Bayesian SAE with an Emphasis on Low- and Middle-Income Countries: Lecture 2: Bayesian Inference and Temporal Smoothing Models

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Motivation

Motivation

- We give a brief introduction to Bayesian inference, which we will use for many of our modeling endeavors.
- Smoothing in space can be achieved using prior distributions - such models can increase precision by effectively increasing the sample size.
- Can be thought of as penalizing spatial surfaces that are too "jumpy".
- As a prelude to introducing spatial smoothing models, we consider the simpler case of temporal smoothing models.

Bayesian Inference

Frequentist Inference

- Suppose we wish to make inference for p unknown parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]$, given n data points $\boldsymbol{y} = [y_1, \ldots, y_n]$.
- In frequentist inference, the parameters $\boldsymbol{\theta}$ are viewed as constants, i.e., not as random variables.
- Inference follows from evaluating procedures over hypothetical datasets generated under the same conditions as the actual data.
- For example, for constant $\boldsymbol{\theta}$ we examine how an estimator,

$$\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{Y})$$

  performs under repeated sampling, so that $\widehat{\boldsymbol{\theta}}(\boldsymbol{Y})$ is a random variable, because $\boldsymbol{Y}$ is.
- Under Bayesian inference, the data $\boldsymbol{y}$ are taken as a fixed set of numbers, i.e., constant, so not random variables, whereas the parameters $\boldsymbol{\theta}$ are treated as random variables.

Bayesian Inference

- Inference is made through the posterior probability distribution of θ after observing $\boldsymbol{y}$, and is determined from Bayes theorem:

$$p(\theta \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \theta) \times \pi(\theta)}{p(\boldsymbol{y})}.$$

- For continuous θ, the normalizing constant is

$$p(\boldsymbol{y}) = \int_{\theta} p(\boldsymbol{y} \mid \theta)\pi(\theta)\, d\theta.$$

- This is the marginal probability of the observed data averaged over the assumed model, i.e., the likelihood and prior.

- To emphasize: Bayesian inferential probabilities are for random θ conditional on fixed data $\boldsymbol{y}$, which is in stark contrast to frequentist inferential probabilities which are over hypothetical sampling of new datasets $\boldsymbol{Y}$ for fixed θ.

Bayesian Inference

Bayesian inference is a convenient framework within which to implement temporal and spatial smoothing models.

- A Sampling Model (Likelihood) describes the distribution of the data, depending on unknown parameters θ.
- A Penalization (Prior) that expresses beliefs about the parameters θ encoding the model, in particular we can mathematically place soft constraints on "neighboring" points.
- Combination occurs via Bayes Theorem:

$$\underbrace{p(\theta \mid y)}_{\text{Posterior}} \propto \underbrace{L(\theta)}_{\text{Likelihood}} \times \underbrace{\pi(\theta)}_{\text{Prior}}.$$

- On the log scale:

$$\underbrace{\log p(\theta \mid y)}_{\text{Updated Beliefs}} = \underbrace{\log L(\theta)}_{\text{Data Model}} + \underbrace{\log \pi(\theta)}_{\text{Penalization}}.$$

Posterior Inferential Summaries

- To summarizes the typically multivariate posterior distribution, $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, marginal distributions for parameters of interest may be considered.

- For example the univariate marginal distribution for a component θ_i is,

$$p(\theta_i \mid \boldsymbol{y}) = \int_{\boldsymbol{\theta}_{-i}} p(\boldsymbol{\theta} \mid \boldsymbol{y}) \, d\boldsymbol{\theta}_{-i}, \tag{1}$$

  where

$$\boldsymbol{\theta}_{-i} = [\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p]$$

  is the vector $\boldsymbol{\theta}$ excluding θ_i, i.e., we are integrating over the $p-1$ additional parameters.

- By evaluating (1) at different values for θ_i we can construct the complete univariate distribution.

Posterior Inferential Summaries

- The posterior marginal distribution may be further summarized.
- In frequentist inference, we have a point estimate and a measure of uncertainty, for example the standard error, both reflections of the sampling distribution.
- In Bayesian inference, the fundamental summary is the posterior distribution, then we choose how to summarize – posterior moments and quantiles are obvious candidates.
- The posterior mean is,

$$\mathrm{E}[\theta_i \mid \boldsymbol{y}] = \int_{\theta_i} \theta_i p(\theta_i \mid \boldsymbol{y})\, d\theta_i. \tag{2}$$

- The posterior standard deviation is,

$$\mathrm{sd}(\theta_i \mid \boldsymbol{y}) = \sqrt{\mathrm{E}\left[(\theta_i - \mathrm{E}[\theta_i \mid \boldsymbol{y}])^2\right]} = \sqrt{\int_{\theta_i} (\theta_i - \mathrm{E}[\theta_i \mid \boldsymbol{y}])^2 p(\theta_i \mid \boldsymbol{y})\, d\theta_i}. \tag{3}$$

Inferential Summaries

- The $100 \times q\%$ posterior quantile, $\theta_i(q)$ $(0 < q < 1)$ is found by solving

$$\int_{-\infty}^{\theta_i(q)} p(\theta_i \mid \boldsymbol{y})\, d\theta_i. \tag{4}$$

- In particular, the posterior median, $\theta_i(0.5)$, will often provide an adequate summary of the location of the posterior marginal distribution.

- A $100 \times p\%$ equi-tailed credible interval $(0 < p < 1)$ is provided by

$$[\, \theta_i(1-p)/2), \theta_i\{(1+p)/2\}\,],$$

For example, a 90% credible interval for parameter i is

$$[\, \theta_i(0.05), \theta_i(0.95)\,].$$

- Posterior quantiles are more useful summaries than posterior mean and standard deviations, when the posterior distribution is not symmetric.

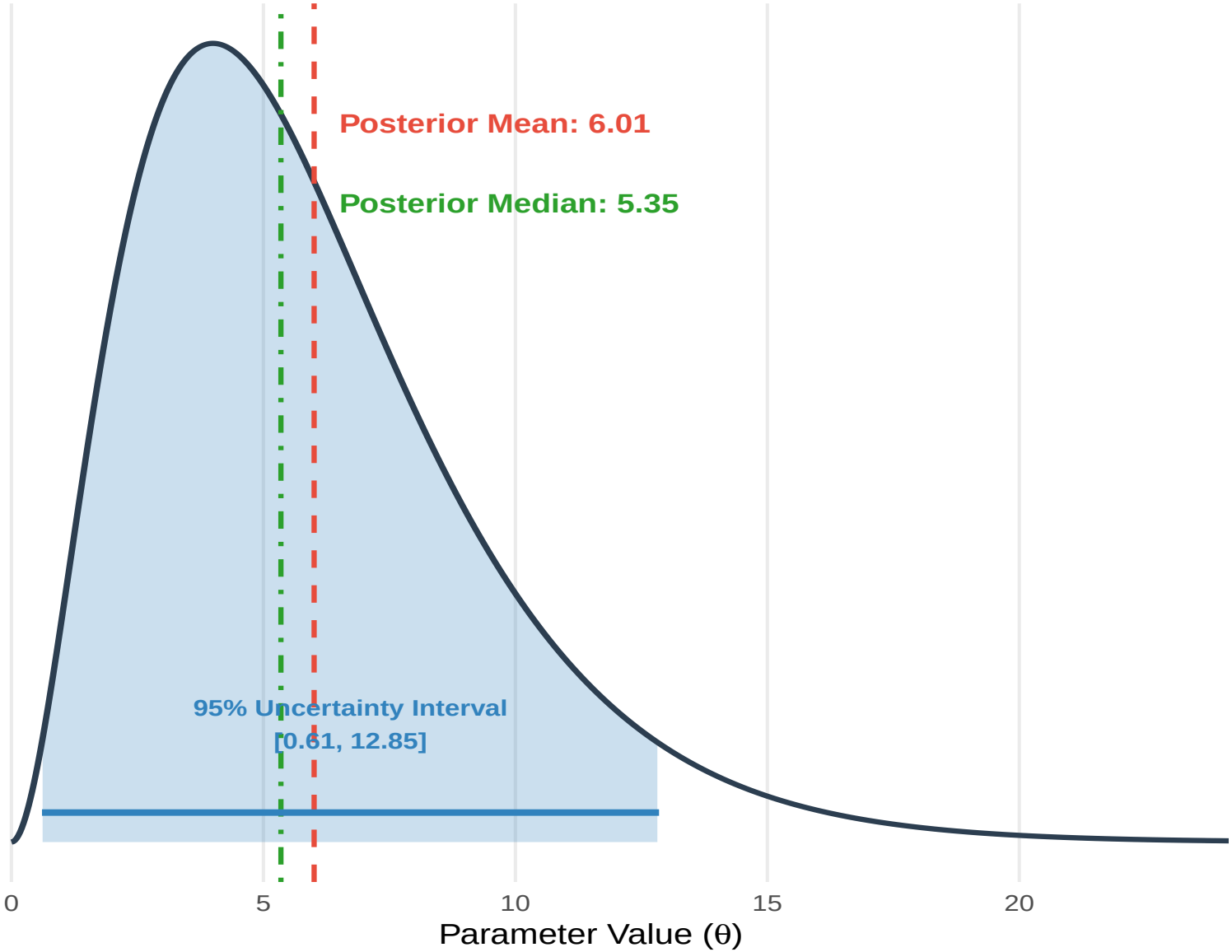

FIGURE 1: Posterior distribution along with mean, posterior median and 95% uncertainty interval.

Frequentist Properties

- Posterior mean and median tend to the MLE in large samples.
- What are the frequentist properties of Bayes credible intervals?
- If we specify a (say) 90% interval, will such intervals contain the true $\theta = \theta_0$, 90% of the time (known as the correct coverage), if we were to repeatedly simulate data from $p(\boldsymbol{y} \mid \theta_0)$?
- With proper priors, Bayes intervals do not have this property, because different choices of θ_0 are weighted differently by the prior.
- Instead, what we can say is that, if we average over the prior, then we will have the correct coverage (so average over data and parameters).
- However, for the priors used in our software, the coverage will tend to the nominal in large samples.

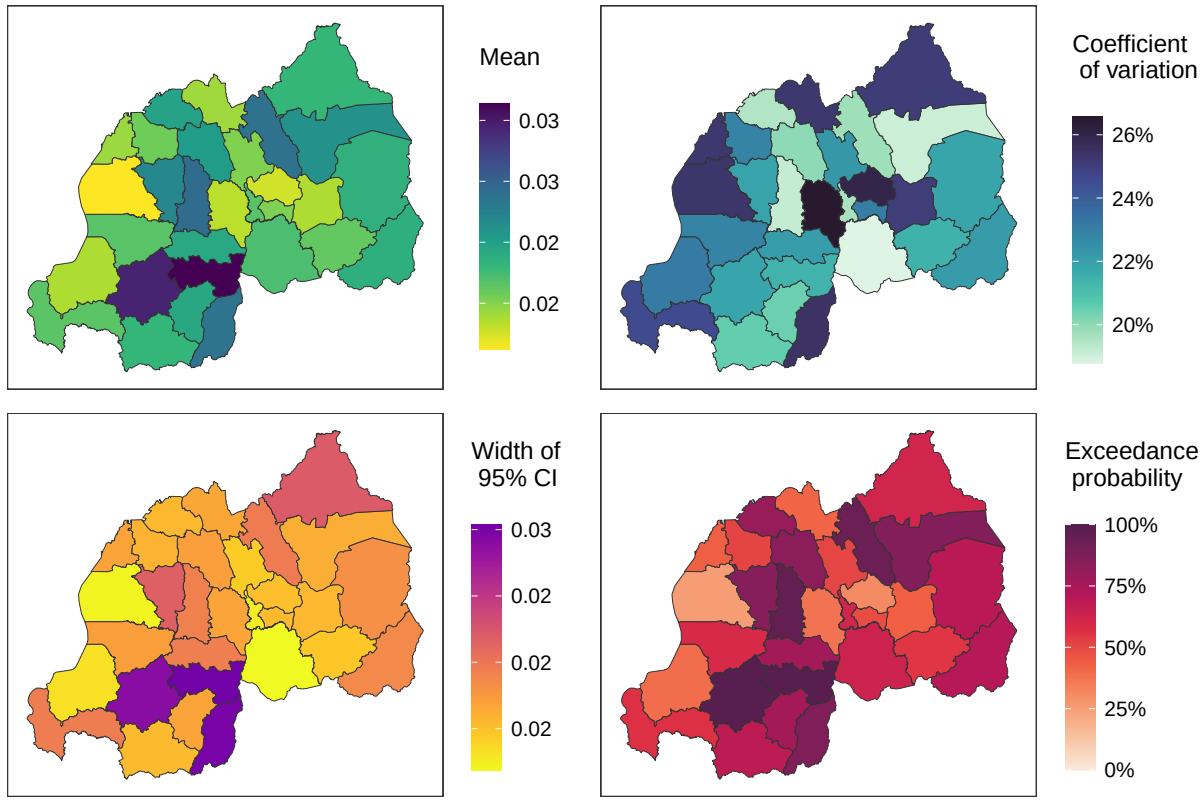

FIGURE 2: Summary measures for area-level (Fay-Herriot) model at Admin2.

- Top left: Posterior mean of the prevalence.
- Top right: Posterior standard deviation divided by the posterior mean, i.e., the posterior coefficient of variation.
- Bottom left: Width of 95% credible interval.
- Posterir probability of exceeding the national level.

Predictive Distributions

- Another useful inferential quantity is the predictive distributions for future observations $\boldsymbol{z}$ which is given, under conditional independence, by

$$\begin{aligned} p(\boldsymbol{z} \mid \boldsymbol{y}) &= \int_{\theta} p(\boldsymbol{z}, \theta \mid \boldsymbol{y})\, d\theta \\ &= \int_{\theta} p(\boldsymbol{z} \mid \theta, \boldsymbol{y}) p(\theta \mid \boldsymbol{y})\, d\theta \\ &= \int_{\theta} \underbrace{p(\boldsymbol{z} \mid \theta)}_{\substack{\text{Conditional} \\ \text{independence}}} p(\theta \mid \boldsymbol{y})\, d\theta \qquad (5) \\ &= \mathrm{E}_{\theta \mid y}\left[p(\boldsymbol{z} \mid \theta)\right] \end{aligned}$$

- Predictive distributions can be used in a cross-validation exercise for model assessment.

Bayesian Inference – In Practice

Bayesian inference is deceptively simple to describe probabilistically, but there have been two major obstacles to its routine use:

- The first is how to specify prior distributions.
- The second is how to do the computation evaluate the integrals required for inference, for example, (1)–(5), given that for most models, these are analytically intractable.

In addition, in contrast to simple frequentist procedures (e.g., least squares) we need to specify a sampling (full probability) model for the data, that is specify the likelihood.

Bayesian Computation

- The computations required for Bayesian inference (integrals) is often not trivial and many be carried out using a variety of analytic, numeric and simulation based techniques.
- Markov chain Monte Carlo (MCMC) is a very popular technique, but is inefficient for spatial data, because of the strong dependencies in the posterior.
- For computation: we use the integrated nested Laplace approximation (INLA), introduced by Rue *et al.* (2009) – rigorously evaluated for SAE models, e.g., Osgood-Zimmerman and Wakefield (2020).

Book-length treatments on INLA:

- Blangiardo and Cameletti (2015) – space-time models.
- Wang *et al.* (2018) – general models.
- Krainski *et al.* (2018) – advanced space-time models.
- Moraga (2019) – introductory text for spatial analysis.
- Gómez-Rubio (2020) – general models.
- Moraga (2023) – introductory text.

Simple Bayes Examples

Example 1: Normal Sampling Model

- Imagine the data model is normal with an unknown mean μ:

$$\bar{y} \mid \mu \sim \mathrm{N}(\mu, \sigma^2/n),$$

where σ^2/n is assumed known ($\sigma/\sqrt{n}$ is the standard error).

- We also imagine the prior is normal:

$$\mu \sim \mathrm{N}(\underbrace{m, v}_{\substack{\text{Specified} \\ \text{constants}}})$$

Values of μ that are (relatively) far from m are penalized.

- The log posterior is:

$$\underbrace{\log p(\mu \mid \boldsymbol{y})}_{\text{Updated Beliefs}} = \underbrace{-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2}_{\text{Data Model}} \underbrace{-\frac{1}{2v}(\mu - m)^2}_{\text{Penalization}}.$$

- Preferred values of μ are consistent with the data $\boldsymbol{y}$ (through $\bar{y}$) and the prior (through m), with the relative weight depending on σ^2/n and v.

Example 1: Normal Sampling Model

The posterior distribution is

$$\mu \mid \boldsymbol{y} \sim \mathrm{N}\left(F \times \overline{y} + (1-F) \times m, F \times \frac{\sigma^2}{n}\right), \qquad F = \frac{nv}{nv+\sigma^2} = \frac{n}{n+\sigma^2 v^{-1}},$$

which makes clear the compromise between data and prior and we have interesting cases:

- $n = 0$ gives $F = 0$ and we recover the prior.
- $v = 0$ (not a good idea) gives $F = 0$ and the posterior is a point mass at the prior mean m.
- If $v^{-1} = 0$ we have an improper prior, and frequentist and Bayesian estimates coincide.
- As $n \to \infty$ then $F \to 1$ (unless $v = 0$), and we hone in on correspondence between Bayes and MLE.
- $\mathrm{var}(\mu \mid \boldsymbol{y}) = F\sigma^2/n \leq \sigma^2/n = \mathrm{var}(\widehat{\mu})$. Prior adds information.

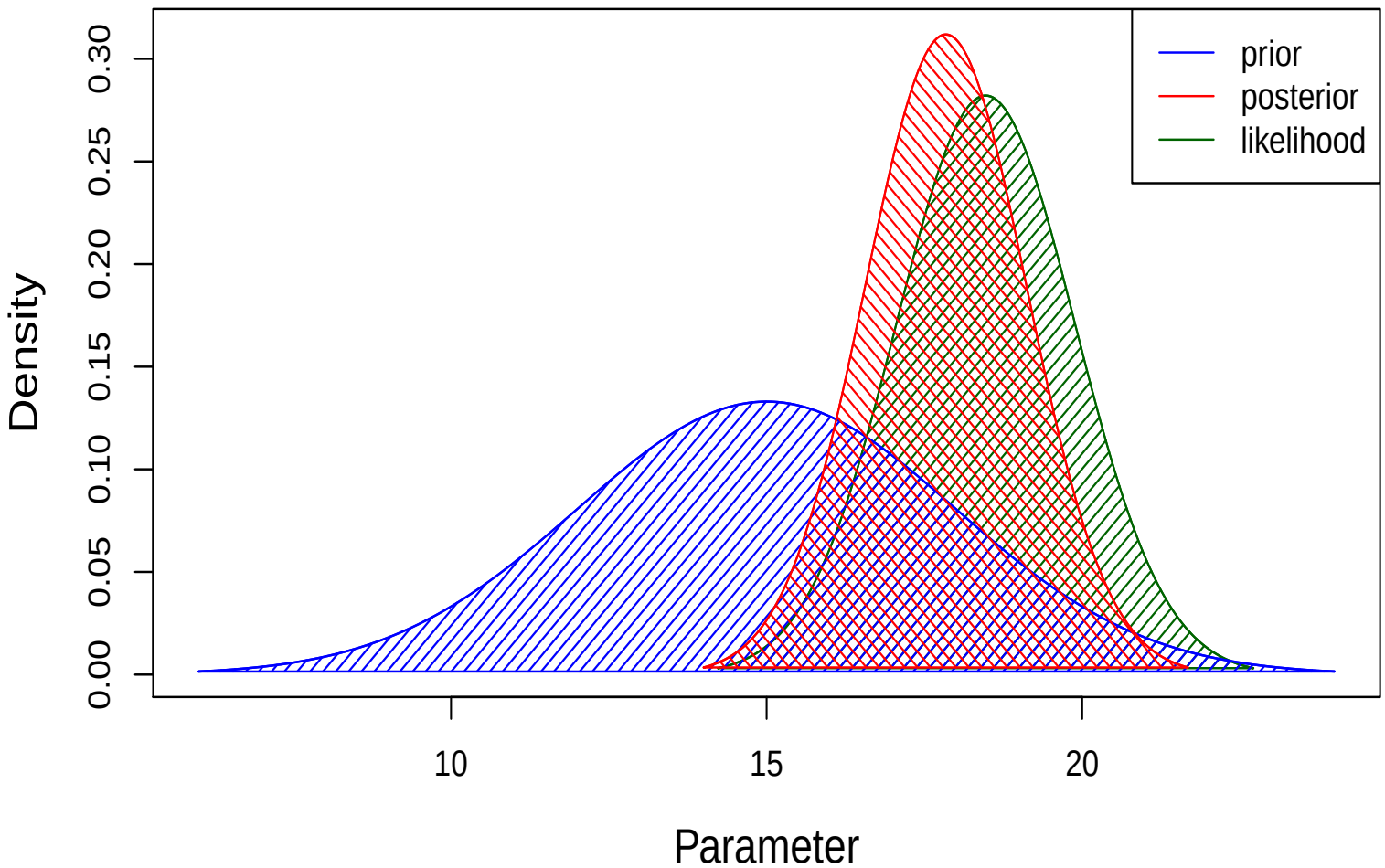

FIGURE 3: Normal data model with $n = 10$, $\overline{y} = 19.3$ and standard error 1.41. The prior for μ has mean m =15 and $v = 3^2$. The posterior for the parameter μ is a compromise between the two sources of information: the posterior mean is 18.5 and the posterior standard deviation is 1.28.

Example 2: Poisson Sampling Model

- Suppose
$$Y \mid \theta \sim \text{Poisson}(E\theta),$$
where $Y = 0, 1, 2, \ldots,$ is the number of disease events, E is the expected number, and $\theta > 0$ is the relative risk.

- For a Bayesian analysis we need a prior for θ.

- Consider the gamma prior, Ga(a, b) which has the form:
$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta), \qquad \theta > 0,$$
which has mean a/b and variance a/b^2.

- To use this as a prior we need to specify a and b to reflect what we believe about the relative risk before we see the data.

Example: Sellafield

We now illustrate the sensitivity of inference to the prior distribution using real data in which $y = 4$ childhood leukemia cases were observed close to the village of Seascale, which was close to the Sellafield nuclear reprocessing plant – based on the population size and a reference risk, we would have expected $E = 0.25$ cases.

- With a gamma prior $\text{Ga}(a, b)$ on θ we obtain a posterior of

$$\begin{aligned} p(\theta \mid y) &\propto L(\theta) \times p(\theta) \\ &\propto \exp(-E\theta)\theta^y \times \theta^{a-1} \exp(-\theta b) \\ &= \theta^{a+y-1} \exp(-\theta[b+E]), \end{aligned}$$

  which is a gamma distribution with parameters $a + y$ and $b + E$, i.e., $\text{Ga}(a + y, b + E)$.

- This is an example of a conjugate analysis, in which the prior and posterior are of the same form.

Example: Sellafield

- The posterior mean is

$$\mathrm{E}[\theta \mid y] = \frac{a+y}{b+E} = F \times \frac{a}{b} + (1-F) \times \frac{y}{E}$$

  where the fraction $F = b/(b+E)$.

- Notice that as E gets bigger, the weight f approaches 0 and $\mathrm{E}[\theta \mid y]$ gets closer to the MLE, which is y/E.

- For a reference analysis we may pick $a = b = 0$ (an improper prior), and in this case the posterior mean coincides with the MLE.

Example: Sellafield

- The quantiles for,

$$y = 4, E = 0.25, \qquad a = b = 0$$

are given in Table 1 and indicate that under this prior there is strong evidence that the relative risk associated with the Seascale area is elevated.

- We also consider an alternative subjective gamma prior with median at 1 and 95% point at 5, these two specifications lead to $a = 0.83, b = 0.53$.

- Posterior quantiles under this prior are in Table 1 – still see strong evidence of increased risk though the quantiles change considerably because data are sparse and prior has influence.

Gamma(a, b) Prior	2.5%	5%	50%	95%	97.5%
$a = b = 0$	4.4	5.5	14.7	31.0	35.1
$a = 0.83, b = 0.53$	2.0	2.4	5.8	11.4	12.8

Table 1: Posterior quantiles.

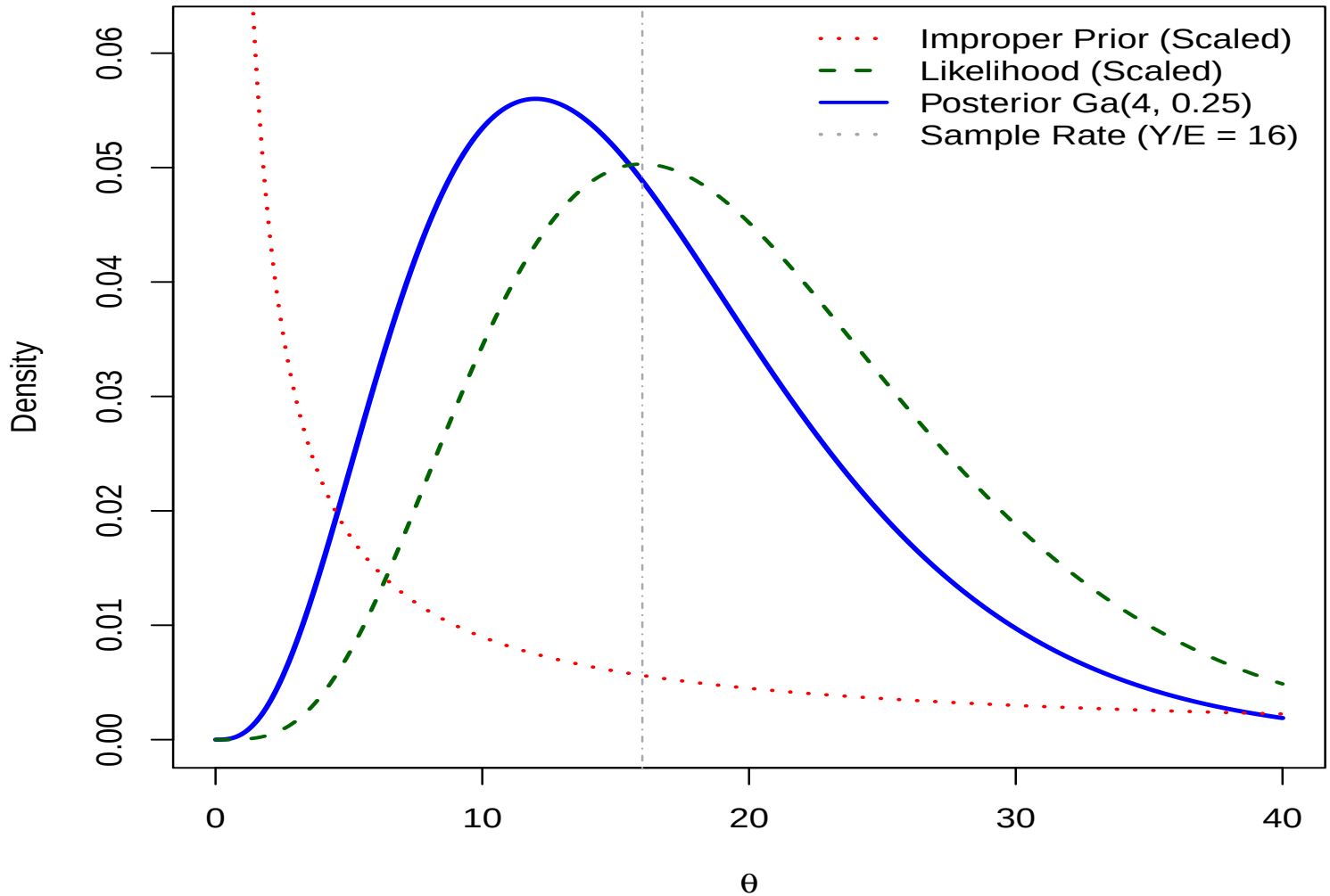

FIGURE 4: Prior (Gamma(a, b) with $a = b = 0$), likelihood and Gamma(4,0.25) posterior.

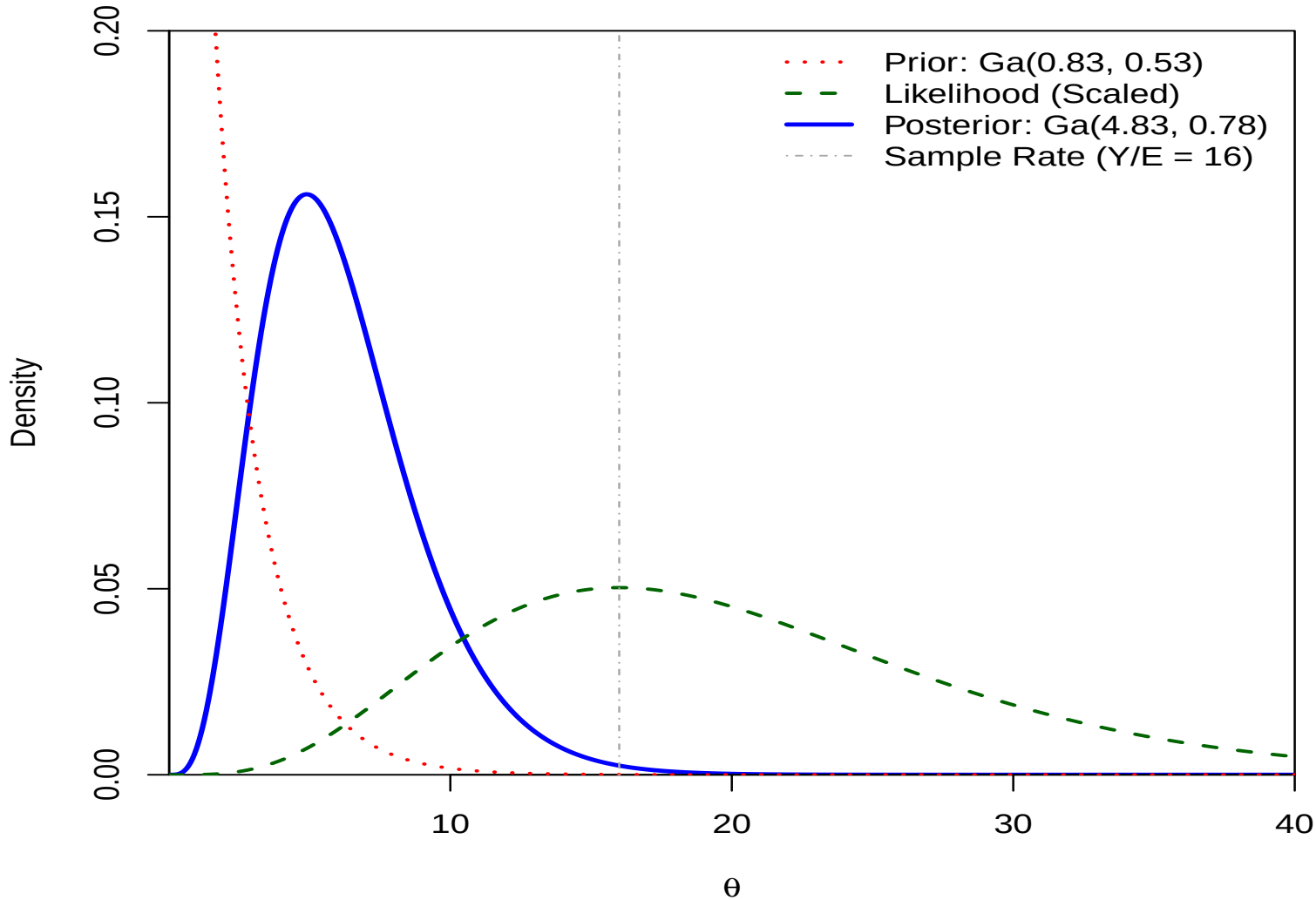

FIGURE 5: Prior (Gamma(a, b) with $a = 0.83, b = 0.53$), likelihood and Gamma(4.83,0.78) posterior.

Bayesian Summary

- To use a Bayesian approach to inference, one must specify a full probability model for the data, and a prior.
- The inferential summary tool is the multivariate posterior distribution, which may be further summarized using marginal distributions, means, medians and quantiles.
- Credible intervals have an obvious interpretation, "the probability that θ lies between a and b is 0.95"; the accuracy of this statement depends on the appropriateness of the model (likelihood and prior).
- The summarization of the posterior often needs to be carried out numerically/via simulation – INLA is extremely fast and accurate.

Temporal Smoothing

Smoothing over Time

We consider the situation in which we have multiple parameters (e.g., monthly prevalence of an infectious disease) indexed by time, and we wish to specify a joint prior that encourages similarity between "close-by" parameters.

Rationale and overview of models for temporal smoothing:

- We often expect that the true underlying prevalence in a population will exhibit some degree of smoothness over time.
- A linear trend in time is unlikely to be suitable for more than a small period of time, and higher degree polynomials can produce erratic fits.
- Hence, local smoothing is preferred.
- Random walk models have proved successful as local smoothers.
- The choice of smoothing parameter is crucial.

Random Walk Models

We use random walk models which encourage the mean of the response at time t, u_t, to be similar to its neighbors.

To simplify, suppose Y_t is the logit prevalence at time t, then we can envisage a model:

$$Y_t = \alpha + u_t + \epsilon_t, \qquad \epsilon_t \sim \mathrm{N}(0, \widehat{V}_i),$$

where the u_t are deviations from the overall mean α.

Two extreme scenarios are:

- $u_t = 0$ so that all the means are the same.
- u_t's are unconstrained and therefore no smoothing (one distinct parameter for each time point, and no interaction between the time points).

Random Walk Models

Random walk models can be described in different (equivalent) ways:

1. Via the joint distribution: $p(u_1, \ldots, u_T)$.

2. One-dimensional distributions conditional on the past: $p(u_t \mid u_{t-1}, u_{t-2}, \ldots)$.

3. One-dimensional distributions conditional on neighbors: $p(u_t \mid \boldsymbol{u}_{-t})$ where $\boldsymbol{u}_{-t} = [u_1, \ldots, u_{t-1}, u_{t+1}, \ldots, u_T]$ is the full collection with the t-th entry removed.

The first and third can be easily generalized to spatial processes.

A Simple Motivating Model

- Again, let Y_t be the logit prevalence and we have

$$Y_t = \alpha + u_t + \epsilon_t, \qquad \epsilon_t \sim \mathsf{N}(0, \widehat{V}_t),$$

where the u_t are deviations from the overall mean.

- A two-stage smoothing model is:

$$\begin{aligned} y_t \mid \alpha, u_t &\sim \mathsf{N}(\alpha + u_t, \widehat{V}_t), \\ u_1, \ldots, u_T \mid \sigma_u^2 &\sim \text{Smoothing-Model} \end{aligned}$$

Smoothing models tie together the $u_1, \ldots, u_T$ by penalizing large deviations.

- The deviations $u_1, \ldots, u_T$ are estimated from:

$$\underbrace{\sum_{t=1}^{T} \frac{(y_t - \alpha - u_t)^2}{\widehat{V}_t}}_{\text{Log-Likelihood}} + \underbrace{\text{Penalization Term on } u_1, \ldots, u_T}_{\text{Depends on Smoothing Parameter } \sigma_u^2}.$$

Random Walk Models

In the first-order random walk model, denoted RW1, the deviations of the logit prevalence at time t, u_t , is modeled as a function of its immediate neighbors via:

$$u_t \mid u_{t-1}, u_{t+1}, \sigma_u^2 \sim \mathrm{N}\left(\frac{1}{2}(u_{t-1} + u_{t+1}), \frac{\sigma_u^2}{2}\right),$$

where σ_u^2 is a smoothing parameter:

- the smoothing parameter is estimated from the data and determines the extent to which deviations from the neighbors are penalized,
- u_t is being encouraged to be close to the mean of its neighbors $\frac{1}{2}(u_{t-1} + u_{t+1})$.

Random Walk Models

- The explicit penalty term for the RW1 model is:

$$p(u_t \mid u_{t-1}, u_{t+1}, \sigma_u^2) \propto \exp\left\{-\frac{1}{2\sigma_u^2}\left[u_t - \tfrac{1}{2}\left(u_{t-1} + u_{t+1}\right)\right]^2\right\}.$$

- Hence:
  - Values of u_t that are close to $\frac{1}{2}(u_{t-1} + u_{t+1})$ are favored (higher density).
  - The relative favourability is governed by σ_u^2 – if this variance is small, then u_t can't stray too far from its neighbors.
- S time step ahead predictions from the RW1 model follow distribution

$$u_{T+S} \mid u_1, \ldots, u_T, \sigma_u^2 \sim \mathrm{N}(u_T, \sigma_u^2 \times S),$$

so that the level is determined by the last value u_T and the variance increases linearly in S

FIGURE 6: Illustration of the RW1 model for smoothing at time 3. The mean of the smoother is the average of the two adjacent points (and is highlighted as •), and deviations from this mean are penalized via the normal distribution shown in **red**. σ_u^2 determines the width of the red normal, and hence the amount of smoothing.

RW1 Model

- Letting $\boldsymbol{u} = [u_1, \ldots, u_T]^\top$, the form of the prior RW1 density is:

$$p(\boldsymbol{u} \mid \sigma_u^2) \propto \exp\left(-\frac{1}{2\sigma_u^2}\sum_{t=1}^{T-1}(u_{t+1} - u_t)^2\right)$$
$$= \exp\left(-\frac{1}{2\sigma_u^2}\sum_{t\sim t'}(u_t - u_{t'})^2\right) = \exp\left(-\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{Q}\boldsymbol{u}\right)$$

  where $t \sim t'$ indicates t is a neighbor of t'.

- the precision is $\boldsymbol{Q} = \boldsymbol{R}/\sigma_u^2$ with

$$\boldsymbol{R} = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}$$

  and zeroes everywhere else – this sparsity leads to big gains in computational efficiency.

RW2 Model

- The second order random walk (RW2) model produces smoother trajectories than the RW1, and has more reasonable short term predictions, which is desirable for many health and demographic indicators.

- The model is defined in terms of second differences:

$$(u_t - u_{t-1}) - (u_{t-1} - u_{t-2}) \mid \sigma_u^2 \sim \mathrm{N}(0, \sigma_u^2),$$

  showing that deviations from linearity are discouraged.

- S time step ahead predictions from the RW2 model follow distribution

$$u_{T+S} \mid u_1, \ldots, u_T, \sigma_u^2 \sim \mathrm{N}\left(u_T + S(u_T - u_{T-1}), \frac{\sigma_u^2}{6} \times S(S+1)(2S+1)\right),$$

  so that the mean is a linear function of the values at the last two time points, and the variance is cubic in the number of periods S, so blows up very quickly.

RW2 Model

- Form of the prior RW2 density is:

$$
\begin{aligned}
p(\boldsymbol{u} \mid \sigma_u^2) &\propto \exp\left(-\frac{1}{2\sigma_u^2}\sum_{t=1}^{T-2}(u_{t+2} - 2u_{t+1} + u_t)^2\right) \\
&= \exp\left(-\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{Q}\boldsymbol{u}\right)
\end{aligned}
$$

- The precision is $\boldsymbol{Q} = \boldsymbol{R}/\sigma_u^2$ with

$$
\boldsymbol{R} = \begin{bmatrix}
1 & -2 & 1 & & & & \\
-2 & 5 & -4 & 1 & & & \\
1 & -4 & 6 & -4 & 1 & & \\
 & 1 & -4 & 6 & -4 & 1 & \\
 & & . & . & . & . & . \\
 & & 1 & -4 & 6 & -4 & 1 \\
 & & & 1 & -4 & 5 & -2 \\
 & & & & 1 & -2 & 1
\end{bmatrix}
$$

and zeroes everywhere else, so again sparse.

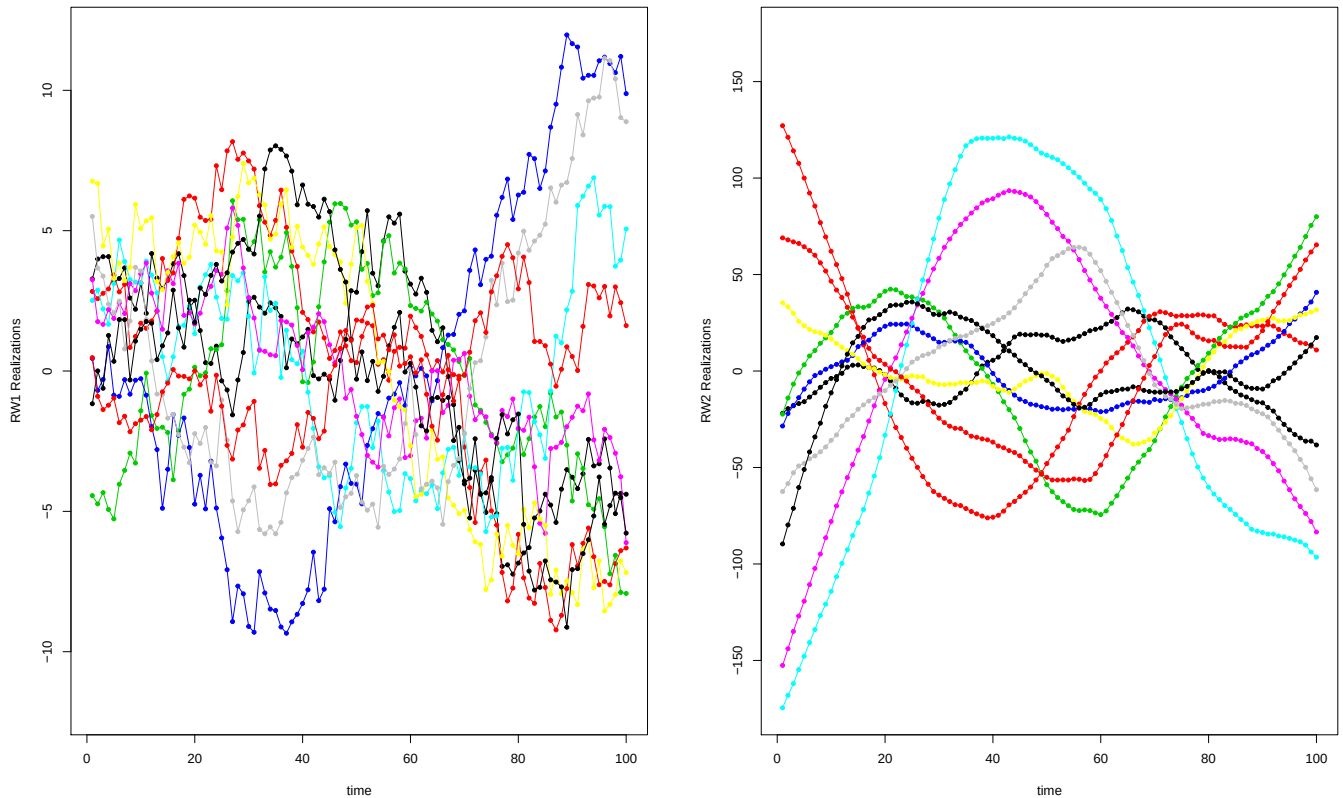

FIGURE 7: Ten simulated realizations from RW1 (left) and RW2 (right) models. In both cases $\sigma_u^2 = 1$.

RW1 Model Applied to Nile Data

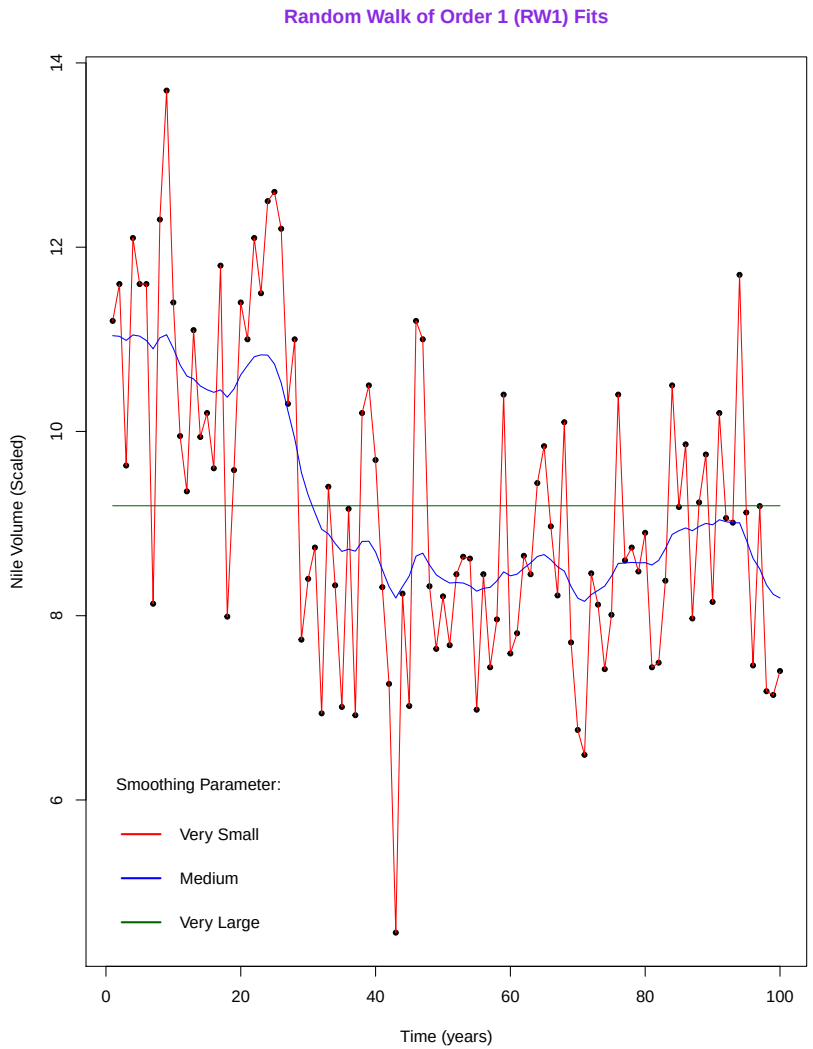

Figure 8: Nile data with RW1 fits under different priors for smoothing parameter $1/\sigma_u^2$.

RW2 Model Applied to Nile Data

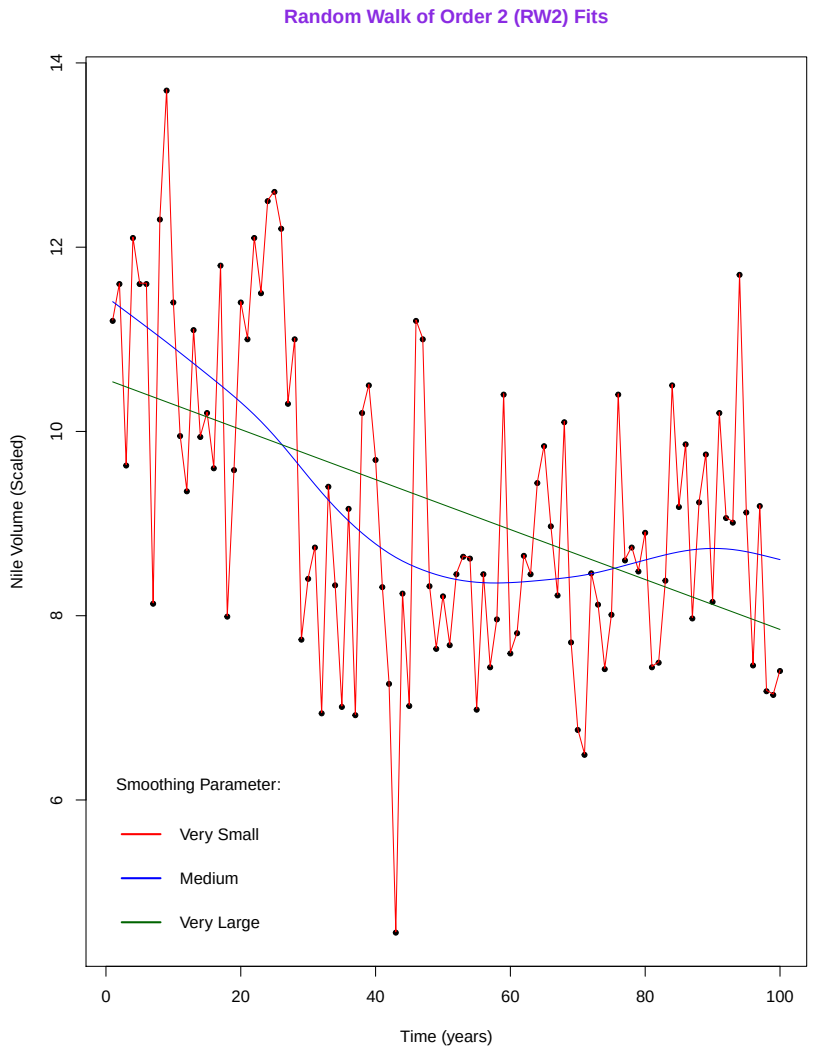

Figure 9: Nile data with RW2 fits under different priors for smoothing parameter $1/\sigma_u^2$.

Temporal Smoothing Model Summary

We have three models:

IID MODEL:

$$u_t \mid \sigma_u^2 \sim \mathsf{N}(0, \sigma_u^2),$$

smooth towards zero.

RW1 MODEL:

$$u_t - u_{t-1} \mid \sigma_u^2 \sim \mathsf{N}(0, \sigma_u^2),$$

smooth towards the previous value.

RW2 MODEL:

$$(u_t - u_{t-1}) - (u_{t-1} - u_{t-2}) \mid \sigma_u^2 \sim \mathsf{N}(0, \sigma_u^2),$$

smooth towards the previous slope.

The random walks (RWs) are examples of Gaussian Markov Random Field (GMRF) models, which have many appealing properties, and are computationally convenient (Rue and Held, 2005).

RW Fitting to Simulated Data

- The model is:

$$\begin{aligned} Y_t \mid p_t &\sim \text{Binomial}(n_t, p_t) \\ \frac{p_t}{1-p_t} &= \exp(\alpha + u_t) \\ [u_1, \ldots, u_T] \mid \sigma_u^2 &\sim \text{RW2}(\sigma_u^2) \\ \sigma_u^2 &\sim \text{Prior on Smoothing Parameter} \\ \alpha &\sim \text{Prior on Intercept} \end{aligned}$$

- Note that we have a linear model on the logistic scale, i.e., we could write

$$\text{logit}(p_t) = \log\left(\frac{p_t}{1-p_t}\right) = \alpha + u_t.$$

- Fit using `R-INLA`.

- On Figures 10 and 11 the fitted values are shown in red – in both examples the reconstruction is reasonable.

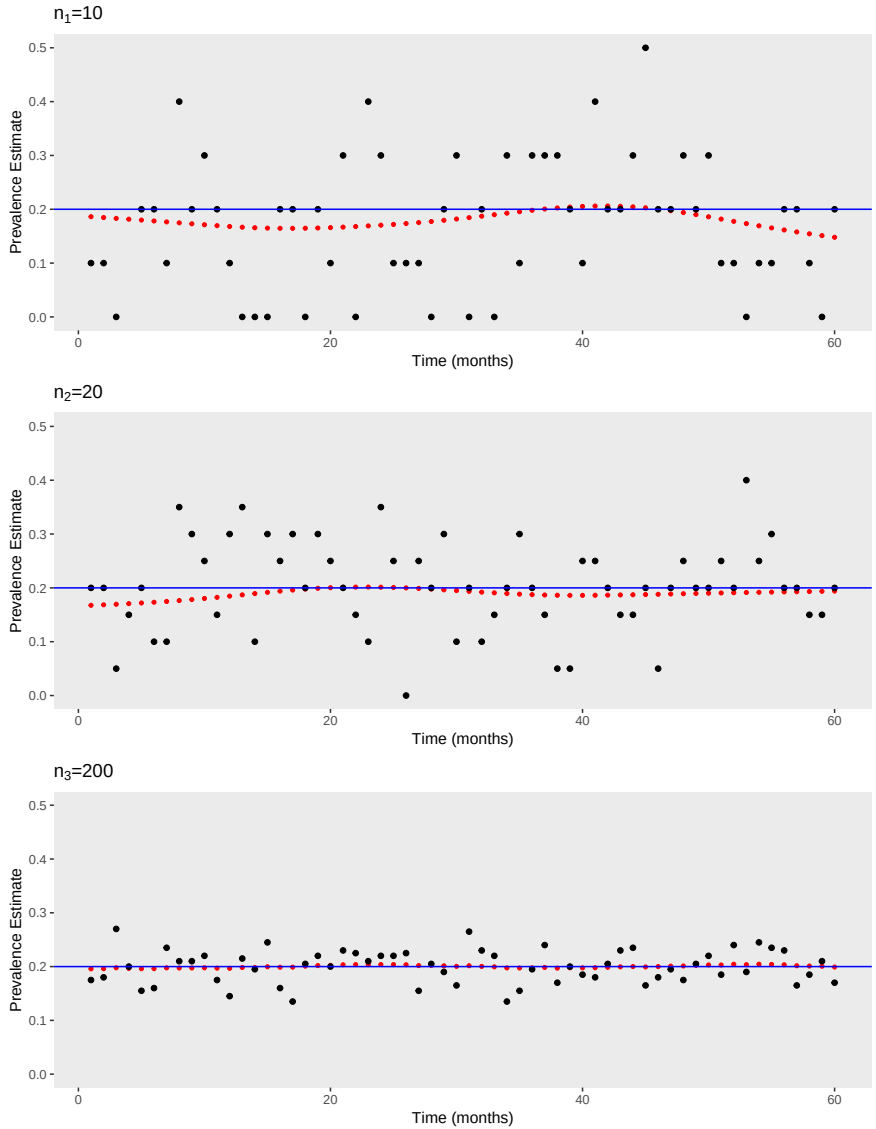

FIGURE 10: Prevalence estimates over time from simulated data, true prevalence $p = 0.2$ (blue solid lines). Smoothed random walk estimates in **red**.

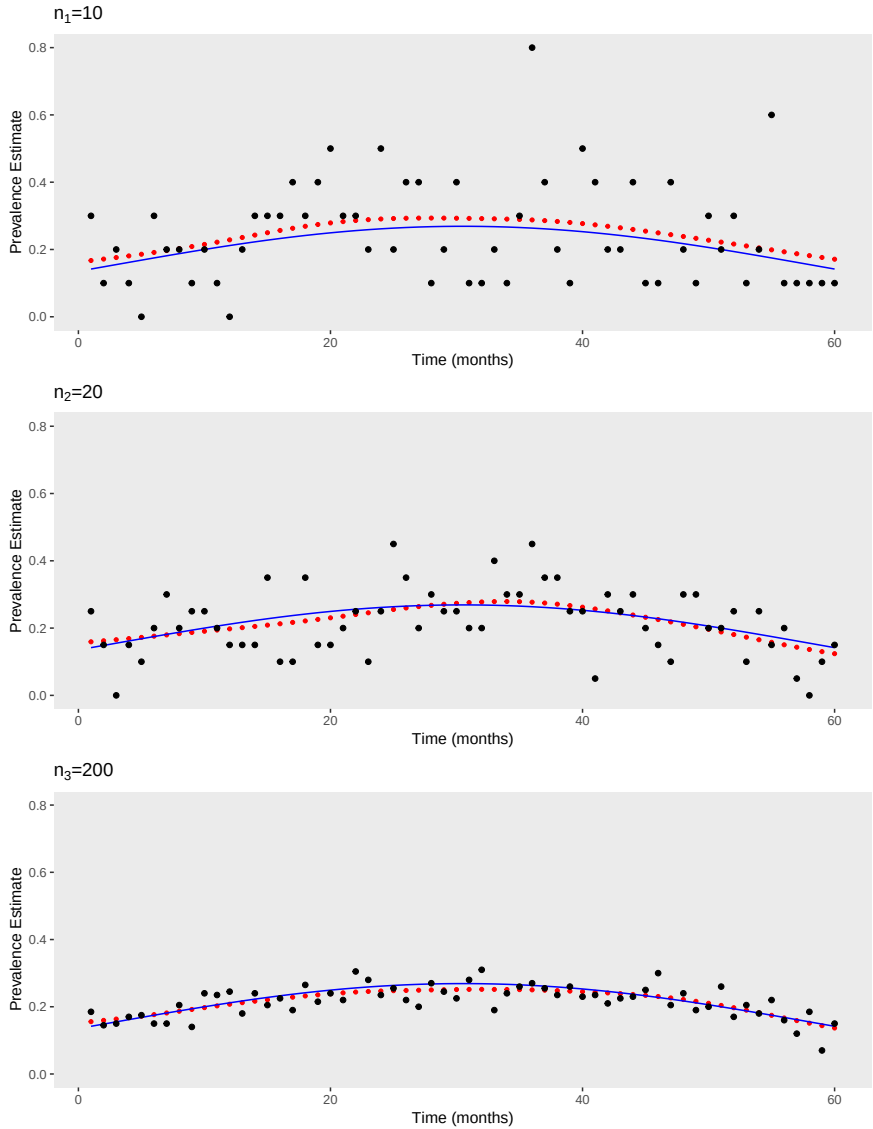

FIGURE 11: Prevalence estimates over time from simulated data, true prevalence corresponds to curved blue solid line. Smoothed random walk estimates in **red**.

RW2 Model Applied to Ecuador Survey Data

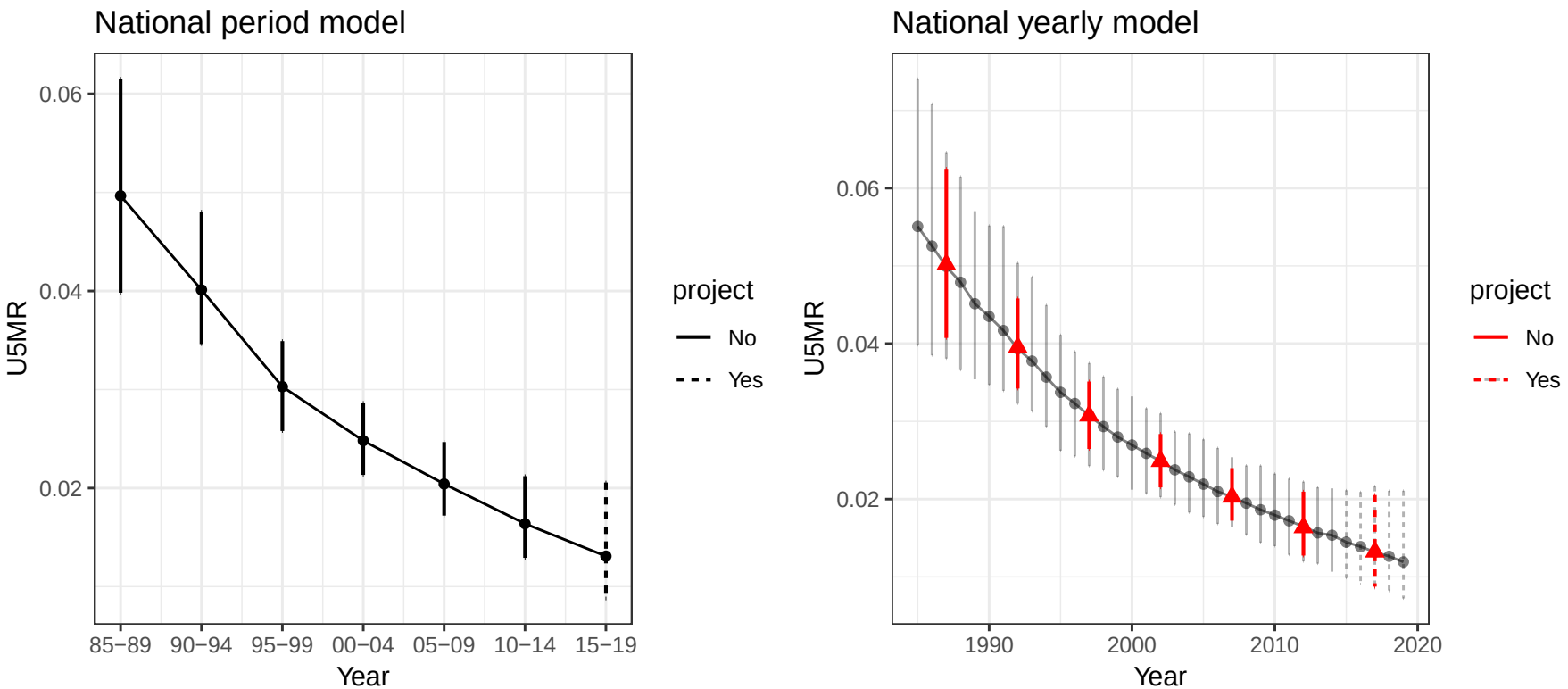

FIGURE 12: Yearly RW2 smoothing of weighted estimates of the under-5 mortality rate (U5MR) in Ecuador, with 95% uncertainty intervals. On the left we apply to data aggregated over 5 years and on the right by 1 year. The dashed lines on the right of each plot are projections.

Discussion

Discussion

- Bayes is a convenient way to incorporate temporal and spatial smoothing through prior distributions.
- Computation can be carried out in a number of ways – don't get hung up on this, as it's far more important to:
  - be able to specify a sampling model (the likelihood) and the penalization (prior), and
  - be able to interpret posterior summaries.
- We will make extensive use of the intrinsic conditional autoregressive (ICAR) model for spatial smoothing, which is the 2D extension of the RW1 model.

References

Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley and Sons.

Gómez-Rubio, V. (2020). *Bayesian Inference with INLA*. CRC Press.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press.

Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. CRC Press.

Osgood-Zimmerman, A. and Wakefield, J. (2020). Template Model Builder (TMB), a flexible alternative to the Integrated Nested Laplace Approximation. *https://arxiv.org/abs/2103.09929*.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Application*. Chapman and Hall/CRC Press, Boca Raton.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested

Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.

Wang, X., Yue, Y., and Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC.