Constrained denoising, empirical Bayes, and optimal transport

Adam Quinn Jaffe

With Nikolaos Ignatiadis and Bodhisattva Sen

Model for data:

$$\Theta \sim G \qquad \text{and } Z = \Theta + \varepsilon \qquad \text{where } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

where G is unknown distribution, and Σ is known likelihood variance.

Let $(\Theta_1, Z_1), \ldots (\Theta_n, Z_n)$ be i.i.d pairs from above.

Goal is *denoising*, estimating latent variables $\Theta_1, \ldots, \Theta_n$ from the observed variables $Z_1, \ldots, Z_n$

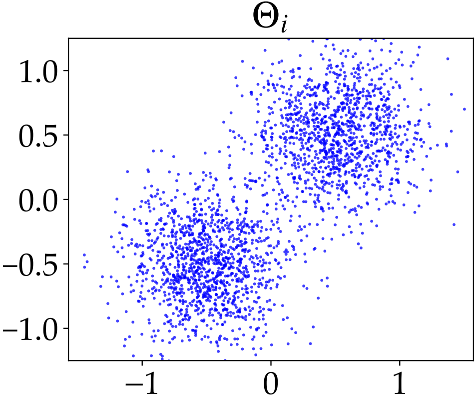
Model for data:

$$\Theta \sim G \qquad \text{and } Z = \Theta + \varepsilon \qquad \text{where } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

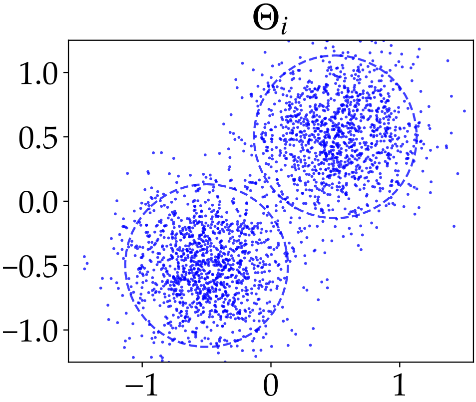
where G is unknown distribution, and Σ is known likelihood variance.

Let $(\Theta_1, Z_1), \dots (\Theta_n, Z_n)$ be i.i.d pairs from above.

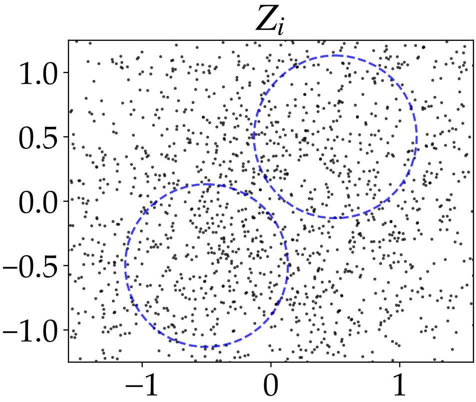
Goal is *denoising*, estimating latent variables $\Theta_1, \dots, \Theta_n$ from the observed variables $Z_1, \dots, Z_n$



Θi
1.0
0.5
0.0
−0.5
−1.0
−1
0
1



Θ_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1


Z_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

Minimize risk:

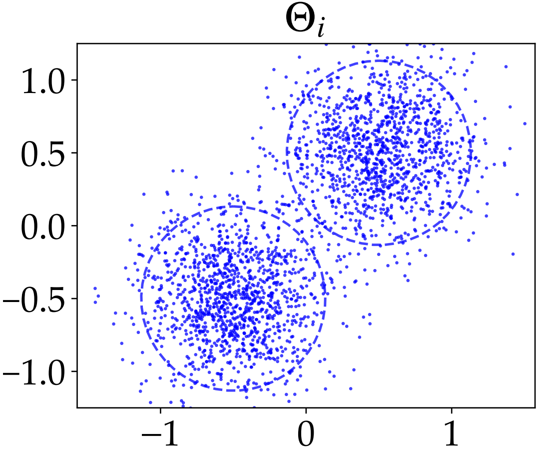
$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \end{cases}$$

Solution is the posterior mean:

$$\delta_{\mathcal{B}}(z) = \mathbb{E}[\Theta \mid Z = z].$$

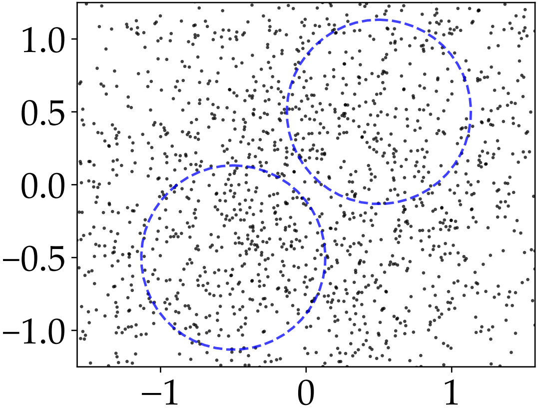
Although $\delta_{\mathcal{B}}$ depends on G, we can try to approximate $\hat{\delta}_{\mathcal{B}}$ empirically.

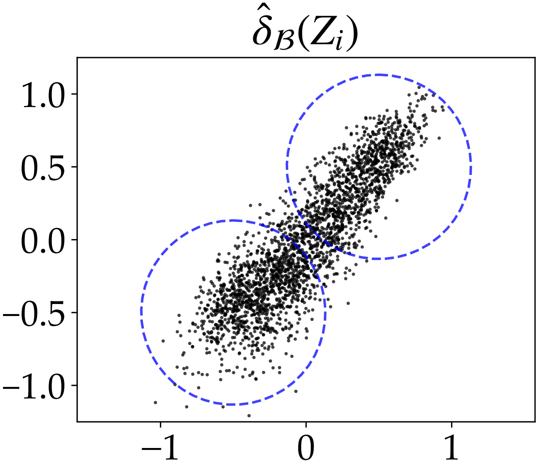
Any $\delta : \mathbb{R}^m \to \mathbb{R}^m$ is called a *denoiser*, $\delta_{\mathcal{B}}$ is *oracle Bayes denoiser*, and $\hat{\delta}_{\mathcal{B}}$ is *empirical Bayes denoiser* (Robbins 1956, Efron 2019).



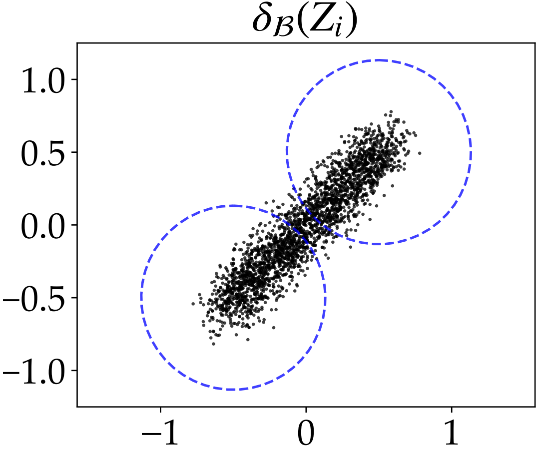
Θ_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

Z_i





$\hat{\delta}_{\mathcal{B}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1


$\delta_{\mathcal{B}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

Shrinkage in $\delta_{\mathcal{B}}$ and $\hat{\delta}_{\mathcal{B}}$:

$$\mathrm{Cov}(\Theta) = \mathrm{Cov}(\delta_{\mathcal{B}}(Z)) + \mathbb{E}[\mathrm{Cov}(\Theta \mid Z)] \succeq \mathrm{Cov}(\delta_{\mathcal{B}}(Z)).$$

Sometimes want distributions of $\hat{\delta}_{\mathcal{B}}(Z_1), \ldots, \hat{\delta}_{\mathcal{B}}(Z_n)$ and $\Theta_1, \ldots, \Theta_n$ to be similar (Louis 1984, Ghosh 1992, Ghosh-Maiti 1999, Loredo 2007).

Not guaranteed from $\hat{\delta}_{\mathcal{B}}(Z_1), \ldots, \hat{\delta}_{\mathcal{B}}(Z_n)$ estimating $\Theta_1, \ldots, \Theta_n$ well!

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \end{cases}$$

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \\ \text{with} & \delta(Z)\stackrel{\mathcal{D}}{=}\Theta \end{cases}$$

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)]=\mathbb{E}[\Theta] \\ \text{and} & \mathrm{Cov}(\delta(Z))=\mathrm{Cov}(\Theta) \end{cases}$$

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \\ \text{with} & \cdots \end{cases}$$

Add constraints:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \cdots \end{cases}$$

Oracle solution?

How to approximate empirically?

Some previous work:

Oracle Bayes and empirical Bayes for variance constraints:

- Gaussian G in dimension $m = 1$ (Louis 1984)
- General G in dimension $m = 1$ (Ghosh 1992)
- Gaussian G in general dimension m (Ghosh-Maiti 1999)

Oracle Bayes for distribution constraints:

- Calculation of excess risk (Freirich 2021)
- Characterization of solutions (García-Trillos-Sen 2024)

I. Problem Statement

II. Oracle Constrained Bayes

III. Empirical Constrained Bayes

IV. Astronomy Application

II. Oracle Constrained Bayes

General constrained denoising problem:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta(Z_i) - \Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \cdots \end{cases}$$

Rewrite the objective:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta(Z_i) - \Theta_i\|^2\right]$$

$$= \underbrace{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2\right]}_{\text{excess risk}} + \underbrace{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta_{\mathcal{B}}(Z_i) - \Theta_i\|^2\right]}_{\text{Bayes risk}}$$

Distribution-constrained denoising problem:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\delta(Z_i) - \Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \delta(Z) \stackrel{\mathcal{D}}{=} \Theta \end{cases} \tag{$\mathcal{DCB}$}$$

$$\cong \begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\delta(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \delta(Z) \stackrel{\mathcal{D}}{=} \Theta \end{cases}$$

Approximately looks like a Monge transport problem from distribution of $\delta_{\mathcal{B}}(Z)$ to distribution of Θ, but with non-standard cost function.

Existence and uniqueness of solutions? How to solve in practice?

Theorem (García-Trillos-Sen, 2024)

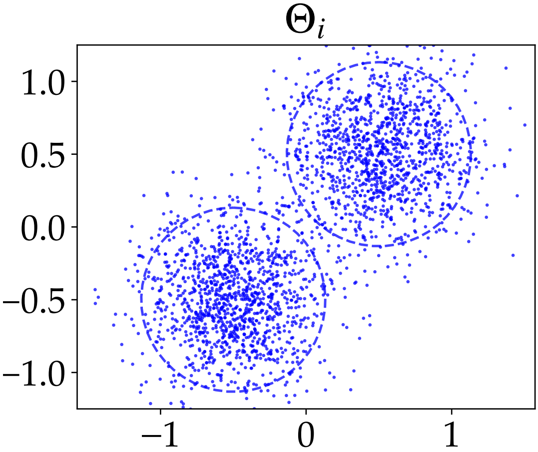
Under suitable regularity conditions, if F and G denote the distributions of Z and Θ, respectively, then the problem

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

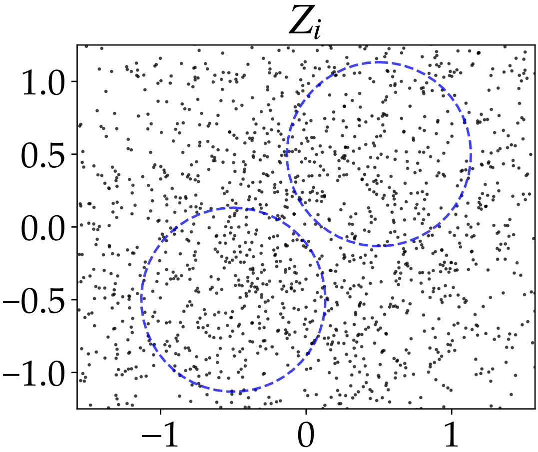
admits a unique solution, this solution is concentrated on the graph of a function $\delta_{\mathcal{DCB}} : \mathbb{R}^m \to \mathbb{R}^m$ which is the unique solution to the problem

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \delta(Z) \overset{\mathcal{D}}{=} \Theta, \end{cases} \qquad (\mathcal{DCB})$$

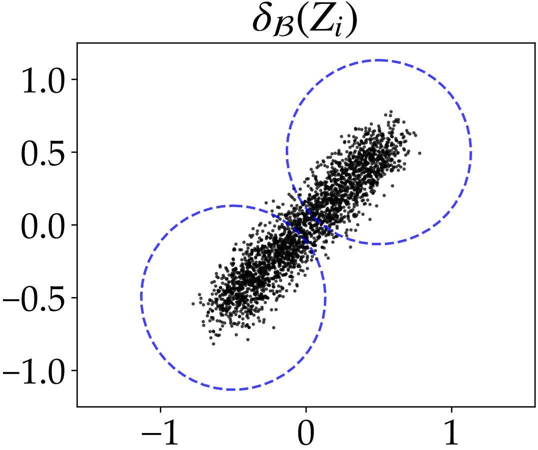
and we have $\delta_{\mathcal{DCB}} = \nabla\phi \circ \delta_{\mathcal{B}}$ for some convex function $\phi : \mathbb{R}^m \to \mathbb{R}$.



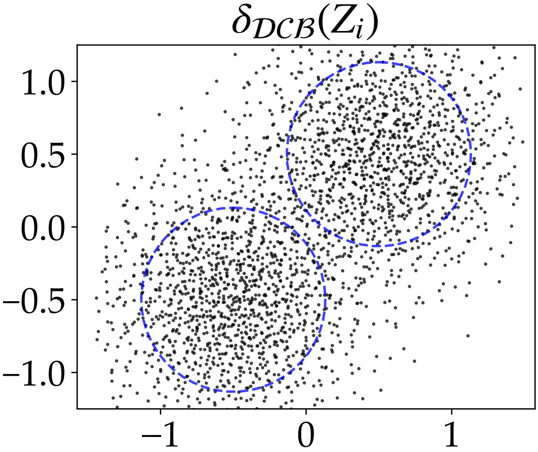
Θ_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1



Z_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1



$\delta_{\mathcal{B}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1


$\delta_{\mathcal{DCB}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

Variance-constrained denoising problem:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta(Z_i) - \Theta_i\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Cov}(\delta(Z)) = \mathrm{Cov}(\Theta) \end{cases} \tag{$\mathcal{VCB}$}$$

$$\cong \begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \|\delta(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Cov}(\delta(Z)) = \mathrm{Cov}(\Theta) \end{cases}$$

Approximately looks like a Monge transport problem, but with non-standard cost function and one marginal partially-specified.

Gaussianization trick:

$$\text{replace} \qquad \begin{cases} \delta_{\mathcal{B}}(Z) \leftarrow \mathcal{N}\left(\mathbb{E}[\delta_{\mathcal{B}}(Z)], \mathrm{Cov}(\delta_{\mathcal{B}}(Z))\right) \\ \delta(Z) \leftarrow \mathcal{N}\left(\mathbb{E}[\Theta], \mathrm{Cov}(\Theta)\right). \end{cases}$$

Then ($\mathcal{VCB}$) reduces to a Monge transport problem with non-standard cost function and with Gaussian marginals.

For $a, b \in \mathbb{R}^m$ and $A, B \succ 0$, consider

$$\begin{cases} \text{minimize} & \int_{\mathbb{R}^m} \|x - x'\|^2 \mathrm{d}\pi(x, x') \\ \text{over} & \pi \in \Gamma(\mathcal{N}(a, A), \mathcal{N}(b, B)) \end{cases}$$

Unique solution (Olkin-Pukelsheim 1982) concentrates on graph of

$$\delta(x) = \mathbf{t}_A^B(x - a) + b$$

where

$$\mathbf{t}_A^B = A^{-1/2}(A^{1/2}BA^{1/2})^{1/2}A^{-1/2}.$$

If A, B commute, then $\mathbf{t}_A^B = B^{1/2}A^{-1/2}$.

Theorem (AQJ-Ignatiadis-Sen, 2025)

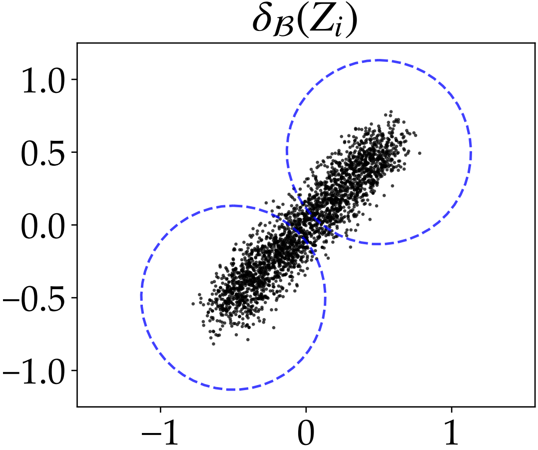
Under suitable regularity conditions, the problem

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z) - \Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \text{Cov}(\delta(Z)) = \text{Cov}(\Theta) \end{cases}$$

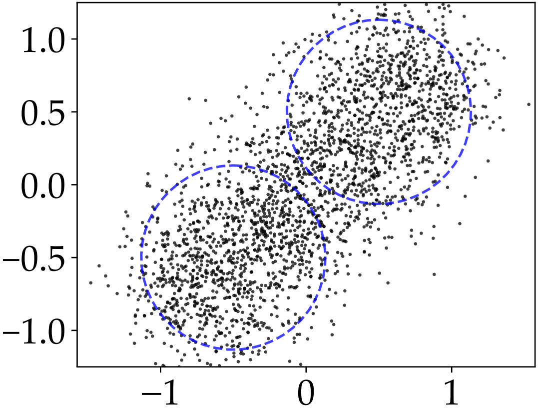
admits a unique solution given by

$$\delta_{\mathcal{VCB}}(\,\cdot\,) = \mathbf{t}^{\text{Cov}(\Theta)}_{\text{Cov}(\delta_{\mathcal{B}}(Z))}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta].$$

Extends some known special cases (Louis 1984, Ghosh 1992, Ghosh-Maiti 1999).


$\delta_{\mathcal{B}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

$\delta_{\mathcal{VCB}}(Z_i)$



III. Empirical Constrained Bayes

Many known methods for achieving small value of

$$\frac{1}{n}\sum_{i=1}^{n}\|\hat{\delta}_{\mathcal{B}}(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2$$

when n large (Brown-Greenshtein 2009, Jiang-Zhang 2009, Soloff et al. 2024, Shen-Wu 2025, ...).

Can we do similar for $\hat{\delta}_{\mathcal{DCB}}$ and $\hat{\delta}_{\mathcal{VCB}}$?

Note we don't seek small overall risk, just comparable risk to the oracle.

Suffices to take an approximation $\hat{\delta}_{\mathcal{B}}$ of $\delta_{\mathcal{B}}$ and apply some further transformations to it!

Distribution constrained denoising?

Kantorovich-type problem:

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

$$\approx \begin{cases} \text{minimize} & \int \|\hat{\delta}_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(\bar{F}_n, \hat{G}_n) \end{cases}$$

Need:

Distribution constrained denoising?

Kantorovich-type problem:

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

$$\approx \begin{cases} \text{minimize} & \int \|\hat{\delta}_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(\bar{F}_n, \hat{G}_n) \end{cases}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$,

Distribution constrained denoising?

Kantorovich-type problem:

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

$$\approx \begin{cases} \text{minimize} & \int \|\hat{\delta}_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(\bar{F}_n, \hat{G}_n) \end{cases}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\hat{G}_n$ approximating G,

Distribution constrained denoising?

Kantorovich-type problem:

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z, \theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

$$\approx \begin{cases} \text{minimize} & \int \|\hat{\delta}_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z, \theta) \\ \text{over} & \pi \in \Gamma(\bar{F}_n, \hat{G}_n) \end{cases}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\hat{G}_n$ approximating G, and some $\bar{F}_n$ approximating F.

Distribution constrained denoising?

Kantorovich-type problem:

$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

$$\approx \begin{cases} \text{minimize} & \int \|\hat{\delta}_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(\bar{F}_n, \hat{G}_n) \end{cases}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\hat{G}_n$ approximating G, and some $\bar{F}_n$ approximating F.

procedure DistributionConstrainedEB$(Z_1, \dots, Z_n)$

input: samples $Z_1, \dots, Z_n$
output: denoising function $\hat{\delta}_{\mathcal{DCB}} : \{Z_1, \dots, Z_n\} \to \mathbb{R}^m$

$\hat{\delta}_{\mathcal{B}}(\,\cdot\,) \leftarrow$ approximation of $\delta_{\mathcal{B}}(\,\cdot\,)$
$\hat{G}_n \leftarrow$ approximation of G

$\hat{c}_n(Z_i, \eta) \leftarrow \|\hat{\delta}_{\mathcal{B}}(Z_i) - \eta\|^2$ for all $1 \le i \le n$ and $\eta \in \mathbb{R}^m$

$\hat{\pi}_{\mathcal{DCB}} \leftarrow$ **minimize** $\int_{\mathbb{R}^m \times \mathbb{R}^m} \hat{c}_n(z, \eta) \mathrm{d}\pi(z, \eta)$
over probability measures $\pi \in \mathcal{P}(\mathbb{R}^m \times \mathbb{R}^m)$
with $\pi(\{Z_i\} \times \mathbb{R}^m) = \frac{1}{n}$ for all $1 \le i \le n$
and $\pi(\mathbb{R}^m \times \mathrm{d}\eta) = \hat{G}_n(\mathrm{d}\eta)$

$\hat{\delta}_{\mathcal{DCB}}(Z_i) \leftarrow \int_{\mathbb{R}^m} \eta \, \mathrm{d}\hat{\pi}_{\mathcal{DCB}}(\eta \,|\, Z_i)$ for all $1 \le i \le n$

return $\hat{\delta}_{\mathcal{DCB}}$

Theorem (AQJ-Ignatiadis-Sen 2025)

Suppose $\hat{\delta}_{\mathcal{B}} : \mathbb{R}^m \to \mathbb{R}^m$ *satisfies*

$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{\delta}_{\mathcal{B}}(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2 = O_{\mathbb{P}}(\alpha_n),$$

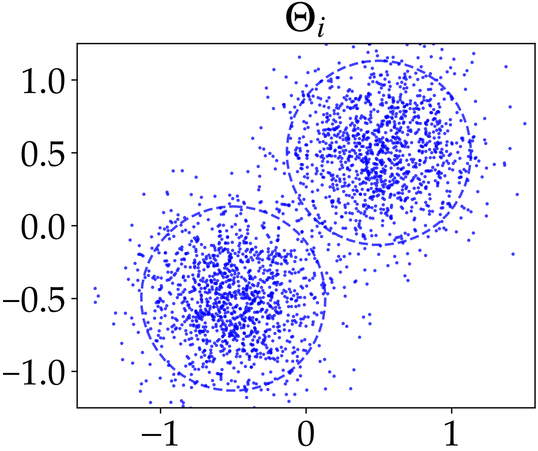
and that $\hat{G}_n$ *satisfies*

$$W_2(\hat{G}_n, G) = O_{\mathbb{P}}(\beta_n),$$

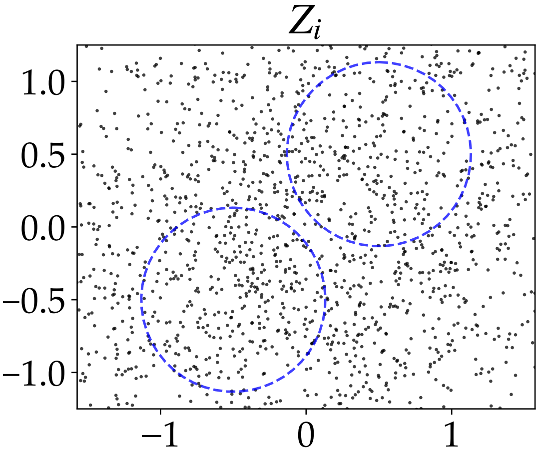
as $n \to \infty$. *Then the denoiser* $\hat{\delta}_{\mathcal{DCB}}$ *described before satisfies*

$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{\delta}_{\mathcal{DCB}}(Z_i) - \delta_{\mathcal{DCB}}(Z_i)\|^2 = O_{\mathbb{P}}\left(\alpha_n^{1/2} \vee \beta_n\right).$$

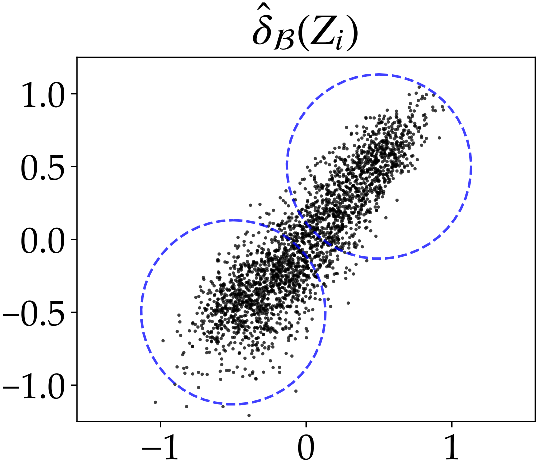
Rate of convergence of $\hat{\delta}_{\mathcal{DCB}}$ typically dominated by the slow rate of convergence of nonparametric deconvolution (Fan 1991).

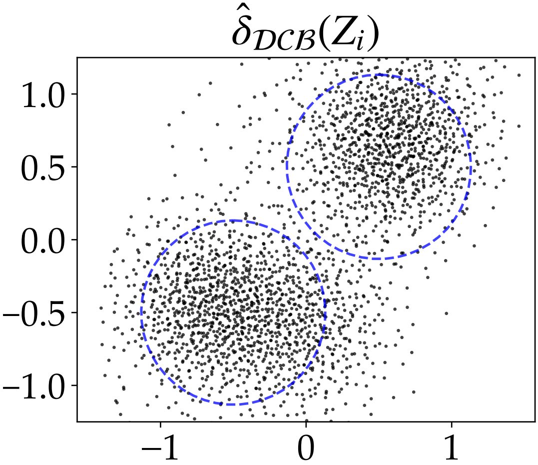


Θ_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1



Z_i
1.0
0.5
0.0
−0.5
−1.0
−1
0
1


$\hat{\delta}_{\mathcal{B}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1


$\hat{\delta}_{\mathcal{DCB}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

Variance constrained denoising?

Explicit formula for optimal denoiser:

$$\mathbf{t}^{\mathrm{Cov}(\Theta)}_{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

$$\approx \mathbf{t}^{\widehat{\mathrm{Cov}(\Theta)}}_{\mathrm{Cov}(\widehat{\delta_{\mathcal{B}}(Z)})}(\hat{\delta}_{\mathcal{B}}(\,\cdot\,) - \widehat{\mathbb{E}[\Theta]}) + \widehat{\mathbb{E}[\Theta]}$$

Need:

Variance constrained denoising?

Explicit formula for optimal denoiser:

$$\mathbf{t}_{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}^{\mathrm{Cov}(\Theta)}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

$$\approx \mathbf{t}_{\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}}^{\widehat{\mathrm{Cov}(\Theta)}}(\hat{\delta}_{\mathcal{B}}(\,\cdot\,) - \widehat{\mathbb{E}[\Theta]}) + \widehat{\mathbb{E}[\Theta]}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$,

Variance constrained denoising?

Explicit formula for optimal denoiser:

$$\mathbf{t}^{\mathrm{Cov}(\Theta)}_{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

$$\approx \mathbf{t}^{\widehat{\mathrm{Cov}(\Theta)}}_{\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}}(\hat{\delta}_{\mathcal{B}}(\,\cdot\,) - \widehat{\mathbb{E}[\Theta]}) + \widehat{\mathbb{E}[\Theta]}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}$ approximating $\mathrm{Cov}(\delta_{\mathcal{B}}(Z))$,

Variance constrained denoising?

Explicit formula for optimal denoiser:

$$\mathbf{t}_{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}^{\mathrm{Cov}(\Theta)}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

$$\approx \mathbf{t}_{\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}}^{\widehat{\mathrm{Cov}(\Theta)}}(\hat{\delta}_{\mathcal{B}}(\,\cdot\,) - \widehat{\mathbb{E}[\Theta]}) + \widehat{\mathbb{E}[\Theta]}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}$ approximating $\mathrm{Cov}(\delta_{\mathcal{B}}(Z))$, some $\widehat{\mathrm{Cov}(\Theta)}$ approximating $\mathrm{Cov}(\Theta)$, and some $\widehat{\mathbb{E}[\Theta]}$ approximating $\mathbb{E}[\Theta]$.

Variance constrained denoising?

Explicit formula for optimal denoiser:

$$\mathbf{t}^{\mathrm{Cov}(\Theta)}_{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}(\delta_{\mathcal{B}}(\,\cdot\,) - \mathbb{E}[\Theta]) + \mathbb{E}[\Theta]$$

$$\approx \mathbf{t}^{\widehat{\mathrm{Cov}(\Theta)}}_{\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}}(\hat{\delta}_{\mathcal{B}}(\,\cdot\,) - \widehat{\mathbb{E}[\Theta]}) + \widehat{\mathbb{E}[\Theta]}$$

Need: some $\hat{\delta}_{\mathcal{B}}$ approximating $\delta_{\mathcal{B}}$, some $\widehat{\mathrm{Cov}(\delta_{\mathcal{B}}(Z))}$ approximating $\mathrm{Cov}(\delta_{\mathcal{B}}(Z))$, some $\widehat{\mathrm{Cov}(\Theta)}$ approximating $\mathrm{Cov}(\Theta)$, and some $\widehat{\mathbb{E}[\Theta]}$ approximating $\mathbb{E}[\Theta]$.

procedure VarianceConstrainedEB$(Z_1, \ldots, Z_n)$

input: samples $Z_1, \ldots, Z_n \in \mathbb{R}^m$
output: denoising function $\hat{\delta}_{\mathcal{VCB}} : \mathbb{R}^m \to \mathbb{R}^m$

$\hat{\delta}_{\mathcal{B}}(\cdot) \leftarrow$ approximation of $\delta_{\mathcal{B}}(\cdot)$

$\hat{M} \leftarrow$ sample covariance matrix of $\hat{\delta}_{\mathcal{B}}(Z_1), \ldots \hat{\delta}_{\mathcal{B}}(Z_n)$
$\hat{\mu} \leftarrow$ sample mean of $Z_1, \ldots, Z_n$
$\hat{S} \leftarrow$ sample covariance matrix of $Z_1, \ldots, Z_n$
$\hat{A} \leftarrow (\hat{S} - \Sigma)_+$

$\hat{\mathbf{t}} \leftarrow \hat{M}^{-1/2}(\hat{M}^{1/2}\hat{A}\hat{M}^{1/2})^{1/2}\hat{M}^{-1/2}$
$\hat{\delta}_{\mathcal{VCB}}(\cdot) \leftarrow \hat{\mathbf{t}}(\hat{\delta}_{\mathcal{B}}(\cdot) - \hat{\mu}) + \hat{\mu}$

return $\hat{\delta}_{\mathcal{VCB}}$

Theorem (AQJ-Ignatiadis-Sen 2025)

Suppose $\hat{\delta}_{\mathcal{B}} : \mathbb{R}^m \to \mathbb{R}^m$ *satisfies*

$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{\delta}_{\mathcal{B}}(Z_i) - \delta_{\mathcal{B}}(Z_i)\|^2 = O_{\mathbb{P}}(\alpha_n),$$

as $n \to \infty$. *Then the denoiser* $\hat{\delta}_{\mathcal{VCB}}$ *described above satisfies*

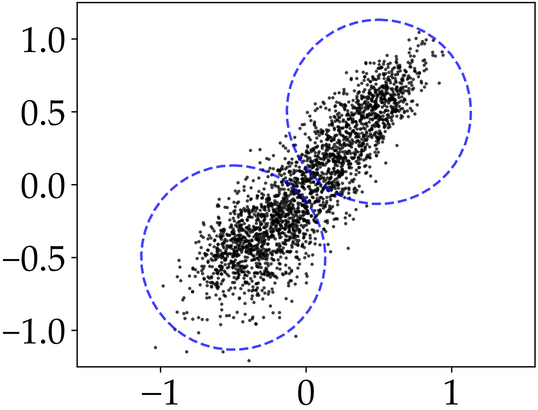
$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{\delta}_{\mathcal{VCB}}(Z_i) - \delta_{\mathcal{VCB}}(Z_i)\|^2 = O_{\mathbb{P}}(\alpha_n)$$

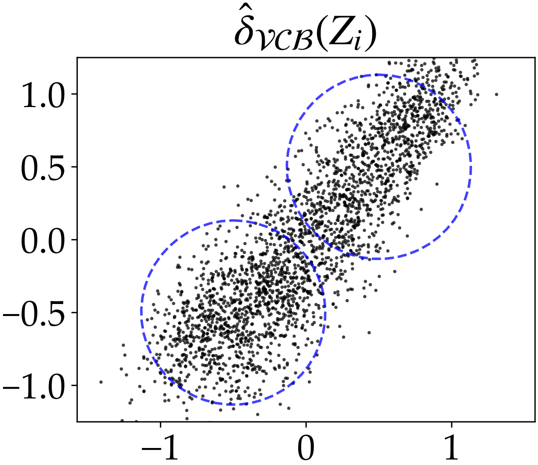
as $n \to \infty$.

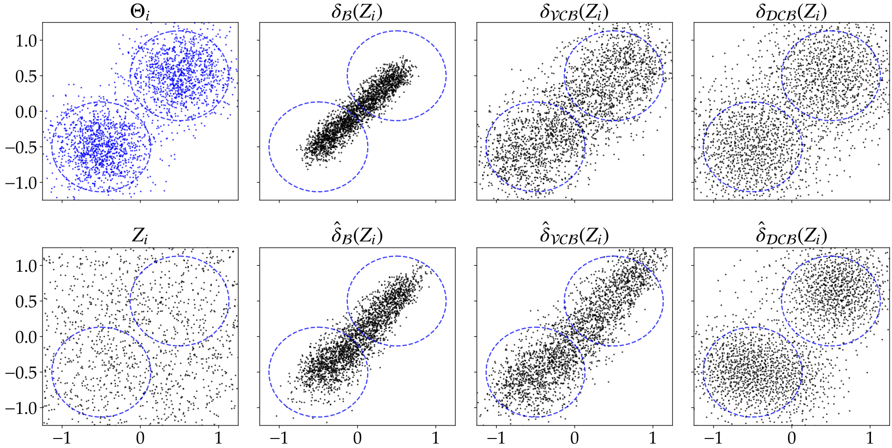
Rate of convergence of $\hat{\delta}_{\mathcal{VCB}}$ as fast as rate of convergence of $\hat{\delta}_{\mathcal{B}}$
(Jiang-Zhang 2009, Soloff et al. 2024, Shen-Wu 2025, ...)

Moment estimation in ($\mathcal{VCB}$) is easier than deconvolution in ($\mathcal{DCB}$).

$\hat{\delta}_{\mathcal{B}}(Z_i)$




$\hat{\delta}_{\mathcal{VCB}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1



Θ_i
$\delta_{\mathcal{B}}(Z_i)$
$\delta_{\mathcal{VCB}}(Z_i)$
$\delta_{\mathcal{DCB}}(Z_i)$
Z_i
$\hat{\delta}_{\mathcal{B}}(Z_i)$
$\hat{\delta}_{\mathcal{VCB}}(Z_i)$
$\hat{\delta}_{\mathcal{DCB}}(Z_i)$
1.0
0.5
0.0
−0.5
−1.0
−1
0
1

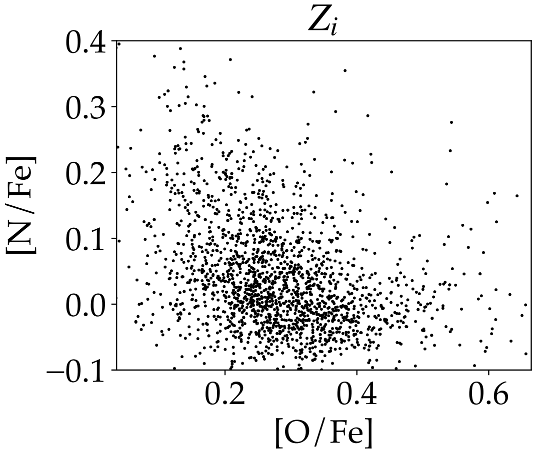
Other discussions in the paper:

- For general constraints: uniqueness, characterization, method for computation, rates of convergence, etc.
- Modifications for non-Gaussian likelihoods, and heteroskedasticity in the likelihood
- Some computational considerations

Notably not in the paper:

- Practically, how to choose between $\hat{\delta}_{\mathcal{B}}$, $\hat{\delta}_{\mathcal{DCB}}$, and $\hat{\delta}_{\mathcal{VCB}}$?
- Scalable computation for large data sets or high dimension?
- Lower bounds to show that the rates of convergence are sharp?

IV. Astronomy Application

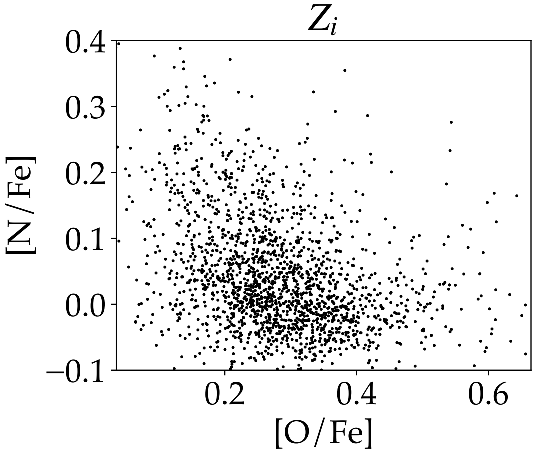

Z_i
0.4
0.3
0.2
0.1
0.0
−0.1
[N/Fe]
0.2
0.4
0.6
[O/Fe]

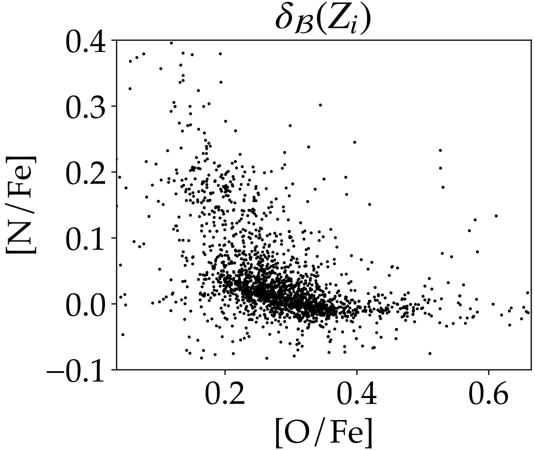
Interested in $n = 2000$ stars in a given catalog.

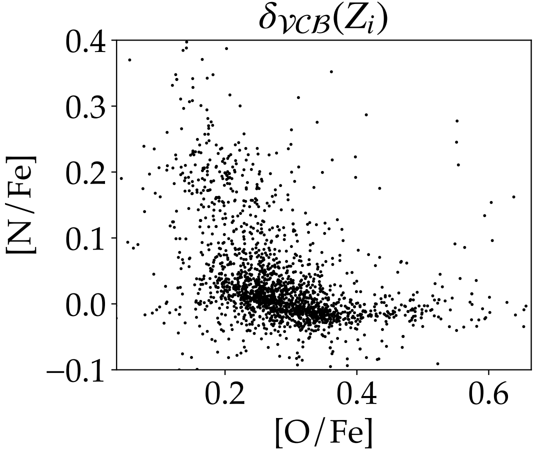
For each star, measure the amount of nitrogen (N) and oxygen (O), relative to the amount of iron (Fe).

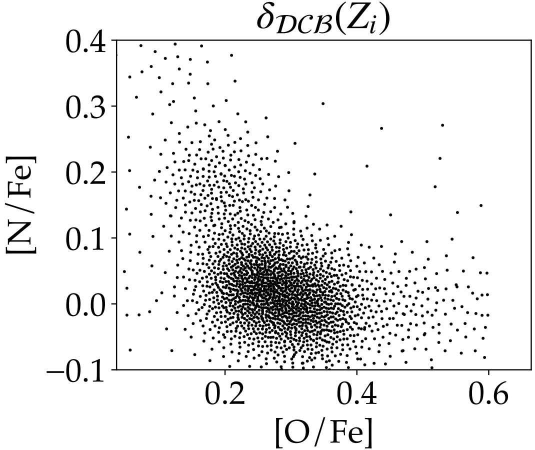
Latent relative abundances $\Theta_1, \ldots, \Theta_n \in \mathbb{R}^2$ are i.i.d from unknown G.

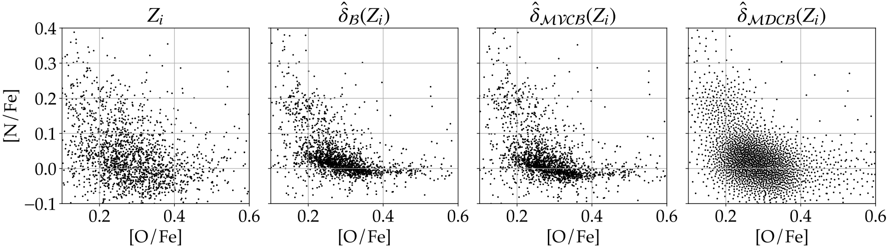
Observations $Z_1, \ldots, Z_n \in \mathbb{R}^2$ have heteroskedastic additive Gaussian noise, but the likelihood covariances $\Sigma_1, \ldots, \Sigma_n$ are known


Z_i
0.4
0.3
0.2
0.1
0.0
−0.1
[N/Fe]
0.2
0.4
0.6
[O/Fe]


$\hat{\delta}_{\mathcal{B}}(Z_i)$
0.4
0.3
0.2
0.1
0.0
−0.1
[N/Fe]
0.2
0.4
0.6
[O/Fe]


$\hat{\delta}_{\mathcal{VCB}}(Z_i)$
[N/Fe]
0.4
0.3
0.2
0.1
0.0
−0.1
[O/Fe]
0.2
0.4
0.6


$\hat{\delta}_{\mathcal{DCB}}(Z_i)$
0.4
0.3
0.2
0.1
0.0
−0.1
[N/Fe]
0.2
0.4
0.6
[O/Fe]



Z_i
$\hat{\delta}_{\mathcal{B}}(Z_i)$
$\hat{\delta}_{\mathcal{MVCB}}(Z_i)$
$\hat{\delta}_{\mathcal{MDCB}}(Z_i)$
[N/Fe]
0.4
0.3
0.2
0.1
0.0
−0.1
0.2
0.4
0.6
[O/Fe]

Thank you!

References

L. D. Brown and E. Greenshtein. Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Ann. Statist.,* 37(4): 1685-1704, 2009.

B. Efron. Bayes, oracle Bayes and empirical Bayes. *Stat. Sci.,* 34(2):177-201, 2019

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.,* 19(3):1257-1272, 1991.

D. Freirich, T. Michaeli, and R. Meir. A theory of the distortion-perception tradeoff in Wasserstein space. *NeurIPS*, 34:25661–25672, 2021.

N. García-Trillos and B. Sen. A new perspective on denoising based on optimal transport. *Inf. Inference,* 13(4):iaae029, 2024.

M. Ghosh. Constrained Bayes estimation with applications. *J. Am. Stat. Assoc.,* 87(418):533–540, 1992.

M. Ghosh and T. Maiti. Adjusted Bayes estimators with applications to small area estimation. *Sankhya*, 61(1): 71-90, 1999.

W. Jiang and C.-H. Zhang. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.*, 37(4):1647-1684, 2009.

T. Loredo. Analyzing data from astronomical surveys: Issues and directions. *Statistical Challenges in Modern Astronomy IV*, 371:121-138, 2007.

T. A. Louis. Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Am. Sat. Assoc.,* 79(386):393–398, 1984.

I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.*, 48:257-263, 1982.

H. Robbins. An empirical Bayes approach to statistics. *Berkeley Symp. on Math. Statist. and Prob.,* 3(1):157-163, 1956.

Y. Shen and Y. Wu. Empirical Bayes estimation: When does g-modeling beat f-modeling in theory (and in practice)? *Ann. Statist.,* to appear, 2025

J. A. Soloff, A. Guntuboyina, and B. Sen. Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *J. R. Stat. Soc. B. Methodol.,* 87(1):1-32, 2024.