Binomialfordelingen

Billeder:
Forside: ©iStock.com/ginevre
Side 15: ©iStock.com/jaroon
Side 16: ©iStock.com/3pod
Desuden egne fotos og illustrationer.

Indholdsfortegnelse

1. Indledning

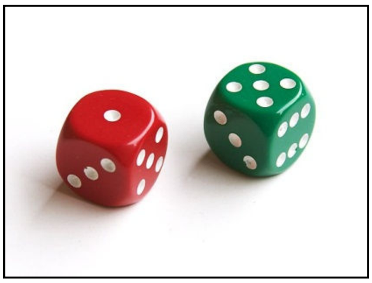
I denne lille note skal vi kigge på *binomialfordelingen*, som er en af de vigtigste fordelinger indenfor sandsynlighedsregning. Før vi går i gang med at beskrive binomialfordelingen, skal vi have nogle grundlæggende begreber på plads. For det første er der begrebet en *stokastisk* variabel, dernæst skal vi kigge på et par vigtige formler indenfor kombinatorikken, der er den gren af matematikken, hvor man lidt løst sagt ”optæller antallet af kombinationer” af et eller andet.

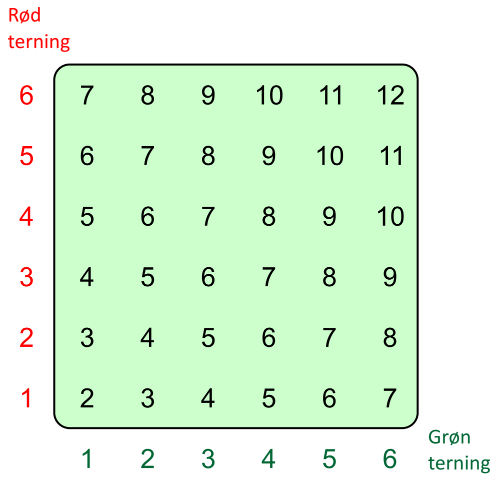
2. Stokastiske variable

Et *stokastisk* eksperiment er et forsøg, hvor man ikke på forhånd kan forudsige udfaldet. Der kan forekomme en række *udfald*, og mængden af disse udgør *udfaldsrummet U*. En *stokastisk variabel X* er en funktion, som til ethvert udfald $u \in U$ knytter et tal $X(u)$. Man er ofte interesseret i *sandsynlighedsfordelingen* for den stokastiske variabel, dvs. sandsynlighederne for de forskellige værdier, som X kan antage. Sandsynlighederne kan angives på lidt forskellig måde. Vi skal se på nogle eksempler.

Eksempel 1

Eksperiment: Et kast med to terninger, en grøn og en rød. Antal øjne betragtes.
Udfaldsrum: $U = \{(1,1), (1,2), \ldots, (1,6), (2,1), (2,2), \ldots, (2,6), \ldots, (6,6)\}$, hvor første koordinaten i talparret angiver antal øjne for den grønne terning, mens andenkoordinaten angiver antal øjne for den røde terning.
Stokastisk variabel: X angiver summen af øjnene af de to terninger.

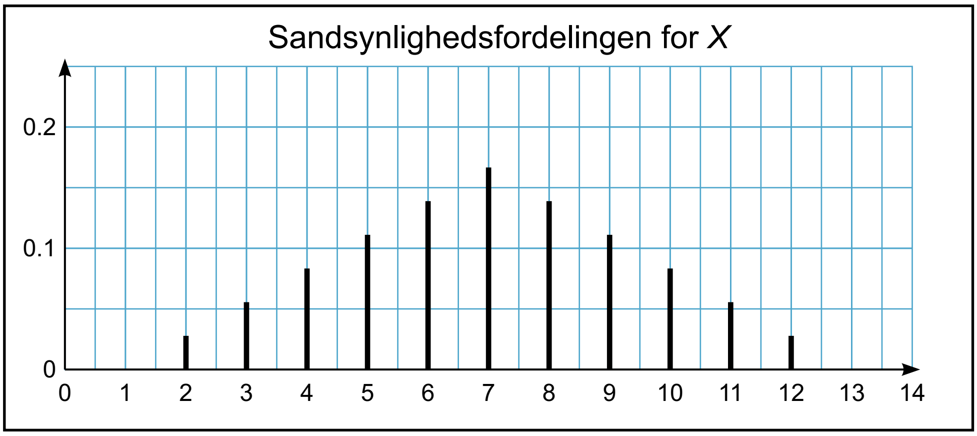




Hvis man med terningerne slår (2,5) svarende til, at den grønne terning viser 2 og den røde terning viser 5, har den stokastiske variabel X værdien $2+5=7$. Af overskuelighedshensyn kan det være hensigtsmæssigt at tegne en figur som ovenfor til højre. Ud for 2 på førsteaksen og 5 på andenaksen er anbragt tallet 7. Tilsvarende med alle de øvrige mulige udfald i udfaldsrummet. Vi ser, at den stokastiske variabel X kan antage værdierne fra 2 til 12. Da den kun kan antage endeligt mange værdier, kaldes X for en *diskret* stokastisk variabel. For at finde *sandsynlighedsfordelingen* for X skal vi beregne sandsynligheden for hver af dens mulige værdier. X har værdien 4 for følgende udfald: $(1,3)$, $(2,2)$ og $(3,1)$. Sandsynligheden for hver af disse er som bekendt $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. Derfor er den ønskede sandsynlighed lig med $\frac{3}{36}$. Vi skriver: $P(X = 4) = \frac{3}{36}$. Vi udtaler det: Sandsynligheden for at X er lig med 4 er lig med $\frac{3}{36}$. Sandsynlighederne for de øvrige værdier af X udregnes tilsvarende og anbringes i en tabel:

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

En kontrol på om man har regnet rigtigt er at alle sandsynlighederne i tabellen skal give 1 tilsammen! Sandsynlighedsfordelingen kan om ønskeligt afbildes i et stolpediagram:



Ved hjælp af sandsynlighedsfordelingen kan man løse forskellige opgaver, for eksempel bestemme sandsynligheden for at summen af terningernes øjne *højst* er lig med 5:

$$
\begin{aligned}
P(X \leq 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\
&= \tfrac{1}{36} + \tfrac{2}{36} + \tfrac{3}{36} + \tfrac{4}{36} = \tfrac{10}{36} = \tfrac{5}{18}
\end{aligned}
$$

□

I eksempel 1 så vi, hvordan vi kunne beregne sandsynligheder, som involverer en stokastisk variabel X, som er tilknyttet et bestemt udfaldsrum. Man kan naturligvis også direkte tale om sandsynlighederne af de enkelte udfald $u_i \in U$, og de betegnes $P(u_i)$. Det er oplagt, at der altid gælder at sandsynlighederne ligger mellem 0 og 1:

(1) $$0 \leq P(u_i) \leq 1$$

samt at sandsynlighederne tilsammen giver 1. At $P(u_1) + P(u_2) + \ldots + P(u_n) = 1$ skriver man ofte på en kompakt måde med et særligt sumtegn:

(2) $$\sum_{i=1}^{n} P(u_i) = 1$$

Det er her underforstået at $U = \{u_1, u_2, \ldots, u_n\}$.

I forbindelse med udfaldsrum taler man også om begrebet en *hændelse*. Med en hændelse H menes en delmængde af udfaldsrummet, som skrives $H \subseteq U$. Man definerer naturligt nok sandsynligheden for hændelsen H som summen af sandsynlighederne af de enkelte udfald i H. Altså hvis $H = \{u_{i_1}, u_{i_2}, \ldots, u_{i_r}\}$, så sættes

(3) $$P(H) = P(u_{i_1}) + P(u_{i_2}) + \ldots + P(u_{i_r})$$

Der skal siges lidt om et par særlige hændelser: Udfaldsrummet er jo en delmængde af sig selv og dermed en hændelse. Det er oplagt, at sandsynligheden for den er 1. I den anden boldgade har vi den delmængde, som ikke indeholder noget udfald overhovedet, den *tomme mængde*, betegnet med det særlige symbol $\varnothing$. Den har sandsynligheden 0. Endelig har vi den *komplementære hændelse til H*, kaldet $\bar{H}$ (udtales H bar). Hermed menes den komplementære mængde til H, der netop består af alle de udfald i U, der *ikke* er i H. Det er her indlysende, at sandsynligheden for den komplementære hændelse fås ved at trække sandsynligheden for hændelsen H fra 1. Vi opsummerer:

(4) $$P(U) = 1, \quad P(\varnothing) = 0 \quad \text{og} \quad P(\bar{H}) = 1 - P(H)$$

Lad os kigge på et eksempel.

Eksempel 2

I eksempel 1 kunne man for eksempel betragte den hændelse H, der består af alle de udfald, hvor den grønne terning viser 3: $H = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$. Da alle udfaldene i U har sandsynligheden $\frac{1}{36}$, får vi af (3):

$$\begin{aligned} P(H) &= P((3,1)) + P((3,2)) + P((3,3)) + P((3,4)) + P((3,5)) + P((3,6)) \\ &= \tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} \\ &= \tfrac{6}{36} = \tfrac{1}{6} \end{aligned}$$

Den komplementære hændelse til H består af alle de udfald, hvor den grønne terning *ikke* viser 3. Sandsynligheden for den er ifølge (4): $P(\bar{H}) = 1 - P(H) = 1 - \frac{1}{6} = \frac{5}{6}$.

□

I det følgende eksempel skal vi se, at man sagtens kan have flere forskellige stokastiske variable, der er knyttet til det samme udfaldsrum. I eksempel 1 var X, som så ofte før set, lig med summen af terningernes øjne. Nu skal vi se et eksempel, hvor kastet med de to terninger betragtes som et spil og hvor X er gevinsten ved spillet.

Eksempel 3

En bankør tilbyder et spil, hvor spilleren slår med to terninger: en grøn og en rød. Hvis der er en 1'er blandt de to terninger, skal spilleren betale 4 kr. til bankøren. I alle andre tilfælde vinder spilleren det beløb i kroner, som svarer til forskellen mellem de to terningers visning. Hvis den ene terning viser 5 og den anden 2, vinder spilleren altså $5 - 2 = 3$ kroner. Lad os være systematiske igen:

Eksperiment: Et kast med to terninger, en grøn og en rød. Antal øjne betragtes.
Udfaldsrum: $U = \{(1,1), (1,2), \dots, (1,6), (2,1), (2,2), \dots, (2,6), \dots, (6,6)\}$, hvor første koordinaten i talparret angiver antal øjne for den grønne terning, mens andenkoordinaten angiver antal øjne for den røde terning.
Stokastisk variabel: X angiver det beløb spilleren vinder i ét enkelt spil.

For at bestemme sandsynlighedsfordelingen for X skal vi først finde ud, hvad X giver for de enkelte udfald i udfaldsrummet. Det er gjort skematisk på figuren nedenfor til venstre. Situationen er set fra spillerens synspunkt, så et tab anføres som et negativt tal. Vi ser, at X kan antage 6 forskellige værdier: –4, 0, 1, 2, 3, og 4. Sandsynligheden for hver af disse værdier fås ved at tælle op, hvor ofte de forekommer i skemaet og så gange med 1/36, som er sandsynligheden for hvert enkelt udfald. Det giver tabellen til højre:

Rød terning						
6	-4	4	3	2	1	0
5	-4	3	2	1	0	1
4	-4	2	1	0	1	2
3	-4	1	0	1	2	3
2	-4	0	1	2	3	4
1	-4	-4	-4	-4	-4	-4
	1	2	3	4	5	6 Grøn terning

Sandsynlighedsfordelingen for X:

x_i	–4	0	1	2	3	4
$P(X = x_i)$	$\frac{11}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

□

3. Middelværdi, varians og spredning

Der er specielt to størrelser, som er med til at sige noget om en stokastisk variabel, og det er *middelværdien* og *variansen* eller *spredningen*. Middelværdien fås ved at tage det *vejede gennemsnit* af værdierne for den stokastiske variabel, mens variansen fås som det vejede gennemsnit af kvadratet på afvigelserne fra middelværdien. Spredningen er kvadratroden af variansen. Lad os se på eksempel 1 og 3 ovenfor.

Eksempel 4

Middelværdien af den stokastiske variabel fra eksempel 1 fås ved at betragte tabellen for sandsynlighedsfordelingen og udregne det vejede gennemsnit:

(5) $$2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{3}{36} + 5 \cdot \tfrac{4}{36} + 6 \cdot \tfrac{5}{36} + 7 \cdot \tfrac{6}{36} + 8 \cdot \tfrac{5}{36} + 9 \cdot \tfrac{4}{36} + 10 \cdot \tfrac{3}{36} + 11 \cdot \tfrac{2}{36} + 12 \cdot \tfrac{1}{36}$$

Denne udregning giver 7. Resultatet er ikke overraskende, da fordelingen er symmetrisk omkring 7. Middelværdien betegnes nogle gange med μ, andre gange med $E(X)$. Sidstnævnte står for "expectation of X" – den forventede værdi af X. Man skriver udregningen (5) på en lidt smart måde ved hjælp af et sumtegn:

(6) $$\mu = E(X) = \sum_{i=1}^{n} x_i P(X=x_i) = 2\cdot\tfrac{1}{36}+3\cdot\tfrac{2}{36}+4\cdot\tfrac{3}{36}+\ldots+12\cdot\tfrac{1}{36} = 7$$

Nu til variansen. På kompakt form skriver vi:

(7) $$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{n}(x_i-\mu)^2\cdot P(X=x_i)\\ &= (2-7)^2\cdot\tfrac{1}{36}+(3-7)^2\cdot\tfrac{2}{36}+(4-7)^2\cdot\tfrac{3}{36}+\ldots+(12-7)^2\cdot\tfrac{1}{36} = 5{,}83\end{aligned}$$

Vi tager altså forskellene mellem den stokastiske variabels værdier og middelværdien, opløfter til 2. potens og vejer med sandsynlighederne. Spredningen fås ved at tage kvadratroden af variansen:

(8) $$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{5{,}83} = 2{,}52$$

Selv om talværdien for spredningen ikke siger noget direkte om fordelingen, så kan man dog sige, at jo større tallet er, jo mere spredte er data.

□

Lad os sammenfatte begreberne ovenfor:

Definition 5 (Middelværdi, varians og spredning for stokastisk variabel)
Givet en diskret stokastisk variabel X. Middelværdien for X er givet ved:

(9) $$\mu = E(X) = \sum_i x_i P(X=x_i)$$

hvor der summeres over de mulige værdier $x_1, x_2, \ldots$, som X kan antage. Variansen for X er givet ved:

(10) $$\mathrm{Var}(X) = \sum_i (x_i-\mu)^2 P(X=x_i)$$

hvor μ angiver middelværdien for X. Spredningen for X er givet ved:

(11) $$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

Eksempel 6

I spillet i eksempel 3 får vi følgende værdier for henholdsvis middelværdi og varians for den stokastiske variabel X:

$$E(X) = \sum_i x_i P(X=x_i) = -4\cdot\tfrac{11}{36} + 0\cdot\tfrac{5}{36} + 1\cdot\tfrac{8}{36} + 2\cdot\tfrac{6}{36} + 3\cdot\tfrac{4}{36} + 4\cdot\tfrac{2}{36} = -\tfrac{1}{9}$$

$$\begin{aligned}\mathrm{Var}(X) &= \sum_i (x_i-\mu)^2 P(X=x_i)\\ &= (-4-(-\tfrac{1}{9}))^2\cdot\tfrac{11}{36} + (0-(-\tfrac{1}{9}))^2\cdot\tfrac{5}{36} + (1-(-\tfrac{1}{9}))^2\cdot\tfrac{8}{36}\\ &\quad + (2-(-\tfrac{1}{9}))^2\cdot\tfrac{6}{36} + (3-(-\tfrac{1}{9}))^2\cdot\tfrac{4}{36} + (4-(-\tfrac{1}{9}))^2\cdot\tfrac{2}{36}\\ &= 7{,}65\end{aligned}$$

Vi ser, at middelværdien er negativ, hvilket ikke er så mærkeligt, da bankører har det med at sørge for, at de selv har de bedste odds! Helt præcist fortæller middelværdien, at spilleren i gennemsnit vil *tabe* 1/9 kr. i hvert spil. Talværdien for variansen er ikke nem at give en god fortolkning af, men vi kan lave lidt om på spilreglerne og studere den virkning, som det har på variansen. Lad os sige, at spilleren stadig taber 4 kr. hvis blot en af terningerne viser 1, at to ens giver summen af øjnene i kr., undtagen hvis de to ens er (1,1), mens alle andre kombinationer hverken giver tab eller gevinst. Situationen er vist på figuren nedenfor til venstre. Middelværdien viser sig at være nøjagtig den samme som i det oprindelige spil, men variansen kan vises at være vokset til 14,87. Det skyldes, at spillet er blevet mere chancebetonet. Der er større præmier, som er fordelt på færre udfald. Men bankøren vil altså i gennemsnit få den samme indtjening!

Rød terning						
6	-4	0	0	0	0	12
5	-4	0	0	0	10	0
4	-4	0	0	8	0	0
3	-4	0	6	0	0	0
2	-4	4	0	0	0	0
1	-4	-4	-4	-4	-4	-4
	1	2	3	4	5	6 Grøn terning

□

4. Lidt kombinatorik

Som hjælperedskab til næste afsnit får vi brug for lidt kombinatorik. *Kombinatorik* er den matematiske disciplin, som beskæftiger sig med at tælle antal kombinationer.

Definition 7

Antag givet n forskellige elementer opstillet på en række. Med betegnelsen en *permutation* af de n elementer menes en ombytning af elementerne, så de står i en ny (evt. samme) rækkefølge. Antallet af mulige permutationer af n forskellige elementer betegnes $n!$ og udtales ”n fakultet”.

På figuren nedenfor er vist permutationer af tallene fra 1 til 7. Da der er i alt 5040 forskellige permutationer, kan vi kun opskrive nogle af dem.

Permutationer $7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$

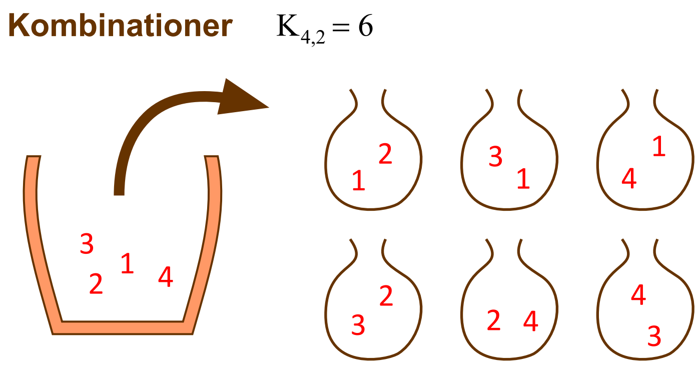
1 2 3 4 5 6 7 , 2 1 3 4 5 6 7 , 2 3 1 4 5 6 7 ,

… , 4 6 1 2 5 7 3 , … , 7 6 5 4 3 2 1

Definition 8

En anden vigtig størrelse er *kombinationer*. Med betegnelsen $K_{n,r}$ vil vi mene antallet af måder, hvorpå man kan udtage r elementer ud af en mængde på n forskellige elementer (antaget $0 \le r \le n$).

Situationen er illustreret nedenfor: Der skal udtages 2 tal fra en krukke med de fire tal 1, 2, 3 og 4. Det kan gøres på 6 forskellige måder som vist. Derfor er $K_{4,2} = 6$. De udtrukne tal er vist i "poser" for at indikere, at man ikke interesserer sig for, i hvilken rækkefølge tallene udtrækkes.

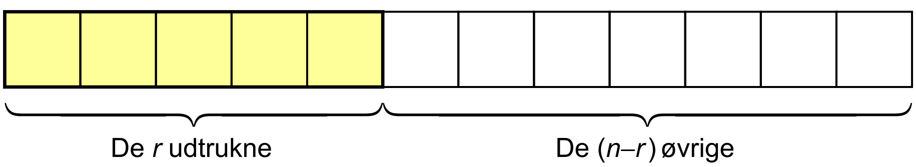


Sætning 9 (Permutationer og kombinationer)

a) $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$

b) $K_{n,r} = \dfrac{n!}{r!(n-r)!}$

Bevis: a) Givet n forskellige elementer. Den første plads kan besættes på n måder. For hver af disse tilfælde kan plads 2 besættes på $(n-1)$ måder. For hver af disse kan plads 3 besættes på $(n-2)$ måder, osv. … indtil den sidste plads, som kun kan besættes på 1 måde. Antallet af muligheder fås dermed ved at gange disse tal sammen.



b) Man kan betragte problemet på følgende måde: Vælg en tilfældig permutation af de n elementer og lad os vedtage, at de r elementer, som står først, skal være de udtrukne. Der er imidlertid en masse tilfælde, som tælles med flere gange. De første r elementer kan permuteres på $r!$ forskellige måder, men de repræsenterer den samme udtrækning. Vi må derfor dividere med $r!$ for at tage høje for de permutationer, som tælles med flere gange. På samme måde vil permutationer af de $n-r$ øvrige elementer heller ikke resultere i nogen ny udtrækning, hvorfor vi må dividere med $(n-r)!$ for at tage højde for de tilfælde, som tælles med flere gange. Antallet af mulige udtrækninger fås derfor til:

$$K_{n,r} = \frac{n!}{r!(n-r)!}$$

□

Eksempel 10

I eksemplet på den første figur i afsnit 4 har vi, at antallet af permutationer af de 7 elementer er $7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$. I eksemplet på den anden figur har vi, at man kan udtage 2 elementer ud af en krukke med 4 forskellige elementer på 6 måder:

$$K_{4,2} = \frac{4!}{2!(4-2)!} = \frac{4!}{2! \cdot 2!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 1 \cdot 2} = 6$$

□

Fakultet og kombinationer i Maple og på TI-30X lommeregneren

Man behøver ikke udregne fakultet og kombinationer manuelt ved hjælp af formlerne i sætning 9. De fleste lommeregnere har funktionerne indbygget. I TI-30X kan man for eksempel benytte en knap med navnet PRB, så vil der vise sig nogle muligheder. I den fremkomne menu er udråbstegnet ! den såkaldte fakultetsfunktion, mens `nCr` er kombinations-funktionen. Angående sidstnævnte, så udregnes $K_{4,2}$ ved at trykke `4 nCr 2` og derefter **ENTER**. I Maple udregnes $K_{4,2}$ ved hjælp af kommandoen `binomial(4,2)`. For at udregne størrelsen 6! i Maple skriver man blot 6! - nemmere bliver det ikke!

Eksempel 11

Terne skal arrangere en badminton turnering med 6 personer, hvor alle skal spille mod alle to og to. Hvor mange kampe skal der arrangeres? *Svar*: Det er klart, at det svarer til antallet af mulige måder at udtrække 2 personer ud af 6 personer, hvilket er $K_{6,2} = 15$.

Eksempel 12

Hvor mange forskellige hænder med 13 kort kan man få ud af et kortspil med 52 kort? Svaret er $K_{52,13} = 635013559600$.

Bemærkning 13

I stedet for $K_{n,r}$ bruges ofte betegnelsen $\binom{n}{r}$, og sidstnævnte udtales: "n over r"

5. Binomialfordelingen

En af de mest anvendte sandsynlighedsfordelinger er den såkaldte *binomialfordeling*. At den anvendes så ofte, har sin årsag i, at den er defineret ud fra nogle abstrakte kriterier, som gør den så generel, at den passer på mange situationer. En binomialfordelt stokastisk variabel X er defineret ved følgende:

Eksperiment: Et *basiseksperiment* udføres n gange. n kaldes for *længden* af eksperimentet. Et basiseksperiment kan *lykkes* eller *mislykkes*. Udfaldene af hver af basiseksperimenterne skal være indbyrdes *uafhængige*. Sandsynligheden for, at et enkelt basiseksperiment lykkes betegnes p, hvormed sandsynligheden for at det mislykkes er $1-p$.
Udfaldsrum: Et udfald u kan beskrives ved en vektor med n koordinater. Den i'te koordinat er lig L, hvis det i'te basiseksperiment lykkes og M, hvis det mislykkes. Eksempel: Hvis $n=5$, er $(\mathrm{L},\mathrm{M},\mathrm{M},\mathrm{L},\mathrm{M})$ det udfald, at første og fjerde basiseksperiment lykkes, mens de øvrige tre mislykkes. Mængden af alle udfald udgør udfaldsrummet.
Stokastisk variabel: X angiver det antal gange, som basiseksperimentet lykkes.

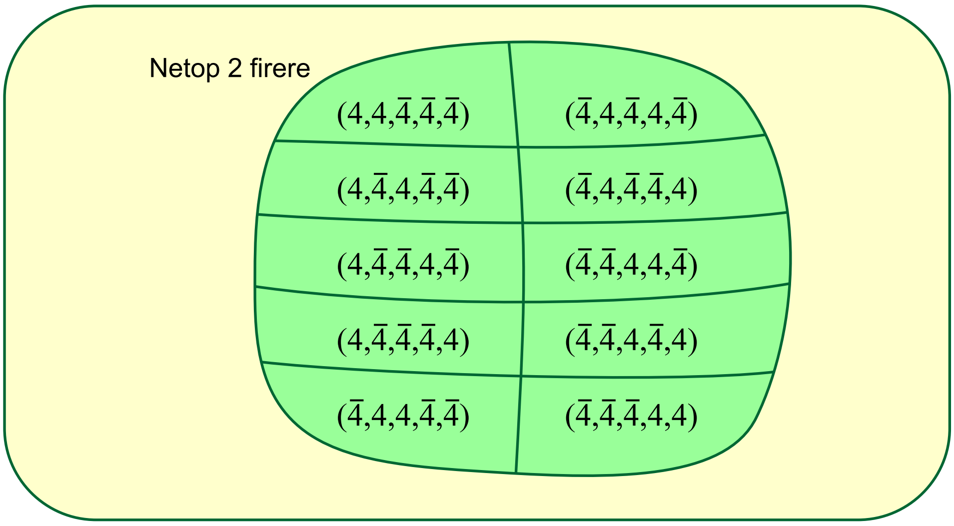
Det generelle i definitionen gør som sagt, at fordelingen finder anvendelse i vidt forskellige situationer: Det kan være, at man slår 8 gange med en terning og vil undersøge sandsynligheden for at få 3 femmere. Det kan være, at man ønsker at finde sandsynligheden for at få højst 2 piger ved 4 fødsler, eller måske vil man undersøge usikkerheden af stikprøver i forbindelse med folketingsvalg.

Vi skal senere se, hvorfor disse eksempler opfylder kriterierne. Først skal vi dog opstille en formel for sandsynligheden for, at basiseksperimentet lykkes r gange:

Sætning 14 (Binomialfordelingen)
Lad X være en binomialfordelt stokastisk variabel. Sandsynligheden for, at X antager værdien r er givet ved følgende udtryk:

(12) $$P(X=r)=K_{n,r}\cdot p^r\cdot(1-p)^{n-r}$$

Bevisskitse: Af hensyn til overskueligheden vil vi argumentere for ovenstående formel ved hjælp af et eksempel: Lad os antage, at vi slår 5 gange med en terning og ønsker at finde sandsynligheden for at få netop 2 firere. Basiseksperimentet er da at slå én gang med en terning. Vi vedtager, at basiseksperimentet lykkes, hvis man får en firer. Det giver en basissandsynlighed på $p=\frac{1}{6}$. Sandsynligheden for at basiseksperimentet mislykkes, dvs. at resultatet ikke er en firer, er dermed $1-p=\frac{5}{6}$. Den stokastiske variabel X skal angive antallet af gange basiseksperimentet lykkes, dvs. hvor mange gange man får en firer i $n=5$ kast. Netop 2 firere kan fremkomme på forskellig vis: Det er hensigtsmæssigt at dele op i en række hændelser. På figuren nedenfor symboliserer skrivemåden $(4,4,\overline{4},\overline{4},\overline{4})$, at man fik firere i de to første kast, og *ikke-firere* i de næste tre kast. Man kan også tænke, at man fik firere i første og tredje kast og ikke-firere i de øvrige, etc.



Der er i alt 10 muligheder, kan vi se, og de udelukker indbyrdes hinanden samtidigt med, at de udgør alle tilfælde med netop 2 firere. Sandsynligheden for netop to firere kan da findes ved at lægge sandsynligheden for hver af de 10 muligheder sammen. Lad os først bestemme sandsynligheden for $(4,4,\overline{4},\overline{4},\overline{4})$:

$$P(4,4,\overline{4},\overline{4},\overline{4}) \;=\; \tfrac{1}{6}\cdot\tfrac{1}{6}\cdot\tfrac{5}{6}\cdot\tfrac{5}{6}\cdot\tfrac{5}{6} \;=\; \left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^3$$

Hvor vi har benyttet, at de enkelte kast er uafhængige af hinanden, så vi får sandsynligheden for hele kombinationen ved at gange deres indbyrdes sandsynligheder sammen. Vi kan gøre det samme med hændelsen $(4,\overline{4},4,\overline{4},\overline{4})$:

$$P(4,\overline{4},4,\overline{4},\overline{4}) \;=\; \tfrac{1}{6}\cdot\tfrac{5}{6}\cdot\tfrac{1}{6}\cdot\tfrac{5}{6}\cdot\tfrac{5}{6} \;=\; \left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^3$$

Vi ser, at resultatet er det samme som før. Man overbeviser sig hurtigt om, at alle disse sandsynligheder er ens, nemlig $\left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^3$. Derfor fås sandsynligheden for netop 2 firere ved at gange $\left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^3$ med antallet af kombinationer:

$$P(X=2) \;=\; 10\cdot\left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^3 \;=\; K_{5,2}\cdot\left(\tfrac{1}{6}\right)^2\cdot\left(\tfrac{5}{6}\right)^{5-2}$$

idet $K_{5,2}=10$ er antallet af kombinationer, svarende til antallet af måder, hvorpå man kan udpege de to pladser, hvorpå firerne skal stå. Resten af pladserne skal indeholde ikke-firere. Vi ser, at resultatet stemmer med (12) for $n=5,\, r=2,\, p=\tfrac{1}{6}$.

□

Eksempel 15

Bestem sandsynligheden for ved 8 kast med en terning at få netop 3 femmere. Angiv desuden hele sandsynlighedsfordelingen for den stokastiske variabel X, som angiver antal femmere ved de 8 kast.

Løsning: Basiseksperimentet er ét kast med en terning, og det udføres $n=8$ gange. Udfaldene af de enkelte basiseksperimenter er klart uafhængige. Dermed kan vi benytte binomialfordelingen. Basissandsynligheden er $p=\tfrac{1}{6}$. Ved brug af formel (12) fås:

$$P(X=3) \;=\; K_{8,3}\cdot\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^5 \;=\; 56\cdot\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^5 \;=\; 0{,}1042$$

Så den søgte sandsynlighed er altså 10,42%. Opgaven kan løses hurtigere i Maple:

restart
with(*Statistics*) :
$X := RandomVariable\left(Binomial\left(8, \frac{1}{6.}\right)\right)$:
ProbabilityFunction(X, 3) = 0.1041904817

Man skal kalde pakken *Statistics*. Herefter har man rådighed over en række funktioner. *RandomVariable* er det engelske ord for stokastisk variabel. I tredje linje defineres således en binomialfordelt stokastisk variabel med $n = 8$ og $p = \frac{1}{6}$. Bemærk at "kommafidusen" med at skrive et af tallene som et kommatal – her et sat et punktum efter 6-tallet i brøken 1/6 – sikrer at vi får svarene i kommatal og ikke eksakte brøker! Dette var $P(X = 3)$. På helt tilsvarende måde kan man udregne sandsynligheden for alle de øvrige mulige værdier, som den stokastiske variabel kan antage, og anbringe resultaterne i et skema. Hermed haves den såkaldte *sandsynlighedsfordeling* for X:

x	0	1	2	3	4	5	6	7	8
$P(X = x)$	0,2326	0,3721	0,2605	0,1042	0,0261	0,0042	0,0004	0,0000	0,0000

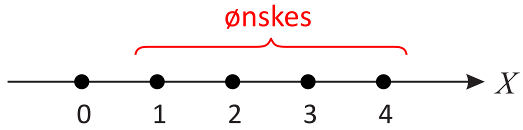
Eksempel 16

Det er en udbredt opfattelse blandt folk, at hvis man får mange børn af samme køn, så er det en ekstrem hændelse. Men er det nu det? Lad os se på eksempler. Hvad er sandsynligheden for ved 4 fødsler at få: a) fire piger?, b) mindst en pige?, c) højst to piger?

Løsning: Basiseksperimentet er en fødsel. Det oplyses, at det at få en pige eller en dreng kan betragtes som uafhængige hændelser. Hermed menes, at sandsynligheden for at få en dreng eller pige er helt uafhængigt at hvad kønnet på eventuelt tidligere børn måtte have været. Statistikker viser, at piger forekommer en smule sjældnere end drenge, nemlig i 49% af tilfældene. Lad os vedtage, at basiseksperimentet lykkes, hvis det bliver en pige, dvs. basissandsynligheden er $p = 0,49$. Lad X angive antallet af piger. Det er klart at $n = 4$, da basiseksperimentet udføres 4 gange.

a) $P(X = 4) = K_{4,4} \cdot 0,49^4 \cdot (1-0,49)^{4-4} = K_{4,4} \cdot 0,49^4 \cdot 0,51^0 = 0,49^4 = 0,0576$, så fire piger ved 4 fødsler sker altså trods alt i knap 6% af alle tilfælde!

b) Hvis der skal være mindst én pige, skal X altså være lig med 1, 2, 3 eller 4. Problemet kan dermed løses ved at udregne $P(X=1)+P(X=2)+P(X=3)+P(X=4)$.

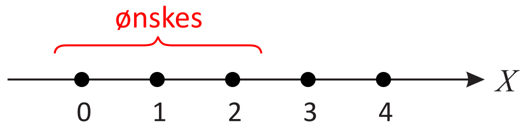


Der er imidlertid en nemmere måde: Det modsatte af at få mindst 1 pige er, at man ingen piger får (komplementære hændelse). Derfor kan opgaven også løses ved:

$$P(X \geq 1) \;=\; 1-P(X=0) \;=\; 1-K_{4,0}\cdot 0,49^0 \cdot 0,51^4 \;=\; 1-0,51^4 \;=\; 0,932$$

så sandsynligheden for at få mindst én pige ved 4 fødsler er altså 93,2%.

c) Hvis man skal have højst 2 piger, så skal X være lig med 0, 1 eller 2. Opgaven kan altså løses ved at udregne $P(X \leq 2) = P(X=0)+P(X=1)+P(X=2)$.



Det er dog hurtigere at bruge Maples facilitet for *kumulerede binomial-sandsynligheder*: *Cumulative Distribution Function*, forkortet *CDF*, til at udregne sandsynligheden direkte. Med den kan man udregne summen af alle sandsynligheder fra en værdi og nedefter. Her er det fra 2 og nedefter:

```
restart
with(Statistics) :
X := RandomVariable(Binomial(4, 0.49)) :
CDF(X, 2) = 0.7023480300
```

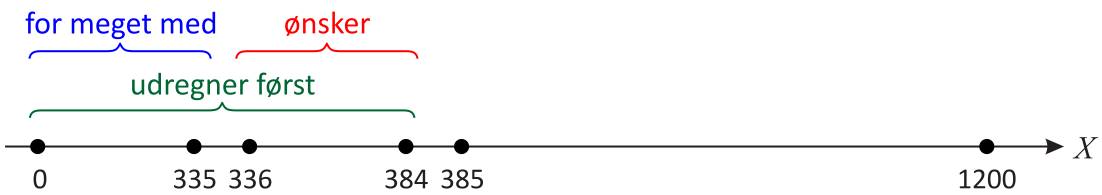
Så sandsynligheden for at få højst 2 piger er altså 70,2%.

□

Eksempel 17

Op til folketingsvalg foretages mange valgprognoser. Typisk udspørges mellem 1000 og 1500 personer om, hvilket parti de vil stemme på. I dette eksempel skal vi forsøge at få en fornemmelse for, hvor usikre sådanne stikprøver er. Vi vil antage, at man udspørger 1200 helt tilfældigt udvalgte personer. Lad os antage, at et parti A på landsplan har 30% af stemmerne, uden at man kender tallet. Hvad er da sandsynligheden for, at en stikprøve på 1200 vil give et resultat, som afviger med mere end 2 procentpoint fra det rigtige tal?

Løsning: Vi kan benytte binomialfordelingen: Basiseksperimentet er, at man udspørger én person. Basiseksperimentet lykkes, hvis personen stemmer på A. Vi kan antage, at personerne stemmer uafhængigt, og basissandsynligheden er $p = 0{,}30$. Den stokastiske variabel X angiver, hvor mange af de 1200 personer, som stemmer på partiet. Lad os først prøve at bestemme sandsynligheden for det modsatte af det der spørges om, dvs. at stikprøvens resultat afviger med *højst* 2% point fra det rigtige, som er 30%. Vi skal altså finde sandsynligheden for at stikprøvens resultat er mindst 28% og højst 32%. Ud af 1200 svarer det til at bestemme sandsynligheden for at der er mindst 336 og højst 384, som stemmer på A, altså $P(336 \le X \le 384)$. De redskaber, vi har til rådighed, stiller os ikke i stand til at bestemme denne værdi direkte. Vi må gøre det i to trin: Først bestemmes sandsynligheden for at X er lig med 384 eller derunder. Så har vi fået for meget med, nemlig sandsynligheden for at X er lig med 335 eller derunder. Dette er illustreret på figuren herunder. Vi har: $P(336 \le X \le 384) = P(X \le 384) - P(X \le 335)$.



I Maple kan udregningen klares således:

```
restart
with(Statistics) :
X := RandomVariable(Binomial(1200, 0.30)) :
CDF(X, 384) − CDF(X, 335) = 0.8773176689
```

Vi har altså $P(336 \le X \le 384) = 0{,}877$. Vi blev imidlertid spurgt om det modsatte: Sandsynligheden for at stikprøven afviger med *mere* end 2% point fra den rigtige procent (populationens) er altså $1 - 0{,}877 = 0{,}123 = 12{,}3\%$. Man skal med andre ord være forsigtig med at konkludere for håndfast ud fra stikprøver. Her gik vi endda ud fra, at gruppen på 1200 var helt tilfældigt udvalgt. Er der en skævhed i sammensætningen, vil det forøge usikkerheden.

Bemærkning 18

Til den opmærksomme læser skal det lige tilføjes, at man egentligt principielt ikke kan anvende binomialformlen i eksempel 17, hvis man skal være nøjeregnende: Når en person har afgivet en stemme, så er sandsynligheden for, at den næste person stemmer på partiet A ikke helt 30%, for en person er jo fjernet fra gruppen! Ud af en stor population, som den danske befolkning er, har dette dog ingen praktisk betydning. Denne problematik er årsagen til, at man undertiden siger, at binomialfordelingen gælder for situationer *med tilbagelægning*.

I det følgende skal vi kigge på middelværdien og variansen for en binomialfordelt stokastisk variabel. Netop i dette tilfælde gælder der nogle ekstra simple og smukke formler, så vi undgår at skulle gå tilbage til de generelle definitioner (9) og (10):

Sætning 19

Lad X være en binomial-fordelt stokastisk variabel med antalsparameter n og basissandsynlighed p. Da er middelværdien og variansen givet ved følgende formler:

a) $E(X) = n \cdot p$

b) $\mathrm{Var}(X) = n \cdot p \cdot (1-p)$

Bevis: Overspringes, da de er lidt tekniske.

□

Eksempel 20

I eksempel 15, hvor der kastes 8 gange med en terning, fås $E(X) = n \cdot p = 8 \cdot \frac{1}{6} = 1{,}33$. Der vil altså i gennemsnit forekomme 1,33 femmere ved 8 kast med en terning. Variansen udregnes til $\mathrm{Var}(X) = n \cdot p \cdot (1-p) = 8 \cdot \frac{1}{6} \cdot \frac{5}{6} = 1{,}11$.

Eksempel 21

Betragt eksempel 17: I stikprøver á 1200 respondenter er middeltallet for antallet af stemmer på parti A ikke overraskende givet ved: $E(X) = n \cdot p = 1200 \cdot 0{,}30 = 360$, hvilket netop svarer til 30% af 1200!

Bemærkning 22 (Uafhængighed)

Som nævnt er uafhængigheden af de enkelte basiseksperimenter en vigtig forudsætning for, at vi har at gøre med en binomialfordeling. Derfor skal man være påpasselig med at overveje, om denne forudsætning kan siges at være opfyldt i den enkelte situation. Ved man statistisk over en lang periode, at Brøndbys fodboldhold plejer at vinde 70% af deres kampe, og man vil vurdere sandsynligheden for, at holdet vinder to af de næste tre kampe, så skal man passe på. For er holdet inde i en god stime, så ved man godt, at sandsynligheden for gevinst øges i næste kamp, fordi selvtilliden er i top. Udfaldene er med andre ord ikke uafhængige!

Opgaver

Nedenstående opgaver er nummereret, så første ciffer angiver det afsnit, som opgaven hører til. Således er opgave 21 opgave 1 i afsnit 2.

Opgave 20

Betragt følgende eksperiment: Et kast med to terninger, en grøn og en rød. Antal øjne betragtes. Lad den stokastiske variabel X angive *forskellen* på øjnene af de to terninger. Benyt idéerne i eksempel 1 til at løse følgende opgaver:

a) Bestem $P(X=3)$.
b) Hvilke værdier kan X antage? Bestem sandsynlighedsfordelingen for X.
c) Bestem sandsynligheden for at terningernes differens højst er 2, dvs. $P(X \leq 2)$.

Opgave 21

Eksperimentet er et kast med tre mønter. Man interesserer sig for, om en mønt viser plat eller krone. Det er en pædagogisk hjælp at tænke på, at mønterne er nummererede. Lad for eksempel (k, p, k) betyde, at mønt 1 viser krone, mønt 2 viser plat og mønt 3 krone.

a) Opskriv de 8 mulige udfald. Overvej hvorfor de er lige mulige?

I det følgende lader vi X være den stokastiske variabel, som angivet antallet af plat ved det pågældende kast med tre mønter.

b) Angiv de mulige værdier for X og bestem sandsynlighedsfordelingen for X.

Opgave 22

Bestem sandsynligheden for en sum på mindst 4 og højst 6 ved ét kast med to terninger?

Opgave 30

En stokastisk variabel X kan antage værdierne 0, 1, 2, 3, 4 og 5 og har følgende sandsynlighedsfordeling:

x_i	0	1	2	3	4	5
$P(X=x_i)$	0,15	0,10	0,30	0,10	0,20	0,15

a) Bestem middelværdien $E(X)$.
b) Bestem variansen $Var(X)$ og spredningen $\sigma(X)$.

Opgave 31

En bookmaker foreslår følgende spil til en spiller: Ét spil består i at kaste tre gange med en mønt. Hvis spilleren udelukkende får plat, skal han betale 20 kr. Hvis spilleren får krone to gange lige efter hinanden, vinder han 5 kr. I alle andre tilfælde vinder spilleren 1 kr. Er det et fornuftigt spil for spilleren i det lange løb?

Hjælp: Opskriv først de 8 mulige udfald ved ét spil, dvs. tre kast – husk at rækkefølgen er væsentlig! Indfør en stokastisk variabel X, som skal være gevinsten ved ét spil. Hvilke værdier kan X antage? Bestem sandsynligheden for hver af disse værdier ved at betragte listen med de 8 mulige udfald. Da du således har sandsynlighedsfordelingen for X, kan du afgøre spørgsmålet ved at udregne middelværdien $E(X)$.

Opgave 32

Nogle børn laver en spillebule. Spillet består i at kaste med to terninger som i eksempel 8. Det vedtages, at boden vil udbetale 8 kr. hvis summen af øjnene er mindst 8, 3 kr., hvis summen er 6 eller 7, mens spilleren må punge ud med 15 kr., hvis summen af øjnene er 5 og derunder.

a) Udregn en sandsynlighedsfordeling for X.
b) Bestem middelværdien for X, dvs. $E(X)$.
c) Er spillet fordelagtig for boden eller spilleren?

Opgave 40

I en fodboldtrup er der 15 spillere. Der ses i det følgende bort fra de specifikke pladser, som spillerne kan spille.

a) På hvor mange måder kan man udtage de 11 spillere, som skal spille den dag? (man tager ikke hensyn til spillernes placering på banen).
b) Samme spørgsmål, hvor man tager hensyn til spillernes placering på banen.

Opgave 41

Lotto er et spil, hvor en spiller kan afkrydse 7 tal på en liste med tallene fra 1 til 36. Ved udtrækning udtrækkes 7 numre tilfældigt fra en tromle.

a) På hvor mange måder kan man udtrække 7 numre ud af 36? Hvad er sandsynligheden dermed for at få 7 rigtige?
b) Hvor mange måder er der at få 6 rigtige på? (Benyt multiplikationsprincippet!)

Opgave 42

Hvor mange måder kan man udtage 5 kort af et spil på 52 kort?

Opgave 50

Løs nedenstående to blandede opgaver.

a) Bestem sandsynligheden for at få netop 4 toere ved 20 slag med en terning.
b) Hvad er sandsynligheden for at få mindst 11 rigtige på en tipskupon, hvis man benytter ”sypigetips”, dvs. man sætter krydset tilfældigt? Det oplyses, at der er i alt 13 kampe på kuponen. *Hjælp*: Benyt binomialformlen: Hvad er basiseksperimentet og basissandsynligheden?

Opgave 51

Man regner med, at 5% af drengene i Danmark er farveblinde. Benyt binomialfordelingen til at løse nedenstående opgaver. Redegør samtidigt for, hvorfor fordelingen kan benyttes i den aktuelle situation. Vi betragter en klasse med 25 drenge.

a) Hvad er sandsynligheden for, at der netop er 1 farveblind dreng i klassen?
b) Hvad er sandsynligheden for, at der i klassen findes mindst én dreng, som er farveblind?
c) På hele skolen er der 300 drenge. Hvad er sandsynligheden for, at der er mere end 25 farveblinde drenge på skolen?
d) Hvad er det gennemsnitlige antal farveblinde drenge på en skole af den nævnte størrelse?

Opgave 52

I den danske befolkning er der følgende fordeling af blodtyper:

	Rhesus positiv	Rhesus negativ
A	37%	7%
B	8%	2%
0	35%	6%
AB	4%	1%

Der udtages en tilfældig gruppe på 50 personer fra den danske befolkning.

a) Hvad er sandsynligheden for, at der netop er fem B Rhesus positive i gruppen?
b) Hvad er middelværdien for antallet af B Rhesus positive i gruppen?
c) Bestem sandsynligheden for, at der er mindst seks med blodtype 0 Rhesus negativ.

Opgave 53 (svær)

Hvor mange gange skal man kaste en mønt for at være 99% sikker på at få mindst én krone? *Hjælp*: Hvad er det modsatte af at få mindst én krone? Hvad skal sandsynligheden for at dette forekommer da være?

Opgave 54 (Chevalier de Mérés problem)

Den franske adelsmand og gambler *Chevalier de Méré* studerede følgende to hændelser:

1) At få mindst én sekser ved fire kast med *en* terning.
2) Mindst én gang at få en dobbelt-sekser ved 24 kast med *to* terninger.

Selv om han havde en anelse om, hvilken hændelse, der var den mest sandsynlige, kunne de Méré ikke redegøre for det. Derfor henvendte han sig til den store franske matematiker *Blaise Pascal* (1623 – 1662). Pascals svar bekræftede Mérés formodning. I det følgende skal du prøve selv at bestemme de to sandsynligheder. *Hjælp*: Se på den omvendte eller modsatte hændelse og benyt: $P(X \geq 1) = 1 - P(X = 0)$.

Opgave 55

På en fabrik produceres der en komponent til en bil. Det vides erfaringsmæssigt, at 12% af komponenterne har en defekt. Der udtages på tilfældig vis 50 komponenter.

a) Hvad er sandsynligheden for, at netop 7 komponenter har defekten?
b) Hvad er sandsynligheden for at få mindst 6 og højst 8 defekte komponenter?
c) Hvad er sandsynligheden for at mindst 4 komponenter er defekte?
d) Hvad er det gennemsnitlige antal defekte komponenter i stikprøven?
e) Hvor stor er spredningen?

Opgave 56

I en multiple-choice prøve er der 25 spørgsmål med hver 5 mulige svar. Antag, at du ikke ved et klap om emnet og svarer helt tilfældigt. Hvad er da sandsynligheden for at få mindst 10 rigtige i prøven? Hvad er middeltallet for antal rigtige svar?