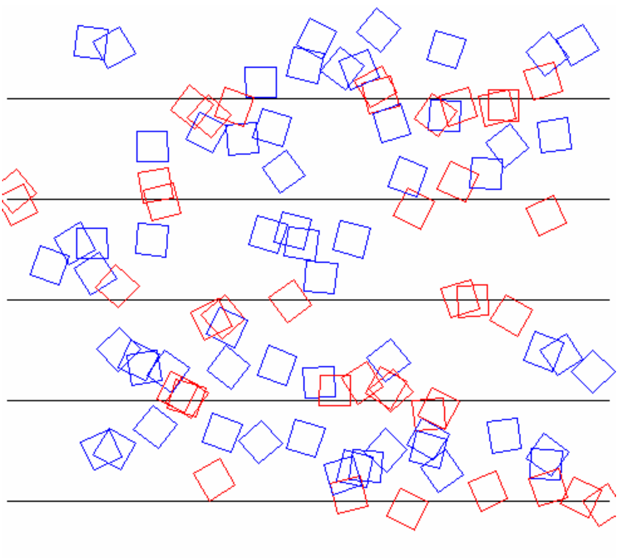
Allan C. Malmberg

CHANCE OG RISIKO

Kan det virkelig passe?



INFA 2006

Allan C. Malmberg

CHANCE OG RISIKO

Kan det virkelig passe?

Faglige udfordringer med løsninger

INFA 2006

Seneste publikationer af samme forfatter:

CHANCE – Et it-læremiljø, INFA 2002
En fagdidaktisk behandling af emner fra sandsynlighed som de kan fremtræde i en indledende undervisning. Bogen henvender sig til grundskolens matematiklærere.

Håndbog i sandsynlighed og statistik, INFA 2003
Bogen behandler væsentlige temaer fra sandsynlighed og statistik. Den henvender sig til cand.pæd.er i matematik. En del af håndbogens opslag vil kunne benyttes af brugere som ikke har en baggrundsviden på kandidatniveau.

Lær om chancer. Sanne og Malene går på opdagelse med computeren, INFA 2005
Bogen viser hvordan de to piger arbejder med emner fra sandsynlighedsregningen. De faglige udfordringer tages op med computeren som et vigtigt værktøj. Bogen henvender sig til elever med særlige interesser og evner for matematik.

Matematik i glimt, INFA 2005
Eksempler på matematiske ræsonnementer og bevisførelser fra forskellige områder af matematikken. Bogen henvender sig til elever med særlige interesser og evner for matematik.

CVU København og Nordsjælland
Titangade 11, København N

ISBN 87-91786-12-6

Forord

Denne bog henvender sig til matematiklærere i grundskolen som i deres uddannelse har stiftet bekendtskab med emner fra sandsynlighedsregning og statistik.

Bogen giver en indføring i fagområdet gennem udvalgte opgaver og deres løsninger. Opgaverne er af to typer, dels opgaver som går tilbage til sandsynlighedsregningens tidlige historie i 1600- og 1700-tallet, dels opgaver som naturligt kan tages op i tilknytning til en indledende undervisning i emner fra fagområdet.

Opgaverne i bogen er udfordrende på to måder. De kan give udfordringer som sætter læserens faglige baggrund på prøve, og de kan give udfordringer til læserens umiddelbare chance- og risikofornemmelse når de fører til resultater som er i klar modstrid med det intuitive.

Bogens emner fremlægges ved brug af "nænsom matematik" hvor indsigt opnås uden at alle faglige detaljer behandles: Udledningerne knytter sig til nogle få konkrete situationer med talværdier som gør det muligt for læseren at overskue situationen. I fortsættelse heraf vil mere generelle resultater blive fremlagt uden tilhørende teoretiske udledninger. - De faglige resultater i arbejdet med bogens opgaver vil ofte blive støttet gennem data fra computersimuleringer.

Skulle læseren få inspiration til at gå i detaljer med opgavesituationerne, vil det opfylde et af bogens formål. Og som ved al beskæftigelse med matematik gælder:

Er opgaven svær, er den umagen værd !

Allan C. Malmberg

Indhold

Indholdsbeskrivelse

1. Mønter og terninger
Her omtales nogle opgaver som traditionelt tages op i en indledende undervisning i sandsynlighed. Opgaverne har løsninger som ofte strider mod skoleelevers umiddelbare chancefornemmelse. – Nogle af opgaverne behandles mere detaljeret i senere afsnit af bogen.

2. Sig det med chancetræer
I dette afsnit ser vi på klassiske småopgaver fra den elementære sandsynlighedsregning. Det er opgaver der undertiden omtales som ”chanceparadokser” idet de kan give anledning til fortolkninger der fører til vidt forskellige resultater. I de fleste tilfælde vil en beskrivelse af den forelagte situation i form af et chancetræ fjerne formuleringens uklarheder og dermed "det paradoksale”.

3. Fødselsdagsproblemet
Et problem der hører til de absolutte klassikere. Der findes næppe en elementær lærebog i sandsynlighed som ikke behandler dette eksempel på en opgave med en overraskende løsning. De teoretiske svar på de stillede opgaver kan let efterprøves med computersimuleringer.

4. Nabotal
Her er endnu et elementært problem med en overraskende løsning. Nabotalproblemet burde have helt samme pædagogiske status som fødselsdagsproblemet, men det har stået lidt i skygge af opgaverne vedrørende fødselsdage. - Afsnittet afsluttes med en omtale af de ugentlige udtrækninger af lottotal.

5. Sammentræf
Et chanceeksperiment består i at fordele n kugler tilfældigt i n celler med én kugle pr. celle. Hvis en kugle placeres i en celle hvis nummer er identisk med kuglens, tales om et sammentræf. Hvor mange sammentræf vil der i gennemsnit forekom-

me i chanceeksperimentet? Og hvad er sandsynligheden for at der ikke forekommer nogen sammentræf i eksperimentet? Også her vil der være tale om problemer med overraskende løsninger.

6. Kommer ulykker ofte tre ad gangen?
Et ordsprog siger "En ulykke kommer sjældent alene" eller "Ulykker kommer ofte tre ad gangen". Vi skal i dette afsnit se om der er noget om snakken. - Endvidere ser vi generelt på fænomenet "Sjældne hændelser" og undersøger hvilke lovmæssigheder der kan ligge bag dem.

7. En spiller går fallit
A og B spiller mod hinanden i en serie af spil. A har en startkapital på a spillemærker, og B har en startkapital på b spillemærker. Efter hvert spil modtager vinderen 1 spillemærke fra taberen. A har i hvert spil en sandsynlighed på p for vinde, og B har en sandsynlighed på q for at vinde. Spillene fortsætter indtil en af de to spillere er gået fallit. - Hvad er sandsynligheden for at spillene slutter med at A går fallit? Hvad er sandsynligheden for at spillene slutter med at B går fallit? Hvor mange spil kræves der i gennemsnit for at bringe spillene til afslutning?

I afsnittet ses også på slumpture hvor en partikel bevæger sig i koordinatsystemet styret af tilfældigheder. Hvad er chancen for at partiklen vender tilbage til udgangspunktet når turen foregår i én dimension, to dimensioner, tre dimensioner?

8. Fuldt sæt
En terning kastes indtil alle seks mulige øjental er forekommet. Hvor mange kast skal der i gennemsnit udføres før et fuldt sæt af øjental foreligger? I afsnittet undersøges en række situationer der kan beskrives ved modellen: Fra en æske med n kugler udtages kugler én for én med tilbagelægning. Hvor mange udtagelser skal foretages før alle kugler har været udtaget?

9. Den bedste strategi

I en pose findes n sedler, og på hver seddel er der skrevet et tal. Fra posen udtages sedler én for én og uden tilbagelægning. Spilmasteren læser det udtrukne tal op, og spilleren skal afgøre om han vil stoppe eller gå videre. Hvis han stopper ved det største af alle n tal, vinder han den udsatte gevinst. Hvis han ikke stopper ved det største tal, har han mistet muligheden for gevinst. - Hvilken strategi skal spilleren bruge for at hans chance for gevinst bliver så stor som mulig? Svaret er overraskende enkelt, og chancen for gevinst er overraskende stor uanset antallet ad sedler.

10. En stok brækkes

En stok brækkes i tre tilfældige dele. Hvad er chancen for at de tre dele kan sammensættes til en trekant? Vi arbejder her med geometrisk sandsynlighed, et emne der havde stor bevågenhed i ældre lærebøger i sandsynlighed. I den geometriske sandsynlighed repræsenteres udfald ved punkter på linjer eller kurver eller ved punkter i plane områder, og sandsynligheder beregnes ved hjælp af længder og arealer.

11. Gennemsnit: uendelig

I sandsynlighedsregningens tidlige periode optrådte overvejelser over sandsynligheder ofte i tilknytning til spil. Man var således optaget af spørgsmålet om en retfærdig pris for at deltage i et spil. En spiludbyder tilbyder et spil mod en betaling af et beløb. Hvor stort skal dette beløb være for at der kan tales om en retfærdig pris for både spiller og spiludbyder?

I dette afsnit tages to situationer op hvor den betragtede sandsynlighedsfordeling ikke har nogen middelværdi, det vil sige at middelværdien er uendelig stor. De fører begge til resultater som sætter ens intuition på en hård prøve.

12. Hvordan skal puljen deles?

I en konkurrence mellem to spillere, A og B, indgår en serie af spil. Det antages at A og B er lige gode til spillet, og at de hver har en sandsynlighed på ½ for at vinde i hvert enkelt spil.

Spillets regler er sådan at den spiller der først når et forud aftalt antal af vundne spil, modtager hele spillepuljen.

Konkurrencen bliver imidlertid afbrudt da A stadig mangler a vundne spil for at vinde konkurrencen, og B mangler b vundne spil for at vinde konkurrencen. I hvilket forhold bør spillepuljen deles mellem de to spillere?

Dette problem som i sandsynlighedsregningens historie kaldes *delingsproblemet*, går helt tilbage til tiden før 1500.

13. Hverdagens skæve tal

Vi skal i dette afsnit se på hverdagens tal som de fremtræder i tabelværker og datasamlinger. Vi vil se på det første ciffer i disse tal, og interessere os for den frekvens de enkelte cifre 1,2,3,...,9 optræder med som førsteciffer. Her gemmer sig nogle overraskende resultater.

14. Bertrands paradoks

Afsnittet behandler et klassisk problem inden for den geometriske sandsynlighed: I en cirkel trækkes en tilfældig korde. Hvad er sandsynligheden for at korden er længere end siden i cirklens indskrevne ligesidede trekant?

I afsnittet fremlægges tre løsninger på det forelagte problem. Derefter gives en forklaring på at der kan optræde flere løsninger som alle ser ud til at være korrekte, men som fører til forskellige resultater.

15. Kast ting på gulvet og find π

Vi ser igen på geometrisk sandsynlighed: En række parallelle linjer med samme indbyrdes afstand er tegnet. En plan figur kastes ned på mønstret af linjer. Hvad er chancen for at figuren skæres af en af de tegnede linjer? - Et klassisk problem som går tilbage til 1700-tallet. Vi ser på kast med nåle, trekanter, cirkler og kvadrater og viser hvordan eksperimenterne kan føre til bestemmelse af talværdien for π.

16. Er det virkelig tilfældigt?
Vi tager udgangspunkt i en dataindsamling: På en skole fører man statistik over hvor mange uheld der sker i skolegården. Det viser sig at der i løbet af 20 skoledage i en måned sker 20 uheld som kræver at skolens forbindingskasse kommer i brug. Statistikken viser at uheldene ikke er jævnt fordelt på de 20 dage. Der er en enkelt dag med 3 uheld, og der er hele 7 dage hvor der slet ikke indtræffer nogen uheld.

Vi ser på hvordan den slags hændelser kan fordele sig når de underkastes tilfældighedens love. Generelt ser vi på situationer der kan beskrives ved "kugler fordeles i celler".

17. Talmønstre med overraskelser
Dette er bogens længste afsnit. Vi ser igen på møntkast. Vi betegner Krone med 1 og Plat med 0, og vi vil se på de talmønstre der fremkommer i en serie af møntkast. I jagten på mønstre holder vi os til mønstre der kun består af to kast eller tre kast. Det vil vise sig at der ligger mange overraskende resultater og venter på os.

Computerprogrammer
Af INFAs computerprogrammer er følgende benyttet i arbejdet med bogens emner:

Lod, Lod2, Kugle123, KugleX,
Buffon, Slump,Træ, First, Stattabel,
Strategi, Rekord, Talmønster.

Samtlige programmer er programmeret af Søren Ravn

1. Mønter og terninger

Vi ser her på nogle simple opgaver som ofte forekommer i elementære lærebøger i sandsynlighedsregning.

1. Møntkast

Møntkast er hyppigt anvendt som eksempler i en indledende undervisning. Selv om situationen med møntkast ser enkel og gennemskuelig ud, kan der godt forekomme fejltagelser.

(1) Der udføres 4 møntkast. I hvert kast noteres om kastet gav Krone (K) eller Plat (P).

Hvilket af følgende tre kasteforløb er mest sandsynligt:

KKKK; KKPP; KPKP

Svaret er at de er lige sandsynlige, de har alle tre en sandsynlighed på 1/16.

(2) Der udføres 4 møntkast. Hvad er mest sandsynligt:

(1) Alle fire kast giver Krone
(2) Af de fire kast giver to Krone og to Plat
(3) "3+1", dvs. tre af en slags, en af den anden slags

De tre sandsynligheder er her: 1/16, 6/16 og 8/16. Bemærk at "3+1" er mere sandsynlig end "2+2".

(3) I et spil kastes 10 mønter. Hvad er sandsynligheden for at netop 5 af mønterne viser Krone? - Mange vil overvurdere denne chance, 5 er jo "det rigtige" antal Krone i et kast med 10 mønter. Men sandsynligheden for netop 5 gange Krone er under 25%.

(4) Der kastes tre mønter. Hvad er sandsynligheden for at de alle viser ens, altså enten KKK eller PPP. Her er et ræsonnement: Af de tre mønter må mindst to vise ens. For den tredje mønt er der en chance på 1/2 for at den viser det samme som de to. Altså er svaret: 1/2.

Men det er forkert. Hvor ræsonneres der galt? Og hvad er det rigtige svar?

2. Den franske encyklopædi i 1700-tallet

I dette værk skrev den franske matematiker d'Alembert (1717 – 83) en række artikler om opgaver inden for sandsynlighed. Fra 1754 stammer artiklen "Plat og Krone". Her fremlægger han det problem at få mindst ét kronekast i to kast med en mønt. Hans løsning er denne: Hvis første kast giver Krone, så er der ingen mening i at udføre det andet kast. Hvis første kast giver Plat, må det andet kast udføres. Der er derfor tre mulige kasteforløb der kan komme på tale:

K; PK; PP

Af disse tre forløb forekommer Krone i de to første, altså er sandsynligheden 2/3 for mindst en Krone.

Forfatteren overser at de tre forløb ikke er ligevægtede. Selv en kort kasteserie med rigtige mønter vil afsløre at hvert af de tre forløb ikke kan tillægges en sandsynlighed på 1/3.

Løsningen er at forløbet K har en sandsynlighed på 1/2 og de to andre forløb har hver en sandsynlighed på 1/4. Sandsynligheden for K eller PK, dvs. mindst ét kronekast er derfor:

$$1/2 + 1/4 = 3/4.$$

Også situationen med tre møntkast blev behandlet i artiklen. Her foreligger ifølge d'Alembert fire mulige kasteforløb:

K;   PK;   PPK;   PPP

Derfor er der, siger d'Alembert, en sandsynlighed på 3/4 for at få mindst ét kronekast. - Det korrekte svar er her: 7/8.

3. En opgave fra 1711.
I en afhandling finder man denne opgave: Hvad er sandsynligheden for at få mindst to seksere i et kast med otte terninger?

Vi beregner sandsynligheden for 0 seksere og sandsynligheden for at få 1 sekser:

$P(0) = (5/6)^8 = 0.2326$ $P(1) = 8 \cdot(1/6)\cdot(5/6)^7 = 0.3721$

Sandsynligheden for at mindst 2 seksere er da: 1 – 0.2326 – 0.3721 = 0.3953, dvs. ca. 40%.

4. Huygens' lærebog fra 1657
I historiens første lærebog i sandsynlighedsregning finder vi følgende opgave: Hvor mange kast skal der udføres med en terning for at der er mindst 50% chance for at få to seksere?

Vi kan beregne det søgte antal efter metoden fra opgave 3. Idet vi sætter det ukendte antal af kast til x skal der gælde:

$1 - P(0) - P(1) \geq 0.5$ dvs. $1 - (5/6)^x - x\cdot(1/6)\cdot(5/6)^{x-1} \geq 0.5$

Ved afprøvning af x-værdier finder vi at x skal være 10 for at sandsynligheden for mindst to seksere kommer over de 50%. For x = 10 er sandsynligheden 0.5155, for x = 9 er den kun 0.4573.

I afsnit 6 vender vi tilbage til en generel løsning af opgaver hvor der skal foretages en beregning af hvor mange forsøg der skal gøres for at opnå 50% chance for en given hændelse.

5. Den rigtige løsning?
I en æske er der to røde kugler og to sorte. Fra æsken udtages to gange en kugle uden tilbagelægning. Hvad er chancen for at de to udtagne kugler har samme farve?

Her er tre forslag til løsning:

1. Der er tre muligheder: To røde kugler, to sorte kugler, en af hver slags. I to af de tre tilfælde er kuglerne af samme farve. Svaret er derfor: **2/3.**

2. Der er fire muligheder: To røde kugler, to sorte kugler, første kugle rød og anden sort, første kugle sort og anden rød. I to af de fire situationer er kuglerne af samme farve. Svaret er derfor: 2/4, dvs. **1/2.**

3. Der er seks muligheder. Blandt 4 kugler skal udvælges 2, det kan gøres på K(4,2) måder, dvs. 6. Hvis vi nummererer de to sorte kugler med 1 og 2, og de to røde kugler med 3 og 4, så er følgende seks udvalg mulige:

   (1,2), (1,3), (1,4), (2,3), (2,4), (3,4)

   I udvalgene (1,2) og (3,4) er de to kugler af samme farve. Svaret er derfor: **1/3.**

Afgør hvilken løsning der er den rigtige, og forklar hvorfor de to andre er forkerte.

Løs også opgaven hvis udtagelsen foretages med tilbagelægning efter første kugleudtagelse.

6. Hvor mange terninger var der i raflebægeret?
I et terningbæger findes røde og sorte terninger. Fra bægeret skal der rystes to terninger ud på bordet. Det oplyses at der vil være nøjagtig 50% chance for at begge de to terninger er

sorte. Hvor mange terninger var der i bægeret, og hvor mange af dem var røde?

Der er flere løsninger, men vi antager at det skal være den enkleste der søges, dvs. den som kræver det mindste antal terninger i bægeret.

Løsning: Man kan foretage nogle gæt, og derefter prøve om de giver den søgte sandsynlighed på 50%. Et rigtigt gæt er: 4 terninger, heraf 1 rød. Et andet gæt er: 21 terninger, heraf 6 røde.

7. Hvilket spil giver størst chance?

Du får tilbudt at deltage i et spil hvor der kastes med mønter. Du får gevinst hvis mindst 60% af mønterne viser Krone, og du kan vælge mellem tre versioner: Du kan kaste 5, 10 eller 25 mønter. Der er altså gevinst hvis du opnår henholdsvis mindst 3, mindst 6 og mindst 15 kronekast. Hvilken af de tre muligheder skal du vælge?

Mange vil mene at det kan være ligegyldigt hvilken version der vælges. Men sådan er det ikke. Jo flere kast du vælger, jo mindre bliver din chance for gevinst. Det hænger sammen med at det forventede er at 50% af kastene giver Krone. Når du derfor skal opnå 60% Kronekast, så vil du blive ramt af de store tals lov som siger at jo længere kasteserie, desto større chance for at antallet af dine Kronekast vil ligge nær ved 50%. Det er i små kasteserier der er størst chance for at afvige fra de 50%.

De tre sandsynligheder for gevinst er da også: Ved 5 kast 50%, ved 10 kast 38% og ved 25 kast 21%.

Det vil altså her være fornuftigt at vælge et spil med 5 kast.

8. Hvor er chancen størst?
I hvilken af disse to situationer er chancen størst:

1. Chancen for at få mindst 3 seksere i 10 terningkast

2. Chancen for at få mindst 6 seksere i 20 terningkast

Mange vil pr. intuition mene at de to chancer er lige store eller omtrent lige store. Men det er ikke sandt, den ene chance er mere end dobbelt så stor som den anden. Hvilken af dem vil du holde på som værende den største?

9. 50% chance
Du deltager i et spil hvor der er en chance i hvert spil på 1/100 for at vinde den store gevinst. Hvor mange spil skal du udføre for at have en chance på 50% for at få den store gevinst?

Et intuitivt svar er her at nogle siger 50 spil og andre siger 100. Men så enkelt er det ikke. Chancen for ikke at få den store gevinst i et spil er 99%. Chancen for ikke at få gevinsten i x spil er da: 0.99^x. For x skal nu gælde: $0.99^x < 0.5$. Det kræver: x=69. Svaret er altså at du må deltage i 69 spil for at have en chance på 50% for (mindst én gang) at modtage den store gevinst.

Hvis du deltager i et endnu sværere spil hvor chancen for den store gevinst kun er 1/1000, er svaret at du må deltage i næsten 700 spil, mere præcist 693, for at have 50% chance for den store gevinst. - Vi vender tilbage til problemet i afsnit 6.

10. Terningopgaver: Lette at formulere ...
Terningkast benyttes ofte i en indledende undervisning i sandsynlighed. Kast med to terninger hører til de mest benyttede eksempler på chanceeksperimenter, og de fleste skoleelever arbejder suverænt med det traditionelle udfaldsrum med 36 ligevægtede udfald.

1. Mønter og terninger

I tilknytning til terningkast kan der formuleres opgaver som er nemme at forstå, men som det ligger uden for skoleelevens faglige muligheder at løse. Her er nogle eksempler på sådanne opgaver og deres facitter.

Hvor mange terningkast skal der i gennemsnit udføres for at:

(1) Få en sekser? (2) Få to seksere?

(3) Få to ens? (4) Få to ens i træk?

(5) Få to seksere i træk?

(6) Alle seks mulige øjental forekommer?

Svarene på de seks opgaver er:

(1): 6 (2): 12 (3): 3.77 (4): 7 (5): 42 (6): 14.7

Svarene kan let efterprøves med computersimuleringer.

Her er data fra 10 000 serier af terningkast. I hver serie er der kastet indtil der foreligger to seksere i træk.

```
Talmønster. Statistik over 10000 eksperimenter.

   Mønster: 66    Talområde: 1..6
   Ventetid på forekomst af 66.


        Gennemsnit:    41.97 kast
        Maximum   :   316    kast
        Minimum   :     2    kast
```

Antallet af kast i serierne går fra 2 til 316 med et gennemsnit på 41.97 kast.

Historisk

Den engelske matematiker De Moivre (1667-1754) udsendte i 1718 lærebogen "The Doctrine of Chances" med undertitlen "En metode til beregning af sandsynligheder for hændelser i spil". Bogen udkom i flere udgaver, den mest omfattende er udsendt i 1756 to år efter forfatterens død.

I bogen, der er på ca. 350 sider, findes en omfattende samling af opgaver som formuleres i tilknytning til terningkast og andre elementære chancespil. Alle opgaver fremlægges med løsninger, og lærebogen har i 1700-tallet været et nyttigt værktøj for alle der ønskede at få en indføring i sandsynlighedsregningen, den nye matematiske disciplin.

I senere afsnit vil vi give nogle henvisninger til opgaver der er taget op i "The Doctrine of Chances".

2. Sig det med chancetræer

I dette afsnit ser vi på nogle klassiske småopgaver fra den elementære sandsynlighed. Det er opgaver der omtales som ”chanceparadokser” idet de kan give anledning til fortolkninger som fører til vidt forskellige resultater. Ofte vil en beskrivelse af den forelagte situation i form af et chancetræ fjerne uklarhederne og dermed "det paradoksale”.

1. De tre bægre
På bordet står tre raflebægere. I det ene er der to falske terninger, i det andet er der to ægte terninger, og i det tredje er der én ægte og én falsk terning. Der udvælges nu tilfældigt en af de seks terninger på følgende måde: Et bæger udvælges tilfældigt, og en af dets to terninger rystes ud. Hvis det viser sig at denne terning er falsk, hvad er da sandsynligheden for at den anden terning i bægeret også er falsk?

Et hyppigt forekommende svar er her: 1/2. Og argumentationen er sådan: Når der rystes en falsk terning ud af det valgte bæger, så kan der enten være tale om bægeret med én falsk terning, eller der kan være tale om bægeret med to falske terninger. Der er altså to muligheder, og af dem svarer den ene til at den anden terning i bægeret også er falsk. Altså er svaret: 1/2.

Svaret bygger på et udfaldsrum med to ligevægtede udfald. Når det oplyses at den valgte terning er falsk, så foreligger der to muligheder:

1. Terningen kommer fra bægeret med én falsk terning og én ægte terning
2. Terningen kommer fra bægeret med to falske terninger.

Men de to udfald er ikke ligevægtede. Der er ikke samme chance for at den falske terning kommer fra de to bægre. Lad os kalde de to falske terninger i udfald 2 for F1 og F2. Udfald 2 kan nu opdeles i:

2a. Terning F1 rystes ud
2b. Terning F2 rystes ud.

I virkeligheden er der tale om tre ligevægtede udfald: 1, 2a og 2b. Og i de to tilfælde 2a og 2b er den anden terning i bægeret også falsk. Det er derimod ikke tilfældet i udfald 1.

Svaret er derfor at i 2 ud af 3 ligevægtede udfald er den anden terning i bægeret også falsk. Det søgte svar er dermed: 2/3.

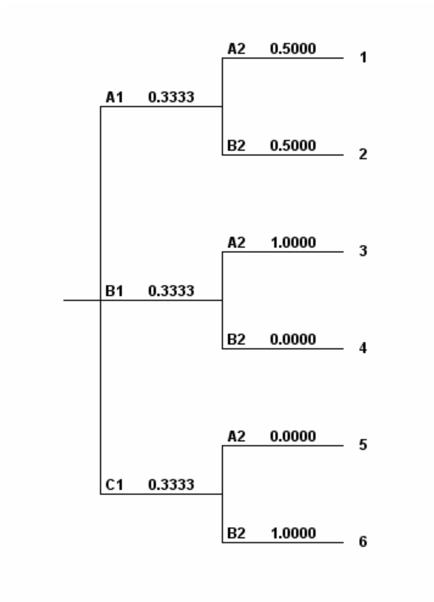
Et chancetræ til illustration
I chancetræet nedenfor svarer de tre grene A1, B1 og C1 til de tre bægre:

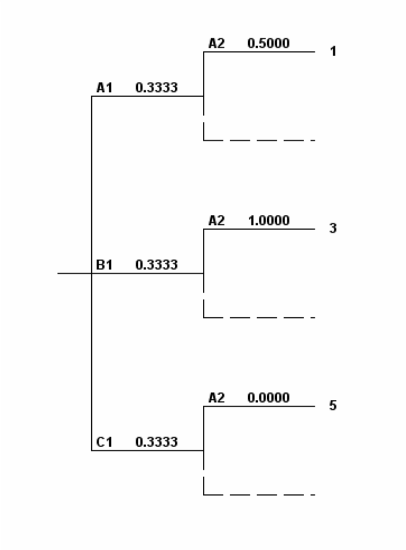
A1: Der udtages en terning fra bægeret der indeholder en falsk og en ægte terning
B1: Der udtages en terning fra bægeret der indeholder to falske terninger
C1: Der udtages en terning fra bægeret der indeholder to ægte terninger

De to grene A2 og B2 svarer til at der udtages henholdsvis en falsk og en ægte terning.

Her er først det totale chancetræ og derefter træet reduceret med hensyn til hændelsen A2: Der udtrækkes en falsk terning.

2. Sig det med chancetræer






```
Beregninger af chancer:
Reduc: A2          0.5000
B1                 0.6667
```

2. En bil og to geder

Dette problem blev kendt for den store offentlighed i 1990 hvor det blev lanceret i forbindelse med en amerikansk tv-konkurrence. Konkurrencedeltageren bliver stillet over for et valg: Her er tre døre. Bag den ene befinder sig en bil, bag hver af de to andre finder du en ged. Hvis du kan udpege den dør der skjuler bilen, får du bilen som gevinst.

Lad os antage at deltageren udpeger dør 1. Konkurrencelederen (som ved hvor bilen befinder sig) åbner derefter fx dør 3, og det viser sig at den skjuler en ged. Bilen må altså være bag dør 1 eller dør 2. Deltageren får nu det tilbud at han kan fastholde sit valg af dør 1, eller han kan skifte til dør 2. Hvad skal han vælge?

Dette problem skabte en heftig debat. Et stort antal af læserbreve (fra matematiske fagfolk og fra andre interesserede) gik ind for at det var uden betydning om deltageren fastholdt sit valg af dør 1, eller at han skiftede til dør 2.

Deres ræsonnement var sådan: Når konkurrencelederen har åbnet dør 3, så er der kun mulighed for at bilen befinder sig bag dør 1 eller dør 2. Det er to muligheder som må tillægges samme chance, altså har begge sandsynligheden ½ for at give gevinst, og det er derfor ligegyldigt om deltageren fastholder sit valg eller skifter til dør 2.

I konkurrencen er det underforstået at lederen ved hvor bilen befinder sig, og at lederen altid vil åbne en dør til en ged, og denne dør vælges blandt de to døre som deltageren ikke har valgt. Vi forudsætter i det følgende at deltageren altid vælger dør 1 (eller vi definerer hans valg af dør til at være dør 1). Hvis bilen befinder sig bag dør 1, så har lederen to valg: at åbne dør 2 eller dør 3. Hvis bilen fx befinder sig bag dør 2, så har lederen kun den mulighed at åbne dør 3, og hvis bilen befinder sig bag dør 3, så må lederen vælge at åbne dør 2. Når lederen har to valg, forudsætter vi at han trækker lod om de to muligheder (lodtrækningen foregår i al hemmelighed, så deltageren ikke ved at den finder sted).

Men der var fagfolk som ikke kunne acceptere at det var ligegyldigt om deltageren fastholdt sit valg eller skiftede. De argumenterede fx sådan:

På forhånd er der en chance på 1/3 for at deltageren har valgt den rigtige dør, altså at bilen findes bag dør 1. Der er altså en chance på 2/3 for at bilen befinder sig bag dør 2 eller 3. At den ikke er bag dør 3, er uden betydning. Når dør 3 er åbnet, så er al sandsynlighed for at bilen er bag en af de to døre overført til dør 2. Deltageren bør altså skifte fra det oprindelige valg af dør 1 til dør 2. Dermed øges hans chance for gevinst fra 1/3 til 2/3.

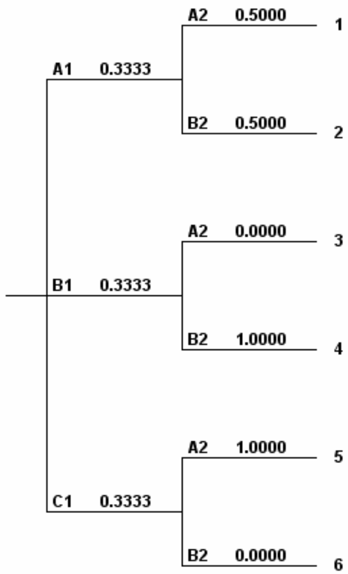
Bemærk at deltageren enten vil få gevinst ved at fastholde sit valg, nemlig hvis bilen befinder sig bag dør 1, eller han vil få gevinst ved at skifte valg, nemlig hvis bilen befinder sig bag dør 2 eller dør 3. Man kan derfor argumentere for et skift på følgende måde: Lad os antage at der spilles et stort antal spil, fx 300. Den deltager der fastholder sit valg, kan forvente at vinde i ca. 100 spil. Men hvis han vælger at skifte i alle spil, så vil han vinde i de øvrige ca. 200 spil. Hans chance for gevinst er altså 2/3 hvis han skifter.

Nedenfor er der et chancetræ af situationen. Her er forklaringen til træet.(Det forudsættes at deltageren har valgt dør 1):

A1: Bilen findes bag dør 1
B1: Bilen findes bag dør 2
C1: Bilen findes bag dør 3
A2: Lederen åbner dør 2
B2: Lederen åbner dør 3

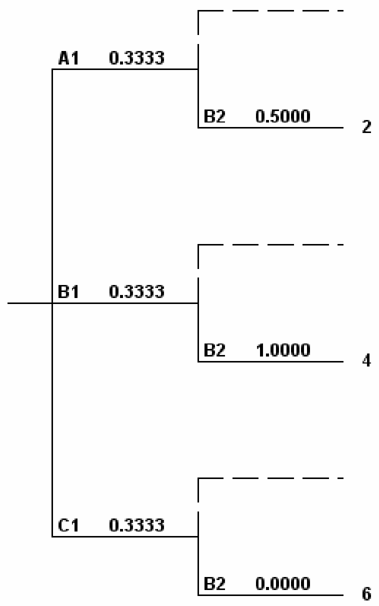
På grenene er sat sandsynligheder: Der er en sandsynlighed på 1/3 (på træet 0.3333) for at bilen er bag dør 1, og der er samme sandsynlighed for at den er bag dør 2 eller bag dør 3.

Endvidere er der, når bilen befinder sig bag dør 1, en sandsynlighed på ½ for at lederen åbner dør 2 og samme sandsynlighed for at dør 3 åbnes. (Hvad dette betyder for konkurrencen vender vi tilbage til). – Når bilen befinder sig bag en dørene 2 og 3, har lederen intet valg, da skal henholdsvis dør 3 og dør 2 åbnes.



Vi reducerer nu træet ud fra hændelsen B2, altså at lederen åbner dør 3. I det reducerede træ beregner vi sandsynligheden for hændelsen B1, dvs. at bilen befinder sig bag dør 2, og at deltageren altså vil vinde hvis han skifter sit valg fra dør 1 til dør 2.

2. Sig det med chancetræer



For hændelsen B1 i det reducerede træ får vi nu sandsynligheden

$$0.3333/ (0.3333+0.1667) = 2/3.$$

Det er altså en fordel for deltageren at skifte fra dør 1 til dør 2. Chancen for gevinst fordobles.

Vi kan benytte chancetræet til at undersøge andre muligheder. Det afgørende punkt er den strategi konkurrencelederen benytter i det tilfælde at bilen findes bag dør 1. Hvis han trækker lod om dør 2 eller dør 3, er – som vi har set – sandsynligheden for gevinst 2/3 ved et skift fra dør 1 til dør 2. Hvis lederen altid vælger dør 3, er sandsynligheden for gevinst kun 1/2.- Og hvis lederen fx benytter et terningkast til afgørelse af valget mellem dør 1 og 2 og ved øjentallene 1 og 2 vælger dør 2, ved de øvrige øjental dør 3, så er deltagerens chance for gevinst 0.6 ved skift til dør 2.

I øvrigt vil det gælde at hvis chancen for at lederen vælger dør 3 varierer fra 1 til 0, så vil deltagerens chance for gevinst ved at skifte til dør 2 variere fra ½ til 1. Dette kan let kontrolleres ved brug af et chancetræ.

Uden kendskab til konkurrencelederens strategi kan der derfor ikke gives noget entydigt svar på hvad deltagerens sandsynlighed for gevinst er ved et skift fra dør 1 til dør 2. Men det vil nok være fornuftigt at skifte: I værste fald ændres chancen for gevinst ikke, den er 1/2 ved begge valg. I bedste fald forøges sandsynligheden fra 1/3 til en værdi der kan nærme sig 1.

Der findes en omfattende litteratur om bil/ged - problemet. I litteraturlisten er blot anført en artikel fra 2006 af S. Götz.

3. To kast med en mønt

En spiller udfører to møntkast, kast nr. 1 og kast nr. 2. Han får ikke kendskab til resultaterne af sine kast, de er alene kendt af spilmasteren.

Spilmasteren stiller ham nu nogle opgaver i bestemmelse af sandsynligheder. Den første lyder:

Jeg kan oplyse dig at du har opnået mindst ét kronekast. Hvad er sandsynligheden for at begge dine mønter viser krone?

Spilleren opstiller et udfaldsrum for de to møntkast. Der er tale om fire ligevægtede udfald: KK, KP, PK, PP. Det første bogstav giver resultatet af første kast. KP betyder altså krone i første kast, plat i andet kast.

Spilmasterens oplysning reducerer mængden af udfald til de tre første: KK, KP, PK. Det vides jo at der forekommer mindst ét kronekast. Disse tre udfald betragter spilleren som ligevægtede. Et af dem, KK, svarer til at begge mønter viser krone. Svaret på det stillede spørgsmål er derfor: **1/3.**

Det andet spørgsmål fra spilmasteren lyder:

Jeg kan oplyse dig at dit første kast var et kronekast. Hvad er sandsynligheden for at begge dine mønter viser krone?

I dette tilfælde reduceres mængden af mulige udfald til: KK, KP. Disse to udfald betragter spilleren som ligevægtede. Et af dem svarer til at begge mønter viser krone. Svaret på det stillede spørgsmål er derfor: **1/2.**

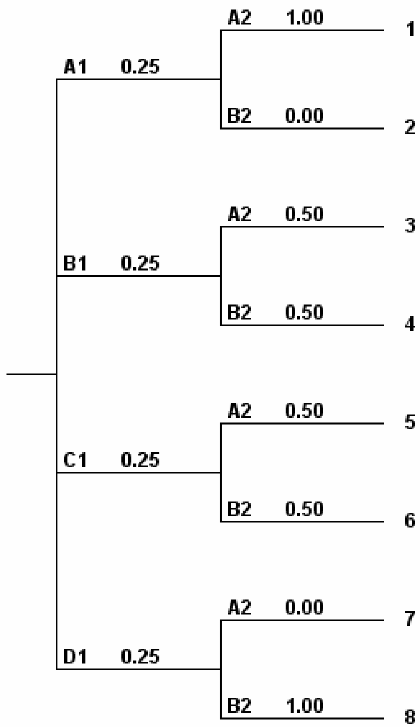
Her er det tredje spørgsmål fra spilmasteren:

Efter at du har kastet de to mønter, udvælges en af dem ved lodtrækning. Den viser krone. Hvad er sandsynligheden for at den anden mønt også viser krone?

Spilleren vælger her at benytte et chancetræ. På træet er benyttet følgende afmærkning:

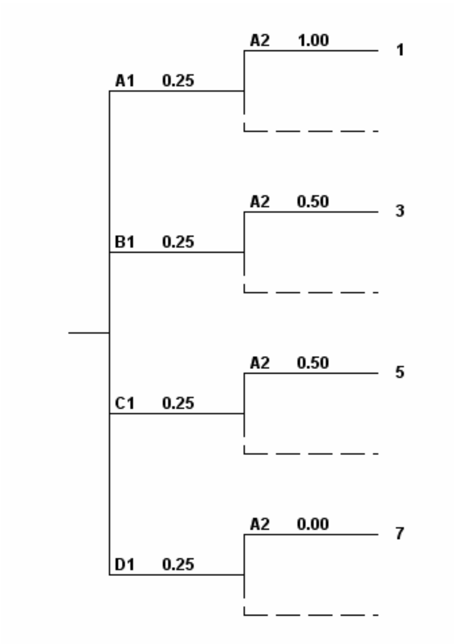
A1: Begge mønter viser krone
B1: Første mønt viser krone, anden mønt viser plat
C1: Første mønt viser plat, anden mønt viser krone.
D1: Begge mønter viser plat
A2: Den tilfældigt valgte mønt viser krone
B2: Den tilfældigt valgte mønt viser plat

2. Sig det med chancetræer



Træet reduceres nu på grundlag af at hændelsen A2 er indtruffet:

2. Sig det med chancetræer



Inden for dette træ beregnes sandsynligheden for A1:

$$0.25/(0.25+0.125+0.125) = \frac{1}{2}$$

Problemstillingen i ”To kast med en mønt” optræder ofte i en anden iklædning: Der er tale om en familie med to børn. Vi ser på følgende tre situationer:

1. Familien har mindst én dreng
2. Familiens ældste barn er en dreng
3. Et af børnene bliver udvalgt ved lodtrækning. Det blev en dreng.

I hver af disse situationer skal vi bestemme sandsynligheden for at familiens to børn begge er drenge. De tre situationer kan med fordel belyses ved brug af chancetræer.

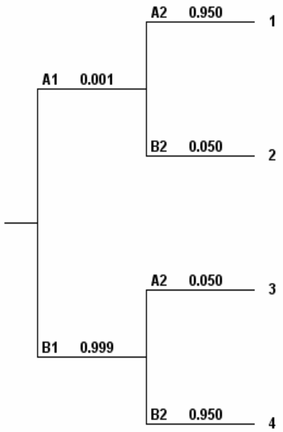
4. Den positive test
En sygdom har en udbredelse på 0.1% i en bestemt befolkningsgruppe, dvs. sygdommen er til stede hos 1 ud af 1000. En test til afgørelse af om sygdommen er til stede hos en person har en sikkerhed på 95%, dvs. hvis sygdommen er til stede, vil testen i 95% af tilfældene være positiv, og hvis sygdommen ikke er til stede, vil testen i 95% af tilfældene være negativ.

Fra befolkningsgruppen udtages en tilfældig person til testning. Det viser sig at testen er positiv. Hvad er sandsynligheden for at sygdommen er til stede hos den testede?

Nedenfor er et chancetræ over situationen.

På træet er benyttet følgende betegnelser:

A1: Sygdommen er til stede
B1: Sygdommen er ikke til stede
A2: Testen er positiv
B2: Testen er negativ



2. Sig det med chancetræer

Det stillede spørgsmål vedrører betinget sandsynlighed, nemlig sandsynligheden for A1 under betingelse af at A2 indtræffer. Denne sandsynlighed P(A1|A2) er:

$$P(A1|A2) = P(A1og\ A2) / P(A2) =$$

$$0.001\cdot 0.95/(0.001\cdot 0.95 + 0.999\cdot 0.05) = 1.87\%$$

Sandsynligheden for at den positivt testede person er ramt af sygdommen er altså mindre end 2%. En risiko som ligger langt under det tal de fleste intuitivt vil gætte på.

5. At svare på et delikat spørgsmål

Vi vil foretage en interview-undersøgelse blandt unge. Vi ønsker at undersøge hvor mange der har prøvet at tage narkotiske stoffer. Vi kan imidlertid ikke uden videre stille et direkte spørgsmål om sagen, det kunne give for mange uærlige svar.

Vi går derfor frem på følgende måde: Den adspurgte får en svarseddel, en mønt og en seddel med det delikate spørgsmål. Han går derefter ind i stemmeboksen og kaster mønten. Hvis den viser Krone, besvarer han det delikate spørgsmål og afkrydser sit svar (Ja eller Nej) på svarsedlen.

Hvis mønten viser Plat, kaster han den ekstra gang. Nu besvarer han på svarsedlen spørgsmålet: Viste mønten Krone i det sidste kast?

Han afleverer derefter sin svarseddel som indeholder ét kryds, enten ved Ja eller ved Nej.

Et ja-svar kan altså opstå på to måder: Ved et svar på det delikate spørgsmål, eller ved et svar på om mønten viste Krone i andet kast. Intervieweren kan ikke vide hvordan svaret er fremkommet.

Men han kan alligevel få oplysninger ud fra tallene.

Her er en analyse:

Første kast: Krone (50% chance)
Det delikate spørgsmål besvares, x% siger Ja.
Frekvensen af Ja-svar er altså her: 0.5x

Første kast: Plat (50%).
Andet kast giver Krone. (50%)
Chancen for rækkefølgen PlatKrone er altså 25%.

Frekvensen af Ja-stemmer i alt er dermed: 0.5x + 0.25.
Vi ønsker at bestemme x.

Taleksempel
Undersøgelsen giver 60% Ja og 40% Nej.

Vi har da: 0.5x + 0.25 = 0.60. Heraf: x= 70%

Altså skønner vi at 70% af de adspurgte svarer Ja til det stillede delikate spørgsmål. Og ingen af deltagerne er blevet afsløret.

Også i denne opgave kan en skitse af et chancetræ give en god oversigt over de foreliggende muligheder.

3. Fødselsdagsproblemet

Problemet: Hvad er chancen for at der i en forsamling af n personer findes to personer med samme fødselsdag?

Under forudsætning af at fordelingen af de n personers fødselsdage kan beskrives ved modellen for ordnet stikprøveudtagelse med tilbagelægning vil der være 365^n mulige ligevægtede fordelinger (vi ser her bort fra skudår). Såfremt der ikke må forekomme dobbeltfødselsdage, er antallet af mulige fordelinger 365^{n-},"365 opløftet i n minus" som er fastlagt ved:

$$365^{n-} = 365 \cdot 364 \cdot 363 \cdot ... \cdot (365 - n+1).$$

Sandsynligheden for at der forekommer dobbeltfødselsdag er derfor:

$$p = 1 - (365^{n-} / 365^n)$$

Nogle talværdier:

n= 10	p= 0.117
n= 15	p= 0.253
n= 20	p= 0.411
n= 22	p= 0.476
n= 23	p= 0.507
n= 30	p= 0.706
n= 50	p= 0.970
n= 70	p= 0.999
n= 80	p> 0.9999
n= 89	p> 0.99999
n= 97	p> 0.999999

Det overrasker de fleste at der allerede ved n=23 er over 50% chance for dobbeltfødselsdag. Og at chancen ved n = 50 er helt oppe på 97%.

3. Fødselsdagsproblemet

Dobbelt- eller tripelfødselsdag
For n = 25 kan beregnes følgende sandsynligheder:

Dobbeltfødselsdag forekommer p= 0.569

Heraf:	Netop ét par af dobbeltfødselsdag:	p= 0.379
	Netop to par af dobbeltfødselsdag:	p= 0.140
	Netop tre par af dobbeltfødselsdag:	p= 0.029
	Netop fire par af dobbeltfødselsdag:	p= 0.004
	Netop ét sæt af tripelfødselsdag	p= 0.009
	Netop ét sæt tripelfødselsdag samt ét sæt dobbeltfødselsdag:	p= 0.006

En simulering
Her er resultatet fra en simulering af fødselsdage i 1000 klasser med hver 25 elever. Simuleringen er foretaget med INFA-programmet Lod2.

```
Antal eksperimenter: 1000

Dobbeltfødselsdag forekommer:   55.5 %
Heraf:  To eller flere par      18.7 %
        Tre ens                  2.0 %
```

En tilnærmet beregning
Sandsynligheden for at der forekommer dobbeltfødselsdag kan med tilnærmelse beregnes af:

$$p \approx 1 - (1 - n/730)^{n-1}$$

3. Fødselsdagsproblemet

Taleksempler:

n = 10: p= 0.117   n= 20: p= 0.410   n=50: p= 0.969

Fraktilværdier
p-fraktilen $x_{p,}$ den x-værdi som afgrænser de nederste p procent af datamaterialet, kan beregnes tilnærmelsesvis ved:

$$x_p = (1 + \sqrt{1 - 8n \cdot \ln p})/2$$

For n=365 og p=0.5 fås: $x_{0.5}$ = 23.0.

Skudår og skæve fordelinger
Indflydelsen fra skudår med en ekstra dag hvis sandsynlighed for at forekomme som fødselsdag er 1/4 af sandsynligheden for de øvrige 365 dage, vil ændre de beregnede sandsynligheder en smule. De rokker imidlertid ikke ved at sandsynligheden for dobbeltfødselsdag ved n = 23 overstiger 0.5.

Heller ikke sæsonmæssige udsving i fødselsdagsfrekvens hen over årets måneder vil ændre ved at sandsynligheden ved n= 23 overstiger 0.5. Ved ekstreme svingninger vil det dog kunne forekomme at sandsynligheden allerede passerer 0.5 for en n-værdi under 23.

Ventetiden på sammenfald ved fordeling af kugler
Der foretages en tilfældig fordeling af kugler i n celler. Kuglerne fordeles én for én. En tilnærmet værdi for den gennemsnitlige ventetid V indtil første sammenfald (to kugler i samme celle) er givet ved:

$$E(V) \approx \sqrt{\pi \cdot n/2} - 1/3$$

Nogle talværdier:

n=12: E(V) = 4.0 n=365: E(V) = 23.6

n=1000: E(V) = 39 n=5000: E(V) = 88

Andre fødselsdagsopgaver

Dreng og pige med samme fødselsdag
I en skoleklasse er der lige mange drenge og piger. Hvor mange elever skal der være for at sandsynligheden for at en dreng og en pige har fødselsdag samme dag overstiger 50%? - Svaret er 32, altså 16 drenge og 16 piger.

Multifødselsdage
Ved en multifødselsdag er mere end to elever inddraget: Der kan være tale om tre der har fødselsdag samme dag, eller der kan være tale om flere par af dobbeltfødselsdage. Hvor mange elever skal der være i klassen for at chancen for multifødselsdag overstiger 50% ? Svaret er: 36.

Tredobbelt og firdobbelt fødselsdag
Hvor mange elever skal der være i klassen for at chancen for en tripelfødselsdag overstiger 50%? Og hvor mange ved en firdobbelt fødselsdag?

Svarene er henholdsvis: 88 elever og 187 elever.

3. Fødselsdagsproblemet

Her er en computersimulering med 1000 klasser på hver 88 elever:

```
LOD-Model:

Der udtages tal i talområdet 1..365
Et eksperiment består i at udtage 88 tal.
Det er tilladt at udtage samme tal flere gange.
Der udføres 1000 eksperimenter.

Resultat:

   Tre ens      518 eksperimenter = 51.8%
```

Et andet problem
Vælg en tilfældig dag i året. Hvor mange personer skal du udspørge for at finde en der har fødselsdag på den valgte dag?

Her er svaret at medianen i antallet af forespørgsler er 253. Der skal altså udspørges 253 personer for at chancen for at den bestemte dag forekommer, er 50%.

Sikkert også et svar der vil overraske de fleste.

4. Nabotal

Vi vil udvælge 3 tal blandt tallene 1,2,3,…,8. Vi kan hurtigt fastslå hvor mange udvalg der er mulige: K(8,3) = 8*7*6/3! = 56.

Spørgsmålet er nu: Hvor mange af disse 56 udvalg indeholder nabotal? Altså fx tallene 5 og 6, eller 2 og 3, eller 7 og 8.

Vi vil finde dette antal ved at beregne hvor mange af de 56 udvalgte der *ikke* indeholder nabotal.

Vi går frem på følgende måde: Der skal udtages 3 tal ud af 8. Der er altså 3 tal der skal udtages og 5 der ikke skal udtages. Vi betegner nu de tal der skal udtages med et X og de tal der ikke skal udtages med et O. De fem tal der ikke skal udtages stiller vi op i rækkefølge:

O O O O O

Vi ved endnu ikke hvilke tal det drejer sig om, men vi ved at der er 5 tal der ikke skal udtages.

Tilsvarende ved vi at der er tre tal der skal udtages: X X X.

Vi anbringer nu de tre kryds i rækken af O'er. I et mellemrum mellem to O'er må der ikke sættes mere end ét kryds. Men endvidere kan der sættes et kryds foran det første O eller efter det sidste. Her er et eksempel på placeringen af de tre kryds:

X O O X O X O O

Vi sætter nu tal på rækken af tegn: Der er 8 tegn i rækken svarende til de 8 tal hvorfra vi udtager 3 tal. Tæller vi forfra i rækken af tegn ser vi at krydsene står på plads nr. 1, nr. 4 og nr. 6. Rækken svarer derfor til at vi har udtaget udvalget som består af tallene 1, 4 og 6.

Du kan se, at ved denne metode vil vi altid få et udvalg som ikke indeholder nabotal. Der vil jo altid være et O mellem to X'er.

Hvor mange måder er der for at vælge pladserne til de tre X'er? Her kan vi bruge hvad vi ved om udvalg, vi skal jo vælge 3 pladser blandt de 6 mulige pladser i rækken af O'er. Der er 6 pladser, nemlig 4 mellemrum og 2 yderpladser. Antallet af sådanne udvalg er givet ved: K(6,3) = 6*5*4/3! = 20.

Der findes altså 20 udvalg *uden* nabotal. I alt var der 56 mulige udvalg. Så chancen for at udtage et udvalg uden nabotal er: 20/56 = 0.36 Men så er chancen for at et tilfældigt udvalg indeholder nabotal: 1 – 0.36 = 0.64. Der er altså 64% chance for at et tilfældigt udvalg indeholder nabotal.

Ved bestemmelse af chancen for udvalg med nabotal går vi altså frem på den måde at vi bestemmer chancen for udvalg ud*en* nabotal. Disse udvalg tæller vi op således: Vi tager en række af O'er svarende til de tal der ikke skal udtages. Derefter placerer vi X'er i denne rækkefølge.

På denne baggrund kan vi nu give følgende almene resultat for sandsynligheden for nabotal når der udvælges k tal blandt tallene 1...n:

$$\text{P(Nabotal)} = 1 - \frac{K(n-k+1,k)}{K(n,k)}$$

Nogle eksempler

Her er nogle talværdier for chancen for nabotal:

n=36	k=7	p= 75.6%
n=100	k=10	p= 62.9%
n= 365	k= 25	p= 82.8%

Andre resultater
Lad A_d være hændelsen: *Afstandene mellem de udtagne tal er alle større end d.* Da gælder:

$$P(A_d) = \frac{K(n - d \cdot k + d, k)}{K(n,k)}$$

Sandsynligheden for nabotal er givet ved: 1– $P(A_1)$.

Endvidere gælder:

$$\text{P(Netop ét par af nabotal)} = \frac{(k-1) \cdot K(n-k+1, k-1)}{K(n,k)}$$

$$\text{P(Netop to par af nabotal)} = \frac{K(k-2,2) \cdot K(n-k+1, k-2)}{K(n,k)}$$

$$\text{P(En stribe på r tal i træk)} = \frac{K(k-r+1,1) \cdot K(n-k+1, k-r+1)}{K(n,k)}$$

Tærskelværdien for nabotal
Ved tærskelværdien T(n) forstår vi det mindst antal tal der skal indgå i udvalget for at sandsynligheden for forekomsten af nabotal i udvalget overstiger 0.5. For tærskelværdien gælder:

$$T(n) \approx \sqrt{n \ln 2} + 1 \approx 5\sqrt{n}/6 + 1$$

For Lotto med n=36 fås: T(n) = 6.

Kombinationer i Lotto(36,7)
I det danske lørdagslotto hvor der blandt 36 tal udtrækkes 7 tal, er der i alt K(36,7) mulige udvalg af syv tal, dvs. 8 347 680 mulige udvalg.

4. Nabotal

Der kan forekomme følgende typer af nabotal: Tostriber, trestriber, firstriber, femstriber, seksstriber og syvstriber. I alt er der mulighed for 15 kombinationer af nabotal som anført i listen nedenfor. Antal af udvalg af de forskellige typer blandt de K(36,7) er:

Ingen nabotal: K(30,7) = 2 035 800

Én tostribe: 6 K(30,6) = 3 562 650

To tostriber: 10 K(30,5) = 1 425 060

Tre tostriber: 4 K(30,4) = 109 620

Én trestribe: 5 K(30,5) = 712 530

To trestriber: 3 K(30,3) = 12 180

En trestribe samt en tostribe: 12 K(30,4) = 328 860

En trestribe samt to tostriber: 3 K(30,3) = 12 180

En firstribe: 4 K(30,4) = 109 620

En firstribe samt en tostribe: 6 K(30,3) = 24 360

En firstribe samt en trestribe: 2 K(30,2) = 870

En femstribe: 3 K(30,3) = 12 180

En femstribe samt en tostribe: 2 K(30,2) = 870

En sekstribe: 2 K(30,2) = 870

En syvstribe: K(30,1) = 30

4. Nabotal

En simulering
Her er resultatet af 10 000 lottotrækninger. Simuleringen er foretaget med INFA-programmet Lod2.

```
Antal eksperimenter: 10000

Mindst to på stribe                    75.6 %
Mindst tre på stribe                   14.6 %
```

Og her har computeren udskrevet 10 lotterækker med afmærkning af nabotal. Af de 10 rækker indeholder de 8 nabotal, nogle af dem endda flere sæt af nabotal.

```
 1:  6     9    15    19    24    31    35

 2: 17 -- 18    20    23    29    32    35

 3:  8    19    23 -- 24    27    30    34

 4:  5 --- 6    16    19    26    34    36

 5:  2     6 --- 7    15    24    33 -- 34

 6: 12    22    25    27 -- 28    34 -- 35

 7:  5    10    16    21    25    29    33

 8:  9    15    24 -- 25    27 -- 28    30

 9:  1     3    18 -- 19    21    26    33

10:  1    10    12    16 -- 17    24    33
```

5. Sammentræf

Der foreligger n kugler med numre fra 1 til n. Endvidere foreligger n celler, ligeledes med numre fra 1 til n. Et chance-eksperiment består i at fordele kuglerne tilfældigt i cellerne med én kugle pr. celle. Hvis en kugle placeres i en celle hvis nummer er identisk med kuglens, forligger et sammentræf.

Problemer: Hvor mange sammentræf vil der i gennemsnit forekomme i chanceeksperimentet? Og hvad er sandsynligheden for at der ikke forekommer nogen sammentræf i eksperimentet?

Problemerne fremsættes i 1708 i en lærebog af den franske matematiker P. R. Montmort (1678-1719). Her kaldes det *Rencontre-problemet*.

Vi ser på nogle enkle eksempler. Lad os antage at n=3. Der foreligger altså 3 kugler som skal fordeles i 3 celler. Her er der seks muligheder for fordelinger, de tre kugler kan jo opstilles i seks rækkefølger:

1	2	3	Antal sammentræf: 3
1	3	2	Antal sammentræf: 1
2	1	3	Antal sammentræf: 1
2	3	1	Antal sammentræf: 0
3	1	2	Antal sammentræf: 0
3	2	1	Antal sammentræf: 1

Vi har altså for **n=3**:

Antal af sammentræf:	0	1	2	3
Sandsynlighed.	2/6	3/6	0	1/6

På samme vis får vi for **n=4:**

5. Sammentræf

Antal sammentræf:	0	1	2	3	4
Sandsynlighed:	9/24	8/24	6/24	0	1/24

Disse eksempler viser to generelle træk: Sandsynligheden for at der forekommer sammentræf for alle n kugler er 1/n!. Der er jo kun én af de n! mulige rækkefølger af kuglerne som giver n sammentræf. For n=3 er sandsynligheden altså: 1/3! = 1/6, og for n=4 er sandsynligheden: 1/4! = 1/24.

Endvidere ser vi at der optræder et nul i listen over sandsynligheder. For n=3 kan det ikke forekomme at der er netop to sammentræf. Hvis to af de tre kugler kommer i ”rette celle”, så må alle tre kugler jo være i rette celle.

Tilsvarende ser vi at for n=4 er sandsynligheden 0 for at der forekommer netop 3 sammentræf.

Hvor mange sammentræf i gennemsnit?

Vi kan beregne gennemsnittet i tilfældene n=3 og n=4 af tabellerne ovenfor. For n=3 får vi at gennemsnittet er:

$$0\cdot 2/6 + 1\cdot 3/6 + 2\cdot 0 + 3\cdot 1/6 = 6/6 = 1$$

Ved et eksperiment med 3 kugler vil der altså i gennemsnit forekomme ét sammentræf.

For n=4 får vi at gennemsnittet er:

$$0\cdot 9/24 + 1\cdot 8/24 + 2\cdot 6/24 + 3\cdot 0 + 4\cdot 1/24 = 24/24 = 1$$

Også ved et eksperiment med 4 kugler vil der i gennemsnit forekomme ét sammentræf.

Beregning af gennemsnittet
Vi vil nu foretage beregningen af gennemsnittet på en anden måde som kan give os resultatet for enhver værdi af n uden at vi behøver at opstille en liste over sandsynligheder for de forskellige antal af sammentræf.

Lad os se på n=4. Vi indfører 4 chancestørrelser: X_1, X_2, X_3, X_4 . De angiver hvor mange sammentræf der finder sted i celle 1, 2, 3 og 4. Til eksempel har vi for celle 1:

X_1 = 1 med sandsynligheden 1/4
X_1 = 0 med sandsynligheden 3/4

Der er jo fire kugler, og kun én af dem giver sammentræf i celle 1.

For de andre celler gælder tilsvarende:

X_j = 1 med sandsynligheden 1/4
X_j = 0 med sandsynligheden 3/4

I hver celle er der jo kun én kugle der kan give sammentræf.

Hver af de fire chancestørrelser har et gennemsnit på 1/4:

$$1 \cdot 1/4 + 0 \cdot 3/4 = 1/4$$

For X, det samlede antal sammentræf i de fire celler gælder nu:

$$X = X_1 + X_2 + X_3 + X_4$$

Heraf har vi at gennemsnittet af X er lig med summen af de fire gennemsnit, dvs. gennemsnittet af X er $4 \cdot 1/4 = 1$.

Helt samme fremgangsmåde kan anvendes for et vilkårligt antal af kugler. Her får vi at gennemsnittet af hver af de n chancestørrelser er 1/n. Ved det betragtede chanceeksperiment vil gennemsnittet af sammentræf, uafhængig af værdien af n, altså være: **1**

Variansen
Vi kan benytte de samme chancestørrelser til beregningen af variansen af antallet af sammentræf. For variansen $V(X_j)$ og covariansen $Cov(X_i, X_J)$ gælder:

$$V(X_j) = \frac{n-1}{n^2} \qquad \text{og} \qquad Cov(X_i, X_J) = \frac{1}{n^2(n-1)}$$

Herudaf kan variansen for $X = X_1 + X_2 + X_3 + ... + X_n$ beregnes:

$$V(X) = 1.$$

Såvel gennemsnit som varians, og dermed standardafvigelse, af antallet af sammentræf er altså 1 uanset værdien af n.

Sandsynligheden for mindst ét sammentræf
Her vil vi nøjes med at fremsætte nogle resultater uden udledning. Vi giver en liste over sandsynligheden for 0 sammentræf:

n=2. Sandsynlighed for 0 sammentræf: 1/2! = 1/2

n=3. Sandsynlighed for 0 sammentræf: 1/2! – 1/3! = 1/3

n=4. Sandsynlighed for 0 sammentræf:

1/2! – 1/3! + 1/4! = 9/24

n=5. Sandsynlighed for 0 sammentræf:

1/2! – 1/3! + 1/4! – 1/5!= 44/120

Listen går videre på denne vis: der skiftes mellem addition og subtraktion af et nyt led.

Med tre decimaler er sandsynlighederne ovenfor:

n=2: 0.500   n=3: 0.333   n=4: 0.375   n=5: 0.367

Det kan vises at sandsynlighederne nærmer sig til den matematiske konstant 1/e som med tilnærmelse har værdien: 0.36788.

Når blot n er større end 3, er sandsynligheden for 0 sammentræf nær ved 37%. Det betyder at sandsynligheden for mindst ét sammentræf er ca. 63%. Og det gælder uanset om n=4 eller n= 1000.

En rekursionsformel

Vi vil se på et eksempel: Vi kender antallet af fordelinger uden sammentræf når to kugler placeres i to celler, og vi kender antallet af fordelinger uden sammentræf når tre kugler placeres i tre celler. De to antal kalder vi a(2) og a(3). Af oversigten vi tidligere opstillede, har vi for de to antal:

$$a(2) = 1 \qquad\qquad a(3) = 2$$

Vi vil nu benytte disse to tal til at beregne a(4), antallet af fordelinger uden sammentræf når fire kugler placeres i fire celler. Vi bruger her en kombinatorisk optælling.

Vi ser på kugle 4. Den må ikke placeres i celle 4. Der er derfor **3** muligheder for placering af kugle 4: celle 1, 2 eller 3. Lad os antage at den placeres i celle 2.

Der er nu to muligheder for det videre forløb:

(1) Kugle 2 kan placeres i celle 4, det vil sige at kugle 2 og 4 bytter celler med hinanden.

(2) Kugle 2 placeres ikke i celle 4

I tilfælde (1) skal de to resterende kugler, 1 og 3, placeres uden at der forekommer sammentræf. Det kan gøres på **a(2)** måder.

I tilfælde (2) skal de tre kugler 1, 2 og 3 placeres sådan at der ikke sker sammentræf for kuglerne 1 og 3, og at kugle 2 ikke placeres i celle 4. Det betyder at der for hver af de tre kugler er en ”forbudt placering”: Kugle 1 må ikke placeres i celle 1, kugle 2 må ikke placeres i celle 4, og kugle 3 må ikke placeres i celle 3. Antallet af fordelinger uden at nogen kugle placeres i en forbudt celle er **a(3)**. Situationen svarer jo helt til fordelingen uden sammentræf af tre kugler i tre celler.

Alt i alt har vi derfor at a(4) kan beregnes ved:

$$a(4) = 3\cdot(a(2) + a(3)) = 3\cdot(1 + 2) = 9.$$

Generelt kan vi på samme måde udlede formlen som gælder for n>3:

$$\mathbf{a(n) = (n-1)\cdot(a(n-2) + a(n-1))}$$

Ved brug af denne formel får vi for de næste antal:

$a(5) = 4\cdot(a(3) + a(4)) = 44$ $\qquad$ $a(6) = 5\cdot(a(4) + a(5)) = 265$

Ud fra de fundne antal kan vi nu beregne sandsynligheden for at der ikke forekommer sammentræf:

$$\frac{a(n)}{n!}$$

En værdi der som nævnt ligger nær ved tallet 1/e $\approx$ 0.368.

Sandsynligheden for netop k sammentræf kan tilnærmes ved:

$$0.368/\,k!.$$

Heraf har vi at sandsynligheden for 0 sammentræf og for 1 sammentræf med tilnærmelse begge er 0.37. Det betyder at medianen i fordelingen af antal sammentræf er 1. Så både gennemsnit og median i fordelingen er 1.

Her er en computersimulering med programmet KugleX.

```
 100 kugler fordelt i 100 celler.
 Antal sammentræf i 100 eksperimenter

 Antal sammentræf  Antal gange
         0              39
         1              29
         2              26
         3               5
         4               0
         5               1



Gennemsnit  :   1.01
Standardafv.:   1.00
Maximum     :   5
Minimum     :   0
```

6. Kommer ulykker ofte tre ad gangen?

Et ordsprog siger "En ulykke kommer sjældent alene" eller "Ulykker kommer ofte tre ad gangen". Vi skal se om der er noget om snakken.

Vi lader computeren efterligne en situation med en ulykke som i gennemsnit forekommer én gang for hver 100 dage. Der kan fx være tale om en alvorlig trafikulykke eller et omfattende strømsvigt eller et nedbrud af et elektronisk system.

Ved undersøgelse af situationen tænker vi os at benytte en æske der indeholder 99 hvide kugler og én rød kugle. Hver dag udtages en tilfældig kugle fra æsken. Hvis kuglen er hvid, er der tale om en dag uden forekomst af ulykken. Hvis den røde kugle udtages, indtræffer ulykken den pågældende dag. - Efter udtagelsen af kuglen lægges den tilbage i æsken, således at den chancemæssige situation er uændret fra dag til dag.

I gennemsnit vil der kræves 100 udtagelser for at få den røde kugle, men der kan forekomme store udsving. Her er data fra 100 eksperimenter indtil den røde kugle blev udtaget.

```
Antal kugler der blev udtaget indtil den
røde kugle blev udtaget
```

182	305	**22**	**14**	120
133	105	4	82	110
523	15	41	163	82
24	**10**	54	75	200
75	376	142	170	159
28	19	87	**9**	**9**
44	18	28	333	108
21	92	129	278	80
59	144	216	89	356
12	295	57	132	16

6. Kommer ulykker ofte tre ad gangen?

46	89	3	30	18
157	60	99	42	177
184	117	102	105	205
53	271	102	77	51
17	81	10	169	20
143	**12**	**7**	62	73
338	41	22	256	29
42	21	2	140	79
306	117	**16**	**10**	53
205	76	1	135	61

I det første eksperiment gik der 182 dage før den røde kugle forelå. I det næste måtte vi vente i hele 305 dage. Men derefter var der to eksperimenter med kun 22 og 14 udtagelser.

Vi ser altså at der efter 2. forekomst af ulykken (med ventetiden 305) kun gik 22 dage før den forekom igen, og efter yderligere 14 dage forekom ulykken endnu en gang. Det betyder, at inden for 36 dage forekom ulykken tre gange. Da vi ved at der i gennemsnit går 100 dage mellem de enkelte forekomster, er det iøjnefaldende at ulykken nu optræder hele tre gange på bare 36 dage.

I oversigten ovenfor har vi fremhævet de situationer hvor to på hinanden følgende tal er små, dvs. højst 25. Vi ser at der fem gange forekommer sådan en klump af ulykker.

Det er den slags ophobninger af ulykker som vil blive bemærket. Så selv om der er mange "normale ventetider" i skemaet, så er det de fem ophobninger der giver anledning til undren. På denne baggrund kan det gamle ordsprog om "ulykker der ofte kommer tre ad gangen" være opstået.

For skemaet ovenfor er den gennemsnitlige ventetid på ulykken 105 dage. Men bemærk hvor meget tallene varierer: Fra 1 dag til 523 dage. Teoretisk er sandsynligheden 22% for en ventetid på 25 dage og derunder. Og sandsynligheden for to på hinanden følgende tal som begge er under 25, er omkring 5%. Så statistisk er der intet mærkeligt i at der forekommer fem ulykkesklumper i de 100 eksperimenter.

Sjældne hændelser

Vi vil nu se på sjældne hændelser og deres forekomst over en tidsperiode. Det kan fx dreje sig om at få en stor gevinst i klasselotteriet. Hvor mange år vil der gå før vor lotteriseddel kommer ud? Og kan det betale sig at spille videre på en seddel som lige har fået en stor gevinst?

En sjælden hændelse kan også være noget ubehageligt. En sjælden ulykke er lige indtruffet. Hvornår kan det frygtes at den indtræffer igen?

Megen overtro knytter sig til forekomsten af sjældne hændelser: *"Når en sjælden hændelse lige er indtruffet, så må det vare længe inden den forekommer igen".*

Vi bruger programmet Kugle 2

Vi vil ved hjælp af kuglemodeller se nærmere på de sjældne hændelser. Vi ser først på en hændelse som har en chance (eller risiko) på 5% for at forekomme inden for en bestemt tidsperiode, fx en periode på 1 år. Ved anvendelse af programmet Kugle2 vil vi nu undersøge denne hændelse nærmere.

En sandsynlighed på 5% svarer til 1 mulighed ud af 20. Vi kan derfor vælge en kuglemodel med 20 kugler hvoraf der er én rød kugle. Kugleudtagelsen foregår med tilbagelægning, således at der ved hver udtagelse er en chance på 5% for at trække en rød kugle.

Kuglemodellen afspejler den forelagte chancesituation på følgende måde: Hver kugleudtagelse svarer til at vi ser på et tidsforløb på et år. Hvis den udtagne kugle er rød, forekommer den sjældne hændelse inden for det pågældende år. Hvis den udtrukne kugle ikke er rød, forekommer hændelsen ikke i det betragtede år.

I aviser kan man se en hændelse der har en chance på 5% for at forekomme inden for et år, omtalt som en hændelse der forekommer en gang for hver 20 år. Vi vil nu se hvor meget hold der er i det.

Vi lader Kugle2 udføre 1000 eksperimenter.

```
 Kugle-Model K2(20,1,ja)

   Statistik over 1000 eksperimenter

      Talområde            Fraktil

        1..1                  1 %
        1..2                  5 %
        1..3                 10 %
        1..6                 25 %
        1..14                50 %
        1..28                75 %
        1..46                90 %
        1..56                95 %
        1..93                99 %

   Gennemsnit:    20.10
   Maximum:       152
   Minimum:         1
```

I oversigten ser vi at gennemsnittet af udtagelser i de 1000 eksperimenter er 20.10. Det svarer til at den sjældne hændelse i gennemsnit forekommer med ca. 20 års mellemrum. Vi ser også at 50%-fraktilen, medianen, er 14, dvs. i halvdelen af eksperimenterne er den sjældne hændelse indtrådt i løbet af de første 14 år.

Men der er store variationer i antallet af år vi måtte vente på den sjældne hændelse: Fra 1 år op til 152 år.

I 10% af eksperimenter indtraf den sjældne hændelse allerede i løbet af de første tre år. Men i 5% af tilfældene måtte vi vente i mere end 56 år før hændelsen indtraf!

Den forventede periode

En hændelse der har en chance på 5% - altså 1 ud af 20 - for at forekomme inden for et år, kan forventes at forekomme i gennemsnit én gang for hver 20 år. Dette kalder vi den *forventede periode* for hændelsens forekomst. I vore 1000 eksperimenter så vi at den gennemsnitlige ventetid var 20.10, altså tæt på den forventede periode. En ny kørsel af 1000 eksperimenter vil sikkert også give et resultat der kun afviger lidt fra tallet 20. Afvigelser på op til nogle få procent fra den forventede periode vil være helt almindelige ved kørsler der omfatter 1000 eksperimenter.

Talområde	Frekvens
..10	40.4 %
40..	14.6 %

Ved hjælp af programmet har vi beregnet hvad chancen er for at vi kan klare os med højst 10 eksperimenter, og hvad risikoen er for at vi må bruge mindst 40 eksperimenter. Vi ser at 40.4% af eksperimenterne krævede 10 udtagelser eller derunder. Det svarer til at den sjældne hændelse i 40% af tilfældene indtraf allerede inden for de første 10 år, altså inden for den første halvdel af den forventede periode.

Vi ser også at 14.6% af eksperimenterne krævede mindst 40 udtagelser. Det betyder at i ca. 15% af tilfældene måtte vi vente mindst 40 år på at den sjældne hændelse indtraf. Vi måtte altså i ca.15% af tilfældene vente længere end det dobbelte af den forventede periode.

Disse kendsgerninger vil vi kalde "*40-15 reglen for sjældne hændelser“:*

6. Kommer ulykker ofte tre ad gangen?

40-15 reglen
Der er ca. 40% chance for at den sjældne hændelse indtræffer allerede inden for den første halvdel af den forventede periode.

Der er ca. 15% chance for at den sjældne hændelse først indtræffer efter forløbet af det dobbelte af den forventede periode.

Ved vurderingen af sjældne hændelsers forekomster kan denne regel være en god hjælp.

Er modellen den rigtige?
Når vi her har belyst sjældne hændelsers forekomst ved anvendelse af kuglemodeller, er der alene tale om hændelser som kan beskrives på passende vis ved modeller af denne art. Det vil sige at hændelsens forekomst bestemmes af tilfældigheder, her ved udtagelse af kugler, og ved hver enkelt udtagelse er der samme chance for at den pågældende hændelse indtræffer. Der er altså ikke tale om situationer hvor chancen for at den betragtede hændelse indtræffer, ændrer sig fra udtagelse til udtagelse.

Det kan ikke altid umiddelbart afgøres om hverdagens sjældne hændelser kan beskrives ved den slags kuglemodeller. Ved ulykker der forårsages af personers træthed eller af materialers nedslidthed, kan det vel tænkes at risikoen for ulykkens forekomst bliver større og større for hver tidsenhed der forløber.

På den anden side kan der også være tale om sjældne hændelser hvor personer får mere og mere erfaring med at håndtere den forelagte situation. Det kan bevirke at hændelsens sandsynlighed bliver mindre og mindre jo længere tid der går.

Der må derfor som altid hvor der gøres brug af matematiske modeller, foretages overvejelser over modellens rimelighed før den anvendes på den praktiske situation.

Teorien bag 40-15 reglen
Sandsynligheden for at hændelsen indtræffer ved udførelsen af et forsøg er p. Sandsynligheden for at hændelsen *ikke* forekommer i de første x forsøg er da: $G(x) = (1-p)^x$. Og heraf har vi at sandsynligheden F(x) for at hændelsen indtræffer mindst én gang i de første x forsøg er givet ved:

$$F(x) = 1 - G(X) = 1 - (1-p)^x$$

For p= 1/20 får vi:

$F(10) = 0.401$ $\qquad G(40) = 0.129$

For p=1/100 får vi:

$F(50) = 0.395$ $\qquad G(200) = 0.134$

Og for p = 1/1000:

$F(500) = 0.394$ $\qquad G(2000) = 0.135$

Det er disse tal der ligger til grund for 40 -15 reglen. Det ses at procenterne 40 og 15 er at betragte som omtrentlige værdier. 40 -15 reglen er således en "tommelfingerregel", men det gør den ikke mindre nyttig. Strengt taget burde reglen nok hedde 40 -14 reglen.

50% chance for den sjældne hændelse
Hvor mange forsøg skal der mon udføres for at den sjældne hændelse har en sandsynlighed på 50% for at indtræffe?

Af computersimuleringen ovenfor så vi at der for 14 forsøg var en sandsynlighed på 50% for at hændelsen indtraf. De 14 svarer til 70% af den teoretiske periode på 20 år.

Det kan vises at dette resultat har almen gyldighed, idet der for værdier af p som er 1/100 eller mindre, gælder at den sjældne hændelse i teorien vil forekomme inden for 69% af den teoretiske periode. For p=1/n har vi altså for antallet af forsøg for hændelsens indtræffen:

Gennemsnit: n   Median: ca. 0.69·n

Og fra 40 -15 reglen har vi yderligere:

Chancen for

Indtræffen inden for de første n/2 forsøg: ca. 40%
Ikke-indtræffen inden for de første 2n forsøg: ca. 15%

Historisk
I "The Doctrine of Chances" (se omtalen i afsnit 1) finder vi følgende opgave: Hvor mange kast skal der udføres med to terninger for at der er chancelighed for at få to seksere?

Med andre ord: Hvor mange kast skal der udføres for at chancen for at få to seksere er 50%. Lærebogen bruger her 69%-reglen (som er udledt i bogen) og konkluderer at der skal udføres 25 (= 69% af 36) kast for at chancen for to seksere er mindst 50%.

Tilsvarende beregnes i bogen hvor mange kast der skal udføres med tre terninger for at der er chancelighed for at få tre seksere. Svaret er her: 69% af 216 = 150 kast.

To gange indtræffer den sjældne hændelse
Bogen udleder også en regel for hvor mange eksperimenter der skal udføres for at der er 50% chance for at den sjældne hændelse indtræffer to gange. Her er svaret at der skal udføres 1.69·n kast, hvor n er den forventede periode. Hvis hændelsen har en sandsynlighed på 1/100 for at forekomme, skal der altså udføres 169 eksperimenter for at chancen er 50% for at hændelsen er forekommet to gange.

6. Kommer ulykker ofte tre ad gangen?

Hvis der er tale om tre seksere i et kast med tre terninger, så skal der i gennemsnit udføres 1.69·216 = 365 kast for to gange at få tre seksere.

Her er en computerkørsel med programmet KugleX af 100 kasteserier indtil der foreligger tre seksere for anden gang.

Talområde	Fraktil
22..22	1 %
22..54	5 %
22..96	10 %
22..143	25 %
22..332	50 %
22..485	75 %
22..754	90 %
22..854	95 %
22..1018	99 %

```
Gennemsnit:    370.98
Maximum   :   1058
Minimum   :     22
```

Igen er der tale om store variationer: Fra 22 kast til 1058 kast. Gennemsnittet ligger nær ved det forventede på 365 kast.

7. En spiller går fallit

Situationen: A og B spiller mod hinanden i en serie af spil. A har en startkapital på a spillemærker, og B har en startkapital på b spillemærker. I hvert spil modtager vinderen 1 spillemærke fra taberen. A har i hvert spil en sandsynlighed på p for vinde, og B har en sandsynlighed på q for at vinde (p+q=1). Uafgjorte spil kan ikke forekomme. Spillene fortsætter indtil en af de to spillere har mistet alle sine spillemærker.

Problem: Hvad er sandsynligheden for at spillene slutter med at A går fallit? Hvad er sandsynligheden for at spillene slutter med at B går fallit? Hvor mange spil kræves der i gennemsnit for at bringe spillene til afslutning?

Vi ser nu på et meget enkelt eksempel. Vi sætter a=1 og b=2. A har altså en startkapital på 1 og B har en startkapital på 2. Vi vil se på sandsynligheden for at A går fallit.

Hvis B vinder 1. spil og A altså taber 1. spil, mister A sin kapital, og han går fallit. Hvis A vinder første spil, kan han ikke gå fallit i 2. spil, han har jo efter 1. spil en kapital på 2.

Vi kan hurtigt indse at A kun kan gå fallit efter et ulige antal spil. Her er de forløb der fører til A's fallit efter 1 spil, 3 spil, 5 spil, 7 spil og 9 spil (A betyder at A vinder spillet, B betyder at B vinder spillet):

B;   ABB; ABABB; ABABABB; ABABABABB;

Sandsynligheden P(A) for at A går fallit kan nu beregnes.

$$P(A) = q + q^2p + q^3p^2 + q^4p^3 + ...$$

Her er tale om en uendelig kvotientrække. For summen af den får vi:

7. En spiller går fallit

$$P(A) = \frac{q}{1 - qp}$$

Spiller B kan kun gå fallit efter et lige antal spil: Her er de forløb som fører til B's fallit efter 2 spil, 4 spil, 6 spil, 8 spil og 10 spil:

AA; ABAA; ABABAA; ABABABAA; ABABABABAA;

Sandsynligheden P(B) for at B går fallit er da:

$$P(B) = p^2 + p^3q + p^4q^2 + p^5q^3 + \dots$$

Denne sum kan omskrives til:

$$P(B) = \frac{p^2}{1 - pq}$$

Vi har nu:

$$P(A) + P(B) = \frac{q + p^2}{1 - pq} = \frac{q + p(1 - q)}{1 - pq} = \frac{q + p - pq}{1 - pq} = \frac{1 - pq}{1 - pq} = 1$$

Sandsynligheden er altså 1 for at en af spillerne går fallit. Det betyder at sandsynligheden for at spillene fortsætter i det uendelige er 0. Men denne hændelse er ikke umulig selv om den har en sandsynlighed på 0: Ved spilleforløb med stadige skift mellem A og B, ABABABABAB..., kommer spillene aldrig til en afslutning. Denne hændelse med sandsynligheden 0 er altså ikke teoretisk umulig, men den kan siges at være praktisk umulig.

Spillets længde

Vi vil nu se på spillets forventede længde, altså gennemsnittet L af det antal spil der må udføres for at bringe spillene til afslutning.

7. En spiller går fallit

Med brug af de tidligere anvendte sandsynligheder har vi for L:

$$L = (1\cdot q + 3\cdot q^2p + 5\cdot q^3p^2 + \ldots\ldots) + (2\cdot p^2 + 4\cdot p^3q + 6\cdot p^4q^2 + \ldots\ldots)$$

hvor den første parentes indeholder bidragene fra de ulige antal spil, og den anden parentes indeholder bidragene fra de lige antal spil. De to summer kan omregnes til:

$$L = \frac{q + pq^2}{(1 - pq)^2} + \frac{2p^2}{1 - pq)^2} = \frac{1 + q}{1 - pq}$$

Hvis p og q begge er 1/2, får vi følgende resultater:

$$P(A) = 2/3 \qquad P(B) = 1/3 \qquad L = 2$$

Den gennemsnitlige spillængde er altså 2, og A's sandsynlighed for at gå fallit er dobbelt så stor som B's.

Vi skal nu uden at gå ind på de matematiske detaljer fremsætte de generelle resultater som gælder for vilkårlige værdier af a og b. Her bliver det nødvendigt at skelne mellem to situationer:

(1) p og q er forskellige (2) p og q er begge 1/2

Vi har i disse situationer:

$$P(A) = \frac{1 - (p/q)^b}{1 - (p/q)^{a+b}} \quad \text{for } p \neq q \qquad P(A) = \frac{b}{a + b} \quad \text{for } p = q$$

$$P(B) = \frac{1 - (q/p)^a}{1 - (q/p)^{a+b}} \quad \text{for } p \neq q \qquad P(B) = \frac{a}{a + b} \quad \text{for } p=q$$

For spillets længde gælder:

7. En spiller går fallit

$$L = \frac{a}{q-p} - \frac{a+b}{q-p} \cdot \frac{1-(q/p)^a}{1-(q/p)^{a+b}} \quad \text{for } p \neq q \qquad L = a \cdot b \quad \text{for } p=q$$

Eksempler

For p = 2/3, q = 1/3, a = 3, b = 2 får vi:

P(A) = 3/31 = 0.097   P(B) = 28/31 = 0.903   L = 4.548

Den gennemsnitlige spillængde er altså ca. 4.5 spil.

For p = q = 1/2 og a = 1, b = 99 får vi:

P(A) = 99/100 = 0.99   P(B) = 1/100 = 0.01   L = 1·99 = 99

Den gennemsnitlige længde af spilleserien er altså 99 spil, selv om A kun har en startkapital på 1.

Formlerne kan illustrere hvor meget en forskel mellem p og q betyder i forhold til en forskel mellem a og b. Til eksempel kan undersøges om det er at foretrække at have en fordel i startkapital frem for en fordel i sandsynlighed for gevinst i de enkelte spil. Er fx a=10 og b=50 samt p=2/3 og q=1/3, så er A's chance for at vinde spillet mere end 99.9%. Det er altså, i hvert fald i dette spil, bedre at være dygtig end at være rig.

Indsatsen forøges
Vi kan også af formlerne se hvad en forøgelse i indsatsen i det enkelte spil betyder for chancerne i spillet. Vi ser på fire situationer. I alle fire situationer er p=0.4 og q=0.6:

(1) a= 90, b=10.   Her får vi: P(A) = 0.983

(2) a= 45, b=5.   Her får vi: P(A) = 0.868

(3) a = 18, b=2. Her får vi: P(A) = 0.556

(4) a = 9, b=1. Her får vi: P(A) = 0.325

Ændringen fra (1) til (2) kan vi beskrive sådan: I (1) er indsatsen 1 i hvert spil, i (2) har vi opdelt startkapitalen i portioner med 2 enheder, så vi nu har en indsats på 2 pr. spil. I (3) er indsatsen i hvert spil 5, og i (4) er den 10. Vi kan af oversigten se hvad det betyder for A's risiko for at gå fallit: Den går fra over 98% i situation (1) ned til ca. 32% i situation (4). Heraf kan vi se at for den spiller der har mindst vinderchance (her 0.4), er det en fordel at indsatsen gøres større, dvs. at startkapitalen opdeles i færre portioner. – Ved rouletten kan det derfor være god strategi at opdele startkapitalen i store portioner og spille "dristigt" i stedet for at trække pinen ud med mange små indsatser.

Slumptur

En anden formulering der ofte kommer på tale ved "En spiller går fallit"-problemer er denne: En partikel befinder sig på x-aksen i punktet (a,0) hvor a er det helt positivt tal. Partiklen går nu ud på en tur der er styret af chancer, en såkaldt *slumptur*: Der foretages en række skridt af længden 1. Der er en sandsynlighed på p for at partiklen går et skridt mod højre, altså til punket (a+1,0), og der en sandsynlighed på q for at partiklen går et skridt mod venstre til punket (a -1,0). Der gælder som før: p+q=1.

På x-aksen er opstillet to absorberende barrierer: En ved x=0 og en ved x=k, hvor k er et helt positivt tal større end a. Når partiklen rammer en af de to barrierer fastholdes den, og slumpturen er afsluttet.

Ved hjælp af de formler der er opstillet for P(A), P(B) og L, kan nu beregnes sandsynligheden for at slumpturen afsluttes ved at partiklen absorberes ved x=0 eller ved x=k, lige som den gennemsnitlige turlængde kan beregnes af formlen for L.

Symmetriske slumpture i 1, 2 og 3 dimensioner
Her skal vi kort omtale en anden form for slumpture. Det er ture uden absorberende barrierer og med samme sandsynlighed for de forskellige skridtmuligheder. Ved en sådan slumptur på x-aksen er derfor p = q = 1/2.

Partiklen starter i punktet x=0 og fortsætter derefter sin vandring på x-aksen i det uendelige. Spørgsmål: Hvad er chancen for at partiklen vender tilbage til udgangspunktet x=0? Og hvor tit vil partiklen vende tilbage?

Svaret er at sandsynligheden for at partiklen vender tilbage til x = 0 er 1. Og i øvrigt er sandsynligheden også 1 for at partiklen undervejs på slumpturen besøger punket (d,0) hvor d er et vilkårligt helt tal.

Der gælder endvidere at ikke blot vil partiklen komme på besøg i ethvert heltalligt punkt på x-aksen, den vil også komme på besøg uendeligt ofte.

Partiklens gennemsnitlige afstand fra udgangspunktet efter n skridt er ca. $0.8 \cdot \sqrt{n}$. Efter 100 skridt er partiklens gennemsnitlige afstand altså ca. 8.

Selv om sandsynligheden er 1 for at partiklen vender tilbage til udgangspunktet, findes der uendeligt mange slumpture som ikke fører partiklen tilbage til x = 0. Til eksempel vil en tur der hele tiden består af to skridt mod højre efterfulgt af et skridt mod venstre, ikke føre partiklen tilbage til 0. Men beregningerne viser at den slags ture har en sandsynlighed på 0.

I to dimensioner starter partiklen i punktet (0,0) og har i hvert skridt 4 muligheder: Et skridt mod højre, et skridt mod venstre, et skridt opad, et skridt nedad. De 4 muligheder er lige sandsynlige. I hvert skridt skal partiklens koordinater (x,y) altså ændres til en af følgende muligheder. x+1, x -1, y+1, y -1.

Hvordan ligger et mon nu med partiklens sandsynlighed for at vende tilbage til udgangspunktet?

Her gælder helt det samme som før ved slumpture på x-aksen: Sandsynligheden er 1 for at partiklen vender tilbage, og den er også 1 for at partiklen besøger et vilkårligt punkt (x,y) i koordinatsystemet. Og lige som i én dimension vil partiklen besøge ethvert punkt uendeligt ofte. – Sender man to partikler på individuelle slumpture fra hver sit udgangspunkt (som ligger et lige antal skridt fra hinanden), så vil de to partikler mødes uendeligt ofte.

Nu bliver partiklen væk
I tre dimensioner har partiklen i hvert skridt 6 muligheder som er lige sandsynlige. Men nu er svaret på de stillede spørgsmål ikke som før. I tre dimensioner har partiklen fået så meget frihed i sin slumptur at det ikke er sikkert at den vender tilbage til udgangspunktet. Det kan vises at sandsynligheden for en tilbagevenden er ca. 34%. Og det gennemsnitlige antal besøg i udgangspunktet er nu blot ca. 0.52.

*

Et gammelt problem
Historisk går problemet med en spillers fallit tilbage til 1600-tallet. En opgave af denne art blev formuleret af Fermat (1601-65) i 1650'erne. To spillere A og B har hver 12 dukater. De deltager i et spil hvor der udføres terningkast med tre terninger. A får 1 dukat fra B hvis øjentalsummen er 11, B får en dukat fra A hvis terningerne tilsammen viser 14. Kast der ikke giver 11 eller 14 som øjentalsum, tæller ikke med. Hvad er A's og B's sandsynlighed for at vinde spillet?

Her foreligger en situation med "En spiller går fallit". Sandsynligheden for at opnå en øjentalsum på 11 ved et kast med tre terninger er 27/216, og sandsynligheden for at opnå en øjentalsum på 14 er 15/216. De to spilleres vinderchancer i de enkelte spil forholder sig altså som 27:15. Da kun kast der giver øjentallene 11 og 14 tæller med, er A's chance i det enkelte kast 27/42 = 9/14 og B's chance er 5/14.

Der er altså tale om en spilsituation hvor p=9/14, q=5/14, a=12, b=12. Ved anvendelse af de opstillede formler kan vi beregne at P(B) > 0.999. Der er altså næsten sikkerhed for at spillet slutter med at B går fallit. For den forventede spillængde får vi: L = 42. Her er kun medregnet de kast der giver enten 11 eller 14 som øjentalsum.

Havde spillerne udført møntkast med lige sandsynligheder for de to spillere, ville den forventede længde af spillet have været 12·12 = 144 spil. Men på grund af den skæve fordeling mellem p og q reduceres den gennemsnitlige spillængde altså til 42.

Her er en computersimulering af 10 000 spil:

```
Slumptur

 A's vinderchance i enkeltkast: 9/14
 A's startkapital: 12
 B's vinderchance i enkeltkast: 5/14
 B's startkapital: 12

 Antal spil: 10000.
 Der spilles i hvert spil indtil A eller B går
 fallit.

 A vandt 9991 spil;  B vandt 9 spil

      Gennemsnit spillængde       :    42.15
      Maximum spillængde          :   214
      Minimum spillængde          :    12
```

8. Fuldt sæt

Problemet: En terning kastes indtil alle seks mulige øjental er forekommet. Hvor mange kast skal der i gennemsnit udføres før et fuldt sæt af øjental foreligger?

Vi lader INFA-programmet Kugle3 simulere 1000 udførelser af eksperimentet. Her er en oversigt:

```
Kugle-Model K3(6,6,ja)

I æsken findes kugler nr. 1..6
          Røde kugler nr. 1..6

Et eksperiment består i at udtage kugler
indtil alle røde kugler er udtaget

Samme kugle kan udtages flere gange.


Antal udtagne   Antal             Antal
      kugler    eksperimenter      i %

     6-8           105            10.50
     9-11          227            22.70
    12-14          242            24.20
    15-17          174            17.40
    18-20          104            10.40

    21-23           63             6.30
    24-26           35             3.50
    27-29           19             1.90
    30-32           16             1.60
    33-36            5             0.50

    37-39            3             0.30
    40-42            4             0.40
    46-52            3             0.30

Gennemsnit: Antal kugler pr. eksperiment:   14.77
```

Oversigten viser at der i gennemsnit måtte anvendes 14.77 kast for at opnå alle seks øjental. Men oversigten viser også at der er stor forskel på længen af kasteserierne. Af programmets optælling får vi: Den korteste serie har en længde på 6 kast, og den længste består af 52 kast. Vi ser også af oversigten at vi i 50% af kasteserierne kan nøjes med 14 kast før et fuldt sæt af øjental foreligger. Median og middeltal ligger altså tæt på hinanden.

Beregning af gennemsnittet

Vi vil nu se hvordan beregningen af den teoretiske værdi af fordelingens middeltal kan udføres.

Vi gør her brug af følgende: Hvis en hændelse har en sandsynlighed på p for at forekomme, så skal der i gennemsnit udføres 1/p forsøg for at få hændelsen til at forekomme. Hvis sandsynligheden for en sekser i et terningkast er 1/6, så skal der i gennemsnit udføres 6 kast for at fremkalde en sekser. Selvfølgelig kan sekseren opnås allerede i første kast, men det kan også forekomme at der må udføres 30 kast eller flere før sekseren forekommer. I det lange løb vil en sekser i gennemsnit kræve 6 kast.

Vi vil nu se på hvor mange kast der i gennemsnit kræves for at opnå et fuldt sæt af øjental.

I første kast opnår vi med sikkerhed et af de seks øjental. Derefter skal vi beregne hvor mange kast der er nødvendige for at få et *nyt* øjental. Chancen for et nyt øjental i kast nr. 2 er 5/6. Der er jo fem øjental vi endnu ikke har set.

Derefter er chancen for et nyt øjental reduceret til 4/6. Der er jo 4 øjental som endnu ikke er forekommet.

Således fortsætter vi beregningerne. Her er en oversigt:

1. øjental. Chance: 1. Kast i gennemsnit: 1 = 6/6
2. øjental. Chance: 5/6 Kast i gennemsnit: 6/5
3. øjental. Chance: 4/6 Kast i gennemsnit: 6/4
4. øjental. Chance: 3/6 Kast i gennemsnit: 6/3
5. øjental. Chance: 2/6 Kast i gennemsnit:: 6/2
6. øjental. Chance: 1/6 Kast i gennemsnit: 6/1

Det samlede antal kast i gennemsnit for at få alle øjental:

$$6/6 + 6/5 + 6/4 + 6/3 + 6/2 + 6/1 = 14.68$$

Vi ser at den teoretiske værdi stemmer fint med det gennemsnit vi opnåede ved simuleringen ovenfor.

Beregningen af fordelingens varians
For terningkast-eksemplet har vi beregnet gennemsnittet af det nødvendige antal kast ved hjælp af udtrykket:

$$\text{Gennemsnit} = 6\cdot(1+1/2 + 1/3 + 1/4 + 1/5 + 1/6) \approx 14.7$$

For variansen kan opstilles følgende udtryk:

$$\text{Varians} = 6\cdot(1/5^2 + 2/4^2 + 3/3^2 + 4/2^2 + 5/1^2) \approx 39.0$$

Den almene model
På tilsvarende måde kan beregningen gennemføres for andre situationer. De kan alle beskrives ved en kuglemodel: I en æske findes n kugler. Fra æsken udtages kugler en ad gangen og med tilbagelægning efter hver udtagelse. Hvor mange kugler skal der i gennemsnit udtages før der foreligger et fuldt sæt af kugler?

Nogle eksempler
n=10. Gennemsnit: 29.3
n=12. Gennemsnit: 37.2

n=30.	Gennemsnit:	120
n=40 .	Gennemsnit:	171
n=50.	Gennemsnit:	225
n=100.	Gennemsnit:	519
n=365.	Gennemsnit:	2365

Andre resultater

En tilnærmet beregning

Gennemsnittet af kugler der skal udtages for at opnå et fuldt sæt kan med tilnærmelse beregnes af:

$$\text{Gennemsnit} \approx n \cdot \ln(n + \tfrac{1}{2}) + n \cdot 0.5772$$

For n=10 giver formlen: 29.3   For n = 100 : 518.7

Hvor mange forskellige kugler efter k forsøg

Det gennemsnitlige antal forskellige kugler der er opnået efter k forsøg, kan beregnes ved:

$$n \cdot (1 - (1 - 1/n)^k)$$

For n=6 og k=6 giver beregningen: 4. Med andre ord: I et kast med 6 terninger vil vi i gennemsnit opnå 4 forskellige øjental.

Her er en simulering med KugleX af 1000 kast med seks terninger:

```
  Antal forsk.    Antal
      øjental     eksperimenter

          2          24
          3         225
          4         521
          5         215
          6          15

Antal forskellige øjental i gennemsnit: 3.97
```

9. Den bedste strategi

I en pose findes n sedler, og på hver seddel er der skrevet et tal. Vi antager at alle tal er forskellige. Fra posen udtages sedler én for én og uden tilbagelægning. Spilmasteren læser det udtrukne tal op, og spilleren skal afgøre om han vil stoppe eller gå videre. Hvis han stopper, får han alle sedler at se, og hvis han er stoppet ved det største af alle n tal, vinder han den udsatte gevinst. Hvis han ikke stopper ved det største tal, har han mistet muligheden for gevinst. – Det skal tilføjes at spilleren ikke får nogen oplysning om hvilke tal der findes på sedlerne. Han ved altså ikke hvor store tallene på sedlerne kan være.

Problemet: Hvilken strategi skal spilleren bruge for at hans chance for gevinst bliver så stor som mulig? Og hvor stor en chance vil han kunne opnå?

Vi ser på nogle enkle situationer inden vi tager fat på den generelle situation. Lad os antage at der blot er to sedler i posen. Spilleren har da to mulige strategier: (1) Vælg den første seddel (2) Vælg den anden seddel. I begge tilfælde vil hans chance for at få det største tal være 1/2. Der er jo lige stor chance for at det største tal er på seddel 1 og på seddel 2.

Vi ser herefter på situationen med 3 sedler. Vi betegner tallene på dem for A, B og C, hvor de er angivet i størrelsesorden med A som det mindste og C som det største. De tre tal kan blive udtrukket i 3! = 6 rækkefølger:

ABC
ACB
BAC
BCA
CAB
CBA

9. Den bedste strategi

Vi kan nu se hvordan det går med en banal strategi: "Vælg altid det første tal". I de seks rækkefølger ovenfor vil vi vinde gevinsten i de to sidste tilfælde, da C jo er det største af de tre tal. Denne strategi giver altså en gevinstchance på 2/6 = 1/3.

Vi kan se at de to andre strategier: "Vælg altid det andet tal" og "Vælg altid det tredje tal" også fører til gevinst i 2 ud af de 6 situationer. Hver af de tre strategier har altså en gevinstchance på 1/3.

Men her er endnu en strategi: "Noter dig det første tal, og stop derefter ved det første tal som er større end det noterede". Denne strategi ville føre til valg af tallet C og dermed til gevinst i følgende af de seks rækkefølger:

ACB, BAC og BCA.

Gevinstchancen vil altså her være 1/2.

På samme måde kan vi undersøge situationen med 4 sedler med tallene A, B, C og D. De fire tal kan udtrækkes i 4! = 24 rækkefølger. Vi udskriver her de 24 muligheder:

ABCD	BACD	CABD	DABC
ABDC	BADC	CADB	DACB
ACBD	BCAD	CBAD	DBAC
ACDB	BCDA	CBDA	DBCA
ADBC	BDAC	CDAB	DCAB
ADCB	BDCA	CDBA	DCBA

Det ses let at en banal strategi som "Vælg altid det første tal" vil have en gevinstchance på 1/4 = 0.25. Det giver gevinst i de 6 rækkefølger der indledes med D.

Vi ser nu på den strategi som var den bedste i tilfældet før med tre sedler: "Noter dig det første tal, og stop derefter ved det første tal som er større end det noterede".

9. Den bedste strategi

Denne strategi vil føre til gevinst i 11 af de 24 rækkefølger:

ABCD	BACD	**CABD**	DABC
ABDC	**BADC**	**CADB**	DACB
ACBD	BCAD	**CBAD**	DBAC
ACDB	BCDA	**CBDA**	DBCA
ADBC	**BDAC**	**CDAB**	DCAB
ADCB	**BDCA**	**CDBA**	DCBA

Strategien har altså en gevinstchance på 11/24 = 0.458

Men lad os også prøve med en anden strategi: "Noter *dig det største af de to første tal*, og stop derefter ved det første tal som er større end det noterede". Denne strategi fører til gevinst i 10 af de 24 rækkefølger:

ABCD	BACD	**CABD**	DABC
ABDC	**BADC**	**CADB**	DACB
ACBD	**BCAD**	**CBAD**	DBAC
ACDB	**BCDA**	**CBDA**	DBCA
ADBC	BDAC	CDAB	DCAB
ADCB	BDCA	CDBA	DCBA

Strategien har altså en gevinstchance på 10/24 = 0.417.

Der er således ikke stor forskel på de to strategiers vinderchancer, men den der nøjes med at tage bestik efter det første tal er lidt bedre end den der venter til de to første tal foreligger.

Ved et større antal sedler vil vi gøre brug af erfaringerne fra 3 og 4 sedler. Vi bygger strategien op efter følgende model: "Noter dig det største af de x første tal, og stop derefter ved det første tal som er større end det noterede". Ved både 3 og 4 sedler havde x værdien 1. Men vi må nok imødese at x vil have en større værdi når antallet af sedler forøges.

Det gælder altså om ved et forelagt antal af sedler, **n**, at finde den værdi **x** som giver en strategi med størst gevinstchance.

9. Den bedste strategi

Vi ser nu på n=10. Vi prøver her med værdien 4 for x. Vi noterer altså det største tal der forekommer på de fire første sedler, og derefter stopper vi ved det første tal som er større end det noterede.

Hvornår vil denne strategi føre til gevinst? Det vil den hvis følgende to betingelser er opfyldt:

(1) Det største af alle 10 tal findes på seddel nr. k, hvor k>4.

(2) Det største tal på sedlerne før seddel nr. k findes på en af de fire første sedler.

Når de to betingelser opfyldt, og kun da, vil strategien føre til at det største tal blandt de 10 vælges.

Vi ser nu på hvad sandsynligheden er for at de to betingelser er opfyldt for en bestemt værdi af k.

I alt kan de 10 tal opstilles i 10! forskellige rækkefølger. Når vi fastholder det største tal på seddel nr. k, så kan de øvrige tal opstilles i 9! forskellige rækkefølger. Endvidere skal gælde betingelse (2), som har sandsynligheden 4/(k-1). Tallet i betingelse (2) kan jo placeres på 4 ud af de sedler som går forud for den med det største tal.

Alt i alt er sandsynligheden for at de to betingelser er opfyldt:

$$\frac{x}{k-1}\cdot\frac{9!}{10!} = \frac{x}{10}\cdot\frac{1}{k-1} = \frac{4}{10}\cdot\frac{1}{k-1}$$

Dette udtryk er for en fast position af det største tal, nemlig på seddel nr. k. Vi må derfor beregne den samlede sandsynlighed når k antager de mulige værdier, nemlig 5, 6, 7, 8, 9, 10.

Vi får her for den søgte sandsynlighed P(4,10):

9. Den bedste strategi

P(n=10, x=4) =

$$\frac{4}{10} \cdot (1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9) = 0.398$$

Vi kan nu let undersøge strategien der består i at sætte x til 5:

$$P(n=10, x=5) = \frac{5}{10} \cdot (1/5 + 1/6 + 1/7 + 1/8 + 1/9) = 0.373$$

Strategien med at udtage 5 tal inden der noteres et tal, er altså dårligere end strategien med at notere efter de første fire tal.

Vi undersøger nu den strategi der består i at notere det største tal blandt de tre første. For gevinstchancen får vi her:

P(n=10, x=3) =

$$\frac{3}{10} \cdot (1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9) = 0.399$$

Denne strategi er altså en anelse bedre end strategien med x=4.

Alt i alt kan det godtgøres at strategien med x=3 er den bedst mulige. Vi skal altså notere os det største tal blandt tallene på de tre første sedler, og derefter stoppe ved det første tal som er større end det noterede.

Her er en tabel over 10 eksperimenter med 10 sedler.

1:	50	51	30	**67**	11	45
	36	17	46	38		
2:	56	21	35	55	29	22
	36	12	**60**	50		
3:	35	56	74	71	45	68

	59	58	54	67		
4:	62	68	41	31	69	73
	64	14	34	45		
5:	45	10	31	**57**	32	29
	30	25	40	26		
6:	59	12	72	28	22	42
	16	31	23	58		
7:	21	13	68	10	20	18
	41	23	27	11		
8:	10	11	45	**74**	70	13
	64	65	14	15		
9:	47	68	21	65	34	17
	45	44	14	36		
10:	21	71	48	46	44	29
	32	59	40	25		

Hvis vi anvender strategien med x=3, så vil den føre til gevinst i følgende af de 10 eksempler: Nr. 1, 2, 5 og 8. Altså i 4 ud af de 10 eksempler.

Den almene situation

Det kan vises at for et vilkårligt antal n af sedler så fås den optimale strategi for en x-værdi som opfylder følgende to krav:

(1) $1/(x+1) + 1/(x+2) + ... + 1/(n-1) < 1$

(2) $1/x + 1/(x+1) + ... + 1/(n-1) > 1$

Og den tilhørende gevinstchance er givet ved:

$$P(n, x) = \frac{x}{n} \cdot \left(1/x + 1/(x+1) + 1/(x+2) + ...+ 1/(n-1)\right)$$

For n = 10 er de to betingelser opfyldt for x = 3. Og vi får som tidligere beregnet.

$$P(n=10, x= 3) = 0.399$$

Her er en tabel over optimale strategier og deres gevinstsandsynligheder:

n	x	P(n,x)
3	1	0.500
4	1	0.458
5	2	0.433
6	2	0.428
10	3	0.399
20	7	0.384
50	18	0.374
100	37	0.371

I situationen med 100 sedler er den optimale strategi altså at notere det største tal på de første 37 sedler, og derefter stoppe ved det første tal som er større end det noterede. Det giver en strategi med gevinstsandsynligheden 37%.

For store værdier af n gælder at der for den optimale strategi skal benyttes en x-værdi som er ca. 0.368 ·n.

Gevinstsandsynligheden kommer samtidig nær til 0.368. (Tallet 0.368 er identisk med 1/e, hvor e er grundtallet for de naturlige logaritmer).

9. Den bedste strategi

Bemærk at for store værdier af n, fx n >20, gælder en ganske enkel regel for valg af strategi: "Noter det største tal på de første 37% af sedlerne, og stop derefter ved det første tal som er større end det noterede". Det giver uanset antallet af sedler en gevinstchance på over 1/3.

Her er resultaterne fra 10 eksperimenter. Der udtages 100 tal, og det største af de første 37 tal findes. Det betegnes Y. Derefter findes Z, stoptallet, det første tal efter de 37 tal som er større end Y. Z er strategiens forslag til det største tal blandt de 100 tal. Hvis Z er lig med Max, har strategien ført til det rigtige resultat.

```
Talområde:                1..999
 Antal tal, der udtages: 100
 Inspektionsgrænsen:       37
 Antal eksperimenter:      10
```

Eksperiment nr.	Y	Z	Max
1	999	979	999
2	976	965	976
3	971	995	996
4	963	**979**	979
5	997	987	997
6	988	**995**	995
7	990	**999**	999
8	997	983	997
9	982	**989**	989
10	995	992	995

Det ses at strategien fører til det rette resultat i eksperimenterne nr. 4, 6, 7 og 9. Altså i fire af de ti eksperimenter.

10. En stok brækkes

Problemet: En stok af længden 100 brækkes ved hjælp af to tilfældigt valgte punkter i tre dele. Hvad er chancen for at de tre dele kan sammensættes til en trekant?

Lad os betegne de to valgte punkter ved X og Y:

```
|--------------------X-------------------------Y---------------------|
0                                                                  100
```

Afstanden fra nulpunktet til X kaldes x, og afstanden fra nulpunktet til Y kaldes y.

Stokken brækkes altså i tre dele hvis længder er: x, $y–x$ og $100 – y$.

Hvis de tre dele skal kunne udgøre siderne i en trekant, så skal alle tre dele have en længde på under halvdelen af stokkens længde, altså på under 50. I en trekant er summen af de to siders længder jo altid større end længden af den tredje side. Ingen af siderne kan derfor være på 50 eller derover.

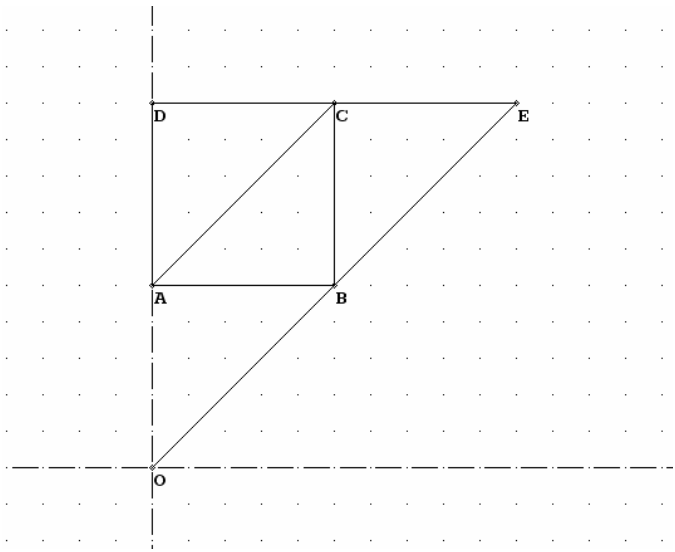
Vi får således følgende betingelser udtrykt ved x og y:

(1) $x < 50$ (2) $y–x < 50$ (3) $100–y < 50$

Vi indtegner nu i et koordinatsystem de tre linjer:

(L1) $x = 50$ (L2) $y = x +50$ (L3) $y = 50$

10. En stok brækkes



Linje L1 er fastlagt ved linjestykket BC, linje L2 ved linjestykket AC og linje L3 ved linjestykket AB. De tre betingelser (1), (2) og (3) er opfyldt for punkter beliggende inde i trekant ABC, og de er ikke opfyldt for andre punkter.

Det område der kan komme på tale ved valget af X og Y, består af de punkter der ligger i trekant ODE. Denne trekant er jo begrænset af betingelserne:

$$0<x<100, \quad 0<y<100 \quad \text{og} \quad y>x$$

Og inden for dette område er det altså kun de punkter (X,Y) som ligger inde i trekant ABC som giver mulighed for at samle de tre stokdele til en trekant.

Vi vil nu give en værdi for chancen for at de tre dele kan samles til en trekant. Her benytter vi arealer på figuren (”geometrisk sandsynlighed”). Hvis der skal kunne dannes en trekant, skal det valgte punkt (X,Y) falde i trekant ABC. Denne trekant har et areal der er lig med en fjerdedel af arealet af det totale område, trekant ODE. Vi sætter derfor den søgte chance til: ¼.

Vi har ikke i detaljer gjort rede for hvordan udvælgelsen af de to punkter X og Y foregår. De to punkter kan fx vælges ved udtrækning af to tal fra intervallet (0,100). Det mindste af de to tal benytter vi til placeringen af X, det største til placeringen af Y.

Vi vælger altså ikke et tilfældigt punkt X på stokken og brækker den ved dette punkt, for derefter at vælge et tilfældigt punkt Y på det største af de to stykker. Dette vil chancemæssigt være en anden situation end den vi har set på. Den vil i øvrigt føre til en chance på ca. 39% for at delene kan samles til en trekant.

Et generelt resultat

Vi antager at stokken har en længde på 1. Sandsynligheden for at de tre stykker som stokken opdeles i alle har en længde mindre end x er givet i følgende tabel:

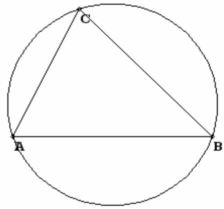
$x \leq 1/3$:	Sandsynlighed: 0
$1/3 < x \leq 1/2$:	Sandsynlighed: $(3x - 1)^2$
$1/2 < x \leq 1$:	Sandsynlighed: $1 - 3(1-x)^2$

For x=1/2 får vi det ovenfor fundne resultat: 1/4

En anden iklædning

På en cirkelperiferi er valgt et fast punkt A. På periferien vælges nu to tilfældige punkter A og B. Herefter tegnes trekant ABC. Hvad er sandsynligheden for at trekanten er en spidsvinklet trekant, altså at alle tre vinkler er mindre end 90° ?

10. En stok brækkes



Trekanten vil være spidsvinklet hvis de tre cirkelbuer AB, BC og CA alle har en længde der er mindre end den halve cirkelperiferi. Vi kan tænke os periferien rullet ud i et linjestykke med A som startpunkt:

A--------------------B---------------------------C------------------------(A)

De to punkter B og C skal da vælges således at hvert af de tre linjestykker AB, BC og C(A) udgør mindre end halvdelen af linjestykket A(A). Men det er jo netop den opgave vi har løst ovenfor. Sandsynligheden for at trekanten bliver spidsvinklet er derfor: 1/4.

Afstanden mellem de to delepunkter
Lad stokkens længde være a. For afstanden y–x mellem de to punkter X og Y gælder da:

Middelværdi: $a/3$ Varians: $a^2/18$

De gennemsnitlige dele af stokken
Om de tre dele stokken opdeles i, gælder i teorien:

Opdelingen vil give dele hvis gennemsnitlige længder forholder sig som 2:5:11

Hvis opdelingen giver en trekant, forholder de gennemsnitlige længder sig som 7:13:16

En anden procedure
Hvis tre linjestykker fastlægges ved at der udvælges tre tilfældige længder i intervallet fra 0 til 100, så er chancen ½ for at de tre linjestykker kan sammensættes til en trekant.

En simulering
Her bringes resultater fra en simulering af chancesituationen. I hvert eksperiment er registreret afstanden mellem de to valgte punkter. Stokkens længde er sat til 1000.

Simuleringen er foretaget med INFA-programmet Lod2.

```
 Statistik over 1000 eksperimenter.
 Variabel: Afstand mellem X og Y

   Talområde            Fraktil
     1..7                 1 %
     1..28                5 %
     1..54               10 %
     1..138              25 %
     1..282              50 %
     1..509              75 %
     1..689              90 %
     1..763              95 %
     1..896              99 %

Gennemsnit:   334.79
Maximum   :   984    Minimum   :     1
```

Det ses at de opnåede data stemmer godt overens med de teoretiske værdier for middelværdien. Af fraktilværdierne fremgår at fordelingens median er 282, altså noget mindre end fordelingens middelværdi.

11. Gennemsnit: uendelig

I sandsynlighedsregningens tidlige periode optrådte overvejelser over sandsynligheder ofte i tilknytning til spil. Man var således optaget af spørgsmålet om en retfærdig pris, en fair pris, for at deltage i et spil. En spiludbyder tilbød et spil mod en betaling af et beløb. Hvad skulle dette beløb være for at der kunne tales om en retfærdig pris for både spiller og spiludbyder?

Fagets udøvere var enige om at man skulle sætte prisen efter den økonomiske forventning en spiller kunne have til spillet. Hvis den gennemsnitlige gevinst ved deltagelse i spillet var et beløb på k, så ville man sætte betaling af et beløb på k som krav for deltagelse i spillet. Hvis man forlangte et større beløb, ville spillet i det lange løb være til fordel for udbyderen. Hvis man satte et beløb mindre end k, ville spillet være til fordel for spilleren.

Det gjaldt altså om at beregne den gennemsnitlige gevinst ved deltagelse i spillet.

Denne metode blev sat på prøve i det såkaldte Sankt Petersborgproblem som blev diskuteret af en række matematikere i 1700-tallet. Metoden førte her til et resultat som var i stærk modstrid med sund fornuft og med spilerfaringer.

I Petersborgproblemet tilbyder spiludbyderen et spil som udføres ved møntkast. Spilleren udfører kast med en mønt indtil han opnår et Kronekast. Så snart et Kronekast foreligger, afbrydes spillet, og spillerens gevinst udbetales.

Hvis han opnår Krone i første kast, er hans gevinst 2. Hvis han opnår Krone i andet kast, er hans gevinst $2^2 = 4$. Hvis han først opnår Krone i kast nr. n, er hans gevinst 2^n. I teorien kan spillet fortsætte i det uendelige.

11. Gennemsnit: uendelig

Vi vil nu se på den gennemsnitlige gevinst i dette spil. Sandsynligheden for at spillet afsluttes efter 1,2,3 ... spil er:

Efter 1 spil: 1/2 Der skal jo opnås Krone i 1. kast.
Efter 2 spil: 1/4 Kasteforløbet skal være: PK
Efter 3 spil: 1/8 Kasteforløbet skal være: PPK.
.
Efter n spil: $1/2^n$ Kasteforløbet skal være: PPPP...K
(n-1 gange P efterfulgt af ét K)

Vi kan nu beregne den gennemsnitlige gevinst:

$G = 2/2 + 4/4 + 8/8 + ...2^n/2^n + ... = 1 + 1 +1 + 1 +.........$

Vi ser at der ikke kan sættes noget tal på den gennemsnitlige gevinst, "G vokser ud over alle grænser", eller med andre ord: G er uendelig stor. Nogle vil sige: G eksisterer ikke.

Konklusionen måtte altså være: Uanset hvor meget spilleren betaler for at deltage i spillet, så er hans betaling ikke stor nok. Vi har jo netop beregnet at den fair pris for deltagelse i spillet er uendelig stor.

Denne pris stemmer ikke med den forventning spillerne har til spillet. Lad os se på nogle udførelser af det. Her er tal fra 100 spil udført med programmet Talmønster:

Antal kast	Antal spil
1	42
2	28
3	14
4	12
5	2
7	1
8	1

Gennemsnit af antal kast: 2.13

11. Gennemsnit: uendelig

Oversigten viser hvor mange kast der blev foretaget i de enkelte spil. Det ses at 1 forekommer mange gange (42), og det største antal kast der forekommer er 8. I spillerens bedste spil vil han altså kaste Plat i syv kast og derefter kaste Krone. Spilleren ville altså i det bedste spil af de 100 få en gevinst på $2^8 = 256$. – Og i 42 spil vil han få en gevinst på 2.

Det kan være svært at overbevise en spiller om at en fair pris for deltagelse i dette spil er uendelig stor. Efter listen over de 100 udførelser af spillet ville en spiller nok sige at en betaling på 4 måtte være det højeste han kunne gå med til. Og medianen i den teoretiske gevinstfordeling er 2. I 50% af spillene kan man jo regne med at få Krone i første kast.

Hvis man vil løse Petersborgproblemet på en sådan måde at resultatet stemmer bedre med ens intuition, så må man fjerne de uendeligt store beløb fra spillet. Uendeligt store beløb er ikke af denne verden, og de teoretiske overvejelser der involverer uendeligt store størrelser kan ikke umiddelbart overføres på virkeligheden.

Man kan derfor modificere spillereglerne så spillængden begrænses til fx 8 kast. Hvis spilleren ikke har opnået Krone i de første 8 kast, så betragtes hans 9. kast som et kronekast. Den gennemsnitlige gevinst i dette spil kan beregnes til G = 10. Selv det beløb vil spilleren dog nok finde for stort.

En mere drastisk begrænsning kunne være at sætte gevinsten ned, fx at ændre udbetalingen fra 2^n til $\sqrt{2^n}$. Derved ville den gennemsnitlige gevinst i spillet blive reduceret til G = 2.41. Så en pris på 2.50 ville måske blive accepteret af spillerne.

Hvis udbetalingen blev ændret fra 2^n til 1.5^n, ville den gennemsnitlige gevinst være: G = 3. Denne pris ville spillerne nok også kunne acceptere.

En uendelig lang ventetid
Et andet eksempel som giver et teoretisk resultat der er i modstrid med ens intuition: Mikkel er ravsamler, og han går hver dag en tur langs stranden for at søge efter rav. Der er gevinst hver dag, somme tider kun en lille klump, til andre tider en større mængde.

Mikkel har fødselsdag i dag, og han stiller sig spørgsmålet: Hvor mange dage vil der mon gå inden jeg får en høst der overstiger den jeg får i dag?

Lad X være ravhøsten på hans fødselsdag, og lad X1, X2, X3,...være ravhøsten på de efterfølgende dage.

Hvis X>X1, så skal Mikkel vente mere end 1 dag før han får en høst der er større end X.

Hvis X>X1 og X>X2, så skal Mikkel vente mere end 2 dage før han får en høst der er større end X.

Og hvis X>X1 og X>X2 og X>X3, så skal Mikkel vente mere end 3 dage før han får en høst der er større end X.

Lad os se nærmere på den sidste betingelse. Der foreligger fire høstresultater: X, X1, X2, X3. De må alle have samme chance for at være den største af de fire (vi antager at der ikke kan forekomme uafgjort mellem X'erne). Chancen for at X er den største høst er da: 1/4.

Hvis vi lader T være ventetiden (målt i dage) indtil der optræder en høst større end X, så har vi:

$$P(T>3) = 1/4$$

På samme måde kan vi se på ravhøsten for de første n dage efter Mikkels fødselsdag. Hvis X er større end hver af de n værdier, så må Mikkel vente mere end n dage før han får en høst der er større end X. Og sandsynligheden for at X er den største værdi er 1/(n+1)

For ventetiden T har vi da:

$$P(T>n) = 1/(n+1)$$

Sandsynligheden for at Mikkel skal vente fx mere end 19 dage på en større høst er altså 1/20.

Af udtrykket for P(T>n) har vi:

$$P(T=n) = P(T>n-1) - P(T>n) = 1/n - 1/(n+1) = 1/(n\cdot(n+1))$$

Altså har vi for ventetiden T:

$$P(T=n) = \frac{1}{n\cdot(n+1)} \quad \text{for } n=1,2,3,...$$

For den gennemsnitlige ventetid får vi da:

$$E(T) = 1\cdot P(T=1) + 2\cdot P(T=2) + 3\cdot P(T=3) + ...$$

Ved indsætning af udtrykket for P(T) får vi:

$$E(T) = 1/2 + 1/3 + 1/4 + \; ....$$

Denne række er bortset fra det manglende første led identisk med den harmoniske række:

$$1/1 + 1/2 + 1/3 + ...$$

som ikke har nogen endelig sum. Med andre ord: Summen er uendelig stor. Det samme gælder da for ventetiden på en høst større end X.

11. Gennemsnit: uendelig

At Mikkel skal igennem uendelig mange rav-dage for at få en høst der er større end X, er i konflikt med vor intuition. Husk at sandsynligheden for at han skal vente mere end 19 dage kun er 1/20.

Igen har det uendelige bevirket at vore fornemmelser for chancer og risikoer kommer på en hård prøve. Det hænger sammen med at vi ønsker at karakterisere chancesituationen ved hjælp den uendelige middelværdi. Lad os se på andre muligheder:

For ventetiden T har vi:

Ventetid højst n dage	Sandsynlighed 1– 1/(n+1)
n=1	1/2 = 0.50
n=2	2/3 = 0.67
n=3	3/4 = 0.75
n=4	4/5 = 0.80
.	
.	
n=10	10/11= 0.91

Heraf ser vi at medianen i fordelingen af ventetider er 1: Der er 50% chance for at ventetiden ikke overstiger 1 dag. Det er jo et helt andet realistisk resultat end den uendelige middelværdi. Vi har her en fordeling med en så sjælden kombination som uendelig middelværdi og medianen 1.

Vi ser også af tabellen at øvre kvartil er 3: Der er 75% chance for at ventetiden ikke overstiger 3 dage.

Vi kunne måske have benyttet 90%-fraktilen i fordelingen af ventetider. Den er 10: Der er 90% chance for at ventetiden ikke overstiger 10 dage.

Konklusion: Er middelværdien uendelig, så benyt en passende fraktilværdi i stedet, brug fx 90%-fraktilen.

Rekorder
Vi kunne også have behandlet Mikkels problem på en anden måde uden at inddrage et uendeligheds-perspektiv. Lad os se på en liste over ti talværdier:

`61   49   `**`82`**`   31   46   `**`88`**`   65   36   60   80`

Listen indledes med 61. Derefter kommer der et større tal i listens tredje tal, 82. Dette tal er *en ny rekord*, det er større end alle de foregående tal. Endnu en ny rekord indtræffer med tallet 88. Derefter er der ikke flere rekorder i listen. Vi ser at denne liste over ti tilfældige tal indeholder to rekorder.

For Mikkel kunne vi formulere hans problem således: Hvor lang er ventetiden på den første rekord, altså det første tal der er større end alle de foregående. Vi nåede frem til at den gennemsnitlige ventetid på den første rekord var uendelig.

Vi vil nu behandle hans problem ved hjælp af overvejelser over rekorder i en talliste. Vi vil se hvor mange rekorder der i gennemsnit kan forventes i en liste med n tal.

Vi betegner listens tal med chancestørrelserne X_1, X_2, X_3,.. Ethvert af tallene efter X_1 kan være en rekord. Hvis X_2 skal være rekord, så skal X_2 være det største af de to første tal. Sandsynligheden herfor er 1/2. De to tal har jo samme chance for at være det største.

Hvis X_3 skal være rekord, skal X_3 være det største af de første tre tal, og sandsynligheden herfor er 1/3. – Og hvis X_j skal være rekord, skal det være det største af de første j tal, og sandsynligheden herfor er 1/j.

Hvis X_j er rekord, giver vi det værdien 1, hvis det ikke er rekord får det værdien 0. Vi har da:

$X_j = 1$ med sandsynlighed 1/j $\qquad$ $X_j = 0$ i øvrigt

For middelværdien $E(X_j)$ har vi da:

$$E(X_j) = 1/j.$$

Denne middelværdi angiver det gennemsnitlige antal rekorder på den j'te plads i tallisten.

Det samlede antal rekorder i hele tallisten får vi nu ved at summere middelværdierne for alle X'er fra X_2 og frem til X_n. Vi betegner det gennemsnitlige antal af rekorder med R_n.

$$R_n = 1/2 + 1/3 + 1/4 + 1/5 + ... + 1/n$$

For n= 10 får vi: $R_{10} = 1.92$.

I gennemsnit må vi altså regne med 2 rekorder i en liste over ti tal. Det kan være til trøst for Mikkel, allerede efter 10 dage kan han have fået en dag med en bedre ravhøst end på startdagen.

For n=20 og n=50 har vi: $R_{20} = 2.60$ og $R_{50} = 3.50$. I en liste med 50 tal vil vi altså kunne vente 3-4 rekorder. Her er en liste fra programmet rekord over 20 tilfældige tal fra området 1..1000:

272	**588**	492	551	27	**882**	764	4	607
168	586	365	157	137	**986**	532	732	155
32	741							

Det ses at listen indeholder tre rekorder.

Også "rigtige data" kan benyttes. Her er listen over nedbør i Danmark i april måned i årene 1990 til 1999:

36 **47** **62** 17 37 40 9 38 **78** 39

Listen indeholder tre rekorder.

11. Gennemsnit: uendelig

Den harmoniske række
Vi har i dette afsnit gjort brug af den harmoniske række:

$$1 + 1/2 + 1/3 + 1/4 + ...$$

Som nævnt har denne række en uendelig sum. At det er tilfældet kan man se af en omskrivning med parenteser:

$$1 + 1/2 + (1/3 + 1/4) + (1/5 + 1/6 + 1/7 + 1/8) + ...$$

Indholdet af den første parentes er større end $2 \cdot 1/4 = 1/2$. Og indholdet af den næste parentes er større end $4 \cdot 1/8 = 1/2$. På denne måde fortsættes med parenteser der indeholder 8, 16, 32,...addender. Hver gang vil indholdet af parentesen overstige 1/2. Den harmoniske række har da en sum der overstiger

$$1 + 1/2 + 1/2 + 1/2 + 1/2 ...$$

Med andre ord: Summen er uendelig.

Men den harmoniske rækkes delsummer vokser meget langsomt. Ved 100 led er summen kun kommet op på lidt over 5:

$$1 + 1/2 + 1/3 + ... + 1/100 = 5.18$$

I en liste med 100 tal kan vi derfor forvente 4-5 rekorder.

I afsnit 8 har vi benyttet en formel for tilnærmet beregning af delsummer i den harmoniske række:

$$1 + 1/2 + 1/3 + ... + 1/n \approx \ln(n + ½) + 0.5772$$

Taleksempler. n=1000: 7.49. n=1 000 000: 14.39

12. Hvordan skal puljen deles?

Problem: I en konkurrence mellem to spillere, A og B, indgår en serie af spil. I hvert spil vil enten A eller B vinde, uafgjort kan ikke forekomme. Det antages at A og B er lige gode til spillet, og at de hver har en sandsynlighed på ½ for at vinde i hvert enkelt spil. Spillets regler er sådan at den spiller der først når et forud aftalt antal af vundne spil, modtager hele spillepuljen.

Konkurrencen bliver imidlertid afbrudt da A stadig mangler a vundne spil for at vinde konkurrencen og B mangler b vundne spil for at vinde konkurrencen. I hvilket forhold bør spillepuljen deles mellem de to spillere?

Dette problem som i sandsynlighedsregningens historie kaldes *delingsproblemet*, går helt tilbage til 1494 hvor det fremsættes i en lærebog af italieneren Lucas Pacioli (1445 - 1517).

Gennem de følgende tre hundrede år blev der af tidens matematikere givet en række forslag til løsninger af delingsproblemet.

Vi ser nu på en situation hvor der kræves 10 sejre for at vinde konkurrencen. Da konkurrencen bliver afbrudt har A 9 sejre, og B har 7 sejre. A mangler altså 1 sejr, og B mangler 3 for at vinde konkurrencen. Her er nogle forslag til puljedelingen fra tiden frem til ca. 1560.

1. Den førende spiller får hele puljen. Dvs. A tildeles den samlede pulje, B modtager intet.
2. Ingen af spillerne modtager noget, konkurrencen er jo ikke fuldført.
3. De to spillere deler puljen i forhold til antallet af vundne spil, dvs. i forholdet 9:7. Det betyder at A modtager 9/16 = 56% af puljen, B modtager 7/16 = 44% af puljen. Denne løsning er fremført af *Pacioli*.

4. *Cardano* (1501-76) fremlagde i 1539 en løsning som gav A 78/133 = 59% og B = 41% af puljen, og *Tartaglia* (1500-57) gav i 1556 en løsning som gav A 60% og B 40% af puljen.

Ingen af de nævnte løsninger byggede på overvejelser over sandsynligheder. Først 100 år senere, fra midten af 1600-tallet, benyttes sandsynlighedsbetragninger ved løsningen af puljedelingsproblemet.

I 1650'erne blev delingsproblemet taget op af tre af tidens store navne inden for matematikken: *Pascal*, *Fermat* og *Huygens*. De fremsatte forskellige metoder til beregningen af puljens deling, men deres metoder byggede på samme ide og førte alle til samme resultat.

Deres ide var at puljen skulle deles i forhold til de to spilleres chance for at sejre i konkurrencen hvis de manglende spil blev gennemført. Lad os se hvad det ville føre til i situationen hvor A mangler 1 sejr og B mangler 3.

Vi indser at yderligere tre spil vil afslutte konkurrencen. Efter tre spil må A enten have vundet det manglende spil eller B må have vundet de tre manglende spil. Der er følgende muligheder for resultaterne i de tre spil:

A vinder alle tre spil:	AAA
A vinder to spil:	AAB, ABA, BAA
A vinder ét spil:	ABB, BAB, BBA
A vinder ingen spil:	BBB

Der er altså 8 mulige forløb af de tre manglende spil. Og da A og B har samme chance for gevinst i hvert spil, så må de 8 muligheder betragtes som lige sandsynlige. Hvert forløb har derfor en sandsynlighed på 1/8.

Af de 8 muligheder vil A vinde konkurrencen i de 7 første i listen ovenfor. Her får A jo den manglende sejr uden at B vinder sine tre manglende sejre. Kun i den sidste situation hvor B vinder alle tre spil, vil B vinde konkurrencen.

Puljen skal derfor efter matematikernes forslag deles i forholdet 7:1. Det betyder at A skal have 7/8 = 87.5% og B skal have 1/8 = 12.5%.

På samme måde kan puljens deling beregnes for andre værdier af a (A's manglende vundne spil) og b (B's manglende vundne spil).

Her er en lille tabel. Tallene i tabellen angiver sandsynligheden for at A, den førende spiller, vinder konkurrencen hvis de manglende spil gennemføres.

a=1
b=2: 3/4 b=3: 7/8 b=4: 15/16 b=5: 31/32 b=6: 63/64

a=2
b=3: 11/16 b=4: 13/16 b=5: 57/64 b=6: 15/16

a=3
b=4: 21/32 b=5: 99/128 b=6: 219/256

a=4
b=5: 163/256 b=6: 191/256

a=5
b=6: 319/512

Det oprindelige problem kan udvides til at omfatte situationer hvor A og B har forskellige chancer p og q for gevinst i de enkelte spil, fx p=3/4 og q=1/4.

Også situationer med flere end to spillere er behandlet i faglitteraturen.

12. Hvordan skal puljen deles?

Historisk
Historiens første lærebog i sandsynlighedsregning er Huygens' afhandling på hollandsk fra 1656. Den foreligger på dansk i 1986 under titlen " Om Regning på Lykkespil". Her finder vi følgende oplæg til delingsproblemet:

"Antag at jeg spiller mod en anden spiller i tre spil, og jeg allerede har vundet 2 og han kun et. I tilfælde af, at jeg ikke ønsker at spille videre, men at dele det, som er indsat, retfærdigt, vil jeg vide hvor meget deraf, der tilkommer mig." (Proposition IV i bogen)

Huygens giver i bogen sin løsning på problemet.

Derefter tager han flere delingsopgaver op:

Proposition V: Antag, at jeg mangler 1 spil, og den jeg spiller imod 3 spil; nu skal man foretage fordelingen.

Proposition VI: Antag, at jeg mangler to spil og han, der spiller mod mig, tre spil.

Proposition VII: Antag, at jeg mangler to spil og han 4 spil.

Derefter behandler Huygens i Proposition VIII en situation med tre spillere:

"Lad os nu antage, at tre personer spiller sammen, og at den første mangler 1 spil, den anden ligeledes 1 spil, men den tredje 2 spil."

I Proposition IX gives til afslutning en generel behandling af delingsproblemet.

Også hos De Moivre i "The Doctrine of Chances" (se omtalen i afsnit 1) finder vi eksempler på delingsproblemer:

12. Hvordan skal puljen deles?

Der indledes med: "Hvis A og B spiller mod hinanden, og A mangler 1 spil og B mangler 2, hvad er så deres respektive chancer for at vinde?"

Derefter bliver opgavesituationerne hurtigt sværere: Spillerne har nu forskellige chancer for at vinde det enkelte spil. Og der udvides med spil hvori der deltager tre spillere.

Delingsproblemet er i detaljer beskrevet i Lars Rasmussen: *Sandsynlighedsbegrebets oprindelse og præhistorie*. Se litteraturlisten bag i bogen.

13. Hverdagens skæve tal

Vi skal i dette afsnit se på hverdagens tal som de fremtræder i tabelværker og datasamlinger. Vi vil især se på det første ciffer i disse tal og interessere os for den frekvens de enkelte cifre 1,2,3,...,9 optræder med som førsteciffer. Det er altså det første betydende ciffer i tallene der er emnet for vore undersøgelser. Til eksempel er 3 førsteciffer i tallet 378 og 5 er førsteciffer i tallet 0.0523.

Ser vi til eksempel på alle trecifrede tal, 100,101,102, ..., 999, så har 100 af disse 900 tal førstecifferet 1. Det vil sige at 1 har en frekvens på 1/9 som førsteciffer. Den samme frekvens har hvert af de andre førstecifre 2, 3, 4,..., 9.

Samme frekvens møder vi hvis vi undersøger 6-cifrede tal eller tal med et andet antal cifre. Hver gang kan vi observere at alle muligheder for førsteciffer optræder med en frekvens på 1/9.

Man kunne så tro at hverdagens tal som vi møder dem i tabeller og datasamlinger også ville vise en sådan ligevægt i frekvensen af førstecifre. Men det er ikke tilfældet.

Allerede i 1880'erne fremkom en artikel *"Hvorfor slides logaritmetabeller ujævnt?"*. Man havde observeret på biblioteker og laboratorier at logaritmetabeller og andre regnetabeller udviste megen slitage af de sider hvor tallene begynder med cifrene 1 eller 2, og kun lidt slitage af de sider hvor tallene begynder med cifrene 8 eller 9. Det så ud til at man i daglig brug af tabellerne oftere foretog tabelopslag for tal med små værdier af førstecifferet end for tal med store værdier af førstecifferet.

Man udførte derefter systematiske undersøgelser af datasamlinger, og kunne her fx iagttage at i ca. 70% af tallene forekommer 1, 2, 3, 4 som førsteciffer, og tallene 5, 6, 7, 8, 9 forekommer kun som førsteciffer i ca. 30% af tallene. Det gentog sig i tabeller af vidt forskellig art: geografiske og demografiske tabeller, finansielle tabeller over produktion, eksport og import og fysiske og kemiske tabelværker over naturkonstanter.

Det afgørende var at der var tale om såkaldte *brede taldata*, dvs. taldata hvor ethvert ciffer principielt havde mulighed for at forekomme som førsteciffer. Således ville man ikke kunne anvende data over skoleelevers højde, eller deres vægt, eller deres skostørrelse. Disse data ville ikke være brede data.

I 1938 fremsatte fysikeren F. Benford et forslag til en teoretisk model til beskrivelse af fordelingen af førstecifre i hverdagstal. Benfords fordeling bygger på empiriske data, ikke på en teori. Hverdagstal er jo noget der er knyttet til praksis, og der kan ikke gives en præcis matematisk beskrivelse af dem.

Benfords fordeling gør brug af logaritmefunktionen til beskrivelse af modellens frekvenser for de forskellige førstecifre.

I modellen tillægges førstecifferet x, hvor x = 1, 2, 3,..., 9, en frekvens på:

$$\log(1 + 1/x).$$

Førstecifferet 1 får dermed frekvensen log 2 = 0.3010. Og førstecifferet 9 får tillagt en frekvens på log(1 + 1/9) = 0.0458.

For den kumulerede frekvens svarende til førstecifferet x får vi følgende enkle udtryk: log(1 + x). Til eksempel: Den kumulerede frekvens for 3:

$$\log(2) + \log(3/2) + \log(4/3) = \log(2 \cdot 3/2 \cdot 4/3) = \log(4)$$

Vi samler modellen i en tabel over frekvens og kumuleret frekvens for de ni førstecifre:

Førsteciffer X	Frekvens log(1 + 1/x)	Kumuleret frekvens log(1 + x)
1	0.301	0.301
2	0.176	0.477
3	0.125	0.602
4	0.097	0.699
5	0.079	0.778
6	0.067	0.845
7	0.058	0.903
8	0.051	0.954
9	0.046	1.000

Af tabellen ser vi fx at førstecifrene 1,2,3,4 optræder med en samlet frekvens på 0.699. Cifferet 1 har alene en frekvens på 0.301.- På en regnestok har vi direkte fysiske størrelser som afspejler førstecifrenes frekvenser.

Benfords fordeling kan ikke udledes teoretisk. Den må begrundes i at den giver resultater som er i god overensstemmelse med de data der findes i virkelighedens databaser over brede data. Benford kaldte selv fordelingen "en naturlov" og sammenlignede den med fysiske love.

Fordeling har været grundlag for et spil: Spilleren vælger en af et stort antal databaser som er indlagt i computeren. De er alle datasamlinger med hverdagstal og med brede data. Fra den valgte base udtrækkes et tilfældigt tal. Såfremt tallet har førsteciffer 1 eller 2, betaler spilleren 10 €, såfremt førstecifferet er 3,4,5,6,7,8 eller 9 modtager han 5 €. Ville du med dit kendskab til Benfords fordeling modtage tilbuddet om at deltage i spillet?

Andre enheder

Hvordan går det med Benfords fordeling hvis alle tal i en database multipliceres med en konstant? Altså hvis der skiftes enhed, for eksempel fra kilometer til engelske miles, eller fra kvadratkilometer til square miles, fra kilogram til pounds, eller fra danske kroner til euro. Vil det påvirke fordelingen af førstecifre?

Her kan vises at Benfords fordeling ikke er påvirket af et sådant skifte i enheder eller for en multiplikation med en fast positiv værdi. Fordelingen af førstecifre er altså uanset en ny måleskala beskrevet ved Benfords fordeling på en sådan måde at den stemmer godt overens med praktiske observationer. - Denne skala-neutralitet hos Benfordfordelingen kan være et yderligere argument for dens anvendelse til beskrivelse af hverdagstal.

"Matematiske tal"

Benfords fordeling beskriver hvordan førsteciffene fordeler sig i hverdagens tal. Men samme fordeling kan også forekomme i de tal vi møder i matematiske databaser.

Lad os til eksempel undersøge de 1000 første potenser af 3, altså 3^1, 3^2, 3^3, ..., $3^{1000.}$ Vi lader computeren undersøge førstecifferet i disse 1000 tal. Her er en tabel over resultatet:

```
Første ciffer i potenser fra 3^1 til 3^1000
```

Første ciffer	Antal tilf.	Frekvens
1	300	0.300
2	177	0.177
3	123	0.123
4	98	0.098
5	79	0.079
6	66	0.066
7	59	0.059
8	52	0.052
9	46	0.046

Det ses at fordelingen af førstecifre stemmer fint overens med Benfords fordeling.

Førsteciffer i brøktal
Nogle overraskelser gemmer sig i undersøgelsen af brøktals førsteciffer. Vi vælger to forskellige tal x og y i intervallet fra 0 til 1. Derefter danner vi brøktallet x/y. Vi danner x/y efter tre forskellige metoder:

(1) x er det første af de to valgte tal, y er det andet
(2) x er det største af de to valgte tal, y er det mindste
(3) x er det mindste af de to tal, y er det største.

Vi udfører i hvert af de tre tilfælde en undersøgelse af 1000 talpar (x,y).

Model 1
Første ciffer i x/y for 1000 par af tilfældige tal. x er det først udtagne tal.

Første ciffer	Antal tilf.	Frekvens
1	305	0.305
2	153	0.153
3	106	0.106
4	74	0.074
5	83	0.083
6	76	0.076
7	62	0.062
8	70	0.070
9	71	0.071

Det ses at fordelingen ved første øjekast kunne minde om Benfords fordeling. Vi vender tilbage til den teoretiske fordeling nedenfor.

Model 2
Første ciffer i x/y for 1000 par af tilfældige tal. x er det største af de to udtagne tal.

Første ciffer	Antal tilf.	Frekvens
1	556	0.556
2	174	0.174
3	89	0.089
4	64	0.064
5	44	0.044
6	24	0.024
7	22	0.022
8	15	0.015
9	12	0.012

Denne fordeling afviger klart fra Benfords fordeling.

Model 3
Første ciffer i x/y for 1000 par af tilfældige tal. x er det mindste af de to udtagne tal.

Første ciffer	Antal tilf.	Frekvens
1	115	0.115
2	103	0.103
3	105	0.105
4	104	0.104
5	129	0.129
6	118	0.118
7	93	0.093
8	114	0.114
9	119	0.119

Også denne fordeling adskiller sig væsentligt fra Benfords fordeling. Her er tale om en fordeling som ligger tæt på en jævn fordeling.

De teoretiske fordelinger
Der kan gives elementære formler for de fordelinger der forekommer i de tre modeller. Med F(c) betegner vi frekvensen for førstecifferet c. Da gælder at de teoretiske værdier for F(c) er givet ved:

Model 1

$$F(c) = \left(1 + \frac{10}{c(c+1)}\right)/18$$

Model 2

$$F(c) = \frac{10}{9c(c+1)}$$

Model 3

$$F(c) = 1/9$$

Bemærk at logaritmefunktionen ikke indgår i disse udtryk for førstecifferets frekvens. På dette punkt afviger de tre udtryk for F(c) altså principielt fra de frekvenser der er givet ved Benfords fordeling.

Konklusion: Det er ikke så ligetil at forudsige hvordan førstecifrene for tallene i en database vil fordele sig.

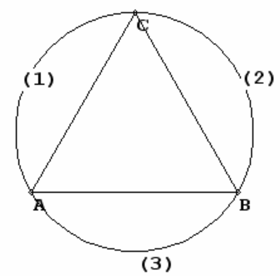
I tabellen nedenfor er angivet frekvenserne for de enkelte cifre i hver af de tre modeller.

Ciffer	**Model 1**	**Model 2**	**Model 3**
1	0.333	0.556	0.111
2	0.148	0.185	0.111
3	0.102	0.093	0.111
4	0.083	0.056	0.111
5	0.074	0.037	0.111
6	0.069	0.026	0.111
7	0.065	0.020	0.111
8	0.063	0.015	0.111
9	0.062	0.012	0.111

Det ses at alle tre fordelinger afviger fra Benfords fordeling.

14. Bertrands paradoks

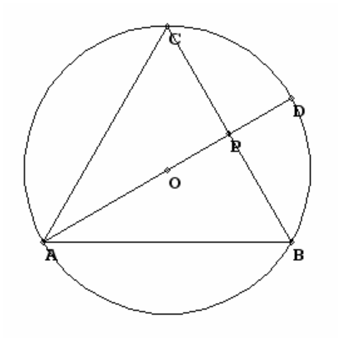
Problemet: I en cirkel trækkes en tilfældig korde. Hvad er sandsynligheden for at korden er længere end siden i cirklens indskrevne ligesidede trekant?



Første løsning
Vi tænker os at punkt A er valgt på cirklens periferi. Vi skal da bestemme sandsynligheden for at en korde gennem A er længere end siderne i den indskrevne trekant ABC, altså længere end siden AB.

Hvis korden skal være længere end AB, må kordens andet endepunkt placeres på den del af cirkelbuen der er betegnet (2). Ved valg af endepunkt på buen (1) eller (3) vil korden blive kortere end AB.

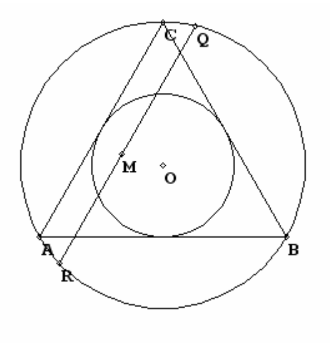
Buen (2) udgør en tredjedel af cirklens periferi. Hvis vi vælger et tilfældig punkt på periferien er der altså en chance på 1/3 for at punket falder på (2). Første løsning er altså at den søgte sandsynlighed er: **1/3.**



Anden løsning
Vi tegner diameteren AD i cirklen. Linjestykkerne OP og PD har begge en længde på den halve radius i cirklen. Den søgte korde tegner vi nu på følgende måde: Vi vælger et vilkårligt punkt på diameteren AD og tegner korden vinkelret på diameteren i det valgte punkt. Hvis vi fx vælger et punkt på linjestykket PD, vil kordens længde blive mindre end BC. Vælger vi derimod et punkt på linjestykket OP, vil korden blive længere end BC.

Tilsvarende overvejelser kan vi gøre hvis det valgte punkt placeres på linjestykket OA. Hvis punktet vælges på den halvdel af linjestykket AD som ligger nærmest O, bliver korden længere end BC. Hvis punktet vælges på den anden halvdel af OA, bliver korden kortere end BC. Alt i alt kan vi til kordens skæring med diameteren vælge halvdelen af diameteren, nemlig den halvdel der ligger nærmest punkt O.

Sandsynligheden for en korde som er længere end siden i den indskrevne trekant, er derfor ved denne løsning: **1/2.**



Tredje løsning
Her vælger vi et punkt inden i trekant ABC's omskrevne cirkel. Dette punkt lader vi være midtpunkt af den korde vi vil tegne. På figuren er det valgte punkt markeret med M. Vi tegner derefter en korde gennem M således at dens midtpunkt bliver M, dvs. vi tegner korden vinkelret på linjestykket OM. Derved fremkommer korden RQ.

Når M vælges inden i den lille cirkel, vil korden være længere end siderne i trekant ABC. Når M vælges uden for den lille cirkel, vil korden være kortere end siderne i trekant ABC.

Arealet af den lille cirkel udgør en fjerdedel af arealet af den store cirkel, radius i den lille cirkel er jo halvt så stor som radius i den store cirkel. Vi benytter dette forhold mellem arealerne som et udtryk for den søgte sandsynlighed: Sandsynligheden for at punktet M falder i den lille cirkel, og dermed sandsynligheden for at den tilfældige korde er længere end siderne i trekant ABC, er: **1/4**.

Den rigtige løsning?
Man vil naturligt stille sig det spørgsmål: *Hvad er den rigtige løsning?* Er sandsynligheden 1/2 eller 1/3 eller 1/4?

Vi vil illustrere situationen med et overskueligt eksempel:

14. Bertrands paradoks

Hvilket udfaldsrum?
Sanne og Malene diskuterer følgende problem: I klasseværelset er der to ledige tomandsborde. To elever kommer ind i klasseværelset og vælger på tilfældig vis siddepladser ved de to borde. Hvad er sandsynligheden for at de sætter sig ved samme bord?

Sanne og Malene kan ikke blive enige om hvad sandsynligheden er. De er nemlig ikke enige om hvilket udfaldsrum der skal bruges. Sanne mener at der er tale om et udfaldsrum med 4 ligevægtede udfald. Hver elev kan vælge mellem bord 1 og bord 2. Sannes udfaldsrum ser derfor således ud:

		2. elev vælger	
		Bord 1	Bord 2
1. elev vælger	Bord 1	X	
	Bord 2		X

Sannes løsning.

Sandsynlighed: 1/2

Udfaldsrummet indeholder 4 ligevægtede udfald, og to af udfaldene (markeret med X) svarer til at de to elever vælger plads ved samme bord. Den søgte sandsynlighed må derfor efter Sannes mening være 1/2.

14. Bertrands paradoks

		2. elev vælger			
		Plads 1	Plads 2	Plads 3	Plads 4
1. elev vælger	Plads 1		X		
	Plads 2	X			
	Plads 3				X
	Plads 4			X	

Malenes løsning

Sandsynlighed: 1/3

Malene er af en anden mening. Hun siger: Der er fire siddepladser, plads nr. 1 og 2 ved bord 1, og plads nr. 3 og 4 ved bord 2. Situationen kan derfor beskrives ved hjælp af et udfaldsrum med 12 ligevægtede udfald som vist på figuren ovenfor.

Vi ser at 4 af de 12 udfald (markeret med X) svarer til at de to elever vælger plads ved samme bord. Den søgte sandsynlighed må derfor efter Malenes mening være 4/12 = 1/3.

Hvem har nu ret, Sanne eller Malene? Ja, det har de begge to. De har nemlig begge løst opgaven helt korrekt, men blot ud fra forskellige antagelser. Der var jo ikke givet oplysning om hvordan de to elevers tilfældige valg af siddepladser foregik. Derfor kunne Sanne og Malene komme til to forskellige løsninger af opgaven.

Hvis valget af siddepladser foretages således at hver elev vælger mellem de to borde, så har Sanne ret. Men hvis eleverne vælger mellem de fire pladser, så har Malene ret.

Vi kan derfor ikke sige hvilken af de to løsninger der er den rigtige. Først når chancesituationen er så klart beskrevet at det fremgår hvordan de tilfældige valg er foretaget, kan det afgøres hvilken løsning der er den rette.

Diskussionen mellem Sanne og Malene viser os at chancesituationer må behandles med varsomhed. En chancesituation kan undertiden opfattes på mere end én måde, og derfor kan der opstå uenighed om hvordan et forelagt problem skal løses.

I sandsynlighedsregningens historie finder man adskillige diskussioner der ligner den Sanne og Malene havde.

Ikke noget paradoks
Bertrands paradoks har fået navn efter den franske matematiker J. Bertrand (1822 -1900). I virkeligheden er der ingen mystik knyttet til Bertrands tre løsninger på det stillede problem. Der er blot tale om tre forskellige løsninger af en opgave som ikke er klart specificeret. Det er således ikke på forhånd fastlagt hvad problemets udfaldsrum er: Skal der som i første løsning vælges et tilfældigt punkt på en cirkelperiferi? Skal der som i anden løsning vælges et tilfældigt punkt på en cirkeldiameter? Eller skal der som i tredje løsning vælges et tilfældigt punkt i et cirkelareal?

Så snart problemet er veldefineret, forsvinder mystikken, og der er kun én løsning på det stillede problem.

Bertrand sagde selv om de tre løsninger: *Der er ingen af dem der er falsk, og der er ingen der er fyldestgørende, der er blot tale om et uklart formuleret problem.*

15. Kast ting på gulvet og find π

Problemet: En række parallelle linjer med samme indbyrdes afstand er tegnet. En plan figur kastes ned på mønstret af linjer. Hvad er chancen for at figuren skærer en af de tegnede linjer?

Vi vil se på en situation hvor den anvendte figur er en femkant. De fem kanter benævner vi k1, k2, k3, k4 og k5. Femkanten kastes ned på mønstret af parallelle linjer. Vi antager at linjerne har en sådan afstand fra hinanden at femkanten ikke kan ramme mere end én af linjerne. Femkanten vil altså skære én linje, eller den vi falde helt inde i et mellemrum mellem to nabolinjer.

Hvis femkanten falder således at kanten k1 skærer en linje, så må linjen også skæres af en af de andre kanter, linjen kan jo ikke kun "gå ind i" femkanten, den må også komme ud igen. Fx kan k3 også skæres. Vi noterer det som (k1k3) Vi kan opdele situationen således:

(k1): (k1k2) eller (k1k3) eller (k1k4) eller (k1k5)

At femkanten skæres af en af de tegnede linjer betyder:

(femkant): (k1k2) eller (k1k3) eller (k1k4) eller (k1k5)
eller (k2k3) eller (k2k4) eller (k2k5)
eller (k3k4) eller (k3k5) eller (k4k5)

For de fem kanter har vi lige som ovenfor:

(k1): (k1k2) eller (k1k3) eller (k1k4) eller (k1k5)
(k2): (k2k1) eller (k2k3) eller (k2k4) eller (k2k5)
(k3): (k3k1) eller (k3k2) eller (k3k4) eller (k3k5)
(k4): (k4k1) eller (k4k2) eller (k4k3) eller (k4k5)
(k5): (k5k1) eller (k5k2) eller (k5k3) eller (k5k4)

De to opstillinger indeholder henholdsvis 10 og 20 udtryk af formen (kikj). Vi kan let se at der i den nederste opstilling blot er tale om at hvert af de 10 udtryk fra den øverste opstilling forekommer to gange. Således er jo fx (k1k2) det samme som (k2k1), og (k3k5) er det samme som (k5k3).

Heraf har vi at chancen for at femkanten falder således at den skærer en af de tegnede linjer kan udtrykkes ved:

P(femkant) = ½(P(k1) +P(k2) +P(k3) +P(k4) +P(k5))

En antagelse
Vi antager nu at chancen for at en kant kommer til at skære en af de tegnede linjer afhænger af kantens længde. Vi sætter:

P(k) = c · lg(k)

hvor c er en konstant som ikke afhænger af kantens længde, og hvor lg(k) er længden af kanten k.

For femkanten har vi nu:

P(femkant) =
½(c·lg(k1) + c·lg(k2) + c·lg(k3) + c·lg(k4) + c·lg(k5))

Dette kan vi skrive således.

P (femkant) = ½ c·Omkredsen af femkanten

En tilsvarende udledning kunne vi have foretaget for en vilkårlig polygon. Vi ville da få at chancen for at polygonen falder sådan at den skærer en af de tegnede linjer kan udtrykkes ved:

P(polygon) = ½ c·Omkredsen af polygonen

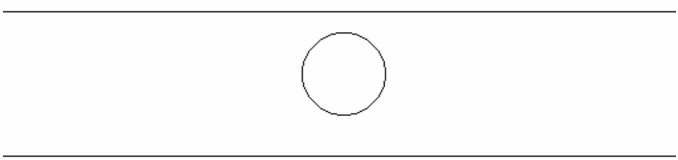
15. Kast ting på gulvet og find π

Den søgte chance afhænger altså ikke af hvor mange kanter der er i den foreliggende polygon. Samme formel gælder for en trekant, firkant, femkant og en syttenkant.

Hvad er c?
For at finde c må vi kende chancen for at en bestemt polygon skærer de tegnede linjer.

Vi benytter her som polygon en cirkel. Det kan se ud som snyd, men da formlen gælder for vilkårlige polygoner, gælder den også for regulære polygoner som i omkreds ligger meget tæt på en forelagt cirkel.

Lad os se på en cirkel med diameter d og lad de tegnede linjer have en indbyrdes afstand på a. Der gælder: a>d, cirklen må kun kunne ramme en af de tegnede linjer.



Hvis der skal være skæring må cirklens centrum falde inden for en afstand af ½d fra en af to nabolinjer. Chancen for at centrum falder således vil være:

$$P(\text{cirklen}) = d/a$$

Af formlen ovenfor har vi derfor:

$$d/a = \tfrac{1}{2}\, c \cdot \pi \cdot d$$

hvoraf:

15. Kast ting på gulvet og find π

$$c = \frac{2}{\pi \cdot a}$$

Formlen for chancen for at polygonen skærer en af de tegnede linjer kan herefter gengives ved:

$$\mathbf{P(polygon)} = \frac{O}{\pi \cdot a}$$

hvor O er polygonens omkreds. Denne formel viser at den forventede skæringsfrekvens kan beregnes som *forholdet mellem to omkredse*: I tælleren omkredsen af "kastetingen", i nævneren omkredsen af en cirkel med diameter a. Altså en cirkel som lige passer ind mellem to linjer i mønstret af de parallelle linjer.

Andre figurer
I de følgende eksempler forudsættes det at den betragtede figur kun kan skære én linje i mønstret af parallelle linjer.

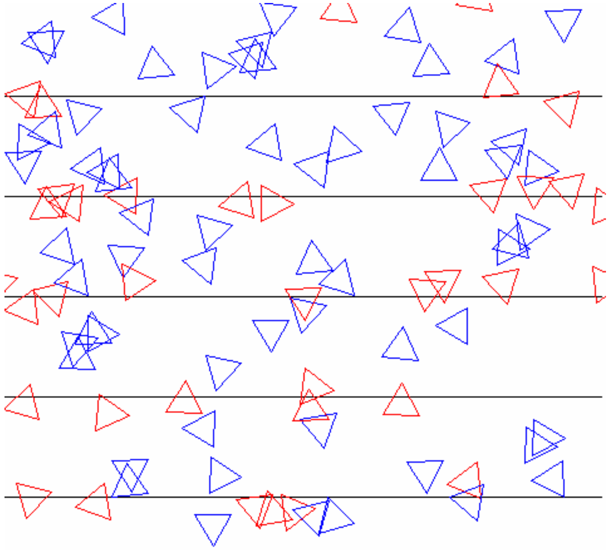
En trekant kastes
Vi ser på en ligesidet trekant med siderne s. Her gælder:
Chancen for at trekanten skærer en af linjerne:

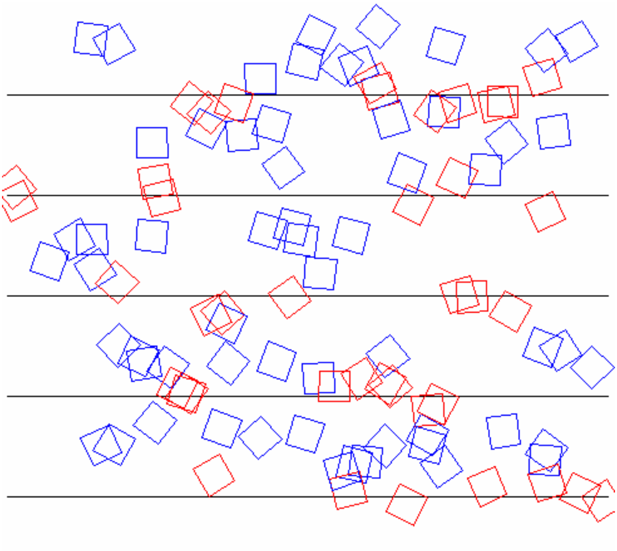
$$\frac{3s}{\pi \cdot a}$$

Et kvadrat kastes
Chancen for at kvadratet med siderne s skærer en af linjerne:

$$\frac{4s}{\pi \cdot a}$$

15. Kast ting på gulvet og find π





15. Kast ting på gulvet og find π

Et rektangel kastes

Chancen for at rektanglet med siderne x og y skærer en af linjerne:

$$\frac{2(x+y)}{\pi \cdot a}$$

Buffons problem
Historisk stammer problemerne med kast af en figur på et mønster af parallelle linjer fra den franske matematiker G.- L. Buffon (1707 – 88). I 1777 fremlagde han følgende problem:

En nål af længden b kastes på et mønster af parallelle linjer der har en indbyrdes afstand af a, hvor a>b. Hvad er chancen for at nålen skærer en af de tegnede linjer?

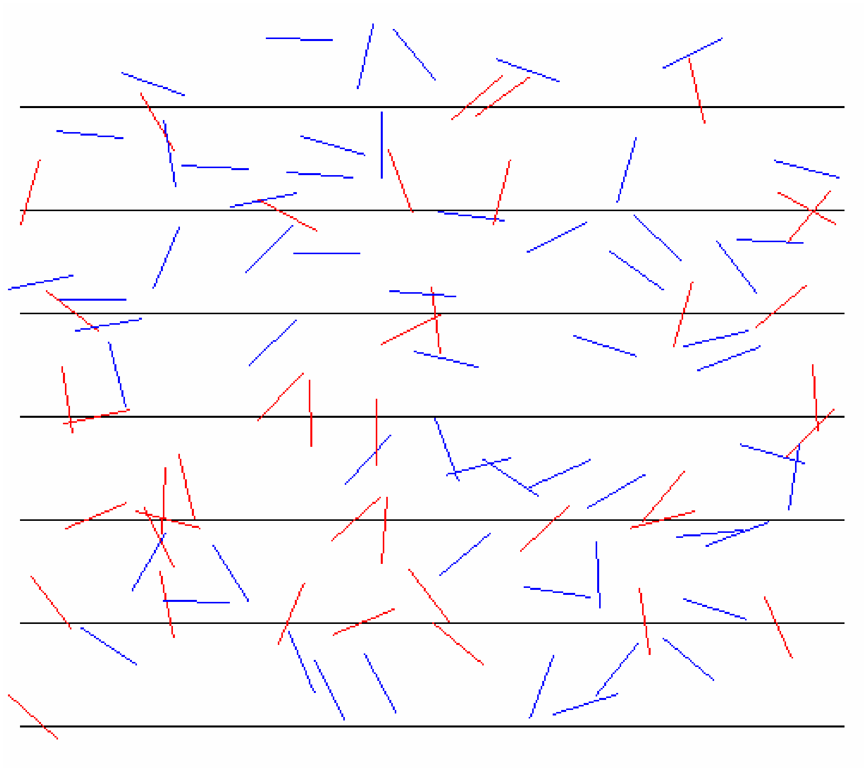
Vi kan benytte vores formel, idet vi betragter nålen som en "tokant" med to sider, k1 og k2, som begge har længden b. Omkredsen af tokanten bliver da: 2b. Svaret på Buffons problem er derfor:

$$\frac{2b}{\pi \cdot a}$$

Buffons problem kan løses ad andre matematiske veje, og de fører alle frem til det ovenfor givne resultat.

Af formlerne ser vi at hvis vi kaster en ring af ståltråd og bag-efter klipper den op og retter den ud som en nål, så bliver chancen for nålens skæring med de tegnede linjer dobbelt så stor som ringens.

15. Kast ting på gulvet og find π



Et dobbeltnet

Buffons problem kan udvides til at omfatte kast i et kvadratnet bestående af kvadrater med siden a eller i et rektangulært net dannet af ét sæt linjer med indbyrdes afstand g og et andet sæt (vinkelret på det første) med en indbyrdes afstand h. Her er chancen for at en nål af længden b skærer en af de tegnede linjer (eller dem begge):

$$\frac{2b}{\pi \cdot g} + \frac{2b}{\pi \cdot h} - \frac{b^2}{\pi \cdot g \cdot h} \qquad \text{hvor} \quad b<g \ \text{ og } \ b<h$$

15. Kast ting på gulvet og find π

Eksperimentel bestemmelse af π
De fundne formler til beregning af chancer ved kast af plane figurer på mønstret af parallelle linjer kan benyttes til en bestemmelse af talværdien for π.

Vi foretager et kast med tændstikker af længden 4.5 cm. De kastes på et mønster af parallelle linjer med en indbyrdes afstand af 10 cm. I et eksperiment på 100 kast forekommer skæring af en af de tegnede linjer i 28 tilfælde. Chancen for skæring er altså her målt til 0.28.

Ved indsættelse i formlen for kast med en nål får vi med b=4.5 og a=10:

$$0.28 = \frac{2 \cdot 4.5}{\pi \cdot 10}$$

Heraf beregnes: π = 3.21.

Denne bestemmelse er følsom over for små ændringer. Ved 27 skæringer var resultatet blevet: π = 3.33 og ved 29 skæringer: π = 3.10.

Gennem årene er der berettet om mange eksperimenter der gør brug af Buffons formel til bestemmelse af π. Nogle eksperimenter giver en π-værdi som kun afviger fra den teoretiske værdi på 7. decimal. De må dog tages med en vis skepsis, det kan være svært at tælle skæringer korrekt når man på forhånd ved hvad resultatet burde blive.

Her er data fra 100 eksperimenter med hver 100 kast af et kvadrat.

```
Antal simuleringer: 100
Antal kvadrater pr. simulering: 100
Kvadratets omkreds: 8
Linjernes afstand: 8
```

15. Kast ting på gulvet og find π

```
Gennemsnitlig frekvens
for skæring: 0.317

Teoretisk frekvens
for skæring: 0.318
```

π `-estimat:   3.152`

Når den observerede frekvens for skæring lægges til grund for beregning af π, bliver resultatet ca. 3.15.

Faglig udvidelse
Der er også opstillet formler for kast med figurer som har en sådan udstrækning at de kan skære to af de parallelle linjer.

Ligeledes er undersøgt kast med ting på et net bestående af ligesidede trekanter.

16. Er det virkelig tilfældigt?

På en skole fører man statistik over hvor mange uheld der sker i skolegården. Det viser sig at der i løbet af 20 skoledage i en måned sker 20 uheld som kræver at skolens forbindingskasse kommer i brug.

Statistikken viser at uheldene ikke er jævnt fordelt på de 20 dage. Der er en enkelt dag med 3 uheld, og der er hele 7 dage hvor der slet ikke indtræffer nogen uheld.

En klasse diskuterer i matematiktimen om uheldene kan være tilfældigt fordelt, eller om der skulle være en eller anden grund til at de klumper sammen. Da man ikke kan blive enige, vælger man at belyse problemet ved hjælp af en computersimulering.

Der udtages 20 tal fra talområdet 1...20. Tallene udtages med tilbagelægning, og de udskrives i sorteret tilstand:

1	1	2	3	3	4	4	7	9	11
12	14	14	15	15	15	16	18	18	20

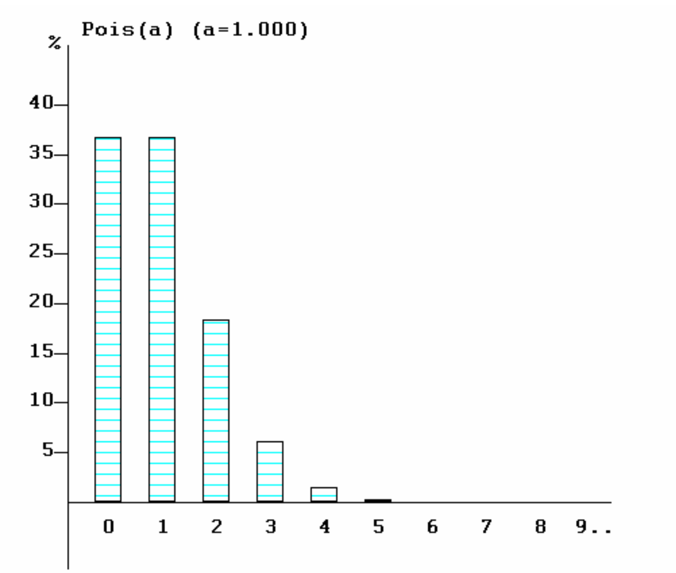
Udskriften viser at der dag nr. 1 var 2 uheld, dag nr. 1 er jo udtrukket 2 gange. Dag nr. 2 er der kun ét uheld, men dag nr. 3 er der igen 2 uheld. Dagene nr. 5 og nr. 6 forekommer ikke i listen, disse dage var der altså ingen uheld.

En lille statistik over antal uheld giver følgende resultat:

Dage med 3 uheld:	1	(det er dag nr.15)
Dage med 2 uheld:	5	
Dage med 1 uheld:	7	
Dage uden uheld:	7	

Dette eksperiment giver altså som resultat at der forekommer én dag med tre uheld, og der er 7 dage uden uheld. De 7 dage der ikke er med i listen er dagene nr. 5, 6, 8, 10, 13, 17 og 19. De ligger ganske godt spredt blandt de 20 dage.

Den slags chancesituationer kan belyses ved hjælp af en sandsynlighedsfordeling som hedder poisson-fordelingen. Den har navn efter den franske matematiker Simon Poisson (1781-1840).



Her er et grafisk billede (fra programmet Stattabel)af den poissonfordeling der svarer til situationen med uheld i skolegården. Der var 20 uheld på 20 dage, dvs. der var i gennemsnit 1 uheld pr. dag. Vi skal da gøre brug af poissonfordelingen med parameteren 1, og det er den der er afbildet i pindediagrammet.

Vi kan af figuren se at 0 uheld har en sandsynlighed på ca. 37%, og det samme gælder 1 uheld. Det svarer til at der af de 20 skoledage i gennemsnit vil være 7 uden uheld og 7 med ét uheld. Og det stemmer jo fint med vores opgørelse fra før. Af poissonfordelingen kan vi se at dage med 2 uheld har en sandsynlighed på lidt under 20%, dvs. at vi kan forvente omkring 4 dage med 2 uheld. Vor optælling gav 5 dage.

Af dage med 3 uheld var der kun 1, svarende til 5% af dagene. Også det stemmer godt med poissonfordelingen.

Fordelingen af uheld i skolegården stemmer således fint med den teoretiske fordeling i poissonfordelingen, eller med andre ord: *Uheldene følger tilfældighedens love*, og der er ingen grund til at stille spørgsmål ved uheldenes tilfældighed.

Frekvensfordelingen for poissonfordelingen med parameter a (lig med fordelingens middeltal) er fastlagt ved:

$$f(x) = e^{--a} \cdot a^x/x!$$

Her er *e* grundtallet for de naturlige logaritmer: e = 2.71828.

For a = 1 og x = 0 får vi: $f(0) = e^{-1} = 0.3679$.

Poissonfordelinger kan benyttes ved situationer der kan beskrives som "Fordeling af kugler i celler". I vort eksempel var der 20 kugler (uheld) der skulle fordeles i 20 celler(dage).

Nogle andre "kugler i celle"-situationer:

Kugler	**Celler**
Trykfejl	Bogsider
Trafikuheld	Dage
Kraftige jordskælv	År
Rosiner	Rosinboller
Mål	Fodboldkampe

16. Er det virkelig tilfældigt?

Hestespark

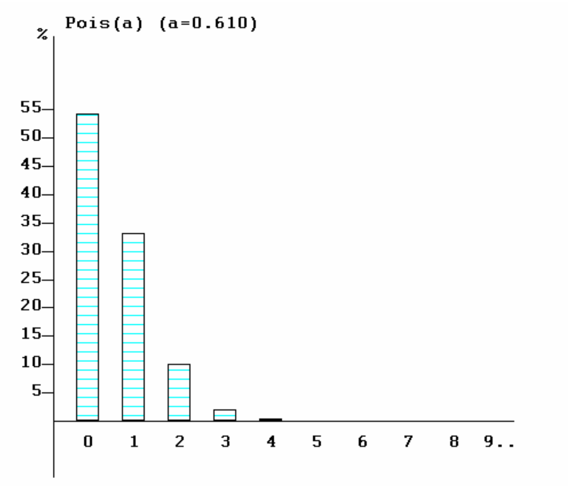
Den russiske matematiker Bortkiewicz (1868 -1931) fremlagde i sin bog fra 1898 "De små tals lov" en række anvendelser af poissonfordelinger. Blandt dem var en statistik fra udvalgte kavaleriregimenter over dødsfald på grund af hestespark.

Over tyve år blev observeret dødsfald i 10 regimenter. Her er tallene for fordelingen af dødsfald på de enkelte år

Antal dødsfald	Antal år	Frekvens
0	109	54.5%
1	65	32.5%
2	22	11%
3	3	1.5%
4	1	0.5%
0ver 4	0	0%

Det giver 122 dødsfald på 200 år, eller 0.61 dødsfald pr.år.

Vi sammenligner nu med poissonfordelingen med parameter 0.61:

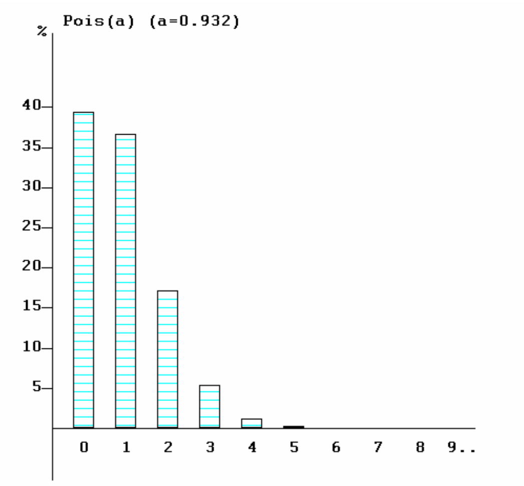


Det ses at de teoretiske frekvenser stemmer fint overens med de observerede.

Bombenedslag
En anden ofte citeret anvendelse vedrører bombningen af London under 2. verdenskrig. Her blev optalt bombenedslag for et område bestående af 576 felter hvert på ½ km x ½ km. For hvert felt blev noteret hvor mange bomber der faldt i feltet. Her er statistikken:

Antal bomber	Antal felter	Frekvens
0	229	39.8%
1	211	36.6%
2	93	16.1%
3	35	6.1%
4	7	1.2%
5 og derover	1	0.2%

I alt var der 537 bombenedslag i de 576 felter, dvs. et gennemsnit på 0.932 pr. felt. Her er poissonfordelingen pois(0.932):



Igen er der tale om en god overensstemmelse mellem de teoretiske frekvenser og de observerede.

*

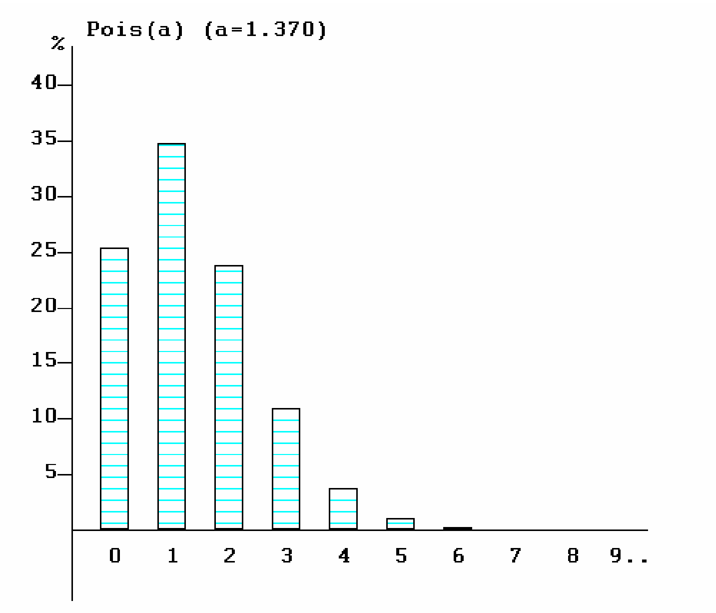
Poissonsandsynligheder, som de er anvendt i dette afsnit, er at opfatte som tilnærmelser for binomialsandsynligheder. Lad os til eksempel se på 500 kugler der fordeles i 365 celler. Det kan fx være fødselsdage for 500 personer som fordeles på årets 365 dage.

Hvis vi vil beregne sandsynligheden for at 0,1,2,...,6 personer blandt de 500 har fødselsdag en bestemt dag, fx 1. marts, kan vi benytte en binomialfordeling bin(n,p) med n=500, p=1/365, eller vi kan benytte en poissonfordeling med parameter 500/365 = 1.37.

For frekvenserne får vi for antallet af personer med fødselsdag 1. marts

Antal personer	Binomial	Poisson
0	0.2537	0.2541
1	0.3484	0.3481
2	0.2388	0.2385
3	0.1089	0.1089
4	0.0372	0.0373
5	0.0101	0.0102
6	0.0023	0.0023

Det ses at der er stor overensstemmelse mellem binomial-sandsynlighederne og poissonsandsynlighederne. Som en tommelfingerregel kan poissonsandsynligheder benyttes som tilnærmelsesværdier for binomialsandsynligheder når n har en stor værdi og p en lille værdi. Det betyder at parameteren a=n·p i poissonfordelingen vil have en "moderat værdi".



Poissonsandsynligheder kan beregnes med INFA-programmet Stattabel, eller de kan aflæses i *Erlang S* hvor en række poissonfordelinger er tabellagt.

17. Talmønstre med overraskelser

Vi ser igen på møntkast. Vi betegner Krone med 1 og Plat med 0, og vi vil se på de talmønstre der fremkommer i en serie af møntkast. Først vil vi holde os til de mønstre der kun består af to kast. Her er fire muligheder: 00, 01, 10 og 11. Når vi foretager to kast med en mønt, så har disse fire muligheder samme sandsynlighed, nemlig 1/4 til hver.

Hvis vi blot udfører serier af to møntkast, så vil vi i en lang forsøgsrække kunne observere at de fire mønstre hvert forekommer i ca. 25% af eksperimenterne.

Dette er der intet overraskende i. Men nu vil vi se på de fire mønstre når de indgår i en længere række af møntkast.

Her er en oversigt over 10 serier hver på 6 møntkast. Vi vil se på de to mønstre 00 og 10.

Vi tænker os at der er to spillere A og B. Spiller A har valgt mønstret 00 og B har valgt 10. Den spiller hvis mønster forekommer først i rækken over kast, har vundet spillet.

1:	0	**1**	**0**	1	1	1
2:	1	1	1	1	**1**	**0**
3:	0	1	1	1	**1**	**0**
4:	**1**	**0**	1	1	0	1
5:	0	1	**1**	**0**	1	0
6:	**1**	**0**	1	0	1	1
7:	1	1	1	**1**	**0**	1

8:	**0**	**0**	0	1	0	0
9:	0	**1**	**0**	1	1	0
10:	**0**	**0**	0	1	0	0

I listen har vi fremhævet det af de to mønstre der først forekommer. I serie 1 er det mønstret 10 der allerede forekommer i kast nr. 2 og 3. I serie 2 vinder B igen, men her forekommer hans mønster først i kast nr. 5 og 6.

I de syv første serier vinder B, først i serie 8 kommer A's mønster først. Og det samme sker igen i serie 10.

Alt i alt ser vi at A vinder 2 gange, og B vinder 8 gange. Og alligevel ved vi at de to mønstre har samme sandsynlighed for at forekomme når vi udfører to møntkast. Det vil vi se lidt nærmere på.

Vi vil sammenligne mønstrene to og to og beregne sandsynlighederne for at hvert af de to mønstre vinder i en serie af møntkast.

Mønstrene 00 og 10. Vi ser på de to mønstre 00 og 10. Hvis mønstret 00 forekommer i de to første møntkast, så vinder A. Hvis mønstret ikke forekommer i de to første møntkast, så vinder B, dvs. A kan ikke vinde efter de to første møntkast. Det ville nemlig betyde at mønstret 00 skulle forekomme før 10. Men forud for 00 må der have været et 1-tal, og derfor ville mønstret 10 forekommer før 00.

Det betyder at A kun kan vinde hvis de to første kast begge er Plat, og sandsynligheden herfor er 1/4. Altså er B's vindersandsynlighed 3/4, og A's vindersandsynlighed er 1/4. Disse sandsynligheder stemmer jo fint med de resultater vi har fra listen over de 10 serier af møntkast.

Mønstrene 11 og 01. Ved sammenligningen af mønstrene vil vi udnytte de symmetrier der foreligger. Vi har sammenlignet mønstrene 00 og 10. Hvis vi ombytter 0 og 1 i de to mønstre, så har vi mønstrene 11 og 01. Her vil vi kunne foretage helt de samme overvejelser som ovenfor. Resultatet ville være: Mønstret 01 har en vindersandsynlighed over for mønstret 11 på 3/4.

Mønstrene 00 og 11. Vi vil også kunne udnytte symmetrien ved sammenligning af mønstrene 00 og 11. De to mønstre må stå helt lige, det må være "lige svært" at opnå to Plat i træk som to Krone i træk. Hvert af de to mønstre har altså en vindersandsynlighed på 1/2 over for det andet mønster. I øvrigt får vi jo det ene mønster fra det andet ved en ombytning af 0 og 1.

Mønstrene 01 og 10. Samme ræsonnement kan vi benytte over for de to mønstre 01 og 10. Det ene fremgår af det andet ved en ombytning af 0 og 1. Også disse mønstre har derfor en vindersandsynlighed på 1/2 over for hinanden.

Mønstrene 00 og 01 og 11 og 10. Vi mangler at se på mønstrene 00 og 01 og de to mønstre der forekommer ved ombytning af 0 og 1: mønstrene 11 og 10. Vi lader A vælge 00 og B vælge 01, og vi deler op efter kasteresultater:

Første kast 0. Næste kast vil enten give 0 (og A vinder) eller 1 (og B vinder). A og B står helt lige i sandsynligheder

Første kast 1. Hvis næste kast giver 0 er situationen helt som ovenfor, de to spillere står lige. Hvis næste kast giver 1, foretages et nyt kast.

Det ses at så snart der forekommer et 0, vil det efterfølgende kast afslutte spillet. Og i dette kast har de to spillere samme sandsynlighed for at vinde.

Med andre ord: De to mønstre står helt lige, og hvert af dem har en vindersandsynlighed på 1/2 over for det andet mønster.

Af symmetrigrunde vil det samme resultat gælde for mønstrene 11 og 10.

Vi samler resultaterne i en tabel. A vælger et mønster, og derefter vælger B et mønster. Tabellen angiver B's vindersandsynligheder.

	A:11	**A:10**	**A:01**	**A:00**
B:11	--	1/2	1/4	1/2
B:10	1/2	--	1/2	3/4
B:01	3/4	1/2	--	1/2
B:00	1/2	1/4	1/2	--

Tabellen viser at uanset A's valg, så kan B foretage et sådant valg at hans vindersandsynlighed er på mindst 1/2. I hver søjle i tabellen forekommer jo sandsynligheder på 1/2 eller derover. Spillet er altså gunstigt for spiller B.

Her er en sammenligning af mønstrene 00 og 01. Efter vore beregninger skulle de begge have en vindersandsynlighed på 1/2 over for det andet mønster

1:	1	1	1	1	**0**	**0**
2:	**0**	**0**	0	1	0	0
3:	1	**0**	**1**	1	0	0
4:	**0**	**0**	1	1	1	0
5:	**0**	**1**	0	0	0	1

6:	**0**	**1**	1	0	1	0
7:	1	1	**0**	**1**	0	1
8:	1	1	1	1	**0**	**0**
9:	**0**	**0**	0	1	1	1
10:	**0**	**0**	1	1	1	1

Her fordeler serierne sig med fire til mønstret 01 og seks til mønstret 00.

Kastegennemsnit
Vi vil nu se på hvor mange kast der i gennemsnit skal udføres før et bestemt mønster foreligger. Vi ser til eksempel på mønstret 10, og vi vil beregne K(10), det gennemsnitlige antal kast der skal udføres for at mønstret 10 foreligger.

Vi gør her brug af såkaldte betingede gennemsnit. Således anvender vi i beregningen af K(10) det betingede gennemsnit K(10|1), nemlig kastegennemsnittet for at få mønstret 10 når det vides at første kast gav resultatet 1. På tilsvarende vis er K(10|11) kastegennemsnittet for at få mønstret 10 når det vides at de to første kast gav resultatet 11.

I udledningen lader vi p være sandsynligheden for et krone-kast, dvs. for resultatet 1, og vi lader q være sandsynligheden for resultatet 0. Når udledningen af K(10) foreligger, kan vi indsætte værdierne ½ for p og q og dermed få K(10) givet en talværdi i vore møntkastserier.

Beregningen af K(10):

$$K(10) \;\; = K(10|1)\cdot p \; + \; K(10|0)\cdot q$$

$$K(10|1) = K(10|11)\cdot p \; + \; K(10|10)\cdot q$$

Vi har her:

K(10|0) = 1 + K(10)
Kommentar: Hvis første kast giver 0, er vi tilbage ved udgangspunktet: at bestemme K(10), og vi har allerede brugt 1 kast som ikke har ført os nærmere til mønstret 10.

K(10|11) = 1 + K(10|1)
Kommentar: Hvis de to første kast giver 11, så har vi brugt et kast (det første) som ikke kan udnyttes. Vi er dermed i situationen K(10|1), men har brugt 1 kast.

K(10|10) = 2
Kommentar: Her har vi opnået det ønskede mønster i løbet af 2 kast.

Ved indsættelse i udtrykkene ovenfor får vi idet vi sætter K(10) til x og K(10|1) til y:

$$x = y \cdot p + (1 + x) \cdot q$$

$$y = (1 + y) \cdot p + 2q$$

Vi har her to ligninger med de to ukendte x og y. Ved omformninger får vi for x, dvs. for K(10):

$$K(10) = \frac{1}{p \cdot q}$$

Ved indsætning af p = q = 1/2 har vi det ønskede tal:

$$K(10) = 4.$$

Af symmetrigrunde må vi da også have: K(01) = 4.

I gennemsnit skal der altså udføres 4 kast før mønstret 10 foreligger. Og det samme gælder for mønstret 01.

Vi lader computeren udføre 1000 kasteserier. I hver serie kastes indtil mønstret 10 forekommer. Her er statistikken over de 1000 forsøg:

```
   Statistik over 1000 eksperimenter
   Mønster: 10     Variabel: Ventetid


      Talområde            Fraktil

        2..2                 1 %
        2..2                 5 %
        2..2                10 %
        2..2                25 %
        2..4                50 %
        2..5                75 %
        2..7                90 %
        2..8                95 %
        2..11               99 %

Gennemsnit:    3.95
Maximum   :   18
Minimum   :    2
```

Det ses at der i gennemsnit skulle udføres 3.95 kast før mønstret 10 forelå. Af fraktilværdierne ses også at medianen i fordelingen af kasteantal er 4.

Mønstrene 00 og 11

Kastegennemsnittene for mønstrene 00 og 11 kan udledes på samme vis med anvendelse af betingede gennemsnit. Vi anfører her resultatet af udledningerne:

$$K(00) = K(11) = \frac{1}{p} + \frac{1}{p^2}$$

Ved indsættelse af p = q = 1/2 får vi: K(00) = K(11) = 6.

Her er en statistik over 1000 kasteserier:

```
Statistik over 1000 eksperimenter.
Variabel: Ventetid    Mønster: 00

   Talområde            Fraktil

     2..2                  1 %
     2..2                  5 %
     2..2                 10 %
     2..3                 25 %
     2..5                 50 %
     2..8                 75 %
     2..12                90 %
     2..15                95 %
     2..22                99 %

  Gennemsnit:    5.95
  Maximum   :   36
  Minimum   :    2
```

Vore resultater viser at mønstrene 00 og 01 har kastegennemsnit på henholdsvis 6 og 4. Alligevel har de to mønstre begge en vindersandsynlighed på1/2 i en indbyrdes konkurrence.

De opnåede resultater udtrykt ved p og q kan fx anvendes på terningkast. Lad 1 svare til en sekser og lad 0 svare til de øvrige øjental. Vi har da: p=1/6 og q=5/6.
Ved indsættelse får vi:

$$K(11) = \frac{1}{p} + \frac{1}{p^2} = 42.$$

I en serie af kast med en terning må vi altså i gennemsnit vente 42 kast på at få to seksere i træk.

Tre-mønstre
Vi vil nu se på mønstre bestående af tre møntkast. Idet vi igen lader 1 stå for Krone og 0 for Plat, så er der 8 mulige mønstre: 111, 110, 101, 100, 011, 010, 001 og 000.

Hvis vi blot ser på serier af tre kast, så har disse 8 mønstre samme sandsynlighed for at forekomme, nemlig 1/8.

Vindersandsynligheder
Mønstrene 111 og 011. Vi foretager nogle sammenligninger af mønstre. Vi lader A vælge 111 og B vælger 011. Hvis de tre første kast giver krone, vinder A. Men hvis blot ét af de første tre kast giver Plat, så vil A ikke kunne vinde. Efter et 0 vinder B jo på to gange 1, medens A har behov for tre gange 1. Sandsynligheden for at A vinder er altså kun 1/8, og B har en vindersandsynlighed på 7/8.

Mønstrene 110 og 011. Nu vælger A mønstret 110 og B vælger 011. Hvis de første kast giver 1, vinder A. Så snart det første 0 forekommer, foreligger jo mønstret 110. Men hvis 0 forekommer i de to første kast, så kan A's mønster ikke komme før B's. Vindersandsynligheden for A er derfor 1/4 og for B er den 3/4.

Mønstrene 100 og 110. A vælger 100, og B vælger 110. Hvis de to første kast er 11, vil B vinde, idet A's mønster ikke kan komme før B's. Hvis første kast giver 0, er vi tilbage i startpositionen. Vi ser nu nærmere på mulighederne:
Lad nu x være sandsynligheden for at B vinder, og lad y være sandsynligheden for at B vinder når de første to kast giver 10. Vi har da:

$$y = P(11)\cdot 1 + P(10)\cdot y = 1/4 + y/4$$

hvoraf vi får: y=1/3.

For x har vi:

$$x = P(11)\cdot 1 + P(10)\cdot y + P(0)\cdot x = 1/4 + 1/4 \cdot 1/3 + 1/2\cdot x$$

hvoraf vi får: x= 2/3. Vindersandsynligheden for B er altså 2/3.

Af de tre eksempler kan vi aflede resultater for andre mønstre ved at foretage ombytning af 0 og 1. Og vi kan gøre brug af at fx 111 og 000 må have samme vinderchance over for hinanden, altså 1/2. – I en tabel samler vi de resultater der gælder for tre-mønstre. Tabellen angiver B's vinderchance over for A's valg.

	A:111	**A:110**	**A:101**	**A:100**	**A:011**	**A:010**	**A:001**	**A:000**
B:111	--	1/2	2/5	2/5	1/8	5/12	3/10	1/2
B:110	1/2	--	2/3	2/3	1/4	5/8	1/2	7/10
B:101	3/5	1/3	--	1/2	1/2	1/2	3/8	7/12
B:100	3/5	1/3	1/2	--	1/2	1/2	3/4	7/8
B:011	7/8	3/4	1/2	1/2	--	1/2	1/3	3/5
B:010	7/12	3/8	1/2	1/2	1/2	--	1/3	3/5
B:001	7/10	1/2	5/8	1/4	2/3	2/3	--	1/2
B:000	1/2	3/10	5/12	1/8	2/5	2/5	1/2	--

Af tabellen fremgår at B altid ville kunne matche A's valg af mønster: Uanset A's valg vil B kunne vælge et mønster som giver ham en vinderchance på mindst 2/3.

B's bedste valg

Der findes en enkel opskrift som kan hjælpe B til at vælge det mønster der har størst vinderchance over for A's valg af mønster.

Lad A's mønster være:	a,b,c
B's bedste valg er da mønstret:	1-b,a,b

Et eksempel: A vælger 011. B skal da vælge. 1-1,0,1, dvs. 001. Af tabellen ovenfor ser vi at B med dette valg har en vindersandsynlighed på 2/3.

Denne let anvendelige opskrift gælder for tre-mønstre.

For mønstre med fire tegn er sagen lidt mere kompliceret. Lad A's mønster være a,b,c,d. B vælger da: x,a,b,c, hvor x er lig med 1-c. Hvis de to mønstre dermed fremgår af hinanden ved ombytning af 0 og 1, skal x i stedet ændres til c.

Et eksempel: A vælger 1001. B vælger da 1100. De to mønstre fremgår ikke af hinanden ved ombytning af 0 og 1.

Et andet eksempel: A vælger 1010. B vælger i første omgang 0101. Men de to mønstre fremgår af hinanden ved ombytning af 0 og 1. B's valg ændres derfor til: 1101.

Generelt gælder for B's jagt på det bedste mønster: Tegnene i A's mønster skubbes én plads mod højre. Derved er det sidste tegn ude i mørket, og det kommer ikke til direkte at indgå i B's mønster. På den tomme førsteplads sættes et 0 eller 1 afhængigt af hvilket B-mønster der giver den største vinderchance. Der er altså kun behov for at sammenligne to mønstre for at få det rigtige B-mønster fastlagt.

Et eksempel med to mønstre med fem tegn: A vælger mønstret 01101. B har da to mønstre at sammenligne med A's: 00110 og 10110. Vi anvender nu INFA-programmet Talmønster som kan foretage en beregning af de teoretiske vinderchancer for et forelagt mønster over for et andet. Her får vi:

A: 01101	B: 00110	B's vinderchance:	9/13	(69.2%)
A: 01101	B: 10110	B's vinderchance:	13/22	(59.1%)

Det rigtige svar på A's mønster 01101 er altså: 00110.

Kastegennemsnit

Også for tre-mønstre kan der foretages beregninger af kastegennemsnit. Se algoritmen nedenfor. Vi skal her angive resultaterne:

K = 8: 110, 100, 011, 001

K = 10: 101, 010

K = 14: 111, 000

Blandt tre-mønstre er der flere eksempler på at mønstre med samme kastegennemsnit godt kan have forskellige vinder-chancer i kamp mod hinanden. Således har mønstrene 001 og 100 begge et kastegennemsnit på 8, men i en indbyrdes konkurrence har 100 en vinderchance på 3/4. – Også møn-strene 110 og 011 har samme kastegennemsnit, men 011 har en vinderchance på 3/4 over for 110.

Derimod kan vi ikke finde eksempel på at et mønster med et større kastegennemsnit har en vinderchance på over 1/2 over for et mønster med et mindre kastegennemsnit. Men sådanne eksempler findes hvis vi går op til fire-mønstre. Så her er igen tale om mønstre der har egenskaber som ikke stemmer med vor intuition. Her er et eksempel:

```
Statistik over 1000 eksperimenter.
Variabel: Ventetid  Mønster: 0101

   Talområde            Fraktil

     4..4                 1 %
     4..4                 5 %
     4..5                10 %
     4..8                25 %
     4..15               50 %
     4..27               75 %
     4..41               90 %
     4..50               95 %
     4..83               99 %

Gennemsnit:    20.20
Maximum   :   146
Minimum   :     4
```

```
Statistik over 1000 eksperimenter.
Variabel: Ventetid   Mønster: 1011
```

Talområde	Fraktil
4..4	1 %
4..4	5 %
4..5	10 %
4..7	25 %
4..14	50 %
4..24	75 %
4..37	90 %
4..48	95 %
4..62	99 %

```
Gennemsnit:   17.93
Maximum   :   98
Minimum   :    4
```

Model:
Der udtages tal i talområdet 0..1.Der udtages tal indtil et af mønstrene 0101 og 1011 er fundet. Der udføres 1000 eksperimenter.

Mønstret som blev fundet først	Antal eksperimenter
0101	639
1011	361

De to computerkørsler af Talmønster viser at 4-mønstret 0101 har et kastegennemsnit på 20, og mønstret 1011 har et kaste-gennemsnit på 18. Alligevel er det mønstret 0101 der har den største vindersandsynlighed i en konkurrence mellem de to mønstre. De teoretiske værdier af de to vindersandsynligheder er 9/14 = 64.3% og 5/14 = 35.7%

Det kan også forekomme at et fire-mønster kan besejre et tre-mønster. Vi ser på de to mønstre 0111 og 111. En beregning giver resultatet:

```
Teoretisk vinderchance
Mønster A: 111  Mønster B: 0111

A's vinderchance:    1/8 ( 12.5 %)
B's vinderchance:    7/8 ( 87.5 %)
```

Det gælder endda at mønstret 111 har kastegennemsnittet 14 medens mønstret 0111 har et kastegennemsnit på 16.

En algoritme til beregning af K

Der findes en simpel algoritme til beregning af K, det gennemsnitlige kasteantal der kræves for at opnå et forelagt mønster. Vi vil illustrere algoritmen ved et eksempel.

Lad der være forelagt mønstret 1010. Vi udfører da nogle sammenligninger:

Første tegn: 1	Sidste tegn: 0	
Første to tegn: 10	Sidste to tegn: 10	Bidrag: 2^2
Første tre tegn: 101	Sidste tre tegn: 010	
Første fire tegn: 1010	Sidste fire tegn: 1010	Bidrag: 2^4

Hvis der er overensstemmelse ved sammenligningerne, noteres der et bidrag. Hvis der er k tegn som stemmer overens, er bidraget 2^k.

I eksemplet ovenfor bliver det samlede bidrag: 4 + 16 = 20.

Kastgennemsnittet for mønstret 1010 er derfor 20.

Denne algoritme kan også benyttes inden for andre talområder. Ved terningkast hvor talområdet er 1..6, skal bidragene beregnes med en vægt på 6^k. Kastegennemsnittet for mønstret 66 er derfor $6^1 + 6^2 = 42$.

Andre typer af mønstre
Vi skal til afslutning af dette emne se på to eksempler på en anden slags mønstre. Vi vil undersøge hvad ventetiden er for et mønster hvor både 0 og 1 er repræsenteret. I møntkast-sprog: Vi venter på at både Krone og Plat er forekommet i kasteserien.

Eksempler på kasteforløb:

KP;   KKKP; PPPPPK

Vi betegner sandsynligheden for 1 med p. Sandsynligheden for 0 er da: 1– p.

Det kan vises at den gennemsnitlige ventetid E(T) for forekomsten af en kastefølge med begge tegn er:

$$E(T) = \frac{1-p+p^2}{p(1-p)}$$

For p=1/2 får vi: E(T) = 3.

I familieplanlægning kan resultatet oversættes til: Hvis en familie ønsker et barn af hver køn, må de i gennemsnit regne med at skulle have 3 børn.

Gentagelse af første tegn
Vi ser nu på en situation hvor vi venter på at første tegn gentages. I møntkastudgaven drejer det sig fx om:

KPK; PP; PKKKKP

Hvad er mon den gennemsnitlige ventetid på denne situation?

Hvis p=0 eller p=1, vil alle kast have samme resultat, så da vil ventetiden på gentagelsen være 2.

Når p har værdier imellem 0 og 1 kan det vises at den gennemsnitlige ventetid er givet ved:

$$E(T) = 3$$

uanset værdien af p. Man kunne ellers tro at ventetiden som er 2 ved p=0 og p=1, ville stige jævnt fra 2 op til et maximum og derefter aftage igen til 2 når p varierer fra 0 til 1. Men det er ikke tilfældet: Den gennemsnitlige ventetid er altid 3 når p er forskellig fra 0 og 1.

I familieplanlægningsudgaven: Hvor mange børn skal familien i gennemsnit have hvis de ønsker at få et barn med samme køn som det førstefødte? Svaret er: 3.

Register

Tallene henviser til bogens afsnit

Litteratur

F. Barth, R. Haller: *Stochastik Leistungskurs*, Ehrenwirth 1987.

E. Czuber: *Wahrscheinlichkeitsrechnung*, B.G. Teubner 1914.

A. de Moivre*: The Doctrine of Chances*, A. Millar London 1756.

W. Feller: *An Introduction to Probability Theory and Its Applications*, Wiley 1950.

S. Götz: *Ziegen*, *Autos and Bayes – eine never-ending story,* Stochastik in der Schule, Band 26, Heft 1, 2006

C. Huygens: *Om Regning på Lykkespil*, Videnskabshistorisk Museum, Århus 1986.

A. Malmberg: *Håndbog i Sandsynlighed og Statistik*, INFA 2003.

F. Mosteller: *Fifty Challenging Problems in Probability with Solutions,* Addison-Wesley 1965.

L. Rasmussen: *Sandsynlighedsbegrebets oprindelse og præhistorie*, Danmarks Lærerhøjskole 1983.

H. Solomon: *Geometric Probability,* CBMS-NSF Conference Series in Applied Mathematics 1978.

D. Stirzaker: *Elementary Probability*, Cambridge University Press 1994.

J.V. Uspensky: *Introduction to Mathematical Probability,* McGraw-Hill 1937.

CHANCE OG RISIKO henvender sig til matematiklærere i grundskolen som i deres uddannelse har stiftet bekendtskab med fagområdet sandsynlighedsregning og statistik.

Bogens emner behandles i form af udfordrende opgaver som fremlægges med løsninger.

Opgaverne i bogen er udfordrende på to måder. De kan give udfordringer som sætter læserens faglige baggrund på prøve, og de kan give udfordringer til læserens umiddelbare chance- og risikofornemmelse når de fører til resultater som er i klar modstrid med det intuitive.

Af samme forfatter foreligger:

CHANCE – Et it-læremiljø, INFA 2002

Håndbog i sandsynlighed og statistik, INFA 2003

Lær om chancer. Sanne og Malene går på opdagelse med computeren, INFA 2005

Matematik i glimt, INFA 2005