MAT0100V
Sannsynlighetsregning og kombinatorikk

Tilfeldige variabler og sannsynlighetsfordelinger (repetisjon)
Hypergeometrisk fordeling (repetisjon)
Binomisk fordeling
Forventningsverdi

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Tilfeldige variabler

Når vi kaster to terninger er det 36 utfall

Vi er ofte ikke interessert i de enkelte utfallene

Vi kan for eksempel bare være interessert i
X = «summen av antall øyne»

X er en *tilfeldig variabel*

(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)

De *mulige verdiene* til X er 2, 3, 4, … , 11, 12

Ved å telle opp antall gunstige utfall for hendelsen «$X = k$» kan vi bestemme $P(X = k)$ for k = 2, 3, … , 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)

«$X = 7$»

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

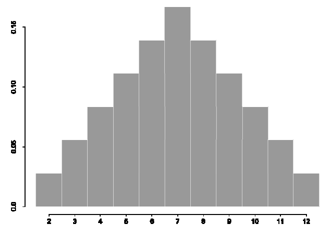
Vi får tabellen:

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



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»

Vi vil finne $P(X = k)$

Antall mulige utvalg $\binom{12}{3}$

Vi kan trekke k defekte på $\binom{4}{k}$ måter

Vi kan trekker 3 – k som er i orden på $\binom{8}{3-k}$ måter

Antall gunstige utvalg for k defekte (og 3 – k som er i orden) er $\binom{4}{k} \cdot \binom{8}{3-k}$

Sannsynlighetsfordelingen til X

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

Når vi setter inn k = 0, 1, 2, 3 får vi:

$$P(X = 0) = 0.255 \qquad P(X = 1) = 0.509$$

$$P(X = 2) = 0.218 \qquad P(X = 3) = 0.018$$

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

Ved å resonnere som i eksemplet finner vi at 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*

Oppgave 79 I en eske er det 6 røde og 4 blå kuler. Vi trekker tilfeldig fire av kulene. La X stå for antall røde kuler vi får.

a) Skriv opp en formel for sannsynlighetsfordelingen til X.

b) Bruk formelen til å finne sannsynligheten for at vi får
   (i) bare røde kuler
   (ii) 3 røde kuler og 1 blå kule
   (iii) 2 røde kuler og 2 blå kule

Oppgave 79 I en eske er det 6 røde og 4 blå kuler. Vi trekker tilfeldig fire av kulene. La X stå for antall røde kuler vi får.

a) Skriv opp en formel for sannsynlighetsfordelingen til X.

b) Bruk formelen til å finne sannsynligheten for at vi får
   (i) bare røde kuler
   (ii) 3 røde kuler og 1 blå kule
   (iii) 2 røde kuler og 2 blå kule

a)

$$P(X=k)=\frac{\binom{6}{k}\cdot\binom{4}{4-k}}{\binom{10}{4}} \qquad k = 0, 1, 2, 3, 4$$

b.i)

$$P(\text{bare røde kuler}) = P(X=4)$$

$$=\frac{\binom{6}{4}\cdot\binom{4}{0}}{\binom{10}{4}} = \frac{15\cdot 1}{210} = 0.071$$

b.ii)

$$P(\text{3 røde kuler og 1 blå kule}) = P(X=3)$$

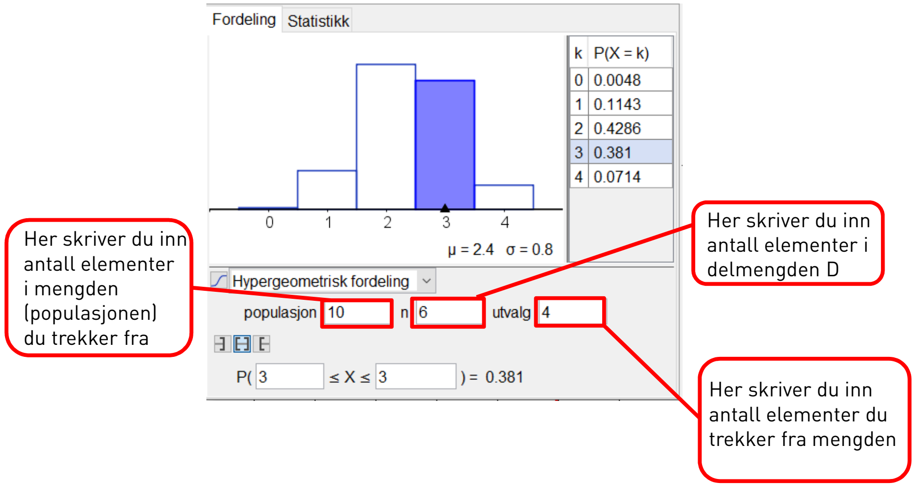
$$=\frac{\binom{6}{3}\cdot\binom{4}{1}}{\binom{10}{4}} = \frac{20\cdot 4}{210} = 0.381$$

b.iii)

$$P(\text{2 røde kuler og 2 blå kuler}) = P(X=2)$$

$$=\frac{\binom{6}{2}\cdot\binom{4}{2}}{\binom{10}{4}} = \frac{15\cdot 6}{210} = 0.429$$

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



Eksempel 7.2:

Når du tipper én lottorekke, krysser du av sju tall fra 1 til 34

Det trekkes tilfeldig ut 7 vinnertall

La *X* være antall riktige vinnertall du tipper

Sannsynlighetsfordelingen blir:

$$P(X = k) = \frac{\binom{7}{k} \cdot \binom{27}{7-k}}{\binom{34}{7}}$$

k	0	1	2	3	4	5	6	7
$P(X = k)$	0.165	0.385	0.315	0.114	0.019	0.0014	0.000035	0.00000019

Binomisk fordeling

Eksempel 7.3: I en søskenflokk er det fire barn

Hva er sannsynligheten for at det er to gutter og to jenter i søskenflokken?

La *X* = «antall gutter i søskenflokken»

Vi vil finne *P*(*X* = 2)

To eldste gutter, to yngste jenter: *GGJJ*

Eldste og yngste gutt, to midterste jenter: *GJJG*

Andre rekkefølger som gir to gutter og to jenter: *GJGJ, JGGJ, JGJG* og *JJGG*

Vi antar at barnas kjønn er uavhengig av hverandre

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

$$P(GJJG) = 0.514 \cdot 0.486 \cdot 0.486 \cdot 0.514 = 0.514^2 \cdot 0.486^2$$

Hver av de fire andre rekkefølgene som gir to gutter og to jenter har også sannsynligheten $0.514^2 \cdot 0.486^2$

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

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

Vi ser på eksempel 7.3

Der fant vi at det var seks rekkefølger som gir to gutter og to jenter

Det kunne vi funnet ut uten å skrive opp alle rekkefølgene

For å velge en bestemt rekkefølge er det samme som å velge 2 plasser blant 4 der det skal stå *G*

G		G	

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

Generelt har vi følgende situasjon:

- Vi gjør *n* forsøk
  (I eksempel 7.3 er hvert barn et «forsøk»)
- I hvert forsøk er det to muligheter:
  Enten inntreffer en bestemt begivenhet *S*
  ellers så gjør den ikke det
  (I eksempel 7.3 er hvert barn enten en gutt eller en jente)
- I hvert forsøk er sannsynligheten lik *p* for at *S* skal inntreffe
  (I eksempel 7.3 er sannsynligheten for gutt 51.4%)
- Forsøkene er uavhengige
  (I eksempel 7.3 har vi antatt uavhengighet siden vi ser bort fra tvillinger, osv.)

La *X* være antall ganger *S* inntreffer i de *n* forsøkene

Ved å resonnere som i eksempel 7.3 finner vi at *X* har sannsynlighetsfordelingen

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

Vi sier at *X* er *binomisk fordelt*

Oppgave 82 (litt endret)
Du kaster 6 kronestykker. La X være antall mynt du får.

a) Skriv opp en formel for sannsynlighetsfordelingen til X.

b) Bruk formelen til å finne sannsynligheten for at du vil få nøyaktig (i) tre mynt; (ii) fire mynt; (iii) fem mynt

a) $P(X=k)$

$$=\binom{6}{k}\cdot\left(\frac{1}{2}\right)^k\cdot\left(1-\frac{1}{2}\right)^{6-k}=\binom{6}{k}\cdot\left(\frac{1}{2}\right)^6=\binom{6}{k}\frac{1}{64}$$

b.i)

$$P(\text{tre mynt})=P(X=3)=\binom{6}{3}\frac{1}{64}=\frac{20}{64}=0.313$$

b.ii)

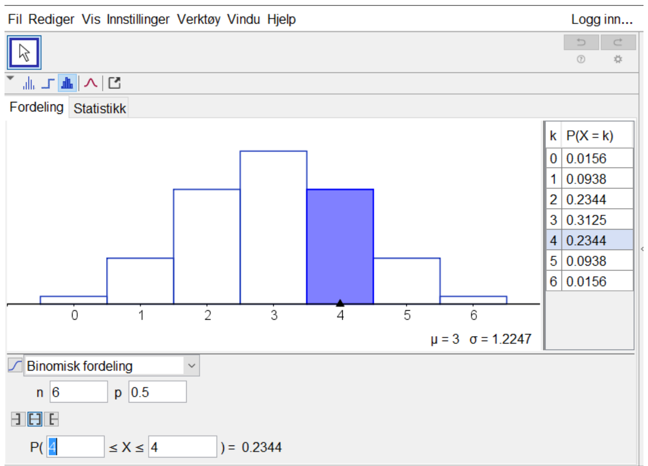
$$P(\text{fire mynt})=P(X=4)=\binom{6}{4}\frac{1}{64}=\frac{15}{64}=0.234$$

b.iii)

$$P(\text{fem mynt})=P(X=5)=\binom{6}{5}\frac{1}{64}=\frac{6}{64}=0.094$$

GeoGebra

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



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

Vi sår 20 frø. Hva er sannsynligheten for at nøyaktig 15 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

$$P(15 \text{ frø spirer}) = P(X = 15)$$

$$= \binom{20}{15} 0.70^{15}\, 0.30^{5} = 0.179$$

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:

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

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

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 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.

$$E(X) = 2 \cdot \frac{1}{6} + 3 \cdot \frac{2}{6} + 4 \cdot \frac{3}{6}$$
$$= \frac{2}{6} + \frac{6}{6} + \frac{12}{6} = \frac{20}{6} = \frac{10}{3}$$

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 administasjonsomkostninger, fortjeneste, o.l.).

b) Finn denne nettopremien.

b) *X* = utbetaling for tilfeldig valgt forsikret

$$E(X) = 0 \cdot 0.93 + 1000 \cdot 0.05 + 3500 \cdot 0.02$$

$$= 120$$