Lotto	12, 87
Lotto	87
Miller, Tristan	139
Modell	46, 61
Multiplikationssatsen	26
Musikspelare	106
Normalfördelning	129
Och	26, 28
$P(A)$	25
Pascal , Blaise	11
Permutationer	80
Platonsk kropp	35
Procent	16
Radioaktivt sönderfall	49, 55
Registreringsskyltar	78
Rekord	156
Riedwyl, Hans	93
Russel, Bertrand	12
Samlarkort	161
Sannolikhet	15, 45
Slump	45
Slumptalstabell	51
Smögåsar	68
Standardavvikelse	132
Stora talens lag	53

Svenska Spel	11, 19, 88, 158
Svälta räv	121
Talesätt	63
Théorie analytique des Probabilités	11
Träd	30
Tumbling Toast Test	68
Utfall	16
Viking Lotto	88
Väntevärde	54
YouTube	149

