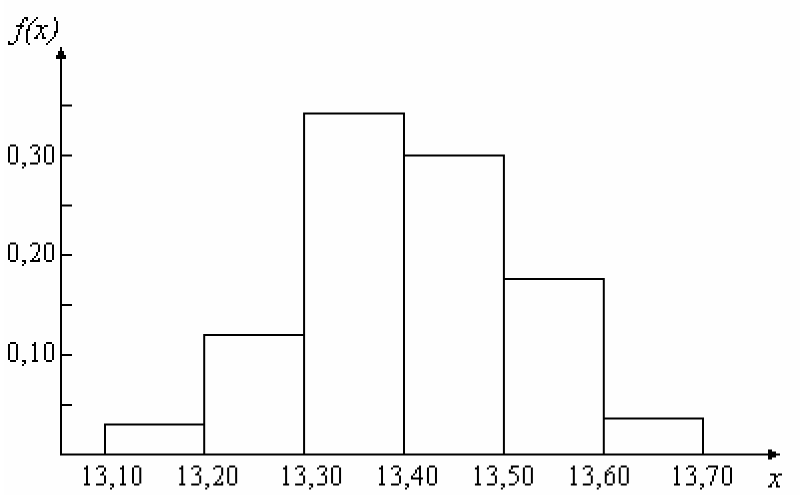
Matematikkens mysterier
- på et obligatorisk niveau

af

Kenneth Hansen

9. Sandsynlighedsregning



Hvad er den typiske størrelse af et nittehoved ?

9. Statistik og sandsynlighedsregning

Indhold

9.0 Indledning

I dette hæfte skal du lære om *statistik* og *sandsynlighedsregning*.

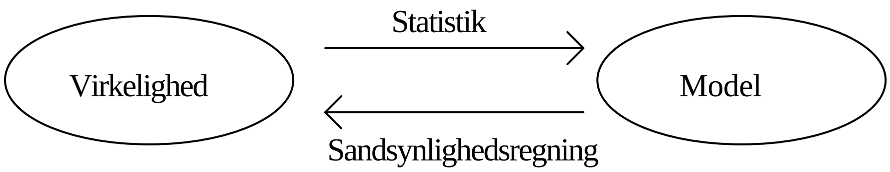
Matematikkens ædleste funktion ude i det virkelige liv er at opstille og formulere matematiske modeller, hvormed vi kan forudsige og kontrollere virkeligheden. Tænk bare på differentialregningen, som bl.a. kan bruges til at beskrive objekters bevægelse - en kanonkugle, en bil, Månen, et rumskib, ... Eller på de eksponentielle vækstmodeller, som kan forudsige størrelsen af Jordens befolkning, antallet af kerner i en mængde radioaktivt stof eller beløbet på en bankkonto.

Statistik og sandsynlighedsregning anvendes ofte ved beskrivelse af meget komplicerede situationer, hvor man ikke kender alle relevante data. F.eks. ved kast med en terning:

Kaster man en terning, så kan man vha. differentialregning og kendskab til Newtons love principielt beregne, hvorledes terningen vil lande. Man skal dog være i besiddelse af nogle data, nemlig en detaljeret beskrivelse af raflebægerets bevægelse og viden om, hvor terningen lå i raflebægeret til at starte med. Man skal også have en meget stor computer, idet de nødvendige udregninger er meget lange og meget komplicerede.

Dette er umuligt at gennemføre i praksis, så man vælger i stedet at beskrive terningekastet som et såkaldt *stokastisk eksperiment*. Udfaldet af eksperimentet er et af tallene 1, 2, 3, 4, 5 eller 6. Disse udfald forekommer med visse *sandsynligheder* - forekommer udfaldet 3 med sandsynligheden 0,25, så betyder det, at man kan forvente, at terningen viser 3 i 25% af tilfældene. Man kan dog ikke vide noget om, hvad terningen vil vise i næste kast.

Kort sagt kan man beskrive samspillet mellem statistik og sandsynlighedsregning på følgende måde:



Statistik hjælper med at "oversætte" virkeligheden til en model - den giver redskaber til beskrivelse af virkeligheden, og den fortæller, hvilke informationer, der er relevante, og hvilke, der ikke er relevante.

Sandsynlighedsregningen opstiller modellen og fortæller, hvorledes denne model kan bruges til at forudsige virkeligheden.

Arbejdsprocessen omkring modelleringen af et stokastisk eksperiment er som følger:

1) *Deskriptiv statistik*: Her indsamler og bearbejder man data fra allerede udførte eksperimenter.

2) *Sandsynlighedsregning*: Her opstiller man sin model og kommer med forskellige forudsigelser.

3) *Hypotesetest*: Her sammenholder man sin model med virkeligheden - passer modellen?

Vi skal starte med den deskriptive statistik. Her vil 3 forskellige eksperimenter blive studeret.

Herefter kommer sandsynlighedsregningen, hvor vi tager de fleste formelle begreber under afsnittet med sandsynlighedsfelter. De opnåede begreber bliver udnyttet i afsnittet om stokastiske variable, idet vi definerer et stokastisk eksperiment svarende til et definitionen på et endeligt sandsynlighedsfelt. Specielt vil vi betragte modeltyperne binomialfordeling og normalfordeling. I normalfordelingstilfældene laver vi en kontinuert model af virkeligheden, og det går godt , hvis antallet af udfald i det stokastiske eksperiment er meget stort. Så selv om vi kun har defineret en stokastisk variabel svarende til et endeligt sandsynlighedsfelt, så er det muligt at bruge sin viden om stokastiske variable og endelige sandsynlighedsfelter et langt stykke af vejen.

Hypotesetest vil vi ikke beskæftige os med, da det ligger udenfor bekendtgørelsens rammer, men det kan udmærket tages som et lille valgfrit projekt.

9.1 Deskriptiv statistik

I dette kapitel studerer vi tre forskellige eksempler. Disse eksempler vil senere bruges til at belyse forskellige sider af sandsynlighedsregningen. De tre eksempler er

1) Hyppighed af alfabetets bogstaver
2) Matematik-prøve-karakterer i 2x
3) Nittehoveders størrelse

Hyppighed af alfabetets bogstaver

Det er velkendt, at alfabetets bogstaver ikke forekommer lige ofte. F.eks. er det danske sprog spækket med e'er, mens bogstavet q sjældent eller aldrig optræder. Det har en vis interesse at undersøge forekomsten af de forskellige bogstaver. Bogtrykkere ville i gamle dage gerne vide, hvor mange *typer*, dvs. små blyklodser med et bogstav på, de skulle have - f.eks. var det dumt at have dobbelt så mange q'ere end e'ere. Og *kryptoanalytikere* - kodebrydere - kan bruge bogstavfordelingen til at bryde og læse kodede meddelelser.

Man har optalt antallet af de forskellige bogstaver i en dansk tekst på i alt 457 bogstaver. Resultatet er angivet i tabellen på næste side.

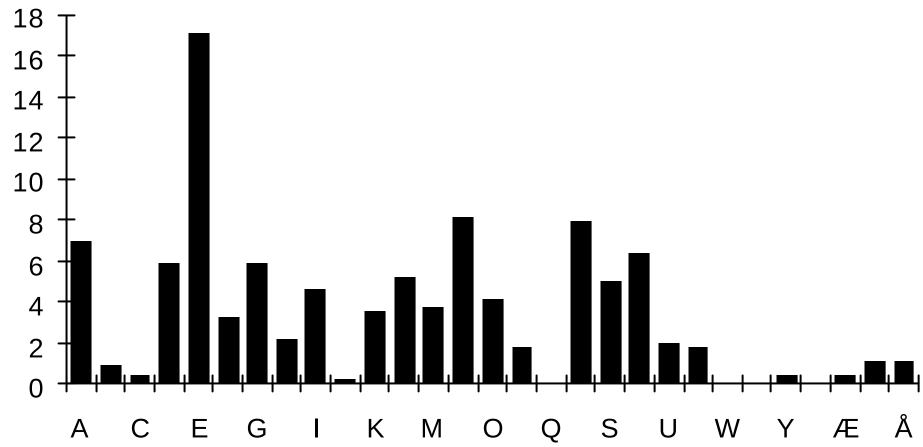
Hyppigheden er antal gange, bogstavet forekommer.

Frekvensen er hyppigheden divideret med det samlede antal bogstaver - eller sagt med andre ord, brøkdelen af det totale antal bogstaver, vores bogstav udgør. Man angiver normalt frekvensen i %.

Frekvensen er umiddelbart mere interessant end hyppigheden. Hyppigheden af et bogstav afhænger jo bl.a. af tekstens længde - f.eks. vil bogstavet “e” forekomme ca. 17 gange i en tekst på 100 bogstaver men 17000 gange i en tekst på 100000 bogstaver. Frekvensen er derimod den samme, nemlig ca. 17 %.

bogstav	hyppighed	frekvens (%)	bogstav	hyppighed	frekvens (%)
A	32	7,0	P	8	1,8
B	4	0,9	Q	0	0,0
C	2	0,4	R	36	7,9
D	27	5,9	S	23	5,0
E	78	17,1	T	29	6,3
F	15	3,3	U	9	2,0
G	27	5,9	V	8	1,8
H	10	2,2	W	0	0,0
I	21	4,6	X	0	0,0
J	1	0,2	Y	2	0,4
K	16	3,5	Z	0	0,0
L	24	5,3	Æ	2	0,4
M	17	3,7	Ø	5	1,1
N	37	8,1	Å	5	1,1
O	19	4,2			

Man afbilder normalt frekvenserne i et *pinde*- eller *stolpediagram*:



Hér er frekvenserne angivet i procent på y-aksen.

Der er mulighed for at foretagere flere statistiske undersøgelser, hvis det, man observerer, er tal. Det er jo muligt at sammenligne tal, addere tal og tage gennemsnit af tal.

Matematikprøve i 2x

I 2x på Pladderballe Gymnasium og HF-Kursus har man haft matematikprøve. Der er 23 elever i klassen, og deres karakterer var

Karakter	00	03	5	6	7	8	9	10	11	13
Antal	0	2	3	1	3	6	3	2	2	1

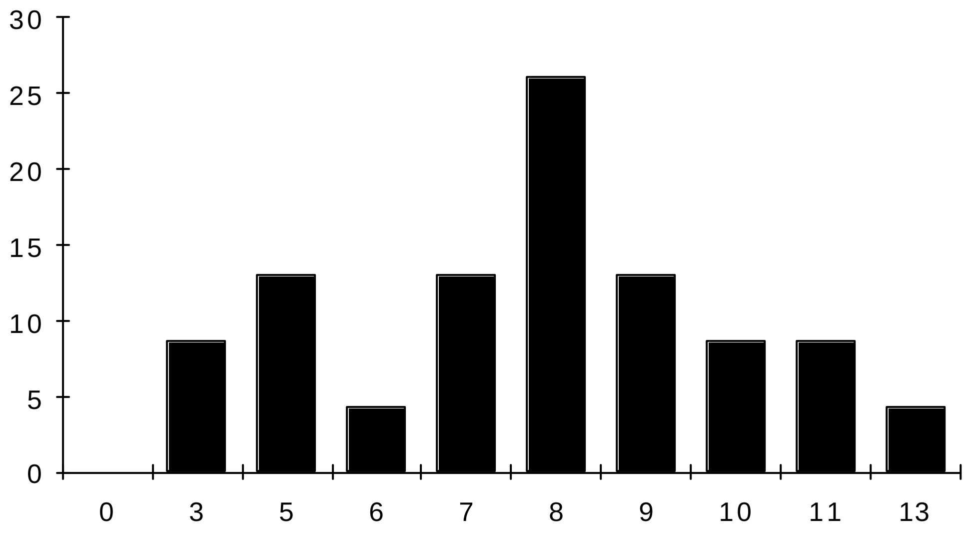
Ud fra dette talmateriale kan vi lave forskellige ting. Til hver mulig karakter kan vi knytte hyppigheden, frekvensen og den *kumulerede frekvens*. Den kumulerede (eller opsamlede) frekvens til f.eks. karakteren 6 er summen af frekvenserne for alle karakterer, som er mindre end eller lig med 6, dvs. for karaktererne 00, 03, 5 og 6.

Disse tal er samlet i nedenstående skema:

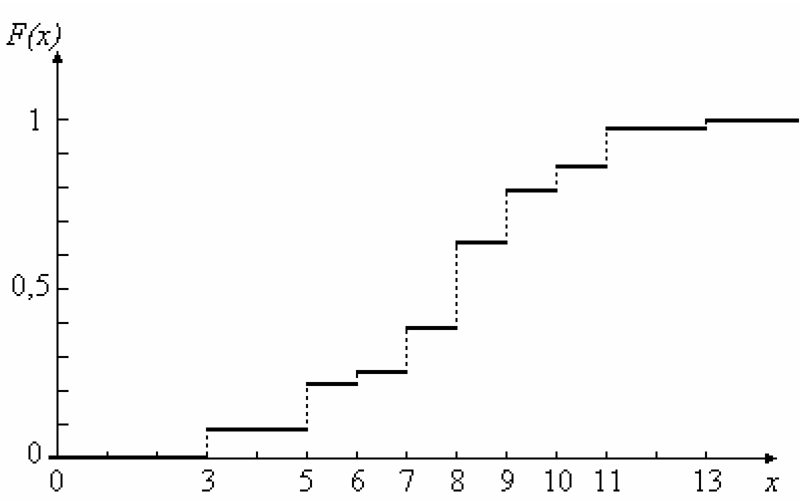
Karakter	00	03	5	6	7	8	9	10	11	13
Hyppighed	0	2	3	1	3	6	3	2	2	1
Frekvens	0,0	8,7	13,0	4,3	13,0	26,1	13,0	8,7	8,7	4,3
Kum. frekvens	0,0	8,7	21,7	26,0	39,0	65,1	78,1	86,8	95,5	99,9

Frekvenserne og de kumulerede frekvenser er angivet i %. Den kumulerede frekvens for den største karakter, som er 13, burde være på 100%, idet alle observationerne ligger under denne karakter. Men pga. afrundingsfejl får vi kun 99,9%.

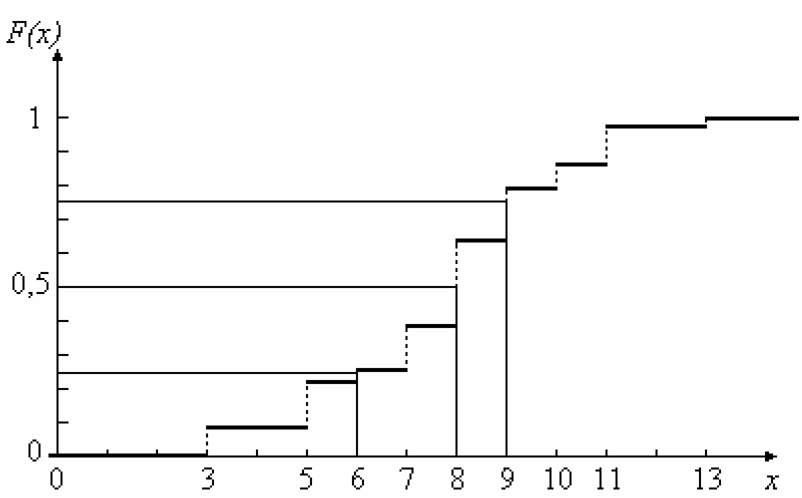
Frekvenserne kan som tidligere repræsenteres grafisk i et stolpediagram:



De kumulerede frekvenser angives normalt ved en *sumkurve* som er grafen for en stykkevis lineær funktion, hvis værdier i de observerede karakterer er lig den kumulerede frekvens. Bemærk, hvordan sumkurven er trappeformet. Det skyldes, at man jo kun har nogle bestemte tal som observationer - en såkaldt *diskret fordeling*.



På figuren nedenfor er vist, hvorledes man kan aflæse *kvartil-sættet*. Dette består af *medianen* eller *2-kvartilen*, som er den “observation”, som 50% af observationerne ligger nedenunder og 50% af observationer ligger over. Her er den 8.



Også *1. kvartilen* eller *25%-percentilen* er indtegnet - den er på 6, og ligeledes kan man se, at *3. kvartilen* eller *75%-percentilen* er på 9.

Kvartilsættet er således på (6 , 8 , 9).

Generelt siger kvartilsættet noget om karakternes fordeling, og især skæve fordelinger kan ofte afsløres ved en studium af kvartilsættet.
Endelig kan man beregne *gennemsnittet* af de givne karakterer. Dette kan gøres på flere måder - man kan addere alle de 23 givne karakterer og dividere med 23. Men

det er nu meget lettere at beregne ved at bruge formlen for vejet gennemsnit for vi kender jo vægtene af de enkelte karakterer, det er jo frekvenserne

$$0\cdot f_{00} + 3\cdot f_{03} + 5\cdot f_5 + 6\cdot f_6 + \ldots + 13\cdot f_{13}$$

hvor f_x er frekvensen af karakteren x. Det ses, at gennemsnittet er 7,74.

I ovenstående eksempel kunne den observerede størrelse kun antage et bestemt antal værdier, nemlig de 10 mulige karakterer. I andre sammenhænge observerer man størrelser, som kan antage alle mulige værdier, og som kan være behæftede med *måleusikkerhed*. Et eksempel herpå er måling af størrelser af nittehoveder:

Diameteren af nittehoveder

Man har målt diameteren (i mm) af 200 nittehoveder. Resultaterne er angivet nedenfor:

13,39	13,42	13,54	13,40	13,55	13,40	13,26	13,44
13,42	13,50	13,32	13,31	13,28	13,52	13,40	13,43
13,38	13,44	13,52	13,53	13,37	13,33	13,24	13,13
13,53	13,53	13,39	13,57	13,51	13,34	13,39	13,47
13,51	13,48	13,62	13,58	13,57	13,33	13,51	13,40
13,30	13,48	13,40	13,57	13,51	13,40	13,52	13,58
13,40	13,34	13,23	13,37	13,48	13,48	13,62	13,25
13,40	13,36	13,45	13,48	13,29	13,58	13,44	13,56
13,28	13,59	13,47	13,46	13,62	13,54	13,20	13,38
13,43	13,35	13,56	13,51	13,47	13,40	13,29	13,20
13,46	13,44	13,42	13,29	13,41	13,39	13,50	13,48
13,53	13,34	13,45	13,42	13,29	13,38	13,45	13,50
13,55	13,33	13,32	13,69	13,46	13,32	13,32	13,48
13,29	13,25	13,44	13,60	13,43	13,51	13,43	13,38
13,24	13,28	13,58	13,31	13,31	13,45	13,43	13,44
13,34	13,49	13,50	13,38	13,48	13,43	13,37	13,29
13,54	13,33	13,36	13,46	13,23	13,44	13,38	13,27
13,66	13,26	13,40	13,52	13,59	13,48	13,46	13,40
13,43	13,26	13,50	13,38	13,43	13,34	13,41	13,24
13,42	13,55	13,37	13,41	13,38	13,14	13,42	13,52
13,38	13,54	13,30	13,18	13,32	13,46	13,39	13,35
13,34	13,37	13,50	13,61	13,42	13,32	13,35	13,40
13,57	13,31	13,40	13,36	13,28	13,58	13,58	13,38
13,26	13,37	13,28	13,39	13,32	13,20	13,43	13,34
13,33	13,33	13,31	13,45	13,39	13,45	13,41	13,45

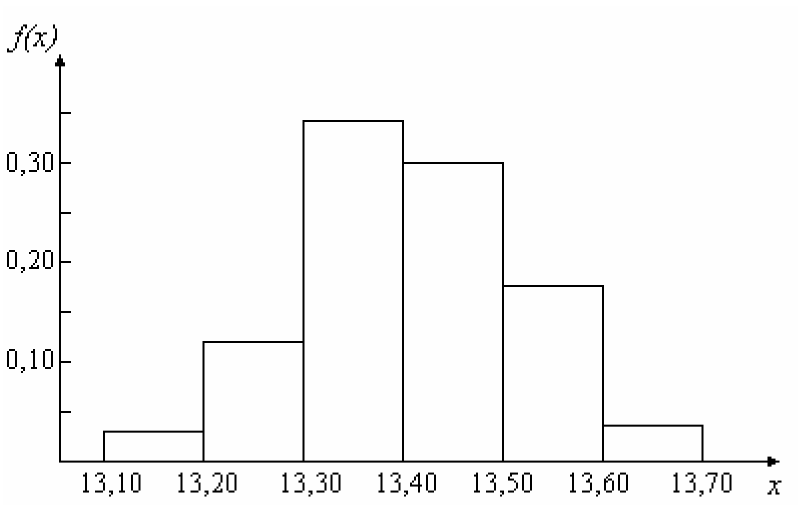
Det ser jo ikke specielt overskueligt ud. Man plejer derfor at gruppere målingerne. Dette gøres ved at man vælger nogle intervaller (som ikke behøver at være lige store) og optæller, hvor mange observationer der ligger i hvert interval. Vi vælger intervallerne

$$]13{,}10\,;13{,}20],\ ]13{,}20\,;13{,}30],\ ...,\ ]13{,}60\,;13{,}70].$$

Bemærk, at alle intervallerne er af formen $]a;b]$ - dette sikrer, at der aldrig er tvivl om, hvor en observation skal placeres. Vi laver en tabel med intervallerne *I*, hyppighederne $h(I)$, frekvenserne $f(I)$ og de kumulerede frekvenser $F(I)$.

I	$h(I)$	$f(I)$	$F(I)$
] 13,10 , 13,20]	6	0,03	0,03
] 13,20 , 13,30]	24	0,12	0,15
] 13,30 , 13,40]	68	0,34	0,49
] 13,40 , 13,50]	60	0,30	0,79
] 13,50 , 13,60]	35	0,175	0,965
] 13,60 , 13,70]	7	0,035	1,000

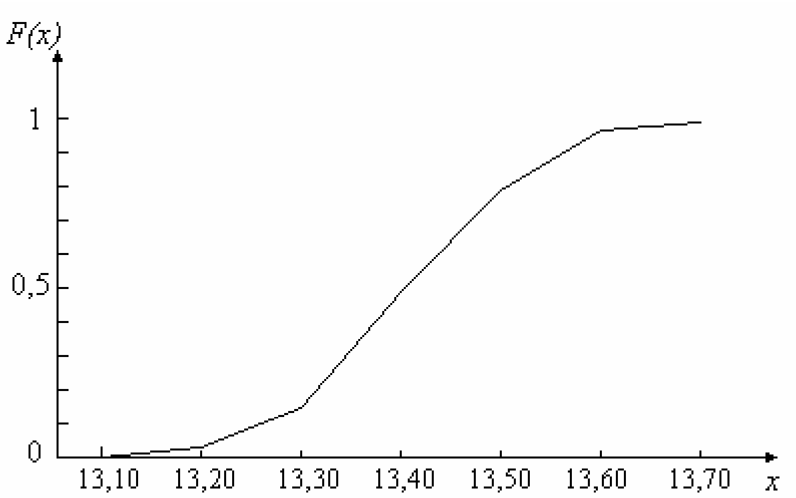
Vi tegner nu frekvensfordelingen i et *histogram*. I et histogram er der til hvert interval (dvs. vi laver kun histogrammer for grupperede observationer) lavet et blok med et areal svarende til intervalfrekvensen. Det betyder, at vi er nødt til at definere en arealenhed. Hvis alle intervaller er lige stor, så viser man arealenheden på 2.aksen, men hvis der er nogle intervaller der er af forskellig størrelse, så bliver man nødt til at fjerne 2.aksen og vise arealenheden ved f.eks. at lave en firkant af størrelse 1%.



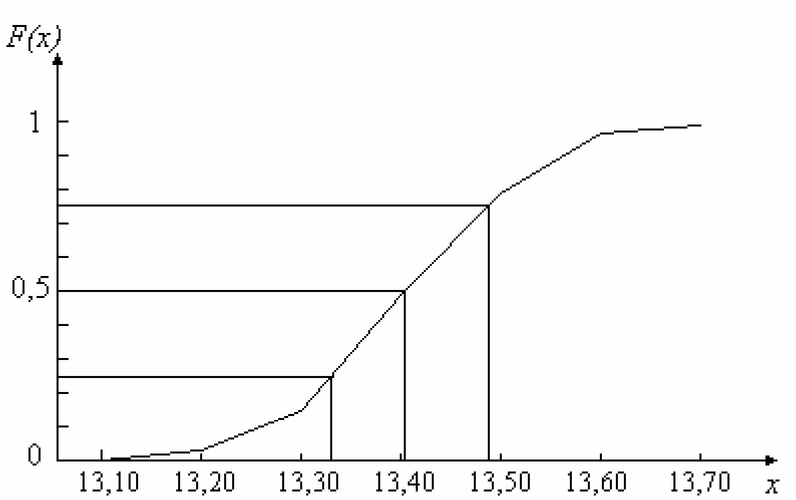
Vi ser at typeintervallet er $]13{,}30\,;13{,}40]$.

Vi vil nu tegne sumkurven. Det kræver, at vi ved hvordan observationerne er fordelt i det enkelte interval! Men det kan jo forekomme lidt dumt at skulle til at lave se på den enkelte observation, når vi møjsommeligt har grupperet observationerne. Vi foretager den rimelige antagelse, at observationerne ligger jævnt fordelt i hvert enkelt interval, det betyder, at den kumulerede frekvens stiger jævnt indenfor det enkelte interval, eller matematisk formuleret - linieært.

Vi indtegner de højre intervalendepunkter som x-værdier og de dertil hørende kumulerede frekvenser, og får nu en sammenhængende sumkurve med følgende udseende.



Vi kan nu aflæse kvartilsættet til (13,33; 13,40; 13,49). Se figuren nedenfor:



Vi kan også aflæse f.eks. hvor procent der er under 13,25 ved at gå ind på sumkurven ved 13,35 og aflæse den tilhørende kumulerede frekvens, som giver 9,5%. Hvis vi nu lod x være nittehovedernes størrelse, så kunne vi sige

sandsynligheden for at $x \leq 13{,}35$ er 9,5%

Denne skrivemåde vil vi gøre meget brug af senere, idet vi vil definere et begreb, der hedder stokastisk variabel, som gør det lettere at regne på disse ting.

Medianen er den midterste observation dvs. 50%-fraktilen, så den er 13,40.

Middeltallet beregnes på samme måde som før ved at vægte intervallerne med deres frekvens (som jo er sandsynligheden for at få intervallet). Da vi siger, at observationerne i intervallerne er jævnt fordelte, så vil vi bruge intervalmidtpunktet til middeltalsberegning.

$$\text{middeltal} = 13{,}15\cdot 0{,}03 + 13{,}25\cdot 0{,}012 + 13{,}35\cdot 0{,}34 + 13{,}45\cdot 0{,}30 + 13{,}55\cdot 0{,}175 + 13{,}65\cdot 0{,}035$$

$$= 13.41$$

Bemærk at middeltallet er meget tæt på medianen. Dette er normalt tilfældet, når man har mange observationer, idet observationerne så har en tendens til at være jævnt fordelt omkring middeltallet. Faktisk så kan man i sådanne tilfælde bestemme matematisk en frekvensfordeling, der utrolig godt stemmer overens med den frekvensfordeling, som man faktisk har. Denne fordeling kaldes en normalfordeling, og den vil blive behandlet mere indgående i sidste afsnit.

Opgaver

1.1 Ved optælling af røde blodlegemer i blodet betragter man under mikroskop et tællekammer fyldt med fortyndet blod. Et tællekammer er et lille fladt rum med glasvægge, som indeholder et ganske bestemt rumfang. Tællekammeret er inddelt i lige store felter.
Man tæller nu antallet af blodlegemer i hvert felt.
I skemaet nedenfor ses resultatet, dvs. antallet af felter med det angivne antal blodlegemer.

x	0	1	2	3	4	5	6	7	8	9	10	11	12
h	0	2	7	14	26	28	32	36	25	12	12	5	1

x er antallet af blodlegemer, og h er den tilsvarende hyppighed.

a) Bestem frekvenserne og de kumulerede frekvenser for ovennævnte observationer.
b) Tegn et pindediagram for frekvensfordelingen.
c) Tegn et trappediagram for den kumulerede frekvensfordeling.
d) Bestem kvartilsættet for observationerne.
e) Bestem middeltallet.

1.2 Tabellen nedenfor viser aldersfordelingen for den grønlandske befolkning pr. 1/1 1989.

Alder (år)	% af befolkningen
0 - 9	18,2 %
10 - 19	14,1 %
20 - 29	23,9 %
30 - 39	17,1 %
40 - 49	12,7 %
50 - 59	7,8 %
60 - 69	3,9 %
70 - 79	1,7 %
80 -	0,5 %

a) Er der tale om en diskret eller en kontinuert fordeling.
b) Tegn histogrammet for denne fordeling.
c) Tegn sumkurven for denne fordeling.
d) Bestem kvartilsættet
e) Hvor stor en del af den grønlandske befolkning er 67 år eller derover?

9.3 Sandsynlighedsfelt

Vi skal nu se, hvorledes en matematiker vil modellere et stokastisk eksperiment. Vi starter med at definere et *endeligt sandsynlighedsfelt*:

Definition 1

Et *endeligt sandsynlighedsfelt* består af

a) en endelig mængde U, kaldet *udfaldsrummet*,
b) en *funktion P* fra U til $[0;1]$, kaldet *sandsynlighedsfunktionen*, opfyldende
c) $P(u_1)+P(u_2)+P(u_3)+...+P(u_n)=1$
(Elementerne i U kaldes *udfald* og betegnes med symbolerne u_1, ..., u_n).

Hvordan skal dette forstås? Vi betragter et stokastisk eksperiment, hvis udfald kan være et af elementerne $u_1, u_2, ..., u_n$. Sandsynligheden $P(u_1)$ skal da fortolkes som frekvensen af udfaldet u_1, når dette stokastiske eksperiment udføres mange, mange gange.

Eksempel

Inkarnerede ludo-spillere vil vide, at kaster man en terning f.eks. 6000 gange, så vil man få øjentallet 1 ca. 1000 gange, øjentallet 2 ca. 1000 gange osv. Kort sagt, frekvensen for udfaldet 1 er $\frac{1}{6}$, frekvensen for udfaldet 2 er ligeledes $\frac{1}{6}$, osv. Sandsynlighedsteoretisk formulerer vi dette på følgende måde:

Udfaldsrummet er $U=\{1,2,3,4,5,6\}$

Sandsynlighedsfunktionen P er fastlagt ved nedenstående tabel:

u	1	2	3	4	5	6
$P(u)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Betingelsen

$$P(1)+P(2)+P(3)+P(4)+P(5)+P(6)=\tfrac{1}{6}+\tfrac{1}{6}+\tfrac{1}{6}+\tfrac{1}{6}+\tfrac{1}{6}+\tfrac{1}{6}=1$$

er opfyldt.

Bemærk, at grunden til, at sandsynligheden $P(u)$ for et udfald er et tal i intervallet $[0;1]$ er jo, at frekvenser jo altid ligger i dette interval.

Ligeledes skal summen af sandsynlighederne for alle udfaldene være 1, idet det samme gælder for frekvenserne.

Normalt arbejder man ikke med udfaldene alene, men derimod med hændelser, som er et mere generelt begreb:

Definition 2

En *hændelse* A er en delmængde af udfaldsrummet U.

Sandsynligheden for en hændelse $P(A)$ er summen af sandsynlighederne for alle udfaldene i A.

Eksempel

Mikkel foretrækker at spille ludo med en falsk terning. Denne kan beskrives ved sandsynlighedsfeltet

$$U = \{1,2,3,4,5,6,7\}$$

u	1	2	3	4	5	6	7
$P(u)$	0,15	0	0,05	0,25	0,1	0,4	0,05

Denne terning er således indrettet således, at man får en 6'er i 40% af kastene. Man kan aldrig på slaget 2, idet $P(2) = 0$; men til gengæld får man, til Mikkels modspilleres store forundring, slaget 7 i 5% af kastene.

I et bestemt spil ludo kan Mikkel slå Mette hjem, hvis han slår 1, 2 eller 4, og han kan slå Kasper hjem, hvis han slår 4, 5, 6 eller 7. Hvad er sandsynligheden for, at Mikkel kan slå nogen hjem?

For at kunne besvare dette spørgsmål, indfører vi nogle hændelser:

M = Mikkel kan slå Mette hjem = {1, 2, 4}
K = Mikkel kan slå Kasper hjem = {4, 5, 6, 7}
H = Mikkel kan slå en eller anden hjem

Vi ser, at

$$H = M \cup K = \{1,2,4,5,6,7\}$$

og

$$P(M) = P(1) + P(2) + P(4) = 0,15 + 0 + 0,25 = 0,40$$

$$P(K) = P(4) + P(5) + P(6) + P(7) = 0{,}25 + 0{,}1 + 0{,}4 + 0{,}05 = 0{,}80$$

$$P(H) = P(1) + P(2) + P(4) + P(5) + P(6) + P(7) =$$
$$0{,}15 + 0 + 0{,}25 + 0{,}1 + 0{,}4 + 0{,}05 = 0{,}95$$

Mikkel har altså 95% chance for at slå en eller anden hjem.
Bemærk, at selvom

$$H = M \cup K$$

så gælder der ikke, at

$$P(H) = P(M) + P(K)$$

- dette skyldes jo, at man ved udregning af summen $P(M) + P(K)$ kommer til at medtage leddet $P(4)$ to gange.

For regning med hændelser gælder der generelt følgende regneregler:

Sætning 3 (FS)

Lad A og B være hændelser i et udfaldsrum U. Så gælder:

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

b) $P(A \cup B) = P(A) + P(B)$, hvis $A \cap B = \emptyset$

c) $P(U) = 1$

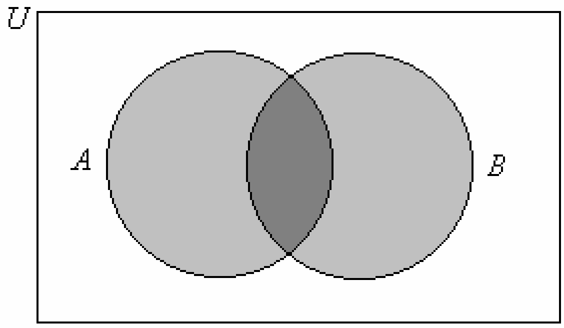
d) $P(\emptyset) = 0$

e) $P(\overline{A}) = 1 - P(A)$

Bevis:

a) Når man udregner $P(A) + P(B)$, så kommer man til at medregne sandsynlighederne for elementerne i $A \cap B$ **to** gange. Dette justeres ved at fratrække sandsynlighederne for disse fælles elementer **en** gang, nemlig i leddet $P(A \cap B)$.

Sætningen er illustreret på figuren nedenfor. Forestil dig, at du skal male hele $A \cup B$ gråt. Dette kan gøres ved først at male området *A* og derefter området *B*. Men hov - det fælles område $A \cap B$ fik dobbelt dosis, og vi må derfor skrabe det ene lag maling af igen!



b) Idet der ingen fælles elementer er i A og B, så er der ingen udfald, hvis sandsynlighed man kommer til at medregne to gange i summen $P(A)+P(B)$.

c) Dette er en reformulering af definitionen på sandsynlighedsfunktionen:

$$P(U) = P(u_1) + P(u_2) + \ldots + P(u_n) = 1$$

d) Den tomme mængde $\emptyset$ indeholder ingen elementer, så sandsynligheden er 0 for at man rammer ind i et af disse.

e) Dette følger af b) - en hændelse A og dens komplementærmængde $\overline{A}$ opfylder

$$A \cup \overline{A} = U \qquad \text{og} \qquad A \cap \overline{A} = \emptyset$$

hvilket giver

$$P(A) + P(\overline{A}) = P(U) = 1$$

eller

$$P(\overline{A}) = 1 - P(A).$$

Eksempel

Mikkel har nu fået anskaffet sig en ny, 10-sidet ludo-terning. Den kan beskrives ved sandsynlighedsfeltet

$$U = \{1,2,3,4,5,6,7,8,9,10\}$$

u	1	2	3	4	5	6	7	8	9	10
$P(u)$	0,1	0,2	0,1	0,05	0,05	0,15	0,05	0,05	0,05	0,2

Mikkel vil finde sandsynligheden for følgende hændelser:

A : Mikkel slår et ulige tal med terningen
B: Mikkel slår højst 4 med terningen
C: Mikkel slår ikke 1 med terningen.

Han får

$$P(A)=P(1)+P(3)+P(5)+P(7)+P(9)=$$
$$0{,}1+0{,}1+0{,}05+0{,}05+0{,}05=0{,}35$$
$$P(B)=P(1)+P(2)+P(3)+P(4)=$$
$$0{,}1+0{,}2+0{,}1+0{,}05=0{,}45$$
$$P(C)=P(2)+P(3)+P(4)+P(5)+P(6)+P(7)+P(8)+P(9)+P(10)=$$
$$0{,}2+0{,}1+0{,}05+0{,}05+0{,}15+0{,}05+0{,}05+0{,}05+0{,}2=0{,}9$$

Men den sidste sandsynlighed kunne regnes ud meget lettere.
Komplementærhændelsen til C er netop hændelsen $\{1\}$, som kun består af udfaldet 1. Derfor

$$P(C)=1-P(\overline{C})=1-P(1)=1-0{,}1=0{,}9.$$

Vi skal nu beskæftige os med de såkaldte *betingede* sandsynligheder:

Eksempel

Privatdetektiv O. P. Dager har ved et nøjere studium af Hr. D. I. Rektør´s tøjvaner fundet frem til nedenstående tabel, der viser til hvilke møder, han har slips på. På arbejdet er der 3 slags møder, nemlig bestyrelsesmøder, planlægningsmøder og medarbejdermøder. Disse forekommer med hyppighederne henholdsvis 10%, 30% og 60% .

Møde	hyppighed	slips
bestyrelsesmøde	0,10	100 %
planlægningsmøde	0,30	50 %
medarbejdermøde	0,60	10 %

I sin rapport til Fru D. I. Rektør skal han fortælle i hvor stor en procentdel af tilfældene, Hr. D. I. Rektør har slips på.

Hr. O. P. Dager regner således:

Procentdel tilfælde med slips på =
procentdel tilfælde (bestyrelsesmøde og med slips) +
procentdel tilfælde (planlægningsmøde og med slips) +
procentdel tilfælde (medarbejdermøde og med slips) =
$100\%\cdot 0{,}10+50\%\cdot 0{,}30+10\%\cdot 0{,}60$ =
31%

Han regner med andre ord, som var det et vejet gennemsnit!

Lad os oversætte eksemplet til sandsynlighedsregning, hvor problemet naturligt hører hjemme. Dvs. vi laver nogle hændelser

B: bestyrelsesmøde
P: planlægningsmøde
M: medarbejdermøde
S: Slips på

Vi kan nu skrive

$$
\begin{aligned}
P(S) = P(S \cap B) + P(S \cap P) + P(S \cap M) = \\
&P(S \textit{ betinget bestyrelsesmøde}) \cdot P(B) + \\
&P(S \textit{ betinget planlægningsmøde}) \cdot P(P) + \\
&P(S \textit{ betinget medarbejdermøde}) \cdot P(M)
\end{aligned}
$$

Det kan skrives mere kompakt, hvis vi lader ordet betinget erstatte med en lodret streg, f.eks. $P(S \textit{ betinget bestyrelsesmøde}) = P(S|B)$:

$$P(S) = P(S|F) \cdot P(F) + P(S|M) \cdot P(M) + P(S|A) \cdot P(A)$$

Vi har med dette eksempel motiveret nedenstående definitioner og sætninger.

Definition 4

Lad A og B være to hændelser, så er den *betingede sandsynlighed* for A under forudsætning af B, skrevet $P(A|B)$, givet ved

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{(forudsat } P(B) \neq 0\text{)}$$

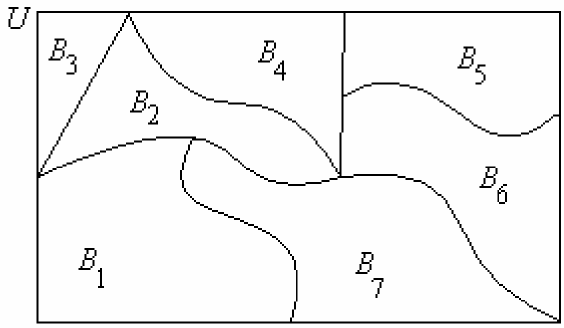
Bemærk at hvis $P(B) = 0$, så er ovenstående definition meningsløs, men i dette tilfælde sætter man bare $P(A|B) = 0$.

Definition 5

En *klassedeling* af en mængde (eller et udfaldsrum) U er en samling af delmængder B_1 , B_2 , B_3 , ... , B_n af U, som opfylder:

1) $B_i \neq \emptyset \quad , \quad i = 1, 2, 3, ..., n$
2) $B_i \cap B_j = \emptyset \qquad \text{for} \qquad i \neq j$
3) $B_1 \cup B_2 \cup B_3 \cup ... \cup B_n = U$

Man skal altså forestille sig, at man har savet U i n forskellige stykker.



Nedenstående kaldes ofte *loven om total sandsynlighed*:

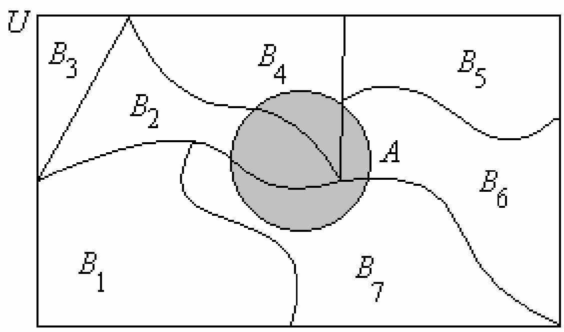
Sætning 6

Lad B_1 , B_2 , … , B_n være en klassedeling af udfaldsrummet U. og lad A være en hændelse i U. Så gælder

$$P(A) = P(A|B_1)\cdot P(B_1) + \quad ..... \quad + P(A|B_n)\cdot P(B_n)$$

Bevis:

Situationen er skitseret nedenfor:



Af tegningen ses

$$A = (A\cap B_1)\cup(A\cap B_2)\cup \quad ..... \quad \cup(A\cap B_n)$$

$$P(A) = P((A\cap B_1)\cup(A\cap B_2)\cup \quad ..... \quad \cup(A\cap B_n))$$

Da $B_i \cap B_j = \emptyset$, så er $(A\cap B_i)\cap(A\cap B_j) = \emptyset$. Af sætning 3b fås

$$P(A) = P(A\cap B_1) + P(A\cap B_2) + \quad ..... \quad + P(A\cap B_n)$$

$$P(A|B_1)\cdot P(B_1) + \quad ..... \quad + P(A|B_n)\cdot P(B_n)$$

Endelig har vi *omvendingsformlen*:

Sætning 7

Lad A og B være hændelser med $P(A) \neq 0$ og $P(B) \neq 0$, så er

$$P(B \mid A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Bevis

Beviset følger direkte af definition på betinget sandsynlighed

$$B \cap A = A \cap B$$

$$\Downarrow$$

$$P(B \cap A) = P(B \cap A)$$

$$\Updownarrow$$

$$P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$$

$$\Updownarrow$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Eksempel

Lad 1000 skruer være fordelt i 4 bokse B_1, B_2, B_3, B_4 med henholdsvis 100, 200, 300 og 400 skruer. De 4 bokse pakkes af 4 forskellige personer, som har en vis sandsynlighed for at pakke defekte skruer ned i en boks på henholdsvis 5%, 3%, 12% og 1%.

Find sandsynligheden for at på en defekt skrue.

Først fastlægger vi de hændelser, som vi er interesseret i

B_1:	Få en skrue fra boks1	$P(B_1) = 0{,}1$
B_2:	Få en skrue fra boks2	$P(B_2) = 0{,}2$
B_3:	Få en skrue fra boks3	$P(B_3) = 0{,}3$
B_4:	Få en skrue fra boks4	$P(B_4) = 0{,}4$
A:	Få en forkert skrue	

Vi kan nu skrive

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3) + P(A|B_n) \cdot P(B_n)$$
$$= 5\% \cdot 0{,}1 + 3\% \cdot 0{,}2 + 12\% \cdot 0.3 + 1\% \cdot 0.4 = 5{,}1\%$$

Find sandsynligheden for at skruen kom fra boks2, når skruen er forkert.

$$P(B_2 | A) = \frac{P(A|B_2) \cdot P(B_2)}{P(A)} = \frac{3\% \cdot 0{,}2}{5{,}1\%} = 0{,}12$$

Eksempel

I en pose er der farvede bolde. Der er 4 farver: Blå, grøn, gul, rød.
Jens tager en bold op, og vi betragter nu forskellige hændelser.

A: Bolden har en af farverne gul eller rød
B: Bolden har en af farverne gul eller grøn
C: Bolden har farven grøn
D: Bolden har en af farverne blå, gul eller rød

Alle farver trækkes med lige stor sandsynlighed, så vi får

$P(A) = 0{,}50$ $\quad$ $P(B) = 0{,}50$
$P(C) = 0{,}25$ $\quad$ $P(D) = 0{,}75$

Lad os se på *D* forudsat, at vi ved at hændelsen *B* eller *C* er indtruffet

$$P(D|B) = \frac{P(D \cap B)}{P(B)} = \frac{P("gul")}{P(B)} = \frac{0{,}25}{0{,}50} = 0{,}50$$

$$P(D|C) = \frac{P(D \cap C)}{P(C)} = \frac{P("Ø")}{P(C)} = \frac{0}{0{,}25} = 0$$

Som vi kan se, så kan viden om, at *B* eller *C* er indtruffet formindske sandsynligheden for D drastisk. Det er jo oplagt, for hvis vi ved, at kuglen er grøn, ja så kan den naturligvis ikke være blå, gul eller rød.

$$P(B|C) = \frac{P(B \cap C)}{P(c)} = \frac{P("grøn")}{P(C)} = \frac{0{,}25}{0{,}25} = 1$$

$$P(B|D) = \frac{P(B \cap D)}{P(D)} = \frac{P("gul")}{P(D)} = \frac{0{,}25}{0{,}75} = \frac{1}{3}$$

Vi ser, at viden om, at *D* er indtruffet, formindsker sandsynligheden for, at *B* indtræffer, og modsat så er det helt sikkert, at *B* indtræffer, når *C* er indtruffet. Vi må konkludere at viden om en hændelse kan formindske eller forøge sandsynligheden for en anden hændelse.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P("gul")}{P(B)} = \frac{0{,}25}{0{,}50} = 0{,}50$$

Men hov $P(A) = 0{,}50$, så om *B* er indtruffet eller ej ændrer ikke på, om *A* indtræffer, m.a.o. så er *A* uafhængig af *B*.

Definition 8

A er *uafhængig* af B hvis $P(A|B) = P(A)$

Dette kan vi omforme ved brug af definitionen på betinget sandsynlighed

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Så får vi, når A er uafhængig af B

$$P(A \cap B) = P(A) \cdot P(B)$$

Man får selvfølgelig den samme ligning frem, hvis B er uafhængig af A, og dermed har vi vist følgende sætning.

Sætning 9 (FS)

A og B er uafhængige hændelser netop hvis

$$P(A \cap B) = P(A) \cdot P(B)$$

Eksempel

En en-krone og en to-krone kastes. Her har vi fire forskellige udfald:

$1P$: 1-kronen viser plat
$1K$: 1-kronen viser krone
$2P$: 2-kronen viser plat
$2K$: 2-kronen viser krone

Idet mønterne er symmetriske, må der gælde, at

$$P(1P) = P(1K) = 0{,}5$$
$$P(2P) = P(2K) = 0{,}5$$

Hvad er sandsynligheden for, at 1-kronen viser plat, og 2-kronen krone, altså for hændelsen $1P \cap 2K$

Nu er en-kronen jo ligeglad med, hvad to-kronen viser, så vi må konkludere, at udfaldene $1P$ og $2K$ er uafhængige!
Derfor:

$$P(1P \cap 2K) = P(1P) \cdot P(2K) = 0{,}5 \cdot 0{,}5 = 0{,}25$$

Opgaver

2.1 Givet: To hændelser *A* og *B* opfyldende

$$P(A) = \tfrac{2}{5} \qquad P(B) = \tfrac{2}{5} \qquad P(A \cap B) = \tfrac{1}{2}$$

Find: $P(A \cup B)$

2.2 Givet: To hændelser *A* og *B* opfyldende

$$P(A) = \tfrac{1}{3} \qquad P(A \cup B) = \tfrac{1}{2} \qquad P(A \cap B) = \tfrac{1}{4}$$

Find: $P(B)$

2.3 På Pladderballe Cykelpumpefabrik A/S har man to cykelpumpemaskiner, *A* og *B*. Maskine *A* står for 60% af cykelproduktionen, mens *B* producerer resten. Maskine *A* producerer en defekt cykelpumpe med 3% chance, mens sandsynligheden for, at en cykelpumpe fra *B* er defekt er på 5%.

a) Bestem sandsynligheden for, at en tilfældig cykelpumpe fra Pladderballe Cykelpumpefabrk A/S er defekt.

b) Casper køber en cykelpumpe og opdager til sin rædsel, at den er defekt. Hvad er sandsynligheden for, at den blev produceret på maskine *B*?

2.4 Jacob er til afsluttende prøve i teoridelen af *Anvendt Cykelpumpning*. Prøven er nu ikke så svær - den består af nogle spørgsmål med 5 svarmuligheder, hvoraf kun ét er det rigtige svar. Desværre har Jacob ikke fået læst på pensum, så han kender kun svarere på 30% af opgaverne. Hvis han ikke kender svaret, så sætter han bare sit kryds helt tilfældigt.

a) Hvad er sandsynligheden for, at Jacob svarer rigtigt på et givet spørgsmål?

b) Hvis Jacob givet det rigtige svar på spørgsmålet, hvad er så sandsynligheden for, at han faktisk kender svaret og ikke bare gætter?

2.5 Betragt udfaldsrummet $U = \{1, 2, 3, 4\}$ med sandsynlighedsfordelingen

$$P(1) = P(2) = P(3) = P(4) = \tfrac{1}{4}$$

Betragt endvidere hændelserne

$$A = \{1, 2\} \qquad B = \{1, 3\} \qquad C = \{1, 4\}$$

a) Bevis, at *A* og *B* er uafhængige.

b) Bevis, at *A* og *C* er uafhængige.

c) Er *B* og *C* uafhængige?

2.6 I brætspillet *Matador* slår man med to terninger og rykker efter summen af deres øjne.

a) Er det rimeligt at antage, at udfaldet af de terninger er uafhængige af hinanden?

b) Opstil et sandsynlighedsfelt, som beskriver de mulige udfald for kast med to terninger.

c) Michael spiller Matador mod Trine og Lene. Michael skal til at slå. Hvis Michael slår 5, rammer han på *Dronningens Tværgade*, som ejes af Trine, og som er bebygget med 4 huse. Slår han 8, så lander han på *Rådhuspladsen*, som Lene har bygget hotel på. I begge tilfælde vil lejeafgiften ruinere Michael.
Bestem sandsynligheden for, at Michael bliver ruineret.

d) Opstil en tabel over sandsynligheden for, at man skal rykke 2 felter, 3 felter, 4 felter, osv...

2.7 I det ikke særligt udbredte brætspil *Århusiansk Matador* spiller man efter de sædvanlige *Matador*-regler, dog skal man slå med en rød og en blå terning og rykke det røde øjenantal minus det blå øjenantal. Hvis dette tal er negativt, så skal man rykke baglæns (og passerer man *Start*, så skal man aflevere 4000 kr. til banken. Omvendt, rykker man baglæns og lander på *Rådhuspladsen*, som er bebygget med hotel, så får man 40000 kr. i anti-leje fra hotelejeren).

Opstil en tabel, som i opgave 2.6.d, over de mulige udfald og deres sandsynligheder.

2.8 På Pladderballe Traktorfabrik er der lige stor sandsynlighed for, at en traktor er produceret på en mandag, en tirsdag, en onsdag, en torsdag eller en fredag. Traktorer lavet på en mandag har 4% risiko for at være defekte, traktorer lavet om tirsdagen, onsdagen eller torsdagen har 1% risiko for at være defekte, og traktorer lavet om fredagen har 2% risiko for at være defekte.

a) Hvad er sandsynligheden for, at en tilfældig traktor er defekt?

b) Mette køber en traktor, da hun ikke vil bruge sin hest til pløjning. Desværre for hesten er traktoren defekt. Hvad er sandsynligheden for, at den blev produceret på en mandag?

2.9 På Pladderballe Gymnasium og HF-kursus består 2y af 70% drenge og 30% piger. 40% af drengene og 60% af pigerne ryger.
Hvad er sandsynligheden for, at en elev, som ryger, er en pige?

9.3 Stokastisk variabel

Et stokastisk eksperiment er et forsøg, som giver en række udfald, der alle kan have en vis sandsynlighed for at forekomme. Ofte kender man kun sandsynligheden for de enkelte udfald ved at have gentaget eksperimentet mange gange.

Eksempel

At kaste med en terning og notere sig øjentallet er et stokastisk eksperiment med seks udfald, der alle har sandsynligheden en sjettedel.

Eksempel

At bide i et tilfældigt nedfaldent æble i hos Jørgens forældres have vil have følgende udfald!

Udfald	spist 0 orm	spist 0,25 orm	spist 0,5 orm	spist 0,75 orm	spist 1 orm
sandsynlighed	0,85	0,06	0,04	0,04	0,01

Generelt kan vi definere

Definition 10

Et *stokastisk eksperiment* er et eksperiment med n mulige udfald $u_1, u_2, \ldots, u_n$, og en sandsynlighedsfunktion P, som opfylder

1) $P(u_i) \in [0;1]$ for $i = 1, 2, \ldots, n$

2) $P(u_i) + P(u_2) + \ldots + P(u_n) = 1$

Vi ser, at definitionen på et stokastisk eksperiment passer godt sammen med definitionen på et endeligt sandsynlighedsfelt.

D.v.s. vi kan bruge vores viden om endelige sandsynlighedsfelter på et stokastisk eksperiment.

I forbindelse med et endeligt sandsynlighedsfelt snakkede vi meget om hændelser, og det skal vi også gøre her. Det viser sig, at det er praktisk at have en form for tæller, når man skal beskrive forskellige hændelser på en kort form. En sådan tæller kaldes en *stokastisk variabel*.

Lad os motivere det lidt mere, før vi definerer en stokastisk variabel.

Ofte ønsker man at behandle sine eksperimenter mere indgående. Jørgen kunne f.eks. ønske sig at vide, hvor meget orm han i gennemsnit spiser pr æble! Og som ovenstående tabel ser ud nu, er det ikke umiddelbart muligt. Men hvis J.I. blot lader udfaldet være repræsenteret ved et tal - nemlig den spiste ormemængde, så bliver talbehandling muligt.

udfald - den spiste ormemængde	0	0,25	0,5	0,75	1
P(udfald)	0,85	0,06	0,04	0,04	0,01

Man kan nu regne gennemsnittet ud (der er jo et vejet gennemsnit, hvor vægten af et udfald er sandsynligheden for udfaldet).

$$\text{gennemsnit} = 0\cdot 85 + 0{,}25\cdot 0{,}06 + 0{,}5\cdot 0{,}04 + 0{,}75\cdot 0{,}04 + 1\cdot 0{,}01 = 0{,}075$$

I statistikken siger man, at man har lavet en stokastisk variabel, der tæller spist ormemængde, og vi kunne nu lave mange flere beregninger, hvis vi ønskede det.

Definition 11

En *stokastisk variabel X* knytter et reelt tal til udfaldene for et stokastisk eksperiment

Bemærk, at man til ét stokastisk eksperiment således godt kan have mange forskellige stokastiske variable.

Definition 12

Sandsynlighedsfordelingen for en stokastisk variabel *X* er en tabel, som til en værdi t_i af den stokastiske variabel knytter summen af sandsynlighederne for de udfald, som giver *X* værdien t_i

t	t_1	t_2	t_3	...	t_n
$P(X=t)$	p_1	p_2	p_3	...	p_n

Vi ser umiddelbart at sandsynlighedsfordelingen for en stokastisk variabel, svarer til frekvenserne vi arbejdede med i statistik afsnittet. Man kalder derfor ofte sandsynlighedsfordelingen for en *frekvensfordeling* eller en *tæthedsfunktion*, og skriver i da fald $f(a) = P(X = a)$. Lige sådan som vi i forrige afsnit fik brug for den kumulerede frekvens, så er den meget anvendelsesbar for en stokastisk variabel, hvor vi kalder den kumulerede frekvens for en *fordelingsfunktion* og i da fald skriver $F(a) = P(X \le a)$.

Definition 13

Lad X være en stokastisk variabel. Så definerer vi

1) *tæthedfunktionen f* ved $f(a) = P(X = a)$
2) *fordelingsfunktionen F* ved $F(a) = P(X \le a)$

Hvis vi vil illustrere frekvensfunktionen, så tegnes et stolpediagram, og hvis vi vil illustrere fordelingsfunktionen, så tegnes en sumkurve, som vi gjorde det i statistik afsnittet i det diskrete tilfælde.

Fordelingsfunktionen er så vigtig, at i afsnittene om binomialfordelingen og normalfordelingen vil vi faktisk kun regne med fordelingsfunktionen, når vi løser et problem.

Definition 14

Lad X være en stokastisk variabel med m udfald, så defineres *middelværdien* af X til at være det vejede gennemsnit $E(X)$ (eller μ)

$$E(X) = X(u_1)\cdot p(u_1) + X(u_2)\cdot p(u_2) + ... + X(u_m)\cdot p(u_m)$$

Sætning 15

Lad X være en stokastisk variabel med værdierne $t_1, t_2, \dots, t_n$.
Så er

$$E(X) = t_1 \cdot P(X = t_1) + ... + t_n \cdot P(X = t_n)$$

Bevis

Ifølge definition 14 får vi

$$E(X) = X(u_1)\cdot p(u_1) + X(u_2)\cdot p(u_2)+...+X(u_m)\cdot p(u_m)$$

Vi samler leddene, så alle de udfald, der fik knyttet tallet t_1 til sig, kommer til at stå forrest, og med t_1 sat udenfor en parentes, og ligeledes med de andre led.

$$E(X) = t_1 \cdot \left(P(v_1) + P(v_2)+...+P(v_k)\right)+...+t_n\left(P(w_1) + P(w_2)+...+P(w_l)\right)$$

hvor $v_1, v_2, \dots, v_k$ alle er udfald, hvortil X knyttes til tallet t_1. Vi ser nu, at

$$P(X = t_1) = P(v_1) + P(v_2)+...+P(v_n)$$

hvoraf sætningen følger.

Eksempel

Lad os udføre et stokastisk eksperiment, nemlig at kaste to terninger. Et udfald kan beskrives som et ordnet par (s,t), hvor s er det første øjental og t det andet øjental. Udfaldsrummet bliver mængden

$$
\begin{aligned}
U = \{ &(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\
&(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\
&(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\
&(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\
&(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\
&(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \}
\end{aligned}
$$

Heldigvis er sandsynlighedsfunktionen mere simpel:

$P((s,t)) = \frac{1}{36}$, for alle værdier af s og t.

Vi kan nu definere to stokastiske variable, X og Y ved

$$X((s,t)) = s + t$$
$$Y((s,t)) = s - t$$

Disse stokastiske variable har tæthedsfunktionerne:

t	2	3	4	5	6	7
$P(X=t)$	1/36	2/36	3/36	4/36	5/36	6/36
t	8	9	10	11	12	
$P(X=t)$	5/36	4/36	3/36	2/36	1/36	

t	-5	-4	-3	-2	-1	0
$P(Y=t)$	1/36	2/36	3/36	4/36	5/36	6/36
t	1	2	3	4	5	
$P(Y=t)$	5/36	4/36	3/36	2/36	1/36	

-dette fandt du jo selv ud af i opgaverne 2.6 og 2.7.

Ved direkte udregning ser man, at middelværdierne er

$E(X) = 7$ og $E(Y) = 0$

(En konsekvens af dette er, at man i Århusiansk Matador ikke kan forvente at komme ret langt væk fra startpositionen...)

For regning med stokastiske variable gælder der:

Sætning 16

Lad X og Y være stokastiske variable, og c en konstant.

1) $E(c) = c$ når c er et tal

2) $E(X) = c \cdot E(X)$

3) $E(X + Y) = E(X) + E(Y)$

4) $E(X + c) = E(X) + E(c)$

Bevis

1) $$E(c) = c(u_1) \cdot P(u_1) + c(u_2) \cdot P(u_2) + \ldots + c(u_m) \cdot P(u_m)$$
$$c \cdot P(u_1) + c \cdot P(u_2) + \ldots + c \cdot P(u_m) =$$
$$c \cdot \left(P(u_1) + P(u_2) + \ldots + P(u_m) \right) = c$$

Her skal c opfattes som en *konstant stokastisk variabel*,

2) $$E(cX) = cX(u_1) \cdot P(u_1) + cX(u_2) \cdot P(u_2) + \ldots + cX(u_m) \cdot P(u_m) =$$
$$c \cdot \left(X(u_1) \cdot p(u_1) + X(u_2) \cdot p(u_2) + \ldots + X(u_m) \cdot p(u_m) \right) =$$
$$c \cdot E(X)$$

3) og 4) er en øvelse

Eksempel

Vi fortsætter terninge-eksemplet fra før. Vi har to stokastiske variable:

$S((s,t)) = s$ og $T((s,t)) = t$

(S betyder 'Hvad er øjentallet på den første terning. Pyt med den anden')

Man kan let beregne, at $E(S) = E(T) = 3{,}5$.

Vi har nu, at

$$X = S + T$$

og derfor

$$E(X) = E(S + T) = E(S) + E(T) = 3{,}5 + 3{,}5 = 7$$

ganske som før.
Tilsvarende er

$$Y = S - T$$

og

$$E(Y) = E(S - T) = E(S) - E(T) = 3{,}5 - 3{,}5 = 0$$

Alle stokastiske variables sandsynlighedsfordeling vil fordele sig omkring middelværdien, men måden den er fordelt på kan variere temmeligt meget.

Hvis man f.eks. lader en stokastisk variabel angive højden af en tilfældig udvalgt person i et rum, og man får at vide at middeltallet er 1,65 m, så kan højdefordelingen jo godt være meget forskellig, som man kan se af nedenstående to sandsynlighedsfordelinger.

Lad os f.eks. sige at man selv er 1,65 m høj, og får at vide at middeltallet er 1,65 m blandt personerne i et lokale. Så er man jo glad, og tror at man nok skal kunne føle sig hjemme i lokalet. Men træder man ind i lokalet med den venstre sandsynlighedsfordeling, så vil man jo hurtig finde udad, at man faktisk er en outsider i det lokale, hvilket vil være knap så meget tilfældet i et lokale med den højre sandsynlighedsfordeling. Og skulle man træde ind i et lokale hvor middeltallet også er den eneste højde, så vil man være lykkelig.

Dette lille eksempel skulle gerne vise, at det ikke er nok at kende middeltallet. Vi er nødt til at have et mål for, hvor meget den stokastiske variabel varierer omkring middeltallet. Et sådant mål kan laves på mange måder, men den mest brugte er den vægtede sum af kvadratet på afvigelserne til middelværdien. Herved opnås, at man giver megen betydning til de værdier, som ligger langt fra middelværdien, mens de værdier som ligger tæt på middelværdien næsten ikke bidrager til summen. Vi får altså et mål, som vi kalder variansen, som ser ud som følger

$$\mathrm{var}(X) = (X(u_1) - \mu)^2 \cdot P(u_1) + (X(u_2) - \mu)^2 \cdot P(u_2) + \ldots + (X(u_m) - \mu)^2 \cdot P(u_m)$$

Det minder jo meget om definitionen på middelværdi

$$E(X) = X(u_1) \cdot P(u_1) + X(u_2) \cdot P(u_2) + \ldots + X(u_m) \cdot P(u_m)$$

Blot er er f.eks. $X(u_1)$ erstattet med $(X(u_1) - m)^2$, så vi foretager følgende definition

Definition 17

Lad X være en stokastisk variabel med middelværdien μ, så definerer vi

1) Variansen af X som $\mathrm{var}(X) = E\left((x-\mu)^2\right)$

2) Spredningen af X som $\sigma(X) = \sqrt{\mathrm{var}(X)}$

Der er tradition for at bruge kvadratroden af variansen som et mål, der kaldes spredningen. Det skyldes især, at man i statistiske sammenhænge ofte skal regne med kvadratroden af variansen, og så kan man jo lige så give det et navn.

Sætning 18 (FS)

Lad X være en stokastisk variabel så er

$$\mathrm{var}(X) = E(X^2) - \left(E(X)\right)^2$$

Bevis:

$$\mathrm{var}(X) = E\left((X-\mu)^2\right)$$

$$\mathrm{var}(X) = E\left((X^2+\mu^2-2\cdot X\cdot\mu\right)$$

$$\mathrm{var}(X) = E(X^2)+E(\mu^2)+E(-2\cdot X\cdot\mu)$$

$$\mathrm{var}(X) = E(X^2)+\mu^2-2\cdot\mu\cdot E(X))$$

$$\mathrm{var}(X) = E(X^2)+\mu^2-2\cdot\mu\cdot\mu$$

$$\mathrm{var}(X) = E(X^2)-\mu^2$$

Eksempel

Den stokastiske variabel X har tæthedsfunktionen

t	-1	0	1	2
$P(X=t)$	0,5	0,2	0,1	0,2

Vi finder middelværdien, variansen og spredningen:

$$E(X) = (-1)\cdot 0{,}5 + 0\cdot 0{,}2 + 1\cdot 0{,}1 + 2\cdot 0{,}2 = -0{,}2$$

$$E(X^2) = (-1)^2 \cdot 0{,}5 + 0^2 \cdot 0{,}2 + 1^2 \cdot 0{,}1 + 2^2 \cdot 0{,}2 = 1{,}4$$

$$\text{var}(X) = E(X^2) - E(X)^2 = 1{,}4 - (-0{,}2)^2 = 1{,}36$$

$$\sigma(X) = \sqrt{\text{var}(X)} = \sqrt{1{,}36} = 1{,}167$$

Sætning 19 (FS)

Lad X være en stokastisk variabel så gælder

1) $\text{var}(aX) = a^2 \cdot \text{var}(X)$

2 $\text{var}(X + b) = \text{var}(X)$

Bevis

1)
$$\text{var}(aX) = E((aX)^2) - \left(E(aX)\right)^2$$
$$\text{var}(aX) = E(a^2 X^2) - \left(aE(X)\right)^2$$
$$\text{var}(aX) = a^2 E(X^2) - a^2 \left(E(X)\right)^2$$
$$\text{var}(aX) = a^2 \cdot \left(E(X^2) - \left(E(X)\right)^2\right)$$
$$\text{var}(aX) = a^2 \cdot \text{var}(X)$$

2)
$$\text{var}(X + b) = E((X + b)^2) - \left(E(X + b)\right)^2$$
$$\text{var}(X + b) = E(X^2 + b^2 + 2xb) - \left(E(X) + b\right)^2$$
$$\text{var}(X + b) = E(X^2) + E(b^2) + E(2xb) - \left(\left(E(X)\right)^2 + b^2 + 2bE(X)\right)$$
$$\text{var}(X + b) = E(X^2) + b^2 + 2bE(x) - \left(E(X)\right)^2 - b^2 - 2bE(X)$$
$$\text{var}(X + b) = E(X^2) - \left(E(X)\right)^2$$
$$\text{var}(X + b) = \text{var}(X)$$

Eksempel

Lad X og Y være stokastiske variable med

$$E(X) = 10 \quad og \quad \text{var}(X) = 2 \quad og \quad Y = 3X + 8$$

Find $E(Y)$ og var(Y) :

$$E(Y) = E(3X+8) = 3 \cdot E(x) + 8 = 3 \cdot 10 + 8 = 38$$
$$\text{var}(Y) = \text{var}(3x+8) = 3^2 \cdot \text{var}(X) = 9 \cdot 2 = 18$$

Eksempel

Lad X og Y være stokastiske variable med

$$E(Y) = 20 \quad og \quad \sigma(Y) = 3 \quad og \quad Y = 5X - 10$$

Find $E(X)$ og var(X)

$$20 = E(Y) = E(5x-10) = 5 \cdot E(X) - 10$$
$$E(X) = 6$$

$$3 = \sigma(Y) = \sigma(5X-10) = |5| \cdot \sigma(X)$$
$$\sigma(X) = 0{,}6$$
$$\text{var}(X) = (\sigma(X))^2 = 0{,}6^2 = 0{,}36$$

Opgaver

3.1 Lad X være en stokastisk variabel med fordelingen

t	-3	-1	0	1	2	3	5	8
$f(t)$	0,1	0,2	0,15	0,2	0,1	0,15	0,05	0,05

Bestem nedenstående størrelser:

a) $P(X < 0)$ b) $P(X \leq 0)$ c) $P(X < -0{,}5)$

d) $P(X > 2)$ e) $1 - P(X > 2)$ f) $P(X \leq 2)$

g) $P(X = -3 \mid X \leq 0)$ h) $P(X \geq 3 | X \geq 0)$ i) $P(X = 2 | x < 1)$

j) $E(X)$ k) $\text{var}(X)$ l) $\sigma(X)$

3.2 Lad X og Y være to stokastiske variable med den fælles fordeling angivet nedenfor. Hermed menes, at tabellen angiver tallene $P(X = s \text{ og } Y = t)$

s/t	-1	0	2	6
-2	1/9	1/27	1/27	1/9
1	2/9	0	1/9	1/9
3	0	0	1/9	4/27

Bestem

a) tæthedsfunktionen for X
b) tæthedsfunktionen for Y
c) $P(Y \text{ er lige})$
d) tæthedsfunktionen for XY
e) $E(X)$, $E(Y)$ og $E(XY)$
f) $\text{var}(X)$, $\text{var}(Y)$ og $\text{var}(XY)$

9.4 Kombinatorik

Kombinatorik handler om forskellige måder, som man kan tælle på. Det er ofte nyttigt at kende nogle tællemetoder, når man skal regne en sandsynlighed ud. Det kommer sig blandt af, at den mest oplagte måde at regne en sandsynlighed for en hændelse ud på, **når alle udfald er lige sandsynlige**, er ved

$$P(succes) = \frac{\text{antal udfald med succes}}{\text{antal mulige udfald}}$$

Det skal forstås sådan, at man er i en situation, hvor man har en række mulige udfald, men kun er interesseret i nogle bestemte udfald, som man følgelig opfatter som succeser. Det må jo så være rimeligt at sige, at sandsynligheden for den situation er antallet af udfald, som er gunstige divideret med det samlede antal mulige udfald.

I øvrigt kalder man et sandsynlighedsfelt, hvori alle udfald er lige sandsynlige, for et *symmetrisk udfaldsrum*.

Eksempelvis er det situationen, når man slår med en terning, og man ønsker at få en 1´er, så er antal gunstige udfald 1 og samlede antal udfald er 6, altså er sandsynligheden for en 1´er en $\frac{1}{6}$.

Så spørgsmålet er da, hvordan man regner disse antal ud på en let måde. Denne ædle videnskab (at regne antal ud) kaldes *kombinatorik*.

Kombinatorikkens fundamentale regel er:

Sætning 20 (multiplikationsprincippet)

Lad et valg bestå af n delvalg, hvor hvert delvalg kan fortages på et antal måder, så er valgets samlede antal måder lig med produktet af delvalgenes antal måder.

Det illustreres lettest ved at bruge et skema af typen

Valg	Valgbeskrivelse	Antal måder valget kan gøres på
1	beskrivelse 1	m_1
2	beskrivelse 2	m_2
. .		. .
n	beskrivelse n	m_n
Samlet valg		$m_1 \cdot m_2 \cdot \ldots \cdot m_n$

Sætning 21

Antallet af rækkefølger n objekter kan placeres i er

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$$

hvor $n!$ læses som “n fakultet”

Bevis

Beviset følger direkte af multiplikationsprincippet, idet vi laver n valg og i hvert valg vælger et objekts placering.

Valg	Valgbeskrivelse	Antal måder valget kan gøres på
1	vælg 1. objekts placering	n
2	vælg 2. objekts placering	n-1
3	vælg 3. objekts placering	n-2.
. .		. .
n	vælg sidste objekts placering	1
Samlet valg		$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$

□

Bemærk at 0!=1, idet man kan placere 0 objekter på netop én måde.

Eksempel

På hvor mange måder kan fire elever være placeret på? Det kan man jo passende bruge et frikvarter på at undersøge! Men man kan selvfølgelig også blot bruge sætningen, for den siger, at svaret er 4!, som er det samme som $4 \cdot 3 \cdot 2 \cdot 1 = 24$

De fleste lommeregnere har også en tast med “!” på, som kan udregne 4 ! direkte.

Facit virker rimeligt, når man tænker på, at den første kan placere sig på 4 steder, den næste på 3 steder, den næste igen på 2 steder og til den sidste er der kun 1 stol.

Sætning 22

Antal måder man kan vælge *r* objekter ud af *n* objekter er givet ved

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

hvor $K(n,r)$ læses som "k n komma r". Den kaldes også en binomialkoefficient.

Bevis

Beviset går ud på at vælge rækkefølgen af n objekter på en bestemt måde vel vidende at resultatet pr sætning 21 skal give n!
Man foretager først et valg af *r* objekter ud af de *n* objekter, og uden at kende antal måder det kan gøres på, så døber vi dette antal for $K(n,r)$.
Derefter så vælges rækkefølgen af de *r* objekter og til sidst så vælges rækkefølgen af de $n-r$ objekter.
Ved brug af multiplikationsprincippet får vi

Valg	Valgbeskrivelse	Antal måder valget kan gøres på
1	vælg r objekter ud af n objekter	$K(n,r)$
2	vælg en rækkefølge blandt de *r* objekter	$r!$
3	vælg en rækkefølge blandt de *n-r* objekter	$(n-r)!$.
Samlet valg		$n!$

Dvs. multiplikationsprincippet giver

$$K(n,r) \cdot r! \cdot (n-r)! = n!$$

Og sætningen følger ved isolering af $K(n,r)$

Sætning 24

1) $K(n,0) = 1$
2) $K(n,1) = n$
3) $K(n,r) = K(n,n-r)$

Bevis

1) $K(n,0) = \dfrac{n!}{0!\cdot(n-0)!} = \dfrac{n!}{n!} = 1$

2) $K(n,1) = \dfrac{n!}{1!\cdot(n-1)!} = \dfrac{n\cdot(n-1)!}{1\cdot(n-1)!} = n$

3) $K(n,r) = \dfrac{n!}{r!\cdot(n-r)!} = \dfrac{n!}{(n-r)!\cdot r!} = K(n,n-r)$

□

Eksempel

I kortspillet *Poker* skal man udvælge 5 kort blandt 52 mulige.
Hvor mange måder kan dette gøres på?

Ifølge sætning 23 kan dette gøres på i alt

$$K(52,5) = \frac{52!}{5!\cdot(52-5)!} = \frac{52!}{5!\cdot 47!} = \frac{8{,}0658\cdot 10^{67}}{120\cdot 2{,}586\cdot 10^{59}} = 2598960$$

En af de såkaldte *pokerhænder* er *fuldt hus*, som består af 3 kort med samme værdi og 2 kort med en anden værdi. F.eks. er hånden

spar 2, klør 2, ruder 2, hjerner 10, klør 10

et fuldt hus.

Hvor mange hænder med fuldt hus findes der?

For at besvare dette skal man lave nogle valg. Disse kan f.eks. være

1: vælg 2 talværdier ud af de 13 mulige
2: vælg 3 farver ud af de 4 mulige til de tre første kort
3: vælg 2 farver ud af 4 mulige til de to sidste kort

I alt får man ifølge multiplikationsprincippet, at der er

$$K(13,2)\cdot K(4,3)\cdot K(4,2) = \frac{13!}{2!\cdot 11!}\cdot\frac{4!}{3!\cdot 1!}\cdot\frac{4!}{2!\cdot 2!} = 1872$$

Sandsynligheden for at fuldt hus, når man spiller poker, er derfor

$$\frac{1872}{2598960} = 0{,}00072 = 0{,}072\%$$

Opgaver

4.1 Den genetiske kode består af blokke af 3 nukleotider. Hvert nukleotid kan være af fire forskellige slags, A, C, T, G.
Hvor mange 'bogstaver' er der i den genetiske kode?

4.2 Rikke har været på udsalg og har købt 50 cykelpumper, hvoraf 10 desværre er defekte. Hun forærer nu Berit 10 tilfældigt udvalgte cykelpumper.

a) Hvad er sandsynligheden for, at Berit får alle 10 defekte cykelpumper?

b) Hvad er sandsynligheden for, at Berit ikke får nogen defekt cykelpumpe?

4.3 Berit har også været på udsalg og har købt 100 opvaskebørster. Desværre er 5 af opvaskebørsterne i stykker. Som tak for cykelpumperne forærer hun Rikke 2 opvaskebørster.
Hvad er sandsynligheden for, at begge Rikkes opvaskebørster er i stykker?

4.4 Bestem antallet af følgende pokerhænder:

a) Royal Straight Flush (Es, konge, dame, knægt, ti i samme farve)

b) Royal Flush (Es, konge, dame, knægt, ti i vilkårlige farver)
(husk at fratrække antallet af Royal Straight Flushs)

c) Fire af en slags

d) Tre af en slags (husk at fratrække antallet af fuldt hus)

e) To par

f) Et par

g) Flush (5 kort i samme farve)

h) Straight (fem kort i rækkefølge, vilkårlige farver)

i) Straight flush (5 kort i rækkefølge i den samme farve)

9.5 Binomialfordeling

Man bruger en binomialfordeling, når man foretager et eksperiment med præcist to udfald, og gentager eksperimentet et vist antal gange. Det kan f.eks. være at slå 10 gange med en terning og i hvert slag at notere sig om det var en etter eller ej. Det kan også være stikprøveundersøgelse, hvor man spørger 100 personer, om de holder Jyllandsposten - ja eller nej. Lad os se på et eksempel.

Eksempel

Hans har kastet mange gange med en terning og fundet ud af, at en $\frac{1}{6}$ af kastene viser "1" , og dermed så må $\frac{5}{6}$ af kastene vise "ikke 1'ere". Hans funderer nu over, hvor mange "1'ere" han sandsynligvis får, hvis han kaster terningen 2 gange.

Det må give en tabel som følgende

Udfald	Kast 1	Kast 2
Udfald 1	*Ikke* 1	*Ikke* 1
Udfald 2	1	*Ikke* 1
Udfald 3	*Ikke* 1	1
Udfald 4	1	1

Da kastene er uafhængige, så ganger man blot sandsynligheden for udfaldet i Kast 1 med sandsynligheden for udfaldet i Kast 2.

$$P(Udfald\ 1) = P(Ikke\ 1)\cdot P(Ikke\ 1) = \frac{5}{6}\cdot\frac{5}{6} = \left(\frac{1}{6}\right)^0\cdot\left(\frac{5}{6}\right)^2$$

$$P(Udfald\ 2) = P(1)\cdot P(Ikke\ 1) = \frac{1}{6}\cdot\frac{5}{6} = \left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1$$

$$P(Udfald\ 3) = P(Ikke\ 1)\cdot P(1) = \frac{5}{6}\cdot\frac{1}{6} = \left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1$$

$$P(Udfald\ 4) = P(1)\cdot P(1) = \frac{1}{6}\cdot\frac{1}{6} = \left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^0$$

Hvis man er ligeglad med rækkefølgen, dvs. hvornår man får "1'ere", så får man

$$P(0\quad \text{"1'}ere\text{"}) = \left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^0$$

$$P(1\quad \text{"1'}ere\text{"}) = \left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1 + \left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1 = 2\cdot\left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1$$

$$P(2\quad \text{"1'}ere\text{"}) = \left(\frac{1}{6}\right)^0\cdot\left(\frac{5}{6}\right)^2$$

Ansporet af vores eksempler, så lad os se på en mere generel situation.

Definition 25

Et *binomialeksperiment* er et eksperiment bestående af n ens og af hinanden uafhængige deleksperimenter. Hvert deleksperiment skal have netop to udfald, 'succes' og 'fiasko', og 'succes' skal indtræde med sandsynligheden p.

Tallet n kaldes ofte *antalsparametren*, og p kaldes *sandsynlighedsparametren*.

Typiske binomialeksperimenter er gentagelser at et simpelt forsøg med to udfald.

Definition 26

En *binomialfordelt stokastisk variabel X* er en stokastisk variabel, som tæller antallet af succeser i et binomialeksperiment.

Vi siger, at 'X er binomialfordelt med antalsparametren n og sandsynlighedsparametren p', og skriver

$$X \approx b(n, p)$$

Sætning 27 (FS)

Lad X være en binomialfordelt stokastisk variabel, $X \approx b(n, p)$.

Så gælder $\quad P(x = r) = K(n,r) \cdot p^r \cdot (1-p)^{n-r}$

for $r = 0, 1, 2, \ldots, n$

Bevis

At $X = r$ betyder at der i de n deleksperimenter skal være placeret r succeser og dermed $n - r$ fiaskoer. Det kan f.eks. være et udfald af typen

$$U = (\underbrace{s, f, f, f, s, s, s, f, s, \ldots.. s, f}_{r \quad successer})$$

Sandsynligheden for dette ene udfald fås via multiplikationsprincippet

$$P(U) = P(s) \cdot P(f) \cdot P(f) \cdot P(f) \cdot \quad \ldots\ldots \quad \cdot P(f)$$

$$= p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot \quad \ldots\ldots \quad \cdot (1-p)$$

$$= p^r \cdot (1-p)^{n-r}$$

Dette er altså sandsynligheden for ét udfald med r succeser; men dem kan der jo være mange af! (I eksempel ??? var der f.eks. 2 udfald med 1 succes). Så på hvor mange forskellige måder kan de r succeser være placeret blandt de n gentagelser. Dette løses ved binomialkoefficienter Samlet får vi da, at der er $K(n,r)$ udfald med r succeser og hver med sandsynligheden $p^r(1-p)^{n-r}$.

Dermed får vi

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

Bemærk at sandsynlighedsfunktionen ofte skrives som $P(X=r)=b(n,p;r)$.

Hvis man skal undersøge om et problem kan beskrives ved en binomialfordelt stokastisk variabel, så skal man undersøge to ting

1) Har man et eksperiment med præcist to udfald - med sandsynligheden p for succes?

2) Gentages eksperimentet et vist antal gange - n ?

I bekræftende fald, så fastlægger vi

X er den stokastiske variabel, som angiver antallet af succeser

Og vi kan da sige at

X er binomialfordelt med parametre n, p : $X\approx(n,p)$

Eksempel

Lad os vende tilbage til kast med en terning, hvor man ønsker at slå "1´ere". Hvis man er ligeglad med rækkefølgen af sine "1'ere", så kan man i stedet for nøjes med at tælle antallet af "1'ere". Det gøres lettest ved indførelsen af en stokastisk variabel X

X: tæller antal "1'ere" ud af n kast.

Sandsynlighederne i eksemplet tidligere, hvor $n=2$ (der er jo 2 kast), bliver da

$$P(X=0)=P(0\quad "1'ere")=K(2,0)\cdot\left(\frac{1}{6}\right)^0\cdot\left(\frac{5}{6}\right)^2=0{,}694$$
$$P(X=1)=P(1\quad "1'ere")=K(2,1)\cdot\left(\frac{1}{6}\right)^1\cdot\left(\frac{5}{6}\right)^1=0{,}278$$
$$P(X=2)=P(2\quad "1'ere")=K(2,2)\cdot\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^0=0{,}028$$

Det svarer jo meget fint i overensstemmelse med det vi fik ved simpelthen at opskrive samtlige mulige udfald. Men hvis vi nu havde 30 kast, dvs. $n=30$, så ville en sådan fremgangsmetode jo blive helt uoverskueligt, og så er det godt at have sin binomialformel.

$$P(X=0)=P(0\quad \text{"1'ere"})=K(30,0)\cdot\left(\frac{1}{6}\right)^{0}\cdot\left(\frac{5}{6}\right)^{30}$$

$$P(X=18)=P(18\quad \text{"1'ere"})=K(30,18)\cdot\left(\frac{1}{6}\right)^{18}\cdot\left(\frac{5}{6}\right)^{12}$$

$$P(X=30)=P(30\quad \text{"1'ere"})=K(30,30)\left(\frac{1}{6}\right)^{30}\cdot\left(\frac{5}{6}\right)^{0}$$

Eller hvis $n=1000$

$$P(X=300)=P(300\quad \text{"1'ere"})=K(1000,300)\cdot\left(\frac{1}{6}\right)^{300}\cdot\left(\frac{5}{6}\right)^{700}$$

Bemærk at summen af eksponenterne hele tiden giver antallet af kast. Det skulle det jo også gerne, da eksponenten angiver antal kast, som har en given sandsynlighed.

Eksempel

25 tilfældige personer spørges, om de har været i bad i dag. Man ved, at i gennemsnit går 70% af alle personer i bad hver dag.

Lad os undersøge om situationen kan beskrives ved en binomialfordeling

1) Vi har et eksperiment med to udfald, der består i at undersøge, om en person har været i bad i dag. Sandsynligheden for succes er $p=0{,}7$.
2) Eksperimentet gentages $n=25$ gange.

Vi fastlægger en stokastisk variabel

X Tæller antal personer, der har været i bad i dag

Vi har nu, X er binomialfordelt med parametre $n=25$ og $p=0{,}7$

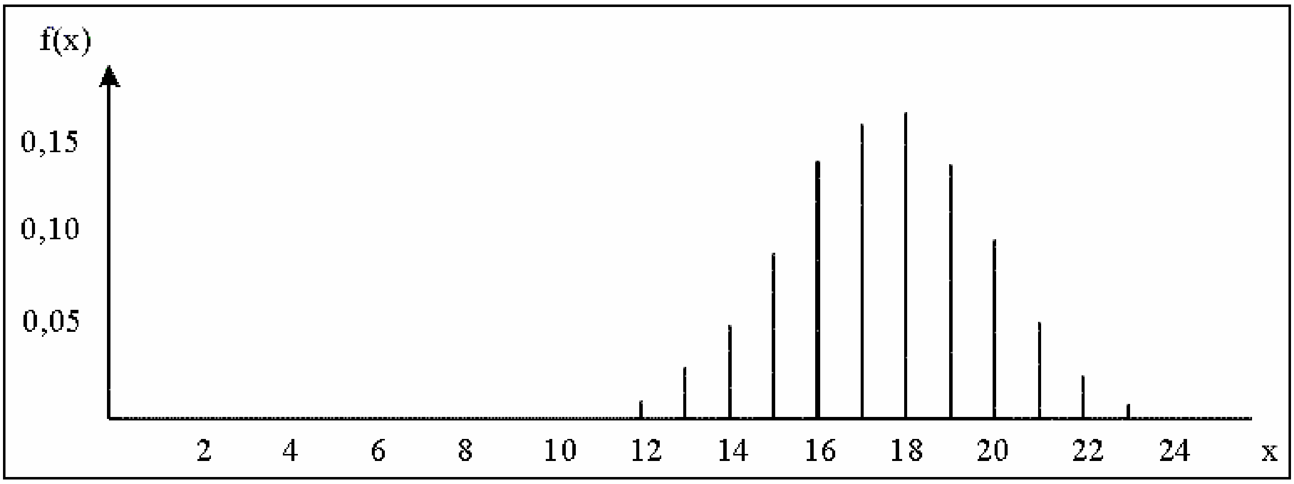
Vi er nu klar til at regne løs

Lad os finde sandsynligheden for at præcist 15 svarer ja

$$P(X=15)=K(25,15)\cdot 0{,}7^{15}\cdot 0{,}3^{10}=3268760\cdot 0{,}7^{15}\cdot 0{,}3^{10}=0{,}091614$$

Så der er altså 9,164% chance for at præcist 15 har været i bad i dag.

Man kan lettere overskue sin sandsynlighedsfordeling, hvis man tegner den. Her er den f.eks. tegnet for n=25 og p=0,7, dog er sandsynligheder under 0,005 ikke med.



Faktisk er man i stand til at vurdere, hvilken x-værdi der har den største sandsynlighed for at indtræffe ved at beregne $n \cdot p = 25 \cdot 0{,}7 = 17{,}5$. Det er nemlig et af de to hele tal, der er tættest på $n \cdot p$. Dvs her er det 17 eller 18, og det passer jo meget godt med ovenstående sandsynlighedsfordeling.

Udregner vi faktisk

$$P(X = 17) = K(25{,}17) \cdot 0{,}7^{17} \cdot 0{,}3^{8} = 0{,}165$$
$$P(X = 18) = K(25{,}18) \cdot 0{,}7^{18} \cdot 0{,}3^{7} = 0{,}171$$

ser vi, at det mest sandsynlige antal 'badere' er 18.

Ofte bliver man bedt om at beregne f.eks. $P(X \leq 5)$, som umiddelbart er let nok da

$$P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

Men det tager lang tid at gøre det i praksis, hvis man skal beregne på samme måde som i foregående eksempel. For at lette arbejdet med at udregnes sådanne kumulerede binomialsandsynligheder, så er de tabellagt i Sigmatabellen fra side 2 til side 25, hvor én søjle svarer til én fordeling, så selv om tabellen er stor, så er det kun et spørgsmål om at finde den rigtige søjle, og det vil vi se på i et eksempel.

Vi husker på, at de kumulerede frekvenser for en stokastisk variabel, svarer til fordelingsfunktionen for den stokastiske variabel, så binomialfordelingstabellerne er faktisk tabeller over fordelingsfunktionerne. Vi kunne derfor i stedet for $P(X \leq 5)$ skrive $F(5)$.

Bemærk at

$$P(X \leq 5) = P(X \leq 4) + P(X = 5)$$
$$P(X = 5) = P(X \leq 5) - P(X \leq 4)$$

Eller mere generelt

$$P(X = r) = P(X \leq r) - P(X \leq r - 1)$$

Så tabellen kan også bruges til at beregne en enkelt binomialsandsynlighed

Skal vi beregne sandsynligheden for et interval, så får vi

$$P(2 \le X \le 4) = P(X = 2) + P(X = 3) + P(X = 4) =$$
$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) - (P(X = 0) + P(X = 1)) =$$
$$P(X \le 4) - P(X \le 1)$$

Eller mere generelt

$$P(a \le X \le b) = P(X \le b) - P(X \le a - 1)$$

Desuden gælder

$P(X \le 5) + P(X \ge 6) = 1$, så
$$P(X \ge 6) = 1 - P(X \le 5)$$

Eller mere generelt

$$P(X \ge r) = 1 - P(X \le r - 1)$$

Eksempel

Vi regner videre på bade-eksemplet, hvor X var binomialfordelt med parametre $n = 25$ og $p = 0{,}7$, og finder ved tabelopslag forskellige sandsynligheder.
Tabellen vi skal bruge står i Sigma på side 12 og 13.

1) Ved Tabelopslag ($n = 25$) ses

$$P(X \le 14) = 0{,}098 = 9{,}8\%$$

j \ p	0,7
14	0,098

2) Ved Tabelopslag ($= 25$) ses

$$P(X \le 15) = 0{,}189 = 18{,}9\%$$

j \ p	0,7
15	0,189

3) $P(X = 15) = P(X \le 15) - P(X \le 14) = 18{,}9\% - 9{,}8\% = 9{,}1\%$
hvilket passer meget godt med udregningen .

4) $P(X \ge 16) = 1 - P(X \le 15) = 1 - 0{,}189 = 0{,}811 = 81{,}1\%$

5) Ved tabelopslag $(= 25)$ ses

j \ p	0,7
16	0,323
20	0,910

$$P(17 \leq X \leq 20) = P(X \leq 20) - P(X \leq 16)$$

$$= 0{,}910 - 0{,}323 = 0{,}587 = 58{,}7\%$$

(Det skal nævnes, at 4) kan løses en anelse mere direkte ved gå ind i tabellen nederst fra højre af; men som man så i 4) er det ikke nødvendigt.

Ved Tabelopslag $(n = 25)$ ses

0,811	16
0,7	p \ j

$$P(X \geq 16) = 0{,}811 = 81{,}1\%$$

Til sidst skal vi kigge lidt på middelværdien og variansen af den binomialfordelt stokastisk fordeling. Nedenstående sætning anføres uden bevis:

Sætning 28 (FS)

Lad X være en binominalfordelt stokastisk variabel med parametrene n,p . Så gælder

1) $E(X) = n \cdot p$
2) $\text{var}(X) = n \cdot p \cdot (1 - p)$
3) $s(X) = \sqrt{n \cdot p \cdot (1 - p)}$

Eksempel

Lad X være binominalfordelt med parametre n=120 og $p = \frac{1}{6}$.

Så er middelværdien $E(X) = n \cdot p = 120 \cdot \frac{1}{6} = 20$

og varianten $\text{var}(X) = n \cdot p \cdot (1 - p) = 120 \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right) = 16{,}67$

og spredningen $\sigma(X) = \sqrt{\text{var}(X)} = \sqrt{16{,}\overline{67}} \approx 4{,}082$.

Eksempel

Lad X være binominalfordelt med parametre n, p, med $\sigma(X) = 5$ og $E(X) = 30$.
Lad os finde n og p.

$$5 = \sigma(X) = \sqrt{\operatorname{var}(X)} = \sqrt{n \cdot p \cdot (1 - p)}$$

eller

$$25 = n \cdot p \cdot (1 - p)$$

Desuden har vi

$$30 = E(X) = n \cdot p$$

eller

$$p = \tfrac{30}{n}$$

Det indsætter vi i den første ligning

$$25 = n \cdot \tfrac{30}{n} \cdot (1 - \tfrac{30}{n}) = 30 - \tfrac{900}{n}$$

$$-5 = -\tfrac{900}{n}$$

$$n = 180$$

Dette indsætter vi i udtrykket for $E(X)$

$$30 = 180 \cdot p$$

og får

$$p = \tfrac{1}{6}$$

Altså, $X \approx b(180, \tfrac{1}{6})$

Opgaver

5.1 Thomas spiller computerspillet *Doom*. Han er ved at blive slået ihjel af en dæmon, og skyder derfor på den. Han har kun 8 skud tilbage, og da han er dårlig til at skyde, så er sandsynligheden for, at han rammer dæmonen kun 0,25 for hvert skud. Dæmonen skal rammes af mindst 2 skud, før den dør.
Bestem sandsynligheden for, at Thomas overlever mødet med dæmonen.

5.2 I 2z ryger 25% af eleverne. Til gruppearbejde i historie udtages en gruppe på 3 tilfældige elever fra 2z. Bestem sandsynligheden for, at

a) alle i gruppen ryger
b) mindst en i gruppen ryger
c) Bestem det forventede antal af rygere i gruppen.

5.3 Af den danske befolkning har 64% blå øjne
Bestem sandsynligheden for, at der i en tilfældig formsamling af 20 danskere er netop 9, der har blå øjne.

5.4 Pladderballe Bodega laver sodavandsrafling. Man slår med tre terninger, og alt efter antallet af seksere sker der følgende

0) 0 seksere - man skal betale fuld pris
1) 1 sekser - man får 2 kroners rabat
2) 2 seksere - man får 5 kroners rabat
3) 3 seksere - man får sodavanden gratis (15 kr. rabat)

a) Bestem sandsynlighederne for at få 0, 1, 2 eller 3 seksere.
b) Bestem den forventede rabat.
c) Bodegaen indkøber sodavand til 10 kr. stykket. Hvad er den gennemsnitlige fortjeneste på hver sodavand?

9.6 Normalfordeling

Til at starte på havde vi et eksempel med nittehoveder, hvor vi havde en stor mængde data, som vi systematiserede ved at dele dataene op i grupper, som hver især havde en vis hyppighed. Vi fandt frekvenserne og ved at tegne disse, så fremkom et klokkeformet histogram, hvor vi kunne aflæse medianen til stort set at være middeltallet.

Denne situation er normal for store mængder af data, at de grupperer sig jævnt faldende på begge sider af middelværdien. Vi vil derfor prøve matematisk at lave en fordeling, som beskriver en sådan situation. Denne fordeling kalder vi selvfølgelig for en normalfordeling.

Man kan finde mange fordelinger, som er mere eller mindre klokkeformede, men fælles for dem alle er kravet om, at fordelingen skal være kontinuert. Et krav, som klart fremkommer af den bløde kurve, som klokkeformen viser. Det betyder, at den stokastiske variabel kan antage uendeligt mange x-værdier, hvilket igen betyder, at der må være uendeligt mange udfald.. Desværre dur vores gamle definition på et stokastisk eksperiment ikke mere, for i den definition var der kun et endeligt antal udfald. Vi er derfor nødt til at lave en nu definition.

Definition 29

Et *kontinuert stokastisk eksperiment* har uendeligt mange udfald opfyldende

1) $P(u) = 0$ for alle udfald u

2) $P(A) \in [0,1]$ hvor A er en mængde af udfald

Hvis $P(A) > 0$ kaldes A en *hændelse*.

3) $P(A) = 1$, hvor A er mængden af alle udfald

Definition 30

En *kontinuert stokastisk variabel* X knytter et reelt tal til alle udfald for et kontinuert stokastisk eksperiment.
X skal opfylde betingelsen $P(X = a) = 0$ for alle tal a

Bemærk, at vi som følge af definitionen får

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Nedenstående definition giver en frekvensfunktion, som er klokkeformet, og som har den egenskab, at 'summen af alle frekvenser giver 1'. Det er en kontinuert funktion, så i 3g vil vi sige, at integralet af f giver 1, hvilket skrives på følgende måde

$$\int_{-\infty}^{\infty} f(t)\,dt = 1$$

Det kan intuitivt opfattes som, at arealet mellem frekvensfunktionens graf og førsteaksen giver 1. Det gjorde det jo også, da vi tegnede frekvensfunktionen af nittehoved-eksemplet i den deskriptive statistik., hvilket man hurtigt kan tjekke efter.

Definition 31

En kontinuert stokastisk variabel X kaldes *standard-normalfordelt*, hvis den har frekvensfunktionen

$$j(t) = \frac{1}{\sqrt{2p}} \cdot e^{-\frac{1}{2}t^2}$$

Denne normalfordeling er særligt fin, fordi den har en simpel middelværdi og varians.

Sætning 32

Lad X være standard normalfordelt, så gælder

1) $E(X)=0$

2) $s(X)=1$

Bevis

Uden nærmere forklaringer er beviset for f.eks. 1 som følger

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}\,dx = \frac{1}{\sqrt{2\pi}}\left[-e^{-\frac{1}{2}x^2}\right]_{-\infty}^{\infty} =$$

$$\lim_{x\to\infty}\left(-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right) - \lim_{x\to-\infty}\left(-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right) = 0$$

Beviset kræver kendskab til kontinuitet, grænseværdi og integralregning.

Udfra standard-normalfordelingen kan man nu definere alle mulige andre normalfordelinger:

Definition 33

En stokastisk variabel X som kan skrives

$$X = aY + b$$

hvor Y er en standard normalfordelt og $a > 0$ kaldes en *normalfordelt* stokastisk variabel. Man skriver

$$X \approx n(b, a)$$

Sætning 34

Lad X være normalfordelt: $X \approx n(b, a)$. Så gælder

1) $E(X) = b$
2) $s(X) = a$

Bevis

1) $E(X) = E(aY + b) = aE(Y) + b = a \cdot 0 + b = b$
2) $s(X) = s(aY + b) = a s(Y) = a \cdot 1 = a$

Bemærk at heraf følger, at hvis X er normalfordelt med middelværdi μ og spredning s, så kan X skrives $X = \sigma Y + \mu$, hvor Y er standard normalfordelt. Alternativt, så er $X \approx n(\mu, \sigma)$.

Omvendt har vi, at

$$Y = \frac{t - m}{s}$$

Dvs. vi kan lave en omregning fra sandsynligheder i X til sandsynligheder i Y, så et godt kendskab til standard normalfordelingen vil give et godt kendskab til alle andre normalfordelinger.

Som ved binomialfordelingen, så er fordelingsfunktionen mere anvendelig. Vi husker, at fordelingsfunktionen er den kumulerede tæthed, dvs. summen af alle tætheder op til et bestemt tal:

$$F(t) = \text{sum af } f(s) \quad \text{for} \quad s \le t$$

Dette kan man også skrive som

$$F(t) = \sum_{s \le t} f(s) \, .$$

I binomialtilfældet ville man skrive

$$F(j) = \sum_{m=0}^{j} P(x = m) = \sum_{m=0}^{j} b(n\, , p\, ; m)$$

I normalfordelingen er summeringen lidt mere besværlig, fordi normalfordelingen er kontinuert. I tilfældet med binomialfordelingen har vi kun et bestemt antal x-værdier, som man kan bruge, og derfor så summerer vi blot deres sandsynligheder sammen. Måden man summerer på, når fordelingen er kontinuert er mere besværlig, og kræver egentlig kendskab til integralregning, hvor man bl.a. i stedet for sumtegnet $\sum$ bruger integraltegnet $\int$, dvs. uden yderlige forklaringer har vi

$$F(j) = \int_{-\infty}^{j} j\,(t)dt = \int_{-\infty}^{j} \frac{1}{\sqrt{2p}} \cdot e^{\frac{-t^2}{2}} dt$$

Da det er en tung måde at skrive fordelingsfunktionen, og vi ikke lærer integralregning før i 3.g, så døber vi fordelingsfunktionen “store φ“ og skriver

$$P(X \le t) = \Phi(\tfrac{t-m}{s})$$

Bemærk, at Φ faktisk er en stamfunktion til φ, dvs. $\Phi'(x) = \varphi(x)$.
En tabel over standardnormalfordelingen findes i Sigmatabellen side 28 og 29.

Vi bruger tabellen og regner på samme måde som ved binominalfordelingen.

Eksempel

Lad X være standard normalfordelt, $X \approx n(0,1)$.
Så får vi ved tabelopslag.

$$P(X \le 0{,}5) = \Phi(0{,}5) = 0{,}691$$

$$P(X \ge 0{,}5) = 1 - P(X \le 0{,}5) =$$
$$1 - \Phi(0{,}5) = 1 - 0{,}691 = 0{,}311$$

På samme måde udregnes følgende

$$P(X \le 0{,}329) = \Phi(0{,}329) = 0{,}999$$

$$P(X \le -1{,}73) = \Phi(-1{,}73) = 0{,}042$$

$$P(X \le -3{,}39) = \Phi(-3{,}39) = 0{,}000$$

Generelt kan vi skrive for en standard normalfordelt variabel.

Sætning 35

Lad X være standard normalfordelt så gælder

$$P(X \leq a) = \Phi(a)$$

$$P(X \geq a) = 1 - \Phi(a)$$

$$P(a \leq X \leq b) = \Phi(b) - \Phi(a)$$

Generelt får man for en normalfordelt stokastisk variabel via omskrivningen på forrige side.

Sætning 36

Lad X være en normalfordelt stokastisk variabel med middelværdi μ og spredning σ : $X \approx n(\mu, \sigma)$

$$P(X \leq a) = \Phi(\frac{a-\mu}{\sigma})$$

$$P(X \geq a) = 1 - \Phi(\frac{a-\mu}{\sigma})$$

$$P(a \leq X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Sætningen betyder, at man er i stand til at regne på en vilkårlig normalfordeling, når blot man har en standardnormalfordelingstabel, og det har vi jo i Sigmatabellen.

Eksempel

Lad X være normalfordelt med middelværdi 20 og spredning 5.
Ved opslag i standardnormalfordelingstabellen fås

$$P(X \leq 10) = \Phi(\tfrac{10-20}{5}) = \Phi(-0{,}2) = 0{,}421$$

På tilsvarende måde udregnes

$$P(5 \leq X \leq 25) = P(X \leq 25) - P(X \leq 5) = \Phi(\tfrac{25-20}{5}) - \Phi(\tfrac{5-20}{5})$$
$$= \Phi(+1) - \Phi(-3) = 0{,}841 - 0{,}001 = 0{,}840$$

Eksempel

Det er også muligt at gå den modsatte vej. Hvis man ved, at sandsynligheden for at X er mindre end a f.eks. er 0,25 - altså man står med

$$P(X \le a) = 0{,}25$$

og gerne vil beregne a.
Vi finder så værdien af a ved at gå ind i tabellen og finde de sandsynligheder, som ligger omkring 0,25.

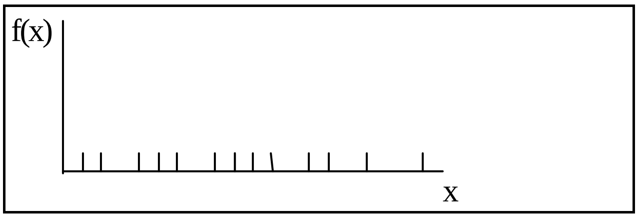
$$P(X \le -0{,}67) = 0{,}251 \qquad P(X \le -0{,}68) = 0{,}248$$

Vi tager den sandsynlighed, der en giver en sandsynlighed tættest på 0,25, dvs. her -0,67

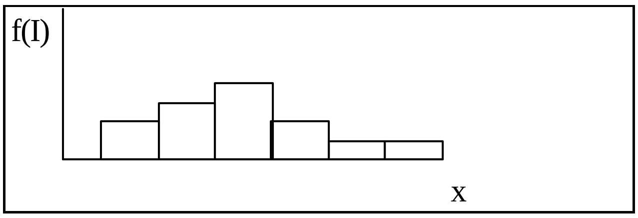
Det er også muligt at bruge en tabel over den inverse normalfordeling, som står i sigmatabellen side 28 nederst. Af den tabel ser man direkte

$$P(X \le a) = 0{,}25 \Rightarrow \text{a = -0,67}$$

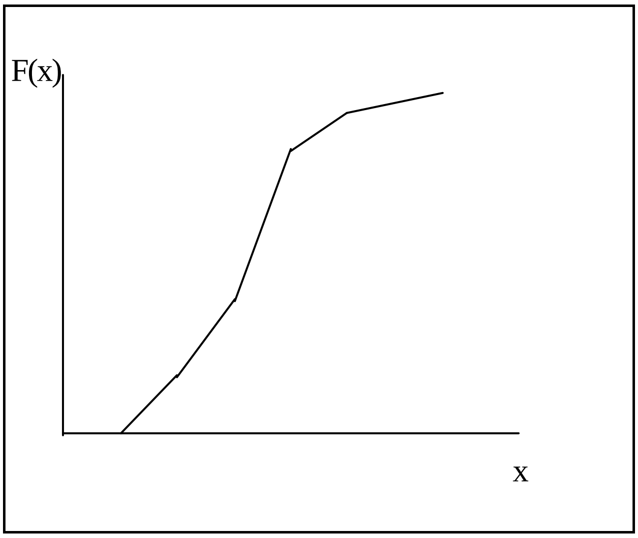
Hvis man skal undersøge om en række observationer er normalfordelte, så er man nødt til at tegne frekvensfunktionen. Typisk vil alle observationer være forskellige og i så fald vil man få



Som regel vil det vise sig at observationer er grupperet omkring en midte, hvis man har mange observationer; men det er svært at afgøre, om det er en normalfordeling. Det bliver lettere at se, hvis man grupperer sine observationer i intervaller og laver et histogram.



Man ser et klokkeformet histogram, hvis observationerne er normalfordelte. Men det næste spørgsmål må så være, hvornår er et histogram klokkeformet? Det kan være svært at vurdere, så derfor tegner man fordelingsfunktionen (dvs. sumkurven)



Hvis sumkurven er s-formet, så er observationer normalfordelte, men det er nu også ret svært at vurdere, hvornår det er tilfældet. Faktisk er den eneste kurveform, der er let at vurdere en ret linie

Vi konstruerer derfor et papir, normalfordelingspapiret, som intuitivt skal forstås sådan, at man trækker forneden og foroven i et almindeligt (men elastisk) koordinatsystem, sådan at man ikke ændrer på midten af koordinatsystemet, mens inddelingerne på *y*-aksen bliver større og større jo længere væk man kommer fra midten til begge sider.

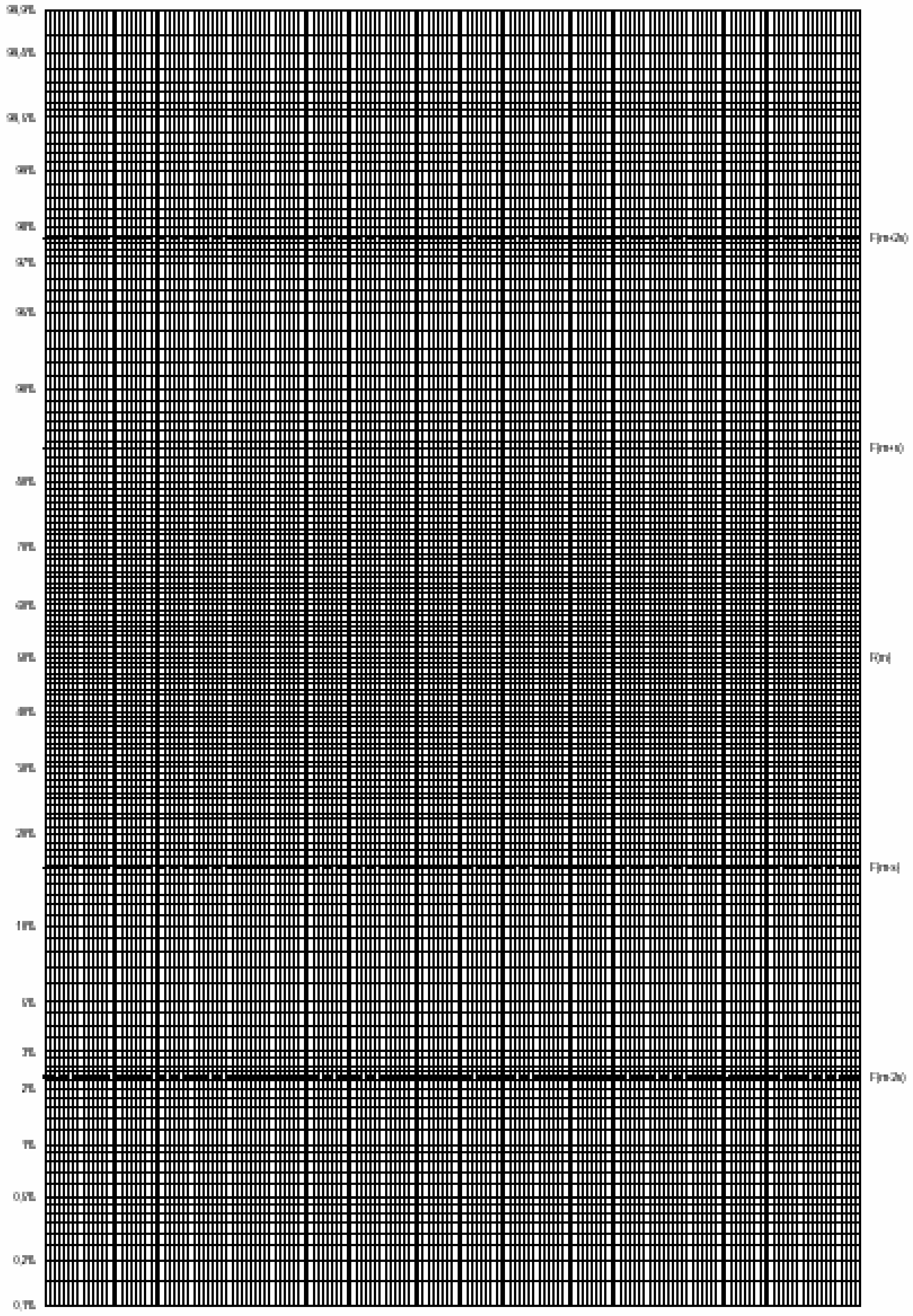
Dette er ganske som ved det enkeltlogaritmiske papir. Her ændrede vi *y*-aksen, således at kun eksponentielle udviklingers graf blev rette linier. Tilsvarende gælder for normalfordelingspapiret:

Sætning 37

Netop kun normalfordelte observationers fordelingsfunktion giver en ret linie på et normalfordelings papir.

Et normalfordelingspapir kaldes også for sandsynlighedspapir. Man skal bemærke, at *y*-aksen er inddelt på forhånd, så man kan ikke ændre på *y*-aksen overhovedet. *y*-værdierne der skal indsættes skal være de kumulerede frekvenser i procent, hvilket også illustreres af %-tegnet øverst på *y*-aksen. *x*-aksen derimod kan man inddele, som man vil. Bemærk, at man ikke har 0% eller 100% med på *y*-aksen. Det skyldes, at inddelingerne bliver større og større jo mere man nærmer sig de to tal, så uanset hvor stor man lavede *y*-aksen, så ville man aldrig kunne få disse to tal med.

Et normalfordelingspapir ser således ud:



Lad os se på et eksempel på brugen af normalfordelingspapir

Eksempel

Pladderballe Affaldsposefabrik fabrik laver affaldsposer, der kan rumme 30 liter. Man laver en forbrugerundersøgelse for at se, om poserne nu faktisk indeholder præcis 30 liter. De kumulerede hyppigheder giver følgende resultater

Interval	=<26	]26;27]	]27;28]	]28;29]	]29;30]	]30;31]	]31;32]	>32
F(I)	0,001	0,01	0,07	0,25	0,57	0,90	0,975	1,00

Rent sandsynlighedsteoretisk har vi en stokastisk variabel X, som angiver rumfanget af en tilfældigt udvalgt pose.

1) Vi vil vise at observationerne er normalfordelte

Vi indtegner de kumulerede frekvenser på et normalfordelingspapir, og indtegner bedste rette linie - se næste side.

Idet punkterne tilnærmelsesvist ligger på en ret linie, kan vi konkludere, at den stokastiske variabel, som angivet rumfanget af en affaldspose, er tilnærmelsesvist normalfordelt.

Bemærk, at punktet for f.eks. den anden observation har koordinaterne (27 , 0,01) - man skal altid bruge den x-værdi, som ligger sidst i intervallet.

2) Find kvartilsættet

Vi aflæser udfor 25%, 50% og 75% liniens x-værdier og får
kvartilsæt = (28,85 , 29,65 , 30,45)

3) Find middeltallet og spredningen

Middeltallet findes ved at aflæse x-værdien svarende til y-værdien markeret med $F(\mu)$ (denne y-værdi er i øvrigt altid 50%).
Vi får $\mu = 29,65$

Spredningen findes ved at aflæse forskellen i x-værdierne svarende til y-værdierne markeret med $F(\mu+\sigma)$ og $F(\mu)$.
Vi får $\sigma = (\mu+\sigma) - \mu = 30,65 - 29,65 = 1,00$

4) Find sandsynligheden for at en pose indeholder under 29 liter

Vi skal altså finde $P(X \leq 29)$
Det fås af fordelingsfunktionen som F(29), hvilket aflæses på normalpapiret til

$$P(X \leq 29) = F(29) = 0{,}29 = 29\%$$

5) Find sandsynligheden for at en pose indeholder mellem 29 og 31 liter.

$$P(29 \leq X \leq 31) = P(X \leq 31) - P(X \leq 29) = F(31) - F(29)$$
$$= 0{,}87 - 0{,}29 = 0{,}58$$

6) Find sandsynligheden for en pose indeholder over 31,5 liter

$$P(X \leq 31{,}5) = 1 - P(X \leq 31{,}5) = 1 - F(31{,}5)$$
$$= 1 - 0{,}94 = 0{,}06$$

Normalfordelingen er kendetegnet ved sin klokkeformede sandsynlighedsfordeling. Hvis man i stedet tegner sandsynlighedsfordelingen for binomialfordelingen for store værdier af *n*, så vil man også se en klokkeformet fordeling. Dette antyder, at man for store *n* kan approximere binomialfordelingen med en normalfordeling.

For at have en betingelse for at bedømme om en sådan approksimation er god, så fastlægger vi

Sætning 39

Lad X være binomialfordelt med parametrene n og p :
$X \approx b(n,p)$. Hvis

$$n \cdot p > 5 \text{ og } n(1-p) > 5$$

så er X approximativt normalfordelt, dvs.

$$P(X \leq a) = \Phi(\frac{a + 0{,}5 - m}{s})$$

$$P(a \leq X \leq b) = \Phi(\frac{b + 0{,}5 - m}{s}) - \Phi(\frac{a - 0{,}5 - m}{s})$$

Eksempel

I kapitlet om binomialfordelingen havde vi et eksempel, hvor man udspurgte 25 mennesker, om de havde været i bad. Den her betragtede stokastiske variabel havde fordelingen $X \approx b(25\,,\,0{,}70)$.

Kan vi normalfordelingsapproximere denne?
Ja, for

$$n \cdot p = 25 \cdot 0{,}7 = 17{,}5 > 5$$

og

$$n \cdot (1 - p) = 25 \cdot 0{,}3 = 7{,}5 > 5 .$$

Vi kan nu bestemme sandsynligheden for, at højst 14 mennesker har været i bad i dag:

$$P(X \leq 14) = \Phi\left(\frac{14 + 0{,}5 - 25 \cdot 0{,}7}{\sqrt{25 \cdot 0{,}7 \cdot 0{,}3}}\right) = \Phi(-1{,}31) = 0{,}095$$

Opgaver

6.1 I Pladderballe har man i perioden 1896-1965 hvert år målt den samlede regnmængde. Målt i tommer gave det følgende resultat:

regnmængde	antal år
16-18	1
18-20	4
20-22	6
22-24	8
24-26	17
26-28	7
28-30	11
30-32	9
32-34	3
34-36	2
36-38	1
38-40	1

a) Gør rede for, at observationerne med tilnærmelse er normalfordelte.
b) Bestem middelværdi og varians.
c) Pladderballe Landboforening vil gerne til at dyrke yams. Desværre kræver yams en årlig nedbør på mindst 26 tommer regn for at kunne give et godt udbytte. Landboforeningen vil dyrke yams, hvis man kan få et godt udbytte mindst hvert andet år. Kan dette lade sig gøre?

6.2 Pladderballe Mejeri har en maskine, som fylder mælk på literkartoner. For et stort antal kartoner med mælk kan man regne med, at mælkemængden i kartonerne er normalfordelt. Middelværdien er 1,005L, og 85% af kartonerne indeholder mindre end 1,012L.

a) Indtegn på normalfordelingspapir den rette linie, som de kumulerede frekvenser skal følge.

b) Aflæs vha. papiret spredningen for mælkemængden.

c) Hvor mange procent af kartonerne indeholder mindst 1,000L mælk?

6.3 Pladderballe Pladefabrik fremstiller plader. I forbindelse med produktionen kontrolleres pladetykkelsen. Målinger har vist, at 1,2% af pladerne har en tykkelse på under 14,0 mm, og at 95,4 % har eb tykkelse under 16,0 mm.
Det antages, at pladetykkelsen er normalfordelt.

a) Bestem middelværdi og spredning for denne normalfordeling.

b) Hvor mange procent af pladerne har en tykkelse på over 16,5 mm?

Facitliste

1.1 a)

x	0	1	2	3	4	5	6
f	0	0,01	0,035	0,07	0,13	0,14	0,16
F	0	0,01	0,045	0,115	0,245	0,385	0,545
x	7	8	9	10	11	12	
f	0,18	0,125	0,06	0,06	0,025	0,005	
F	0,725	0,85	0,91	0,97	0,995	1,000	

d) (5, 6, 8) e) 6,205

1.2 a) kontinuert
b) (14,8 , 27,4 , 41,3) c) 3,4 %

2.1 3/10

2.2 5/12

2.3 a) 3,8% b) 39,5%

2.4 a) 44% b) 68,2%

2.5 c) Ja

2.6 a) Ja
b)

$$\begin{aligned} U = \{ &(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ &(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ &(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ &(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ &(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ &(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \} \end{aligned}$$

Alle udfald har sandsynligheden 1/36
c) 1/4
d)

t	2	3	4	5	6	7
$P(X=t)$	1/36	2/36	3/36	4/36	5/36	6/36
t	8	9	10	11	12	
$P(X=t)$	5/36	4/36	3/36	2/36	1/36	

2.7

t	-5	-4	-3	-2	-1	0
$P(Y=t)$	1/36	2/36	3/36	4/36	5/36	6/36
t	1	2	3	4	5	
$P(Y=t)$	5/36	4/36	3/36	2/36	1/36	

2.8 a) 1,8% b) 44,4 %

2.9 39,1%

3.1

a)	0,3	b)	0,45	c)	0,3
d)	0,25	e)	0,75	f)	0,75
g)	0,222	h)	0,357	i)	0
j)	1	k)	6,5	l)	2,550

3.2 a)

t	-2	1	3
$P(X=t)$	8/27	4/9	7/27

b)

t	-1	0	2	6
$P(Y=t)$	1/3	1/27	7/27	10/27

c)

t	-12	-4	-1	0	2	6	18
$P(XY=t)$	1/9	1/27	2/9	1/27	2/9	2/9	4/27

d) 17/27 83/27 146/27

e) 1628/729 3830/729 32414/729

4.1 64

4.2 a) $9{,}735 \cdot 10^{-11}$ b) 0,0825

4.3 0,00202

4.4

a)	4	b)	1020	c)	624	d)	54912
e)	5148	f)	17160	g)	5148	h)	10240
i)	40						

5.1 63,3 %

5.2	a)	1,57%	b)	57,8%	c)	0,75 (eventuelt 1)

5.3 3,98%

5.4	a)	57,9% , 34,7% , 6,9% , 0,46%		
	b)	1,11 kr	c)	8,89 kr
6.1	b)	26,6 4,5	c)	Ja
6.2	b)	0,007 L	c)	25%
6.3	a)	15,70 1,08	b)	0,3%