The Penalizing Model Complexity Prior framework - PC-priors

INLA Team

November, 2019

Outline

Changing the hyperprior in the R-package INLA

Which prior should we choose?

A new concept: Penalised complexity priors

Changing the prior: Internal scale

- Hyperparameters are represented internally with more well-behaved transformations, e.g. correlation ρ and precision τ are internally represented as

$$\theta_1 = \log(\tau)$$
$$\theta_2 = \log\left(\frac{1+\rho}{1-\rho}\right)$$

- The prior must be set on the parameter in **internal scale**
- Initial values for the mode-search must be set in **internal scale**

Changing the prior: Code

```
hyper = list(prec = list(prior = "loggamma",
                         param = c(1, 0.1),
                         initial = 4,
                         fixed = FALSE))

formula = y ~ f(idx, model = "iid", hyper = hyper) + ...
```

```
# For the iid model, default options can be seen with
inla.doc("iid")
```

Changing the prior: Available models

Some of the available choices, see also `www.r-inla.org`,

- `"pc.prec"` ← NEW
- `"gaussian"`
- `"loggamma"`
- `"flat"`
- `"logtgaussian"`

You can get information about the paramterisation of the `loggamma` prior, say, using

```
1 inla.doc("loggamma")
```

Changing the prior: Available models

Some of the available choices, see also `www.r-inla.org`,

- `"pc.prec"` ← NEW
- `"gaussian"`
- `"loggamma"`
- `"flat"`
- `"logtgaussian"`

You can get information about the paramterisation of the `loggamma` prior, say, using

```
1 inla.doc("loggamma")
```

It is also possible to use your own prior (on internal scale) with

- `"expression:"`: `R` expression that calculates log-prior
- `"table:"`: Tabulated values that are interpolated

How to assign your own prior (table)

It is possible to provide a table of suitable support values x (internal scale) and the corresponding log-density values y as string:

$$\text{“}table:\ \ x\_1\ \ldots\ \ x\_n\ y\_1\ \ldots\ \ y\_n\text{”}$$

```
# use suitable support points x
lprec = seq(-10, 10, len=100)

# link the x and corresponding y values into a string which
    begins with "table:""
prior.table = INLA:::inla.paste(c("table:",cbind(lprec,
    prior.function(lprec))))
```

This is consequently assigned as

```
hyper = list(prec = list(prior = prior.table))
```

About the choice of prior distributions

The issue of setting prior distributions on model parameters is a difficult issue in applied Bayesian statistics, in particular for parameters further down the model hierarchy, such as precision or correlation parameters.

About the choice of prior distributions

The issue of setting prior distributions on model parameters is a difficult issue in applied Bayesian statistics, in particular for parameters further down the model hierarchy, such as precision or correlation parameters.

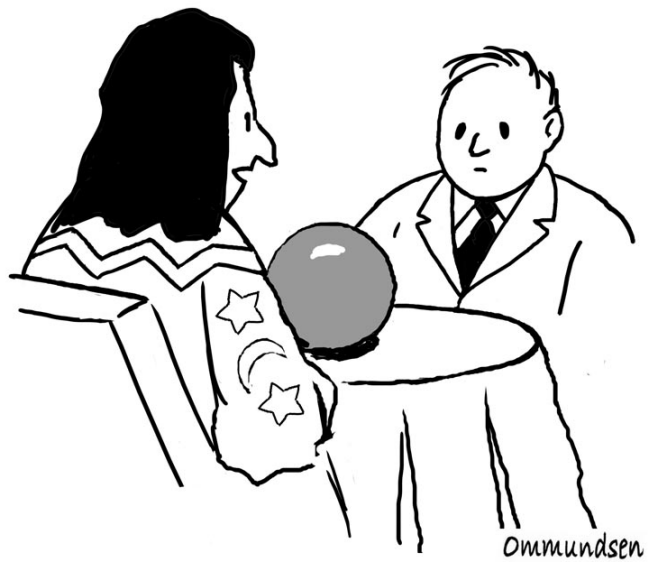
What is the current practice?

About the choice of prior distributions

The issue of setting prior distributions on model parameters is a difficult issue in applied Bayesian statistics, in particular for parameters further down the model hierarchy, such as precision or correlation parameters.

What is the current practice?

- Choose priors based on computational convenience.
- Choose priors used in the literature and hope to avoid criticism.
- Ignore the problem and hope that the data will dominate the prior.



"Is this needed for a Bayesian analysis?"

About prior choices

Simpson, Rue, Martins, Riebler and Sørbye (2015)

"Prior selection is the fundamental issue in Bayesian statistics. Priors are the Bayesian's greatest tool, but they are also the greatest point for criticism: the arbitrariness of prior selection procedures and the lack of realistic sensitivity analysis are a serious argument against current Bayesian practice."

Reference:
Simpson, D. P., Rue, H., Martins, T. G., Riebler, A. and Sørbye, S. H. (2015). Penalising model component complexity: A principled practical approach to constructing priors. arXiv:1403.4630.

Assignment of hyperpriors

The scaling problem of model components, where the structure matrix has not full rank, i.e intrinsic GMRFs (IGMRFs)

- Models for splines (`rw1`, `rw2`)

Assignment of hyperpriors

The scaling problem of model components, where the structure matrix has not full rank, i.e intrinsic GMRFs (IGMRFs)

- Models for splines (`rw1`, `rw2`)
- The Besag-model for area/regional models (`besag`)

Assignment of hyperpriors

The scaling problem of model components, where the structure matrix has not full rank, i.e intrinsic GMRFs (IGMRFs)

- Models for splines (`rw1`, `rw2`)
- The Besag-model for area/regional models (`besag`)
- and others...

Assignment of hyperpriors

The scaling problem of model components, where the structure matrix has not full rank, i.e intrinsic GMRFs (IGMRFs)

- Models for splines (`rw1`, `rw2`)
- The Besag-model for area/regional models (`besag`)
- and others...

Assignment of hyperpriors

The scaling problem of model components, where the structure matrix has not full rank, i.e intrinsic GMRFs (IGMRFs)

- Models for splines (`rw1`, `rw2`)
- The Besag-model for area/regional models (`besag`)
- and others...

Problem:

- These models are unscaled and their properties change with locations/dimension/graph.

Sørbye and Rue, 2014, Spat Stat

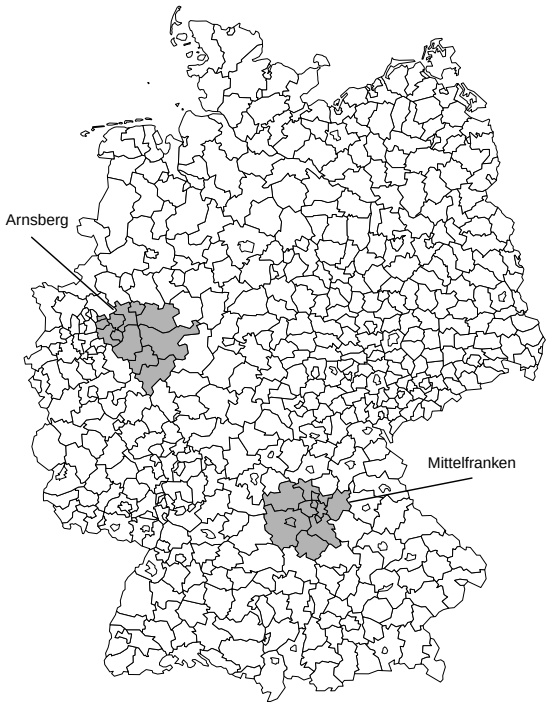
Illustration RW1: Marginal variance

Consider the generalised variance

$$\sigma^2_{\text{GV}} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{\tau_u}\left[\boldsymbol{R}^{-1}\right]_{ii}\right)\right) = \frac{1}{\tau}\exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(\left[\boldsymbol{R}^{-1}\right]_{ii})\right)$$

```
> rw1(5)
[1,]  1 -1  .  .  .
[2,] -1  2 -1  .  .
[3,]  . -1  2 -1  .
[4,]  .  . -1  2 -1
[5,]  .  .  . -1  1
> geom.mean(diag(ginv(rw1(5))))
[1] 0.73
> geom.mean(diag(ginv(rw1(50))))
[1] 7.55
> geom.mean(diag(ginv(rw1(500))))
[1] 75.580
```

Illustration: Besag model



We consider two administrative regions:

- Mittelfranken
- Arnsberg

Both regions have 12 districts.

Can we use the same hyperpriors?

What is the issue?

[Sørbye and Rue (2014), Spat Stat]

The marginal variances $\tau_u^{-1}[\boldsymbol{R}^-]_{ii}$ depend on the graph structure!
Consider the generalised variance

$$\sigma^2_{\text{GV}} = \frac{1}{\tau_u} \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\left(\left[\boldsymbol{R}^-\right]_{ii}\right)\right)$$

What is the issue?

[Sørbye and Rue (2014), Spat Stat]

The marginal variances $\tau_u^{-1}[\boldsymbol{R}^-]_{ii}$ depend on the graph structure! Consider the generalised variance

$$\sigma^2_{\text{GV}} = \frac{1}{\tau_u} \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\left(\left[\boldsymbol{R}^-\right]_{ii}\right)\right)$$

If $\tau_u = 1$ we get

- $\sigma^2_{\text{GV}} \approx 0.40$ for Arnsberg
- $\sigma^2_{\text{GV}} \approx 0.29$ for Mittelfranken
- $\sigma^2_{\text{GV}} \approx 0.56$ for whole Germany

What is the issue?

[Sørbye and Rue (2014), Spat Stat]

The marginal variances $\tau_u^{-1}[\boldsymbol{R}^-]_{ii}$ depend on the graph structure! Consider the generalised variance

$$\sigma^2_{\text{GV}} = \frac{1}{\tau_u} \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\left(\left[\boldsymbol{R}^-\right]_{ii}\right)\right)$$

If $\tau_u = 1$ we get

- $\sigma^2_{\text{GV}} \approx 0.40$ for Arnsberg
- $\sigma^2_{\text{GV}} \approx 0.29$ for Mittelfranken
- $\sigma^2_{\text{GV}} \approx 0.56$ for whole Germany

A prior for τ_u will not mean the same thing for every problem!

How do we scale?

In order to unify the interpretation of the prior for τ_u and make it transferable, we need to scale the model so that

$$\sigma^2_{\text{GV}} = \frac{1}{\tau_u}$$

Now, τ_u is the precision of the (marginal) deviation from a constant level, independently from the underlying graph.

How do we scale?

In order to unify the interpretation of the prior for τ_u and make it transferable, we need to scale the model so that

$$\sigma^2_{\text{GV}} = \frac{1}{\tau_u}$$

Now, τ_u is the precision of the (marginal) deviation from a constant level, independently from the underlying graph.

Caution: This scaling issues applies to all IGMRFs!

```
1   formula = f(.,model="..", hyper=..., scale.model=T)
```

The R-package `INLA` contains a function to scale any singular precision matrix `R`

```
inla.scale.model(R, ...)
```

How to choose our parameters?

- Assume $\tau \sim$ Gamma(a, b) where $\mathrm{E}(\tau) = a/b$.
- We can say something about the scale of the effect with

$$\sigma = \sqrt{1/\tau}$$

For example:

$$\mathrm{Prob}(\sigma > U) = \alpha$$

From this we can derive parameter b, if we fix a value for a, say.

Sørbye and Rue, 2014, Spat Stat; Papoila et al., 2014, Biom J

- This isn't enough: Why are we using a Gamma distribution, why not half-Cauchy . . . ?

Penalised complexity (PC) priors

Simpson et al. (2015) introduced a new concept of defining priors that are robust, invariant to reparameterisations and principle based.

Main idea: Occam's razor—a principle of parsimony

Simpler model formulations should be preferred until there is enough support for a more complex model.

Our background: R-INLA

Building models adding up model components

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(...;\boldsymbol{\theta}_1) + f_2(...;\boldsymbol{\theta}_2) + \cdots$$

- Many model components represent a flexible extension of a base model.

Our background: R-INLA

Building models adding up model components

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(...;\boldsymbol{\theta}_1) + f_2(...;\boldsymbol{\theta}_2) + \cdots$$

- Many model components represent a flexible extension of a base model.
- Put a prior on the *distance* between the flexible model and the base model.

Our background: R-INLA

Building models adding up model components

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(...;\boldsymbol{\theta}_1) + f_2(...;\boldsymbol{\theta}_2) + \cdots$$

- Many model components represent a flexible extension of a base model.
- Put a prior on the *distance* between the flexible model and the base model.
- Important: Mode should be at a distance equal to zero.

Our background: R-INLA

Building models adding up model components

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(...;\boldsymbol{\theta}_1) + f_2(...;\boldsymbol{\theta}_2) + \cdots$$

- Many model components represent a flexible extension of a base model.
- Put a prior on the *distance* between the flexible model and the base model.
- Important: Mode should be at a distance equal to zero.
- Transform the prior back to the parameter of interest.

1. Principle: Occam's razor

- Many model components represent a flexible extension of a base model. For each model component $\boldsymbol{x}$ we define a flexible model

$$f = \pi(\boldsymbol{x}|\xi)$$

  where ξ is interpreted as a flexibility parameter.
- f is a flexible version of a base model

$$g = \pi(\boldsymbol{x}|\xi = \xi_0)$$

Examples for base models

Case	Parameter	ξ	Base model
IID	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (no random effect)

Examples for base models

Case	Parameter	ξ	Base model
IID	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (no random effect)
Student-t	ν (dof)	$\xi = 1/\nu$	$\xi = 0$ (Gaussian)

Examples for base models

Case	Parameter	ξ	Base model
IID	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (no random effect)
Student-t	ν (dof)	$\xi = 1/\nu$	$\xi = 0$ (Gaussian)
IGMRF	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (constant, line, plane)
AR1	ρ (correlation)	$\xi = \rho$	$\xi = 0$ (no time-dependence)
			$\xi = 1$ (no change in time)
Correlation matrix	$\boldsymbol{Q}$	$\boldsymbol{\xi} = \boldsymbol{Q}$	$\boldsymbol{\xi} = \boldsymbol{I}$ (no correlation)

Examples for base models

Case	Parameter	ξ	Base model
IID	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (no random effect)
Student-t	ν (dof)	$\xi = 1/\nu$	$\xi = 0$ (Gaussian)
IGMRF	τ (precision)	$\xi = 1/\tau$	$\xi = 0$ (constant, line, plane)
AR1	ρ (correlation)	$\xi = \rho$	$\xi = 0$ (no time-dependence)
			$\xi = 1$ (no change in time)
Correlation matrix	$\boldsymbol{Q}$	$\boldsymbol{\xi} = \boldsymbol{Q}$	$\boldsymbol{\xi} = \boldsymbol{I}$ (no correlation)

Side comment: In a BYM model we would have nested base models:

Base model $= 0 \rightarrow$ iid $\rightarrow$ dependence $=$ more flexible model

1. Occams razor

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ

1. Occams razor

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

1. Occams razor

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

1. Occams razor

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause overfitting (force complexity) if, loosely,

$$\pi_\xi(\xi = 0) = 0$$

2. Principle: Measure of complexity

Use Kullback-Leibler discrepancy to measure the increased complexity introduced by $\xi > 0$,

$$\mathrm{KLD}(f\|g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

for flexible model f and base model g.

2. Principle: Measure of complexity

Use Kullback-Leibler discrepancy to measure the increased complexity introduced by $\xi > 0$,

$$\mathrm{KLD}(f\|g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

for flexible model f and base model g.

Example

Assume that the flexible model f is a $\mathcal{N}(\mu; 1)$ where $\mu > 0$. The base model g refers to $\mu = 0$. Then

$$\mathrm{KLD}(f\|g) = \frac{\mu^2}{2}$$

3. Principle: Constant-rate penalisation

Main idea

Assign priors to "distances" between models, instead of assigning priors to the parameters.

- Define the (uni-directional) "distance"

$$d(\xi) = \sqrt{2\,\mathrm{KLD}(\xi)}$$

3. Principle: Constant-rate penalisation

Main idea

Assign priors to "distances" between models, instead of assigning priors to the parameters.

- Define the (uni-directional) "distance"

$$d(\xi) = \sqrt{2\,\mathrm{KLD}(\xi)}$$

- Assign an exponential distribution to $d(\xi)$:

$$\pi(d(\xi)) = \lambda \exp\left(-\lambda d(\xi)\right), \qquad \lambda > 0$$

which has mode at $d(\xi) = 0$.

3. Principle: Constant-rate penalisation

Main idea

Assign priors to "distances" between models, instead of assigning priors to the parameters.

- Define the (uni-directional) "distance"

$$d(\xi) = \sqrt{2\,\mathrm{KLD}(\xi)}$$

- Assign an exponential distribution to $d(\xi)$:

$$\pi(d(\xi)) = \lambda \exp\left(-\lambda d(\xi)\right), \qquad \lambda > 0$$

  which has mode at $d(\xi) = 0$.

- Do the change-of-variables to get a prior for the parameter of interest.

4. Principle: User-defined scaling

- Determine λ based on some knowledge of the model component, for example in terms of prior mass in the tail.

4. Principle: User-defined scaling

- Determine λ based on some knowledge of the model component, for example in terms of prior mass in the tail.
- A natural criterion for IGMRFs is

$$P(\sigma > U) = P\left(\frac{1}{\sqrt{\tau}} > U\right) = \alpha$$

where U is an upper limit for the standard deviation and α is a small probability.

4. Principle: User-defined scaling

- Determine λ based on some knowledge of the model component, for example in terms of prior mass in the tail.
- A natural criterion for IGMRFs is

$$P(\sigma > U) = P\left(\frac{1}{\sqrt{\tau}} > U\right) = \alpha$$

  where U is an upper limit for the standard deviation and α is a small probability.
- The scale U determines the magnitude of the effect of a model component and how informative the prior will be.

Example: Precision of a Gaussian

Analytic result in this case (type-2 Gumbel):

$$\pi(\tau) = \frac{\theta}{2}\tau^{-3/2}\exp\left(-\theta/\sqrt{\tau}\right), \qquad \mathsf{E}(\tau) = \infty,$$

where $\mathsf{Prob}(\sigma > U) = \alpha$ gives

$$\theta = -\frac{\ln(\alpha)}{u}$$

Example: Precision of a Gaussian

Analytic result in this case (type-2 Gumbel):

$$\pi(\tau) = \frac{\theta}{2}\tau^{-3/2}\exp\left(-\theta/\sqrt{\tau}\right), \qquad \mathsf{E}(\tau) = \infty,$$

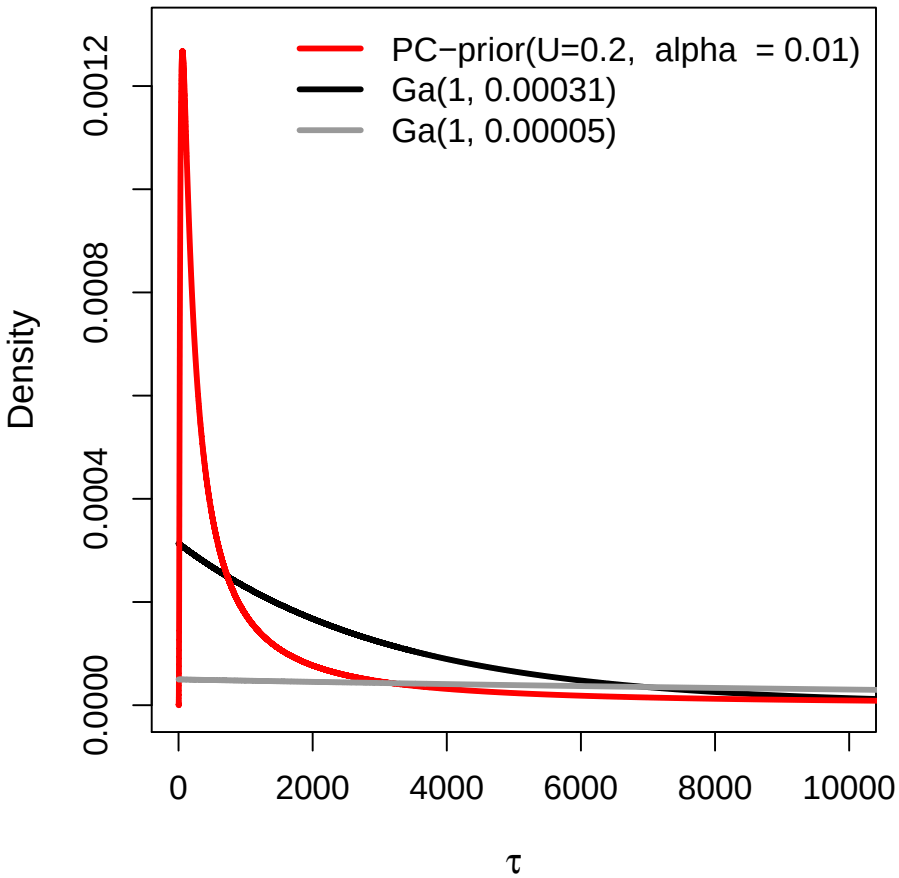
where $\mathsf{Prob}(\sigma > U) = \alpha$ gives

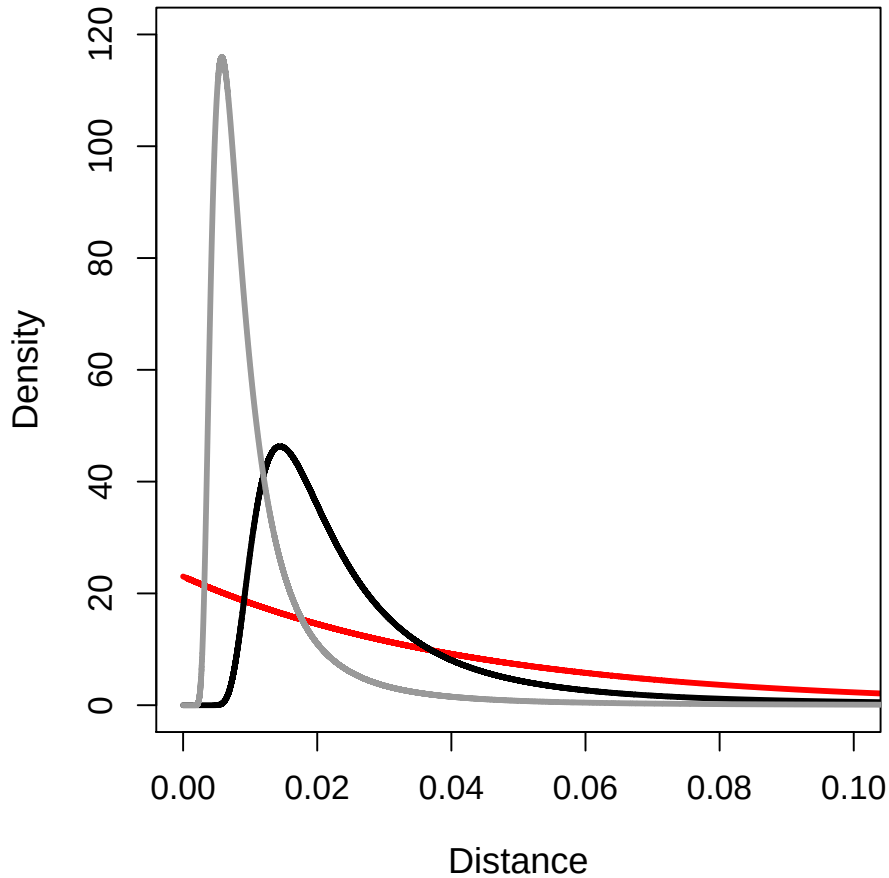
$$\theta = -\frac{\ln(\alpha)}{u}$$

Alternative interpretation

$$\pi(\sigma) = \lambda \exp(-\lambda\sigma)$$

Comparison to a gamma prior





The parameter λ can be chosen based on $P(1/\sqrt{\tau_b} > U) = \alpha$
$\Rightarrow \lambda = -\log(\alpha)/U$.

How to do this in INLA

Specifying the pc-prior in the f-function:

```
hyper = list(precision =
    list(prior = "pc.prec",
    param = c(u, alpha)))
```

Documentation:

```
inla.doc("pc.prec")
```

Disease mapping

Assume

$$Y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i))$$

where the log relative risk is decomposed into

$$\eta_i = \mu + u_i + v_i + f(c_i)$$

- μ is the overall level (intercept).
- $v_i \sim \mathcal{N}(0, \tau_v^{-1})$ represent non-spatial overdispersion.
- u_i are random effects with spatial structure.
- $f(c_i)$ denotes a non-linear covariate effect.

The spatially structured effect

To incorporate a spatial structure into a model, the so called Besag model is often used.

$$p(\boldsymbol{u} \mid \kappa_u) \propto \kappa_u^{(n-1)/2} \exp\left(-\frac{\kappa_u}{2} \sum_{i \sim j} (u_i - u_j)^2\right)$$
$$= \kappa_u^{(n-1)/2} \exp\left(-\frac{\kappa_u}{2} \boldsymbol{u}^T \mathbf{R} \boldsymbol{u}\right).$$

where R is called structure matrix and defined as

$$R_{ij} = \begin{cases} n_i & i = j \\ -1 & i \sim j \\ 0 & \text{otherwise.} \end{cases}$$

Here, $i \sim j$ denotes that i and j are neighbouring regions.
We are in a Bayesian framework, so how do we choose priors?

Choice of prior distributions

In this model, we have three precision parameters for which we need a prior distribution. Let us look at the precision parameter τ_u of the spatially structured effect.

- The Besag model is an intrinsic Gaussian Markov random field (IGMRF), i.e. the precision matrix has not full rank
- The model penalises local deviation from a constant level.
- The hyperprior will control this local deviation.

Disease mapping (II)

Next challenge: How should we think about the model?

Disease mapping (II)

Next challenge: How should we think about the model?

Base model $= 0 \rightarrow$ iid $\rightarrow$ dependence $=$ more flexible model

Disease mapping (II)

Next challenge: How should we think about the model?

Base model $= 0 \rightarrow$ iid $\rightarrow$ dependence $=$ more flexible model

Rewrite the model as

$$\eta = \mu + f(c_i) + \frac{1}{\sqrt{\tau_b}}\left(\sqrt{1-\gamma}\cdot v_i + \sqrt{\gamma}\cdot {u_i}^*\right)$$

where $\cdot^*$ is a unit-variance standardised model.

- Marginal precision τ_b.

Disease mapping (II)

Next challenge: How should we think about the model?

Base model $= 0 \rightarrow$ iid $\rightarrow$ dependence $=$ more flexible model

Rewrite the model as

$$\eta = \mu + f(c_i) + \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1-\gamma} \cdot v_i + \sqrt{\gamma} \cdot {u_i}^* \right)$$

where $\cdot^*$ is a unit-variance standardised model.

- Marginal precision τ_b.
- γ gives it interpretation:
  independence ($\gamma = 0$), maximal spatial dependence ($\gamma = 1$)

Disease mapping (II)

Next challenge: How should we think about the model?

Base model $= 0 \rightarrow$ iid $\rightarrow$ dependence $=$ more flexible model

Rewrite the model as

$$\eta = \mu + f(c_i) + \frac{1}{\sqrt{\tau_b}}\left(\sqrt{1-\gamma}\cdot v_i + \sqrt{\gamma}\cdot {u_i}^*\right)$$

where $\cdot^*$ is a unit-variance standardised model.

- Marginal precision τ_b.
- γ gives it interpretation:
  independence ($\gamma = 0$), maximal spatial dependence ($\gamma = 1$)
- Parameters control different features.

PC prior for the precision parameter

To construct the prior for τ_b we use:

- Base model: No spatial effect, i.e $\boldsymbol{\Sigma}_{\text{base}}(\tau_b) = \mathbf{0}$ $(\tau_b \to \infty)$
- Flexible model: $\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1-\gamma} v_i + \sqrt{\gamma} u_i^\star \right) \Rightarrow$

$$\boldsymbol{\Sigma}_{\text{flex}}(\tau_b) = \tau_b^{-1} \left((1-\gamma) \boldsymbol{I} + \gamma \boldsymbol{R}_\star^- \right)$$

so that

$$\text{KLD}(\tau_b) = \text{KLD}(\mathcal{N}(0, \boldsymbol{\Sigma}_{\text{flex}}(\tau_b)) \mid \mathcal{N}(0, \boldsymbol{\Sigma}_{\text{base}})).$$

PC prior for the precision parameter

- Applying all principles we get a type-2 Gumbel distribution:

$$\pi(\tau_b) = \frac{\lambda}{2}\tau_b^{-3/2}\exp\left(-\lambda/\sqrt{\tau_b}\right)$$

which has a very heavy tail (no expectation!)

Note: This prior also corresponds to an exponential prior for the standard deviation.

PC prior for the precision parameter

- Applying all principles we get a type-2 Gumbel distribution:

$$\pi(\tau_b) = \frac{\lambda}{2}\tau_b^{-3/2}\exp\left(-\lambda/\sqrt{\tau_b}\right)$$

  which has a very heavy tail (no expectation!)

  Note: This prior also corresponds to an exponential prior for the standard deviation.

- It can be shown that any prior with finite expectation shrinks towards a complex model. $\Rightarrow$ enforces complexity!

PC prior for the mixing parameter

The PC-prior is thereby derived analogously as for the precision, where we use:

- Base model ($\gamma = 0$): $\mathbf{v}$ $\boldsymbol{\Sigma}_{\text{base}}(\gamma) = \mathbf{I}$

PC prior for the mixing parameter

The PC-prior is thereby derived analogously as for the precision, where we use:

- Base model ($\gamma = 0$): $\mathbf{v}$ $\boldsymbol{\Sigma}_{\text{base}}(\gamma) = \mathbf{I}$
- Flexible model: $\sqrt{1-\gamma} \cdot \mathbf{v} + \sqrt{\gamma} \cdot \mathbf{u}^*$
$$\Rightarrow \boldsymbol{\Sigma}_{\text{flex}}(\gamma) = (1-\gamma) \cdot \mathbf{I} + \gamma \mathbf{R}_\star^-$$

PC prior for the mixing parameter

The PC-prior is thereby derived analogously as for the precision, where we use:

- Base model ($\gamma = 0$): $\mathbf{v}$ $\boldsymbol{\Sigma}_{\text{base}}(\gamma) = \mathbf{I}$
- Flexible model: $\sqrt{1-\gamma} \cdot \mathbf{v} + \sqrt{\gamma} \cdot \mathbf{u}^*$
$$\Rightarrow \boldsymbol{\Sigma}_{\text{flex}}(\gamma) = (1-\gamma) \cdot \mathbf{I} + \gamma \mathbf{R}_\star^-$$
- The resulting prior is not a standard distribution.

PC prior for the mixing parameter

The PC-prior is thereby derived analogously as for the precision, where we use:

- Base model ($\gamma = 0$): $\mathbf{v}$ $\boldsymbol{\Sigma}_{\text{base}}(\gamma) = \mathbf{I}$
- Flexible model: $\sqrt{1-\gamma} \cdot \mathbf{v} + \sqrt{\gamma} \cdot \mathbf{u}^*$
  $\Rightarrow \boldsymbol{\Sigma}_{\text{flex}}(\gamma) = (1-\gamma) \cdot \mathbf{I} + \gamma \mathbf{R}_\star^-$
- The resulting prior is not a standard distribution.
- It can be defined using $\Pr(\gamma < U) = \alpha$.

Germany-example: The model is able to learn

We also have a covariate, which is conjectured to have a non-linear effect $\Rightarrow$ We use an independent PC prior on a spline precision.

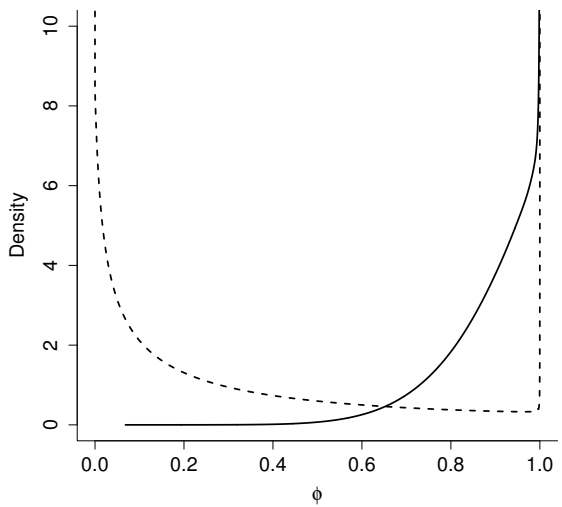
- $U = 0.2/0.31, \alpha = 0.01$ for the precision parameters
- $U = 1/2, \alpha = 2/3$ for the mixing weight γ.

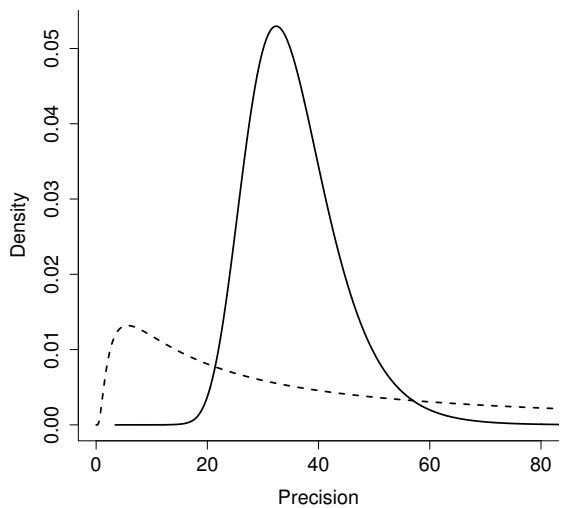
Germany-example: The model is able to learn

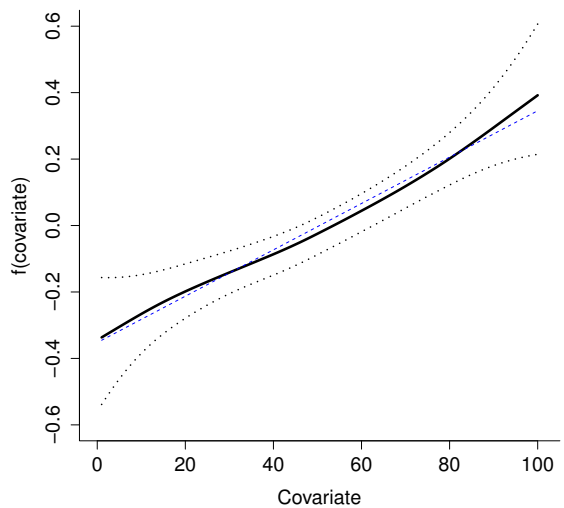
We also have a covariate, which is conjectured to have a non-linear effect $\Rightarrow$ We use an independent PC prior on a spline precision.

- $U = 0.2/0.31, \alpha = 0.01$ for the precision parameters
- $U = 1/2, \alpha = 2/3$ for the mixing weight γ.

Results:







Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising.

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising.
- Easy and natural interpretation as a well defined shrinkage to a base-model

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising.
- Easy and natural interpretation as a well defined shrinkage to a base-model
- We can chose the degree of "informativeness"

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising.
- Easy and natural interpretation as a well defined shrinkage to a base-model
- We can chose the degree of "informativeness"
- A lot of work to integrate these ideas into `R-INLA`

Ongoing work

- Work out PC-priors for more models and implement these as new defaults in INLA.
- Reparameterize models to make them interpretable:

```
1 inla.doc("bym2")
```

- Develop a novel system for assigning priors, in which the user only has to choose one overall scaling parameter!

References

- D. P. Simpson, H. Rue, T. G. Martins, A. Riebler, and S. H. Sørbye (2015) *Penalising model component complexity: A principled, practical approach to constructing priors.* arxiv:1403.4630
- T. G. Martins and H. Rue. *Prior for flexibility parameters: the Student's t case.* Technical report S8-2013, Department of mathematical sciences, NTNU, Norway, 2013.
- S. H. Sørbye, and H. Rue (2014) *Scaling intrinsic Gaussian Markov random field priors in spatial modelling*, Spatial Statistics.

Thank you for your attention!

If you have any doubts or questions, please write us:

`elias@r-inla.org`

`help@r-inla.org`