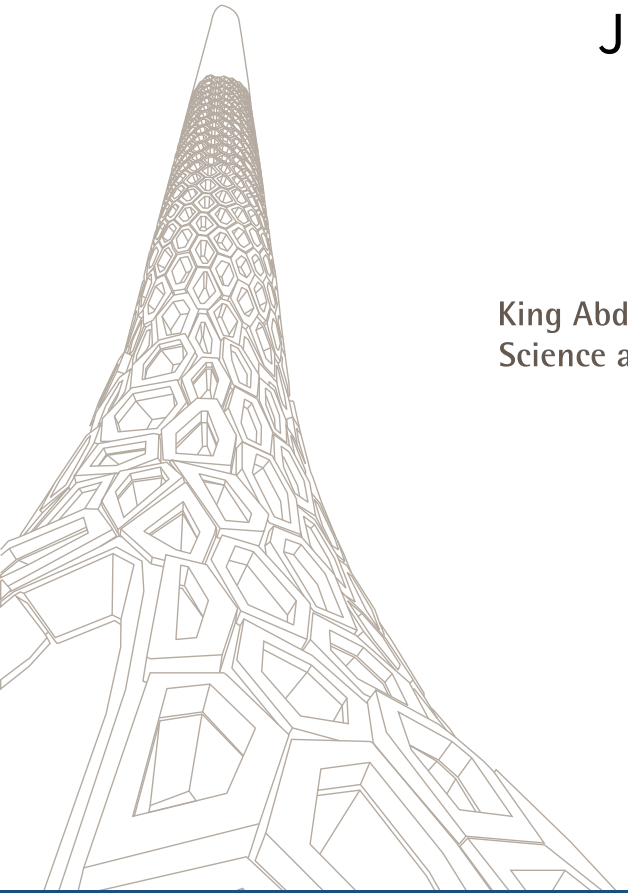
Bayesian inference with Laplace and VB

Janet van Niekerk
Janet.vanNiekerk@kaust.edu.sa



جامعة الملك عبدالله
للعلوم والتقنية

King Abdullah University of
Science and Technology

September 2024

QR code for the slides

https://github.com/JanetVN1201/Geomed_24

Outline

BayesComp group at KAUST

Bayesian inference

Data $\boldsymbol{y}$ (with covariates $\boldsymbol{Z}$), depend on $\boldsymbol{X}$ and $\boldsymbol{\theta}$ such that, $E[Y] = h(\boldsymbol{A}(\boldsymbol{Z})\boldsymbol{X})$.

Bayes' theorem:

$$\begin{aligned} q(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})p(\boldsymbol{X}, \boldsymbol{\theta}) \\ \text{Posterior} &\propto \text{Likelihood} \times \text{Prior} \end{aligned}$$

Computational aspects

- Analytical methods - conjugacy (pre-computer era)
- Approximate methods - Laplace (can be inaccurate)
- Exact methods - MCMC (very slow for complex models or large data)

Now, due to complexity and data size approximate methods are gaining popularity - INLA, VB, EP etc.

INLA - 2009 [Rue et al., 2009]
2021+ [Van Niekerk et al., 2023]

What is INLA?

INLA - Integrated Nested Laplace Approximations

- Determinisitic approximations instead of sampling
- LGM - Latent Gaussian models
- Three internal strategies - Gaussian, simplified Laplace, Laplace (pre 2021)
- R package "INLA"

This work proposes a fourth strategy that is now the default in INLA.

Gaussian approximation from the Laplace method I

Suppose we have a twice differentiable function $\log f(\boldsymbol{X})$, then the Gaussian approximation of $\log f(\boldsymbol{X})$ from the Laplace method is then derived from

$$\log f(\boldsymbol{X}) = \log f(\boldsymbol{X}_0) - \frac{1}{2}(\boldsymbol{X} - \boldsymbol{X}_0)^\top \boldsymbol{H}|_{\boldsymbol{X}=\boldsymbol{X}_0}(\boldsymbol{X} - \boldsymbol{X}_0) + \text{higher order terms},$$

where $\boldsymbol{X}_0$ is the mode of $\log f(\boldsymbol{X})$ and $\boldsymbol{H}$ is the negative Hessian matrix of $\log f(\boldsymbol{X})$. Then

$$\tilde{f}(\boldsymbol{X}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{X}_0)^\top \boldsymbol{H}|_{\boldsymbol{X}=\boldsymbol{X}_0}(\boldsymbol{X} - \boldsymbol{X}_0)\right), \tag{1}$$

so that $\boldsymbol{X} \dot{\sim} N(\boldsymbol{X}_0, \boldsymbol{H}^{-1}|_{\boldsymbol{X}=\boldsymbol{X}_0})$.

Gaussian approximation from the Laplace method II

To find the mode we solve for $\boldsymbol{X}_0$ in the system

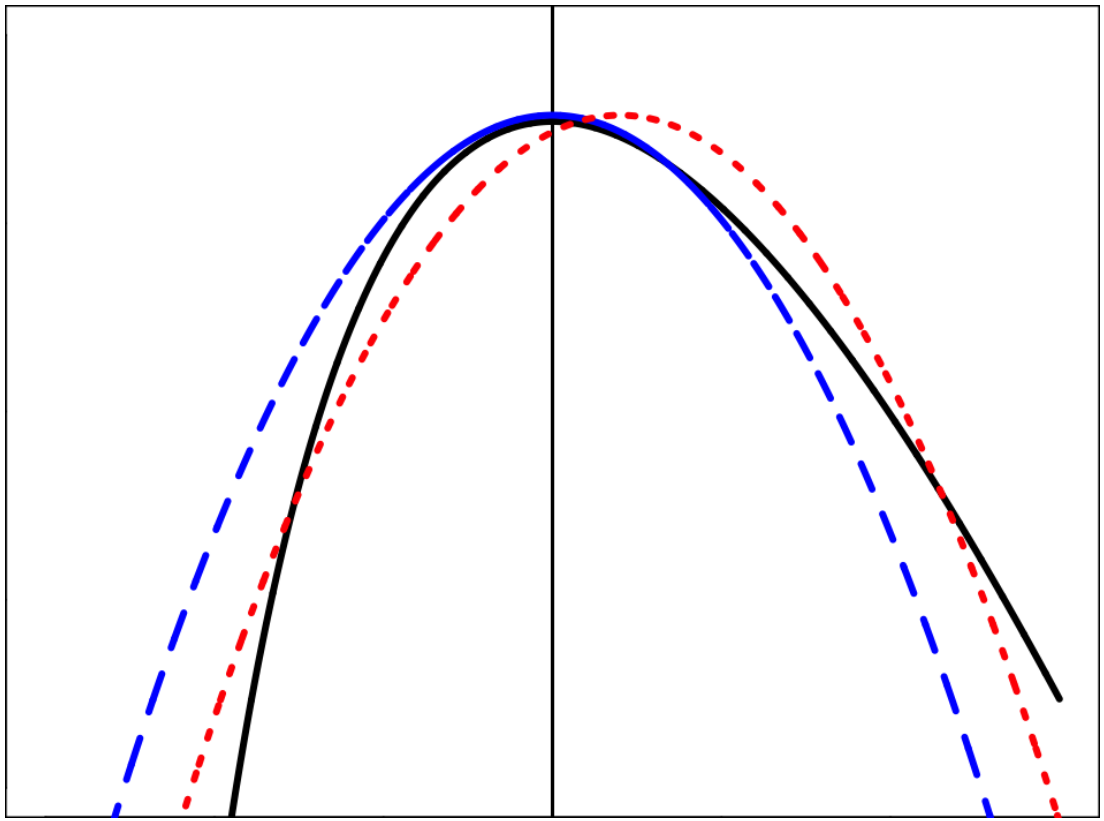
$$\boldsymbol{H}|_{\boldsymbol{x}=\boldsymbol{x}_0}\boldsymbol{X}_0 = \boldsymbol{\gamma}|_{\boldsymbol{x}=\boldsymbol{x}_0} + \boldsymbol{H}|_{\boldsymbol{x}=\boldsymbol{x}_0}\boldsymbol{X}_0, \tag{2}$$

where $\boldsymbol{\gamma}|_{\boldsymbol{x}=\boldsymbol{x}_0}$ is the gradient of $\log f(\boldsymbol{X})$ evaluated at $\boldsymbol{X} = \boldsymbol{X}_0$. Now let $\boldsymbol{Q}_0 = \boldsymbol{H}|_{\boldsymbol{x}=\boldsymbol{x}_0}$ and $\boldsymbol{b}_0 = \boldsymbol{\gamma}|_{\boldsymbol{x}=\boldsymbol{x}_0} + \boldsymbol{H}|_{\boldsymbol{x}=\boldsymbol{x}_0}\boldsymbol{X}_0$, then the system can be written as

$$\boldsymbol{Q}_0\boldsymbol{X}_0 = \boldsymbol{b}_0. \tag{3}$$

Gaussian approximation from the Laplace method III

Maybe the approximation at the mode does not give the best approximation?



Variational Inference

Optimization problem - minimize the KLD between the prior and a family of posteriors.
For a Gaussian approximation to $f(\boldsymbol{X})$, we can solve for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in

$$\arg_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \min \mathrm{KLD}(N(\boldsymbol{\mu},\boldsymbol{\Sigma})||f(\boldsymbol{X}))$$

The ELBO is then derived from this KLD as the optimization target that should be maximized. Various ways to "do" the optimization has been developed and still ongoing.
Mean-field VI is a simplification...

Model definition - GAMM

Suppose we have response data $\boldsymbol{y}_{n\times 1}$ (conditionally independent) with density function $\pi(y|\boldsymbol{X},\boldsymbol{\theta})$ and link function $h(.)$, that is linked to some covariates $\boldsymbol{Z}$ through linear predictors

$$\boldsymbol{\eta}_n = \beta_0 + \boldsymbol{Z}_\beta\boldsymbol{\beta} + \sum f^k(\boldsymbol{Z}_f) = \boldsymbol{A}\boldsymbol{X}$$

The inferential aim is to estimate the latent field $\boldsymbol{X}_m = \{\beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$, and the hyperparameters $\boldsymbol{\theta}$.

Information theoretic point of view - Zellner (1988)[1]

Based on prior information $\mathcal{I}$, data $\boldsymbol{y}$ and parameters $\boldsymbol{P} = \{\boldsymbol{X}, \boldsymbol{\theta}\}$, define the following:

1. $\pi(\boldsymbol{P}|\mathcal{I})$ is the prior model
2. $q(\boldsymbol{P}|\mathcal{D})$ is the learned model from the prior information and the data where $\mathcal{D} = \{\mathcal{I}, \boldsymbol{y}\}$
3. $l(\boldsymbol{P}|\boldsymbol{y}) = f(\boldsymbol{y}|P)$ is the likelihood
4. $p(\boldsymbol{y}|\mathcal{I})$ is the marginal model for the data where $p(\boldsymbol{y}|\mathcal{I}) = \int f(\boldsymbol{y}|\boldsymbol{P})\pi(\boldsymbol{P}|\mathcal{I})d\boldsymbol{P}$

The input information in the learning of $\boldsymbol{P}$ is given by $\pi(\boldsymbol{P}|\mathcal{I})$ and $l(\boldsymbol{P}|\boldsymbol{y})$. An information processing rule (IPR) then delivers $q(\boldsymbol{P}|\mathcal{D})$ and $p(\boldsymbol{y}|\mathcal{I})$ as output information.

[1] Zellner, A., 1988. Optimal information processing and Bayes's theorem. The American Statistician, 42(4), pp.278-280.

Bayes Rule as an efficient IPR

A stable and efficient IPR would provide the same amount of output information than received through the input information, thus being information conservative. Zellner shows that $q(\boldsymbol{P}|\mathcal{D})$ minimizes

$$
\begin{aligned}
- & \int \left[\log \pi(\boldsymbol{P}|\mathcal{I}) + \log l(\boldsymbol{P}|\boldsymbol{y})\right] q(\boldsymbol{P}|\mathcal{D}) d\boldsymbol{P} + \int \left[\log q(\boldsymbol{P}|\mathcal{D}) + \log p(\boldsymbol{y}|\mathcal{I})\right] q(\boldsymbol{P}|\mathcal{D}) d\boldsymbol{P} \\
= & \; E_{q(\boldsymbol{P}|\mathcal{D})}\left[-\log l(\boldsymbol{P}|\boldsymbol{y})\right] + \mathrm{KLD}\left[q(\boldsymbol{P}|\mathcal{D})||\pi(\boldsymbol{P}|\mathcal{I})\right]. \qquad (4)
\end{aligned}
$$

Variational form of Bayes' theorem

Finding the best fit from a certain family $Q = \{q(\boldsymbol{P})\}$, for prior $\pi(\boldsymbol{P})$,

$$\arg\min_{p \in Q} \left(\mathrm{E}_{p(\boldsymbol{P})} \left[-\sum_{i=1}^{n} \log f(y_i|\boldsymbol{P}) \right] + \mathrm{KLD}(p||\pi) \right) \tag{5}$$

This enables us to do Variational Inference without the need for an ELBO or other simplifying assumptions.

Laplace method with low-rank Variational Bayes correction I

For a GAMM, data $\boldsymbol{y}$, model parameters $\boldsymbol{X}$ and prior $\pi(\boldsymbol{X})$,

- Gaussian approximation using Laplace method to $q(\boldsymbol{X}|\boldsymbol{y})$

$$\boldsymbol{Q}_0\boldsymbol{X}_0 = \boldsymbol{b}_0$$

- Correct the Laplace method's mean with VB[2], $\boldsymbol{X}_0^*$ such that

$$\boldsymbol{Q}_0(\boldsymbol{X}_0 + \boldsymbol{\delta}) = \boldsymbol{b}_0^*$$

But what if the dimension of $\boldsymbol{X}$ is large?

Laplace method with low-rank Variational Bayes correction II

Do an implicit correction of the mean, solve for $\boldsymbol{\delta}$ such that

$$\boldsymbol{Q}_0 \boldsymbol{X}_0^* = \boldsymbol{b}_0 + \boldsymbol{\delta}$$

This means that there is a map for the change the i^{th} element of δ will cause in the j^{th} element of $\boldsymbol{X}_0$. This map is constructed from $\boldsymbol{Q}_0$.

If $dim(X) = m$ then we can have that the non-zero entries in δ is at most p, $p << m$. The optimization is then in p dimensions and not in m dimensions.

Who? What? Why? When? etc...

[2] van Niekerk, J. and Rue, H., 2024. Low-rank variational Bayes correction to the Laplace method. Journal of Machine Learning Research, 25(62), pp.1-25.

Example I

$$y_i \sim \text{Poisson}(\exp(\eta_i)), \quad \eta_i = \beta_0 + \beta_1 x_i + u_i,$$

with a sum-to-zero constraint on $\boldsymbol{u}$, to ensure identifiability of β_0. We want to perform full Bayesian inference for the latent field $\boldsymbol{\psi} = \{\beta_0, \beta_1, \boldsymbol{u}\}$, and the linear predictors $\boldsymbol{\eta} = \{\eta_1, \eta_2, ..., \eta_n\}$. We assume the following illustrative priors,

$$\beta_0 \sim t(5), \quad \beta_1 \sim U(-3, 3) \quad \text{and} \quad \boldsymbol{u} \sim N(\boldsymbol{0}, 0.25\boldsymbol{I})$$

Example II

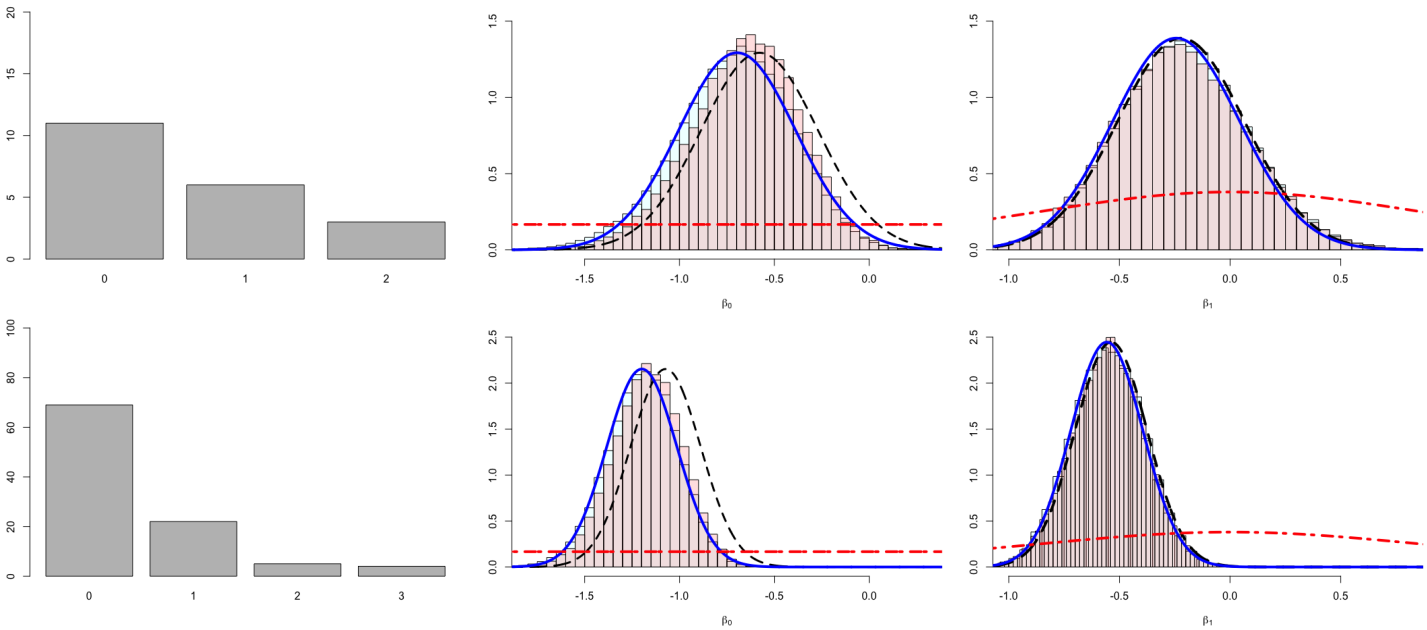


Figure: Poisson counts simulated from (18) (left) and the marginal posterior of β_0 (center) and β_1 (right) from MCMC (blue histogram), HMC (red histogram), the Laplace method (dashed line) and VBC (solid line) based on the prior (broken line) for $n = 20$ (top) and $n = 100$ (bottom)

Example III

	n=100			
	LM	VBC	MCMC	HMC
β_0	-1.073	-1.199	-1.196	-1.180
β_1	-0.538	-0.567	-0.552	-0.552
u_1	0.177	0.174	0.175	0.174
u_8	-0.046	-0.052	-0.049	-0.044
u_{15}	-0.074	-0.079	-0.077	-0.073
Time(s)	9.48	17.36	384.12	169.57

Table: Posterior means from the Laplace method, VBC, MCMC and HMC

VBC in INLA...

GAMM $\rightarrow$ LGM

Assume

$$\boldsymbol{X}|\boldsymbol{\theta} \sim N(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is a sparse matrix ($\boldsymbol{X}$ is a GMRF).

$p(\boldsymbol{X}, \boldsymbol{\theta}) = p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ can be non-Gaussian.

Ingredients

Likelihood - $\prod_{i=1}^{n} \pi(y_i|\boldsymbol{X},\boldsymbol{\theta}), \eta = \boldsymbol{AX}$
Prior for the latent - $\pi(\boldsymbol{X}|\boldsymbol{\theta})$
Prior for the hyperparameters - $\pi(\boldsymbol{\theta})$
Goal:

- $q(X_j|\boldsymbol{y})$
- $q(\theta_k|\boldsymbol{y})$

Posterior approximations by INLA

For

$$\pi(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{y}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{X}|\boldsymbol{\theta})\prod_{i=1}^{n}\pi(y_i|(\boldsymbol{A}\boldsymbol{X})_i,\boldsymbol{\theta})$$

1. $$\tilde{q}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{y})}{\pi_G(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{\boldsymbol{X}=\boldsymbol{\mu}(\boldsymbol{\theta})}$$
2. $$\tilde{q}(\theta_j|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-j}$$
3. $$\tilde{q}(\boldsymbol{X}_j|\boldsymbol{y}) = \int \tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta},\boldsymbol{y})\tilde{q}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},$$

$\tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta},\boldsymbol{y})$ depends on the approximation used in Stage 1.

⏩ Skip

INLA 1.0

- $\boldsymbol{X} = \{\boldsymbol{\eta}, \beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$
- Laplace strategy: For each j,

$$\tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta},\boldsymbol{y}) = \left.\frac{\pi(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{y})}{\pi_G(\boldsymbol{X}_{-j}|X_j,\boldsymbol{\theta},\boldsymbol{y})}\right|_{\boldsymbol{X}_{-j}=\boldsymbol{\mu}_{-j}}$$

Wishes and dreams

1. How can we get a good and cheap approximation $\tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta},\boldsymbol{y})$ using $\pi_G(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y})$?
   Non-Gaussian? Other family?
2. How can we remove $\boldsymbol{\eta}$ from $\boldsymbol{X}$ and still produce full posteriors of $\boldsymbol{\eta}$?
   Huge data? Prediction? Stability?

Gaussian approximation $\pi_G(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y})$

Laplace method - fit the best Gaussian at the mode of a curve where the variance is derived from the inverse Hessian at the mode.

$$\begin{aligned}\log\left(\pi(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y})\right) &\propto -\frac{1}{2}\boldsymbol{X}^\top \boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{X} + \sum_{i=1}^{n}\left(b_i X_i - \frac{1}{2}c_i X_i^2\right)\\ &= -\frac{1}{2}\boldsymbol{X}^\top\left(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{D}\right)\boldsymbol{X} - \boldsymbol{b}^\top\boldsymbol{X}\end{aligned}$$

hence

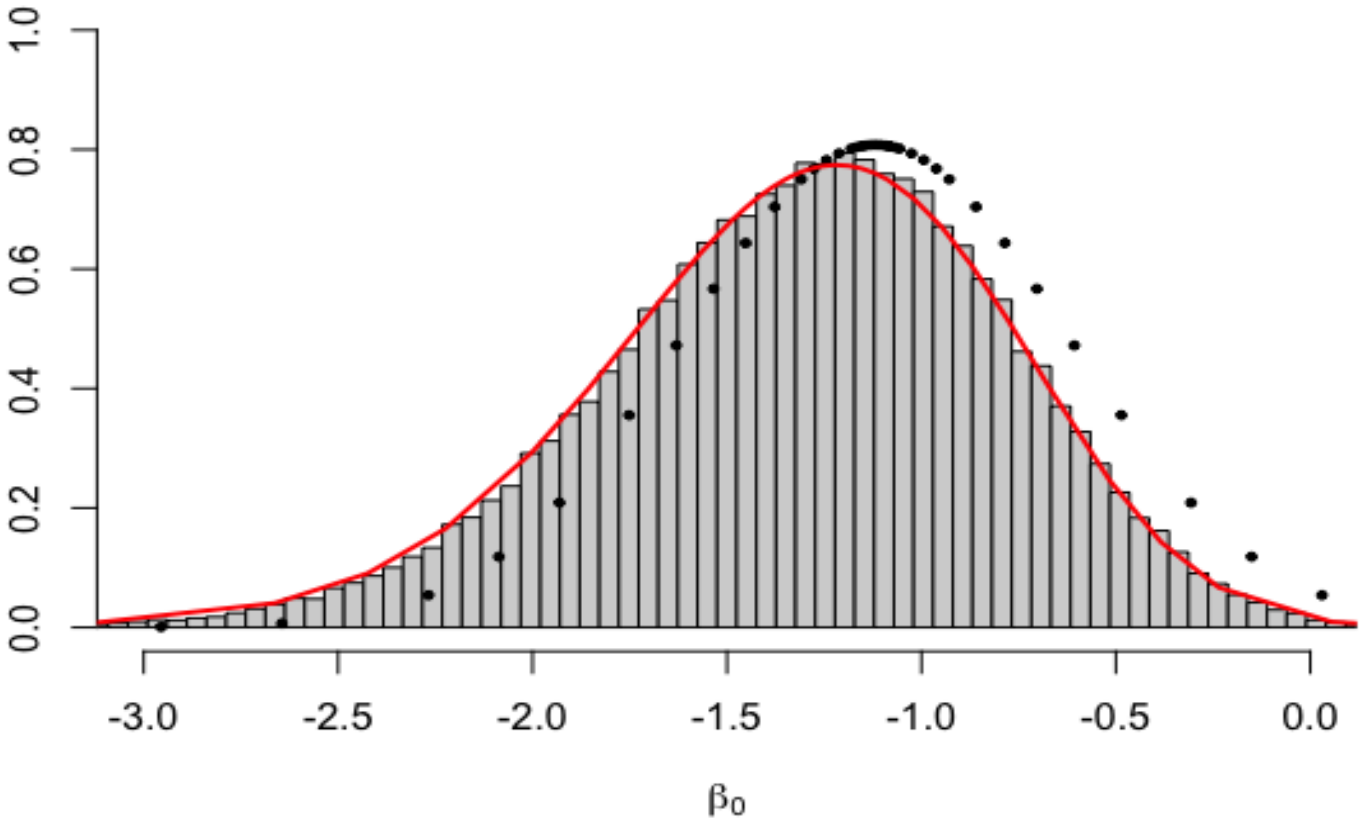
$$\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y} \sim N(\boldsymbol{\mu}, (\boldsymbol{Q}(\boldsymbol{\theta}) + diag(\boldsymbol{c}))^{-1}) \tag{6}$$

with

$$(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{D})\boldsymbol{\mu} = \boldsymbol{b}$$

Poisson example

$$Y_i|\beta_0 = -1, \beta_1 = -0.5 \quad \sim \quad \text{Poisson}(\exp(\beta_0 + \beta_1 X_i))$$



How can we use this?

We apply this to the Gaussian approximation in the denominator of Stage 1 ▸▸ Stage 1.

Recall that $(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{D})\boldsymbol{\mu} = \boldsymbol{Q}\boldsymbol{\mu} = \boldsymbol{b}$.

Now let's formulate $\boldsymbol{Q}\boldsymbol{\mu}^* = \boldsymbol{b} + \boldsymbol{\lambda}$, so that

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + \boldsymbol{M}\boldsymbol{\lambda}$$

So now we solve for,

$$\arg_{\boldsymbol{\lambda}} \min_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} \left(\mathrm{E}_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} \left[-\sum_{i=1}^{n} \log f(y_i|\boldsymbol{x}_i,\boldsymbol{\theta}) \right] + \mathrm{KLD}(p||\pi) \right)$$

where $\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta} \sim N(\boldsymbol{\mu}^*, \boldsymbol{Q}^{-1})$.

Low-rank correction $\rightarrow$ Only correct some $b's$, change to all $\mu's$.

VB corrected marginal posterior of η_i

$$
\begin{aligned}
\eta_j|\boldsymbol{\theta},\boldsymbol{y} &\sim N(\mu_j(\boldsymbol{\theta}),\sigma_j^2(\boldsymbol{\theta})) \\
\mu_j(\boldsymbol{\theta}) &= (\boldsymbol{A}\boldsymbol{\mu}^*(\boldsymbol{\theta}))_j \\
\text{Cov}(\boldsymbol{\eta}|\boldsymbol{\theta},\boldsymbol{y}) &= \boldsymbol{A}\text{Cov}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y})\boldsymbol{A}^\top \\
\tilde{\pi}(\eta_j|\boldsymbol{y}) &\approx \sum_{k=1}^{K} \pi_G(\eta_j|\boldsymbol{\theta}_k,\boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})\delta_k.
\end{aligned}
$$

Overdispersed Poisson regression I

$$y_i \sim \text{Poisson}(\exp(\eta_i)), \quad \eta_i = \beta_0 + \beta_1 x_i + u_i, \tag{7}$$

for $i = 1, 2, ..., n$, where $u_i|\tau \sim N(0, \tau^{-1})$, $\log \tau \sim \text{loggamma}(1, 5 \times 10^{-5})$, $\beta_0 \sim N(0, 1)$ and $\beta_1 \sim N(0, 1)$. The data is simulated based on $\beta_0 = -1, \beta_1 = -0.5, \tau = 1, n = 1000$ and the parameters to infer are $\boldsymbol{\psi} = \{\beta_0, \beta_1, u_1, u_2, ..., u_n\}$, the linear predictors $\{\eta_1, \eta_2, ..., \eta_n\}$ i.e. $\boldsymbol{x} = \{\boldsymbol{\psi}, \boldsymbol{\eta}\}$, and the set of hyperparameters $\boldsymbol{\theta} = \{\tau\}$.

Overdispersed Poisson regression II

	GA	INLA	INLA-VBC	HMC
β_0	-0.972	-0.664	-0.972	-0.934
β_1	-0.484	-0.532	-0.531	-0.529
τ	1.056	1.056	1.056	1.037
Time(s)	5.067	18.299	5.718	207.445

Table: Posterior means from the Gaussian strategy (GA), Laplace strategy (INLA), INLA-VBC and MCMC

Tokyo example

The Tokyo dataset in the R INLA library contains information on the number of times the daily rainfall measurements in Tokyo was more than 1*mm* on a specific day *t* for two consecutive years. In order to model the annual rainfall pattern, a stochastic spline model with fixed precision is used to smooth the data.

$$\begin{aligned} y_i|\mathcal{X} &\sim Bin\left(n_i, p_i = \frac{\exp(\alpha_i)}{1+\exp(\alpha_i)}\right) \\ (\alpha_{i+1} - 2\alpha_i + \alpha_{i-1})|\tau &\overset{\text{iid}}{\sim} N(0, \tau^{-1}), \end{aligned}$$

where $i = 1, 2, \ldots, 366$ on a torus, and $n_{60} = 1$ else $n_i = 2$.

Results

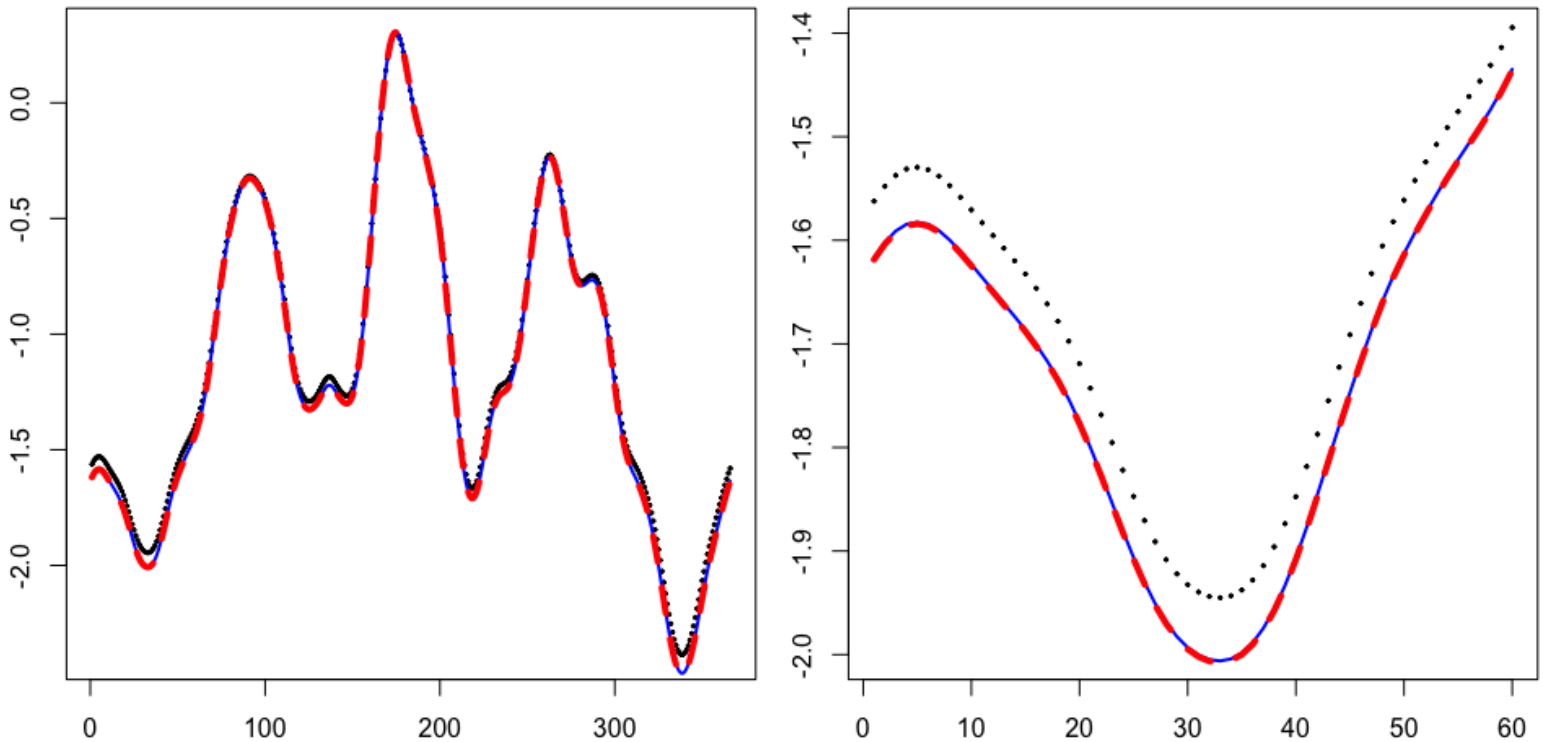


Figure: Posterior mean of α (left) (zoomed for the first two months (right)) from the Laplace method (points), VBC (solid line) and INLA (broken line)

Spatial survival example I

Consider the R dataset `Leuk` that features the survival times of 1043 patients with acute myeloid leukemia (AML) in Northwest England between 1982 to 1998.

$$h(t, \boldsymbol{s}) = h_0(t) \exp(\boldsymbol{\beta X} + \boldsymbol{u}(\boldsymbol{s})),$$

with

$$\eta_i(s) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{WBC}_i + \beta_3 \text{TPI}_i + u(s).$$

which implies a latent field of size $m = 39158$.

Spatial survival example II

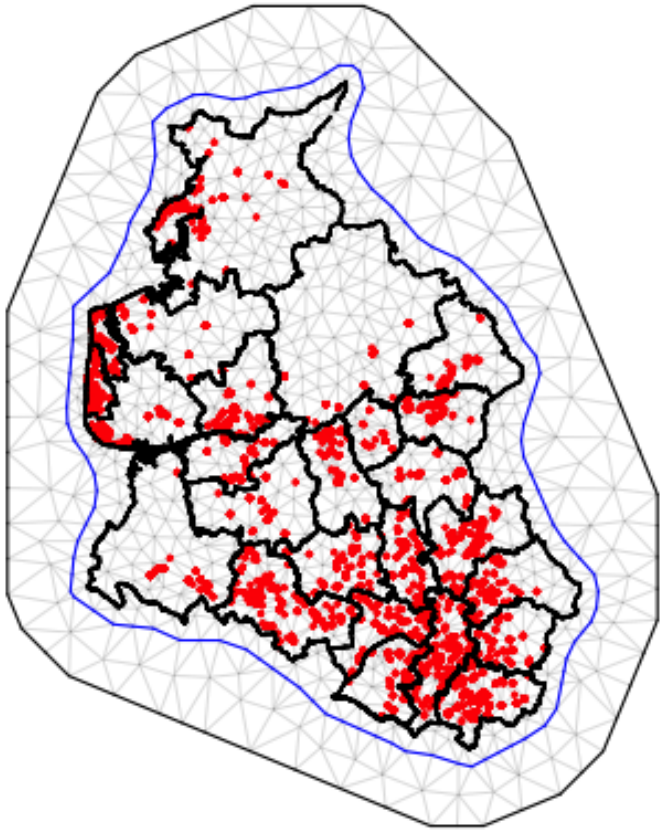


Figure: Exact residential locations of patients with AML

Results I

	GA	INLA	INLA-VBC
β_0	-2.023	-2.189	-2.189
β_1	0.596	0.597	0.597
β_2	0.242	0.241	0.241
β_3	0.108	0.108	0.108
τ	0.340	0.340	0.340
σ_u	0.223	0.223	0.223
r	0.202	0.202	0.202
Time(s)	25.9	1276	26.3

Table: Posterior means from the Gaussian strategy, INLA and INLA-VBC - all fixed effects are significant

Results II

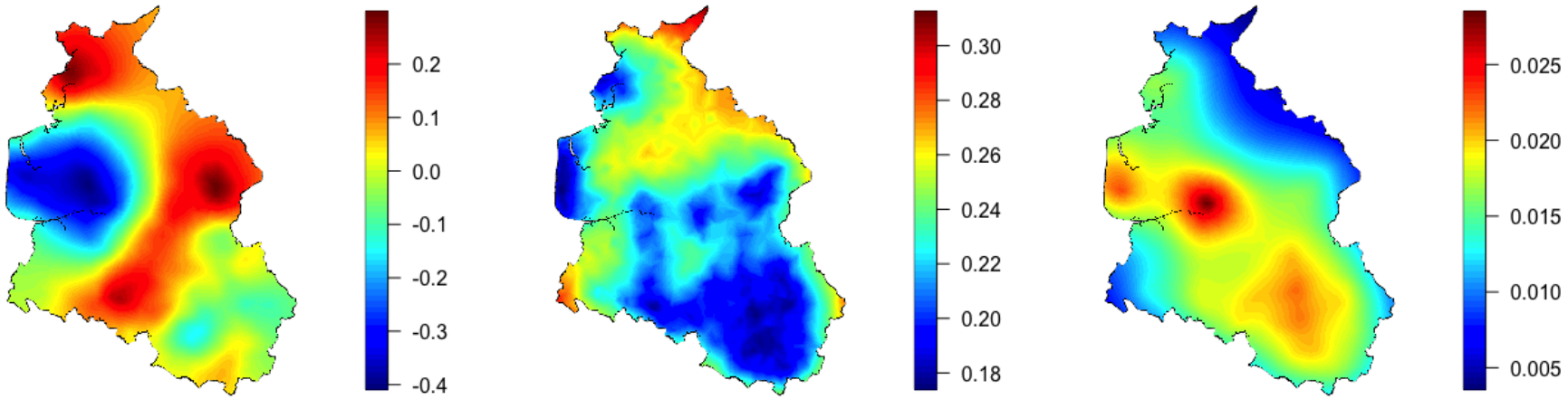


Figure: Posterior mean (left) and posterior standard deviation (center) of $\boldsymbol{u}(\boldsymbol{s})$ from INLA-VBC with the absolute difference between the posterior means of $\boldsymbol{u}(\boldsymbol{s})$ from the Gaussian strategy and INLA-VBC (right)

cs-fMRI model

Functional magnetic resonance imaging (fMRI) is a noninvasive neuro-imaging technique used to localize regions of specific brain activity during certain tasks. For T timepoints and N vertices per hemisphere resulting in data $\boldsymbol{y}_{TN\times 1}$ with the latent Gaussian model as follows:

$$
\begin{aligned}
\boldsymbol{y}|\beta,\boldsymbol{b},\boldsymbol{\theta} &\sim N(\boldsymbol{\mu}_y,\boldsymbol{V}), \quad \boldsymbol{\mu}_y = \sum_{k=0}^{K}\boldsymbol{X}_k\beta_k + \sum_{j=1}^{J}\boldsymbol{Z}_j\boldsymbol{b}_j \\
\beta_k &= \boldsymbol{\Psi}_k\boldsymbol{w}_k \quad \text{(SPDE prior on } \beta_k\text{)} \\
\boldsymbol{w}_k|\boldsymbol{\theta} &\sim N(\boldsymbol{0},\boldsymbol{Q}^{-1}_{\tau_k,\kappa_k}) \\
\boldsymbol{b}_j &\sim N(\boldsymbol{0},\delta\boldsymbol{I}) \quad \text{(Diffuse priors for } \boldsymbol{b}_j\text{)} \\
\boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}),
\end{aligned}
$$

where we have K task signals and J nuisance signals.[3]

[3] Van Niekerk, J., Krainski, E., Rustand, D. and Rue, H., 2023. A new avenue for Bayesian inference with INLA. Computational Statistics & Data Analysis, 181, p.107692.

cs-fMRI model

The data consists of a 3.5-min fMRI for each subject, consisting of 284 volumes, where each subject performs 5 different motor tasks interceded with a 3 second visual cue. Each hemisphere of the brain contained 32492 surface vertices. From these, 5000 are resampled to use for the analysis. This results in a response data vector $\boldsymbol{y}$ of size $\mathbf{2\,523\,624}$, with an SPDE model defined on a mesh with 8795 triangles.

The inference based on the modern formulation of INLA was computed in 148 seconds.

cs-fMRI model

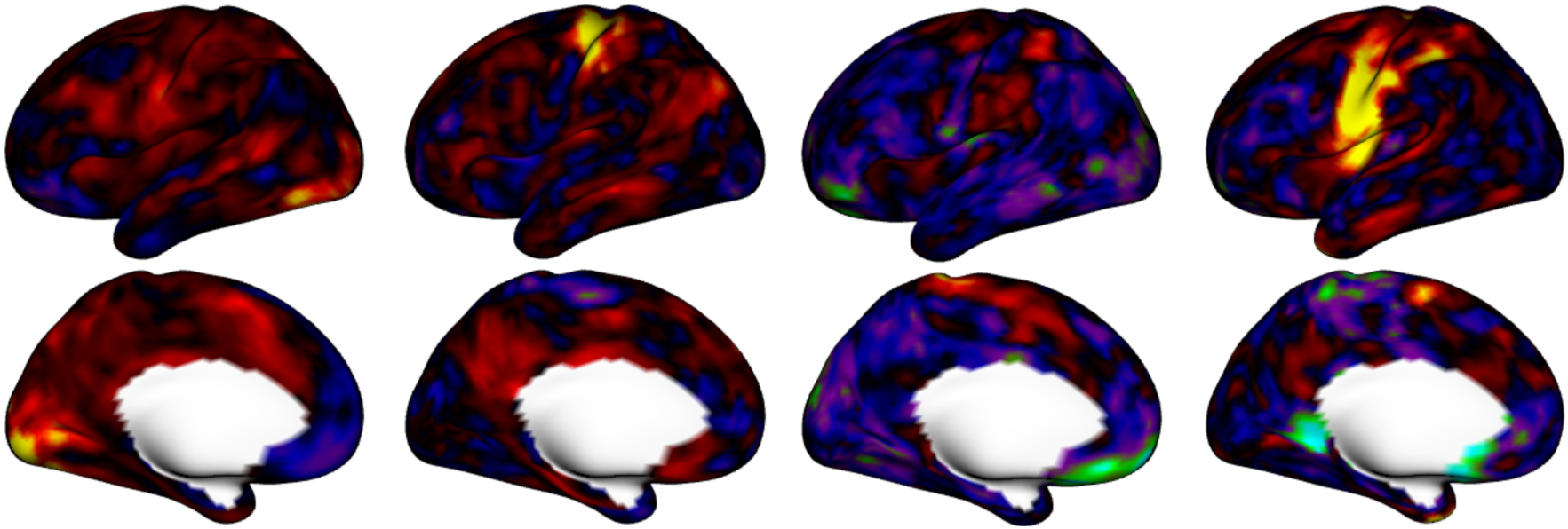


Figure: Activation areas for the different tasks in the left hemisphere - visual cue, right hand motor, right foot motor, tongue motor task (from left to right)

Further details

`www.r-inla.org`

New default setting in INLA (VB) (previously inla.mode = "experimental")

- INLA can fit many different statistical models and complex models can be built using multiple "building blocks"/random effects.
- Remove the linear predictors from the latent field $\rightarrow$ accurate posterior inference with VB correction (I - VB - LA)
- New applications that aren't feasible with INLA 1.0


Rue, H., Martino, S., and Chopin, N. (2009).
Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(2):319–392.


Van Niekerk, J., Krainski, E., Rustand, D., and Rue, H. (2023).
A new avenue for Bayesian inference with INLA.
Computational Statistics & Data Analysis, 181:107692.

Thank you · شكرا

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology