Robust Inference via the Blended Paradigm

John Lewis, Steven MacEachern, Yoonkyung Lee

The Ohio State University With Support of the Nationwide Center for Advanced Customer Insights

Sunday, July 29, 2012

Introduction to the Blended Paradigm

- Traditional Bayesian Framework:
  - $X \sim f(X|\theta)$ and $\theta \sim \pi(\theta)$
  - Pass to the posterior: $\pi(\theta|X) \propto f(X|\theta)\pi(\theta)$
- Deficiencies:
  - We may not fully believe the likelihood $f(X|\theta)$
  - Asymptotics do not fix this problem
  - More modeling may not fix this problem
- Acknowledge inadequacy of our model
- Drive Bayesian update with the 'good' portion of the data

Introduction to the Blended Paradigm

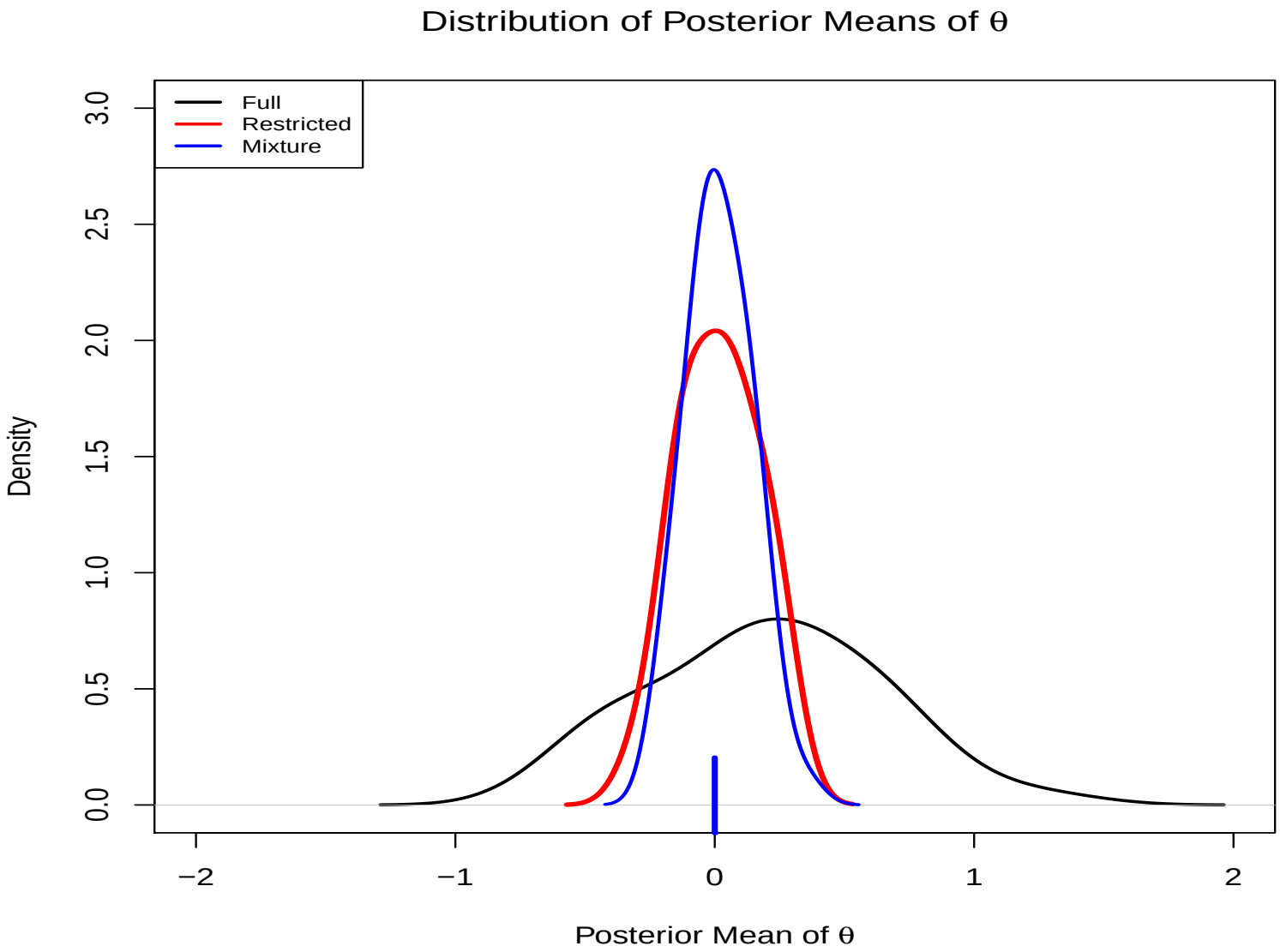
- Blended Paradigm Framework:
  - Here, $T(X)$ is a 'robust' summary of the data.
    - 'Robust' means insensitivity to the deficiencies in the model for inferences of interest
    - Model for $f(X|\theta)$ implies model for $f(T(X)|\theta)$
  - $T(X) \sim f(T(X)|\theta)$ and $\theta \sim \pi(\theta)$
  - Pass to the posterior: $\pi(\theta|T(X)) \propto f(T(X)|\theta)\pi(\theta)$
  - Implementation via MCMC
- Goal : To obtain better inference through a wise choice of $T(X)$

Proof of Concept-Outliers

- Simulation study
  - True model: $X_i \sim (1-p)N(\boldsymbol{\theta}, \sigma^2) + pN(\boldsymbol{\theta}, 100\sigma^2), \quad i = 1, 2, \cdots, 100$
  - $(\boldsymbol{\theta}, \sigma^2) = (0, 1), \quad p = 0.2$
  - We regard the large variance component as outliers
  - Alternately, a thick tailed distribution
- Model for Analysis:
  - Standard normal theory model
  - $X_i \sim N(\boldsymbol{\theta}, \sigma^2)$
  - $\boldsymbol{\theta} \sim N(5, 4)$, $\sigma^2 \sim IG(5, 5)$
  - Note that the prior is off-center from the truth

- Compare two likelihoods
  - Full Likelihood (Traditional Bayes): $T(X) = (X_1, \cdots, X_{100})$
  - Restricted Likelihood (Order Statistics): $T(X) = (X_{(31)}, X_{(32)}, \cdots, X_{(70)})$
- The 'True' mixture likelihood (p known) is fit as a reference
- Details of the simulation
  - Sample the posteriors via MH
  - Replicate the simulation 50 times
  - For each replicate, obtain estimates of the posterior means under each model

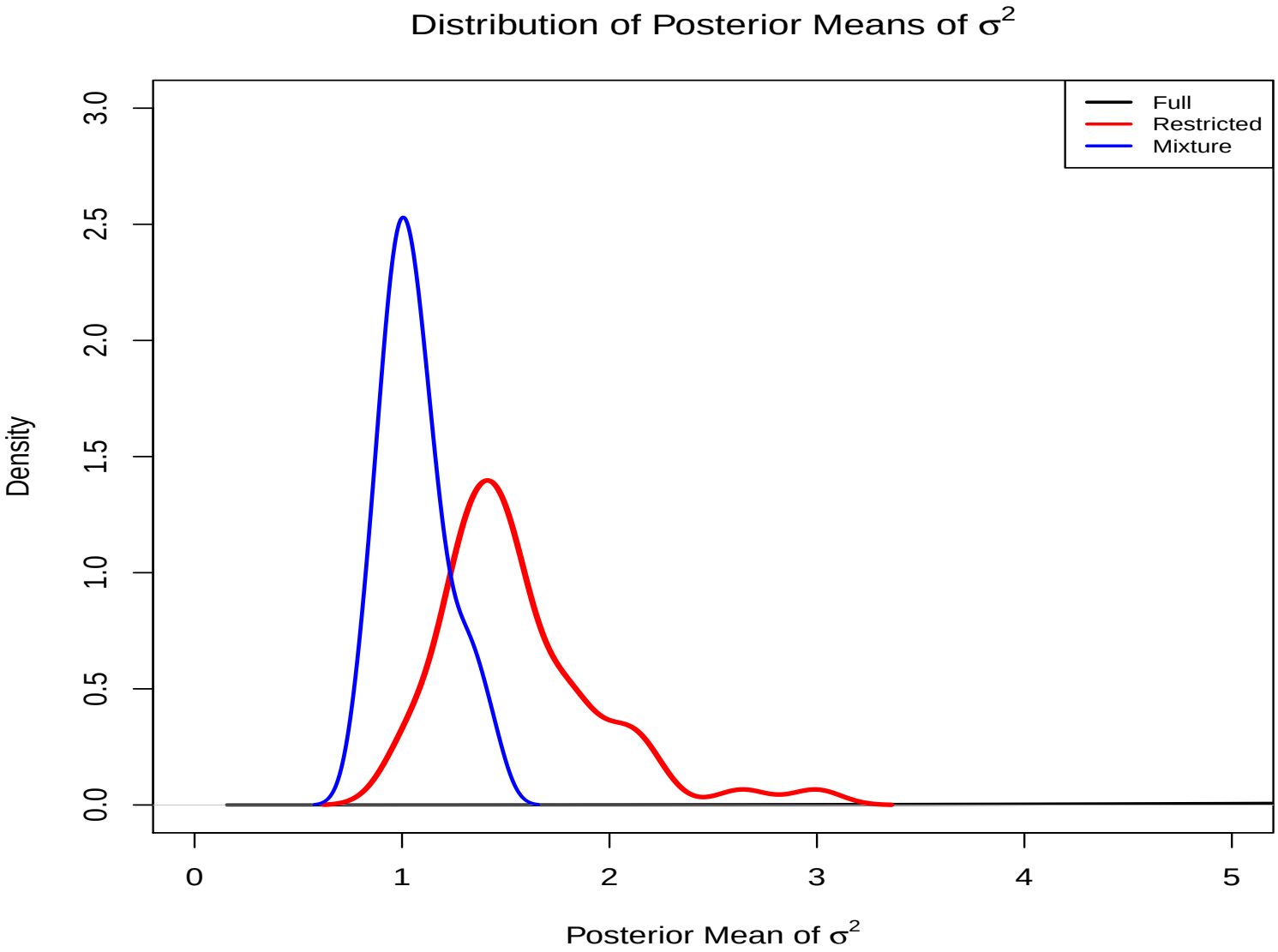
Simulation Results



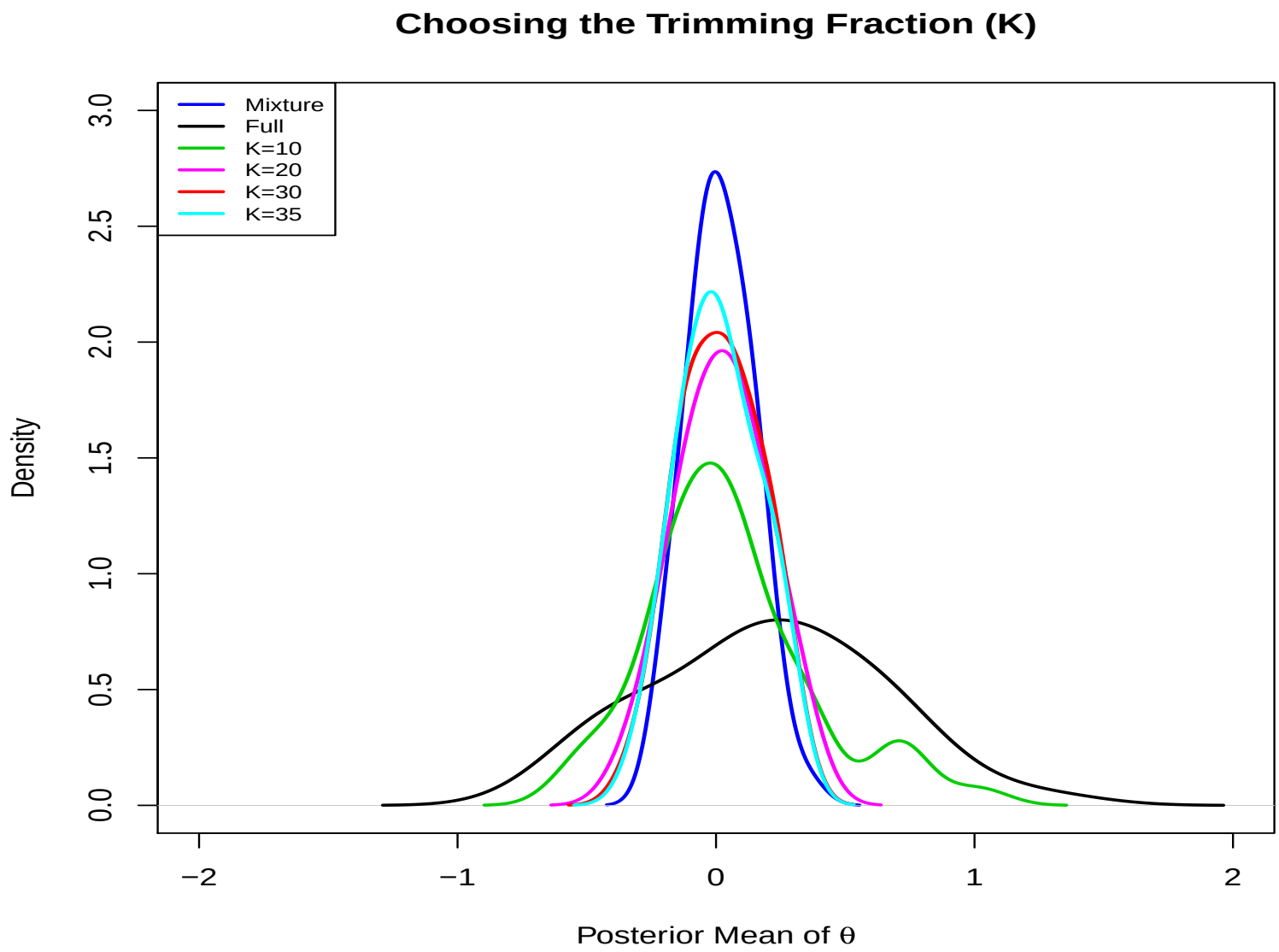
Simulation Results

- Using the restricted likelihood we have:
  - Smaller bias
  - Smaller variance
- Related to the nuisance parameter σ^2
  - Under the mixture likelihood: $\sigma^2 = 1$
  - Under the full likelihood: $\sigma^2 = 20.8$

Simulation Results



Choosing the Timming Fraction (K)



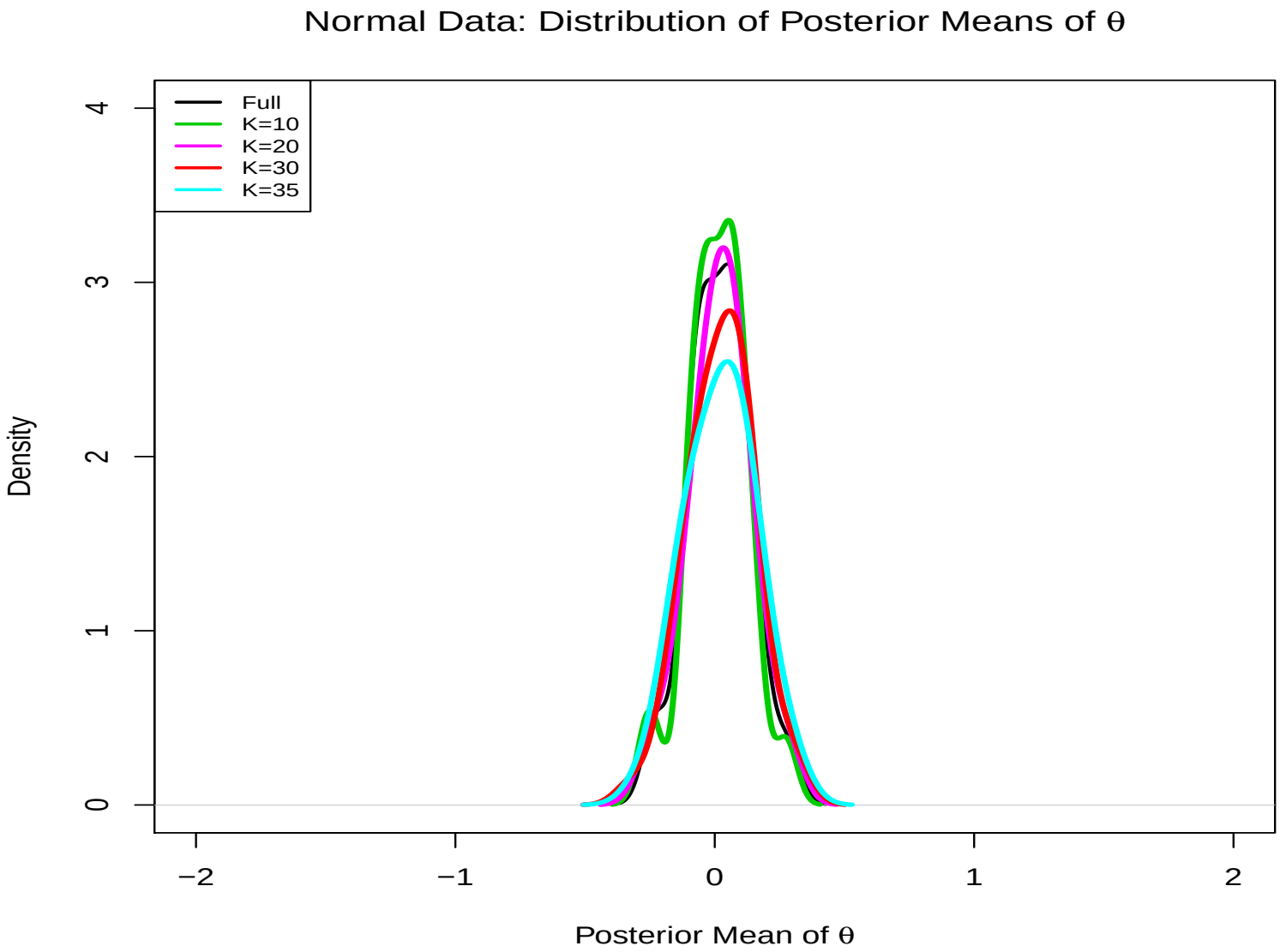
Choosing the Timming Fraction (K)

- Choosing the trimming fraction matters – a little
  - 20% outliers on average
  - Should trim more than 10% from each tail
  - Trimming less moves towards traditional Bayes
  - Trimming more moves towards $T(X) = \text{Median}(X)$ (not the true mixture likelihood)

What is Sacrificed?

- Simulation Study:
  - True Model: $X_i \sim N(\theta, \sigma^2)$
  - $(\theta,\ \sigma^2) = (0, 1)$
- Model for Analysis:
  - $X_i \sim N(\theta, \sigma^2)$
  - $\theta \sim N(5, 4)$, $\sigma^2 \sim IG(5, 5)$
  - Use both the full and restricted likelihoods

What is Sacrificed?



Conclusions

- Blended Paradigm offers a method of inference when we are unsure of our likelihood
- Acknowledges the inadequacy of our model
- Drive the Bayesian update with the 'good' portion of the likelihood
- Simulations show the potential for reducing bias and variance
- This talk focused on'trimming' summaries (i.e. order statistics)
- The idea extends to other classical robust statistics