An introduction to optimality

Statistical Consultary

SHENDUO ZHANG

Institute of Mathematics
Xi'an Jiaotong University

November 8, 2020

Summary

Minimaxity

Classical Linear Model

Linear model:

$$Y = X\beta + \xi, \xi \sim N(0, \sigma^2 I_V)$$

W, V linear space, $\beta \in W$, $X : W \mapsto V$, $Y, \xi \in V$, σ is known.

$$\Downarrow$$

Gaussian shift model:

$$Y = \mu + \xi, \xi \sim N(0, \sigma^2 I_V)$$

$\mu = X\beta$, $\mu \in L = \mathrm{Im}(X) \subset V$, σ is known.

Classical Linear Model

Linear model:

$$Y = X\beta + \xi, \xi \sim N(0, \sigma^2 I_V)$$

W, V linear space, $\beta \in W$, $X : W \mapsto V$, $Y, \xi \in V$, σ is known.

$$\Downarrow$$

Gaussian shift model:

$$Y = \mu + \xi, \xi \sim N(0, \sigma^2 I_V)$$

$\mu = X\beta$, $\mu \in L = \mathrm{Im}(X) \subset V$, σ is known.

Least square estimator

$$\hat{\beta} = \arg\min_{\beta \in W} \|Y - X\beta\|^2$$

$$\Downarrow$$

$$\hat{\mu} = P_L Y = X\hat{\beta}$$

$$\Downarrow$$

$$X^* X \hat{\beta} = X^* Y$$

$$\Downarrow$$

$$\hat{\beta} \in X^+ Y + \ker(X)$$

where X^+ is the MP inverse.

Least square estimator

$$\hat{\beta} = \arg\min_{\beta \in W} \|Y - X\beta\|^2$$

$$\Downarrow$$

$$\hat{\mu} = P_L Y = X\hat{\beta}$$

$$\Downarrow$$

$$X^* X \hat{\beta} = X^* Y$$

$$\Downarrow$$

$$\hat{\beta} \in X^+ Y + \ker(X)$$

where X^+ is the MP inverse.

Least square estimator

$$\hat{\beta} = \arg\min_{\beta \in W} \|Y - X\beta\|^2$$

$$\Downarrow$$

$$\hat{\mu} = P_L Y = X\hat{\beta}$$

$$\Downarrow$$

$$X^* X \hat{\beta} = X^* Y$$

$$\Downarrow$$

$$\hat{\beta} \in X^+ Y + \ker(X)$$

where X^+ is the MP inverse.

Least square estimator

$$\hat{\beta} = \arg\min_{\beta \in W} \|Y - X\beta\|^2$$

$$\Downarrow$$

$$\hat{\mu} = P_L Y = X\hat{\beta}$$

$$\Downarrow$$

$$X^* X \hat{\beta} = X^* Y$$

$$\Downarrow$$

$$\hat{\beta} \in X^+ Y + \ker(X)$$

where X^+ is the MP inverse.

$\hat{\mu}$ is a good estimator

Theorem (Gauss-Markov)

Suppose $< Y, d >$ *for some* $d \in V$ *is a linear unbiased for linear functinoal* $< \mu, c >$, $\mu \in L$. *Then,*

$$\mathbf{Var}(< Y, d >) \geq \mathbf{Var}(< \hat{\mu}, c >), \mu \in L \tag{1}$$

and more over, $< \hat{\mu}, c >$ *is the unique linear unbiased estimator with the smallest possible variance.*

General statistical framework

1. $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\theta)$ where the data was drawn. $\mathbb{P}$ is index by θ.
2. $\theta \in \Theta$ where θ is the estimation target, Θ is the candidate family.
3. $d(\cdot, \cdot) : \Theta \times \Theta \mapsto [0, \infty)$ is a semi-distance to judge the performance of estimation.

Risk of $\hat{\mu}$

Definition (Risk)

$$R(\theta; \hat{\theta}) = \mathbb{E}_{\theta} d\left(\theta - \hat{\theta}\right).$$

Definition (Maximum Risk)

$$r(\theta; \hat{\theta}) = \sup_{\theta \in \Theta} R(\theta; \hat{\theta}) = \mathbb{E}_{\theta} d\left(\theta - \hat{\theta}\right)^2.$$

Risk is not accessible, maximum risk is with enough regulation.

$$R(\mu; \hat{\mu}) = r(\mu; \hat{\mu}) = \mathbb{E}_{\theta} ||\mu - \hat{\mu}||^2 = \sigma^2 \dim(L). \tag{2}$$

Risk of $\hat{\mu}$

Definition (Risk)

$$R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right).$$

Definition (Maximum Risk)

$$r(\theta;\hat{\theta}) = \sup_{\theta\in\Theta} R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right)^2.$$

Risk is not accessible, maximum risk is with enough regulation.

$$R(\mu;\hat{\mu}) = r(\mu;\hat{\mu}) = \mathbb{E}_\theta ||\mu - \hat{\mu}||^2 = \sigma^2 \dim(L). \tag{2}$$

Risk of $\hat{\mu}$

Definition (Risk)

$$R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right).$$

Definition (Maximum Risk)

$$r(\theta;\hat{\theta}) = \sup_{\theta\in\Theta} R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right)^2.$$

Risk is not accessible, maximum risk is with enough regulation.

$$R(\mu;\hat{\mu}) = r(\mu;\hat{\mu}) = \mathbb{E}_\theta \|\mu - \hat{\mu}\|^2 = \sigma^2 \dim(L). \tag{2}$$

Risk of $\hat{\mu}$

Definition (Risk)

$$R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right).$$

Definition (Maximum Risk)

$$r(\theta;\hat{\theta}) = \sup_{\theta\in\Theta} R(\theta;\hat{\theta}) = \mathbb{E}_\theta d\left(\theta - \hat{\theta}\right)^2.$$

Risk is not accessible, maximum risk is with enough regulation.

$$R(\mu;\hat{\mu}) = r(\mu;\hat{\mu}) = \mathbb{E}_\theta \|\mu - \hat{\mu}\|^2 = \sigma^2 \dim(L). \tag{2}$$

$\hat{\mu}$ is the best estimator

Theorem (Minimaxity of projection estimator)

$\forall$ estiamtor $T(Y)$ of μ,

$$\sup_{\mu \in L} \mathbb{E}_\mu \|T(Y) - \mu\|^2 \geq \sup_{\mu \in L} \mathbb{E}_\mu \|\hat{\mu} - \mu\|^2 = \sigma^2 \dim L. \tag{3}$$

In another words,

$$\sup_{\mu \in L} R(\mu; \hat{\mu}) = \inf_T \sup_{\mu \in L} R(\mu; T). \tag{4}$$

Definition (Minimax estimator)

An estimator T^* is called minimax estimator if

$$\sup_{\theta \in \Theta} R(\theta, T^*) = \inf_T \sup_{\theta \in \Theta} R(\theta, T) \tag{5}$$

$\hat{\mu}$ is the best estimator

Theorem (Minimaxity of projection estimator)

$\forall$ *estiamtor* $T(Y)$ *of* μ,

$$\sup_{\mu \in L} \mathbb{E}_\mu \|T(Y) - \mu\|^2 \geq \sup_{\mu \in L} \mathbb{E}_\mu \|\hat{\mu} - \mu\|^2 = \sigma^2 \dim L. \tag{3}$$

In another words,

$$\sup_{\mu \in L} R(\mu; \hat{\mu}) = \inf_T \sup_{\mu \in L} R(\mu; T). \tag{4}$$

Definition (Minimax estimator)

An estimator T^* is called minimax estimator if

$$\sup_{\theta \in \Theta} R(\theta, T^*) = \inf_T \sup_{\theta \in \Theta} R(\theta, T) \tag{5}$$

Reduction to finding Bayes estimator

Definition (Bayes risk)

Let T be an estimator of θ, $R(\theta;T)$ be the risk of T,

$$R_\Pi(T) = \int_\Theta R(\theta;T)\Pi(\mathrm{d}\theta) = \int_\Theta R(\theta;T)\pi(\theta)\mathrm{d}\theta \tag{6}$$

is called the average risk with respect to prior Π.

Definition (Bayes estimator)

The estimator T_Π is Bayes estimator with respect to Π if and only if $\forall T$, estimator of θ, one has

$$R_\Pi(T) \geq R_\Pi(T_\Pi) \tag{7}$$

Reduction to finding Bayes estimator

Definition (Bayes risk)

Let T be an estimator of θ, $R(\theta; T)$ be the risk of T,

$$R_\Pi(T) = \int_\Theta R(\theta; T)\Pi(\mathrm{d}\theta) = \int_\Theta R(\theta; T)\pi(\theta)\mathrm{d}\theta \tag{6}$$

is called the average risk with respect to prior Π.

Definition (Bayes estimator)

The estimator T_Π is Bayes estimator with respect to Π if and only if $\forall T$, estimator of θ, one has

$$R_\Pi(T) \geq R_\Pi(T_\Pi) \tag{7}$$

Definition (Bayes estimator)

The estimator T_Π is Bayes estimator with respect to Π if and only if $\forall T$, estimator of θ, let Pi be a probability distribution on Θ, one has

$$R_\Pi(T) \geq R_\Pi(T_\Pi) \tag{8}$$

Theorem

Suppose exists an estimator $T(\mathbf{X})$ and a sequence of prior distribution $\{\Pi_k\}$ such that

$$R_{\Pi_k}(T_{\Pi_k}) \to \sup_{\theta\in\Theta} R(\theta; T) \quad (k \to \infty) \tag{9}$$

Then T is a minimax estimator.

Reduction to finding Bayes estimator

Theorem

Suppose exists an estimator $T(\mathbf{X})$ and a sequence of prior distribution $\{\Pi_k\}$ such that

$$R_{\Pi_k}(T_{\Pi_k}) \to \sup_{\theta\in\Theta} R(\theta; T) \quad (k \to \infty) \tag{10}$$

Then T is a minimax estimator.

Proof.

$\forall$ esimator $\tilde{T}$,

$$\sup_{\theta\in\Theta} R(\theta;\tilde{T}) \geq R_{\Pi_k}(\tilde{T}) \geq R_{\Pi_k}(T_{\Pi_k}) \to \sup_{\theta\in\Theta} R(\theta; T) \tag{11}$$

□

Explicit construction of prior

$$\Pi_k \sim N(0, \tau_k^2 I_L), \tau_k^2 \to \infty \tag{12}$$

$$T_{\Pi_k}(Y) = \frac{\tau_k^2}{\sigma^2 + \tau_k^2} \hat{\mu} \tag{13}$$

$$R_{\Pi_k}(\mu; T_{\Pi_k}) = \frac{\tau_k^2}{\sigma^2 + \tau_k^2} \sigma^2 d + \left(\frac{\sigma^2}{\sigma^2 + \tau_k^2} \right)^2 \tau_k^2 d \to \sigma^2 d \tag{14}$$

$\hat{\mu}$ is not the best estimator

Theorem (James-Stein)

When $\dim L \geq 3$, *there* $\exists$ *an estimator* $T(Y)$ *of* μ *such that* $\forall \mu \in L$,

$$\mathbb{E}_\mu \|T(Y) - \mu\|^2 < \mathbb{E}_\mu \|\hat{\mu} - \mu\|^2 = \sigma^2 \dim L \tag{15}$$

Admissibility

Admissible estimator

Definition (Admissible estimator)

An estimator $\hat{\theta}$ is called inadmissible if there $\exists T(Y)$ of θ such that

$$\mathbb{E}_\theta \|T(Y) - \theta\|^2 \leq \mathbb{E}_\mu \left\|\hat{\theta} - \theta\right\|^2 \tag{16}$$

with the strict inequality for some $\theta \in \Theta$. Otherwise, it's called admissible.

In another word, projection estimator $\hat{\mu}$ is an inadmissible estimator of linear model when $\dim L \geq 3$.

Admissible estimator

Definition (Admissible estimator)

An estimator $\hat{\theta}$ is called inadmissible if there $\exists T(Y)$ of θ such that

$$\mathbb{E}_\theta \|T(Y) - \theta\|^2 \leq \mathbb{E}_\mu \left\|\hat{\theta} - \theta\right\|^2 \tag{16}$$

with the strict inequality for some $\theta \in \Theta$. Otherwise, it's called admissible.

In another word, projection estimator $\hat{\mu}$ is an inadmissible estimator of linear model when $\dim L \geq 3$.

James-Stein

$$T(Y) = \hat{\mu} + \sigma^2 g(\hat{\mu}) \tag{17}$$

where $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a smooth function.

Check its risk,

$$\begin{aligned}
\mathbb{E}_\mu \|T(Y) - \mu\|^2 &= \mathbb{E}_\mu \Big\| \hat{\mu} + \sigma^2 g(\hat{\mu}) - \mu \Big\|^2 \\
&= \mathbb{E}_\mu \Big(\|\mu - \hat{\mu}\|^2 + \sigma^4 \|g(\hat{\mu})\|^2 + 2\sigma^2 \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle \Big) \\
&= \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^2 \mathbb{E}_\mu \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle
\end{aligned}$$

James-Stein

$$T(Y) = \hat{\mu} + \sigma^2 g(\hat{\mu}) \tag{17}$$

where $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a smooth function.
Check its risk,

$$\begin{aligned}
\mathbb{E}_\mu \|T(Y) - \mu\|^2 &= \mathbb{E}_\mu \left\| \hat{\mu} + \sigma^2 g(\hat{\mu}) - \mu \right\|^2 \\
&= \mathbb{E}_\mu \left(\|\mu - \hat{\mu}\|^2 + \sigma^4 \|g(\hat{\mu})\|^2 + 2\sigma^2 \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle \right) \\
&= \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^2 \mathbb{E}_\mu \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle
\end{aligned}$$

Stein lemma

Theorem (Stein identity)

Let $X \sim N(\theta, \sigma^2 I_d)$, *let* g *be a differentiable fucntion satisfying* $\mathbb{E}|\nabla g(X)| < \infty$. *Then*

$$\mathbb{E}\langle g(X), X - \theta\rangle = \sigma^2 \mathbb{E}\nabla \cdot g(X) \tag{18}$$

Stein lemma

Proof.

In the case when $d = 1$,

$$\mathbb{E}g(X)(X-\theta) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\infty}^{\infty} g(x)(x-\theta)e^{-(x-\theta)^2/(2\sigma^2)}\mathrm{d}x$$

Integration by parts,

$$\mathbb{E}g(X)(X-\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\left[-\sigma^2 g(x)e^{-(x-\theta)^2/(2\sigma^2)}\right]\Big|_{\infty}^{\infty} + \sigma^2\int_{\infty}^{\infty} g'(x)e^{-(x-\theta)^2/(2\sigma^2)}\mathrm{d}x$$

The first term is 0 by writing g as integral of g' and apply bounded convergence theorem, and the second term is $\sigma^2\mathbb{E}g'(X)$. □

James-Stein

$$T(Y) = \hat{\mu} + \sigma^2 g(\hat{\mu}) \tag{19}$$

$$R(\mu, T) = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^2 \mathbb{E}_\mu \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle \tag{20}$$

$$\mathbb{E}\langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle = \sigma^2 \mathbb{E} \nabla \cdot g(X) \tag{21}$$

$$\Downarrow$$

$$R(\mu, T) = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E} \nabla \cdot g(\hat{\mu}) \tag{22}$$

James-Stein

$$T(Y) = \hat{\mu} + \sigma^2 g(\hat{\mu}) \tag{19}$$

$$\begin{aligned} R(\mu, T) &= \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 \\ &\quad + 2\sigma^2 \mathbb{E}_\mu \langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle \end{aligned} \tag{20}$$

$$\mathbb{E}\langle \mu - \hat{\mu}, g(\hat{\mu}) \rangle = \sigma^2 \mathbb{E} \nabla \cdot g(X) \tag{21}$$

$$\Downarrow$$

$$R(\mu, T) = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E} \nabla \cdot g(\hat{\mu}) \tag{22}$$

Choice of g

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 + \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}\nabla\cdot g(\hat{\mu}) \tag{23}$$

$$g = \nabla \log \psi \tag{24}$$

ψ is a smooth, non-constant, positive function $\psi : \mathbb{R}^d \mapsto \mathbb{R}$.
We have the following equation

$$g(x) = \frac{\nabla\psi(x)}{\psi(x)} \tag{25}$$

$$\nabla\cdot g(x) = \frac{\Delta\psi(x)\psi(x) - \|\nabla\psi\|^2}{\psi^2(x)} = \frac{\Delta\psi(x)}{\psi(x)} - \|g(x)\|^2 \tag{26}$$

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 - \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}_\mu\frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{27}$$

Choice of g

$$\mathbb{E}_\mu \|T(Y)-\mu\|^2 = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 + \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E} \nabla \cdot g(\hat{\mu}) \tag{23}$$

$$g = \nabla \log \psi \tag{24}$$

ψ is a smooth, non-constant, positive function $\psi : \mathbb{R}^d \mapsto \mathbb{R}$.

We have the following equation

$$g(x) = \frac{\nabla \psi(x)}{\psi(x)} \tag{25}$$

$$\nabla \cdot g(x) = \frac{\Delta \psi(x) \psi(x) - \|\nabla \psi\|^2}{\psi^2(x)} = \frac{\Delta \psi(x)}{\psi(x)} - \|g(x)\|^2 \tag{26}$$

$$\mathbb{E}_\mu \|T(Y)-\mu\|^2 = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 - \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E}_\mu \frac{\Delta \psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{27}$$

Choice of g

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 + \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}\nabla\cdot g(\hat{\mu}) \tag{23}$$

$$g = \nabla\log\psi \tag{24}$$

ψ is a smooth, non-constant, positive function $\psi : \mathbb{R}^d \mapsto \mathbb{R}$.
We have the following equation

$$g(x) = \frac{\nabla\psi(x)}{\psi(x)} \tag{25}$$

$$\nabla\cdot g(x) = \frac{\Delta\psi(x)\psi(x) - \|\nabla\psi\|^2}{\psi^2(x)} = \frac{\Delta\psi(x)}{\psi(x)} - \|g(x)\|^2 \tag{26}$$

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 - \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}_\mu\frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{27}$$

Choice of g

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 + \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}\nabla\cdot g(\hat{\mu}) \tag{23}$$

$$g = \nabla \log \psi \tag{24}$$

ψ is a smooth, non-constant, positive function $\psi : \mathbb{R}^d \mapsto \mathbb{R}$.
We have the following equation

$$g(x) = \frac{\nabla\psi(x)}{\psi(x)} \tag{25}$$

$$\nabla\cdot g(x) = \frac{\Delta\psi(x)\psi(x) - \|\nabla\psi\|^2}{\psi^2(x)} = \frac{\Delta\psi(x)}{\psi(x)} - \|g(x)\|^2 \tag{26}$$

$$\mathbb{E}_\mu\|T(Y)-\mu\|^2 = \mathbb{E}_\mu\|\mu-\hat{\mu}\|^2 - \sigma^4\mathbb{E}_\mu\|g(\hat{\mu})\|^2 + 2\sigma^4\mathbb{E}_\mu\frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{27}$$

Choice of g

$$\mathbb{E}_\mu \|T(Y) - \mu\|^2 = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 - \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E}_\mu \frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{28}$$

1. $\sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 > 0$
2. $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ that is smooth, positive
3. $\Delta\psi = 0$

In another words, we are interested to find a smooth harmonic function $\psi \neq$constant and is positive.

Choice of g

$$\mathbb{E}_\mu \|T(Y)-\mu\|^2 = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 - \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E}_\mu \frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{28}$$

1. $\sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 > 0$
2. $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ that is smooth, positive
3. $\Delta\psi = 0$

In another words, we are interested to find a smooth harmonic function $\psi \neq$constant and is positive.

Choice of g

$$\mathbb{E}_\mu \|T(Y) - \mu\|^2 = \mathbb{E}_\mu \|\mu - \hat{\mu}\|^2 - \sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 + 2\sigma^4 \mathbb{E}_\mu \frac{\Delta\psi(\hat{\mu})}{\psi(\hat{\mu})} \tag{28}$$

1. $\sigma^4 \mathbb{E}_\mu \|g(\hat{\mu})\|^2 > 0$
2. $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ that is smooth, positive
3. $\Delta\psi = 0$

In another words, we are interested to find a smooth harmonic function $\psi \neq$constant and is positive.

Choice of g

When $d \geq 3$, such function ψ exists and satisfies for some $\mu \in \mathbb{R}^d$,

$$\mathbb{E}_\mu \|g(\hat{\mu})\| > 0. \tag{29}$$

Explicit form of JS-estimator

$$\psi(x) = \|x\|^{-(d-2)}. \tag{30}$$

$$T_{JS}(Y) = \hat{\mu} - (d-2)\frac{\sigma^2 \hat{\mu}}{\|\hat{\mu}\|} = \hat{\mu}\left(1 - \frac{\sigma^2(d-2)}{\|\hat{\mu}\|^2}\right) \tag{31}$$

for $d > 2$.

Explicit form of JS-estimator

$$\psi(x) = \|x\|^{-(d-2)}. \tag{30}$$

$$T_{JS}(Y) = \hat{\mu} - (d-2)\frac{\sigma^2 \hat{\mu}}{\|\hat{\mu}\|} = \hat{\mu}\left(1 - \frac{\sigma^2(d-2)}{\|\hat{\mu}\|^2}\right) \tag{31}$$

for $d > 2$.

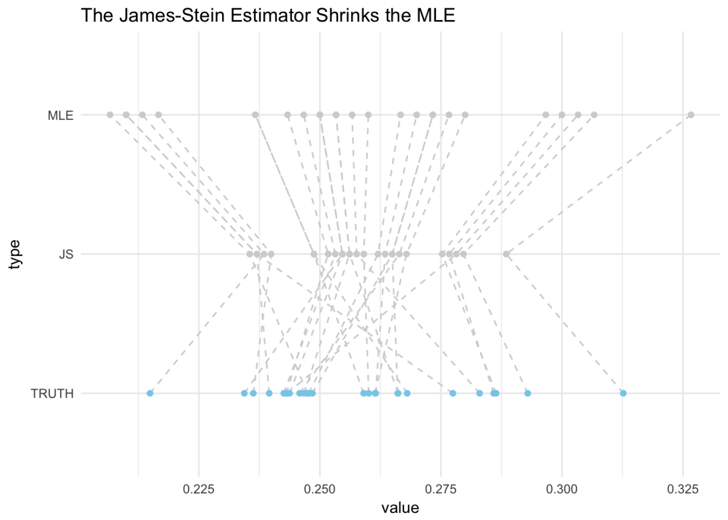
Risk of Jame-Stein

$$\mathbb{E}_\mu \|T_{JS} - \mu\|^2 = \sigma^2 d - \sigma^2 (d-2)^2 \mathbb{E} \frac{1}{\chi^2_{d, \frac{\|\mu\|}{\sigma}}} \tag{32}$$

where $\chi^2_{d, \frac{\|\mu\|}{\sigma}}$ is the non-central Chi-square distribution.

Projection estimator is indeed the best estimator.

Simulation


The James-Stein Estimator Shrinks the MLE
MLE
JS
TRUTH
type
0.225
0.250
0.275
0.300
0.325
value

Reference

- Stein, C. Inadmissibility of the usual estimator for the mean of a multivariate distribution. (1956).
- Tsybakov, A. *Introduction to non-parametric estimation.* (Springer, 2003).
- Wichura, M. J. *The coordinate-free approach to linear models.* (Cambridge Press, 2006).

The End