Penalising model component complexity: A principled practical approach to constructing priors

Håvard Rue
Department of Mathematical Sciences
NTNU, Norway

March 27, 2014

Joint work with

Thiago Martins

Daniel Simpson

Andrea Riebler

Sigrunn Sørbye

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) "issue" with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any "reasonable choice" will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) "issue" with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any "reasonable choice" will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) “issue” with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any “reasonable choice” will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) “issue” with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any “reasonable choice” will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) "issue" with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any "reasonable choice" will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

Bayes theorem and priors

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

To be honest, Bayesian statistics, has an (practical) "issue" with priors:

- Do not really know what to do priors in practise
- Hope/assume that data will dominate the prior, so that any "reasonable choice" will do fine
- Objective/reference priors are hard to compute and often not available, but we may not want to use them in any case
- Hierarchical models make it all more difficult
- There are exceptions, so it is not uniformly bad!

What is a prior?

Wikipedia:

> *In Bayesian statistical inference, a prior probability distribution, often called simply the prior, of an uncertain quantity p is the probability distribution that would express one's uncertainty about p before the "data" is taken into account.*

How do I chose my prior?

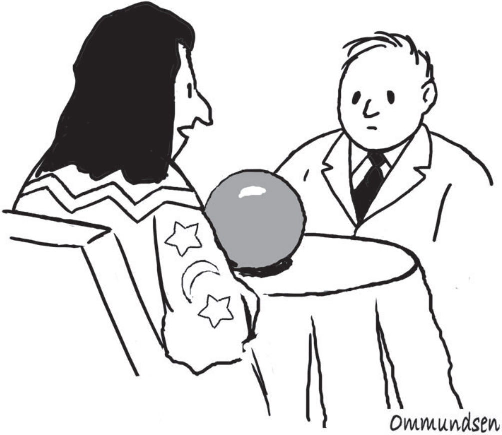
- **How?** Use probability-densities to express your uncertainty.
- **But how?** Use probability-densities to express your uncertainty.
- **Please just tell me what to do?** Use probability-densities to express your uncertainty.

How do I chose my prior?

- **How?** Use probability-densities to express your uncertainty.
- **But how?** Use probability-densities to express your uncertainty.
- **Please just tell me what to do?** Use probability-densities to express your uncertainty.

How do I chose my prior?

- **How?** Use probability-densities to express your uncertainty.
- **But how?** Use probability-densities to express your uncertainty.
- **Please just tell me what to do?** Use probability-densities to express your uncertainty.



"Is this needed for a Bayesian analysis?"

Yes, it is needed...

There are a lot of good reasons for why "we" need to move forward with the "prior"-issue:

- *(standard arguments goes here)*
- Marginal likelihood (easy in `R-INLA`)
- Prevent overfitting

Yes, it is needed...

There are a lot of good reasons for why "we" need to move forward with the "prior"-issue:

- *(standard arguments goes here)*
- Marginal likelihood (easy in `R-INLA`)
- Prevent overfitting

Yes, it is needed...

There are a lot of good reasons for why "we" need to move forward with the "prior"-issue:

- *(standard arguments goes here)*
- Marginal likelihood (easy in `R-INLA`)
- Prevent overfitting

Our background: R-INLA (www.r-inla.org)

- Building models adding up model components

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

- Also likelihoods have hyper-parameters
- We (as developers) can leave the responsibility to the user and require ALL priors to be specified by the user
- Does not solve the fundamental problem, nor does it make the world a better place to be
- Would be nice and important to come up with "good" default priors (up to a notion of *scale*) for most parameters

Our background: R-INLA (www.r-inla.org)

- Building models adding up model components

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

- Also likelihoods have hyper-parameters
- We (as developers) can leave the responsibility to the user and require ALL priors to be specified by the user
- Does not solve the fundamental problem, nor does it make the world a better place to be
- Would be nice and important to come up with "good" default priors (up to a notion of *scale*) for most parameters

Our background: R-INLA (www.r-inla.org)

- Building models adding up model components

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

- Also likelihoods have hyper-parameters
- We (as developers) can leave the responsibility to the user and require ALL priors to be specified by the user
- Does not solve the fundamental problem, nor does it make the world a better place to be
- Would be nice and important to come up with "good" default priors (up to a notion of *scale*) for most parameters

Our background: R-INLA (www.r-inla.org)

- Building models adding up model components

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

- Also likelihoods have hyper-parameters
- We (as developers) can leave the responsibility to the user and require ALL priors to be specified by the user
- Does not solve the fundamental problem, nor does it make the world a better place to be
- Would be nice and important to come up with "good" default priors (up to a notion of *scale*) for most parameters

Our background: R-INLA (www.r-inla.org)

- Building models adding up model components

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

- Also likelihoods have hyper-parameters
- We (as developers) can leave the responsibility to the user and require ALL priors to be specified by the user
- Does not solve the fundamental problem, nor does it make the world a better place to be
- Would be nice and important to come up with "good" default priors (up to a notion of *scale*) for most parameters

Sometimes, **we** are to be blamed!

Intrinsic Gaussian model (components)

- Popular in applications
- Often have a precision matrix of form

$$\boldsymbol{Q} = \tau \boldsymbol{R}$$

  where τ is the *precision parameter*
- $\boldsymbol{R}$ has not full rank but an interesting null-space

Sometimes, **we** are to be blamed!

Intrinsic Gaussian model (components)

- Popular in applications
- Often have a precision matrix of form

$$\boldsymbol{Q} = \tau \boldsymbol{R}$$

  where τ is the *precision parameter*
- $\boldsymbol{R}$ has not full rank but an interesting null-space

Sometimes, **we** are to be blamed!

Intrinsic Gaussian model (components)

- Popular in applications
- Often have a precision matrix of form

$$\boldsymbol{Q} = \tau \boldsymbol{R}$$

  where τ is the *precision parameter*
- $\boldsymbol{R}$ has not full rank but an interesting null-space

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Examples of intrinsic models

- Models for splines (`rw1`, `rw2`)
- Thin-plate splines (dimension > 1, `rw2d`)
- The "CAR" model/Besag-model for area/regional models (`besag`)
- and others...

Problem:

- The "problem" is that these models are *unscaled* and change with locations/dimension/graph.
- Setting prior for the precision parameter τ is a mess...

Example2

rw1-model

$$x'(t) = \text{noise}(t)$$

Null-space

$$\mathbf{1}$$

rw2-model

$$x''(t) = \text{noise}(t)$$

Null-space

$$\mathbf{1},\ \boldsymbol{t}$$

Example2

rw1-model

$$x'(t) = \text{noise}(t)$$

Null-space

$$\mathbf{1}$$

rw2-model

$$x''(t) = \text{noise}(t)$$

Null-space

$$\mathbf{1},\ \boldsymbol{t}$$

Example (dimension)

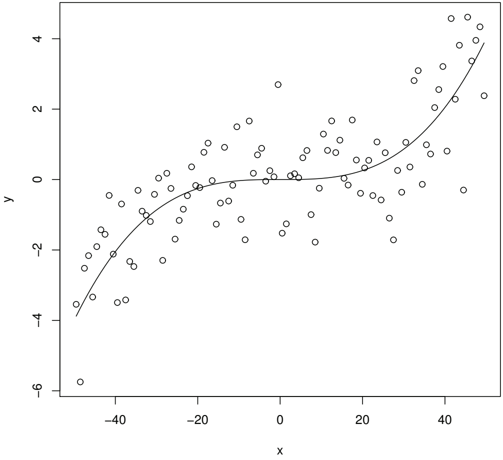
```
> inla.rw(5)
5 x 5 sparse Matrix of class "dgTMatrix"

[1,]  1 -1  .  .  .
[2,] -1  2 -1  .  .
[3,]  . -1  2 -1  .
[4,]  .  . -1  2 -1
[5,]  .  .  . -1  1

> mean(diag(inla.ginv(inla.rw(5, sparse=FALSE), rankdef=1)))
[1] 0.8
> mean(diag(inla.ginv(inla.rw(50, sparse=FALSE), rankdef=1)))
[1] 8.33
> mean(diag(inla.ginv(inla.rw(500, sparse=FALSE), rankdef=1)))
[1] 83.333
```

Example (order)

```
> mean(diag(inla.ginv(inla.rw(100, order = 1, sparse=FALSE),
                      rankdef=1)))
[1] 16.665
> mean(diag(inla.ginv(inla.rw(100, order = 2, sparse=FALSE),
                      rankdef=2)))
[1] 2381.19
```

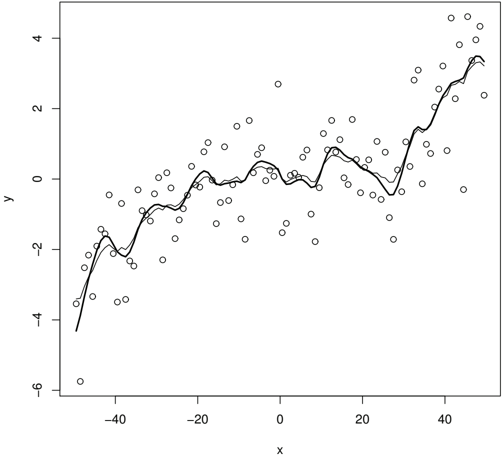
Example: Smoothing

Data



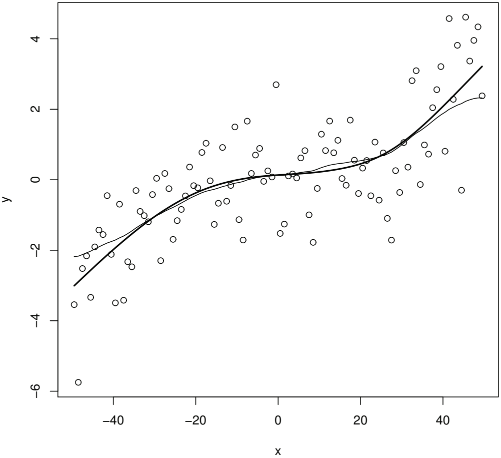
Example: Smoothing

Unscaled (fixed precision)



Example: Smoothing

Scaled (same fixed precision)



How to scale?

- Scale so that $\sigma_*^2 = 1$, where (f.ex)

$$\sigma_*^2 = \exp\left(\text{mean}\left(\mathit{log}\left(\text{diag}\left(\boldsymbol{R}^-\right)\right)\right)\right)$$

- If we know the null-space of $\boldsymbol{R}$ we can compute diag($\boldsymbol{R}^-$) using sparse matrix algebra.
- In `R-INLA`

```
f(..., scale.model=TRUE)                        ## case-spesific
inla.setOption(scale.model.default = TRUE) ## global
```

How to scale?

- Scale so that $\sigma_*^2 = 1$, where (f.ex)

$$\sigma_*^2 = \exp\left(\text{mean}\left(log\left(\text{diag}\left(\boldsymbol{R}^-\right)\right)\right)\right)$$

- If we know the null-space of $\boldsymbol{R}$ we can compute diag($\boldsymbol{R}^-$) using sparse matrix algebra.
- In R-INLA

```
f(..., scale.model=TRUE)                    ## case-spesific
inla.setOption(scale.model.default = TRUE) ## global
```

How to scale?

- Scale so that $\sigma_*^2 = 1$, where (f.ex)

$$\sigma_*^2 = \exp\left(\text{mean}\left(log\left(\text{diag}\left(\boldsymbol{R}^-\right)\right)\right)\right)$$

- If we know the null-space of $\boldsymbol{R}$ we can compute diag$(\boldsymbol{R}^-)$ using sparse matrix algebra.
- In R-INLA

```
f(..., scale.model=TRUE)                        ## case-spesific
inla.setOption(scale.model.default = TRUE) ## global
```

Choosing prior parameters

- Assume

$$\tau \sim \text{Gamma}(a, b)$$

  where $\text{E}(\tau) = a/b$.

- Can say something about the **scale** of the effect (family dependent): with

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

- Can answer a question like

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

- Need one more criteria/question...

Choosing prior parameters

- Assume
$$\tau \sim \text{Gamma}(a, b)$$
where $\text{E}(\tau) = a/b$.
- Can say something about the **scale** of the effect (family dependent): with
$$\sigma = \sqrt{1/\tau}$$
- Can answer a question like
$$\text{Prob}(\sigma > U) = \alpha$$
- Need one more criteria/question...

Choosing prior parameters

- Assume
$$\tau \sim \text{Gamma}(a, b)$$
where $\text{E}(\tau) = a/b$.
- Can say something about the **scale** of the effect (family dependent): with
$$\sigma = \sqrt{1/\tau}$$
- Can answer a question like
$$\text{Prob}(\sigma > U) = \alpha$$
- Need one more criteria/question...

Choosing prior parameters

- Assume

$$\tau \sim \text{Gamma}(a, b)$$

  where $\text{E}(\tau) = a/b$.

- Can say something about the **scale** of the effect (family dependent): with

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

- Can answer a question like

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

- Need one more criteria/question...

But we want more...

Summary so far

- Must standardise model components
- Can have an opinion of the *scale* of the effect

Big questions:

- Can we 'say' something about the prior density itself?
- How to approach the issue of possible overfitting?

But we want more...

Summary so far

- Must standardise model components
- Can have an opinion of the *scale* of the effect

Big questions:

- Can we 'say' something about the prior density itself?
- How to approach the issue of possible overfitting?

But we want more...

Summary so far

- Must standardise model components
- Can have an opinion of the *scale* of the effect

Big questions:

- Can we 'say' something about the prior density itself?
- How to approach the issue of possible overfitting?

But we want more...

Summary so far

- Must standardise model components
- Can have an opinion of the *scale* of the effect

Big questions:

- Can we 'say' something about the prior density itself?
- How to approach the issue of possible overfitting?

But we want more...

Summary so far

- Must standardise model components
- Can have an opinion of the *scale* of the effect

Big questions:

- Can we 'say' something about the prior density itself?
- How to approach the issue of possible overfitting?

Not all parameters are easy to interpret

Martyn Plummer (the author of JAGS):[1]

> *However, nobody can express an informative prior in terms of the precision, ...*

Would be nice to think about priors without having to care about the parameterisation (*invariance*)

[1] http://martynplummer.wordpress.com/2012/09/02/stan/

Not all parameters are easy to interpret

Martyn Plummer (the author of JAGS):[1]

> *However, nobody can express an informative prior in terms of the precision, ...*

Would be nice to think about priors without having to care about the parameterisation (*invariance*)

[1] http://martynplummer.wordpress.com/2012/09/02/stan/

Penalised complexity priors (PC priors)

- A new approach
- Based on a few principles
- Parameterisation invariant+++

Penalised complexity priors (PC priors)

- A new approach
- Based on a few principles
- Parameterisation invariant+++

Penalised complexity priors (PC priors)

- A new approach
- Based on a few principles
- Parameterisation invariant+++

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
    - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
    - Student-t is nested within the Gaussian
    - Spline models are nested within its null-space: linear effect or constant effect
    - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
    - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

- Prefer simplicity over complexity
- Many model components are nested within a simpler model
  - $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$ is nested within $\tau = \infty$
  - Student-t is nested within the Gaussian
  - Spline models are nested within its null-space: linear effect or constant effect
  - AR(1) is nested within $\rho = 0$ (no dependence in time) or $\rho = 1$ (no changes in time).
  - and so on

Principle I: Occam's razor

Consider the case where the more flexible model

$$\pi(x|\xi), \qquad \xi \geq 0$$

is nested within a base model $\pi(x|\xi = 0)$.

- 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 if, loosly,

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

Principle I: Occam's razor

Consider the case where the more flexible model

$$\pi(x|\xi), \qquad \xi \geq 0$$

is nested within a base model $\pi(x|\xi = 0)$.

- 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 if, loosly,

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

Principle I: Occam's razor

Consider the case where the more flexible model

$$\pi(x|\xi), \qquad \xi \geq 0$$

is nested within a base model $\pi(x|\xi = 0)$.

- 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 if, loosly,

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

Principle I: Occam's razor

Consider the case where the more flexible model

$$\pi(x|\xi), \qquad \xi \geq 0$$

is nested within a base model $\pi(x|\xi = 0)$.

- 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 if, loosly,

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

Principle I: Occam's razor

Consider the case where the more flexible model

$$\pi(x|\xi), \qquad \xi \geq 0$$

is nested within a base model $\pi(x|\xi = 0)$.

- 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 if, loosly,

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

Principle II: Measure of complexity

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

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

for flexible model f and base model g.

Gives a measure of the information lost when the base model is used to approximate the more flexible models

Principle II: 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.

Gives a measure of the information lost when the base model is used to approximate the more flexible models

Principle III: Constant rate penalisation

Define

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

as the (uni-directional) "distance" from flexible-model to the base model. Need the square-root to get the dimension right (*meter* not *meter*^{2})

Constant rate penalisation:

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

with mode at $d = 0$

Invariance: OK

Principle III: Constant rate penalisation

Define

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

as the (uni-directional) "distance" from flexible-model to the base model. Need the square-root to get the dimension right (*meter* not *meter*^{2})

Constant rate penalisation:

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

with mode at $d = 0$

Invariance: OK

Principle III: Constant rate penalisation

Define

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

as the (uni-directional) "distance" from flexible-model to the base model. Need the square-root to get the dimension right (*meter* not *meter*^{2})

Constant rate penalisation:

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

with mode at $d = 0$

Invariance: OK

Principle IV: User-defined scaling

The rate λ is determined from knowledge of the *scale* or some interpretable transformation $Q(\xi)$ of ξ:

$$\Pr(Q(\xi) > U) = \alpha$$

Example I

- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu,1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

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

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example I

- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu,1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

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

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example I

- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu,1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

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

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example I

- Base model $\mathcal{N}(0, 1)$
- Flexible model $\mathcal{N}(\mu, 1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:
$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$
- Can determine λ from a question like
$$\text{Prob}(\mu > u) = \alpha$$

Example I

- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu,1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

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

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example II

- Base model: Binomial$(\text{size} = 1, \text{prob} = 1/2)$
- Flexible model: Binomial$(\text{size} = 1, \text{prob} = p)$
- This gives

$$d(p) = \sqrt{2p\log(2p) + 2(1-p)\log(2(1-p))}$$

and the PC prior:

$$\pi(p) = \lambda \exp\left(-\lambda d(p)\right) \frac{1}{d(p)} \log\left(\frac{p}{1-p}\right)$$

Example II

- Base model: Binomial$(\text{size} = 1, \text{prob} = 1/2)$
- Flexible model: Binomial$(\text{size} = 1, \text{prob} = p)$
- This gives

$$d(p) = \sqrt{2p\log(2p) + 2(1-p)\log(2(1-p))}$$

and the PC prior:

$$\pi(p) = \lambda \exp\left(-\lambda d(p)\right) \frac{1}{d(p)} \log\left(\frac{p}{1-p}\right)$$

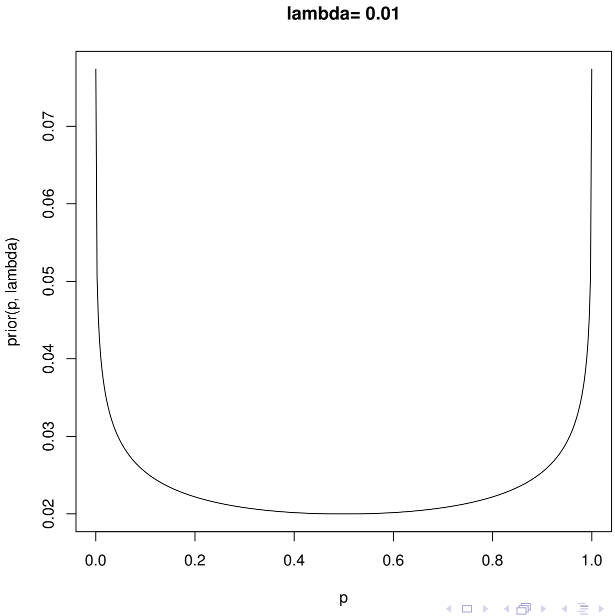
Example II

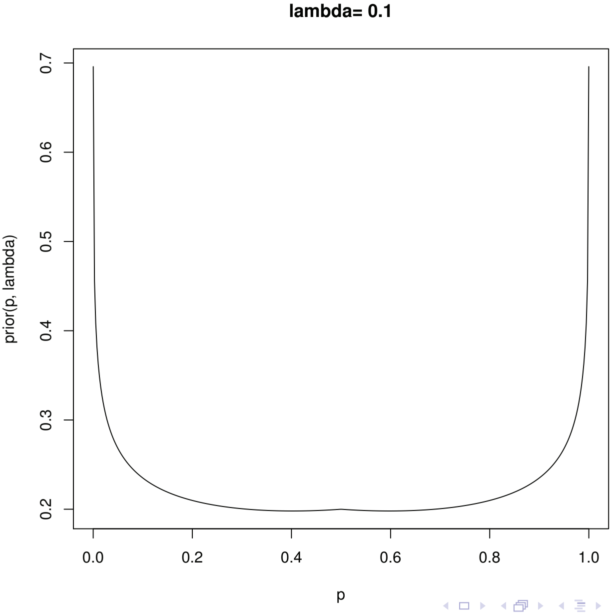
- Base model: Binomial$(\text{size} = 1, \text{prob} = 1/2)$
- Flexible model: Binomial$(\text{size} = 1, \text{prob} = p)$
- This gives

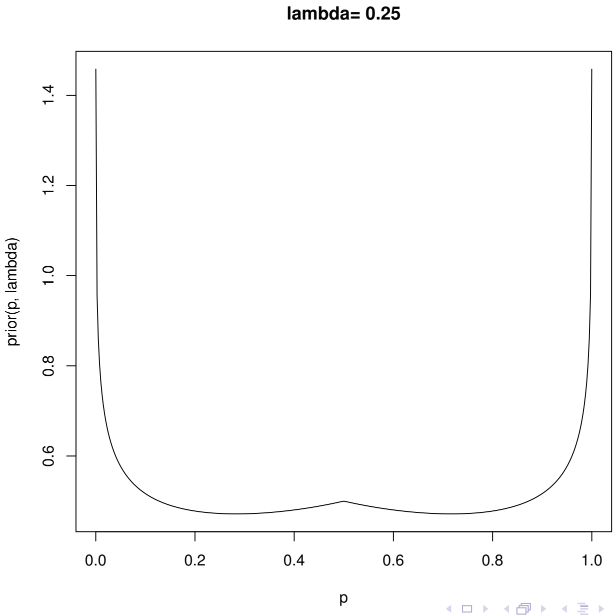
$$d(p) = \sqrt{2p\log(2p) + 2(1-p)\log(2(1-p))}$$

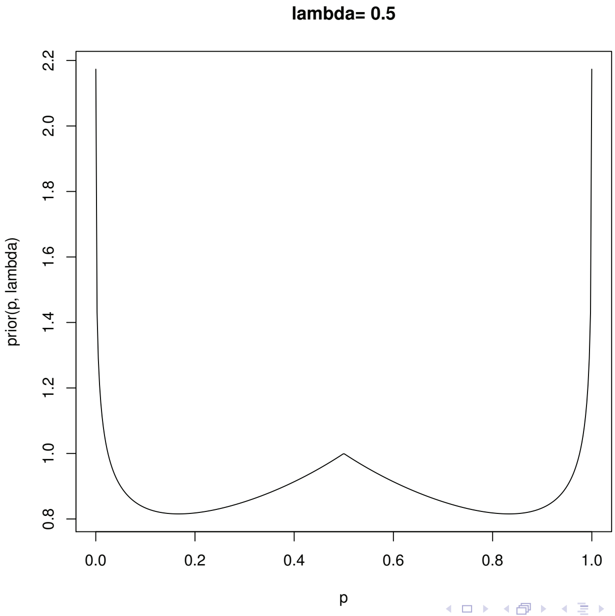
and the PC prior:

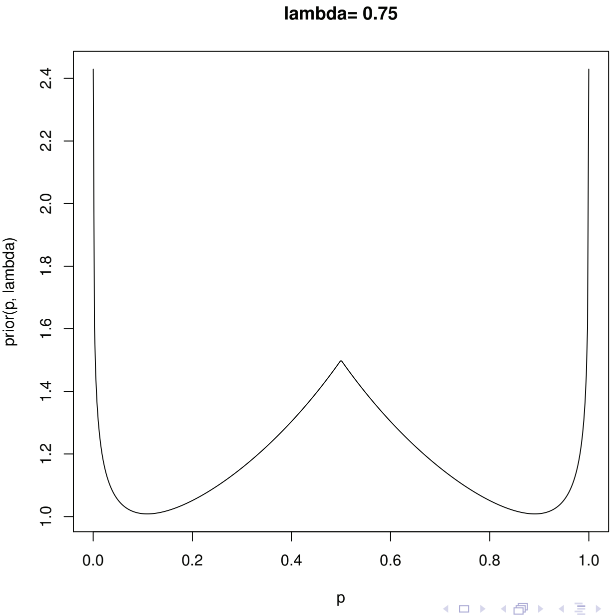
$$\pi(p) = \lambda \exp\left(-\lambda d(p)\right) \frac{1}{d(p)} \log\left(\frac{p}{1-p}\right)$$

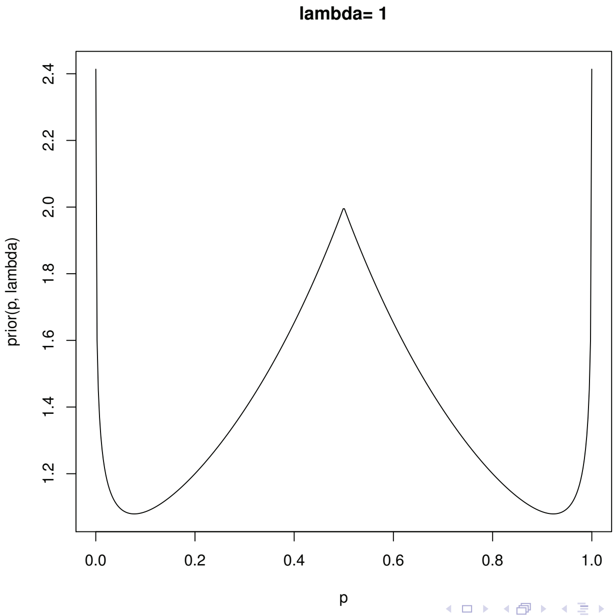

lambda= 0.01
prior(p, lambda)
0.02
0.03
0.04
0.05
0.06
0.07
0.0
0.2
0.4
0.6
0.8
1.0
p

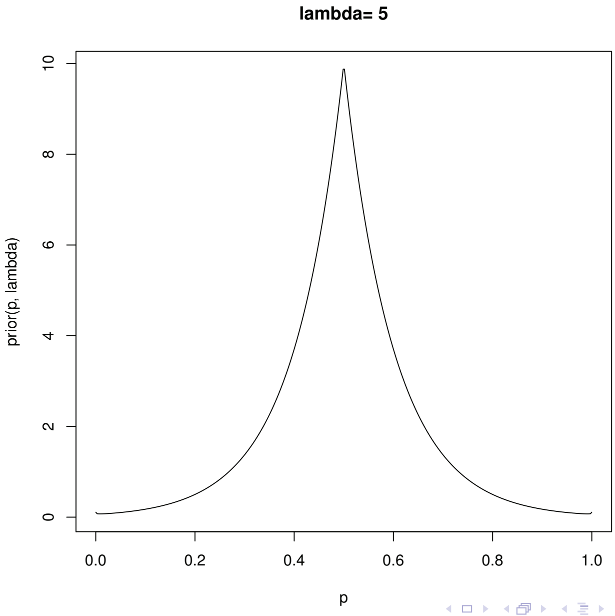

lambda= 0.1
prior(p, lambda)
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.2
0.4
0.6
0.8
1.0
p

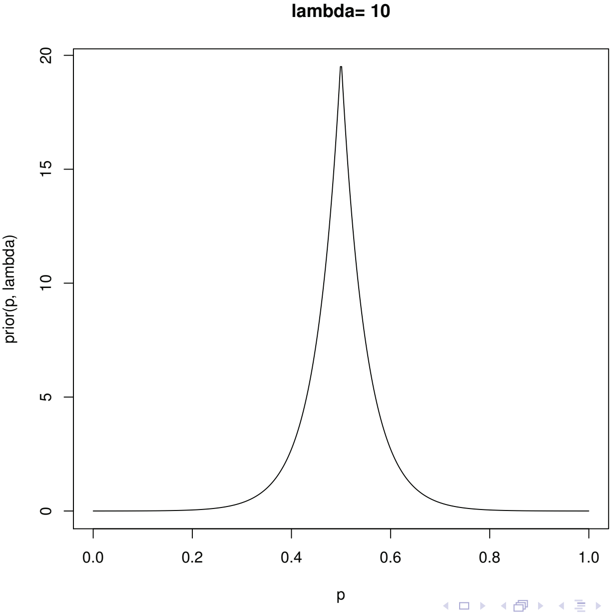

lambda= 0.25
prior(p, lambda)
0.6
0.8
1.0
1.2
1.4
0.0
0.2
0.4
0.6
0.8
1.0
p

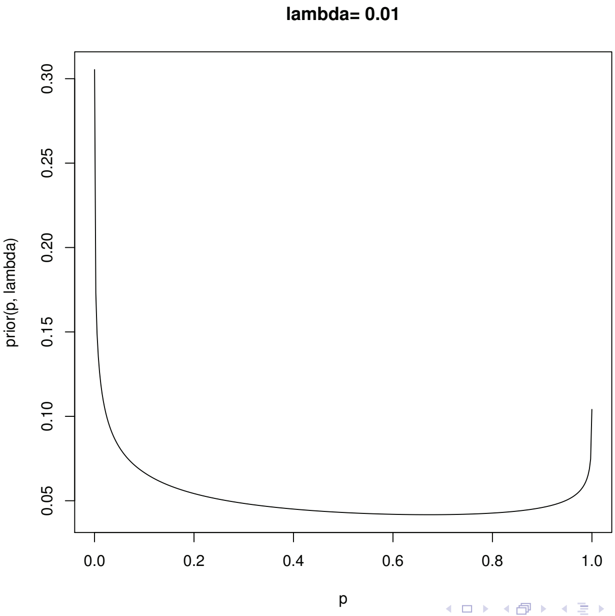

lambda= 0.5
prior(p, lambda)
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.0
0.2
0.4
0.6
0.8
1.0
p

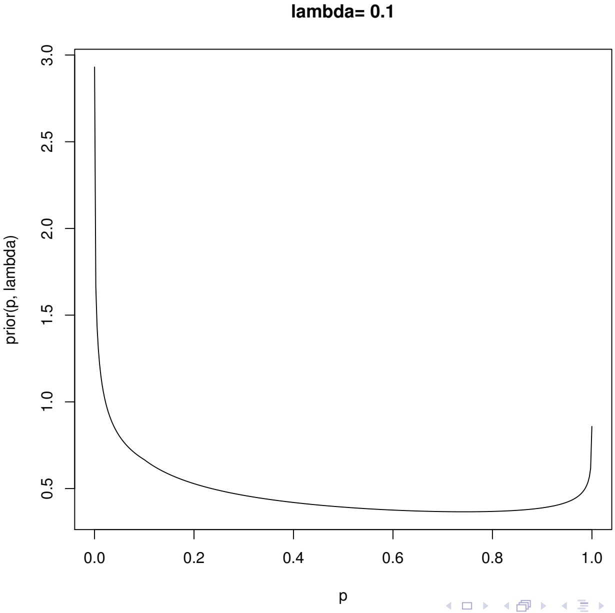

lambda= 0.75
prior(p, lambda)
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
0.0
0.2
0.4
0.6
0.8
1.0
p

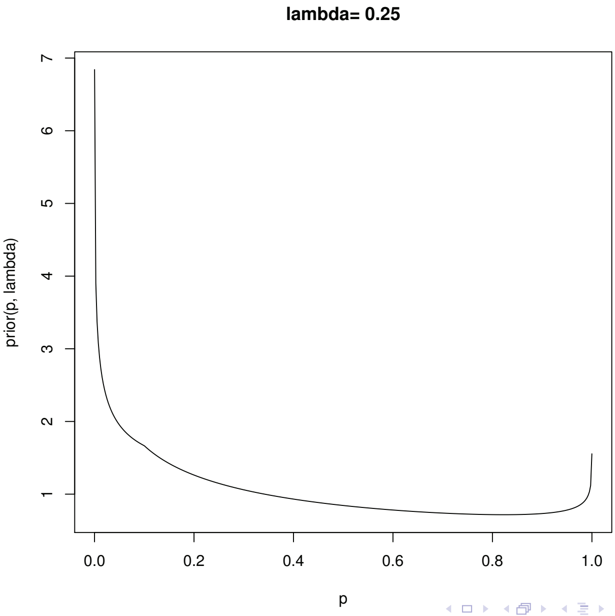

lambda= 1
prior(p, lambda)
2.4
2.2
2.0
1.8
1.6
1.4
1.2
0.0
0.2
0.4
0.6
0.8
1.0
p

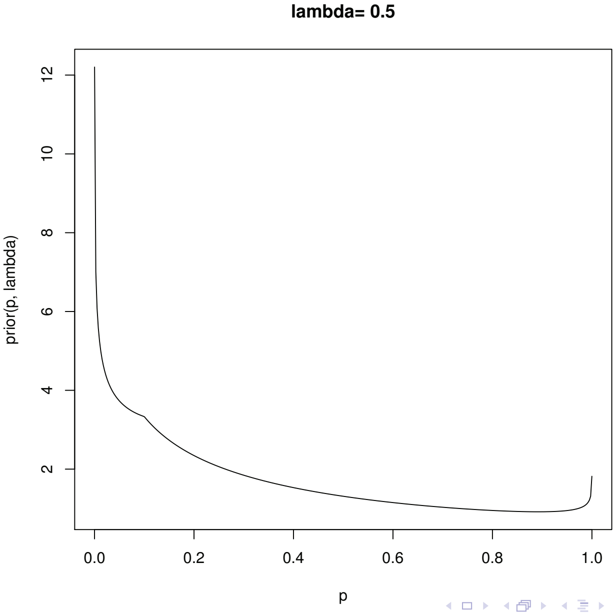

lambda= 5
prior(p, lambda)
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
p

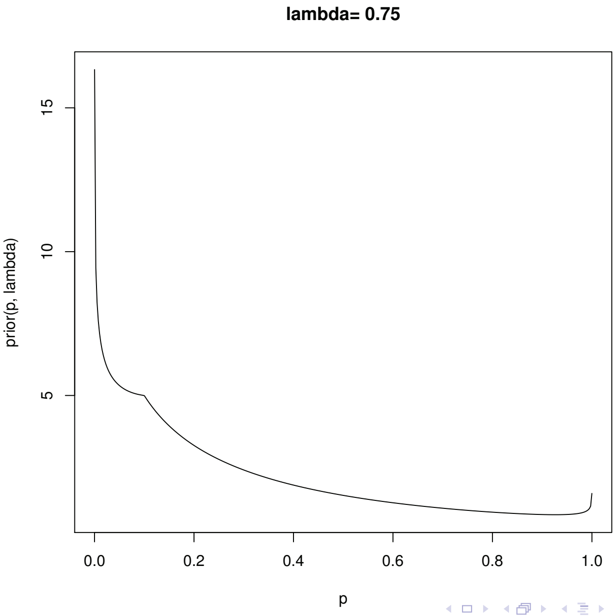

lambda= 10
prior(p, lambda)
p
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0

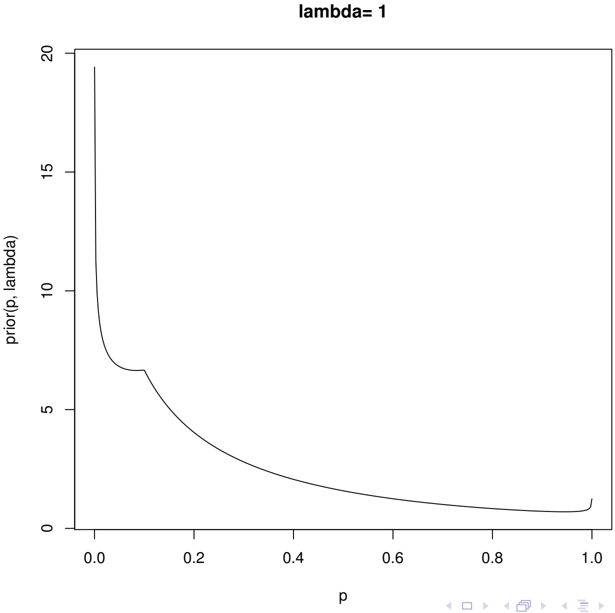

lambda= 0.01
prior(p, lambda)
0.05
0.10
0.15
0.20
0.25
0.30
0.0
0.2
0.4
0.6
0.8
1.0
p

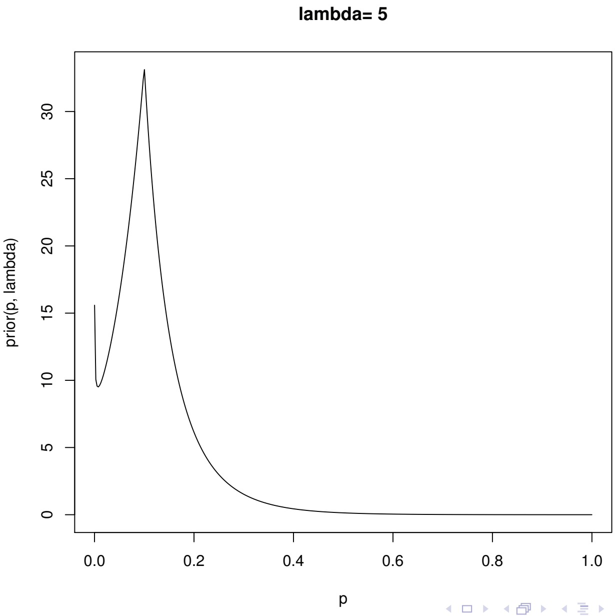

lambda= 0.1
prior(p, lambda)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
p

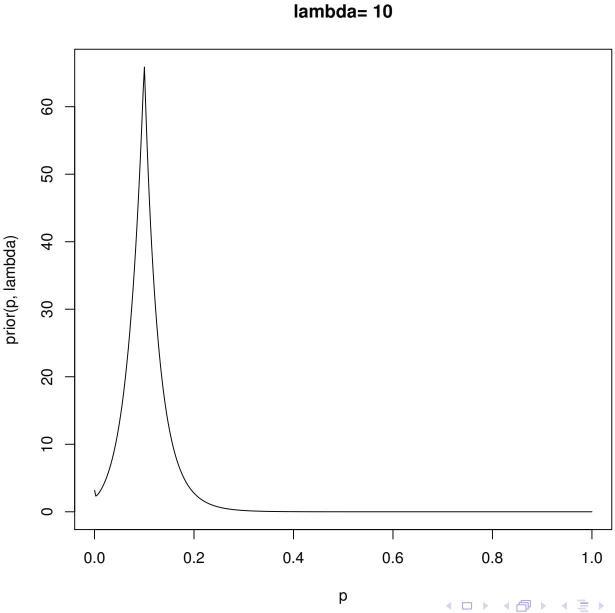

lambda= 0.25
prior(p, lambda)
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
p


lambda= 0.5
prior(p, lambda)
12
10
8
6
4
2
0.0
0.2
0.4
0.6
0.8
1.0
p


lambda= 0.75
prior(p, lambda)
15
10
5
0.0
0.2
0.4
0.6
0.8
1.0
p


lambda= 1
prior(p, lambda)
p
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0


lambda= 5
prior(p, lambda)
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
p


lambda= 10
prior(p, lambda)
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
p

Small $\lambda\xi > 0$: Tilted Jeffreys' prior

For small $\lambda\xi > 0$, we have that

$$\pi(\xi) = I(\xi)^{1/2} \exp\left(-\lambda m(\xi)\right) + \text{higher order terms}$$

where $I(\xi)$ is the Fisher information and

$$m(\xi) = \int_0^{\xi} \sqrt{I(s)}ds$$

is the distance defined by the metric tensor $I(\xi)$ on the Riemannian manifold.

Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.

Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.

Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.

A useful but negative result

Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

A useful but negative result

Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

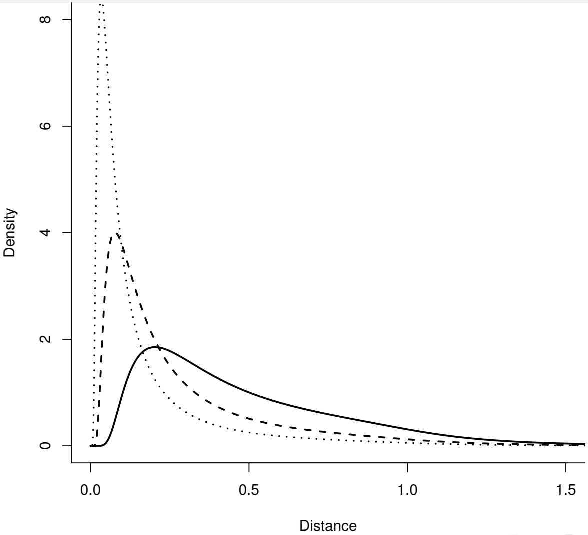
- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

A useful but negative result

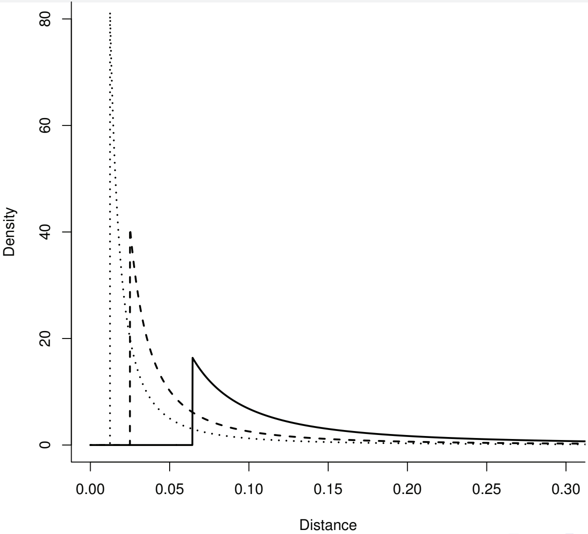
Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

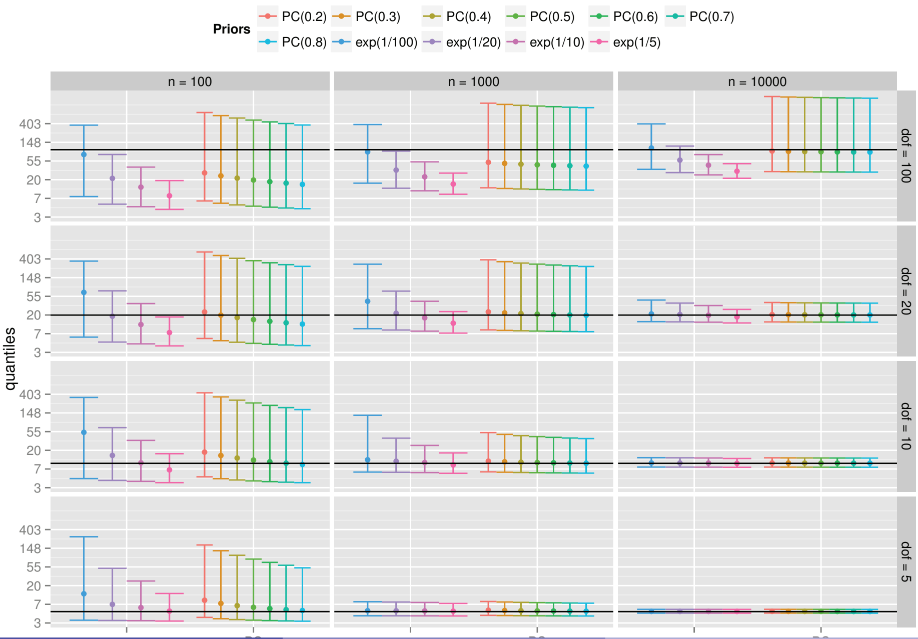
- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

The exp-prior with mean 5, 10, 20, converted to a prior for the distance



The uniform prior with upper$= 20, 50, 100$, converted to a prior for the distance




Priors
PC(0.2)
PC(0.3)
PC(0.4)
PC(0.5)
PC(0.6)
PC(0.7)
PC(0.8)
exp(1/100)
exp(1/20)
exp(1/10)
exp(1/5)
n = 100
n = 1000
n = 10000
dof = 100
dof = 20
dof = 10
dof = 5
quantiles
403
148
55
20
7
3

Experience with the PC prior

- Robust wrt prior settings and true value of ν
- Excellent learning properties!
- Behave like we want it to do!

The precision of a Gaussian

PC prior for the precision κ when $\kappa = \infty$ defines the base model

- "random effects"/iid-model
- The smoothing parameter in spline models
- etc...

Result Let $\pi_\kappa(\nu)$ be a prior for $\kappa > 0$ where $E(\kappa) < \infty$, then $\pi_d(0) = 0$ and the prior overfits.

The precision of a Gaussian

PC prior for the precision κ when $\kappa = \infty$ defines the base model

- "random effects"/iid-model
- The smoothing parameter in spline models
- etc...

Result Let $\pi_\kappa(\nu)$ be a prior for $\kappa > 0$ where $E(\kappa) < \infty$, then $\pi_d(0) = 0$ and the prior overfits.

The precision case (II)

Analytic result in this case (type-2 Gumbel)

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

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

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

Alternative interpretation

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

The precision case (II)

Analytic result in this case (type-2 Gumbel)

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

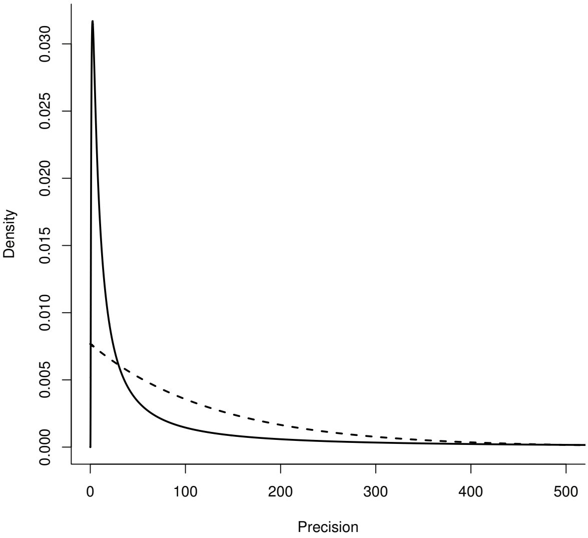
$\mathrm{Prob}(\sigma > u) = \alpha$ gives

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

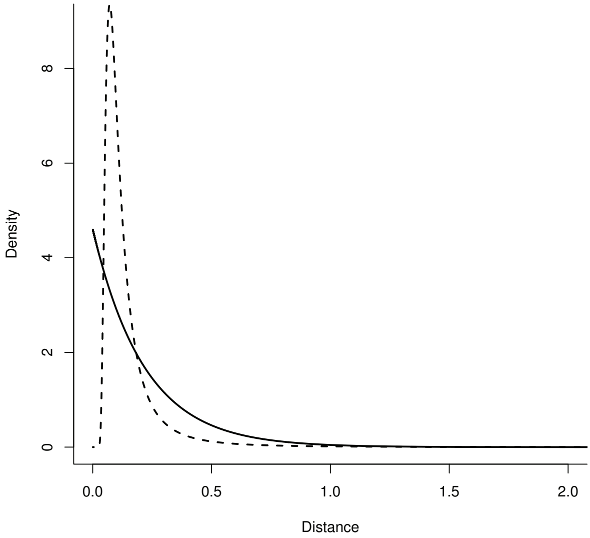
Alternative interpretation

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

Comparison with a similar Gamma-prior



Comparison with a similar Gamma-prior



Experience

As for ν in the Student-t

Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are almost orthogonal

Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are almost orthogonal

Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are almost orthogonal

The AR(1) case

$$x_t = \rho x_{t-1} + \epsilon_t$$

Parameterise using

- Lag-1 correlation ρ
- Marginal precision

Base model:

- $\rho = 0$ which is no *dependence* in time
- $\rho = 1$ which is no *change* in time

The AR(1) case

$$x_t = \rho x_{t-1} + \epsilon_t$$

Parameterise using

- Lag-1 correlation ρ
- Marginal precision

Base model:

- $\rho = 0$ which is no *dependence* in time
- $\rho = 1$ which is no *change* in time

The AR(1) case

$$x_t = \rho x_{t-1} + \epsilon_t$$

Parameterise using

- Lag-1 correlation ρ
- Marginal precision

Base model:

- $\rho = 0$ which is no *dependence* in time
- $\rho = 1$ which is no *change* in time

The AR(1) case

$$x_t = \rho x_{t-1} + \epsilon_t$$

Parameterise using

- Lag-1 correlation ρ
- Marginal precision

Base model:

- $\rho = 0$ which is no *dependence* in time
- $\rho = 1$ which is no *change* in time

The AR(1) case

$$x_t = \rho x_{t-1} + \epsilon_t$$

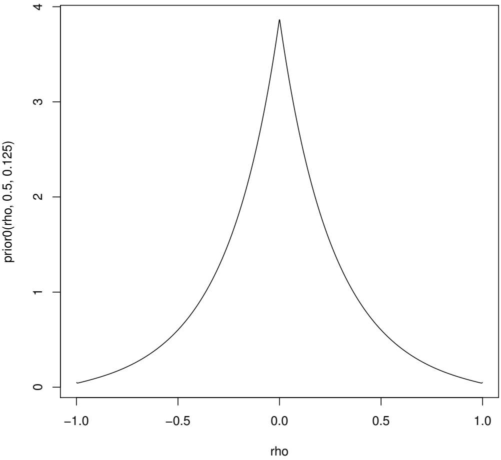
Parameterise using

- Lag-1 correlation ρ
- Marginal precision

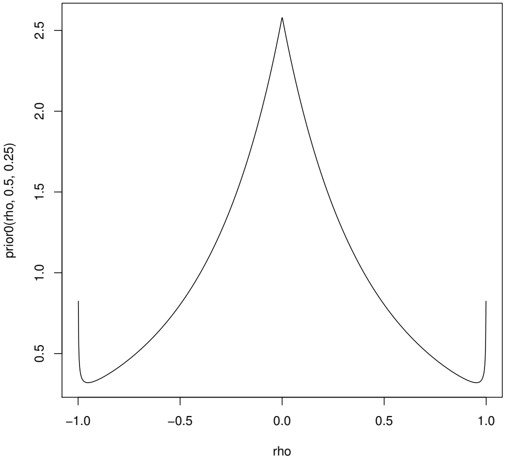
Base model:

- $\rho = 0$ which is no *dependence* in time
- $\rho = 1$ which is no *change* in time

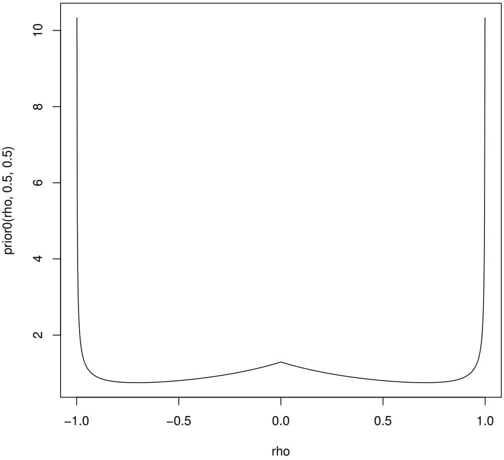
Base model $\rho = 0$



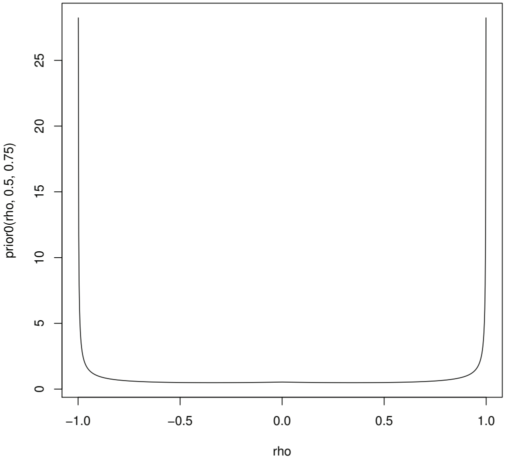
Base model $\rho = 0$



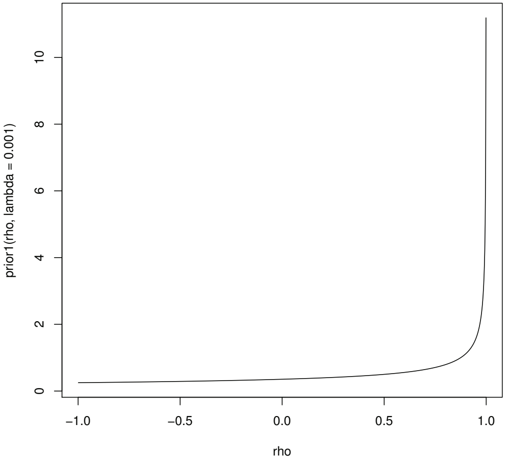
Base model $\rho = 0$



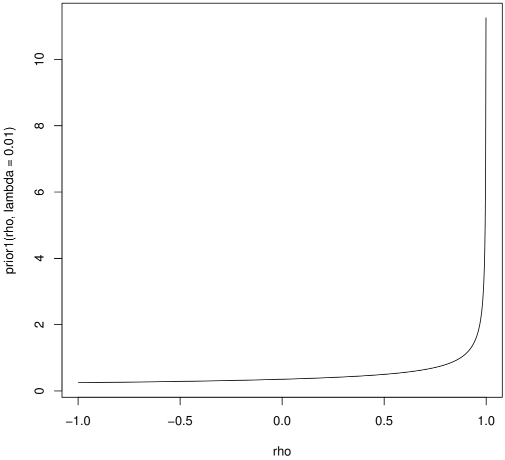
Base model $\rho = 0$



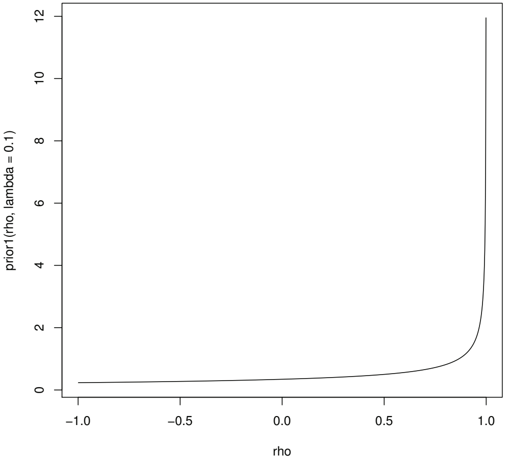
Base model $\rho = 1$



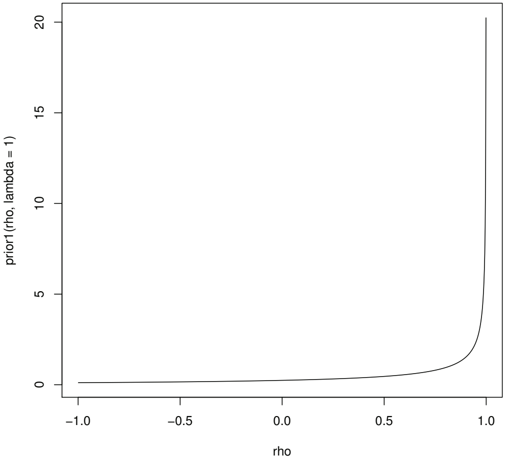
Base model $\rho = 1$



Base model $\rho = 1$



Base model $\rho = 1$



Cox proportional hazard model with time dependent frailty

Hazard for individual i

$$h_i(t, \boldsymbol{z}) = h_{\text{baseline}}(t) \exp\left(\boldsymbol{z}_i^T \boldsymbol{\beta} + u(t, i)\right)$$

Frailty

$$u(t, i) \;\sim\; \text{AR}(1)(t)$$

and replicated in i

Cox proportional hazard model with time dependent frailty

Hazard for individual i

$$h_i(t, \boldsymbol{z}) = h_{\text{baseline}}(t) \exp\left(\boldsymbol{z}_i^T \boldsymbol{\beta} + u(t, i)\right)$$

Frailty

$$u(t, i) \; \sim \; \text{AR}(1)(t)$$

and replicated in i

Prior details...

- Baseline hazard: RW1(κ) with PC prior on κ (stdev ≈ 0.15).
- Time dependent frailty: AR(1) model with
  - PC prior on marginal precision (stdev ≈ 0.3).
  - PC prior for ρ. Base model $\rho = 1$ and Prob($\rho > 1/2$) = 0.75

Prior details...

- Baseline hazard: RW1(κ) with PC prior on κ (stdev ≈ 0.15).
- Time dependent frailty: AR(1) model with
  - PC prior on marginal precision (stdev ≈ 0.3).
  - PC prior for ρ. Base model $\rho = 1$ and $\text{Prob}(\rho > 1/2) = 0.75$

Prior details...

- Baseline hazard: RW1(κ) with PC prior on κ (stdev ≈ 0.15).
- Time dependent frailty: AR(1) model with
  - PC prior on marginal precision (stdev ≈ 0.3).
  - PC prior for ρ. Base model $\rho = 1$ and Prob$(\rho > 1/2) = 0.75$

Prior details...

- Baseline hazard: RW1(κ) with PC prior on κ (stdev ≈ 0.15).
- Time dependent frailty: AR(1) model with
  - PC prior on marginal precision (stdev ≈ 0.3).
  - PC prior for ρ. Base model $\rho = 1$ and Prob$(\rho > 1/2) = 0.75$

Results: `survival::cgd`

Data are from a placebo controlled trial of gamma interferon in chronic granulotomous disease (CGD).

Uses the complete data on time to first serious infection observed through end of study for each patient, which includes the initial serious infections observed through the 7/15/89 interim analysis data cutoff, plus the residual data on occurrence of initial serious infections between the interim analysis cutoff and the final blinded study visit for each patient.

R-code (I)

```
formula <- inla.surv(time, status) ~ 1 +
    treat + inherit2 + age + height + weight +
    propylac + sex + region +
    f(baseline.hazard.idx, model = "ar1", replicate = id,
      hyper = list(
              prec = list(
                      prior = "pc.prec",
                      param = c(u.frailty, a.frailty)),
              rho = list(
                      prior = "pc.rho1",
                      param = c(upper.rho, alpha.rho))))
```

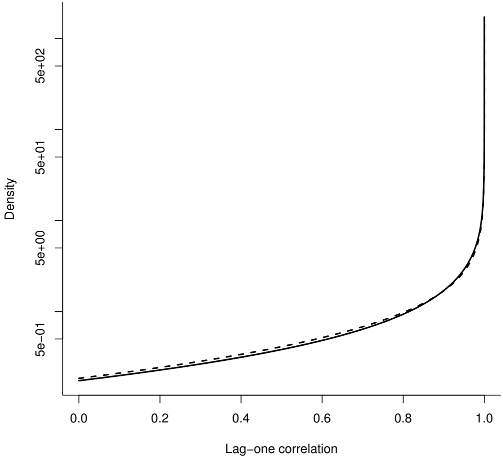
R-code (II)

```
result <- inla(formula,
               family = "coxph",
               data = cgd,
               control.hazard = list(
                       model = "rw1",
                       n.intervals = 25,
                       scale.model = TRUE,
                       hyper = list(
                               prec = list(
                                       prior = "pc.prec",
                                       param = c(u.bh, a.bh)))))
```

Results

Lag-one correlation:

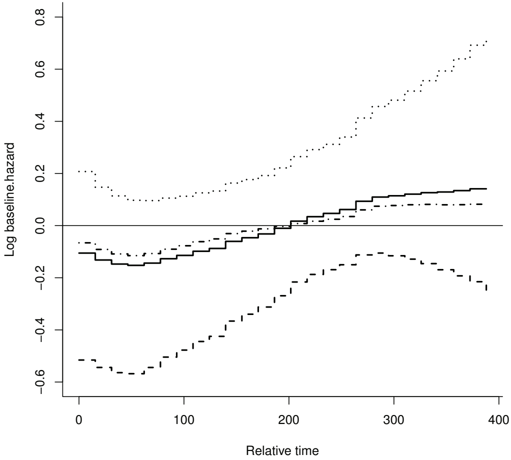
- Log posterior (solid)
- Log prior (dashed)



Results

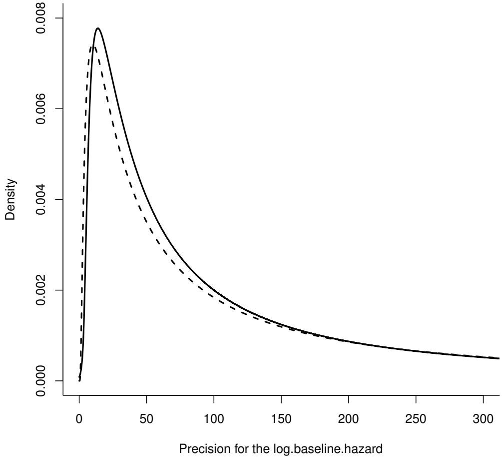
Log baseline hazard:

- Mean (solid)
- Median
- Lower/upper quantile



Results

Posterior for precision for the log baseline hazard



Summary of results

No sign of any time-dependent baseline hazard. This is somewhat contrary to a previous study

STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:1263–1274
Published online 29 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1995

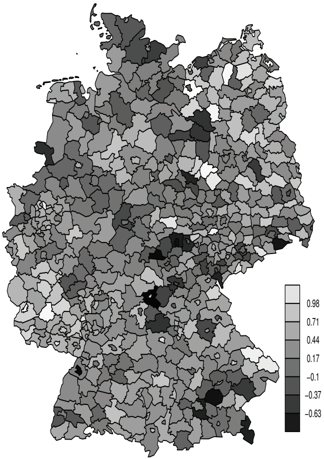
Bayesian inference for recurrent events data using time-dependent frailty

Samuel O. M. Manda[1,*,†] and Renate Meyer[2]

[1] *Biostatistics Unit, School of Medicine, University of Leeds, 24 Hyde Terrace, Leeds LS2 9LN, U.K.*
[2] *Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1, New Zealand*

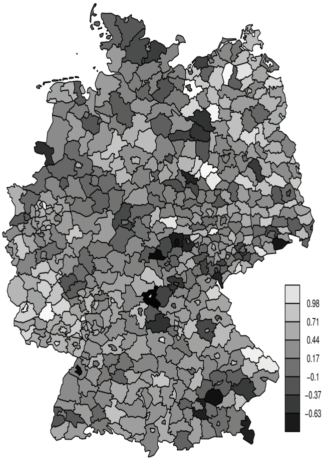
Disease mapping: The BYM-model

- Data $y_i \sim$ Poisson$(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



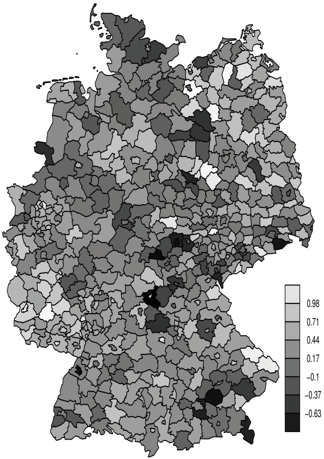
Disease mapping: The BYM-model

- Data $y_i \sim$ Poisson($E_i exp(\eta_i)$)
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



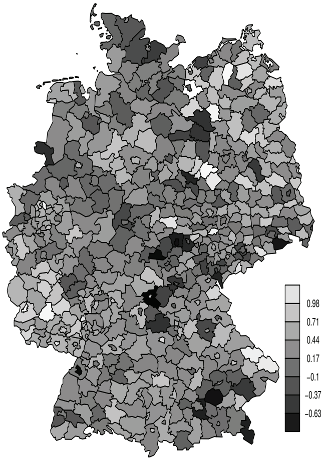
Disease mapping: The BYM-model

- Data $y_i \sim$ Poisson$(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



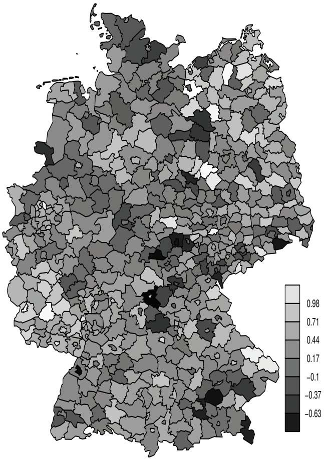
Disease mapping: The BYM-model

- Data $y_i \sim \text{Poisson}(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



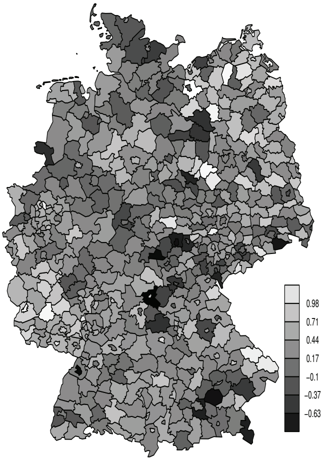
Disease mapping: The BYM-model

- Data $y_i \sim \text{Poisson}(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



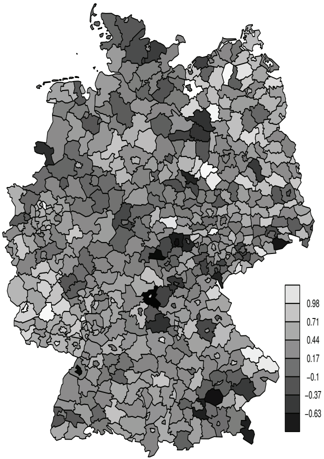
Disease mapping: The BYM-model

- Data $y_i \sim \text{Poisson}(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



Disease mapping: The BYM-model

- Data $y_i \sim \text{Poisson}(E_i exp(\eta_i))$
- Log-relative risk $\eta_i = u_i + v_i$
- Structured/spatial component $\boldsymbol{u}$
- Unstructured component $\boldsymbol{v}$
- Precisions κ_u and κ_v
- Common to use independent Gamma-priors
- Confusion about priors in this case: spatial model is not scaled



Disease mapping (II)

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

Rewrite the model as

$$\eta = \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\gamma}v^{*} + \sqrt{\gamma}u^{*}\right)$$

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

- Marginal precisions τ.
- γ gives it interpretation: independence ($\gamma = 0$), maximal dependence ($\gamma = 1$)
- (almost) orthogonal parameters, use the PC priors for τ and γ separately.

Disease mapping (II)

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

Rewrite the model as

$$\eta = \frac{1}{\sqrt{\tau}} \left(\sqrt{1-\gamma} v^{*} + \sqrt{\gamma} u^{*} \right)$$

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

- Marginal precisions τ.
- γ gives it interpretation: independence ($\gamma = 0$), maximal dependence ($\gamma = 1$)
- (almost) orthogonal parameters, use the PC priors for τ and γ separately.

Disease mapping (II)

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

Rewrite the model as

$$\eta = \frac{1}{\sqrt{\tau}} \left(\sqrt{1-\gamma} v^* + \sqrt{\gamma} u^* \right)$$

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

- Marginal precisions τ.
- γ gives it interpretation: independence ($\gamma = 0$), maximal dependence ($\gamma = 1$)
- (almost) orthogonal parameters, use the PC priors for τ and γ separately.

Disease mapping (II)

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

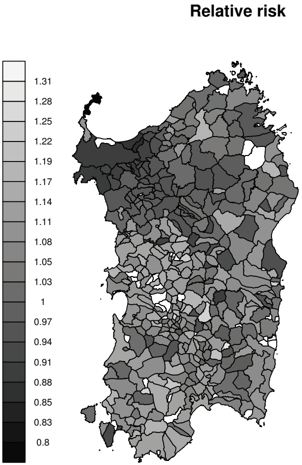
Rewrite the model as

$$\eta = \frac{1}{\sqrt{\tau}} \left(\sqrt{1-\gamma} v^{*} + \sqrt{\gamma} u^{*} \right)$$

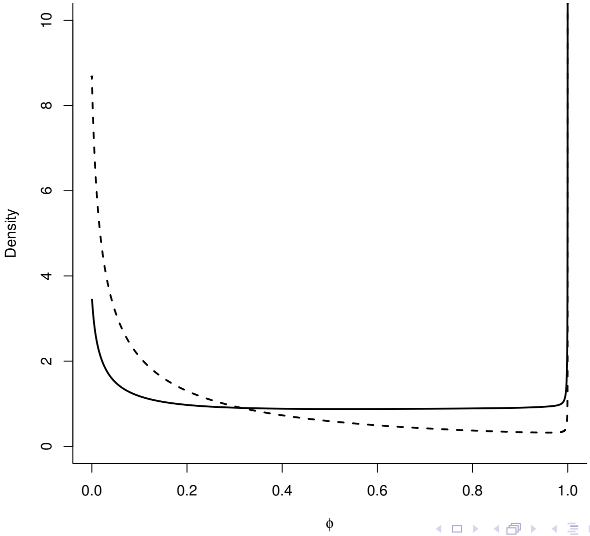
where $\cdot^{*}$ is a unit-variance standardised model.

- Marginal precisions τ.
- γ gives it interpretation: independence ($\gamma = 0$), maximal dependence ($\gamma = 1$)
- (almost) orthogonal parameters, use the PC priors for τ and γ separately.

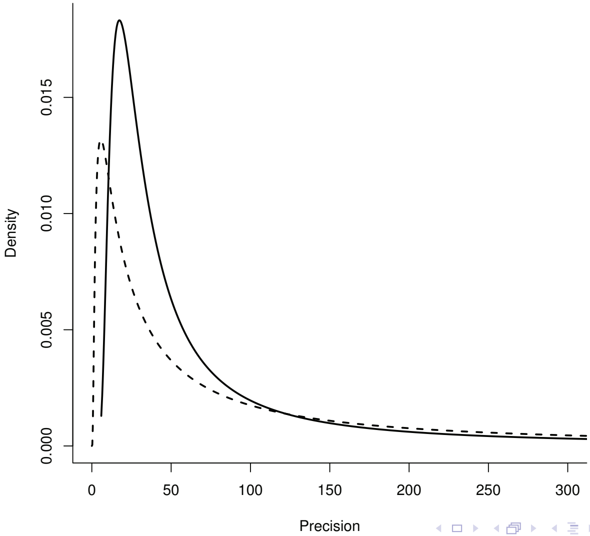
Sardinia-example



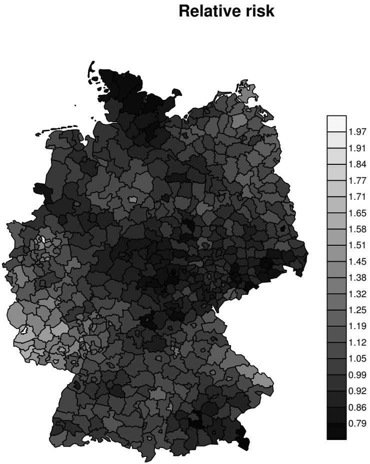
Sardinia-example



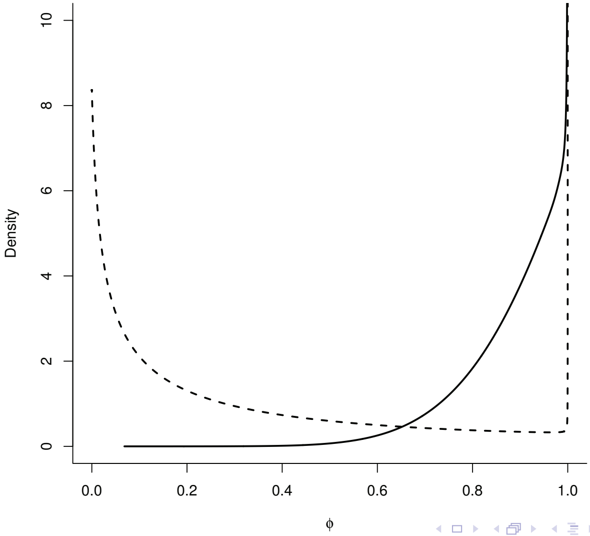
Sardinia-example



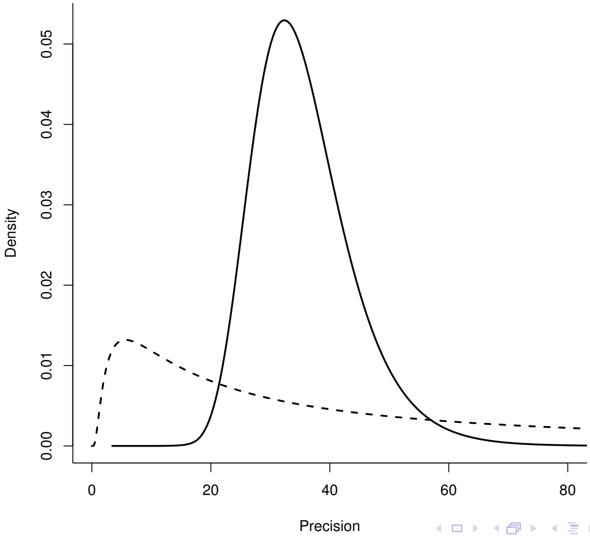
Germany-example



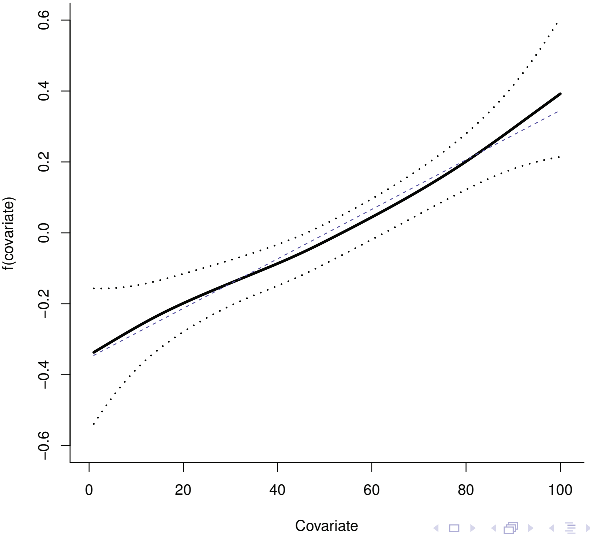
Germany-example



Germany-example



Germany-example



Multivariate cases

- The general multivariate case is harder, due to the

  $$\textit{many(-parameters)} \longrightarrow \textit{one(-distance)}$$

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\boldsymbol{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

Multivariate cases

- The general multivariate case is harder, due to the

  $$\textit{many(-parameters)} \longrightarrow \textit{one(-distance)}$$

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\mathbf{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

Multivariate cases

- The general multivariate case is harder, due to the

  $$\textit{many(-parameters)} \longrightarrow \textit{one(-distance)}$$

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\boldsymbol{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

Multivariate cases

- The general multivariate case is harder, due to the

  $$\textit{many(-parameters)} \longrightarrow \textit{one(-distance)}$$

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\boldsymbol{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

Multivariate cases

- The general multivariate case is harder, due to the

  many(-parameters) $\longrightarrow$ *one(-distance)*

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\boldsymbol{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

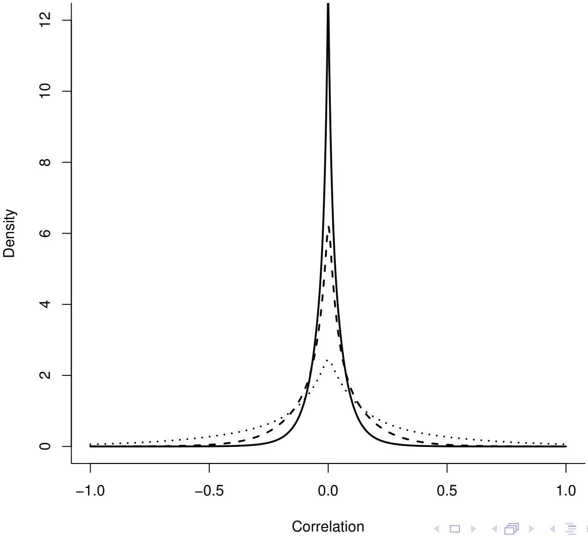
Multivariate cases

- The general multivariate case is harder, due to the

  $$\textit{many(-parameters)} \longrightarrow \textit{one(-distance)}$$

  problem
- With some (serious) skills, we can work out the PC prior for
  - General covariance matrix (base model $\boldsymbol{\Sigma}_0$)
  - General correlation matrix (base model $\boldsymbol{I}$ + linear transform)
  - Toeplitz correlation matrix (AR(p), base model $\boldsymbol{I}$)

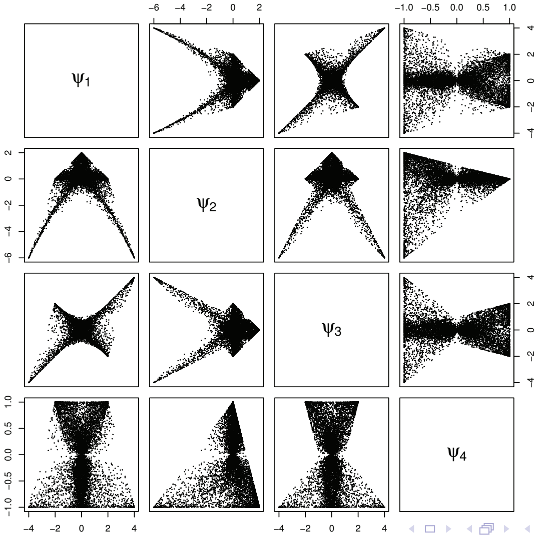
Prior marginal for a 3×3 correlation matrix



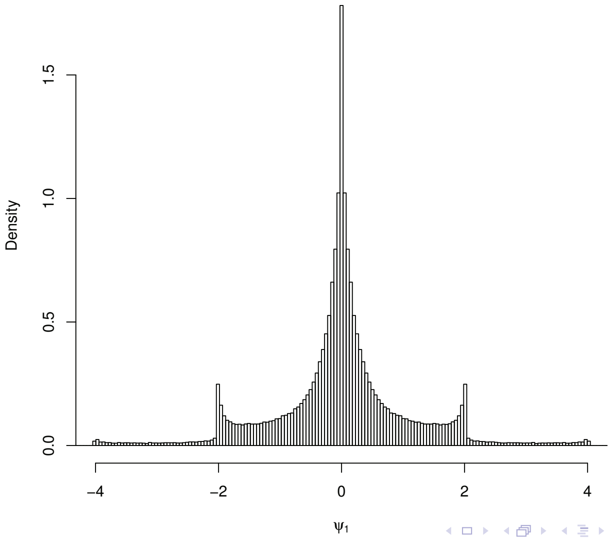
AR(4) model

$$x_t = \psi_1 x_{t-1} + \psi_2 x_{t-2} + \psi_3 x_{t-3} + \psi_4 x_{t-4} + \epsilon_t$$

Samples from the PC prior for the AR(4) model



PC prior marginal for ψ_1 in an AR(4) model



Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

Discussion: PC priors

- The new principled constructive approach to construct priors seems very promising, we are all very excited!
- Easy and very natural interpretation + a well defined shrinkage.
- We can chose the degree of "informativeness".
- Finally, I know what I'm doing wrt priors!!!
- Exciting extentions will grow out this (not discussed)
- Not all cases are easy...
- A lot of work to integrate this into `R-INLA`
- I belive this approach has a great future

References

- T. G. Martins, D. P. Simpson, A. Riebler, H. Rue, and S. H. Sørbye (2014) *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, to appear.