SANDSYNLIGHEDSREGNING
OG
KOMBINATORIK

x-klasserne

Gammel Hellerup Gymnasium

November 2023 ; Michael Szymanski ; mz@ghg.dk

Indholdsfortegnelse

SANDSYNLIGHEDSREGNING........3
SANDSYNLIGHEDSFELT........3
DE STORE TALS LOV........14
Sandsynligheder og frekvenser:........14
STOKASTISK VARIABEL........15
Middelværdi, spredning og varians........16
Væddemål........17
Sætninger om middelværdi, spredning og varians........18
Normeret stokastisk variabel........22
Uafhængige stokastiske variable........23
Tæthedsfunktion og fordelingsfunktion........26
KOMBINATORIK........30
Princippet om inklusion og eksklusion........30
Skuffeprincippet (Dirichlets skuffeprincip eller dueslagsprincippet)........34
CENTRALE BEGREBER: Permutationer og kombinationer........35
Multiplikationsprincipperne:........36
Additionsprincipperne........38
Permutationer........39
Kombinationer........43
Pascals Trekant (Den Aritmetiske Trekant)........45
Chu-Vandermondes identitet........48
BINOMIALFORDELING........49
NEGATIV BINOMIALFORDELING........55
DEN HYPERGEOMETRISKE FORDELING........60
DEN NEGATIVE HYPERGEOMETRISKE FORDELING........62
DEN MULTINOMIALE FORDELING........64
DEN MULTIVARIABLE HYPERGEOMETRISKE FORDELING........66
POISSON-FORDELING........68

SANDSYNLIGHEDSREGNING

Begrebet *sandsynlighed* er som udgangspunkt knyttet til situationer eller forsøg, hvor der - fordi der spiller nogle tilfældigheder ind - optræder mere end ét muligt udfald. Vi kalder sådanne situationer eller forsøg for *stokastiske eksperimenter*, mens eksperimenter med ét (forudsigeligt) udfald kaldes *deterministiske eksperimenter*.

Man kan så udvide sandsynlighedsbegrebet til at omfatte alle situationer ved at tildele et sikkert udfald sandsynligheden 100% og et umuligt udfald sandsynligheden 0%. I disse tilfælde bruger vi dog betegnelsen *hændelse* i stedet for *udfald*.

SANDSYNLIGHEDSFELT

Et stokastisk eksperiment kan beskrives ved et *sandsynlighedsfelt*, der enten er *endeligt* eller *uendeligt*.

Endelige sandsynlighedsfelter (sandsynlighedsfordelinger)

Definition 1: Et ***endeligt sandsynlighedsfelt*** (U,P) består af et *udfaldsrum* $U=\{u_1,u_2,u_3,...,u_n\}$, hvor $n\in\{2,3,4,5,...\}$, der er mængden af samtlige mulige *udfald*, samt en *sandsynlighedsfunktion* $P:U\mapsto\,]0,1[$, der angiver *sandsynligheden* for de enkelte udfald. Der gælder:

$$\sum_{i=1}^{n} P(u_i) = P(u_1)+P(u_2)+P(u_3)+...+P(u_n)=1$$

Man kan også anvende ordet ***sandsynlighedsfordeling*** i stedet for *sandsynlighedsfelt*.

Eksempel 1: Et stokastisk eksperiment består i at kaste en mønt 3 gange og for hvert kast notere, om det giver plat eller krone. Det endelige sandsynlighedsfelt, der beskriver dette eksperiment, er:

U	kkk	kkp	kpk	kpp	pkk	pkp	ppk	ppp
P(u)	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=1$ ← Kontrol.

Eksempel 2: Et stokastisk eksperiment består i at kaste en almindelig terning (kubus) og notere øjentallet. Det endelige sandsynlighedsfelt bliver så:

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

Opgaverne 200*

I det sidste eksempel ville man kunne bestemme middelværdien for sandsynlighedsfeltet, mens det ikke er tilfældet for det første eksempel (overvej selv hvorfor!). Vi venter med at beregne middelværdier, til vi har fået indført begrebet *stokastisk variabel*.

Eksemplerne 1 og 2 er begge eksempler på såkaldte *symmetriske sandsynlighedsfordelinger*, hvilket følger af definition 2:

Definition 2: Et endeligt sandsynlighedsfelt, hvor sandsynlighederne for hvert udfald er ens (dvs. $P(u_i)=P(u_j)\ \forall u_i,u_j\in U$), kaldes et *symmetrisk sandsynlighedsfelt*.

Et eksempel på et endeligt sandsynlighedsfelt, der IKKE er symmetrisk er:

Eksempel 3 (et IKKE-symmetrisk sandsynlighedsfelt): Det stokastiske eksperiment, der består i at kaste to terninger og notere summen af øjentallene, giver følgende endelige sandsynlighedsfelt:

U	2	3	4	5	6	7	8	9	10	11	12
$P(u)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Kontrol: $\sum_{i=1}^{11} P(u_i) = \frac{1}{36}+\frac{1}{18}+\frac{1}{12}+\frac{1}{9}+\frac{5}{36}+\frac{1}{6}+\frac{5}{36}+\frac{1}{9}+\frac{1}{12}+\frac{1}{18}+\frac{1}{36}=1$

Bemærk altså, at ovenstående IKKE er et symmetrisk sandsynlighedsfelt, selvom man ret hurtigt kan bemærke en form for symmetri i skemaet.

Diskrete uendelige sandsynlighedsfelter

Definition 3: Et ***diskret uendeligt sandsynlighedsfelt*** (U,P) består af et *udfaldsrum* $U=\{u_1,u_2,u_3,...\}$ med tælleligt mange udfald samt en *sandsynlighedsfunktion* $P:U\mapsto\,]0;1[$, der angiver *sandsynligheden* for de enkelte udfald. Der gælder:

$$\sum_{i=1}^{\infty} P(u_i) = P(u_1)+P(u_2)+P(u_3)+...=1$$

Eksempel 4: Et stokastisk eksperiment består i at blive ved med at kaste en mønt, indtil man første gang får plat, og man noterer ved hvert kast, om det blev plat eller krone. Dette giver følgende uendelige sandsynlighedsfelt:

U	p	kp	kkp	kkkp	kkkkp	kkkkkp	...
$P(u)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	...

Kontrol: $\sum_{i=1}^{\infty} P(u_i) = \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\frac{1}{64}+...=1$

Eksempel 5: Et stokastisk eksperiment består i at blive ved med at kaste en terning, indtil man slår en 4'er, og man noterer antallet af kast. Dette giver følgende uendelige sandsynlighedsfordeling:

U	1	2	3	4	5	...
$P(u)$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{5^2}{6^3}$	$\frac{5^3}{6^4}$	$\frac{5^4}{6^5}$	...

Med en sætning fra forløbet om uendeligheder får man:

Kontrol: $\sum_{i=1}^{\infty} P(u_i) = \frac{1}{6}+\frac{5}{36}\cdot\left(1+\frac{5}{6}+\left(\frac{5}{6}\right)^2+\left(\frac{5}{6}\right)^3+...\right)=\frac{1}{6}+\frac{\frac{5}{36}}{1-\frac{5}{6}}=\frac{1}{6}+\frac{\frac{5}{36}}{\frac{1}{6}}=\frac{1}{6}+\frac{5}{6}=1$

Opgaverne 201*

Man kan også have **kontinuerte** uendelige sandsynlighedsfelter:

Eksempel $\frac{23}{4}$: Forestil dig et område med arealet *A*, hvor man udvælger punkter helt tilfældigt. De enkelte punkter må nødvendigvis tildeles sandsynligheden 0, og derfor har man ingen sandsynlighedsfunktion. I stedet vil man arbejde med såkaldte *tæthedsfunktioner*, der i dette tilfælde vil kunne fortælle os, at sandsynligheden for at udvælge et punkt inden for et delområde med arealet *d* vil være $P = \frac{d}{A}$.
Vi indfører tæthedsfunktioner (i én dimension) i Definition 10.

Hændelser

Definition 4: En delmængde af et udfaldsrum kaldes en *hændelse*.

Af definitionen følger:

Sætning 1: Sandsynligheden for en hændelse er summen af sandsynligheder for de udfald, som hændelsen består af.

Eksempel 6: I eksempel 4 kunne en hændelse *H* bestå i, at man får plat i andet eller tredje kast, dvs. $H = \{kp, kkp\}$. Sandsynligheden for denne hændelse er:

$$P(H) = P(kp) + P(kkp) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

Eksempel 7: I eksempel 3 kunne en hændelse være at slå mindst 9 med de to terninger, dvs.

$H = \{9,10,11,12\}$ og $P(H) = P(9) + P(10) + P(11) + P(12) = \frac{1}{9} + \frac{1}{12} + \frac{1}{18} + \frac{1}{36} = \frac{5}{18}$

Opgaverne 202*

Da en delmængde af *U* både kan være den tomme mængde og hele *U*, har man følgende særlige hændelser:

Hvis $H = \emptyset$, er $P(H) = 0$, hvilket kaldes for en *umulig* hændelse.

Hvis $H = U$, er $P(H) = 1$, hvilket kaldes for en *sikker* hændelse.

Og hvis man tager udgangspunkt i en hændelse *H*, har man følgende:

Den *komplementære* hændelse til *H* skrives $\overline{H}$ og består af alle de udfald i *U*, der ikke er med i *H*.

Den umulige hændelse og den sikre hændelse er komplementære hændelser.

Desuden gælder følgende vigtige sætning:

Sætning 2: $P(H) + P(\overline{H}) = 1$

Eksempel 8: I eksempel 7 er den komplementære hændelse til H hændelsen $\overline{H} = \{2,3,4,5,6,7,8\}$, der består i højst at slå 8 med to terninger. Sandsynligheden er $P(\overline{H}) = 1 - P(H) = 1 - \frac{5}{18} = \frac{13}{18}$

I en hel del situationer er det meget nemmere at bestemme sandsynligheden for den komplementære hændelse end for den søgte hændelse, og så kan sætning 2 benyttes.

Eksempel 9: I eksempel 5 kunne man se på hændelsen H, der består i, at man får mindst 2 kast. Hvis man ønsker at finde sandsynligheden for denne hændelse, er det nemmere først at se på komplementærhændelsen $\overline{H}$, der består i at opnå netop 1 kast, hvorefter man har:

$$P(H) = 1 - P(\overline{H}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Opgaverne 203*

Hændelser i et symmetrisk sandsynlighedsfelt

I et symmetrisk sandsynlighedsfelt, hvor sandsynlighederne for hvert udfald som bekendt er ens, kan man ret nemt beregne sandsynligheden for en hændelse:

Sætning 3: I et symmetrisk sandsynlighedsfelt, hvor udfaldsrummet indeholder *n* udfald, er sandsynligheden for hændelsen *H* bestående af *r* udfald givet ved:

$$P(H) = \frac{r}{n}$$

Dette skrives sommetider: $P(hændelse) = \frac{Antal\ gunstige\ udfald}{Antal\ mulige\ udfald}$

Eksempel 10: Se på terningkastet fra eksempel 2, hvor man har et symmetrisk sandsynlighedsfelt, og hvor antallet af elementer i udfaldsrummet er $n = 6$. Hændelsen $H = \{3,4,5,6\}$, der består i at slå mindst en 3'er, indeholder 4 elementer, så man har $P(H) = \frac{4}{6} = \frac{2}{3}$

Sætning 3 kan godt virke meget begrænset i sine anvendelsesmuligheder af forudsætningen om, at det skal være en symmetrisk sandsynlighedsfordeling, men ofte kan man i ikke-symmetriske tilfælde betragte situationen fra en anden indfaldsvinkel, der giver et andet udfaldsrum, og dermed implicit arbejde med et "bagvedliggende" symmetrisk sandsynlighedsfelt.

Eksempel 11: I eksempel 3, hvor man kastede to terninger og så på summen af øjentallene, fik man som bekendt et ikke-symmetrisk sandsynlighedsfelt. Men hvis man ændrer udfaldsrummet til følgende med 36 udfald, ...

U	1	2	3	4	5	6	Øjental for terning B
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)	
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)	
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)	
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)	
Øjental for terning A							

... får man et symmetrisk sandsynlighedsfelt med $P(u_i) = \frac{1}{36}$.

Hændelsen bestående i, at summen giver 4, er så $H_1 = \{(3,1),(2,2),(1,3)\}$, og da *H* indeholder 3 elementer, er $P(H_1) = \frac{3}{36} = \frac{1}{12}$.

Hændelsen bestående i, at summen giver 8, er $H_2 = \{(6,2),(5,3),(4,4),(3,5),(2,6)\}$ og indeholder altså 5 elementer, dvs. $P(H_2) = \frac{5}{36}$.

Betinget sandsynlighed:

Vi har hidtil kun set på hændelser hver for sig. Vi skal nu til at se på flere hændelser i forbindelse med hinanden forstået på den måde, at vi skal se på sandsynligheden for, at hændelse A indtræffer *under forudsætning af* at hændelsen B **er** indtruffet.

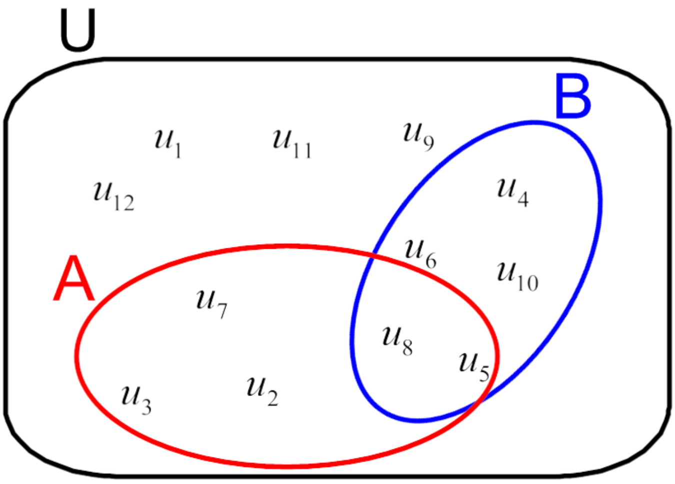
Dette skrives: $P(A|B)$

Det læses: *Sandsynligheden for hændelse A, givet hændelse B.*
Eller: *Den betingede sandsynlighed for hændelse A, givet hændelse B.*

Eksempel 12: Eksempler på sådanne situationer kunne være:

- Hvad er sandsynligheden for at få et ulige øjental ved et kast med en terning, givet at øjentallet er over 3?
- Hvad er sandsynligheden for at være farveblind, givet at du er en mand?
- Hvis man kaster en mønt tre gange, hvad er så sandsynligheden for at få krone i tredje kast, givet at man har fået plat i andet kast?

Vi ønsker at finde en formel til at beregne disse betingede sandsynligheder og ser derfor på følgende udfaldsrum $U = \{u_1, u_2, u_3, ..., u_{12}\}$ bestående af 12 udfald:



$A = \{u_2, u_3, u_5, u_7, u_8\}$ og $B = \{u_4, u_5, u_6, u_8, u_{10}\}$ er forskellige hændelser.

Fællesmængden $A \cap B = \{u_5, u_8\}$ udgør i sig selv en hændelse, og den er væsentlig i denne sammenhæng, for når vi skal se på sandsynligheden for, at hændelse A indtræffer under forudsætning af, at hændelse B **er** indtruffet, så svarer det til, at vi har begrænset udfaldsrummet fra *U* til *B* og nu ser på sandsynligheden for, at hændelsen $A \cap B$ indtræffer.

Eller med andre ord: Det er givet, at en af hændelserne u_4, u_5, u_6, u_8 og u_{10} er indtruffet, og vi skal nu finde sandsynligheden for, at det er en af hændelserne u_5 eller u_8, der er indtruffet.

Dvs. $P(A|B)$ er sandsynligheden for, at udfaldet ligger i fællesmængden $A \cap B$, i forhold til sandsynligheden for, at udfaldet ligger i *B*:

Sætning 4: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Bemærk, at $A \cap B = B \cap A$ (kommutativitet), for venstresiden består jo af de elementer, der både ligger i *A* og *B*, mens højresiden består af de elementer, der både ligger i *B* og *A*, dvs. det er de samme elementer, der indgår på begge sider.

Da omskrivninger af sætning 4 giver $P(A\cap B)=P(A|B)\cdot P(B)$ og $P(B\cap A)=P(B|A)\cdot P(A)$, og da $P(A\cap B)=P(B\cap A)$, har man altså $P(A|B)\cdot P(B)=P(B|A)\cdot P(A)$ eller omskrevet:

Sætning 5 (Bayes' sætning): $P(A|B)=P(B|A)\cdot\dfrac{P(A)}{P(B)}$

Eksempel 13:
Hvad er sandsynligheden for et ulige øjental ved et kast med 1 terning, givet at øjentallet er over 3?
Vi har altså de to hændelser $A=\{1,3,5\}$ og $B=\{4,5,6\}$, hvor $A\cap B=\{5\}$.

Det er et symmetrisk sandsynlighedsfelt med 6 mulige udfald, så alle udfald har sandsynligheden $\frac{1}{6}$

og man har så: $$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{1}{3}$$

Man kunne også have set på den modsatte situation:
Hvad er sandsynligheden for at få et øjental over 3, givet at kastet gav et ulige øjental?

$$P(B|A)=\frac{P(B\cap A)}{P(A)}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{1}{3}$$

Bayes' sætning anvendt til denne udregning: $P(B|A)=P(A|B)\cdot\dfrac{P(B)}{P(A)}=\dfrac{1}{3}\cdot\dfrac{\frac{3}{6}}{\frac{3}{6}}=\dfrac{1}{3}$

Hvis sandsynlighederne for de to hændelser *A* og *B* er lige store, fortæller Bayes' sætning os altså, at de to betingede sandsynligheder er lige store.

Men hvis f.eks. sandsynligheden for en hændelse *A* er dobbelt så stor som sandsynligheden for en hændelse *B*, så vil den betingede sandsynlighed for hændelse *A*, givet hændelse *B*, også være dobbelt så stor som den betingede sandsynlighed for hændelse *B*, givet hændelse *A*.

Lad os nu se på et tilfælde, hvor sandsynlighederne for hændelserne *A* og *B* ikke er lige store:

Eksempel 14:
Hvad er sandsynligheden for et ulige øjental ved et kast med 1 terning, givet at øjentallet er over 4?
Vi har de to hændelser $A=\{1,3,5\}$ og $B=\{5,6\}$, hvor $A\cap B=\{5\}$.

Man har så: $$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{6}}{\frac{2}{6}}=\frac{1}{2}$$

Hvad er sandsynligheden for at få et øjental over 4, givet at kastet gav et ulige øjental?

$$P(B|A)=\frac{P(B\cap A)}{P(A)}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{1}{3}$$

Bayes' sætning anvendt til dette: $P(B|A)=P(A|B)\cdot\dfrac{P(B)}{P(A)}=\dfrac{1}{2}\cdot\dfrac{\frac{2}{6}}{\frac{3}{6}}=\dfrac{1}{2}\cdot\dfrac{2}{3}=\dfrac{1}{3}$

Bayes' sætning kan anvendes til andet end terningkast. Den er relevant i mange sammenhænge:

Eksempel 14b: Vi ser på følgende opdigtede situation: 60% af alle voksne danskere er tatoverede, 4% af alle voksne er slemme forbrydere og 98% af alle voksne slemme forbrydere er tatoverede.

En slem forbrydelse er blevet begået, og vi pågriber en tilfældig voksen dansker, hvorefter vi opdager, at hun er tatoveret, hvorefter vi straks gør hende til mistænkt med begrundelsen, at 98% af alle slemme forbrydere er tatoverede.

Tallet 98% er korrekt, men er det det rigtige tal at basere sin begrundelse på?

Vi lader T være hændelsen "voksen tatoveret dansker", dvs. $P(T)=0,60$.

Vi lader F være hændelsen "voksen slem forbryder", dvs. $P(F)=0,04$.

Med disse betegnelser kan "98% af alle voksne slemme forbrydere er tatoverede" oversættes til $P(T|F)=0,98$.

Men bemærk nu, at vi har pågrebet en tilfældig voksen dansker og opdaget, at hun er tatoveret. Og vi ønsker nu at finde sandsynligheden for, at hun er en slem forbryder. Dette svarer til $P(F|T)$ og ikke $P(T|F)$. Vi anvender Bayes' sætning til at finde sandsynligheden:

$$P(F|T)=P(T|F)\cdot\frac{P(F)}{P(T)}=0,98\cdot\frac{0,04}{0,60}=0,065=6,5\%$$

Det er altså ikke særlig sandsynligt, at vi har fået fat i en slem forbryder, men hvis man forveksler $P(F|T)$ og $P(T|F)$, er der tale om en fejlslutning.

Opgaverne 204*

Prosecuter's Fallacy (Anklagerens fejlslutning) er en fællesbetegnelse for en række fejlslutninger, der – som navnet henviser til – kan optræde i retssager, men også helt generelt i dagligdagen. Vi kan benytte Bayes' sætning til at belyse en enkelt af dem:

Vi lader B være hændelsen "Beviset er fundet" og U hændelsen "Den anklagede er uskyldig". Hermed er den betingede sandsynlighed $P(B|U)$ altså sandsynligheden for at have fundet det pågældende bevis, givet at den anklagede er uskyldig, mens $P(U|B)$ er sandsynligheden for, at den anklagede er uskyldig, givet det pågældende bevis optræder.

Læg godt mærke til forskellen!

Prosecuter's Fallacy optræder, hvis man forveksler disse to sandsynligheder med hinanden. I en retssag er det $P(U|B)$, der er væsentlig (hvis denne sandsynlighed er meget lille, ser sagen skidt ud for den anklagede), mens $P(B|U)$ ofte kan være nemmere at regne ud. Bayes' sætning giver os:

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

Det afgørende i denne sammenhæng er brøken på højresiden i ovenstående. For selv hvis anklageren er i stand til at fremlægge et argument for, at $P(B|U)$ er meget lille, så behøver $P(U|B)$ ikke også at være så lille, da brøken kan være meget stor.

Lad os se på et ekstremt eksempel:

Eksempel 14c: Vi antager, at vi har en population på 5 millioner mennesker, og en forbrydelse har fundet sted. Der er efterladt et DNA-spor. Politiet vælger tilfældigt blandt hele populationen en mistænkt og foretager en DNA-test. Der er overensstemmelse mellem den mistænktes DNA og det efterladte DNA-spor, og det gælder kun for 1 ud af 10.000 personer.

Så politiet konkluderer, at $P\left(B|U\right)=\frac{1}{10000}=0,0001$, dvs. at sandsynligheden for at få dette bevis, givet at personen er uskyldig, er 0,01%, og derfor må personen være skyldig.

Men ud af de 5 millioner mennesker er 4.999.999 uskyldige, dvs. $P\left(U\right)\approx 1$, og da næsten alle i populationen er uskyldige, er $P\left(B\right)\approx P\left(B|U\right)=0,0001$, dvs. Bayes' sætning giver $P\left(U|B\right)\approx 1$, dvs. sandsynligheden for, at personen er uskyldig, givet dette bevis optræder, er faktisk næsten 100%.

Der er anvendt "$\approx$"et par steder i ovenstående eksempel, da vi ikke er interesserede i det eksakte resultat. Når vi har lært om multiplikations- og additionsprincipperne, vil vi se, at man – teoretisk set - ville kunne finde $P\left(B\right)$ved $P\left(B\right)=P\left(B|\overline{U}\right)\cdot P\left(\overline{U}\right)+P\left(B|U\right)\cdot P\left(U\right)$.

En pointe ved Eksempel 14c er, at man ikke kan benytte f.eks. en DNA-test, hvis den optræder som det eneste bevis, da man så har $P\left(U\right)\approx 1$. Men hvis det optræder sammen med andre beviser, kan det være et stærkt bevis.

Med ovenstående udtryk kan Bayes' sætning altså omskrives til følgende sammenhæng, hvor vi først rigtigt kan forstå nævneren i brøken, når vi har gennemgået noget kombinatorik:

Sætning 5 – udvidet: $P\left(A|B\right)=P\left(B|A\right)\cdot\frac{P\left(A\right)}{P\left(B|A\right)\cdot P\left(A\right)+P\left(B|\overline{A}\right)\cdot P\left(\overline{A}\right)}$

Eksempel 14d: Til en fest testes for, om deltagerne har indtaget stoffer.

Der anvendes et apparatur, hvor 99% af alle, der har indtaget stoffer, vil blive testet positive (man siger, at *sensitiviteten* er 0,99). Kun 3% af dem, der ikke har indtaget stoffer, vil blive testet positive (man siger, at *specificiteten* er 0,97). Vi antager, at 2% af alle deltagere i festen har indtaget stoffer.

Vi indfører nu følgende hændelser:

B: Testen er positiv (dvs. apparatet siger "Biiiiiiip")

S: Personen har indtaget stoffer.

$\overline{S}$: Personen har ikke indtaget stoffer (den komplementære hændelse til ovenstående)

Hermed er altså:

$P\left(B|S\right)=0,99$

$P\left(S\right)=0,02$ $\qquad$ $P\left(\overline{S}\right)=0,98$

$P\left(B|\overline{S}\right)=0,03$

En tilfældig person **bliver testet positiv**, og vi ønsker nu at finde sandsynligheden for, at personen har indtaget stoffer. Vi skal altså bestemme $P\left(S|B\right)$:

$$P\left(S|B\right)=P\left(B|S\right)\cdot\frac{P\left(S\right)}{P\left(B|S\right)\cdot P\left(S\right)+P\left(B|\overline{S}\right)\cdot P\left(\overline{S}\right)}=0,99\cdot\frac{0,02}{0,99\cdot 0,02+0,03\cdot 0,98}=0,40=40\%$$

Det mest sandsynlige er altså, at personen ikke har indtaget stoffer.

Opgaverne 205*

Vi ser nu på et eksempel, hvor sandsynlighederne for hændelserne A og B er lige store, og hvor vi derfor nøjes med at beregne $P(A|B)$, da vi ved, at $P(B|A)=P(A|B)$:

Eksempel 15: Hvis man kaster en mønt tre gange, hvad er så sandsynligheden for at få krone i tredje kast, givet at man har fået plat i andet kast?
Det er et symmetrisk sandsynlighedsfelt med udfaldsrummet
$U=\{ppp, ppk, pkp, pkk, kpp, kpk, kkp, kkk\}$ og $P(u_i)=\frac{1}{8}$.

Vi har desuden:

Hændelsen at få krone i tredje kast: $A=\{ppk, pkk, kpk, kkk\}$ $\quad P(A)=\frac{4}{8}=\frac{1}{2}$

Hændelsen at få plat i andet kast: $B=\{ppp, ppk, kpp, kpk\}$ $\quad P(B)=\frac{4}{8}=\frac{1}{2}$

$A\cap B=\{ppk, kpk\}$ $\quad P(A\cap B)=\frac{2}{8}=\frac{1}{4}$

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{4}}{\frac{1}{2}}=\frac{2}{4}=\frac{1}{2}$$

Ovenstående er et eksempel på såkaldt *uafhængige* hændelser. Man kan hurtigt overbevise sig selv om, at det at få plat i andet kast ikke kan have nogen indvirkning på, om man får krone i tredje kast, eller sagt på en anden måde: **Sandsynligheden for at få krone i tredje kast er uafhængig af, om man har fået plat eller ej i andet kast**.

Vi definerer nu:

Definition 5: Hændelse A siges at være *uafhængig af* hændelse B, hvis $P(A|B)=P(A)$.

Opgaverne 206*

Øvelse 1: Undersøg i hvilke af situationerne i eksemplerne 13, 14 og 15, at den ene af hændelserne er uafhængig af den anden.

Vi ser nu på følgende vigtige - og måske overraskende - sætning:

Sætning 6: Hvis hændelse A er uafhængig af hændelse B, så er hændelse B også uafhængig af hændelse A, og man taler derfor om *de uafhængige hændelser A og B*.

Bevis 6: Antag, at hændelse A er uafhængig af hændelse B. Der gælder så ifølge Definition 5 og Sætning 4:

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

Dette udsagn kan omskrives til:

$$P(A)=\frac{P(B\cap A)}{P(B)} \Leftrightarrow P(B)=\frac{P(B\cap A)}{P(A)}$$

dvs. ifølge sætning 4: $P(B)=P(B|A)$

Ifølge Definition 5 har vi altså, at B er uafhængig af A.

Ved at nærstudere beviset kan man desuden se følgende sætning:

Sætning 7: Hændelserne A og B er uafhængige, netop hvis $P(A \cap B) = P(A) \cdot P(B)$

Øvelse 2: Tjek, om du - evt. ved nærlæsning af beviset for sætning 6 - kan se, at der gælder "netop hvis" (dvs. en biimplikation) i Sætning 7.

Eksempel 16: Vi kaster en mønt fire gange og vil gerne undersøge, om følgende hændelser er uafhængige:
A: Man får krone i andet kast.
B: Man får mindst to plat i alt.
Sandsynlighedsfeltet er symmetrisk og består af et udfaldsrum med 16 mulige udfald:

$$U = \{pppp, pppk, ppkp, ppkk, pkpp, pkpk, pkkp, pkkk, kppp, kppk, kpkp, kpkk, kkpp, kkpk, kkkp, kkkk\}$$

$$A = \{pkpp, pkpk, pkkp, pkkk, kkpp, kkpk, kkkp, kkkk\} \text{ 8 udfald}$$

$$B = \{pppp, pppk, ppkp, ppkk, pkpp, pkpk, pkkp, kppp, kppk, kpkp, kkpp\} \text{ 11 udfald}$$

$$A \cap B = \{pkpp, pkpk, pkkp, kkpp\} \text{ 4 udfald}$$

$$P(A) \cdot P(B) = \frac{8}{16} \cdot \frac{11}{16} = \frac{11}{32}$$

$$P(A \cap B) = \frac{4}{16} = \frac{1}{4}$$

Da $P(A \cap B) \neq P(A) \cdot P(B)$, er hændelserne IKKE uafhængige.

Da hændelse B i eksempel 16 er mere sandsynlig end hændelse A, så ved vi desuden ifølge Bayes' sætning, at det er mere sandsynligt at få mindst to plat, givet at andet kast var krone, end det er at få krone i andet kast, givet at man får mindst to plat.

Eksempel 17: En mønt kastes 3 gange, og vi vil gerne undersøge, om følgende hændelser er uafhængige:
A: Man får krone i første kast.
B: Man får det samme i alle tre kast.

Vi har mængderne:

$$U = \{ppp, ppk, pkp, pkk, kpp, kpk, kkp, kkk\}$$

$$A = \{kpp, kpk, kkp, kkk\}$$

$$B = \{ppp, kkk\}$$

$$A \cap B = \{kkk\}$$

Udregningerne giver:

$$P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

$$P(A \cap B) = \frac{1}{8}$$

Da $P(A \cap B) = P(A) \cdot P(B)$, er hændelserne *A* og *B* uafhængige.

Opgaverne 207*

Sætning 7 er ekstremt vigtig. Anvendelsen vil oftest være, at man ønsker at finde sandsynligheden for, at flere hændelser (*A, B, C, …*) alle indtræffer, og at man kan gennemskue, at de er uafhængige af hverandre. Sætning 7 siger så, at vi kan finde denne søgte sandsynlighed ved at multiplicere sandsynlighederne for de enkelte hændelser.

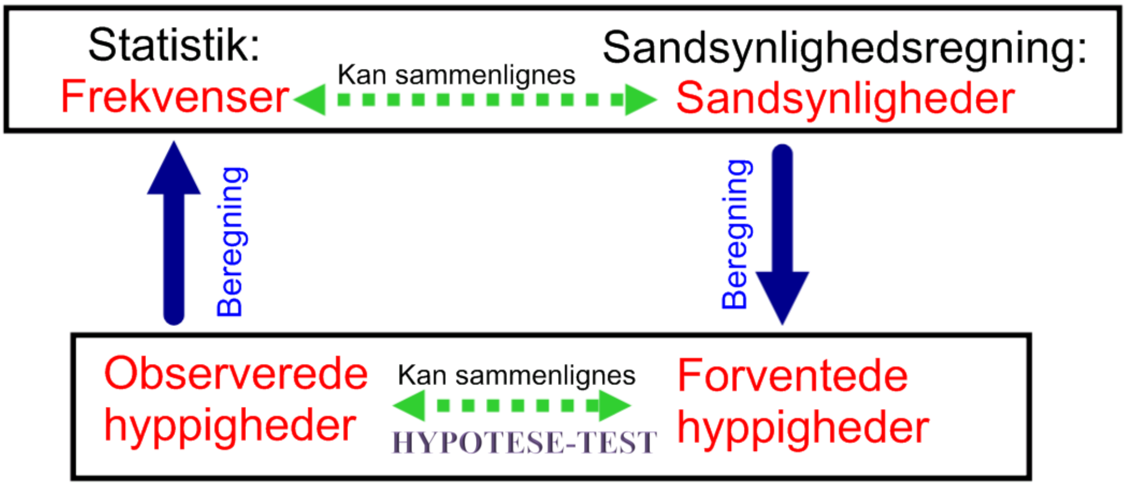
DE STORE TALS LOV

Sandsynligheder og frekvenser:

Sandsynligheder og frekvenser er begge størrelser, der ligger i intervallet $[0,1]$ eller mellem 0% og 100%, men det er vigtigt at skelne mellem disse to beslægtede begreber.

Sandsynligheder hører - ikke overraskende - hjemme inden for sandsynlighedsregning, mens *frekvenser* hører hjemme inden for statistik.
Sandsynligheder er noget, man tænker sig frem til (ofte ved inddragelse af kombinatorik), mens *frekvenser* fremkommer ved beregning på indsamlet datamateriale.

Dette kan illustreres med følgende skema:



Sætning 8 (De store tals lov): Ved gentagelse af et eksperiment *n* gange gælder

$$f(A) \to P(A) \;\; for \;\; n \to \infty \, ,$$

hvor $f(A)$ er frekvensen af hændelsen *A*, og $P(A)$ er sandsynligheden for hændelsen *A*.
Eller lidt løst sagt: Jo flere gange man udfører et eksperiment, jo tættere kommer frekvensen for en hændelse i store træk på sandsynligheden for denne hændelse.
Eller: De observerede hyppigheder vil i store træk komme **relativt** tættere på de forventede hyppigheder, jo flere gange man udfører et eksperiment.

Bemærk, at dette ikke er en almindelig matematisk sætning. Sætningen kobler statistik og sandsynlighedsregning sammen, og den afviger fra vores allerede kendte sætninger om grænseværdier ved, at vi ikke kan sætte tal på, hvor stor *n* skal være, for at $|f(A)-P(A)|<\varepsilon$ for et givet ε. Vi kan kun udtale os om sandsynligheder for den slags. Dvs. vi er nødt til, hvis vi skal have en præcis formulering, der ville kunne bruges til udregninger, at indføre sandsynlighedsbegrebet i formuleringen. Det kan skrives på forskellige måder. En af dem er:

$$\forall \varepsilon_1, \varepsilon_2 \in \,]0,1] \, \exists M \in \mathbb{N} : n > M \Rightarrow p\left(\left|f_n(A)-P(A)\right| \geq \varepsilon_1\right) < \varepsilon_2$$

Dvs. uanset, hvor tæt, ε_1, vi kræver, at frekvensen, $f_n(A)$, for hændelsen *A* ved *n* gentagelser af et eksperiment skal komme på sandsynligheden, $P(A)$, for hændelsen *A*, så kan risikoen for, at det ikke sker, gøres mindre end enhver positiv størrelse, ε_2, hvis bare vi sørger for, at *n* er større end et givet – tilpas stort tal - *M*.

Sætningen fungerer dog også som en rent statistisk sætning. Hvis f.eks. 14,7% af alle danskere vil stemme på partiet Q, så kan man regne med, at jo flere danskere, man udtager i en stikprøve, jo tættere vil frekvensen af Q-stemmere i stikprøven i store træk komme på 14,7%.

Når man inden for statistik skal foretage et såkaldt *hypotesetest*, er det observerede og forventede **hyppigheder**, der indgår i testet, da antallet af observationer er væsentligt for vurderingen.

STOKASTISK VARIABEL

Vi ønsker nu at indføre begreberne middelværdi, varians og spredning som nogle størrelser, der skal fortælle noget om vores eksperimenter. Men disse størrelser kræver tal, der kan regnes på, og vi har allerede set eksempler på udfaldsrum - f.eks. $U = \{ppp, ppk, pkp, pkk, kpp, kpk, kkp, kkk\}$ - der ikke består af tal.

Vi vil derfor indføre begrebet *stokastisk variabel*, der er en funktion, der sætter tal på de enkelte udfald.

I ovennævnte tilfælde med vores plat og krone kunne en stokastisk variabel f.eks. være en funktion, der angav det samlede antal plat, eller det kunne være en funktion, der gav værdien 10, hvis det andet kast gav krone, og ellers -5.

Helt generelt indfører vi altså:

Definition 6: En ***stokastisk variabel*** X i et endeligt sandsynlighedsfelt er en funktion $X : U \mapsto \mathbb{R}$, der til ethvert udfald i udfaldsrummet knytter et reelt tal.

Eksempel 18: Hvis man kaster en mønt tre gange, kan en stokastisk variabel være den funktion, der til hvert af de 8 udfald knytter antallet af krone:

U	kkk	kkp	kpk	kpp	pkk	pkp	ppk	ppp
$X(u)$	3	2	2	1	2	1	1	0

Hvis man knytter en sandsynlighedsfunktion på, får man:

t	0	1	2	3
$P(X=t)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Her er *t* altså de værdier, som den stokastiske variabel kan antage.

Eksempel 19: Et lykkehjul med forskellige farver kan drejes, og eksperimentet bestående i *at dreje én gang* kan beskrives ved sandsynlighedsfeltet:

U	Grøn	Blå	Sort	Gul	Violet	Hvid
$P(u)$	0,08	0,05	0,15	0,30	0,02	0,40

Lykkehjulet anvendes i et tivoli, så man kan vinde penge, hvis man er heldig, og vores stokastiske variabel skal i dette tilfælde være en funktion, der "oversætter" farven til et beløb (i kroner). Det kunne f.eks. være:

U	Grøn	Blå	Sort	Gul	Violet	Hvid
$X(u)$	10	20	5	2	50	0

Opgaverne 208*

Når vi har indført begrebet stokastisk variabel, kan vi indføre følgende størrelser, der fortæller noget om den stokastiske variabel (og dermed om det eksperiment, som den stokastiske variabel er en del af):

Middelværdi, spredning og varians

Definition 7: I et endeligt sandsynlighedsfelt med den stokastiske variabel X, der kan antage værdierne $\{x_1, x_2, x_3, \ldots, x_m\}$, indføres størrelserne:

Middelværdi (μ eller $E(X)$): $$\mu = \sum_{i=1}^{m} x_i \cdot P(X = x_i)$$

Varians: $$\operatorname{var}(X) = \sum_{i=1}^{m} (x_i - \mu)^2 \cdot P(X = x_i)$$

Spredning: $$\sigma = \sqrt{\operatorname{var}(X)}$$

Eksempel 20: Vi vil bestemme middelværdi, varians og spredning for den stokastiske variabel indført i Eksempel 18:

$$\mu = \sum_{i=1}^{4} x_i \cdot P(X = x_i) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \underline{\underline{\frac{3}{2}}}$$

$$\operatorname{var}(X) = \sum_{i=1}^{4} (x_i - \mu)^2 \cdot P(X = x_i) = \left(0 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} =$$

$$\frac{9}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{9}{4} \cdot \frac{1}{8} = \frac{12}{16} = \underline{\underline{\frac{3}{4}}}$$

$$\sigma = \sqrt{\operatorname{var}(X)} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2} \approx 0{,}866$$

Vi vil altså i gennemsnit få 1,5 krone, når vi udfører eksperimentet.

Eksempel 21: Vi udregner middelværdien i Eksempel 19:

$$\mu = \sum_{i=1}^{6} x_i \cdot P(X = x_i) = 10 \cdot 0{,}08 + 20 \cdot 0{,}05 + 5 \cdot 0{,}15 + 2 \cdot 0{,}30 + 50 \cdot 0{,}02 + 0 \cdot 0{,}40 =$$

$$0{,}8 + 1 + 0{,}75 + 0{,}6 + 1 + 0 = 4{,}15$$

Dvs. man vinder i gennemsnit 4,15 pr. spil (som sikkert koster et sted mellem 10 og 20 kroner pr. spil).

I Gym-pakken ligger kommandoerne *middel*, *varians* og *spredning*, der kan bruges til at udregne disse værdier, når den stokastiske variabels sandsynlighedstabel er angivet som en $n \times 2$-tabel. Dette gøres ved at adskille de to rækker med en lodret linje som vist nedenfor:

$restart$

$with(Gym):$

$$X := \left\langle 0,\ 1,\ 2,\ 3 \,\middle|\, \frac{1}{8},\ \frac{3}{8},\ \frac{3}{8},\ \frac{1}{8} \right\rangle = \begin{bmatrix} 0 & \frac{1}{8} \\ 1 & \frac{3}{8} \\ 2 & \frac{3}{8} \\ 3 & \frac{1}{8} \end{bmatrix}$$

Anvend tastaturet, eller find linjen i Maple under 'relational'.
Det ser ud til, at både *VerticalBar* og *VerticalLine* kan bruges.

$middel(X) = 1.50000000000000$

$varians(X) = 0.750000000000000$

$spredning(X) = 0.866025403784439$

Opgaverne 209*

Væddemål

I situationer, hvor det er muligt at fastsætte sandsynligheder for de mulige udfald af et væddemål samt knytte reelle tal til de enkelte udfald, kan man bruge middelværdien af den stokastiske variabel, der beskriver væddemålet, til at afgøre, om det er et favorabelt væddemål.

På nuværende tidspunkt mangler vi noget kombinatorik for at kunne se på nogle sofistikerede væddemål, så det følgende eksempel er forholdsvis simpelt:

Eksempel 22a: *Hvis du kan slå to 6'ere i ét slag med disse to terninger, får du 50 kr., men hvis du ikke kan, skal du betale mig 2 kr.*
Vil du gå med til det væddemål?

Sandsynligheden for at slå to 6'ere i ét slag med to terninger er $\frac{1}{36}$, så den stokastiske variabel *X*, der beskriver væddemålet er:

x	50	-2
$P(x)$	$\frac{1}{36}$	$\frac{35}{36}$

Middelværdien for denne stokastiske variabel er:

$$E(X)=\sum_{i=1}^{n} x_i \cdot P(X=x_i)=50\cdot\frac{1}{36}+(-2)\cdot\frac{35}{36}=-\frac{5}{9}<0$$

Da middelværdien er negativ, kan det ikke betale sig for dig at indgå dette væddemål.

Dette stemmer med tommelfingerreglen, der siger, at hvis nogen tilbyder dig et væddemål, skal du sige nej, da du må gå ud fra, at det ikke kan betale sig for dig, for hvorfor skulle du ellers få tilbuddet?

Eksempel 22b: Det koster 10 kroner at købe et lod i tombolaen, hvor der er 5000 sedler. Der er én seddel med hovedgevinsten på 15000 kroner, 50 sedler med en gevinst på 200 kroner og 100 sedler med en gevinst på 50 kroner.
Vi vil regne på, om det kan betale sig at købe et lod:

X / Gevinst	0	50	200	15000
P	$\frac{4849}{5000}$	$\frac{1}{50}$	$\frac{1}{100}$	$\frac{1}{5000}$

$$E(X)=\sum_{i=1}^{n} x_i \cdot P(X=x_i)=0\cdot\frac{4849}{5000}+50\cdot\frac{1}{50}+200\cdot\frac{1}{100}+15000\cdot\frac{1}{5000}=0+1+2+3=6$$

Da et lod koster 10 kroner, og den gennemsnitlige gevinst er 6 kroner, kan det sjovt nok ikke betale sig at købe et lod.

Man kunne også med det samme medregne loddets pris, hvilket gøres med den stokastiske variabel $(X-10)$, hvor man får:

X - 10 / Indtægt	-10	40	190	14990
P	$\frac{4849}{5000}$	$\frac{1}{50}$	$\frac{1}{100}$	$\frac{1}{5000}$

$$E(X-10)=\sum_{i=1}^{n} (x_i-10)\cdot P(X=x_i)=-10\cdot\frac{4849}{5000}+40\cdot\frac{1}{50}+190\cdot\frac{1}{100}+14990\cdot\frac{1}{5000}=-4$$

Med denne stokastiske variabel, der altså trækker 10 fra hver af *X*'s funktionsværdier, er det afgørende, om middelværdien er under 0, og da den er det, kan det ikke betale sig at købe et lod.

Øvelse 3: *Vi slår plat og krone med to mønter hver. Jeg kaster først. Hvis du får flere plat end mig, får du 10 kr. Hvis du ikke får flere plat end mig, skal du betale 5 kr.*
Vil du indgå dette væddemål?

Tommelfingerreglen siger, at du skal afvise væddemålet. Men prøv selv at regne efter og se, at middelværdien er $-\frac{5}{16}$.

Opgaverne 210*

Sætninger om middelværdi, spredning og varians

I Eksempel 22b så vi et eksempel på, hvordan man kan danne en ny stokastisk variabel $(X-10)$ ud fra den stokastiske variabel X ved at trække 10 fra alle funktionsværdier. Og måske blev det bemærket, at middelværdien for den nye stokastiske variabel også blev 10 mindre end den oprindelige.

Det er et af resultaterne i følgende sætning:

Sætning 9: I et endeligt sandsynlighedsfelt med de stokastiske variable X og Y gælder følgende, når a og b er reelle tal:

Middelværdi:

$$a)\ E(a\cdot X+b)=a\cdot E(X)+b$$

$$b)\ E(X+Y)=E(X)+E(Y)$$

Spredning:

$$c)\ \sigma(a\cdot X+b)=|a|\cdot\sigma(X)$$

Varians:

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

$$e)\ \mathrm{var}(a\cdot X+b)=a^2\cdot\mathrm{var}(X)$$

Bemærk (se evt. Definition 6), at stokastiske variable er funktioner, dvs. $(X+Y)$ er en sumfunktion, og X^2 er en produktfunktion (da $X^2=X\cdot X$), mens $(a\cdot X+b)$ er en funktion multipliceret og efterfølgende adderet med konstanter. Præcis som vi kender det fra alle andre funktioner.

Dvs. $(X+Y)$ er den funktion, der til hvert udfald knytter summen af de tal, som de stokastiske variable X og Y hver især knytter til det pågældende udfald. Og X^2 kvadrerer X's funktionsværdier.

Eksempel 23a: Vi antager, at vi har nedenstående endelige sandsynlighedsfelt:

U	Blå	Grøn	Rød	Sort
P	0,4	0,3	0,2	0,1

Vi ser følgende to stokastiske variable X og Y i sandsynlighedsfeltet:

U	Blå	Grøn	Rød	Sort
X	2	2	4	5
Y	6	7	7	3

Da man altså f.eks. har $X(sort)=5$ og $Y(sort)=3$, gælder:

$$(X+Y)(sort)=5+3=8,\quad X^2(sort)=5^2=25\quad \text{og}\quad (4\cdot X+7)(sort)=4\cdot5+7=27$$

Eksempel 23b: Vi kan ud fra tabellerne i Eksempel 23a se, at man for den stokastiske variabel X får sandsynlighedsfunktionen:

t	2	4	5
$P(X=t)$	0,7	0,2	0,1

Disse sandsynligheder gælder også for X^2 og $a \cdot X + b$, da disse stokastiske variable jo tager udgangspunkt i X's funktionsværdier. Nogle eksempler er:

P	0,7	0,2	0,1		μ	σ
X	2	4	5		2,7	1,1
X^2	4	16	25		8,5	7,26
$3 \cdot X$	6	12	15		8,1	3,3
$X+15$	17	19	20		17,7	1,1
$3 \cdot X + 15$	21	27	30		23,1	3,3

Kan anvendes til at beregne var(X)

Tjek, at dette er i overensstemmelse med Sætning 9 a) og c).

Ifølge Sætning 9d har man altså:

$$\operatorname{var}(X) = E(X^2) - E(X)^2 = 8,5 - 2,7^2 = 1,21$$

Bemærk, at vi ikke kan opstille en tabel som i Eksempel 23b, når vi inddrager den stokastiske variabel *Y*, fordi den har andre sandsynligheder end *X*:

Eksempel 23c: Med informationerne fra Eksempel 23a har man:

t	2	4	5
$P(X=t)$	0,7	0,2	0,1

t	3	6	7
$P(Y=t)$	0,1	0,4	0,5

Da sandsynlighedsfunktionerne er forskellige, kan man ikke opstille en tabel som i Eksempel 23b. Hvis man skal kunne opskrive både X og Y i ét skema (og dermed også forskellige sammensætninger af disse, som f.eks $X+Y$, $X \cdot Y$ og $2 \cdot X - 3 \cdot Y$), er man derfor nødt til at gå helt tilbage og inddrage udfaldsrummet, da dette er fælles for de to stokastiske variable:

U	Blå	Grøn	Rød	Sort
P	0,4	0,3	0,2	0,1
X	2	2	4	5
Y	6	7	7	3
$X+Y$	8	9	11	8
$X \cdot Y$	12	14	28	15
$2 \cdot X - 3 \cdot Y$	-14	-17	-13	1

$$E(X) = \sum_{i=1}^{4} X(u_i) \cdot P(u_i) = 2 \cdot 0,4 + 2 \cdot 0,3 + 4 \cdot 0,2 + 5 \cdot 0,1 = 2,7$$

$$E(Y) = \sum_{i=1}^{4} Y(u_i) \cdot P(u_i) = 6 \cdot 0,4 + 7 \cdot 0,3 + 7 \cdot 0,2 + 3 \cdot 0,1 = 6,2$$

$$E(X+Y) = \sum_{i=1}^{4} (X+Y)(u_i) \cdot P(u_i) = 8 \cdot 0,4 + 9 \cdot 0,3 + 11 \cdot 0,2 + 8 \cdot 0,1 = 8,9$$

$$E(X \cdot Y) = \sum_{i=1}^{4} (X \cdot Y)(u_i) \cdot P(u_i) = 12 \cdot 0,4 + 14 \cdot 0,3 + 28 \cdot 0,2 + 15 \cdot 0,1 = 16,1$$

Det bemærkes, at $E(X+Y)$ er i overensstemmelse med Sætning 9b, men det er også værd at bemærke, at da $2,7 \cdot 6,2 = 16,74 \neq 16,1$, dvs. der gælder ikke generelt en tilsvarende sætning for multiplikation (men vi skal senere se, hvornår dette gælder).

Opgaverne 211*

Efter disse eksempler er vi klar til at bevise sætningerne:

Bevis 9: En stor del af beviserne består i at kunne regne med sumtegn.

9.a: Vi benytter Definition7 på middelværdi og bemærker, at det er selve værdierne af den stokastiske variabel, der bliver ændret med *a* og *b*, mens der ikke sker noget med sandsynlighedsfunktionen, der jo er tilknyttet det oprindelige sandsynlighedsfelt:

$$E(a\cdot X+b)=\sum_{i=1}^{m}(a\cdot x_i+b)\cdot P(X=x_i)=\sum_{i=1}^{m}a\cdot x_i\cdot P(X=x_i)+\sum_{i=1}^{m}b\cdot P(X=x_i)=$$

$$a\cdot\sum_{i=1}^{m}x_i\cdot P(X=x_i)+b\cdot\sum_{i=1}^{m}P(X=x_i)=a\cdot E(X)+b\cdot 1=a\cdot E(X)+b$$

Der blev i næst sidste skridt benyttet $\sum_{i=1}^{m}P(X=x_i)=1$, da sandsynlighederne for alle udfald og dermed også for alle værdier af den stokastiske variabel summerer op til 1.

9.e: Vi benytter Definition 7 på varians:

$$\mathrm{var}(a\cdot X+b)=\sum_{i=1}^{m}\left(a\cdot x_i+b-E(a\cdot X+b)\right)^2\cdot P(X=x_i)=$$

$$\sum_{i=1}^{m}\left(a\cdot x_i+b-a\cdot E(X)-b\right)^2\cdot P(X=x_i)=$$

$$\sum_{i=1}^{m}\left(a\cdot(x_i-E(X))\right)^2\cdot P(X=x_i)=$$

$$a^2\cdot\sum_{i=1}^{m}\left(x_i-E(X)\right)^2\cdot P(X=x_i)=a^2\cdot\mathrm{var}(X)$$

9.c: Dette følger direkte af Sætning 9.e samt Definition 7.

9.b: Når vi skal bevise denne sætning, er det vigtigt at være opmærksom på, at vi er nødt til at gå helt tilbage til udfaldsrummet, når vi arbejder med sumtegnene, for som vi så i Eksempel 23c, kunne forskellige udfald godt give ens værdier for *X*, men forskellige værdier for $X+Y$ (f.eks. *blå* og *grøn*).

$$E(X+Y)=\sum_{i=1}^{n}(X+Y)(u_i)\cdot P(u_i)=\sum_{i=1}^{n}\left(X(u_i)+Y(u_i)\right)\cdot P(u_i)=$$

$$\sum_{i=1}^{n}\left(X(u_i)\cdot P(u_i)+Y(u_i)\cdot P(u_i)\right)=\sum_{i=1}^{n}X(u_i)\cdot P(u_i)+\sum_{i=1}^{n}Y(u_i)\cdot P(u_i)=E(X)+E(Y)$$

9.d: Her skal man gøre sig klart, at man med X^2 mener den stokastiske variabel, der til hvert udfald giver kvadratet på den værdi, som den stokastiske variabel *X* giver.

$$\mathrm{var}(X)=\sum_{i=1}^{m}\left(x_i-E(X)\right)^2\cdot P(X=x_i)=$$

$$\sum_{i=1}^{m}\left(x_i^2+E(X)^2-2\cdot x_i\cdot E(X)\right)\cdot P(X=x_i)=$$

$$\sum_{i=1}^{m}x_i^2\cdot P(X=x_i)+\sum_{i=1}^{m}E(X)^2\cdot P(X=x_i)-\sum_{i=1}^{m}2\cdot x_i\cdot E(X)\cdot P(X=x_i)=$$

$$E(X^2)+E(X)^2\cdot\sum_{i=1}^{m}P(X=x_i)-2\cdot E(X)\cdot\sum_{i=1}^{m}x_i\cdot P(X=x_i)=$$

$$E(X^2)+E(X)^2\cdot 1-2\cdot E(X)\cdot E(X)=E(X^2)-E(X)^2$$

Eksempel 24: Sætning 9.d. blev tidligere - inden computeren overtog arbejdet - anvendt, når man skulle bestemme variansen, fordi det ofte var hurtigere at foretage denne udregning (se også Eksempel 23b nederst). Lad os se på eksempel 18 og sammenligne med eksempel 20:

U	kkk	kkp	kpk	kpp	pkk	pkp	ppk	ppp
$X(u)$	3	2	2	1	2	1	1	0
$X^2(u)$	9	4	4	1	4	1	1	0

x_i	0	1	2	3
$P(X=x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

x_i^2	0	1	4	9
$P(X=x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

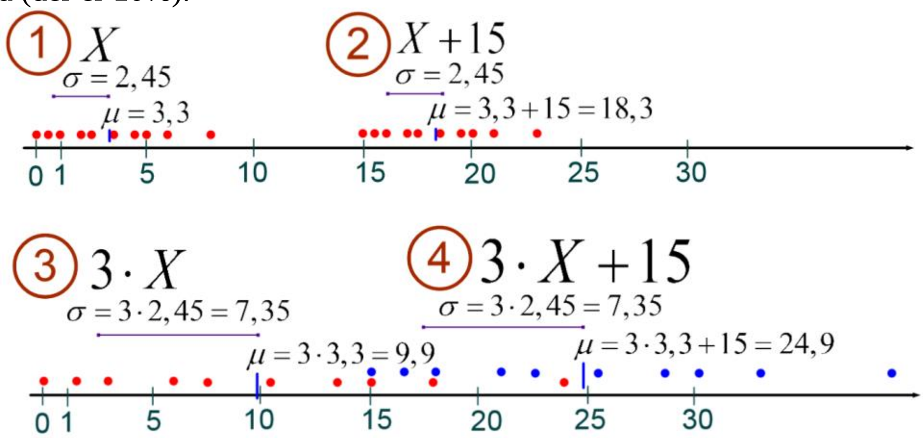
$$E(X)=\sum_{i=1}^{4}x_i\cdot P(X=x_i)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{3}{8}+\frac{6}{8}+\frac{3}{8}=\frac{12}{8}=\frac{3}{2}$$

$$E(X^2)=\sum_{i=1}^{4}x_i^2\cdot P(X=x_i)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+4\cdot\frac{3}{8}+9\cdot\frac{1}{8}=\frac{3}{8}+\frac{12}{8}+\frac{9}{8}=3$$

$$\mathrm{var}(X)=E(X^2)-(E(X))^2=3-\left(\frac{3}{2}\right)^2=3-\frac{9}{4}=\frac{3}{4}$$

Resultatet stemmer med eksempel 20 (men man kan også bemærke, at der i dette tilfælde ikke var det store regnearbejde at spare).

Sætning 9 angiver nogle regler for regning med middelværdier, spredning og varians. Da $\sigma=\sqrt{var(X)}$, fortæller spredningen og variansen på sin vis det samme om den stokastiske variabel. De fortæller det bare med forskellige værdier. Vi støder på dem igen inden for statistik. Vores regler skulle gerne stemme overens med vores intuitive forståelse af begreberne middelværdi og spredning. For se på nedenstående figur, hvor vi betragter en stokastisk variabel *X* samt nogle ændringer af denne. De røde og blå prikker repræsenterer de mulige værdier for den pågældende stokastiske variabel, og for at gøre det nemmere at overskue tildeles hver værdi samme sandsynlighed (der er 10%).



1: Den stokastiske variabel *X* har middelværdien 3,3 og spredningen 2,45. Spredningen er angivet med det violette linjestykke. Spredningen er et udtryk for, hvor spredt punkterne (vægtet) ligger omkring middelværdien.
2: Her ses på en ny stokastisk variabel, hvor der er lagt 15 til hver værdi. Dette øger middelværdien med 15, men spredningen påvirkes ikke.
3: Her er hver værdi for den stokastiske variabel multipliceret med 3. Dette påvirker både middelværdi og spredning (se figuren). Bemærk, hvordan punkterne bliver mere spredt.
4: Angivet med blå prikker.

Normeret stokastisk variabel

Enhver stokastisk variabel *X* kan *normeres*:

Definition 8: Lad *X* være en stokastisk variabel med middelværdi μ og spredning σ.
Den stokastiske variabel $N=\frac{X-\mu}{\sigma}$ kaldes så for den tilsvarende *normerede* stokastiske variabel.

Eksempel 25: Vi ser igen på Eksempel 18, hvor vi i Eksempel 20 beregnede, at $\mu=\frac{3}{2}$ og $\sigma=\frac{\sqrt{3}}{2}$:

X	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Den tilsvarende normerede stokastiske variabel $N=\frac{X-\frac{3}{2}}{\frac{\sqrt{3}}{2}}=\frac{2\cdot X-3}{\sqrt{3}}$ er så:

N	$\frac{2\cdot 0-3}{\sqrt{3}}=-\sqrt{3}$	$\frac{2\cdot 1-3}{\sqrt{3}}=-\frac{1}{\sqrt{3}}$	$\frac{2\cdot 2-3}{\sqrt{3}}=\frac{1}{\sqrt{3}}$	$\frac{2\cdot 3-3}{\sqrt{3}}=\sqrt{3}$
$P(U=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Sætning 10: En normeret stokastisk variabel har middelværdien 0 og spredningen 1.

Eksempel 26: Vi benytter Sætning 9 til at beregne middelværdi og spredning for den normerede stokastiske variabel fra Eksempel 25. Først benyttes Sætning 9.a:

$$E(U)=E\left(\frac{2\cdot X-3}{\sqrt{3}}\right)=E\left(\frac{2}{\sqrt{3}}\cdot X-\frac{3}{\sqrt{3}}\right)=\frac{2}{\sqrt{3}}\cdot E(X)-\frac{3}{\sqrt{3}}=\frac{2}{\sqrt{3}}\cdot\frac{3}{2}-\frac{3}{\sqrt{3}}=\frac{3}{\sqrt{3}}-\frac{3}{\sqrt{3}}=0$$

Dvs. middelværdien af den stokastiske variabel *U* er 0.

Sætning 9.c giver os: $\sigma(U)=\sigma\left(\frac{2\cdot X-3}{\sqrt{3}}\right)=\sigma\left(\frac{2}{\sqrt{3}}\cdot X-\frac{3}{\sqrt{3}}\right)=\frac{2}{\sqrt{3}}\cdot\sigma(X)=\frac{2}{\sqrt{3}}\cdot\frac{\sqrt{3}}{2}=1$

Dvs. at den stokastiske variabel *U* har spredningen 1.

Bevis 10: Vi kan udnytte Sætning 9 til at bevise Sætning 10:
Vores udgangspunkt er, at den stokastiske variabel *X* har middelværdien μ og spredningen σ (der er en ikke-negativ størrelse). Af Sætning 9.a følger:

$$E(N)=E\left(\frac{X-\mu}{\sigma}\right)=E\left(\frac{1}{\sigma}\cdot X-\frac{\mu}{\sigma}\right)=\frac{1}{\sigma}\cdot E(X)-\frac{\mu}{\sigma}=\frac{1}{\sigma}\cdot\mu-\frac{\mu}{\sigma}=0$$

Dvs. middelværdien for den stokastiske variabel *U* er 0.

Af Sætning 9.c følger: $\sigma(N)=\sigma\left(\frac{X-\mu}{\sigma(X)}\right)=\sigma\left(\frac{1}{\sigma(X)}\cdot X-\frac{\mu}{\sigma(X)}\right)=\frac{1}{\sigma(X)}\cdot\sigma(X)=1$

Dvs. den stokastiske variabel *N* har spredningen 1.

Normering er en metode til at gøre forskellige stokastiske variable sammenlignelige, på samme måde som indekstal indføres for at kunne sammenligne prisudviklinger af forskellige størrelser.

Opgaverne 212*

Uafhængige stokastiske variable

I Definition 5 og sætningerne 6 og 7 så vi på uafhængige hændelser. Vi kan på samme måde tale om uafhængige stokastiske variable.

Man skal huske på, at en stokastisk variabel er en funktion, der til ethvert element i udfaldsrummet knytter et reelt tal, dvs. den kan godt knytte det samme tal til flere forskellige udfald, og en stokastisk variabel kan så være en metode til at få dannet forskellige hændelser, nemlig hvis man lader de forskellige hændelser hver især bestå af de udfald, som den stokastiske variabel tildeler samme værdi.

Eksempel 27: Vi ser igen på situationen fra Eksempel 18 (en mønt kastes tre gange og den stokastiske variabel X angiver antal 'krone').

U	kkk	kkp	kpk	kpp	pkk	pkp	ppk	ppp
$X(u)$	3	2	2	1	2	1	1	0

I dette tilfælde danner den stokastiske variabel hændelserne:

$$H_1 = \{ppp\}$$
$$H_2 = \{kpp, pkp, ppk\}$$
$$H_3 = \{kkp, kpk, pkk\}$$
$$H_4 = \{kkk\}$$

Vi benytter derfor nu Sætning 7 som udgangspunkt for følgende definition:

Definition 9: De stokastiske variable X og Y kaldes *uafhængige*, hvis det gælder, at:

$$P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \quad \forall x_i \in Vm(X), y_j \in Vm(Y)$$

Eksempel 28: Vi ser endnu engang på situationen med 3 kast med en mønt. Vi ser på tilfældet:

X: Angiver antallet af krone.
Y: Angiver antallet af plat.

Vores klare fornemmelse må være, at disse to stokastiske variable IKKE er uafhængige, for hvis man kender værdien for den ene, kender man den også for den anden ud fra ligningen $y = 3 - x$.

Vi ser på et par situationer:

a) X giver 2, og Y giver 2.

Vi kan ikke have begge dele på én gang, så $P(X = 2 \wedge Y = 2) = 0$.

Vi har $P(X = 2) = \frac{3}{8}$, og $P(Y = 2) = \frac{3}{8}$, dvs. $P(X = 2) \cdot P(Y = 2) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}$

Da $P(X = 2 \wedge Y = 2) \neq P(X = 2) \cdot P(Y = 2)$, er de stokastiske variable IKKE uafhængige, og derfor behøver vi egentlig ikke at se på flere udregninger, men her kommer en mere for forståelsens skyld.

b) X giver 1, og Y giver 2. Her har vi:

$P(X = 1 \wedge Y = 2) = \frac{3}{8}$ og $P(X = 1) \cdot P(Y = 2) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}$

Dvs. $P(X = 1 \wedge Y = 2) \neq P(X = 1) \cdot P(Y = 2)$

Eksempel 29: Tre kast med en mønt og følgende stokastiske variable:

X: Antal krone.

Y: Antal kast

Bemærk, at *Y* er en ret kedelig stokastisk variabel, der hele tiden giver 3.

Dermed skulle det også være oplagt, at disse to stokastiske variable er uafhængige.

Der ses på nogle situationer (tænk selv over sandsynlighederne):

a) $X = 0$ og $Y = 3$:

$$P(X=0 \wedge Y=3) = \frac{1}{8} \qquad P(X=0) = \frac{1}{8} \qquad P(Y=3) = 1$$

$$Dvs.\ P(X=0 \wedge Y=3) = P(X=0) \cdot P(Y=3)$$

b) $X = 1$ og $Y = 3$:

$$P(X=1 \wedge Y=3) = \frac{3}{8} \qquad P(X=1) = \frac{3}{8} \qquad P(Y=3) = 1$$

$$Dvs.\ P(X=1 \wedge Y=3) = P(X=1) \cdot P(Y=3)$$

c) $X = 2$ og $Y = 3$:

$$P(X=2 \wedge Y=3) = \frac{3}{8} \qquad P(X=2) = \frac{3}{8} \qquad P(Y=3) = 1$$

$$Dvs.\ P(X=2 \wedge Y=3) = P(X=2) \cdot P(Y=3)$$

d) $X = 3$ og $Y = 3$:

$$P(X=3 \wedge Y=3) = \frac{1}{8} \qquad P(X=3) = \frac{1}{8} \qquad P(Y=3) = 1$$

$$Dvs.\ P(X=3 \wedge Y=3) = P(X=3) \cdot P(Y=3)$$

Vi har nu gennemgået alle kombinationer og set, at udtrykket er sandt i alle tilfælde, og dermed er disse to stokastiske variable uafhængige.

Eksempel 30: Igen 3 kast med en mønt, dvs. $U = \{kkk, kkp, kpk, kpp, pkk, pkp, ppk, ppp\}$.

X: Antal krone.

Z: Antal skift mellem plat til krone.

Vi ser på 4 ud af de 12 mulige kombinationer (overvej selv sandsynlighederne).

a) $X = 0 \wedge Z = 0 \quad P(X=0 \wedge Z=0) = \frac{1}{8} \quad P(X=0) = \frac{1}{8} \quad P(Z=0) = \frac{1}{4}$ Falsk

b) $X = 0 \wedge Z = 1 \quad P(X=0 \wedge Z=1) = 0 \quad P(X=0) = \frac{1}{8} \quad P(Z=1) = \frac{1}{2}$ Falsk

c) $X = 0 \wedge Z = 2 \quad P(X=0 \wedge Z=2) = 0 \quad P(X=0) = \frac{1}{8} \quad P(Z=2) = \frac{1}{4}$ Falsk

h) $X = 2 \wedge Z = 1 \quad P(X=2 \wedge Z=1) = \frac{1}{4} \quad P(X=2) = \frac{3}{8} \quad P(Z=1) = \frac{1}{2}$ Falsk

Endnu engang har vi altså IKKE uafhængige stokastiske variable.

Bemærk, at Definition 9 ikke er en sætning (muligvis er navnet alene nok til at give én en mistanke om, at det forholder sig sådan). Det er en definition, der sikrer, at vi kan bruge ordet *uafhængighed* på samme måde, som vi gjorde i forbindelse med hændelser. Når vi senere anvender denne sætning, er det IKKE som i eksemplerne ovenfor ved at udregne sandsynligheder og teste **om** *X* og *Y* er uafhængige. Vi anvender den modsat ved at **antage** eller **ræsonnere** os frem til, at *X* og *Y* er uafhængige, hvorefter vi kan udnytte, at $P(X = x_i \wedge Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$.

Opgaverne 213*

Sætning 11: For uafhængige stokastiske variable X og Y gælder: $E(X\cdot Y)=E(X)\cdot E(Y)$

Vi husker, at den stokastiske variabel $X\cdot Y$ er den stokastiske variabel, der til ethvert udfald knytter produktet af de værdier, som de enkelte stokastiske variable X og Y hver især knytter. Det er derfor nødvendigt, at de to stokastiske variable X og Y er baseret på det samme sandsynlighedsfelt, hvilket de strengt taget ikke behøver at være i Definition 9.

Men ovenstående er ikke så væsentligt, da man altid kan tilpasse udfaldsrummet og dermed sandsynlighedsfeltet. F.eks. kan et kast med en terning og en mønt beskrives med to udfaldsrum $U_1=\{1,2,3,4,5,6\}$ og $U_2=\{k,p\}$, men hvis det er nødvendigt, kan disse slås sammen til ét $U=\{1k,1p,2k,2p,3k,3p,4k,4p,5k,5p,6k,6p\}$.

I beviset for Sætning 11 anvendes dobbelte sumtegn, der skal forstås på den måde, at for hver værdi af i udregnes hele sumtegnet med j. Tænk over, hvorfor hvert skridt i beviset er gyldigt. Specielt er det vigtigt at forstå det første lighedstegn ned til mindste detalje.

Bevis 11: Vores udgangspunkt er, at vi har to **uafhængige** stokastiske variable X og Y, dvs. vi ved, at $P(X=x_i \wedge Y=y_j)=P(X=x_i)\cdot P(Y=y_j)\quad \forall x_i\in Vm(X), y_j\in Vm(Y)$. Den stokastiske variabel X's værdimængde består af k elementer, mens $Vm(Y)$ indeholder l elementer.

Vi benytter så Definition 7 (på middelværdi) og regner løs:

$$E(X\cdot Y)=\sum_{i=1}^{k}\left(\sum_{j=1}^{l}x_i\cdot y_j\cdot P(X=x_i\wedge Y=y_j)\right)=$$

$$\sum_{i=1}^{k}\left(\sum_{j=1}^{l}x_i\cdot y_j\cdot P(X=x_i)\cdot P(Y=y_j)\right)=$$

$$\sum_{i=1}^{k}\left(x_i\cdot P(X=x_i)\cdot\sum_{j=1}^{l}y_j\cdot P(Y=y_j)\right)=$$

$$\sum_{i=1}^{k}\left(x_i\cdot P(X=x_i)\cdot E(Y)\right)=E(Y)\cdot\sum_{i=1}^{k}x_i\cdot P(X=x_i)=E(Y)\cdot E(X)$$

Eksempel 31: I Eksempel 29 så vi på de ret oplagte uafhængige stokastiske variable, hvor X var antal 'krone', og Y var antal kast. Y har oplagt middelværdien 3, og Sætning 11 giver så:

$$E(X\cdot Y)=E(X)\cdot E(Y)=\frac{3}{2}\cdot 3=\frac{9}{2}$$

Det er ikke verdens mest overraskende resultat, da den stokastiske variabel $X\cdot Y$ giver os antal 'krone' multipliceret med 3.

Eksempel 32: En rød og en sort terning kastes. Vi ønsker at finde middelværdien for den stokastiske variabel, der angiver produktet af øjentallene.

Vi lader X og Y være de stokastiske variable, der angiver øjentallene for henholdsvis den røde og den sorte terning. X og Y er oplagt uafhængige (øjentallene på de to terninger afhænger ikke af hinanden), og de har begge middelværdien $E(X)=E(Y)=3{,}5$, da alle øjentallene 1, 2, 3, 4, 5 og 6 har samme sandsynlighed.

$X\cdot Y$, der angiver produktet af øjentallene, har dermed ifølge Sætning 11 middelværdien:

$$E(X\cdot Y)=E(X)\cdot E(Y)=3{,}5\cdot 3{,}5=12{,}25$$

I opgave 2112 a) blev dette tal udregnet ud fra Definition 7. Anvendelsen af Sætning 11 gør udregningen væsentlig kortere.

Opgaverne 214*

Der gælder også en sætning om variansen i forbindelse med uafhængige stokastiske variable:

Sætning 12: Hvis X og Y er uafhængige stokastiske variable, er $Var(X+Y)=Var(X)+Var(Y)$.

Bevis 12: Vi benytter Sætning 9a $(magenta)$, 9b $(blå)$ samt 9d $(rød)$, hvor varianserne kan udregnes ud fra middelværdier, og da vores forudsætning er, at X og Y er uafhængige stokastiske variable, kan vi også benytte Sætning 11 $(grøn)$.

Vi får så:

$$
\begin{aligned}
Var(X+Y) = {} & E\left((X+Y)^2\right)-\left(E(X+Y)\right)^2 = \\
& E\left(X^2+Y^2+2\cdot X\cdot Y\right)-\left(E(X)+E(Y)\right)^2 = \\
& E\left(X^2\right)+E\left(Y^2\right)+E(2\cdot X\cdot Y)-\left(E(X)\right)^2-\left(E(Y)\right)^2-2\cdot E(X)\cdot E(Y) = \\
& E\left(X^2\right)-\left(E(X)\right)^2+E\left(Y^2\right)-\left(E(Y)\right)^2+2\cdot E(X\cdot Y)-2\cdot E(X)\cdot E(Y) = \\
& Var(X)+Var(Y)+2\cdot E(X)\cdot E(Y)-2\cdot E(X)\cdot E(Y)=Var(X)+Var(Y)
\end{aligned}
$$

Eksempel 33: Vi ser endnu engang på Eksempel 29, hvor vi havde uafhængige stokastiske variable. Variansen for Y er 0, da den tildeler alle udfald værdien 3, og vi har så ifølge Sætning 12:

$$Var(X+Y)=Var(X)+Var(Y)=Var(X)+0=Var(X)=\frac{3}{4}$$

Opgaverne 215*

Vi får brug for sætningerne 11 og 12, når vi i forbindelse med kombinatorik skal bestemme middelværdier og spredninger for forskellige standardfordelinger.

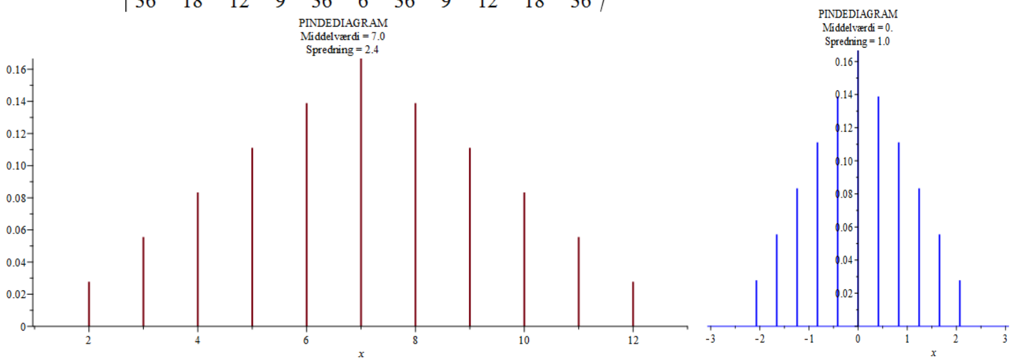
Tæthedsfunktion og fordelingsfunktion

Når vi arbejder med vores diskrete sandsynlighedsfelter med tilknyttet stokastisk variabel X, har vi vores sandsynlighedsfunktion $P(X=x_i)$, der til hver værdi af den stokastiske variabel knytter sandsynligheden for denne værdi. Dette kan afbildes i et *pindediagram* med Gym-pakken:

Eksempel 34: Dataene fra Eksempel 3 afbildes i Maple (se pindediagrammet til venstre):

$$M := \left\langle 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \middle| \frac{1}{36}, \frac{1}{18}, \frac{1}{12}, \frac{1}{9}, \frac{5}{36}, \frac{1}{6}, \frac{5}{36}, \frac{1}{9}, \frac{1}{12}, \frac{1}{18}, \frac{1}{36} \right\rangle :$$

plotPindediagram(M)



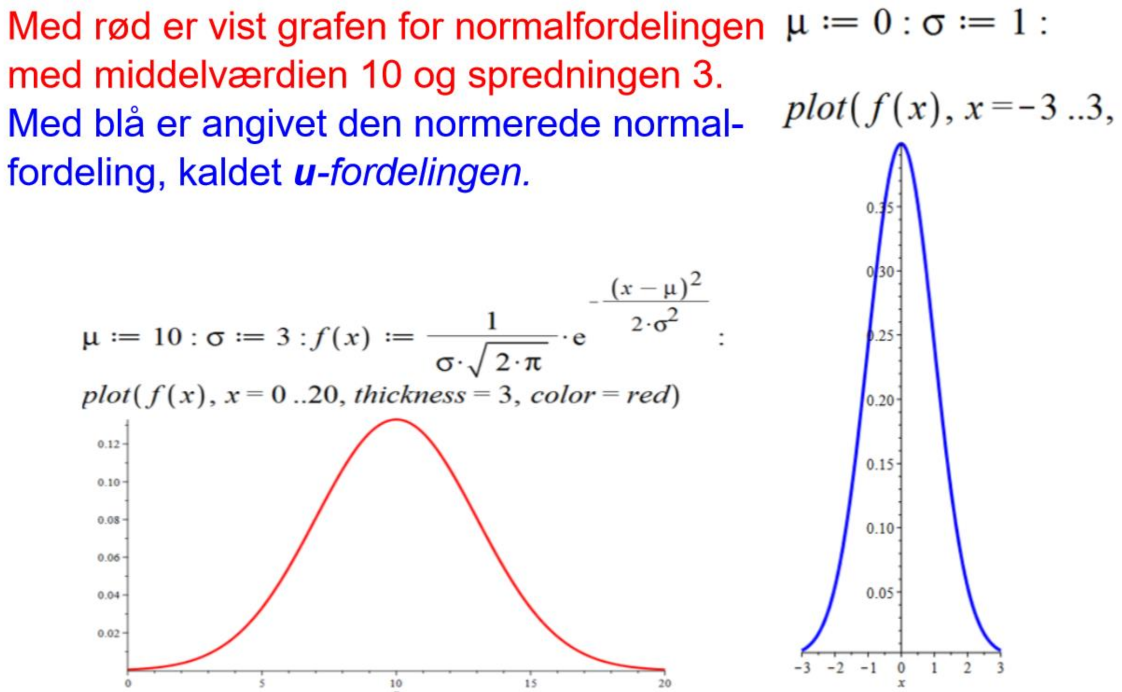
Desuden ses til højre (pindediagrammet med de blå pinde) den tilsvarende normerede stokastiske variabel $\left(\frac{M-7}{2,4}\right)$. Bemærk, at sandsynlighederne er de samme, men at pindene er forskudt mod venstre (så middelværdien bliver 0) og skubbet sammen (så spredningen bliver 1).

Vi skal senere inden for statistik arbejde med kontinuerte funktioner, og her kan man ikke arbejde med sandsynlighedsfunktioner på samme måde, da sandsynligheden for et konkret udfald i en kontinuert fordeling vil være 0 (da der er uendelig mange udfald inden for ethvert interval).
I disse tilfælde arbejder man med *tæthedsfunktioner*.

Definition 10: For et (kontinuert) udfaldsrum $U = I$, hvor I er et interval, kaldes en funktion f for *tæthedsfunktionen*, hvis der gælder $P(u \in [a,b]) = \int_a^b f(x)\,dx$ for alle $a,b \in I$ med $b > a$, hvor $P(u \in [a,b])$ er sandsynligheden for, at udfaldet ligger inden for intervallet $[a,b]$.
Desuden skal $\int_U f(x)\,dx = 1$, dvs. integration over hele udfaldsrummet skal give 1.

Eksempel 35a: De mest kendte og mest direkte eller indirekte anvendte tæthedsfunktioner er helt klart *normalfordelinger* (eller *Gauss-fordelinger*).
Med middelværdien μ og spredningen σ er forskriften:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$$



Kig på både første- og andenaksen. Bemærk, at man ved normeringen (fra rød til blå graf) både har fået gjort grafen smallere og højere i modsætning til diskrete sandsynlighedsfelter (Eksempel 34), hvor den kun bliver smallere. Det skyldes, at man på andenaksen ikke længere har sandsynligheder, men sandsynligheds**tætheder** (det går vi mere i dybden med, når vi vender tilbage til normalfordelinger i forbindelse med statistik).
Man kan også forstå det ud fra Definition 10, hvor det fremgår, at arealet under graferne skal være 1, da tæthedsfunktionen er en sandsynlighedsfunktion.

Lad os se på, hvordan man kan bruge Definition 10 til at beregne sandsynligheder med udgangspunkt i en normalfordeling.

Eksempel 35b: Vi ser på normalfordelingen med middelværdien 100 og spredningen 20:

$$f(x) = \frac{1}{20 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-100)^2}{2 \cdot 20^2}}$$

Hvis vi ønsker at finde sandsynligheden for, at et udfald ligger i intervallet [90,105], dvs. $P(u \in [90,105])$, har vi ifølge Definition 10:

$$P(u \in [90,105]) = \int_{90}^{105} f(x)\,dx$$

Beregningen udføres i Maple:

$$f(x) := \frac{1}{20 \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-100)^2}{2 \cdot 20^2}} :$$

$$P = \int_{90.}^{105} f(x)\,dx = 0.2901687870$$

Dvs. der er 29% chance for, at udfaldet er mellem 90 og 105.

Da $U = \mathbb{R}$, kan man tjekke, at $P(u \in U) = 1$ (hvilket jo pr. definition SKAL gælde) ved:

$$P(u \in U) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Opgaverne 216*

Begrebet *tæthedsfunktion* er egentlig knyttet til kontinuerte udfaldsrum, men man kan også anvende det på diskrete udfaldsrum. Man taler så om *diskrete tæthedsfunktioner*, men hvis du kigger på vores definitioner, kan du se, at det er det samme som de *sandsynlighedsfunktioner*, vi indførte i definitionerne 1 og 3.

Det følgende begreb anvendes – hvilket også klart fremgår af definitionen – både for diskrete og kontinuerte udfaldsrum:

Definition 11: Fordelingsfunktion.

For et sandsynlighedsfelt med tilknyttet stokastisk variabel *X* og sandsynlighedsfunktion P, er *fordelingsfunktionen F* den funktion, der til enhver værdi $x \in \mathbb{R}$ angiver sandsynligheden for **højst** at opnå værdien *x*, dvs:

$$F(x) = P(X \leq x).$$

For et (kontinuert) udfaldsrum $U = I$ (hvor *I* er et interval) med tæthedsfunktionen f, er *fordelingsfunktionen F* givet ved:

$$F(x) = \int_a^x f(t)\,dt \quad ,$$

hvor *a* er intervallet *I*'s infimum (nedre grænse). Hvis der ikke er et infimum, erstattes *a* af $-\infty$.

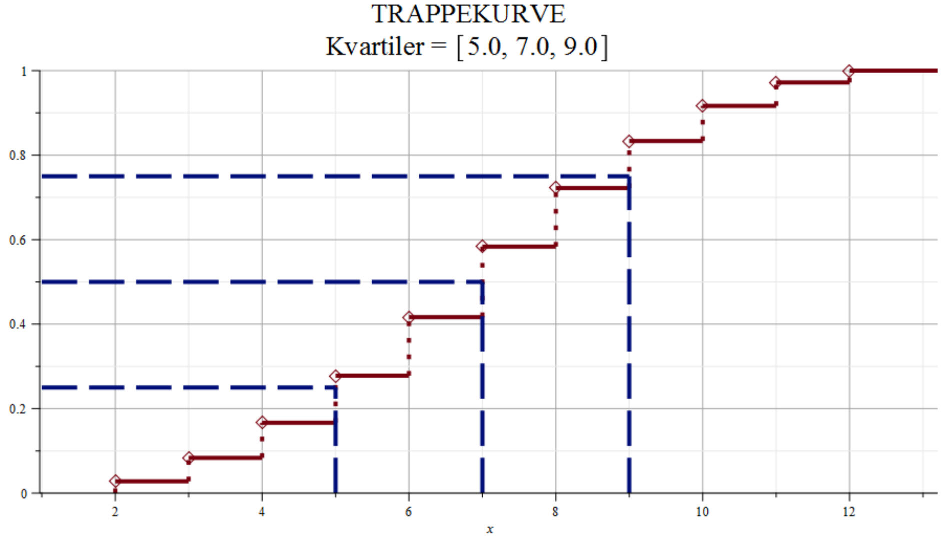
Eksempel 36a (endeligt/diskret sandsynlighedsfelt): Fra Eksempel 3 har vi følgende sandsynlighedsfunktion $P(x)$ og *fordelingsfunktion* $F(x)$:

X	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

Eksempel 36b: En fordelingsfunktion for et sandsynlighedsfelt med tilknyttet stokastisk variabel vil afbildes som et såkaldt trappediagram:

$$M := \left\langle 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \middle| \frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36} \right\rangle :$$

plotTrappekurve(*M*)

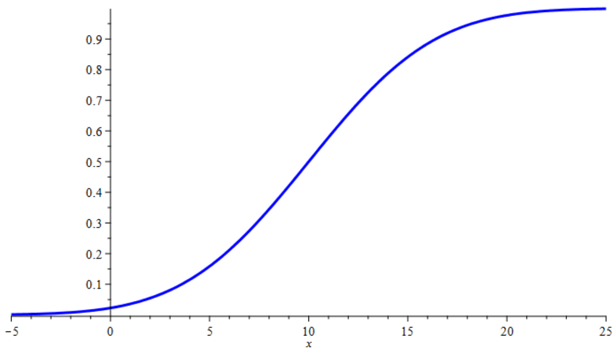


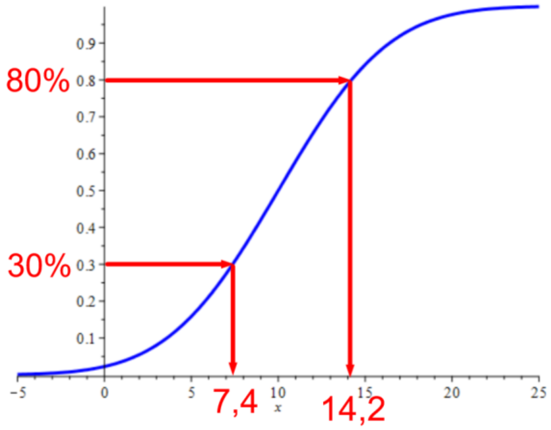
De blå linjer, der udgår fra frekvenserne 25%, 50% og 75% på andenaksen, anvendes til at aflæse det såkaldte kvartilsæt, som består af de tilsvarende aflæsninger på førsteaksen (5, 7 og 9). Det vender vi tilbage til, men du kan godt allerede overveje, hvordan *nedre kvartilen* 5, *medianen* 7 og *øvre kvartilen* 9 skal fortolkes.

Eksempel 37 (kontinuert sandsynlighedsfelt): Vi ser på fordelingsfunktionen for normalfordelingen med middelværdien 10 og spredningen 5:

$$F(x) := \int_{-\infty}^{x} \frac{1}{5\cdot\sqrt{2\cdot\pi}} \cdot \mathrm{e}^{-\frac{(a-10)^2}{2\cdot 5^2}} \, \mathrm{d}a :$$

plot(*F*(*x*), *x* = −5 ..25, *thickness* = 3, *color* = *blue*)





Bemærk, at F (selvfølgelig) er en voksende funktion, og $F(x) \to 1$ for $x \to \infty$.

fsolve(*F*(*x*) = 0.3) = 7.377997436
fsolve(*F*(*x*) = 0.8) = 14.20810617

På grafen til højre er illustreret, at 80% af alle udfald vil være under (eller lig) 14,2, mens 30% af udfaldene vil være under (eller lig) 7,4.

Det er desuden vist, hvordan man med *fsolve* kan bestemme disse værdier i Maple.

KOMBINATORIK

Kombinatorik beskæftiger sig med endelige eller tællelige diskrete strukturer. Som du husker fra forløbet om uendeligheder, vil det altså sige, at strukturen enten skal være endelig eller skal kunne nummereres med de naturlige tal (hvilket f.eks. ikke var tilfældet med intervallet [0,1], der heller ikke er en diskret struktur).

Man kan også sige, at kombinatorik omhandler det at tælle (elementer i) mængder.

Vi får derfor brug for begrebet *kardinaltal*, der blev introduceret i forbindelse med uendelighedsbegrebet. Der anvendte vi det hovedsageligt til at skelne mellem forskellige slags uendeligheder ved at betegne disse med $\aleph_0, \aleph_1, \aleph_2, ...$

Men nu arbejder vi hovedsageligt med endelige mængder, og her bliver det noget simplere, da *kardinaltallet* for en endelig mængde simpelthen er antallet af elementer i mængden. Vi bruger samme notation, som vi kender fra numerisk værdi, længde af linjestykke og længde af vektor. Dvs. vi har:

Hvis $A = \{-8, -3, 0, 7, 26\}$, er $|A| = 5$, hvilket udtales "*Mængden A har kardinaltallet 5*".

Hvis $B = \{abc, bdc, eab, bca\}$, er $|B| = 4$, dvs. "*B har kardinaltallet 4*".

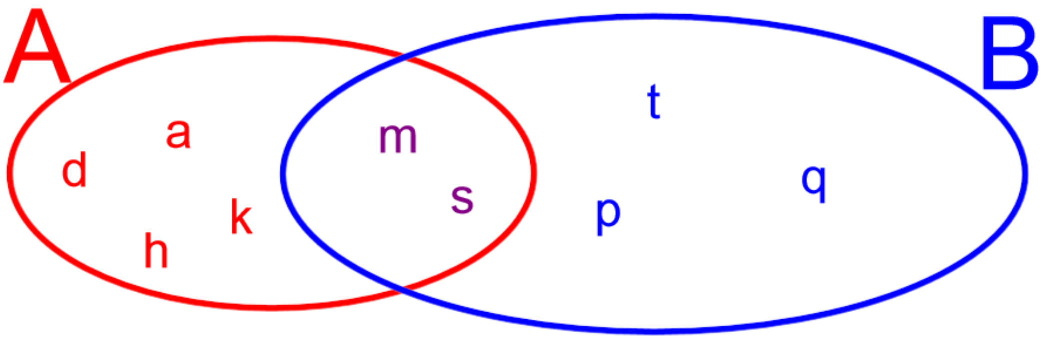
Hvis $C = \{2, 4, 6, 8, ..., 100\}$, er $|C| = 50$.

Og vi har: $|\varnothing| = 0$.

Vores første princip illustrerer klart udtalelsen om, at man inden for kombinatorik beskæftiger sig med at tælle elementer i mængder, for det er netop en tællemetode. Det er:

Princippet om inklusion og eksklusion.

Vi ser først på to mængder *A* og *B*, hvor vi ønsker at tælle antallet af elementer i foreningsmængden $A \cup B$, dvs. alle de størrelser, der enten ligger i *A* eller i *B* eller i begge.



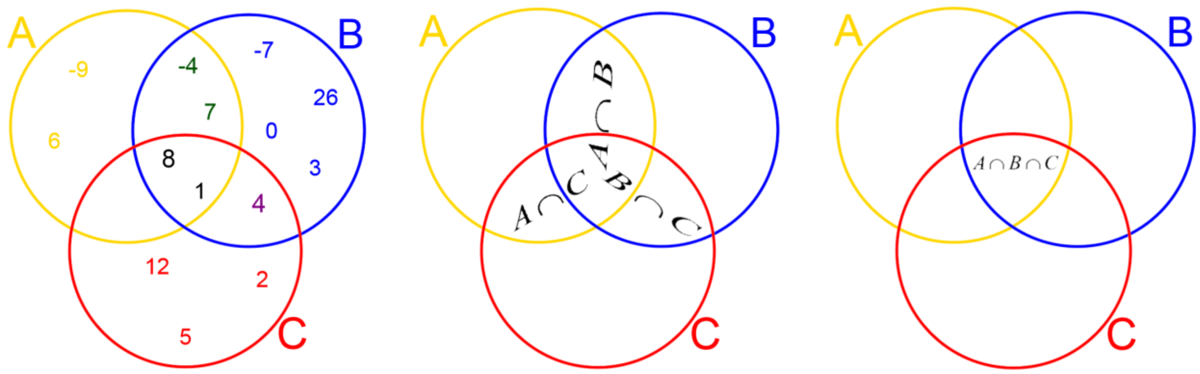
I ovenstående tilfælde har vi $A \cup B = \{a, d, h, k, m, s, p, q, t\}$ og altså $|A \cup B| = 9$.

Vi kan hurtigt se, at vores antal 9 ikke fremkommer ved at lægge antallet af elementer i *A* sammen med antallet af elementer i *B*, for der er 6 elementer i *A* og 5 elementer i *B*, og summen af disse tal er 11. De 2 "ekstra" elementer skyldes, at vi er kommet til at tælle elementerne *m* og *s* med to gange, fordi de både ligger i *A* og i *B*. Men det kan vi rette op på ved at trække dem fra igen én gang. Det gøres ved at trække antallet af elementer i fællesmængden $A \cap B$ fra, da $A \cap B = \{m, s\}$.

Vi har dermed generelt:

$$|A \cup B| = \underbrace{|A| + |B|}_{\text{Inklusion}} - \underbrace{|A \cap B|}_{\text{Eksklusion}}$$

Vi ser nu på tre mængder *A*, *B* og *C*, hvor vi ønsker at bestemme $|A \cup B \cup C|$:



I det konkrete tilfælde til venstre har man:

$$A = \{-9,-4,1,6,7,8\} \qquad |A| = 6$$

$$B = \{-7,-4,0,1,3,4,7,8,26\} \qquad |B| = 9$$

$$C = \{1,2,4,5,8,12\} \qquad |C| = 6$$

$$A \cap B = \{-4,1,7,8\} \qquad |A \cap B| = 4$$

$$A \cap C = \{1,8\} \qquad |A \cap C| = 2$$

$$B \cap C = \{1,4,8\} \qquad |B \cap C| = 3$$

$$A \cap B \cap C = \{1,8\} \qquad |A \cap B \cap C| = 2$$

$$A \cup B \cup C = \{-9,-7,-4,0,1,2,3,4,5,6,7,8,12,26\} \qquad |A \cup B \cup C| = 14$$

Tjek, at du har styr på, hvad der menes med de forskellige fællesmængder og foreningsmængden. Det er altså tallet 14, vi skal komme frem til.

Hvis vi går frem på samme måde, som vi gjorde i tilfældet med 2 mængder, kan vi prøve at lægge antallet af elementer i de enkelte mængder sammen og derefter fratrække antallene af elementer i de tre parvise fællesmængder ($A \cap B, A \cap C$ og $B \cap C$). Vi får så $6+9+6-4-2-3=12$, dvs. vi får 2 mindre end det ønskede. Dette skyldes elementerne i $A \cap B \cap C$. For først blev de talt med tre gange (da de tilhører alle mængderne *A*, *B* og *C*). Men derefter blev de trukket fra tre gange, da de også tilhører alle fællesmængderne $A \cap B$, $A \cap C$ og $B \cap C$. De skal derfor lægges til igen, og det gøres ved at lægge elementerne i $A \cap B \cap C$ til.

Vores generelle udtryk bliver altså:

$$|A \cup B \cup C| = \underbrace{|A| + |B| + |C|}_{\text{Inklusion}} \underbrace{-|A \cap B| - |A \cap C| - |B \cap C|}_{\text{Eksklusion}} \underbrace{+|A \cap B \cap C|}_{\text{Inklusion}}$$

Sammenlign udtrykket ovenfor med mængdediagrammerne. Tjek, at du kan se, at hvert element ender med at være blevet talt med netop én gang.

Med 4 mængder får man (bemærk, hvordan man skifter mellem inklusion og eksklusion):

$$\begin{aligned} |A \cup B \cup C \cup D| = {} & |A| + |B| + |C| + |D| \\ & -|A \cap B| - |A \cap C| - |A \cap D| - |B \cap C| - |B \cap D| - |C \cap D| \\ & +|A \cap B \cap C| + |A \cap B \cap D| + |A \cap C \cap D| + |B \cap C \cap D| \\ & -|A \cap B \cap C \cap D| \end{aligned}$$

For at tjekke dette udtryk skal du tage hvert område for sig og bemærke, at det netop bliver regnet med én gang.

Tag f.eks. et element i $A \cap B \cap C \cap D$. Det ligger i alle fire mængder *A*, *B*, *C* og *D* og dermed også i samtlige fællesmængder. Derfor bliver det først talt med 4 gange, så bliver det fratrukket 6 gange, så lægges det til 4 gange og endelig trækkes det fra 1 gang. Altså er det i alt talt med én gang.

Tag som et andet eksempel et element, der ligger i $A \cap B \cap C$, men ikke i *D*. Det bliver først talt med 3 gange, derefter fratrækkes det 3 gange (da det ligger i $A \cap B, A \cap C$ og $B \cap C$, men ikke i fællesmængder, hvor *D* indgår), så lægges det til én gang og endelig fratrækkes det IKKE igen, da det ikke tilhører $A \cap B \cap C \cap D$. Igen er det altså talt med netop én gang.

Øvelse 4:

a) Tjek, at en størrelse tilhørende *C*, men ikke *A*, *B* eller *D*, bliver talt med netop én gang.

b) Tjek, at en størrelse, der tilhører $B \cap D$, men hverken *A* eller *C*, bliver talt med netop én gang.

Princippet med skiftevis inklusion og eksklusion kan udvides til vilkårligt mange mængder.

Eksempel 38: Vi ønsker at besvare et spørgsmål inden for talteori og arbejder altså med hele tal:
Hvor mange tal i mængden $M = \{1,2,3,...,1000\}$ *er ikke divisible med nogen af tallene 3, 7 og 10?*
Vi indfører nu følgende mængder:

M_3 er mængden bestående af de tal i *M*, der er divisible med 3.

M_7 er mængden bestående af de tal i *M*, der er divisible med 7.

M_{10} er mængden bestående af de tal i *M*, der er divisible med 10.

Hermed er $M_3 \cup M_7 \cup M_{10}$ mængden bestående af de tal i *M*, der er divisible med mindst ét af tallene 3, 7 og 10, og vores spørgsmål bliver altså besvaret, når vi har fundet antallet af elementer i komplementærmængden $M_{svar} = M \setminus (M_3 \cup M_7 \cup M_{10})$, hvilket gøres ved:

$$|M_{svar}| = |M| - |M_3 \cup M_7 \cup M_{10}|$$

Princippet om inklusion og eksklusion fortæller os:

$$|M_3 \cup M_7 \cup M_{10}| = |M_3| + |M_7| + |M_{10}| - |M_3 \cap M_7| - |M_7 \cap M_{10}| - |M_3 \cap M_{10}| + |M_3 \cap M_7 \cap M_{10}|$$

$M_3 \cap M_7$ består af de tal, som både 3 og 7 er divisorer til, dvs. alle de tal, som $3 \cdot 7 = 21$ er divisor til (da både 3 og 7 skal være med i primfaktoropløsningen). Det er altså tallene i mængden $\{21,42,63,84,...,987\}$. Dem er der 47 af, da $\frac{1000}{21} = 47,6$.

$M_3 \cap M_7 \cap M_{10}$ består af de tal, som både 3, 7 og 10 er divisorer til, dvs. alle de tal som $3 \cdot 7 \cdot 10 = 210$ er divisor til (primfaktoropløsningen skal indeholde tallene 2, 3, 5 og 7). Det er de 4 tal i mængden $\{210,420,630,840\}$. Antallet 4 ses også ved, at $\frac{1000}{210} = 4,76$.

Da $\frac{1000}{3} = 333,3$, er der 333 elementer i M_3. Og på denne måde kan antallet af elementer i de enkelte mængder bestemmes, hvorved man får:

$$|M_3 \cup M_7 \cup M_{10}| = |M_3| + |M_7| + |M_{10}| - |M_3 \cap M_7| - |M_7 \cap M_{10}| - |M_3 \cap M_{10}| + |M_3 \cap M_7 \cap M_{10}| =$$
$$333 + 142 + 100 - 47 - 14 - 33 + 4 = 485$$

Og dermed er: $|M_{svar}| = |M| - |M_3 \cup M_7 \cup M_{10}| = 1000 - 485 = 515$

Opgaverne 300*

Eksempel 39: I en klasse uden tomme stole har de n elever faste pladser. En dag vælger man at placere alle elever helt tilfældigt. På hvor mange måder kan eleverne placeres, så ikke en eneste sidder på sin rigtige plads?
Vi mangler nogle begreber for at kunne sætte tal på dette eksempel. Disse begreber kommer senere, hvilket selvfølgelig vil øge anvendelsesmulighederne for princippet om inklusion og eksklusion.
Vi lader A være mængden bestående af samtlige forskellige måder, man kan placere eleverne på.
Vi nummererer eleverne og lader A_i være mængden af forskellige placeringer, hvor elev nummer i sidder på den rigtige position.
Foreningsmængden $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$ kommer så til at bestå af alle de placeringer, hvor mindst én elev sidder rigtigt. Antallet af elementer i denne mængde kan findes ved hjælp af princippet om inklusion og eksklusion, og man kan så svare på spørgsmålet ved at trække dette antal fra det samlede antal måder at placere eleverne.

Koblingen mellem kombinatorik og sandsynlighedsregning er Sætning 3, da vi i symmetriske sandsynlighedsfelter kan bestemme sandsynligheden for en hændelse ved at tælle antallet af elementer i udfaldsrummet og antallet af elementer i hændelsen.

Da vi desuden har set, hvordan man sommetider kan omforme et ikke-symmetrisk endeligt sandsynlighedsfelt til et symmetrisk sandsynlighedsfelt ved at angive udfaldsrummet på en anden måde (Eksempel 11), vil Sætning 3 sammen med kombinatorik være et stærkt matematisk redskab.

Udtrykt med sandsynligheder lyder Eksempel 39:

Eksempel 39 (med sandsynligheder): Hvad er sandsynligheden P for, at eleverne ved en tilfældig placering bliver sat, så ikke en eneste sidder rigtigt?
Da samtlige placeringer er lige sandsynlige, har man et symmetrisk sandsynlighedsfelt, og vores sandsynligheder kan derfor bestemmes ved at dividere kardinaltallet for den pågældende mængde med kardinaltallet for A.
F.eks. findes sandsynligheden for, at elev nummer 4 kommer til at sidde rigtigt, ved $P_4 = \dfrac{|A_4|}{|A|}$.
Og den søgte sandsynlighed er $P = \dfrac{|A| - |A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n|}{|A|}$.

Opgaverne 301*

Et andet – meget simpelt og meget anvendeligt – princip inden for kombinatorik er …

Skuffeprincippet (Dirichlets skuffeprincip eller dueslagsprincippet)

Dirichlets skuffeprincip siger, at hvis du har flere sokker end skuffer, kan du ikke placere sokkerne i skufferne, uden at mindst én skuffe kommer til at indeholde mere end én sok.
Eller tilsvarende: Hvis du har flere duer end dueslag, kan du ikke placere duerne i dueslagene, så ingen duer kommer til at dele dueslag med en anden.
Eller generelt: Hvis n genstande skal placeres i m beholdere, hvor $n > m$, vil der være mindst én beholder med mere end én genstand.

Eksempel 40: I en klasse vil der altid være mindst 2 personer, der har hilst på det samme antal klassekammerater om morgenen (en hilsen går begge veje).
For hvis der er n personer i klassen, har alle disse mulighed for at hilse på 0 til $n-1$personer, dvs. n muligheder, men hvis en person har hilst på 0 personer (ingen), kan der ikke samtidig være en, der har hilst på $n-1$ personer (alle) – og omvendt. Der er altså $n-1$muligheder for antal hilsner, og da der er n personer i klassen, fortæller skuffeprincippet os, at mindst 2 personer har hilst på det samme antal.

Eksempel 41: I en mængde indeholdende mindst 7 forskellige heltal vil der altid være mindst to af tallene, der giver samme principale rest ved division med 6. For antallet af mulige principale rester er 6 (tallene 0, 1, 2, 3, 4 og 5), og dette antal er mindre end antallet af tal i mængden, hvorfor skuffeprincippet fører til den angivne konklusion.

Eksempel 42: En anden formulering af Eksempel 41 er, at i en mængde indeholdende mindst 7 heltal vil der altid være mindst én differens mellem to tal i mængden, der er divisibel med 6.

Eksempel 43: Vi skal senere behandle det såkaldte *Fødselsdagsparadoks*, der drejer sig om, at sandsynligheden for, at der i et selskab er to personer med samme fødselsdag, er overraskende stor (i forhold til de flestes intuition). Vi kan indtil videre nøjes med at konstatere, at skuffeprincippet fortæller os, at hvis et selskab består af mindst 367 personer, vil der med sikkerhed være mindst to personer med samme fødselsdag.

Opgaverne 302*

CENTRALE BEGREBER: Permutationer og kombinationer

Inden for kombinatorik er det vigtigt at kunne skelne mellem følgende to ord:

Kombinere: kombi'nere ; v
sammensætte, forene forskellige dele til et hele
OPRINDELSE: lat. combinare.
Permutere: permu'tere ; v
ombytte, omstille
OPRINDELSE: af lat. permutare ændre, bytte, udveksle, per- + mutare flytte noget, ændre, bytte.

Vi definerer nu:

Definition 12: Lad $A = \{a_1, a_2, a_3, \ldots, a_n\}$ være en mængde med *n* elementer (en *n-mængde*), og lad *r* være et helt tal, hvorom det gælder, at $0 \le r \le n$.

a) En ***permutation*** af mængden *A* er en opstilling af de *n* elementer i en bestemt rækkefølge (også kaldet en *ordnet* mængde).

b) Antallet af permutationer af *A* skrives P_n.

c) En ***r - permutation*** fra mængden *A* er en ordnet delmængde bestående af *r* (forskellige) elementer fra *A*.

d) Antallet af *r*-permutationer fra mængden *A* skrives $P(n,r)$.

Når man i definitionen tillader $r = n$, kan man også kalde en *permutation* af en *n-mængde* for en *n-permutation*. Og en *0-permutation* er den (ordnede) tomme mængde, som der kun er én af, dvs. $P(n,0) = 1$.

Eksempel 44: Lad $A = \{a,b,c,d,e\}$ være en mængde med 5 elementer.

Fire forskellige permutationer af *A* er så: *adceb* *ecabd* *edcba* *aebdc*
Følgende er **IKKE** permutationer: *adabc* *bceed* *aaaaa*
Forskellige 4-permutationer fra *A* er: *beda* *acbd* *acdb* *edcb*
Forskellige 3-permutationer fra *A* er: *ace* *deb* *abc* *edc*
Forskellige 1-permutationer fra *A* er: *a* *c* *d* *b*

Definition 13: Lad *A* være en *n-mængde*, og lad *r* være et helt tal, hvorom det gælder $0 \le r \le n$:

a) En ***kombination*** fra *A* er en delmængde af *A*.

b) Antallet af kombinationer fra *A* betegnes K_n.

c) En ***r - kombination*** fra *A* er en kombination med *r* elementer fra *A*.

d) Antallet af r-kombinationer fra *A* betegnes $K(n,r)$.

Eksempel 45: Lad $A = \{a,b,c,d,e\}$.

4 forskellige kombinationer fra A er: $\{a,c,d\}$ $\{a,b,c,d,e\}$ $\varnothing$ $\{b,a\}$
Følgende kombinationer er **ens**: $\{a,c,d\}$ $\{c,a,d\}$ $\{d,c,a\}$
Forskellige 2-kombinationer: $\{a,b\}$ $\{a,c\}$ $\{b,c\}$ $\{e,b\}$ $\{c,d\}$

Opgaverne 303*

Vi ønsker at bestemme udtryk for $P_n, K_n, P(n,r)$ og $K(n,r)$, men inden det kan lade sig gøre, skal vi have introduceret nogle principper:

Multiplikationsprincipperne:

Sætning 13 (multiplikationsprincippet for valgmuligheder):
Ved et samlet valg bestående af n ordnede delvalg med antallet af valgmuligheder $v_1, v_2, v_3, \ldots, v_n$, hvor antallet af valgmuligheder v_i i det enkelte delvalg altså er uafhængigt af de foregående delvalg, er det samlede antal valgmuligheder:

$$V_{samlet} = v_1 \cdot v_2 \cdot v_3 \cdot \ldots \cdot v_n \ .$$

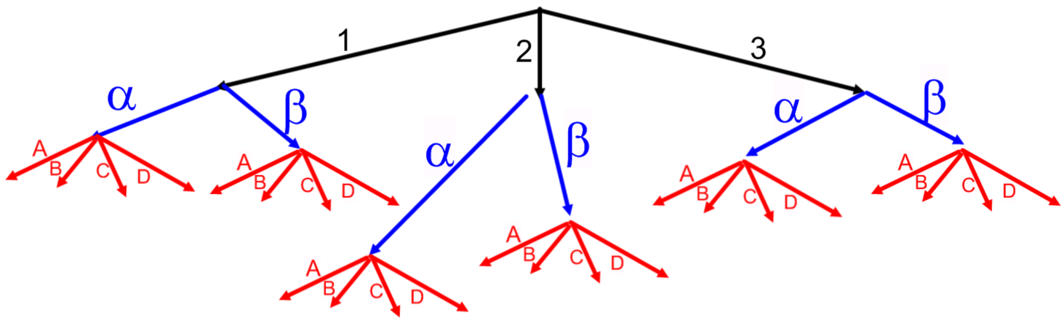
Eller udtrykt med begreber fra sandsynlighedsregning: Udfaldsrummet, der beskriver de forskellige samlede valgmuligheder, vil indeholde V_{samlet} elementer, hvor $V_{samlet} = v_1 \cdot v_2 \cdot v_3 \cdot \ldots \cdot v_n$.

Med *ordnede* delvalg menes, at hvis det samme delvalg træffes flere gange, er placeringen væsentlig. Dvs. hvis man f.eks. kaster en terning 4 gange og dermed træffer fire ens delvalg, skal der i den pågældende situation være forskel på, om man får øjentallene 3,1,6,2 eller 6,1,3,2.
Dermed ville sætningen ikke kunne bruges, hvis man kastede 4 ens terninger én gang, da man i så fald ikke ville kunne skelne 3,1,6,2 og 6,1,3,2 fra hinanden.

Eksempel 46: En knøs har 2 par sko, 5 par strømper, 3 bukser, 6 skjorter og 4 bluser. Han har ingen sans for sammensætning af tøjet, så valget af de enkelte dele er uafhængige af hinanden. På hvor mange måder kan han klæde sig på (når en påklædning består af 1 af hver slags beklædningsdel)?
Der skal træffes 5 uafhængige delvalg med antal valgmuligheder 2, 5, 3, 6 og 4.
Det samlede antal valgmuligheder er derfor $V_{samlet} = 2 \cdot 5 \cdot 3 \cdot 6 \cdot 4 = 720$.
Han har altså 720 forskellige måder at klæde sig på.

Eksempler på tre af de 720 elementer i udfaldsrummet:
$(sko_2\, , strømpepar_4 , buks_1 , skjorte_5\, , bluse_1)$
$(sko_1\, , strømpepar_2 , buks_1 , skjorte_6\, , bluse_3)$
$(sko_2\, , strømpepar_1 , buks_3 , skjorte_1\, , bluse_2)$

Bevis 13: Beviset for Sætning 13 vil typisk være et såkaldt *tælletræ*, hvor vi her ser på et konkret eksempel med tre delvalg med antal valgmuligheder 3, 2 og 4:



$$U = \{1\alpha A, 1\alpha B, 1\alpha C, 1\alpha D, 1\beta A, 1\beta B, 1\beta C, 1\beta D, 2\alpha A, 2\alpha B, 2\alpha C, 2\alpha D, 2\beta A, \ldots, 3\beta D\}$$

Der er $V_{samlet} = 3 \cdot 2 \cdot 4 = 24$ forskellige samlede valgmuligheder svarende til de 24 røde pilespidser. Udfaldsrummet med 24 elementer er angivet.

Opgaverne 304*

Det andet multiplikationsprincip er egentlig bare en variation af det første, hvor man i stedet for valgmuligheder (og dermed naturlige tal) taler om sandsynligheder for permutationer (og dermed om tal mellem 0 og 1):

Sætning 14 (multiplikationsprincippet for sandsynligheden for permutationer af hændelser): Lad $A_1A_2A_3...A_n$ være en permutation af *n* hændelser $A_1, A_2, A_3, ..., A_n$. Sandsynligheden for, at hændelserne alle indtræffer <u>i den angivne rækkefølge</u>, er:

$$P(A_1A_2A_3...A_n) = P(A_1)\cdot P(A_2)\cdot P(A_3)\cdot ...\cdot P(A_n)$$

hvor $P(A_i)$ er sandsynligheden for, at hændelsen A_i indtræffer, når hændelserne $A_1, A_2, A_3, ..., A_{i-1}$ allerede er indtruffet.

Eksempel 47: Man kaster én terning 5 gange og ønsker at finde sandsynligheden for at slå en 6'er alle 5 gange. Da udfaldet af et terningkast ikke afhænger af tidligere kast, har man 5 uafhængige hændelser, hver med sandsynligheden $\frac{1}{6}$ for at indtræffe. Hermed bliver den søgte sandsynlighed:

$$P = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = \frac{1}{6^5} = \frac{1}{7776}$$

Eksempel 48: Man kaster én terning 5 gange og ønsker at finde sandsynligheden for at slå en 2'er i de 3 første kast og en 1'er i de 2 sidste kast. Igen har man 5 uafhængige hændelser med sandsynligheden $\frac{1}{6}$ for at indtræffe, så man får: $P = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = \frac{1}{6^5} = \frac{1}{7776}$.

Bemærk, at dette <u>IKKE</u> er det samme som sandsynligheden for at slå 3 2'ere og 2 1'ere med 5 terninger, da sætningen lægger vægt på rækkefølgen (jævnfør ordet: permutation).

Eksempel 49: Man kaster en mønt tre gange og ønsker at finde sandsynligheden for udfaldet *pkk*. Denne permutation består af tre uafhængige hændelser hver med sandsynligheden 0,5, så man har:

$$P(pkk) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$$

Eksempel 50: Vi omformulerer Eksempel 46 til nu at se på sandsynligheden for at få en konkret permutation, f.eks.: $(sko_2, strømpepar_4, buks_1, skjorte_5, bluse_1)$.

Hvis antallet af valgmuligheder for den enkelte type beklædning er v_i, og hvis hvert valg kan beskrives med et symmetrisk sandsynlighedsfelt (dvs. at ingen beklædningsdel foretrækkes frem for andre af samme slags), vil sandsynligheden for et bestemt valg være $P = \frac{1}{v_i}$. Da vi havde 5 uafhængige delvalg med antal valgmuligheder 2, 5, 3, 6 og 4, giver det altså:

$$P = \frac{1}{2}\cdot\frac{1}{5}\cdot\frac{1}{3}\cdot\frac{1}{6}\cdot\frac{1}{4} = \frac{1}{720}$$

Opgaverne 305*

Ovenstående eksempel kan fungere som en anvisning på, hvordan man kommer fra det første multiplikationsprincip til det andet.

Det er væsentligt at bemærke, at multiplikationsprincipperne er en slags *"både-og"-principper*. De forklarer, hvordan man skal regne på situationer, hvor både en hændelse **og** en anden hændelse **og** en tredje hændelse **og** ... skal indtræffe.
Bemærk, at sandsynligheden bliver mindre, jo flere hændelser der kobles på.

Det har vist sig, at mennesker ikke altid har en intuitiv opfattelse af multiplikationsprincipperne. F.eks. vil en del mennesker - specielt hvis problemet ikke formuleres så direkte - vurdere sandsynligheden for, at person A er bankdirektør og kører i Mercedes, højere end sandsynligheden for, at person A er bankdirektør.

Vi skal nu se på *"enten-eller"*-principperne (additionsprincipperne):

Additionsprincipperne

Inden vi ser på additionsprincipperne, skal vi have defineret et begreb:

Definition 14: To hændelser A og B kaldes ***disjunkte***, hvis de ikke har nogle udfald tilfælles, dvs. hvis $A \cap B = \emptyset$

Eksempel 53: En terning kastes, og man ser på de fire hændelser:

A: Øjentallet er lige.
B: Øjentallet er ulige.
C: Øjentallet er mindst 5.
D: Øjentallet er 1.

A og B er disjunkte hændelser, da øjentallet ikke både kan være lige og ulige. De er desuden komplementære hændelser, da de desuden tilsammen udgør hele udfaldsrummet U.

$A \cap B = \emptyset \quad \Leftrightarrow$ A og B er disjunkte.

$A \cap B = \emptyset \ \wedge \ A \cup B = U \ \Leftrightarrow$ A og B er komplementære.

A og C er **ikke** disjunkte hændelser, da $A \cap C = \{6\}$

C og D er disjunkte hændelser, da $C \cap D = \emptyset$.

Opgaverne 306*

Sætning 15 (Additionsprincippet for valgmuligheder):

Hvis man skal vælge et element i en af de disjunkte mængder $A_1, A_2, A_3, ..., A_n$ indeholdende henholdsvis $v_1, v_2, v_3, ..., v_n$ elementer, er antallet V af valgmuligheder:

$$V = v_1 + v_2 + v_3 + ... + v_n \, .$$

Eksempel 51: En pige har om fredagen lov til at vælge en usund ting. Hun kan vælge mellem 5 slags is, 4 slags chokolade, 3 slags vingummi og 1 flødebolle.

Hendes samlede antal valgmuligheder er $V_{samlet} = 5+4+3+1=13$

Eksempel 52: En dreng skal i gymnasiet og har 4 gymnasier at vælge imellem, der udbyder henholdsvis 3, 1, 2 og 5 relevante studieretninger.

Han har i alt $V_{samlet} = 3+1+2+5=11$ valgmuligheder.

Opgaverne 307*

Sætning 16 (Additionsprincippet for sandsynligheder):

Sandsynligheden P for, at én af de <u>disjunkte</u> hændelser $A_1, A_2, A_3, ..., A_n$ indtræffer, er:

$$P = P(A_1) + P(A_2) + P(A_3) + ... + P(A_n)$$

Eksempel 54: En mønt kastes tre gange. Vi ønsker at finde sandsynligheden for, at vi får netop én krone <u>eller</u> tre krone.

Vi lader hændelse A_1 bestå i at få netop én krone, og hændelse A_2 er at få tre krone.

Vi kender sandsynlighederne fra tidligere ($P(A_1) = \frac{3}{8}$ og $P(A_2) = \frac{1}{8}$), og da det er disjunkte hændelser (man kan jo ikke både få netop én krone og tre krone), får man:

$$P = P(A_1) + P(A_2) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

Opgaverne 308*

Vi er nu næsten fremme ved de centrale sætninger inden for kombinatorik. Vi mangler blot at få defineret et enkelt begreb:

Permutationer

Definition 15: Lad $n \in \mathbb{N}$. Med skrivemåden $n!$, der læses "*n fakultet*", forstås

$$n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$$

Desuden fastsættes det, at $0! = 1$

Eksempel 55: $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$

$1! = 1$

Eksempel 56: I Maple indtastes fakultet meget simpelt, da man kan benytte tastaturets udråbstegn (man kan også finde symbolet under 'common symbols'):

$6! = 720$

$20! = 2432902008176640000$

$0! = 1$

$36! = 371993326789901217467999448150835200000000$

$7.43! = 12136.43843$

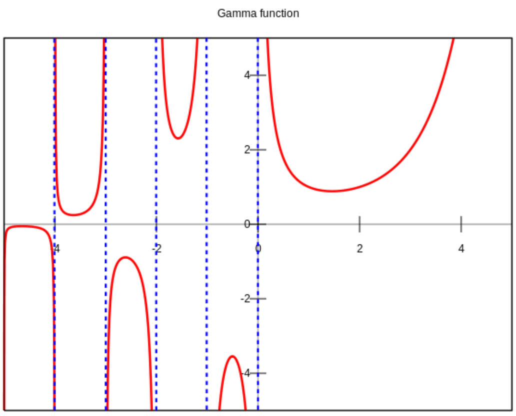
Opgaverne 309*

Som det ses, følger Maple også definitionen $0! = 1$.

Helt generelt taler man om *Det tomme produkt*, hvilket sættes til at give det neutrale element ved multiplikation (dvs. 1).
Man har også *Den tomme sum*, der sættes til det neutrale element ved addition (dvs. 0).

Man kan desuden bemærke, at Maple godt kan udregne $r!$ for et decimaltal, selvom vores definition ikke tillader det. Det skyldes, at fakultet-begrebet kan udvides til den såkaldte *gammafunktion* Γ, der er defineret for alle reelle tal bortset fra negative heltal (se nedenstående):

$(-6)!$Error, numeric exception: division by zero
$(-1.2)! = -5.821148569$
$(-1)!$Error, numeric exception: division by zero
$(-0.5)! = 1.772453851$
$0! = 1$
$0.13! = 0.9399314497$
$1! = 1$
$1.02! = 1.008621087$
$1.9! = 1.827355081$
$2! = 2$
$e! = e! \xrightarrow{\text{at 10 digits}} 4.260820474$
$3! = 6$
$\pi! = \pi! \xrightarrow{\text{at 10 digits}} 7.188082733$
$3.2! = 7.756689536$
$4! = 24$



Bemærk, at $\Gamma(2) = 1$, mens $1! = 1$. Der gælder $\Gamma(n) = (n-1)!$ for naturlige tal *n*.

Sætning 17: Antallet af permutationer af en *n*-mængde er: $P_n = n!$

Eksempel 57: Et nederlandsk fodboldhold stiller op i en 4-3-3 formation, hvor der altså er 11 forskellige pladser at spille. Træneren har udtaget de 11 spillere, der skal spille, men skal nu finde ud af, hvilke pladser de skal spille. Hvor mange forskellige holdopstillinger kan hun vælge?
Vi har en 11-mængde, og en holdopstilling svarer til en permutation af denne, dvs:

$$P_{11} = 11! = 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 39916800$$

Eksempel 58: En sliksulten elev har købt en slikpose med en skumbanan, et skumjordbær, en saltlakrids og en vingummi med citronsmag. Hvor mange forskellige rækkefølger kan eleven vælge at spise slikket i?
Dette er en 4-mængde, og en spiserækkefølge svarer til en permutation, dvs:

$$P_4 = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

Eksempel 59: En anden sliksulten elev har ikke købt nogen slikpose. I hvor mange forskellige rækkefølger kan eleven spise det slik, der ikke er der?
Det er en 0-mængde (den tomme mængde), så der er $0! = 1$ mulig rækkefølge (der svarer til ikke at spise noget).

Opgaverne 310*

Bevis 17: Vi ønsker at bruge multiplikationsprincippet (Sætning 13).
Vi skal træffe *n* delvalg.
Vores første delvalg består i at vælge ét element blandt de *n* elementer i vores mængde og placere dette som første element i vores ordnede mængde.
I en permutation af *n* forskellige elementer må det samme element ikke indgå flere gange, så når vi skal træffe vores næste delvalg, har vi kun $n-1$ muligheder for at vælge det andet element i vores ordnede mængde. Og dette gælder uanset hvilket element, der blev valgt første gang (dvs. antal valgmuligheder i delvalget er uafhængigt af udfaldet af det første delvalg).
Ved tredje delvalg har vi ud fra samme argument $n-2$ valgmuligheder, og således fortsættes, indtil der ved vores sidste delvalg kun er ét element tilbage i den oprindelige mængde, som vi ikke har placeret i den ordnede mængde.
Sætning 13 giver derfor, at det samlede antal valgmuligheder (dvs. antal forskellige permutationer) er $P_n = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot ... \cdot 2 \cdot 1 = n!$

Sætning 18: Antallet af *r*-permutationer fra en *n*-mængde er: $P(n,r) = \dfrac{n!}{(n-r)!}$

Eksempel 60: Den nederlandske træner fra tidligere har nu 14 spillere til rådighed til sin foretrukne 4-3-3 opstilling. Hvor mange forskellige holdopstillinger kan hun vælge?

$$P(14,11) = \frac{14!}{(14-11)!} = \frac{14!}{3!} = 14529715200$$

Eksempel 61: Der er 5 hylder i et skab, og man skal finde en hylde til bukser, bluser og strømper (en til hver). Hvor mange måder kan det gøres på?

$$P(5,3) = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60$$

Eksempel 62: I en klasse på 25 elever skal vælges en elevrådsrepræsentant, en elevrådsrepræsentantsekretær samt en suppleant.

Dette kan gøres på $P(25,3)=\dfrac{25!}{(25-3)!}=\dfrac{25!}{22!}=25\cdot 24\cdot 23=13800$ forskellige måder.

Opgaverne 311*

Bevis 18: Vi har at gøre med en *n*-mængde. Vores permutation skal bestå af *r* elementer fra denne mængde opstillet i rækkefølge.

Igen benytter vi Sætning 13 (multiplikationsprincippet for delvalg). Vi skal træffe *r* delvalg.

Ved det første delvalg har vi *n* valgmuligheder, dvs. der kan stå *n* forskellige elementer på vores permutations første plads.

Vores andet delvalg (permutationens anden plads) foretages blandt $n-1$ muligheder.

Vores tredje delvalg foretages blandt $n-2$ muligheder.

Og vores r'te (det sidste) delvalg foretages blandt $n-r+1$ muligheder.

Vi får derfor følgende (undervejs ændres udtrykket til en brøk, der forlænges):

$$P(n,r)=n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-r+2)\cdot(n-r+1)=$$

$$\frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-r+2)\cdot(n-r+1)\cdot(n-r)\cdot(n-r-1)\cdot\ldots\cdot 2\cdot 1}{(n-r)\cdot(n-r-1)\cdot\ldots\cdot 2\cdot 1}=\frac{n!}{(n-r)!}$$

Det er ikke det helt store arbejde at opskrive brøken $\dfrac{n!}{(n-r)!}$ i Maple, men hvis man vil anvende en kommando til det, skal man hente pakken *combinat*:

$with(combinat):$

$numbperm(5,3)=60$

$\mathrm{P(5,3)}=\dfrac{5!}{(5-3)!}=60$

$permute(5,3)=$

[[1, 2, 3], [1, 2, 4], [1, 2, 5], [1, 3, 2], [1, 3, 4], [1, 3, 5], [1, 4, 2], [1, 4, 3], [1, 4, 5], [1, 5, 2], [1, 5, 3], [1, 5, 4], [2, 1, 3], [2, 1, 4], [2, 1, 5], [2, 3, 1], [2, 3, 4], [2, 3, 5], [2, 4, 1], [2, 4, 3], [2, 4, 5], [2, 5, 1], [2, 5, 3], [2, 5, 4], [3, 1, 2], [3, 1, 4], [3, 1, 5], [3, 2, 1], [3, 2, 4], [3, 2, 5], [3, 4, 1], [3, 4, 2], [3, 4, 5], [3, 5, 1], [3, 5, 2], [3, 5, 4], [4, 1, 2], [4, 1, 3], [4, 1, 5], [4, 2, 1], [4, 2, 3], [4, 2, 5], [4, 3, 1], [4, 3, 2], [4, 3, 5], [4, 5, 1], [4, 5, 2], [4, 5, 3], [5, 1, 2], [5, 1, 3], [5, 1, 4], [5, 2, 1], [5, 2, 3], [5, 2, 4], [5, 3, 1], [5, 3, 2], [5, 3, 4], [5, 4, 1], [5, 4, 2], [5, 4, 3]]

Bemærk, at P(5,3) i ovenstående IKKE er en kommando i Maple. Det er skrevet som tekst, da Maple ikke genkender denne notation. Det er udtrykket med fakultetstegnene, som Maple udregner.

Som det ses, svarer *numbperm(n,r)* til $P(n,r)$, mens kommandoen *permute(n,r)* giver os de konkrete permutationer. Bemærk, at mængden $\{1,2,3\}$ optræder 6 gange, fordi det er permutationer og ikke kombinationer.

Eksempel 63 (Fødselsdagsparadokset): Vi kan benytte Sætning 18 om permutationer til at arbejde med det såkaldte *Fødselsdagsparadoks*, der blev omtalt i Eksempel 43.

Vi antager, at man kan have fødselsdag på 365 forskellige dage (og ser altså bort fra den 29. februar), og at sandsynlighedsfeltet er symmetrisk, hvilket det selvfølgelig ikke helt vil være i praksis, da der ikke vil være præcis lige mange mennesker, der har fødselsdag hver dag.

Fødselsdagsparadokset kunne så lyde: *28 elever begynder i en x-klasse. Hvad er sandsynligheden for, at mindst to af eleverne har fødselsdag den samme dag?*

Vi bruger sætningerne 2, 3, 13 og 18 til at svare på spørgsmålet. Vi begynder med at se på vores hændelse og den komplementære hændelse:

H: Der er mindst to elever, der har fødselsdag samme dag.

$\overline{H}$: Ingen elever har fødselsdag samme dag.

Det er nemmest at finde sandsynligheden for den komplementære hændelse, da man ellers skal beregne sandsynligheden for 2 med samme fødselsdag, 3 med samme fødselsdag, … , 28 med samme fødselsdag. Når man har fundet sandsynligheden for den komplementære hændelse, kan vi bruge Sætning 2 til at bestemme sandsynligheden for vores søgte hændelse.

Vores antagelse om, at sandsynlighedsfeltet er symmetrisk, gør, at vi kan finde sandsynlighederne ved at se på antal gunstige udfald delt med antal mulige udfald (Sætning 3).

Antal mulige udfald:

Hver elev skal placeres på en dag (elevens fødselsdag). Der skal altså træffes et samlet valg bestående af 28 delvalg, hvor hvert delvalg har 365 valgmuligheder (for hver elev har mulighed for at have fødselsdag på en hvilken som helst dag). Sætning 13 giver os så:

$$N_{mulige} = V_{samlet} = 365^{28}$$

Antal gunstige udfald (i hændelsen $\overline{H}$):

Et muligt udfald er netop en 28-permutation fra vores 365-mængde, for vi skal udtage 28 forskellige dage blandt de 365 dage, og disse skal hver især knyttes til et bestemt bogstav. Det bliver altså en ordnet delmængde, hvor f.eks. 47,13,248, … svarer til, at elev *a* har fødselsdag dag 47, elev *b* har fødselsdag dag 13, elev *c* har fødselsdag dag 248, … . Sætning 18 giver så:

$$N_{gunstige} = P(n,r) = \frac{365!}{(365-28)!}$$

Vi har dermed: $P(H) = 1 - P(\overline{H}) = 1 - \dfrac{N_{gunstige}}{N_{mulige}} = 1 - \dfrac{365!}{365^{28}\cdot(365-28)!} \approx 0{,}654 = 65{,}4\%$

Der er altså godt 65% chance for, at der i denne *x*-klasse med 28 elever vil være mindst to elever med fødselsdag samme dag.

Denne chance er væsentlig større, end de fleste umiddelbart ville have troet, og deraf kommer navnet *Fødselsdagsparadokset*.

Da resultatet er overraskende, er det selvfølgelig oplagt at anvende i forbindelse med væddemål.

Opgaverne 312*

Eksempel 64 (Fødselsdagsparadokset igen): Man kan også anvende Sætning 14 i stedet for Sætning 13. Her bliver argumentationen:
Vi ønsker at finde sandsynligheden for, at alle elever har forskellige fødselsdage, hvorefter vi kan anvende Sætning 2 om komplementære hændelser.

Elev *a* skal have placeret sin fødselsdag, og sandsynligheden for at ramme en dag, der ikke allerede er besat, er $\frac{365}{365}=1$.

Så skal elev *b* have placeret sin fødselsdag. Der er nu 364 ledige dage ud af 365 dage, så sandsynligheden for at ramme en ledig dag er $\frac{364}{365}$.

Elev *c* har så sandsynligheden $\frac{363}{365}$ for at ramme en ledig dag.

Og sådan fortsættes, indtil elev *å* har sandsynligheden $\frac{338}{365}$.

Dette fører (selvfølgelig) til samme resultat som i Eksempel 63.

Helt generelt gælder for et selskab på *r* personer: $P(H)=1-\frac{365!}{365^r\cdot(365-r)!}$

I *x*-klasserne er der ca. 80 elever. Formlen giver derfor, at sandsynligheden for, at mindst to elever i *x*-klasserne har fødselsdag samme dag, er 99,991% (regn selv efter!).

Du kan også prøve at finde ud af, hvor mange personer, der skal være i et selskab, før der er mere end 50% chance for, at der er mindst to med samme fødselsdag.

Kombinationer

Sætning 19: Antallet af kombinationer fra en *n*-mængde er: $K_n=2^n$

Eksempel 65: Man har et ubegrænset antal ens brikker til rådighed og ønsker nu at placere et vilkårligt antal af dem på felterne på et skakbræt, således at der på hver felt står enten 0 eller 1 brik.
Skakbrættets felter udgør nu en 64-mængde (hvert felt er et element i mængden), og vi skal udtage de felter, hvor der skal placeres en brik. Dette kan gøres på $K_{64}=2^{64}=18446744073709551616$ måder.

Eksempel 66: En gymnasieklasse med 23 elever bliver inviteret på skitur. På hvor mange forskellige måder kan gruppen af elever, der tager af sted, sammensættes?
Her er der tale om kombinationer og ikke permutationer, da der ikke er nogen ordning af elever, og gruppen - der godt kan være en tom gruppe - kan så sammensættes på:

$$K_{23}=2^{23}=8388608 \text{ måder}$$

Opgaverne 313*

Bevis 19: Vi benytter Sætning 13 (multiplikation af valgmuligheder) på situationen. Vi har en mængde med *n* elementer, og vi kan danne en konkret kombination ved at kigge på et element ad gangen og stille spørgsmålet: "*Skal dette element med i kombinationen eller ej?*". Da der er *n* elementer, skal vi træffe *n* delvalg, der alle har 2 valgmuligheder ("*ja*" eller "*nej*"). Hermed giver sætning 13 os $K_n=2^n$

Sætning 20: Antallet af *r*-kombinationer fra en *n*-mængde er: $K(n,r)=\dfrac{n!}{(n-r)!\cdot r!}$

Eksempel 67: Den nederlandske træner fra tidligere skal nu i første omgang ud af de 14 spillere blot udtage de 11, der skal starte på banen. På hvor mange måder kan det gøres?

$$K(14,11)=\frac{14!}{(14-11)!\cdot 11!}=\frac{14!}{3!\cdot 11!}=364$$

Eksempel 68: Den nederlandske træner har nemmere ved at vælge de 3 spillere, der ikke skal med i startopstillingen. Hvor mange måder kan det gøres på?

$$K(14,3)=\frac{14!}{(14-3)!\cdot 3!}=\frac{14!}{11!\cdot 3!}=364$$

Eksempel 69: Hvor mange forskellige lotto-kuponer findes der (der er 36 tal, hvoraf 7 skal vælges)?

$$K(36,7)=\frac{36!}{(36-7)!\cdot 7!}=\frac{36!}{29!\cdot 7!}=8347680$$

Eksempel 70: Det er svært nok at vinde i lotto, men nu indføres et nyt spil, hvor man blandt de 36 tal ikke blot skal udvælge 7 tal, men hele 29. Hvor mange forskellige kuponer findes der i denne nye slags lotto?

$$K(36,29)=\frac{36!}{(36-29)!\cdot 29!}=\frac{36!}{7!\cdot 29!}=8347680$$

Opgaverne 314*

Eksemplerne 67 og 68 samt 69 og 70 illustrerer en vigtig pointe, man også kan se ved at betragte selve formlen:

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

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

Der er lige mange $(n-r)$-kombinationer og *r*-kombinationer fra en *n*-mængde, eller med andre ord: **Det er lige meget, om du vælger de elementer, der skal med i din delmængde, eller dem, der ikke skal med.**

Vi skal snart se på de såkaldte binomialfordelinger, og i den forbindelse kommer $K(n,r)$ til at spille en rolle. Man bruger i den forbindelse ofte en anden skrivemåde:

Definition 16: $K(n,r)$ kaldes en *binomialkoefficient* og skrives $K(n,r)=\dbinom{n}{r}$, der udtales *binomialkoefficienten n over r*.

Det er denne notation, der anvendes i Maple (hvis man ikke vil bruge kommandoen 'binomial'):

$$K(10,4)=\frac{10!}{(10-4)!\cdot 4!}=210$$

$$\binom{10}{4}=210$$

$$\text{binomial}(10,4)=210$$

Bemærk, at K(10,4) ikke er en kommando i Maple. Det er udtrykket med fakultetstegnene, som Maple udregner.

Opgaverne 315*

Bevis 20: I beviset benytter vi ud over den ofte benyttede Sætning 13 også sætningerne 17 og 18. Vi skal finde antallet af r-kombinationer fra en n-mængde, men vælger at anskue situationen fra en anden synsvinkel.

Vi vil igen forsøge at finde antallet af r-permutationer fra en n-mængde $P(n,r)$ (et resultat vi allerede kender fra Sætning 18). I stedet for at skabe de enkelte permutationer med det samme forestiller vi os nu, at vi gør følgende:

1) Vi udtager først de r elementer, der skal indgå i permutationen.
2) Derefter permuterer vi de r udtagne elementer.

Punkt 1) svarer netop til at udtage r-kombinationer, mens punkt 2) er behandlet i sætning 17.

Vores "valg" med $P(n,r)$ valgmuligheder består altså af to delvalg, hvor det første har $K(n,r)$ valgmuligheder og det andet P_r muligheder.

Dermed giver sætning 13: $P(n,r) = K(n,r)\cdot P_r \Leftrightarrow K(n,r) = \dfrac{P(n,r)}{P_r} = \dfrac{n!}{(n-r)!\cdot r!}$

Bevis 21: Da vi skulle bevise Sætning 19 ($K_n = 2^n$), benyttede vi en tankegang, hvor vi tog et element ad gangen og så på de to valgmuligheder: Enten skal elementet med i kombinationen eller elementet skal ikke med.

Vi kunne også have anvendt en anden indfaldsvinkel:

Vi skal finde det samlede antal kombinationer fra en n-mængde, og denne gang tæller vi først 0-kombinationerne, så 1-kombinationerne, så 2-kombinationerne, osv. indtil n-kombinationerne. Sætning 15 om addition af valgmuligheder giver så:

$$K_n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n}$$

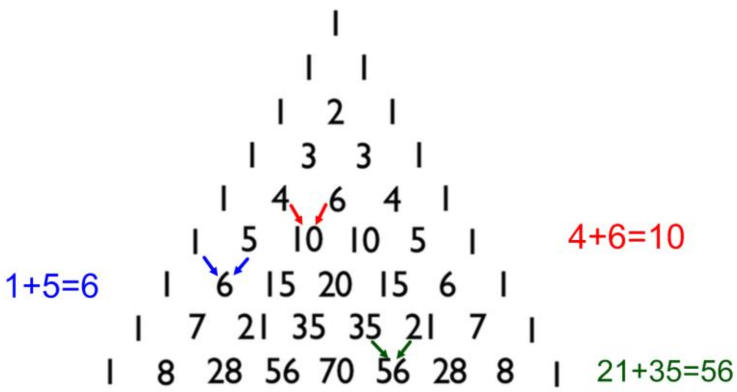
Sammen med Sætning 19 giver dette resultat os altså følgende sætning:

Sætning 21: $2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n}$

Opgaverne 316*

Pascals Trekant (Den Aritmetiske Trekant)

Binomialkoefficienterne optræder i *Den Aritmetiske Trekant*, der er bedre kendt som *Pascals Trekant*.



Tallene er fremkommet ved, at man danner trekantens form med 1-tallerne, hvorefter de andre tal dannes én række ad gangen ved at tage summen af de to nærmeste tal i ovenstående række.

Bemærk, at $\binom{n}{n} = \dfrac{n!}{(n-n)!\cdot n!} = \dfrac{n!}{0!\cdot n!} = \dfrac{n!}{n!} = 1$ og $\binom{n}{0} = \dfrac{n!}{(n-0)!\cdot 0!} = \dfrac{n!}{n!\cdot 0!} = \dfrac{n!}{n!} = 1$.

Pascal trekant opskrevet med binomialkoefficienter ser ud på følgende måde:

$$\binom{0}{0}$$
$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$
$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$
$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$
$$\vdots$$
$$\binom{n-1}{0} \quad \binom{n-1}{1} \quad \binom{n-1}{2} \quad \cdots \quad \binom{n-1}{n-3} \quad \binom{n-1}{n-2} \quad \binom{n-1}{n-1}$$
$$\binom{n}{0} \quad \binom{n}{1} \quad \binom{n}{2} \quad \binom{n}{3} \quad \binom{n}{4} \quad \cdots \quad \binom{n}{n-2} \quad \binom{n}{n-1} \quad \binom{n}{n}$$

Vi har allerede argumenteret for 1-tallerne i trekanten, men mangler at argumentere for alle de andre tal. Dette gøres med sætningen:

Sætning 22: For $n \geq 2$ og $1 \leq k < n$ gælder: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

Bevis 22: Vi benytter sætning 20 og laver en række omskrivninger:

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \frac{(n-1)!}{((n-1)-(k-1))! \cdot (k-1)!} + \frac{(n-1)!}{(n-1-k)! \cdot k!} =$$

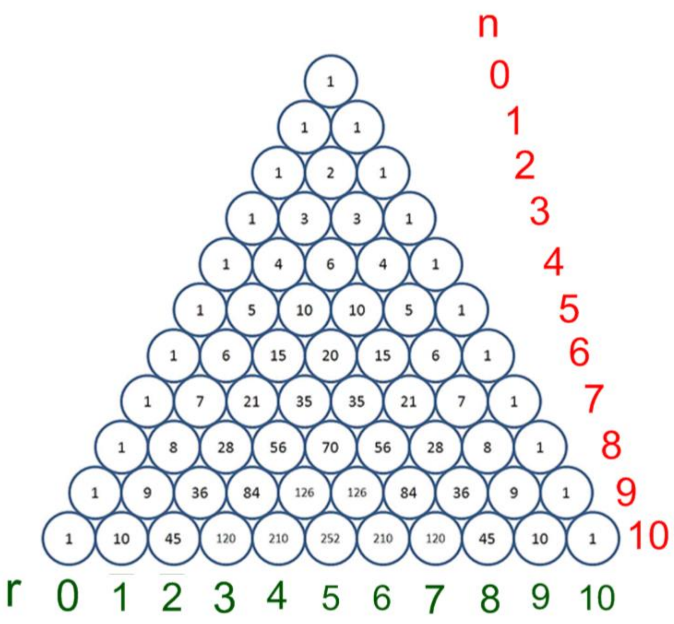
$$\frac{(n-1)!}{(n-k)! \cdot (k-1)!} + \frac{(n-1)!}{(n-1-k)! \cdot k!} = \frac{(n-1)! \cdot k}{(n-k)! \cdot k \cdot (k-1)!} + \frac{(n-k) \cdot (n-1)!}{(n-k) \cdot (n-1-k)! \cdot k!} =$$

$$\frac{(n-1)! \cdot k}{(n-k)! \cdot k!} + \frac{(n-k) \cdot (n-1)!}{(n-k)! \cdot k!} = \frac{(n-1)! \cdot k + (n-k) \cdot (n-1)!}{(n-k)! \cdot k!} =$$

$$\frac{(n-1)! \cdot (k+n-k)}{(n-k)! \cdot k!} = \frac{n \cdot (n-1)!}{(n-k)! \cdot k!} = \frac{n!}{(n-k)! \cdot k!} = \binom{n}{k}$$

Bevis 22 (alternativt): Man kan også bevise Sætning 22 ved en sproglig argumentation.
Vi ser på en *n*-mængde *A* og ønsker at finde antallet af *k*-kombinationer fra denne mængde. Vi fjerner nu et eller andet element α fra *A* og får dermed *n-1*-mængden *B*. Hvis vi ser på alle de *k*-kombinationer fra *A*, der IKKE indeholder elementet α, består de af *k* elementer fra *B*, og der findes derfor $\binom{n-1}{k}$ af denne slags kombinationer. Alle de *k*-kombinationer fra *A*, der indeholder α, indeholder $k-1$ elementer fra *B*, og der findes derfor $\binom{n-1}{k-1}$ af disse. Da en *k*-kombination fra *A* **enten** indeholder α **eller** ikke indeholder α, har man dermed Sætning 22 (de to antal adderes, da det er "enten-eller").

Man kan altså anvende Pascals trekant til at aflæse binomialkoefficienter:



Vi kan f.eks. aflæse: $\binom{8}{3} = 56$ og $\binom{10}{5} = 252$

Opgaverne 317*

Eksempel 71: Pascals trekant angiver også koefficienterne foran de enkelte led, hvis man opskriver $(x+y)^n$ med leddene i rækkefølge med faldende eksponent på potensen x^r. F.eks:

$$(x+y)^7 \overset{\text{expand}}{=} x^7 + 7\,x^6\,y + 21\,x^5\,y^2 + 35\,x^4\,y^3 + 35\,x^3\,y^4 + 21\,x^2\,y^5 + 7\,x\,y^6 + y^7$$

Bemærk, at koefficienterne 1, 7, 21, 35, 35, 21, 7 og 1 netop er tallene i rækken for $n = 7$ i Pascals trekant. Dette kan indses ved at tænke på $(x+y)^7$ som et produkt med 7 faktorer nummeret 1, 2, 3, 4, 5, 6, 7. Koefficienten foran f.eks. $x^5 y^2$ fremkommer på følgende måde:

Hvert led i hver faktor skal ganges sammen. $x^5 y^2$ fremkommer ved, at man i 2 af faktorerne har valgt leddet *y*, mens man i de fem andre faktorer har valgt leddet *x*. Og man kan netop vælge disse 2 faktorer på $\binom{7}{2} = 21$ forskellige måder.

Vi har hermed argumenteret for følgende sætning:

Sætning 23 (Binomialformlen): $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} \cdot x^{n-i} \cdot y^i$

Eksempel 72: Ud fra Pascals Trekant kan vi altså ifølge Sætning 23 opskrive f.eks.:

$$(x+y)^9 = x^9 + 9x^8 y + 36x^7 y^2 + 84x^6 y^3 + 126x^5 y^4 + 126x^4 y^5 + 84x^3 y^6 + 36x^2 y^7 + 9xy^8 + y^9$$

Opgaverne 318*

Da vi har en computer til rådighed, har vi ikke problemer med at regne med store mængder, men tidligere var næste sætning – *Chu-Vandermondes identitet* – en vigtig sætning, da den kunne benyttes til at beregne store binomialkoefficienter ud fra mindre:

Sætning 24: For naturlige tal *n*, *m* og *r*, hvor $r \le n \wedge r \le m$, gælder:

$$\binom{n+m}{r} = \binom{n}{r} \cdot \binom{m}{0} + \binom{n}{r-1} \cdot \binom{m}{1} + \binom{n}{r-2} \cdot \binom{m}{2} + \ldots + \binom{n}{1} \cdot \binom{m}{r-1} + \binom{n}{0} \cdot \binom{m}{r}$$

Hvilket også kan skrives: $\binom{n+m}{r} = \sum_{i=0}^{r} \binom{n}{r-i} \cdot \binom{m}{i}$

Bevis 24: I en gymnasieklasse er der *n* drenge og *m* piger. Man ønsker at udtage et udvalg bestående af *r* elever, hvor $r \le n \wedge r \le m$, dvs. udvalget må hverken indeholde flere medlemmer, end der er drenge eller piger.

Vi har $n+m$ elever i klassen, så udvalget på *r* elever kan udtages på $\binom{n+m}{r}$ måder.

Men vi kunne også betragte det på følgende måde:
Først ser vi på de mulige udvalg, hvor der er *r* drenge og *0* piger.
Dette er et samlet valg opdelt i de to delvalg, hvor det første består i at udtage *r* elever blandt de *n* drenge, hvilket kan gøres på $\binom{n}{r}$ måder, og det andet består i at udtage *0* elever blandt de *m* piger, hvilket kan gøres på $\binom{m}{0}$ måder. Sætning 13 giver så, at dette samlede valg kan foretages på $\binom{n}{r} \cdot \binom{m}{0}$ forskellige måder.

Derefter ses på de mulige udvalg, hvor der er $r-1$ drenge og 1 pige. Det er igen to delvalg, når vi først udtager $r-1$ medlemmer blandt drengene og 1 medlem blandt pigerne, og det samlede antal muligheder er $\binom{n}{r-1} \cdot \binom{m}{1}$.

Således fortsættes, indtil vi til sidst ser på udvalget bestående af udelukkende piger, og når alle disse muligheder lægges sammen (Sætning 15), fremkommer Sætning 24.

Eksempel 73: Bemærk, at der undervejs i Bevis 24 blev benyttet en tankegang, der kan være nyttig i nogle typer opgaver (se evt. afsnittet om den hypergeometriske fordeling).
Man har en krukke med 30 kugler. 12 blå, 10 røde og 8 grønne.
Hvad er sandsynligheden for at trække 3 blå, 2 røde og 1 grøn, hvis man trækker 6 kugler fra krukken (uden tilbagelægning)?
Opgaven drejer sig om sandsynligheder, men vi begynder med at se på valgmuligheder. Vi skal træffe 3 delvalg (blå, rød, grøn), og vores antal gunstige valgmuligheder er derfor:

$$V_{gunstige} = \binom{12}{3} \cdot \binom{10}{2} \cdot \binom{8}{1} = 79200.$$

De mulige valgmuligheder er 6 kugler blandt 30: $\binom{30}{6} = 593775$

Dvs. sandsynligheden er ifølge Sætning 3: $P = \dfrac{79200}{593775} = 0{,}1333838575$

Opgaverne 319*

BINOMIALFORDELING

Vi har tidligere set generelt på begrebet *fordelingsfunktion*. Vi skal nu se på en konkret fordeling med store anvendelsesmuligheder.

Definition 17: Et *binomialeksperiment* (eller *bernoulli-eksperiment*) er et stokastisk eksperiment, der består af en række **uafhængige** og **identiske** deleksperimenter, der kun har **2 mulige udfald**:
'Succes' og 'fiasko'.

Eksempel 74: En terning kastes 1000 gange, og man er interesseret i at slå 6'ere. Her er 'succes' det at slå en 6'er, mens 'fiasko' er ikke at slå en 6'er.

Eksempel 75: 5000 mennesker undersøges for kræft. Her er 'succes' det at få konstateret kræft, mens 'fiasko' er ikke at få konstateret kræft.

Eksempel 76: En undersøgelse blandt 1500 danskere af vælgertilslutningen til de politiske partier indeholder muligheden for at vælge ét blandt 10 partier eller vælge 'ved ikke'. Dette ligner ikke umiddelbart et binomialeksperiment, da der er 11 mulige udfald, men man kan opdele eksperimentet i 10 binomialeksperimenter ved f.eks. at sige:
"*Stemte personen på SF eller stemte personen ikke på SF*" osv.

Opgaverne 320*

Det er en helt central del af et binomialeksperiment, at de enkelte deleksperimenter skal være **uafhængige** og **identiske/ens**.

Man kan derfor IKKE anvende binomialfordelingen, hvis man f.eks. står med en krukke med blå og røde kugler og trækker én kugle ad gangen og noterer farven uden at lægge kuglen tilbage i krukken, da de enkelte deleksperimenter ikke er identiske (der er ikke det samme antal kugler i krukken).

Strengt taget kan man heller ikke anvende binomialfordelingen, hvis man laver en stikprøveundersøgelse bl.a. de danske vælgere (Eksempel 76), da man aldrig udspørger den samme person mere end én gang (dvs. man arbejder "uden tilbagelægning"). Når man alligevel anvender binomialfordelingen i denne situation, skyldes det, at det skal anvendes til statistik, og at fejlen er ubetydeligt lille, så længe stikprøvens størrelse er meget mindre end populationens størrelse.

Sætning 25 (binomialfordelingen): Ved et binomialeksperiment bestående af *n* deleksperimenter med succes-sandsynligheden *p*, er sandsynligheden $P(r)$ for succes i *r* af eksperimenterne:

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

Definition 18: Et binomialeksperiment med *n* deleksperimenter og successandsynligheden *p* betegnes $b(n,p)$.

Eksempel 77: Hvad er sandsynligheden for at slå netop tre 6'ere ved et kast med 5 almindelige terninger?

$$P(3) = K(5,3) \cdot \left(\frac{1}{6}\right)^3 \cdot \left(1-\frac{1}{6}\right)^2 = \frac{125}{3888} \approx 3{,}2\%$$

Eksempel 78: Hvis 14,7% af danskerne stemmer på DF, hvad er så sandsynligheden for, at der blandt 1500 tilfældigt udvalgte danskere er netop 220, der stemmer på DF?

$$P(220) = K(1500,220) \cdot 0{,}147^{220} \cdot (1-0{,}147)^{1280} = 0{,}02908586 \approx 2{,}9\%$$

Opgaverne 321*

Bevis 25: I beviset anvendes Definition 13 (knyttet til Sætning 20) og sætningerne 14 og 16 om multiplikation og addition af sandsynligheder.

Vi ser på n delforsøg med successandsynligheden p.

Så er fiaskosandsynligheden $1-p$, da vi kun har disse to mulige udfald, og da summen af sandsynligheder for alle udfald skal være 1.

Vi opstiller nu disse n delforsøg i rækkefølge og ser på en konkret situation med r succeser og dermed $n-r$ fiaskoer:

Sandsynligheden for netop dette samlede udfald $\longrightarrow$ $P_{samlet} = p\cdot p\cdot(1-p)\cdot(1-p)\cdot p\cdot(1-p)\cdot p\cdot \ldots \cdot p =$

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

Sætning 14 giver os sandsynligheden for netop dette samlede udfald (dvs. succes og fiasko i lige netop denne rækkefølge).

Men r succeser kan fremkomme på forskellige måder (en anden rækkefølge af succes og fiasko). Antallet af måder er $K(n,r)$, da man kan forestille sig, at man skal udtage r-mængder bestående af numrene på de delforsøg, der giver succes.

Sandsynligheden for hver af de $K(n,r)$ disjunkte hændelser er $p^r\cdot(1-p)^{n-r}$, så sandsynligheden for, at en af disse disjunkte hændelser indtræffer, er:

$$P_{r\,succes} = \underbrace{p^r\cdot(1-p)^{n-r} + p^r\cdot(1-p)^{n-r} + p^r\cdot(1-p)^{n-r} + \ldots + p^r\cdot(1-p)^{n-r}}_{K(n,r)\,led} = K(n,r)\cdot p^r\cdot(1-p)^{n-r}$$

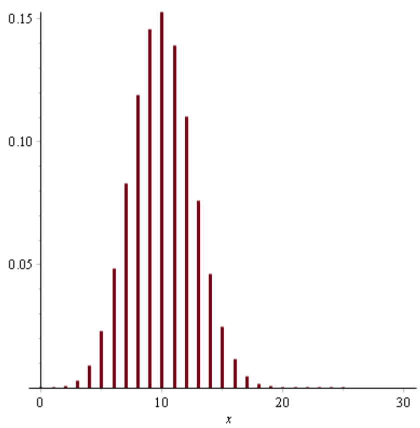
Med Gym-pakken til Maple kan man få tegnet både pindediagrammer og fordelingsfunktioner (trappediagrammer) for binomialfordelingen. Her er det vigtigt at bemærke, at de to størrelser, der karakteriserer binomialfordelinger, er antallet af delforsøg n og successandsynligheden p.

Antallet af succeser fungerer som vores uafhængige variabel og er altså ikke med til at karakterisere den pågældende fordeling.

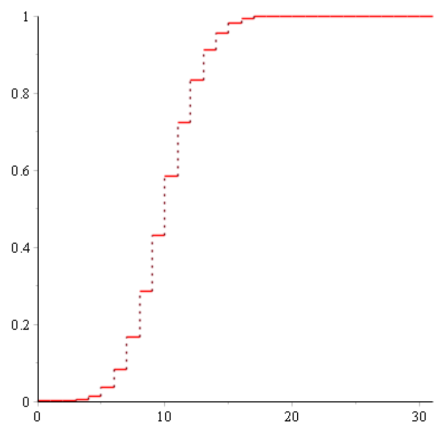
Eksempel 79: Vi vil gerne se et pindediagram og et trappediagram for $b\left(30,\frac{1}{3}\right)$ i Maple:

$with(Gym):$

$pindediagramBIN\left(30,\ \frac{1}{3}\right)$



$trappekurveBIN\left(30,\ \frac{1}{3}\right)$



Eksempel 80: Gym-pakken kan også beregne konkrete sandsynligheder eller sandsynligheder for at få mindst et bestemt antal succeser (såkaldte kumulerede sandsynligheder).
Vi vil gerne udregne sandsynlighederne for at få henholdsvis 10, 6 og 23 succeser:

$$binpdf\left(30, \frac{1}{3}, 10\right) = 0.1530152432$$

$$binpdf\left(30, \frac{1}{3}, 6\right) = 0.04838426662$$

$$binpdf\left(30, \frac{1}{3}, 23\right) = 0.000001265631974$$

Disse værdier kan sammenlignes med pindediagrammet i Eksempel 79.

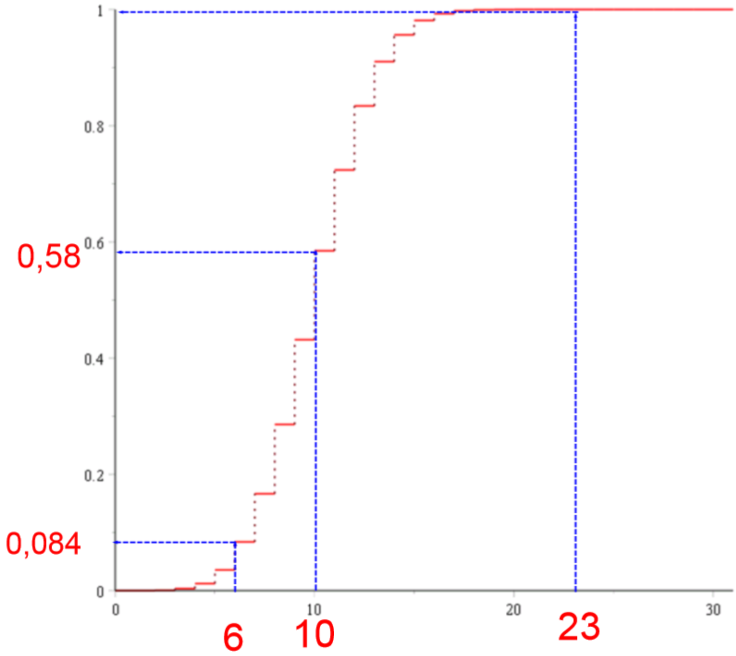
Vi ser nu på sandsynlighederne for at få **højst** ovenstående antal succeser:

$$bincdf\left(30, \frac{1}{3}, 10\right) = 0.5847595988$$

$$bincdf\left(30, \frac{1}{3}, 6\right) = 0.08383843844$$

$$bincdf\left(30, \frac{1}{3}, 23\right) = 0.9999997910$$

Disse værdier kan sammenlignes med trappediagrammet i Eksempel 79.



Hvis vi vil have sandsynligheden for at få **mindst** et bestemt antal succeser, kan vi udnytte Sætning 2 om komplementære hændelser. Så sandsynligheden for at få **mindst** 11 succeser er:

$$P(X \geq 11) = 1 - P(X \leq 10) = 1 - bincdf\left(30, \frac{1}{3}, 10\right) = 0.4152404012$$

Det bemærkes, at $0,5847595988 + 0,4152404012 = 1$

Opgaverne 322*

For binomialfordelinger gælder følgende vigtige sætning, hvoraf første del gerne skulle virke intuitivt rigtig:

Sætning 26: Binomialfordelingen $b(n, p)$ har:

$$\mu = n \cdot p \qquad \text{og} \qquad \sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

Bevis 26: Første del af sætningen bevises ved at opdele binomialeksperimentet i de n identiske deleksperimenter med successandsynligheden p og så udnytte Sætning 9b.

De enkelte deleksperimenter beskrives ved den stokastiske variabel X_i, der angiver antallet af succes ved én udførelse af eksperimentet, dvs. X_i kan antage værdierne 0 og 1 ($x_1 = 0$ og $x_2 = 1$). Det samlede binomialeksperiments stokastiske variabel X bliver så $X = X_1 + X_2 + X_3 + \ldots + X_n$, da man netop skal lægge antallet af succeser i de enkelte deleksperimenter sammen for at få det samlede antal succeser.

Hvert deleksperiment består i at udføre et forsøg, der har sandsynligheden p for succes og sandsynligheden $1-p$ for fiasko. Middelværdien for antal succeser er derfor:

$$E(X_1) = \sum_{j=1}^{2} x_j \cdot P(X_1 = x_j) = 0 \cdot (1-p) + 1 \cdot p = p$$

Alle deleksperimenterne er identiske, og Sætning 9b giver så:

$$\mu = \sum_{i=1}^{n} E(X_i) = \underbrace{p + p + p + \ldots + p}_{n\ led} = n \cdot p$$

Til at bestemme variansen for binomialfordelingen udnytter vi, at de enkelte deleksperimenter er **uafhængige**, dvs. vi kan benytte Sætning 12. Men først skal vi have bestemt variansen af de enkelte deleksperimenter (der som allerede vist har middelværdien p). Dette kunne gøres ved hjælp af Sætning 9d, men her udregnes det ud fra selve definitionen på begrebet varians (Definition 7):

$$Var(X_1) = \sum_{i=1}^{2} (x_i - \mu)^2 \cdot P(X_1 = x_i) =$$

$$(0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p \cdot (1-p) \cdot (p + (1-p)) = p \cdot (1-p)$$

Sætning 12 giver så første lighedstegn i nedenstående:

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = \underbrace{p \cdot (1-p) + p \cdot (1-p) + p \cdot (1-p) + \ldots + p \cdot (1-p)}_{n\ led} = n \cdot p \cdot (1-p).$$

Dermed er:

$$\sigma = \sqrt{Var(X)} = \sqrt{n \cdot p \cdot (1-p)}$$

Inden vi ser et eksempel på anvendelsen af denne sætning, skal vi have indført et sidste begreb og behandlet det i forbindelse med binomialfordelingen.

Definition 19: I følgende definition anvendes entalsbetegnelser. Hvis der er flere størrelser, der opfylder betingelserne, er der flere typetal eller typeintervaller.

a) Inden for statistik er *typetallet* observationen med den største hyppighed.

b) Inden for statistik er *typeintervallet* observationsintervallet med den største tæthed.

c) Inden for sandsynlighedsregning er *typetallet* den værdi af den stokastiske variabel, der har størst sandsynlighed.

Sætning 27: I en binomialfordeling $b(n, p)$ gælder:

a) Hvis $n \cdot p$ giver et helt tal, er $n \cdot p$ typetallet.

b) Hvis $n \cdot p + p$ giver et helt tal, er der to typetal, nemlig $n \cdot p + p$ og $n \cdot p + p - 1$.

c) Hvis ingen af betingelserne a) eller b) er opfyldt, er typetallet det største hele tal, der er mindre end $n \cdot p + p$.

Bemærkning: Da p ligger mellem 0 og 1, følger a) af c), så på sin vis er a) en overflødig del af sætningen. Men den er simplere formuleret end c) og medtages derfor alligevel.

Bevis 27: Vi vil bevise sætningen ved at se på forholdene mellem sandsynlighederne for to successive antal succeser (dvs. to antal, der følger lige efter hinanden, f.eks. 8 og 9).

Pointen er, at så længe forholdet $\frac{P(r)}{P(r-1)}$ er større end 1, vil sandsynligheden være større for *r* succeser end $r-1$ succeser. Når brøken giver under 1, vil sandsynligheden være større for $r-1$ succeser end *r* succeser. Hvis brøken giver 1, er de to sandsynligheder lige store:

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

Så hvis vi ser på, hvornår sandsynligheden for *r* succeser er større end for $r-1$succeser:

$$\frac{(n-r+1)\cdot p}{r\cdot(1-p)} > 1 \Leftrightarrow (n-r+1)\cdot p > r\cdot(1-p) \Leftrightarrow n\cdot p - r\cdot p + p > r - r\cdot p \Leftrightarrow n\cdot p + p > r$$

Tilsvarende udregning kan foretages med lighedstegn og med ulighedstegnet pegende den anden vej, og man har altså:

Hvis $n\cdot p + p > r$ er sandsynligheden for *r* succeser større end for $r-1$succeser.
Hvis $n\cdot p + p = r$ er sandsynligheden for *r* succeser den samme som for $r-1$succeser.
Hvis $n\cdot p + p < r$ er sandsynligheden for *r* succeser mindre end for $r-1$succeser.

Ud fra disse tre sammenhænge følger Sætning 27 (overvej selv hvorfor!).

Eksempel 81: Vi ser på binomialfordelingen $b\left(50, \frac{1}{10}\right)$.

Middelværdien er: $\mu = n\cdot p = 50\cdot\frac{1}{10} = 5$

Spredningen er: $\sigma = \sqrt{n\cdot p\cdot(1-p)} = \sqrt{50\cdot\frac{1}{10}\cdot\left(1-\frac{1}{10}\right)} = \sqrt{5\cdot 0{,}9} = \sqrt{4{,}5} = 2{,}12132$

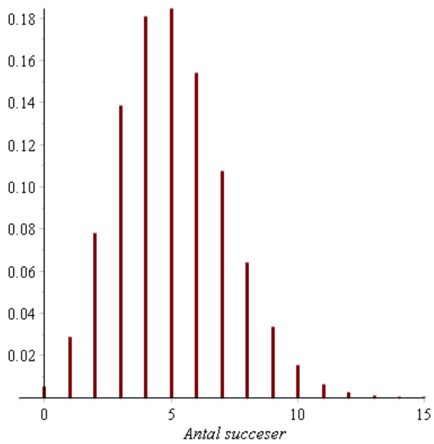
Da $n\cdot p = 5$ er et helt tal, er typetallet 5, dvs. det mest sandsynlige antal succeser er 5.
For at anskueliggøre dette ses på sandsynlighederne for 4, 5 og 6 succeser:

$binpdf(50, 0.1, 4) = 0.1809045009$
$binpdf(50, 0.1, 5) = 0.1849246009$
$binpdf(50, 0.1, 6) = 0.1541038341$

Det bemærkes, at sandsynligheden for 5 succeser er større end for 4 og 6.

Et pindediagram over binomialfordelingen tegnes:

$pindediagramBIN\left(50, \frac{1}{10}\right) =$



Dette er et eksempel, hvor middelværdien og typetallet er ens. Bemærk desuden, at spredningen er 2,1 og sammenlign dette tal med pindediagrammet (gå 2,1 ud til begge sider fra middelværdien 5), så du får en fornemmelse af begrebet spredning.

Eksempel 82: Vi ser på binomialfordelingen $b\left(65, \frac{5}{6}\right)$.

Middelværdien er $\mu = n \cdot p = 65 \cdot \frac{5}{6} = \frac{325}{6} = 54{,}1667$

Spredningen er $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{65 \cdot \frac{5}{6} \cdot \left(1 - \frac{5}{6}\right)} = 3{,}0046$

Da $n \cdot p + p = 65 \cdot \frac{5}{6} + \frac{5}{6} = 66 \cdot \frac{5}{6} = 11 \cdot 5 = 55$ er et helt tal, er der to typetal, nemlig 55 og 54.

For at anskueliggøre dette ses på sandsynlighederne for 53, 54, 55 og 56 succeser:

$binpdf\left(65, \frac{5}{6}, 53\right) = 0.1176651853$

$binpdf\left(65, \frac{5}{6}, 54\right) = 0.1307390947$

$binpdf\left(65, \frac{5}{6}, 55\right) = 0.1307390947$

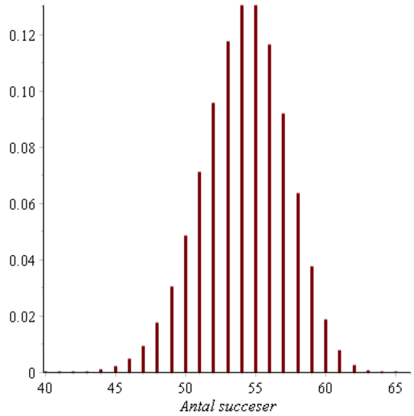
$binpdf\left(65, \frac{5}{6}, 56\right) = 0.1167313346$

Bemærk, at sandsynlighederne for 54 og 55 succeser er (præcis) lige store.

Bemærk også, at dette ikke gælder for 53 og 56 succeser.

Pindediagrammet er:

$pindediagramBIN\left(65, \frac{5}{6}\right) =$



Eksempel 83: Vi ser på binomialfordelingen $b(100; 0{,}175)$.

Middelværdien er $\mu = n \cdot p = 100 \cdot 0{,}175 = 17{,}5$

Spredningen er $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{100 \cdot 0{,}175 \cdot 0{,}825} = 3{,}79967$

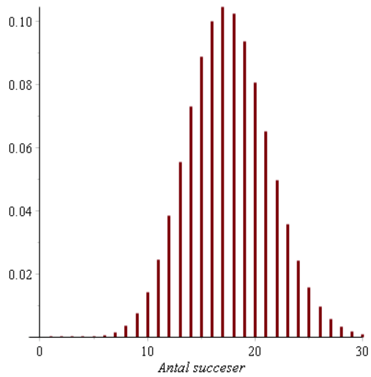
Da hverken $n \cdot p = 17{,}5$ eller $n \cdot p + p = 100 \cdot 0{,}175 + 0{,}175 = 17{,}5 + 0{,}175 = 17{,}675$ er hele tal, er typetallet det største hele tal, der er mindre end 17,675, dvs. 17.

For at anskueliggøre dette ses på sandsynlighederne for 16, 17 og 18 succeser:

$binpdf(100, 0.175, 16) = 0.09994027634$
$binpdf(100, 0.175, 17) = 0.1047502361$
$binpdf(100, 0.175, 18) = 0.1024577226$

Pindediagrammet tegnes:

$pindediagramBIN(100, 0.175) =$



Opgaverne 323*

NEGATIV BINOMIALFORDELING

Den *negative binomialfordeling* (også kaldet *Pascalfordelingen*) baserer sig lige som binomialfordelingen på binomialeksperimenter, men fra en anden synsvinkel. Det er i modsætning til binomialfordelingen en uendelig sandsynlighedsfordeling.

Definition 20: *Den negative binomialfordeling* er sandsynlighedsfordelingen for antallet n af succeser i et binomialeksperiment, hvor deleksperimenterne udføres, indtil et fra start fastsat antal fiaskoer r er nået.
Den negative binomialfordeling er bestemt af successandsynligheden p og det fastsatte antal r af fiaskoer.
Den angives ved $NB(r,p)$.

Det fremgår implicit af denne definition, at r er et naturligt tal. Men vi skal snart se, at vi kommer til at anvende binomialkoefficienter og dermed fakultetstegn, når vi skal angive sandsynlighederne. Da vi indførte fakultetstegnet, blev det nævnt, at man kunne udvide dette begreb til at omhandle reelle tal (gammafunktionen). På samme måde kan man udvide den negative binomialfordeling til at tillade reelle r-værdier. I så fald kaldes det *Polya-fordelingen*. Vi holder os dog til *Pascalfordelingen* (naturlige tal).

Eksempel 84: *Vi bliver ved med at kaste en terning, indtil vi slår en 6'er.*
Hvis dette skal være en negativ binomialfordeling, skal vi altså se på sandsynlighederne for de forskellige antal kast, **inden** vi slår vores 6'er, og vi må sætte successandsynligheden til $p=\frac{5}{6}$ (da vi fortsætter, hvis vi ikke slår en 6'er), og det fastsatte antal fiaskoer er $r=1$, da vi stopper første gang, vi slår en 6'er.

Vores negative binomialfordeling kommer derfor til at hedde $NB\left(1,\frac{5}{6}\right)$.

Vi finder sandsynlighederne for de forskellige antal succeser:

$P(0)=\frac{1}{6}$, da 0 succeser fremkommer, hvis vi slår en 6'er i første kast.

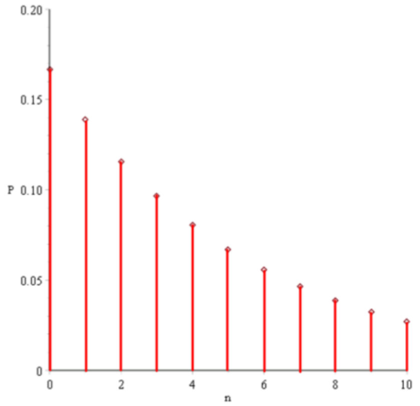
$P(1)=\frac{5}{6}\cdot\frac{1}{6}=\frac{5}{36}$, da 1 succes fremkommer, hvis vi ikke slår en 6'er i første kast, men i andet.

$P(2)=\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}=\frac{5^2}{6^3}$. Først to ikke-seksere og i tredje kast en 6'er.

$P(n)=\frac{5^n}{6^{n+1}}$. Først n ikke-seksere og i $n+1$'te kast en 6'er.

Den negative binomialfordeling $NB\left(1,\frac{5}{6}\right)$ er altså:

n	0	1	2	3	4	5	6	7	8	9	…
$P(n)$	0,167	0,139	0,116	0,096	0,080	0,067	0,056	0,047	0,039	0,032	…



Når vi lægger sandsynlighederne sammen, får vi en uendelig række, som vi genkender fra vores forløb om uendeligheder, nemlig en kvotientrække. Vi kan derfor udregne rækkesummen:

$$\sum_{n=0}^{\infty}P(n)=\frac{1}{6}+\frac{5}{6^2}+\frac{5^2}{6^3}+\ldots=\frac{\frac{1}{6}}{1-\frac{5}{6}}=1$$

Bemærk, at udseendet af pindediagrammet i Eksempel 84, hvor sandsynlighederne hele vejen mindskes for større *n*-værdier, gælder for samtlige negative binomialfordelinger med $r=1$, dvs. *p*-værdien kan ikke ændre dette (overvej dette!).

Dette specialtilfælde af den negative binomialfordeling $(r=1)$ kaldes også for *den geometriske fordeling*.

Opgaverne 324*

Lad os nu se på fordelingens udseende i et tilfælde, hvor $r>1$:

Eksempel 85: *Fra et kortspil uden jokere trækkes (MED tilbagelægning og blanding) ét kort ad gangen, indtil der er trukket 5 hjerter.*

Vores negative binomialfordeling giver os sandsynlighederne for de forskellige antal trukne kort, der ikke var hjerter, når vi stopper efter den 5. trukne hjerter.

Vi ser altså på den negative binomialfordeling $NB\left(5,\frac{3}{4}\right)$.

Hvis du overvejer situationen, kan du muligvis allerede uden udregninger se, at vi ikke længere får den største sandsynlighed for 0 trukne ikke-hjerter.

Lad os se på nogle sandsynligheder:

$P(0)=\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\left(\frac{1}{4}\right)^5=0,0977\%$, da vi her trækker 5 hjerter i træk.

Hvis vi skal ende med at have trukket ét kort, der ikke er hjerter, skal vi have trukket 6 kort i alt (5 hjerter og 1 ikke-hjerter). Dette minder rigtig meget om vores binomialfordeling, for det svarer jo til 6 deleksperimenter med 1 'succes'. MEN! Der er en lille, væsentlig forskel. Vi ved, at vores 'succes' ikke må være placeret som sidste trækning, for så ville vi allerede være stoppet efter at have trukket hjerter i de første 5 træk. Vores 'succes' har altså kun 5 pladser at vælge imellem. Vi får dermed følgende udregning (jævnfør binomialfordelingen – Sætning 25):

$$P(1)=\binom{5}{1}\cdot\left(\frac{3}{4}\right)^1\cdot\left(\frac{1}{4}\right)^5=\frac{5!}{1!\cdot(5-1)!}\cdot\left(\frac{3}{4}\right)^1\cdot\left(\frac{1}{4}\right)^5=5\cdot\left(\frac{3}{4}\right)^1\cdot\left(\frac{1}{4}\right)^5=0,3662\%$$

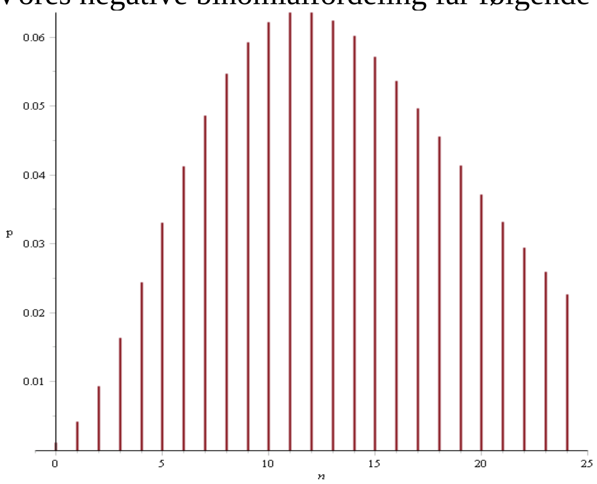
Hvis vi skal have trukket to kort, der ikke er hjerter, skal vi have trukket 7 kort i alt (5 hjerter og 2 ikke-hjerter). Igen må vi ifølge vores regler for trækningerne nødvendigvis ende med en hjerter, dvs. vores 2 ikke-hjerter skal være placeret på de første 6 pladser. Vi får dermed:

$$P(2)=\binom{6}{2}\cdot\left(\frac{3}{4}\right)^2\cdot\left(\frac{1}{4}\right)^5=\frac{6!}{2!\cdot(6-2)!}\cdot\left(\frac{3}{4}\right)^2\cdot\left(\frac{1}{4}\right)^5=0,8240\%$$

Samme argumentation giver os:

$$P(n)=\binom{5+n-1}{n}\cdot\left(\frac{3}{4}\right)^n\cdot\left(\frac{1}{4}\right)^5$$

Vores negative binomialfordeling får følgende udseende:



Det er altså mest sandsynligt at trække enten 11 eller 12 ikke-hjerter. Eller tilsvarende: Det er mest sandsynligt, at man ender med at have trukket 16 eller 17 kort.

Eksempel 85 gav en anvisning på, hvordan man kan behandle den generelle negative binomialfordeling $NB(r,p)$.

$P(n)$ er sandsynligheden for *n* succeser og *r* fiaskoer i $(n+r)$ deleksperimenter, hvor der ikke må være en succes i sidste deleksperiment, dvs. de *n* succeser skal fordeles på $(n+r-1)$ pladser. Dette giver os følgende sætning:

Sætning 28: Sandsynligheden for *n* succeser i den negative binomialfordeling $NB(r,p)$ er:

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

Opgaverne 325*

I den næste sætning bruges heltalsfunktionen angivet ved *int*. For positive tal angiver den det største heltal, der er mindre end eller lig med argumentet.

Dvs. på **positive tal** virker den ved at "afskære" eventuelle decimaler af et tal, f.eks. $\text{int}(29.753)=29$. Bemærk altså, at heltalsfunktionen IKKE afrunder. Da heltalsfunktionen afskærer decimaler, vil den for negative tal angive det mindste heltal, der er større end eller lig med argumentet. I Maple findes de to beslægtede funktioner *floor* og *ceil* (ceiling), hvor *floor* for alle tal angiver det største heltal, der er mindre end eller lig argumentet, mens *ceil* angiver det mindste heltal, der er større end eller lig argumentet. Afprøv begge dele og tjek, at du forstår betydningen af dem. For positive tal virker *floor* og *int* ens, så *floor* kunne også have været anvendt i nedenstående:

Sætning 29: For den negative binomialfordeling $NB(r,p)$ gælder:

a) Det gennemsnitlige antal succeser er: $n_{gen}=\dfrac{p\cdot r}{1-p}$

b) Det gennemsnitlige antal deleksperimenter er: $A_{gen}=\dfrac{r}{1-p}$

c) Typetallene/typetallet (det mest sandsynlige antal succeser) er:
$\dfrac{p\cdot r-1}{1-p}$ og $\dfrac{p\cdot(r-1)}{1-p}$, hvis tallene er hele tal (differensen mellem de to brøker er 1).
$\text{int}\left(\dfrac{p\cdot(r-1)}{1-p}\right)$, hvis tallet ikke er et helt tal.

d) Det mest sandsynlige antal deleksperimenter er:
$\dfrac{r-1}{1-p}$ og $\dfrac{r-p}{1-p}$, hvis tallene er hele tal (differensen mellem de to brøker er 1).
$\text{int}\left(\dfrac{r-p}{1-p}\right)$, hvis tallet ikke er et helt tal.

Hvis $r=1$, er det kun det største af tallene i Sætning 29.c/d, der gælder. Hvis $r=1$, får man f.eks. i c) resultaterne -1 og 0, og her er det så kun 0, der er typetal.

Bemærk, at forskellen mellem antal deleksperimenter og antal succeser er *r* (overvej dette!). Når du har lavet nedenstående to øvelser, behøver vi altså kun at bevise b) og c) i Sætning 29.

Øvelse 5: Vis, at der med udtrykkene fra Sætning 29 gælder: $n_{gen}+r=A_{gen}$.

Øvelse 6: Vis, at $\dfrac{p\cdot r-1}{1-p}+1=\dfrac{p\cdot(r-1)}{1-p}$, $\dfrac{p\cdot r-1}{1-p}+r=\dfrac{r-1}{1-p}$ og $\dfrac{p\cdot(r-1)}{1-p}+r=\dfrac{r-p}{1-p}$

Bevis 29: Vi vil benytte de store tals lov (Sætning 8) til at bevise Sætning 29.b.

Vi forestiller os, at vi udfører vores eksperiment m gange. Dette giver os en masse forskellige antal deleksperimenter $f_1, f_2, f_3, \ldots, f_m$. Hvis det f.eks. var vores terningkast fra Eksempel 84, så kunne det være, at vi i første eksperiment fik 4 deleksperimenter (6'eren kom i fjerde kast), i andet eksperiment 1 deleksperiment (6'eren kom i første kast), i tredje kast 9 deleksperimenter (6'eren kom i niende kast), osv.

Vi lægger alle disse antal deleksperimenter sammen og får det samlede antal N af deleksperimenter i de m eksperimenter:

$$f_1 + f_2 + f_3 + \ldots + f_m = N$$

Da hvert eksperiment fortsætter, indtil vi har fået r fiaskoer, har vi det samlede antal fiaskoer:

$$N_{fiasko} = m \cdot r$$

Dermed er frekvensen af fiaskoer: $$\frac{N_{fiasko}}{N} = \frac{m \cdot r}{N}$$

Men hvis vores m er tilpas stort (vi forestiller os, at vi laver grænseovergangen $m \to \infty$, da gennemsnittet jo netop er baseret på "uendelig" mange gentagelser), så fortæller de store tals lov os, at denne frekvens svarer til sandsynligheden for fiasko, og den er $1-p$.

Vi har altså: $$1-p = \frac{m \cdot r}{N}$$

I dette udtryk isoleres $\frac{N}{m}$, da det netop er det gennemsnitlige antal deleksperimenter i hvert eksperiment, fordi N er det samlede antal deleksperimenter og m det samlede antal eksperimenter:

$$1-p = \frac{m \cdot r}{N} \Leftrightarrow \frac{N}{m} = \frac{r}{1-p}$$

Tjek med udtrykket i Sætning 29.b og se, at beviset er gennemført.

For at bevise Sætning 29.c benyttes samme fremgangsmåde som i Bevis 27.

Vi opstiller forholdet mellem to successive sandsynligheder og ser på, hvornår det kommer under 1.

$$\frac{P(n)}{P(n-1)} = \frac{\binom{n+r-1}{n} \cdot p^n \cdot (1-p)^r}{\binom{n-1+r-1}{n-1} \cdot p^{n-1} \cdot (1-p)^r} = \frac{\frac{(n+r-1)!}{n! \cdot (n+r-1-n)!} \cdot p}{\frac{(n+r-2)!}{(n-1)! \cdot (n+r-2-(n-1))!}} =$$

$$\frac{(n+r-1)! \cdot (r-1)! \cdot (n-1)!}{n! \cdot (r-1)! \cdot (n+r-2)!} \cdot p = \frac{n+r-1}{n} \cdot p = \left(1 + \frac{r-1}{n}\right) \cdot p = p + \frac{r-1}{n} \cdot p$$

Vi er interesserede i, om denne brøk er større eller mindre end 1 for de forskellige n-værdier. For hvis brøken er større end 1, vil sandsynligheden for n succeser være større end for $n-1$ succeser. Og hvis brøken er mindre end 1, er vi "kommet for langt", for så er sandsynligheden for n succeser mindre end for $n-1$ succeser. Hvis brøken har værdien 1, er sandsynlighederne for n og $n-1$ succeser lige store.

Vi udregner:

$$p + \frac{r-1}{n} \cdot p = 1 \Leftrightarrow p - 1 = -\frac{r-1}{n} \cdot p \Leftrightarrow n = -\frac{r-1}{p-1} \cdot p \Leftrightarrow n = \frac{p \cdot (r-1)}{1-p}$$

$$p + \frac{r-1}{n} \cdot p > 1 \Leftrightarrow p - 1 > -\frac{r-1}{n} \cdot p \Leftrightarrow n < -\frac{r-1}{p-1} \cdot p \Leftrightarrow n < \frac{p \cdot (r-1)}{1-p}$$

$$p + \frac{r-1}{n} \cdot p < 1 \Leftrightarrow p - 1 < -\frac{r-1}{n} \cdot p \Leftrightarrow n > -\frac{r-1}{p-1} \cdot p \Leftrightarrow n > \frac{p \cdot (r-1)}{1-p}$$

Undervejs i ulighederne blev det benyttet, at $p-1<0$. Tjek, at det ønskede er bevist.

Eksempel 86: Vi anvender Sætning 29 på eksemplerne 84 og 85.

I Eksempel 84 havde vi $NB\left(1,\frac{5}{6}\right)$.

$n_{gen}=\frac{p\cdot r}{1-p}=\frac{\frac{5}{6}\cdot 1}{1-\frac{5}{6}}=5$. Dvs. man får i gennemsnit 5 ikke-seksere, før man første gang får en 6'er.

$A_{gen}=\frac{r}{1-p}=\frac{1}{1-\frac{5}{6}}=6$. Dvs. man skal i gennemsnit kaste 6 gange for at få den første 6'er.

$\frac{p\cdot(r-1)}{1-p}=\frac{\frac{5}{6}\cdot(1-1)}{1-\frac{5}{6}}=0$. Dvs. typetallet er 0.

I Eksempel 85 havde vi $NB\left(5,\frac{3}{4}\right)$

$n_{gen}=\frac{p\cdot r}{1-p}=\frac{\frac{3}{4}\cdot 5}{1-\frac{3}{4}}=15$.

Dvs. man får i gennemsnit 15 ikke-hjertere, før man når 5 hjerter.

$A_{gen}=\frac{r}{1-p}=\frac{5}{1-\frac{3}{4}}=20$. Dvs. man kommer i gennemsnit til at trække 20 kort.

$\frac{p\cdot r-1}{1-p}=\frac{\frac{3}{4}\cdot 5-1}{1-\frac{3}{4}}=\frac{\frac{15}{4}-1}{\frac{1}{4}}=11$ og $\frac{p\cdot(r-1)}{1-p}=\frac{\frac{3}{4}\cdot(5-1)}{1-\frac{3}{4}}=12$. Dvs. typetallene er 11 og 12.

Opgaverne 326*

Vi skal nu se på en fordeling, der på en del punkter minder om binomialfordelingen, og som i nogle sammenhænge derfor erstattes af denne, da binomialfordelingen er simplere at regne med og er en god tilnærmelse.

DEN HYPERGEOMETRISKE FORDELING

Definition 21: Udgangspunktet for den hypergeometriske fordeling er en population på *n* objekter, hvoraf *k* symboliserer succes og $n-k$ symboliserer fiasko.
Fra de *n* objekter udtrækkes **uden** tilbagelægning *t* objekter, og den hypergeometriske fordeling angiver så sandsynligheden for at få *r* succeser.

Vi har altså at gøre med en række IKKE-uafhængige stokastiske eksperimenter, da man ikke lægger tilbage efter hver trækning, og dermed bliver $P\left(succes_{i+1} \middle| succes_i\right) \neq P\left(succes_{i+1}\right)$.

Sandsynlighederne $P\left(succes_i\right)$ for at trække en succes i den *i*'te udtrækning $\left(i = 1, 2, 3, \ldots, t\right)$ er alle ens, nemlig $P\left(succes_i\right) = \frac{k}{n}$, men den betingede sandsynlighed for en trække en succes i en udtrækning, givet at der blev udtrukket en succes i udtrækningen før, er mindre, da man i så fald har fjernet et af succes-objekterne, dvs. $P\left(succes_{i+1} \middle| succes_i\right) < \frac{k}{n}$ (jf. Definition 5 om uafhængighed).

Eksempel 87: Standardeksemplet er en krukke indeholdende to forskelligt farvede kugler. Det kunne være en krukke med 60 kugler, hvoraf 15 er grønne og 45 sorte. Og spørgsmålet kunne være: "*Hvad er sandsynligheden for at trække netop 3 grønne kugler ved trækning af 10 kugler?*"
Her kunne grøn symbolisere 'succes', og den stokastiske variabel *X* kunne så angive antal succeser ved optrækning af 10 kugler.
Vi skulle så finde $P\left(X = 3\right)$.

Opgaverne 327*

Eksempel 88: I en meningsmåling udgør den stemmeberettigede del af befolkningen populationen *n*, og stikprøven udgør så de *t* mennesker, der udspørges. At ville stemme på partiet Imperiet (T) angiver 'succes', og 'fiasko' er så ikke at ville stemme på partiet T. Den hypergeometriske fordeling angiver så sandsynlighederne for de forskellige antal blandt de *t* mennesker, der vil stemme på partiet T.

Bemærk, at vi i Eksempel 78 behandlede ovenstående situation som en binomialfordeling, og det er også det, man normalt vil gøre, da populationen er meget større end stikprøven.
Men hvis man skal have det helt præcist, skal det altså være den hypergeometriske fordeling.

Sætning 30: Lad $h\left(n, k, t\right)$ være det hypergeometriske eksperiment, der består af *n* objekter, hvoraf *k* er succeser, og hvorfra der trækkes *t* objekter uden tilbagelægning.
X er den stokastiske variabel, der angiver antallet *r* af succeser.

Ud fra ovenstående gælder $n \geq t \geq r$ og $n \geq k \geq r$, og man har så: $P\left(X = r\right) = \dfrac{\binom{k}{r} \cdot \binom{n-k}{t-r}}{\binom{n}{t}}$

Bevis 30: Vi har allerede i Eksempel 73 set den tankegang, der skal bruges til at beregne sandsynlighederne. De findes ved at benytte sætning 3, dvs. finde antal gunstige udfald og dele det med antal mulige udfald. De mulige udfald er alle udtrækninger af *t* elementer fra en *n*-mængde, dvs. $\binom{n}{t}$, mens de gunstige udfald ifølge Sætning 13 findes ved at finde antallet af mulige udtrækninger af *r* succeser blandt de *k* succeser og multiplicere det med antallet af mulige udtrækninger af de $t-r$ fiaskoer blandt de $n-k$ elementer, der er fiaskoer.

Eksempel 89 (udregning af sandsynligheden beskrevet i Eksempel 87): Vi har $n = 60$ og $k = 15$.
Da $t = 10$ har man:

$$P(X=3) = \frac{\binom{15}{3} \cdot \binom{60-15}{7}}{\binom{60}{10}} = \frac{8021650}{29290609} \xrightarrow{\text{at 5 digits}} 0.27386$$

Dvs. at sandsynligheden for, at der er netop 3 grønne blandt de 10 optrukne kugler, er 27%.

Eksempel 90: Vi har en vælgerskare på 4 millioner mennesker, hvoraf 1 million vil stemme på partiet Imperiet (med partibogstavet T). Vi udvælger en stikprøve på 2500 personer og ønsker at finde sandsynligheden for, at præcis 600 af dem vil stemme på T.

Vi prøver både at benytte den korrekte hypergeometriske fordeling og binomialfordelingen, der kun er en tilnærmelse, da man ikke "lægger tilbage", dvs. man kan ikke spørge den samme person flere gange:

Den hypergeometriske fordeling:

$$P(X=600) = evalf\left(\frac{\binom{1000000}{600} \cdot \binom{4000000-1000000}{2500-600}}{\binom{4000000}{2500}}\right) = 0.009531233544$$

Binomialfordelingen (successandsynligheden er 25%):

$$P(X=600) = \binom{2500}{600} \cdot 0.25^{600} \cdot (1-0.25)^{2500-600} = 0.009532307446$$

Afvigelsen ligger på 4. betydende ciffer, så vi kan se, at det ikke er urimeligt at anvende binomialfordelingen i netop denne situation.

Opgaverne 328*

DEN NEGATIVE HYPERGEOMETRISKE FORDELING

Muligvis kommer det ikke som den store overraskelse, at der også er en negativ hypergeometrisk fordeling:

Definition 22: Se på en samling af *n* objekter, hvoraf *k* kan betegnes som 'succes'.
De resterende $n-k$ objekter betegnes som fiasko.
Der trækkes blandt objekterne **uden** tilbagelægning, indtil et på forhånd fastsat antal *t* af fiaskoer er nået ($0<t\leq n-k$).
Den negative hypergeometriske fordeling angiver sandsynlighederne $P(r)$ for antallet *r* af succeser i trækningen.

Bemærk, at det er en **endelig** sandsynlighedsfordeling, for man har $0\leq r\leq k$.

Den negative hypergeometriske fordeling karakteriseres ved *n, k* og *t*, og den skrives $NHG(n,k,t)$.

Sætning 31: For den negative hypergeometriske fordeling $NHG(n,k,t)$ gælder:

$$P(r)=\frac{\binom{k}{r}\cdot\binom{n-k}{t}}{\binom{n}{r+t}}\cdot\frac{t}{r+t}$$

Bevis 31: Hvis man har fået *r* succeser, har man trukket $r+t$ objekter, for man trækker jo, indtil man har nået *t* fiaskoer. Dvs. vi kan bruge vores hypergeometriske fordeling (Sætning 30) med $r+t$ trækninger og med den væsentlige tilføjelse, at vi skal have sorteret alle de trækninger fra, der ender med en succes (for reglen er jo, at man slutter, når man når *t* fiaskoer).
Vi ser nu på de $r+t$ objekter, der er trukket. Sandsynligheden for, at det sidste objekt er en fiasko, er $\frac{t}{r+t}$. Vi bruger så vores multiplikationsprincip for sandsynligheder ("både-og"), hvor vi siger, at vi skal **både** have trukket *r* succeser og *t* fiaskoer, **og** vi skal have en fiasko i sidste trækning. Det giver os:

$$P(r)=\frac{\binom{k}{r}\cdot\binom{n-k}{(r+t)-r}}{\binom{n}{r+t}}\cdot\frac{t}{r+t}=\frac{\binom{k}{r}\cdot\binom{n-k}{t}}{\binom{n}{r+t}}\cdot\frac{t}{r+t}$$

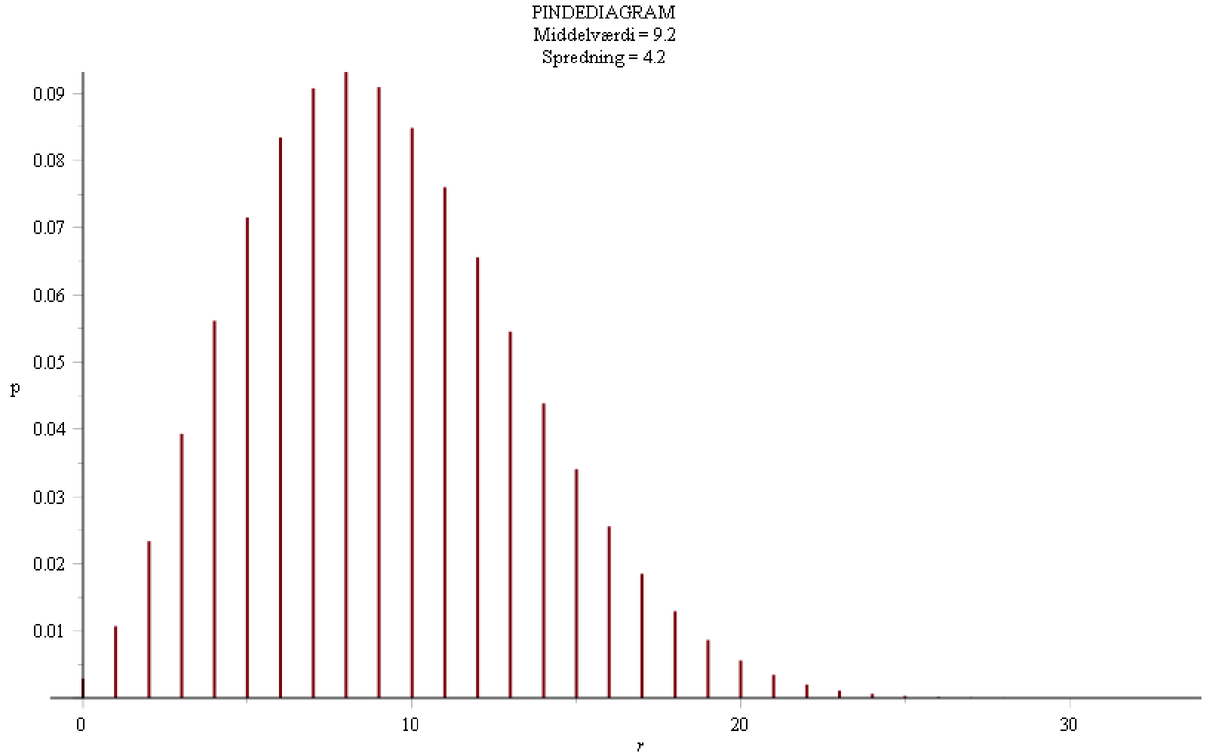
Eksempel 91: En slikpose indeholder 50 stykker slik, hvoraf 17 er lakridser. En person vil gerne have 5 lakridser og bliver derfor ved med at tage stykker op, indtil det ønskede antal er nået (naturligvis kigger personen ikke ned i posen eller lægger stykker tilbage).
Bemærk, at da det er lakridserne, der er afgørende for, hvornår personen stopper, skal disse sættes som fiaskoer, selvom det er dem, personen foretrækker. Dvs. posen indeholder 33 succeser og 17 fiaskoer. Vores negative hypergeometriske fordeling giver os så sandsynlighederne for det antal ikke-lakridser, der trækkes.

Eksempler:

$P(0)=\dfrac{\binom{33}{0}\cdot\binom{17}{5}}{\binom{50}{5}}\cdot\dfrac{5}{0+5}=0,292\%$, hvilket er sandsynligheden for at trække udelukkende lakridser.

$P(1)=\dfrac{\binom{33}{1}\cdot\binom{17}{5}}{\binom{50}{5}}\cdot\dfrac{5}{1+5}=1,071\%$. Her trækkes 5 lakridser og 1 ikke-lakrids.

Et pindediagram over denne negative hypergeometriske fordeling er:



Det er altså mest sandsynligt (9,3%), at man trækker 8 stykker, der ikke er lakridser, for at få 5 lakridser (dvs. 13 stk. slik i alt).
Og den angivne middelværdi fortæller os, at vi må forvente i gennemsnit at skulle trække godt 14 stykker slik for at få 5 lakridser.

Uden bevis kan det nævnes, at gennemsnittet for den negative hypergeometriske fordeling $NHG(n,k,t)$ er $\mu=\dfrac{k\cdot t}{n-k+1}$. Tjek, at det stemmer med Eksempel 91.

Vi kan også bruge gennemsnittet til næste eksempel:

Eksempel 92:
Hvor mange kort skal man i gennemsnit trække fra et kortspil (52 kort) for at få 2 esser?
Her er esserne fiaskoer, da de er afgørende for, hvornår man stopper. Vores fordeling bliver altså $NHG(52,48,2)$.

Gennemsnittet for fordelingen er derfor:

$$\mu=\frac{k\cdot t}{n-k+1}=\frac{48\cdot 2}{52-48+1}=\frac{96}{5}=19,2$$

Dvs. der skal i gennemsnit trækkes godt 21 kort (2 esser og 19 ikke-esser).

Opgaverne 329*

DEN MULTINOMIALE FORDELING

Binomialfordelingen og den hypergeometrisk fordeling beskæftiger sig begge med *n* deleksperimenter, hvor der kun er to mulige udfald – succes og fiasko. Forskellen på de to er, at binomialfordelingen beskæftiger sig med uafhængige og identiske deleksperimenter – svarende til ”med tilbagelægning”, hvis man tænker på trækning af kugler fra krukker eller kort fra kortspil – mens den hypergeometriske fordeling er ”uden tilbagelægning”, hvilket bryder med uafhængigheden.

Begge fordelinger kan udvides til fordelinger, hvor deleksperimentet har mere end to mulige udfald. Binomialfordelingen bliver i så fald til *den multinomiale fordeling*, mens den hypergeometriske fordeling bliver til *den multivariable hypergeometriske fordeling*.

Definition 23: Et *multinomialeksperiment* er et stokastisk eksperiment, der består af *n* udførelser af identiske deleksperimenter med mere end to mulige udfald, dvs. $U_{del} = \{u_1, u_2, u_3, ..., u_s\}$, hvor $s > 2$.

De stokastiske variable $X_1, X_2, X_3, ..., X_s$ angiver antallene af udfaldene $u_1, u_2, u_3, ..., u_s$ i de *n* udførelser.

Vi skriver sandsynligheden $P(X_1 = r_1 \wedge X_2 = r_2 \wedge X_3 = r_3 \wedge ... \wedge X_s = r_s)$ som $P(r_1, r_2, r_3, ..., r_s)$

Sætning 32: Ved et multinomialeksperiment med *n* deleksperimenter, der hver har *s* mulige udfald med sandsynlighederne $p_1, p_2, p_3, ..., p_s$, hvor $\sum_{i=1}^{s} p_i = 1$, er den multinomiale fordeling givet ved følgende sandsynlighedsfunktion, hvis $\sum_{i=1}^{s} r_i = n$:

$$P(r_1, r_2, r_3, ..., r_s) = \frac{n!}{r_1! \cdot r_2! \cdot r_3! \cdot ... \cdot r_s!} \cdot p_1^{r_1} \cdot p_2^{r_2} \cdot p_3^{r_3} \cdot ... \cdot p_s^{r_s} = n! \cdot \prod_{i=1}^{s} \frac{p_i^{r_i}}{r_i!}$$

Hvis $\sum_{i=1}^{s} r_i \neq n$, er sandsynligheden 0 (overvej selv, hvorfor).

Eksempel 93: I en krukke med 100 kugler er der 37 blå, 29 røde, 21 grønne og 13 sorte kugler. 20 gange trækkes fra krukken en kugle, der efterfølgende lægges tilbage.

Sandsynligheden for at trække 7 blå, 6 røde, 4 grønne og 3 sorte kugler er:

$$P(7,6,4,3) = \frac{20!}{7! \cdot 6! \cdot 4! \cdot 3!} \cdot 0{,}37^7 \cdot 0{,}29^6 \cdot 0{,}21^4 \cdot 0{,}13^3 = 0{,}011233 \approx 1{,}1\%$$

Sandsynligheden for at trække 8 blå, 5 røde, 0 grønne og 7 sorte kugler er:

$$P(8,5,0,7) = \frac{20!}{8! \cdot 5! \cdot 0! \cdot 7!} \cdot 0{,}37^8 \cdot 0{,}29^5 \cdot 0{,}21^0 \cdot 0{,}13^7 = 0{,}000045102 \approx 0{,}0045\%$$

Sandsynligheden for at trække 5 blå, 7 røde, 4 grønne og 9 sorte kugler er 0, da $5+7+4+9 \neq 20$

Opgaverne 330*

Bemærk, hvordan Sætning 32 er en udvidelse af Sætning 25, da man ved at indføre skrivemåden $r_{succes} = r$ og $r_{fiasko} = n - r$ samt $p_{succes} = p$ og $p_{fiasko} = 1 - p$ i binomialfordelingen kunne få Sætning 25 til at hedde:

$$P(r_{succes}, r_{fiasko}) = K(n,r) \cdot p^r \cdot (1-p)^{n-r} = \frac{n!}{r_{succes}! \cdot r_{fiasko}!} \cdot p_{succes}^{r_{succes}} \cdot p_{fiasko}^{r_{fiasko}}$$

Derfor er det ikke overraskende, at beviset for Sætning 32 minder om beviset for Sætning 25

Bevis 32: Vi ser på n deleksperimenter med de mulige udfald $u_1, u_2, u_3, ..., u_s$, der i de enkelte deleksperimenter indtræffer med sandsynlighederne $p_1, p_2, p_3, ..., p_s$, og som i de n udførelser hver især indtræffer $r_1, r_2, r_3, ..., r_s$ gange.

Sætning 14 kan så bruges til at beregne sandsynligheden $P_{permutation}$ for, at de enkelte udfald indtræffer i en bestemt rækkefølge:

$$P_{permutation} = \underbrace{p_1 \cdot p_1 \cdot ... \cdot p_1}_{r_1\ faktorer} \cdot \underbrace{p_2 \cdot p_2 \cdot ... \cdot p_2}_{r_2\ faktorer} \cdot ... \cdot \underbrace{p_s \cdot p_s \cdot ... \cdot p_s}_{r_s\ faktorer} = p_1^{r_1} \cdot p_2^{r_2} \cdot ... \cdot p_s^{r_s}$$

Det er væsentligt at bemærke, at dette udtryk er det samme uanset hvilken permutation, der ses på. For så mangler vi blot at tælle hvor mange af sådanne permutationer, der findes:

Først skal vi blandt de n deleksperimenter have placeret de r_1 deleksperimenter, der gav udfaldet u_1. Det kan gøres på $K(n, r_1) = \dfrac{n!}{r_1! \cdot (n - r_1)!}$ måder, da vi udvælger r_1 placeringer blandt n mulige (præcis som i beviset for binomialfordelingen).

Derefter (og hermed opfyldes betingelsen om *ordnede* delvalg i Sætning 13 om multiplikation af valgmuligheder) skal vi blandt de resterende $n - r_1$ placeringer have udvalgt de r_2 placeringer, hvor udfaldet u_2 indtraf (bemærk, at **antallet** af valgmuligheder er uafhængigt af, hvordan vi placerede u_1 - hvilket sikrer, at vi om lidt kan benytte Sætning 13). Dette kan gøres på

$K(n - r_1, r_2) = \dfrac{(n - r_1)!}{r_2! \cdot (n - r_1 - r_2)!}$ måder.

Når vi efterfølgende skal have placeret u_3 på r_3 forskellige placeringer, er der $n - r_1 - r_2$ tilbage at vælge imellem, dvs. det kan gøres på $K(n - r_1 - r_2, r_3) = \dfrac{(n - r_1 - r_2)!}{r_3! \cdot (n - r_1 - r_2 - r_3)!}$ måder.

Og således fortsættes, indtil u_s skal placeres (det kan kun gøres på 1 måde, men dette 1-tal kan ifølge vores system skrives som): $K(n - r_1 - r_2 - ... - r_{s-1}, r_s) = \dfrac{(n - r_1 - r_2 - ... - r_{s-1})!}{r_s! \cdot (n - r_1 - r_2 - ... - r_s)!}$.

Sætning 13 (multiplikation for valgmuligheder) giver så, at det samlede antal permutationer er:

$$\frac{n!}{r_1! \cdot (n - r_1)!} \cdot \frac{(n - r_1)!}{r_2! \cdot (n - r_1 - r_2)!} \cdot \frac{(n - r_1 - r_2)!}{r_3! \cdot (n - r_1 - r_2 - r_3)!} \cdot ... \cdot \frac{(n - r_1 - r_2 - ... - r_{s-1})!}{r_s! \cdot (n - r_1 - r_2 - ... - r_s)!} = \frac{n!}{r_1! \cdot r_2! \cdot r_3! \cdot ... \cdot r_s!}$$

Og da vores valg af permutation er et *enten-eller* valg, giver Sætning 16 os:

$$P(r_1, r_2, r_3, ..., r_s) = \frac{n!}{r_1! \cdot r_2! \cdot r_3! \cdot ... \cdot r_s!} \cdot p_1^{r_1} \cdot p_2^{r_2} \cdot p_3^{r_3} \cdot ... \cdot p_s^{r_s}$$

Eksempel 94: 30 terninger kastes.

Sandsynligheden for at få 6 1'ere, 4 2'ere, 5 3'ere, 2 4'ere, 7 5'ere og 6 6'ere er:

$$P(6,4,5,2,7,6) = \frac{30!}{6! \cdot 4! \cdot 5! \cdot 2! \cdot 7! \cdot 6!} \cdot \left(\frac{1}{6}\right)^6 \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{1}{6}\right)^5 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^6 =$$

$$\frac{30!}{6! \cdot 4! \cdot 5! \cdot 2! \cdot 7! \cdot 6!} \cdot \left(\frac{1}{6}\right)^{30} = 0{,}0000797268 \approx 0{,}0080\%$$

Sandsynligheden for at få 5 af hver øjental er:

$$P(5,5,5,5,5,5) = \frac{30!}{(5!)^6} \cdot \left(\frac{1}{6}\right)^{30} = 0{,}000401823 \approx 0{,}040\%$$

Sandsynligheden for 30 6'ere er: $P(0,0,0,0,0,30) = \dfrac{30!}{(0!)^5 \cdot 30!} \cdot \left(\dfrac{1}{6}\right)^{30} = \left(\dfrac{1}{6}\right)^{30} = 4{,}5 \cdot 10^{-22}\%$

DEN MULTIVARIABLE HYPERGEOMETRISKE FORDELING

Vi udvider nu den hypergeometriske fordeling ved at sige:

Definition 24: I en population på n objekter er objekterne opdelt efter egenskaberne $u_1, u_2, u_3, ..., u_s$, ($s > 2$), hvor $k_1, k_2, k_3, ..., k_s$ angiver antallet af objekter med den pågældende egenskab, og hvor $\sum_{i=1}^{s} k_i = n$.

Fra de n objekter udtrækkes **uden** tilbagelægning t objekter. De stokastiske variable $X_1, X_2, X_3, ..., X_s$ angiver antallet af objekter i udtrækningen med hver af egenskaberne $u_1, u_2, u_3, ..., u_s$ og den multivariable hypergeometriske fordeling er så fordelingen beskrevet ved $P\left(X_1 = r_1 \wedge X_2 = r_2 \wedge X_3 = r_3 \wedge ... \wedge X_s = r_s\right)$, der skrives $P\left(r_1, r_2, r_3, ..., r_s\right)$

Sætning 33: I den multivariable hypergeometriske fordeling gælder for $\sum_{i=1}^{s} r_i = t$:

$$P\left(r_1, r_2, r_3, ..., r_s\right) = \frac{\binom{k_1}{r_1} \cdot \binom{k_2}{r_2} \cdot \binom{k_3}{r_3} \cdot ... \cdot \binom{k_s}{r_s}}{\binom{n}{t}} = \frac{\prod_{i=1}^{s} \binom{k_i}{r_i}}{\binom{n}{t}} \quad ; n \geq t \; ; k_i \geq r_i$$

Bevis 33: Beviset er magen til beviset for Sætning 30 – bare med nogle ekstra faktorer.

Eksempel 95: I en krukke med 100 kugler er der 37 blå, 29 røde, 21 grønne og 13 sorte kugler. Der trækkes 20 kugler fra krukken (uden tilbagelægning).

Sandsynligheden for at trække 7 blå, 6 røde, 4 grønne og 3 sorte kugler er:

$$P(7,6,4,3) = \frac{\binom{37}{7} \cdot \binom{29}{6} \cdot \binom{21}{4} \cdot \binom{13}{3}}{\binom{100}{20}} = 1,56\%$$

Sandsynligheden for at trække 8 blå, 5 røde, 0 grønne og 7 sorte kugler er:

$$P(8,5,0,7) = \frac{\binom{37}{8} \cdot \binom{29}{5} \cdot \binom{21}{0} \cdot \binom{13}{7}}{\binom{100}{20}} = 0,00147\%$$

Sandsynligheden for at trække 2 blå, 2 røde, 1 grønne og 15 sorte kugler er 0, da der kun er 13 sorte kugler i krukken.

Eksempel 96: Fra et kortspil uden jokere trækkes 5 kort til en hånd.

Sandsynligheden for at trække 2 hjerter, 1 spar og 2 klør er:

$$P(2,1,2,0)=\frac{\binom{13}{2}\cdot\binom{13}{1}\cdot\binom{13}{2}\cdot\binom{13}{0}}{\binom{52}{5}}=3,04\%$$

Sandsynligheden for at trække 2 billedkort, 1 es og 2 talkort (2-10) er:

$$P(2,1,2)=\frac{\binom{12}{2}\cdot\binom{4}{1}\cdot\binom{36}{2}}{\binom{52}{5}}=6,40\%$$

Hvis man skal finde sandsynligheden for at trække 2 billedkort og 3 ruder, skal man være opmærksom på, at dette ikke er en brugbar opdeling efter egenskaber, da et billedkort også kan være en ruder, dvs. nogle kort besidder begge egenskaber.

Opgaverne 331*

POISSON-FORDELING

Vi slutter af med en meget anvendelig fordeling, der desværre i disse noter udledes ud fra binomialfordelingen på en ikke helt tilfredsstillende måde, da der bl.a. henvises til en sætning, vi ikke har bevist.

Definition 25: Vi ser på en stokastisk proces, hvor en hændelse indtræffer med en fast, gennemsnitlig hyppighed uafhængigt af, hvornår eller hvor den sidst er indtruffet.
Poissonfordelingen angiver så sandsynligheden for de forskellige antal indtrufne hændelser i et givet interval.
Poissonfordelinger karakteriseres ved parameteren λ, der angiver, hvor mange gange hændelsen i gennemsnit indtræffer i det givne interval, og vi angiver så den pågældende fordeling ved $Pois(\lambda)$.

Bemærk, at vi altså allerede – uden overhovedet at have angivet fordelingen – kender gennemsnittet for poissonfordelingen $Pois(\lambda)$, for det er λ.

Bemærk også, at vi faktisk er nødt til at have fundet (eller anslået) dette gennemsnit, før vi overhovedet er i stand til at begynde at regne på fordelingen.
Bemærk desuden, at længden af vores givne interval ikke kommer til at indgå i udtrykkene, for vores λ er knyttet til denne længde, og hele informationen ligger altså i λ.
Og bemærk endelig, at det givne interval ikke behøver at være et tidsrum. Det kan f.eks. også være en længde, et areal eller et rumfang.

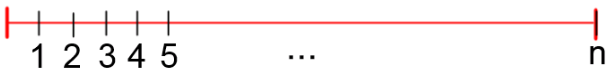
Eksempel 97: Her er nogle eksempler, hvor poissonfordelinger kommer i spil.

a) Se på en radioaktiv kilde med tilpas lang halveringstid til, at vi kan antage aktiviteten for at være konstant i tid. Hvis vi 30 gange måler antal henfald i intervaller af 10 sekunder, kan vi finde det gennemsnitlige antal henfald pr. 10 sekunder. Når dette gennemsnit er fundet, kan vi finde sandsynlighederne for de forskellige antal henfald inden for 10 sekunder.
b) En elendig dartspiller kaster dartpile mod en dartskive. Pilene rammer fuldstændig tilfældigt på skiven. Antallet af pile på skiven delt med skivens areal vil fortælle os, hvor mange pile der i gennemsnit vil være inden for et vist givent areal. Kig nu på arealer på skiven af denne givne størrelse. Pilene vil være poissonfordelt, dvs. sandsynlighederne for de forskellige antal pile inden for de givne arealer vil kunne aflæses af poissonfordelingen.
c) På et langt stykke DNA findes en række genmutationer, og dette giver os et gennemsnit for, hvor mange genmutationer der er på et givet stykke. Antal genmutationer på stykker af den givne længde vil være poissonfordelt.

Sætning 34: Sandsynligheden for, at den givne hændelse indtræffer *k* gange i det givne interval i poissonfordelingen $Pois(\lambda)$, er givet ved:

$$P(k) = \frac{\lambda^k \cdot \mathbf{e}^{-\lambda}}{k!}$$

Bevis 34: Vi ser på vores givne interval (uden at lægge os fast på, om det repræsenterer tid, længde, areal, rum eller noget helt femte), hvor hændelsen i gennemsnit indtræffer λ gange:



$$p = \frac{\lambda}{n}$$

Intervallet opdeles i n stykker, da det i første omgang gør det muligt for os at anvende vores (diskrete) binomialfordeling på situationen. Hvis vi betragter det som en binomialfordeling, har vi n deleksperimenter, og vi kan bestemme successandsynligheden p (sandsynligheden for hændelsen indtræffer inden for ét stykke), da vi også kender gennemsnittet λ. Vores Sætning 26 giver:

$$p = \frac{\lambda}{n}$$

Hermed har vi sandsynlighedsfordelingen (Sætning 25):

$$P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \frac{n!}{k! \cdot (n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Dette udtryk omskrives:

$$P(k) = \frac{n!}{k! \cdot (n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} =$$

$$\frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!} \cdot \frac{1}{n^k \cdot \left(\frac{n-\lambda}{n}\right)^k} \cdot \left(1-\frac{\lambda}{n}\right)^n =$$

$$\frac{\lambda^k}{k!} \cdot \left(\overbrace{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}^{\text{Her er k faktorer}} \right) \cdot \frac{1}{(n-\lambda)^k} \cdot \left(1-\frac{\lambda}{n}\right)^n =$$

$$\frac{\lambda^k}{k!} \cdot \frac{n}{n-\lambda} \cdot \frac{n-1}{n-\lambda} \cdot \frac{n-2}{n-\lambda} \cdot \ldots \cdot \frac{n-k+1}{n-\lambda} \cdot \left(1-\frac{\lambda}{n}\right)^n$$

Bemærk, at ovenstående udtryk altså fremkommer, når vi anvender binomialfordelingen på situationen.

Pointen er nu, at vi IKKE kan anvende binomialfordelingen på situationen, for i binomialfordelingen udfører vi kun vores deleksperiment n gange, dvs. ved alle de sorte streger i intervallet. Og vores situation går netop ud på, at vores deleksperimenter så at sige skal udføres hele tiden. Dette kan vi imidlertid opnå ved at foretage grænseovergangen $n \to \infty$. Dvs. vores binomialfordeling vil komme tættere og tættere på vores poissonfordeling, jo større vi gør n.

Prøv at sammenligne udtrykket for binomialfordelingen med udtrykket for poissonfordelingen i Sætning 34 og tjek, at vi har vist sætningen, når vi om lidt har argumenteret for, at den sorte del af udtrykket fra binomialfordelingen går mod 1, og den magentafarvede del går mod $\mathbf{e}^{-\lambda}$ for $n \to \infty$.

Vi ser først på det sidste. Vi genkender udtrykket fra Funktioner del 2, hvor det blev nævnt, at

$\lim\limits_{n\to\infty} \left(1+\frac{x}{n}\right)^n = \mathbf{e}^x$. Dvs. vi har $\lim\limits_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = \mathbf{e}^{-\lambda}$.

At alle faktorerne i den sorte del af udtrykket går mod 1 for $n \to \infty$ ses ved en omskrivning af en af faktorerne (der tages et enkelt eksempel, og du skal så tjekke, at du kan se, at det samme kan gøres med de andre):

$$\frac{n-2}{n-\lambda} = \frac{n-\lambda+\lambda-2}{n-\lambda} = 1 + \frac{\lambda-2}{n-\lambda} \to 1 \ \mathit{for}\ n \to \infty \quad ,$$

da tælleren i den sidste brøk er en konstant, mens nævneren går mod uendelig.

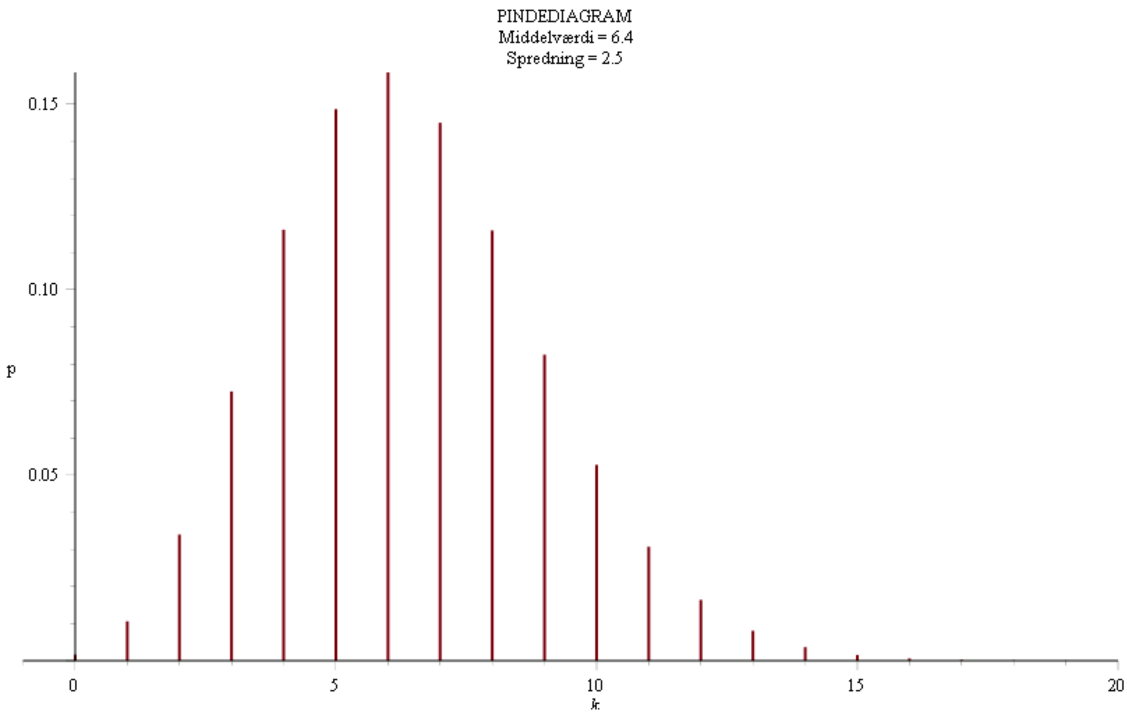
Eksempel 98: Inden et forsøg med en radioaktiv kilde måles baggrundsstrålingen. Vi måler 20 gange i vores interval på 10 sekunder og får et gennemsnit på 6,4 målinger. Vores poissonfordeling er altså $Pois(6.4)$.

Vi ser først på et par eksempler på udregninger af sandsynligheder:

$$P(3) = \frac{\lambda^3 \cdot \mathbf{e}^{-\lambda}}{3!} = \frac{6,4^3 \cdot \mathbf{e}^{-6,4}}{3!} = 7,26\%$$

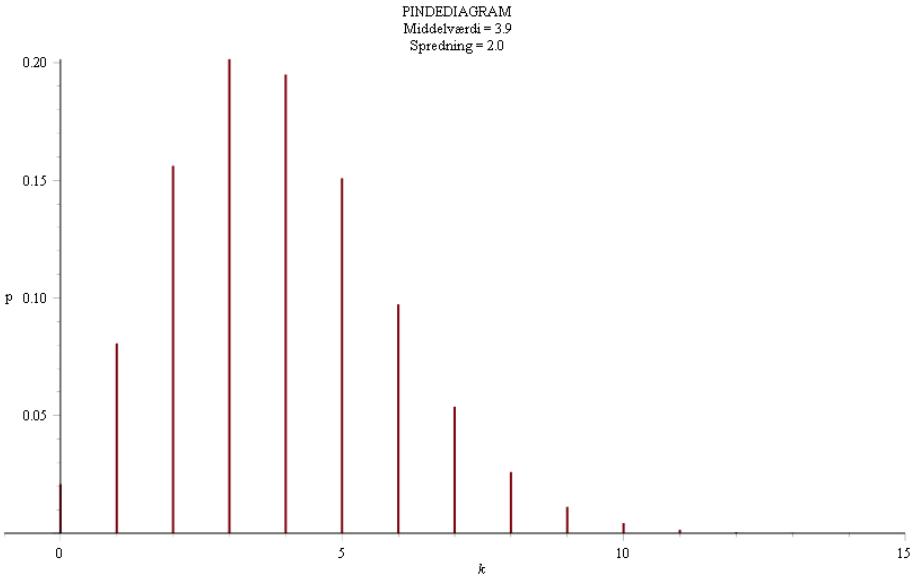
$$P(9) = \frac{\lambda^9 \cdot \mathbf{e}^{-\lambda}}{9!} = \frac{6,4^9 \cdot \mathbf{e}^{-6,4}}{9!} = 8,25\%$$

Et pindediagram over poissonfordelingen er:



Eksempel 99: På en sti belagt med ens fliser observeres 387 fugleklatter på et stykke med 100 fliser. Antallet af fugleklatter på hver flise må forventes at følge poissonfordelingen $Pois(3.87)$.

Et pindediagram er:



Ud over, at λ angiver middelværdien for en poissonfordeling, så angiver den faktisk også variansen:

Sætning 35: For poissonfordelingen $Pois(\lambda)$ gælder:

$$\text{var}(X) = \lambda \qquad \sigma = \sqrt{\lambda}$$

Opgaverne 332*

Vi behøver kun at vise udtrykket for variansen, for udtrykket for spredningen følger direkte af dette. Vi udnytter følgende identitet, som vi viste i forbindelse med taylorrækker:

$$\mathbf{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Vi kan nu bevise sætningen:

Bevis 35: Vi benytter Sætning 9.d ($\text{var}(X) = E(X^2) - E(X)^2$). Vi får så:

$$\text{var}(X) = \sum_{k=0}^{\infty} \frac{\lambda^k \cdot \mathrm{e}^{-\lambda}}{k!} \cdot k^2 - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot k^2 - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \cdot k^2 - \lambda^2 = \text{ [Første led i summationen giver 0, så det fjernes]}$$

$$\mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \cdot k - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot (k+1) - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot k + \mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot 1 - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \cdot k + \mathrm{e}^{-\lambda} \cdot \lambda \cdot \mathrm{e}^{\lambda} - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \lambda^2 \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} + \lambda - \lambda^2 =$$

$$\mathrm{e}^{-\lambda} \cdot \lambda^2 \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \lambda - \lambda^2 = \mathrm{e}^{-\lambda} \cdot \lambda^2 \cdot \mathrm{e}^{\lambda} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Når man siger, at usikkerheden på et tælletal n er $\sqrt{n}$, så antager man altså, at tælletallene er poissonfordelt, og man regner med, at det pågældende tal er tilpas tæt på middelværdien. For egentlig er det jo $\sqrt{\lambda}$, der er usikkerheden (spredningen).

Dette fortæller os også, at den relative usikkerhed på tælletallet er $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$, dvs. den relative usikkerhed mindskes, når tælletallet øges. Det er derfor, man som udgangspunkt skal forsøge at opnå så store tælletal som muligt.