STK1100 våren 2018

Hypergeometrisk og negativ binomisk fordeling

Svarer til avsnitt 3.6 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Hypergeometrisk fordeling

Vi starter med et eksempel:

I en skål er 12 kuler: 4 blå og 8 røde

Vi trekker tilfeldig tre kuler

X = «antall blå kuler vi trekker»

Vi vil bestemme $P(X = x)$ for $x = 0, 1, 2, 3$

Antall mulige utvalg er $\binom{12}{3}$ Antall mulig / Antall gunstig

Vi kan trekke x blå kuler på $\binom{4}{x}$ måter

Vi kan trekker $3 - x$ røde kuler på $\binom{8}{3-x}$ måter

Antall gunstige utvalg for x blå kuler er $\binom{4}{x} \cdot \binom{8}{3-x}$

Punktsannsynligheten til X er gitt ved:

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

for x = 0, 1, 2, 3

Generelt har vi følgende situasjon:

- Vi har en mengde med N elementer
  (I eksemplet er dette mengden av de 12 kulene)
- Elementene i mengden er enten av typen S eller av typen F. Det er M elementer av typen S og $N - M$ elementer av typen F
  (I eksemplet er kulene blå eller røde. Det er 4 blå kuler og 12 – 4 = 8 røde kuler.)
- Vi trekker tilfeldig n elementer fra mengden uten tilbakelegging
  (I eksemplet trekker vi tre kuler)

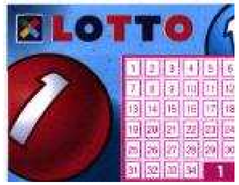
Mengden vi trekker fra kalles ofte en endelig populasjon og elementene i mengden er individer i populasjonen

La X være antall elementer vi trekker som er av typen S

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

$$h(x;n,M,N) = P(X=x) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$$

Vi sier at X er *hypergeometrisk fordelt*



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
Punktsannsynligheten for X er:

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

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

(MATLAB-kommandoer er gitt på neste slide)

Hypergeometrisk fordeling med MATLAB

Punktsannsynlighet:

$$P(X=x) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$$

MATLAB: hygepdf(x,N,M,n)

Kumulativ fordeling:

$$P(X\leq x) = \sum_{y\leq x}\frac{\binom{M}{y}\cdot\binom{N-M}{n-y}}{\binom{N}{n}}$$

MATLAB: hygecdf(x,N,M,n)

Forventning og varians

Momentgenererende funksjon er til liten hjelp for å finne forventning og varians for hypergeometrisk fordeling

Vi vil seinere vise at

$$E(X) = n\cdot\frac{M}{N}$$

$$V(X) = \frac{N-n}{N-1}\cdot n\cdot\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)$$

Negativ binomisk fordeling

Vi starter med et eksempel:

Vi kaster en terning til vi har fått 3 seksere

X = «antall ganger vi får fem eller mindre»

Vi vil bestemme $P(X=x)$ for $x = 0, 1, 2, 3, \ldots$

Vi får $X = x$ hvis vi

- i de x+2 første kastene får to seksere og fem eller mindre de andre x gangene
- i kast nummer x+3 får en sekser

Dermed er

$$P(X=x) = \binom{x+2}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^x \frac{1}{6}$$

Generelt har vi følgende situasjon:

- Vi har en rekke uavhengige forsøk
- I hvert forsøk er det to muligheter: Enten inntreffer S ellers så inntreffer F
- I hvert forsøk er sannsynligheten lik p for at S skal inntreffe og 1-p for at F skal inntreffe
- Vi stopper når S har inntruffet r ganger

Vi er interessert i

X = «antall F-er vi får»

Merk at at forsøk nr
$Y = X + r$

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

$$nb(x; r, p) = P(X=x) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

for x = 0, 1, 2, …..

Vi sier at X er *negativt binomisk fordelt*

MATLAB kommandoer:

Punktsannsynlighet: nbinpdf(x,r,p)
Kumulativ fordeling: nbincdf(x,r,p)

Forventning og varians

Momentgenererende funksjon:

Geometrisk
Binomialformel

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \cdot \underbrace{P(X=x)}_{\binom{x+r-1}{r-1} p^r (1-p)^x} = \frac{p^r}{\left(1-e^t(1-p)\right)^r}$$

Kumulantgenererende funksjon:

$$R_X(t) = \ln\{M_X(t)\} = r\ln(p) - r\ln\{1-e^t(1-p)\}$$

Forventning

$$\mu = E(X) = R_X'(0) = \frac{r(1-p)}{p}$$

Varians:

$$\sigma^2 = V(X) = R_X''(0) = \frac{r(1-p)}{p^2}$$

Kast en terning til vi har fått 3 seksere

F. eks.

F F S F F S F F S

X = antall ganger 5 eller mindre (antall F-er)

$X = 6$

Antall kast $= X + 3$

Vil finne $P(X = x)$ for 0, 1, 2, 3

For å få $X = x$ må vi ha følgende resultat

$\underbrace{\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot}_{x+2 \text{ kast}}$ S

↑

3 S-er

x F-er

2 S-er

Sannsynlighet for x F-er og 2 S-er i x+2 kast er (fra binomisk fordeling)

$$\binom{x+2}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^x$$

Sannsynlighet for S i x+3 kast er $\frac{1}{6}$

Dermed er

$$P(X=x) = \binom{x+2}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^x\frac{1}{6}$$

Generelt får vi $X=x$ hvis resultatet er som følgende

$$\underbrace{\cdot\ \cdot\ \cdot\ \cdot\ \cdot}\ 6$$

$x-(r-1)$ (r-te seks)

antall K-er antall seksere

Sannsynlighet p

Sannsynlighet

$$\binom{x-r-1}{r-1}p^{r-1}(1-p)^x$$

Dermed

$$P(X=x) = \binom{x-r-1}{r-1}p^r(1-p)^x$$

Eksempel

Anta at MDG har oppslutning av 10% av studenter

Spør tilfeldig studenter hva de ville stemt hvis det var valg

Hva er sannsynligheten for at vi må spørre minst 100 før vi får 10 som vil stemme MDG

X = antall som ikke stemmer MDG

$X \sim$ Negativ binomisk med $p = 0.1$ $\quad r = 10$

Vi er interessert i å finne $P(Y \geq 100)$

$$P(Y \geq 100) = P(X + 10 \geq 100) = P(X \geq 90) = 1 - P(X \leq 89) = 0{,}465$$

↑ Matlab

```
1-nbincdf(89, 10, 0.1)
```

```
ans = 0.4645
```

(12)

For negativ binomisk fordeling kan vise at

$$\mu_x(t) = \frac{p^r}{1-e^t(1-p)^r}$$

Enklest måte å se på $R_\alpha(t) = \ln\{\mu_x(\alpha)\}$

Får at

$$R_\alpha(t) = r\ln p - r\ln(1-e^t(1-p))$$

$$R'_\alpha(t) = +r\,\frac{e^t(1-p)}{1-e^t(1-p)}$$

$$R''(t) = \frac{r\,e^t(1-p)}{1+e^t(1-p)^2}$$

Dermed

$$E(X) = R_x'(0) = r\,\frac{1-p}{1-(1-p)} = \frac{r(1-p)}{p}$$

$\vdots$