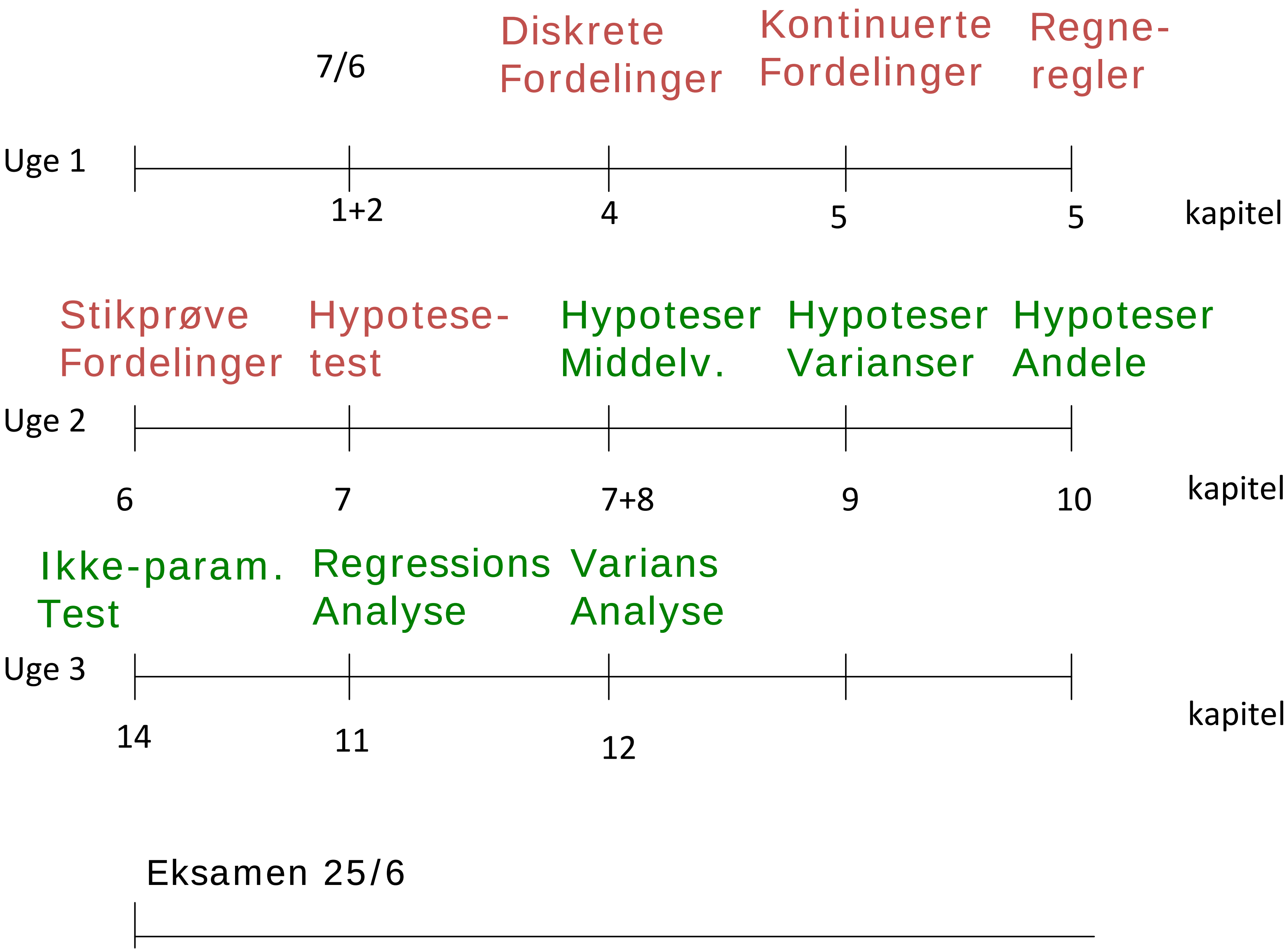
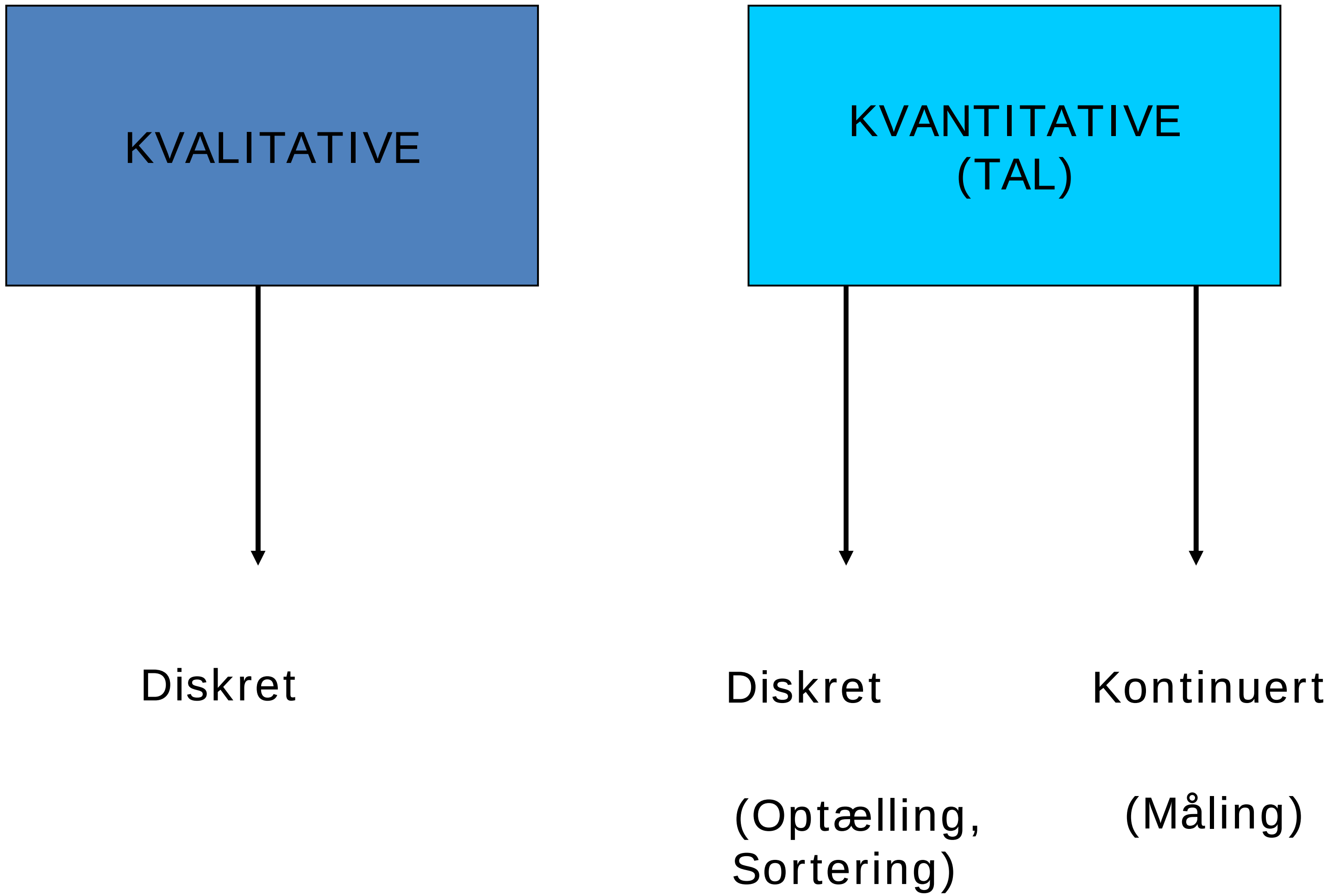
02403 Introduktion til Statistik

Klaus Kaae Andersen

8. juni 2010


7/6
Diskrete Fordelinger
Kontinuerte Fordelinger
Regne-regler
Uge 1
1+2
4
5
5
kapitel
Stikprøve Fordelinger
Hypotese-test
Hypoteser Middelv.
Hypoteser Varianser
Hypoteser Andele
Uge 2
6
7
7+8
9
10
kapitel
Ikke-param. Test
Regressions Analyse
Varians Analyse
Uge 3
14
11
12
kapitel
Eksamen 25/6

Forskellige typer data



Kapitel 4: Diskrete fordelinger

- Definition af en stokastisk variabel
- Tæthedsfunktion for diskrete variable
- Fordelingsfunktion for diskrete variable

Diskrete generiske fordelinger:

- Binomial fordelingen
- Poisson fordelingen

- Middelværdi af diskrete stokastiske variable
- Varians af diskrete stokastiske variable

Stokastiske Variable

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

Det klassiske sandsynlighedskoncept

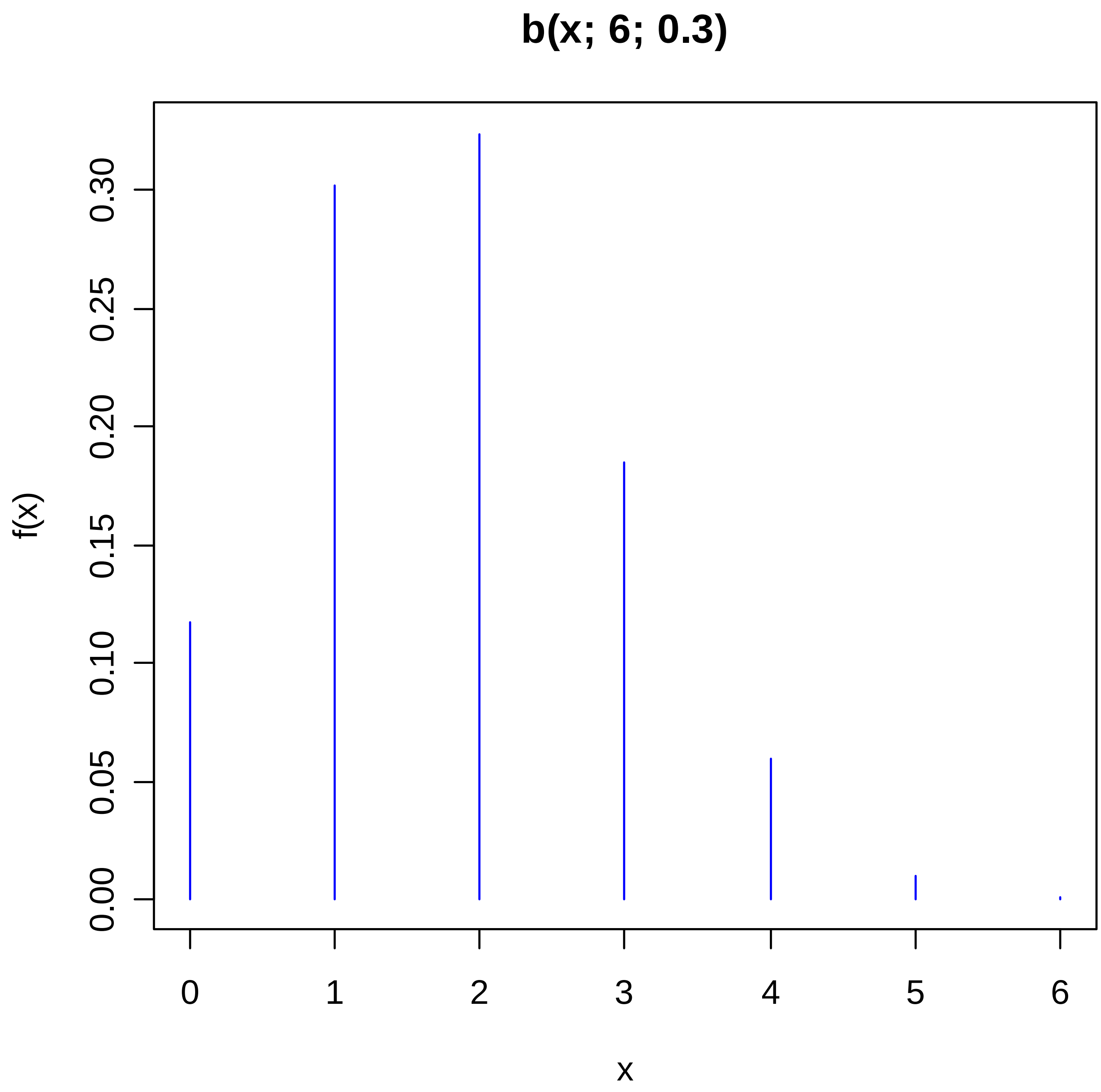
Det klassiske sandsynlighedskoncept er defineret:

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

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

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)


b(x; 6; 0.3)
f(x)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
6
x

Tæthedsfunktion for en diskret variabel

For en diskret variabel kan vi skrive tæthedsfunktionen som:

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

Der gælder:

$$f(x) > 0 \quad \text{for } x \in S$$
$$f(x) = 0 \quad \text{for } x \ni S$$
$$\sum f(x) = 1$$

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

Generiske Fordelinger

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

Binomial fordelingen

- Vi betrager n uafhængige forsøg, hvor n er konstant
- I hvert enkelt forsøg kan udfaldet/hændelsen blive enten succes eller fiasko
- Sandsynligheden for succes er p (og er den samme for alle n forsøg)
- Er dette opfyldt, siger vi at den stokastiske variabel, X, er binomial fordelt
  $X \sim b(x; n, p)$

Binomial fordelingen

- Vi betragter en stokastisk variabel, X, der 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:
  $F(x) = P(X \leq x)$ tabel eller lommeregner
- Fordelingsfunktionen $F(x) = P(X \leq x)$ skrives $B(x; n, p)$

Binomial Fordelingen

- Antag, at sandsynligheden for at være bærer af en sygdom er p = 0.30

1. Angiv sandsynligheden for, at blandt n=10 tilfældigt udvalgte personer højst er 2 personer, der bærer sygdommen 0.38
2. Angiv sandsynligheden for, at mindst en person bærer sygdommen 1-0.028 = 0.971

Binomial Fordelingen

- Antag, at sandsynligheden for at få karakteren 12 til statistikeksamen er $p = 0.10$

1. Angiv sandsynligheden for, at blandt $n=20$ tilfældigt udvalgte personer mindst er 3 personer, der får karakteren 12 0.32

Binomial Fordelingen

- Du er til et fødselsdagsselskab med 50 andre mennesker.

1. Angiv sandsynligheden for, at blandt de n=50 er mindst en person, der har samme fødselsdag som du

0.128

Poisson fordelingen

- Poisson fordelingen anvendes ofte som en fordeling (model) for tælletal, 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

Poisson fordelingen

- $X \sim P(\lambda)$
- tæthedsfunktion:
  $f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$
- Fordelingsfunktion:
  $F(x) = P(X \leq x)$ tabel eller lommeregner

Poisson Fordelingen

- Incidens af en sygdom er 2.1 tilfælde / dag

1. Angiv sandsynligheden for, at der ikke er nye sygdomstilfælde på en vilkårlig udvalgt dag

0.122

2. Angiv sandsynligheden for, at der højst er 5 nye sygdomstilfælde i løbet af en uge

0.0034

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

Binomial fordelingen

- Middelværdi:
  $\mu = n \cdot p$
- Varians:
  $\sigma^2 = n \cdot p \cdot (1 - p)$

Poisson fordelingen

- Middelværdi:
  $\mu = \lambda$
- Varians:
  $\sigma^2 = \lambda$

Kapitel 4: Diskrete fordelinger

- Definition af en stokastisk variabel
- Tæthedsfunktion for diskrete variable
- Fordelingsfunktion for diskrete variable

Diskrete generiske fordelinger:

- Binomial fordelingen
- Poisson fordelingen

- Middelværdi af diskrete stokastiske variable
- Varians af diskrete stokastiske variable

- Næste gang: kontinuerte fordelinger