

Kursus 02403

Introduktion til Statistik

Forelæsning 2: Kapitel 4, Diskrete fordelinger

Jan Kloppenborg Møller

DTU Informatik

Bygning 305 - rum 210

Danmarks Tekniske Universitet

2800 Lyngby – Danmark

e-mail: jkm@imm.dtu.dk

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

Stokastiske Variable

- Hvad er en stokastisk variabel?
- Stokastiske variable angives ved store bogstaver, f.eks. X, Y, Z
- Udfald fra stokastiske variable angives ved tilsvarende små bogstaver, f.eks. x, y, z
- Vi skelner mellem diskrete og kontinuerte stokastiske variable

Fra kapitel 3:

Det klassiske sandsynlighedskoncept er defineret:

Hvis der findes n lige sandsynlige udfald, hvorfra et må ske, og hændelsen s betegnes som 'succes', så er sandsynligheden for succes givet ved:

$$\frac{s}{n}$$

s =antal *gunstige* udfald n =antal *mulige* udfald

Tæthedsfunktion

- Tæthedsfunktionen for en stokastisk variabel betegnes ved $f(x)$
- Ofte bruger man også betegnelsen frekvensfunktion eller hyppighed om tæthedsfunktionen
- $f(x)$ siger noget om hyppigheden af udfaldet x for den stokastiske variabel X
- Et godt plot af $f(x)$ er et bar chart (diskret) eller et histogram (kontinuert)

Tæthedsfunktion for en diskret variabel

For en diskret variabel kan vi skrive tæthedsfunktionen som:

$$f(x) = P(X = x)$$

Der gælder:

$$f(x) > 0 \quad \text{for } x \in S$$

$$f(x) = 0 \quad \text{for } x \notin S$$

$$\sum f(x) = 1$$

Fordelingsfunktion

- Fordelingsfunktion for en stokastisk variabel betegnes ved $F(x)$.

Fordelingsfunktionen svarer til den kumulerede tæthedsfunktion:

$$F(x) = P(X \leq x)$$

$$F(x) = \sum_{t=-\infty}^x f(t)$$

- Et godt plot for fordelingsfunktionen er den kumulative fordeling

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

Konkrete Statistiske fordelinger

- Der findes en række statistiske fordelinger, som kan bruges til at beskrive og analysere forskellige problemstillinger med
- Vi betragter først diskrete fordelinger
 - Binomial fordelingen
 - Den Hypergeometriske fordeling
 - Poisson fordelingen

Binomial fordelingen

- Vi betragter n uafhængige forsøg, hvor n er konstant
- I hvert enkelt forsøg kan udfaldet/hændelsen blive enten succes eller fiasko (1,0)
- Sandsynligheden for succes er p (og er den samme for alle n forsøg)
- Vi ønsker at beskrive antallet af succeser

Binomial fordelingen

Vi siger at en stokastisk variabel, X , er binomial fordelt

$$X \sim b(x; n, p)$$

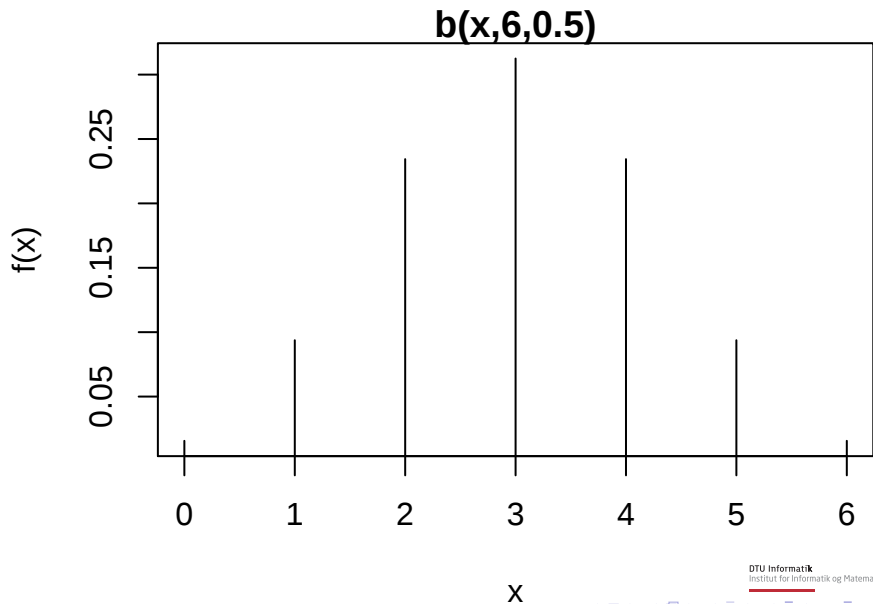
Tæthedsfunktion for binomial fordelingen:

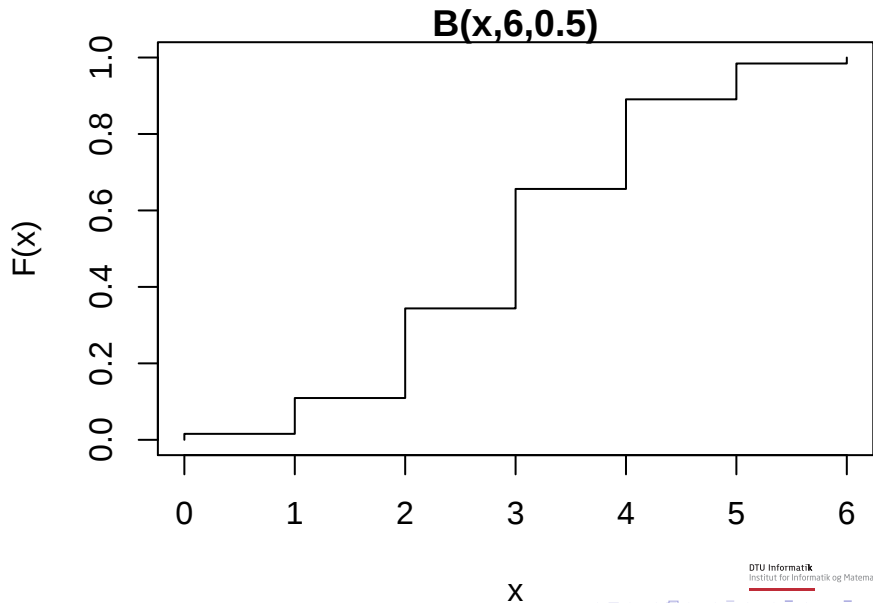
$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

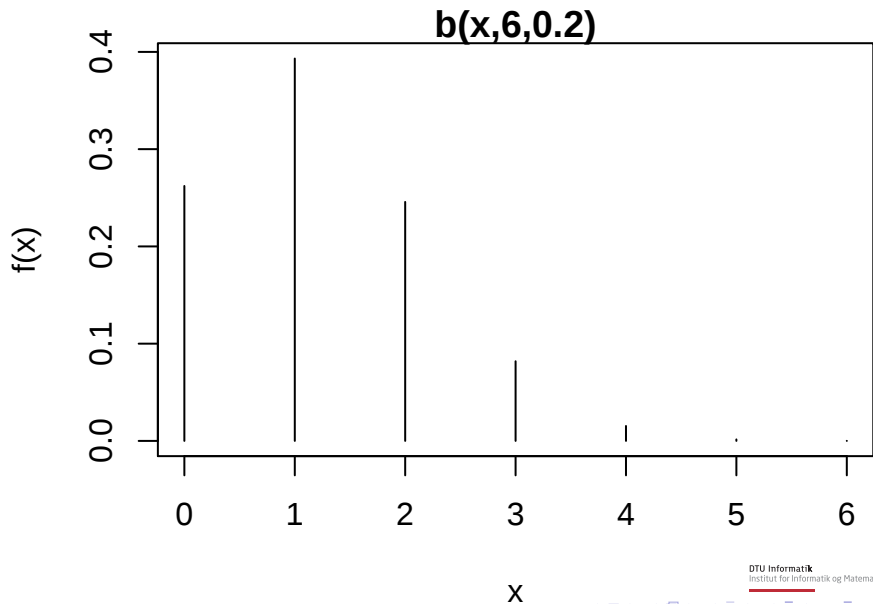
Fordelingsfunktion for binomial fordelingen:

$$F(x) = P(X \leq x) \quad \text{tabel 1, side 505 (7ed: 576) (6ed: 565)} \rightarrow$$

Fordelingsfunktionen $F(x) = P(X \leq x)$ skrives $B(x; n, p)$







Eksempel 1

I et kundecenter i et telefonselskab søger man at forbedre kundetilfredsheden. Især er det vigtigt, at når der indrapporteres en fejl, bliver fejlen udbedret i løbet af samme dag.

Antag at sandsynligheden for at en fejl bliver udbedret i løbet af samme dag er $p = 0.7$.

Eksempel 1

I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

Eksempel 1

Hvad er sandsynligheden for at højst 1 fejl ikke bliver udbedret?

Eksempel 2

I en forsendelse af 15 hard disks har 5 mindre skrammer.

Hvis der udtages en tilfældig stikprøve på 3 hard disks, hvad er sandsynligheden for at 1 af dem har skrammer?

Binomial fordelingen

- Binomial fordelingen anvendes også for at analysere stikprøver med tilbagelægning
- Når man vil analysere stikprøver uden tilbagelægning anvendes den hypergeometriske fordeling

Den hypergeometriske fordeling

- En population af størrelse N
- Vi tager en stikprøve af størrelse n
- Der er a defekte i populationen
- Der er $N - a$ ikke-defekte i populationen

Den hypergeometriske fordeling

Vi siger at en stokastisk variabel, X , er hypergeometrisk fordelt $X \sim h(x; n, a, N)$

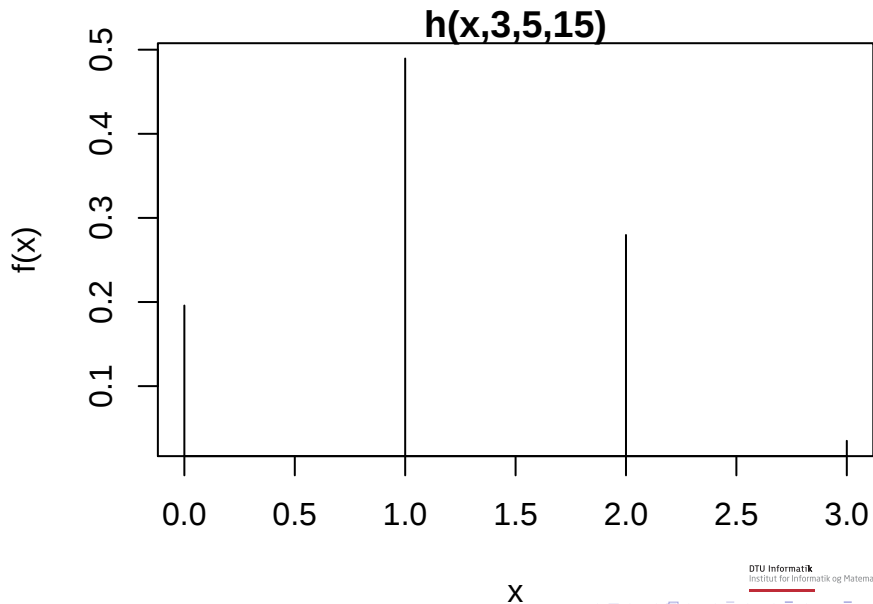
Tæthedsfunktion for den hypergeometriske fordeling

$$f(x) = P(X = x) = \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}}$$

Eksempel 2

I en forsendelse af 15 hard disks har 5 mindre skrammer.

Hvis der udtages en tilfældig stikprøve på 3 hard disks, hvad er sandsynligheden for at 1 af dem har skrammer?



Den hypergeometriske fordeling

Den hypergeometriske fordeling kan approksimeres ved binomial fordelingen hvis populationen N er stor og stikprøven n er lille

Eksempel 3

Det antages, at der i gennemsnit bliver indlagt 0.3 patienter pr. dag på Københavnske hospitaler som følge af luftforurening.

Hvad er sandsynligheden for at der på en vilkårlig dag bliver indlagt højst 2 patienter som følge af luftforurening?

Poissonfordelingen

- Poisson fordelingen anvendes ofte som en fordeling (model) for tælleletal, hvor der ikke er nogen naturlig øvre grænse
- Poisson fordelingen kan ofte karakteriseres som intensitet, dvs på formen antal/enhed
- Parameteren λ angiver intensiteten i poisson fordelingen

Poissonfordelingen

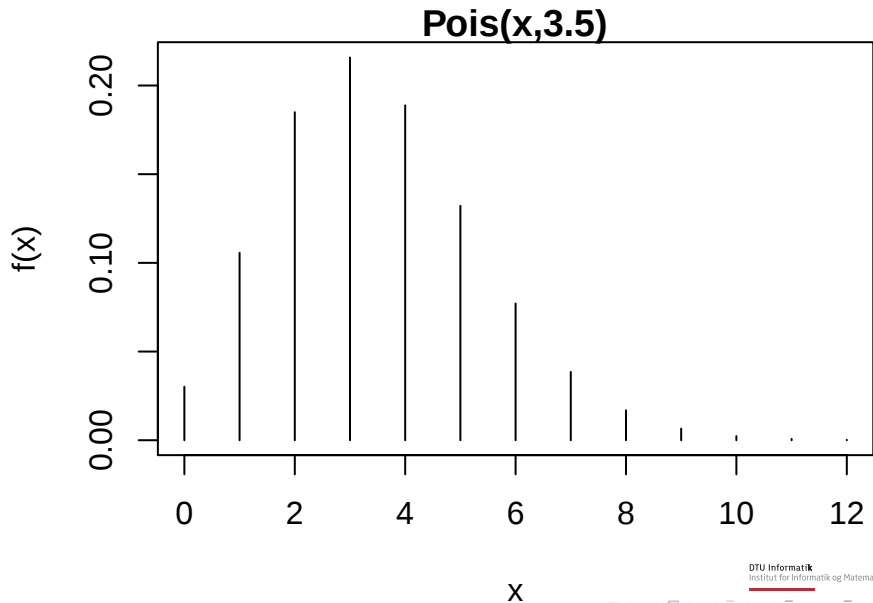
$$X \sim P(\lambda)$$

tæthedsfunktion:

$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Fordelingsfunktion:

$$F(x) = P(X \leq x) \quad \text{tabel 2, side 510 (7ed: 581) (6ed: 570)}$$



Eksempel 3

Det antages, at der i gennemsnit bliver indlagt 0.3 patienter pr. dag på Københavnske hospitaler som følge af luftforurening.

Hvad er sandsynligheden for at der på en vilkårlig dag bliver indlagt højst 2 patienter som følge af luftforurening?

Eksempel 3

Hvad er sandsynligheden for at der på en vilkårlig dag bliver indlagt præcis 2 patienter?

Eksempel 3

Hvad er sandsynligheden for at der på en vilkårlig dag bliver indlagt mindst 1 patient?

Eksempel 3

Hvad er sandsynligheden for at der i en periode på 3 dage bliver indlagt højst 1 patient?

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

Middelværdi af en diskret stokastisk variabel

Middelværdien af en diskret stokastisk variabel beregnes ved:

$$\mu = \sum_S x \cdot f(x)$$

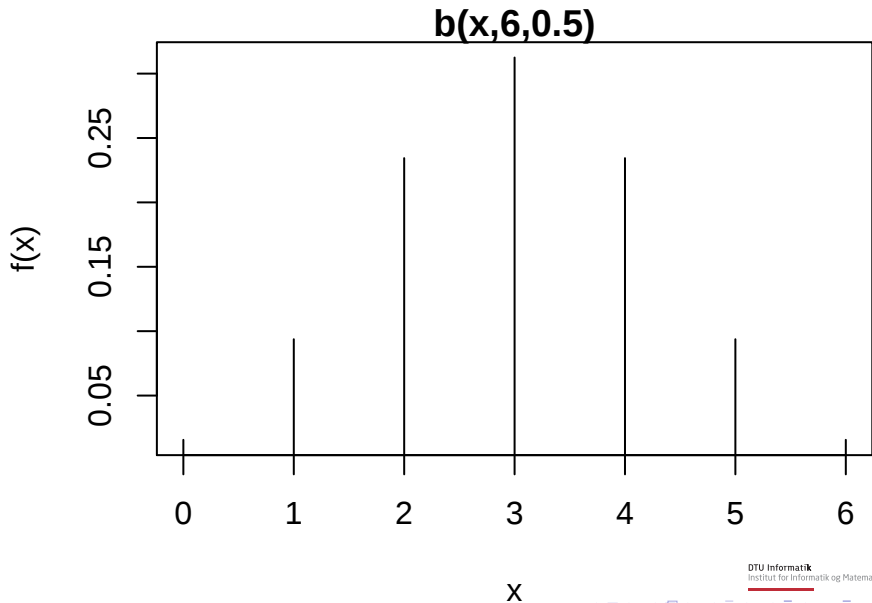
hvor S er udfaldsrummet for X

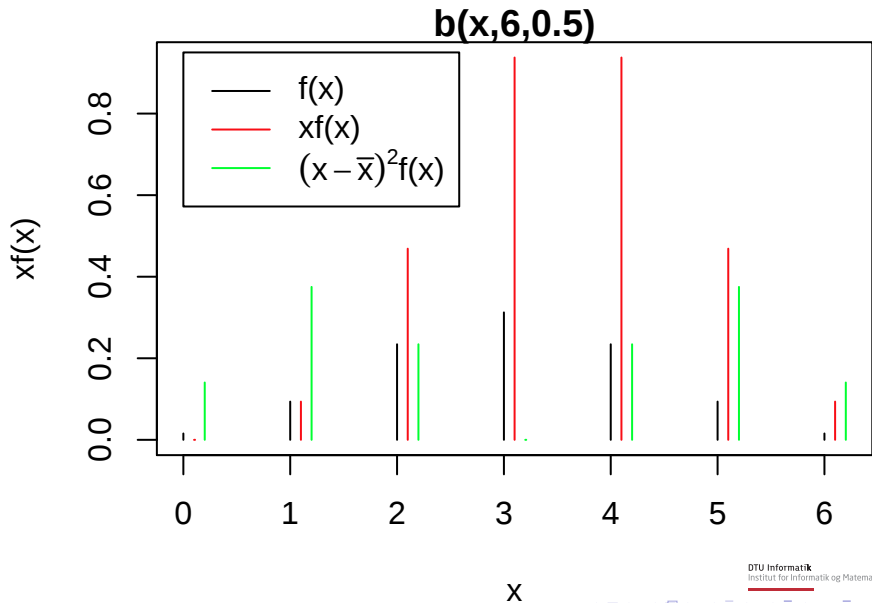
Varians af en diskret stokastisk variabel

Variansen af en diskret stokastisk variabel beregnes ved:

$$\sigma^2 = \sum_S (x - \mu)^2 \cdot f(x)$$

hvor S er udfaldsrummet for X





Middelværdi og varians for kendte diskrete fordelinger

- Binomial fordelingen:

- Middelværdi:

$$\mu = n \cdot p$$

- Varians:

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

Middelværdi og varians for kendte diskrete fordelinger

- Den hypergeometriske fordeling:
- Middelværdi:

$$\mu = n \cdot \frac{a}{N}$$

- Varians:

$$\sigma^2 = \frac{na \cdot (N-a) \cdot (N-n)}{N^2 \cdot (N-1)} \cdot a$$

Middelværdi og varians for kendte fordelinger

- Poisson fordelingen

- Middelværdi:

$$\mu = \lambda$$

- Varians:

$$\sigma^2 = \lambda$$

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

R (R Note afsnit 3)

R	Betegnelse
binom	Binomialfordelingen
pois	Poissonfordelingen

- d Tæthedsfunktion $f(x)$ (probability density function).
- p Fordelingsfunktion $F(x)$ (cumulative distribution function).
- r Tilfældige tal fra den anførte fordeling. (Forelæsning 10)
- q Fraktil (quantile) i fordeling.

Eksempel:

$B(5; 10, 0.6)$

`pbinom(5,10,0.6)`

Oversigt

- 1 Stokastiske variable og fordelinger
 - Definition af en stokastisk variabel
 - Tæthedsfunktion
 - Fordelingsfunktion
- 2 Konkrete Statistiske fordelinger
 - Binomial fordelingen
 - Eksempel 1
 - Den hypergeometriske fordeling
 - Eksempel 2
 - Poissonfordelingen
 - Eksempel 3
- 3 Middelværdi og varians
 - Middelværdi og varians for kendte diskrete fordelinger
- 4 R (R Note afsnit 3)

I morgen:

- Kontinuerte fordelinger - kapitel 5