

Variable selection in logistic regression with a latent Gaussian field models for analysis of epigenetic data

Hubin A.A., Storvik G.O.

Grini P.E., Lingjærde O.C., Butenko M.A.

University of Oslo

aliaksah@math.uio.no



UiO : Universitetet i Oslo

Belarusian State University, Minsk

28.03.2016

Introduction. Biological motivation

- More precise estimation of the **methylation probability** of locations, which is represented by a number a binary events for all reads per given location
- Discovery of methylated and unmethylated regions and corresponding **local** and **global** structures:
 - Represented by nucleotides sequences patterns (**CPG-islands**)
 - Represented by such structures as **genes on the whole**, **promoters**, **coding regions** and their sequences
- Linking genetic and epigenetic data to **phenotypic responses** (levels of **expression** of genes, presence of **transposons**, etc.)
- Finding other **covariates** (special genetic signatures) significantly influencing methylation patterns along the genome

Introduction. Responses visualization

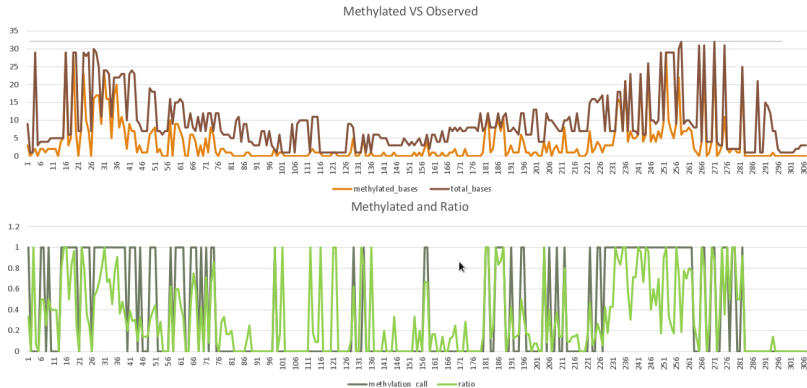


Figure: Total reads and methylated reads for some part of the genome

The model (GLMM): Hierarchical Bayesian Model

The model: Logistic Regression With a Gaussian Latent Field Model (*Logistic Regression With a Random Effect Model*)

$$\Pr(y_t = y | n_t = n, p_t) = \binom{n}{y} p_t^y (1 - p_t)^{n-y} \quad (1)$$

$$p_t = \frac{e^{\beta_0 + \sum_{i=1}^M \beta_i X_{t,i} + \delta_t}}{1 + e^{\beta_0 + \sum_{i=1}^M \beta_i X_{t,i} + \delta_t}} \quad (2)$$

$$\delta_t = \rho \delta_{t-1} + \epsilon_t \quad (3)$$

$$\epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (4)$$

- $y_t \in \{1, \dots, T\}$ is the number of methylated reads per loci t
- $n_t \in \mathbb{N}$ is the total number of reads per loci t
- $\beta_i \in \mathbb{R}, i \in \{0, \dots, M\}$ are regression coefficients of the covariates of the model
- δ_t is a Gaussian random effect of $AR(1)$ type with a parameter $\rho \in \mathbb{R}$
- ϵ_t is the error term of $AR(1)$

The model

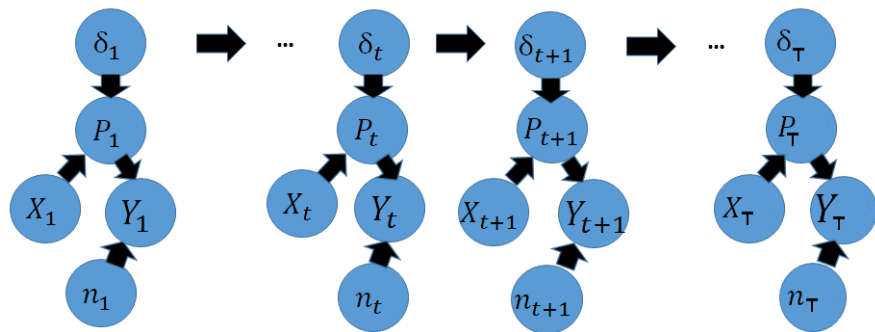


Figure: The model

T is extremely large \Rightarrow Big Data

The model: Hyper-parameters of the model

We use a fully Bayesian approach, hence specify priors

$$\beta_i \sim N(\mu_\beta, \sigma_\beta^2) \quad (5)$$

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \sim N_2(\mu_{\rho, \epsilon}, \Sigma_{\rho, \epsilon}) \quad (6)$$

- $\psi_1 = \log \frac{1}{\sigma_\epsilon^2} (1 - \rho^2)$ and $\psi_2 = \log \frac{1+\rho}{1-\rho}$ are scaled hyper-parameters of the latent model

The model: Model Selection

Let $\Theta = \{\vec{\beta}, \rho, \sigma_\epsilon^2\}$ define parameters of the model and $\mathbb{M} : \vec{\gamma}$ define a model itself, i.e. which covariates are addressed, then:

$$p_t = \frac{e^{\gamma_0 \beta_0 + \sum_{i=1}^p \gamma_i \beta_i X_{t,i} + \delta_t}}{1 + e^{\gamma_0 \beta_0 + \sum_{i=1}^p \gamma_i \beta_i X_{t,i} + \delta_t}} \quad (7)$$

$$\beta_i | \gamma_i \sim \mathbb{I}(\gamma_i = 1) N(\mu_\beta, \sigma_\beta^2) \quad (8)$$

$$\gamma_i \sim \text{Binom}(1, q) \quad (9)$$

- $\gamma_i \in \{0, 1\}, i \in \{0, \dots, p\}$ are latent indicators, defining if covariate i is included into the model
- q is the prior probability of including any covariate into the model, which corresponds to the spike and slab model

Let:

- $\mathbb{M} = \vec{\gamma}$ be further addressed as simply a model
- $\Theta|\mathbb{M}$ define parameters conditioned on fixed models
- $\exists 2^{p+1}$ different models

Goals:

- $\Pr(\mathbb{M}, \Theta|\mathbb{D})$ posterior distribution of parameters and models
- $\Pr(\mathbb{M}|\mathbb{D})$ marginal posterior distribution of the models
- Set of estimated models performing well in terms of some model selection criteria (MAP, WAIC, DIC, MLIK)

- **Note that** $\Pr(\mathbb{M}, \Theta | \mathbb{D}) = \Pr(\Theta | \mathbb{M}, \mathbb{D}) \Pr(\mathbb{M} | \mathbb{D})$
- $\Pr(\Theta | \mathbb{M}, \mathbb{D})$ and $\log \Pr(\mathbb{D} | \mathbb{M})$ can be efficiently obtained by INLA
- **Note that** $\Pr(\mathbb{M} = M | \mathbb{D}) = \frac{e^{\log \Pr(\mathbb{D} | \mathbb{M} = M) + \log \Pr(\mathbb{M} = M)}}{\sum_{M' \in \Omega_{\mathbb{M}}} e^{\log \Pr(\mathbb{D} | \mathbb{M} = M') + \log \Pr(\mathbb{M} = M')}}$
- $\hat{\Pr}(\mathbb{M} = M | \mathbb{D}) = \frac{e^{\log \Pr(\mathbb{D} | \mathbb{M} = M) + \log \Pr(\mathbb{M} = M)}}{\sum_{M' \in \mathbb{V}} e^{\log \Pr(\mathbb{D} | \mathbb{M} = M') + \log \Pr(\mathbb{M} = M')}}$
- \mathbb{V} is the subspace of $\Omega_{\mathbb{M}}$ to be efficiently explored
- Note that for $\Pr(\mathbb{M} = M) = \Pr(\mathbb{M} = M') \forall M, M' \in \Omega_{\mathbb{M}}$:
- $\Pr(\mathbb{M} = M | \mathbb{D}) \gg \Pr(\mathbb{M} = M' | \mathbb{D})$ if $\log \Pr(\mathbb{D} | \mathbb{M} = M) > \log \Pr(\mathbb{D} | \mathbb{M} = M')$ often \implies
- **Near modal values in terms of MLIK are particularly important** for construction of reasonable $\mathbb{V} \subset \Omega_{\mathbb{M}}$, **missing them can dramatically influence** posterior in the original space $\Omega_{\mathbb{M}}$

Assume

Observation model: $\pi(y|\eta)$

Parameter model: $\pi(\eta|v) \sim N_u(\mu(v), Q(v))$

Hyperparameter: $v \sim f(v)$

The models are assumed to satisfy some properties:

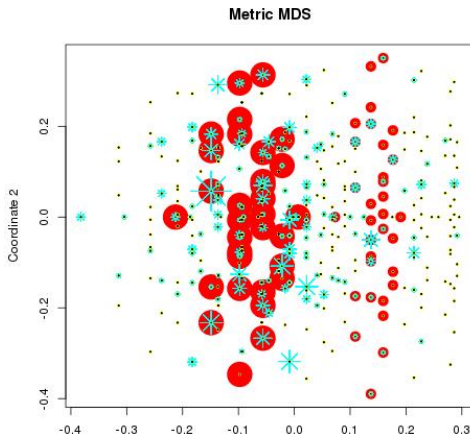
- The parameter can be of big size but with a sparse precision matrix
- The dimension of the hyperparameter vector v is relatively small
- Laplace approximation method of the posterior density can be used

INLA efficiently calculates:

- The marginal posterior distribution of parameters which can be summarized by means, variances and quantiles
- Model selection criteria DIC, WAIC, MLIK (exactly $\log \Pr(\mathbb{D}|\mathbb{M} = M)$)
- Predictive measures (CPO, PIT)

Model selection problems

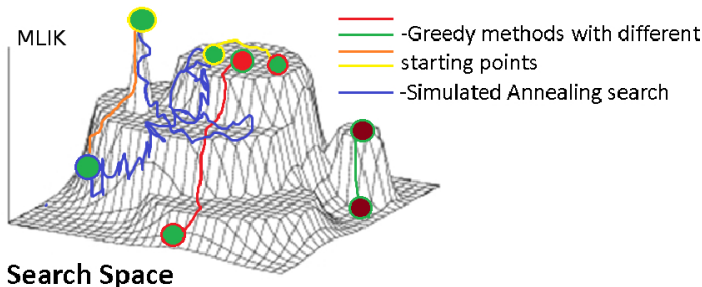
- Proceed with **efficient exploration of \mathbb{V}** in the subspace of $\Omega_{\mathbb{M}}$ **to estimate $\Pr(\mathbb{M} = M|\mathbb{D})$** , $\underset{M \in \Omega_{\mathbb{M}}}{\operatorname{argmax}} \Pr(\mathbb{M} = M|\mathbb{D})$, **and** $\underset{M \in \Omega_{\mathbb{M}}}{\operatorname{argmax}} \operatorname{WAIC}(M)$
- Main challenges are **multimodality** in $\Omega_{\mathbb{M}}$ and its **size**



Possible ways to explore $\mathbb{V} \subset \Omega_{\mathbb{M}}$

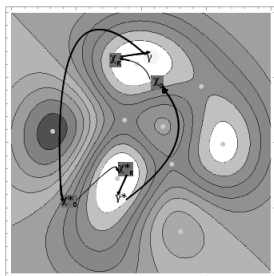
Main challenges are multimodality in $\Omega_{\mathbb{M}}$ and its size.

- Full enumeration of $\Omega_{\mathbb{M}}$ - infeasible for large dimensions
- Random walk in $\Omega_{\mathbb{M}}$ including simple MCMC - does not take advantage of the structure of $\Omega_{\mathbb{M}} \implies$ too slow
- Greedy optimization - end up in local optima
- SA - ends up with random descent with almost no chance to change the mode
- Random walk with mode jumping proposals seems to be a good idea

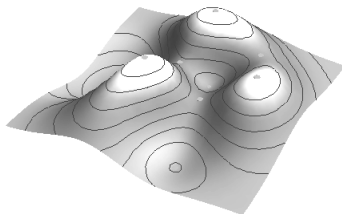


Treatments for multimodality to explore $\mathbb{V} \subset \Omega_{\mathbb{M}}$

- **Greedly optimized local improvements** (in presentation)
- *Simulated annealing based local improvements* (in paper)
- *MCMC based local improvements* (in paper)
- Other local metaheuristics (TA, ant colony optimization, local genetic algorithms, etc) (not addressed)
- *Combinations of them* (in paper)



- Minima
- Saddles
- Maxima



MCMC with locally optimized proposals

Tjelmeland and Hegstad [6] suggested continuous mode jumping proposals, **Storvik [5]** considers a more general setup, **we suggest mode jumping proposals** in the **discrete parameter spaces**.

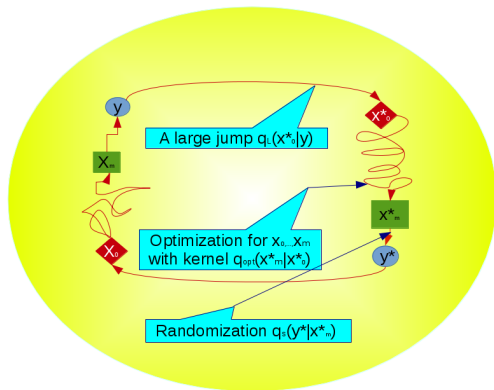


Figure: Locally optimized with randomization proposals

MCMC balance with mode jumping proposals

forward move	comment	backward move
$y \sim \pi(y)$	initial state	$y^* \sim \pi(y^*)$
$x_0^* \sim q_L(x_0^* y)$	large jump	$x_0 \sim q_L(x_0 y^*)$
$x_m^* \sim q_{\text{opt}}(x_m^* x_0^*)$	optimization	$x_m \sim q_{\text{opt}}(x_m x_0)$
$y^* \sim q_s(y^* x_m^*)$	randomization	$y \sim q_s(y x_m)$
$(x^*, y^*) \sim w(x^*, y^* y)$	thus	$(x, y) \sim w(x, y y^*)$
$x y, x^*, y^* \sim h(x y, x^*, y^*)$	choose	$x^* y, x, y^* \sim h(x^* y, x, y^*)$

$$\begin{aligned}
 & \blacktriangleright \pi(y, x)A(y, x; y^*, x^*) = \pi(y)w(y^*, x^*|y)h(x|y^*, x^*, y)r_m(y, x; y^*, x^*) \\
 & = \pi(y)w(y^*, x^*|y)h(x|y^*, x^*, y) \min \left\{ 1, \frac{\pi(y^*)w(y, x|y^*)h(x^*|y, x, y^*)}{\pi(y)w(y^*, x^*|y)h(x|y^*, x^*, y)} \right\} \\
 & = \pi(y^*)w(y, x|y^*)h(x^*|y, x, y^*) \min \left\{ \frac{\pi(y)w(y^*, x^*|y, x)h(x|y^*, x^*, y)}{\pi(y^*)w(y, x|y^*)h(x^*|y, x, y^*)}, 1 \right\} \\
 & \quad = \pi(y^*, x^*)A(y^*, x^*; y, x) \blacktriangleleft \quad (10)
 \end{aligned}$$

Application of MCMC with mode jumping proposals

Let $y = \mathbb{M}_j$, $y^* = \mathbb{M}_k$, $x^* = \{\mathbb{M}_{k_0}\}, \dots, \{\mathbb{M}_{k_{K-1}}\}$, and $x = \{\mathbb{M}_{j_0}\}, \dots, \{\mathbb{M}_{j_{K-1}}\}$ and $h(\cdot)$ be in the form (12) then (10) becomes:

$$r_m(\mathbb{M}_j, \mathbb{M}_k) = \min \left\{ 1, \frac{\Pr(D|\mathbb{M}_k) \Pr(\mathbb{M}_k) q_s(\mathbb{M}_j|\mathbb{M}_{j_{K-1}}) q_s(\mathbb{M}_{k_{K-1}}|\mathbb{M}_k)}{\Pr(D|\mathbb{M}_j) \Pr(\mathbb{M}_j) q_s(\mathbb{M}_k|\mathbb{M}_{k_{K-1}}) q_s(\mathbb{M}_{j_{K-1}}|\mathbb{M}_j)} \right\}. \quad (11)$$

with

$$h(\mathbb{M}_{j_0}, \dots, \mathbb{M}_{j_{K-1}}|\mathbb{M}_k, \mathbb{M}_j, \mathbb{M}_{k_0}, \dots, \mathbb{M}_{k_{K-1}}) = q_L(\mathbb{M}_{j_0}|\mathbb{M}_k) \times \\ \times \prod_{i \in \{1, \dots, K-2\}} Q(\mathbb{M}_{j_i}|\mathbb{M}_{j_{i-1}}) q_s(\mathbb{M}_{j_{K-1}}|\mathbb{M}_j) \quad (12)$$

where $Q(\cdot|\cdot)$ is the transition kernel of the local optimization algorithm and $q_s(\cdot|\cdot)$ is the kernel of randomization at the end.

Allowed transitions

Proposal $q(\mathbb{M}^* \mathbb{M})$	Label
$\frac{\prod_{i \in \{j(1), \dots, j(S)\}} \rho_i}{\binom{p}{S}(\eta - \zeta + 1)}$	<i>Random change with random size of the neighborhood</i>
$\frac{\prod_{i \in \{j(1), \dots, j(S)\}} \rho_i}{\binom{p}{S}}$	<i>Random change with fixed size of the neighborhood</i>
$\frac{1}{\binom{p}{S}(\eta - \zeta + 1)}$	<i>Swap with random size of the neighborhood</i>
$\binom{p}{S}^{-1}$	<i>Swap with fixed size of the neighborhood</i>
$\frac{1 - \mathbb{I}(\sum_i^p \mathbb{M}_i = p)}{p - \sum_i^p \mathbb{M}_i + \mathbb{I}(\sum_i^p \mathbb{M}_i = p)}$	<i>Uniform addition of a covariate</i>
$\frac{1 - \mathbb{I}(\sum_i^p \mathbb{M}_i = 0)}{\sum_i^p \mathbb{M}_i + \mathbb{I}(\sum_i^p \mathbb{M}_i = p)}$	<i>Uniform deletion of a covariate</i>

Table: Types of proposals suggested for the moves between the models during MCMC procedure. Here S is either deterministic or random $S \sim \text{Unif}\{\zeta, \dots, \eta\}$ size of the neighborhood; $\rho_i, i \in \{j(1), \dots, j(S)\}$ is the probability of inclusion of variable $\mathbb{M}_i, i \in \{j(1), \dots, j(S)\}$, which can be either deterministic or adaptive when $\rho_i, i \in \{j(1), \dots, j(S)\}$ are adaptively updated approximations of the marginal inclusion probabilities; $\mathbb{I}(\cdot)$ is the identity function; p is the total number of covariates.

MCMC with mode jumping proposals

Notice that

Locally annealed, locally optimized, locally simulated and locally multiple try simulated proposals and **their combination** are all **of this type of extension of the original space** and therefore their detailed balanced equation is proven in (10).

Also notice

Also note that within this setting of locally optimized MCMC we get an **alternative MCMC based approximations** for posterior probabilities of the **models**, namely $\tilde{\text{Pr}}(\mathbb{M} = M | \mathbb{D}) = \frac{\sum_{i=1}^W \mathbb{I}(M_i = M)}{W} \xrightarrow[W \rightarrow \infty]{d} \text{Pr}(\mathbb{M} = M | \mathbb{D})$ and $\arg\max_{i \in 1, \dots, W} \text{WAIC}(M_i) \xrightarrow[W \rightarrow \infty]{} \arg\max_{M \in \Omega_{\mathbb{M}}} \text{WAIC}(\mathbb{M} = M)$. Whist simultaneously $\mathbb{V} = \bigcup_{i=1}^W M_i \xrightarrow[W \rightarrow \infty]{} \Omega_{\mathbb{M}}$. This allows us to verify the results and show that the strategies are **efficient for MCMC in discrete non-concave spaces**.

The illustration. Data

chrom	pos	methyated	total	CHG	CG	CHH	DT1	DT2	DT3	DT4	DT5	DT6_20
1	2073472	4	11	1	0	0	0	1	0	0	0	0
1	2073476	3	18	0	0	1	0	0	0	1	0	0
1
1	2076202	7	12	0	0	1	1	0	0	0	0	0

Figure: The illustrations. Small example (9 variables)

The illustration. MDS

Modes are important: the standard MCMC procedure (right) misses two in this example. Visualization is challenging

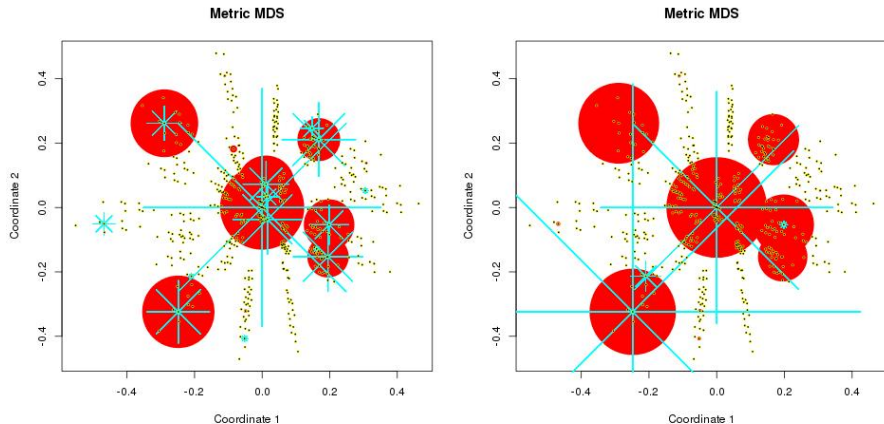


Figure: MDS plots with posterior modes of all found solutions for the approaches

The illustration. 9 variables

Mode jumping proposals - better MCMC approximations. Modes have always overestimated probabilities (right figure) when some are missed

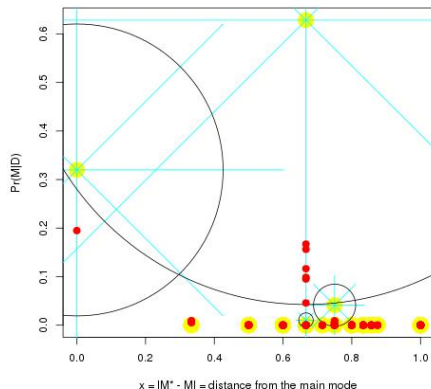
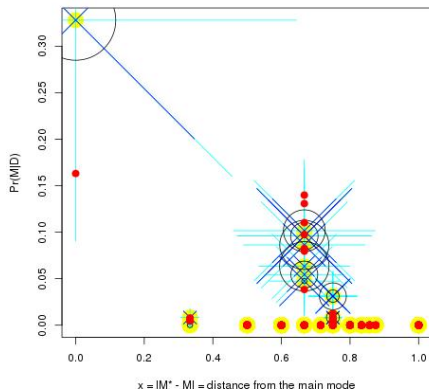


Figure: Posterior probability versus distance from the global mode

The illustration. Combination of optimizers

We now apply a mixture of local optimizers with greedily optimized frequencies or kernel of their appearance learned during the burn-in

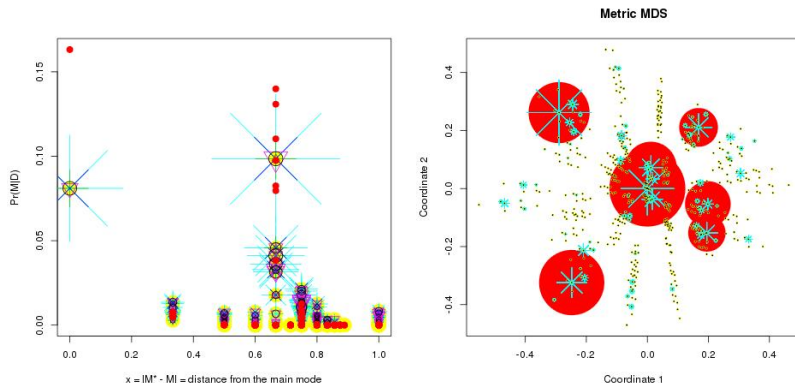
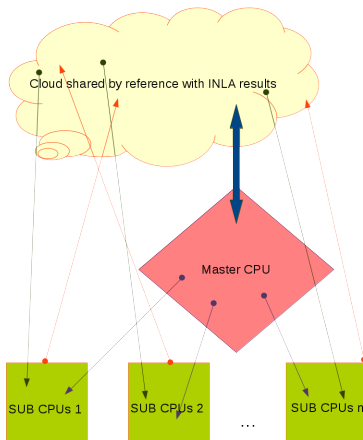


Figure: Combination of locally optimized proposals

Multicore and shared memory issues



1. Share the work done by reference
2. Before assigning a job to a CPU check if the job is already done
3. Thus avoid re-completing jobs & minimize communication times
4. Important to control writing to the shared memory efficiently

Figure: Multiprocessing architecture

Bi-modal data. Bayesian linear regression with g-prior

- $p = 15$ and $T = 100$;
- $X_i \sim N(0, 1), i \in \{1, \dots, p\}$;
- $\text{corr}(X_2, X_9) = 0.99$, others are uncorrelated \Rightarrow two modes;
- $\beta = (0.48, 8.72, 1.76, 1.87, 0, 0, 0, 0, 4.00, 0, 0, 0, 0, 0, 0)$, $\phi = 1$;
- Analyzed by a linear Bayesian regression with a g-prior, $g = T$;
- Marginal likelihoods are available analytically as:
$$\Pr(\mathbf{y}|\gamma) \propto (1 + g)^{(T-P-1)/2} (1 + g[1 - R_\gamma^2])^{-(T-1)/2}.$$
- Here R_γ^2 is the usual coefficient of determination of a linear regression model

Bi-modal data. Bayesian linear regression with g-prior

Par	True	TOP	EMJMCMC				BAS		MC ³		RS	
Δ	π_j	-	RM	MC	RM	MC	eplnp	unif	MC	RM	MC	RM
γ_{12}	0.09	0.29	2.11	5.31	1.19	5.73	1.23	1.35	2.77	4.27	2.14	3.83
γ_{14}	0.10	0.28	2.13	6.99	1.13	6.25	1.14	1.26	2.92	4.31	2.59	3.95
γ_{10}	0.11	0.28	2.31	7.41	1.31	7.74	1.15	1.31	3.06	4.31	2.40	4.07
γ_8	0.12	0.27	1.97	6.44	1.09	7.80	0.97	1.12	2.77	4.01	2.23	3.87
γ_6	0.13	0.25	2.25	8.87	1.27	8.46	1.05	1.28	3.12	4.74	2.72	4.31
γ_7	0.14	0.25	2.06	7.75	1.29	8.51	1.05	1.19	3.45	4.52	2.50	4.17
γ_{13}	0.15	0.24	2.42	9.98	1.36	8.79	1.15	1.24	3.50	4.87	2.44	4.38
γ_{11}	0.16	0.24	2.36	9.38	1.22	8.31	1.13	1.29	3.64	4.71	3.01	4.52
γ_{15}	0.17	0.23	1.96	9.38	1.08	9.73	0.78	0.93	3.92	4.27	3.32	3.84
γ_5	0.48	0.00	1.22	15.66	0.50	12.90	0.27	0.40	3.69	1.41	4.35	1.59
γ_9	0.51	0.10	1.15	16.35	0.38	12.92	0.37	0.39	16.70	5.62	6.93	2.08
γ_2	0.54	0.07	1.46	20.69	0.58	15.38	0.39	0.40	16.56	5.25	6.91	1.46
γ_1	0.74	0.18	2.15	6.43	1.06	5.97	1.20	0.92	4.10	3.55	4.51	3.90
γ_3	0.91	0.25	1.61	3.03	0.92	3.33	1.57	1.31	2.96	3.66	3.42	4.10
γ_4	1.00	0.01	0.00	6.08	0.00	2.66	0.00	0.00	0.01	0.01	0.17	0.01
$\mathbf{I}(\mathbb{M})$	0.00	1.52	7.94	39.79	3.55	36.15	3.16	3.40	33.61	25.35	29.43	22.12
Cap	1.00	0.99	0.89	0.89	0.95	0.95	0.95	0.95	0.72	0.72	0.74	0.74
Eff	2^{15}	3276	1906	1906	3212	3212	3276	3276	400	400	416	416
Tot	2^{15}	3276	3276	3276	6046	6046	3276	3276	3276	3276	3276	3276

Table: Square root of the mean squared error (RMSE) from the 100 simulated runs of every algorithm on the simulated data ($p = 15$); the values reported in the table are $\text{RMSE} \times 10^2$ for $\Delta = \gamma_j$ and $\text{RMSE} \times 10^5$ for $\mathbf{I}(\mathbb{M})$.

- $p = 15$ and $T = 47$;
- Numerous correlations \Rightarrow multiple modes;
- Analyzed by a linear Bayesian regression with a g-prior, $g = T$;
- Marginal likelihoods are available analytically as:
$$\Pr(\mathbf{y}|\boldsymbol{\gamma}) \propto (1 + g)^{(T-P-1)/2} (1 + g[1 - R_{\boldsymbol{\gamma}}^2])^{-(T-1)/2}.$$
- Here $R_{\boldsymbol{\gamma}}^2$ is the usual coefficient of determination of a linear regression model

U.S. crime data. Bayesian linear regression with g-prior

Par	True	TOP	EMJMCMC				BAS		MC ³		RS	
Δ	π_j	-	RM	MC	RM	MC	epInp	unif	MC	RM	MC	RM
γ_8	0.16	3.51	6.57	10.68	5.11	10.29	5.21	6.36	6.49	3.49	5.87	3.31
γ_{13}	0.16	3.34	7.46	10.54	5.60	10.19	6.26	7.20	8.62	3.39	8.83	3.05
γ_{14}	0.19	3.24	8.30	12.43	6.30	12.33	6.20	7.39	6.58	2.55	6.22	2.46
γ_{12}	0.22	3.27	6.87	13.61	5.57	13.64	3.10	5.21	5.81	6.23	4.93	5.27
γ_5	0.23	2.56	6.30	13.45	4.59	13.65	1.84	4.02	6.07	13.05	5.13	12.77
γ_9	0.23	3.27	9.49	16.21	7.40	16.21	9.27	8.37	5.99	2.99	5.70	2.60
γ_7	0.29	2.31	4.37	13.63	3.45	12.73	2.28	2.96	4.74	9.61	3.46	9.70
γ_4	0.30	1.57	6.18	19.22	3.79	17.31	0.99	1.20	13.24	21.84	13.53	21.48
γ_6	0.33	1.92	8.61	19.71	6.14	19.49	3.11	5.30	10.19	7.47	10.99	7.12
γ_1	0.34	2.51	11.32	22.68	7.29	20.50	8.43	7.20	22.89	25.19	23.63	24.71
γ_3	0.39	0.43	3.95	11.13	2.38	6.99	5.02	3.78	21.48	30.24	21.39	29.94
γ_2	0.57	1.58	5.92	13.21	3.82	9.03	13.78	8.66	30.81	37.57	29.27	37.15
γ_{11}	0.59	0.58	3.57	13.49	2.37	15.94	4.04	2.18	11.88	21.79	11.16	21.31
γ_{10}	0.77	3.25	7.62	7.28	5.97	4.78	15.45	10.46	21.83	19.18	20.53	19.65
γ_{15}	0.82	3.48	9.23	4.45	6.89	5.85	14.50	11.75	69.68	76.81	69.19	76.30
$\mathbf{I}(\mathbb{M})$	1.00	11.44	16.83	24.92	10.00	22.22	12.47	11.65	34.39	45.68	34.03	44.18
Cap	1.00	0.86	0.58	0.58	0.71	0.71	0.66	0.67	0.1	0.1	0.1	0.1
Eff	2^{15}	3276	1909	1909	3237	3237	3276	3276	829	829	1071	1071
Tot	2^{15}	3276	3276	3276	5936	5936	3276	3276	3276	3276	3276	3276

Table: Square root of the mean squared error (RMSE) from the 100 simulated runs of every algorithm on the U.S. crime data ($p = 15$); the values reported in the table are $\text{RMSE} \times 10^2$ for $\Delta = \gamma_j$ and $\text{RMSE} \times 10^5$ for $\mathbf{I}(\mathbb{M})$.

Multimodal data. Bayesian logistic model with AIC-prior

- $p = 20$ and $T = 2000$;
- X_1 and X_3 - factors from a group with 3 levels;
- X_4 and X_6 - correlated to them factors from group with 3 levels;
- X_7 and X_8 - two jointly dependent $\sim \text{Expo}(\lambda = 0.3)$;
- X_9, X_{10}, X_{11} - jointly dependent $\text{Unif}[-1, 10]$;
- $X_{12}, X_{13}, X_{14}, X_{15}$ - MVN with zero mean and some covariance;
- X_{16} - sinus tranformation of the radiant representation of the IDs;
- X_{17} - the quadratic trend: squared IDs of observations;
- $X_{19} = (-4 + 5X_1 + 6X_3)X_{15}$ and $X_{20} = (-4 + 5X_1 + 6X_3)X_{11}$;
- $X_2 = (X_{10} + X_{14})X_9$ and $X_5 = (X_{11} + X_{15})X_{12}$;
- $\beta = (99, -4, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1.2, 0, 37.1, 0, 0, 50, -5 \times 10^{-5}, 10, 3, 0)$;
- $\eta \sim N(\beta'X, 0.5)$
- Analyzed by Bayesian logistic regression with AIC-prior;
- Marginal likelihoods are proportional to AIC;

Multimodal data. Bayesian logistic model with AIC-prior

Par	True	TOP	EMJMCMC				BAS	MCBAS	RS	
Δ	π_j	-	RM	MC	RM	MC	RM	RM	RM	MC
γ_6	0.29	0.00	7.38	15.54	4.54	16.62	6.47	3.67	6.01	2.11
γ_8	0.31	0.00	6.23	15.50	3.96	16.94	5.58	3.02	5.37	2.55
γ_{12}	0.35	0.00	4.86	14.62	2.78	13.66	4.22	2.12	3.91	2.37
γ_{15}	0.35	0.00	4.55	15.24	2.56	15.45	4.66	1.64	3.40	2.56
γ_2	0.36	0.00	4.90	16.52	2.92	17.39	5.42	2.45	3.65	2.61
γ_{20}	0.37	0.00	4.82	14.35	2.66	14.08	3.32	1.80	4.15	2.18
γ_3	0.40	0.00	9.25	20.93	5.65	22.18	9.75	4.82	6.76	2.83
γ_{14}	0.44	0.00	3.14	17.54	1.58	16.24	3.73	1.30	1.33	2.93
γ_{10}	0.44	0.00	4.60	18.73	2.29	17.90	4.87	1.30	1.51	2.42
γ_5	0.46	0.00	3.10	17.17	1.53	16.97	4.06	1.51	1.09	2.85
γ_9	0.61	0.00	3.68	16.29	1.63	13.66	3.89	1.39	2.19	2.35
γ_4	0.88	0.00	5.66	6.70	3.74	6.26	6.60	5.57	7.61	2.15
γ_{11}	0.91	0.00	5.46	6.81	3.95	6.90	4.66	3.14	4.32	1.57
γ_1	0.97	0.00	1.90	1.74	1.35	1.34	2.43	1.96	2.30	1.1
γ_{13}	1.00	0.00	0.00	0.43	0.00	0.32	0.00	0.00	0.00	0.37
γ_7	1.00	0.00	0.00	0.57	0.00	0.41	0.00	0.00	0.00	0.33
γ_{16}	1.00	0.00	0.00	0.41	0.00	0.33	0.00	0.00	0.00	0.23
γ_{17}	1.00	0.00	0.00	0.43	0.00	0.39	0.00	0.00	0.00	0.23
γ_{18}	1.00	0.00	0.00	0.47	0.00	0.35	0.00	0.00	0.00	0.24
γ_{19}	1.00	0.00	0.00	0.52	0.00	0.36	0.00	0.00	0.00	0.41
$I(M)$	1.00	0.00	1.36	2.95	0.69	2.72	1.21	0.63	1.54	2.42
Cap	1.00	1.00	0.72	0.72	0.85	0.85	0.74	0.85	0.68	0.68
Eff	1048576	10000	5148	5148	9988	9988	10000	10000	1889	1889
Tot	1048576	10000	9998	9998	19849	19849	10000	10000	10000	10000

Table: Square root of the mean squared error (RMSE) from the 100 simulated runs of every algorithm on the simulated data ($p = 20$); the values reported in the table are $RMSE \times 10^2$ for $\Delta = \gamma_i$ and $RMSE \times 10^5$ for $I(M)$

WAIC is yet another story...

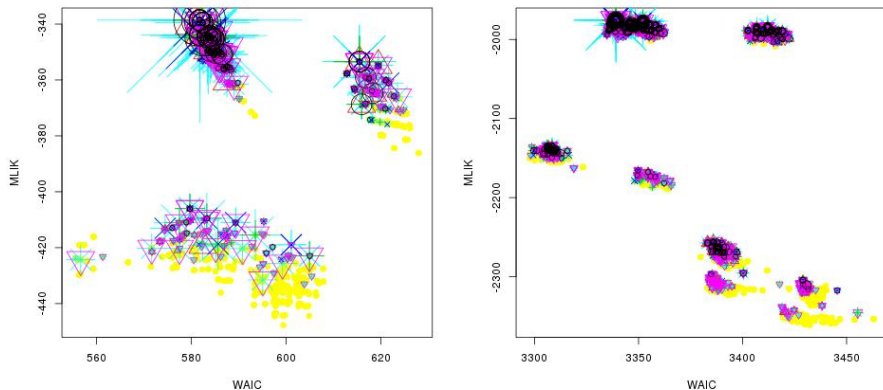


Figure: MLIK against WAIC

Concluding remarks

- We suggest using a model based approach for inference on methylation pattern along the genome
- We benefit of capturing local spatial correlation
- We suggest using different variables to improve precision of inference
- We carry out efficient choice of the subsets of these variables with respect to posterior marginal model probability and other criteria by means of mode jumping MCMC strategies
- Approach might be computationally expensive, since the nature of such a search is NP-hard, thence we efficiently address both mode jumping and parallel computation providing reasonably fast communication of the central processing units involved
- Model selection procedure developed is not problem specific and can be easily adopted to any problem where marginal likelihoods of the models are available. In particular it gives a general model selection tool within a popular INLA approach

References



M. Clyde, J. Ghosh, and M. Littman.

Bayesian adaptive sampling for variable selection and model averaging.
Journal of Computational and Graphical Statistics, 20(1):80–101, 2011.



C. P. Robert and G. Casella.

Monte Carlo Statistical Methods.
Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.



H. Rue, S. Martino, and N. Chopin.

Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations.
Journal of the Royal Statistical Society, 71(2):319–392, 2009.



G.O. Storvik.

On the flexibility of metropolis-hastings acceptance probabilities in auxiliary variable proposal generation.
Scandinavian Journal of Statistics, 38:342–358, 2011.

The End.



Thanks!