FORELESNING 4 — ELE 3719 MATEMATIKK

EIVIND ERIKSEN

JAN 24 2013

BI

PLAN:

1. Diskrete stokastiske fordelinger
   a. Uniform
   b. Binomisk
   c. Geometrisk
   d. Poisson

Lærebok:

[R] 2.2

---

1. Diskrete stokastiske fordelinger

X : diskret stokastisk variabel

* $p(a) = f_X(a) = p(X=a)$: funksjon, sannsynlighetstettheten til X

* $F_X(b) = p(X \le b) = \sum_{a \le b} f_X(a)$: funksjon, kumulativ fordelingsfunksjon til X (cdf)

Krav til f_X:

i) $f_X(a) \geq 0$ for alle a

ii) $\sum_a f_X(a) = 1$

Krav til F_X:

i) $F_X(b)$ ikke-avtagende funksjon

ii) $\lim_{b \to \infty} F_X(b) = 1$

iii) $\lim_{b \to -\infty} F_X(b) = 0$

a) Uniform: Uniform(n), $n = 1, 2, 3, \ldots$

Mulige verdier for X: $1, 2, 3, \ldots, n$

Sannsynlighetstetthed: $f_X(a) = \frac{1}{n}$ for $a = 1, 2, \ldots, n$.

Eks:

Vi kaster en terning

X = antall øyne

$X \sim \text{Uniform}(6)$

x	$p(X=x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

b) Binomisk: Binom(n, p) $\begin{cases} n = 1, 2, 3, \ldots \\ 0 \leq p \leq 1 \end{cases}$

Spesialtilfelle: Bernoulli forsøk ($n=1$)

$$X = \begin{cases} 1, & \text{"suksess"} \\ 0, & \text{"fiasko"} \end{cases} \qquad \left.\begin{matrix} p(X=1) = p \\ p(X=0) = 1-p \end{matrix}\right\} \text{Binom}(1, p)$$

Generelt: Binomisk forsøksrekke

- et Bernoulli forsøk gjentas n ganger
- variabelen X = antall suksess
- forsøkene er uavhengige av hverandre

$\Big\}$ $X \sim \text{Binom}(n, p)$

n = antall gjentag.

p = sannsynlig. for suksess i et enkelt forsøk

Eks: Vi kaster mynt/kron 10 ganger

X = antall mynt

$n = 10$ gjentagelser

$p = 1/2$ sannsynlighet for mynt = suksess i et enkelt kast

$X \sim \text{Binom}(n, p)$

Mulige verdier for X: $0, 1, 2, \dots, n$

Sannsynlighetstetthet:

$$\boxed{f_X(i) = P(X=i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}}$$

Eks: $X \sim \text{Binom}(10, 1/2)$

$$p(X=7) = f_X(7) = \binom{10}{7} \cdot \left(\frac{1}{2}\right)^7 \cdot \left(\frac{1}{2}\right)^3 = \binom{10}{7} \cdot \left(\frac{1}{2}\right)^{10}$$

$$= 120/2^{10} \simeq \underline{\underline{0.117}}$$

Binomialkoeffisienter:

$$\binom{n}{r} := \frac{n!}{(n-r)! \cdot r!} = \frac{\cancel{1} \cdot \cancel{2} \cdot \cancel{3} \cdots \cancel{(n-r)} \cdots n}{\cancel{1} \cdot \cancel{2} \cdot \cancel{3} \cdots \cancel{(n-r)} \cdot 1 \cdot 2 \cdot 3 \cdots r}$$

$$= \frac{n \cdot (n-1) \cdots (n-r+1)}{r \cdot (r-1) \cdots 1}$$

$$\binom{10}{7} = \frac{10 \cdot 9 \cdot 8 \cdot \cancel{7 \cdot 6 \cdot 5 \cdot 4}}{\cancel{7 \cdot 6 \cdot 5 \cdot 4} \cdot 3 \cdot 2 \cdot 1} = \frac{10 \cdot \overset{3}{\cancel{9}} \cdot \overset{4}{\cancel{8}}}{\cancel{3} \cdot \cancel{2} \cdot 1} = \underline{\underline{120}}$$

På kalkulator: $\binom{n}{r} = {}_nC_r$

Husk: $0! = 1! = 1$

Hvorfor gir dette riktige sannsynligheter?

MMMMMMMKKK → sannsynlighet $p \cdot p \cdots p \cdot (1-p)(1-p)(1-p)$
MKMKMMKMMM ↗ $= p^7 \cdot (1-p)^3$

⋮

antall omstendigheter $= \binom{n}{r} = \binom{10}{7} = 120$

en rekke utfall som gir $X=7$
antall $= \binom{n}{r}$

total sannsynlighet for $X=7$

$$\binom{10}{7} p^7 (1-p)^3 = \underline{120} \cdot p^7 \cdot (1-p)^3$$

$$(a+b)^n = (a+b)(a+b)(a+b)\cdots(a+b)$$
$$= a^n + n\,a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \dots$$

$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$
$$\vdots$$

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \\ & & & & \vdots & & & & \end{array}$$

$$\Downarrow$$

$$(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^i b^{n-i} = b^n + n\,a\,b^{n-1} + \binom{n}{2}a^2 b^{n-2} + \dots$$

$$\left.\begin{array}{l} a = p \\ b = 1-p \end{array}\right\} \quad \left.\begin{array}{rl} (a+b)^n & = (p+(1-p))^n = 1^n = 1 \\ & = \displaystyle\sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i} \\ & = \displaystyle\sum_{i=0}^{n} f_X(i) \end{array}\right\} \sum_{i=0}^{n} f_X(i) = 1$$

Dette beviser at $X \sim \text{Binom}(n,p)$ gir en sannsynlighets-fordeling.

$\uparrow$

(dvs $f_X(i) = \binom{n}{i}p^i(1-p)^{n-i}$, $i = 0, 1, \dots, n$)

<u>Merk:</u> To måter å beskrive en sannsynlighetsfordeling

a) Formel: $f_X(i) = \binom{n}{i}p^i(1-p)^{n-i}$, $i = 0, 1, 2, \dots, n$

b) Beskrive hva X "egentlig er", dvs antall suksesser i en binomisk forsøksrekke

c) Geometrisk Geom(p) $(0 \leq p \leq 1)$ BI

Vi gjentar et Bernoulliforsøk helt til første suksess forekommer. (p = sannsynlighet for suksess i et enkelt forsøk)

X = antall gjentagelser
Forsøkene er uavhengige.

Eks: Vi kaster mynt/kron helt til vi får mynt.

X = antall kast $\quad X \sim \text{Geom}(p = 1/2)$

Mulige verdier for X: $1, 2, 3, 4, \ldots$

Sannsynlighetstetthet:

$$f_X(i) = P(X = i) = (1-p)^{i-1} \cdot p$$

Eks: $p(X=4) = (1-p)^3 \cdot p = (1/2)^4 = \underline{\underline{\frac{1}{16}}}$

Sannsynlighetsfordeling siden vi har:

$$\sum_{i=1}^{\infty} f_X(i) = p + (1-p)\cdot p + (1-p)^2 \cdot p + \ldots$$

$$= p \cdot \frac{1}{1-(1-p)} = \frac{p}{p} = \underline{\underline{1}}$$

(og $f_X(i) \geq 0$)

d) Poisson: Poisson (λ) ($\lambda > 0$)

BI

Mulige verdier for X: 0, 1, 2, ...
Sannsynlighetstetthet:

$$f_X(i) = p(X=i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

Er dette en sannsynlighetsfordeling?

i) $f_X(i) = \frac{\lambda^i}{i!} e^{-\lambda} \geq 0$ ok.

ii)
$$\sum_{i=0}^{\infty} f_X(i) = \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} + \ldots$$
$$= e^{-\lambda} \cdot \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right)$$
$$= e^{-\lambda} \cdot e^{\lambda} = 1 \quad \text{ok}$$

Husk: Potensrekken til eksponentialfunksjonen er

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots + \frac{x^n}{n!} + \ldots$$

Bruk av Poissonfordeling?

X = antall forekomster av noe i et gitt tidsintervall

λ = forventet antall forekomster per tidsintervall.

Eks:
X = antall skipsankomster i en havn per time

λ = forventet (gj.snittlig) antall ankomster per time

Hvis $X \sim \text{Binom}(n, p)$ og $\begin{cases} n \text{ er stor} \\ p \text{ er liten} \end{cases}$

da er $X \simeq Y$, der $Y \sim \text{Poisson}(\lambda = n \cdot p)$

$$\left.\begin{array}{l} X \sim \text{Binom}(n, p) \\ Y \sim \text{Poisson}(\lambda = np) \end{array}\right\} \rightarrow p(X=i) \simeq p(Y=i)$$

hvis n stor, p liten

Teoretisk begrunnelse:

Hvis: X = antall forekomster av en hendelse E per tidsintervall T:

Eks:
E = skip ankommer
T = 1 time

Vi tenker slik:

Vi deler tidsintervallet T opp i mange små tidsintervall av lengde T/n, der n er stor. Anta at følgende forutsetninger er tilstede:

i) Sannsynligheten p for hendelsen E i hvert tidsintervall er konstant (og liten)

ii) At hendelsen E forekommer i et tidsintervall er uavhengig av om den forekommer i et annet intervall

Da har vi en binomisk forsøksrekke, og $X \sim \text{Binom}(n, p)$. Forventet antall forekomster i hele T er $n \cdot p$. Dermed er X tilnærmet Poisson med $\lambda = np$ siden vi har

$$\boxed{p(X=i) = \binom{n}{i} p^i (1-p)^{n-i} \simeq e^{-\lambda} \cdot \frac{\lambda^i}{i!} = e^{-np} \cdot \frac{(np)^i}{i!}}$$

Eks: $n = 100$, $p = 0.01$

$\Downarrow$

$\lambda = np = 1$

$X \sim \text{Binom}(n, p)$ $\quad p(X=2) = \binom{100}{2} \cdot 0.01^2 \cdot 0.99^{98} \simeq \underline{0.18486}$

$Y \sim \text{Poisson}(\lambda)$ $\quad p(Y=2) = e^{-1} \cdot \frac{1^2}{2!} = \frac{e^{-1}}{2} \simeq \underline{0.18394}$

Tilnærmingen er god når $\begin{cases} n = 100 & (n \text{ stor}) \\ p = 0.01 & (p \text{ liten}) \end{cases}$

Eks:

Antall soldater som døde pga spark fra en hest

(Se datasett neste side).

k	antall
0	109
1	65
2	22
3	3
4	1
	200

antall som døde totalt:

$65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1 = 122$

forventet antall døde per korpsår:

$$\lambda = \frac{122}{200} = \underline{0.61}$$

X = antall døde pga spark fra hest i et tilfeldig korpsår

$X \sim \text{Poisson}(0.61)$ gir god tilpasning til dataene!

k	faktisk antall	$P(X=k)$	$P(X=k) \cdot 200$
0	109	0.543	108.7
1	65	0.331	66.3
2	22	0.101	20.2
3	3	0.021	4.1
4	1	0.003	0.6

forventet antall døde om $X \sim \text{Poisson}(0.61)$

bruker $P(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ med $\lambda = 0.61$

[Previous] [Next] [Up] [Top] **Categorical Data Analysis with Graphics**
Michael Friendly

Part 1: Plots for discrete distributions

Contents

Discrete frequency distributions often involve counts of occurrences such as accidental fatalities, words in passages of text, or blood cells with some characteristic. Typically such data consist of a table which records that *n sub k* of the observations pertain to the basic outcome value k, $k = 0, 1, \ldots$.

The table below shows two such data sets:

- von Bortkiewicz's (1898) data on death of soldiers in the Prussian army from kicks by horses and mules. The data pertain to 10 army corps, each observed over 20 years. In 109 corps-years, no deaths occurred; 65 corps-years had one death, etc. (Figure 1)
- Mosteller & Wallace's (1964) data on the occurrence of the word *may* in 262 blocks of text (each about 200 words long) from issues of the *Federalist Papers* known to be written by James Madison. In 156 blocks, the word *may* did not occur; it occurred once in 63 blocks, etc. (Figure 2)

```
Deaths by Horsekick              Occurrences of 'may'

   k     nk                         k     nk
  ----------                       ----------
   0     109                        0     156
   1      65                        1      63
   2      22                        2      29
   3       3                        3       8
   4       1                        4       4
        -----                       5       1
        N=200                       6       1
                                         -----
                                         N=256
```