Constrained denoising, empirical Bayes, and optimal transport

Adam Quinn Jaffe

With Nikolaos Ignatiadis and Bodhisattva Sen

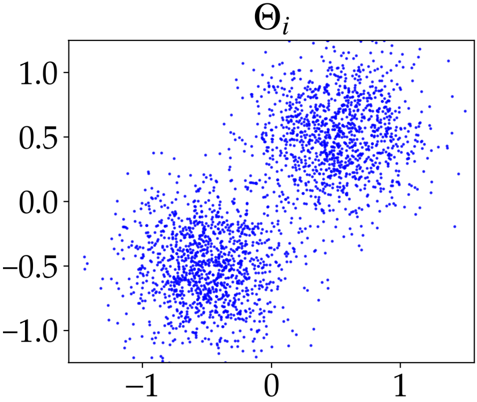
Model for data:

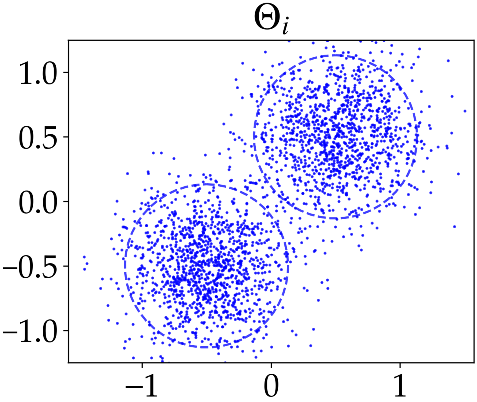
$$\Theta \sim G \qquad \text{and } Z = \Theta + \varepsilon \qquad \text{where } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

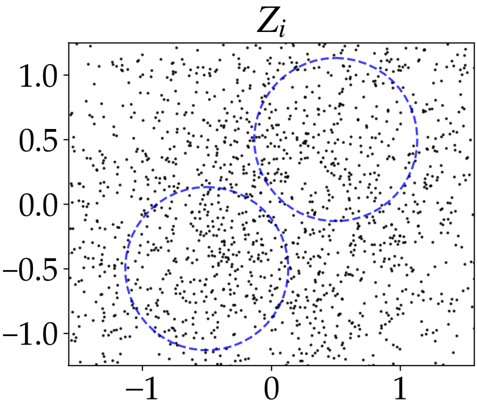
where G is unknown distribution, and Σ is known likelihood variance.

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 $Z_1, \ldots, Z_n$


Θ_i
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:

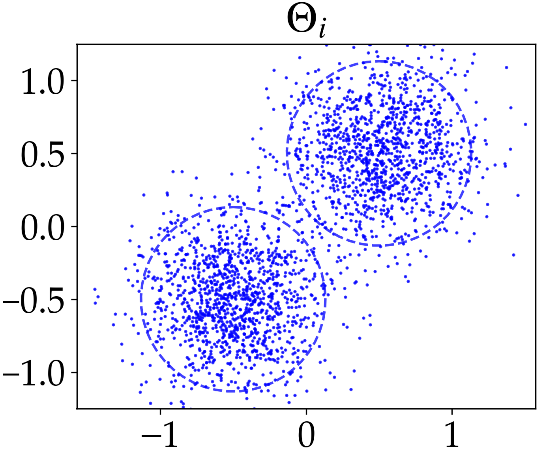
$$\begin{cases} \text{minimize} & \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\|\delta(Z_i)-\Theta_i\|^2\right] \\ \text{over} & \delta:\mathbb{R}^m\to\mathbb{R}^m \end{cases}$$

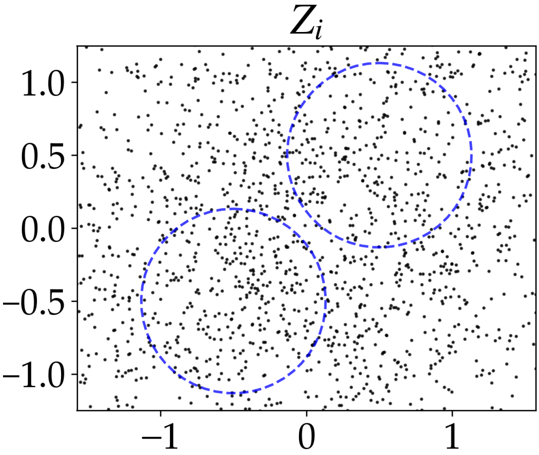
Solution is the posterior mean:

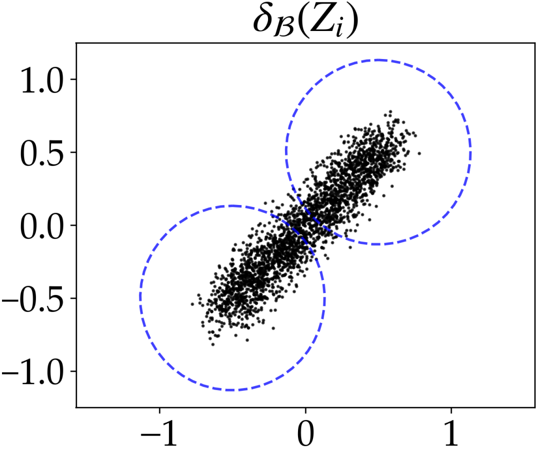
$$\delta_{\mathcal{B}}(z) = \mathbb{E}[\Theta \mid Z = z].$$

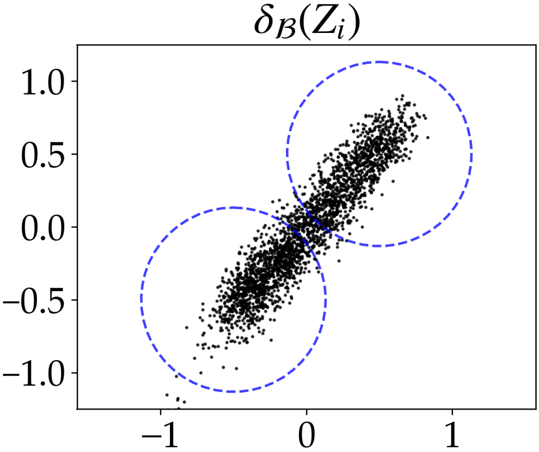
Although $\delta_{\mathcal{B}}$ depends on G, we can try to approximate $\hat{\delta}_{\mathcal{B}}$ empirically.

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
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


$\hat{\delta}_{\mathcal{B}}(Z_i)$
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 the same as $\hat{\delta}_{\mathcal{B}}(Z_1), \ldots, \hat{\delta}_{\mathcal{B}}(Z_n)$ 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:

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

Add constraints:

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

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?

Empirical Bayes approximation?

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. Optimal Transport

III. Oracle Constrained Bayes

IV. Empirical Constrained Bayes

V. Applications

II. Optimal Transport

For probability measures μ, μ' on $\mathbb{R}^k$, how to move mass from μ to μ'?

Measuring cost as squared distance leads to the *Monge problem*:

$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|T(X) - X\|^2\right] \\ \text{over} & T : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & X \overset{\mathcal{D}}{=} \mu \text{ and } T(X) \overset{\mathcal{D}}{=} \mu' \end{cases}$$

Example. If $\mu = \mathcal{N}(a, \sigma^2)$ and $\mu' = \mathcal{N}(a', \sigma^2)$, then $T(x) = x - a + a'$.

Example. If $\mu = \mathrm{U}[0, b]$ and $\mu' = \mathrm{U}[0, b']$, then $T(x) = (b'/b)x$.

The Monge problem is hard since constraints are non-convex

For probability measures μ, μ' on $\mathbb{R}^k$, how to move mass from μ to μ'?

Measuring cost as squared distance leads to the *Monge problem*:

$$\begin{cases} \text{minimize} & \int_{\mathbb{R}^k} \|T(x) - x\|^2 \mathrm{d}\mu(x) \\ \text{over} & T : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & T_\# \mu = \mu' \end{cases}$$

Example. If $\mu = \mathcal{N}(a, \sigma^2)$ and $\mu' = \mathcal{N}(a', \sigma^2)$, then $T(x) = x - a + a'$.

Example. If $\mu = \mathrm{U}[0, b]$ and $\mu' = \mathrm{U}[0, b']$, then $T(x) = (b'/b)x$.

The Monge problem is hard since constraints are non-convex

For probability measures μ, μ' on $\mathbb{R}^k$, how to move mass from μ to μ'?

Previously, we sent all mass at point x to point $T(x)$, but instead we can send mass at point x to a distribution over points $\pi_x(\,\cdot\,)$.

This leads to the *Kantorovich problem*:

$$\begin{cases} \text{minimize} & \int_{\mathbb{R}^k \times \mathbb{R}^k} \|x' - x\|^2 \mathrm{d}\pi(x, x') \\ \text{over} & \pi \in \Gamma(\mu, \mu') \end{cases}$$

where $\Gamma(\mu, \mu')$ is the space of *couplings* of μ, μ'.

The Kantorovich problem is just a linear program!

Theorem (Brenier, 1991)

If μ has a density with respect to Lebesgue measure, then the problem

$$\begin{cases} \text{minimize} & \int \|x' - x\|^2 \mathrm{d}\pi(x, x') \\ \text{over} & \pi \in \Gamma(\mu, \mu') \end{cases}$$

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

$$\begin{cases} \text{minimize} & \int \|T(x) - x\|^2 \mathrm{d}\mu(x) \\ \text{over} & T : \mathbb{R}^k \to \mathbb{R}^k \\ \text{over} & T_{\#}\mu = \mu', \end{cases}$$

and we have $T = \nabla\phi$ for some convex function $\phi : \mathbb{R}^m \to \mathbb{R}$.

Brenier's theorem says that the Kantorovich problem is a tight convex relaxation of the nonconvex Monge problem:

- Both problems have the same value
- Both problems have a unique solution
- There is a natural correspondence between the solutions:
  - If T is a solution to the Monge problem, then the distribution of $(X, T(X))$ for $X \sim \mu$ is a solution to the Kantorovich problem.
  - If π is a solution to the Kantorovich problem, then the *barycentric projection* $T(x) := \mathbb{E}[X' \mid X = x]$ for $(X, X') \sim \pi$ is a solution to the Monge problem

Thus, the Monge problem can be solved via the Kantorovich problem

Optimal transport can be understood as...

- a concrete problem in applied mathematics
- a probabilistic formulation of some differential equations
- a useful metric on spaces of probability measures for statisticians
- a method for solving optimization problems with non-convex distributional constraints

III. 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:

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

Distribution-constrained denoising problem:

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

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

Approximately looks like a Monge transport problem from distribution of 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)

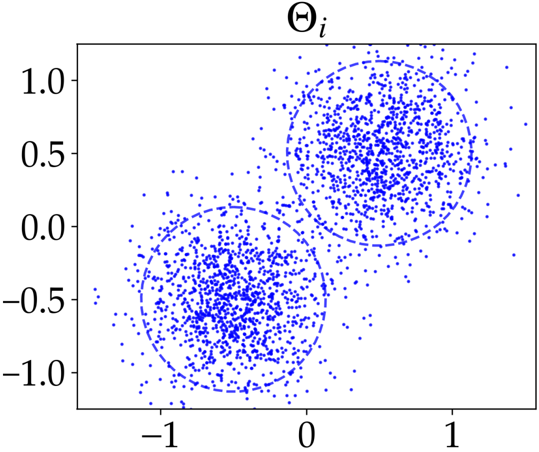
Under suitable regularity conditions, if F and G denote the distributions of Z and Θ, respectively, then the problem

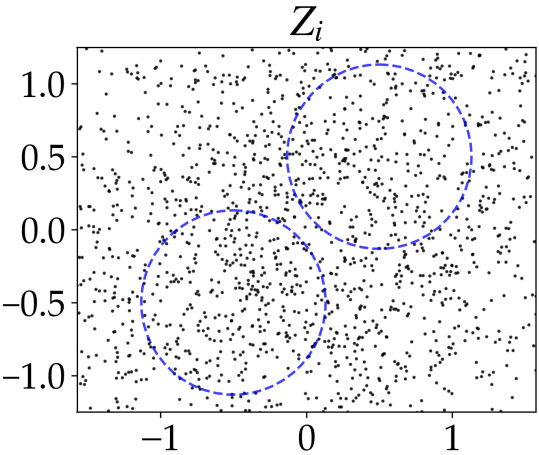
$$\begin{cases} \text{minimize} & \int \|\delta_{\mathcal{B}}(z) - \theta\|^2 \mathrm{d}\pi(z,\theta) \\ \text{over} & \pi \in \Gamma(F, G) \end{cases}$$

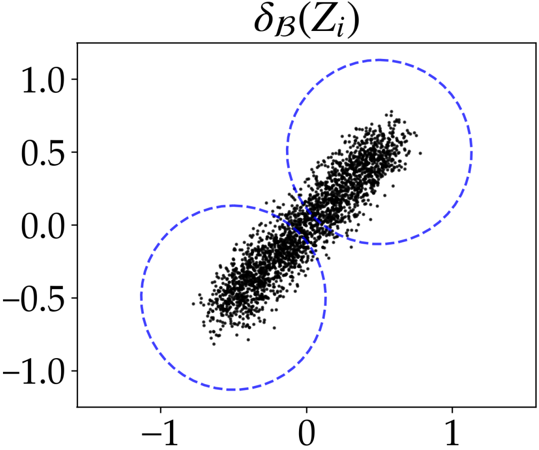
admits a unique solution, this solution is concentrated on the graph of a function $\delta_{\mathcal{DCB}} : \mathbb{R}^m \to \mathbb{R}^m$ which is the unique solution to the problem

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

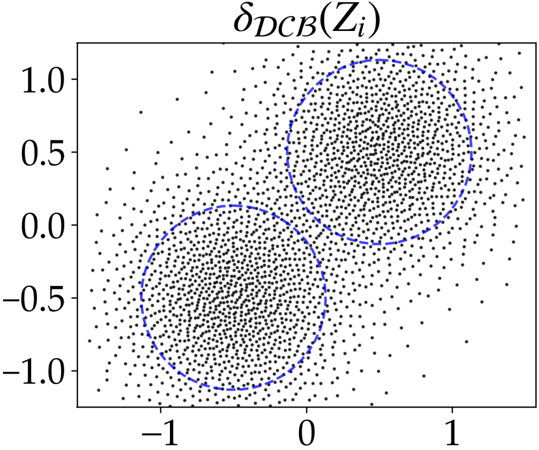
and we have $\delta_{\mathcal{DCB}} = \nabla\phi \circ \delta_{\mathcal{B}}$ for some convex function $\phi : \mathbb{R}^m \to \mathbb{R}$.


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.

Explicit formula for optimal transport between Gaussians:

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

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

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

Theorem (AQJ-Ignatiadis-Sen, 2025)

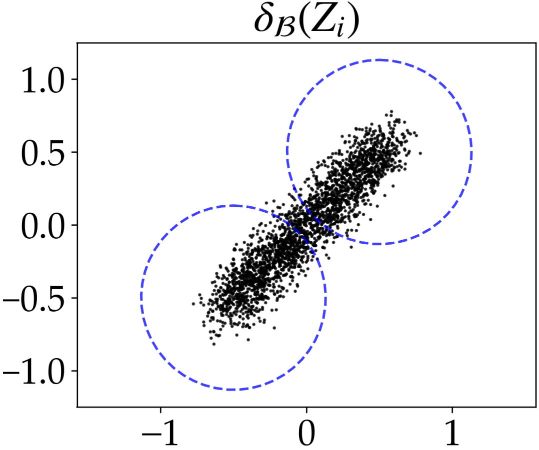
Under suitable regularity conditions, the problem

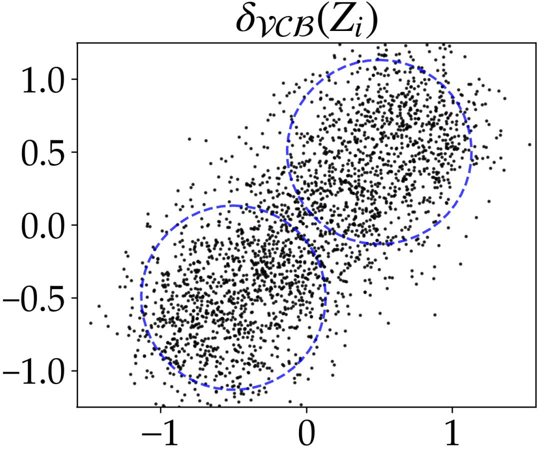
$$\begin{cases} \text{minimize} & \mathbb{E}\left[\|\delta(Z)-\Theta\|^2\right] \\ \text{over} & \delta : \mathbb{R}^m \to \mathbb{R}^m \\ \text{with} & \mathbb{E}[\delta(Z)] = \mathbb{E}[\Theta] \\ \text{and} & \mathrm{Cov}(\delta(Z)) = \mathrm{Cov}(\Theta) \end{cases}$$

admits a unique solution given by

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

Extends several previously 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


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

IV. Empirical Constrained Bayes

Many known methods for achieving small value of the following:

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

Example. Under parametric likelihood and conjugate prior, have

$$\delta_{\mathcal{B}}(z)=\alpha z+(1-\alpha)\mathbb{E}[\Theta]$$

for some $\alpha \in \mathbb{R}$ that can be estimated via method of moments.

Example. Under Gaussian likelihood and nonparametric prior

$$\delta_{\mathcal{B}}(z):=z+\sigma^2\nabla\log f(z)$$

for marginal density f of Z that can be estimated via KDE

Example. Set $\hat{\delta}_{\mathcal{B}}(z):=\mathbb{E}_{\hat{G}_n}[\Theta \mid Z=z]$ for some estimate $\hat{G}_n$ of G.

Want to achieve small value of the following:

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

for $* \in \{\mathcal{DCB}, \mathcal{VCB}\}$.

In other words, we want comparable risk to the oracle.

Further, we want a modular strategy to transform any $\hat{\delta}_{\mathcal{B}}$ into some $\hat{\delta}_*$.

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 $1 \leq i \leq n$ 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 $1 \leq i \leq n$
                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 $1 \leq i \leq n$

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

Theorem (AQJ-Ignatiadis-Sen 2025)

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

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

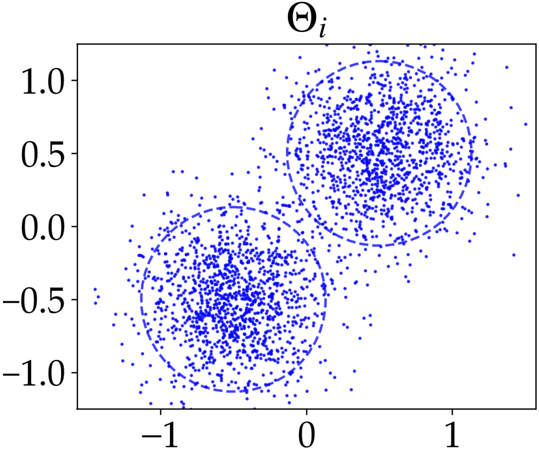
and that $\hat{G}_n$ *satisfies*

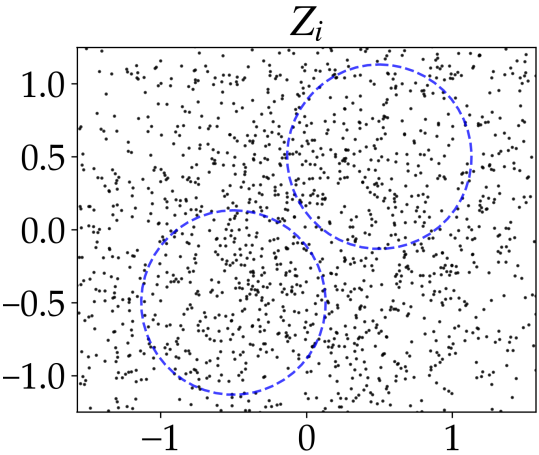
$$W_2^2(\hat{G}_n, G) = O_{\mathbb{P}}(\beta_n),$$

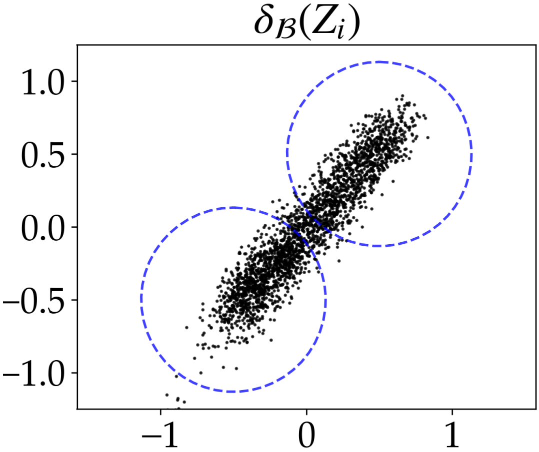
as $n \to \infty$*. Then the denoiser* $\hat{\delta}_{\mathcal{DCB}}$ *described before satisfies*

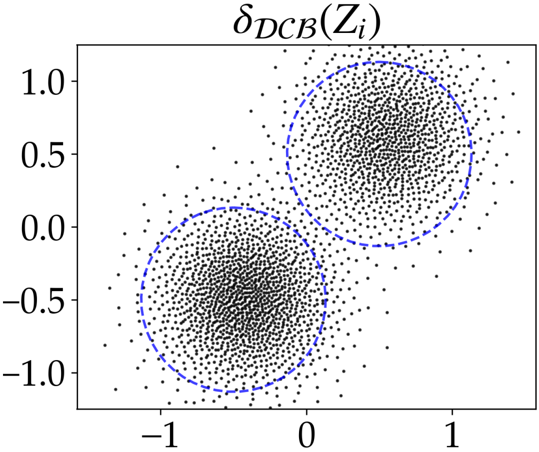
$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{\delta}_{\mathcal{DCB}}(Z_i) - \delta_{\mathcal{DCB}}(Z_i)\|^2 = O_{\mathbb{P}}\left(\alpha_n^{1/2} \vee \beta_n\right).$$

In nonparametric problems, rate of convergence of $\hat{\delta}_{\mathcal{DCB}}$ is dominated by the slow rate of convergence of 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:

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

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

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 $\mathbb{E}[\Theta]$.

```
procedure VarianceConstrainedEB($Z_1, \dots, Z_n$)

    input: samples $Z_1, \dots, Z_n \in \mathbb{R}^m$
    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 $\hat{\delta}_{\mathcal{B}}(Z_1), \dots \hat{\delta}_{\mathcal{B}}(Z_n)$
    $\hat{\mu} \leftarrow$ sample mean of $Z_1, \dots, Z_n$
    $\hat{S} \leftarrow$ sample covariance matrix of $Z_1, \dots, Z_n$
    $\hat{A} \leftarrow (\hat{S} - \Sigma)_+$

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

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

Theorem (AQJ-Ignatiadis-Sen 2025)

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

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

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

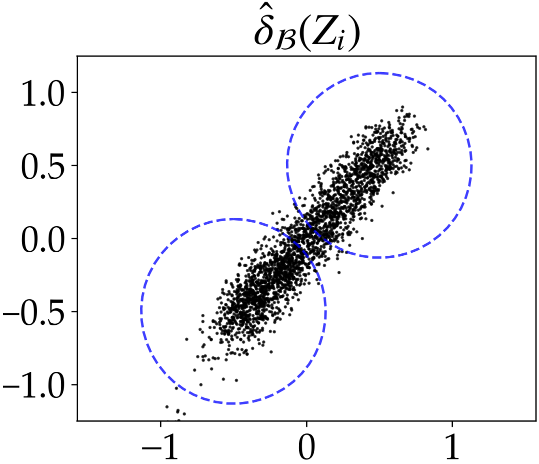
$$\left\|\hat{S}_n - S\right\|^2 = O_{\mathbb{P}}(\gamma_n)$$

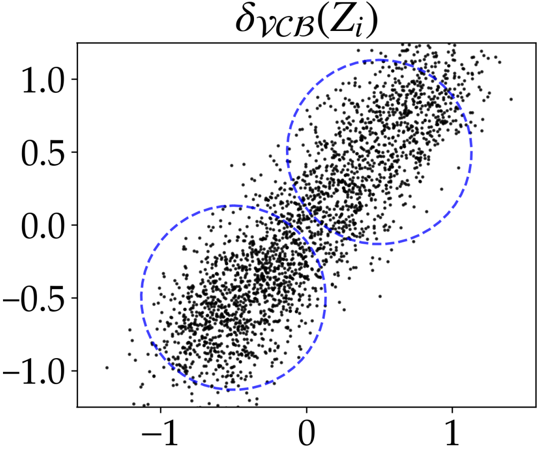
as $n \to \infty$. *Then the denoiser* $\hat{\delta}_{\mathcal{VCB}}$ *described above satisfies*

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

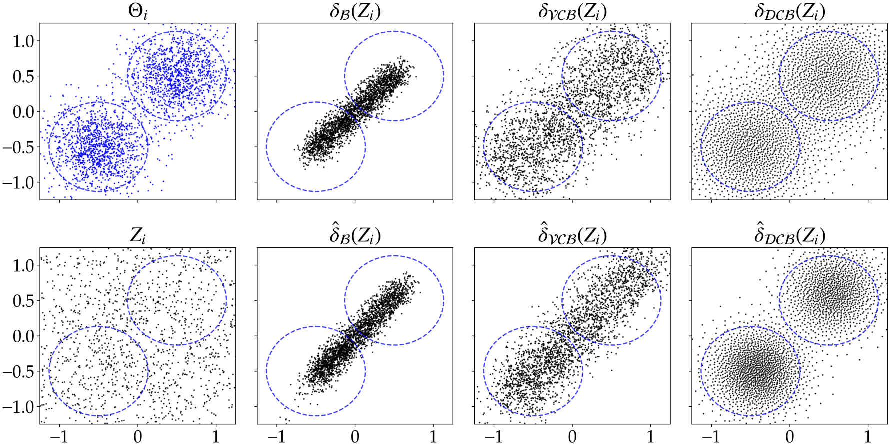
as $n \to \infty$.

Rate of convergence of $\hat{\delta}_{\mathcal{VCB}}$ is dominated by whichever of EB denoising or covariance estimation is harder.


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



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


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:

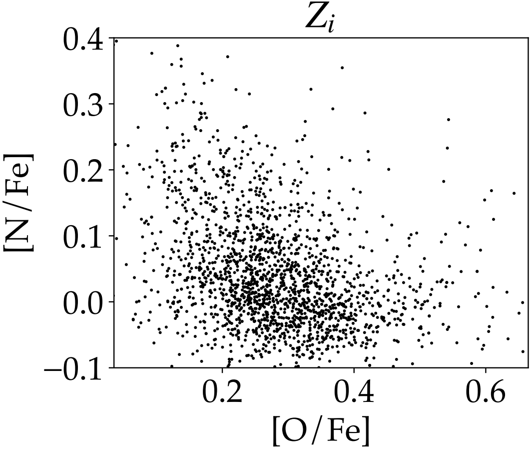
- For general constraints: uniqueness, characterization, method for computation, rates of convergence, etc.
- Modifications for non-Gaussian likelihoods, and heteroskedasticity in the likelihood
- Some computational considerations

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?

V. Applications

Astronomy

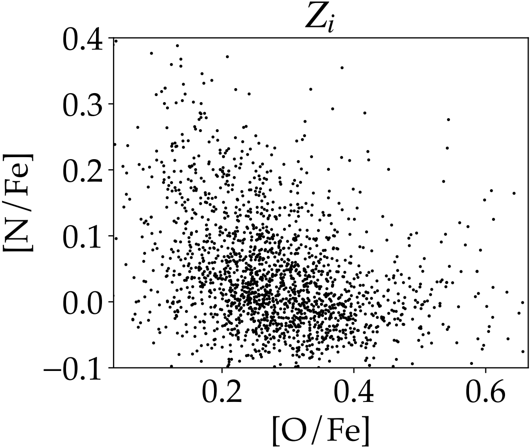

Z_i
0.4
0.3
0.2
0.1
0.0
−0.1
[N/Fe]
0.2
0.4
0.6
[O/Fe]

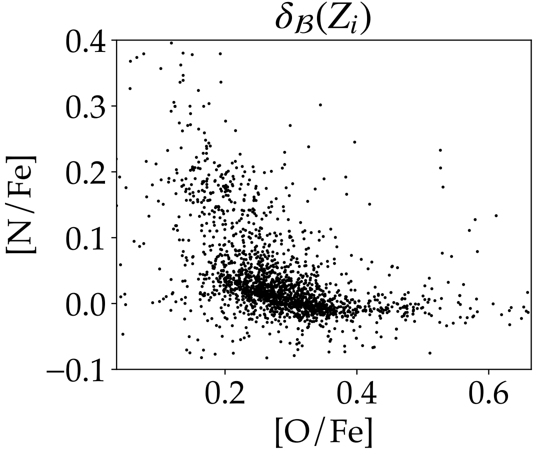
Interested in chemical composition of $n = 2000$ stars in a given catalog.

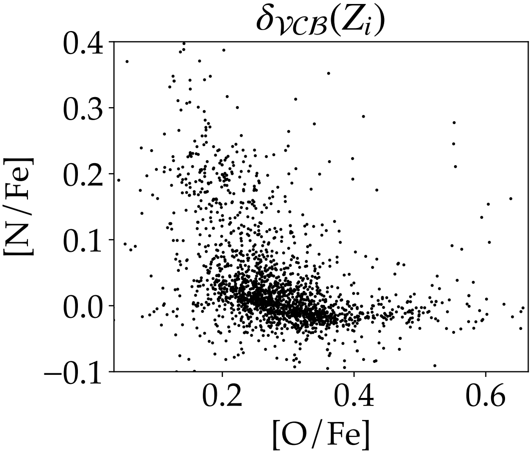
For each star, measure the amount of nitrogen (N) and oxygen (O), relative to the amount of iron (Fe).

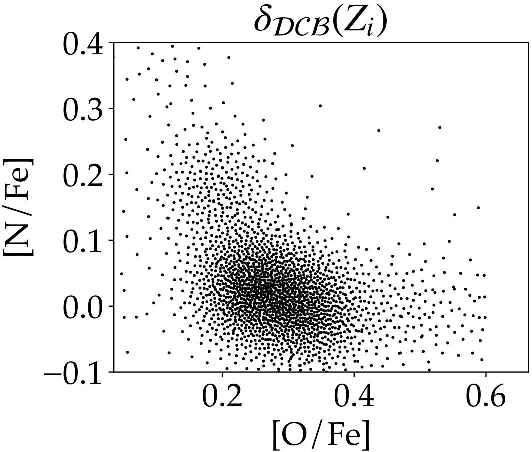
Latent relative abundances $\Theta_1, \ldots, \Theta_n \in \mathbb{R}^2$ are i.i.d from unknown G.

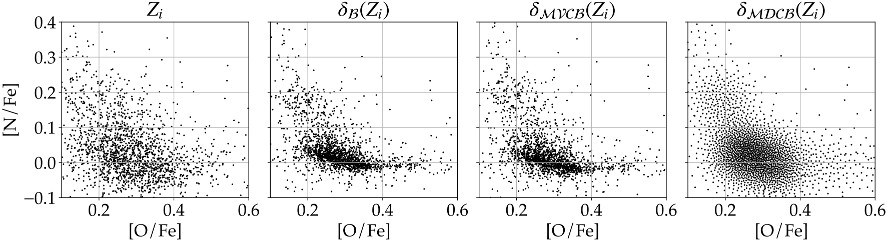
Observations $Z_1, \ldots, Z_n \in \mathbb{R}^2$ have heteroskedastic additive Gaussian noise, but the likelihood covariances $\Sigma_1, \ldots, \Sigma_n$ are known


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

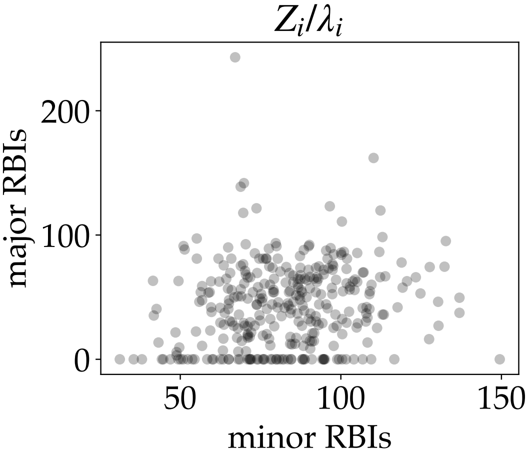

$\hat{\delta}_{\mathcal{B}}(Z_i)$
[N/Fe]
0.4
0.3
0.2
0.1
0.0
−0.1
[O/Fe]
0.2
0.4
0.6


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


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

Baseball


Z_i/λ_i
major RBIs
minor RBIs
0
100
200
50
100
150

Interested in the joint distribution of minor- and major-league batting skill for $n = 324$ rookie baseball players.

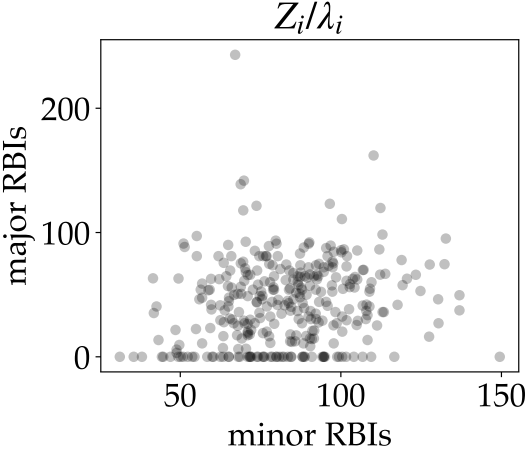
For each player, measure his runs batted in (RBIs) in last season of minor league and first season of major league.

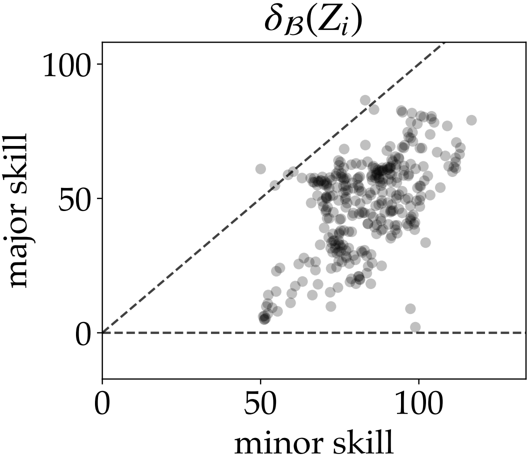
Latent bivariate skill $\Theta_1, \ldots, \Theta_n \in [0, \infty)^2$ are i.i.d from unknown G.

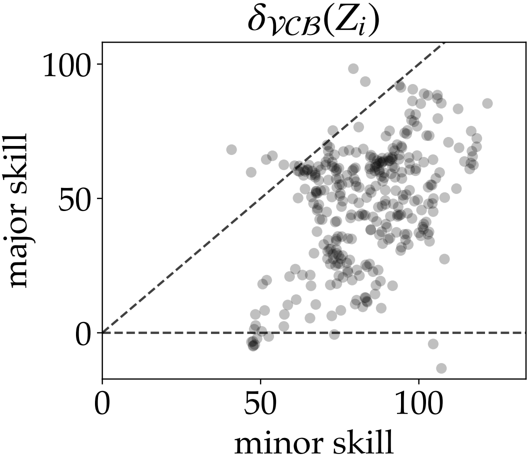
Observations $Z_1, \ldots, Z_n \in \mathbb{N}^2$ come from likelihood

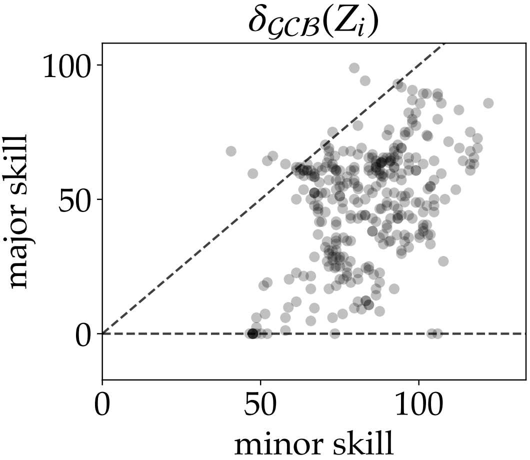
$$(Z_i \mid \Theta_i = \theta_i) \overset{d}{=} \mathrm{Poi}(\lambda_i^{\text{minor}} \theta_i^{\text{minor}}) \otimes \mathrm{Poi}(\lambda_i^{\text{major}} \theta_i^{\text{major}})$$

with known exposures $\lambda_1, \ldots, \lambda_n \in (0, \infty)^2$, i.e., number of at-bats.

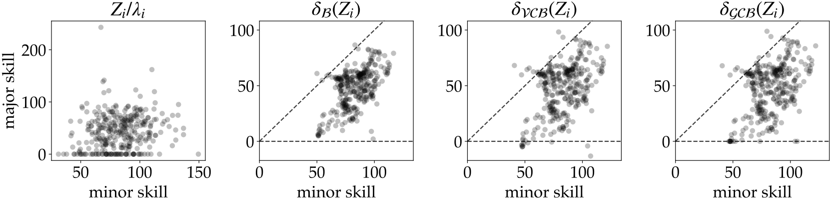

Z_i/λ_i
major RBIs
minor RBIs
0
100
200
50
100
150


$\hat{\delta}_{\mathcal{B}}(Z_i)$
major skill
minor skill
100
50
0
0
50
100

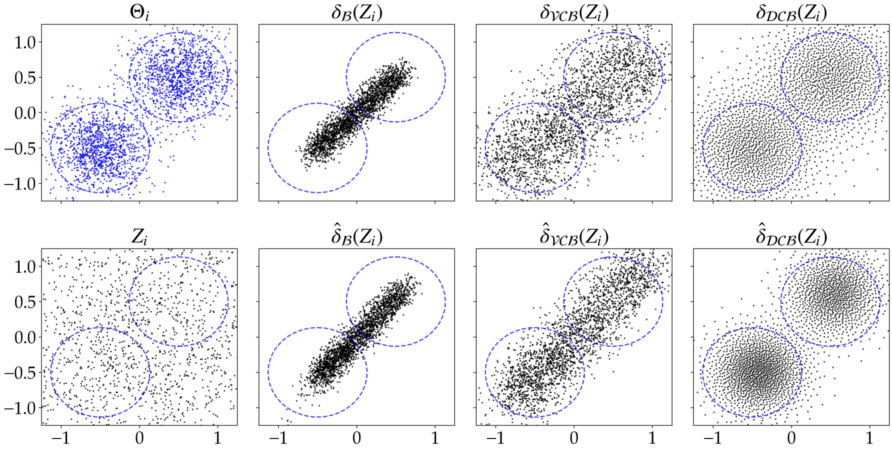

$\hat{\delta}_{\mathcal{VCB}}(Z_i)$
major skill
100
50
0
0
50
100
minor skill



$\hat{\delta}_{\mathcal{GCB}}(Z_i)$
100
50
0
major skill
0
50
100
minor skill


Z_i/λ_i
$\hat{\delta}_{\mathcal{B}}(Z_i)$
$\hat{\delta}_{\mathcal{VCB}}(Z_i)$
$\hat{\delta}_{\mathcal{GCB}}(Z_i)$
major skill
minor skill
0
50
100
150
200

Thank you!


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

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.