

# Averaged Variational Inference for Hierarchical Modelling of

## Genetic Association

William van Rooij

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

william.vanrooij@epfl.ch



### 1. Introduction

In **expression quantitative trait locus** (eQTL) studies, we estimate the association between **single nucleotide polymorphisms** (SNPs) and a gene expression phenotype, a **trait**. It is a *small n, large p* situation where

- $n$  is the number of observations,
- $p$  is the number of SNPs.

In a *small n, large p* situation, the traditional methods such as **Markov Chain Monte Carlo** (MCMC) do not apply because of the cost of the computation being too large. As an alternative, we use **variational inference** from Blei et al. [2017].

Moreover, the data is highly correlated with a block structure, which complicates inference and interpretation. Here, we seek to enhance regression approaches in such difficult settings.

### 2. Hierarchical Regression Model

Let  $\mathbf{X} = (X_1, \dots, X_p)$  represent the SNPs, and  $\mathbf{y}$  represent the trait observed.  $\mathbf{y}$  is linearly related with the predictors  $\mathbf{X}$  and has residual precision  $\tau$ . We suppose that they follow the following hierarchical model, for all  $s = 1, \dots, p$ :

$$\begin{aligned} \mathbf{y} \mid \boldsymbol{\beta}, \tau &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \tau^{-1}\mathbf{I}_n), \\ \beta_s \mid \gamma_s, \tau, \sigma^{-2} &\sim \gamma_s \mathcal{N}(0, \sigma^2 \tau^{-1}) + (1 - \gamma_s) \delta_0, \\ \gamma_s \mid \omega_s &\sim \text{Bernoulli}(\omega_s), \\ \omega_s &\sim \text{Beta}(a_s, b_s). \end{aligned}$$

where  $\delta_0$  is the Dirac distribution, and  $a_s$  and  $b_s$  are chosen to enforce sparsity.  $\gamma$  is a binary matrix that indicates which SNP is associated with the trait, i.e.,

$$\gamma_s = 1 \iff \text{SNP } s \text{ is associated with the trait.}$$

### 3. Variational Inference

Given a family of densities  $\mathcal{D}$  over the parameters  $\boldsymbol{\theta}$ , variational inference approximates our density of interest  $p(\boldsymbol{\theta} \mid \mathbf{y})$  by a distribution  $q \in \mathcal{D}$  minimizing the “reverse” **Kullback–Leibler divergence**, by Kullback and Leibler [1951],

$$\text{KL}(q \parallel p) = \int q(\boldsymbol{\theta}) \log \left\{ \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \mathbf{y})} \right\}.$$

We introduce the **evidence lower bound**,

$$\mathcal{L}(q) = \mathbb{E}_q[\log p(\boldsymbol{\theta}, \mathbf{y})] - \mathbb{E}_q[\log q(\boldsymbol{\theta})] = \int q(\boldsymbol{\theta}) \log \frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta},$$

where  $\mathbb{E}_q$  represents the expectation with respect to  $q$ .

The lower bound verifies

$$\text{KL}(q \parallel p) = \log(p) - \mathcal{L}(q).$$

As the Kullback–Leibler divergence can be difficult to minimize, variational inference maximizes the lower bound instead.

When applied to highly correlated data, variational inference **underestimates** posterior variances, as explained in Blei et al. [2017].

- The lower bound  $\mathcal{L}(q)$  tends to be **highly multimodal**,
- the mean-field independence assumption,
- the reverse Kullback–Leibler divergence optimisation,

tend to concentrate mass on a single mode and neglect the rest.

To better handle the multimodality, we build a method consisting of averaging over multiple parameter initialisations with weights equal to the posterior model probability corresponding to the obtained mode, based on Ruffieux et al. [2017].

### 4. Averaged LOCUS

Assume that the data  $\mathbf{y}$  has been obtained from a one of  $K$  models  $M_k$ ,  $k = 1, \dots, K$ . We want to estimate the association between the SNPs and the trait, i.e., we want to estimate  $\gamma$ .

We perform a weighted average accounting for the likelihood that the data corresponds to each model. The more the model corresponds to the observed data, the more weight it will have in the average, i.e., for all  $s = 1, \dots, p$ ,

$$\mathbb{E}[\gamma_s \mid \mathbf{y}] = \sum_{k=1}^K \mathbb{E}[\gamma_s \mid M_k, \mathbf{y}] p(M_k \mid \mathbf{y}).$$

The posterior probability for model  $M_k$  is

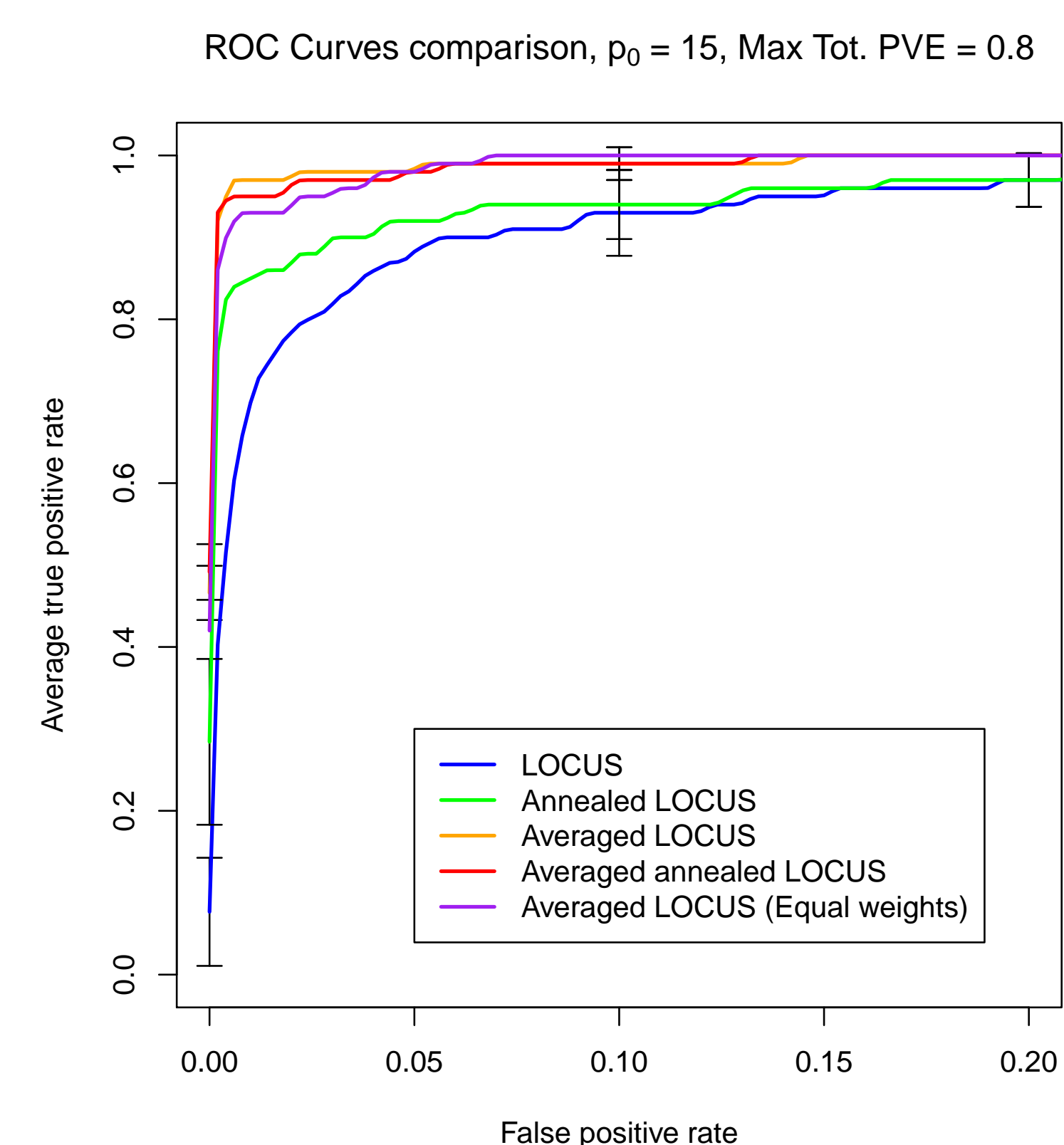
$$p(M_k \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid M_k) p(M_k)}{\sum_{j=1}^K p(\mathbf{y} \mid M_j) p(M_j)} \approx \frac{\mathcal{L}(q) p(M_k)}{\sum_{j=1}^K \mathcal{L}(q) p(M_j)}$$

where  $p(\mathbf{y} \mid M_k)$  is the likelihood under model  $M_k$ , approximated by  $\mathcal{L}(q)$ , and  $p(M_k)$  is the prior probability of model  $M_k$ .

### 5. Performance

We compare the performance of variable selection of five methods,

- classical variational inference “LOCUS”,
- averaged variational inference “averaged LOCUS”,
- their annealing augmented equivalents “annealed LOCUS” and “averaged annealed LOCUS”,
- averaged variational inference with equal weights “averaged LOCUS (Equal weights)”



**Figure 1:** Comparison of ROC curves between LOCUS, averaged LOCUS, their respective annealed versions, and averaged LOCUS with equal weights, colored in blue, orange, green, red, and purple respectively. The data involve 300 observations, 15 SNPs out of 500 SNPs are associated with the trait. The response variance explained by the SNPs is below 80%.

The three averaging methods have similar variable selection performance, which is better than the standard LOCUS method’s. The annealing step does not improve the performance of variable selection.

### 6. Conclusion

The averaged variational inference better handles the multimodality of the posterior and the shows the uncertainty due to the strong correlation structure in the data.

The averaged methods can be implemented in parallel, drastically diminishing the runtime. The runtime of the averaged annealing LOCUS is smaller than the runtime of averaged LOCUS.

### References

- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112:859–877.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86.
- Ruffieux, H., Davison, A. C., Hager, J., and Irincheeva, I. (2017). Efficient inference for genetic association studies with multiple outcomes. *Biostatistics*, 18:618–636.