

# Bayesian modelling

## Deterministic approximations

Léo Belzile

Last compiled Wednesday Mar 19, 2025

## Rationale for deterministic approximations

Markov chain Monte Carlo methods require tuning and can be markedly slow when the dimension of the parameter space grows.

The curse of dimensionality affects the performance of MCMC.

We consider simple approximations to the marginal likelihood, posterior moments, or posterior density that only require **numerical optimization**.

# Landau notation

We need notation to characterize the growth rate of functions: when  $n \rightarrow \infty$

- big-O:  $x = O(n)$  means that  $x/n \rightarrow c \in \mathbb{R}$  as  $n \rightarrow \infty$
- little-o:  $x = o(n)$  implies  $x/n \rightarrow 0$  as  $n \rightarrow \infty$ .

# Taylor series expansion

Consider a concave function  $h(x)$  assumed twice continuously differentiable with mode at  $x_0$ . Then, a Taylor series expansion around  $x_0$  gives

$$h(x) = h(x_0) + h'(x_0)(x - x_0) + h''(x_0)(x - x_0)^2/2 + R$$

where the remainder  $R = O\{(x - x_0)^3\}$ .

# Multivariate Taylor series expansion

Similarly, for  $h(\mathbf{x})$  a smooth vector valued function with mode at  $\mathbf{x}_0 \in \mathbb{R}^p$ , we can write

$$\begin{aligned} h(\mathbf{x}) &= h(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top h'(\mathbf{x}_0) \\ &\quad + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^\top h''(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) + R. \end{aligned}$$

Under regularity conditions, the mode  $\mathbf{x}_0$  is such that  $h'(\mathbf{x}_0) = \mathbf{0}_p$ .

# Laplace approximation

The **Laplace approximation** is used to approximate integrals of non-negative functions  $g(\mathbf{x})$  that are  $O(n)$  of the form

$$I_n = \int_{\mathbb{R}^p} g(\mathbf{x}) d\mathbf{x} = \int \exp\{h(\mathbf{x})\} d\mathbf{x}.$$

The idea is that we can, ignoring terms above third order and assuming  $\mathbf{x}_0$  satisfies  $h'(\mathbf{x}_0) = \mathbf{0}_p$ , approximate  $g(\cdot)$  by a multivariate Gaussian density.

# Laplace approximation to integrals

If we perform a Taylor series expansion of the log of the integrand, then

$$I_n \approx (2\pi)^{p/2} |\mathbf{H}(\mathbf{x}_0)|^{-1/2} \exp\{h(\mathbf{x}_0)\} + \mathcal{O}(n^{-1})$$

where  $|\mathbf{H}(\mathbf{x}_0)|$  is the determinant of the Hessian matrix of  $-h(\mathbf{x})$  evaluated at the mode  $\mathbf{x}_0$ .

# Laplace approximation

- The idea behind the Laplace approximation is to approximate the log of the density (since the latter must be non-negative).
- Compared to sampling-based methods, the Laplace approximation requires **optimization**.
- The Laplace approximation is not invariant to reparametrization: in practice, it is best to perform it on a scale where the likelihood is as close to quadratic as possible in  $g(\boldsymbol{\theta})$  and back-transform using a change of variable.



# Laplace approximation to the marginal likelihood

Consider a simple random sample of size  $n$  from a distribution with parameter vector  $\boldsymbol{\theta} \in \mathbb{R}^p$ .

Write (Raftery, 1995)

$$p(\mathbf{y}) = \int_{\mathbb{R}^p} p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

and take

$$h(\boldsymbol{\theta}) = \log p(\mathbf{y} \mid \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}).$$

# Laplace approximation to the marginal likelihood

Evaluating at the maximum a posteriori  $\hat{\boldsymbol{\theta}}_{\text{MAP}}$  and letting  
–  $\mathbf{H}$  denote the Hessian matrix of second partial derivatives  
of the unnormalized log posterior, we get (Tierney & Kadane,  
1986)

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\hat{\boldsymbol{\theta}}_{\text{MAP}}) + \log p(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{\text{MAP}}) \\ &\quad + \frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{H}(\hat{\boldsymbol{\theta}}_{\text{MAP}})| + \mathcal{O}(n^{-1}) \end{aligned}$$

## Example with exponential likelihood

Consider an exponential likelihood  $Y_i \mid \lambda \sim \text{expo}(\lambda)$  with conjugate gamma prior  $\lambda \sim \text{gamma}(a, b)$ . The exponential model has information  $i(\lambda) = n/\lambda^2$  and the mode of the posterior is

$$\hat{\lambda}_{\text{MAP}} = \frac{n + a - 1}{\sum_{i=1}^n y_i + b}.$$

# Marginal likelihood approximation for exponential likelihood

We can also obtain an estimate of the marginal likelihood, which is equal for the conjugate model

$$p(\mathbf{y}) = \frac{\Gamma(n + a)}{\Gamma(a)} \frac{b^a}{\left(b + \sum_{i=1}^n y_i\right)^{n+a}}.$$

For the sample of size 62, the exponential model marginal likelihood is  $-276.5$ .

# Numerical approximation to marginal likelihood

```
1 data(waiting, package = "hecbayes")
2 a <- 0.01; b <- 0.01
3 n <- length(waiting); s <- sum(waiting)
4 map <- (n + a - 1)/(s + b) #posterior mode
5 logpost <- function(x){
6   sum(dexp(waiting, rate = x, log = TRUE)) +
7   dgamma(x, a, b, log = TRUE)
8 }
9 # Hessian evaluated at MAP
10 H <- -c(numDeriv::hessian(logpost, x = map))
11 # Laplace approximation
12 marg_lik_laplace <- 0.5*log(2*pi) - log(H) + logpost(map)
```

The Laplace approximation gives  $-281.9$ .

# Posterior expectation using Laplace method

If we are interested in computing the posterior expectation of a positive real-valued functional  $g(\boldsymbol{\theta}) : \mathbb{R}^p \rightarrow \mathbb{R}_+$ , we may write

$$\mathbf{E}_{\boldsymbol{\Theta}|\mathbf{Y}}\{g(\boldsymbol{\theta}) \mid \mathbf{y}\} = \frac{\int g(\boldsymbol{\theta})p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

# Posterior expectation via Laplace

We can apply Laplace's method to both numerator and denominator. Let  $\hat{\boldsymbol{\theta}}_g$  and  $\hat{\boldsymbol{\theta}}_{\text{MAP}}$  of the integrand of the numerator and denominator, respectively, and the negative of the Hessian matrix of the log integrands

$$J_g = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \{ \log g(\boldsymbol{\theta}) + \log p(\mathbf{y} \mid \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \},$$

$$J = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \{ \log p(\mathbf{y} \mid \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \}.$$

# Posterior expectation approximation

Putting these together

$$\mathbb{E}_{\Theta|Y}\{g(\boldsymbol{\theta}) \mid \mathbf{y}\} \approx \frac{|J(\hat{\boldsymbol{\theta}}_{\text{MAP}})|^{1/2}}{|J_g(\hat{\boldsymbol{\theta}}_g)|^{1/2}} \frac{g(\hat{\boldsymbol{\theta}}_g)p(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_g)p(\hat{\boldsymbol{\theta}}_g)}{p(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{\text{MAP}})p(\hat{\boldsymbol{\theta}}_{\text{MAP}})}.$$

While the Laplace method has an error  $O(n^{-1})$ , the leading order term of the expansion cancels out from the ratio and the above has error of  $O(n^{-2})$ .



## Example of posterior mean for exponential likelihood

Consider the posterior mean  $\mathbf{E}_{\Lambda|Y}(\lambda)$  and let  $s = \sum_{i=1}^n y_i$ .

Then,

$$\hat{\lambda}_g = \frac{(n + a)}{s + b}$$
$$|J_g(\hat{\lambda}_g)|^{1/2} = \left( \frac{n + a}{\hat{\lambda}_g^2} \right)^{1/2} = \frac{s + b}{(n + a)^{1/2}}$$

# Posterior mean

Simplification gives the approximation

$$\frac{\exp(-1)}{s + b} \frac{(n + a)^{n+a+1/2}}{(n + a - 1)^{n+a-1/2}}$$

which gives 0.03457, whereas the true posterior mean is  $(n + a)/(s + b) = 0.03457$ .

## Aside on prior and likelihood relative contribution

Usually,

- $p(\boldsymbol{\theta}) = O(1)$  and
- $p(\mathbf{y} \mid \boldsymbol{\theta}) = O(n)$

Thus, provided the prior does not impose unnecessary support constraints, we could alternatively

- replace the MAP by the MLE, and
- the Hessian  $-\mathbf{H}(\hat{\boldsymbol{\theta}}_{\text{MAP}})$  by the Fisher information  $n\mathbf{r}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$ .

# Frequentist approximations

If we use these approximations instead, we get

$$\begin{aligned} \log p(\mathbf{y}) = & \log p(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{\text{MLE}}) - \frac{p}{2} \log n + \log p(\hat{\boldsymbol{\theta}}_{\text{MLE}}) \\ & - \frac{1}{2} \log |\boldsymbol{\nu}(\hat{\boldsymbol{\theta}}_{\text{MLE}})| + \frac{p}{2} \log(2\pi) + \mathbf{O}(n^{-1/2}) \end{aligned}$$

where the error is now  $\mathbf{O}(n^{-1/2})$  due to replacing the true information by its sample counterpart.

## Reducing the approximation rate

Ignoring all but the two first terms leads to  $O(1)$  approximation error, unless we consider the setting where we take a prior centered at the MLE with unit Fisher information precision (equivalent to  $n = 1$  phantom observation). Then, due to cancellation of terms in the expansion,

$$\boldsymbol{\theta} \sim \text{Gauss}_p\{\hat{\boldsymbol{\theta}}_{\text{MLE}}, \boldsymbol{\nu}^{-1}(\hat{\boldsymbol{\theta}}_{\text{MLE}})\},$$

with approximation error of  $O(n^{-1/2})$ .

## Frequentist approximation to the marginal likelihood

This gives the approximation, whose quality improves with increasing sample size  $n$ :

$$-2 \log p(\mathbf{y}) \approx \text{BIC} = -2 \log p(\mathbf{y} \mid \boldsymbol{\theta}) + p \log n$$

The unnormalized weight  $\exp(-\text{BIC}/2)$  is an approximation for the marginal likelihood sometimes used for model comparison in Bayes factor.

# Bayesian model averaging via BIC

We consider the `diabetes` model from Park & Casella (2008) and models with 10 predictors plus intercept.

```
1 library(leaps)
2 data(diabetes, package = "lars")
3 search <- leaps::regsubsets(
4   x = diabetes$x,
5   y = diabetes$y,
6   intercept = TRUE,
7   method = "exhaustive",
8   nvmax = 10, nbest = 99, really.big = TRUE)
9 models_bic <- summary(search)$bic
10 # Renormalize BIC and keep only models with some weight
11 bic <- models_bic - min(models_bic)
12 models <- which(bic < 7)
13 bma_weights <- exp(-bic[models]/2)/sum(exp(-bic[models]/2))
```

# Weights and model components

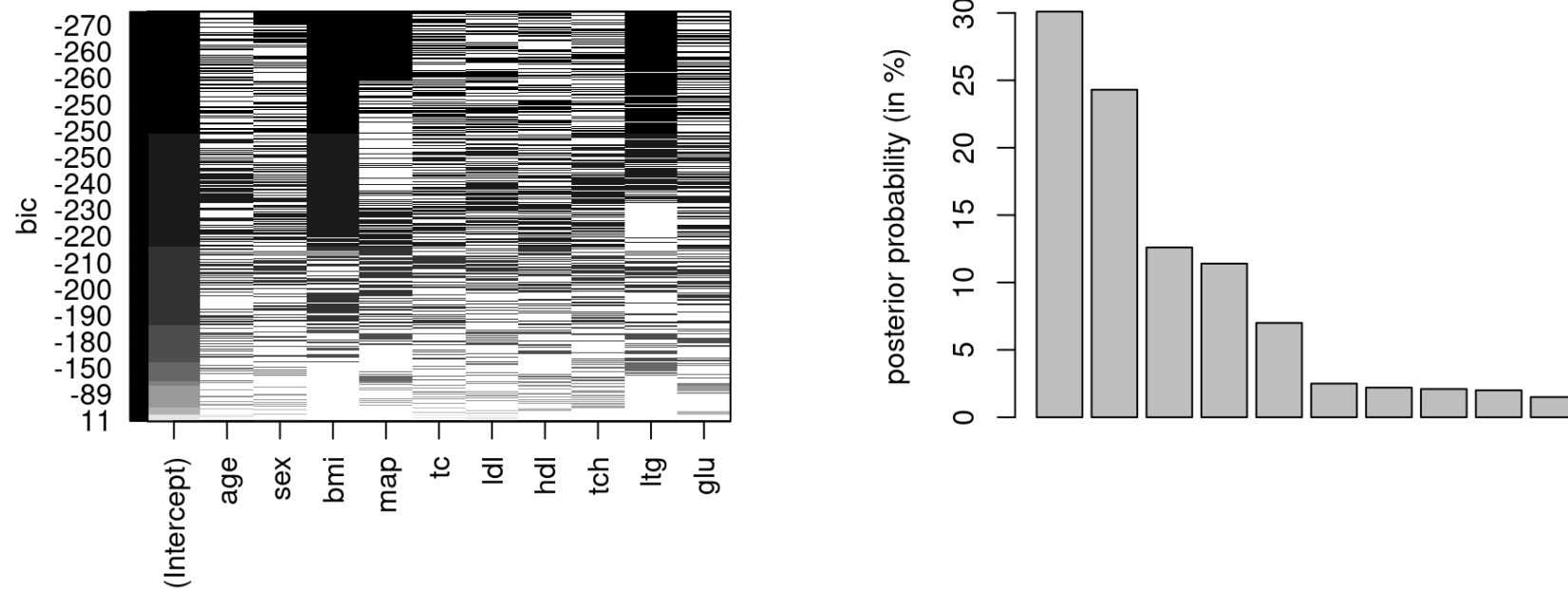


Figure 1: BIC as a function of the linear model covariates (left) and Bayesian model averaging approximate weights (in percentage) for the 10 models with the highest posterior weights according to the BIC approximation.



# Gaussian approximation to the posterior

We can also use similar ideas to approximate the posterior. Suppose that we can Taylor expand the log prior and log density around their respective mode, say  $\hat{\theta}_0$  and  $\hat{\theta}_{\text{MLE}}$ , with  $J_0(\hat{\theta}_0)$  and  $J(\hat{\theta}_{\text{MLE}})$  denoting negative of the corresponding Hessian matrices.

# Approximation to posterior

Together, these yield

$$\log p(\boldsymbol{\theta}) \approx \log p(\hat{\boldsymbol{\theta}}_0) - \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)^\top J_0(\hat{\boldsymbol{\theta}}_0) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0)$$

$$\begin{aligned} \log p(\mathbf{y} \mid \boldsymbol{\theta}) &\approx \log p(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{\text{MLE}}) \\ &\quad - \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{MLE}})^\top J(\hat{\boldsymbol{\theta}}_{\text{MLE}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{MLE}}) \end{aligned}$$

# Gaussian approximation to posterior

The approximate posterior must be Gaussian with precision  $J_n^{-1}$  and mean  $\mu_n$ , where

$$J_n = J_0(\hat{\boldsymbol{\theta}}_0) + J(\hat{\boldsymbol{\theta}}_{\text{MLE}})$$
$$\hat{\boldsymbol{\theta}}_n = J_n^{-1} \left\{ J_0(\hat{\boldsymbol{\theta}}_0) \hat{\boldsymbol{\theta}}_0 + J(\hat{\boldsymbol{\theta}}_{\text{MLE}}) \hat{\boldsymbol{\theta}}_{\text{MLE}} \right\}$$

and note that  $J_0(\hat{\boldsymbol{\theta}}_0) = O(1)$ , whereas  $J_n$  is  $O(n)$ .

# Gaussian large-sample approximation to MLE

Suppose that the prior is continuous and positive in a neighborhood of the maximum.

Assume further that the regularity conditions for maximum likelihood estimator holds. Then, in the limit as  $n \rightarrow \infty$

$$\boldsymbol{\theta} \mid \mathbf{y} \sim \text{Gauss}_p \left\{ \hat{\boldsymbol{\theta}}_{\text{MLE}}, J^{-1}(\hat{\boldsymbol{\theta}}_{\text{MLE}}) \right\}$$

## Corollary

In large samples, the inference obtained from using likelihood-based inference and Bayesian methods will be equivalent

- credible intervals will also have guaranteed frequentist coverage.

Misspecified model: Bayesian will return the model from the family that minimizes the Kullback–Leibler divergence with the true data generating process.

# Regularity conditions

- The maximizer must be uniquely identified from the data and must not be on boundary, so that we can perform a two-sided Taylor series expansion around  $\theta_0$ .
- We need to be able to apply the law of large numbers to get the variance (reciprocal Fisher information) and apply a central limit theorem to the score.
- The third-order derivative of the likelihood is bounded: we can get away with weaker, but this is easiest to check (and ensures that the higher order terms of the Taylor series expansion vanishes asymptotically).

# Gaussian approximation to gamma posterior

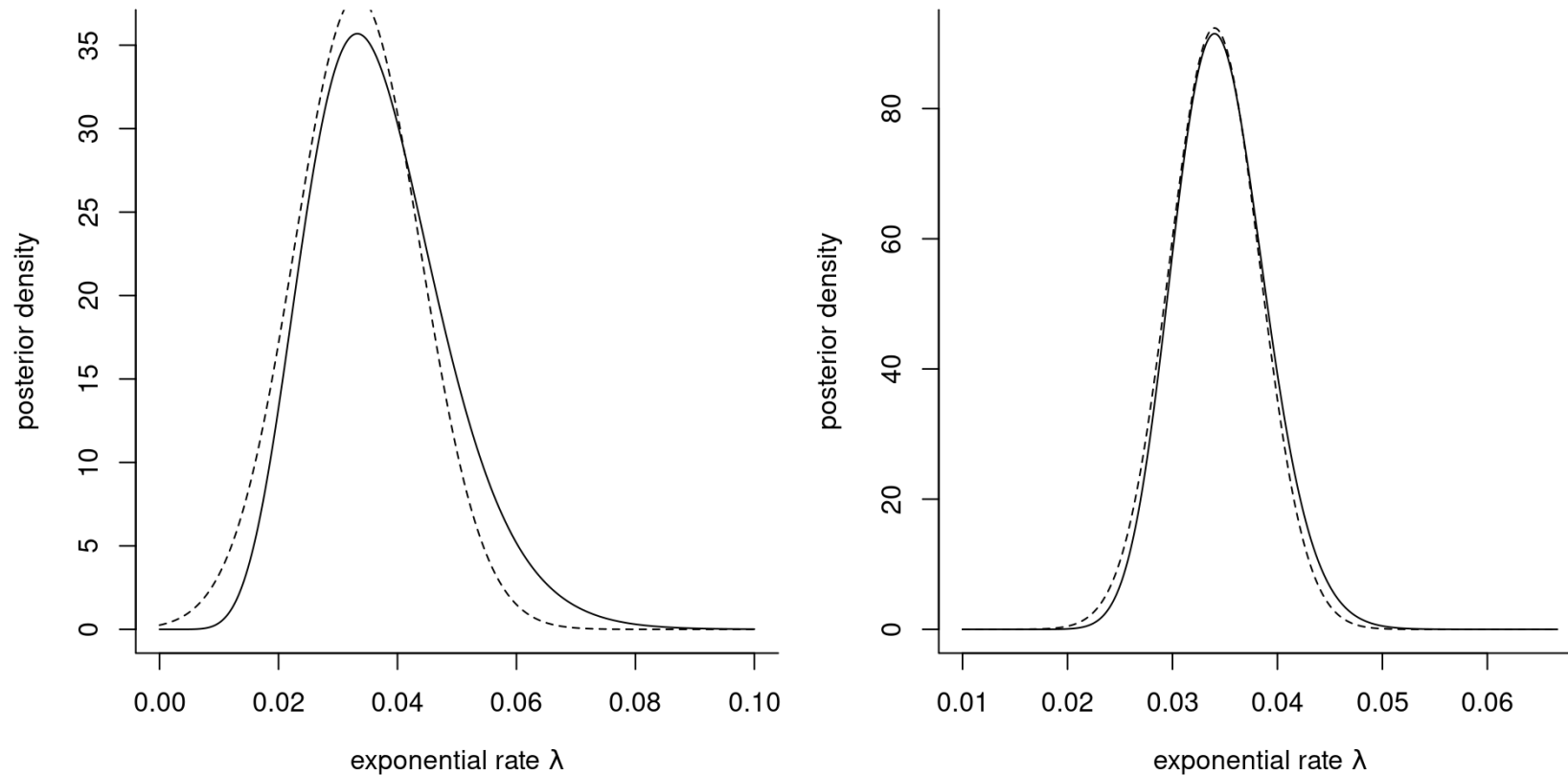


Figure 2: Gaussian approximation (dashed) to the posterior density (full line) of the exponential rate  $\lambda$  for the `waiting` dataset with an exponential likelihood and a gamma prior with  $a = 0.01$  and  $b = 0.01$ . The plots are based on the first 10 observations (left) and the whole sample of size  $n = 62$  (right).

# Structured models

Models in genomics data or spatio-temporal applications including random effects (e.g., spline smoothers) are predominant.

They often contain several thousands or millions of latent parameters.

Inference becomes **unfeasible** in reasonable time using methods we covered so far.



# Integrated nested Laplace approximation

Consider a model with response  $\mathbf{y}$  which depends on covariates  $\mathbf{x}$  through a latent Gaussian process  $\beta$ .

Typically, the prior of coefficients  $\beta \in \mathbb{R}^p$ . The dimension  $p$  can be substantial (several thousands) with a comparably low-dimensional hyperparameter vector  $\theta \in \mathbb{R}^m$ .

Consider data that are conditionally independent given  $\beta$  and  $\beta \sim \text{Gauss}_p(\mathbf{0}_p, \mathbf{Q}^{-1})$  for simplicity.

# Gaussian approximations considered

Then,

$$p(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^\top \mathbf{Q} \boldsymbol{\beta} + \sum_{i=1}^n \log p(y_i \mid \beta_i, \boldsymbol{\theta}) \right\}$$

If  $\boldsymbol{\beta}$  is a Gaussian Markov random field, its precision matrix  $\mathbf{Q}$  will be sparse. A Gaussian approximation to this model would have precision matrix  $\mathbf{Q} + \text{diag}(\mathbf{c})$  for some vector  $\mathbf{c}$  obtained from the second-order expansion of the likelihood.

This allows one to use dedicated algorithms for sparse matrices.

# INLA: targets of inference

Interest typically then lies in marginal parameters

$$p(\beta_i | \mathbf{y}) = \int p(\beta_i | \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}$$

$$p(\theta_i | \mathbf{y}) = \int p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-i}$$

where  $\boldsymbol{\theta}_{-i}$  denotes the vector of hyperparameters excluding the  $i$ th element  $\theta_i$ .

# Philosophy of INLA

The INLA method builds Laplace approximations to the integrands  $p(\beta_i | \boldsymbol{\theta}, \mathbf{y})$  and  $p(\boldsymbol{\theta} | \mathbf{y})$ , and evaluates the integral using quadrature rules over a coarse grid of values of  $\boldsymbol{\theta}$ .

The marginal posterior  $p(\boldsymbol{\theta} | \mathbf{y})$  is approximated by writing

$$p(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{y}) \propto p(\boldsymbol{\beta} | \boldsymbol{\theta}, \mathbf{y})p(\boldsymbol{\theta} | \mathbf{y})$$

and performing a Laplace approximation for fixed value of  $\boldsymbol{\theta}$  for the term  $p(\boldsymbol{\beta} | \boldsymbol{\theta}, \mathbf{y})$ , whose mode we denote by  $\hat{\boldsymbol{\beta}}$ .

# INLA approximation (step 1)

This yields

$$\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \frac{p(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta} \mid \mathbf{y})}{p_G(\hat{\boldsymbol{\beta}} \mid \mathbf{y}, \boldsymbol{\theta})} = \frac{p(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta} \mid \mathbf{y})}{|\mathbf{H}(\hat{\boldsymbol{\beta}})|^{1/2}}$$

## Note on approximation

The Laplace approximation  $p_G(\hat{\beta} \mid \mathbf{y}, \boldsymbol{\theta})$  has kernel

$$p_G(\beta \mid \mathbf{y}, \boldsymbol{\theta}) \propto |\mathbf{H}(\hat{\beta})|^{1/2} \exp\left\{-\frac{(\beta - \hat{\beta})^\top \mathbf{H}(\hat{\beta})(\beta - \hat{\beta})}{2}\right\}$$

since it is evaluated at  $\hat{\beta}$ , we retrieve only the determinant of the negative Hessian of  $p(\beta \mid \boldsymbol{\theta}, \mathbf{y})$ , namely  $\mathbf{H}(\hat{\beta})$ . Note that the latter varies with  $\boldsymbol{\theta}$ .

# Numerical integration

To obtain  $p(\theta_i \mid \mathbf{y})$ , we then proceed with

1. finding the mode of  $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$  using a Newton's method, approximating the gradient and Hessian via finite differences.
2. Compute the negative Hessian at the mode to get an approximation to the covariance of  $\boldsymbol{\theta}$ . Use an eigendecomposition to get the principal directions  $\mathbf{z}$ .

## Numerical integration (2)

3. In each direction of  $\mathbf{z}$ , consider drops in  $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$  as we move away from the mode and define a coarse grid based on these, keeping points where the difference in  $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$  relative to the mode is less than some small  $\delta$ .
4. Retrieve the marginal by numerical integration using the central composition design outline above. We can also use directly avoid the integration and use the approximation at the posterior mode of  $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$ .



## Approximation of marginal of Gaussian latent effect

For the marginal  $p(\beta_i | \mathbf{y})$  term, Rue et al. (2009) proceed instead with the marginal for  $\beta_i$  by building an approximation of it based on maximizing  $\beta_{-i} | \beta_i, \boldsymbol{\theta}, \mathbf{y}$  to yield  $\hat{\boldsymbol{\beta}}_{(i)}$  whose  $i$ th element is  $\beta_i$ , yielding

$$\tilde{p}(\beta_i | \boldsymbol{\theta}, \mathbf{y}) \propto \frac{p(\hat{\boldsymbol{\beta}}_{(i)}, \boldsymbol{\theta} | \mathbf{y})}{\tilde{p}(\hat{\boldsymbol{\beta}}_{(i), -i} | \beta_i, \boldsymbol{\theta}, \mathbf{y})},$$

with a suitable renormalization of  $\tilde{p}(\hat{\boldsymbol{\beta}}_{(i), -i} | \beta_i, \boldsymbol{\theta}, \mathbf{y})$ .

## Remark on approximation

While we could use the Laplace approximation  $p_G(\hat{\beta} \mid \mathbf{y}, \theta)$  and marginalize the latter directly, this leads to evaluation of the Laplace approximation to the density far from the mode, which is often inaccurate.

# Numerical challenges to Laplace approximation

One challenge is that  $p$  is **very large**, so calculation of the Hessian  $\mathbf{H}$  is costly to evaluate.

Having to evaluate it repeatedly for each marginal  $\beta_i$  for  $i = 1, \dots, p$  is prohibitive since it involves factorizations of  $p \times p$  matrices.

## Further approximations

To reduce the computational costs, Rue et al. (2009) propose to use the approximate mean to avoid optimizing and use the conditional of the Gaussian approximation with mean  $\hat{\boldsymbol{\beta}}$  and covariance  $\boldsymbol{\Sigma} = \mathbf{H}^{-1}(\hat{\boldsymbol{\beta}})$ ,

$$\boldsymbol{\beta}_{-i} \mid \beta_i, \boldsymbol{\theta}, \mathbf{y} \approx \text{Gauss}_{p-1} \left( \tilde{\boldsymbol{\beta}}_{(i)}, \mathbf{M}_{-i,-i}^{-1} \right);$$

$$\tilde{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}}_{-i} + \boldsymbol{\Sigma}_{i,i}^{-1} \boldsymbol{\Sigma}_{i,-i} (\beta_i - \hat{\beta}_i),$$

This only requires a rank-one update.

## Further approximations

Wood (2019) suggest to use a Newton step to correct  $\tilde{\beta}_{(i)}$ , starting from the conditional mean.

The second step is to exploit the local dependence on  $\beta$  using the Markov structure to build an improvement to the Hessian. Various strategies are proposed in Rue et al. (2009) and Wood (2019).

Nowadays, the INLA software uses a low-rank variational correction to Laplace method, proposed in van Niekerk & Rue (2024).

# The INLA software

The **INLA R package** provides an interface to fit models with Gaussian latent random effects. While the software is particularly popular for spatio-temporal applications using the SPDE approach, we revisit two examples in the sequel where we can exploit the Markov structure.

# Stochastic volatility model with INLA

Financial returns  $Y_t$  typically exhibit time-varying variability. The **stochastic volatility** model is a parameter-driven model that specifies

$$Y_t = \exp(h_t/2) Z_t$$
$$h_t = \gamma + \phi(h_{t-1} - \gamma) + \sigma U_t$$

where  $U_t \stackrel{\text{iid}}{\sim} \text{Gauss}(0, 1)$  and  $Z_t \stackrel{\text{iid}}{\sim} \text{Gauss}(0, 1)$ . The [INLA documentation](#) provides information about which default prior and hyperparameters are specified. We use a  $\text{gamma}(1, 0.001)$  prior for the precision.

# Stochastic volatility code

```
1 library(INLA)
2 data(exchangerate, package = "hecbayes")
3 # Compute response from raw spot exchange rates at noon
4 y <- 100*diff(log(exchangerate$dexrate))
5 time <- seq_along(y)
6 data <- data.frame(y = y, time = time)
7 f_stochvol <- y ~ f(time, model = "ar1",
8                     param = list(prec = c(1, 0.001)))
9 mod_stochvol <- inla(f_stochvol, family = "stochvol", data = data)
10 # Obtain summary
11 summary <- summary(mod_stochvol)
12 # plot(mod_stochvol)
13 marg_prec <- mod_stochvol$marginals.hyperpar[[1]]
14 marg_phi <- mod_stochvol$marginals.hyperpar[[2]]
```



# Marginal posterior approximations

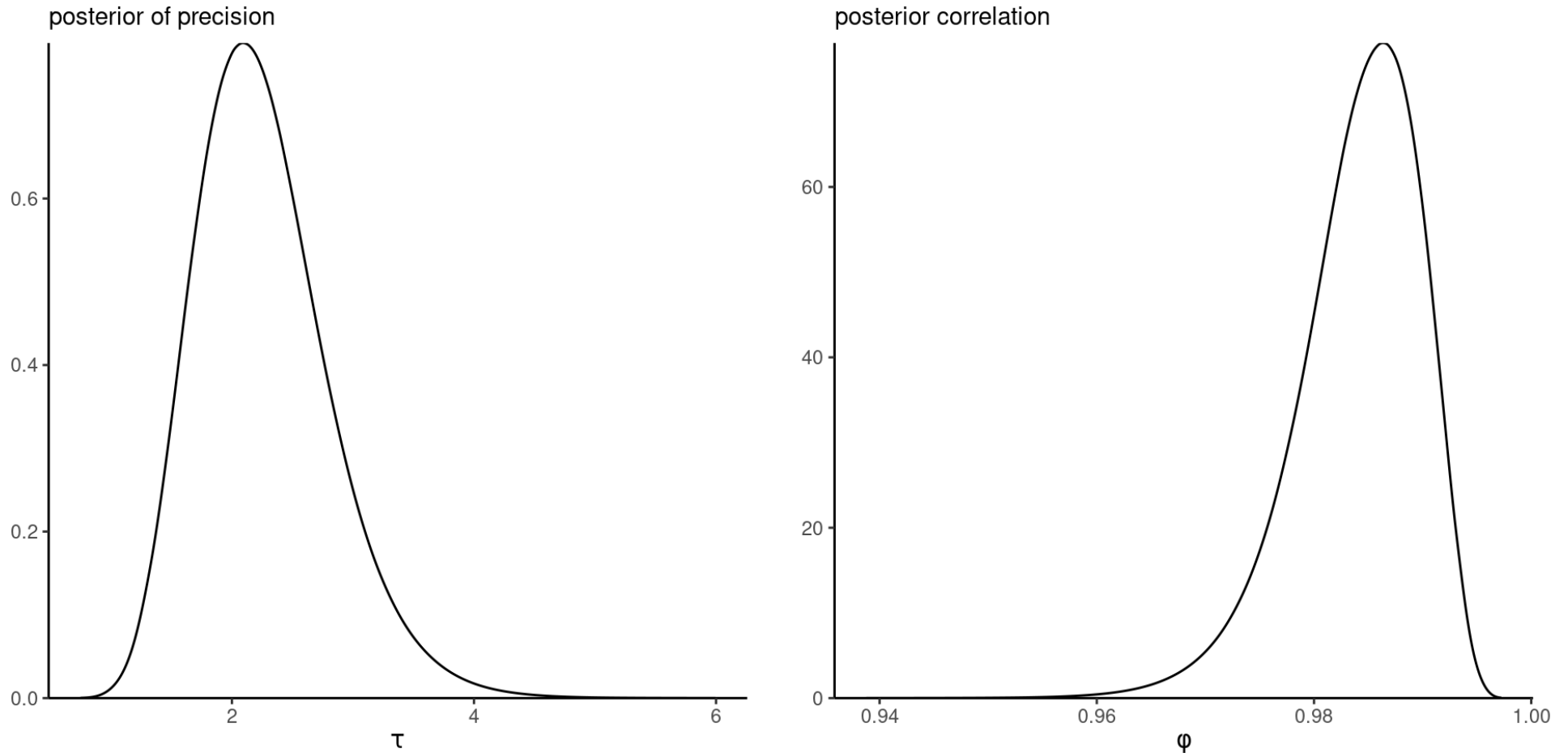


Figure 3: Marginal densities of precision and autocorrelation parameters from the Gaussian stochastic volatility model.

## Comment on stochastic volatility

Figure 3 shows that the correlation  $\phi$  is nearly one, leading to random walk behaviour and high persistence over time (this is also due to the frequency of observations).

This strong serial dependence in the variance is in part responsible for the difficulty in fitting this model using MCMC.

# Marginal approximations

We can use the marginal density approximations to obtain quantiles for summary of interest, marginal posterior moments, etc.

The software also includes utilities to transform the parameters using the change of variable formula.

# Marginal summaries

```

1 # Compute density, quantiles, etc. via inla.*marginal
2 # Change of variable to get variance from precision
3 marg_var <- INLA::inla.tmarginal(
4   fun = function(x) { 1 / x },
5   marginal = marg_prec)
6 INLA::inla.qmarginal(marg_var, p = c(0.025, 0.5, 0.975))

```

```
[1] 0.2864905 0.4543037 0.7396820
```

```

1 # Posterior marginal mean and variance of phi
2 mom1 <- INLA::inla.emarginal(
3   fun = function(x){x},
4   marginal = marg_phi)
5 mom2 <- INLA::inla.emarginal(
6   fun = function(x){x^2},
7   marginal = marg_phi)
8 c(mean = mom1, sd = sqrt(mom2 - mom1^2))

```

```

      mean      sd
0.984052723 0.005762531

```

# Tokyo binomial time series

We revisit the Tokyo rainfall example, but this time fit the model with `INLA` rather than MCMC.

We specify the mean model without intercept and fit a logistic regression, with a second-order cyclic random walk prior for the coefficients, and the default priors for the other parameters.

# Code syntax

```
1 data(Tokyo, package = "INLA")
2 # Formula (removing intercept)
3 formula <- y ~ f(time, model = "rw2", cyclic = TRUE) - 1
4 mod <- INLA::inla(
5   formula = formula,
6   family = "binomial",
7   Ntrials = n,
8   data = Tokyo)
```

The marginal posterior with pointwise 95% credible intervals on the next slide show nearly identical results to marginals from the probit model.

# Posterior of random effect prior

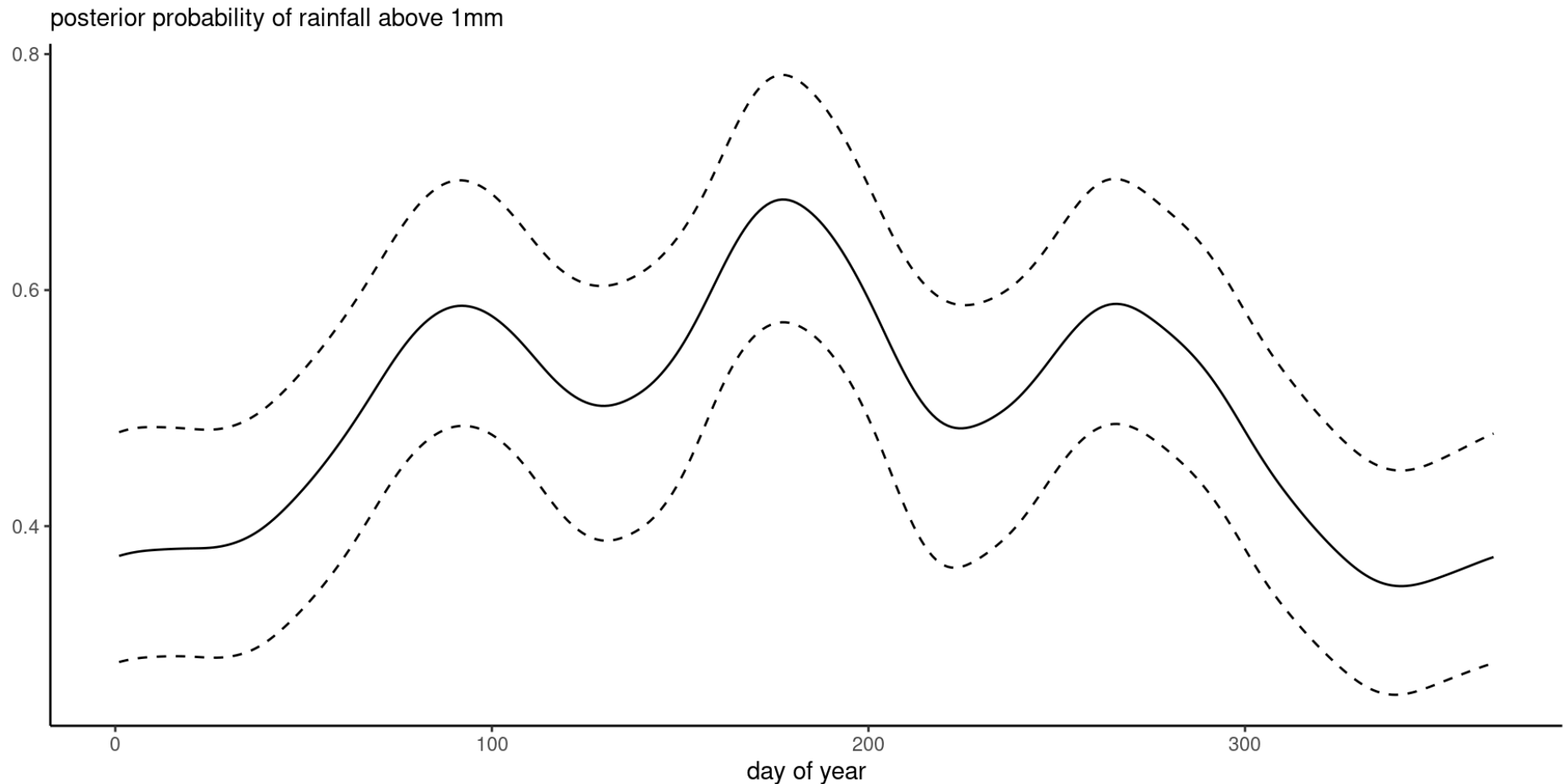


Figure 4: Posterior probability per day of the year with posterior median and 95% credible interval for the Tokyo rainfall binomial time series.

# References

- Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681–686. <https://doi.org/10.1198/0162145080000000337>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>
- Rue, H., Martino, S., & 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. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393), 82–86. <https://doi.org/10.1080/01621459.1986.10478240>
- van Niekerk, J., & Rue, H. (2024). Low-rank variational Bayes correction to the Laplace method. *Journal of Machine Learning Research*, 25(62), 1–25. <http://jmlr.org/papers/v25/21-1405.html>
- Wood, S. N. (2019). Simplified integrated nested Laplace approximation. *Biometrika*, 107(1), 223–230. <https://doi.org/10.1093/biomet/asz044>