MAT0100V
Sannsynlighetsregning og kombinatorikk

Forventning, varians og standardavvik

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)

«$X = 7$»

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

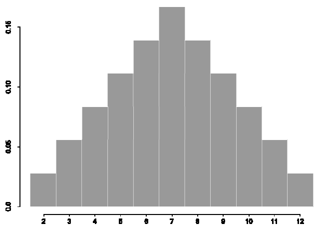
Vi får tabellen:

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



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 sannsynlig-hetene 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*

G		G	

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 $p = 0.70$

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

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

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

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.

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

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.

Sannsynlighetsfordelingen til X

k	0	1000	3500
$P(X=k)$	0.93	0.05	0.02

a) Ved store talls lov vil de forsikrede i gjennomsnitt få en erstatning på omtrent $E(X)$ kroner fra selskapet.

Hvis hver forsiktet betaler $E(X)$ kroner for forsikringen, vil innbetalingene til selskapet og utbetalingene fra selskapet være omtrent like store.

Derfor er $E(X)$ en rettferdig nettopremie

b)

$$E(X) = 0\cdot 0.93 + 1000\cdot 0.05 + 3500\cdot 0.02 = 120$$

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 · 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 · 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.

a) X = antall mynt

$$E(X)=np=75\cdot\frac{1}{2}=37.5$$

b) X = antall seksere

$$E(X)=np=30\cdot\frac{1}{6}=5$$

c) X = antall gutter

$$E(X)=np=20\cdot 0.514=10.3$$

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:

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

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:

$$E(Y) = E(-30+20X) \quad = -30 + 20E(X)$$

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

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.

a)

$$E(X) = 23 \cdot 0.10 + 24 \cdot 0.30 + 25 \cdot 0.40 + 26 \cdot 0.20 = 24.7$$

b) $E(V) = E(4.0 + 0.8X) = 4.0 + 0.8E(X)$

$$= 4.0 + 0.8 \cdot 24.7 = 23.8$$

Varians

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

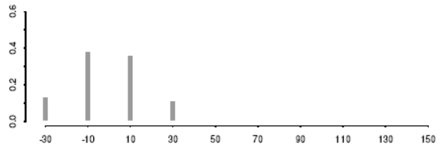
Vi ønsker oss også et summarisk mål som sier noe om hvor mye verdien til en tilfeldig variabel vil variere fra forsøk til forsøk

Variansen er et slikt mål

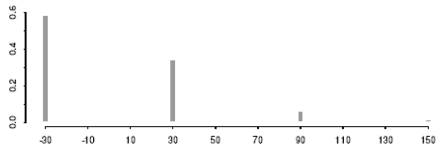
Vi bruker igjen rulett som motivasjon

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

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

Ofte bruker en σ^2 for å betegne varians

Fortsettelse følger på samlingen 23-24. april