STK1100 våren 2023

Binomisk fordeling

Svarer til avsnitt 3.5 i læreboka

Matematisk institutt
Universitetet i Oslo

Binomisk fordeling

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*

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

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

Generelt har vi et binomisk forsøk:

- Vi gjør *n* forsøk
  (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

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

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

for $x = 0,1,2,\ldots,n$

Vi sier at X er *binomisk fordelt*

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

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

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

$$P(X \geq 15) = 1 - P(X \leq 14) = 0.416$$

Python-kommandoer er gitt på neste slide

Binomisk fordeling med Python

Punktsannsynlighet:

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

Python: import scipy.stats as stats

stats.binom.pmf(x,n,p)

Kumulativ fordeling:

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

Python: stats.binom.cdf(x,n,p)

Tabell A.1 i boka gir $B(x;n,p)$ for noen verdier av *n* og *p*. Vi vil ikke bry oss om disse tabellene

Eksempel: Meningsmåling

Anta at Arbeiderpartiet har oppslutning fra 19% av velgerne

Et meningsmålingsinstitutt spør et tilfeldig utvalg på 1000 personer hva de ville ha stemt hvis det var valg

Hva er sannsynligheten for at Arbeiderpartiets oppslutning på meningsmålingen blir mellom 17 og 21 prosent?

X = antall som ville ha stemt Arbeiderpartiet

$$X \sim Bin(1000, 0.19)$$

Arbeiderpartiets oppslutning på meningsmålingen blir mellom 17% og 21% hvis $170 \leq X \leq 210$

Sannsynligheten for dette er

$$P(170 \leq X \leq 210) = P(X \leq 210) - P(X \leq 169)$$
$$= 0.9494 - 0.0477 = 0.$$

Eksempel: Firebarnsmødre født 1935-1991

Passer binomisk fordeling med $n = 4$ og $p = 0.514$?

	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

```
import numpy as np
import scipy.stats as stats
x = np.arange(0,5)
stats.binom.pmf(x,4,0.514)
```

Kilde: http://www.ssb.no/a/samfunnsspeilet/utg/200903/01/

Forventning og varians

Momentgenererende funksjon:

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

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

Momenter:

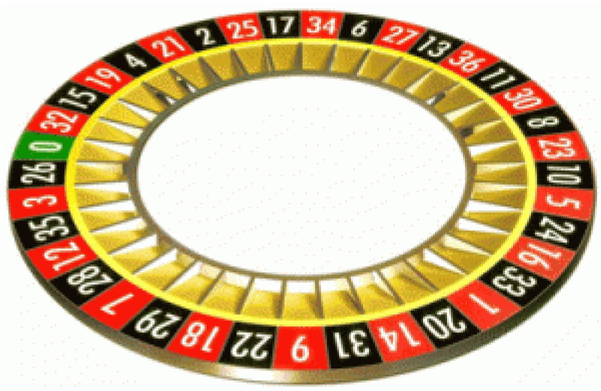
$$\mu = E(X) = M_X'(0) = np$$

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

Varians:

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

Eksempel: Rulett



En person spiller rulett *n* ganger

Hver gang satser spilleren
10 euro på *k* felt

Vi har tidligere sett på situasjonen der *n* er 3 og *k* er 6 eller 18

X = antall ganger spilleren vinner

$X \sim Bin(n, p)$ med p=$\frac{k}{37}$

$$E(X) = np = n\frac{k}{37}$$

$$V(X) = np(1-p) = n\frac{k}{37}\left(1-\frac{k}{37}\right)$$

Nettogevinsten til spilleren er

$$Y = 10\frac{36}{k}X - 10n$$

Vi får dermed at

$$E(Y) = 10\frac{36}{k}E(X) - 10n = -\frac{10}{37}n$$

$$V(Y) = \left(10\frac{36}{k}\right)^2 V(X) = \left(\frac{360}{37}\right)^2 \frac{37-k}{k}n$$

Hvis vi setter inn $n = 3$, $k = 18$ og $k = 6$, får vi resultatene vi har sett på tidligere