KUVERTPROBLEMET – EN ANMÄRKNING OM VARIABLER OCH KONSTANTER INOM SANNOLIKHETSTEORIN

Jan Odelstad

1

I nr 1 1986 av Filosofisk Tidskrift återger redaktören Lars Bergström ett sannolikhetsteoretiskt problem som jag i fortsättningen ska kalla kuvertproblemet. Jag tror att detta problem hänger samman med distinktionen mellan variabel och godtycklig konstant (parameter) tillämpad inom sannolikhetsteorin och jag ska försöka utveckla denna tankegång närmare. Låt mig först återge problemet som det är formulerat i Filosofisk Tidskrift.

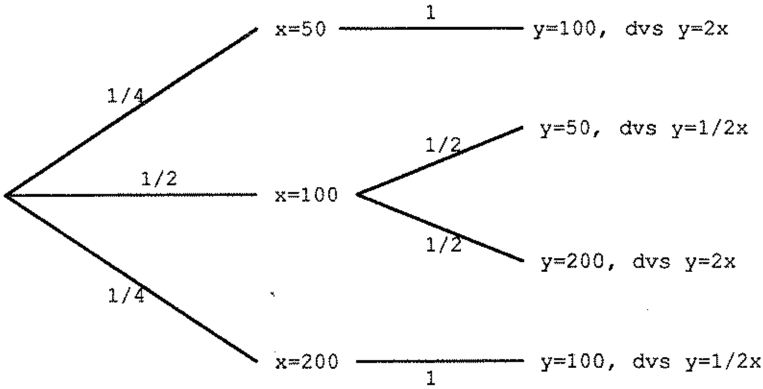
> Häromdagen dök följande problem upp på redaktionens bord (förmedlat av Wlodek Rabinowicz i Uppsala, som i sin tur fått det från Peter Gärdenfors i Lund, som i sin tur fått det från Sandy Zabell i USA). Antag att någon låter dig välja mellan två kuvert, som innehåller pengar, och att du får veta att det ena innehåller dubbelt så mycket som det andra.Då kanske man resonerar så här: Antag att det vänstra kuvertet innehåller x kronor.Då innehåller det högra kuvertet antingen 2x kr eller x/2 kronor.Bägge dessa fall är lika sannolika, inget talar mer för det ena än för det andra. Det betyder att sannolikheten är 1/2 i bägge fallen. Och det betyder i sin tur att det förväntade värdet av att ta det högra kuvertet är
>
> $$1/2 \cdot 2x + 1/2 \cdot x/2 = 1{,}25x \text{ kr}$$
>
> Det förväntade värdet av att ta det vänstra kuvertet är ju däremot
>
> $$1/2 \cdot x + 1/2 \cdot x = x \text{ kr}$$
>
> Alltså är det förväntade värdet större för det högra kuvertet. Men på ett helt analogt sätt kan man – om man i stället startar med det högra kuvertet – visa att det förväntade värdet är större för det vänstra kuvertet. Men detta innebär en motsägelse. Var ligger felet?

2

Låt oss börja med att ändra på problemet något. Antag att någon låter Dig välja mellan två kuvert och Du får veta att ett av dem innehåller 100 kr och det andra antingen 50 eller 200 kr. (Du får alltså inte veta vilket av kuverten som innehåller 100 kr och vilket som innehåller 50 eller 200 kr.) Det ena kuvertet innehåller då dubbelt så mycket som det andra, ty antingen innehåller kuverten 100 kr respektive 50 kr eller 200 kr respektive 100 kr. Det kan nu vara frestande att resonera analogt med i citatet från Filosofisk Tidskrift:

> Antag att det vänstra kuvertet innehåller x kronor.Då innehåller det högra kuvertet antingen 2x kr eller x/2 kronor. Bägge dessa fall är lika sannolika, inget talar mer för det ena än för det andra. Det betyder att sannolikheten är 1/2 i bägge fallen. Och det betyder i sin tur att det förväntade värdet av att ta det högra kuvertet är
>
> $$1/2 \cdot 2x + 1/2 \cdot x/2 = 1{,}25x \text{ kr}$$
>
> Det förväntade värdet av att ta det vänstra kuvertet är ju däremot
>
> $$1/2 \cdot x + 1/2 \cdot x = x \text{ kr}$$
>
> Alltså är det förväntade värdet större för det högra kuvertet. Men på ett helt analogt sätt kan man – om man i stället startar med det högra kuvertet – visa att det förväntade värdet är större för det vänstra kuvertet. Men detta innebär en motsägelse. Var ligger felet?

Påståendet att det högra kuvertet innehåller 2x kr med sannolikheten 1/2 och 1/2x kr med sannolikheten 1/2 behöver kanske motiveras. Antag att det högra kuvertet innehåller y kr. Observera nu att om x är 100 kr (d.v.s. det vänstra kuvertet innehåller 100 kr), så är y 50 kr med sannolikheten 1/2 och 200 kr med sannolikheten 1/2 (d.v.s. det högra kuvertet innehåller antingen 50 kr eller 200 kr och det med lika stor sannolikhet). Är x 50 kr så är y 100 kr och om x är 200 kr så är y 100 kr. Att x är 100 kr inträffar med sannolikheten 1/2 och att x är 50 kr respektive 200 kr inträffar var och en med sannolikheten 1/4. Situationen kan illustreras med följande träddiagram.



Således gäller att $y=2x$ inträffar med sannolikheten 1 när $x=50$ och med sannolikheten 1/2 när $x=100$, d.v.s. sannolikheten för att $y=2x$ är

$$1/4 \cdot 1 + 1/2 \cdot 1/2 = 1/2$$

Vidare inträffar $y=1/2x$ med sannolikheten 1 när $x=200$ och med sannolikheten 1/2 när $x=100$, d.v.s. sannolikheten för att $y=1/2x$ är

$$1/4 \cdot 1 + 1/2 \cdot 1/2 = 1/2$$

Sammanfattningsvis gäller alltså att $y=2x$ med sannolikheten 1/2 och $y=1/2x$ med sannolikheten 1/2, d.v.s. det högra kuvertet innehåller 2x kr med sannolikheten 1/2 och 1/2x kr med sannolikheten 1/2.

Utgående från detta resonemang är det lätt att förstå var felet sitter i detta reviderade kuvertproblem. Det förväntade värdet av att ta det vänstra kuvertet är (se träddiagrammet)

$$1/4 \cdot 50 + 1/2 \cdot 100 + 1/4 \cdot 200 = 12{,}5 + 50 + 50 = 112{,}5$$

Det förväntade värdet av att ta det högra är (se träddiagrammet igen)

$$1/4 \cdot 1 \cdot 100 + 1/2 \cdot 1/2 \cdot 50 + 1/2 \cdot 1/2 \cdot 200 + 1/4 \cdot 1 \cdot 100 =$$

$$= 25 + 12{,}5 + 50 + 25 = 112{,}5$$

Det förväntade värdet av att ta det vänstra kuvertet är således lika stort som av att ta det högra.

Det innebär med andra ord att det förväntade värdet av x är 112,5 och det förväntade värdet av y är 112,5. Ty x och y är här variabla storheter; x är ju beloppet i det vänstra kuvertet och det kan vara 50, 100 eller 200, medan y är beloppet i det högra kuvertet som också kan anta värdena 50, 100 eller 200. Variabla storheter av detta slag brukar inom sannolikhetsteorin kallas för stokastiska variabler (på engelska "random variables"). Karl Menger har framhållit att en variabel storhet i många sammanhang ska förstås som en funktion, se t.ex. Menger (1959). Detta anses ofta som kontroversiellt, men då det gäller stokastiska variabler är det helt accepterat. Den numera gängse definitionen av en stokastisk variabel lyder ungefär på följande sätt (Se t.ex. Blom, 1969, sid 3:2):

> En stokastisk variabel är en funktion definierad på ett utfallsrum.

En stokastisk variabel kallas *diskret* om den kan anta ett ändligt eller uppräkneligt antal olika värden. För mina syften i denna uppsats är det lämpligt att särskilja just fallet att den stokastiska variabeln antar ett ändligt antal värden. Låt oss därför säga att en stokastisk variabel är *ändlig* om den endast kan anta ett ändligt antal värden. En ändlig stokastisk variabel antar vart och ett av sina värden med en viss sannolikhet. Antag att den stokastiska variabeln X kan anta värden $a_1,..,a_n$. Det innebär att X är definierat på ett utfallsrum U med icke-tomma delmängder $A_1,..,A_n$ sådana att de är parvis disjunkta och tillsammans uttömmer U och sådana att $X(e)=a_i$ för alla $e \in A_i$. Om sannolikheten för att A_i inträffar är p_i, så antar alltså X värdet a_i med sannolikheten p_i, d.v.s. $P(X=a_i)=P(A_i)$. Definiera en funktion f_X från X:s värdeområde på följande sätt:

$$f_X(a_i) = P(X=a_i)$$

d.v.s.

$$f_X(a_i) = P(A_i)$$

Således är f_X:s värde för a_i sannolikheten för att X antar värdet a_i. f_X brukar kallas X:s *sannolikhets-* eller *frekvensfunktion*. *Väntevärdet* E(X) av en ändlig stokastisk variabel X definieras som

$$E(X) = \sum_a a \cdot f_X(a)$$

där summationen alltså sker över alla värden a i X:s värdeområde.

För x och y i vårt exempel gäller alltså att de är ändliga stokastiska variabler. Såväl x:s som y:s värdeområde är {50,100,200}. Vidare gäller att

$$f_x(50) = f_y(50) = 1/4$$

$$f_x(100) = f_y(100) = 1/2$$

$$f_x(200) = f_y(200) = 1/4$$

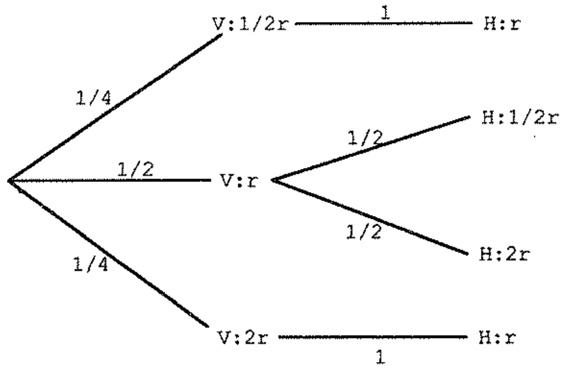
Således är

$$E(x) = E(y) = 1/4 \cdot 50 + 1/2 \cdot 100 + 1/4 \cdot 200 = 112{,}5$$

Nu torde felet i det tidigare resonemanget framgå klart. Det förväntade värdet av att ta det vänstra kuvertet är inte x kr, eftersom x är en stokastisk variabel, utan 112,5 kr, vilket är x:s väntevärde. Med tanke på att x och y är stokastiska variabler är det uppenbart orimligt att det förväntade värdet av en handling skulle vara x eller 1,25x. Det förväntade värdet i ett speciellt fall är ett tal och inte en funktion. E(x)=x och E(y)=1,25x är således helt orimliga påståenden på grund av den logiska formen hos x och E( ). Däremot är ju E(y)=1,25 E(x) något meningsfullt, men vi har inte visat att detta gäller.

3

Låt oss så ändra på det ursprungliga problemet på ett annat sätt. Antag att någon låter Dig välja mellan två kuvert och Du får veta att det ena innehåller ett visst belopp, låt oss säga r kr, och det andra antingen det dubbla (2r kr) eller hälften (1/2r kr). Däremot får Du inte veta vilket av kuverten som innehåller r kr och vilket som innehåller 2r kr eller 1/2r kr. Vi gör samma antaganden som tidigare om lika sannolikhet för de olika fallen. Situationen kan då beskrivas med hjälp av följande träddiagram, där V:k är en förkortning för att det vänstra kuvertet innehåller k kr och H:k analogt att det högra kuvertet innehåller k kr.



Det förväntade värdet av att välja det vänstra kuvertet är således

$$1/4 \cdot 1/2r + 1/2 \cdot r + 1/4 \cdot 2r = 1{,}125r$$

och det förväntade värdet av att välja det högra är

$$1/4 \cdot r + 1/2 \cdot 1/2 \cdot 1/2r + 1/2 \cdot 1/2 \cdot 2r + 1/4r = 1{,}125r.$$

Det är självklart inget konstigt att uttrycka det förväntade värdet av handlingar i termer av r, ty r är här ingen variabel storhet. r är i stället vad som ibland kallas en *godtycklig konstant* eller en *parameter*. Inom ramen för den situation vi betraktar antar den inte olika värden, d.v.s. är ingen stokastisk variabel. Är x beloppet i det vänstra kuvertet och y beloppet i det högra så är alltså $E(x)=E(y)=1{,}125r$.

I den aktuella situationen antar r alltså inte olika värden. Däremot kan man tänka sig att r antar olika värden i olika situationer eller sammanhang. r är alltså konstant *i* situationen men kan anta olika värden i *olika* situationer.

Man bör skilja mellan olika sorters konstanter. Terminologin kring detta är varierande. I Nordisk Familjebok 2:a upplagan kan man under uppslagsordet "konstant" läsa följande:

> **Konstant** (af lat. *constare*, förbli oförändrad), beständig, oföränderlig, som icke växlar, varaktig, oafbruten, ihållande. - *Mat.*, en storhet, som betraktas som oföränderlig till sitt värde. Konstanterna delas i *absoluta konstanter*, hvilka icke under några förhållanden förändras, t.ex. π (förhållandet mellan periferien och diametern i en cirkel), *e* (basen för det naturliga logaritmsystemet) och i allmänhet alla talvärden, samt *relativa konstanter* hvilka visserligen i en viss

kalkyl eller i förhållande till de egentliga variablerna anses oföränderliga, men icke dess mindre i en annan kalkyl eller i särskilda förhållanden kunna undergå förändring. Så är t.ex. i cirkelns ekvation,

$$(x-a)^2 + (y-b)^2 = r^2,$$

radien r konstant i förhållande till koordinaterna x och y, men kan betraktas som föränderlig, om cirkeln är framställd som typ för en serie cirklar, som alla ha samma medelpunkt. - En relativ konstant plägar understundom kallas *variabel parameter*, hvarmed utmärkes, att den väl är föränderlig, men dock vid sin förändring oberoende af de ursprungliga variablerna. (I.F.)

(I.F. står här för den framstående svenske matematikern Ivar Fredholm, 1866-1927.)

Under uppslagsordet "parameter" kan man i samma upplaga av Nordisk Familjebok bl.a. läsa följande:

Parameter (af grek. *parametrein*, afmäta), *mat*... 2.En storhet, som i förhållande till de i en ekvation eller expression ingående variabla storheterna är konstant, men dock kan antaga en kontinuerlig serie olika värden (jfr Konstant) och därigenom ger upphof till en serie af ekvationer eller expressioner. (I.F)

I Encyclopædia Britannica 15:e upplagan finner man bl.a. följande under "parameter".

parameter, in mathematics, a variable for which the range of possible values identifies a collection of distinct cases in a problem. Any equation expressed in parameters is a parametric equation. The general equation of a straight line in slope-intercept form, $y = mx + b$, in which m and b are parameters, is an example of a parametric equation. When values are assigned to the parameters, such as the slope $m = 2$ and the y-intercept $b = 3$, and substitution is made, the resulting equation, $y = 2x + 3$, is that of a special straight line and is no longer parametric.

I Webster's Third New International Dictionary (1971) kan man bl.a. läsa följande under "parameter":

an arbitrary constant characterizing by each of its particular values some particular member of a system (as of expressions, curves, surfaces, functions) <a parameter is a quantity which have various values each fixed within the limits of a stated case or discussion — T.F.Weldon>

Nu används också "parameter" med en annan innebörd än den som behandlas i dessa citat, nämligen i betydelsen hjälpvariabel. När man i den analytiska geometrin anger kurvor, ytor m.m. i parameterform, d.v.s. genom en parameterekvation, är det med denna innebörd av parameter. Men det är således inte parameter i betydelsen hjälpvariabel som är av intresse i denna uppsats utan i betydelsen relativ (godtycklig) konstant.

Vad som är viktigt i kuvertproblemet är alltså att skilja mellan en stokastisk variabel som varierar i den aktuella situationen eller det aktuella experimentet och en parameter som är konstant i situationen eller experimentet men kan anta olika värden i olika "fall" (d.v.s. i olika situationer eller experiment som dock är relaterade till varandra t.ex. genom att utgöra olika specialfall av ett generellare problem).

Nu skulle någon kanske vilja mena att då man antar i det första reviderade kuvertproblemet (då beloppen 100 resp. 50 eller 200 stoppas i kuverten) att det vänstra kuvertet innehåller x kr, så är x inte att förstå som en variabel storhet; x står inte för beloppet i det vänstra kuvertet som kan vara 50, 100 eller 200. Utan x står för det vi finner när vi öppnat det vänstra kuvertet. (Vi öppnar det vänstra kuvertet och finner - låt oss säga x kr.) x är då snarast en parameter, kan man tycka. Men vad gäller då om det högra kuvertet? Jo, i det finns (1) 2x med sannolikheten 1 om x=50, (2) 1/2x med sannolikheten 1/2 och 2x med sannolikheten 1/2 om x=100, samt (3) 1/2x med sannolikheten 1 om x=200. Vi får alltså särskilja tre fall som inte kan behandlas analogt oberoende av vad x är och som inträffar med sannolikheterna 1/4, 1/2 och 1/4 respektive. Detta särskiljande av tre fall är alltså inget annat än att återge x dess variabilitet, d.v.s. är ekvivalent med att se x som en variabel storhet. Avgörande här är att x antar olika värden *i* den situation vi betraktar (till skillnad från r i det andra reviderade kuvertproblemet som är konstant i den situation vi betraktar men varierar *mellan* olika analoga situationer).

4

Låt oss nu se på det ursprungliga problemet. För tydlighets skull gör jag det något mer detaljerat. Antag att någon låter Dig välja mellan två kuvert som innehåller pengar. Denna någon har slumpmässigt valt ett belopp som placerats i det vänstra kuvertet. Därefter har denna någon slumpmässigt valt att placera antingen halva eller dubbla beloppet i det högra kuvertet. Det ena kuvertet innehåller alltså dubbelt så mycket som det andra. Låt x vara beloppet i det vänstra kuvertet. x är då en stokastisk variabel med Re^+, de positiva reella talen, som värdeområde. Vidare tänks sannolikhetsfördelningen över möjliga värden för x vara likformig ("jämn"). x är här alltså inte ändlig utan kan anta ett oändligt antal värden. Frekvensfunktionen för x kan därför inte definieras som tidigare utan går över x:s fördelningsfunktion. Låt oss kortfattat skissera detta.

Om X är en stokastisk variabel, låt $P(X \leq t)$ vara sannolikheten att X är mindre eller lika med t. Fördelningsfunktionen för X betecknas F_X och definieras av att

$$F_X(t) = P(X \leq t) \ .$$

Observera att $F_X(t) \to 1$ då $t \to \infty$, ty $P(X \leq t) \to 1$ då $t \to \infty$.

Frekvensfunktionen för X är derivatan av $F_X(t)$ med avseende på t och betecknas $f_X(t)$. (F_X måste uppfylla vissa villkor för att frekvensfunktionen ska existera, men jag förbigår här dessa tekniska detaljer.) Väntevärdet för X kan nu definieras som

$$\int_{-\infty}^{\infty} t f_X(t) dt \ .$$

Låt oss så återvända till exemplet. Vilken är fördelningsfunktionen för x? För varje positivt reellt tal t är $F_x(t)=0$. Ty antag att $F_x(a)=k \neq 0$ för något positivt reellt tal a. Eftersom sannolikhetsfördelningen ska vara likformig ska gälla att $F_x(2a)-F_x(a)=F_x(a)=k$. Således gäller att $F_x(na)=nk$ där n positivt reellt tal. Om $n_1>1/k$ så gäller

$$F_x(n_1 a) = n_1 \cdot k > 1/k \cdot k = 1$$

d.v.s.

$$F_x(n_1 a) > 1 \ .$$

Men $F_x(t) \leq 1$ för alla positiva reella tal t, vilket innebär en motsägelse. Således är $F_x(t)=0$ för varje reellt tal t, d.v.s. F_x är konstantfunktionen 0. Men för att F_x ska vara en fördelningsfunktion för en stokastisk variabel ska $F_x(t) \to 1$ då $t \to \infty$, vilket inte gäller eftersom F_x är konstantfunktionen 0. Situationen är alltså inte beskrivbar inom ramen för den matematiska teorin.

Vi betraktar så det högra kuvertet. Låt y vara beloppet i detta. y är då en stokastisk variabel som kan anta alla positiva reella tal som värden. Vidare är liksom för x sannolikhetsfördelningen över möjliga värden likformig, vilket innebär att $F_y(t)=F_y(2t)-F_y(t)$. Detta innebär, analogt med vad som gäller för x, att fördelningsfunktionen för y antar värdet 0 för varje argument. Inte heller detta fall låter sig alltså beskrivas inom ramen för den matematiska teorin (som jag här skisserat den).

Innebär det inte en allvarlig brist att den matematiska sannolikhetsteorin inte kan hantera kuvertproblemet? I så fall är det dock en brist av matematiskt slag, en matematisk ofullkomlighet, ty problemet är, i sin "oändliga" form, ett rent teoretiskt (matematiskt) problem. Bland de få saker man säkert kan veta om vår värld är att vi aldrig kommer att stöta på det i verkliga livet. Och det beror inte bara på en brist på rika generösa typer utan fastmer på att man i den konkreta verkligheten aldrig stöter på en äkta slump-situation som måste beskrivas av en stokastisk variabel vars fördelningsfunktion är konstantfunktionen 0. Sådana situationer är alltid matematiska idealiseringar och ett internt problem för den matematiska teorin.

Men säger inte våra intuitioner oss att väntevärdet måste vara större än 0 för både x och y? Ty skulle vi inte välja att ta vilket som helst av kuverten framför att inte ta något? Nu tror jag emellertid att man ska vara försiktig med att extrapolera sina intuitioner anpassade till konkreta, "ändliga", situationer långt ut i oändligheten. Vetenskapshistorien är full av exempel på hur fel man då kan hamna. (Låt mig f.ö. komma med brasklappen att mer avancerad matematisk sannolikhetsteori som inte är begränsad till Riemann-integrering kanske tillhandahåller en lösning på kuvertproblemet — jag vet inte.)

Alltnog, hur kuvertproblemet ska behandlas inom den matematiska teorin är inte väsentligt för den fråga som främst intresserar oss här, nämligen vad det är som leder till *paradoxen* i kuvertproblemet som det formuleras i Filosofisk Tidskrift. Svaret på detta är alltså att man behandlar en stokastisk variabel (beloppet i det vänstra kuvertet) som om det vore en parameter (nämligen då man beräknar väntevärdena). Jag tar det som intäkt för att distinktionen variabel – parameter är värd att uppmärksamma även inom sannolikhetsteorin.

5

Nu är kuvertproblemet som det formuleras i Filosofisk Tidskrift inte säkert det problem Peter Gärdenfors har haft i tankarna. Enligt samtal med Gärdenfors är detta kanske i stället ungefär följande. Antag att någon låter Dig välja mellan två kuvert som innehåller pengar, och Du får veta att det ena innehåller dubbelt så mycket som det andra. Liksom i paragraf 4 antar vi mer specifikt att denna någon har slumpmässigt valt ett belopp som placerats i det vänstra kuvertet, och därefter valt att placera halva eller dubbla beloppet i det högra kuvertet. Du väljer det ena kuvertet – låt oss säga det vänstra – och öppnar det. Därefter får Du möjlighet att ångra Dig och i stället ta det högra. Hur bör Du göra?

Antag att Du finner att det vänstra kuvertet innehåller – låt oss säga – x kr. Det högra innehåller då 0,5x kr med sannolikheten 1/2 och 2x kr med sannolikheten 1/2. Det förväntade värdet av det högra är således $1/2 \cdot 0{,}5x + 1/2 \cdot 2x = 1{,}25x$ medan det förväntade värdet hos det vänstra är x kr. Det förväntade värdet är alltså större för det högra kuvertet. Men på ett helt analogt sätt kan man - om man i stället hade valt det högra kuvertet - visa att det förväntade värdet är större för det vänstra kuvertet. Du bör alltså alltid ångra Dig och välja det kuvert som Du inte först valde, och det oavsett vad Du finner i det först öppnade kuvertet. Är detta inte paradoxalt?

Skillnaden mellan detta problem och det som formuleras i Filosofisk Tidskrift är att nu öppnar man det ena kuvertet och kontrollerar vad det innehåller. En viktig fråga är vad det innebär för den logiska formen hos x; är x, d.v.s. det belopp Du finner i det vänstra kuvertet, en variabel eller en parameter?

I studiet av det "nya" kuvertproblemet använder vi oss av samma metod som tidigare och betraktar några olika situationer mer eller mindre besläktade med den vi är intresserade av.

Antag först att någon låter Dig välja mellan två kuvert och Du får veta att det vänstra innehåller 100 kr och det högra antingen 50 kr eller 200 kr och att sannolikheten är lika för båda möjligheterna. Det ena innehåller alltså dubbelt så mycket som det andra. Vilket kuvert bör Du välja? Det förväntade värdet av att välja det högra är alltså $1/2 \cdot 50 + 1/2 \cdot 200 = 125$ kr. Det förväntade värdet av det vänstra är ju 100 kr, så Du bör välja det högra.

Låt oss nu ändra på exemplet något så att det får följande lydelse: Antag att någon låter Dig välja mellan två kuvert och Du får veta att det vänstra innehåller ett visst belopp, säg x kr, och det högra med lika stor sannolikhet antingen hälften, d.v.s. 0,5x kr, eller det dubbla, d.v.s. 2x kr. Vilket kuvert bör Du välja? Det förväntade värdet av att välja det högra är naturligtvis

$$1/2 \cdot 0{,}5x + 1/2 \cdot 2x = 1{,}25x \text{ kr}$$

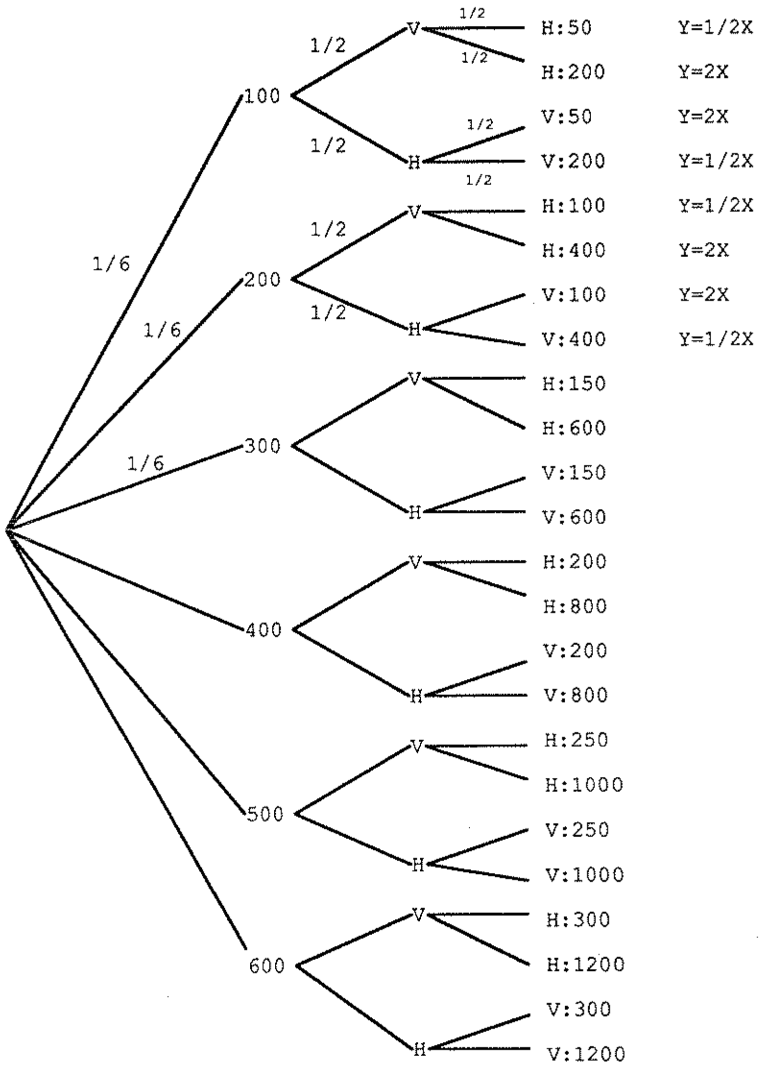
ty x är här uppenbarligen en parameter. Av samma skäl är det förväntade värdet av det vänstra x kr. Du bör alltså välja det högra, som har störst förväntat värde. Observera att vi inte kan "starta med det högra kuvertet" och "visa att det förväntade värdet är större för det vänstra kuvertet".

Detta exempel visar att i valet mellan två kuvert, det vänstra med ett fixt belopp och det högra med antingen hälften eller det dubbla, vardera med sannolikheten 1/2, ska Du välja det högra, d.v.s. inte det som har ett fixt belopp.

6

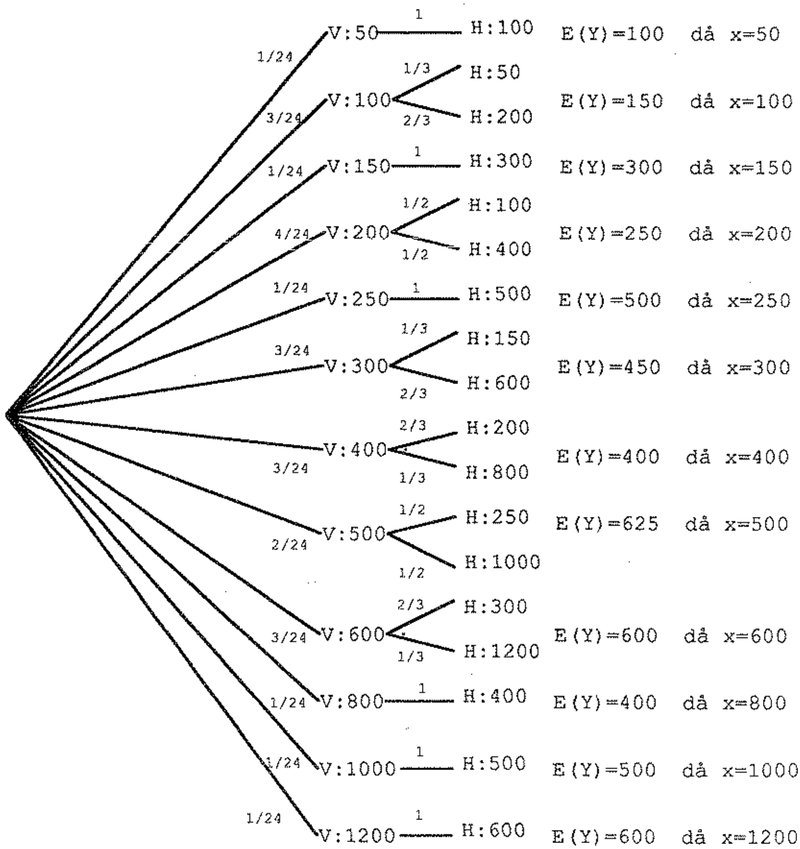
Det hade nu varit naturligt ur en viss synpunkt att betrakta följande exempel: Antag att någon låter Dig välja mellan två kuvert. Du får veta att det ena innehåller 100 kr och det andra antingen hälften eller det dubbla, vardera möjligheten med sannolikheten 1/2. Men det är lite för enkelt för mina syften. Låt oss därför se på ett något mer komplicerat exempel.

Antag att N väljer ett belopp genom att kasta tärning. Kommer i "ögon" upp så väljer N $i \times 100$ kr ($1 \leq i \leq 6$). Därefter avgör N genom att singla slant i vilket kuvert, det vänstra eller det högra, som beloppet ska läggas. Slutligen låter N slantsingling diktera om hälften eller det dubbla ska läggas i det andra kuvertet. Låt X vara beloppet i det vänstra kuvertet och Y beloppet i det högra. Situationen kan beskrivas med hjälp av träddiagrammet:



Observera att sannolikheten för att Y=1/2X är 1/2 och sannolikheten för att Y=2X också är 1/2. Du får nu välja ett av kuverten och öppna det. Därefter har Du möjlighet att ångra Dig och ta det andra.

Antag nu att Du väljer det ena av kuverten, låt oss säga det vänstra, och öppnar det. Låt oss säga att Du finner x kr. Vad är sannolikheten för att det andra innehåller 2x respektive 1/2x? Det beror naturligtvis på vad x är. Om x är 50 kr är sannolikheten 1 för att det andra innehåller 100 kr. Är x 100 kr så är sannolikheten 1/3 för att det högra innehåller 50 kr och 2/3 för att det innehåller 200 kr. Det förväntade värdet av att ta det högra kuvertet, d.v.s. väntevärdet av Y, är alltså 1/3·50+2/3·200=150 kr om x=100 kr. O.s.v. Situationen beskrivs av följande träddiagram:



Du skall alltså ångra Dig om x=50,100,150,200,250,300 eller 500. Däremot inte om x=800,1000 eller 1200. Om x=400 eller 600 är det förväntade värdet av att ta det högra kuvertet lika stort som att ta det vänstra.

7

Om Du i exemplet med tärningen ska ångra Dig och ta det andra kuvertet beror alltså på vad Du finner i det Du öppnar. Ty sannolikheten för att det andra kuvertet innehåller hälften respektive dubbelt så mycket beror på vad Du finner i det kuvert Du öppnar. I det exempel vi utgick ifrån, vilket är beskrivet i början av paragraf 5, är situationen inte sådan; vad Du än finner i det första kuvertet Du öppnar så är sannolikheten för att det andra innehåller hälften respektive dubbelt så mycket densamma (lika med 1/2). Givet att det är så bör Du alltid ångra Dig och välja det kuvert Du inte har öppnat. Situationen är analog med den då Du har att välja mellan två kuvert, där det ena innehåller ett fixt belopp och det andra hälften eller det dubbla, vardera med sannolikheten 1/2. Du ska då välja det som inte innehåller det fixerade beloppet. Men exemplet är helt teoretiskt och Du kan vara absolut säker på att aldrig i det praktiska livet stöta på en sådan situation. Ty exemplet förutsätter att sannolikheten att beloppet i det vänstra (eller högra) kuvertet (innan något kuvert öppnats) ligger inom vilket som helst ändligt intervall är 0. Således är inte bara frekvensfunktionen för X, där X är beloppet i det vänstra kuvertet, utan också fördelningsfunktionen för X lika med konstantfunktionen 0 (se sektionen 4). En dylik situation är, som tidigare påpekats, en ren matematisk fiktion. Att vi kan uppleva förhållandet att vi alltid ska ångra oss som paradoxalt beror på att vi tillämpar intuitioner hämtade från konkreta, praktiskt möjliga situationer – där vad vi finner i det öppnade kuvertet i åtminstone något fall har betydelse för sannolikheten för att hälften eller det dubbla finns i det andra kuvertet – i en situation av rent matematiskt slag där de alls inte hör hemma.

Jag kommer i detta sammanhang att tänka på följande exempel som David Hilbert lär ha gett under sina föreläsningar om oändligheten.

> Vi tänker oss ett hotell med ett ändligt antal rum och antar att alla rummen är upptagna. En ny gäst anländer och frågar efter ett rum. "Tyvärr", säger hotellvärden, "är alla rum upptagna." Låt oss nu tänka oss ett hotell med ett oändligt antal rum och att alla rummen är upptagna. Även till detta hotell kommer en gäst och frågar efter rum.
>
> "Men naturligtvis!" utropar hotellvärden, och han flyttar den person som tidigare innehaft rum N1 till rum N2, personen i rum N2 till rum N3, personen från rum N3 till rum N4 o.s.v. Och den nya gästen får rum N1, som blev ledigt på grund av dessa omflyttningar.
>
> Låt oss nu tänka oss ett hotell med ett oändligt antal rum, alla upptagna, och ett oändligt antal nya gäster som kommer in och frågar efter rum.
>
> "Naturligtvis, mina herrar", säger hotellvärden, "vänta bara ett ögonblick."
>
> Han flyttar nu innehavaren av rum N1 till N2, innehavaren av N2 till N4, innehavaren av N3 till N6 o.s.v., o.s.v.
>
> Nu blir alla udda rum lediga och det oändliga antalet nya gäster kan lätt placeras i dessa.

Detta finns att läsa i Gamow (1949) (sid 21 i pocketupplagan från 1962). Gamow skriver i en fotnot att det är hämtat "ur den opublicerade och för övrigt aldrig skrivna men kända volymen 'Den fullständiga samlingen av Hilberthistorier' av R.Courant."

Gamow kommenterar Hilberts exempel på följande sätt:

> Det är kanske inte så lätt att tänka sig de förhållanden som Hilbert beskrivit, inte ens i dessa hotellbristens tider, men detta exempel visar utan tvivel att när vi sysslar med oändliga tal, möter vi egenskaper som rätt mycket skiljer sig från dem vi är vana vid i vanlig aritmetik.

8

Åter till exemplet med tärningen. I detta kan vi uppenbarligen inte se x, beloppet vi funnit i det vänstra kuvertet, som en parameter. För olika värden på x får vi olika fall som inte kan behandlas analogt. x kan därför inte behållas konstant i den situation vi betraktar utan x varierar i denna. Det är f.ö. ingen egentlig skillnad mellan x och X. I det just diskuterade exemplet med kuverten däremot ger alla värden på x (där x alltså är vad vi finner i det vänstra kuvertet) fall som kan behandlas analogt och vi kan därför genomföra resonemanget för dem alla med ett godtyckligt värde på x. Vi tycks därför kunna se x som en godtycklig konstant (parameter) i den situation vi betraktar och låta variabiliteten hos x ligga mellan olika situationer. Den situation vi betraktar blir då ett av den familj av analoga fall som erhålls för olika värden på x.

9

Finns det någon sensmoral i kuvertproblemet är det kanske följande. Skilj på variabel och parameter också inom sannolikhets- och beslutsteorin. Och extrapolera inte intuitioner utvecklade på basis av ändliga, konkreta fall till oändliga, abstrakta. Med detta i minnet kan man undvika många skenproblem och kan inrikta sig på verkliga i stället.

Appendix

Min mening om det "kuvertproblem" som formuleras i början av paragraf 5 är alltså följande: Givet förutsättningen att vad Du än finner i det första kuvertet är sannolikheten för att det andra innehåller hälften resp. dubbelt så mycket densamma (lika med 1/2), så bör Du ångra Dig och ta det andra kuvertet. Man kan fråga sig om inte detta har betydelse för hur man bör välja även då man inte får ångra sig. Antag att man just ska till att välja och gör tanke-experimentet att man väljer det vänstra kuvertet och öppnar det. Finge man nu ångra sig och ta det högra så bör man göra det. Bör man då inte vid det val man just ska till att göra välja det högra? Men gör man samma tankeexperiment med det högra så finner man naturligtvis att man bör ångra sig och ta det vänstra. Bör man inte då vid det val man just skall göra ta det högra? Som Buridans åsna står man där.

Till det skulle jag vilja säga följande.

(1) Ett formellt påpekande. Låt X vara en stokastisk variabel som kan anta ett ändligt antal värden $r_1,\dots,r_n$. Sätt $R=\{r_1,\dots,r_n\}$. Låt Y vara en stokastisk variabel som är definierad för samma utfallsrum som X och med samma sannolikhetstilldelning till elementarhändelserna och som kan anta ett

ändligt antal värden $s_1,\dots,s_n$. Sätt $S=\{s_1,\dots,s_n\}$. Om $r\in R$ så är alltså $X=r$ en händelse i utfallsrummet för Y. Det är alltså meningsfullt att bilda $P(Y=s\mid X=r)$ där $s\in S$ och $r\in R$. Vi kan därför beräkna väntevärdet för Y givet att $X=r$, som vi betecknar $E(Y\mid X=r)$. Uppenbarligen gäller

$$E(Y\mid X=r) = \sum_{s\in S} s\cdot P(Y=s\mid X=r) .$$

Nu är ju

$$P(Y=s\mid X=r) = P(Y=s \;\&\; X=r)/P(X=r) .$$

Alltså är

$$\sum_{r\in R}(P(X=r)\cdot E(Y\mid X=r)) = \sum_{r\in R}\left(P(X=r)\cdot\sum_{s\in S}(s\cdot P(Y=s \;\&\; X=r)\;/\;P(X=r))\right) =$$

$$=\sum_{r\in R}\left(\sum_{s\in S}P(X=r)\cdot s\cdot P(Y=s \;\&\; X=r)/\;P(X=r)\right) = \sum_{r\in R}\left(\sum_{s\in S}s\cdot P(Y=s \;\&\; X=r)\right) =$$

$$=\sum_{s\in S}\left(\sum_{r\in R}s\cdot P(Y=s \;\&\; X=r)\right) = \sum_{s\in S}s\left(\sum_{r\in R}P(Y=s \;\&\; X=r)\right) = \sum_{s\in S}s\cdot P(Y=s) = E(Y).$$

Vi har därvid utnyttjat att

$$\sum_{r\in R}P(Y=s \;\&\; X=r) = P(Y=s)$$

ty om $r_1\neq r_2$ så $X=r_1$ och $X=r_2$ disjunkta händelser varför

$$P(Y=s \;\&\; X=r_1) + P(Y=s \;\&\; X=r_2) = P((Y=s \;\&\; X=r_1) \text{ el } (Y=s \;\&\; X=r_2)) =$$
$$=P(Y=s \;\&\; (X=r_1 \text{ el } X=r_2)) .$$

Således gäller

$$E(Y) = \sum_{r\in R}(P(X=r)\cdot E(Y\mid X=r)) .$$

(2) Nu gäller ju

$$E(X) = \sum_{r\in R}(P(X=r)\cdot r) .$$

Kombineras det med resultatet i (1) erhålls att om $E(Y\mid X=r)>r$

för varje $r \in R$, så är $E(Y)>E(X)$.

(3) Analogt med i (1) gäller

$$E(X)=\sum_{s\in S}(P(Y=s)\cdot E(X\mid Y=s))$$

Om $E(Y\mid X=r)>r$ för varje $r\in R$ så är det uppenbart att $E(X\mid Y=s)<s$ för något $s\in S$. (Annars skulle vi ha $E(X)\geq E(Y)$ och $E(Y)>E(X)$.)

(4) Låt oss så återgå till exemplet vi betraktar. Låt X vara beloppet i det vänstra kuvertet och Y beloppet i det högra. Låt R vara de möjliga värdena på X och S de möjliga värdena på Y. För varje $r\in R$ gäller alltså $E(Y\mid X=r) = 1{,}25r$. Ovan har vi visat

$$E(Y)=\sum_{r\in R}(P(X=r)\cdot E(Y\mid X=r))$$

medan

$$E(X)=\sum_{r\in R}(P(X=r)\cdot r)$$

Eftersom $E(Y\mid X=r)>r$ för alla $r\in R$ så är $E(Y)>E(X)$. Men nu är ju $E(X\mid Y=s) = 1{,}25s$ för varje $s\in S$. Eftersom

$$E(X)=\sum_{s\in S}(P(Y=s)\cdot E(X\mid Y=s))$$

och

$$E(Y)=\sum_{s\in S}(P(Y=s)\cdot s)$$

samt $E(X\mid Y=s)>s$ för alla $s\in S$ gäller $E(X)>E(Y)$. Vi har alltså erhållit en paradox.

Jag tror att det är ungefär så man tänker då man upplever att det är något paradoxalt när man har att välja mellan två kuvert och inser att hur man än gör kommer man att ångra sig. Att man inte får tänka så är ju uppenbart. Eftersom vi här har att göra med oändliga summor kan vi inte dra slutsatsen från $E(Y\mid X=r)>r$ för alla $r\in R$ att

$$\sum_{r\in R}(P(X=r)\cdot E(Y\,|\,X=r)) > \sum_{r\in R}(P(X=r)\cdot r)\,.$$

Sådana fel var vanliga på 1600- och 1700-talen och var en av drivkrafterna bakom gränsvärdesbegreppets utveckling. Som Gamow framhållit: "...när vi sysslar med oändliga tal, möter vi egenskaper som rätt mycket skiljer sig från de vi är vana vid i vanlig aritmetik." I själva verket är E(X) och E(Y) inte definierade i detta exempel.

Referenser

Blom, G. (1969): *Sannolikhetsteori och statistik med tillämpningar*. Studentlitteratur, Lund.

Gamow, G. (1949): *Ett två tre ... oändligheten*. Pocketupplaga 1962, Aldus/Bonniers, Stockholm. Det engelska originalet publicerat 1947.

Menger, K. (1959): 'Mensuration and other mathematical connections of observable material'. I Churchman & Ratoosh (red.): *Measurement: Definitions and Theories*, Wiley, New York.