A Statistical Decision Theoretical Perspective on Two-Stage Approach to Parameter Estimation

Braghadeesh Lakshminarayanan and Cristian R. Rojas

Division of Decision and Control Systems,
KTH Royal Institute of Technology,
Stockholm, Sweden

61st IEEE CDC, Cancun
December 8, 2022

Outline

Outline

Introduction

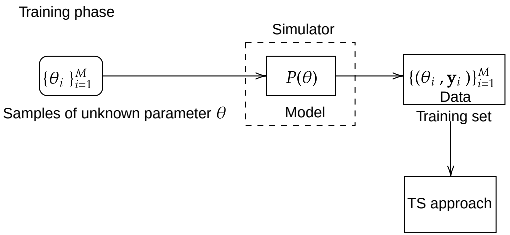
- Parameter estimation: approximating the unknown parameters of a mathematical model describing a real phenomenon

Introduction

- Parameter estimation: approximating the unknown parameters of a mathematical model describing a real phenomenon
- Two-Stage (TS) approach: one of the methods of estimating unknown parameters

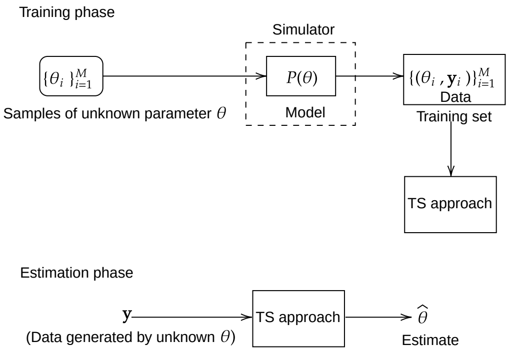
Introduction

- Parameter estimation: approximating the unknown parameters of a mathematical model describing a real phenomenon
- Two-Stage (TS) approach: one of the methods of estimating unknown parameters
  - An inverse supervised learning setup



Introduction

- Parameter estimation: approximating the unknown parameters of a mathematical model describing a real phenomenon
- Two-Stage (TS) approach: one of the methods of estimating unknown parameters
  - An inverse supervised learning setup



Introduction

- Advantages of TS:

Introduction

- Advantages of TS:
  - No explicit likelihood computation
  - Final estimation problem can be a simple convex optimization program

Introduction

- Advantages of TS:
  - No explicit likelihood computation
  - Final estimation problem can be a simple convex optimization program
- However, there is a lack of theoretical justification for TS approach

Introduction

- Advantages of TS:
  - No explicit likelihood computation
  - Final estimation problem can be a simple convex optimization program
- However, there is a lack of theoretical justification for TS approach

Our Contribution

Constructed statistical framework to justify the working principle of TS approach

Introduction

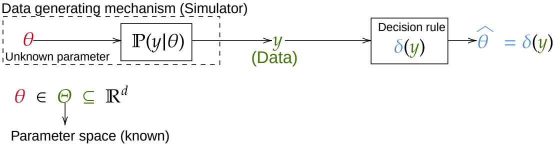
- Advantages of TS:
  - No explicit likelihood computation
  - Final estimation problem can be a simple convex optimization program
- However, there is a lack of theoretical justification for TS approach

Our Contribution

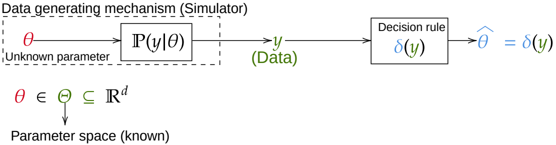
Constructed statistical framework to justify the working principle of TS approach

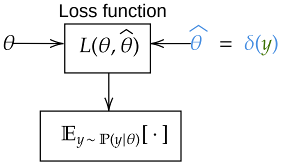
- Assumption: Data generation is an independent and identically distributed (i.i.d.) process.

Problem Statement



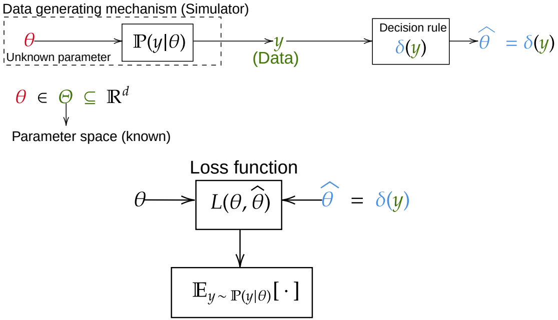
Problem Statement





Risk: $R(\theta, \delta) := \mathbb{E}_{y \sim \mathbb{P}(y|\theta)}[L(\theta, \widehat{\theta})]$

Problem Statement

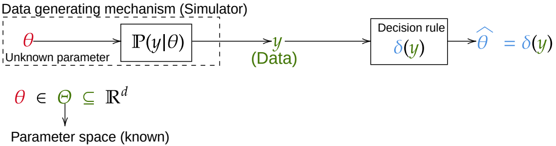


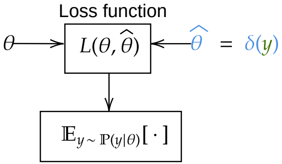
Risk: $R(\theta, \delta) := \mathbb{E}_{y \sim \mathbb{P}(y|\theta)}[L(\theta, \widehat{\theta})]$

Goal

Design optimal decision rules that minimize $R(\theta, \delta)$

Problem Statement





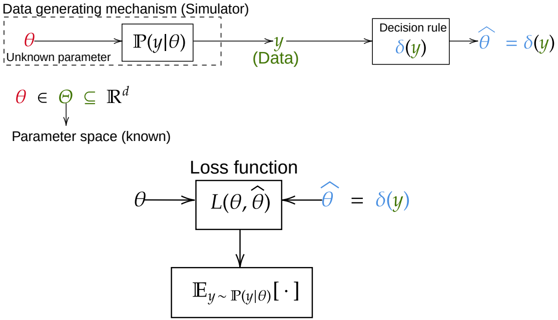
Risk: $R(\theta, \delta) := \mathbb{E}_{y \sim \mathbb{P}(y|\theta)}[L(\theta, \widehat{\theta})]$

Goal

Design optimal decision rules that minimize $R(\theta, \delta)$

- Bayes estimator: Minimize $\mathbb{E}_{\theta \sim \pi}[R(\theta, \delta)]$

Problem Statement



Risk: $R(\theta, \delta) := \mathbb{E}_{y\sim\mathbb{P}(y|\theta)}[L(\theta, \widehat{\theta})]$

Goal

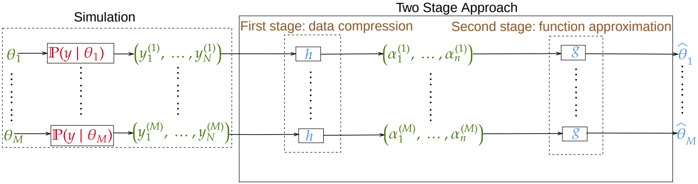
Design optimal decision rules that minimize $R(\theta, \delta)$

- Bayes estimator: Minimize $\mathbb{E}_{\theta\sim\pi}[R(\theta, \delta)]$
- Minimax estimator: Minimize $\max_{\theta\in\Theta} R(\theta, \delta)$

Outline

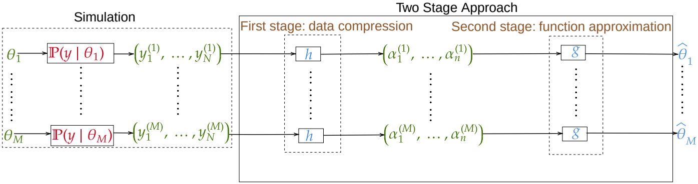
Two-Stage Approach

- Proposed by S. Garatti and S. Bittanti. "Estimation of white-box model parameters via artificial data generation: a two-stage approach". In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 11409–11414



Two-Stage Approach

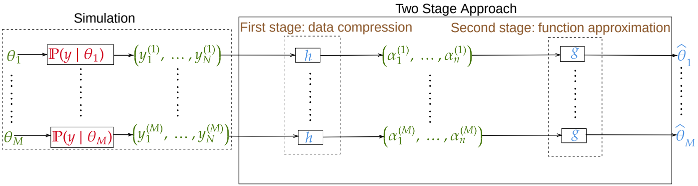
- Proposed by S. Garatti and S. Bittanti. “Estimation of white-box model parameters via artificial data generation: a two-stage approach”. In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 11409–11414



- $\hat{\theta}_i = g(h(\mathbf{y}^{(i)}))$: estimates of θ_i for $i = 1, \dots, M$

Two-Stage Approach

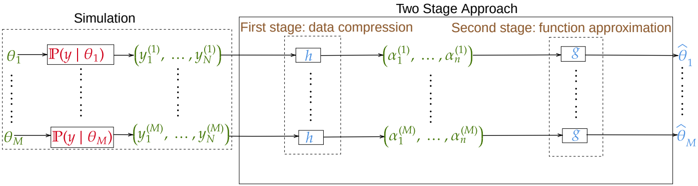
- Proposed by S. Garatti and S. Bittanti. "Estimation of white-box model parameters via artificial data generation: a two-stage approach". In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 11409–11414



- $\hat{\theta}_i = g(h(\mathbf{y}^{(i)}))$: estimates of θ_i for $i = 1, \ldots, M$
- h is fixed and g is chosen from function space G to minimize the training error

Two-Stage Approach

- Proposed by S. Garatti and S. Bittanti. "Estimation of white-box model parameters via artificial data generation: a two-stage approach". In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 11409–11414

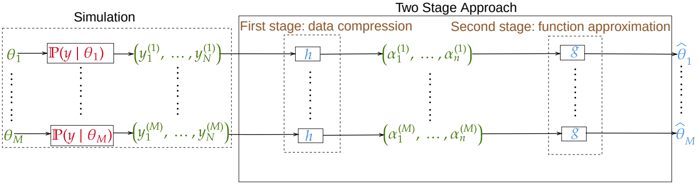


- $\hat{\theta}_i = g(h(\mathbf{y}^{(i)}))$: estimates of θ_i for $i = 1, \dots, M$

- h is fixed and g is chosen from function space G to minimize the training error
  - $g^* = \arg\min_{g \in G} \frac{1}{M_y} \frac{1}{M_\theta} \sum_{i=1}^{M_\theta} \sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)})))$

Two-Stage Approach

- Proposed by S. Garatti and S. Bittanti. "Estimation of white-box model parameters via artificial data generation: a two-stage approach". In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 11409–11414



- $\hat{\theta}_i = g(h(\mathbf{y}^{(i)}))$: estimates of θ_i for $i = 1, \dots, M$
- h is fixed and g is chosen from function space G to minimize the training error
  - $g^* = \arg\min_{g \in G} \frac{1}{M_y} \frac{1}{M_\theta} \sum_{i=1}^{M_\theta} \sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)})))$
  - G can be a linear regressor, deep neural network, gradient boosted regression tree, etc.

Outline

Bayesian Framework for Two-Stage Approach

- Assume a prior π over Θ

Bayesian Framework for Two-Stage Approach

- Assume a prior π over Θ
- *Bayes risk:*

$$R_{\text{Bayes}}(\delta) = \int_{\Theta} R(\theta, \delta)\pi(\theta)\, d\theta$$

Bayesian Framework for Two-Stage Approach

- Assume a prior π over Θ
- *Bayes risk:*

$$R_{\text{Bayes}}(\delta) = \int_\Theta R(\theta, \delta)\pi(\theta)\, d\theta$$

$$= \underbrace{\int_\Theta \left[\int_\mathcal{Y} L(\theta, \delta(y))\, \mathbb{P}(y|\theta) dy \right] \pi(\theta) d\theta}_{\text{can be approximated by a Monte-Carlo sum!}}$$

Bayesian Framework for Two-Stage Approach

- Assume a prior π over Θ
- *Bayes risk:*

$$R_{\text{Bayes}}(\delta) = \int_{\Theta} R(\theta, \delta)\pi(\theta)\, d\theta$$

$$= \underbrace{\int_{\Theta} \left[\int_{\mathcal{Y}} L(\theta, \delta(y))\, \mathbb{P}(y|\theta) dy \right] \pi(\theta) d\theta}_{\text{can be approximated by a Monte-Carlo sum!}}$$

- Approximate Bayes risk:

$$R_{\text{Bayes}}(\delta) \approx \frac{1}{M_y} \frac{1}{M_\theta} \sum_{i=1}^{M_\theta} \sum_{j=1}^{M_y} L(\theta_i, \delta(y_j^{(i)}))$$

Bayesian Framework for Two-Stage Approach

- Assume a prior π over Θ
- *Bayes risk:*

$$R_{\text{Bayes}}(\delta) = \int_\Theta R(\theta, \delta)\pi(\theta)\, d\theta$$

$$= \underbrace{\int_\Theta \left[\int_{\mathcal{Y}} L(\theta, \delta(y))\, \mathbb{P}(y|\theta) dy \right] \pi(\theta) d\theta}_{\text{can be approximated by a Monte-Carlo sum!}}$$

- Approximate Bayes risk:

$$R_{\text{Bayes}}(\delta) \approx \frac{1}{M_y} \frac{1}{M_\theta} \sum_{i=1}^{M_\theta} \sum_{j=1}^{M_y} L(\theta_i, \delta(y_j^{(i)}))$$

Goal

Find δ that minimize $R_{\text{Bayes}}(\delta)$

Bayesian Framework for Two-Stage Approach

- **Caveat**: $\mathbb{P}(y|\theta)$ may be subject to a measure concentration phenomenon. ($\mathbb{P}(y|\theta)$ may be concentrated over a small subset of $\mathbb{R}^N$)

Bayesian Framework for Two-Stage Approach

- **Caveat**: $\mathbb{P}(y|\theta)$ may be subject to a measure concentration phenomenon. ($\mathbb{P}(y|\theta)$ may be concentrated over a small subset of $\mathbb{R}^N$)
- Standard function approximation would require highly complex models.

Bayesian Framework for Two-Stage Approach

- **Caveat**: $\mathbb{P}(y|\theta)$ may be subject to a measure concentration phenomenon. ($\mathbb{P}(y|\theta)$ may be concentrated over a small subset of $\mathbb{R}^N$)
- Standard function approximation would require highly complex models.

Solution

Suitable decomposition of δ as $g \circ h$.

Bayesian Framework for Two-Stage Approach

- Final approximate Bayes risk:

$$\frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

Bayesian Framework for Two-Stage Approach

- Final approximate Bayes risk:

$$\frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

- By fixing h, we obtain $\tilde{\delta}_{\text{Bayes}}$ by solving for g

Bayesian Framework for Two-Stage Approach

- Final approximate Bayes risk:

$$\frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

- By fixing h, we obtain $\tilde{\delta}_{\text{Bayes}}$ by solving for g

$$g^* = \arg\min_{g\in G} \frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

Bayesian Framework for Two-Stage Approach

- Final approximate Bayes risk:

$$\frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

- By fixing h, we obtain $\tilde{\delta}_{\text{Bayes}}$ by solving for g

$$g^* = \arg\min_{g\in G} \frac{1}{M_y}\frac{1}{M_\theta}\sum_{i=1}^{M_\theta}\sum_{j=1}^{M_y} L(\theta_i, g(h(y_j^{(i)}))).$$

Bayes TS estimator!

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$$

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$$

Goal

Find δ that minimize $R_{\text{minimax}}(\delta)$

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$$

Goal

Find δ that minimize $R_{\text{minimax}}(\delta)$

$$\max_{\theta \in \Theta} R(\theta, \delta) = \max_{\theta \in \Theta} \int_{\mathcal{Y}} L\left(\theta, \delta(\mathbf{y})\right) \mathbb{P}(\mathbf{y}|\theta) d\mathbf{y}$$

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$$

Goal

Find δ that minimize $R_{\text{minimax}}(\delta)$

$$\max_{\theta \in \Theta} R(\theta, \delta) = \max_{\theta \in \Theta} \underbrace{\int_{\mathcal{Y}} L\left(\theta, \delta(\mathbf{y})\right) \mathbb{P}(\mathbf{y}|\theta) d\mathbf{y}}_{\approx \frac{1}{M_{\mathbf{y}}} \sum_{j=1}^{M_{\mathbf{y}}} L(\theta, \delta(\mathbf{y}_{j,\theta}))}$$

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta\in\Theta} R(\theta,\delta)$$

Goal

Find δ that minimize $R_{\text{minimax}}(\delta)$

$$\max_{\theta\in\Theta} R(\theta,\delta) = \max_{\theta\in\Theta} \underbrace{\int_{\mathcal{Y}} L\left(\theta, \delta(\mathbf{y})\right) \mathbb{P}(\mathbf{y}|\theta) d\mathbf{y}}_{\approx \frac{1}{M_{\mathbf{y}}} \sum_{j=1}^{M_{\mathbf{y}}} L(\theta,\delta(\mathbf{y}_{j,\theta}))}$$

Approximate minimax optimization problem:

Minimax Framework for Two-Stage Approach

- *Minimax risk*:

$$R_{\text{minimax}}(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$$

Goal

Find δ that minimize $R_{\text{minimax}}(\delta)$

$$\max_{\theta \in \Theta} R(\theta, \delta) = \max_{\theta \in \Theta} \underbrace{\int_{\mathcal{Y}} L\left(\theta, \delta(\mathbf{y})\right) \mathbb{P}(\mathbf{y}|\theta) d\mathbf{y}}_{\approx \frac{1}{M_{\mathbf{y}}} \sum_{j=1}^{M_{\mathbf{y}}} L(\theta, \delta(\mathbf{y}_{j,\theta}))}$$

Approximate minimax optimization problem:

$$\min_{\delta \in \Delta} \max_{\theta \in \Theta} \frac{1}{M_{\mathbf{y}}} \sum_{j=1}^{M_{\mathbf{y}}} L(\theta, \delta(\mathbf{y}_{j,\theta}))$$

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Solution

Scenario approach (Calafiore and Campi, 2006)

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Solution

Scenario approach (Calafiore and Campi, 2006)

- Sample $\theta_i \overset{i.i.d.}{\sim} s(\theta), \quad i = 1, \ldots, M_\theta$

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Solution

Scenario approach (Calafiore and Campi, 2006)

- Sample $\theta_i \overset{i.i.d.}{\sim} s(\theta), \quad i = 1, \ldots, M_\theta$
- Tractable optimization problem:

$$\min_{\delta \in \Delta} \max_{i=1,\ldots,M_\theta} L_i(\delta)$$

$$L_i(\delta) = \sum_{j=1}^{M_{\mathbf{y}}} L(\theta_i, \delta(\mathbf{y}_{ij})), \quad \mathbf{y}_{ij} = \mathbf{y}_{j,\theta_i}$$

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Solution

Scenario approach (Calafiore and Campi, 2006)

- Sample $\theta_i \overset{i.i.d.}{\sim} s(\theta), \quad i = 1, \ldots, M_\theta$
- Tractable optimization problem:

$$\min_{\delta \in \Delta} \max_{i=1,\ldots,M_\theta} L_i(\delta)$$

$$L_i(\delta) = \sum_{j=1}^{M_{\mathbf{y}}} L(\theta_i, \delta(\mathbf{y}_{ij})), \quad \mathbf{y}_{ij} = \mathbf{y}_{j,\theta_i}$$

- Epigraph formulation:

$$\begin{aligned} \min_{\delta \in \Delta, t \in \mathbb{R}} \quad & t \\ \text{s.t.} \quad & L_i(\delta) \leq t, \quad i = 1, \ldots, M_\theta. \end{aligned}$$

Minimax Framework for Two-Stage Approach

- **Caveat**: Θ could be potentially infinite!
- How to make approximate minimax optimization tractable?

Solution

Scenario approach (Calafiore and Campi, 2006)

- Sample $\theta_i \overset{i.i.d.}{\sim} s(\theta), \quad i = 1, \ldots, M_\theta$
- Tractable optimization problem:

$$\min_{\delta \in \Delta} \max_{i=1,\ldots,M_\theta} L_i(\delta)$$

$$L_i(\delta) = \textstyle\sum_{j=1}^{M_{\mathbf{y}}} L(\theta_i, \delta(\mathbf{y}_{ij})), \quad \mathbf{y}_{ij} = \mathbf{y}_{j,\theta_i}$$

- Epigraph formulation:

$$\begin{aligned} \min_{\delta \in \Delta, t \in \mathbb{R}} \quad & t \\ \text{s.t.} \quad & L_i(\delta) \leq t, \quad i = 1, \ldots, M_\theta. \end{aligned}$$

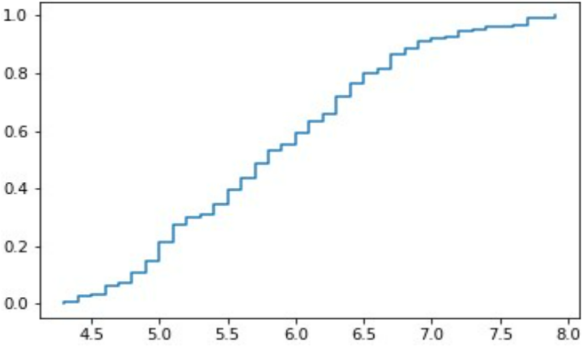
Solved using CVXPY!

Choice of First Stage

- A fixed number of *quantiles* of the order statistic is taken as the compressed data.

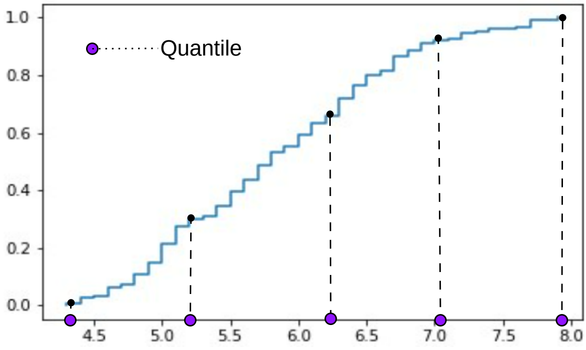
Choice of First Stage

- A fixed number of *quantiles* of the order statistic is taken as the compressed data.



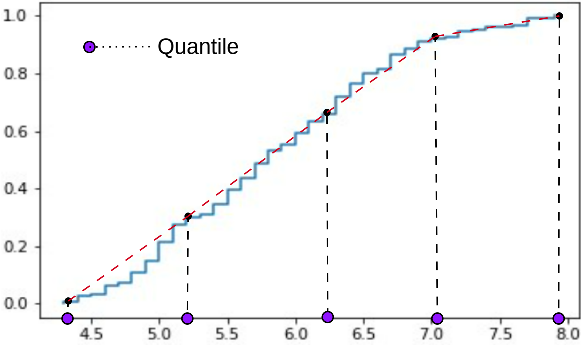
Choice of First Stage

- A fixed number of *quantiles* of the order statistic is taken as the compressed data.



Choice of First Stage

- A fixed number of *quantiles* of the order statistic is taken as the compressed data.



Choice of First Stage

- $$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Choice of First Stage

- $\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \xrightarrow{\text{sort}} \underbrace{\begin{pmatrix} y_{(1)} \\ \vdots \\ y_{(N)} \end{pmatrix}}_{\text{order statistic}}$

Choice of First Stage

- $$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \xrightarrow{\text{sort}} \underbrace{\begin{pmatrix} y_{(1)} \\ \vdots \\ y_{(N)} \end{pmatrix}}_{\text{order statistic}} \xrightarrow{\text{every } q^{th} \text{ quantile}} \underbrace{\begin{pmatrix} y_{(1)} \\ y_{(\frac{N}{q})} \\ y_{(\frac{2N}{q})} \\ \vdots \\ y_{(N)} \end{pmatrix}}_{\text{compressed data}} = (\alpha_1, \ldots, \alpha_n)^T$$

Choice of First Stage

- $$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \xrightarrow{\text{sort}} \underbrace{\begin{pmatrix} y_{(1)} \\ \vdots \\ y_{(N)} \end{pmatrix}}_{\text{order statistic}} \xrightarrow{\text{every } q^{th} \text{ quantile}} \underbrace{\begin{pmatrix} y_{(1)} \\ y_{(\frac{N}{q})} \\ y_{(\frac{2N}{q})} \\ \vdots \\ y_{(N)} \end{pmatrix}}_{\text{compressed data}} = (\alpha_1, \ldots, \alpha_n)^T$$
- *Compressed data*: $\alpha := (\alpha_1, \ldots, \alpha_n)^T$, where α_is are quantiles in increasing order.

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$
  - α is the compressed data from first stage

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$
  - α is the compressed data from first stage
  - ϕ is a non-linear function $\phi\colon \mathbb{R}^n \to \mathbb{R}^m$ with $n < m \ll N$

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$
  - α is the compressed data from first stage
  - ϕ is a non-linear function $\phi \colon \mathbb{R}^n \to \mathbb{R}^m$ with $n < m \ll N$
- Deploy linear regression with $\phi(\alpha)$ as features

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$
  - α is the compressed data from first stage
  - ϕ is a non-linear function $\phi\colon \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $n < m \ll N$
- Deploy linear regression with $\phi(\alpha)$ as features
- Both n and $\phi(\cdot)$ are user choices

Choice of Second Stage

- Construct non-linear features $\phi(\alpha)$
  - α is the compressed data from first stage
  - ϕ is a non-linear function $\phi\colon \mathbb{R}^n \to \mathbb{R}^m$ with $n < m \ll N$
- Deploy linear regression with $\phi(\alpha)$ as features
- Both n and $\phi(\cdot)$ are user choices
  - Selected according to the specific estimation problem.

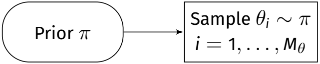
Algorithm: Bayes Two-Stage Estimator

Training

Prior π

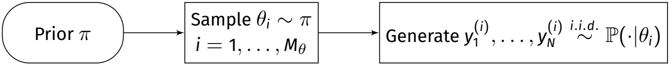
Algorithm: Bayes Two-Stage Estimator

Training



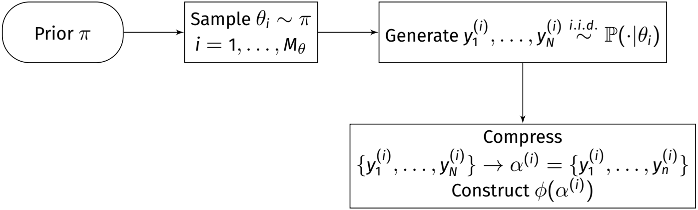
Algorithm: Bayes Two-Stage Estimator

Training



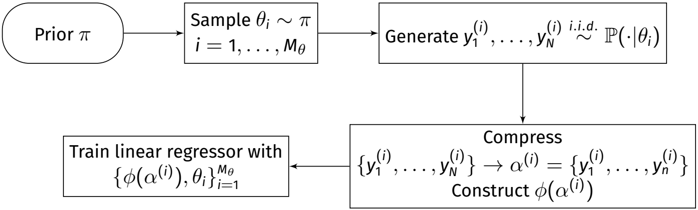
Algorithm: Bayes Two-Stage Estimator

Training



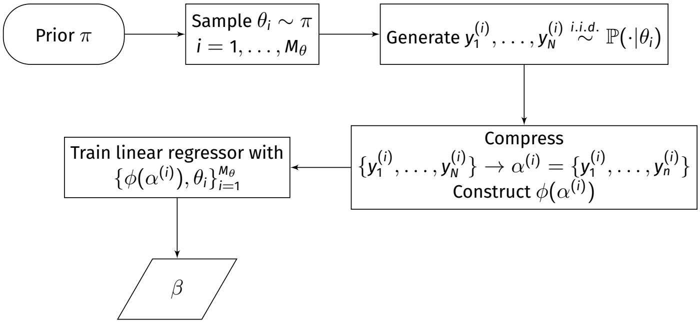
Algorithm: Bayes Two-Stage Estimator

Training



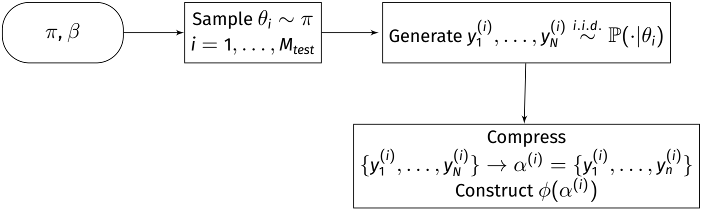
Algorithm: Bayes Two-Stage Estimator

Training



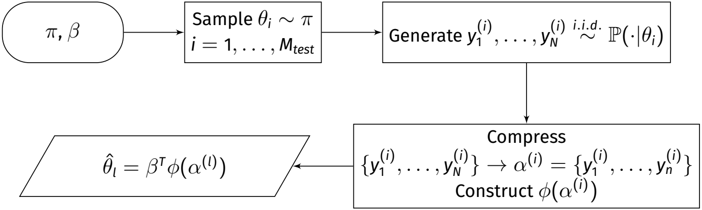
Algorithm: Bayes Two-Stage Estimator

Estimation



Algorithm: Bayes Two-Stage Estimator

Estimation



Algorithm: Minimax Two-Stage Estimator

Training

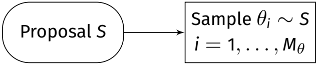
Algorithm: Minimax Two-Stage Estimator

Training

Proposal S

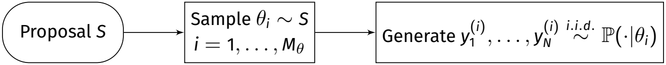
Algorithm: Minimax Two-Stage Estimator

Training



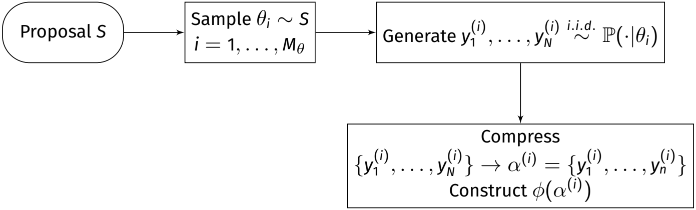
Algorithm: Minimax Two-Stage Estimator

Training



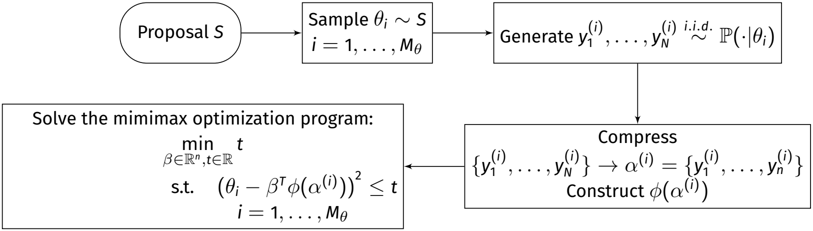
Algorithm: Minimax Two-Stage Estimator

Training



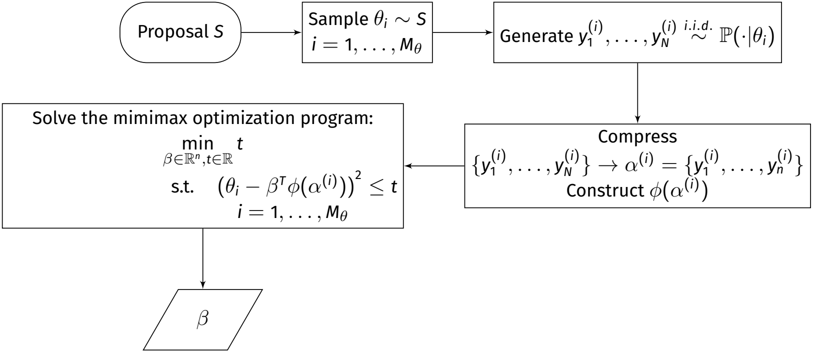
Algorithm: Minimax Two-Stage Estimator

Training



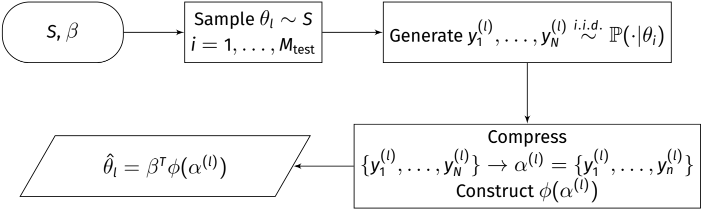
Algorithm: Minimax Two-Stage Estimator

Training



Algorithm: Minimax Two-Stage Estimator

<u>Estimation</u>



Outline

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

$$f(A) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1}\exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0.$$

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

$$f(A) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1} \exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0.$$

- $M_\theta = 1000$, $N = 10000$, $n = 10$, and $M_{test} = 1000$

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

$$f(A) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1} \exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0.$$

- $M_\theta = 1000$, $N = 10000$, $n = 10$, and $M_{test} = 1000$
- Choice of ϕ:
  - Scale:

$$\phi_i(\alpha) = \begin{cases} \alpha_i, & \text{if } 1 \leq i \leq n \\ \frac{\alpha_{i-n+1}}{\alpha_1}, & \text{if } n+1 \leq i \leq 2n-1. \end{cases}$$

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

$$f(A) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1} \exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0.$$

- $M_\theta = 1000$, $N = 10000$, $n = 10$, and $M_{test} = 1000$
- Choice of ϕ:
  - Scale:

$$\phi_i(\alpha) = \begin{cases} \alpha_i, & \text{if } 1 \leq i \leq n \\ \dfrac{\alpha_{i-n+1}}{\alpha_1}, & \text{if } n+1 \leq i \leq 2n-1. \end{cases}$$

  - Shape: Let

$$\psi_j(\alpha) = \begin{cases} \alpha_j, & \text{if } 1 \leq j \leq n \\ \dfrac{\alpha_{j-n}}{\alpha_n}, & \text{if } n+1 \leq j \leq 2n-1. \end{cases}$$

Example

- Aim: To estimate scale (η) and shape (γ) parameters of Weibull process:

$$f(A) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1} \exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0.$$

- $M_\theta = 1000$, $N = 10000$, $n = 10$, and $M_{test} = 1000$
- Choice of ϕ:
  - Scale:

$$\phi_i(\alpha) = \begin{cases} \alpha_i, & \text{if } 1 \leq i \leq n \\ \dfrac{\alpha_{i-n+1}}{\alpha_1}, & \text{if } n+1 \leq i \leq 2n-1. \end{cases}$$

  - Shape: Let

$$\psi_j(\alpha) = \begin{cases} \alpha_j, & \text{if } 1 \leq j \leq n \\ \dfrac{\alpha_{j-n}}{\alpha_n}, & \text{if } n+1 \leq j \leq 2n-1. \end{cases}$$

    - $\phi(\alpha)$ consist of monomials of the $\psi_j(\alpha)$'s up to order 2

Example

Bayes Two-Stage estimator

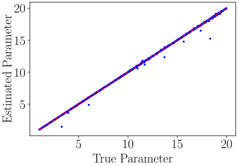
- Independent priors $\mathcal{U}[1, 20]$ for η and γ

Example

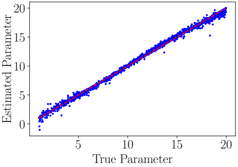
Bayes Two-Stage estimator

- Independent priors $\mathcal{U}[1, 20]$ for η and γ

Scale



Shape



Example

- Independent *uninformative* priors for η and γ

Example

- Independent *uninformative* priors for η and γ

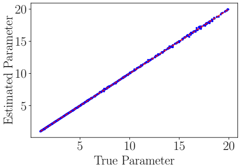
$$f(x) = \begin{cases} \dfrac{\frac{1}{\log \frac{b}{a}}}{x}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Example

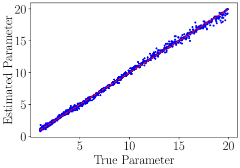
- Independent *uninformative* priors for η and γ

$$f(x) = \begin{cases} \frac{\frac{1}{\log \frac{b}{a}}}{x}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Scale



Shape



Example

Minimax Two-Stage Estimator

Example

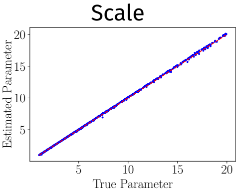
Minimax Two-Stage Estimator

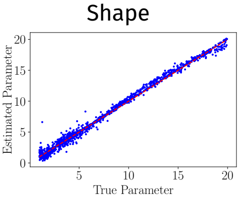
- Proposal distribution S: $\mathcal{U}[1, 20]$

Example

Minimax Two-Stage Estimator

- Proposal distribution S: $\mathcal{U}[1, 20]$





Outline

Conclusion

Conclusion

- Provided statistical decision-theoretical derivation of TS that leads to Bayes and minimax formulations

Conclusion

- Provided statistical decision-theoretical derivation of TS that leads to Bayes and minimax formulations
- Suggested a specific structure for the second stage of TS, which leads to simple convex programs for both Bayes and minimax formulations

Conclusion

- Provided statistical decision-theoretical derivation of TS that leads to Bayes and minimax formulations
- Suggested a specific structure for the second stage of TS, which leads to simple convex programs for both Bayes and minimax formulations
- Illustrated the performance of the novel Bayes and minimax TS formulations

Thank You

Appendix

Appendix: CRLB-MSE Comparison

True Values		CRLB		MSE, Bayesian (Uniform Prior)		MSE, Bayesian (Uninformative Prior)		MSE, Minimax	
η	γ	η	γ	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\gamma}$
2	2	1.11×10^{-4}	2.43×10^{-4}	2.58×10^{-4}	5.77×10^{-2}	1.42×10^{-4}	2.06×10^{-2}	2.17×10^{-4}	16×10^{-2}
2	8	6.93×10^{-6}	3.89×10^{-3}	1.11×10^{-5}	5.61×10^{-2}	1.27×10^{-5}	4.44×10^{-2}	4.28×10^{-4}	13.29×10^{-2}
4	2	4.43×10^{-4}	2.43×10^{-4}	6.74×10^{-4}	1.05×10^{-1}	6.07×10^{-4}	2.43×10^{-2}	8.38×10^{-4}	14.67×10^{-2}
4	8	2.77×10^{-5}	3.89×10^{-3}	3.84×10^{-5}	6.40×10^{-2}	4.33×10^{-5}	3.96×10^{-2}	1.72×10^{-3}	8.59×10^{-2}
8	2	1.77×10^{-3}	2.43×10^{-4}	2.26×10^{-3}	1.89×10^{-1}	2.27×10^{-3}	2.35×10^{-2}	3.307×10^{-3}	18.605×10^{-2}
8	8	1.11×10^{-4}	3.89×10^{-3}	1.58×10^{-4}	7.901×10^{-2}	1.76×10^{-4}	4.51×10^{-2}	6.77×10^{-3}	8.39×10^{-2}

Table: MSE of Bayes and minimax TS estimators of the scale and shape parameters, and their corresponding CRLBs.

Appendix: Minimax TS Estimator for Weibull Process

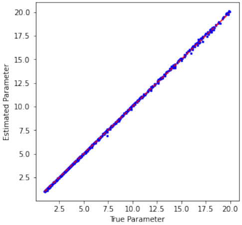
- Proposal distribution S: $\mathcal{U}[1, 20]$;
  True distribution of θ:

$$f(\theta) = \begin{cases} \frac{\frac{1}{\log \frac{b}{a}}}{\theta}, & \text{if } a \leq \theta \leq b \\ 0, & \text{otherwise} \end{cases}$$

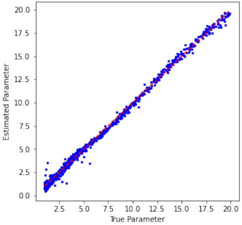
  where $a = 1,\ b = 20$

- d=3; n=5

Scale



Shape



Appendix: Minimax TS Estimator for Weibull Process

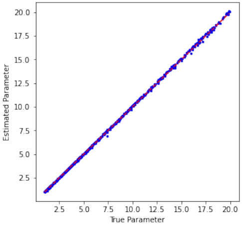
- Proposal distribution S:

$$f(\theta) = \begin{cases} \frac{\frac{1}{\log \frac{b}{a}}}{\theta}, & \text{if } a \leq \theta \leq b \\ 0, & \text{otherwise} \end{cases}$$

where $a = 1,\ b = 20$
True distribution of θ: $\mathcal{U}[1, 20]$

- d=3; n=5

Scale



Shape