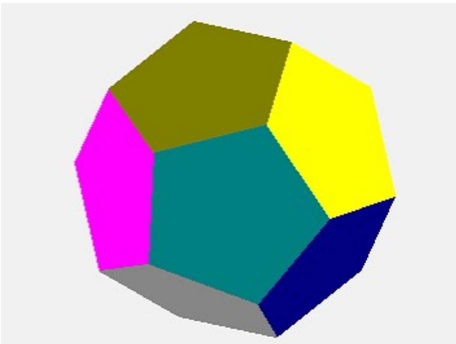
Elementær Matematik

Sandsynlighedsregning



Ole Witt-Hansen Køge Gymnasium 2008

Indhold

Kap 1. Kombinatorik

1. Multiplikations- og addtionsprincippet

I kombinatorikken beskæftiger man sig med metoder til en systematisk optælling af antallet af forskellige måder, hvorpå en udvælgelse af elementer fra en mængde kan finde sted. Kombinatorikken bygger på to simple principper, som kaldes for multiplikationsprincippet (også kaldet for både-og princippet) og additionsprincippet (også kaldet for enten-eller princippet).

Først vil vi fastslå en helt elementær ting.

- Hvis man har en mængde på *n* forskellige elementer, kan man udvælge netop ét element på *n* forskellige måder.

For eksempel, kan man I en klasse med 24 elever udvælge en elev på 24 forskellige måder.

Skal man nu udvælge *både* en elev på x-klassen (24 elever) *og* en elev fra y-klassen (18 elever) ræsonnerer man som følger:
Hver gang man har udvalgt en elev fra x-klassen (24 muligheder), så har man 18 muligheder for at vælge en elev fra y-klassen. Adderer man de 18 muligheder 24 gange giver det i alt $24 \cdot 18 = 432$ muligheder.

- Dette er et eksempel på indholdet af *multiplikationsprincippet*: Skal man *både* udvælge et element fra en mængde med *n*-elementer *og* et element fra en anden mængde med *m*-elementer kan det gøres på $n \cdot m$ forskellige måder.

Multiplikationsprincippet kaldes også for *både- og* princippet, af den grund, at hvis man ved formuleringen af udvælgelsen kan sige *både…og,* så skal man *multiplicere* antallet af valgmuligheder ved de to valg.

Lad os nu antage, at man *enten* skal udvælge en elev fra x-klassen *eller* en elev fra y-klassen. I dette tilfælde er udvælgelsen af en elev blandt 24+18, så der er 24+18 = 42 muligheder.

- Dette er et eksempel på indholdet af *additionsprincippet*: Skal man *enten* udvælge et element fra en mængde med *n*-elementer *eller* et element fra en anden mængde med *m*-elementer kan det gøres på $n+m$ forskellige måder.

Additionsprincippet kaldes også for *enten - eller* princippet, af den grund, at hvis man ved formuleringen af udvælgelsen kan sige *enten...eller*, så skal man *addere* antallet af valgmuligheder ved de to valg.

Sker udvælgelsen fra flere end to mængder, skal man naturligvis blot multiplicere (eller addere) antallet af muligheder for udvælgelse fra hver af mængderne med hinanden.

Eksempler:

1. Skal man f.eks. vælge en menu (både forret, hovedret og dessert) fra et menu med 4 forretter,7 hovedretter og 6 desserter, kan det gøres på $4 \cdot 7 \cdot 6 = 168$ måder. Har man kun råd til en ret (enten forret, hovedret eller dessert) er der $4+7+6 = 17$ muligheder

2. En cykellås har 6 taster, der hver kan placeres i tre stillinger (Ud, Neutral, Ind). Hvor mange forskellige muligheder for lås giver det?
   Da vi *både* skal vælge hvordan den 1. tast skal stå *og* den anden, *og* den tredje..., og da der er 3 muligheder for hver giver multiplikationsprincippet $3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^6 = 729$ koder. (Den kode, hvor ingen af tasterne røres, er nok ikke tilrådelig i praksis)

3. En tipskupon udfyldes ved at sætte et kryds i et af tre felter, dette gentaget 13 gange. Hvor mange muligheder er der for at udfylde en kupon?.
   Ligesom ved cykellåsen har vi tre muligheder for hver kamp, og vi skal *både* vælge 1. *og* i 2. og... i 13. kamp. Ifølge multiplikationsprincippet er der derfor $3 \cdot 3 \cdot 3 \cdot ... \cdot 3 = 3^{13} = 1.594.323$ muligheder

.

2. Permutationer

Har man n-forskellige elementer i en mængde, (tænk på eleverne i en klasse med n=24), kaldes en *rækkefølge* på q-elementer (tænk på sidste række i klassen, hvor q=8) for en *q-permutation* af en n-mængde.
Vi ønsker at opstille en formel for antallet af forskellige q-permutationer af en n-mængde. Dette antal vil vi betegne med $P(n,q)$ eller $P_{n,q}$.

Først vil vi beregne $P(n,n)$ – altså antallet af forskellige rækkefølger af n elementer - som vi blot kalder antallet af permutationer af en n-mængde.

Vi tænker os altså at vi har n (=24) pladser og vi finder antallet af forskellige måder, hvorpå disse n pladser kan besættes med n elementer.

n	n-1	n-2	…						3	2	1

Da *både* den første *og* den anden *og* den tredje, osv. plads skal besættes, skal vi anvende multiplikationsprincippet.
Den første plads kan besættes på n (=24) måder. Når den er besat kan den næste plads besættes på n-1 (=23) måder. I alt kan de to første pladser ifølge multiplikationsprincippet besættes på $n(n-1)$ (= $24 \cdot 23$) måder. Den tredje på n-2 (=22) måder....den næstsidste på 2 måder og den sidste på 1 måde. I alt ifølge multiplikationsprincippet

$$P(n,n) = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 3 \cdot 2 \cdot 1 \qquad P(24,24) = 24 \cdot 23 \cdot 22 \cdot ... 3 \cdot 2 \cdot 1$$

For produktet af hele positive tal fra 1 til n bruger man betegnelsen $n!$ og det læses "*n-udråbstegn*" eller "*n-fakultet*". Af hensyn til gyldigheden af nogle senere formler definerer man 0! =1.
$n!$ kan udregnes på lommeregneren. F.eks. er $6! = 720$ og $10! = 3.628.800$. Vi har altså vist formlen:

$$P(n,n) = n!$$

Det er nu relativt simpelt at udlede en formel for $P(n,q)$, hvor $n \geq q$. Altså antallet af q-rækkefølger af en n-mængde. (q = 8, n = 24. Sidste række i klassen)

Første plads på n (=24) måder 2. plads på n-1 (=23) måder, 3. plads på n-2 (=22) måder...sidste plads på n-(q-1) = n-q+1 (= 17 = 24-(8-1)). Ifølge multiplikationsprincippet finder vi derfor:

$P(n,q) = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-q+1)$ og som eksempel $P(24,8) = 24\cdot 23\cdot 22\cdot\ldots 18\cdot 17 = 2{,}9654\cdot 10^{10}$

Første gang man ser det, overraskes man som regel over hvor stort et tal $P(24,8)$ egentlig er.

For at få en mere kompakt formel, multiplicerer man højresiden med $(n-q)!$ - med $16!$ i eksemplet $P(24,8)$ - i tæller og nævner. Hermed får man

$$P(n,q) = \frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-q+1)\cdot(n-q)!}{(n-q)!} = \frac{n!}{(n-q)!} \quad P(24{,}8) = \frac{24\cdot 23\cdot\ldots 17\cdot 16!}{16!} = \frac{24!}{16!}$$

Vi finder altså formlen:

$$P(n,q) = \frac{n!}{(n-q)!}$$

Bemærk, at med definitionen 0! = 1 bliver formlen også korrekt for $q = n$. $P(n,n) = n!/0! = n!$

3. Kombinationer

Har man en mængde A med n elementer, kaldet en n-mængde (tænk på en klasse med 24 elever), kan man udvælge en delmængde af A på q-elementer, kaldet en q-delmængde. (et udvalg på 4 elever).
En sådan q-delmængde er *ikke* en permutation, fordi rækkefølgen af elementerne i q-delmængden er underordnet. (Rækkefølgen af de 4 elever er underordnet).

En q-delmængde af en n-mængde kaldes også for en *kombination*

- Antallet af forskellige kombinationer med q-elementer udtaget af en n-mængde (antallet af forskellige udvalg på 4 elever, der kan dannes i en klasse med 24 elever) betegnes $K(n,q)$.

Vi ønsker at finde en formel for $K(n,q)$.

Vi bemærker, at det trivielt gælder, at $K(n,1) = n$ og $K(n,n) = 1$

Vi gennemfører først ræsonnementet med udvalget på 4 elever.
Udvalget skal - efter de er valgt - konstituere sig med en formand, en næstformand, en kasserer og en referent. Vi vil nu beregne antallet af muligheder for et udvalg på to forskellige måder.

Udvalget kan nedsættes på $K(24,4)$ forskellige måder, ifølge definitionen ovenfor, men vi kender foreløbig ikke tallet $K(24,4)$. Når et af de $K(24,4)$ forskellige udvalg er nedsat, kan det konstituere sig på $4\cdot 3\cdot 2\cdot 1$ måder (4 muligheder for formand, dernæst 3 for næstformand osv.). Det konstituerede udvalg kan derfor vælges på $K(24,4)\cdot 4\cdot 3\cdot 2\cdot 1 = K(24,4)\cdot 4!$ forskellige måder.

Vi kunne imidlertid også tænke os at udvalget blev nedsat ved direkte valg af formand, næstformand, osv. Da rækkefølgen nu er væsentlig – det er f.eks. et nyt udvalg, hvis man f.eks. bytter om på formand og referent - er antallet af muligheder $P(24,4)$.

Antallet af forskellige udvalg må være det samme, hvad enten det vælges direkte eller konstituerer sig. Derfor må der gælde.

$$K(24,4)\cdot 4\cdot 3\cdot 2\cdot 1 = P(24,4) \iff K(24,4)=\frac{P(24,4)}{4\cdot 3\cdot 2\cdot 1}=\frac{24!}{(24-4)!4!} = 10.626$$

Dette er den ønskede formel for $K(24,4)$. Ræsonnementet kan gennemføres ordret med n og q i stedet for 24 og 4.
Skal vi udvælge en q- ***permutation***, kan det gøres på $P(n,q)$ forskellige måder. Men vi kan også få en q-permutation ved først at udvælge en kombination (en q-delmængde), hvilket kan gøres på $K(n,q)$ forskellige måder og så permutere de q elementer. Der er $q!$ forskellige permutationer. Der må således gælde:

$$\boxed{K(n,q)\cdot q! = P(n,q) \iff K(n,q)=\frac{P(n,q)}{q!}=\frac{n!}{(n-q)!q!}}$$

Vi har nu opnået en generel formel for $K(n,q)$, som vi vil gøre en del brug af i det følgende.

Til praktisk beregning af $P(n,q)$ og især $K(n,q)$, anvender man ofte en formel, hvor man dividerer $(n\text{-}q)!$ op i tælleren: Dette giver formlerne:

$$\boxed{P(n,q)=n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-q+1) \qquad og \qquad K(n,q)=\frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-q+1)}{1\cdot 2\cdot 3\cdot\ldots\cdot q}}$$

Bemærk, at i den sidste formel er der det samme antal faktorer q i tæller og nævner.

Eksempler

1. Lotto. Når man spiller lotto, skal man vælge 7 tal blandt 36. Vi vil finde, hvor mange forskellige lottokuponer, der kan dannes. En lottokupon er det samme, som en 7-delmængde af en 36 mængde, så svaret er simpelthen

$$K(36,7)=\frac{36!}{(36-7)!7!}=8.347.680$$

2. en klasse med 10 drenge og 12 piger, skal der nedsættes et udvalg bestående af 2 drenge og 2 piger. På hvor mange måder kan udvalget nedsættes.
De to drenge kan vælges på K(10,2) = 45 måder og de to piger på K(12,2) = 66 måder. Ifølge multiplikationsprincippet, er antallet af forskellige udvalg herefter K(10,2) K(12,2) = 2970

3. Af en forsamling på 8 kvinder og 12 mænd, skal der nedsættes et udvalg på 5 medlemmer. På hvor mange måder kan det gøres når,
   a) Der kan vælges frit mellem de 20 medlemmer.
   b) Udvalget skal have mindst en kvinde og en mand.
   c) Udvalget skal have mindst to kvinder og to mænd.

a) K(20,5) = 15.504

b) Vi finder antallet ved at subtrahere de udvalg, som består af lutter mænd eller lutter kvinder:
K(20,5) – K(8,5) –K(12,5) = 15.504 – 56 – 792 = 14.656.

c) Der er kun 2 muligheder, enten er der 3 kvinder og 2 mænd eller 2 kvinder og 3 mænd. Vi anvender da såvel multiplikations- som additionsprincippet.

$K(12,3)\ K(8,2) + K(12,2)\ K(8,3) = 985$

Kap 2. Endeligt Sandsynlighedsfelt

1. Indledning

I dagligsprog anvender man ofte begrebet "sandsynligheden for…", og ofte synonymt med begrebet "chancen for…". Man angiver da sandsynligheden enten som en ægte brøk mellem 0 og 1 eller man angiver den i procent.
Når man anvender begrebet sandsynlighed, er det altid forbundet med en uvished om, hvorvidt en hændelse vil indtræffe eller ej. Chancen for, at man vinder på rouletten i et Casino. Sandsynligheden for, at togene kører uden forsinkelse til morgen. Sandsynligheden er udtryk for en vurdering af hvorvidt en hændelse vil indtræffe eller ej.

I matematikken er sandsynlighedsregningen en "model" af et "eksperiment", hvor udfaldet er delvis ukendt. I matematikken beskæftiger man sig ikke med at fastsætte sandsynligheder for et bestemt eksperiment, f.eks. sandsynligheden for at slå en sekser ved kast med en terning. Men så snart nogle primære sandsynligheder er fastsat, kan man drage en række vidtgående – og ofte overraskende – konsekvenser. Dette vil vi nu beskæftige os med i det følgende.
Først vil vi definere grundlaget for en matematisk model af et "tilfældigt eksperiment". Det kan for eksempel være kast med en terning eller at trække et kort fra en kortbunke.

2. Endeligt sandsynlighedsfelt

Et endeligt sandsynlighedsfelt (U,P) er en formalisering af et "tilfældigt eksperiment". Det består af en endelig mængde U og en funktion P med definitionsmængde U, og som har egenskaberne.

- $U = \{u_1, u_2, u_3, \ldots, u_n\}$ kaldes for *udfaldsrummet*, og elementerne i U betegnes *udfald*.
- $P(u)$ betegnes "Sandsynligheden for udfaldet u" (indtræffer).
- For alle $u \in U$ gælder: $0 \leq P(u) \leq 1$. samt at: $P(u_1) + P(u_2) + P(u_3) + \ldots + P(u_n) = 1$

Formuleret i ord, så er sandsynligheden for et udfald altid et tal mellem 0 og 1, og summen af sandsynlighederne for alle udfald i udfaldsrummet er 1.

Disse betingelser er de eneste, der skal være opfyldt for et endeligt sandsynlighedsfelt. Teorien beskæftiger sig derimod ikke med fastsættelsen af sandsynlighederne, eller hvorvidt de har noget som helst at gøre med udførelsen af et virkeligt eksperiment.

Eksempel

1. Kast med en terning. Udfaldsrummet er her U = {1,2,3,4,5,6}. Vi antager, at alle udfald har samme sandsynlighed. Ifølge den 2. betingelse, må der derfor gælde: 6 P(u) = 1 eller P(u) =1/6 for samtlige udfald.
2. Kast Med to mønter. Her kan vi skrive udfaldene, idet p=plat og k=krone: U = {(p,p), (p,k), (k,p), (k,k)} Idet vi antager, at disse 4 udfald har samme sandsynlighed, er sandsynligheden for hvert af udfaldene ¼.
3. Ved spil på en roulette i et casino er der 37 felter med numre 0…36. Udfaldsrummet er U = {0, 1, 2, 3…36}. Det er en almindelig antagelse, at sandsynligheden er den samme P(u) = 1/37 for at kuglen lander på et af disse 37 felter.

4. Kast med to terninger. (En rød og en grøn). Vi kan skrive udfaldene som (antal øjne på rød terning, antal øjne på grøn terning). Med denne notation, bliver udfaldsrummet U

   U = {(1,1), (1,2), ….(1,6), (2,1), (2,2),…(2,6)…..(6,5), (6,6)}

   Da der er 6 muligheder for, hvad den røde terning viser og for hver af disse 6 muligheder, 6 muligheder for den grønne terning, giver dette i alt 6· 6 = 36 udfald i udfaldsrummet. Vi antager, at alle disse udfald har samme sandsynlighed, og sandsynligheden for ethvert af udfaldene er derfor 1/36

3. Hændelser

En hændelse H er formelt defineret som en delmængde af udfaldsrummet. Hændelser skrives altid med store bogstaver. Udfald skrives med små bogstaver. Hvis et udfaldsrum er defineret ved $U = \{u_1, u_2, u_3, ..., u_n\}$, så kan hændelser være $A = \{u_2, u_7, u_8\}$, $B = \{u_4\}$ eller $C = \{u_1, u_7\}$.

Man formulerer ofte en hændelse i ord. I eksemplet med kast med en terning, kan man f.eks. om hændelsen. A: at man får et lige øjental, B: at man ikke får en sekser, C: at man får et øjental, som er større end 4. De 3 hændelser kan skrives: $A = \{2, 4, 6\}$, $B = \{1, 2, 3, 4, 5\}$ og $C = \{5,6\}$.

I sandsynlighedsregningen er opgaven ofte den at afgøre (og ofte ved hjælp af kombinatorik) , hvilke udfald, der tilhører en bestemt hændelse. I eksemplet ovenfor er det trivielt, men det er ofte langt fra tilfældet.
Rent Intuitivt, vil man nok i eksemplet ovenfor udregne sandsynlighederne for de 3 hændelser til:

$$P(A) = 3/6 = ½, \quad P(B) = 5/6 \text{ og } P(C) = 2/6 = 1/3$$

Hvilket også er korrekt, og som fører til følgende *definition*.

- Ved sandsynligheden for en hændelse $P(H)$ forstår man summen af sandsynlighederne for de udfald, som tilhører hændelsen H

  Hvis $U = \{u_1, u_2, u_3, ..., u_n\}$, og $A = \{u_2, u_7, u_8\}$, så er $P(A) = P(u_2) + P(u_7) + P(u_8)$

Det er let at godtgøre, at de udregnede sandsynligheder for hændelserne A, B og C i eksemplet med kast med en terning er i overensstemmelse med denne definition.

Da såvel udfaldsrummet U , som den tomme mængde $Ø$ er delmængder af udfaldsrummet fører det til to specielle hændelser.

- U kaldes for den *sikre* hændelse. $P(U)$ =1, ifølge definitionen
- $Ø$ kaldes for den *umulige* hændelse. $P(Ø)$ =0, ifølge definitionen

3.1 Brug af summationstegn

Når man opererer med hændelser, er det praktisk – for ikke at sige nødvendigt – at anvende det såkaldte summationstegn Σ. Summationstegnet er en forkortet skrivemåde for en sum af led.

Nedenfor er vist nogle eksempler på dette.

$$\sum_{n=2}^{10} \frac{1}{n} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... + \frac{1}{10} \qquad P(H) = \sum_{u \in H} P(u)$$

Det første summationstegn angiver summen af brøkerne 1/2 til 1/10. n=2 kaldes for nedre grænse og 10 for øvre grænse. Summen udregnes ved at sætte n=2 i udtrykket der står efter (man siger "under") summationstegnet og sætte et plustegn, derefter sætte n=3, og sådan fremdeles, indtil n=10.

I det andet udtryk er der ikke angivet nedre og øvre grænse. I stedet skal det forstås således, at man summerer sandsynlighederne over alle udfald som tilhører *H*. Hvilket netop er sandsynligheden for hændelsen *H*.

3.2 Symmetrisk sandsynlighedsfelt

Vi har i eksemplerne ovenfor betragtet udfaldsrum, hvor alle udfald har samme sandsynlighed. Et sådant sandsynlighedsfelt kaldes *symmetrisk*.
Hvis der er n udfald i et symmetrisk udfaldsrum, må sandsynligheden for ethvert udfald være $\frac{1}{n}$, da summen af de n sandsynligheder skal være 1.

- I et symmetrisk sandsynlighedsfelt med n-udfald er $P(u) = \frac{1}{n}$

I et symmetrisk sandsynlighedsfelt er det specielt simpelt at finde sandsynligheden for en hændelse *H*. Ifølge definitionen, skal man nemlig addere sandsynligheden for de elementer, som tilhører hændelsen. Hvis udfaldsrummet har n elementer, og hændelsen *H* har q elementer, skal man udregne summen:

$$P(H) = \sum_{u \in H} P(u) = \frac{1}{n} + \frac{1}{n} + \frac{1}{n} + ... \frac{1}{n} = \frac{q}{n}$$

- Man udregner sandsynligheden for en hændelse, som antallet af udfald i hændelsen divideret med antallet af udfald i udfaldsrummet.

I denne forbindelse anvender man en bestemt sprogbrug. Udfaldene i udfaldsrummet, kaldes de *mulige* udfald. Et udfald, som tilhører hændelsen kaldes et *gunstigt* udfald (for hændelsen). Man kan da opskrive sandsynligheden for en hændelse på en måde, som er let at huske.

$$P(H) = \frac{\text{gunstige udfald}}{\text{mulige udfald}} \quad \text{eller blot} \quad P(H) = \frac{\text{gunstige}}{\text{mulige}} \quad \text{eller} \quad P(H) = \frac{n(H)}{n(U)}$$

I det sidste udtryk læses $n(H)$ og $n(U)$ som antallet af udfald i henholdsvis *H* og *U*.

Eksempler

1. Vi tænker os, at vi trækker et kort fra et almindeligt kortspil med 52 kort. Vi vil bestemme sandsynligheden for følgende hændelser: A: Vi trækker en spar. B: vi trækker et billedkort. C: Vi trækker et es.
   Idet n(U) = 52, og for de 3 hændelser gælder: n(A) = 13, n(B) = 12 og n(C) = 4, finder man sandsynlighederne:

$$P(A) = \frac{n(A)}{n(U)} = \frac{13}{52} = \frac{1}{4} \qquad P(B) = \frac{n(B)}{n(U)} = \frac{12}{52} = \frac{3}{13} \qquad P(C) = \frac{n(C)}{n(U)} = \frac{4}{52} = \frac{1}{13}$$

2. Vi ser nu på kast med to terninger (en grøn og en rød). Vi ønsker at finde sandsynlighederne P(2), P(3), …., P(12) for at summen af øjentallene er 2, 3, 4, …,12. Vi har før set på udfaldsrummet. Det kan opskrives som følger:

   U = {(1,1), (1,2), ….(1,6), (2,1), (2,2),…(2,6)…..(6,5), (6,6)}.

   Udfaldsrummet er symmetrisk, består af 36 elementer og sandsynligheden for et af disse udfald er 1/36. Beregningen af sandsynlighederne P(2),…,P(12), går derfor blot ud på at optælle, hvor mange udfald, der er i hver af hændelserne. Heraf fås:

$P(2) = \frac{1}{36}$	Udfaldet {(1,1)}
$P(3) = \frac{2}{36} = \frac{1}{18}$	Udfaldene {(1,2), (2,1)}
$P(4) = \frac{3}{36} = \frac{1}{12}$	Udfaldene {(1,3), (3,1), (2,2)}
$P(5) = \frac{4}{36} = \frac{1}{9}$	Udfaldene {(1,4), (4,1), (2,3), (3,2)}
$P(6) = \frac{5}{36}$	Udfaldene {(1,5), (5,1), (2,4), (4,2),(3,3)}
$P(7) = \frac{6}{36} = \frac{1}{6}$	Udfaldene {(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)}

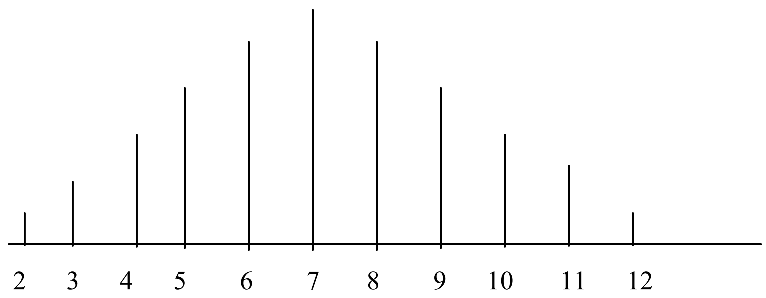
   På helt tilsvarende måder finder man sandsynlighederne:

$$P(8) = \frac{5}{36}, \quad P(9) = \frac{4}{36} = \frac{1}{9}, \quad P(10) = \frac{3}{36} = \frac{1}{12}, \quad P(11) = \frac{2}{36} = \frac{1}{18} \quad \text{og} \quad P(12) = \frac{1}{36}$$

   Vi kan nu definere et nyt (ikke symmetrisk) sandsynlighedsfelt) med udfaldsrum U = {2,3,4,…,12} og sandsynligheder givet ved tabellen:

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

   Endelig kan man tegne et pindediagram over sandsynlighedsfordelingen.



3. **Fødselsdagsproblemet:**
   Dette er et klassisk eksempel, som findes i de fleste lærebøger om sandsynlighedsregning. Hvis man har en klasse på 24 elever. Hvad er så sandsynligheden for at mindst to elever har fødselsdag på samme dag. Mere generelt, hvis man har n personer, hvor n<365, hvad er så sandsynligheden for at mindst to har fødselsdag på

samme dag. I tilfældet n=24 er det interessant (f.eks. med henblik på et væddemål) at undersøge, hvorvidt denne sandsynlighed er større end ½.
Vi bestemmer først antallet af elementer i udfaldsrummet. Da hver person har 365 muligheder for fødselsdag er der $365 \cdot 365 \cdot 365 \cdot \ldots \cdot 365 = 365^n$ muligheder for placering af fødselsdage på de n personer.
For at udregne den ønskede sandsynlighed, udregner vi i stedet sandsynligheden P(H) for den komplementære hændelse, altså at de n personer alle har fødselsdag på forskellige dage. Sandsynligheden for mindst et sammenfald, kan så udregnes som 1-P.
De gunstige udfald for H er $365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365-n+1)$ (n-faktorer). 365 muligheder for den første, 364 for den anden osv. P(H) udregnes da som tidligere, som antal gunstige udfald divideret med antal mulige udfald.

$$P(H) = \frac{365 \cdot (365-1) \cdot (365-2) \cdot \ldots (365-n+1)}{365 \cdot 365 \cdot 365 \cdot 365 \cdot \ldots \cdot 365} = \frac{365}{365} \cdot \frac{365-1}{365} \cdot \frac{365-2}{365} \cdot \ldots \frac{365-n+1}{365}$$

Med lidt tålmodighed, kan man på en lommeregner udregne dette udtryk for n = 24. Man finder at sandsynligheden er lidt mindre end ½, o dermed, at der (lidt overraskende) er mere end 50% chance for sammenfald af fødselsdage i en klasse med 24 elever. For n=22 er der derimod mindre end 50% chance for sammenfald af fødselsdage.
Selv om det er lidt over det faglige niveau for disse noter, vil vi vise, hvorledes man simplere kan beregne sandsynligheden ovenfor. For at udlede formlen dividerer vi 365 op i tælleren i hver af brøkerne i udtrykket ovenfor og tager den naturlige logaritme til begge sider.

$$P(H) = \frac{365}{365} \cdot \frac{365-1}{365} \cdot \frac{365-2}{365} \cdot \ldots \frac{365-n+1}{365} = 1 \cdot (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdot \ldots (1 - \frac{n-1}{365})$$

$$\ln P(H) = \ln(1 - \frac{1}{365}) + \ln(1 - \frac{2}{365}) + \ldots + \ln(1 - \frac{n-1}{365})$$

For ln funktionen gælder der følgende tilnærmelsesformel, når h er numerisk meget mindre end 1 (mindre end 1/10) ln(1+h) = h. Anvendes denne formel på de n-1 led finder man:

$$\ln P(\overline{H}) = -\frac{1}{365} - \frac{2}{365} - \frac{3}{365} - \ldots \frac{n-1}{365} = -\frac{1}{365}(1 + 2 + 3 + \ldots(n-1))$$

$$\ln P(\overline{H}) = -\frac{1}{365} \frac{n(n-1)}{2}$$

Vi søger da at bestemme n således at P(H) < ½ ⇔ ln P(H) < ln½ ⇔ ln P(H) < -0,6931
Dette giver uligheden

$$-\frac{1}{365} \frac{n(n-1)}{2} < -0{,}6931 \Leftrightarrow n(n-1) > 505{,}963$$

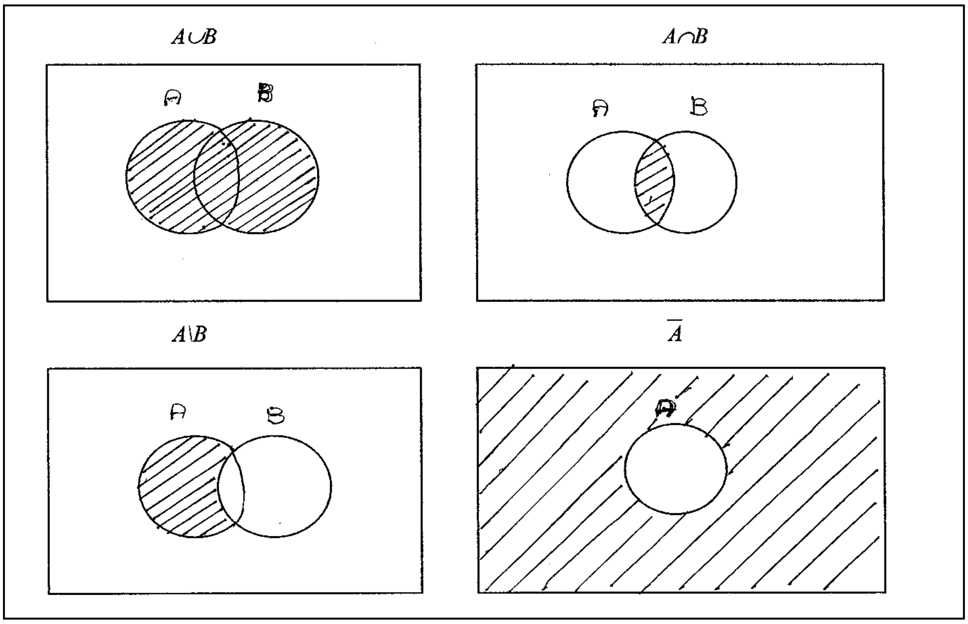
Udregner man nu n(n-1) for n=21, 22,23, 24 finder man: 420, 462, 506, 552.
Konklusionen er altså, at der skal være mindst 23 elever i en klasse for at sandsynligheden for sammenfald af fødselsdage er større end ½. Udregner man sandsynligheden finder man: 1-P(H) = 0,5305.

4. Regning med hændelser og sandsynligheder

For to delmængder A og B af en mængde U kan man som bekendt danne nogle nye mængder:

- *Foreningsmængden* $A \cup B$, som er de elementer, som *enten* tilhører A *eller* tilhører B.
- *Fællesmængden* $A \cap B$, som er de elementer, som *både* tilhører A *og* B.
- *Differensmængden* $A \backslash B$, som er de elementer i A, som ikke tilhører B.
- *Komplementærmængden* til A, $\overline{A}$ som er de elementer i U, som *ikke* tilhører A.

Nedenfor er de 4 begreber illustreret grafisk.



Hvis A og B er hændelser I et udfaldsrum U, svarer de 4 mængder: $A \cup B$, $A \cap B$, $A \setminus B$ og $\overline{A}$ til hændelser. Man udtrykker dette på følgende måde:

Forenings-hændelsen $A \cup B$, indtræffer, hvis mindst en af hændelserne A og B indtræffer

Fælles-hændelsen $A \cap B$, indtræffer, hvis både A og B indtræffer samtidig.

Differens-hændelsen $A \setminus B$, indtræffer, hvis A indtræffer, og B ikke indtræffer.

Komplementær-hændelsen $\overline{A}$, indtræffer, hvis A ikke indtræffer.

Der gælder nogle regneregler for sandsynligheder for hændelser. Vi opskriver først disse sandsynligheder ved hjælp af summationstegn, der jo som nævnt betyder, at vi adderer sandsynlighederne for de udfald, som tilhører en hændelse.

$$P(A) = \sum_{u \in A} P(u) \ , \quad P(B) = \sum_{u \in B} P(u) \ , \quad P(A \cup B) = \sum_{u \in A \cup B} P(u) \ , \quad P(A \cap B) = \sum_{u \in A \cap B} P(u) \ ,$$
$$P(A \setminus B) = \sum_{u \in A \setminus B} P(u) \ , \quad P(\overline{A}) = \sum_{u \in CA} P(u)$$

Vi kan så udlede følgende regneregler:

$$P(A \cup B) = \sum_{u \in A \cup B} P(u) = \sum_{u \in A} P(u) + \sum_{u \in B} P(u) - \sum_{u \in A \cap B} P(u)$$

Når vi skal summere over antallet af udfald i foreningsmængden, ved at summere over elementerne i A plus elementerne i B, så tæller vi elementerne i fællesmængden af A og B to gange, hvorfor vi må trække dem fra. Vi får derfor formlen:

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

Sandsynligheden for at *enten A eller B* indtræffer, er sandsynligheden for at *A* indtræffer plus sandsynligheden for at *B* indtræffer minus sandsynligheden for at de indtræffer samtidig.

Hvis $A \cap B = \emptyset$, altså hvis fællesmængden er tom, siges de to hændelser at udelukke hinanden. I dette tilfælde bliver formlen simplere, idet $P(\emptyset) = 0$.

$$P(A \cup B) = P(A) + P(B)$$

På helt tilsvarende vis finder man:

$$P(\overline{A}) = \sum_{u \in CA} P(u) = \sum_{u \in U} P(u) - \sum_{u \in A} P(u)$$

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

Sandsynligheden for at A *ikke* indtræffer er 1 minus sandsynligheden for at den indtræffer.

$$P(A \setminus B) = \sum_{u \in A \setminus B} P(u) = \sum_{u \in A} P(u) - \sum_{u \in A \cap B} P(u)$$

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

Sandsynligheden for at *A* indtræffer, men *B* indtræffer ikke, er sandsynligheden for at *A* indtræffer – minus sandsynligheden for at *A* og *B* indtræffer samtidig.

Eksempler

1. Vi betragter først eksemplet, hvor vi trækker et kort fra en kortbunke. Vi vil finde sandsynligheden for følgende hændelser: A: Man trækker et billedkort eller en spar. B: Man trækker ikke et billedkort. C: Man trækker et billedkort, men ikke en hjerter. Ifølge formlerne ovenfor gælder:

   P(A) = P(Billedkort) + P(spar) – P(Billedkort i spar) = $\frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{22}{52}$

   P(B) = 1 – P(Billedkort) = $1 - \frac{12}{52} = \frac{40}{52} = \frac{10}{13}$

   P(C) = P(Billedkort) - P(Billedkort i hjerter) = $\frac{12}{52} - \frac{3}{52} = \frac{9}{52}$

2. Vi betragter dernæst eksemplet med samtidig kast med to terninger. Vi ser på følgende hændelser: A: De to terninger viser samme øjental. B: De to terninger viser forskellige øjental. C: Den røde terning viser mere end den grønne. D: Vi slår ikke en sekser. D: Vi slår mindst en sekser.

   $$P(A) = \frac{6}{36} = \frac{1}{6} \quad \text{og} \quad P(B) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$$

   I nøjagtig halvdelen af de 36-6 udfald, hvor terningerne viser forskelligt, viser den røde mere end den grønne, så svaret er:

   $$P(C) = \frac{1}{2} P(B) = \frac{5}{12} \left(= \frac{15}{36}\right)$$

Der er 5·5 =25 udfald, hvor ingen af terninger viser 6 øjne. Derfor er P(D) = $\frac{25}{36}$.

$$P(E) = 1 - P(D) = 1 - \frac{25}{36} = \frac{11}{36}$$

Stikprøveudtagning uden tilbagelægning. Dette er en klassisk disciplin indenfor sandsynlighedsregningen, som oftest bliver formuleret ved at man tager et antal kugler af forskellige farver fra en urne, og beregner sandsynligheden for forskellige farvekombinationer. Vi skal se på to eksempler på en sådan problematik.

3. I en klasse er der 10 drenge og 12 piger, og der skal nedsættes et udvalg på 4 elever. Vi vil bestemme sandsynligheden for følgende hændelser. A: udvalget består af lutter piger. B: Udvalget består af to drenge og to piger. C: Der er mindst en dreng i udvalget.
   For at udregne disse sandsynligheder, vil vi erindre om formlen for sandsynligheden for en hændelse H

   $$P(H) = \frac{Gunstige\ \ udfald}{Mulige\ \ udfald}$$

   De mulige udfald er K(22,4) = $\frac{22!}{4!(22-4)!} = \frac{22 \cdot 21 \cdot 20 \cdot 19}{4 \cdot 3 \cdot 2 \cdot 1} = 11 \cdot 7 \cdot 5 \cdot 19 = 7.315$

   De 4 piger kan vælges på K(12,4) = 495 måder. Herefter får vi:

   $$P(4\ piger) = P(A) = \frac{K(12,4)}{K(22,4)} = \frac{495}{7315} = 0.0677$$

   Helt på samme måde finder man:

   $$P(2\ drenge+2piger) = P(B) = \frac{K(10,2) \cdot K(12,2)}{K(22,4)} = \frac{45 \cdot 66}{7315} = 0{,}4060$$

   P(Mindst en dreng) = P(C) = 1-P(4 piger) = 1 - 0,0667 = 0,9333

4. I en skuffe ligger der uordnet 6 blå, 8 grå og 10 sorte sokker. En person tager (uden at kontrollere farven) 3 sokker (i det håb) at få to af samme farve. Vi vil udregne sandsynlighederne for hændelserne. A: Man får 3 forskellige sokker. B: Man får enten 3 blå eller 3 grå eller 3 sorte sokker. C: Man får mindst 2 sokker i samme farve.

   De mulige udfald er K(6+8+10,3) = K(24,3) = $\frac{24 \cdot 23 \cdot 22}{3 \cdot 2 \cdot 1} = 4 \cdot 23 \cdot 22 = 2024$

   $$P(3\ forskellige) = \frac{6 \cdot 8 \cdot 10}{K(24,3)} = \frac{480}{2024} = 0{,}2372$$

   $$P(3\ i\ samme\ farve) = \frac{K(6,3)+K(8,3)+K(10,3) \cdot}{K(24,3)} = \frac{20+56+120}{2024} = \frac{196}{2024} = 0{,}0968$$

   P(Mindst 2 i samme farve) = 1 - P(3 forskellige) = 1 - 0,2372 = 0,7628

5. Sandsynligheder ved lottospil. Der vælges 7 tal ud af 36 mulige. Der udtrækkes på tilfældig måde 7 vindertal og 2 tillægstal. Man får præmie for 7 rigtige, 6 rigtige plus et tillægstal, 6rigtige, 5 rigtige og 4 rigtige.
   Følgelig er sandsynligheden for 7 rigtige:

   $$P(7) = \frac{1}{K(36,7)} = p_7 = 1{,}1979 \cdot 10^{-7}$$

   Når vi skal udregne sandsynligheden for 6 rigtige plus et tillægstal, ræsonnerer vi på følgende måde:

De 6 rigtige kan vælges ud af 7 på K(7,6) = 7 forskellige måder, og tillægstallet kan vælges på 2 måder.
Da vi både skal have 6 rigtige og et tillægstal rigtigt, skal de to tal multipliceres for at finde antal gunstige udfald.

P(6 rigtige + 1tt) = K(7,6)*2/K(36,7)=14/K(36,7)=14*p_7 = $1{,}677*10^{-6}$

På samme måde kan vi finde sandsynligheden for 6 rigtige, idet antal gunstige er K(7,6) gange med antal måder det forkerte tal kan vælges på: 36 –7 –2 =27 (det må hverken være et af de 7 rigtige eller et tillægstal).

P(6 rigtige) =K(7,6)*27/K(36,7) = 189*p_7 = $2{,}624*10^{-5}$

Sandsynligheden for 5 rigtige findes som antallet af måder at udtage 5 rigtige ud af 7 lig med K(7,5), gange antallet af muligheder for udvælgelse af de to sidste tal, som kan vælges blandt 36-7 =29 tal.
Dette antal er K(29,2).

P(5 rigtige) = K(7,5)*K(29,2)/K(36,7) = 8526*p_7 = $1{,}02*10^{-3}$ =1,02 ‰

På helt samme måde opskriver vi sandsynligheden for 4 rigtige, idet de gunstige er K(7,4) *K(29,3) 4 rigtige ud valgt blandt 7 gange 3 forkerte udvalgt blandt 29.

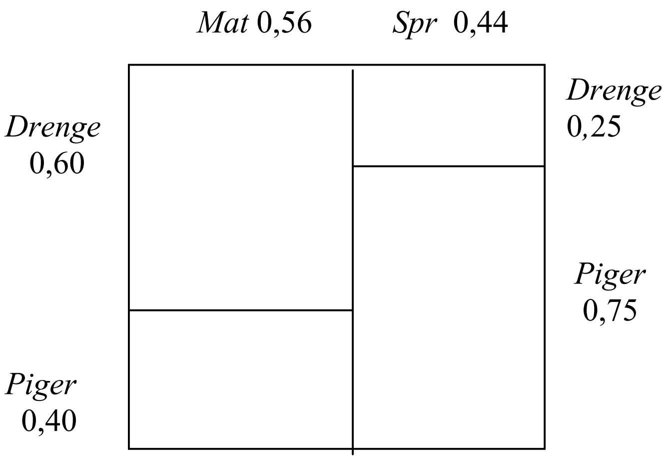
P(4 rigtige) = K(7,4)*K(29,3)/K(36,7) = 127.890*p_7 = 0,0153 = 1,53%

Af dette fremgår, at chancen for at få mere end 4 rigtige er uhyre ringe. Til gengæld er der en rimelig chance for at få 4 rigtige. Dette er helt bevidst for dem der har planlagt spillet. Erfaringen viser nemlig, at hvis man aldrig vinder, holder man op med at spille efter en vis tid.

Kap 3. Betinget sandsynlighed

1. Definition af betinget sandsynlighed

På et gymnasium er 56% af eleverne matematikere og 44% sproglige. Blandt matematikere er 60 % drenge, mens der blandt de sproglige er 25% drenge. Dette er søgt illustreret nedenfor.



Vi betragter det eksperiment, at man på tilfældig måde udtager en elev, og vi indfører nogle betegnelser for hændelser

M: Der udtrækkes en matematiker.
S: Der udtrækkes en sproglig.
D: Der udtrækkes en dreng.
K: Der udtrækkes en pige.

Vi kan da umiddelbart opskrive nogle sandsynligheder: $P(M) = 0{,}56$ og $P(S) = 0{,}44$.
Sandsynlighederne $P(D)$ og $P(K)$ kan vi imidlertid ikke umiddelbart opskrive.
Udtrækkes en elev kun blandt matematikerne er sandsynligheden for at udtrække en dreng imidlertid 0,60 ifølge ovenstående. Dette kan vi udtrykke således:
Sandsynligheden for at udtrække en dreng, når det er givet, at det er en matematiker er 0,60.

Dette kalder man i matematikken for den *betingede sandsynlighed for dreng, givet matematiker*. Sandsynligheden skrives på følgende måde:

$$P(D \mid M) = 0{,}60$$

Den lodrette streg mellem de to hændelser læses som "givet". $P(D \mid M)$ læses da som: "S*andsynligheden for dreng, givet matematiker*".
Helt på den samme måde kan vi skrive: $P(K \mid M) = 0{,}40$, $P(D \mid S) = 0{,}25$ og $P(K \mid S) = 0{,}75$

Vi skal nu se på hvorledes man kan udregne $P(D \mid M)$ i et symmetrisk sandsynlighedsfelt. Som sædvanlig betegner vi med $n(H)$ antallet af elementer i en mængde H. (= antal udfald i en hændelse H)

Hændelsen "En matematisk dreng" er $D \cap M$. $P(D \mid M)$ udregnes da som sædvanlig som gunstige/mulige, altså som: Antallet af matematiske drenge divideret med antallet af matematikere.

$$P(D \mid M) = \frac{n(D \cap M)}{n(M)}$$

Dividerer vi nu i denne formel i tæller og nævner med $n(U)$ = antallet af udfald i udfaldsrummet fås:

$$P(D \mid M) = \frac{n(D \cap M)}{n(M)} = \frac{\frac{n(D \cap M)}{n(u)}}{\frac{n(M)}{n(U)}} = \frac{P(D \cap M)}{P(M)}$$

Vi definerer nu for to vilkårlige hændelser A og $H \neq Ø$ i et vilkårligt (ikke nødvendigvis symmetrisk) sandsynlighedsfelt: Den betingede sandsynlighed for hændelsen *A givet H* (dvs. når H er indtruffet), som:

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)}$$

Sandsynligheden for A: *givet H* er sandsynligheden for at både A og H indtræffer, divideret med sandsynligheden for at H indtræffer.
Ofte anvender man formlen ved at gange igennem med $P(H)$.

$$P(A \cap H) = P(A \mid H)\, P(H)$$

Da $A \cap H = H \cap A$, gælder der åbenbart $P(A \cap H) = P(H \cap A)$ og hermed

$$P(A \mid H)\, P(H) = P(H \mid A)\, P(A) \quad \Leftrightarrow \quad P(A \mid H) = \frac{P(H \mid A) P(A)}{P(H)}$$

Denne formel kaldes for Bayes formel. Rent formelt er Bayes formel helt enkel, men det underlige er, at der i formlen bliver byttet om på kausaliteten (hvad der er årsagen til hvad) af de to hændelser. Det er nemlig meningsløst at tale om sandsynligheden for en hændelse, som er indtruffet.

Hvis f.eks. sandsynligheden for en ulykke er betinget af glat føre, så kan man ved formlen udregne sandsynligheden for at det har været glat føre, hvis ulykken er indtruffet. Pointen er så den, at det faktisk er kendt om det var glat føre på ulykkestidspunktet, og at det derfor er meningsløst, at tale om sandsynligheden for at det indtræffer. Statistisk set, vil det dog nok passe, at den dertil svarende procent af alle ulykker er sket i glat føre.

Eksempler

1. Der trækkes et kort fra et kortspil. Det oplyses, at det er en hjerter (H). Find sandsynligheden for at det er et billedkort (B). Opgaven kan løses på to måder: Antallet af billedekort i hjerter er 3 og der er 13 hjerter, derfor er sandsynligheden: P(B | H) = $\frac{3}{13}$. Opgaven kan imidlertid også løse ved formlen ovenfor:

$$P(B|H)=\frac{P(B\cap H)}{P(H)}=\frac{\frac{3}{52}}{\frac{13}{52}}=\frac{3}{13}$$

2. En familie har 3 børn.
   a) Find sandsynligheden for, at der blandt børnene er en dreng og en pige (A)
   Der er i alt $2^3 = 8$ muligheder. (På samme måde som 3 kast med en mønt). Der er 2 udfald (lutter drenge eller lutter piger), hvor det ikke er tilfældet. Sandsynligheden er derfor P(A) = $\frac{6}{8} = \frac{3}{4}$.
   b) Samme spørgsmål, når det oplyses, at de har en pige (P). Vi skal altså bestemme sandsynligheden P(A | P)

$$P(A \mid P) = \frac{P(A\cap P)}{P(P)} = \frac{\frac{6}{8}}{\frac{7}{8}} = \frac{6}{7}$$

   De to sandsynligheder følger af at A∩P = A og n(A) = 6, samt n(P) = 7 (8-lutter drenge).

3. Dette eksempel, angår egentlig ikke betinget sandsynlighed. Det er derimod en klassisk gåde, som også har været stillet i matematiklærernes fagblad, og som ofte giver anledning til fejlslutninger. Det kan formuleres som følger. Der er 3 ens æsker. De to af dem er tomme, mens der er en guldmønt i den tredje. En person bliver bedt om at vælge en æske. Efter dette valg er foretaget, bliver en af de to resterende æsker, men som ikke indeholder guldmønten, fjernet.
   Spørgsmålet er da: Skal man vælge om? Skal man ikke vælge om? Er det underordnet om man vælger om eller ej?
   De fleste er måske tilbøjelige til at svare, at det er underordnet om man vælger om eller ej, men dette er forkert af følgende grund. Ved det første valg har man udpeget æsken med mønten sandsynligheden 1/3, og følgelig er sandsynligheden for at mønten befinder sig i en af de to andre æsker 2/3. Disse sandsynligheder ændres *ikke* ved, at man fjerner en af de to tomme æsker, så sandsynligheden for at mønten befinder sig i en af de to æsker, som ikke er valgt i første omgang er fortsat 2/3. Da man kun kan vælge den ene er sandsynligheden for at mønten befinder sig der stadig 2/3. Man skal følgelig *altid* vælge om. Sandsynligheden for at vælge æsken med guldmønten er dobbelt så stor, som hvis man ikke vælger om.

2. Udvikling på hændelser

Vi mangler stadig, at besvare nogle spørgsmål fra det indledende eksempel med eleverne i et gymnasium. Vi vil nu finde sandsynligheden for at den udtrukne elev er en dreng. Vi kender sandsynlighederne $P(D \mid M) = 0{,}60$ og $P(D \mid S) = 0{,}25$ samt sandsynlighederne $P(M) = 0{,}56$ og $P(S) = 0{,}44$.
Vi "udvikler" nu hændelsen *D* på de to hændelser *M* og *S*:

$$D = (D\cap M)\cup(D\cap S)$$

Drengene er de som går i matematisk gymnasium samt de som går i sproglig gymnasium.
Da de to hændelser: $(D\cap M)$ og $(D\cap S)$ er disjunkte (de har ingen fælles elementer), kan vi anvende den forkortede version af regning med hændelser:

$$P(D) = P(D\cap M) + P(D\cap S)$$

På højresiden af udtrykket ovenfor anvender vi dernæst formlen for betinget sandsynlighed:

$$P(D) = P(D \mid M)\cdot P(M) + P(D \mid S)\cdot P(S)$$

Denne formel kan kun anvendes, når $M \cup S = U$ og $M \cap S = \emptyset$. Man siger i dette tilfælde at M og S udgør en klassedeling af U. Mere generelt siger man at mængderne $H_1, H_2, H_3, \ldots, H_n$ udgør en klassedeling af en mængde U, hvis $U = H_1 \cup H_2 \cup H_3 \cup \ldots \cup H_n$ og $H_i \cap H_j = \emptyset$ for alle $i \neq j$.

Den generelle formel bliver herefter:

$$P(A) = P(A \mid H_1)\cdot P(H_1) + P(A \mid H_2)\cdot P(H_2) + P(A \mid H_3)P(H_3) + \ldots + P(A \mid H_n)P(H_n)$$

Eksempler

1. Vi kan nu jf. ovenfor beregne P(D) = 0,60·0,56+0,25·0,44 = 0,446

2. Der udtrækkes en elev. Det oplyses, at det er en dreng. Find sandsynligheden for at det er en sproglig. Vi ønsker altså at bestemme sandsynligheden P(S | D). Vi kender kun den omvendte sandsynlighed, så vi anvender Bayes formel:
$$P(S \mid D) = \frac{P(D|S)\cdot P(S)}{P(D)} = \frac{0{,}25\cdot 0{,}44}{0{,}446} = 0{,}247$$

3. Der er 75% sandsynlighed for at en spiller lyver. Han kaster en terning, og bliver spurgt om det var en sekser. a) Hvad er sandsynligheden for at har svarer ja? b) Hvad er sandsynligheden for at det var en sekser, hvis han svarede ja. Vi udvikler på hændelserne "sekser" "Ikke sekser"

   P(ja) = P(ja | Ikke sekser) P(Ikke sekser) + P(ja | sekser) P(sekser)

   P(ja) = P(Han lyver) P(Ikke sekser) + P(Han lyver ikke) P(sekser)

   a) $P(ja) = \frac{3}{4}\cdot\frac{5}{6} + \frac{1}{4}\cdot\frac{1}{6} = \frac{16}{24} = \frac{2}{3}$

   b) $P(Sekser \mid ja) = \frac{P(ja|Sekser)P(Sekser)}{P(ja)} = \frac{\frac{1}{4}\cdot\frac{1}{6}}{\frac{2}{3}} = \frac{1}{16}$

3. Uafhængige hændelser

Vi har ikke endnu givet en generel formel til beregning af sandsynligheden for "både og hændelsen", altså sandsynligheden $P(A \cap B)$, når $P(A)$ og $P(B)$ er kendte. De fleste vil mene, at man nok skal multiplicere de to sandsynligheder,

- men dette er *kun* korrekt hvis de to hændelser A og B er uafhængige af hinanden.

Eksempel

Hvis man trækker et kort fra et kortspil og ser på hændelserne A: Et billedkort. B: Et rødt billedkort. C: En hjerter. Så er såvel $A \cap C$, som $B \cap C$: Et billedkort i hjerter.
Da der er 3 billedkort i hjerter er $P(A \cap C) = P(B \cap C) = 3/52$. Imidlertid er :
$P(A)\cdot P(C) = 12/52\cdot 13/52 = 12/52\ 1/4 = 3/52 = P(A \cap C)$, mens
$P(B)\cdot P(C) = 6/52\cdot 13/52 \neq P(B \cap C)$, så de to hændelser A og C er uafhængige af hinanden, mens B og C afhænger af hinanden, hvilket vel også er klart nok. (Der er det samme antal billedkort i alle farver)

Vender vi imidlertid tilbage til definitionen af betinget sandsynlighed, så gælder der:

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

Hvilket er den korrekte måde at udregne $P(A \cap B)$. Begrebet uafhængige hændelser, hviler på følgende definition.

- En hændelse A siges at være uafhængig af hændelsen B, hvis $P(A) = P(A \mid B)$.

Indsættes dette i udtrykket for betinget sandsynlighed, finder man:

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

Ofte kan det være svært at gennemskue om to hændelser er uafhængige af hinanden eller ej.
Spørgsmålet kan da afgøres ved at udregne $P(A \cap B)$ og $P(A)\, P(B)$ og sammenligne.
Da formlen $P(A \cap B) = P(A)\, P(B)$ er symmetrisk i A og B (samme formel, hvis man ombytter A og B), så følger det, at hvis A er uafhængig af B, så er B også uafhængig af A.
Man siger derfor, at de to hændelser er *uafhængige*.

Eksempler

1. Det er en udbredt fejlslutning, at hvis man f.eks. kaster en mønt, og har slået plat 5 gange, så er der større sandsynlighed for at slå krone i næste kast. Det er ikke tilfældet. Ethvert kast er uafhængigt af de foregående og har samme sandsynlighed for at give krone. Noget helt tilsvarende gør sig gældende, hvis man spiller på rød eller sort i et casino eller hvis man mener, at der er større sandsynlighed for at få en dreng, hvis man har 3 piger i forvejen.
   Det er derimod korrekt, at der er (langt) større sandsynlighed for at få krone mindst en gang i en serie på 5 kast med en mønt end ingen krone.
   Helt præcist gælder der: P(5 plat) = $\frac{gunstige}{mulige} = \frac{1}{2^5} = \frac{1}{32}$ P(Mindst 1 krone) = 1 – P(5 plat) = $1 - \frac{1}{32} = \frac{31}{32}$

2. Sandsynligheden for at få en sekser er 1/6. I gennemsnit skulle man derfor få en sekser i 6 forsøg. Vi vil bestemme sandsynligheden for ikke at få en sekser i 6 kast. P(ej sekser) = 5/6. Da de 6 kast er uafhængige af hinanden, skal vi blot multiplicere 5/6 med sig selv 6 gange. P(ej sekser i 6 kast) = $\left(\frac{5}{6}\right)^6 = 0{,}335$

3. Det kan godt være svært umiddelbart at gennemskue, hvorvidt to hændelser er uafhængige eller ej. I forbindelse med 2 kast med en terning, betragter vi følgende hændelser: A: Første kast viser 4 øjne. B: Summen af de to øjental er 7. C: Summen af de to øjental er 5. Vi finder:

   $$P(A) = \frac{1}{6} \quad P(B) = \frac{6}{36} = \frac{1}{6} \quad P(C) = \frac{4}{36} = \frac{1}{9} \quad P(A \cap B) = \frac{1}{36} \quad P(A \cap C) = \frac{1}{36}$$

   Af dette slutter vi, at A og B er uafhængige, mens A og C afhænger af hinanden.

Kap 4. Binomialfordelingen

1. Gentagelse af et eksperiment

Vi vil nu se på den situation, at man gentager et eksperiment et vist antal gange, og hvor hver udførelse er uafhængig af de foregående. Ved hver udførelse, vil vi kun interesserer os for, hvorvidt en bestemt hændelse indtræffer eller ej.
Helt konkret vil vi betragte 12 kast med en terning, og for hvert kast se på hændelsen: Terningen viser 6 øjne. Resultaterne fra dette eksempel, kan imidlertid direkte overføres til ethvert andet forsøg, som blot opfylder kravene om uafhængighed.

1.1. Et eksempel på en binomialfordeling

Binomialfordelingen er en sandsynlighedsfordeling, der fremkommer, når et "forsøg" gentages et antal gange uafhængigt af hinanden, og hvor man kun interesserer sig for, hvor mange gange en bestemt hændelse indtræffer.
Mere konkret skal vi se på et eksempel, hvor man kaster en terning 12 gange, og hvor vi vil udregne sandsynligheden for at få netop 2. seksere i 12 kast.

Ved kast med en terning har vi et endeligt sandsynlighedsfelt (U_1, P_1) med udfaldsrummet $U_1 = \{1,2,3,4,5,6\}$. Betegner A hændelsen $\{6\}$, og $\overline{A} = \{1,2,3,4,5\}$ den komplementære hændelse, gælder åbenbart: $P_1(A) = 1/6$ og $P_1(\overline{A}) = 5/6$.

Forestiller vi os, at forsøget går ud på at kaste en terning 12 gange, består udfaldsrummet U i dette sandsynlighedsfelt af udfald, der hver er en sekvens af 12 udfald fra udfaldsrummet U_1 .
$U = (u_1, u_2, u_3, ...u_{11}, u_{12})$, hvor hver af u_i 'erne er et af udfaldene 1..6. Udfaldsrummet U har 6^{12} mulige udfald, (nemlig 6 for hver gentagelse), og man skriver ofte $U = U_1 x U_1 x ... x U_1 = U_1^{12}$.

Vi vil først betragte hændelsen A_4 : "En sekser i det 4. kast" i udfaldsrummet U. Da udfaldet af det 4. kast er uafhængig af de 11 øvrige kast vil sandsynligheden for dette være $P(A_4) = P_1(6) = 1/6$.
Tilsvarende finder vi for A_7 = "En sekser i det 7. kast": $P(A_7) = 1/6$.
Vi ser nu på hændelsen $A_4 \cap A_7$: "Både en sekser i det 4. og i det 7. kast".
Bemærk nu - og det er vigtigt - at vi har antaget at hændelserne A_4 og A_7 er *uafhængige* af hinanden, så sandsynligheden for "både og hændelsen" kan beregnes som produktet af de to sandsynligheder:

$$P(A_4 \cap A_7) = P(A_4) \cdot P(A_7) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Ved den samme argumentation, kan vi f.eks. bestemme sandsynligheden for hændelsen:
$A_4 \cap A_5 \cap A_7 \cap \overline{A}_8 \cap \overline{A}_{10}$ som
$P(A_4 \cap A_5 \cap A_7 \cap \overline{A}_8 \cap \overline{A}_{10}) = 1/6 \cdot 1/6 \cdot 1/6 \cdot 5/6 \cdot 5/6 = (1/6)^3 \cdot (5/6)^2$
Nemlig 1/6 for hver gang vi får en sekser og 5/6, hver gang vi *ikke* får en sekser.

Vi er interesseret i hændelsen: " Netop to seksere i 12 kast". En mulighed (blandt mange) kunne være: sekser i de to første kast og ikke sekser i de 10. følgende kast. Sandsynligheden for denne hændelse er:

$$P(\textit{sekser i } 1. \textit{ og } 2. \textit{ kast; ej sekser i } 3 \text{ -}12 \textit{ kast}) = (1/6)^2 * (5/6)^{10}$$

De to seksere kunne også være i 4. og 7. kast eller et hvilket som helst par af de 12 kast. Pointen er nu den, at ligegyldig hvor de to seksere indtræffer vil sandsynligheden for hændelsen "netop to seksere' være den samme, nemlig en faktor 1/6 for hver gang vi slår en sekser og en faktor 5/6 for hver gang vi ikke slår en sekser.

For at beregne sandsynligheden for "Netop to seksere i 12 kast", skal vi derfor addere alle de identiske sandsynligheder. Da de er ens behøver vi derfor blot at vide hvor mange der er.

Vi skal derfor stille spørgsmålet: På hvor mange måder kan vi udvælge (eller anbringe) to seksere på 12 pladser, eller på hvor mange måder kan vi udtage en 2-delmængde (f.eks. {1,2} eller {4,7}), af en 12-mængde. Men svaret på dette er kendt, det er $K(12,2)$.

Vi kan da opskrive den søgte sandsynlighed:

$$P(Netop\ 2.\ seksere\ i\ \ 12\ kast\ med\ en\ terning) = K(12,2)*(1/6)^2*(5/6)^{10}$$

Tallene $K(n,q)$ kaldes også for *binomialkoefficienter*, og man har en alternativ skrivemåde for $K(n,q)$.

$$K(n,q) = \binom{n}{q}$$

Det sidste symbol læses som: "n over q" (som altså ikke betyder n divideret med q)
Sandsynligheden for at en hændelse indtræffer netop q gange i n forsøg, kaldes for en binomialsandsynlighed, og alle disse sandsynligheder kaldes til sammen for binomialfordelingen. Man har tradition for at skrive binomialsandsynligheder, med anvendelse af det nye symbol for binomialkoefficienterne. Hvis X betegner antallet af seksere i 12 kast, finder vi da:

$$\mathrm{P(X = 2)} = \binom{12}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{10} = \binom{12}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^{10}$$

Sandsynligheden kan i øvrigt udregnes til $P(X = 2) = 0,2961$.
På helt samme måde kan man finde sandsynlighederne $P(X = j)\ ;\ j = 0,1,2,...12$

$$\mathrm{P(X = j)} = \binom{12}{j}\left(\frac{1}{6}\right)^j\left(\frac{5}{6}\right)^{12-j} = \binom{12}{j}\left(\frac{1}{6}\right)^j\left(1-\frac{1}{6}\right)^{12-j} \quad ;\ \mathrm{j = 0,\ 1,\ 2,\ ...,12}$$

På denne måde har vi åbenbart fået defineret et endeligt sandsynlighedsfelt med udfaldsrum $U = \{0,\ 1,\ 2,...12\}$, svarende til at man får 0 , 1, ..12. seksere i 12 kast. For at vise, at det virkelig drejer sig om et sandsynlighedsfelt, mangler vi at bevise, at summen af sandsynlighederne $P(X = 0) + P(X = 1) +...+P(X = 12) = 1$. Dette vil blive gjort nedenfor, når vi ser på det mere generelle tilfælde.

1.2. Den generelle binomialfordeling

Vi tænker os nu n uafhængige udførelser af et eksperiment. Vi interesserer os for om en hændelse A indtræffer i det j' te forsøg eller ej. I sandsynlighedsregningen har man tradition for at angive den komplementære hændelse ved at sætte en streg over det symbol, som betegner hændelsen. $\overline{A}$ betegner således den komplementære hændelse til A. Vi sætter P(A) = p , som kaldes for *primærsandsynligheden*. Hermed er $P(\overline{A}) = 1\text{-}p$, sandsynligheden for at A ikke indtræffer.

Sandsynligheden for at A indtræffer i "j" af forsøgene og ikke indtræffer i de resterende "n-j" forsøg, kan beregnes helt på samme måde som i eksemplet med terningen ovenfor, som $p^j\,(1\text{-}p)^{n\text{-}j}$.

Sandsynligheden for at A indtræffer netop j gange (uafhængigt af rækkefølgen) er nu sandsynligheden $p^j\,(1\text{-}p)^{n\text{-}j}$ gange med antallet af forskellige rækkefølger de j gange A indtræffer. Men dette antal er det samme som antallet af j-delmængder man kan udtage af en n-mængde, som er $K(n,j) = \binom{n}{j}$. Heraf fås formlen for binomialsandsynligheden.

$$\mathrm{P(X = j)} = \binom{n}{j} p^j (1-p)^{n-j} \quad ;\, j = 0,\ 1,\ 2,\ ...,n$$

Denne sandsynlighedsfordeling kaldes for *binomialfordelingen*.

For at godtgøre, at disse sandsynligheder faktisk udgør et endeligt sandsynlighedsfelt, mangler vi blot at vise at summen af sandsynlighederne for j = 0 til n er lig med 1.
Umiddelbart synes det ikke særlig simpelt at føre bevis for dette, og det lader sig heller ikke gøre uden kendskab til den såkaldte binomialformel. Vi vil derfor anvende det næste afsnit til at omtale *binomialformlen* og den dertil knyttede *Pascals trekant*.

2. Pascals trekant og binomialformlen

Pascals trekant, er opkaldt efter Blaise Pascal, en fransk matematiker, som levede i 1600-tallet. Han regnes for grundlæggeren af sandsynlighedsregningen, ligesom han var den første som konstruerede en mekanisk regnemaskine. Pascal trekant er blot en opstilling af binomialkoefficienterne $K(n,q) = \binom{n}{q}$ i et skema.

$$
\begin{array}{lccccccccccc}
n=0 & & & & & \binom{0}{0} \\
n=1 & & & & \binom{1}{0} & & \binom{1}{1} \\
n=2 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} \\
n=3 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} \\
n=4 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}
\qquad
\begin{array}{ccccccccc}
 & & & & 1 \\
 & & & 1 & & 1 \\
 & & 1 & & 2 & & 1 \\
 & 1 & & 3 & & 3 & & 1 \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}
$$

....

$$
\begin{array}{ll}
n-1 & \binom{n-1}{0}\cdots\cdots\binom{n-1}{q-1}\;\binom{n-1}{q}\cdots\cdots\binom{n-1}{n-1} \\
n & \binom{n}{0}\cdots\cdots\binom{n}{q-1}\;\binom{n}{q}\;\binom{n}{q+1}\cdots\cdots\binom{n}{n}
\end{array}
$$

Til venstre har vi opstillet binomialkoefficienterne $K(n,q)=\binom{n}{q}$ op til n=4, og tilføjet rækker med n-1 og n, ordnet i en trekant. Til højre har vi udregnet binomialkoefficienterne op til n=4.
Man bemærker, at der står et-taller langs med siderne og at, ethvert tal, som ikke står langs siderne er summen af de to nabo-tal i rækken oven over.

Dette kunne tyde på at der gælder formlen:

$$\binom{n}{q}=\binom{n-1}{q-1}+\binom{n-1}{q} \qquad \text{eller} \qquad K(n,q) = K(n-1,q-1) + K(n-1,q)$$

At denne formel faktisk er korrekt, kan vises algebraisk (ved at sætte højresiderne på fælles brøkstreg og reducere), men formlen vises lettere ved et ræsonnement.

Lad os antage at vi har en mængde U med n elementer. $U = \{a_1 , a_2 , a_3 ,..., a_n \}$. Vi vil nu dele q-kombinationerne i denne mængde, hvoraf der findes $K(n,q)$ forskellige op i to grupper.

1. De q-kombinationer, hvor elementet a_1 *er* med: De resterende q-1 elementer skal udtages af n-1 elementer, hvilket kan gøres på $K(n\text{-}1,q\text{-}1)$ forskellige måder.
2. De q-kombinationer, hvor elementet a_1 *ikke* er med. De q elementer skal da udtages af n-1 elementer, hvilket kan gøres på $K(n\text{-}1,q)$ forskellige måder, hvoraf formlen følger.

Vi ser dernæst på et såkaldt binomium: $(a + b)^n$, hvor a og b er tal, og n er et helt positivt tal eller 0.
Vi ønsker at finde en formel til udregning af et sådant binomium. Vi viser først resultatet for n=0 til 4 nedenfor

$(a + b)^0 = \quad 1$
$(a + b)^1 = \quad 1 \cdot a + 1 \cdot b$
$(a + b)^2 = \quad 1 \cdot a^2 + 2ab + 1 \cdot b^2$
$(a + b)^3 = \quad 1 \cdot a^3 + 3a^2b + 3ab^2 + 1 \cdot b^3$
$(a + b)^4 = \quad 1 \cdot a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1 \cdot b^4$

Det man opdager er, at koefficienterne til hvert af leddene, nøjagtig svarer til en række i Pascals trekant. Dette er ikke noget tilfælde, og det kan indses, ved f.eks. at betragte udregningen af

$$(a + b)^5 = (a + b)\,(a + b)\,(a + b)\,(a + b)\,(a + b)$$

Hvert af leddene, der er et resultat af udregningen, vil fremkomme ved at gange med a eller b, fra hver af de 5 parenteser. Hvert led vil derfor være et udtryk af formen $a^j\, b^{5-j}$, j=0,..,5. F.eks. $a^3\, b^2$, hvor vi har ganget med a i 3 af parenteserne og med b i 2 af parenteserne.
Hvis vi spørger, hvor mange led, der bliver af formen $a^3\, b^2$,så er svaret, at vi skal vælge b fra netop 2 af parenteserne. På hvor mange måder kan vi udvælge 2 parenteser (elementer) af en 5-mængde? Standardsvaret på dette er $K(5,2) = \binom{5}{2}$. Så der bliver altså $K(5,2)$ led af formen $a^3\, b^2$. Noget helt tilsvarende kan opnås for de øvrige led. Herefter kan vi opskrive:

$$(a+b)^5 = \binom{5}{0}a^5b^0 + \binom{5}{1}a^4b^1 + \binom{5}{2}a^3b^2 + \binom{5}{3}a^2b^3 + \binom{5}{4}a^1b^4 + \binom{5}{5}a^0b^5$$

Formlen kan uden videre generaliseres til $(a + b)^n$ og skrives da lettes ved anvendelse af summationstegn.

$$\boxed{(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j}$$

Denne formel kaldes for *binomialformlen*.

Det erindres, at (binomial)koefficienterne til leddene $a^j\, b^{n-j}$, netop svarer til den n'te række i Pascals trekant.

3. Binomialfordelingen fortsat

Vi manglede, at godtgøre at binomialfordelingen er et endeligt sandsynlighedsfelt, hvilket indebærer at vise, at summen af binomialsandsynlighederne for j = 0 til n er lig med 1. Dette er imidlertid nemt at vise, når man anvender binomialformlen

Antag nemlig at vi har to tal s og t , hvor s + t = 1, og dermed t = 1-s. Vi vil da udregne $(s+t)^n = 1^n = 1$. Ved anvendelse af binomialformlen finder vi:

$$1 = 1^n = (s+t)^n = \binom{n}{0}s^n t^0 + \binom{n}{1}s^{n-1}t^1 + \ldots + \binom{n}{j}s^{n-j}t^j + \ldots + \binom{n}{n}s^0 t^n$$

Hvis man i denne formel sætter t = p og dermed s = 1 - p, får man:

$$1 = \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} + \ldots + \binom{n}{j} p^j (1-p)^{n-j} + \ldots + \binom{n}{n} p^n (1-p)^0$$

Man kan da se, at binomialformlen netop er summen af sandsynlighederne $P(X = 0) + P(X = 1) + \ldots + P(X = n)$, som ifølge ovenstående er lig med 1.

Man betegner ofte det, at primærhændelsen indtræffer som "succes" (f.eks. at man slår en sekser), og at den ikke indtræffer betegnes som "fiasko", men man skal i øvrigt ikke tillægge disse ord deres sædvanlige betydning. Primærhændelsen kan f.eks. være at man bliver dræbt i trafikken eller man pådrager sig en dødelig sygdom.
Sandsynligheden $P(X = j)$, som altså betegner sandsynligheden for at få netop j-succeser i n forsøg, når primærsandsynligheden er p, betegnes ofte mere udførligt med $b(j;\ n,\ p)$ eller blot med b_j, hvis n og p er underforstået.

Ofte kan man være interesseret i det antal "succeser", der har den største sandsynlighed, altså det j, for hvilket $b_j = b(j;\ n,\ p)$ er størst. For at undersøge dette udregner vi forholdet b_j / b_{j-1}.

$$\frac{b_j}{b_{j-1}} = \frac{\binom{n}{j} p^j (1-p)^{n-j}}{\binom{n}{j-1} p^{j-1} (1-p)^{n-j+1}} = \frac{p(n-j+1)}{(1-p)j}$$

Det sidste udtryk fremkommer ved at opskrive udtrykkene for binomialkoefficienterne og forkorte.
Vi kan nu opskrive en betingelse for at rækken b_0 , b_1 , b_2 , ... , b_n. er voksende.

$$b_j > b_{j-1} \Leftrightarrow \frac{b_j}{b_{j-1}} > 1 \Leftrightarrow \frac{p(n-j+1)}{(1-p)j} > 1 \Leftrightarrow p(n-j+1) > (1-p)j \Leftrightarrow j < np + p$$

Heraf sluttes: $j \le np \Rightarrow b_j > b_{j-1} \Rightarrow$ rækken er voksende.

På helt tilsvarende vis, ved at vende ulighedstegnet i udregningen ovenfor, finder man at rækken er aftagende hvis og kun hvis $j > np+p$, hvoraf man slutter, at

$j \ge np+1 \Rightarrow b_j < b_{j-1}$, *eller* $j \ge np \Rightarrow b_{j+1} < b_j \Rightarrow$ rækken er aftagende

Af dette ses, at rækken b_0 , b_1 , b_2 , ... , b_n. er voksende for $j \le np$ og aftagende for $j \ge np$.

- Størsteværdien må derfor være for $j = np$, hvis np er et heltal og ellers en af de heltallige naboer til np.

3.1 Stolpediagrammer for binomialfordelingen

For at få et overblik over binomialfordelingen, illustrerer man det ofte ved et stolpediagram. Nedenfor er vist et stolpediagram, svarende til n =12 og p=1/3. Bemærk, at den største sandsynlighed befinder sig ved j=np=4.

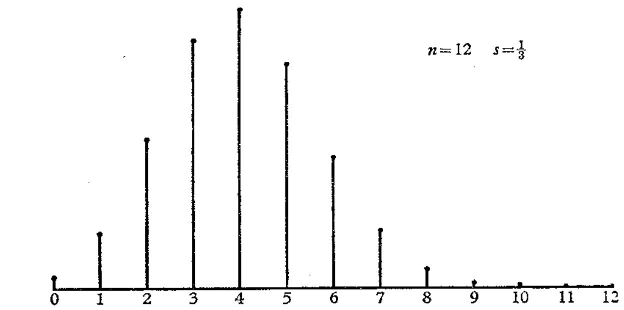


Fig. 14

44

3.2 Kumulerede sandsynligheder

Det kan (selv efter matematikregnerens fremkomst) godt være omstændeligt, at udregne binomialsandsynligheder. Tidligere var man henvist til tabelopslag og interpolation i tabeller. Man anvender stadig tabeller, men de er og har altid været indrettet på en bestemt måde, det de kun indeholder de *kumulerede*, dvs. summerede sandsynligheder.
Vi betragter igen 12 kast med en terning, hvor vi ser på hændelsen 6 øjne. Sandsynligheden for at vi får netop 2 seksere i 12 kast, skriver vi $P(X=2)$. Her er det underforstået, hvilet forsøg og hvilken primærhændelse det drejer sig om. Tilsvarende kan man skrive:

$P(X \leq 4)$ Sandsynligheden for at man får højst 4 seksere

Dette kaldes for en kumuleret sandsynlighed, og det er kun de kumulerede sandsynligheder, der er tabuleret.

Eksempler

1. Vi vil finde sandsynligheden for at slå netop 2 seksere i 12 kast. I tabellen finder vi under n=12, p=1/6
   $P(X \leq 2) = 0{,}6774$ og $P(X \leq 1) = 0{,}3813$. Heraf finder man $P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0{,}2961$

2. Samme forsøg. Vi vil finde sandsynligheden for at man slår mindst 2 seksere.
   P(Mindst 2 seksere) $= P(X \geq 2) = 1 - P(X \leq 1) = 1- 0{,}3813 = 0{,}6187$

3. En familie planlægger 4 børn. Lad primærhændelsen være "en dreng" med sandsynligheden ½. Vi vil bestemme den mest sandsynlige fordeling på drenge af piger, samt hvor stor denne sandsynlighed er.
   Den mest sandsynlige antal drenge er $j = np = 4 \cdot ½ = 2$. Sandsynligheden for at man får 2 drenge og 2 piger er:
   $P(X = 2) = K(4,2)\,(½)^2\,(½)^2 = 6/16 = 3/8$.

3.3 Stikprøveudtagning med tilbagelægning

Denne form for stikprøveudtagning er baseret på binomialfordelingen. Man har en mængde, hvor man udtager et element fra og undersøger det. For at det er det samme forsøg man gentager, skal man lægge elementet tilbage (eventuelt med mulighed for at tage det samme element igen). Heraf navnet: Stikprøveudtagning med tilbagelægning. I praksis udtages elementerne fra så store populationer, at tilbagelægning er underordnet.

Eksempel

1. Lad os antage, at man køber nogen frø, hvor der står at 80% af dem spirer ved korrekt behandling.

Man planter 50 frø, og 34 af dem spirer. Er dette statistisk set acceptabelt?
Antallet af frø, der spirer er binomialfordelt med n=50 og p=0,80. Vi vil finde sandsynligheden for at der spirer 34 eller derunder. Ved tabelopslag finder man: $P(X \leq 34) = 0{,}0308$ I dette tilfælde vil man i almindelighed forkaste hypotesen (antagelsen) at "80% af frøene spirer".
Dette er et eksempel på en hypotesetest eller en kvalitetskontrol ved stikprøve udtagning. Når man laver en hypotesetest (en kvalitetskontrol), vælger man et signifikansniveau. I almindelighed 90%, 95% eller 99%. Vælger vi signifikansniveauet på 90%, betyder det, at ved stikprøven skal der være 90% sandsynlighed eller derover for udfaldet af stikprøven.
Vi skal altså for n=50 og p=0,80 bestemme det største j, således at $P(X \leq j) \leq 0{,}10$, altså, så der er 90% chance for at vi får et udfald større end j. Ved opslag i tabellen finder vi $P(X \leq 35) = 0{,}0607$ og $P(X \leq 36) = 0{,}1106$.
Vælger vi et signifikansniveau på 0,95 finder vi $P(X \leq j) \leq 0{,}05$, altså, så der er 95% chance for at vi får et udfald større end j. Ved opslag i tabellen finder vi $P(X \leq 34) = 0{,}0308$ og $P(X \leq 35) = 0{,}0607$. På signifikansniveauet 95% skal vi således kræve, at 35 planter eller derover spirer for at stikprøven kan accepteres.
Hvis en kvalitetskontrol fejler, kan det naturligvis skyldes statistiske tilfældigheder, og man kan da vælge at foretage en ny stikprøvekontrol med et større antal enheder eller man kan afvise de angivne specifikationer. Sandsynligheden for at en stikprøvekontrol fejler er et ret omfattende emne inden for statistik.

2. Et politisk parti påstår at have en vælgertilslutning på 20%. For at undersøge dette spørger man 40 personer om de ville stemme på partiet. Man ville forvente 8, men kun 6 svarer ja. Kan man på dette grundlag med signifikansniveauet 90% afvise påstanden? Vi har n = 40 og p = 0,20. og vi skal bestemme det største j, så $P(X \leq j) \leq 0{,}10$. Vi finder i tabellen: $P(X \leq 4) = 0{,}0759$ og $P(X \leq 5) = 0{,}1613$. Hermed skal tilslutningen være på 4 eller derunder for at man kan forkaste hypotesen på dette grundlag

3.4 Stikprøveudtagning uden tilbagelægning

Stikprøveudtagning uden tilbagelægning illustreres ofte ved en urne, hvor der er et antal røde og blå kugler. Man udtager en stikprøve på et antal kugler og noterer sig antallet af røde kugler. stikprøveudtagning med tilbagelægning var baseret på binomialfordelingen, mens stikprøveudtagning uden tilbagelægning er baseret på den såkaldte hypergeometriske fordeling.

Eksempel

1. I en urne er der 8 røde og 12 blå kugler. Man udtager en stikprøve på 5, og man vil bestemme Sandsynligheden for j = 0, 1 ,2 ,3 ,4 , 5 røde kugler. Der er K(20,5) muligheder for at udtage stikprøven. Vi vil f.eks. beregne P(3 røde kugler) = P(3). De 3 røde kugler kan udtages på K(8,3) måder og de blå kugler på K(12,2) måder. Heraf finder vi:

$$P(3) = \frac{K(8,3)K(12,2)}{K(20,5)} = \frac{56\cdot 66}{15504} = 0{,}2384$$

3.5 Den hypergeometriske fordeling

Hvis man har en mængde med n elementer, hvoraf q er af en slags *n - q* af en anden slags, og man udtager en stikprøve på *r* elementer, kan man helt på samme måde opskrive sandsynligheden for at få netop j af den ene slags. Man skriver i almindelighed disse sandsynligheder med brug af binomialkoefficienterne hermed fås:

$$P(j) = \frac{\binom{q}{j}\binom{n-q}{r-j}}{\binom{n}{r}} \quad j = 0,1,2 \ldots r$$

Sandsynlighederne *P*(*j*) kaldes for den *hypergeometriske fordeling*.

Kap 5. Stokastiske variable

1. Hvad er en stokastisk variabel

Lad der være givet et endeligt sandsynlighedsfelt (U,P). En stokastisk variabel er formelt defineret som en *funktion med definitionsmængde U*.

Stokastiske variable angives traditionelt med de store bogstaver X,Y og Z. For en stokastisk variabel på et endeligt sandsynlighedsfelt er *værdimængden* for X også en endelig mængde. Man har tradition for at betegne elementer i værdimængden med bogstavet t. Hvis der er n elementer i værdimængden, kan man derfor skrive $Vm(X) = \{t_1, t_2, t_3, \ldots, t_n\}$
Begrebet stokastisk variabel er lettest at anskue, ved at betragte nogle eksempler.
Vi har allerede set eksempler på stokastiske variable. Hvis udfaldsrummet består af tal, kan selve udfaldene betragtes som stokastiske variable.

Eksempler

1. Ved kast med en terning kan øjentallet på terningen betragtes som en (triviel) stokastisk variabel. Man kan nemlig skrive X(1) = 1, X(2) = 2.. osv, Men man kan definere en vilkårlig anden stokastisk variabel: Y(u) = u-3,5 eller Z(u) = u^2.

2. Vi har allerede set på anvendelse af en stokastisk variabel, da vi omtalte binomialfordelingen. Gentager man et eksperiment med udfaldsrum U n gange, skrives udfaldsrummet UxUxUx…xU = U^n. I dette udfaldsrum angiver X antallet af "succeser" (f.eks. antallet af seksere i 20 kast med en terning).
Værdimængden for X er {0, 1, 2, …,n}.
$P(X=j) = \binom{n}{j} p^j (1-p)^{n-j}$; j = 0, 1, 2, ...,n er sandsynligheden for netop j-"succeser" i n forsøg.
3. Der tilbydes et spil, hvor man kaster en terning. Ved hvert af udfaldene {1,2,3,4,5} vinder man 1 kr., men hvis man slår en sekser taber man 6 kr. Gevinsten ved dette spil er en stokastisk variabel, hvor X(1) = X(2) = X(3) = X(4) = X(5) = 1 og X(6) = -6. P(X=1) = 5/6 og P(X=-6) = 1/6. Værdimængden for X er {1,-6}.
Er dette et fordelagtigt spil?

4. På en roulette er der 37 felter {0,1,2,…36}. Hvis kuglen lander på det nummer, man spiller på, får man udbetalt 36 gange sin indsats. Vi betragter en stokastisk variabel, som er gevinsten ved dette spil. Lad os antage at kuglen lander på 23. X(0) = X(1) = X(2) =… X(22) = -1 , X(23) = 35 , X(24) = X(25) =… = X(36) = -1. Værdimængden for X er {-1,35}. Er dette et fordelagtigt spil for spilleren (næppe!)

1.1 Middelværdi for en stokastisk variabel

Middelværdi kan også opfattes som et teoretisk gennemsnit af en størrelse, der er underkastet statistiske tilfældigheder. Middelværdien af den stokastiske variabel X skrives $E(X)$. Ofte anvendes det græske bogstav μ (my) til at betegne middelværdi. Begrebet illustreres lettest med et eksempel.

Eksempel
Man tænker sig et spil, hvor der bliver kastet to mønter. Hvis mønterne viser forskelligt taber man 1 krone, hvis de begge viser plat taber man 2 kroner, men hvis de begge viser krone vinder man 3 kroner. Vi definerer nu en stokastisk variabel, som er gevinsten ved dette spil. For nemheds skyld kalder vi de fire udfald for (*a, b, c, d*). Nedenfor er udfaldsrummet og den stokastiske variabel vist i en tabel.

Udfald *u*	*a* = (pl,pl)	*b* = (pl,kr)	*c* = (kr,pl)	*d* = (kr,kr)
P(u) =	¼	¼	¼	¼
X(u) =	-2	-1	-1	3

Vi tænker os nu at man spiller dette spil et stort antal n (= 100) gange. Vi vil beregne den opnåede gevinst ved de n spil. Lad os antage at udfaldet a er indtruffet n_a gange, b er indtruffet n_b gange, osv. Der gælder: $n_a + n_b + n_c + n_d = n$
Gevinsten G er derfor:

$$G = (-2)\cdot n_a + (-1)\cdot n_b + (-1)\cdot n_c + (3)\cdot n_d = X(a)\cdot n_a + X(b)\cdot n_b + X(c)\cdot n_c + X(d)\cdot n_d$$

Gevinsten pr. spil – den gennemsnitlige gevinst – er G/n. Hvis vi dividerer n op i hvert led i det sidste udtryk, fås:

$$\frac{G}{n} = X(a)\cdot\frac{n_a}{n} + X(b)\cdot\frac{n_b}{n} + X(c)\cdot\frac{n_c}{n} + X(d)\cdot\frac{n_d}{n}$$

Nu er forholdene n_a/n , n_b/n, n_c/n, n_d/n frekvensen (den relative hyppighed) af udfaldene a, b, c , d, som jo netop svarer til sandsynlighederne P(a), P(b), P(c), P(d) for disse udfald i den teoretiske model. Hvis middelværdien af den stokastiske variabel, skal svare til gennemsnittet af den stokastiske variabel i en forsøgsrække, så begrunder det følgende definition.

$$E(X) = X(a)\cdot P(a) + X(b)\cdot P(b) + X(c)\cdot P(c) + X(d)\cdot P(d)$$

$$E(X) = (-2)\cdot 1/4 + (-1)\cdot 1/4 + (-1)\cdot 1/4 + 3\cdot 1/4 = -1/4$$

Den gennemsnitlige gevinst pr. spil er derfor teoretisk –25 øre.

Definition:

- Lad der være givet et endeligt sandsynlighedsfelt (U,P) og en stokastisk variabel X defineret på dette udfaldsrum. Middelværdien af X er defineret som:

$$E(X) = X(u_1)\cdot P(u_1) + X(u_2)\cdot P(u_2) + X(u_3)\cdot P(u_3) + ... + X(u_n)\cdot P(u_n)$$

Eller opskrevet med summationstegn:

$$E(X) = \sum_{u\in U} X(u)\cdot P(u)$$

Det er dog sjældent, at man anvender denne formel til at udregne $E(X)$.

Af formlen følger dog næsten umiddelbart nogle regneregler for middelværdier. Lad X og Y være stokastiske variable defineret i et sandsynlighedsfelt (U,P) og lad c være et vilkårligt reelt tal.

$$E(X+Y) = E(X) + E(Y) \quad , \quad E(c\cdot X) = c\cdot E(X) \quad , \quad E(c) = c$$

Vi beviser kun den første af de 3 formler. De øvrige bevises helt på samme måde.

$$E(X+Y) = \sum_{u\in U}(X(u)+Y(u))\cdot P(u) = \sum_{u\in U}(X(u))\cdot P(u) + \sum_{u\in U}(Y(u))\cdot P(u) = E(X) + E(Y)$$

De udfald, hvor den stokastiske variabel antager en værdi t i værdimængden er en hændelse i udfaldsrummet. Hvis $Vm(X) = \{t_1, t_2, t_3, ..., t_n\}$, så definerer vi hændelsen A_k ved:

$$A_k = \{u \in U \mid X(u) = t_k\}$$

Sandsynligheden for hændelsen $P(A_k)$, vil vi herefter af indlysende grunde skrive som $P(X = t_k)$.

Hændelserne A_1, A_2, A_3, ..., A_n udgør en klassedeling af udfaldsrummet, og vi kan derfor dele summen i udregningen af $E(X)$ op på disse hændelser:

$$E(X) = \sum_{u \in U} X(u) \cdot P(u) = \sum_{u \in A_1} X(u) \cdot P(u) + \sum_{u \in A_2} X(u) \cdot P(u) + \ldots + \sum_{u \in A_n} X(u) \cdot P(u)$$

Da $X(u)$ har den samme værdi t_k på hver af hændelserne A_k finder man:

$$E(X) = \sum_{u \in U} X(u) \cdot P(u) = t_1 \cdot \sum_{u \in A_1} P(u) + t_2 \cdot \sum_{u \in A_2} P(u) + \ldots + t_n \cdot \sum_{u \in A_n} P(u)$$

$$E(X) = \sum_{u \in U} X(u) \cdot P(u) = t_1 \cdot P(X = t_1) + t_2 \cdot P(X = t_2) + \ldots + t_n \cdot P(X = t_n)$$

Vi finder derfor følgende formel til beregning af $E(X)$, som man i praksis anvender.

$$E(X) = \sum_{k=1}^{n} t_k P(X = t_k)$$

Eksempler

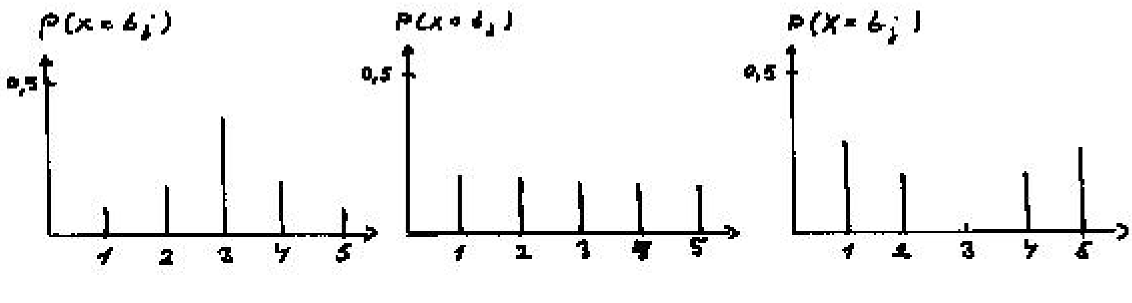
1. I eksemplet 3 ovenfor med terningspillet er Vm(X)={1,-6} og P(X=1) =5/6 og P(X=6)=1/6. Heraf fås: E(X) = 1· 5/6 + (-6)· 1/6 = -1/6. Spillet er ikke fordelagtigt.

2. I eksempel 4 med Casino er Vm(X)={-1,35} og P(X = -1) =36/37 og P(X = 35)=1/37. Heraf fås: E(X) = (-1)· 36/37 + 35· 1/37 = -1/37. Spillet er ikke fordelagtigt.

3. Vi vil bestemme middelværdien af øjentallet ved kast med en terning. U = {1, 2, 3, 4, 5, 6} og P(u) =1/6 for alle u. Vi sætter X = u og finder: E(X) = 1/6·1+1/6·2+…+ 1/6·6 = 1/6·(1+2+3+4+5+6) = 21/6 = 3½.

1.2 Varians og spredning af en stokastisk variabel

Nedenfor er vist sandsynlighedsfordelingen, for tre stokastiske variable, som alle har middelværdien 3, men som alligevel er meget forskellige. Man siger at spredningen er forskellig på de tre variable, idet spredningen vokser, når man går fra venstre mod højre. Blandt andet for at kunne skelne sådanne fordelinger fra hinanden indfører man begrebet spredning på en stokastisk variabel.



Mens middelværdien, kan beskrives som et teoretisk estimat af gennemsnittet af en stokastisk variabel, er spredningen lidt vanskeligere at forklare. Spredningen er mest relevant, når man har en sandsynlighedsfordeling, hvor de fleste udfald samler sig om middelværdien.

Lad os antage, at en slikbutik sælger bolsjer i kræmmerhuse. Middelværdien af antallet af bolsjer i kræmmerhusene er 35, men det kan variere fra 30 til 40. Vi ønsker at bestemme middelværdien af afvigelsen fra middelværdien af antallet af bolsjer.
En mulighed er at man finder middelværdien af $|X(u) - E(X)|$, men det gør man ikke af flere grunde. I stedet finder man middelværdien af $(X(u) - E(X))^2 = (X(u) - \mu)^2$. Denne størrelse kaldes for *variansen* eller *spredningskvadratet* og betegnes $Var(X)$ eller $\sigma(X)^2$. Kvadratroden af variansen betegnes *spredningen* og betegnes $\sigma(X)$. Formelt defineres variansen af stokastisk variabel ved udtrykket:

$$Var(X) = \sigma(X)^2 = E(\,(X - \mu)^2\,)$$

Der gælder følgende vigtige formel, som følger af regnereglerne for stokastiske variable.

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

Hermed:

$$Var(X) = \sigma^2(X) = E(\,(X - \mu)^2\,) = E(X^2) - \mu^2$$

I langt de fleste tilfælde er det lettere at anvende denne formel end definitionen.

$E(X^2)$ udregnes på samme måde som $E(X)$. $E(X^2) = \sum_{k=1}^{n} t_k{}^2 P(X = t_k)$

Der gælder følgende regneregler for varians og spredning, som kan bevises helt tilsvarende til regnereglerne for middelværdi.

$$Var(c) = 0 \qquad Var(c \cdot X) = c^2 \cdot Var(X) \qquad Var(X + c) = Var(X)$$
$$\sigma(c) = 0 \qquad \sigma(c \cdot X) = |c| \cdot \sigma(X) \qquad \sigma(X + c) = \sigma(X)$$

Vi nøjes med at vise at $Var(X + c) = Var(X)$:
$E(X + c) = \mu + c$ og dermed $Var(X + c) = E((X + c) - (\mu + c))^2) = E((X - \mu)^2) = Var(X)$.

I almindelighed gælder der derimod *ikke*, at $\sigma(X + Y)^2 = \sigma(X)^2 + \sigma(Y)^2$. Dette gælder kun, hvis $E(Y \cdot X) = E(X) \cdot E(Y)$. Det sidste er tilfældet, hvis X og Y er uafhængige stokastiske variable.

Lemma
I det følgende får vi brug for følgende lille sætning.
Hvis X og Y er to stokastiske variable på det samme udfaldsrum og hændelserne $X = t_i$ og $Y = s_j$ er uafhængige for alle $t_i \in Vm(X)$ og $s_j \in Vm(Y)$, altså hvis $P(X = t_i \wedge Y = s_j) = P(X = t_i) \cdot P(Y = s_j)$, så er $E(X \cdot Y) = E(X) \cdot E(Y)$

$$E(X \cdot Y) = \sum_{t_i \in Vm(X), s_j \in Vm(Y)} t_i s_j P(X = t_i \wedge Y = s_j) = \sum_{t_i \in Vm(X), s_j \in Vm(Y)} t_i s_j P(X = t_i) P(Y = s_j) =$$
$$\sum_{t_i \in Vm(X)} t_i P(X = t_i) \sum_{s_j \in Vm(Y)} s_j P(Y = s_j) = E(X) \cdot E(Y)$$

Lad nu X og Y være to *uafhængige* stokastiske variable. Vi vil vise at $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$.

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

Ved at anvende denne sætning, kan vi vise en vigtig sætning vedrørende middelværdi og gennemsnit af en række uafhængige målinger af værdien af en stokastisk variabel. Lad X være en stokastisk variabel med middelværdi μ og spredning σ Foretager vi n målinger $\{x_1, x_2, x_3, \ldots, x_n\}$ af denne og danner gennemsnittet.

$$s = \frac{1}{n}(x_1 + x_2 + x_3 + \ldots + x_n)$$

så svarer middelværdien af den stokastiske variabel

$$X_G = \tfrac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)$$

Ved brug af regnereglerne ovenfor finder man:

$$E(X_G) = \tfrac{1}{n}(E(X_1)+E(X_2)+E(X_3)+\ldots+E(X_n)) = \tfrac{1}{n}n\mu = \mu$$

Hvilket næppe kan overraske. For spredningen finder man derimod

$$\sigma^2(X_G) = \sigma^2(\tfrac{1}{n}(X_1+X_2+X_3+\ldots+X_n)) = \tfrac{1}{n^2}\sigma^2((X_1+X_2+X_3+\ldots+X_n)) =$$
$$\tfrac{1}{n^2}(\sigma^2(X_1)+\sigma^2(X_2)+\sigma^2(X_3)+\ldots+\sigma^2(X_n)) = \tfrac{1}{n^2}(n\sigma^2) = \tfrac{1}{n}\sigma^2$$

Og hermed $\sigma(X_G) = \frac{1}{\sqrt{n}}\sigma$. Spredningen på et gennemsnit af n målinger er således $\frac{1}{\sqrt{n}}$ af spredningen på hver af målingerne.

Eksempler

Vi er nu i stand til at udregne middelværdi og spredning for de 3 stolpediagrammer, som vi introducerede før, og som vi nummererer (1), (2) and (3): Sandsynlighederne står I den øverste række-

X_1: $P(1) = 0.1$, $P(2) = 0.2$, $P(3) = 0.4$, $P(4) = 0.2$, $P(5) = 0.1$
$\mu_1 = E(X_1) = 1\cdot 0.1+2\cdot 0.2+3\cdot 0.4+4\cdot 0.2+5\cdot 0.1 = 3.0$
$E(X_1^2) = 1^2\cdot 0.1+2^2\cdot 0.2+3^2\cdot 0.4+4^2\cdot 0.2+5^2\cdot 0.1 = 9.3$
$Var(X_1^2) = E(X_1^2) - \mu_1^2 = 9.3 - 9 = 0.3$

X_2: $P(1) = 0.2$, $P(2) = 0.2$, $P(3) = 0.2$, $P(4) = 0.2$, $P(5) = 0.2$
$\mu_1 = E(X_1) = 1\cdot 0.2+2\cdot 0.2+3\cdot 0.2+4\cdot 0.2+5\cdot 0.2 = 3.0$
$E(X_2^2) = 1^2\cdot 0.2+2^2\cdot 0.2+3^2\cdot 0.2+4^2\cdot 0.2+5^2\cdot 0.2 = 11.0$
$Var(X_2^2) = E(X_2^2) - \mu_2^2 = 11.0 - 9 = 2.0$

X_3: $P(1) = 0.3$, $P(2) = 0.2$, $P(3) = 0.0$, $P(4) = 0.2$, $P(5) = 0.3$
$\mu_3 = E(X_3) = 1\cdot 0.3+2\cdot 0.2+3\cdot 0.0+4\cdot 0.2+5\cdot 0.3 = 3.0$
$E(X_2^2) = 1^2\cdot 0.3+2^2\cdot 0.2+3^2\cdot 0.0+4^2\cdot 0.2+5^2\cdot 0.3 = 11.8$
$Var(X_3^2) = E(X_3^2) - \mu_3^2 = 11.8 - 9 = 2.8$

Som man ser er varianserne fordelt som påstået.

1. Vi vil dernæst beregne varians og spredning på øjentallet ved kast med en terning:
   Var(x) = $E(X^2) - E(X)^2 = 1^2\cdot 1/6 + 2^2\cdot 1/6 + 3^2\cdot 1/6 + 4^2\cdot 1/6 + 5^2\cdot 1/6 + 6^2\cdot 1/6 - (3{,}5)^2 = 2{,}92$
   $\sigma(X) = \sqrt{\mathrm{Var}(x)} = 1{,}70$

2. Gennemsnit og standardafvigelse

Lad os antage, at vi har en række observationer $x_1, x_2, x_3, \ldots, x_n$, svarende til stokastiske variable X_1. $X_2, \ldots X_n$ med den samme middelværdi og spredning.
Lad os antage at vi ikke kender, sandsynlighedsfordelingen for X_i, og dermed heller ikke middelværdien og spredningen for X_i.
Man beregner da to *estimater* af middelværdi og spredning efter følgende formler:

$$\bar{x} = \tfrac{1}{n}\sum_{k=1}^{n} x_k \qquad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$$

Svarende til de stokastiske variable

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

Begrundelsen for disse valg er (naturligvis), at $E(\bar{x}) = \mu$ og $E(s^2) = \sigma^2$
Dette vil nu bevise:
Det første er næsten trivielt, mens den anden del kræver lidt algebra.

$$E(\bar{X}) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}(n\mu) = \mu$$

Den anden halvdel viser vi ved først, at udregne $E((X_i - \bar{X})^2)$.
Vi minder om, at $\sigma^2 = E(X^2) - \mu^2$ så $E(X^2) = \sigma^2 + \mu^2$.

$$E((X_i - \bar{X})^2) = E(X_i^{\,2}) + E(\bar{X}^2) - 2E(X_i\bar{X}) = \sigma^2 + \mu^2 + \frac{1}{n^2}E((\sum_{i=1}^{n} X_i)^2) - 2\frac{1}{n}E(X_i\sum_{k=1}^{n} X_k)$$

Da X_i *og* X_k er antaget at være uafhængige, er

$$E(X_iX_k) = E(X_i)E(X_k) = \mu^2 \quad for \quad i \neq k \quad og \quad E(X_i^{\,2}) = \sigma^2 + \mu^2$$

Summen $(\sum_{i=1}^{n} X_i)$ har n led og $(\sum_{i=1}^{n} X_i)^2$ har derfor n^2 led. Heraf n led X_i^2, og derfor $n^2 - n$ led $X_i X_k$, hvor i er forskellig fra k. Samler vi nu det hele finder man:

$$E((X_i - \bar{X})^2) = \sigma^2 + \mu^2 + \frac{1}{n^2}((n^2 - n)E(X_iX_k) + nE(X_i^{\,2})) - \frac{2}{n}((n-1)E(X_iX_k) + E(X_i^{\,2}))$$

$$E((X_i-\overline{X})^2)=\sigma^2+\mu^2+\frac{1}{n^2}((n^2-n)\mu^2+n(\sigma^2+\mu^2))-\frac{2}{n}((n-1)\mu^2+\sigma^2+\mu^2)$$

$$E((X_i-\overline{X})^2)=\sigma^2+\mu^2+\frac{1}{n}\sigma^2+\mu^2-\frac{2}{n}\sigma^2-2\mu^2$$

$$E((X_i-\overline{X})^2)=\frac{n-1}{n}\sigma^2$$ Heraf følger nemlig:

$$E(s^2)=E\left(\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2\right)=\frac{1}{n-1}\sum_{k=1}^{n}E((X_k-\bar{X})^2)=\frac{1}{n-1}n\frac{n-1}{n}\sigma^2=\sigma^2$$ Hvilket skulle vises.

3. Middelværdi og spredning af binomialfordelingen

Vi tænker os en binomialfordeling med n gentagelser og med primærsandsynlighed p. For binomialfordelingen er disse to begreberne middelværdi og spredning ofte meget vigtige. Det er imidlertid ikke så enkelt at udregne middelværdi og spredning for binomialfordelingen direkte fra definitionen. I stedet indfører man en stokastisk variabel X_i ved definitionen.

$X_i = 1$, hvis primærhændelsen indtræffer i det *i'te* forsøg og 0 ellers

Vi finder umiddelbart $E(X_i) = 1\cdot p + 0\cdot(1-p) = p$. Idet primærhændelsen indtræffer med sandsynlighed p i det i'te forsøg.

Hvis X er den stokastiske variabel, der betegner antallet af "succeser" (antal gange primær hændelsen indtræffer), så er det indlysende, at $X = X_1 + X_2 + X_3 + \ldots + X_n$, hvor X_k betegner den stokastiske variabel, som vi indførte ovenfor. Ifølge regneregler for middelværdier finder vi så:

$$E(X) = E(X_1 + X_2+ X_3+\ldots+ X_n) = E(X_1) + E(X_2) + E(X_3) +\ldots+ E(X_n) = n\cdot p$$

Eksempel

1. Det er formodentlig ikke overraskende, at hvis man kaster en terning 100 gange, så er den gennemsnitlige antal seksere $100/6 \approx 16$.
2. I en familie med 4 børn ($n = 4$, $p = ½$) er det gennemsnitlige antal drenge $4\cdot ½ = 2$.

Når vi skal udregne spredningen på en binomialfordeling, får vi brug for at udregne spredningen på en af de stokastiske variable X_j. Der gælder at $X_j^2 = 1^2 = 1$, når der er succes i i'te og ellers 0.

$$Var(X_j) = E(X_j^2) - E(X_j)^2 = 1\cdot p - p^2 = p(1-p)$$

Da de stokastiske variable $X_1, X_2, X_3, \ldots, X_n$ er uafhængige, kan vi simpelthen bestemme variansen af $X = X_1 + X_2 + X_3 + \ldots + X_n$, som summen af varianserne på hver af dem. Da X_j'er alle har den samme varians finder man: $Var(X) = n\cdot Var(X_j) = np(1-p)$. Og dermed

$$\sigma(X)=\sqrt{np(1-p)}$$

Man definerer frekvensen af "succes" i n forsøg, som $Y = \frac{1}{n}X$. Ifølge regneregler for middelværdier og spredning finder man følgende udtryk for frekvensen:

$$E(Y) = \tfrac{1}{n}E(X) = p \qquad \text{og} \qquad \sigma(Y) = \tfrac{1}{n}\sigma(X) = \tfrac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}$$

Eksempel
Kaster man en terning 100 gange med henblik på at slå en sekser, så er spredningen på antallet af seksere

$$\sigma(X) = \sqrt{100 \cdot \tfrac{1}{6}(1 - \tfrac{1}{6})} = \tfrac{10}{6}\sqrt{5} = 3{,}73$$

4. Chebychevs ulighed

Chebychevs ulighed er grundlaget for det, der ofte løst betegnes som "De store tals lov", og som ofte fejlagtigt formuleres, som at frekvensen af en hændelse i et stort antal forsøg (f.eks. kast med en terning med henblik på en sekser) vil konvergere mod sandsynligheden for hændelsen. Det kan man imidlertid ikke bevise, men man kan bevise, at *sandsynligheden* for at få en afvigelse fra middelværdien, går mod nul, når antallet af forsøg går imod uendelig.

Hvis X er en vilkårlig stokastisk variabel med middelværdi μ og spredning σ. Vi er altså interesseret i en vurdering af afvigelsen fra middelværdien $|X - \mu|$ målt i enheder af σ.
Vi definerer en stokastisk variabel

$$Y_c = c^2 \qquad \text{for} \qquad |X - \mu| \geq c \quad \text{og ellers } 0.$$

Ud fra definitionen af Y_c følger umiddelbart: $Y_c \leq (X - \mu)^2$ og hermed

$$\sigma^2 = E((X-\mu)^2) \geq E(Y_c)$$

$$E(Y_c) = c^2 P(|X-\mu| \geq c) + 0 \cdot P(|X-\mu| < c) = c^2 P(|X-\mu| \geq c)$$

Af disse to relationer finder man Chebychevs ulighed:

$$c^2 P(|X-\mu| \geq c) \leq \sigma^2 \quad \Leftrightarrow \quad P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Som udtrykker, at sandsynligheden for at få en afvigelse fra middelværdien, som er større end et vilkårligt tal c, er mindre end eller lig spredningen divideret med c i 2 potens.
Umiddelbart virker Chebychevs ulighed ikke særlig imponerende, men vi vil nu anvende på den stokastiske variabel Y, som er frekvensen af succes i n forsøg. Vi udledte tidligere

$$E(Y) = \tfrac{1}{n}E(X) = p \qquad \text{og} \qquad \sigma(Y) = \sqrt{\tfrac{p(1-p)}{n}}$$

Heraf finder vi:

$$P(|Y-p| \geq c) \leq \frac{\sigma^2}{c^2} = \frac{p(1-p)}{nc^2} \leq \frac{1}{4nc^2}$$

Den sidste ulighed er opnået fordi $p(1\text{-}p)$ har størsteværdi ¼ for $p=½$
Vælges nu f.eks. c=0,01, så ser man at *sandsynligheden* for at få en frekvens, der afviger fra p med mere end 0,01 er $\frac{1}{4nc^2}$. Kræver vi at denne sandsynlighed er mindre end 0,1, skal vi løse uligheden:

$$\frac{1}{4nc^2} \le \frac{1}{10} \Leftrightarrow n \ge \frac{10}{4\cdot(0{,}01)^2} \Leftrightarrow n \ge 25000$$

Et resultat, der nok har større teoretisk end praktisk interesse. I praksis kan man i almindelighed nøjes med langt færre forsøg for at op nå den samme sikkerhed.

5. Grafisk afbildning. Fordelingsfunktion

Sandsynlighedsfordelingen for en stokastisk variabel, altså sandsynlighederne $P(X = t_i)$, hvor $t_i \in Vm(X)$, kaldes også frekvensfunktionen og betegnes $f(t)$. $f(t) = P(X = t)$ er kun defineret for $t \in Vm(X)$.
Fordelingsfunktionen $F(t) = P(X \le t) = \sum_{t_i \le t} P(X = t_i)$ kan derimod defineres for ethvert reelt tal. I et endeligt sandsynlighedsfelt, hvor værdimængden for X er $\{t_1 , t_2 , t_3 \ldots t_n \}$ vil $F(t)$ være konstant i hver af intervallerne $t_i \le t < t_{i+1}$, $F(t) = 0$ for $t < t_1$ og $F(t) = 1$ *for* $t \ge t_n$.

Nedenfor er vist et par eksempler på fordelingsfunktioner for en binomialfordeling.

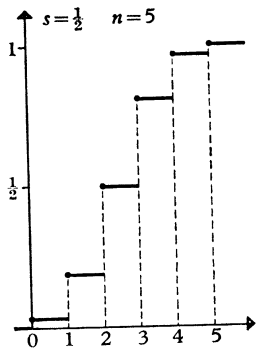


Fig. 3

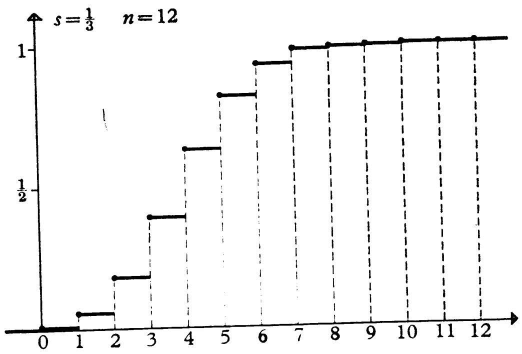


Fig. 4

Som tidligere omtalt er det ikke frekvensfunktionerne, som er tabuleret, men fordelingsfunktionerne.

Kap 6. Normalfordelingen

1. Kontinuerte sandsynlighedsfordelinger

Mange stokastiske variable, har den egenskab, at de kan antage alle værdier i et interval. Det gælder f.eks. højden eller vægten af en person, størrelsen eller vægten af en bestemt komponent i en industriproduktion. Man taler i denne sammenhæng om en kontinuert sandsynlighedsfordeling. Hvis en stokastisk variabel, kan antage uendelig mange værdier, giver det ikke mening, at tale om sandsynligheden for at opnå en bestemt værdi. I stedet definerer man en *tæthedsfunktion* $f(t)$, hvor sandsynligheden for at nå en værdi i et lille interval dt omkring t er $dP = f(t)dt$. For at $f(t)$ er en tæthedsfunktion, skal den være normaliseret. Summen af alle sandsynligheder skal være 1. Det er ikke væsentligt, at forstå nedenstående formel, men for kontinuerte sandsynlighedsfordelinger skriver man dette:

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

Fordelingsfunktionen er herefter defineret som:

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx$$

$F(t)$ er således arealet under grafen for $f(t)$ fra minus uendelig til t.
Der er ingen grund til at forsøge, at forstå disse formler, men da begrebet sandsynlighed for en kontinuert sandsynlighedsfordeling er baseret på disse formler er de medtaget. Vi vil ikke anvende formlerne direkte.

2. Normalfordelingen

Hvis man i binomialfordelingen med primærsandsynlighed p, middelværdi $\mu=np$ og spredning $\sigma = \sqrt{np(1\text{-}p)}$ lader n gå imod uendelig, samtidig med at p går imod 0, på en sådan måde, at $\mu=np$ holdes konstant, nærmer man sig en kontinuert sandsynlighedsfordeling.
Man kan vise (men det kræver matematik langt ud over gymnasieniveau), at binomialsandsynligheden vil nærme sig asymptotisk til følgende funktion, som kaldes normalfordelingen (tæthedsfunktionen) med middelværdi μ og spredning σ.

$$P(X = t) = \binom{n}{t} p^t (1-p)^{n-t} \to \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad \text{for n} \to \propto \;\; p \to 0 \;\; (np \text{ konstant})$$

Det viser sig, at *næsten* alle sandsynlighedsfordelinger i en grænse, nærmer sig til normalfordelingen, hvilket er en begrundelse for navnet på denne fordeling.
Normalfordelingen har imidlertid en anden egenskab, som måske er endnu vigtigere. Hvis man har en stokastisk variabel X med en vilkårlig sandsynlighedsfordeling, middelværdi μ og spredning σ,

så vil gennemsnittet af n målinger: $<x>=\frac{1}{n}(x_1+x_2+x_3+...+x_n)$ i grænsen hvor n går imod uendelig være normalfordelt med samme middelværdi og spredning $\sigma_n=\frac{1}{\sqrt{n}}\sigma$.
Det er især denne sidste egenskab, der er årsagen til, at man antager at en stokastisk variabel er normalfordelt, hvis man i øvrigt ikke ved andet.

3. Anvendelser af normalfordelingen

Standardnormalfordelingen er en normalfordeling som har middelværdi 0, og spredning 1. For Standard normalfordelingen har vi tæthedsfunktion og fordelingsfunktion

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} \quad og \quad \Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-\frac{x^2}{2}}dx$$

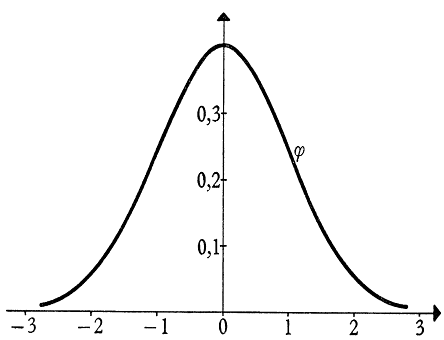
Hvis en stokastisk variabel er normalfordelt med middelværdi μ og spredning σ, så vil *den reducerede variabel* $\frac{t-\mu}{\sigma}$ have middelværdi 0 og spredning 1. Dette følger direkte af regneregler for middelværdier og spredning, idet

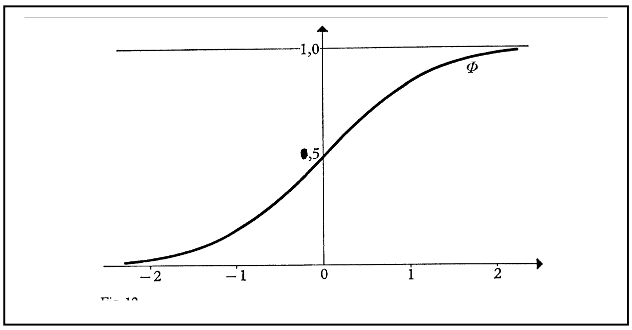
$$E(\tfrac{t-\mu}{\sigma})=\tfrac{1}{\sigma}(E(t)-E(\mu))=\tfrac{1}{\sigma}(\mu-\mu)=0 \quad og \quad \sigma(\tfrac{t-\mu}{\sigma})=\tfrac{1}{\sigma}(\sigma(t)-\sigma(\mu))=\tfrac{1}{\sigma}(\sigma-0)=1$$

Ønsker vi da at bestemme sandsynligheden for at vi får et udfald ≤ a, så skal vi bestemme *P*(*X* ≤ *a*) = *F*(*a*), hvor *F*(*t*) er normalfordelingsfunktionen svarende til middelværdi μ og spredning σ. Dette kan imidlertid udregnes som

$$F(a)=\Phi(\tfrac{a-\mu}{\sigma}) \quad \text{, hvor } \Phi \text{ er standard-normalfordelingsfunktionen}$$

Φ-funktionen er en tabuleret funktion, og den findes også på nogle (statistiske) lommeregnere. Nedenfor er først vist grafen for tæthedsfunktionen og dernæst fordelingsfunktionen for standard-normalfordelingen.





Når man skriver sandsynligheder for kontinuerte sandsynlighedsfordelinger er det underordnet om man skriver F(t) = P(X < t) eller F(t) = P(X ≤ t). Sandsynligheden P(X = t) = 0.
For Φ funktionen gælder følgende værdier, som også fører til nogle resultater for en vilkårlig normalfordelingsfunktion:

$$F(\mu) = P(X < \mu) = \Phi(\tfrac{\mu-\mu}{\sigma}) = \Phi(0) = \tfrac{1}{2}$$

Der er altid 50% chance for at få et udfald mindre end eller lig middelværdien.

$$F(\mu - \sigma) = P(X < \mu - \sigma) = \Phi(\tfrac{\mu-\sigma-\mu}{\sigma}) = \Phi(-1) = 0{,}1587$$

Der er således knap 16% sandsynlighed for at få et udfald, som er mindre end μ-σ.
Tilsvarende fås:

$$F(\mu + \sigma) = P(X < \mu + \sigma) = \Phi(\tfrac{\mu+\sigma-\mu}{\sigma}) = \Phi(1) = 0{,}8413$$

Som det ses, er summen af de to sidste resultater lig med 1. Dette er ikke noget tilfælde. På grund af symmetrien af φ funktionen omkring 0, gælder der nogle små sætninger om Φ-funktionen, som vi blot nævner: $\Phi(-t) = 1-\Phi(t)$ og dermed: $\Phi(t) - \Phi(-t) = 2\cdot\Phi(t)-1$

Vil man bestemme sandsynligheden for at en Stokastisk variabel ligger i intervallet fra a til b, gøres det som følger:

$$P(a \le t \le b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Eksempel

En bestemt fødevare, sælges i 500 g pakker. Producenten oplyser, at den faktiske vægt er normalfordel med middelværdi 500 og spredning 25.

a) Hvor mange % af pakkerne, vil man herefter forvente vejer mindre end 475 g.
b) Hvor mange % af pakkerne, vil man herefter forvente vejer mindre end 450 g.
c) Find det interval omkring middelværdien,, hvor 90% af pakkernes vægt befinder sig.
d) Aftageren kræver, at 90% af pakkerne har en vægt mellem 490 g og 510 g. Hvor stor skal spredningen være for at producenten kan opfylde dette?

a) $P(X \le \mu - \sigma) = F(\mu - \sigma) = \Phi(-1) = 0{,}1587 = 15{,}9\%$
b) $P(X \le \mu - 2\sigma) = F(\mu - 2\sigma) = \Phi(-2) = 0{,}0228 = 2{,}3\%$ (Resultatet fundet ved tabelopslag)
c) $P(\mu - \alpha\cdot\sigma \le X \le \mu + \alpha\cdot\sigma) = F(\mu - 2\sigma) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha) - 1 = 0{,}90$.
Heraf finder man $\Phi(\alpha) = 0{,}95$. Ved tabel opslag i invers normalfordeling $\alpha = 1{,}65$
90% befinder sig altså i intervallet fra $500 - 25\cdot 1{,}65$ til $500 + 25\cdot 1{,}65$, dvs. intervallet [459, 541]
d) Vi skal åbenbart løse ligningen (jf. sp. c) $500-\sigma\cdot 1.65 = 490$, som giver $\sigma = 6.1$

3.1 Anvendelse af normalfordelingspapir

Hvis man skal undersøge, og et statistisk materiale er normalfordelt, udregner man de kumulerede frekvenser og skal da undersøge om dette passer med normalfordelingens Φ-funktion. Dette er imidlertid ikke så ligetil, hvis man hverken kender middelværdi eller spredning.
I stedet anvender man det såkaldte *normalfordelingspapir*. Normalfordelingspapir er i princippet fremstillet på samme måde som enkeltlogaritmisk papir, blot med den forskel at 2. aksen ikke er en logaritmisk skal, men en Φ-funktions skala, forstået på den måde, at alle fordelingsfunktioner, svarende til en normalfordeling:

$$F(t) = P(X \le t) = \Phi(\tfrac{t-\mu}{\sigma})$$

er *rette linier*, når de afbildes på normalfordelingspapir. Valget af enheder på 1. aksen er underordnet.

Afsætter man de udregnede kumulerede frekvenser på normalfordelingspapir, og ligger punkterne med tilnærmelse på en ret linie, kan man med god grund antage at den stokastiske variabel er normalfordelt. I eksemplet nedenfor er vist, hvorledes man ud fra den rette linie let kan beregne middelværdi og spredning for talmaterialet.

Eksempel

Vi har tidligere vist, at $F(\mu) = P(X \leq \mu) = \Phi(\frac{\mu-\mu}{\sigma}) = \Phi(0) = \frac{1}{2}$. Middelværdien kan derfor umiddelbart aflæses som 1. koordinaten til det punkt, som har 2.koordinaten ½. Spredningen kan derefter findes ud fra en af ligningerne:

$$F(\mu - \sigma) = P(X \leq \mu - \sigma) = \Phi(\tfrac{\mu-\sigma-\mu}{\sigma}) = \Phi(-1) = 0{,}1587$$

$$F(\mu + \sigma) = P(X \leq \mu + \sigma) = \Phi(\tfrac{\mu+\sigma-\mu}{\sigma}) = \Phi(1) = 0{,}8413$$

idet μ-σ og μ+σ kan aflæses som 1. koordinaten til de punkter, som har 2.koordinaterne 0,1587 og 0,8413

Ofte aflæser man kun de to sidstnævnte værdier (som er markeret på 2. aksen af normalfordelingspapiret), og bestemmer μ og σ ud fra de to fremkommende ligninger.
Hvis man f.eks. aflæser: $\mu-\sigma = 482$ og $\mu+\sigma = 518$, så finder man umiddelbart $\mu = 500$ og $\sigma = 18$.

Opgaver

Kombinatorik

1. Morsealfabetet består af "tegnene" • og —. En kode (bogstav eller ciffer) består af fra et til 5 tegn. Hvor mange forskellige koder kan der dannes?

2. I en skakturnering er der 8 deltagere. Alle skal spille mod alle. Hvor mange partier skal der spilles?

3. Hvor mange forskellige bordplaner er mulige, når et selskab på 7 herrer og 6 damer skal bænkes omkring et cirkulært bord, og to damer ikke må sidde ved siden af hinanden? Hvor mange forskellige bordplaner er mulige, når der er 6 herrer og 6 damer.

4. Ved en prøvesmagning af øl, stilles 5 glas op i rækkefølge. Hvor mange muligheder er der for rækkefølger?

5. Hvor mange 3-cifrede tal findes der, hvor alle 3 cifre er forskellige? Hvor mange 3-cifrede tal findes der, hvor netop 2 cifre er ens?

6. Registreringsnummeret (Nummerpladen) på en bil består af to bogstaver (A-Z) og 5 cifre uden for anstillede nuller. F.eks. OS 52911 eller RU 45710. Hvor mange forskellige registreringsnumre kan der dannes? For to givne bogstaver, hvor mange registreringsnumre kan der dannes, hvor alle cifrene er forskellige?

7. Som omtalt ovenfor er der 3^{12} forskellige måder at udfylde en tipskupon på. En af disse har 13 rigtige. Hvor mange har 12, 11 og 10 rigtige?

Endeligt sandsynlighedsfelt

8. Lad P være en sandsynlighedsfunktion på mængden U = {1,2,3,4,5}. Find P(5), når
   a) P(1) = P(2) = 0,1 og P(3) = P(4) = 0,2
   b) P(1) = P(2) = P(3) = P(4) = P(5)
   c) P(1) =0,4 og P(2) = P(3) = P(4) = P(5)

9. Ved kast med 3 terninger, skal man finde sandsynligheden for at de 3 terninger ikke viser samme øjental.

10. Hvis en familie planlægger at få 4 børn, skal man bestemme sandsynligheden for
    a) at man får 4 piger eller 4 drenge
    b) at man får mindst en pige og en dreng.
    c) at man får to (søde) piger og to (uartige) drenge.

11. Ved kast med to terninger, skal man finde sandsynligheden for at man slår netop 1 sekser.

12. En bilist skal på vej til arbejde passere 4 stoplys. Det antages, at sandsynligheden for rødt og grønt lys begge er ½.
    a) Hvor mange forskellige muligheder er der for, rødt og grønt lys langs med ruten?
    b) Hvad er sandsynligheden for 4 røde lys?
    c) Hvad er sandsynligheden for 3 grønne og 1 rødt lys?
    d) Hvad er sandsynligheden for 2 grønne og 2 røde lys?

13. Find ud fra resultaterne i opgave 7, sandsynligheden for at man ved at udfylde en tipskupon, får 13, 12 11 eller 10 rigtige.

14. Poker. Fra et spil kort trækkes 5 kort. Hvor mange muligheder er der for dette?
    Hvor mange muligheder er der for at få 5 i samme farve (en Flush).
    Hvad er sandsynligheden for at få en flush?

15. I en urne er der 6 røde og 4 hvide kugler. Der trækkes 3 kugler fra urnen. Find sandsynligheden for
    a) at der trækkes lutter røde kugler
    b) at der trækkes mindst 2 hvide kugler
    c) at der trækkes netop 1 hvid kugle

16. (svær opgave). Ved en prøvesmagning af øl, stilles 5 glas op i rækkefølge, hvor man skal gætte hvilket mærke de for skellige glas er. Under forudsætning af, at man gætter, skal man beregne sandsynligheden for at man gætter 0, 1, 2 3, og 5 rigtige.

Betinget sandsynlighed

17. To kort trækkes fra en bunke bestående af esserne og kongerne fra et sædvanligt kortspil. Find sandsynligheden for at begge kort er esser, når det oplyses at a) det ene kort er et es, b) det ene kort er et sort es, c) det ene kort er spar es

18. En læge mener, at der er 80% sandsynlighed for at en patient lider af sygdommen S. Han underkaster patienten en prøve, hvor der er 90% sandsynlighed for en positiv reaktion, hvis patienten lider af sygdommen S og 15% sandsynlighed for en positiv reaktion, hvis patienten lider ikke af sygdommen S.
    Hvilken sandsynlighed mener lægen der er for en positiv reaktion?
    Det viser sig, at patienten reagerer positivt. Hvilken sandsynlighed mener lægen nu der er for at patienten har sygdommen S?

19. 20% af befolkningen er venstrehåndede. af de venstrehåndede er 45% højre/venstre konfuse, mens det for de højrehåndede kun er 15%. Find sandsynligheden for at en tilfældig udvalgt person er højre/venstre konfus. En tilfældig udvalgt person viser sig at være højre/venstre konfus. Find sandsynligheden for at personen er højrehåndet.

20. Olga læger en bestemt kabale, som går op gennemsnitlig hver tyvende gang. Hun har lagt mærke til, at hvis hun laver et bestemt lille snyderi, er der 50% chance for at den går op. Olgas kabaler går op gennemsnitlig hver tiende gang. Hvad er sandsynligheden for at Olga snyder, når hun lægger kabale?

Binomialfordeling

21. En terning kastes 40 gange. Find sandsynligheden for
    a) at man slår netop 6 seksere.
    b) at man slår mindre end 4 seksere.
    c) at man slår mellem 4 og 7 seksere

22. Hvad er størst: Sandsynligheden for at få netop en sekser i 6 kast med en terning eller sandsynligheden for at få netop 2 seksere i 12 kast med en terning?

23. Ifølge Statistisk årbog er 73,3% af 43 årige mænd i live som 68 årige. 20 herrer er samlet til deres 25 års studenterjubilæum.
    Hvad er sandsynligheden for at den samme kreds vil kunne samles ved deres 50 års jubilæum?
    Hvad er det mest sandsynlige antal, der kan samles ved 50 års jubilæet?
    Hvad er sandsynligheden for at netop dette antal samles?

24. Under en epidemi af en bestemt kvægsygdom regnede man med, at sandsynligheden for smitte var 25%. 10 dyr blev vaccineret med en ny vaccine, og der viste sig, at kun et af disse dyr blev smittet. Er dette resultat signifikant, hvis man forlanger, at der blandt uvaccinerede dyr skal være under 5% sandsynlighed for et så gunstigt eller gunstigere resultat?

Stokastisk variabel

25. Sandsynlighedsfordelingen for en stokastisk variabel er givet ved

x_i	4	5	6	7	8
$P(X = x_i)$	0.01	0,08	0,26	0,41	0,24

a) Tegn et stolpediagram, samt det grafiske billede for fordelingsfunktionen for X
b) Find middelværdi og spredning for X

26. Fra en population på 10 elementer, hvoraf 4 er defekte, udtages uden tilbagelægning en stikprøve på 3 elementer. Find middelværdien af antallet af defekte elementer.
Find dernæst middelværdien af antallet af defekte elementer i stikprøven, når de udtages med tilbagelægning.
Beregn spredningen på antallet af defekte elementer i begge tilfælde.

27. Ved et lotteri med 50000 sedler er der 1 præmie på 5000 kr., 3 præmier på 1000 kr., 10 præmier på 100 kr., og 20 præmier på 50 kr. Når hver seddel koster 10 kr. skal man finde midddelværdi og spredning for gevinsten på en vilkårlig seddel .

Normalfordelingen

28. En normalfordelt stokastisk variabel X har middelværdi 432 og spredningen 5.Find sandsynligheden for, at X antager en værdi i intervallet [430;440], og Find sandsynligheden for, at X antager en værdi i intervallet [420;450]. Hvor stor skal k, for er der 90% sandsynlighed for, at X antager en værdi i intervallet [432-k;432+k].

29. Ved 50 målinger af en fysisk størrelse fandt man nedenstående målinger:

2,26 1,63 2,23 2,12 2,77 2,37 3,30 ,2,20 2,03 2,55 1,40 2,16 1,69 2,74 2,50 2,35 2,12 2,64 2,02 2,30 2,39 2,60 2,40 2,92 2,32 1,98 2,07 2,44 2,94 2,79 3,02 2,15 1,73 1,99 2,86 1,52 2,36 2,40 2,63 1,78 1,89 1,92 2,71 2,13 1,95 2,47 1,73 2,65 2,52 2,31

Vi at dette talmateriale er tilnærmelsesvis normalfordelt, og beregn middelværdi og spredning.

Indeks