# Matematisk Statistik

## 8. Forelæsning 25.02.2021

Klassisk statistik:

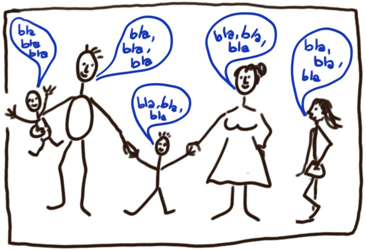
- Fordelingsfamilier
- Estimation i klassisk statistik
  - Momentmetoden
  - Maksimum likelihood metoden
- Estimatorernes egenskaber (det meste nok næste gang)

Hvad gør den “klassiske statistiker”, når den skala analysere data?

1. Bestemmer en **model**:
   en **fordelingsfamilie** der generelt passer til problemet / data.
   Modellen skal indeholde så mange parameter som nødvendigt, og så få som muligt.
2. Fitter modellen, dvs., fisker den passende fordeling i familien ved at **estimere** parameteren af modellen.

Teoretisk udgangspunkt:

- Observerede data $\mathbf{x} = (x_1, \ldots, x_n)$ er en realisering af en vektor $\mathbf{X} = (X_1, \ldots, X_n)$.
- $X_1, \ldots, X_n$ er uafhængige og identisk fordelt.
- Fordelingen af $X_i, i = 1, \ldots, n$ er medlem i en fordelingsfamilie, i.e., en mængde af sandsynlighedsmål, givet ved en fordelingsfunktion.
- Fordelinger identificeres ved en parameter $\theta \in \Theta$, hvor $\Theta$ er parameterrummet.
- Klassisk inferens:
  find den fordeling i familien, der synes mest passende som ophavsmand af de observerede data.
  ☞ bestem $\theta \in \Theta$.




$\Theta =$ 

$\underline{X} = (bla_1, bla_2, bla_3)$

$\underline{X}_1 = (\text{"op"}, \text{"endelig"}, \text{"ryd"})$

$\underline{X}_2 = (\text{"måål"}, \text{"mål"}, \text{"mååål"})$

# Eksempler på fordelingsfamilier, I

**Familier til kontinuerte data**:

- Normalfordelinger $\{N(\mu,\sigma^2) : \underbrace{(\mu,\sigma)}_{\theta} \in \underbrace{\mathbb{R} \times \mathbb{R}^+}_{\Theta}\}$

  Tæthedsfunktion $f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$,

  udfaldsrum: $x \in \mathbb{R}$.

- Exponentialfordelinger $\{\mathrm{Exp}(\lambda) : \lambda \in \mathbb{R}^+\}$

  Tæthedsfunktion $f(x;\lambda) = \lambda \exp(-\lambda x)$,

  udfaldsrum: $x \in \mathbb{R}^+$.

- kender I flere?

# Eksempler på fordelingsfamilier, II

**Familier til diskrete data**:

- Binomialfordelinger $\{\mathrm{Binom}(n,p) : n \in \mathbb{N}, p \in [0,1]\}$
  sandsynlighedsfunktion: $f(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$,
  udfaldsrum: $x \in \mathbb{N}_0$.

  I de fleste anvendelser er antalsparameteren $n$ kendt, så vi arbejder effektivt med familien $\{\mathrm{Binom}(n,p) : p \in [0,1]\}$, med kendt $n$.

- kender I flere?

Betragt en fordelingsfamilie $\{F(.;\theta), \theta \in \Theta\}$, og lad $r$ være dimensionen af $\Theta$.
Vi ønsker at fitte modellen til stikprøven $(x_1, \ldots, x_n)$.

**Momentestimatoren** $\hat{\theta}(x_1, \ldots, x_n) \in \Theta$ er den værdi, der opfylder, at, for $X \sim F(.;\theta)$,

$$\mathrm{E}[X] = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \mathrm{E}[X^2] = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \ldots, \quad \mathrm{E}[X^r] = \frac{1}{n}\sum_{i=1}^{n} x_i^r.$$

Husk,ved kontinuerte fordelinger på $\mathbb{R}$

$$\mathrm{E}[X^k] = \int_{-\infty}^{\infty} x^k f(x;\theta)\,\mathrm{d}x\,,$$

og ved diskrete fordelinger på $\mathbb{Z}$

$$\mathrm{E}[X^k] = \sum_{x \in \mathbb{Z}} x^k f(x;\theta).$$

**Eksempel 6.1:**

Lad $x_1 = 3, x_2 = 4, x_3 = 3, x_4 = 7$ være udfald af uafhængige, identisk Poisson($\lambda$) fordelte stokastiske variabler. Estimer $\lambda$ fra stikprøven.

**Eksempel 6.8:**

$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, \beta]$.

Estimér $\beta$, fra stikprøven: $x_1 = 4, x_2 = x_3 = 1, x_4 = 9$.

**Eksempel 6.11:**

$X_1, X_2, \ldots, X_n$ er uafhængige og identisk fordelt, med fordeling givet ved tæthedsfunktion

$$f(x) = \lambda e^{-\lambda(x-\delta)}, \text{ hvor } \lambda > 0 \text{ og } \delta > 0.$$

Estimér $\lambda$ og $\delta$, fra stikprøven: $x_1 = 3.5, x_2 = 3.9, x_3 = 4.0, x_4 = 4.7$.

## Maksimum likelihood estimation

(se bogen: 6.1)

**Idé**: fra hele fordelingsfamilien, valg den fordeling, hvor den observerede stikprøve $x_1, \ldots, x_n$ har den størreste "sandsynlighed".

**Ingredienser / begreber:**

- Fordelingsfamilie parametriseret med parameter $\theta$; tæthed eller sandsynlighed for $X = x$ er givet ved $f(x;\theta)$.
- Likelihoodfunktion (Definition 6.2 i MSRR)

$$L(\theta \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

  maksimeres af **maksimum likelihood estimatet** $\hat{\theta}_{\text{mle}}$ (tit skriver vi bare $\hat{\theta}$)
- log likelihood

$$\ln L(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

- Score funktion

$$u(\theta) = \frac{\mathrm{d} \ln L}{\mathrm{d}\theta}, \quad \text{ved flere dimensioner:} \quad u(\theta_1, \ldots, \theta_r) = \left( \frac{\partial \ln L}{\partial \theta_1}, \ldots, \frac{\partial \ln L}{\partial \theta_r} \right)$$

## Eksempel 6.1: Poissonfordelingen

Stikprøven $x_1 = 3, x_2 = 4, x_3 = 3, x_4 = 7$ antages at stamme fra en Poisson($\lambda$)-fordeling. Hvad er maksimum-likelihood estimatet?

Sandsynligheder for Poisson($\lambda$):

$$f(x;\lambda) = \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!}$$

Likelihoodfunktion:

$$\begin{aligned} L(\lambda \mid x_1 = 3, x_2 = 4, x_3 = 3, x_4 = 7) &= \frac{\lambda^3 \mathrm{e}^{-\lambda}}{3!} \frac{\lambda^4 \mathrm{e}^{-\lambda}}{4!} \frac{\lambda^3 \mathrm{e}^{-\lambda}}{3!} \frac{\lambda^7 \mathrm{e}^{-\lambda}}{7!} \\ &= \lambda^{17} \mathrm{e}^{-4\lambda} \Big/ (3!4!3!7!) \end{aligned}$$

Find ekstrempunkter ved at sætte den første aflede lige 0:

$$\begin{aligned} \frac{\mathrm{d}\, L(\lambda \mid \mathbf{x})}{\mathrm{d}\lambda} &= \left(17\lambda^{16} \mathrm{e}^{-4\lambda} + \lambda^{1}7(-4\mathrm{e}^{-4\lambda})\right) \Big/ (3!4!3!7!) \\ &= (17 - 4\lambda)\lambda^{16} \mathrm{e}^{-4\lambda} \Big/ (3!4!3!7!) \end{aligned}$$

# Eksempel 6.1: Poissonfordelingen, fortsat

$$\frac{\mathrm{d}\,L(\lambda\mid\mathbf{x})}{\mathrm{d}\lambda} = (17-4\lambda)\lambda^{16}\mathrm{e}^{-4\lambda}\Big/(3!4!3!7!)$$

altså

$$\left.\frac{\mathrm{d}\,L(\lambda\mid\mathbf{x})}{\mathrm{d}\lambda}\right|_{\lambda=\hat{\lambda}} = 0 \quad \Leftrightarrow \quad \hat{\lambda} = \frac{17}{4}.$$

Egentligt burde kontrolleres, om $\frac{\mathrm{d}^2\,L(\lambda|\mathbf{x})}{\mathrm{d}\lambda^2} < 0$ for $\lambda = \hat{\lambda}$....

**Idé**: fra hele fordelingsfamilien, valg den fordeling, hvor den observerede stikprøve $x_1, \dots, x_n$ har den størreste "sandsynlighed".

**Ingredienser / begreber:**

- Fordelingsfamilie parametriseret med parameter $\theta$; tæthed eller sandsynlighed for $X = x$ er givet ved $f(x;\theta)$.
- Likelihoodfunktion (Definition 6.2 i MSRR)

$$L(\theta \mid x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

  maksimeres af **maksimum likelihood estimatet** $\hat{\theta}_{\text{mle}}$ (tit skriver vi bare $\hat{\theta}$)
- log likelihood

$$\ln L(\theta \mid x_1, \dots, x_n) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

- Score funktion

$$u(\theta) = \frac{\mathrm{d} \ln L}{\mathrm{d}\theta}, \quad \text{ved flere dimensioner:} \quad u(\theta_1, \dots, \theta_r) = \left(\frac{\partial \ln L}{\partial \theta_1}, \dots, \frac{\partial \ln L}{\partial \theta_r}\right)$$

## Eksempel 6.1: Poissonfordelingen, fortsat videre

Likelihood:

$$L(\lambda \mid \mathbf{x}), = \lambda^{17} \mathrm{e}^{-4\lambda} \big/ (3!4!3!7!),$$

log-Likelihood:

$$\ln L(\lambda \mid \mathbf{x}) = 17 \ln \lambda - 4\lambda - \ln(3!4!3!7!)$$

score funktion:

$$u(\lambda) = \frac{\mathrm{d} \ln L(\lambda \mid \mathbf{x})}{\mathrm{d} \lambda} = \frac{17}{\lambda} - 4,$$

score funktionens aflede:

$$u'(\lambda) = -\frac{17}{\lambda^2} < 0.$$

## Eksempel 6.4: uniform fordeling

Betragt en stikprøve $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, \beta]$, med tæthed

$$f(x; \beta) = \begin{cases} 1/\beta, & 0 \le x \le \beta, \\ 0, & \text{ellers.} \end{cases}$$

Likelihoodfunktionen er

$$L(\beta) = \begin{cases} \frac{1}{\beta}\frac{1}{\beta}\cdots\frac{1}{\beta}, & 0 \le X_i \le \beta \text{ for alle } X_i \text{ i stikprøven} \\ 0, & \text{hvis der findes en } i \text{ med } X_i < 0 \text{ eller } X_i > \beta \end{cases}$$

$L(\beta)$ bliver maksimalt, hvis $\beta$ vælges så småt som muligt, således at ingen værdi $X_i$ er større end $\beta$:

$$\hat{\beta}_{\text{mle}} = \max\{X_1, X_2, \dots, X_n\}.$$

☞ Det er et *globalt* maksimum af $L(\beta)$, derfor bruges her *ikke* score-funktionen.

Obs: hvis man definerer uniform fordelingen på en åben interval $(0, \beta)$, findes der ingen maksimum likelihood estimator, se eksempel 6.5.

**Theorem 6.1.3** Let $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ be a random sample from a normal distribution with mean $\mu$ and standard deviation $\sigma$. The maximum likelihood estimates of $\mu$ and $\sigma$ are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \tag{6.3}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{6.4}$$

**Bevis:** Vi går direkte til log-likelihood, ved at bruge

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

$$\implies \ln f(x;\mu,\sigma^2) = -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2}(x-\mu)^2$$

Med $\ln(L(\mu,\sigma^2)) = \sum_{i=1}^{n} \ln f(x;\mu,\sigma^2)$ fås

$$\ln(L(\mu,\sigma)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}(x_i - \mu)^2. \tag{6.5}$$

**Theorem 6.1.3** Let $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ be a random sample from a normal distribution with mean $\mu$ and standard deviation $\sigma$. The maximum likelihood estimates of $\mu$ and $\sigma$ are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \tag{6.3}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{6.4}$$

**Bevis:** Vi går direkte til log-likelihood, ved at bruge

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

$$\Longrightarrow \ln f(x;\mu,\sigma^2) = -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2}(x-\mu)^2$$

Med $\ln(L(\mu,\sigma^2)) = \sum_{i=1}^{n} \ln f(x;\mu,\sigma^2)$ fås

$$\ln(L(\mu,\sigma)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}(x_i - \mu)^2. \tag{6.5}$$

## Eksempel: ML estimation i normalfordelingen, fortsat

Setting the partial derivatives of the log-likelihood with respect to $\mu$ and $\sigma$ equal to 0 gives a system of equations:

$$\frac{\partial(\ln(L(\mu,\sigma)))}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0, \tag{6.6}$$

$$\frac{\partial(\ln(L(\mu,\sigma)))}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0. \tag{6.7}$$

Fra (6.7) fås ML estimater

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

og, ved at indsætte dette i (6.8),

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Vindhastigheden modelleres i litteraturen ved en Weibull fordeling:

$$f(x;k,\lambda) = \frac{kx^{k-1}}{\lambda^k} e^{-(x/\lambda)^k}, \quad x > 0.$$

En Weibull fordelte stokastisk variabel kan fås ved en transformation af $Z \sim \mathrm{Exp}(1)$:

$$X = \frac{1}{\lambda} Z^{1/k}.$$

Med

$$\ln f(x;k,\lambda) = \ln k - k \ln \lambda + (k-1) \ln x - \left(\frac{x}{\lambda}\right)^k$$

er log-Likelihood:

$$\ln L(k,\lambda) = n \ln k - nk \ln \lambda + (k-1) \sum_{i=1}^{n} \ln x_- \sum_{i=1}^{n} \left(\frac{x}{\lambda}\right)^k.$$

Thus, the log-likelihood is

$$\ln(L(k,\lambda)) = n\ln(k) - kn\ \ln(\lambda) + (k-1)\sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k.$$

We next compute the partial derivatives of $\ln(L(k,\lambda))$ with respect to $k$ and $\lambda$ and set them equal to 0:

$$\frac{\partial(\ln(L(k,\lambda)))}{\partial k} = \frac{n}{k} - n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k \ln\left(\frac{x_i}{\lambda}\right) = 0, \qquad (6.9)$$

$$\frac{\partial(\ln(L(k,\lambda)))}{\partial \lambda} = \frac{-kn}{\lambda} + \frac{k}{\lambda^{k+1}}\sum_{i=1}^{n}x_i^k = 0. \qquad (6.10)$$

From Equation (6.10), we find

$$\lambda^k = \frac{1}{n}\sum_{i=1}^{n}x_i^k, \qquad (6.11)$$

and substituting this into Equation (6.9), we obtain

$$\frac{1}{k} + \frac{1}{n}\sum_{i=1}^{n}\ln(x_i) - \frac{1}{\alpha}\sum_{i=1}^{n}x_i^k\ln(x_i) = 0, \qquad (6.12)$$

where $\alpha = \sum_{i=1}^{n} x_i^k$.

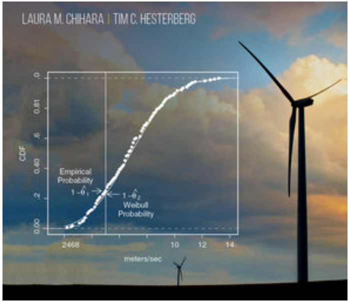
Ligningen løses numerisk.

Autorens (Chiharas) data (udsnit):

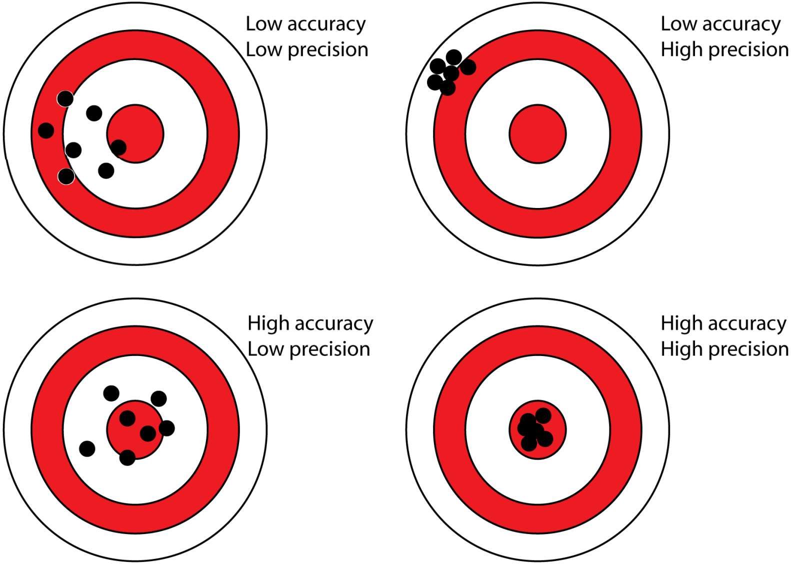
**Table 6.1** Sample of wind speeds (m/s) from Carleton College turbine.

| Feb 14 | Feb 15 | Feb 16 | Feb 17 | Feb 18 | Feb 19 |
|---|---|---|---|---|---|
| 7.8 | 8.9 | 9.7 | 7.7 | 6.4 | 3.1 |

Fordelingsfunktion, fittet til data:



# Præcision og nøjagtighed



 (se bogen: 6.3.1)

### Definition

**Bias** af en estimator $\hat{\theta}$ til parameteren $\theta$ er givet ved

$$\mathrm{Bias}\,\hat{\theta} = \mathrm{E}\,\hat{\theta} - \theta.$$

Estimatoren $\hat{\theta}$ kaldes for **middelret** eller **central** (unbiased) hviss $\mathrm{E}\,\hat{\theta} = \theta$, dvs, $\mathrm{Bias}\,\hat{\theta} = 0$.

 (se bogen: side 164)

Betragt en Bernoulli eksperiment, bestående af $n$ uafhængige stokastiske variabler $X_1, \ldots, X_n$, der antager værdien 1 med sandsynlighed $p$, og er lig 0 ellers.

Andelen $\hat{p}$ af positive udfald i $n$ uafængige Bernoulli eksperimenter,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

er en estimator til andelen $p$ i populationen.

**Proposition 6.3.1** If $X_1, X_2, \ldots, X_n$ are Bernoulli random variables with parameter $p$, then $\mathrm{E}[\hat{p}] = p$.

*Proof*. Let $X = \sum_{i=1}^{n} X_i$. Then

$$\mathrm{E}[\hat{p}] = \mathrm{E}\left[\frac{X}{n}\right] = \frac{np}{n} = p.$$

□

**Theorem 6.3.2** Let $X_1, X_2, \ldots, X_n$ be independent random variables from a distribution with unknown $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then an unbiased estimator of $\sigma^2$ is $S^2 = (1/(n-1)) \sum_{i=1}^{n} (X_i - \overline{X})^2$.

Værktøj til beviset: (se side 164 MSRR)

- **Proposition A.2.1** $\mathrm{Var}[X] = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2$.
- **Theorem A.4.1** Let $X_1, X_2, \ldots, X_n$ be identically distributed random variables with mean $\mu$ and variance $\sigma^2$.
  Then $\mathrm{E}[\overline{X}] = \mu$.
  If, in addition, $X_1, X_2, \ldots, X_n$ are independent, then $\mathrm{Var}[\overline{X}] = \sigma^2/n$.
- samt identiteten

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \left( \sum_{i=1}^{n} X_i^2 \right) - n\overline{X}^2. \tag{6.16}$$

(se også side 164 MSRR)

## Stikprøvevarians

(se bogen: Def. 6.3, s. 165)

### Definition

Lad $X_1, X_2, \dots X_n$ være uafhængige, identisk fordelte stokastiske variabler fra en fordeling med varians $\sigma^2 < \infty$. Stikprøvefunktionen

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X)^2$$

kaldes **stikprøvevarians** (sample variance).

Stikprøvevariansen er en middelret estimator af $\sigma^2$.

**Example 6.12** Let $X_1, X_2, \ldots, X_n$ be i.i.d. from Unif$[0, \beta]$. We have already seen that the MLE of $\beta$ is $\hat{\beta}_{\text{mle}} = X_{\text{max}}$.

With $f_{\text{max}}(x) = (n/\beta^n)x^{n-1}$ (see Corollary 4.2.2), we have

$$\begin{aligned} \mathrm{E}[X_{\text{max}}] &= \int_0^\beta x \, \frac{n}{\beta^n} x^{n-1} \, \mathrm{d}x \\ &= \frac{n}{n+1}\beta, \end{aligned}$$

derfor er $\hat{\beta}_{\text{mle}} = X_{\text{max}}$ ikke middelret , men

$$\hat{\beta} = \frac{n+1}{n} X_{\text{max}}$$

er middelret.

En anden middelret estimator er den fra moment metoden,

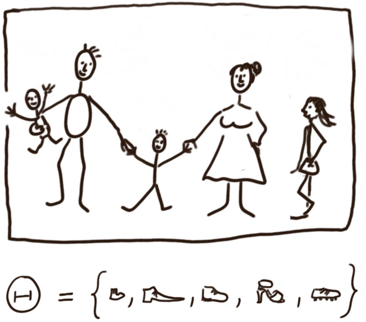
$$\hat{\beta}_{\text{mom}} = 2\bar{X}.$$

# Estimation af transformerede parametre

(se bogen: 6.3.5, Prop. 6.3.5)

Betragt en fordelingsfamilie med parameter $\theta$.

I nogle situationer udtrykkes samme familien vha en anden parameter $\rho = h(\theta)$, hvor $h$ er en **bijektiv** funktion.



$$h(\Theta) = \{23, 45, 33, 39, 36\}$$

Eksempel (R):

```
rgamma(n, shape, rate = 1, scale = 1/rate).
```

$$\hat{\theta} \text{ er } \begin{cases} \text{ML-estimator} \\ \text{moment-estimator} \end{cases} \text{af } \theta \quad \Longrightarrow \quad \hat{\rho} = h(\hat{\theta}) \text{ er } \begin{cases} \text{ML-estimator} \\ \text{moment-estimator} \end{cases} \text{af } \rho$$

**Normalfordelingen** kan parametriseres ved $(\mu, \sigma)$ frem for $(\mu, \sigma^2)$ — se den måde R gør det på, fx `rnorm(n, mean =` $\mu$`, sd =` $\sigma$`)`.

ML estimatorer til $\sigma^2$ og $\sigma$ er givet ved

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \Longrightarrow \quad \hat{\sigma} = \sqrt{\widehat{\sigma^2}} = \left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{1/2}$$

**Eksponentialfordelingen** parametriseres nogle gange med skala $\beta = 1/\lambda$ frem for rate $\lambda$. Så skrives tætheden

$$f(x; \lambda) = \lambda e^{-\lambda x} \quad \text{som} \quad \tilde{f}(x; \beta) = \frac{1}{\beta} e^{-x/\beta}.$$

ML estimator til $\lambda$ og $\beta = 1/\lambda$:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} \quad \Longrightarrow \quad \hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Antag, at $\hat{\theta}$ er en god estimator for $\theta$. Vi vil dog gerne estimere $g(\theta)$ for en given funktion $g$. Er $g(\hat{\theta})$ også en god estimator?

Svar:

- Hvis $\hat{\theta}$ er ML-estimator for $\theta$, så er $g(\hat{\theta})$ også ML estimator for $g(\theta)$.
- Hvis $\hat{\theta}$ er en middelret estimator for $\theta$, så er $g(\hat{\theta})$ generelt set *ikke* middelret for $g(\theta)$.

**Jensens ulighed**[a]

Lad $g$ være en streng konveks funktion og $X$ en ikke degenereret stokastisk variabel[b]. Så er

$$g(\mathrm{E}\,X) < \mathrm{E}\,g(X).$$

[a] Johan Jensen, 1908

[b] en stok. variabel kaldes degenereret, hvis den kun antager 1 værdi

 (se bogen: side 151)

Jensens ulighed betyder for en unbiased estimator $\hat{\theta}$, hvis $g$ er streng konveks:

$$\mathrm{E}\, g(\hat{\theta}) > g(\theta).$$

Eksempel (MSRR, side 166): For Unif$[0, \beta]$ fordelingen er $\hat{\beta} = 2\bar{X}$ middelret. Men $(\hat{\beta})^2 = 4\bar{X}^2$ er ikke middelret for $\beta^2$:

$$\begin{aligned}
\mathrm{E}[4\bar{X}^2] &= 4\,\mathrm{E}[\bar{X}^2] = 4(\mathrm{Var}[\bar{X}] + \mathrm{E}[\bar{X}]^2) \\
&= 4\left(\frac{\beta^2}{12n} + \left(\frac{\beta}{2}\right)^2\right) = \beta^2 + \frac{\beta^2}{3n}
\end{aligned}$$

Altså $\mathrm{E}[\hat{\beta}^2] > \beta^2$.

## Definition

En estimator $\hat{\theta}$ kaldes **asymptotisk middelret**, hvis

$$\lim_{n\to\infty} \mathrm{E}\hat{\theta}(X_1,\ldots,X_n) = \theta,$$

for uafhængige $\mathbf{X}_1, \mathbf{X}_2, \cdots \sim F(.,\theta)$.

## Eksempler

- ML estimatoren til $N(\mu,\sigma^2)$ normalfordelingens varians er asymptotisk middelret:

$$\mathrm{E}\hat{\sigma}^2 = \frac{n-1}{n}\sigma^2 \xrightarrow[n\to\infty]{} \sigma^2.$$

- ML estimator for parameter $\beta$ i Unif$[0,\beta]$-fordelingen er $X_{\max}$. Den er asymptotisk middelret:

$$\lim_{n\to\infty} \mathrm{E}[X_{\max}] = \lim_{n\to\infty} \frac{n}{n+1}\beta = \beta$$