Statistisk signifikans och Armageddon

Olle Häggström

1 Inledning

Tidningar och andra media svämmar över av statistiska upplysningar. Någon viss cancerform är vanligare bland rökare än bland ickerökare, svenska folket dricker allt mer läsk, eller väljarstödet för regeringspartiet har minskat. Ofta sägs resultaten vara ”statistiskt säkra”, men hur många läsare har egentligen klart för sig vad det begreppet står för? Min avsikt med denna text är att förklara just det. Dock skall jag undvika att kalla det ”statistiskt säker”, och istället använda uttrycket **statistiskt signifikant**, som (oftast) betyder samma sak, men som undviker allt vilseledande tal om säkerhet. Statistiska resultat är så gott som aldrig helt säkra.

De statistiska begreppen är lättast att begripa i ett konkret exempel. Strax, i Avsnitt 2, skall jag gå igenom dem i samband med ett exempel inbegripandes slantsingling. Den som önskar komplettera med en mer systematisk och grundlig genomgång av begreppen kan med fördel vända sig till valfri grundläggande lärobok i matematisk statistik, förslagsvis Alm och Britton [AB] eller Olofsson och Andersson [OA].

Det avslutande Avsnitt 3 kan ses som överkurs. Där skall jag ta med läsaren på ett jättesprång, från det enkla och renodlade slantsinglingssammanhanget till något rent grandiost. Jag syftar på det så kallade **domedagsargumentet**, som är ett statistiskt resonemang som (av dess anhängare) påstås indikera mänsklighetens snara undergång. Med stöd i de grundläggande statistiska begreppen från Avsnitt 2 skall jag försöka reda ut huruvida det ligger något i domedagsargumentet.

2 Slantsingling

Vi som undervisar i grundläggande sannolikhetsteori och statistik på universitet och högskolor använder oss ofta av exempel som involverar slantsingling, tärningskast, roulettehjul eller pokerhänder. Detta skall inte uppfattas som att vi gillar spel och dobbel, eller ens att vi tycker sådant är särskilt intressant. Skälet till att vi gärna hämtar exempel från hasardspel är att de fungerar bra som pedagogiska hjälpmedel. Detta beror i sin tur på att den matematiska modelleringen av slumpmekanismen blir relativt enkel för dessa spel, och att det är jämförelsevis lätt att godta dessa enkla modellantaganden utan att behöva ta hänsyn till alla de komplikationer av olika slag som tenderar att dyka upp i mer allvarliga exempel hämtade från verkligheten.

Här vill jag använda slantsingling som ett exempel. Då vi singlar en slant får vi något av de två möjliga utfallen krona och klave. Vanligtvis antar man att krona och klave har samma sannolikhet, nämligen 1/2 vardera:

$$\mathbf{P}(\text{krona}) \;=\; \mathbf{P}(\text{klave}) \;=\; \tfrac{1}{2}\,. \tag{1}$$

Symbolen $\mathbf{P}$ (som i franskans *probabilité* och engelskans *probability*) betecknar här sannolikhet.

Låt oss emellertid anta att vi inte är säkra på om myntet har den rätta symmetriegenskapen för att sannolikheterna skall bli 1/2. Kanske är myntet en aning skevt eller obalanserat? För att tillåta den möjligheten ersätter vi modellantagandet (1) med det mer generella antagandet

$$\begin{cases} \mathbf{P}(\text{krona}) = q \\ \mathbf{P}(\text{klave}) = 1-q \end{cases} \quad \text{för något } q \text{ i intervallet } 0 \le q \le 1\,. \tag{2}$$

Om $q = 1/2$ är myntet rättvist, annars inte. Modellantagandet (2) lämnar båda möjligheterna öppna.

Om vi vill veta huruvida myntet är rättvist bör vi alltså försöka få reda på vilket värde q har. I brist på gudomliga uppenbarelser om saken får vi ta till en experimentell metod: singla slanten ett lämpligt antal n gånger, notera utfallen, och gör sedan en statistisk analys. Någon helt säker kunskap om det exakta värdet på q kan vi inte få (annat än om vi singlar den oändligt många gånger, vilket vi knappast har tid med), men vi kan få viktiga ledtrådar.

Antag att vi bestämmer oss för att singla slanten $n = 10$ gånger, och håller reda på antalet krona X. Det visar sig att vi får utfallet krona $X = 2$ gånger av 10 möjliga. Detta är våra data. Vilka slutsatser kan dras?

På basis av utfallet vill vi ta ställning till om detta ger anledning att misstro myntets symmetri. Med andra ord, kan utfallet anses tala emot $q = 1/2$? En första tanke skulle kunna vara att om $q = 1/2$ så borde utfallet bli symmetriskt, dvs lika många krona som klave ($X = 5$), och att eftersom det utfallet vi faktiskt fick ($X = 2$) avviker detta, så måste myntet vara skevt. Ett problem med detta resonemang är att även om myntet skulle vara symmetriskt ($q = 1/2$), så är det varje gång vi singlar slanten fullt möjligt (i själva verket 50 procents chans) att få krona, samt lika möjligt och lika troligt att få klave, oavsett tidigare utfall. Det kan alltså hända även med ett symmetriskt mynt att man får krona i alla tio kasten. Vad göra?

Vår frågeställning (kan utfallet $X = 2$ anses tala emot att myntet är symmetriskt?) är lyckligtvis av precis det slag som den statistiska metod som kallas **signifikanstest** kan ge svar på. För att förklara metoden behöver vi först lite terminologi:

- Med **nollhypotesen** menas den hypotes som vi vill ta ställning till genom att ange huruvida data talar emot den eller inte. I det här fallet är nollhypotesen att myntet är symmetriskt: $q = 1/2$.

- Givet data definieras **p-värdet** som hur stor sannolikheten skulle vara, om nollhypotesen var sann, att få minst så extrema data som vi faktiskt fick. (Här måste man vara noga med att definiera exakt vad som menas med extrema data – mer om det i slantsinglingsexemplet strax.)

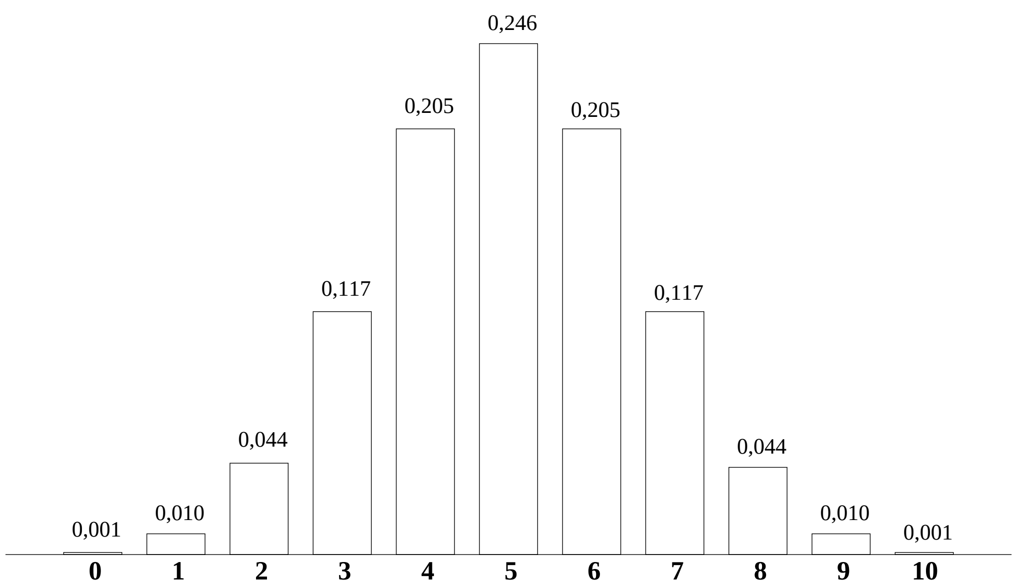
- Om p-värdet underskrider en på förhand given gräns kallad **signifikansnivån** (oftast 0,05) så sägs utfallet vara **statistiskt signifikant**.

Logiken i ett signifikanstest är följande. Om statistisk signifikans erhålles så måste (minst) en av två saker gälla: antingen har något osannolikt inträffat (nämligen statistisk signifikans, vilket har sannolikhet högst lika med den valda signifikansnivån), eller också är nollhypotesen falsk. Om en händelse är osannolik så väntar vi oss inte att den skall inträffa, så utfallet brukar därför anses ge stöd för möjligheten att nollhypotesen är falsk. Och ju lägre p-värde, desto starkare stöd.

Låt oss se hur detta fungerar på slantsinglingsexemplet. Vi följer traditionen och sätter signifikansnivån till 0,05. För att bestämma p-värdet behöver vi veta hur sannolikhetsfördelningen för antalet krona X ser ut om nollhypotesen $q = 1/2$ är sann. Här är inte rätt plats att härleda fördelningen (men den intresserade läsaren kan vända sig till [AB] eller [OA]), utan jag nöjer mig med att berätta att X är **binomialfördelad** med parametrarna 10 och 1/2, vilket innebär att x

$$\mathbf{P}(X = k) = \left(\tfrac{1}{2}\right)^{10} \frac{10!}{k!(10-k)!} \quad \text{för } k = 0, 1, 2, \ldots, 10\,.$$

Avrundat till närmaste tusendel får vi följade värden på $\mathbf{P}(X = 0)$, $\mathbf{P}(X = 1)$, $\ldots$, $\mathbf{P}(X = 10)$.



Vi ser alltså att det observerade värdet $X = 2$ har sannolikhet 0,044. Blir p-värdet alltså 0,044?

Svar nej – p-värdet definierades ju inte som den ur nollhypotesen beräknade sannolikheten att få *exakt* det observerade värdet, utan som motsvarade sannolikhet att få ett *minst så extremt* värde som vi faktiskt observerade. I det här fallet är både $X = 1$ och $X = 0$ att betrakta som mer extrema utfall. Deras sannolikheter måste räknas in, så vår nästa kandidat till p-värde blir $\mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2)$ =0,001+0,010+0,044=0,055. Är vi därmed framme vid rätt p-värde?

Inte riktigt ännu. Fördelningen för X är ju symmetrisk runt värdet 5, och eftersom vi inte sagt något bestämt om i vilken riktning vi tänker oss att q skulle kunna avvika från 1/2, så måste testproceduren respektera denna symmetri. Kontentan av det är att $X = 8$ måste räknas som ett lika avvikande utfall som $X = 2$. Summeringen som skall ge p-värdet måste därför inkludera $\mathbf{P}(X = 8)$, liksom givetvis också $\mathbf{P}(X = 9)$ och $\mathbf{P}(X = 10)$. Vi är äntligen framme vid det riktiga p-värdet:

$$\begin{aligned} \text{p-värde} &= \mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2) \\ &\quad + \mathbf{P}(X = 8) + \mathbf{P}(X = 9) + \mathbf{P}(X = 10) \\ &= 0{,}001 + 0{,}010 + 0{,}044 + 0{,}044 + 0{,}010 + 0{,}001 \\ &= 0{,}110\,. \end{aligned} \tag{3}$$

Återstår endast att jämföra detta med signifikansnivån 0,05. För att vi skall få lov att utropa statistisk signifikans behöver p-värdet *underskrida* signifikansnivån. Men eftersom 0,110>0,05 är alltså experimentutfallet inte att betrakta som någon statistiskt signifikant avvikelse från nollhypotesen. Med andra ord: våra data ger inte anledning att misstro att myntet är symmetriskt.

Därmed kan exemplet sägas vara uttömt, men tolkning av signifikanstest är ett område där det är lätt att gå vilse och där missförstånd florerar till och med bland forskare (se gärna Ziliak och McCloskey [ZM] för en engagerad och kritisk lägesbeskrivning, eller [H] för en kortare och något mer balanserad redogörelse). Några kommentarer är därför på sin plats.

1. Betyder den uteblivna statistiska signifikansen att vi kan dra slutsatsen att nollhypotesen är sann och att myntet alltså är symmetriskt? *Svar nej!* Den uteblivna statistiska signifikansen betyder bara att erhållna data (2 krona av 10) inte talar särskilt starkt emot nollhypotesen q=0,5. Men det finns en uppsjö andra hypoteser som data inte heller talar emot, som t.ex. den att q=0,2, eller den att q=0,3, etc.

2. Antag att vi istället hade fått $X = 1$ krona av 10 möjliga. Då hade vi med hjälp av en summering liknande den i (3) fått p-värdet 0,022. Eftersom 0,022<0,05 hade det räknats som statistisk signifikans. Kan vi då utesluta möjligheten att nollhypotesen är sann? *Svar nej!*

Statistisk signifikans betyder i detta fall att om nollhypotesen är sann så har en händelse med sannolikhet högst 0,05 inträffat (nämligen händelsen att vi fick statistisk signifikans). Men även osannolikha händelser inträffar emellanåt. Dessutom bör vi ha klart för oss att det inte är något magiskt med just talet 0,05. Orsaken till att 0,05 så ofta används som signifikansnivå står främst att finna i tradition och historiska tillfälligheter.

Man träffar ibland på föreställningen att statistisk signifikans på signifikansnivå 0,05 skulle svara mot att nollhypotesen kan förkastas ”bortom rimlig tvivel”, men den tolkningen är helt på tok. Exakt hur låg signifikansnivå som skulle krävas för ett sådant uttryckssätt är väl i någon mån en smaksak och kan bero på olika omständigheter som t.ex. hur allvarliga konsekvenserna av att göra fel är (t.ex. skulle man kunna hävda att vi bör kräva lägre signifikansnivå i en rättegång där den åtalade riskerar dödsstraff, än i en där det som värst kan bli fråga om böter), men själv skulle jag nog i allmänhet dra mig för uttrycket ”bortom rimlig tvivel” i samband med signifikans på nivåer högre än 0,0001. En mer passande reaktion på statistisk signifikans på nivå 0,05 skulle kunna vara ”detta ger skäl till visst ifrågasättande av nollhypotesen, och saken kan vara värd att undersöka vidare”.

3. Kan vi, baserat på vårt slantsinglingsförsök, avgöra hur sannolik nollhypotesen är? Kan vi t.ex. ta det erhållna p-värdet 0,110 som ett mått på hur sannolik nollhypotesen är? *Svar nej!*

Vad vi kanske allra helst skulle vilja ha när vi testar en nollhypotes är ett värde på nollhypotesens sannolikhet, givet våra data. Teorin för hypotestest och statistisk signifikans erbjuder dessvärre inget sådant. Om vi ändå försöker oss på att tolka p-värdet som en sådan sannolikhet, så begår vi det fel som på engelska kallas för **fallacy of the transposed conditional**: att förväxla sannolikheten för nollhypotesen givet data med sannolikheten för data givet nollhypotesen. Med gängse beteckning för betingad sannolikhet skriver vi $\mathbf{P}(A|B)$ för sannolikheten för en händelse A givet att vi vet att en annan händelse B inträffar. Vi kan i allmänhet *inte* kasta om betingningen och påstå att

$$\mathbf{P}(A|B) = \mathbf{P}(B|A)\,. \tag{4}$$

Som ett motexempel till (4), betrakta experimentet att välja en svensk av manskön på måfå (var och en med samma sannolikhet), låt A vara händelsen att den valde mannen är flintskallig, och låt B vara händelsen att den valde mannen är statsminister. Å ena sidan får vi då

$$\mathbf{P}(A|B) = 1 \tag{5}$$

eftersom Sveriges statsminister (i skrivande stund) de facto *är* flintskallig. Å andra sidan får vi

$$\mathbf{P}(B|A) \approx \frac{1}{500\,000} \tag{6}$$

eftersom blott en av Sveriges uppskattningsvis 500 000 flintskalliga herrar är statsminister. Diskrepensen mellan sannolikheterna i (5) och (6) visar att omkastningen i (4) inte är tillåten, och det är väsentligen detta som är skälet till att p-värdet inte kan tolkas som sannolikheten för nollhypotesen givet data.

Om vi verkligen insisterar på att få ett värde på sannolikheten för nollhypotesen givet de data vi har, så duger inte den klassiska teori för signifikanstest vi tillgripit här. Vi har i så fall inget annat val än att ta till så kallade **Bayesianska statistiska metoder.** Sådana metoder inbegriper att vi antar en så kallad **a priori-fördelning** av sannolikheter som representerar hur mycket vi från början (innan vi sett data) tror på nollhypotesen och på diverse mothypoteser. Därefter kan man med hjälp av erhållna data och det som kallas **Bayes sats** beräkna en **a posteriori-fördelning** av sannolikheter för nollhypotesen och de olika mothypoteserna. Själva beräkningssteget är okontroversiellt. Mer omdiskuterat är det inledande steget med val av a priori-fördelning. Detta val påverkar även a posteriori-fördelningen, och (hävdar kritikerna) öppnar för godtycke och subjektivitet. Jag skall inte här fördjupa mig i 1900-talets utdragna debatt mellan Bayesianska och mer klassiskt inriktade statistiker, utan nöjer mig med att konstatera att den alltmer förhärskande uppfattningen bland yngre statistiker är att båda metoderna är bra att ha i sin statistiska verktygslåda, och att det beror på den konkreta situationen vilken av dem som är att föredra. För mer om Bayesiansk statistik hänvisar jag (återigen) till [AB] och [OA].

4. Utöver de skäl för försiktighet i tolkningen av statistisk signifikans som angetts i punkt 2 ovan finns ännu ett, nämligen det som i vetenskapsteori kallas **Duhem–Quines tes**. Denna tes stipulerar att vetenskapliga teorier aldrig kan testas helt separat från andra teorier. En astronom som vill testa någon viss lag för planetrörelser behöver t.ex. förlita sig på att det ljus han tar emot med sitt teleskop följer optikens lagar. En opinionsundersökare som ringer folk för att testa om regeringspartiet minskat i popularitet behöver t.ex. anta att regeringsanhängare är lika benägna att svara i telefon som oppositionsanhängare. Och så vidare.

Om vi skärskådar slantsinglingsexemplet inser vi att våra beräkningar av fördelningen för antal krona X bygger inte bara på nollhypotesantagandet att $q = 1/2$, utan på en hel del andra antaganden. Viktigt är t.ex. att de olika slantsinglingarna antas ge statistiskt oberoende utfall. Annars skulle vi kunna ha ett mynt som visserligen är utmärkt symmetriskt, men en stereotyp rörelse för att plocka upp och singla slanten som tenderar att ge samma utfall som i föregående kast. Detta skulle resultera i långa serier av indentiska utfall, och en fördelning för X som lägger betydligt mindre sannolikhet på de centrala värdena kring $X = 5$, och mer utåt kanterna. Därmed (dvs om inte olika slantsinglingar kan antas ge statistiskt oberoende utfall) faller vår p-värdesberäkning.

Läxan vi bör lära oss av detta är att vi, innan vi alltför tvärsäkert konstaterar att vi har belägg som talar emot nollhypotesen, noga bör tänka igenom vilka andra (ofta outtalade) antaganden vi gjort, och om den observerade avvikelsen istället kan tänkas bero på att någon av dem fallerar. För mer om Duhem–Quines tes, se t.ex. Stanford [S].

3 Domedagsargumentet

Vi har sett att redan en så till synes enkel sak som att statistiskt analysera utfallet av ett antal slantsinglingar kräver en hel del eftertanke. Det hela blir sjufalt mer komplicerat då vi lämnar spelbordet och ger oss på några av de i verkliga livet mer angelägna frågeställningar som dem jag nämnde i inledningen: hur står det till med sambandet mellan rökning och prostatacancer, med förändringarna i svenska folkets lemonadkonsumtion, och med regeringspartiets eventuellt sviktande väljarstöd? Jag skall hoppa över alla dessa komplicerade fall ur en statistikers vardag, och gå raka vägen på en ännu större och (vågar jag påstå) ännu viktigare fråga. Ty vad kan väl vara viktigare än mänsklighetens överlevnad?

3.1 Argumentet i dess grundform

Det så kallade **domedagsargumentet** formulerades ursprungligen av astrofysikern Brandon Carter, och populariserades i en bok av John Leslie [L]. Argumentet är i korthet följande.

Låt N beteckna totala antalet människor som har levat, lever, eller någonsin kommer att leva. Naturligtvis vet vi inte värdet på N, men vi kan ändå tillåta oss vissa matematiska beräkningar med den okända storheten (lite grand som vi räknade på den okända parametern q i slantsinglingsexemplet). Vi kan vidare tänka oss att vi rangordnar mänskligheten kronologiskt efter födelsedatum, så att den förste människan på jorden har nummer 1, den andre har nummer 2, och så vidare till den siste människa som någonsin föds, som har nummer N. (För oss som inte tror på berättelsen om Adam och Eva är givetvis begreppet "den första människan" en smula suspekt, men vi kan ändå tänka oss att någonstans i mänsklighetens (för-)historia dra en (givetvis något godtycklig) gräns för exakt vilka exemplar av människoapor vi räknar som människor.) Om vi nu väljer en människa på måfå ur hela mänsklighetens historia och betecknar dennes position i rangordningen med n, så kan n alltså ta vilket värde som helst mellan 1 och N – vart och ett med samma sannolikhet $1/N$. Kvoten n/N kommer därmed att ta något av värdena $\frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1$, var och ett med samma sannolikhet. Eftersom N är ett stort tal (det måste ju givetvis vara större än antalet *idag* levande människor, som är mer än 7 miljarder) så är sannolikhetsfördelningen för n/N därmed i stort sett jämnt utsmetad på intervallet mellan 0 och 1. Det betyder att för godtyckligt α mellan 0 och 1 gäller

$$\mathbf{P}\left(\frac{n}{N} \leq \alpha\right) \approx \alpha\,.$$

Approximationen är för stora N såpass god att vi utan att begå något svårt fel kan ersätta $\approx$ med ett likhetstecken. Vi kan välja α =0,05 (av skäl som har att göra med signifikansnivåtradition jag nämnde i Avsnitt 2 och som jag strax skall förklara närmare), och får då

$$\mathbf{P}\left(\frac{n}{N} \leq 0{,}05\right) \;=\; 0{,}05\,. \tag{7}$$

Låt oss nu tänka oss att du, kära läsare, är den på måfå valda personen ur mänsklighetens historia. Det är svårt att säga något bestämt om N utan att se in i framtiden, medan däremot antalet människor n som fötts fram till och med dig borde vara lättare att uppskatta. Mycket riktigt finns forskning rörande hur många människor som hittills funnits. Leslie [L] citerar uppskattningen 60 miljarder (så att det för varje nu levande människa finns cirka sju döda), vilket givetvis är en mycket osäker skattning, men troligtvis ändå rätt storleksordning. Låt oss för resonemangets skull acceptera Leslies siffra, och bortse från det fåtal miljarder som (lite grand beroende på hur gammal du är) som är födda efter dig, så att din placering n i raden av människor uppställda efter födelsetidpunkt blir $n = 60\,000\,000\,000$. Om vi pluggar in den siffran i sambandet (7) får vi

$$\mathbf{P}\left(\frac{60\,000\,000\,000}{N} \leq 0{,}05\right) \;=\; 0{,}05 \tag{8}$$

vilket om vi löser för N inne i parentesen ger

$$\mathbf{P}\left(N \geq 1\,200\,000\,000\,000\right) \;=\; 0{,}05 \tag{9}$$

eller med andra ord

$$\begin{aligned} &\mathbf{P}(\text{mänskligheten går under innan totala} \\ &\text{antalet människor når upp till } 1\,200\,000\,000\,000) \;=\; 0{,}95. \end{aligned} \tag{10}$$

1 200 000 000 000, dvs 1200 miljarder, kan låta som ett väldigt stort tal, men hur lång tid motsvarar det? Det beror givetvis på hur stor befolkningen kommer att vara i framtiden. Om vi tänker oss att jordens befolkning stabiliserar sig kring 8 miljarder, och med en medellivslängd på 80 år, såblir antalet födslar per år

$$\frac{8\,000\,000\,000}{80} = 100\,000\,000$$

och de $1\,200\,000\,000\,000 - 60\,000\,000\,000 = 1140$ miljarder födslar som återstår (om totalat antalet människor N som någonsin föds skall uppgå till 1200 miljarder) kommer att räcka till en tidsrymd om

$$\frac{1\,140\,000\,000\,000}{100\,000\,000} = 11\,400 \text{ år.}$$

Mänskligheten är alltså (troligen) inne på sluttampen av sin existens! Detta är domedagsargumentet. Antagandena om medellivslängd och stabiliserad folkmängd kan givetvis varieras, men slutsatsen kvarstår att totala antalet människor som kommer att födas i framtiden troligtvis inte blir väsentligt mer än 1140 miljarder.

3.2 Kritik av argumentet

Är domedagsargumentet övertygande? En viss misstänksamhet kände jag omedelbart då jag första gången hörde talas om det: kan man verkligen förutsäga framtiden med hjälp av en så enkel sifferexcersis och inga andra data än en uppskattning av hur många människor som hittills har levet? Det verkar faktiskt en smula otroligt.

När jag betraktar argumentet mer i detalj, och med den professionella statistikerns kritiska blick, så finner jag mycket riktigt ett mycket suspekt steg i resonemanget, nämligen steget från sambandet (7) till sambandet (8) där vi satte in siffran $n = 60\,000\,000\,000$. Att sätta in en siffra i en formel är vi vana vid att få göra, men här uppstår ett problem med den statistiska tolkningen.

Å ena sidan, betrakta sannolikhetsuttycket $\mathbf{P}(n/N \leq 0{,}05)$ i vänsterledet till (7). För att sannolikheten skall kunna bli något annat än 0 eller 1 (t.ex. 0,05 som vi ser i högerledet) måste minst en av variablerna innanför parentesen vara slumpmässig. Det finns två variabler att välja på: n och N. Om vi backar och ser hur vi definierade den slumpmodell som leder fram till (7), så ser vi att den enda slumpmekanism som är definierad är för n. N är vad det är, men givet N så väljs n på måfå bland talen $1, 2, \ldots, N$.

Å andra sidan, betrakta sannolikhetsuttrycket $\mathbf{P}(60\,000\,000\,000/N \leq 0{,}05)$ i vänsterledet till (8). Även här behövs minst en slumpmässig storhet innanför parentesen för att sannolikheten skall kunna bli mellan 0 eller 1. Nu när vi ersatt n med $60\,000\,000\,000$ (som ju inte är en slumpmässig storhet utan ett fixt tal) så återstår endast att betrakta N som slumpmässig, vilket ju också är vad som görs i domedagsargumentets slutsats (10). Men, som sagt, i den sannolikhetsmodell som hela kalkylen byggde på så finns ingen specificerad slumpmekanism som genererar N, så här har uppenbarligen något gått snett i argumentationen.

Det som skett är en variant på det som jag i Avsnitt 2, punkt 3 i listan över varningsord rörande tolkningar av statistisk signifikans, omtalade som *fallacy of the transposed conditional*. Sambandet (7) handlar om fördelningen för n givet ett fixt värde på N, medan vi i (8) och de följande kalkylerna vantolkat sambandet som om det handlade om fördelningen för N givet ett fixt värde (nämligen 60 miljarder) på n.

3.3 En modifiering av argumentet

För att försöka rädda domedagsargumentet från ovanstående kritik ligger det nära till hands att försöka omforma det till ett signifikanstest, i enlighet med de riktlinjer jag skisserade i Avsnitt 2. Bilda nollhypotesen att totala antalet människor (i det förflutna, nu och i framtiden) N är minst 1200 miljarder. Storheten n, dvs antalet människor fram till idag (eller till läsarens födelse), kan vi då använda till att göra ett signifikanstest av nollhypotesen $N \geq 1200$ miljarder. Vi påminner oss sambandet (7) –

$$\mathbf{P}\left(\frac{n}{N} \leq 0{,}05\right) = 0{,}05$$

– och noterar att nollhypotesen $N \geq 1200$ miljarder medför att

$$\mathbf{P}\left(\frac{n}{1\,200\,000\,000\,000} \leq 0{,}05\right) \leq 0{,}05$$

eller med andra ord

$$\mathbf{P}\,(n \leq 60\,000\,000\,000) \leq 0{,}05\,. \qquad (11)$$

Jämför detta med hur vi tidigare försökte oss på att sätta in ett siffervärde på n i (7) för att erhålla (8). Den gången gjorde vi fel, eftersom vi satte in ett siffervärde på den enda kvantitet n för vilken en slumpmekanism hade specificerats i den matematiska modellen, så att det som fanns kvar inom parentesen inte längre inbegrep någon slump, och påstaendet att dess sannolikhet var 0,05 blev meninslöst och vilseledande. Att istället sätta in ett siffervärde på N undviker detta fel, eftersom N bara stod för ett tal, utan någon slumpmekanism bakom.

Sambandet (11) kan vi lägga till grund för vårt signifikanstest av nollhypotesen $N \geq 1200$ miljarder. Testet utformas så att vi observerar n – antalet människor födda fram till och med observatörens (läsarens) födelse, och om $n \leq 60\,000\,000\,000$ så deklarerar vi statistisk signifikans på nivån 0,05. Och eftersom vi observerat just $n = 60\,000\,000\,000$, så har vi statistisk signifikans – en statistiskt signifikant avvikelse från vad vi skulle förvänta oss under nollhypotesen.

I enlighet med diskussionerna i Avsnitt 2 tolkar vi detta som att *antingen* är nollhypotesen falsk (så att antalet människor någonsin inte kommer att nå upp till 1200 miljarder), *eller också* har något osannolikt (dvs med sannolikhet högst 0,05) inträffat. I övrigt bör vi iaktta den försiktighet i tolkningen av statistisk signifikans som jag framhöll i mina varningsord numrerade 1 till 4 i Avsnitt 2.

3.4 Kritik av det modifierade argumentet

Ovanstående tolkning av domedagsargumentet i termer av statistisk signifikans är avgjort bättre, enligt min mening, än den naiva sannolikhetstolkningen (10). Men den kan ändå kritiseras.

Proceduren för statistiska signifikanstest har historiskt motiverats på följande vis. Eftersom sannolikheten att felaktigt deklarera statistisk signifikans är maximalt 0,05, så kan vi räkna med att om vi gör oberoende upprepningar av försöket så kommer vi i det långa loppet inte att riskera att göra ett sådant felslut i mer än högst 5% av fallen.

Denna tolkning av sannolikheter kallas **frekventistisk**, eftersom den handlar om frekvensen tillfällen då något sker i en (verklig eller, oftast, tänkt) serie upprepningar. Av detta skäl har den statistiska begreppsapparat som signifikanstest ingår i kommit att kallas **frekventistisk statistik** (vilket framför allt brukar kontrasteras mot de Bayesianska statistiska metoder som jag hänvisade till i punkt 3 i min lista över varningsord i Avnitt 2).

Men hur blir det egentligen med den frekventistiska tolkningen i fallet med domedagsargumentet? Vad händer om vi upprepar samma försök gång på gång? Ja, om vi skall göra det här och nu, då får vi ju samma data n varje gång (eventuellt med vissa obetydliga variationer beroende om vi utser dig eller mig till den ”slumpvis valda personen ur mänsklighetens historia”). Och om vi får samma data, så får vi samma utfall: antingen signifikans varje gång, eller frånvaro av signifikans varje gång. Här finns alltså inte tillstymmelse till den garanti om maximalt 5%-ig felfrekvens som den frekventistiska motiveringen av signifikansktest talar om.

Det nyckelvillkor i den frekventistiska motiveringen som fallerar i domedagstillämpningen är att försöken skall vara *oberoende*. För att få verkligt oberoende försök måste vi välja den figur vi kan kalla referenspersonen – dvs hon eller han vars placering i mänsklighetens stora födelselista ligger till grund för n – på måfå ur hela mänskligheten, inklusive alla människor som inte är födda ännu. Det är svårt då vi inte har någon kristallkula som låter oss se in i framtiden. I själva verket förutsätter det att vi kan göra antaganden om hur många människor som kommer att finnas, men då blir ju hela domedagsargumentet till ett cirkelresonemang.

Den som i detta läge verkligen ändå önskar försvara statistisk signifikansversionen av domedagsargumentet kan prova med att hävda att logiken för statistisk signifikans – dvs slutsatsen att antingen är nollhypotesen falsk eller också har något osannolikt inträffat – ändå gäller, trots att den frekventistiska tolkningen av vad som i detta fall menas med sannolikhet inte håller. Det finns ju trots allt andra tolkningar av sannolikhetsbegreppet som föreslagits genom historien (så t.ex. beskrivs, inom vissa inriktningar av Bayesianism, sannolikheter ofta i termer av individers subjektiva uppfattningar, ibland definierade av deras benägenhet att ingå i vissa slags vadslagningar; se gärna Binmore [B]).

Vår diskussion för emellertid uppmärksamheten till frågan om just *du*, läsaren, verkligen kan anses vara likvärdig (ur domedagssynpunkt) med en på måfå vald människa bland alla som någonsin levat eller kommer att leva. I härledningen av det grundläggande sannolikhetssambandet (7), som våra fortsatta domedagsresonemang bygger vidare på, antog vi ju just att referenspersonen är en på måfå vald person ur mänsklighetens hela historia (inklusive framtiden). I brist på någon metod att välja ut en sådan person valde vi istället dig. Men kan vi verkligen göra det, och fortfarande hävda att (7) stämmer? Detta är ytterst problematiskt. Exempelvis skulle det mycket väl kunna vara så att domedagsargumentet är en tankegång som bara uppstår i något visst ganska tidigt skede av mänsklighetens utveckling, och att framtida människor går vidare till nya och mer givande tankegångar som vi idag inte vet något om. Om så är fallet kommer domedagsargumentet *alltid* (i betydelsen varje gång någon försöker sig på att använda det), att resultera i grava underskattningar av hur många framtida människor som är att vänta.

Sammanfattningsvis kan jag om domedagsargumentet inte säga annat att det vilar på skumma antaganden och bräcklig grund. Det kan naturligtvis hända att någon snillrik person så småningom kommer på ett övertygande sätt att försvara domedagsargumentet, men intill dess tycker jag att vi bör betrakta det som ogiltigt. Det finns tillräckligt gott om andra, mycket mer konkreta, skäl att frukta för mänsklighetens överlevnad (som t.ex. risken för globalt kärnvapenkrig eller för global spridning av någon supereffektiv dödssmitta; se Bostrom och Cirkovic [BC] för en systematisk genomgång av sådana katastrofscenarier och uppskattningar av hur troliga de är) för att vi skall ha fullt upp utan att behöva spä på med ett suspekt abstrakt argument. Dessa mer konkreta skäl till oro har dessutom den fördelen (jämfört med domedagsargumentet) att vi har möjlighet att göra något åt dem.

Referenser

[AB] Alm, Sven Erick och Britton, Tom (2008) *Stokastik*, Liber, Stockholm.

[B] Binmore, Ken (2008) *Rational Decisions*, Princeton University Press, Princeton.

[BC] Bostrom, Nick och Cirkovic, Milan M. (2008) *Global Catastrophic Risks*, Oxford University Press, Oxford.

[H] Häggström, Olle (2012) Why the empirical sciences need statistics so desperately, preprint, `http://www.math.chalmers.se/~olleh/haggstrom_proc_ems.pdf`

[L] Leslie, John (1998) *The End of the World: The Science and Ethics of Human Extinction*, Routledge, New York.

[OA] Olofsson, Peter och Andersson, Mikael (2012) *Probability, Statistics and Stochastic Processes* (2nd ed), Wiley-Blackwell, Chicester.

[S] Stanford, Kyle (2009) Underdetermination of scientific theory, *The Stanford Encyclopedia of Philosophy, Winter 2009 Edition* (ed. Edward N. Zalta), `http://plato.stanford.edu/entries/scientific-underdetermination/`

[ZM] Ziliak, Stephen och McCloskey, Deirdre (2008) *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives*, The University of Michigan Press, Ann Arbor, MI.