# Sandsynlighedsregning
# 7. forelæsning


Bo Friis Nielsen

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


## Dagens emner afsnit 4.5 og 4.6

- (Kumulerede) fordelingsfunktion $\mathsf{P}(X \leq x) = F(x)$
  - Fraktiler $F(x_p) = p$ - specielt median $F(x_m) = 0.5$
- Fordeling for maksimum og minimum $F_{max}(x) = \prod_{i=1}^{n} F_i(x)$ og $F_{min}(x) = 1 - \prod_{i=1}^{n}(1 - F_i(x))$
- Fordelinger af ordnede variable

$$f_{(k)}(x) = nf(x)\begin{pmatrix} n-1 \\ k-1 \end{pmatrix}(F(x))^{k-1}(1-F(x))^{n-k}$$

- Betafordelingen

$$f(x) = \frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}$$

med $B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\mathrm{d}x = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}\left(= \frac{(r-1)!(s-1)!}{(r+s-1)}, \text{ for } (r,s) \text{ heltallige}\right)$

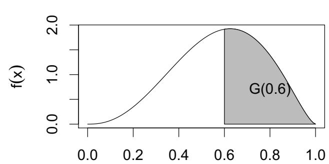
## Fordelingsfunktionen (c.d.f.) $F(x)$

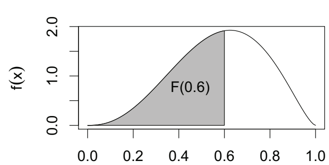
For levetidsfordelinger indførte vi overlevelsesfunktionen

$$G(t) = \mathsf{P}(T > t) = \int_t^{\infty} f(s)\, ds$$

For **alle** stokastiske variable $X$ indfører vi nu **fordelingsfunktionen**

$$F(x) = \mathsf{P}(X \leq x)$$





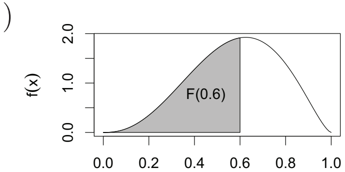

## $F$ for kontinuerte og diskrete variable

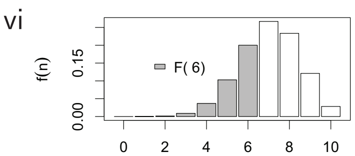
For kontinuerte variable er $F(\cdot)$ kontinuert (!) og

$$F(x) = \int_{-\infty}^{x} f(\xi)\, d\xi$$

For diskrete fordelinger på $\mathbf{Z}$ har vi

$$F(x) = \sum_{n=-\infty}^{x} \mathsf{P}(X = n)$$



Vi betragter en eksponential$(\lambda)$ fordelt variabel $X$.

## Spørgsmål 1

Fordelingsfunktionen for $X$ er

1 ☐ $F(x) = \lambda e^{-\lambda x}$

2 ☐ $F(x) = 1 - \lambda e^{-\lambda x}$

3 ☐ $F(x) = 1 - e^{-\lambda x}$

4 ☐ $F(x) = e^{-\lambda x}$

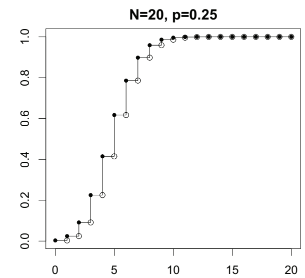
5 ☐ $F(x) = \int_0^x e^{-\lambda x} \mathrm{d}x$

6 ☐ Ved ikke

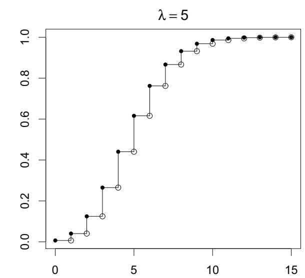


# Eksempler på fordelingsfunktioner

Binomialf. $\sum_{t=0}^{x} \begin{pmatrix} n \\ t \end{pmatrix} p^t (1-p)^{n-t}$

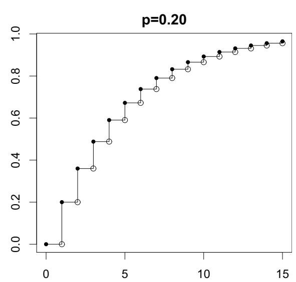



Poissonf. $\sum_{n=0}^{x} \frac{\lambda^n}{n!} e^{-\lambda}$

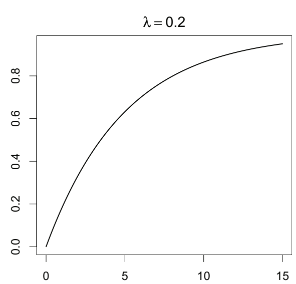



# ... flere fordelingsfunktioner

Geometrisk f. på $\mathbf{N}$ $1-(1-p)^x$

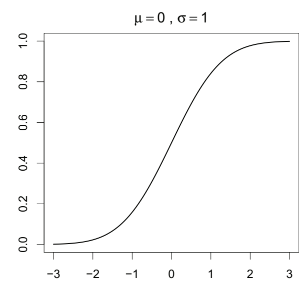



Exponentialf. $1 - \exp(-\lambda x)$

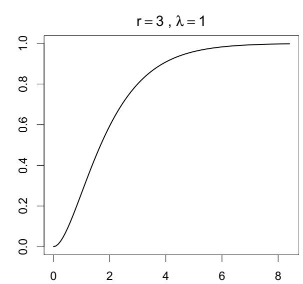



# ... og endnu flere

Normalf. $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t$




Gammaf. $1 - \sum_{n=0}^{r-1} \frac{(\lambda x)^n}{n!} e^{-\lambda x}$



# Fraktiler (percentiles)

Median $m$:

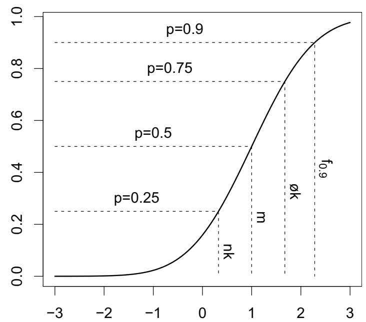
$$\mathsf{P}(X \leq m) = 0.5$$

Nedre/øvre kvartil

$$\mathsf{P}(X \leq nk) = 0.25$$

$$\mathsf{P}(X \leq øk) = 0.75$$

Generelt: $p$-fraktilen $f_p$ er givet ved




$$\mathsf{P}(X \leq f_p) = p$$

(Lidt besværligt for diskrete fordelinger)

# Simulering af stokastiske variable

Givet en kontinuert voksende fordelingsfunktion $F$.

*Spm:* Hvordan kan vi sample en stokastisk variable $X$ med fordelingsfunktion $F$?

*Svar:* Simulér først $U$ fra en uniform fordeling på $[0, 1]$.

Bestem dernæst

$$X = F^{-1}(U)$$

d.v.s. $x$ er $u$-fraktilen i $F$.

Hvorfor virker det? Fordi:

$$\mathsf{P}(X \leq x) = \mathsf{P}(F(X) \leq F(x)) = \mathsf{P}(U \leq F(x)) = F(x)$$



# Brug af dette

- Matlab genererer tilfældige tal mellem 0 og 1 med *rand*
- Transformation med $-\frac{1}{\lambda}\ln(U)$
- Lav histogram over disse
- Tilsvarende histogram over variable genereret med *exprnd*



# Fordeling af maximum og minimum

- Hvor høj er den højeste i en gruppe?
- Hvornår punkterer det første dæk?
- Hvad er den største bølgehøjde en bro vil blive udsat for?
- Største fil på hjemmeside
- Hvornår kommer den sidste, der skal med på skituren?

## Fordelingsfunktionen for maximum

Start med $n$ variable $X_i$, $i = 1, \dots, n$.

... med tilhørende fordelingsfunktioner $F_i(x)$.

Lad $X_{max}$ være maksimum:

$$X_{max} = \max_i X_i$$

$X_{max}$ er da en stokastisk variabel med fordelingsfunktion

$$F_{max}(x) = \mathsf{P}(X_{max} \le x) = \mathsf{P}(\forall i: \; X_i \le x)$$

Hvis $\{X_i, \; i = 1, \dots, n\}$ er uafhængige får vi

$$F_{max}(x) = \prod_{i=1}^{n} F_i(x) \qquad (\text{ for } F_i(x) = F(x)\text{: } \; F_{max}(x) = F(x)^n)$$



## Maksimum af to terningekast

Lad $X$ og $Y$ være antal øjne på to uafhængige terningekast.

Lad $Z = \max\{X, Y\}$.

| $z$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathsf{P}(Z = z)$ | $\frac{1}{36}$ | $\frac{3}{36}$ | $\frac{5}{36}$ | $\frac{7}{36}$ | $\frac{9}{36}$ | $\frac{11}{36}$ |

Dermed får vi

| $z$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathsf{P}(Z \le z)$ | $\frac{1}{36}$ | $\frac{4}{36}$ | $\frac{9}{36}$ | $\frac{16}{36}$ | $\frac{25}{36}$ | $\frac{36}{36}$ |

Bemærk! Dette er lige netop $(F_X(z))^2$ med $F_X(x) = x/6$



## Fordelingsfunktionen for minimum

Samme $n$ variable $X_i$, samme fordelingsfunktioner $F_i(x)$.

Lad $X_{min}$ være minimum:

$$X_{min} = \min_i X_i$$

$X_{min}$ har fordelingsfunktionen

$$F_{min}(x) = \mathsf{P}(X_{min} \le x) = 1 - \mathsf{P}(X_{min} > x) = 1 - \mathsf{P}(\forall i: \; X_i > x)$$

Hvis $\{X_i, \; i = 1, \dots, n\}$ er uafhængige får vi nu

$$1 - F_{min}(x) = \prod_{i=1}^{n} (1 - F_i(x)) \Leftrightarrow F_{min}(x) = 1 - \prod_{i=1}^{n} (1 - F_i(x))$$

For $F_i(x) = F(x)$: $\qquad F_{min}(x) = 1 - (1 - F(x))^n$



## Minimum af to uniformt fordelte variable

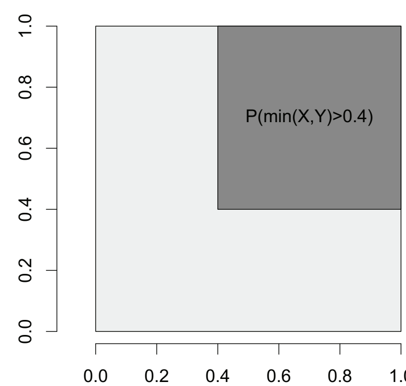
Lad $X, Y$ være uafhængige og uniformt fordelt på $[0, 1]$.

Lad $Z = \min\{X, Y\}$.
Da har vi

$$\mathsf{P}(Z \le z) = 1 - (1 - z)^2 \; .$$

Mere generelt, for $X_i$, $i = 1, \dots, n$, er uafhængige og $U([0, 1])$, får vi:

$$\mathsf{P}(\min_i X_i \le x) = 1 - (1 - x)^n$$



## Eksempel/opgave

En kæde består af $n$ led. Trækstyrken for hvert led har fordelingsfunktionen

$$F(x) = \frac{1}{25}x^2 \qquad 0 < x < 5$$

Vi antager nu, at trækstyrken for hele kæden skal være mindst $\frac{1}{10}$kg med en sandsynlighed på mindst 0,95. Hvor mange led må den da højst bestå af?



## En kæde er ikke stærkere end det svageste led ...

Vi definerer $X_i$ til at være trækstyrken af det $i$'te led.

Trækstyrken af kæden er da $X_{min} = \min_i X_i$ med fordelingsfunktion:

$$\mathsf{P}(X_{min} \leq x) = 1 - (1 - F(x))^n$$

Vi forlanger

$$\mathsf{P}(X_{min} \leq 0,1) = 1 - (1 - F(0,1))^n \leq 0,05$$

Løs m.h.t. $n$:

$$n \leq \frac{\log 0,95}{\log(1 - F(0,1))} \Leftrightarrow n \leq 128$$

hvor $F(0,1) = \frac{1}{25} \cdot 0,1^2$

## En approksimation for fordelingen af minimum

Fordelingsfunktionen for $X_{min} = \min_i X_i$ i kædeeksemplet var

$$\mathsf{P}(X_{min} \leq x) = 1 - (1 - F(x))^n$$

Hvis $n$ er stor og $F(x)$ er lille kan vi approksimere

$$(1 - F(x))^n = \exp(n \log(1 - F(x))) \approx \exp(-nx^2/25)$$

Der er $X_{min}$ ca. Weibull-fordelt, opg. 4.3.4.

For en generel $F$ kan vi approksimere $F(x)$ med det dominerende led i dens Taylor-række, og dermed stadig få en Weibull-fordeling.



## Fordeling af $k$'te mindste side 326

Lad $X_i$, være i.i.d. $(i = 1, \ldots, n)$ med fordelingsfunktion $F(x)$.

Lad $X_{(k)}$ benævne den $k$'te mindste.

$(X_{max} = X_{(n)} \qquad X_{min} = X_{(1)})$

*Spm:* Bestem $\mathsf{P}(x < X_{(k)} \leq x + dx) = f_k(x)\, dx$ !

*Svar:* Under denne hændelse må der gælde at:

- èn af $X_i$'erne ligger i $[x, x + dx]$
- præcis $k - 1$ af dem er mindre end eller lig $x$
- og resten $(n - k)$ er større eller lig $x$.

## Multi/polynomialfordelingen

- $n$ emner
- $m$ kategorier
- Et emne tilhører kategori $j$ med sandsynlighed $p_j$ uafhængigt af de andre emner. Lad $N_j$ være antal emner i kategori $j$

$$\mathsf{P}\left(N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m\right) = \frac{n!}{n_1!n_2!\cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

- Lad $N_1$ være antallet af $X_i \leq x$, og $N_2$ antallet af $X_i$ i intervallet $)x; x + \mathrm{d}x]$, og endeligt $N_3$ af $X_i > x + \mathrm{d}x$.

$$\mathsf{P}\left(N_1 = k-1, N_2 = 1, N_3 = n-k\right)$$
$$= \frac{n!}{(k-1)!1!(n-k)!} F(x)^{k-1} (f(x)\mathrm{d}x)^1 (1 - F(x+\mathrm{d}x)^{n-k}$$

således at

$$f_k(x) = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$$

*Alternativ udledning:*

- èn af $X_i$'erne ligger i $[x, x + dx]$
- præcis $k - 1$ af dem er mindre end eller lig $x$
- og resten $(n - k)$ er større eller lig $x$.

$$f_k(x)\mathrm{d}x = n f(x)\mathrm{d}x \cdot \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} F(x)^{k-1} \cdot (1 - (F(x))^{n-k}$$

således at

$$f_k(x) = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$$



Fire studerende har aftalt, at mødes på Cafe Mig og Annie kl. 0.00. Antag at hver persons ankomsttidspunkt kan beskrives med en normalfordeling med middelværdi kl. 0.00 og med en standardafvigelse på 5 minutter, uafhængigt af de andres ankomsttidspunkt.

### Spørgsmål 2

Sandsynligheden for, at den anden person kommer indenfor de første 10 sekunder efter kl. 0.00, er (approksimativt):

1 ☐ $\frac{1}{5\sqrt{2\pi}} \frac{1}{6}$

2 ☐ $4 \cdot 3 \frac{1}{5\sqrt{2\pi}} \frac{1}{6} \frac{1}{2} \left(\frac{1}{2}\right)^2$

3 ☐ $\Phi\left(\frac{\frac{1}{6}}{5}\right) - \frac{1}{2}$

4 ☐ $(\Phi\left(\frac{\frac{1}{6}}{5}\right) - \frac{1}{2})^2$

5 ☐ $4 \frac{1}{5\sqrt{2\pi}} \frac{1}{6} \left(\frac{1}{2}\right)^2$

6 ☐ Ved ikke

hvor $\Phi$ betegner fordelingsfunktionen for en standardiseret normalfordelt variabel.

## Ordning af uniforme variable

Hvis $X_i$ er U(0,1) (d.v.s. uniformt fordelt på [0,1]) er $F(x) = x$. Vi fandt det generelle resultat:

$$f_k(x) = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$$

Dermed bliver tæthedsfunktionen for $X_{(k)}$:

$$f_k(x) = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} x^{k-1} (1 - x)^{n-k}$$

(Tætheder side 328)

# Et hjælperesultat

Fordi $f_k$ er en tæthedsfunktion har vi

$$\int_0^1 f_k(x)\ dx = 1$$

og dermed

$$\int_0^1 x^{k-1}(1-x)^{n-k}\mathrm{d}x = \frac{1}{n\binom{n-1}{k-1}} = \frac{\Gamma(k)\Gamma(n-k+1)}{\Gamma(n+1)}$$

(idet $\Gamma(n) = (n-1)!$ for $n$ heltallig)



# Betafordelingen

Lidt mere generelt siges en stokastisk variabel med tæthed

$$f(x) = \frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}, \qquad 0 < x < 1$$

for vilkårlige $r > 0, s > 0$ at være $\mathrm{beta}(r,s)$ fordelt.

Her er

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\mathrm{d}x = \frac{\Gamma(r)\ \Gamma(s)}{\Gamma(r+s)}$$

Dermed er $X_{(k)}$ fra før $\mathrm{beta}(k, n-k+1)$ -fordelt.



# Opgave 4.6.2

Lad $X$ være beta$(r,s)$-fordelt.

*Spm:* Find $\mathsf{E}(X^m)$ for $m \geq 1$.

*Svar:*

$$\mathsf{E}(X^m) = \int_0^1 x^m f(x)\mathrm{d}x = \frac{1}{B(r,s)}\int_0^1 x^m x^{r-1}(1-x)^{s-1}\mathrm{d}x$$

og dermed

$$\mathsf{E}(X^m) = \frac{B(r+m,s)}{B(r,s)} = \frac{\Gamma(r+m)\ \Gamma(r+s)}{\Gamma(r+m+s)\ \Gamma(r)}$$



# Opg. 4.6.2. (f): Første momenter i en beta-fordeling

$$\mathsf{E}(X) = \frac{\Gamma(r+1)\ \Gamma(r+s)}{\Gamma(r+s+1)\ \Gamma(r)} = \frac{r}{r+s}$$

$$\mathsf{E}(X^2) = \frac{\Gamma(r+2)\ \Gamma(r+s)}{\Gamma(r+s+2)\ \Gamma(r)} = \frac{r(r+1)}{(r+s)(r+s+1)}$$

$$Var(X) = \mathsf{E}(X^2) - (\mathsf{E}(X))^2 = \frac{rs}{(r+s)^2(r+s+1)}$$

Vi har to æsker, en ulige æske med 1 sort og 3 hvide marmorkugler, og en lige æske med 2 sorte og 2 hvide marmorkugler. Man vælger en æske tilfældigt og dernæst vælges en marmorkugle tilfældigt fra denne æske.

### Spørgsmål 3

Givet vi trækker en hvid marmorkugle, hvad er da sandsynligheden for at den kommer fra den lige æske

1 ☐ $\frac{2}{5}$

2 ☐ $\frac{1}{2}$

3 ☐ $\frac{3}{5}$

4 ☐ $\frac{2}{3}$

5 ☐ Ingen af de ovenstående

6 ☐ Ved ikke

## Nye begreber i dag

DTU

- Fordelingfunktionen $F(x) = \mathrm{P}(X \leq x)$
- Fraktiler: $f_p$ er givet ved $\mathrm{P}(X \leq f_p) = p$
- ... specielt median, øvre og nedre kvartil (p=0.5, 0.75, 0.25)
- Simulation udfra fraktilerne
- Ekstremværdier
- Ordnede stokastiske variable
- Betafordelingen.



## Afsnit 4.5 og 4.6

DTU

- (Kumulerede) fordelingsfunktion $\mathrm{P}(X \leq x) = F(x)$
  - Fraktiler $F(x_p) = p$ - (specielt median $F(x_m) = 0.5$)
- Fordeling for maksimum og minimum $F_{max}(x) = \prod_{i=1}^{n} F_i(x)$ og $F_{min}(x) = 1 - \prod_{i=1}^{n}(1 - F_i(x))$
- Fordelinger af ordnede variable

$$f_{(k)}(x) = nf(x)\begin{pmatrix} n-1 \\ k-1 \end{pmatrix}(F(x))^{k-1}(1-F(x))^{n-k}$$

- Betafordelingen

$$f(x) = \frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}$$

med $B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\mathrm{d}x = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}\left(= \frac{(r-1)!(s-1)!}{r+s-1)}, \text{ for } (r,s) \text{ heltallige}\right)$