Applications of the Coordinate Ascent Variational Inference (CAVI) Algorithm

Gilad Turok Austin Tao

Columbia University – Department of Applied Mathematics

November 16th, 2022

Outline

Table of Contents

About Us/Why We are Interested in This

Gilad

- Senior in Columbia College
- CS + Applied Math Double Major
- Comp Bio Research Assistant
- Fascinated with finding structure in genomic data
- Fun Fact: Keegan Michael prank called me

Austin

- Senior in SEAS
- Applied Math major, CS minor
- Became interested after COMS W4771: Machine Learning
- Currently in COMS E4762: ML for Functional Genomics

Bibliography

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (Apr. 2017). “Variational Inference: A Review for Statisticians”. In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: 10.1080/01621459.2017.1285773.

Gordon-Rodriguez, E., G. Loaiza-Ganem, A. Potapczynski, and J. P. Cunningham (2022). *On the Normalizing Constant of the Continuous Categorical Distribution*. DOI: 10.48550/ARXIV.2204.13290.

Kingma, D. P. and M. Welling (2019). “An Introduction to Variational Autoencoders”. In: *Foundations and Trends® in Machine Learning* 12.4, pp. 307–392. DOI: 10.1561/2200000056.

Prince, S. (2012). *Computer Vision: Models Learning and Inference*. Cambridge University Press.

Wang, G., A. Sarkar, P. Carbonetto, and M. Stephens (2020). “A simple new approach to variable selection in regression, with application to genetic fine-mapping”. In: *bioRxiv*. DOI: 10.1101/501114. eprint: https://www.biorxiv.org/content/early/2020/06/27/501114.full.pdf.

Zhang, W., H. Najafabadi, and Y. Li (2021). “SparsePro: an efficient genome-wide fine-mapping method integrating summary statistics and functional annotations”. In: *bioRxiv*. DOI: 10.1101/2021.10.04.463133. eprint: https://www.biorxiv.

Table of Contents

Variational Inference

- Create probabilistic models that explain how observed data $\mathbf{X}$ is generated from intermediary, hidden variables $\mathbf{z}$
- Want to estimate latent variables $\mathbf{z}$ with Variational Inference [1] to learn about our model

[1] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe (Apr. 2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: 10.1080/01621459.2017.1285773

Variational Inference: Motivating Example

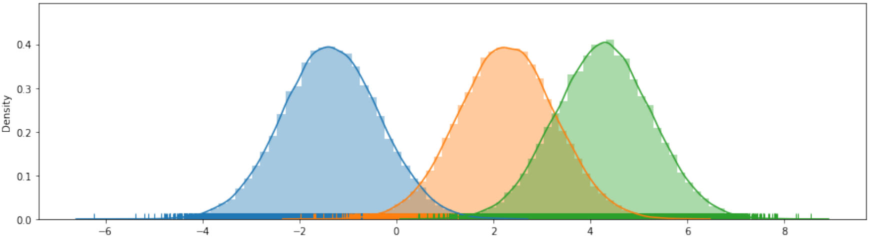

Figure: Gaussian Mixture Model

Which data points belong to which clusters?

Variational Inference: True Posterior p

- Given observed variables $\mathbf{X}$ and joint probability model $p(\mathbf{z}, \mathbf{X})$ we seek the posterior
$$p(\mathbf{z} \mid \mathbf{X}) = \frac{p(\mathbf{z}, \mathbf{X})}{p(\mathbf{X})}$$
- Posterior captures latent variables $\mathbf{z}$ conditioned on observed variables $\mathbf{X}$ e.g. cluster assignments for GMM

Variational Inference: True Posterior p

- Given observed variables $\mathbf{X}$ and joint probability model $p(\mathbf{z}, \mathbf{X})$ we seek the posterior

$$p(\mathbf{z} \mid \mathbf{X}) = \frac{p(\mathbf{z}, \mathbf{X})}{p(\mathbf{X})}$$

- Posterior captures latent variables $\mathbf{z}$ conditioned on observed variables $\mathbf{X}$ e.g. cluster assignments for GMM
- Must calculate the posterior's denominator,

$$p(\mathbf{X}) = \int p(\mathbf{z}, \mathbf{X}) d\mathbf{z}$$

Variational Inference: True Posterior p

- Given observed variables $\mathbf{X}$ and joint probability model $p(\mathbf{z}, \mathbf{X})$ we seek the posterior

$$p(\mathbf{z} \mid \mathbf{X}) = \frac{p(\mathbf{z}, \mathbf{X})}{p(\mathbf{X})}$$

- Posterior captures latent variables $\mathbf{z}$ conditioned on observed variables $\mathbf{X}$ e.g. cluster assignments for GMM
- Must calculate the posterior's denominator,

$$p(\mathbf{X}) = \int p(\mathbf{z}, \mathbf{X}) d\mathbf{z}$$

- Integral is intractable because it is unsolvable or requires exponential computation time

Variational Inference: Approximate Posterior q

- Instead want to approximate the posterior using $q(\mathbf{z})$, which only depends on latent variables $\mathbf{z}$
- Want approximate posterior q to be "close" to true posterior p

Variational Inference: Approximate Posterior q

- Instead want to approximate the posterior using $q(\mathbf{z})$, which only depends on latent variables $\mathbf{z}$
- Want approximate posterior q to be "close" to true posterior p
- Closeness is measured with KL divergence

$$\begin{aligned} KL(q(\mathbf{z}) \,\|\, p(\mathbf{z} \mid \mathbf{X})) &= \mathbb{E}_{\mathbf{z}}[\log q(\mathbf{z})] - \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z} \mid \mathbf{X})] \\ &= \mathbb{E}_{z}[\log q(\mathbf{z})] - \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z}, \mathbf{X})] + \log p(\mathbf{X}) \end{aligned}$$

Suppress notation that expectation is w.r.t. $\mathbf{z}$

Variational Inference: Approximate Posterior q

- Instead want to approximate the posterior using $q(\mathbf{z})$, which only depends on latent variables $\mathbf{z}$
- Want approximate posterior q to be "close" to true posterior p
- Closeness is measured with KL divergence

$$\begin{aligned} KL(q(\mathbf{z}) \parallel p(\mathbf{z} \mid \mathbf{X})) &= \mathbb{E}_{\mathbf{z}}[\log q(\mathbf{z})] - \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z} \mid \mathbf{X})] \\ &= \mathbb{E}_{z}[\log q(\mathbf{z})] - \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z}, \mathbf{X})] + \log p(\mathbf{X}) \end{aligned}$$

Suppress notation that expectation is w.r.t. $\mathbf{z}$

- However, computing KL divergence requires $\log p(\mathbf{X})$ which is intractable

Variational Inference: ELBO

- Because KL divergence is intractable, we optimize the ELBO $\mathcal{L}$, the Evidence Lower Bound instead
- The ELBO $\mathcal{L}$ is equivalent to negative KL divergence up to an additive constant

Variational Inference: ELBO

- Because KL divergence is intractable, we optimize the ELBO $\mathcal{L}$, the Evidence Lower Bound instead
- The ELBO $\mathcal{L}$ is equivalent to negative KL divergence up to an additive constant

The ELBO

$$\begin{aligned}\mathcal{L}(z) &:= \mathbb{E}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}[\log q(\mathbf{z})] \\ &= \log p(\boldsymbol{X}) - KL(q(\mathbf{z}) \,\|\, p(\mathbf{z} \mid \boldsymbol{X}))\end{aligned}$$

Table of Contents

CAVI Algorithm

- Constructed optimization problem: maximize ELBO $\mathcal{L}(\mathbf{z})$ to minimize $KL(q(\mathbf{z}) \,||\, p(\mathbf{z} \mid \mathbf{X}))$

$$q^*(\mathbf{z}) = \underset{q(\mathbf{z})\in\mathcal{D}}{\arg\max}\, \mathcal{L}(z) = \underset{q(\mathbf{z})\in\mathcal{D}}{\arg\max}\, \mathbb{E}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}[\log q(\mathbf{z})]$$

- Approximate posterior $q^*(\mathbf{z})$ best approximates the true posterior $p(\mathbf{z} \mid \mathbf{X})$

CAVI Algorithm

- Constructed optimization problem: maximize ELBO $\mathcal{L}(\mathbf{z})$ to minimize $KL(q(\mathbf{z}) \,||\, p(\mathbf{z} \mid \mathbf{X}))$

$$q^*(\mathbf{z}) = \underset{q(\mathbf{z}) \in \mathcal{D}}{\arg\max}\, \mathcal{L}(z) = \underset{q(\mathbf{z}) \in \mathcal{D}}{\arg\max}\, \mathbb{E}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}[\log q(\mathbf{z})]$$

- Approximate posterior $q^*(\mathbf{z})$ best approximates the true posterior $p(\mathbf{z} \mid \mathbf{X})$
- Optimization is performed one coordinate at a time for tractability i.e. *Coordinate Ascent Variational Inference* algorithm
- Want to derive ideal CAVI update for latent variable z_k while holding other latent variables $z_{\setminus k}$ constant until convergence

CAVI Algorithm: Ideal Update Derivation

$$
\begin{aligned}
\mathcal{L}(z) &= \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}_{\mathbf{z}}[\log q(\mathbf{z})] \\
&= \int_z q(\mathbf{z}) \log p(\mathbf{z}, \mathbf{X}) dz - \Big[\mathbb{E}_{\mathbf{z}_k}[\log q(\mathbf{z}_k)] + \sum_{k\prime \neq k} \mathbb{E}_{\mathbf{z}_{k\prime}}[\log q(\mathbf{z}_{k\prime})]\Big] \\
&= \int_{\mathbf{z}_k} q(\mathbf{z}_k) \left[\int_{\mathbf{z}_{\setminus k}} q(\mathbf{z}_{\setminus k}) [\log p(\mathbf{z}, \mathbf{X})] d\mathbf{z}_{\setminus k} \right] d\mathbf{z}_k \\
&\quad - \int_{\mathbf{z}_k} q(\mathbf{z}_k) \log q(\mathbf{z}_k) d\mathbf{z}_k + C \\
&= \int_{\mathbf{z}_k} q(\mathbf{z}_k) \mathbb{E}_{\mathbf{z}_{\setminus k}}[\log p(\mathbf{z}, \mathbf{X})] d\mathbf{z}_k - \int_{\mathbf{z}_k} q(\mathbf{z}_k) \log q(\mathbf{z}_k) d\mathbf{z}_k + C
\end{aligned}
$$

CAVI Algorithm: Ideal Update Derivation

$$
\begin{aligned}
\mathcal{L}(z) &= \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}_{\mathbf{z}}[\log q(\mathbf{z})] \\
&= \int_z q(\mathbf{z}) \log p(\mathbf{z}, \mathbf{X}) d\mathbf{z} - \Big[\mathbb{E}_{\mathbf{z}_k}[\log q(\mathbf{z}_k)] + \sum_{k\prime \neq k} \mathbb{E}_{\mathbf{z}_{k\prime}}[\log q(\mathbf{z}_{k\prime})]\Big] \\
&= \int_{\mathbf{z}_k} q(\mathbf{z}_k) \left[\int_{\mathbf{z}_{\setminus k}} q(\mathbf{z}_{\setminus k}) [\log p(\mathbf{z}, \mathbf{X})] d\mathbf{z}_{\setminus k} \right] d\mathbf{z}_k \\
&\quad - \int_{\mathbf{z}_k} q(\mathbf{z}_k) \log q(\mathbf{z}_k) d\mathbf{z}_k + C \\
&= \int_{\mathbf{z}_k} q(\mathbf{z}_k) \mathbb{E}_{\mathbf{z}_{\setminus k}}[\log p(\mathbf{z}, \mathbf{X})] d\mathbf{z}_k - \int_{\mathbf{z}_k} q(\mathbf{z}_k) \log q(\mathbf{z}_k) d\mathbf{z}_k + C
\end{aligned}
$$

Define $\log \tilde{p}_k(\mathbf{z}_k, \mathbf{X}) := \mathbb{E}_{z_{\setminus k}}[\log p(\mathbf{z}, \mathbf{x})]$

$$
\begin{aligned}
\mathcal{L}(\mathbf{z}) &= \int_{\mathbf{z}_k} q(\mathbf{z}_k) \log \frac{\tilde{p}_k(\mathbf{z}_k, \mathbf{X})}{q(\mathbf{z}_k)} d\mathbf{z}_k + C \\
&= -KL(q(\mathbf{z}_k) \,||\, \tilde{p}(\mathbf{z}_k, \mathbf{X})) + C
\end{aligned}
$$

CAVI Algorithm: Ideal Update Derivation

Rearranged ELBO $\mathcal{L}$ and Def of $\tilde{p}$

$$\mathcal{L}(\mathbf{z}) = -KL(q(\mathbf{z}_k) \,||\, \tilde{p}(\mathbf{z}_k, \mathbf{X})) + C$$

$$\log \tilde{p}_k(\mathbf{z}_k, \mathbf{X}) := \mathbb{E}_{z_{\setminus k}}[\log p(\mathbf{z}, \mathbf{x})] + C$$

CAVI Algorithm: Ideal Update Derivation

Rearranged ELBO $\mathcal{L}$ and Def of $\tilde{p}$

$$\mathcal{L}(\mathbf{z}) = -KL(q(\mathbf{z}_k) \,||\, \tilde{p}(\mathbf{z}_k, \mathbf{X})) + C$$
$$\log \tilde{p}_k(\mathbf{z}_k, \mathbf{X}) := \mathbb{E}_{z_{\setminus k}}[\log p(\mathbf{z}, \mathbf{x})] + C$$

- To maximize ELBO $\mathcal{L}$ w.r.t. latent variable $\mathbf{z}_k$, want to minimize KL divergence
- KL divergence is minimized when distributions are equal

$$q(\mathbf{z}_k) = \tilde{p}(\mathbf{z}_k, \mathbf{X})$$
$$\log q(\mathbf{z}_k) = \log \tilde{p}(\mathbf{z}_k, \mathbf{X}) \propto \mathbb{E}_{z_{\setminus k}}[\log p(\mathbf{z}, \mathbf{x})]$$

CAVI Algorithm: Pseudocode[2]

Algorithm CAVI

while *ELBO* is **not** converged **do**:
  for $k \in \{1, \ldots, K\}$ **do**:
    Set $q(z_k) \propto \exp\{\mathbb{E}_{\setminus k}[\log p(z_k, \mathbf{z}_{\setminus k}, \mathbf{X})]\}$
  end
  Compute $\mathcal{L}(z) = \mathbb{E}[\log p(\mathbf{z}, \mathbf{X})] - \mathbb{E}[\log q(\mathbf{z})]$
end

[2] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe (Apr. 2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: 10.1080/01621459.2017.1285773

CAVI Algorithm: Demo

Demo in Google Colab [here]

Table of Contents

SparsePro: The Fine-Mapping Problem

- **Genotype:** genetic information e.g. single-nucleotide polymorphisms (SNPs)
- **Phenotype:** observed physical trait e.g. height
- **Goal:** Determine which genotypes cause a phenotypic trait

[3] W. Zhang, H. Najafabadi, and Y. Li (2021). "SparsePro: an efficient genome-wide fine-mapping method integrating summary statistics and functional annotations". In: *bioRxiv*. DOI: 10.1101/2021.10.04.463133. eprint: https://www.biorxiv.org/content/early/2021/11/02/2021.10.04.463133.full.pdf

[4] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens (2020). "A simple new approach to variable selection in regression, with application to genetic fine-mapping". In: *bioRxiv*. DOI: 10.1101/501114. eprint: https://www.biorxiv.org/content/early/2020/06/27/501114.full.pdf

SparsePro: The Fine-Mapping Problem

- **Genotype:** genetic information e.g. single-nucleotide polymorphisms (SNPs)
- **Phenotype:** observed physical trait e.g. height
- **Goal:** Determine which genotypes cause a phenotypic trait
- **Assumption:** Phenotypic trait is the result of the sum of K individual effects
- **Literature:** Refer to [3] and [4] below

[3] W. Zhang, H. Najafabadi, and Y. Li (2021). "SparsePro: an efficient genome-wide fine-mapping method integrating summary statistics and functional annotations". In: *bioRxiv*. DOI: 10.1101/2021.10.04.463133. eprint: https://www.biorxiv.org/content/early/2021/11/02/2021.10.04.463133.full.pdf

[4] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens (2020). "A simple new approach to variable selection in regression, with application to genetic fine-mapping". In: *bioRxiv*. DOI: 10.1101/501114. eprint: https://www.biorxiv.org/content/early/2020/06/27/501114.full.pdf

SparsePro: Model Definition

Assume generative process from genotype data $\boldsymbol{X}$ to phenotype data $\boldsymbol{y}$

Sum of Single Effects Model

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{S}\boldsymbol{\beta} + \epsilon = \sum_{k=1}^{K} \boldsymbol{X}\mathbf{S}_k \beta_k + \epsilon$$

SparsePro: Model Definition

Assume generative process from genotype data $\boldsymbol{X}$ to phenotype data $\boldsymbol{y}$

Sum of Single Effects Model

$$\mathbf{y} = \boldsymbol{XS\beta} + \boldsymbol{\epsilon} = \sum_{k=1}^{K} \boldsymbol{X}\mathbf{S}_k \beta_k + \boldsymbol{\epsilon}$$

- Genotype data $\boldsymbol{X}_{N \times G}$: G SNP measurements for N ppl
- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, \tau_{\beta_k}^{-1})$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$
- Phenotype data $\boldsymbol{y}_{N \times 1}$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N \times G}$: 3 SNP measurements for 2 people
- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1} \mathbb{I})$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N\times G}$: 3 SNP measurements for 2 people

$$\boldsymbol{X}_{2\times 3} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N \times G}$: 3 SNP measurements for 2 people
- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$

$$\boldsymbol{X}_{2 \times 3} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N\times G}$: 3 SNP measurements for 2 people
- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$

$$\boldsymbol{X}_{2\times 3} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad S_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N \times G}$: 3 SNP measurements for 2 people
- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$
- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$
- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$
- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$

$$\boldsymbol{X}_{2\times 3} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad S_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\beta_1 = 0.7 \quad \beta_2 = -0.2 \quad \beta_3 = 1.3$$

SparsePro: Example

- Genotype data $\boldsymbol{X}_{N\times G}$: 3 SNP measurements for 2 people

$$\boldsymbol{X}_{2\times 3} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

- Prior probability of gth SNP being causal $\boldsymbol{\pi}_g = \frac{1}{G}$

$$\boldsymbol{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

- Sparse projection $S_k \sim \mathrm{Cat}(\boldsymbol{\pi})$

$$S_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad S_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Causal effect size $\beta_k \sim \mathcal{N}(0, 1)$

$$\beta_1 = 0.7 \quad \beta_2 = -0.2 \quad \beta_3 = 1.3$$

- Noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau_y^{-1}\mathbb{I})$

$$\mathbf{y} = \mathbf{XS}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7 \\ -0.2 \\ 1.3 \end{bmatrix} + \boldsymbol{\epsilon}$$

SparsePro: Connection to Variational Inference

- **Observed Variables** (X): genotype data $\mathbf{X}$ and phenotype data $\mathbf{y}$
- **Latent Variables** (z): sparse indicator vectors $\{\mathbf{S}_1 \dots \mathbf{S}_K\}$ and causal effect sizes $\{\beta_1 \dots \beta_K\}$

SparsePro: Connection to Variational Inference

- **Observed Variables** (X): genotype data $\mathbf{X}$ and phenotype data $\mathbf{y}$
- **Latent Variables** (z): sparse indicator vectors $\{\mathbf{S}_1 \dots \mathbf{S}_K\}$ and causal effect sizes $\{\beta_1 \dots \beta_K\}$
- **Goal**: interested in the intractable true posterior $p(\boldsymbol{S}, \boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y})$
- **Variational Inference**: instead approximate the posterior with $q(\boldsymbol{S}, \boldsymbol{\beta})$ using Variational Inference and the CAVI algorithm

$$q(\boldsymbol{S}, \boldsymbol{\beta}) = \prod_k q(\beta_k \mid \mathbf{S}_k) q(\mathbf{S}_k)$$

Table of Contents

Optimal CAVI Update: Overview

Optimal CAVI Update for SparsePro

$$
\log q^*(z_k) \propto \mathbb{E}_{\setminus k}[\log p(z, x)]
$$
$$
\log q^*(\beta_k, \mathbf{S}_k) \propto \mathbb{E}_{\setminus k}[\log p(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\beta})]
$$

Optimal CAVI Update: Overview

Optimal CAVI Update for SparsePro

$$\log q^*(z_k) \propto \mathbb{E}_{\setminus k}[\log p(z, x)]$$
$$\log q^*(\beta_k, \mathbf{S}_k) \propto \mathbb{E}_{\setminus k}[\log p(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\beta})]$$

- **Goal 1:** compute $\mathbb{E}_{\setminus k}[\log p(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\beta})]$
- **Goal 2:** identify distributions $\log q^*(\beta_k, \mathbf{S}_k)$ with parameters that match $\mathbb{E}_{\setminus k}[\log p(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\beta})]$
- **Note:** ideal CAVI update is w.r.t. the kth latent variables $\beta_k, \mathbf{S}_k$. All terms with a different k are treated as constant and dropped

Optimal CAVI Update

The Log Joint Probability Function

$$\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \beta) = \log \left[p(\mathbf{y}|\mathbf{X}, \mathbf{S}, \beta) \prod_{k\prime} p(\mathbf{S}_{k\prime}) \prod_{k\prime} p(\beta_{k\prime}) \right]$$

Optimal CAVI Update

The Log Joint Probability Function

$$\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta}) = \log\left[p(\mathbf{y}|\mathbf{X}, \mathbf{S}, \boldsymbol{\beta}) \prod_{k\prime} p(\mathbf{S}_{k\prime}) \prod_{k\prime} p(\boldsymbol{\beta}_{k\prime}) \right]$$

$$\begin{aligned}
\mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})] = & \int \log p(\mathbf{y}|\mathbf{X}, \mathbf{S}, \boldsymbol{\beta}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
& + \int \sum_{k\prime} \log p(\mathbf{S}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
& + \int \sum_{k\prime} \log p(\boldsymbol{\beta}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k})
\end{aligned}$$

Optimal CAVI Update: Defining New Variables

Notation Shorthand

$$\mathcal{T}_1 := \int \log p(\mathbf{y}|\mathbf{X}, \mathbf{S}, \boldsymbol{\beta}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k})$$

$$\mathcal{T}_2 := \int \sum_{k\prime} \log p(\mathbf{S}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k})$$

$$\mathcal{T}_3 := \int \sum_{k\prime} \log p(\boldsymbol{\beta}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k})$$

$$\mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})] = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3$$

Optimal CAVI Update: Computing $\mathcal{T}_1$

- Let $\boldsymbol{\mu}_y := \mathbf{XS}\boldsymbol{\beta}$ such that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \tau_y \mathbb{I})$[5]

$$\log p(\mathbf{y} \mid \mathbf{X}, \mathbf{S}, \boldsymbol{\beta}; \boldsymbol{\mu}_y, \tau_y) = \frac{N}{2} \log\left(\frac{1}{2\pi}\right) - \frac{\tau_y}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^T(\mathbf{y} - \boldsymbol{\mu}_y)$$

[5] S. Prince (2012). *Computer Vision: Models Learning and Inference*. Cambridge University Press

Optimal CAVI Update: Computing $\mathcal{T}_1$

Property of Inner Product

$$(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b}) = \sum_n (a_n - b_n)^2$$

$$\begin{aligned}
\mathcal{T}_1 &= \int \left(\frac{N}{2}\log\left(\frac{1}{2\pi}\right) - \frac{\tau_y}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^T(\mathbf{y}-\boldsymbol{\mu}_y) \right) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto -\frac{\tau_y}{2} \int \sum_{k\prime} (\mathbf{y}-\boldsymbol{\mu}_y)^2_{k\prime} \; q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto \tau_y \mathbf{y}^T \mathbf{X}_g \beta_k - \frac{\tau_y}{2} \beta_k^2 \mathbf{X}_g^T \mathbf{X}_g - \tau_y \beta_k \mathbf{X}_g^T \mathbf{X} \tilde{\boldsymbol{\beta}}
\end{aligned}$$

Let $\mathbf{X}_g := \mathbf{X}\mathbf{S}_k$ and $\tilde{\boldsymbol{\beta}} := \mathbb{E}_{q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k})}\left[\sum_{\substack{k\prime=1 \\ k\prime \neq k}}^{K} \mathbf{S}_{k\prime} \boldsymbol{\beta}_{k\prime}\right]$

Optimal CAVI Update: Computing $\mathcal{T}_2$

Categorical PDF[6] and Sparse Indicator Vector $\boldsymbol{S}_k$

$$f(X;\boldsymbol{\pi}) = \prod_{i=1}^{n} \pi_i^{\mathbb{1}[\mathbf{X}_i = 1]}$$

$$\mathbf{S}_{kg} = 1 \quad \mathbf{S}_{k\setminus g} = 0 \qquad \text{by def of causal SNP}$$

$$\begin{aligned}
\mathcal{T}_2 &= \int \sum_{k\prime} \log p(\mathbf{S}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto \int \log \prod_{g\prime} \boldsymbol{\pi}_{g\prime}^{\mathbb{1}[\boldsymbol{S}_{kg\prime}=1]} \, q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto \int \sum_{g\prime} \boldsymbol{S}_{kg\prime} \log \boldsymbol{\pi}_{g\prime} \, q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto \log(\boldsymbol{\pi}_g)
\end{aligned}$$

[6] E. Gordon-Rodriguez, G. Loaiza-Ganem, A. Potapczynski, and J. P. Cunningham (2022). *On the Normalizing Constant of the Continuous Categorical Distribution*. DOI: 10.48550/ARXIV.2204.13290

Optimal CAVI Update: Computing $\mathcal{T}_3$

Normal PDF

$$f(\beta_k; \mu, \tau) := \frac{\tau}{2\pi}^{1/2} \exp(-\frac{\tau}{2}(\beta_k - \mu)^2)$$

$$\begin{aligned}
\mathcal{T}_3 &= \int \sum_{k\prime} \log p(\boldsymbol{\beta}_{k\prime}) q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto \int \left[\frac{1}{2} \log \left(\frac{\tau_{\boldsymbol{\beta}_k}}{2\pi} \right) - \frac{\tau_{\boldsymbol{\beta}_k}}{2} \beta_k^2 \right] q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto -\frac{\tau_{\boldsymbol{\beta}_k}}{2} \beta_k^2 \int q(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) d(\mathbf{S}_{\setminus k}, \boldsymbol{\beta}_{\setminus k}) \\
&\propto -\frac{\tau_{\boldsymbol{\beta}_k}}{2} \beta_k^2
\end{aligned}$$

Optimal CAVI Update

Expectation of Log-Joint

$$\mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$
$$= -\frac{\tau_{\boldsymbol{\beta}_k}}{2}\boldsymbol{\beta}_k^2 - \frac{\tau_y}{2}\boldsymbol{\beta}_k^2 \mathbf{X}_g^T \mathbf{X}_g + \tau_y \boldsymbol{\beta}_k \mathbf{X}_g (\mathbf{y}^T - \mathbf{X}_g^T \tilde{\boldsymbol{\beta}}) + \log(\boldsymbol{\pi}_g)$$

Optimal CAVI Update: Recognizing Distributions

- Recall optimal CAVI update requires

$$\log q(\beta_k, \mathbf{s}_k) \propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$

Optimal CAVI Update: Recognizing Distributions

- Recall optimal CAVI update requires

$$\log q(\beta_k, \mathbf{s}_k) \propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$

- Expand distributions

$$\log q(\beta_k, \mathbf{s}_k) = \log q(\beta_k \mid \mathbf{s}_k) + \log q(\mathbf{s}_k)$$

- Explicitly parameterize distributions

$$q(\beta_k \mid \mathbf{s}_k) \sim \mathcal{N}(\mu^*_{kg}, (\tau^*_{kg})^{-1}) \qquad q(\mathbf{s}_k) \sim \mathrm{Cat}(\boldsymbol{\gamma}^*_{kg})$$

Optimal CAVI Update: Recognizing Distributions

$$\log q(\beta_k, \mathbf{s}_k) = \log q(\beta_k \mid \mathbf{s}_k; \mu^*_{kg}, (\tau^*_{kg})^{-1}) + \log q(\mathbf{s}_k; \gamma^*_{kg})$$

$$\begin{aligned}
&\mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})] \\
&\quad = C - \frac{\tau_{\boldsymbol{\beta}_k}}{2}\boldsymbol{\beta}_k^2 - \frac{\tau_y}{2}\boldsymbol{\beta}_k^2 \mathbf{X}_g^T \mathbf{X}_g + \tau_y \boldsymbol{\beta}_k \mathbf{X}_g(\mathbf{y}^T - \mathbf{X}_g^T \tilde{\boldsymbol{\beta}}) + \log(\boldsymbol{\pi}_g)
\end{aligned}$$

Optimal CAVI Update: Recognizing Distributions

$$\log q(\beta_k, \mathbf{s}_k) = \log q(\beta_k \mid \mathbf{s}_k; \mu^*_{kg}, (\tau^*_{kg})^{-1}) + \log q(\mathbf{s}_k; \boldsymbol{\gamma}^*_{kg})$$

$$\begin{aligned}\mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})] \\ = C - \frac{\tau_{\boldsymbol{\beta}_k}}{2}\boldsymbol{\beta}_k^2 - \frac{\tau_y}{2}\boldsymbol{\beta}_k^2 \mathbf{X}_g^T \mathbf{X}_g + \tau_y \boldsymbol{\beta}_k \mathbf{X}_g(\mathbf{y}^T - \mathbf{X}_g^T \tilde{\boldsymbol{\beta}}) + \log(\boldsymbol{\pi}_g)\end{aligned}$$

Goal: Define variational parameters

$$\mu^*_{kg}, \tau^*_{kg}, \boldsymbol{\gamma}^*_{kg}$$

such that

$$\log q(\beta_k, \mathbf{s}_k) \propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$

Optimal CAVI Update: Recognizing Distributions

- Spoiler: ideal variational parameters have the form

$$\tau_{kg}^{*} := \tau_{y}\mathbf{X}_{g}^{T}\mathbf{X}_{g} + \tau_{\beta_k}$$

$$\mu_{kg}^{*} := \frac{\tau_y}{\tau_{kg}^{*}}\mathbf{X}_{g}^{T}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k})$$

$$\boldsymbol{\gamma}_{kg}^{*} := \mathrm{softmax}\Big(\log \pi_g - \frac{1}{2}\log\frac{\tau_{kg}^{*}}{2\pi} + \frac{1}{2}\tau_{kg}^{*}(\mu_{kg}^{*})^2\Big)$$

Optimal CAVI Update: Recognizing Distributions

- Spoiler: ideal variational parameters have the form

$$
\begin{aligned}
\tau^*_{kg} &:= \tau_y \mathbf{X}_g^T \mathbf{X}_g + \tau_{\beta_k} \\
\mu^*_{kg} &:= \frac{\tau_y}{\tau^*_{kg}} \mathbf{X}_g^T (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k}) \\
\boldsymbol{\gamma}^*_{kg} &:= \mathrm{softmax}\Big(\log \pi_g - \frac{1}{2} \log \frac{\tau^*_{kg}}{2\pi} + \frac{1}{2} \tau^*_{kg} (\mu^*_{kg})^2 \Big)
\end{aligned}
$$

- Proof to follow

Optimal CAVI Update: Recognizing Distributions

Normal PDF and Def of μ^*_{kg}

$$f(X;\mu,\tau^{-1}) := \frac{\tau}{2\pi}^{1/2}\exp(-\frac{\tau}{2}(X-\mu)^2)$$

$$\mu^*_{kg} := \frac{\tau_y}{\tau^*_{kg}}\mathbf{X}_g^T(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k})$$

Optimal CAVI Update: Recognizing Distributions

Normal PDF and Def of μ^*_{kg}

$$f(X;\mu,\tau^{-1}) := \frac{\tau}{2\pi}^{1/2}\exp(-\frac{\tau}{2}(X-\mu)^2)$$

$$\mu^*_{kg} := \frac{\tau_y}{\tau^*_{kg}}\mathbf{X}_g^T(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k})$$

$$\log q(\beta_k \mid \mathbf{s}_k\ ;\ \mu^*_{kg},(\tau^*_{kg})^{-1}) = \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} - \frac{1}{2}(\tau^*_{kg})(\beta_k-\mu^*_{kg})^2$$

$$= \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} - \frac{\tau^*_{kg}\beta_k^2}{2} + \beta_k\tau_y\mathbf{X}_g^T(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k}) + \frac{1}{\tau^*_{kg}}\Big(\tau_y\mathbf{X}_g^T(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k})\Big)^2$$

Optimal CAVI Update: Recognizing Distributions:

Categorical PDF, Def of $\boldsymbol{\gamma}^*_{kg}$, and Sparse Indicator Vector $\boldsymbol{S}_k$

$$f(X;\gamma) = \prod_{i=1}^{n} \gamma_i^{1[\mathbf{X}_i=1]} \quad \boldsymbol{\gamma}^*_{kg} := \mathrm{softmax}\left(\log \pi_g - \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} + \frac{1}{2}\tau^*_{kg}(\mu^*_{kg})^2 \right)$$

$\mathbf{S}_{kg} = 1 \quad \mathbf{S}_{k \setminus g} = 0 \quad$ by def of causal SNP

Optimal CAVI Update: Recognizing Distributions:

Categorical PDF, Def of $\boldsymbol{\gamma}^*_{kg}$, and Sparse Indicator Vector $\boldsymbol{S}_k$

$$f(X;\gamma) = \prod_{i=1}^{n} \gamma_i^{1[\mathbf{X}_i=1]} \qquad \boldsymbol{\gamma}^*_{kg} := \mathrm{softmax}\Big(\log \pi_g - \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} + \frac{1}{2}\tau^*_{kg}(\mu^*_{kg})^2 \Big)$$

$\mathbf{S}_{kg} = 1 \quad \mathbf{S}_{k\setminus g} = 0 \qquad$ by def of causal SNP

$$\begin{aligned}
\log q(\mathbf{s}_k \,;\, \gamma^*_{kg}) &= \log \prod_{g\prime=1}^{G} (\gamma^*_{kg})^{\mathbb{1}[\mathbf{S}_{kg\prime}=1]} \\
&= \log\left(\mathrm{softmax}\Big(\log \pi_g - \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} + \frac{1}{2}\tau^*_{kg}(\mu^*_{kg})^2 \Big)\right) \\
&= C + \log \pi_g - \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} + \frac{1}{2}\tau^*_{kg}(\mu^*_{kg})^2
\end{aligned}$$

Optimal CAVI Update: Recognizing Distributions

Goal of Proof

$$\log q(\beta_k, \mathbf{s}_k) \propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$

Optimal CAVI Update: Recognizing Distributions

Goal of Proof

$$\log q(\beta_k, \mathbf{s}_k) \propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]$$

$$\begin{aligned}
\log q(\beta_k, \mathbf{S}_k) &= \log q(\beta_k \mid \mathbf{s}_k) + \log q(\mathbf{s}_k) \\
&= \left[\frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} - \frac{\tau^*_{kg}}{2}\left(\beta_k^2 - 2\beta_k\mu^*_{kg} + (\mu^*_{kg})^2\right)\right] \\
&\quad + \left[C + \log\pi_g - \frac{1}{2}\log\frac{\tau^*_{kg}}{2\pi} + \frac{1}{2}\tau^*_{kg}(\mu^*_{kg})^2\right] \\
&= C - \frac{\tau_{\boldsymbol{\beta}_k}}{2}\boldsymbol{\beta}_k^2 - \frac{\tau_y}{2}\boldsymbol{\beta}_k^2\mathbf{X}_g^T\mathbf{X}_g + \tau_y\boldsymbol{\beta}_k\mathbf{X}_g(\mathbf{y}^T - \mathbf{X}_g^T\tilde{\boldsymbol{\beta}}) + \log(\boldsymbol{\pi}_g) \\
&\propto \mathbb{E}_{\setminus k}[\log p(\mathbf{y}, \mathbf{X}, \mathbf{S}, \boldsymbol{\beta})]
\end{aligned}$$

Optimal CAVI Update: Recognizing Distributions

Therefore, the ideal CAVI updates for the latent variables $\beta_k, \mathbf{S}_k$ of the kth causal effect is

$$\tau^*_{kg} := \tau_y \mathbf{X}_g^T \mathbf{X}_g + \tau_{\beta_k}$$

$$\mu^*_{kg} := \frac{\tau_y}{\tau^*_{kg}} \mathbf{X}_g^T (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{\setminus k})$$

$$\boldsymbol{\gamma}^*_{kg} := \mathrm{softmax}\Big(\log \pi_g - \frac{1}{2} \log \frac{\tau^*_{kg}}{2\pi} + \frac{1}{2} \tau^*_{kg} (\mu^*_{kg})^2 \Big)_g$$

for ideal update distributions

$$\log q(\beta_k, \mathbf{s}_k) = \log q(\beta_k \mid \mathbf{s}_k; \mu^*_{kg}, (\tau^*_{kg})^{-1}) + \log q(\mathbf{s}_k; \boldsymbol{\gamma}^*_{kg})$$

Table of Contents

Summary

- Variational Inference:
  - Want to estimate latent variables for probability models that explain observed data
  - Seek posterior $p(\mathbf{z} \mid \mathbf{X})$, but can't compute it so we approximate with q
  - Measure how close p and q are with KL divergence, but can't compute it so we use ELBO $\mathcal{L}$
- CAVI:
  - Optimize $\mathcal{L}$ using CAVI
- SparsePro:
  - Model to solve the Fine-Mapping Problem
  - Linear regression $\mathbf{y} = \mathbf{XS}\boldsymbol{\beta}$
- Tech Nugget
  - Proved and calculated the ideal CAVI update for SparsePro
  - Found ideal parameters for ideal update

Future of the Field

- Generic Variational Inference: no model-specific computation, BBVI
- Scalable Variational Inference: works with large datasets, SVI
- Amortized Variational Inference: variational auto-encoder [7]
- Different Divergence Measures: instead of KL divergence, use α-divergence (Daudel and Douc, 2021)

[7] D. P. Kingma and M. Welling (2019). "An Introduction to Variational Autoencoders". In: *Foundations and Trends® in Machine Learning* 12.4, pp. 307–392. DOI: 10.1561/2200000056

Personal Takeaways/Future Plans

Gilad

- Continue my Comp Bio research → enjoy it a lot
- Applying to Masters programs in CS, Applied Math, Stat, or Comp Bio
- Learned how to be meticulous in mathematical notation and derivations

Austin

- Continue doing research in variational inference among other statistics/ML topics
- Applying to graduate school for Masters programs in ML/Applied Math/Statistics