MAT0100V
Sannsynlighetsregning og kombinatorikk

Forventning, varians og standardavvik
Tilnærming av binomiske sannsynligheter
Konfidensintervall

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Tilfeldige variabler

Når vi kaster to terninger er det 36 utfall

Vi ser på X = «sum antall øyne»

De *mulige verdiene* til X er 2, 3, … , 11, 12

Ved å telle opp antall gunstige utfall for hendelsen «$X = k$» kan vi bestemme $P(X = k)$ for k = 2, … , 12

(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)

«$X = 7$»

$P(X = 7) = 6/36$

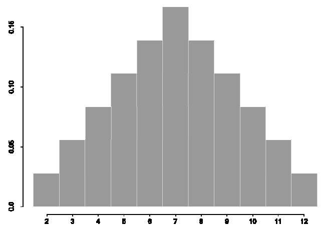
Vi får tabellen:

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

Tabellen gir *sannsynlighetsfordelingen* til X

Summen av sannsynlighetene i tabellen er lik én

Vi kan vise sannsynlighets-fordelingen med et stolpediagram



Oppgave 76 I en eske ligger det fire lapper som har numrene 1, 2, 3 og 4. Du trekker tilfeldig to lapper uten tilbakelegging.

a) Skriv opp alle utfallene for dette forsøket. Hva er sannsynligheten for hvert av utfallene?

b) La X være det høyeste tallet du får på en av lappene. Hvilke verdier kan denne tilfeldige variabelen få? Skriv opp de utfallene som gir hver av disse verdiene.

c) Bestem sannsynlighetsfordelingen til X.

Hypergeometrisk fordeling

Eksempel 7.1: I en kartong er det 12 sikringer

Fire av dem er defekte, resten er i orden

Vi trekker tilfeldig tre sikringer

X = «antall defekte sikringer vi trekker»

Da har X sannsynlighetsfordelingen

$$P(X = k) = \frac{\binom{4}{k} \cdot \binom{8}{3-k}}{\binom{12}{3}}$$

Generelt har vi følgende situasjon:

- Vi har en mengde med N elementer
  (I eksempel 7.1 er dette mengden av de 12 sikringene)
- Elementene i mengden kan deles inn i to delmengder D og $\bar{D}$
  Det er m elementer i D og $N - m$ elementer i $\bar{D}$
  (I eksempel 7.1 er de to delmengdene de defekte og de ikke-defekte sikringene. Det er 4 defekte sikringer og 12 - 4 = 8 som er i orden)
- Vi trekker tilfeldig n elementer fra mengden
  (I eksempel 7.1 trekker vi 3 sikringer)

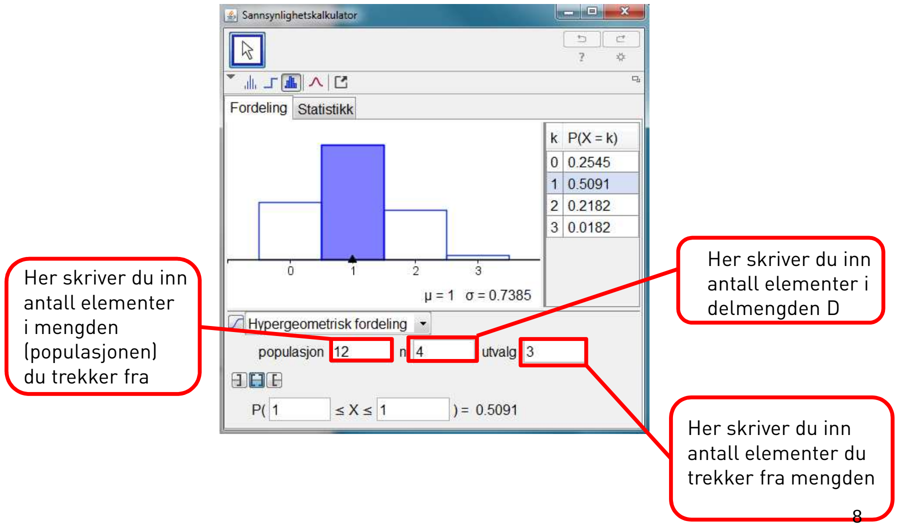
La X være antall elementer vi trekker fra D

X har sannsynlighetsfordelingen

$$P(X = k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}}$$

Vi sier at X er *hypergeometrisk fordelt*

Vi kan bruke sannsynlighetskalkulatoren i GeoGebra til å bestemme $P(X=k)$



Oppgave 80 Per trekker tilfeldig tre bokstaver fra alfabetet. La X være antall konsonanter han får.

a) Skriv opp en formel for sannsynlighetsfordelingen til X.

b) Bruk formelen til å finne sannsynligheten for at Per vil få (i) bare konsonanter; (ii) to konsonanter og én vokal; (iii) én konsonant og to vokaler; (iv) bare vokaler.

Kontroller svarene i b med GeoGebra

Binomisk fordeling

Eksempel 7.3: I en søskenflokk er det fire barn

X = «antall gutter i søskenflokken»

Vi vil finne $P(X = 2)$

Det er seks rekkefølger av barna som gir to gutter:
GGJJ, *GJGJ*, *GJJG*, *JGGJ*, *JGJG* og *JJGG*

Vi antar at barnas kjønn er uavhengig av hverandre

$$P(GGJJ) = 0.514 \cdot 0.514 \cdot 0.486 \cdot 0.486 \;= 0.514^2 \cdot 0.486^2$$

Hver av de fem andre rekkefølgene som gir to gutter har også sannsynlighet $0.514^2 \cdot 0.486^2$

Vi finner $P(X=2)$ ved å legge sammen sannsynlighetene for de seks rekkefølgene som gir to gutter og to jenter:

$$P(X = 2) = 6 \cdot 0.514^2 \cdot 0.486^2 \;= 0.374$$

Det er seks rekkefølger som gir to gutter og to jenter

Merk at å velge en bestemt rekkefølge av to gutter og to jenter, er det samme som å velge 2 plasser av 4 der det skal stå *G*

$$\boxed{G}\quad \boxed{}\quad \boxed{G}\quad \boxed{}$$

Det kan vi gjøre på $\binom{4}{2} = 6$ måter

Generelt har vi følgende situasjon:

- Vi gjør n forsøk
- I hvert forsøk er det to muligheter:
  Enten inntreffer en bestemt hendelse S
  ellers så gjør den ikke det
- I hvert forsøk er sannsynligheten lik p for at S skal inntreffe
- Forsøkene er uavhengige

La X være antall ganger S inntreffer i de n forsøkene

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

X er *binomisk fordelt*

Eksempel 7.4. En bestemt type frø spirer med 70% sannsynlighet

Vi sår 20 frø. Hva er sannsynligheten for at nøyaktig *k* frø vil spire?

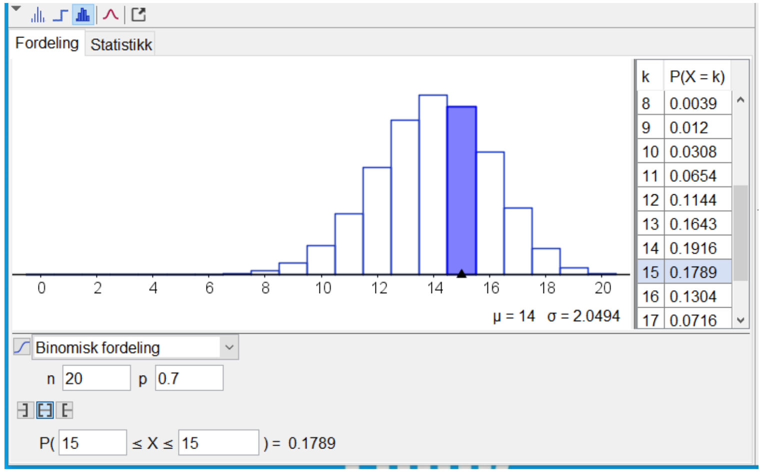
La *X* være antall før som spirer

Hvis frøene spirer uavhengig av hverandre, er *X* binomisk fordelt med *n* = 20 og *p* = 0.70

$$P(X = k) = \binom{20}{k} 0.70^k \, 0.30^{20-k}$$

GeoGebra

Vi kan bruke sannsynlighetskalkulatoren i GeoGebra til å bestemme *P*(*X*=*k*)



Oppgave 83 I en klasse holdes en flervalgsprøve med 10 spørsmål. Prøven gjennomføres ved at elevene krysser av ved ett av tre alternativ (hvorav ett er riktig) for hvert spørsmål. Læreren bestemmer seg for å gi karakteren 6 hvis alle spørsmålene er riktig besvart og karakteren 5 hvis 8 eller 9 spørsmål er riktig besvart.

Per har ikke lest på leksene og krysser helt tilfeldig av for hvert spørsmål.

a) La X være antall riktige svar Per får. Forklar hvorfor X er binomisk fordelt med $n = 10$ og $p = 1/3$.

b) Hva er sannsynligheten for at Per får karakteren 6?

c) Hva er sannsynligheten for at Per får karakteren 5?

Kontroller svarene i b og c med GeoGebra

Oppgave 85 I en urne er det tre røde kuler og to hvite kuler. To kuler trekkes tilfeldig ut, henholdsvis uten og med tilbakelegging. La X angi antall røde kuler vi trekker.

a) Hva gir binomisk fordeling for X, trekking med eller uten tilbakelegging?

b) Beregn $P(X = x)$ for $x = 0, 1, 2$ både for trekning med og uten tilbakelegging. Bruk GeoGebra.

Anta så at det er 300 røde kuler og 200 hvite kuler i urnen. Vi lar fortsatt X angi antall røde kuler i to trekninger.

c) Beregn også nå $P(X = x)$ for $x = 0, 1, 2$ for trekning med og uten tilbakelegging. Sammenlign med det du fant i punkt a. Bruk GeoGebra.

Resultatet i punkt c illustrerer at hvis vi trekker et (relativt sett) lite utvalg fra en (relativt sett) stor populasjon, så er det ubetydelig forskjell på trekning med og uten tilbakelegging.

Ved en meningsmåling om holdningen til norsk EU-medlemskap trekkes det tilfeldig (og uten tilbakelegging) et utvalg på 1000 av befolkningen over 18 år. La X være antallet i utvalget som er mot norsk EU-medlemskap.

d) Hvorfor er det rimelig å anta at X er binomisk fordelt med $n = 1000$ og p lik (den ukjente) andelen av den voksne befolkningen som er mot norsk EU-medlemskap?

Forventningsverdi

Sannsynlighetsfordelingen til en tilfeldig variabel X gir sannsynligheten for de ulike verdiene X kan anta

Vi ønsker i tillegg et summarisk mål som forteller oss hvor fordelingen er «plassert» på tallinja

Forventningsverdien er et slikt summarisk mål

Vi vil bruke rulett som motivasjon (avsnitt 8.1)

Ruletthjulet har 37 felt som er nummerert fra 0 til 36

Når ruletthjulet snurrer slippes en liten kule oppi

Kula blir liggende på ett av de 37 nummererte feltene når hjulet stopper

Feltene 1 - 36 er **røde** eller **sorte**, mens 0 er **grønt**

Spillerne setter sin innsats på grupper av felt (det er ikke lov å satse på 0)

Hvis en spiller satser et beløp på k felt og kula stopper på et av dem, vinner spilleren og hun får utbetalt $36/k$ ganger innsatsen

Vi ser på en «forsiktig» spiller som satser 10 euro på 18 felt (f. eks. de røde)

Spilleren får 20 euro hvis hun vinner og ingenting hvis hun taper. Uansett beholder kasinoet innsatsen på 10 euro

Spillerens nettogevinst i en spilleomgang er 10 euro hvis hun vinner, og den er -10 euro hvis hun taper

Kvinnen spiller tre omganger på denne måten

La Y være hennes samlede nettogevinst i de tre omgangene

Sannsynlighetsfordelingen til Y :

$$P(Y=-30) = \left(\frac{19}{37}\right)^3 = 0.135$$ (taper 3 ganger)

$$P(Y=-10) = 3\cdot\frac{18}{37}\cdot\left(\frac{19}{37}\right)^2 = 0.385$$ (vinner 1 gang og taper 2 ganger)

$$P(Y=10) = 3\cdot\left(\frac{18}{37}\right)^2\cdot\frac{19}{37} = 0.365$$ (vinner 2 ganger og taper 1 gang)

$$P(Y=30) = \left(\frac{18}{37}\right)^3 = 0.115$$ (vinner 3 ganger)

Anta at kvinnen kveld etter kveld spiller tre omganger rulett. Hva blir hennes gjennomsnittlige nettogevinst i «det lange løp»?

Anta at nettogevinstene de 10 første kveldene blir -10, 10, 30, 10, 10, 10, -10, -30, -10 og 10

Gjennomsnittlig nettogevinst:

$$\frac{1}{10}(-10+10+30+10+10+10-10-30-10+10)$$

$$= -30\cdot\frac{1}{10} - 10\cdot\frac{3}{10} + 10\cdot\frac{5}{10} + 30\cdot\frac{1}{10}$$

Relative frekvenser av de mulige verdiene av nettogevinsten

Gjennomsnittlig nettogevinst etter N kvelder:

$$-30\cdot r_N(-30) - 10\cdot r_N(-10) + 10\cdot r_N(10) + 30\cdot r_N(30)$$

Relative frekvenser av de mulige verdiene av nettogevinsten

Hvis spilleren spiller veldig mange kvelder, vil de relative frekvensene nærme seg de tilsvarende sannsynlighetene, og gjennomsnittet vil nærme seg

$$-30\cdot P(Y=-30) - 10\cdot P(Y=-10)$$
$$+10\cdot P(Y=10) + 30\cdot P(Y=30) = -0.81$$

Denne summen kaller vi forventningsverdien til Y
Den skriver vi $E(Y)$

Ruletteksempelet motiverer definisjonen:

En tilfeldig variabel X har mulige verdier $x_1, x_2, \ldots, x_m$. Da er forventningsverdien

$$E(X) = x_1\cdot P(X=x_1) + \ldots.. + x_m\cdot P(X=x_m)$$

Vi sier ofte forventning i stedet for forventningsverdi

Den greske bokstaven μ («my») brukes for å betegne forventningsverdi

Forventningen er «tyngdepunktet» i fordelingen

Eksempel 8.1: Vi kaster to terninger, og lar *X* være summen av antall øyne

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

Forventningsverdien blir:

$$E(X) = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36}$$
$$+ 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + 10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36}$$
$$= \frac{252}{36} = 7$$

Oppgave 87 Du kaster én terning. La X angi antall øyne du får. Finn forventningsverdien til X.

Oppgave 88 I en eske ligger det fire lapper som har numrene 1, 2, 3 og 4. Du trekker tilfeldig to lapper uten tilbakelegging. La X være det høyeste tallet du får på en av lappene. I oppgave 76 fant vi at X har sannsynlighetsfordelingen

k	2	3	4
$P(X=k)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

Finn forventningsverdien til X.

Store talls lov

Ruletteksemplet motiverer også store talls lov:

Vi har et forsøk med en tilfeldig variabel *X*. Hvis vi gjentar forsøket mange ganger, vil gjennomsnittet av verdiene til *X* nærme seg forventningsverdien *E*(*X*)

Store talls lov er blant annet grunnlaget for kasinodrift og forsikringsvirksomhet

Oppgave 90 En rekke personer forsikrer sine sykler i et forsikringsselskap. Vi antar for enkelhets skyld at en person bare kan få erstatning av selskapet av to grunner: *(i)* sykkelen blir stjålet og dukker ikke opp igjen, eller *(ii)* den blir stjålet, men kommer senere til rette delvis ramponert. I det første tilfellet får den forsikrede en erstatning på 3500 kroner (etter at egenandel er trukket fra), mens han i det andre tilfellet får 1000 kroner. Vi antar at sannsynligheten for *(i)* er 2%, mens sannsynligheten for *(ii)* er 5%.

a) Forklar hvorfor *forventet* erstatningsutbetaling er en rettferdig nettopremie for forsikringen (dvs. når vi ser bort fra administrasjonsomkostninger, fortjeneste, o.l.).
b) Finn denne nettopremien.

Forventning for binomisk fordeling

Eksempel 8.3: I en søskenflokk er det fire barn

X = «antall gutter i søskenflokken» er binomisk fordelt med n=4 og p=0.514

Av formelen for binomisk fordeling får vi:

$P(X=0)=0.056 \quad P(X=1)=0.236$

$P(X=2)=0.374 \quad P(X=3)=0.264 \quad P(X=4)=0.070$

Forventningsverdien blir

$$E(X)=0\cdot 0.056+1\cdot 0.236+2\cdot 0.374+3\cdot 0.264+4\cdot 0.070$$
$$=2.06$$

I eksemplet fant vi $E(X)$ = 2.06

Merk at $4 \cdot 0.514 = 2.06$

Forventningen er antall barn ganger sannsynligheten for at et barn er en gutt

Generelt:

Hvis X er binomisk fordelt, er $E(X) = np$

Eksempel 8.4. En bestemt type frø spirer med 70% sannsynlighet. Vi sår 20 frø
Forventet antall frø som spirer er $20 \cdot 0.70 = 14$

Oppgave 92 Vi vil se på noen binomiske situasjoner:

a) Du kaster et kronestykke 75 ganger. Finn forventet antall mynt.
b) Du kaster en terning 30 ganger. Finn forventet antall seksere.
c) Ved et sykehus blir det en uke født 20 barn. Finn forventet antall gutter.

Forventningen til *a* + *bX*

La X være en tilfeldig variabel knyttet til et forsøk. Da er $Y = a + bX$ en ny tilfeldig variabel knyttet til det samme forsøket

I det lange løp er gjennomsnittlig X-verdi lik $E(X)$

Dermed er gjennomsnittlig Y-verdi lik $a + b\,E(X)$

Det gir:

$$E(a+bX) = a + b\,E(X)$$

Eksempel 8.5: Vi ser på den «forsiktige» rulett-spilleren som tre ganger satser 10 euro på 18 felt

La X være antall ganger hun vinner

X er binomisk fordelt med n = 3 og p = 18/37

$$E(X) = 3 \cdot \frac{18}{37} = \frac{54}{37}$$

Samlet nettogevinst: $Y = -30 + 20X$

Dermed:

$$E(Y) = E(-30 + 20X) = -30 + 20E(X)$$
$$= -30 + 20 \cdot \left(\frac{54}{37}\right) = -\frac{30}{37} = -0.81$$

Oppgave 96 Du kaster én terning. La X være antall øyne du får, og la Y være det dobbelte av antall øyne.

a) Bestem sannsynlighetsfordelingen til Y.

b) Bruk sannsynlighetsfordelingen fra oppgave a til å finne forventningsverdien til Y.

c) Hva er sammenhengen mellom $E(X)$ og $E(Y)$?

Oppgave 97 En bedrift produserer pastiller. La X være antall pastiller i en eske. Tabellen gir sannslynlighetsfordelingen til X:

k	23	24	25	26
$P(X = k)$	0.10	0.30	0.40	0.20

a) Finn forventningsverdien til X.

En pastill veier 0.8 gram. Esken pastillene ligger i, veier 4.0 gram. Vekten av en eske med pastiller er da $V = 4.0 + 0.8 \cdot X$ gram.

b) Finn forventet vekt av en eske med pastiller.

Varians

Forventningsverdien til en tilfeldig variabel X forteller oss hva gjennomsnittlig X-verdi vil bli i det lange løp

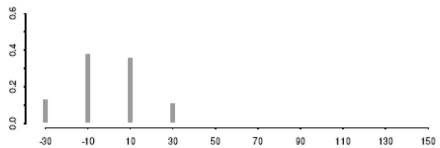
Vi ønsker oss også et summarisk mål som sier noe om hvor mye verdien til en tilfeldig variabel vil variere fra forsøk til forsøk

Variansen er et slikt mål

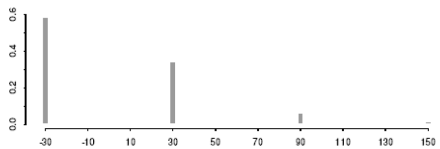
Vi bruker igjen rulett som motivasjon

Vi ser på den «forsiktige» spilleren som tre ganger satser 10 euro på 18 felt og på en annen litt «dristigere» spiller som tre ganger satser 10 euro på 6 felt

Figuren viser fordelingen for nettogevinsten for de to spillerne:



«Forsiktig» spiller (Y)



«Dristig» spiller (Z)

Nettogevinsten *Y* for den «forsiktige» spilleren og nettogevinsten *Z* for den «dristige» spilleren har begge forventningsverdi $\mu = -30/37 = -0.81$

Men fordelingen til *Z* er mer «spredt ut» enn fordelingen til *Y*

For å få et mål på hvor mye fordelingen til *Y* er «spredt ut» tar vi utgangspunkt i kvadratavvikene mellom *Y*-verdiene og forventningsverdien

Hvis *Y* får verdien -30 er kvadratavviket

$$(-30-\mu)^2 = (-30+30/37)^2 = 852.0$$

Hvis den «forsiktige» spilleren om og om igjen spiller tre omganger rulett, gir det samme argumentet som vi brukte i forbindelse med forventningsverdi, at det gjennomsnittlige kvadratavviket vil nærme seg

$$(-30-\mu)^2 \cdot P(Y=-30) + (-10-\mu)^2 \cdot P(Y=-10)$$
$$+(10-\mu)^2 \cdot P(Y=10) + (30-\mu)^2 \cdot P(Y=30) = 300$$

Denne summen kaller vi variansen til *Y*
Den skriver vi Var(*Y*). Altså Var(*Y*)=300

For den «dristige » spilleren får vi tilsvarende at Var(*Z*)=1467

Ruletteksempelet motiverer definisjonen:

En tilfeldig variabel *X* har mulige verdier

$x_1, x_2, \ldots, x_m$ og forventningsverdi μ

Da er variansen

$$\text{Var}(X) = (x_1-\mu)^2 \cdot P(X=x_1) + \ldots. + (x_m-\mu)^2 \cdot P(X=x_m)$$

Ofte bruker en σ^2 for å betegne varians

Eksempel 9.1: Vi kaster to terninger, og lar *X* være summen av antall øyne

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

Vi har funnet $E(X) = 7$

Variansen blir:

$$\text{Var}(X) = (2-7)^2 \cdot \frac{1}{36} + (3-7)^2 \cdot \frac{2}{36} + (4-7)^2 \cdot \frac{3}{36}$$
$$+ \ldots. + (11-7)^2 \cdot \frac{2}{36} + (12-7)^2 \cdot \frac{1}{36}$$
$$= \frac{210}{36} = \frac{35}{6}$$

Oppgave 98 Du kaster én terning. La X angi antall øyne du får. Vi fant i Oppgave 87 at $E(X) = 7/2$. Bestem $\mathrm{Var}(X)$.

Oppgave 99 I en eske ligger det fire lapper som har numrene 1, 2, 3 og 4. Du trekker tilfeldig to lapper uten tilbakelegging. La X være det høyeste tallet du får på en av lappene. I oppgave 76 fant vi at X har sannsynlighetsfordelingen

k	2	3	4
$P(X=k)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

I oppgave 88 fant vi at $E(X) = 10/3$. Finn variansen til X.

Standardavvik

Nettogevinsten til den «forsiktige» spilleren har varians 300

Benevningen for variansen er «kvadrateuro»

Et mål for spredning som har «riktig» benevning er standardavvik:

Standardavviket til en tilfeldig variabel *X* er gitt ved $SD(X) = \sqrt{\mathrm{Var}(X)}$

Ofte bruker en σ for å betegne standardavvik

Nettogevinsten til den «forsiktige» spilleren har standardavvik 17.30 euro

Varians for binomisk fordeling

Eksempel 9.2: I en søskenflokk er det fire barn

X = «antall gutter i søskenflokken» er binomisk fordelt med *n*=4 og *p*=0.514

Har fra før at

$$P(X=0) = 0.056 \quad P(X=1) = 0.236$$
$$P(X=2) = 0.374 \quad P(X=3) = 0.264 \quad P(X=4) = 0.070$$
$$\mu = E(X) = 2.06$$

Variansen blir

$$\mathrm{Var}(X) = (0-\mu)^2 \cdot 0.056 + (1-\mu)^2 \cdot 0.236 + (2-\mu)^2 \cdot 0.374$$
$$+ (3-\mu)^2 \cdot 0.264 + (4-\mu)^2 \cdot 0.070 = 1.00$$

I eksemplet fant vi (avrundet) Var(*X*) = 1.00

Merk at 4 · 0.514 · (1 - 0.514) = 1.00 (avrundet)

Kan vise at vi har generelt:

Hvis *X* er binomisk fordelt, er Var(*X*) = *np*(1-*p*)

Eksempel 9.3. En bestemt type frø spirer med 70% sannsynlighet. Vi sår 20 frø

Variansen til antall frø som spirer er 20 · 0.70 · 0.30 = 4.20

Oppgave 102 Vi vil se på noen binomiske situasjoner:

a) Du kaster et kronestykke 75 ganger. Finn varians og standardavvik for antall mynt.

b) Du kaster en terning 30 ganger. Finn varians og standardavvik for antall seksere.

c) Ved et sykehus blir det en uke født 20 barn. Finn varians og standardavvik for antall gutter.

Variansen til *a* + *bX*

La *X* være en tilfeldig variabel med forventningsverdi μ_X

Y = *a* + *bX* har forventningsverdi

$$\mu_Y = a + b\mu_X$$

Kvadratavviket for *Y* blir

$$(Y-\mu_Y)^2 = (a+bX-\{a+b\mu_X\})^2 = b^2(X-\mu_X)^2$$

Det motiverer resultatet:

$$\boxed{\mathrm{Var}(a+bX) = b^2\,\mathrm{Var}(X)}$$

Eksempel 9.4: Vi ser på den «forsiktige» rulett-spilleren som tre ganger satser 10 euro på 18 felt

La *X* være antall ganger hun vinner

X er binomisk fordelt med *n* = 3 og *p* = 18/37

$$\mathrm{Var}(X) = 3 \cdot \frac{18}{37} \cdot \frac{19}{37} = 0.749$$

Samlet nettogevinst: $Y = -30 + 20X$

Dermed:

$$\mathrm{Var}(Y) = \mathrm{Var}(-30+20X) = 20^2\,\mathrm{Var}(X)$$

$$= 400 \cdot 0.749 = 300$$

Oppgave 106 Du kaster en terning. La X være antall øyne du får, og la Y være det dobbelte av antall øyne. Du fant sannsynlighetsfordelingen til Y og $E(Y)$ i oppgave 96.

a) Bruk sannsynlighetsfordelingen til Y til å finne $\mathrm{Var}(Y)$.

b) Hva er sammenhengen mellom $\mathrm{Var}(X)$ og $\mathrm{Var}(Y)$?

Oppgave 107 Se på oppgave 97. La X være antallet pastiller i en eske og la $V = 4.0 + 0.8 \cdot X$ være vekten til esken med innhold (i gram).

a) Finn variansen til X.

b) Finn variansen til V. Hvilken benevning har denne variansen?

c) Finn standardavviket til V. Hvilken benevning har dette standardavviket?

Oppgave 108 La X være en tilfeldig variabel med forventningsverdi μ og standardavvik σ. Vi innfører den *standardiserte* variablen $Z = (X-\mu)/\sigma$. Vis at $E(Z) = 0$ og $SD(Z) = 1$.

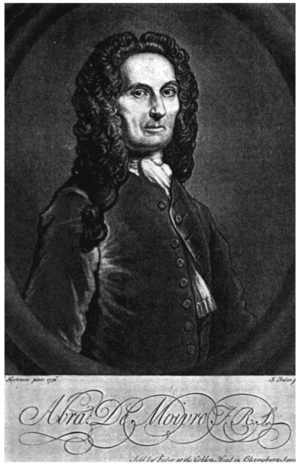
Tilnærming av binomiske sannsynligheter

Tidligere var det vanskelig å bruke formelen for binomisk fordeling til å regne ut sannsynligheter når *n* er stor

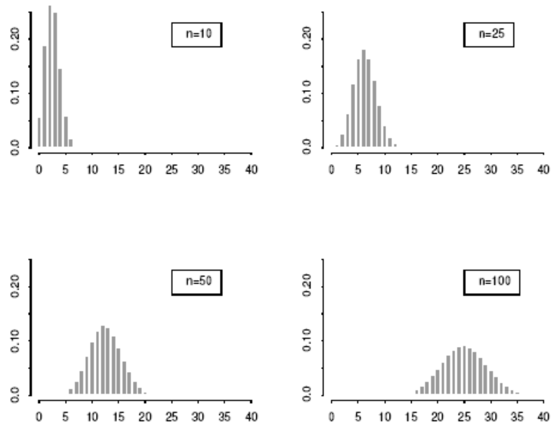
Alt i 1733 viste Abraham de Moivre hvordan en kan finne tilnærmingsverdier for binomiske sannsynligheter

Selv om det nå er enklere å bestemme binomiske sannsynligheter, er denne tilnærmelsen fortsatt viktig

(www.york.ac.uk/depts/maths/histstat)



Binomisk fordeling for p = 0.25 og n = 10, 25, 50, 100



Fordelingen forskyves mot høyre og blir mer «spredt ut» når *n* øker

For å finne en tilnærming «forskyver» vi fordelingene slik at de får « tyngdepunktet» i origo, og vi «skalerer» dem slik at de får samme spredning

Vi ser derfor på den standardiserte variabelen

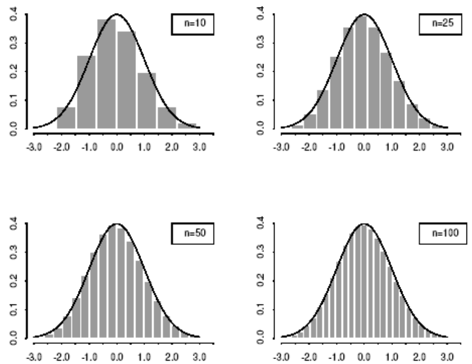
$$Z = \frac{X - E(X)}{SD(X)} = \frac{X - np}{\sqrt{np(1-p)}}$$

Vi har at E(Z) = 0 og SD(Z) = 1

Vi merker oss at hvis $X = k$ så er $Z = \frac{k-np}{\sqrt{np(1-p)}}$

Vi får derfor fordelingen til Z av fordelingen til X

Stolpediagram for fordelingen til Z



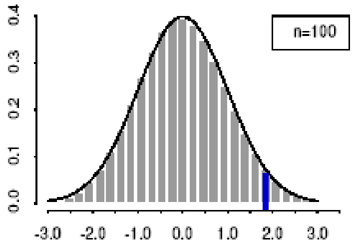
Arealet av en stolpe svarer til sannsynligheten for at Z får den aktuelle verdien

Stolpediagrammene nærmer seg standard-normalfordelingsfunksjonen $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Vil bruke de Moivres tilnærming til å finne $P(X \leq 33)$ når *n* = 100 og *p* = 0.25

Vi merker oss at

$$P(X \leq 33) = P\left(Z \leq \frac{33-100\cdot 0.25}{\sqrt{100\cdot 0.25\cdot 0.75}}\right) = P(Z \leq 1.85)$$



Vi skal egentlig summere arealene av alle søylene til venstre for 1.85

Summen av arealene av søylene er omtrent like stor som arealet under f(x) til venstre for 1.85

Arealet under standardnormalfordelings-funksjonen til venstre for **1.85** finner vi av tabellen bak i kompendiet:

	Siste desimal i z									
z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

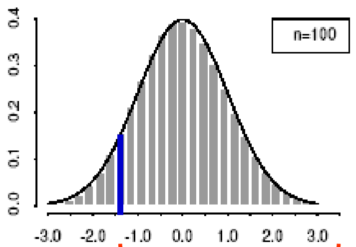
De Moivres tilnærming gir at

$$P(X \leq 33) = P(Z \leq 1.85) \approx 0.968$$

Vil så bruke de Moivres tilnærming til å finne $P(X \geq 19)$ når *n* = 100 og *p* = 0.25

Vi merker oss at

$$P(X \geq 19) = P\left(Z \geq \frac{19-100\cdot 0.25}{\sqrt{100\cdot 0.25\cdot 0.75}}\right) = P(Z \geq -1.39)$$



Vi skal egentlig summere arealene av alle søylene til høyre for -1.39

Summen av arealene av søylene er omtrent like stor som arealet under f(x) til høyre for -1.39

Arealet under standardnormalfordelings-funksjonen til venstre for **-1.39** finner vi av tabellen bak i kompendiet:

	Siste desimal i z									
z	0	1	2	3	4	5	6	7	8	9
−3	0.0013	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0000	0.0000
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

Arealet til høyre for -1.39 er lik én minus arealet til venstre for -1.39

Derfor

$$P(X \geq 19) = P(Z \geq -1.39) \approx 1 - 0.082 = 0.918$$

Oppgave 109 Tenk deg at du kaster en mynt 10000 ganger. Hva er sannsynligheten for at du får krone a) høyst 4930 ganger; b) minst 5100 ganger?

Gjør også oppgaven med GeoGebra uten å tilnærme og sammenlign svarene

Eksempel 10.3: Vi tenker oss at Arbeiderpartiet på et tidspunkt har oppslutning av 32.0% av velgerne

Et meningsmålingsinstitutt spør et tilfeldig utvalg på 1000 personer over 18 år hvilket parti de ville stemt på hvis det hadde vært valg

Hva er sannsynligheten for at mellom 300 og 340 av dem ville ha stemt på Arbeiderpartiet?

Med andre ord: hva er sannsynligheten for at Arbeiderpartiets oppslutning på meningsmålingen vil bli mellom 30.0% og 34.0% ?

La X være antallet av de spurte som ville ha stemt på Arbeiderpartiet

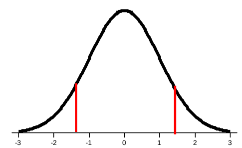
Siden det trekkes uten tilbakelegging, er strengt tatt X hypergeometrisk fordelt

Men da antallet som trekkes ut er lite i forhold til antall over 18 år i hele befolkningen, kan vi regne som om X er binomisk fordelt med n = 1000 og p = 0.32

Nå har vi at

$$P(300 \le X \le 340)$$
$$= P\left(\frac{300 - 1000 \cdot 0.32}{\sqrt{1000 \cdot 0.32 \cdot 0.68}} \le Z \le \frac{340 - 1000 \cdot 0.32}{\sqrt{1000 \cdot 0.32 \cdot 0.68}}\right)$$
$$= P(-1.36 \le Z \le 1.36)$$

Arealet under standard-normalfordeligsfunksjonen mellom -1.36 og 1.36 er lik arealet til venstre for 1.36 minus arealet til venstre for -1.36



	Siste desimal i z									
z	0	1	2	3	4	5	6	7	8	9
−3	0.0013	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0000	0.0000
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

De Moivres tilnærming gir at

$$P(300 \le X \le 340) = P\left(-1.36 \le Z \le 1.36\right)$$
$$\approx 0.913 - 0.087 = 0.826$$

Oppgave 110 En bestemt type frø spirer med 80% sannsynlighet. På et gartneri sår de 2000 frø og ser om de spirer. Hva er sannsynligheten for at a) høyst 1570 frø vil spire; b) minst 1630 frø vi spire c) mellom 1570 og 1630 frø vil spire?

Gjør også oppgaven med GeoGebra uten å tilnærme og sammenlign svarene

Sannsynlighetsregning og statistikk

Vi har sett på tilfeldige variabler og deres sannsynlighetsfordelinger. Det er en del av sannsynlighetsregningen

Vi vil nå se på hvordan sannsynlighetsregningen danner grunnlaget for statistiske metoder

Vi nøyer oss med å se på binomiske situasjoner

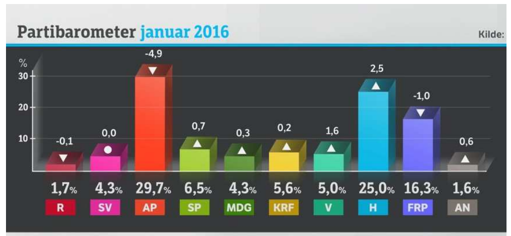
I sannsynlighetsregningen kjenner vi verdien til *p*

I statistikken gjør vi ikke det. Der er poenget nettopp å kunne si noe om verdien til *p* når vi har observert *X*

Estimering og konfidensintervall

Vi er ofte interessert i å anslå («estimere») verdien av *p* ut fra resultatet av et forsøk, og også å si noe om hvor presist anslaget er

Eksempel 11.1:
Av 721 personer som ville ha stemt hvis det hadde vært valg, ville 180 ha stemt på Høyre



Høyres oppslutning er 180/721=0.250, dvs 25.0%

Hvor sikkert er dette anslaget?

Generelt ser vi på en stor populasjon der en ukjent andel *p* har et bestemt "kjennetegn"

I eksemplet er populasjonen alle over 18 år som ville ha stemt hvis det var valg, og kjennetegnet er at en person ville stemt på Høyre

Vi trekker et tilfeldig utvalg på *n* individer fra populasjonen. Størrelsen av utvalget er liten i forhold til størrelsen av hele populasjonen

La *X* være antall i utvalget som har kjennetegnet

Vi kan regne som om *X* er binomisk fordelt med *p* lik den ukjente andelen i populasjonen som har kjennetegnet

Til å anslå («estimere») *p* bruker vi andelen i utvalget som har kjennetegnet, dvs.

$$\hat{p} = \frac{X}{n}$$

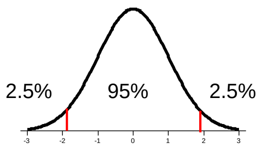
Merk at $\hat{p}$ («p hatt») er en tilfeldig variabel

I eksempelet fikk $\hat{p}$ verdien 0.250

For å kunne si noe om hvor presist et anslag er, må vi ta hensyn til hvor mye verdien av $\hat{p}$ vil variere fra undersøkelse til undersøkelse bare på grunn av tilfeldige variasjoner

Av de Moivres resultat finner vi at

$$P\left(-1.96 \le \frac{X-np}{\sqrt{np(1-p)}} \le 1.96\right) \approx 0.95$$



Nå er $\frac{X-np}{\sqrt{np(1-p)}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$

Dermed

$$P\left(-1.96 \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le 1.96\right) \approx 0.95$$

Kan vise at vi kan erstatte *p* med $\hat{p}$ i nevneren:

$$P\left(-1.96 \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le 1.96\right) \approx 0.95$$

Ulikhetene kan omformes slik at vi får *p* alene i midten:

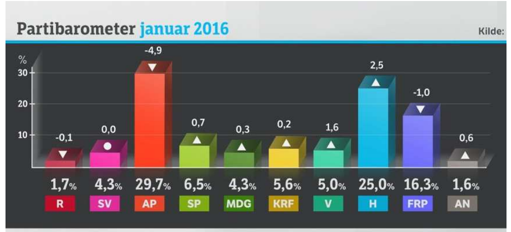
$$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 0.95$$

Det er altså tilnærmet 95% sannsynlig at undersøkelsen vil gi et resultat som er slik at *p* blir liggende i intervallet

$$\left[\ \hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \ \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ \right]$$

Dette intervallet kaller vi et (tilnærmet) 95% konfidensintervall for *p*

Eksempel 11.2:
Vi ser igjen på meningsmålingen



Vårt estimat for Høyres oppslutning er:

$$\hat{p} = \frac{180}{721} = 0.250$$

95% konfidensintervall:

$$\left[\ 0.250 - 1.96\sqrt{\frac{0.250\cdot(1-0.250)}{721}}\ ,\ \ 0.250 + 1.96\sqrt{\frac{0.250\cdot(1-0.250)}{721}}\ \right]$$

Dvs.: $[\ 0.218\ ,\ 0.282\ ]$ (dette gir en «feilmargin»)

Fakta om januarmålingen

- Resultatet må tolkes innenfor feilmarginer som varierer mellom 0,9 og 3,3 prosentpoeng.
- Arbeiderpartiet har en feilmargin på 3,3 prosentpoeng, med en øvre grense på 33 prosent og en nedre grense på 26,4 prosent.
- Høyre har en feilmargin 3,2 prosentpoeng, med en øvre grense på 28,1 prosent og en nedre grense på 21,8 prosent.
- 955 personer er blitt spurt og 721 har avgitt partipreferanse.
- Undersøkelsen er gjort mellom 5. og 11. januar.