Course 02402/02323 Introducerende Statistik

Forelæsning 2: Stokastisk variabel og diskrete fordelinger

Klaus K. Andersen og Per Bruun Brockhoff

DTU Compute, Statistik og Dataanalyse
Danmarks Tekniske Universitet
2800 Lyngby – Danmark
e-mail: klaus@cancer.dk

Oversigt

Oversigt

Stokastisk variabel

En stokastisk variabel (random variable) repræsenterer udfaldet af et eksperiment der endnu ikke er udført

DTU Compute
Institut for Matematik og Computer Science

Stokastisk variabel

En stokastisk variabel (random variable) repræsenterer udfaldet af et eksperiment der endnu ikke er udført

- Et terningekast
- Antallet af seksere i 10 terningekast
- km/l for en bil
- Måling af sukkerniveau i blodprøve
- ...

Diskret eller kontinuert

- Vi skelner mellem diskret og kontinuert
- Diskret kan tælles:
  - Hvor mange der bruger briller herinde
  - Antal mange flyvere letter den næste time
- Kontinuert:
  - Vindmåling
  - Tiden det tog at komme til DTU

Diskret eller kontinuert

- Vi skelner mellem diskret og kontinuert
- Diskret kan tælles:
  - Hvor mange der bruger briller herinde
  - Antal mange flyvere letter den næste time
- Kontinuert:
  - Vindmåling
  - Tiden det tog at komme til DTU

I dag er det diskret næste uge er det kontinuert.

Stokastisk variabel

Før eksperimentet er udført stokastisk variabel haves

$$X \text{ (eller } X_1, \ldots, X_n\text{)}$$

noteret med stort bogstav.

Stokastisk variabel

Før eksperimentet er udført stokastisk variabel haves

$$X \text{ (eller } X_1, \ldots, X_n\text{)}$$

noteret med stort bogstav.
Så udføres eksperimentet, og vi har da en *realisation* eller *observation*

$$x \text{ (eller } x_1, \ldots, x_n\text{)}$$

noteret med småt bogstav.

Simuler et terningekast

Vælg et tal fra $(1, 2, 3, 4, 5, 6)$ med lige sandsynlighed for hvert udfald

Diskrete fordelinger

- Selve konceptet er simpelthen at beskrive eksperimentet før det er udført
- Hvad kan vi gøre når vi endnu ikke kender udfaldet!?

Diskrete fordelinger

- Selve konceptet er simpelthen at beskrive eksperimentet før det er udført
- Hvad kan vi gøre når vi endnu ikke kender udfaldet!?
- Løsning: brug en tæthedsfunktion

DTU Compute
Institut for Matematik og Computer Science

Oversigt

Tæthedsfunktion

En stokastisk variabel har en *tæthedsfunktion* (probability density function (pdf))

Definition

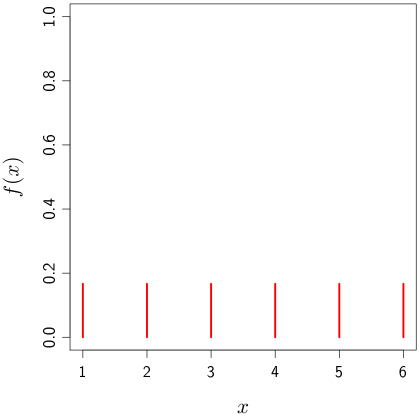
$$f(x) = P(X = x)$$

Sandsynligheden for at X bliver udfaldet x når eksperimentet udføres

DTU Compute
Institut for Matematik og Computer Science

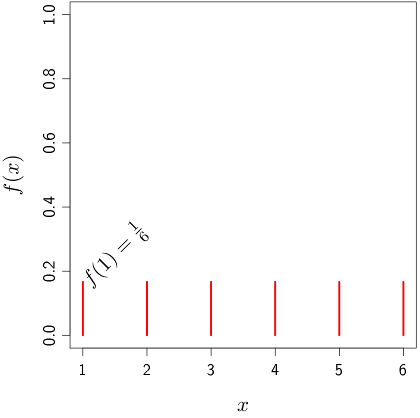
Tæthedsfunktion

En fair ternings tæthedsfunktion



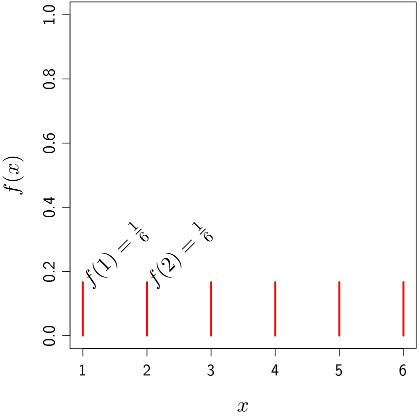
Tæthedsfunktion

En fair ternings tæthedsfunktion



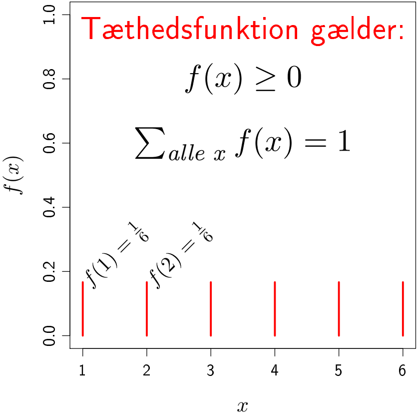
Tæthedsfunktion

En fair ternings tæthedsfunktion



Tæthedsfunktion

En fair ternings tæthedsfunktion



Stikprøve

Hvis vi kun har en observation kan vi da se fordelingen?

Stikprøve

Hvis vi kun har en observation kan vi da se fordelingen? Nej
men hvis vi har n observationer, så har vi en *stikprøve* (a sample)

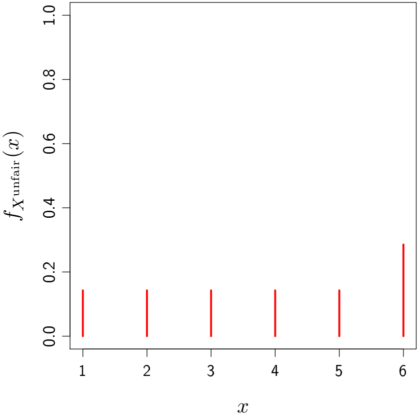
$$\{x_1, x_2, ..., x_n\}$$

og da kan vi begynde at "se" fordelingen.

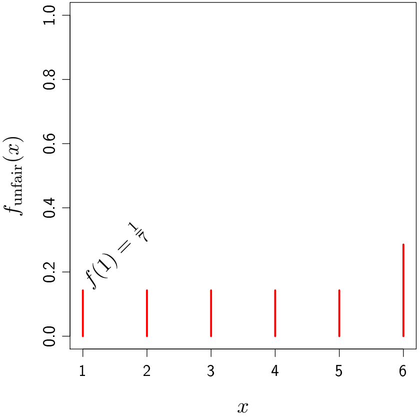
Simulate n rolls with a fair dice

```
## Antal simulerede realiseringer
n <- 30

## Træk uafhængigt fra mængden (1,2,3,4,5,6) med ens sandsynlighed
xFair <- sample(1:6, size=n, replace=TRUE)

## Tæl antallet af hvert udfald
table(xFair)

## Plot den empiriske tæthedsfunktion (pdf)
plot(table(xFair)/n, ylim=c(0,1), lwd=10, xlab="x", ylab="f(x)")

## Tilføj den rigtige tæthedsfunktion til plottet
lines(rep(1/6,6), type="h", lwd=3, col="red")

## legend
legend("topright", c("Empirical pdf","pdf"), lty=1, col=c(1,2), lwd=c(5,2))
```

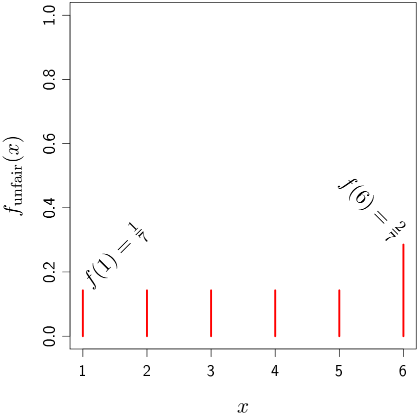
En unfair ternings tæthedsfunktion



En unfair ternings tæthedsfunktion



En unfair ternings tæthedsfunktion



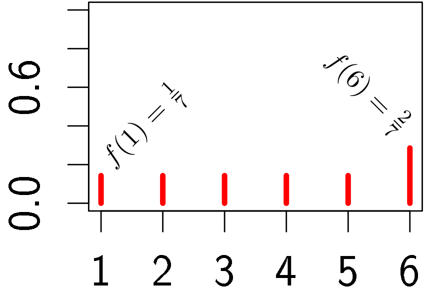
Simuler n kast med en ikke-fair terning

```
## Antal simulerede realiseringer
n <- 30

## Træk uafhændigt fra mængden (1,2,3,4,5,6) med højere
## sandsynlighed for en sekser
xUnfair <- sample(1:6, size=n, replace=TRUE, prob=c(rep(1/7,5),2/7))

## Plot den empiriske tæthedsfunktion
plot(table(xUnfair)/n, lwd=10, ylim=c(0,1), xlab="x", ylab="Density")

## Tilføj den rigtige tæthedsfunktion
lines(c(rep(1/7,5),2/7), lwd=4, type="h", col=2)

## En legend
legend("topright", c("Empirical pdf","pdf"), lty=1, col=c(1,2), lwd=c(5,2))
```

Nogle spørgsmål

Find nogle sandsynligheder for X^{unFair}:

- Sandsynligheden for at få en firer?
- Sandsynligheden for at få en femmer eller en sekser?
- Sandsynligheden for at få mindre end tre?



Oversigt

Fordelingsfunktion (distribution function eller cumulative density function (cdf))

Definition

Fordelingsfunktionen (cdf) er tæthedsfunktionen akkumuleret

$$F(x) = P(X \leq x) = \sum_{j \text{ hvor } x_j \leq x} f(x_j)$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$P(X < 3)$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$P(X < 3) = P(X \leq 2)$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$\begin{aligned} P(X < 3) &= P(X \leq 2) \\ &= F(2) \text{ \textit{fordelingsfunktionen}} \end{aligned}$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$
\begin{aligned}
P(X < 3) &= P(X \leq 2) \\
&= F(2) \text{ \textit{fordelingsfunktionen}} \\
&= P(X = 1) + P(X = 2)
\end{aligned}
$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$\begin{aligned} P(X < 3) &= P(X \leq 2) \\ &= F(2) \text{ \textit{fordelingsfunktionen}} \\ &= P(X = 1) + P(X = 2) \\ &= f(1) + f(2) \text{ \textit{tæthedsfunktionen}} \end{aligned}$$

Fair terning eksempel

Lad X repræsentere et kast med en fair terning
Udregn sandsynligheden for at få udfald under 3:

$$\begin{aligned} P(X < 3) &= P(X \leq 2) \\ &= F(2) \quad \textit{fordelingsfunktionen} \\ &= P(X = 1) + P(X = 2) \\ &= f(1) + f(2) \quad \textit{tæthedsfunktionen} \\ &= \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \end{aligned}$$

DTU Compute
Institut for Matematik og Computer Science

Fair terning eksempel

Udregn sandsynligheden for at få udfald over eller lig 3:

Fair terning eksempel

Udregn sandsynligheden for at få udfald over eller lig 3:

$$P(X \geq 3)$$

Fair terning eksempel

Udregn sandsynligheden for at få udfald over eller lig 3:

$$P(X \geq 3) = 1 - P(X \leq 2)$$

Fair terning eksempel

Udregn sandsynligheden for at få udfald over eller lig 3:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - F(2) \ \textit{fordelingsfunktionen} \end{aligned}$$

Fair terning eksempel

Udregn sandsynligheden for at få udfald over eller lig 3:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - F(2) \ \textit{fordelingsfunktionen} \\ &= 1 - \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Oversigt

Konkrete Statistiske fordelinger

- Der findes en række statistiske fordelinger, som kan bruges til at beskrive og analysere forskellige problemstillinger med
- I dag er det diskrete fordelinger:
  - Binomial fordelingen
  - Den hypergeometriske fordeling
  - Poisson fordelingen

DTU Compute
Institut for Matematik og Computer Science

- Et eksperiment med to udfald (succes eller ikke-succes) gentages
- X er antal succeser efter n gentagelser

- Et eksperiment med to udfald (succes eller ikke-succes) gentages
- X er antal succeser efter n gentagelser
- Så følger X binomial fordelingen

$$X \sim B(n, p)$$

- n antal gentagelser
- p sandsynligheden for succes i hver gentagelse

Binomial fordelingens tæthedsfunktion giver sandsynligheden for x antal succeser

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

Binomial fordeling eksempel

```
## Sandsynlighed for success
p <- 0.1
## Antal gentagelser af succes og ikke-succes eksperimentet
nRepeat <- 30
## Simuler Bernoulli eksperiment nRepeat gange
tmp <- sample(c(0,1), size=nRepeat, prob=c(1-p,p), replace=TRUE)
## x er nu
sum(tmp)

## Lav tilsvarende med funktion til simulering af binomial fordeling
rbinom(1, size=30, prob=p)

###############
## Fair terning eksempel

## Antal simulerede realiseringer
n <- 30
## Træk uafhængigt fra mængden (1,2,3,4,5,6) med ens sandsynlighed
xFair <- sample(1:6, size=n, replace=TRUE)
## Tæl sammen for mange seksere
sum(xFair == 6)

## Lav tilsvarende med rbinom()
rbinom(n=1, size=30, prob=1/6)
```

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl

DTU Compute
Institut for Matematik og Computer Science

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen

DTU Compute
Institut for Matematik og Computer Science

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen

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 0.7.

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

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X = x) = f(x; n, p)$

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X = 6) = f(6; n, p)$

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X = 6) = f(6; n, p)$
- Step 4)
  - Hvad er antal trækninger?
  - Hvad er succes-sandsynligheden?

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 0.7.
I løbet af en dag indrapporteres 6 fejl. Hvad er sandsynligheden for at samtlige fejl udbedres?

- Step 1) Hvad skal repræsenteres: X er <u>antal udbedrede fejl</u>
- Step 2) Hvilken fordeling: X følger <u>binomial fordelingen</u>
- Step 3) Hvilken sandsynlighed: $\underline{P(X = 6) = f(6; n, p)}$
- Step 4)
  - Hvad er antal trækninger? $\underline{n = 6}$
  - Hvad er succes-sandsynligheden? $\underline{p = 0.7}$

DTU Compute
Institut for Matematik og Computer Science

Eksempel 1

Hvad er sandsynligheden for at 2 eller færre fejl bliver udbedret samme dag?

Eksempel 1

Hvad er sandsynligheden for at 2 eller færre fejl bliver udbedret samme dag?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen

Eksempel 1

Hvad er sandsynligheden for at 2 eller færre fejl bliver udbedret samme dag?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X\ ?\ ?)$

Eksempel 1

Hvad er sandsynligheden for at 2 eller færre fejl bliver udbedret samme dag?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X \leq 2) = F(2; n, p)$

Eksempel 1

Hvad er sandsynligheden for at 2 eller færre fejl bliver udbedret samme dag?

- Step 1) Hvad skal repræsenteres: X er antal udbedrede fejl
- Step 2) Hvilken fordeling: X følger binomial fordelingen
- Step 3) Hvilken sandsynlighed: $P(X \leq 2) = F(2; n, p)$
- Step 4)
  - Hvad er antal trækninger? $n = 6$
  - Hvad er succes-sandsynligheden? $p = 0.7$

Hypergeometrisk fordeling

- X er igen antal succeser, men nu er det *uden tilbagelægning ved gentagelsen*

Hypergeometrisk fordeling

- X er igen antal succeser, men nu er det *uden tilbagelægning ved gentagelsen*
- X følger da den hypergeometriske fordeling

$$X \sim H(n, a, N)$$

- n er antallet af trækninger
- a er antallet af succeser i populationen
- N elementer store population

Hypergeometrisk fordeling

- Sandsynligheden for at få x succeser er

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

- n er antallet af trækninger
- a er antallet af succeser i populationen
- N elementer stor population
- i `R`, e.g. function `dhyper`:
  - `k` svarer til n
  - `n` svarer til $N - a$
  - `m` svarer til a

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling
- Step 3) Hvilken sandsynlighed:
  $P(X\ ?\ ?)$

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling
- Step 3) Hvilken sandsynlighed:
  $P(X \geq 1) =$

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling
- Step 3) Hvilken sandsynlighed:
  $P(X \geq 1) = 1 - P(X = 0) = 1 - f(0; n, a, N)$

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling
- Step 3) Hvilken sandsynlighed:
  $P(X \geq 1) = 1 - P(X = 0) = 1 - f(0; n, a, N)$
- Step 4)
  - Hvad er antal trækninger?
  - Hvor mange succeser er der?
  - Hvor mange er der i alt?

DTU Compute
Institut for Matematik og Computer Science

Eksempel 2

I en forsendelse af 10 hard disks har 2 mindre skrammer.

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

- Step 1) Hvad skal repræsenteres: X er antal med skrammer
- Step 2) Hvilken fordeling: X følger den hypergeometriske fordeling
- Step 3) Hvilken sandsynlighed:
  $P(X \geq 1) = 1 - P(X = 0) = 1 - f(0; n, a, N)$
- Step 4)
  - Hvad er antal trækninger? $n = 3$
  - Hvor mange succeser er der? $a = 2$
  - Hvor mange er der i alt? $N = 10$

Binomial vs. hypergeometrisk

- Binomial fordelingen anvendes også for at analysere stikprøver med tilbagelægning (Tænk på en terningekast)
- Når man vil analysere stikprøver uden tilbagelægning anvendes den hypergeometriske fordeling (Tænk på træk fra en hat)

Poissonfordelingen

- Poisson fordelingen anvendes ofte som en fordeling (model) for tælletal, hvor der ikke er nogen naturlig øvre grænse
- Poisson fordelingen karakteriseres ved en intensitet, dvs. på formen antal/enhed
- Parameteren λ angiver intensiteten
- Typisk hændelser per tidsinterval
- Intervallerne mellem hændelserne er uafhængige, dvs. processen er hukommelsesløs

Poissonfordelingen

X følger Poisson fordelingen

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

DTU Compute
Institut for Matematik og Computer Science

Eksempel 3.1

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?

DTU Compute
Institut for Matematik og Computer Science

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger Poisson fordelingen

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger Poisson fordelingen
- Step 3) Hvilken sandsynlighed: $P(X\ ?\ ?)$

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger Poisson fordelingen
- Step 3) Hvilken sandsynlighed: $P(X \leq 2)$

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger Poisson fordelingen
- Step 3) Hvilken sandsynlighed: $P(X \leq 2)$
- Step 4) Hvad er raten:

Eksempel 3.1

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?

- Step 1) Hvad skal repræsenteres: X er antal patienter pr. dag
- Step 2) Hvilken fordeling: X følger Poisson fordelingen
- Step 3) Hvilken sandsynlighed: $P(X \leq 2)$
- Step 4) Hvad er raten: $\lambda = 0.3$ patienter per dag

Eksempel 3.2

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 præcis 2 patienter?

Eksempel 3.2

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 præcis 2 patienter?

- Step 3) Hvilken sandsynlighed: $P(X\ ?\ ?)$

Eksempel 3.2

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 præcis 2 patienter?

- Step 3) Hvilken sandsynlighed: $P(X\ ?\ ?) P(X = 2)$

Eksempel 3.3

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 mindst 2 patienter?

Eksempel 3.3

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 mindst 2 patienter?

- Step 3) Hvilken sandsynlighed:

Eksempel 3.3

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 mindst 2 patienter?

- Step 3) Hvilken sandsynlighed: $P(X\ ?\ ?)$

Eksempel 3.3

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 mindst 2 patienter?

- Step 3) Hvilken sandsynlighed: $P(X \geq 2)$

Eksempel 3.3

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 mindst 2 patienter?

- Step 3) Hvilken sandsynlighed: $P(X \geq 2) = 1 - P(X \leq 1)$

Eksempel 3.4

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

Eksempel 3.4

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

- Step 1) Hvad skal repræsenteres:
  - Fra X antal per dag
  - Til $X^{3\text{dage}}$ som er *patienter per 3 dage*

DTU Compute
Institut for Matematik og Computer Science

Eksempel 3.4

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

- Step 1) Hvad skal repræsenteres:
  - Fra X antal per dag
  - Til X^{3dage} som er *patienter per 3 dage*
- Step 2) Hvilken fordeling følger X^{3dage}: Poisson fordelingen

Eksempel 3.4

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

- Step 1) Hvad skal repræsenteres:
  - Fra X antal per dag
  - Til X^{3dage} som er *patienter per 3 dage*
- Step 2) Hvilken fordeling følger X^{3dage}: Poisson fordelingen
- Step 3) Hvilken sandsynlighed: $P(X^{\text{3dage}} = 1)$
- Step 4) Skaler raten
  - Fra $\lambda = 0.3$ patienter/dag til $\lambda_{3dage} = 0.9$ patienter/3dage

R	Betegnelse
`binom`	Binomial
`hyper`	hypergeometrisk
`pois`	poisson

- `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.

Husk at hjælp til funktion mm. fåes ved at sætte '?' foran navnet.

Eksempel binomial fordelt: $P(X \leq 5) = F(5; 10, 0.6)$

```
pbinom(q=5, size=10, prob=0.6)
## Få hjælpen med
?pbinom
```

Oversigt

Middelværdi (mean) og forventningsværdi (expectation)

Stokastisk variabels middelværdi

$$\mu = E(X) = \sum_{\text{alle x}} x f(x)$$

- Det "rigtige gennemsnit"
- Fortæller hvor "midten" af X er

Middelværdi eksempel

Middelværdi af en terning

$$
\begin{aligned}
\mu = E(X) = & \\
& = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\
& = 3.5
\end{aligned}
$$

```r
## Antal simulerede realiseringer
n <- 30
## Træk uafhængigt fra mængden (1,2,3,4,5,6) med ens sandsynlighed
xFair <- sample(1:6, size=n, replace=TRUE)

## Udregn empirisk middelværdi (sample mean, læg mærke til
## i R hedder funktionen 'mean')
mean(xFair)
```

Jo flere observationer, jo tættere kommer man på den rigtige middelværdi

$$\lim_{n \to \infty} \hat{\mu} = \mu$$

- Prøv det i R

Varians

Definition

$$\sigma^2 = \text{Var}(X) = \sum_{\text{alle x}} (x-\mu)^2 f(x)$$

- Et mål for spredningen
- Den "rigtige spredning" af X (modsat empirisk varians (sample variance))

Varians eksempel

Varians af terningekast

$$\begin{aligned}\sigma^2 = E[(X-\mu)^2] =& \\ &= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} \\ &\quad + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} \\ &\approx 2.92\end{aligned}$$

Varians eksempel

```r
## Antal simulerede realiseringer
n <- 30
## Træk uafhængigt fra mængden (1,2,3,4,5,6) med ens sandsynlighed
xFair <- sample(1:6, size=n, replace=TRUE)

## Udregn empirisk varians (sample variance, læg mærke til
## i R hedder funktionen 'var')
var(xFair)
```

DTU Compute
Institut for Matematik og Computer Science

Middelværdi og varians for kendte diskrete fordelinger

Binomial fordelingen:

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

DTU Compute
Institut for Matematik og Computer Science

Middelværdi og varians for kendte diskrete fordelinger

Den hypergeometriske fordeling:

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

Middelværdi og varians for kendte fordelinger

Poisson fordelingen:

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

Terninge eksempel, se forskel på empirisk middelværdi og middelværdi

```
## Gentag 10 gange: Tæl sammen for mange seksere på 30 slag
antalSeksere <- rbinom(n=10, size=30, prob=1/6)

## Endelig kan vi se på empirisk middelværdi (sample mean)
mean(rbinom(n=10, size=30, prob=1/6))
## Den (rigtige) middelværdi (mean)
n * 1/6
```

Oversigt