FORELESNING 8

MET 3431

STATISTIKK

EIVIND ERIKSEN

FEB 14 2012

1. Sannsynligheter og sannsynlighetsregning [T] Kap. 4.1-4.2
2. Stokastiske variable, binomisk fordeling [T]. Kap. 5.1-5.3

Regelen for uvanlige hendelser

Rare event rule

Anta at:

- Vi har en bestemt antagelse om noe
- Vi observerer noe i dataene som er svært uvanlig hvis antagelsen er sann
- Konklusjon: Da er det noe galt med antagelsen!

Analogi

- Vi antar at tiltalte er uskyldig
- Det er svært usannsynlig å ha disse bevisene dersom tiltalte er uskyldig:
- Da er antagelsen om uskyld gal!!

Regelen for uvanlige hendelser

Example

- IT-avdelingen anslår at rundt 15% av BI studentene har en IPhone
- Men i en stikkprøve av 1939 BI studenter har 22% Iphone
- Hvis bare 15% av populasjonen har Iphone, så skal det mye til for at andelen i en stikkprøve med 1939 studenter er 22% ...
- Konklusjon: antagelsen om 15% er gal!

Regelen for uvanlige hendelser og statistisk inferens

Rare event rule

- Vi har en bestemt antagelse om populasjonen. Den kalles en *hypotese*.
- Dataene i stikkprøven er svært uvanlige i forhold til hypotesen:
- Konklusjon: Da er det noe galt med hypotesen!

Statistisk inferens

Med sannsynlighetsregning begynner vi med andre del av kurset: Statistisk inferens. Det er kunsten å bruke stikkprøven til å si noe generelt om hele populasjonen.

Sannsynligheten til en hendelse

Sannsynlighet

Det er fundamentalt viktig å kunne regne ut sannsynligheten til en hendelse.

Å tolke sannsynligheten

Men det er like viktig å kunne tolke sannsynlighetsverdier.

Definisjoner

= Begivenhet

<u>Tilfeldig prosedyre / forsøk</u>: Mulige utfall

Utfallet er ikke forutsigbart.

Hendelse (eng: event)

En samling av utfall fra en tilfeldig prosedyre.

Eksempel: Du triller to terninger. 'Begge terninger viser det samme' er en hendelse.

= utfall

Enkelhendelse (eng: simple event)

En hendelse som består av ett spesielt utfall.

Eksempel: Du triller to terninger. 'Begge terninger viser 1' er en enkelhendelse

= Samlingen av alle mulige utfall i et tilfeldig forsøk

Utfallsrom (eng: sample space)

Alle enkelhendelser som prosedyren kan resulterere i.

Notasjon for sannsynlighetsregning

Notasjon

- P angir sannsynlighet (eng: probability)
- A, B, og C angir spesifikke hendelser
- $P(A)$ er sannsynligheten for at hendelse A inntreffer

probability   hendelsen A

Eks: Vi kaster to terninger

A = begge terninger viser det samme

$p(A) = ?$ $p(A) = \frac{6}{36} = \frac{1}{6} \simeq 0.167 = 16.7\%$

Beregne sannsynlighet

Alle utfall er like sannsynlige

Da beregner man sannsynligheten som

$$P(A) = \frac{\text{Antall måter A kan inntreffe på}}{\text{Totalt antall enkelthendelser}} = \frac{\text{antall gunstige utfall}}{\text{antall mulige utfall}}$$

Example

Du triller to terninger. La A være hendelsen: Summen av terningene er 3".

$$P(A) = \frac{2}{36} = \frac{1}{18} \simeq 0.056 = 5.6\%$$

Beregne sannsynlighet

Sannsynligheten er et tall mellom 0 og 1

- Sannsynligheten til en umulig hendelse er 0 — $P(A)=0$ betyr A skjer <u>ikke</u>
- Sannsynligheten til en hendelse som helt sikkert inntreffer, er 1 — $P(A)=1$ betyr A <u>skjer</u>
- Enhver hendelse har en sannsynlighet mellom 0 og 1.

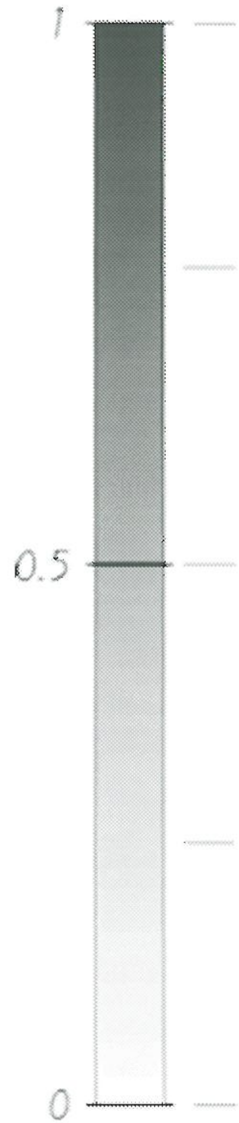
Example

Du triller to terninger.

- La A være hendelsen: 'Summen av terningene er 30'. $P(A) = 0$
- La A være hendelsen: 'Summen av terningene er mer enn 1'. $P(A) = 1$
- La A være hendelsen: 'Summen av terningene er mer enn 2'. $P(A) = \frac{35}{36}$
- Det er *meget sannsynlig* at hendelsen 'Summen av terningene er mer enn 2' inntreffer. Det skjer 35 av 36 ganger...

Tolke sannsynligheter

Summen av terningene høyst 12



Sikkert

Sannsynlig

50% 50

Lite sannsynlig

Umulig

Summen av terningene mer enn 12

Komplementet til en hendelse

$\boxed{p(A) + p(\bar{A}) = 1}$

Komplement

- Noen ganger er det lettere å beregne sannsynligheten til *komplementet*
- For en hendelse A er komplementet alle utfall som *ikke* er i A.
- Det betegnes $\overline{A}$
- Da har vi $P(A) = 1 - P(\overline{A})$

summen er to $\rightarrow$ $\bar{A}$ | A = summen er minst 3

alle utfall = utfallsrommet

$p(\bar{A}) = 1 - p(A)$

$p(A) = 1 - p(\bar{A}) = 1 - \frac{1}{36} = \underline{\frac{35}{36}}$

Example

Hva er sannsynligheten for at B: 'summen av terninger ikke er 3'?

- Komplementet er $\overline{B}$: 'summen av terningene er 3'
- På side 9 så vi at : $P(\overline{B}) = \frac{1}{18}$
- $P(\cancel{A}\,B) = 1 - P(\overline{B}) = 1 - \frac{1}{18} = \frac{17}{18} \simeq= 0.944$
- Det er 94.4% sjanse for at summen av to terninger ikke blir 3

Kapittel 5: Sannsynlighetsfordelinger

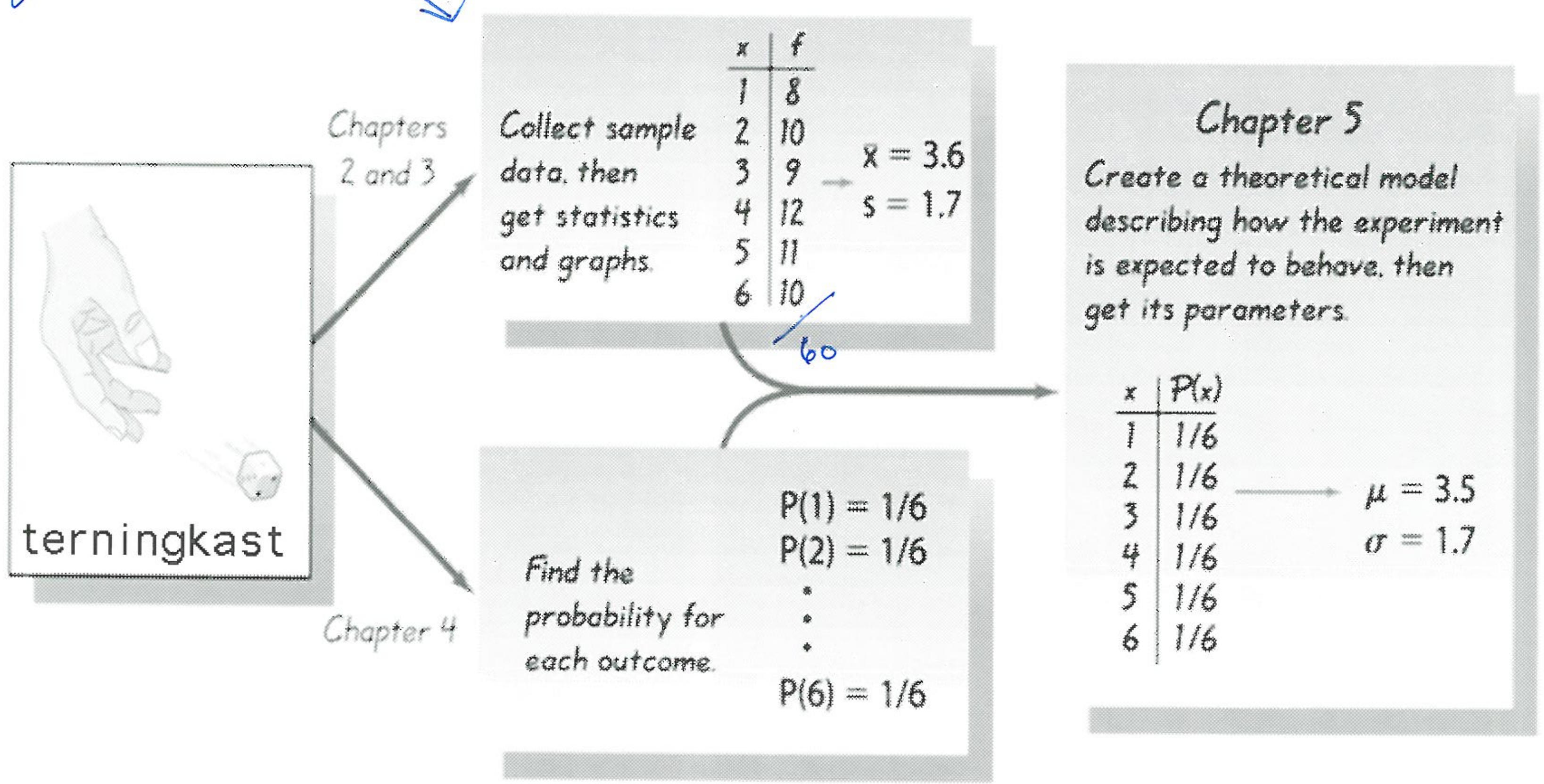
Om kapittel 5

- Vi kombinerer deskriptiv statistikk fra kapittel 3 med sannsynlighet fra kapittel 4
- Da får vi *diskrete sannsynlighetsfordelinger* som er tema i kapittel 5
- Sannsynlighetsfordelinger beskriver hva som sannsynligvis vil skje, ikke hva som faktisk skjedde.

$$\bar{x} = \frac{\overbrace{1+1+\dots+1}^{8} + \overbrace{2+2+\dots+2}^{10} + \dots}{60}$$

$x_1 \quad x_2 \quad \dots$

frekvenstabell sannsynlighetsfordeling



sannsynligheter

Eks: Vi kaster to terninger

$X =$ summen av terningene

Utfall		$X =$
1	1	$X = 2$
1	2	⋮
1	3	
1	4	$X = 5$
1	5	⋮
1	6	
2	1	
2	2	
2	3	$X = 5$
⋮	⋮	⋮

$$P(X=5) = \frac{4}{36} = \frac{1}{9} \approx 0.111$$

Section 5-2: Tilfeldige variable

Stokastisk variabel

//

Tilfeldig variabel (eng: random variable)

En tilfeldig variabel er et tall som er et utfall av en tilfeldig prosess.

utfall	X=
1	1
2	2
3	3
4	4
5	5
6	6

Examples

- Antall 'øyne' ved et terningkast
- Lojalitet hos en tilfeldig valgt bankkunde, målt på en skala fra 1 til 10

Fordelingen til en tilfeldig variabel

Sannsynlighetsfordelingen til en diskret tilfeldig variabel angir sannsynligheten for hver tenkelige verdi av den tilfeldige variabelen. Dette gjøres vha en tabell eller en formel.

Sannsynlighetsfordeling = teoretisk/matematisk modell

Den gir sannsynligheter ($P(X=5) = \dots$)

To typer tilfeldige variable

Diskret tilfeldig variabel

Dersom variabelen har et endelig eller tellbart antall utfall.

Example

Antall 'øyne' ved et terningkast

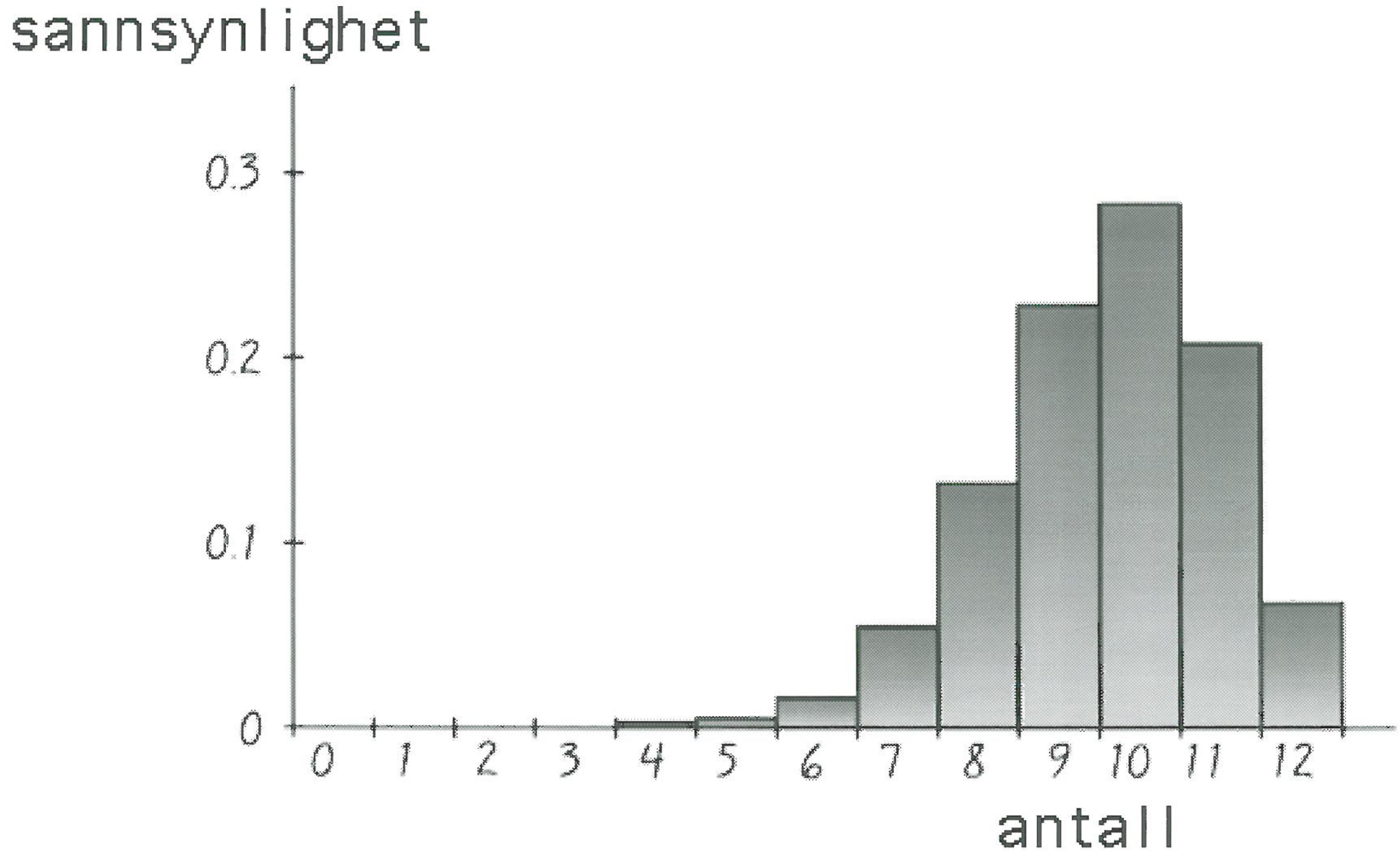
Kontinuerlig tilfeldig variabel

Dersom variabelen har uendelig mange utfall målt på en kontinuerlig skala uten gap.

Example

Vekten på en tilfeldig valgt pasient

Sannsynlighetshistogram



Figur: Antall meksikansk-amerikanske medlemmer i jury

Krav til en sannsynlighetsfordeling

Diskret sannsynlighetsfordeling:

X tilfeldig variabel, diskret dvs mulige verdier for X er typisk 1, 2, 3, ...

$p(1) = p(X=1)$, $p(2) = p(X=2)$, $p_3 = p(X=3)$, ...

To krav må oppfylles

1. $\sum P(x) = 1 \Leftrightarrow p(x_1) + p(x_2) + \ldots = 1$
2. $0 \leq P(x) \leq 1$

X = antall øyne

$p(1) = p(2) = \dots = p(6) = \frac{1}{6}$

Regn ut

$$\mu = E(X) = x_1 \cdot p(x_1) + \dots + x_6 \cdot p(x_6)$$

$$= 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + 4 \cdot p(4) + 5 \cdot p(5) + 6 \cdot p(6)$$

$$= \underline{3.5}$$

Forventning og standardavvik til en tilfeldig variabel

Forventning: $\mu = E(X) = \overline{X}$

Forventning / Gjennomsnitt

Populasjonsgjennomsnittet μ til en tilfeldig variabel:

$$\mu = \sum x \cdot P(x) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \ldots$$

Standardavvik til en tilfeldig variabel

Standardavviket σ i populasjonen:

$$\sigma = \sqrt{\sum (x - \mu)^2 \cdot P(x)}$$