

# Sandsynlighedsregning

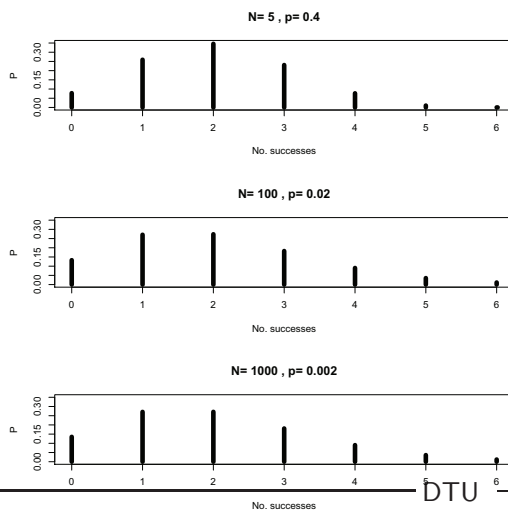
## 5. forelæsning

Bo Friis Nielsen

Matematik og Computer Science  
 Danmarks Tekniske Universitet  
 2800 Kgs. Lyngby – Danmark  
 Email: bfni@dtu.dk

### Poisson-fordelingen

En binomialfordelingen med mange forsøg, hvert med lille sandsynlighed.



$n$  stor  
 $p$  lille  
 $np$  moderat.

### Dagens emner afsnit 3.5 og 4.1

Poissonfordelingen



$$P(X = x) = \frac{\mu^x}{x!} e^{-\mu}, \quad E(X) = \mu, \quad \text{Var}(X) = \mu$$

- Kontinuerte stokastiske variable
  - ◇ Tæthed (density):  $f(x) \geq 0$ ,  $\int f(x)dx = 1$ ,  
 $P(X \in dx) = f(x)dx$ ,  $P(a < X \leq b) = \int_a^b f(x)dx$
  - ◇ Middelværdi, varians (momenter):  $E(X) = \int x f(x)dx$   
 $E(g(X)) = \int g(x) f(x)dx$ ,  $E(X^k) = \int x^k f(x)dx$ ,
- Normalfordelingen:  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$   $Z = \frac{X-\mu}{\sigma}$ 
  - ◇ Den centrale grænseværdisætning
- Standardisering  $Z = \frac{X-E(X)}{SD(X)}$ ,  $E(Z) = 0$ ,  $SD(Z) = 1$

### Fra binomialfordelingen til Poissonfordelingen



$$\begin{aligned} X \sim \text{Bin}(n, p), \quad P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{(np)^x}{x!} \frac{n!}{n^x (n-x)!} (1-p)^n (1-p)^{-x} \\ &= \frac{(np)^x}{x!} (1-p)^n (1-p)^{-x} \prod_{k=0}^{x-1} \frac{n-k}{n} \\ &= \frac{(np)^x}{x!} \left(1 - \frac{np}{n}\right)^n (1-p)^{-x} \prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right), \quad \mu = np \\ &= \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x} \prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right) \rightarrow \frac{\mu^x}{x!} e^{-\mu} \end{aligned}$$

## Eksempler på Poissonfordelte variable (?)



- Antal børn født i et kalenderår i Danmark med en given sjælden sygdom.
- Antal radioaktive henfald i en prøve indenfor et bestemt tidsrum.
- Antal komponenter i en batch der er defekte.

## Sandsynlighedsfordelingen for $X \sim \text{Pois}(\mu)$



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

kan findes som grænsen af  $\text{Bin}(n, \mu/n)$  for  $n \rightarrow \infty$

## Momenter i Poisson-fordelingen

Vi får middelværdi:

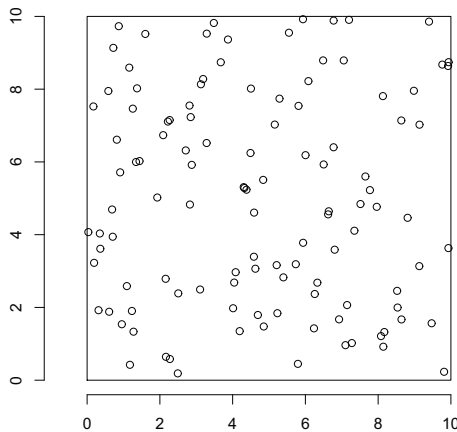
$$E(X) = \mu \quad \text{og varians:}$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2 = \mu$$

Eller

$$\text{Var}(X) = E(X(X - 1)) + E(X) - \mu^2 = \mu$$

## Poisson punktprocessen (scatter)



## Antagelser (p. 229)



Antagelse 1: Ikke multiple forekomster, hvert punkt eller forekomst har særskilt placering

Antagelse 2: Tilfældighed. Alle disjunkte delområder med samme areal bliver "ramt" med samme sandsynlighed og uafhængigt af hinanden.

"Poisson Scatter Theorem" (p. 230):

Punktprocessen har en intensitet (tæthed)  $\lambda$  så:

Antallet af forekomster i et område  $B$  er  $\sim \text{Pois}(\lambda|B|)$

Antallet af forekomster i disjunkte områder er uafhængige.

## Overvejning



Hvis to Poisson-punktprocesser overlejres, er resultatet en ny Poisson punktproces. Intensiteten i den nye proces er summen af intensiteterne i de to komponenter.

## Udtynding

Hvis hvert punkt fjernes fra en Poisson-punktproces med sandsynlighed  $1 - p$  (bibeholdes med sandsynlighed  $p$ ), er resultatet en ny Poisson-punktproces med intensitet  $p \cdot \lambda$ .

## Middelværdi for Poissonfordeling



$$E(X) = \sum_x x f(x) = \sum_x x P(X = x) = \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu}$$

- Lidt manipulation med summen giver

$$\begin{aligned} \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} &= \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} e^{-\mu} = \mu e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} \\ &= \mu \sum_{y=0}^{\infty} \frac{\mu^y}{y!} e^{-\mu} = \mu \cdot 1 = \mu \end{aligned}$$

Samme fremgangsmåde for  $E(X(X-1))$  giver os  $\text{Var}(X) = \mu$

## Ligefordelingen (uniform) side 264-265



- Vi ankommer til et busstoppested et øde sted. Vi ved, at der er en time mellem busserne, men kender ikke køreplanen og ved ikke, hvad klokken er.
- Vi vil gerne vide, hvad risikoen er for, at vi skal vente mindst tre kvarter.
- Hvordan kunne vi gribe det an?

## Beregning i ligefordelingen



*Spm:* For  $a < x < y < b$  find  $P(x < X < y)$

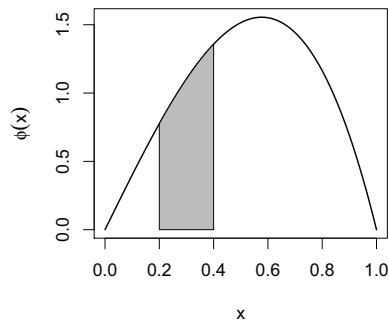
*Svar:*

$$P(x < X < y) = \frac{y-x}{b-a} = c(y-x) = \int_x^y f(u) du$$

Sandsynligheden er altså den del af intervallet fra  $a$  til  $b$ , der udgøres af intervallet fra  $x$  til  $y$ .

Vi kan udtrykke dette som integralet af en normeringskonstant over et delinterval.

## Tætheden $f(x)$ af en kontinuert S.V.



Sandsynligheder findes som  
arealet under kurven (s. 260)

$$P(a < X < b) = \int_a^b f(x) dx$$

NB! Sammenfatning s. 262-263

En funktion  $f(t)$  er en tæthed hvis

1.  $f(t) \geq 0$  for alle  $t$
2.  $f$  er *normeret*, d.v.s.

$$P(-\infty < X < +\infty) = \int_{-\infty}^{\infty} f(t) dt = 1$$

Antag at  $X$  er en stokastisk variabel med tæthed

$$f(x) = \begin{cases} c & \text{hvis } a < x < b \\ 0 & \text{ellers} \end{cases}$$

### Spørgsmål 1

Lad  $U = (X - a)/(b - a)$ . Værdimængden for  $U$  er

- 1   $(0; 1)$
- 2   $(0; b/(b - a))$
- 3   $(-\infty; \infty)$
- 4   $(a/(b - a); b/(b - a))$
- 5  Oplysningerne i teksten er ikke tilstrækkelige
- 6  Ved ikke

## Ligefordelingen

$$f(x) = \begin{cases} c & \text{hvis } a < x < b \\ 0 & \text{ellers} \end{cases} = c \mathbf{1}_{]a,b[}(x)$$

hvor  $\mathbf{1}_A(x)$  er en indikatorfunktion, med værdien 1, når  $x \in A$  (vi undertrykte argumentet  $\omega$  for en stokastisk indikatorfunktion).

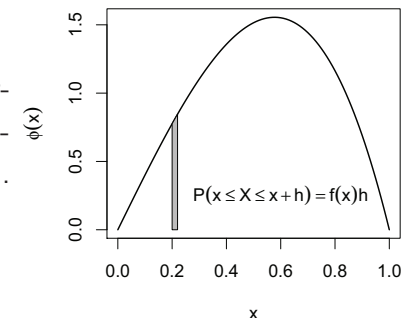
Hvad er så  $c$ ?

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x) dx \\ &= \int_{-\infty}^a f(x) dx + \int_a^b f(x) dx + \int_b^{\infty} f(x) dx = 0 + (b - a) \cdot c + 0 \\ c &= \frac{1}{b - a} \end{aligned}$$

## Tætheder fortolket som sandsynligheder

For en kontinuert stokastisk variabel er alle punktsandsynligheder 0.

Sandsynligheden for at observere en observation i en omegn  $h$  af  $x$  findes som  $hf(x)$ .



## Middelværdi og varians (side 261)



Generelt erstattes beregninger af summer (for diskrete variable) med integraler (for kontinuerte variable)

Eksempelvis

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx$$

Vi har stadig den meget vigtige beregningsformel

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

## Transformation for ligefordelingen



$$U = \frac{X - a}{b - a} \Leftrightarrow X = a + (b - a)U$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x f(x) dx \\ &= \frac{1}{2} \left[ \frac{x^2}{b - a} \right]_a^b = \frac{a + b}{2} \end{aligned}$$

Alternativt

$$E(X) = E(a + (b - a)U) = E(a) + (b - a)E(U) = a + (b - a)\frac{1}{2}$$

## Variansberegning



$$\text{Var}(X) = \text{Var}(a + (b - a)U) = (b - a)^2 \text{Var}(U)$$

$$\text{Var}(X) = (b - a)^2 (E(U^2) - (E(U))^2)$$

$$E(U^2) = \int_0^1 u^2 du = \frac{1}{3}$$

$$\text{Var}(X) = (b - a)^2 \left( \frac{1}{3} - \left( \frac{1}{2} \right)^2 \right) = \frac{(b - a)^2}{12}$$

Antag at  $X$  er en stokastisk variabel med tæthed  $f(x) = cx(1 + x)$  for  $-1 < x < 0$ , og 0 ellers.

### Spørgsmål 2

Værdien af  $c$  er

- 1  -6
- 2  1
- 3  2
- 4  6
- 5   $f(x)$  kan ikke være en tæthed
- 6  Ved ikke

### Opgave 4.1.3 (modificeret)



Lad  $X$  være en stokastisk variabel med tæthed  $f(x) = cx(1+x)$  for  $-1 < x < 0$ , og 0 ellers.

*Spm:* Find værdien af  $c$ .

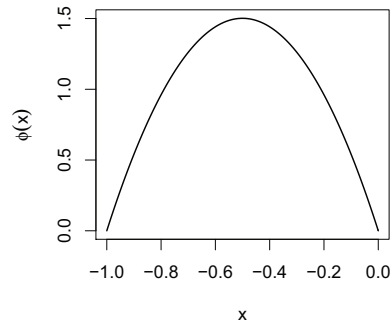
*Svar:* Vi kræver

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

... og finder

$$\int_{-1}^0 cx(1+x) dx = -\frac{c}{6}$$

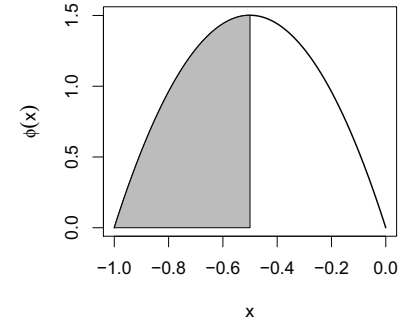
altså  $c = -6$ .



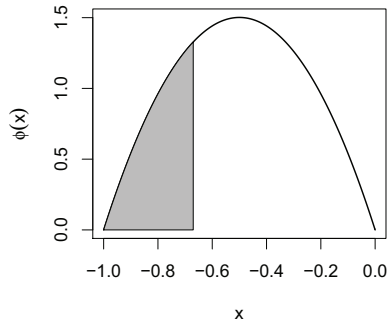
*Spm:* Bestem værdien af  $P(X \leq -\frac{1}{2})$

*Svar:* Da fordelingen er symmetrisk om  $-\frac{1}{2}$  får vi

$$P\left(X \leq -\frac{1}{2}\right) = \frac{1}{2}$$



*Spm:* Bestem værdien af  $P(X \leq -\frac{2}{3})$



*Svar:*



Vi benytter sammenhængen mellem tæthed og sandsynlighed side 260 eller side 263

$$P\left(X \leq -\frac{2}{3}\right) = \int_{-1}^{-\frac{2}{3}} cx(1+x) dx = c \left[ \frac{1}{2}x^2 + \frac{1}{3}x^3 \right]_{x=-1}^{x=-\frac{2}{3}}$$

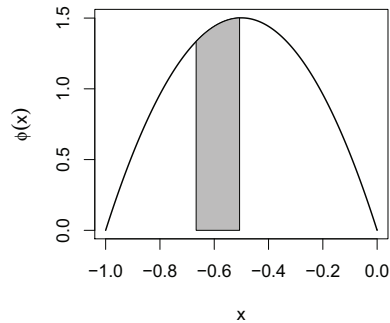
Med  $c = -6$  får vi

$$P\left(X \leq -\frac{2}{3}\right) = \frac{7}{27}$$

Spm: Hvad er  $P\left(-\frac{2}{3} \leq X \leq -\frac{1}{2}\right)$

Svar:

$$P\left(-\frac{2}{3} \leq X \leq -\frac{1}{2}\right) = \frac{1}{2} - \frac{7}{27} = \frac{13}{54}$$



Spm: Hvad er middelværdi og varians af  $X$ ?



Svar: Middelværdien er  $-\frac{1}{2}$  ud fra symmetrien.

Vi finder  $E(X^2)$

$$E(X^2) = -6 \int_{-1}^0 x^2 \cdot x(1+x) dx = \frac{6}{20}$$

...og dermed

$$\text{Var}(X) = \frac{6}{20} - \frac{1}{4} = \frac{1}{20}$$

## En række generiske modeller



- Forskellige ingeniørtekniske problemstillinger:
  - ◊ Hvor mange ulykker sker der i et vejkryds
  - ◊ Tidsafstanden mellem to pakker i et kommunikationsnetværk
  - ◊ Vægt/mængde angivelse af et kemisk produkt
  - ◊ Antal positive respondenter i en markedsanalyse
- og varianter heraf

## Den centrale grænseværdisætning side 268, (196)



- I naturen/teknikken observeres ofte en "klokkeformet" fordeling
- Dette understøttes teoretisk netop af CGS
- En sum af mange bidrag er (tilnærmelsesvist) normal fordelt når intet enkeltbidrag "dominerer" summen
- Formen (formlen!) for normalfordelingen fremkommer netop gennem udledningen - beviset - for CGS

## Normalfordeling - udtryk og egenskaber side 266-267, 477, 484-485



Hvis  $X \sim N(\mu, \sigma^2)$ , så:

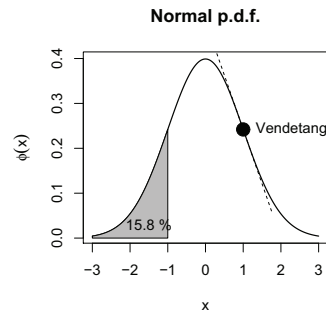
$X$  har tætheden

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Middelværdi  $E(X) = \mu$

Varians  $\text{Var}(X) = \sigma^2$

Spredning  $SD(X) = \sigma$



For  $\mu = 0$  og  $\sigma^2 = 1$  bruger vi  $f(x) = \phi(x)$

Vi betragter en stokastisk variabel  $X$ , der er standard normal fordelt.



### Spørgsmål 3

Sandsynligheden  $P(0 < X < 0,001)$  findes (eventuelt approksimativt) til

- 1  0,001
- 2   $\Phi(0,01) - \frac{1}{2}$
- 3   $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}0,001^2}$
- 4   $\Phi(0)$
- 5   $\frac{0,001}{\sqrt{2\pi}}$
- 6  Ved ikke

Hvor  $\Phi$  som sædvanlig betegner fordelingsfunktionen for en standardiseret normalfordelt variabel.

## Den standardiserede normalfordeling



$$X \sim N(\mu, \sigma^2) : P(a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

De normalfordelte variable bevares under affine transformationer:

$$c_1X + c_2 \sim N(c_1\mu + c_2, c_1^2\sigma^2)$$

Vi kan *standardisere*  $X$ :

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \Phi(z)$$

### Opgave 4.1.8



Vægtmålinger af en metalklump formodes at være uafhængige og identisk fordelte (I.I.D.), med middelværdien af 12 gram og standardafvigelse 1,1 gram.

*Spørgsmål:*

Hvad er sandsynligheden for at en enkelt måling er mellem 11,8 og 12,2 gram under antagelse af, at de enkelte målinger kan beskrives ved en normalfordeling?



Svar:



Lad  $W$  betegne vægten på en måling.  $W \in normal(12, 1.1^2)$

Spørgsmålet kan nu formuleres: Bestem  $P(11,8 < W < 12,2)$ .

Transformér til standardiseret normalfordeling:

$$\begin{aligned}
P(11,8 < W < 12,2) &= P(11,8 - 12 < W - 12 < 12,2 - 12) \\
&= P\left(\frac{11,8 - 12}{1,1} < \frac{W - 12}{1,1} < \frac{12,2 - 12}{1,1}\right) \\
&= \Phi(0,1818) - \Phi(-0,1818) = 2 \cdot 0,0714 = 0,1428
\end{aligned}$$

Spørgsmål:



Beregn sandsynligheden for at gennemsnittet af 100 målinger ligger i intervallet [11,8 gram;12,2 gram]

(er normalfordelingsantagelsen nødvendig her?)

Svar:

Lad  $W_i$  betegne resultatet af den  $i$ 'te måling.

Lad  $\bar{W}$  være gennemsnittet  $\bar{W} = \frac{\sum_{i=1}^{100} W_i}{100}$

Spørgsmålet er altså:

$$P(11,8 < \bar{W} < 12,2)$$

Ifølge C.G.S. (s. 196) er  $\bar{W}$  approksimativt normalfordelt

Sum $S = \sum_{i=1}^n W_i$ : $S \sim N(n\mu, n\sigma^2)$	Gennemsnit $\bar{W} = S/n$ : $\bar{W} \sim N\left(\mu, \frac{1}{n}\sigma^2\right)$
---	---

Standardisér: Indfør



$$Z = \frac{\bar{W} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{W} - \mu}{\sigma} \sim N(0,1)$$

Omformulér spørgsmålet til at gælde  $Z$ :

$$P(a < \bar{W} < b) = P\left(\sqrt{n} \frac{a - \mu}{\sigma} < \sqrt{n} \frac{\bar{W} - \mu}{\sigma} < \sqrt{n} \frac{b - \mu}{\sigma}\right)$$

$$= P\left(\sqrt{n} \frac{a - \mu}{\sigma} < Z < \sqrt{n} \frac{b - \mu}{\sigma}\right) = \Phi\left(\sqrt{n} \frac{b - \mu}{\sigma}\right) - \Phi\left(\sqrt{n} \frac{a - \mu}{\sigma}\right)$$

Vi finder

$$P(11,8 < \bar{W} < 12,2) = P(-1,818 < Z < 1,818) = 0,9312$$

## Afsnit 3.5 og 4.1

Poissonfordelingen



$$P(X = x) = \frac{\mu^x}{x!} e^{-\mu}, \quad E(X) = \mu, \quad \text{Var}(X) = \mu$$

- Kontinuerte stokastiske variable
  - ◇ Tæthed (density):  $f(x) \geq 0$ ,  $\int f(x)dx = 1$ ,  
 $P(X \in dx) = f(x)dx$ ,  $P(a < X \leq b) = \int_a^b f(x)dx$
  - ◇ Middelværdi, varians (momenter):  $E(X) = \int x f(x)dx$   
 $E(g(X)) = \int g(x) f(x)dx$ ,  $E(X^k) = \int x^k f(x)dx$ ,
- Normalfordelingen:  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$   $Z = \frac{X-\mu}{\sigma}$ 
  - ◇ Den centrale grænseværdisætning
- Standardisering  $Z = \frac{X-E(X)}{SD(X)}$ ,  $E(Z) = 0$ ,  $SD(Z) = 1$