Supplementary Materials: Probabilistic Modeling of Imaging, Genetics and Diagnosis

Nematollah K. Batmanghelich, Adrian Dalca, Gerald Quon, Mert Sabuncu, Polina Golland, for the Alzheimer's Disease Neuroimaging Initiative*

I. Further Explanation about the Experiments on ADNI

Fig.1 shows the probability densities of the hyper-parameters used for the ADNI dataset.

List of the SNPs and the corresponding genes are provided in the Table I.

II. Variational Bayes to Approximate the Posteriors

The pseudo-code for our inference algorithm is shown in Algorithm 1, where $\boldsymbol{\gamma}$ is a M-dimensional vector of Bayes Factor computed as follows:

$$\log p(\mathbf{X}_{:m}|b_m;\mathbf{G},\pi) = \sum_{n=1}^{N}\log\mathcal{N}(x_{nm};0,1)+ b_m\left(\log p(\mathbf{X}_{:m}|b_m=1;\mathbf{G},\pi) - \sum_{n=1}^{N}\log\mathcal{N}(x_{nm};0,1)\right) \quad (1)$$

The computationally difficult term in the equation is the marginal likelihood, $p(\mathbf{x}_m|\mathbf{b}_m = 1; G, \pi)$. Exact computation of the marginal likelihood is computationally intractable. However, it is common to approximate it with a lower bound of the variational energy [1]–[3]. We follow the variational mean-field method proposed by Carbonetto *et al.* [4] with a slight modification to approximate $p(\mathbf{X}_{:m}|\mathbf{b}_m = 1; G, \pi)$. To be self-contained, we first briefly summarize the method in [4]:

- We descretize the hyper-parameter space of the imaging part of the model, *i.e.,* $\pi' := \{\log_{10}\alpha, \sigma_0^2, \sigma_\omega^2\} \subset \pi$ into uniform grids, namely $[\alpha(\min), \alpha(\max)] \times [\sigma_0^2(\min), \sigma_0^2(\max)] \times [\sigma_\omega^2(\min), \sigma_\omega^2(\max)]$. Let us call the grid points $\pi'(1), \cdots, \pi'(L)$, where every tuple $\pi' = (\alpha(i), \sigma_0^2(i), \sigma_\omega^2(i))$ is a set of hyper-parameter values.
- Since the space of the hyper-parameters is low dimensional, importance sampling is a simple and effective way to integrate out the hyper-priors with a reasonably small number of samples. The proposal distribution is chosen to be a uniform distribution over a sufficiently large range, *i.e.,* $\tilde{p}(\pi'(i)) = \tilde{p}(\pi'(1))$, where $\tilde{p}(\cdot)$ is the proposal distribution.
- Given a set of hyper-parameter values $\pi'(i)$, a mean-field approach is used to approximate the marginal likelihood, $p(\mathbf{x}_m|\mathbf{b}_m = 1; G, \pi')$, via the variational lower bound. Briefly, the mean-field method maximizes the following objective function:

$$\begin{aligned}
&\log p(\mathbf{x}|b=1;\mathbf{G},\pi') \\
&\geq F(\pi';\boldsymbol{\varsigma},\boldsymbol{\nu},\boldsymbol{\tau}) \equiv \mathbb{E}_q\left[\log\frac{p(\mathbf{x},\boldsymbol{\omega},\mathbf{a}|\mathbf{G},\pi')}{q(\boldsymbol{\varsigma},\boldsymbol{\nu},\boldsymbol{\tau})}\right] \\
&= -N\log(\sigma_0) - \frac{\|\mathbf{x}-\mathbf{G}\mathbf{r}\|^2}{2\sigma_0^2} \\
&-\frac{1}{2\sigma_0^2}\sum_{k=1}^{S}\left(\mathbf{G}^T\mathbf{G}\right)_{kk}\mathbb{V}ar_q[\omega_k] \\
&-\sum_{k=1}^{S}\tau_k\log\left(\frac{\tau_k}{\alpha}\right) - \sum_{k=1}^{S}(1-\tau_k)\log\left(\frac{1-\tau_k}{1-\alpha}\right) \\
&+\sum_{k=1}^{S}\frac{\tau_k}{2}\left[1+\log\left(\frac{\varsigma_k^2}{\sigma_0^2\sigma_\omega^2}\right) - \frac{\varsigma_k^2+\nu_k^2}{\sigma_0^2\sigma_\omega^2}\right]
\end{aligned} \quad (2)$$

  where $\boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$ are the parameters of the approximate posterior, q and $\mathbf{r} := \mathbb{E}[\boldsymbol{\omega}_q] = \boldsymbol{\tau} \odot \boldsymbol{\nu}$ and $\mathbb{V}ar_q[\omega_k] = \tau_k(\varsigma_k^2 + \nu_k^2) - (\tau_k\nu_k)^2$.
- Finally, the importance weights, $\zeta(i)$'s, are normalized. For each of $\boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau}$, a weighted sum over all $\pi'(i)$ is computed as an approximation to integrating out the hyper-parameters.

This procedure needs is run for every brain region. The pseudo-code of the algorithm is shown in Algorithm 2 [4].

To integrate out the hyper-priors, the main idea in [4] is to use importance sampling to compute the following integral:

$$\text{PIP}(s,m) = \int p(a_{sm}=1|\mathbf{G},\mathbf{x}_m,\pi')p(\pi'|\mathbf{G},\mathbf{x}_m)d\pi', \quad (3)$$

where $\text{PIP}(s,m)$ denotes the Posterior Inclusion Probability for SNP s and region m. Carbonetto *et al.* [4] suggest to replace it with the following importance sampling estimate

$$\text{PIP}(s,m) = \frac{\sum_{i=1}^{L} p(a_{sm}=1|\mathbf{G},\mathbf{x}_m,\pi'(i))\zeta(\pi'(i))}{\sum_{i=1}^{L}\zeta(\pi'(i))} \quad (4)$$

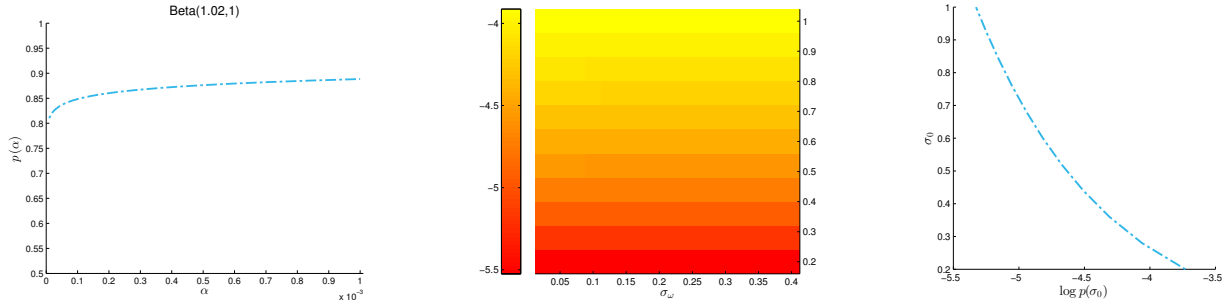
where $\zeta(\pi'(i))$ is the normalized importance weight for $\pi'(i)$. According to the importance sampling procedure,

$$\zeta(\pi') = \frac{p(\mathbf{x}_m|\mathbf{G},\pi')p(\pi')}{\tilde{p}(\pi')}, \quad (5)$$

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

SNP	Gene	SNP	Gene	SNP	Gene	SNP	Gene
rs10106827	NECAB1,TMEM55A	rs12476069	COL6A3,MLPH	rs9393059	FOXQ1,HUS1B	rs71327107	EP300,RBX1
rs10487075	C7orf62,STEAP2-AS1	rs12535226	EGFR,LANCL2	rs2906657	PILRA,ZCWPW1	rs74322721	DDHD1,FERMT2
rs10504488	EYA1,XKR9	rs12778247	ANXA8,ZNF488	rs293168	NDUFA4,NXPH1	rs75340942	HABP2,TCF7L2
rs10812555	C9orf11,LINC00032	rs12997264	ATG16L1,INPP5D	rs34380708	KIAA0317,LTBP2	rs7536931	CR1,CR1L
rs111863968	ATG16L1,INPP5D	rs13040601	CBLN4,DOK5	rs3764648	ABCA7,HMHA1	rs76222305	LRRTM1,SUCLG1
rs113814152	CHRNA2,PTK2B	rs13138250	FGFRL1,IDUA	rs3779632	CHRNA2,PTK2B	rs76448372	AMICA1,SCN2B
rs114773661	CRBN,SUMF1	rs13314819	BBX,CCDC54	rs4133300	KCNJ3,NR4A2	rs76822114	GC,SLC4A4
rs114956101	KCNK17,KCNK5	rs145767144	MAF,WWOX	rs4916928	.,FAM20C	rs76978231	CSMD1,MCPH1
rs115815527	ASB5,SPCS3	rs146373627	ECHDC3,USP6NL	rs56034708	CDC7,TGFBR3	rs77271157	CLU,EPHX2
rs11662059	ACAA2,LIPG	rs146643250	DNAH5,TRIO	rs57677986	ADAM10,FAM63B	rs77287774	ZEB1-AS1,ZNF438
rs117119586	MTDH,TSPYL5	rs147030865	ATP8B4,SLC27A2	rs59776273	CORIN,NFXL1	rs7812465	PLEKHF2,TP53INP1
rs117281307	CTTNBP2,NAA38	rs16849237	RHOU,TMEM78	rs6020063	B4GALT5,SLC9A8	rs78180796	PTPRM,RAB12
rs117547283	POM121L1P,PRAME	rs17108960	CBX5,SMUG1	rs622354	OR10G7,VWA5A	rs79079416	CSNK1G1,KIAA0101
rs117655211	ATP8B4,DTWD1	rs17781348	GAK,TMEM175	rs62389386	CLK4,COL23A1	rs792806	CA10,KIF2B
rs1178036	GNRH2,PTPRA	rs1806522	C3orf27,RPN1	rs6571632	EGLN3,SPTSSA	rs8030340	RSL24D1,UNC13C
rs117984432	ANKRD11,SPG7	rs1834554	MS4A4A,MS4A4E	rs6685242	CD46,CR1L	rs8707	MAP3K12,PCBP2
rs118091716	GRM3,SEMA3D	rs1912718	ATOH1,GRID2	rs6934812	CD2AP,TNFRSF21	rs2824734	LINC00320,TMPRSS15
rs118192075	IMPACT,OSBPL1A	rs2048330	LINC00210,RRP15	rs6949677	C7orf70,CYTH3	rs28592859	HLA-DQA1,HLA-DQB1
rs11875667	CETN1,COLEC12	rs2048330	LINC00210,RRP15	rs7027316	IFNE,MTAP	$\epsilon 3/\epsilon 4$	APOE
rs12002176	C9orf170,DAPK1	rs2136987	CCKAR,RBPJ	rs7068614	ECHDC3,USP6NL		
rs12137076	GNG4,LYST	rs2701623	DTX1,RASAL1	rs7087150	CCNY,GJD4		
rs12198405	CD2AP,TNFRSF21	rs79914380	LMO3,MGST1	rs7129687	EED,PICALM		

TABLE II: Detected SNPs and the corresponding genes.



(a) (b) (c)

Fig. 1: Prior probability density for different hyper-parameters: (a) Density of α for prior distribution of Beta$(1.02, 1)$. The prior assigns almost uniform weight to all values. (b) log of PVE density as a function of σ_ω (x-axis) and σ_0 (y-axis). PVE functions as a prior for σ_ω given a value of σ_0. (c) log of the density of σ_0.

where the numerator is proportional to $p(\pi'|\mathbf{x}_m, \mathbf{G})$ and the denominator is the proposal distribution, a uniform distribution in our experiments. Since the marginal likelihood $p(\mathbf{x}_m|\mathbf{G}, \pi')$ cannot be computed directly, it is approximated by the highest lower bound:

$$\log p(\mathbf{x}_m|\mathbf{G}, \pi') \geq F(\pi'; \boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau}). \tag{6}$$

To approximate the $p(\mathbf{x}_m|\mathbf{G})$, we can apply the following procedure:

$$\begin{aligned} p(\mathbf{x}_m|\mathbf{G}) &= \int p(\mathbf{x}_m, \pi'|\mathbf{G}) dp(\pi'|\mathbf{G}) = \mathbb{E}_{\pi'|\mathbf{G}}\left[p(\mathbf{x}_m, \pi'|\mathbf{G})\right] \\ &\geq \mathbb{E}_{\pi'|\mathbf{G}}\left[e^{F(\pi'; \boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau})}\right] \\ &\geq \exp\left[\mathbb{E}_{\pi'|\mathbf{G}}\left[F(\pi'; \boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau})\right]\right], \end{aligned} \tag{7}$$

where the last line in Eq. (7) follows from the convexity of exponential function.

Similar to Eq. (4), the idea is to replace the expectation with the importance sampling approximation:

$$\mathbb{E}_{\pi'|\mathbf{G}}\left[F(\pi'; \boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau})\right] \approx \frac{\sum_{i=1}^{L} F(\pi'; \boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \zeta(\pi'(i))}{\sum_{i=1}^{L} \zeta(\pi'(i))}. \tag{8}$$

REFERENCES

[1] M. Beal and Z. Ghahramani, "The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian Statistics*, vol. 7, pp. 1–10, 2003.

[2] R. B. Grosse, "Model selection in compositional spaces," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[3] C. Ji, H. Shen, and M. West, "Bounded approximations for marginal likelihoods," 2010.

[4] P. Carbonetto and M. Stephens, "Scalable Variational Inference for Bayesian Variable Selection in Regression, and its Accuracy in Genetic Association Studies," *Bayesian Analysis*, vol. 7, pp. 73–108, 2012.

Algorithm 1: Variational Learning to Approximate Posterior Relevance of Brain Regions

Parameters: (a) prior for the regions $\boldsymbol{\alpha}$, (b) number of iterations: T

Data: (a) Diagnosis $\mathbf{y}$, (b) Imaging Data $\mathbf{X}$, (c) Genotype $\mathbf{G}$

Output: Parameters of the posterior distribution: $\boldsymbol{\theta}$

Approximate Bayes Factors ($\boldsymbol{\gamma}$)

  for $m \leftarrow 1$ ***to*** M **do**

    Set γ_m to an approximation of Eq. (1) (see Algorithm 2)

for $i \leftarrow 1$ ***to*** T **do**

  Draw a set from the current estimate of the posterior distribution: $\mathbf{b}^t \sim q_{\tilde{\boldsymbol{\theta}}^t}$;

  Approximate the marginal conditional likelihood $p(\mathbf{y}|\mathbf{b}^t; \mathbf{X}, \pi)$;

  Set $\hat{\mathbf{g}}^t = \tilde{\mathbf{b}}^t \left(\log p(\mathbf{y}|\mathbf{b}^t) + \boldsymbol{\gamma}^T \mathbf{b}^t + |\mathbf{b}^t| \log \alpha \right)$;

  Set $\hat{\mathbf{C}}^t = \tilde{\mathbf{b}}^t (\tilde{\mathbf{b}}^t)^T$;

  Set $\mathbf{g}^{t+1} = (1-w)\mathbf{g}^t + w\tilde{\mathbf{g}}^t$;

  Set $\mathbf{C}^{t+1} = (1-w)\mathbf{C}^t + w\tilde{\mathbf{C}}^t$;

  Solve $\mathbf{C}^{t+1}\tilde{\boldsymbol{\theta}}_{t+1} = \mathbf{g}^{t+1}$;

  if $t > N/2$ **then**

    Set $\bar{\mathbf{g}} = \bar{\mathbf{g}} + \hat{\mathbf{g}}^t$;

    Set $\bar{\mathbf{C}} = \bar{\mathbf{C}} + \hat{\mathbf{C}}^t$;

return *the solution of a linear system of equations* $\bar{\mathbf{C}}\boldsymbol{\theta} = \bar{\mathbf{g}}$;

Algorithm 2: Variational Learning to Approximate $p(\mathbf{x}|b=1;\mathbf{G})$

Data: (a) Imaging features for one region $\mathbf{x}$, (b) Genotype $\mathbf{G}$

Parameters: Set of hyper-parameters, $\pi'(1), \cdots, \pi'(L)$

Output

  (a) Approximate marginal likelihood $\hat{\gamma} \approx \log p(\mathbf{x}|b=1;\mathbf{G})$;

  (b) Variational estimate of posterior inclusion probability $\hat{\tau}_s := \mathbb{E}_q[a_s|\mathbf{x}, b=1;\mathbf{G}], \quad \forall 1 \leq s \leq S$;

  (c) Variational estimate of posterior variance $(\hat{\boldsymbol{\varsigma}} := \mathbb{E}_q[\boldsymbol{\omega}|\mathbf{x}, b=1;\mathbf{G}])$;

  (d) Variational estimate of posterior variance $(\hat{\boldsymbol{\nu}} := \mathbb{V}ar_q[\boldsymbol{\omega}|\mathbf{x}, b=1;\mathbf{G}])$;

for $i \leftarrow 1$ ***to*** L **do**

  Initialize $(\boldsymbol{\varsigma}_{\text{Init}}, \boldsymbol{\nu}_{\text{Init}}, \boldsymbol{\tau}_{\text{Init}})$ randomly ;

  $(\boldsymbol{\varsigma}(i), \boldsymbol{\nu}(i), \boldsymbol{\tau}(i), Z(i)) \leftarrow$ **Mean-Field**$(\mathbf{G}, \mathbf{x}, \pi'(i))$;

Set $(\boldsymbol{\varsigma}_{\text{Init}}, \boldsymbol{\nu}_{\text{Init}}, \boldsymbol{\tau}_{\text{Init}})$ to the parameters associated with highest Z ;

for $i \leftarrow 1$ ***to*** L **do**

  $(\boldsymbol{\varsigma}(i), \boldsymbol{\nu}(i), \boldsymbol{\tau}(i), Z(i)) \leftarrow$ **Mean-Field**$(\mathbf{G}, \mathbf{x}, \pi'(i))$;

  Compute importance weight $\zeta(i) \leftarrow Z(i)p(\pi'(i))/\tilde{p}(\pi'(i))$;

Normalize importance weights: $\zeta(i) \leftarrow \zeta(i)/(\sum_i \zeta(i))$;

Average Over Hyper-parameters

  $\hat{\boldsymbol{\nu}} \leftarrow \sum_{i=1}^{L} \zeta(i)\boldsymbol{\nu}(i)$;

  $\hat{\boldsymbol{\varsigma}} \leftarrow \sum_{i=1}^{L} \zeta(i)\boldsymbol{\varsigma}(i)$;

  $\hat{\boldsymbol{\tau}} \leftarrow \sum_{i=1}^{L} \zeta(i)\boldsymbol{\tau}(i)$;

  $\hat{\gamma} \leftarrow \sum_{i=1}^{L} \zeta(i) \left(\log Z(i)\right)$;

return $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\varsigma}}, \hat{\boldsymbol{\tau}}, \hat{\gamma})$;

Mean-Field Subroutine

  Input: (a) Genotype $(\mathbf{G})$, (b) Response $(\mathbf{y})$, (c) Hyper-parameters (π')

  $(\boldsymbol{\varsigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \leftarrow (\boldsymbol{\varsigma}_{\text{Init}}, \boldsymbol{\nu}_{\text{Init}}, \boldsymbol{\tau}_{\text{Init}})$;

  repeat

    Choose $s \in \{1, \cdots, S\}$;

    $\varsigma_s^2 \leftarrow \sigma_0^{-2} \left((\mathbf{G}^T\mathbf{G})_{ss} + 1/\sigma_\omega^2 \right)$;

    $\nu_s \leftarrow \varsigma_s^2 \sigma_0^{-2} \left((\mathbf{G}^T\mathbf{y})_s - \sum_{j \neq s} (\mathbf{G}^T\mathbf{G})_{js} \tau_j \nu_j \right)$;

    $\frac{\tau_s}{1-\tau_s} \leftarrow \frac{\alpha}{1-\alpha} \times \frac{\varsigma_s}{\sigma_0 \sigma_\omega} \times \exp(\frac{1}{2}\nu_s^2/\varsigma_s^2)$

  until *Convergence*;

  Set $\log Z$ to the approximate lower bound by Eq. (2) ;

  return $\boldsymbol{\varsigma}$, $\boldsymbol{\nu}$, $\boldsymbol{\tau}$, Z ;