ISIPTA'09 July 14th, 2009

# A Minimum Distance Estimator in an Imprecise Probability Model

Robert Hable
Department of Mathematics
University of Bayreuth

ISIPTA'09 July 14th, 2009

# A Minimum Distance Estimator in an Imprecise Probability Model

## Computational Aspects and Applications

Robert Hable
Department of Mathematics
University of Bayreuth

# Robert Hable

Department of Mathematics, University of Bayreuth (Germany)

- Diploma Thesis in Robust Asymptotic Statistics (Helmut Rieder, University of Bayreuth)
- Ph.D. Thesis: "Data-Based Decisions under Complex Uncertainty" (Thomas Augustin, LMU Munich)
- Now: University of Bayreuth
  Working group: Andreas Christmann (Machine Learning) and Helmut Rieder (Robust Statistics)
- Research Interests:
  - Mathematical Foundations of Imprecise Probabilities
  - (Statistical) Decision Theory under Imprecise Probabilities
  - Robust Statistics
  - Nonparametric Statistics

# Estimation: Classical Statistics and Generalizations

Data $x_1, \ldots, x_n$ are modeled by random variables

$$X_1\,, \ldots\,, X_n \;\sim\; P_0 \qquad\qquad \text{i.i.d.}$$

$P_0$ the true distribution (precise probability measure)

- classical statistics
  - $(P_\theta)_{\theta\in\Theta}$ : a parametric model (precise distributions)
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$

# Estimation: Classical Statistics and Generalizations

Data $x_1, \ldots, x_n$ are modeled by random variables

$$X_1\,, \ldots\,, X_n \;\sim\; P_0 \qquad\qquad \text{i.i.d.}$$

$P_0$ the true distribution (precise probability measure)

- classical statistics
  - $(P_\theta)_{\theta \in \Theta}$ : a parametric model (precise distributions)
  - $\exists\, \theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$
- robust statistics
  - $(P_\theta)_{\theta \in \Theta}$ : a parametric model (precise distributions)
  - $\mathcal{U}_\theta$ : neighborhood about $P_\theta \;\; \forall\, \theta \in \Theta$
  - $\exists\, \theta_0 \in \Theta$ such that $P_0 \in \mathcal{U}_{\theta_0}$

# Estimation: Classical Statistics and Generalizations

Data $x_1, \ldots, x_n$ are modeled by random variables

$$X_1\,, \ldots\,, X_n \;\sim\; P_0 \qquad \text{i.i.d.}$$

$P_0$ the true distribution (precise probability measure)

- classical statistics
  - $(P_\theta)_{\theta\in\Theta}$ : a parametric model (precise distributions)
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$
- robust statistics
  - $(P_\theta)_{\theta\in\Theta}$ : a parametric model (precise distributions)
  - $\mathcal{U}_\theta$ : neighborhood about $P_\theta$ $\forall\,\theta \in \Theta$
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 \in \mathcal{U}_{\theta_0}$
- an imprecise probability model
  - $\mathcal{M}_\theta$ credal set (coherent upper prevision $\overline{P}_\theta$) $\forall\,\theta \in \Theta$
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 \in \mathcal{M}_{\theta_0}$

# Estimation: Classical Statistics and Generalizations

Data $x_1, \ldots, x_n$ are modeled by random variables

$$X_1\,,\,\ldots\,,\,X_n \quad \sim \quad P_0 \qquad\qquad \text{i.i.d.}$$

$P_0$ the true distribution (precise probability measure)

- classical statistics
  - $(P_\theta)_{\theta\in\Theta}$ : a parametric model (precise distributions)
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$
- robust statistics
  - $(P_\theta)_{\theta\in\Theta}$ : a parametric model (precise distributions)
  - $\mathcal{U}_\theta$ : neighborhood about $P_\theta \;\; \forall\,\theta\in\Theta$
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 \in \mathcal{U}_{\theta_0}$
- an imprecise probability model
  - $\mathcal{M}_\theta$ credal set (coherent upper prevision $\overline{P}_\theta$) $\forall\,\theta\in\Theta$
  - $\exists\,\theta_0 \in \Theta$ such that $P_0 \in \mathcal{M}_{\theta_0}$

**Goal:** Estimate $\theta_0$

# An Imprecise Probability Model – Some More Assumptions

# An Imprecise Probability Model – Some More Assumptions

**. . . may be found on the poster!**

# Minimum Distance Estimator

- Simple **idea**:
  - Construct empirical measure $\mathbb{P}_n$ from the data
  - Chose that credal set $\mathcal{M}_\theta$ which lies next to $\mathbb{P}_n$

# Minimum Distance Estimator

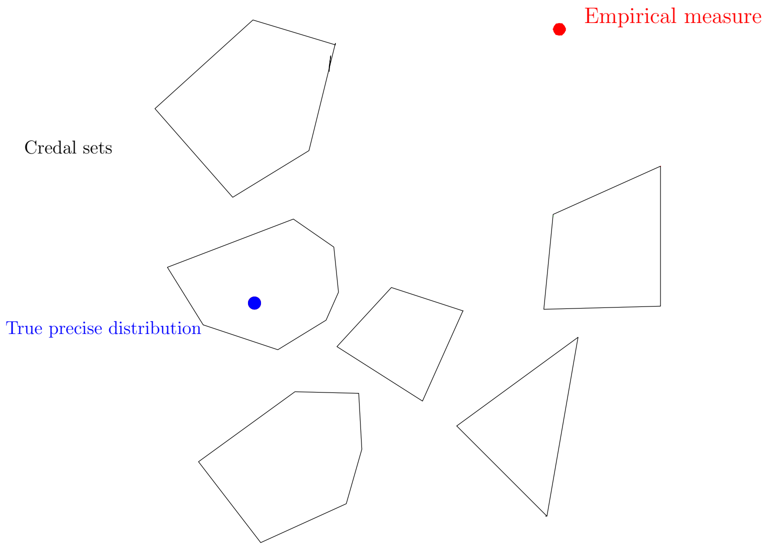
- Simple **idea**:
  - Construct empirical measure $\mathbb{P}_n$ from the data
  - Chose that credal set $\mathcal{M}_\theta$ which lies next to $\mathbb{P}_n$

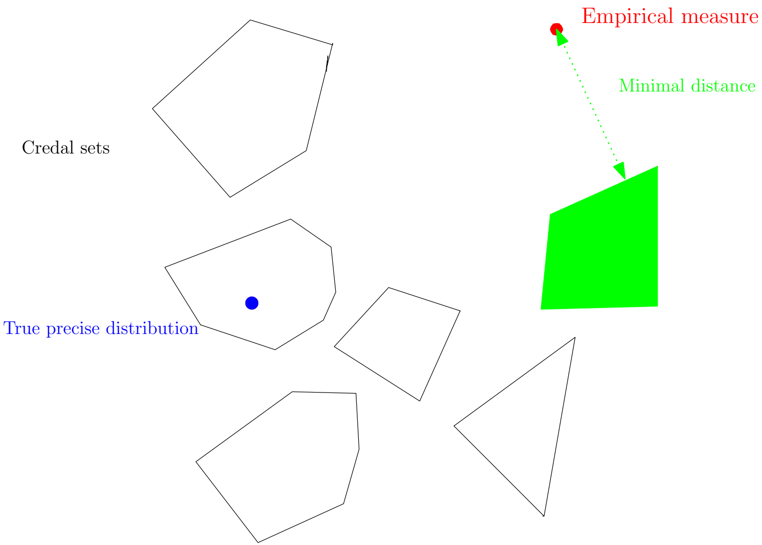
- **Definition**:

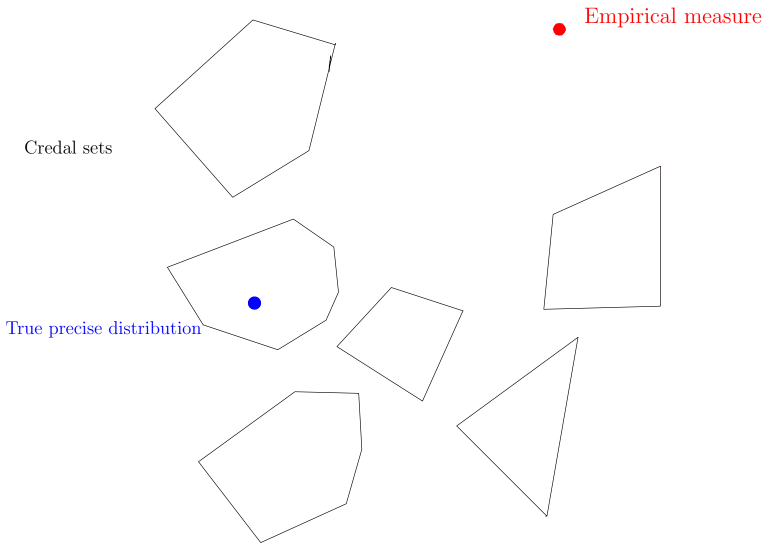
  Estimator $\hat{\theta}_n$ minimizes

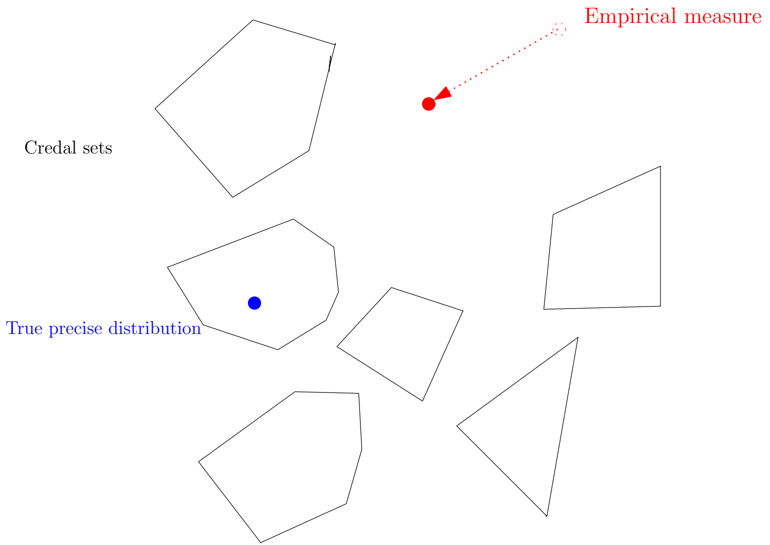
$$\theta \;\mapsto\; d\big(\mathbb{P}_n, \mathcal{M}_\theta\big) \;=\; \inf_{P_\theta \in \mathcal{M}_\theta} d\big(\mathbb{P}_n, P_\theta\big)$$

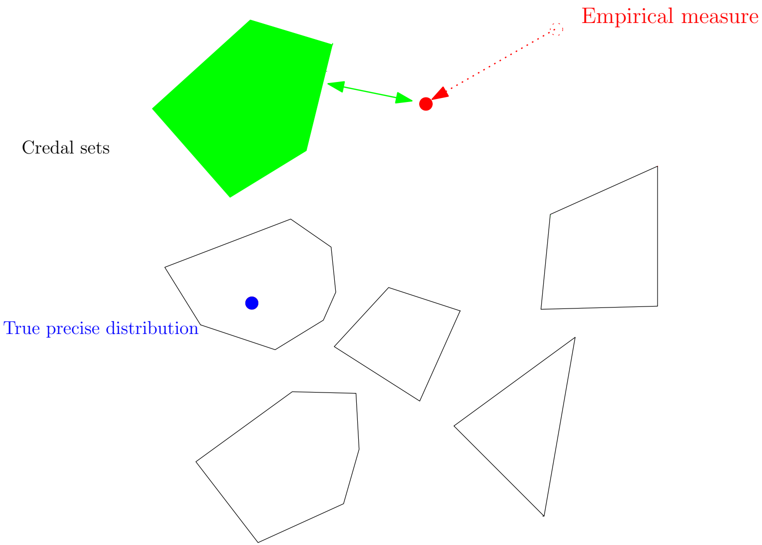
  Distance $d$ : total variation

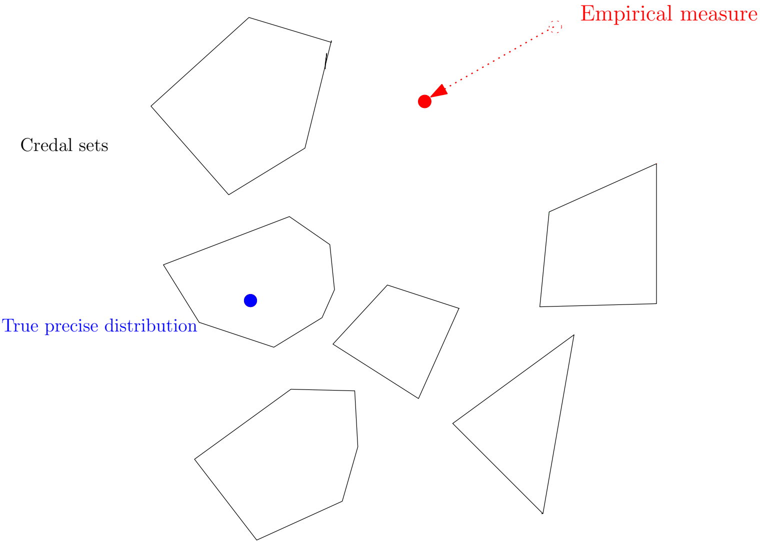

Empirical measure
Credal sets
True precise distribution

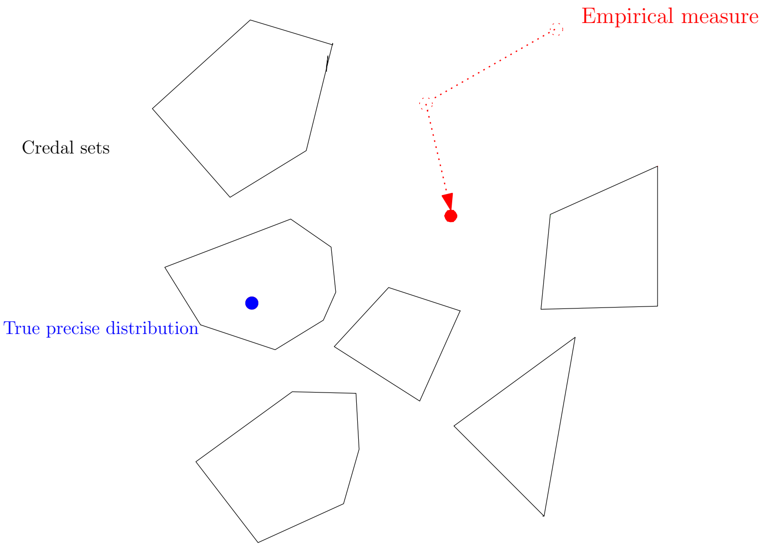

Empirical measure
Minimal distance
Credal sets
True precise distribution

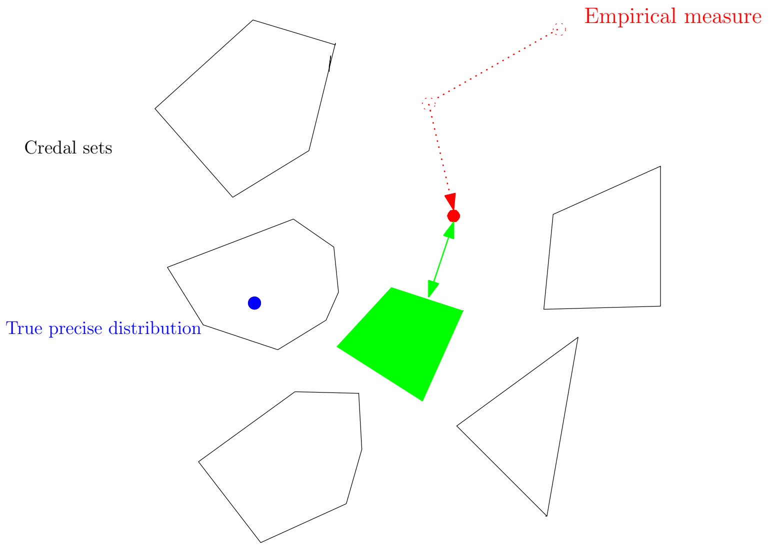

Empirical measure
Credal sets
True precise distribution

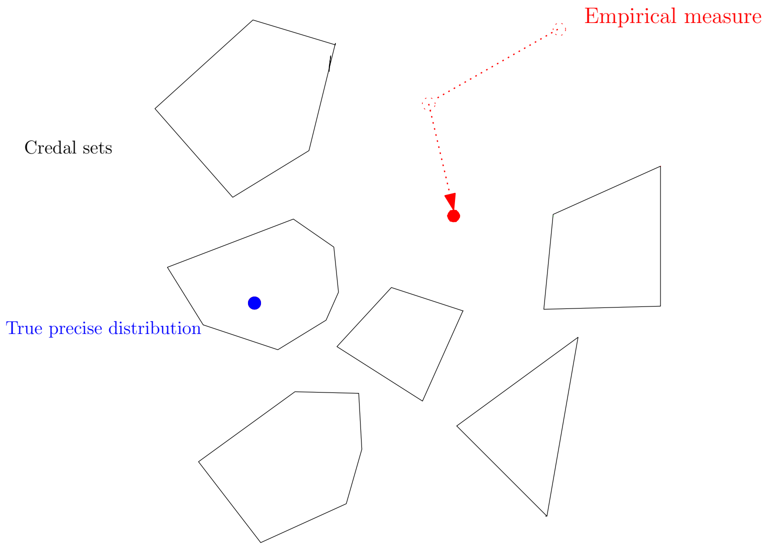

Empirical measure
Credal sets
True precise distribution

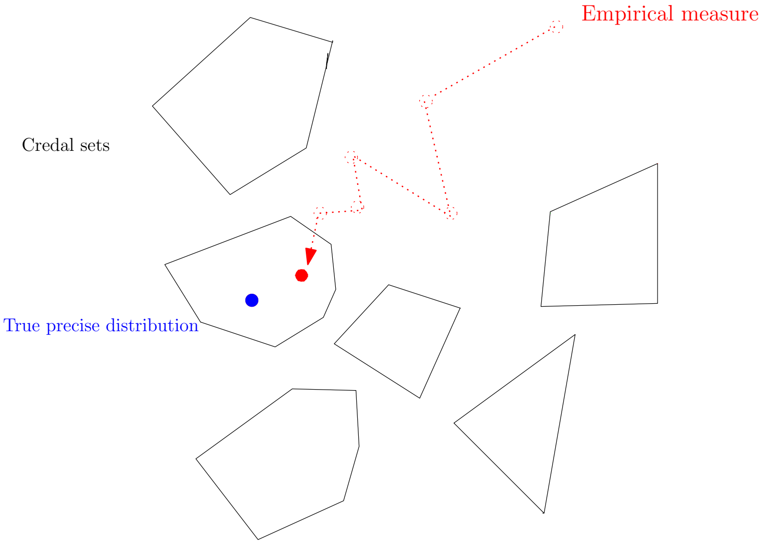

Empirical measure
Credal sets
True precise distribution

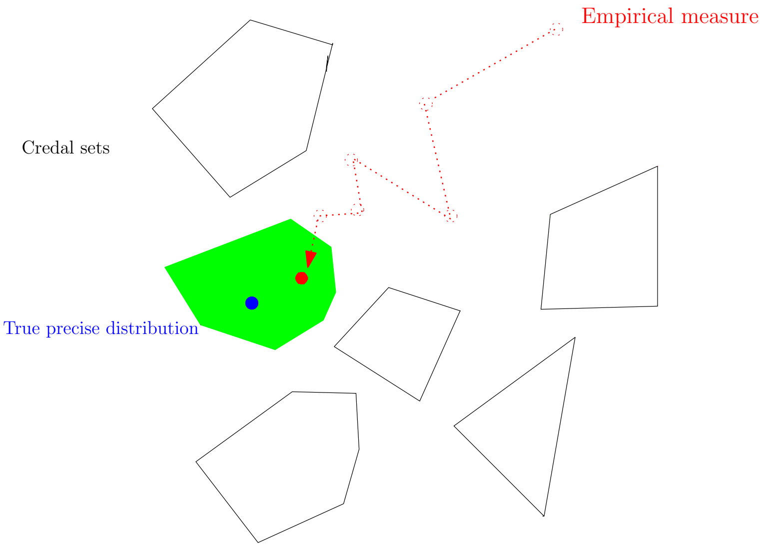

Empirical measure
Credal sets
True precise distribution


Empirical measure
Credal sets
True precise distribution


Empirical measure
Credal sets
True precise distribution


Empirical measure
Credal sets
True precise distribution


Empirical measure
Credal sets
True precise distribution


Empirical measure
Credal sets
True precise distribution

# Calculations

- Replace original sample space $(\mathcal{X}, \mathcal{B})$ by a suitable discrete sample space $(\mathcal{X}, \mathcal{C})$
- Approximate calculation of $d\left(\mathbb{P}_n, \mathcal{M}_\theta\right)$ by linear programming

# Calculations

- Replace original sample space $(\mathcal{X}, \mathcal{B})$ by a suitable discrete sample space $(\mathcal{X}, \mathcal{C})$
- Approximate calculation of $d\big(\mathbb{P}_n, \mathcal{M}_\theta\big)$ by linear programming:

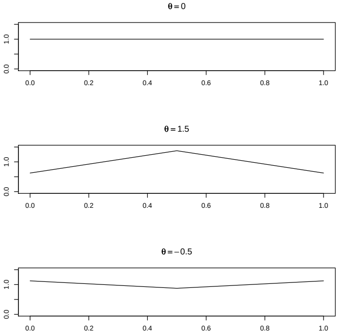
$$\sum_{j \in \mathcal{J}_1} q_j - \gamma_j \quad \longrightarrow \quad \text{max!}$$

$$\begin{aligned}
&\textstyle\sum_{j=1}^{r} q_j = 1 \\
&\textstyle\sum_{j=1}^{r} q_j h_k(c_j) \leq \overline{P}_\theta[f_k] + \varepsilon_\theta^{(k)} \ \forall\, k \in \{1, \dots, s\} \\
&q_j - \gamma_j \ \leq \ \tfrac{n_j}{n} \ \forall\, j \in \mathcal{J}_1 \\
&(q_1, \dots, q_r) \ \in \ \mathbb{R}^r, \ q_j \geq 0 \quad \forall\, j \in \{1, \dots, r\}, \\
&(\gamma_j)_{j \in \mathcal{J}_1} \ \subset \ \mathbb{R}, \ \gamma_j \geq 0 \quad \forall\, j \in \mathcal{J}_1
\end{aligned}$$

- Implemented as R-Package `imprProbEst`; see Hable (2008).

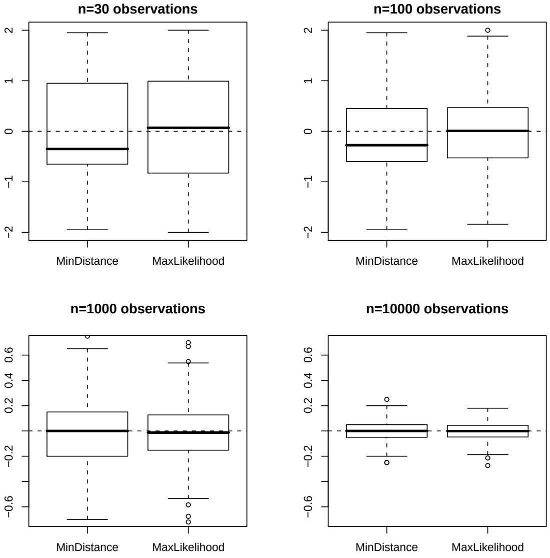
# Simulation Study: Model 1

Data from an ideal, precise model with densities such as




- **Maximum likelihood estimator:** complete information about precise model
- **Minimum distance estimator:** only imprecise information about precise model

# Simulation Study: Model 1



# Simulation Study: Model 2

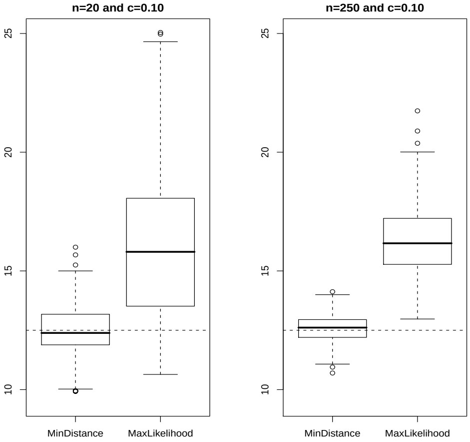
Model: Approximate Poisson Distribution

- **90%** of the data (in average) from **Poisson** distribution
- **10%** of the data (in average) from **uniform** distribution

# Simulation Study: Model 2

Model: Approximate Poisson Distribution

- **90%** of the data (in average) from **Poisson** distribution
- **10%** of the data (in average) from **uniform** distribution

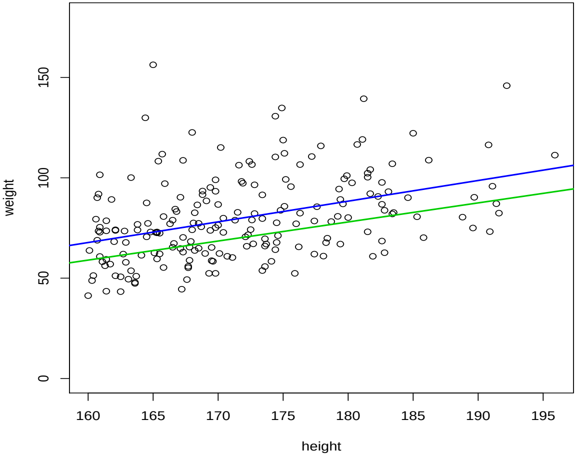


# Real Data Set: Linear Regression

NHANES: persons' body weight depending on persons' height

# Real Data Set: Linear Regression

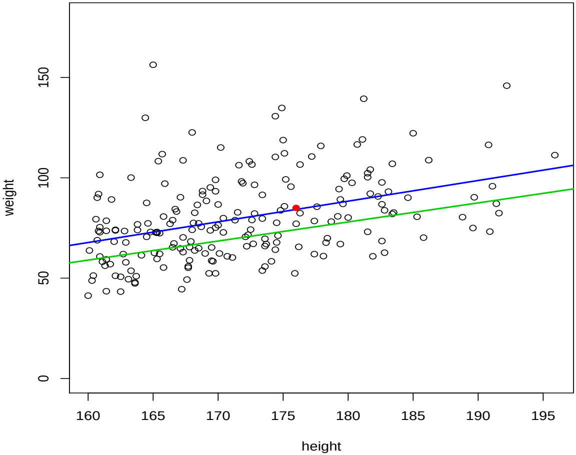
NHANES: persons' body weight depending on persons' height




Minimum Distance Estimator Least Squares Estimator

# Real Data Set: Linear Regression

NHANES: persons' body weight depending on persons' height




Minimum Distance Estimator     Least Squares Estimator

# References


- Hable, R. (2008). *imprProbEst: Minimum distance estimation in an imprecise probability model*. Contributed R-Package on CRAN, Version 1.0, 2008-10-23; maintainer Hable, R.
- Hable, R. (2009a). Data-based decisions under complex uncertainty. *Ph.D. Thesis*. `http://edoc.ub.uni-muenchen.de/9874/`
- Hable, R. (2009b). A minimum distance estimator in an imprecise probability model – computational aspects and applications. *6th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'09)*.


The handout to this talk is also available on my homepage

**http://www.staff.uni-bayreuth.de/˜btms04/index.html**