10 Två böcker om slumpens matematik[1]

Till de mer minnesvärda inläggen i den svenska kärnkraftsdebatt som föregick folkomröstningen 1980, hör Tage Danielssons monolog *Om sannolikhet*, som inleds med följande ord:

> Alltså, sannolikhet, va, det betyder väl nå't som är likt sanning. Och då är det synd, tycker jag, att man tydligen från och med i år – på grund av de rådande konjunkturerna – inte har råd med några riktiga sanningar längre, utan att man måste nöja sig med sannolikhetskalkyler.

Han raljerade elegant med vissa kärnkraftsförespråkares försök att karaktärisera den nyss inträffade olyckan i Harrisburg som osannolik, och han lärde oss att vara kritiska mot felaktiga eller slarvigt utförda sannolikhetsargument. Men det är inte all sannolikhetskalkyl som är felaktig, och vi har nu begåvats med två svenska böcker om slumpen och slumpens matematik, varav den ena – Allan Guts *Sant eller sannolikt* – redan i titeln anspelar på Tage Danielssons inlägg.

I sammanhanget inställer sig genast två frågor. För det första, finns slumpen (eller är naturen tvärtom fundamentalt deterministisk)? Och för det andra, kan man verkligen räkna på det som är slumpmässigt?

Den första frågan hör strängt taget inte till sannolikhetsteorin – som är den gren av matematiken som behandlar slumpen – utan snarare till filosofin eller fysiken. Såväl kvantfysik som kaosteori lämnar utrymme för olika tolkningar i frågan om slumpens existens. En viktig insikt är emellertid att sannolikhetsteorin är till stor hjälp oavsett om verklig slump existerar, eftersom vår okunskap om de exakta detaljerna i ett komplicerat sammanhang ofta bäst behandlas stokastiskt (dvs sannolikhetsteoretiskt). De senaste 100 åren har givit oss otaliga exempel på detta, främst inom naturvetenskaperna, men även inom teknik och samhällsvetenskaper. Ett extremt exempel är position och hastighet hos samtliga molekyler i en gasbehållare; att specificera dessa exakt är naturligtvis omöjligt, medan däremot ett stokastiskt angreppssätt visar sig mycket fruktbart.

På nästa fråga – den om huruvida slumpen medger matematiska beräkningar – ger Allan Gut, som är professor i matematisk statistik vid Uppsala

[1] Publicerad under rubriken ”Liksom av en händelse – om matematisk sannolikhet och randomisering” i *Dagens Forskning*, 9–10 september 2002.

universitet, ett rungande ja-svar i sin bok. Från de allra enklaste exempel, som att visa att två kast med ett rättvist mynt ger utfallet ”krona, krona” med sannolikhet 0,25, demonstrerar han hur man kan jobba sig vidare mot alltmer avancerade och för verkliga tillämpningar relevanta kalkyler.

De mest tacksama illustrationerna av grundläggande sannolikhetsteori hittar man i olika former av spel och dobbel. Med denna teori kan man besvara frågor som exempelvis *Hur stor är sannolikheten att få minst en sexa vid kast med tre tärningar?* och *Kan man under ett långt livs idogt bridgespelande räkna med att någon gång få en hand med 13 kort av samma färg?*. Gut ägnar ett helt kapitel åt illustrativa – och sedelärande – lösningar på dylika problem, och han passar i sammanhanget på att argumentera mot den statliga dubbelmoral som utsugningen av de individer som drabbats av spelberoende innebär.

Nära besläktad med sannolikhetsteorin är den statistiska inferensteorin, dvs teorin för hur man drar statistiska slutsatser ur osäkra data. I ett kapitel betitlat *Lögn och statistik* (kanske bokens viktigaste!) behandlar Gut olika vanligt förekommande metoder att vilseleda med hjälp av statistik. Som exempel kan nämnas manipulation av koordinataxlar och tendentiösa val av vilka siffror som skall presenteras respektive utelämnas. Sådana metoder för missbruk av statistik är viktiga för var och en att vara medveten om, men ger samtidigt det akademiska statistikämnet dåligt rykte – oförtjänt, eftersom ämnet syftar till att ta fram metoder som ger korrekta, dvs varken missvisande eller tendentiösa, resultat. Mycket av tankarna i detta kaptiel är av relativt tidlös karaktär och förekommer redan i Darrell Huffs klassiker *Hur man ljuger med statistik* (1955, sv övers 1958). En nyare trend som Gut noterar är emellertid förekomsten av alltmer sofistikerade program för statistisk databehandling; en användare utan de rätta kunskaperna i statistikteori kan med hjälp av dessa lätt krama fram en rad slutsatser ur sina data, utan att ha en aning om huruvida de erforderliga förutsättningarna är uppfyllda i den aktuella tillämpningen. Detta är analogt med Internets kolossala informationskälla som vem som helst kan ösa ur, men för vilken särskilda kunskaper krävs för att plocka fram den information som är riktig och relevant.

Även merparten av den grundläggande sannolikhetsteorin är av såpass gammalt datum att det mesta som sägs i Guts bok kunde ha sagts för flera årtionden sedan. Detta kompenserar författaren i bokens sista kapitel med ett beundransvärt försök att på ett lättfattligt sätt presentera sin egen forskningsspecialitet: så kallade stoppade slumpvandringar. Ändå blir detta kapitel för de flesta läsare säkerligen det mest svårforcerade.

Sammanfattningsvis kan om Guts bok sägas att den ger en bred och intressant inblick i slumpens matematik och att den torde kunna uppskattas av en vid läsekrets. Emellertid känns det som om författaren borde ha anslagit något mera tid åt sitt bokprojekt, då ett antal av exemplen framstår som ogenomtänkt utvalda och/eller slarvigt genomförda.

Sannolikhetsteorin bjuder på en rad paradoxer, som utmanar intuitionen och har lärorika förklaringar. Ett av dem är det kända ”bilen och getterna”-problemet, som Gut behandlar utförligt. Ett annat är det så kallade kuvertproblemet: Antag att Alice lägger en för Bob okänd summa pengar X i ett kuvert, och att hon lägger den dubbla summan $2X$ i ett likadant kuvert. Dessa blandas noggrant och Bob får öppna det ena och därefter välja antingen att behålla den summan, eller att ta det andra kuvertet. Om han exempelvis finner 20 kr i det kuvert han öppnar, kan det andra innehålla 10 eller 40 kr, dvs ”i genomsnitt” $(10+40)/2=25$ kr. Det tycks därför som om han i medeltal tjänar på att byta. Men motsvarande resonemang fungerar oavsett vilken summa han hittar i det kuvert han öppnar, så det verkar alltså som om han redan utan att se efter vet att han bör välja det andra kuvertet. Men detta förefaller orimligt eftersom kuverten av symmetriskäl är i genomsnitt lika mycket värda. Häri består paradoxen, som jag skall återkomma till längre fram.

*

På irrfärd i slumpens värld av Magnus Arnér – tekn lic i matematisk statistik och verksam vid Saab Automobile – är utformad som en essäsamling. Flera av essäerna tar upp berömda sannolikhetsteoretiska och statistiska paradoxer, medan andra utgör nedslag i dessa ämnens historia.

Den kanske mest intressanta av Arnérs essäer är den om Gregor Mendel och hans experiment med ärtor från 1865, som ligger till grund för den moderna genetiken. Med teorin för statistisk hypotesprövning – som ej var tillgänglig på Mendels tid – kan vi idag undersöka om data från hans experiment är förenliga med de ärftlighetslagar han formulerade. Det överraskande resultatet är att data passar alltför bra, så bra att man med stor säkerhet kan hävda att de måste vara manipulerade. Något medvetet forskningsbedrägeri handlar det antagligen inte om, utan troligare om en (med dagens ögon!) naiv sållning av data.

Jämfört med Guts aningen strama framställning är tonen i Arnérs bok ledigare och mer entusiastisk. Med (själv-)ironisk distans spinner han gärna vidare på den gängse nidbilden av statistikern som torr och gråklädd kalenderbitare.

Det mesta i Arnérs bok är intressant och vederhäftigt, men i vissa stycken brister den tyvärr något i stringens. Hans analys av ovan nämnda kuvertproblem är helt på tok. Han börjar med – den orimliga – ansatsen att en viss symmetriegenskap (nämligen den att innehållen i de båda kuverten har samma väntevärde) då inget av kuverten öppnats, måste bevaras efter öppningen av det ena. Därifrån härleder han att den så kallade a priori-fördelningen för X bör ha en viss form, men trots att detta resulterar i en så kallat oegentlig fördelning (improper prior) undgår han att dra den rätta slutsatsen: vilken modell man än väljer att anta för fördelningen för X, så kommer alltid vissa utfall av vad Bob ser i det öppnade kuvertet att ge information som bryter den önskade symmetrin. Med ett liknande resonemang kan man för övrigt visa att en annan symmetri (den att båda kuverten har lika stor sannolikhet att innehålla det högre beloppet) också måste brytas för vissa sådana utfall, vilket förklarar paradoxen.[2]

Varken Arnér eller Gut drar sig för att använda matematiska formler där så behövs. Detta kan naturligtvis verka avskräckande på en del läsare, men det är också möjligt att skumma förbi formlerna utan att all behållning går förlorad. Jag bedömer Arnérs bok som den något mer krävande, men den torde likväl vara läsbar för en bred allmänhet.

Oaktat de smärre reservationer jag redovisat ovan, finner jag att båda böckerna är värdefulla tillskott till den svenska populärvetenskapliga litteraturen. Vidare hoppas jag att de upptäcks av gymnasieskolans matematiklärare, då de bör kunna fungera utmärkt som bredvidläsning för elever med rätt sorts håg och fallenhet.

[2] Ett par år efter publiceringen av denna text återkom jag i min bok *Slumpens skördar* till en mer utförlig behandling av paradoxen (Häggström, 2004).