STK1100 våren 2015

Binomisk fordeling

Svarer til avsnitt 3.5 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Vi starter med et eksempel:

I en søskenflokk er det fire barn som ikke er tvillinger, trillinger eller firlinger

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 sannsynlig-hetene for de seks rekkefølgene som gir to gutter og to jenter:

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

Vi så 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*

Eldste | Nest eldste | Nest yngste | yngste

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

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

Generelt har vi et binomisk forsøk:

- Vi gjør n forsøk $n=4$
  (I eksemplet er hvert barn et «forsøk»)
- I hvert forsøk er det to muligheter:
  Enten inntreffer S ellers så inntreffer F
  (I eksemplet er hvert barn enten en gutt eller en jente)
- Forsøkene er uavhengige
  (I eksemplet har vi antatt uavhengighet siden vi ser bort fra tvillinger, osv.)
- I hvert forsøk er sannsynligheten lik p for at S skal inntreffe og $1-p$ for at F skal inntreffe
  (I eksemplet er sannsynligheten for gutt 51.4% og sannsynligheten for jente 48.6%)

La X være antall ganger S inntreffer i et binomisk forsøk $\underbrace{SSSS}_{x}\ \underbrace{FFFF}_{n-x}$

Ved å resonnere som i eksemplet finner vi at X har punktsannsynlighet

$$b(x;n,p) = P(X=x) = \underbrace{\binom{n}{k}}_{\text{antall rekkefølger}} p^k (1-p)^{n-k}$$

for $x = 0,1,2,....,n$

Vi sier at X er *binomisk fordelt*

Vi skriver $X \sim \text{Bin}(n,p)$

Eksempel: En type frø spirer med 70% sannsynlighet

Vi sår 20 frø.
Hva er sannsynligheten for at 15 frø vil spire?
Hva er sannsynligheten for at minst 15 frø vil spire?

La X være antall frø som spirer

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

$$P(X=15) = \binom{20}{15} 0.70^{15} 0.30^{5} = 0.179$$

$$P(X \geq 15) = 1 - P(X \leq 14) = 0.416 \quad P(X=15)+P(X=16)+\dots+P(X=19)$$

(MATLAB-kommandoer er gitt på neste slide)

Binomisk fordeling med MATLAB

Punktsannsynlighet:

$$b(x;n,p) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

MATLAB: binopdf(x,n,p)

Kumulativ fordeling:

$$B(x;n,p) = P(X \leq x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}$$

MATLAB: binocdf(x,n,p)

Tabell A.1 bak i boka gir $B(x;n,p)$ for noen verdier av n og p

Vi vil ikke bry oss om disse tabellene

```
>> binopdf(15,20,0.70)

ans =

    0.1789

>> 1-binocdf(14,20,0.70)

ans =

    0.4164

>>
```

$P(X = 15) = 0{,}1789$

$P(X \geq 15) = 0{,}4164$

Eks. <u>Meningsmåling</u>

Anta at AP har oppslutning fra 25% av velgerene

Spør tilfeldige 1000

Hva er sannsynligheten AP's oppslutning i meningsmålingen blir høyst 22.5%

Setter X = antall som ville stemme AP

$X \sim Bin(\underset{n}{1000}, \underset{p}{0.25})$

AP's oppslutning blir høyst 22.5% hvis $X \leq 225$

Sannsynligheten for dette er

$P(X \leq 225) \underset{\substack{\uparrow \\ \text{MatLab}}}{=} 0.036$

Eksempel: Firebarnsmødre født 1935-1991

Passer binomisk fordeling (med p = 0.514)?

	Antall kvinner		
	Absolutt	Prosent	
Mødre som har født minst fire barn			
I alt	95 993	100,0	
4 jenter	6 048	6,3	5.9
3 jenter og 1 gutt	22 484	23,4	23.6
2 jenter og 2 gutter	34 095	35,5	37.4
1 jente og 3 gutter	25 446	26,5	26.4
4 gutter	7 920	8,3	7.0

Binomisk fordeling gir litt for få familier der barna har samme kjønn

http://www.ssb.no/a/samfunnsspeilet/utg/200903/01/tab-2009-06-15-03.html

Forventning og varians

Momentgenererende funksjon:

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \cdot p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = \left(pe^t + 1 - p\right)^n$$

Momenter:

$$\mu = E(X) = M_X'(0) = np$$

$$E(X^2) = M_X''(0) = n(n-1)p^2 + np$$

Varians:

$$\sigma^2 = E(X^2) - \mu^2 = np(1-p)$$