High dimensional statistics

Part 2: Shrinkage and Stein's phenomenon

Hugo Richard (research.hugo.richard@gmail.com)

ENPC - February 2022

A small recap of Stein's phenomenon

Available on previous slides

- Gaussian sequence model: $\mathbf{y} \in \mathbb{R}^n, \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$, we want to estimate $\boldsymbol{\mu}$.
- Linear estimators $\hat{\mu}(\mathbf{y}) = A\mathbf{y}$, ($A = I_n$ is the ML solution)
- Error $\mathbb{E}_{\mathbf{y}}[\|\hat{\mu}(\mathbf{y}) - \boldsymbol{\mu}\|^2]$
- Akaike information criterion (AIC) for linear estimators
- Taking $A = sI_n$ and minimizing with respect to s gives $\hat{\mu}(\mathbf{y}) = (1 - \frac{n\sigma^2}{\|\mathbf{y}\|^2})\mathbf{y}$

A small recap of Stein's phenomenon

On the board last time

- AIC for non-linear estimators = Stein's unbiased risk estimate
- Apply it to estimators of the form $\hat{\mu}(\mathbf{y}) = (1 - \frac{n\sigma^2}{\|\mathbf{y}\|^2})\mathbf{y}$
- Conclude that MLE is not admissible

Let's re-do these steps.

Gaussian integration by part

Lemma (Scalar to scalar function, standard Gaussian)
if $z \sim \mathcal{N}(0, 1)$ *and* $f : \mathcal{R} \to \mathcal{R}$ *is differentiable such that* $\mathbb{E}[|f(z)|] < \infty$ *and* $\mathbb{E}[|f'(z)|] < \infty$ *then* $\mathbb{E}[zf(z)] = \mathbb{E}[f'(z)]$.

Gaussian integration by part

Lemma (Scalar to scalar function, standard Gaussian)
if $z \sim \mathcal{N}(0, 1)$ and $f : \mathcal{R} \to \mathcal{R}$ is differentiable such that $\mathbb{E}[|f(z)|] < \infty$ and $\mathbb{E}[|f'(z)|] < \infty$ then $\mathbb{E}[zf(z)] = \mathbb{E}[f'(z)]$.

Proof.

$$\begin{aligned}
\mathbb{E}[zf(z)1(z > 0)] &= \int_0^\infty z\phi(z)\left(\int_0^z f'(t)dt\right)dz && (1)\\
&= \int_0^\infty f'(t)\left(\int_t^\infty z\phi(z)dz\right)dt && (2)\\
&= \int_0^\infty f'(t)\phi(t)dt && (3)\\
&= \mathbb{E}[1(z > 0)f'(z)] && (4)
\end{aligned}$$

Do the same with $z \leq 0$. □

Gaussian integration by part

Lemma (Scalar to scalar function, scaled Gaussian)
if $z \sim \mathcal{N}(0, \sigma^2)$ *and* $f : \mathcal{R} \to \mathcal{R}$ *is differentiable such that* $\mathbb{E}[|f(z)|] < \infty$ *and* $\mathbb{E}[|f'(z)|] < \infty$ *then*
$\mathbb{E}[zf(z)] = \sigma^2 \mathbb{E}[f'(z)]$.

Gaussian integration by part

Lemma (Scalar to scalar function, scaled Gaussian)
if $z \sim \mathcal{N}(0, \sigma^2)$ *and* $f : \mathcal{R} \to \mathcal{R}$ *is differentiable such that* $\mathbb{E}[|f(z)|] < \infty$ *and* $\mathbb{E}[|f'(z)|] < \infty$ *then* $\mathbb{E}[zf(z)] = \sigma^2 \mathbb{E}[f'(z)]$.

Proof.
For $\sigma \neq 1$, take $g = z/\sigma$ and $\tilde{f}(u) = f(\sigma u)$

$$\mathbb{E}[g\tilde{f}(g)] = \mathbb{E}[\tilde{f}'(g)] \quad (5)$$

$$\iff \mathbb{E}[zf(z)] = \sigma^2 \mathbb{E}[f'(z)] \quad (6)$$

□

Gaussian integration by part

Lemma (vector to scalar function)
if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$ *and* $g : \mathcal{R}^n \rightarrow \mathcal{R}$ *is differentiable, integrable and so are the partial derivatives of* g*, then:*
$\mathbb{E}[(y_i - \mu_i)g(\mathbf{y})] = \sigma^2 \mathbb{E}[\frac{\partial g}{\partial y_i}(\mathbf{y})]$

Gaussian integration by part

Lemma (vector to scalar function)
if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$ *and* $g : \mathcal{R}^n \rightarrow \mathcal{R}$ *is differentiable, integrable and so are the partial derivatives of* g*, then:*
$\mathbb{E}[(y_i - \mu_i)g(\mathbf{y})] = \sigma^2 \mathbb{E}[\frac{\partial g}{\partial y_i}(\mathbf{y})]$

Proof.
Take $z = (y_1 - \mu_1)$ and $f(z) = g(\mu_1 + z, y_2, \ldots, y_n)$.

$$\mathbb{E}_z[zf(z)|y_{2:n}] = \sigma^2 \mathbb{E}_z[f'(z)|y_{2:n}] \tag{7}$$

$$\Rightarrow \mathbb{E}_{y_1}[(y_1 - \mu_1)g(\mathbf{y})|y_{2:n}] = \sigma^2 \mathbb{E}_{y_1}[\frac{\partial g}{\partial y_1}(\mathbf{y})|y_{2:n}] \tag{8}$$

$$\Rightarrow \mathbb{E}_{y_{2:n}}[\mathbb{E}_{y_1}[(y_1 - \mu_1)g(\mathbf{y})|y_{2:n}]] = \sigma^2 \mathbb{E}_{y_{2:n}}[\mathbb{E}_{y_1}[\frac{\partial g}{\partial y_1}(\mathbf{y})|y_{2:n}]]$$

□

Gaussian integration by part

Lemma (vector to vector function)
if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$ *and* $\mathbf{f} : \mathcal{R}^n \to \mathcal{R}^n$ *is such that all coordinate functions are integrable with integrable partial derivatives then*
$\mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{f}(\mathbf{y})] = \sigma^2 \mathbb{E}[\sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y})]$

Gaussian integration by part

Lemma (vector to vector function)
if $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$ *and* $\mathbf{f} : \mathcal{R}^n \rightarrow \mathcal{R}^n$ *is such that all coordinate functions are integrable with integrable partial derivatives then* $\mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{f}(\mathbf{y})] = \sigma^2 \mathbb{E}[\sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y})]$

Proof.

$$\mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{f}(\mathbf{y})] = \sum_{i=1}^n \mathbb{E}[(y_i - \mu_i) f_i(\mathbf{y})] \tag{9}$$

$$= \sum_{i=1}^n \sigma^2 \mathbb{E}[\frac{\partial f_i}{\partial y_i}(\mathbf{y})] \tag{10}$$

□

Stein's unbiased risk estimate

Theorem
Take $\hat{\boldsymbol{\mu}} = \mathbf{f}(\mathbf{y})$ *and* $\hat{r}_f = \|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 + 2\sigma^2 \sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y}) - n\sigma^2$,
then $\mathbb{E}[\hat{r}_f] = \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2]$

Proof.

$$\mathbb{E}[\|\hat{\boldsymbol{\mu}} - \mathbf{y} + \mathbf{y} - \boldsymbol{\mu}\|^2] \tag{11}$$

$$= \mathbb{E}[\|\hat{\boldsymbol{\mu}} - \mathbf{y}\|^2 + \|\mathbf{y} - \boldsymbol{\mu}\|^2 + 2\langle \hat{\boldsymbol{\mu}} - \mathbf{y} | \mathbf{y} - \boldsymbol{\mu}\rangle] \tag{12}$$

$$= \mathbb{E}[\|\mathbf{f}(\mathbf{y}) - \mathbf{y}\|^2 + \|\boldsymbol{\epsilon}\|^2 + 2\langle \hat{\boldsymbol{\mu}} | \epsilon\rangle - 2\langle \mathbf{y} | \epsilon\rangle] \tag{13}$$

$$= \mathbb{E}[\|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 + n\sigma^2 + 2\sigma^2 \sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y}) - 2n\sigma^2] \tag{14}$$

□

MLE is not admissible

Theorem

Take $\mathbf{f}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2})\mathbf{y}$ *Then*

$\mathbb{E}[\|\hat{\mu} - \mu\|^2] = \mathbb{E}[\hat{r}_f] = n\sigma^2 - \mathbb{E}[\frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - q - 4)]$

Proof.

- $\|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 = \|\frac{\sigma^2 q}{\|\mathbf{y}\|^2} y\|^2 = \frac{\sigma^4 q^2}{\|\mathbf{y}\|^2}$

□

MLE is not admissible

Theorem

Take $\mathbf{f}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2})\mathbf{y}$ *Then*

$$\mathbb{E}[\|\hat{\mu} - \mu\|^2] = \mathbb{E}[\hat{r}_f] = n\sigma^2 - \mathbb{E}[\frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - q - 4)]$$

Proof.

- $\|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 = \|\frac{\sigma^2 q}{\|\mathbf{y}\|^2} y\|^2 = \frac{\sigma^4 q^2}{\|\mathbf{y}\|^2}$
- $\frac{\partial f_i}{\partial y_i}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}^2\|}) + \frac{2\sigma^2 q y_i^2}{\|\mathbf{y}\|^4}$

□

MLE is not admissible

Theorem

Take $\mathbf{f}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2})\mathbf{y}$ *Then*

$$\mathbb{E}[\|\hat{\mu} - \mu\|^2] = \mathbb{E}[\hat{r}_f] = n\sigma^2 - \mathbb{E}[\frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - q - 4)]$$

Proof.

- $\|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 = \|\frac{\sigma^2 q}{\|\mathbf{y}\|^2} y\|^2 = \frac{\sigma^4 q^2}{\|\mathbf{y}\|^2}$
- $\frac{\partial f_i}{\partial y_i}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}^2\|}) + \frac{2\sigma^2 q y_i^2}{\|\mathbf{y}\|^4}$
- $2\sigma^2 \sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y}) = 2n\sigma^2(1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2}) + \frac{4\sigma^2 q}{\|\mathbf{y}\|^2} = 2n\sigma^2 - \frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - 4)$

□

MLE is not admissible

Theorem

Take $\mathbf{f}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2})\mathbf{y}$ *Then*

$\mathbb{E}[\|\hat{\mu} - \mu\|^2] = \mathbb{E}[\hat{r}_f] = n\sigma^2 - \mathbb{E}[\frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - q - 4)]$

Proof.

- $\|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 = \|\frac{\sigma^2 q}{\|\mathbf{y}\|^2} y\|^2 = \frac{\sigma^4 q^2}{\|\mathbf{y}\|^2}$
- $\frac{\partial f_i}{\partial y_i}(\mathbf{y}) = (1 - \frac{\sigma^2 q}{\|\mathbf{y}^2\|}) + \frac{2\sigma^2 q y_i^2}{\|\mathbf{y}\|^4}$
- $2\sigma^2 \sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y}) = 2n\sigma^2(1 - \frac{\sigma^2 q}{\|\mathbf{y}\|^2}) + \frac{4\sigma^2 q}{\|\mathbf{y}\|^2} = 2n\sigma^2 - \frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - 4)$
- $\hat{r}_f = \|\mathbf{y} - \mathbf{f}(\mathbf{y})\|^2 + 2\sigma^2 \sum_{i=1}^n \frac{\partial f_i}{\partial y_i}(\mathbf{y}) - n\sigma^2 = n\sigma^2 - \frac{\sigma^4 q}{\|\mathbf{y}\|^2}(2n - q - 4)$

□