Gaussian Kronecker product Markov random fields

Andrea Riebler

Joint work with Leonhard Held and Håvard Rue

Outline

Introduction

Applications

- Extrapolation of time trends in registry data
- Modelling seasonal patterns in multiple longitudinal profiles

Summary and outlook

Introduction

– In biomedical or public health research, longitudinal or spatio-temporal health outcomes are commonplace.

– If multiple outcomes exist for one time-point and/or unit, standard GMRF priors do not account for a potential dependence between outcomes.

– Here, interaction models or Gaussian Kronecker product Markov random fields would be needed.

$\Rightarrow$ It is desirable to have Kronecker products in INLA

Introduction

– In biomedical or public health research, longitudinal or spatio-temporal health outcomes are commonplace.

– If multiple outcomes exist for one time-point and/or unit, standard GMRF priors do not account for a potential dependence between outcomes.

– Here, interaction models or Gaussian Kronecker product Markov random fields would be needed.

$\Rightarrow$ It is desirable to have Kronecker products in INLA

Introduction

- In biomedical or public health research, longitudinal or spatio-temporal health outcomes are commonplace.
- If multiple outcomes exist for one time-point and/or unit, standard GMRF priors do not account for a potential dependence between outcomes.
- Here, interaction models or Gaussian Kronecker product Markov random fields would be needed.

$\Rightarrow$ It is desirable to have Kronecker products in INLA

You all know what a GMRF is ...

Definition

A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$

with zero-mean, say, "precision" matrix $\mathbf{Q}$ and

$$\pi(\boldsymbol{x}) \propto (|\mathbf{Q}|^\star)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x}\right)$$

where $Q_{ij} = 0 \longleftrightarrow x_i \perp x_j | \boldsymbol{x}_{-ij}$ is called a

Gaussian Markov random field.

($| \cdot |^\star$ denotes the generalized determinant, i.e. the product of non-zero eigenvalues.)

You all know what a GMRF is ...

Definition

A random vector $\tilde{\boldsymbol{x}} = (x_{11}, \dots, x_{n1}, \dots, x_{1R}, \dots, x_{nR})^\top$

with zero-mean, say, "precision" matrix $\mathbf{Q} = \mathbf{Q}_1 \otimes \mathbf{Q}_2$ and

$$\pi(\tilde{\boldsymbol{x}}) \propto (|\mathbf{Q_1} \otimes \mathbf{Q_2}|^\star)^{1/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{x}}^\top \{\mathbf{Q}_1 \otimes \mathbf{Q}_2\}\tilde{\boldsymbol{x}}\right)$$

where $Q_{ij} = 0 \longleftrightarrow x_i \perp x_j | \boldsymbol{x}_{-ij}$ is called a

Gaussian Kronecker product Markov random field.

($| \cdot |^\star$ denotes the generalized determinant, i.e. the product of non-zero eigenvalues.)

You all know what a GMRF is ...

Definition

A random vector $\tilde{\boldsymbol{x}} = (x_{11}, \ldots, x_{n1}, \ldots, x_{1J}, \ldots, x_{nJ})^\top$

with zero-mean, say, "precision" matrix $\mathbf{Q} = \mathbf{Q}_1 \otimes \mathbf{Q}_2$ and

$$\pi(\tilde{\boldsymbol{x}}) \propto (|\mathbf{Q}_1 \otimes \mathbf{Q}_2|^\star)^{1/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{x}}^\top \{\mathbf{Q}_1 \otimes \mathbf{Q}_2\}\tilde{\boldsymbol{x}}\right)$$

where $Q_{ij} = 0 \longleftrightarrow x_i \perp x_j | \boldsymbol{x}_{-ij}$ is called a

Gaussian Kronecker product Markov random field.

($| \cdot |^\star$ denotes the generalized determinant, i.e. the product of non-zero eigenvalues.)

- $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are two lower-dimensional precision matrices.
- $\mathbf{Q}_1$ and $\mathbf{Q}_2$ can be both regular and singular.
- The Kronecker product is the interaction of $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

Some Kronecker product models are there ...

In autumn 2009, we included the `group`-ing option in INLA, which allows

- grouping with an AR1, or
- uniform correlation matrix (`exchangeable`):

Let $\mathbf{C}$ be a $R \times R$ correlation matrix with $R > 1, \rho \neq 1$:

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, \text{ then } \quad \mathbf{C}^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix}$$

$$\text{with} \qquad a = -\frac{(R-2)\cdot\rho + 1}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}$$

$$b = \frac{\rho}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}.$$

The determinant is

$$|\mathbf{C}^{-1}| = |\mathbf{C}|^{-1} = [(1 + (R-1)\rho)(1-\rho)^{R-1}]^{-1}.$$

Some Kronecker product models are there ...

In autumn 2009, we included the `group`-ing option in INLA, which allows

- grouping with an AR1, or
- uniform correlation matrix (`exchangeable`):

Let **C** be a $R \times R$ correlation matrix with $R > 1, \rho \neq 1$:

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, \text{ then } \quad \mathbf{C}^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix}$$

$$\text{with} \qquad a = -\frac{(R-2)\cdot\rho + 1}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}$$

$$b = \frac{\rho}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}.$$

The determinant is

$$|\mathbf{C}^{-1}| = |\mathbf{C}|^{-1} = [(1 + (R-1)\rho)(1-\rho)^{R-1}]^{-1}.$$

Some Kronecker product models are there ...

In autumn 2009, we included the `group`-ing option in INLA, which allows

- grouping with an AR1, or
- uniform correlation matrix (`exchangeable`):

  Let **C** be a $R \times R$ correlation matrix with $R > 1, \rho \neq 1$:

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, \text{ then } \quad \mathbf{C}^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix}$$

with

$$a = -\frac{(R-2)\cdot\rho + 1}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}$$

$$b = \frac{\rho}{(\rho - 1)\{(R-1)\cdot\rho + 1\}}.$$

The determinant is

$$|\mathbf{C}^{-1}| = |\mathbf{C}|^{-1} = [(1 + (R-1)\rho)(1-\rho)^{R-1}]^{-1}.$$

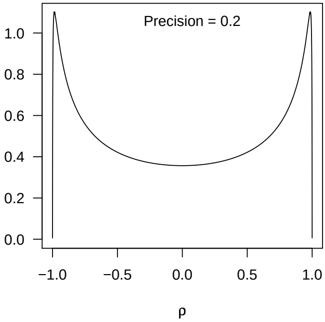
Prior on correlation parameter

Reparameterize ρ using the general Fisher's z-transformation:

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1 - \rho}\right),$$

and assign a $\mathcal{N}(0, \tau^{-1})$ prior to $\rho^\star$.

(Fisher 1958, page 219)



– This prior automatically ensures that $\rho \in (-1/(R-1), 1)$, which is required to ensure positive definiteness of **C**, is fulfilled.

– In addition, $P(\rho > 0) = 0.5$, independent of R.

(see Riebler et al. (2012))

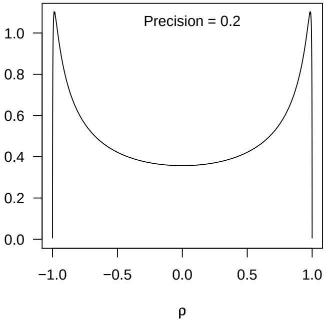
Prior on correlation parameter

Reparameterize ρ using the general Fisher's z-transformation:

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1-\rho}\right),$$

and assign a $\mathcal{N}(0, \tau^{-1})$ prior to $\rho^\star$.

(Fisher 1958, page 219)



– This prior automatically ensures that $\rho \in (-1/(R-1), 1)$, which is required to ensure positive definiteness of **C**, is fulfilled.

– In addition, $P(\rho > 0) = 0.5$, independent of R.

(see Riebler et al. (2012))

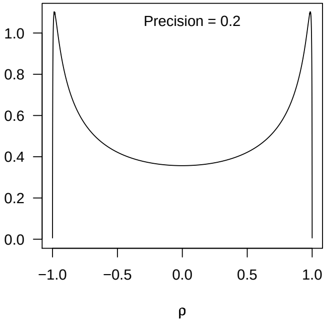
Prior on correlation parameter

Reparameterize ρ using the general Fisher's z-transformation:

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1-\rho}\right),$$

and assign a $\mathcal{N}(0, \tau^{-1})$ prior to $\rho^\star$.

(Fisher 1958, page 219)



- This prior automatically ensures that $\rho \in (-1/(R-1), 1)$, which is required to ensure positive definiteness of **C**, is fulfilled.
- In addition, $P(\rho > 0) = 0.5$, independent of R.

(see Riebler et al. (2012))

Outline

Multivariate age-period-cohort models

Applied to multiple mortality or morbidity tables:

y_{ijr}: Number of cases in age group i at calendar time j in region r
n_{ijr}: Number of persons at risk in age group i at time j in region r

$$y_{ijr} | \eta_{ijr} \sim \text{Poisson}(n_{ijr} \lambda_{ijr})$$
$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{i(,r)} + \varphi_{j(,r)} + \psi_{k(i,j)(,r)} + z_{ijr},$$

with region-specific intercept, (region-specific) age, period and cohort effects and overdispersion parameters.

We assume the usual sum-to-zero constraints.

Note:

- Differences of region-specific effects, e.g. $\Delta_i = \theta_{i,r_1} - \theta_{i,r_2}$, $r_1 \neq r_2$ are identifiable.
- Adjusted differences $\Delta_\mu + \Delta_i$, with $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$, can be interpreted as (average) log relative risk. Riebler and Held (2010)

Multivariate age-period-cohort models

Applied to multiple mortality or morbidity tables:

y_{ijr}: Number of cases in age group i at calendar time j in region r
n_{ijr}: Number of persons at risk in age group i at time j in region r

$$y_{ijr}|\eta_{ijr} \sim \text{Poisson}(n_{ijr}\lambda_{ijr})$$
$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{i(,r)} + \varphi_{j(,r)} + \psi_{k(i,j)(,r)} + z_{ijr},$$

with region-specific intercept, (region-specific) age, period and cohort effects and overdispersion parameters.

We assume the usual sum-to-zero constraints.

Note:

- Differences of region-specific effects, e.g. $\Delta_i = \theta_{i,r_1} - \theta_{i,r_2}$, $r_1 \neq r_2$ are identifiable.
- Adjusted differences $\Delta_\mu + \Delta_i$, with $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$, can be interpreted as (average) log relative risk.

Riebler and Held (2010)

Multivariate age-period-cohort models

Applied to multiple mortality or morbidity tables:

y_{ijr}: Number of cases in age group i at calendar time j in region r
n_{ijr}: Number of persons at risk in age group i at time j in region r

$$
\begin{aligned}
y_{ijr}|\eta_{ijr} &\sim \text{Poisson}(n_{ijr}\lambda_{ijr}) \\
\eta_{ijr} &= \log(\lambda_{ijr}) = \mu_r + \theta_{i(,r)} + \varphi_{j(,r)} + \psi_{k(i,j)(,r)} + z_{ijr},
\end{aligned}
$$

with region-specific intercept, (region-specific) age, period and cohort effects and overdispersion parameters.

We assume the usual sum-to-zero constraints.

Note:

- Differences of region-specific effects, e.g. $\Delta_i = \theta_{i,r_1} - \theta_{i,r_2}$, $r_1 \neq r_2$ are identifiable.
- Adjusted differences $\Delta_\mu + \Delta_i$, with $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$, can be interpreted as (average) log relative risk. Riebler and Held (2010)

Multivariate age-period-cohort models

Applied to multiple mortality or morbidity tables:

y_{ijr}: Number of cases in age group i at calendar time j in region r
n_{ijr}: Number of persons at risk in age group i at time j in region r

$$y_{ijr}|\eta_{ijr} \sim \text{Poisson}(n_{ijr}\lambda_{ijr})$$
$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{i(,r)} + \varphi_{j(,r)} + \psi_{k(i,j)(,r)} + z_{ijr},$$

with region-specific intercept, (region-specific) age, period and cohort effects and overdispersion parameters.

We assume the usual sum-to-zero constraints.

Note:

- Differences of region-specific effects, e.g. $\Delta_i = \theta_{i,r_1} - \theta_{i,r_2}$, $r_1 \neq r_2$ are identifiable.
- Adjusted differences $\Delta_\mu + \Delta_i$, with $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$, can be interpreted as (average) log relative risk. Riebler and Held (2010)

Multivariate age-period-cohort models

Applied to multiple mortality or morbidity tables:

y_{ijr}: Number of cases in age group i at calendar time j in region r
n_{ijr}: Number of persons at risk in age group i at time j in region r

$$
\begin{aligned}
y_{ijr}|\eta_{ijr} &\sim \text{Poisson}(n_{ijr}\lambda_{ijr}) \\
\eta_{ijr} &= \log(\lambda_{ijr}) = \mu_r + \theta_{i(,r)} + \varphi_{j(,r)} + \psi_{k(i,j)(,r)} + z_{ijr},
\end{aligned}
$$

with region-specific intercept, (region-specific) age, period and cohort effects and overdispersion parameters.

We assume the usual sum-to-zero constraints.

Note:

- Differences of region-specific effects, e.g. $\Delta_i = \theta_{i,r_1} - \theta_{i,r_2}$, $r_1 \neq r_2$ are identifiable.
- Adjusted differences $\Delta_\mu + \Delta_i$, with $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$, can be interpreted as (average) log relative risk.

Riebler and Held (2010)

Incorporation of correlation

We propose:

- Correlated overdispersion parameters across regions.
- Correlated smoothing priors for region-specific age, period and/or cohort effects.

Second-order random walks (RW2s) of region-specific period effects $\varphi_1, \ldots,$ φ_R, say, can be correlated using the stacked vector $\tilde{\varphi} = (\varphi_1^\top, \ldots, \varphi_R^\top)^\top$:

$$f(\tilde{\varphi}|\mathbf{C}_\varphi, \kappa_\varphi) \propto |\kappa_\varphi \mathbf{C}_\varphi^{-1}|^{(J-2)/2} \exp\left(-\frac{1}{2}\tilde{\varphi}^\top \left\{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}^{RW2}\right\}\tilde{\varphi}\right).$$

Advantages:

- More precise relative risk estimates
- Imputation of missing data for one particular region may be improved.

Riebler et al. (2012)

Incorporation of correlation

We propose:

- Correlated overdispersion parameters across regions.
- Correlated smoothing priors for region-specific age, period and/or cohort effects.

  Second-order random walks (RW2s) of region-specific period effects $\varphi_1, \ldots,$ φ_R, say, can be correlated using the stacked vector $\tilde{\varphi} = (\varphi_1^\top, \ldots, \varphi_R^\top)^\top$:

$$f(\tilde{\varphi}|\mathbf{C}_\varphi, \kappa_\varphi) \propto |\kappa_\varphi \mathbf{C}_\varphi^{-1}|^{(J-2)/2} \exp\left(-\frac{1}{2}\tilde{\varphi}^\top \left\{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}^{RW2}\right\}\tilde{\varphi}\right).$$

Advantages:

- More precise relative risk estimates
- Imputation of missing data for one particular region may be improved.

Riebler et al. (2012)

Incorporation of correlation

We propose:

- Correlated overdispersion parameters across regions.
- Correlated smoothing priors for region-specific age, period and/or cohort effects.

  Second-order random walks (RW2s) of region-specific period effects $\boldsymbol{\varphi}_1, \dots,$ $\boldsymbol{\varphi}_R$, say, can be correlated using the stacked vector $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\varphi}_1^\top, \dots, \boldsymbol{\varphi}_R^\top)^\top$:

$$f(\tilde{\boldsymbol{\varphi}} | \mathbf{C}_\varphi, \kappa_\varphi) \propto |\kappa_\varphi \mathbf{C}_\varphi^{-1}|^{(J-2)/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \left\{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}^{RW2}\right\} \tilde{\boldsymbol{\varphi}}\right).$$

Advantages:

- More precise relative risk estimates
- Imputation of missing data for one particular region may be improved.

Riebler et al. (2012)

Incorporation of correlation

We propose:

- Correlated overdispersion parameters across regions.
- Correlated smoothing priors for region-specific age, period and/or cohort effects.

  Second-order random walks (RW2s) of region-specific period effects $\boldsymbol{\varphi}_1, \ldots,$ $\boldsymbol{\varphi}_R$, say, can be correlated using the stacked vector $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\varphi}_1^\top, \ldots, \boldsymbol{\varphi}_R^\top)^\top$:

$$f(\tilde{\boldsymbol{\varphi}}|\mathbf{C}_\varphi, \kappa_\varphi) \propto |\kappa_\varphi \mathbf{C}_\varphi^{-1}|^{(J-2)/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \left\{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}^{RW2}\right\}\tilde{\boldsymbol{\varphi}}\right).$$

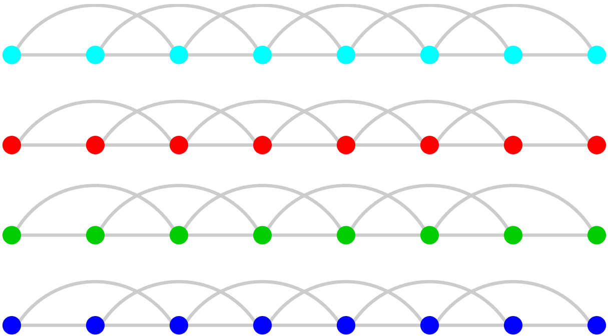
Advantages:

- More precise relative risk estimates
- Imputation of missing data for one particular region may be improved.

Riebler et al. (2012)

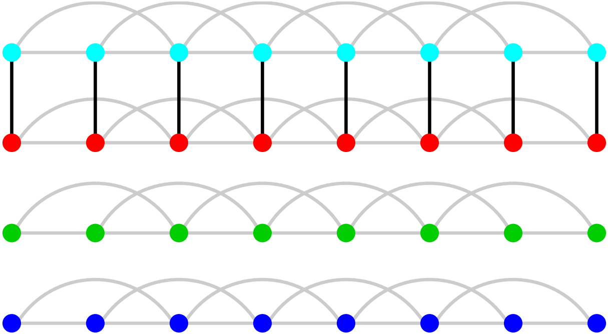
Graph for a correlated RW2

Illustration for $R = 4$:



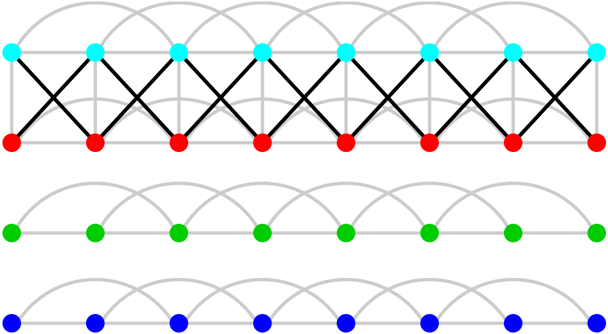
Graph for a correlated RW2

Illustration for $R = 4$:



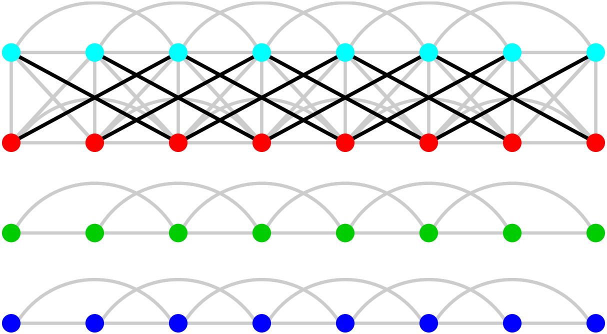
Graph for a correlated RW2

Illustration for $R = 4$:



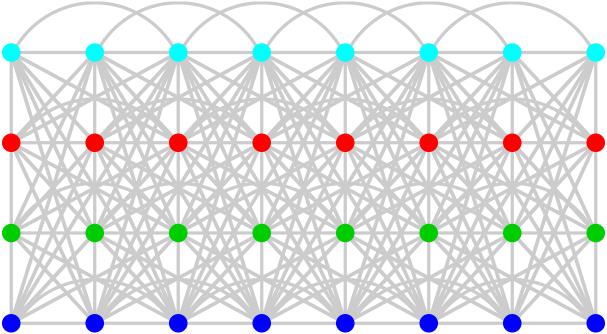
Graph for a correlated RW2

Illustration for $R = 4$:



Graph for a correlated RW2

Illustration for $R = 4$:



Motivation

In Switzerland, cancer is registered on a cantonal level, so that data collection started at different times in the individual cantons.

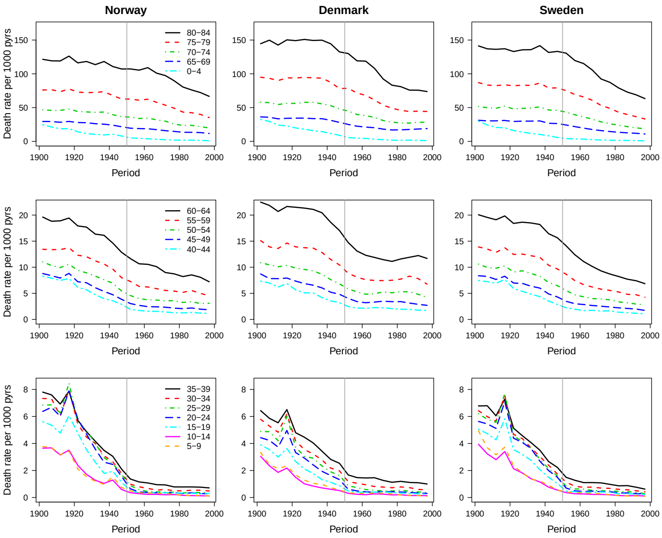
- First Swiss cancer registration system in Geneva in 1970.
- Today, most cantons have cancer registers.
- Until 2013, the whole Swiss population should be captured by a cancer registration system.

Correlated multivariate APC models can borrow strength from cantons with a longer collection period, when projecting missing data for cantons with a younger registration system.

We analyzed this ability in a cross-prediction study

(Details in Riebler et al. (2012))

Female mortality in Scandinavia 1900–2000



$R = 3$ regions, $J = 20$ five-year periods, $I = 17$ age groups.

Cross-prediction study

- For either the first or second half of the 20th century all observations from one particular country are treated as missing.
- Then, the omitted data are predicted.
- Comparison to a univariate APC model and an established demographic forecasting model.

Results:

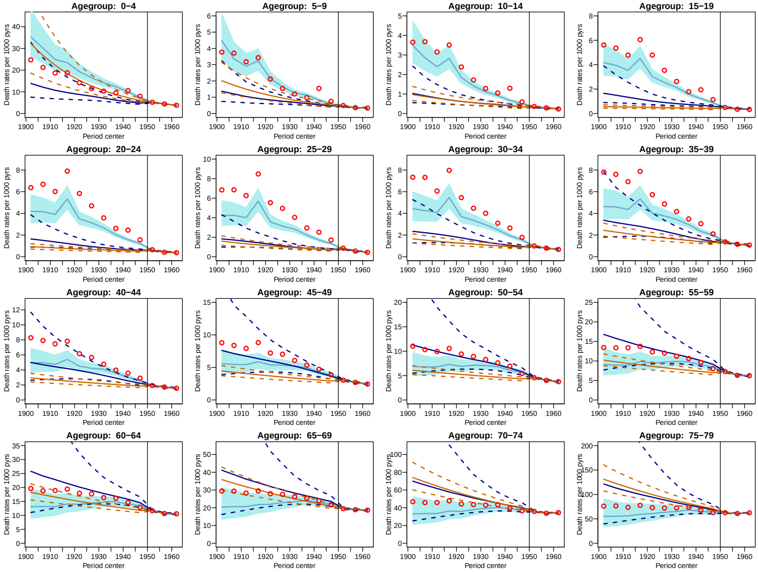
In five of six scenarios the correlated APC model was the best model regarding the proper Dawid-Sebastiani scoring rule. (Gneiting and Raftery, 2007)

INLA-call

```
> library(INLA)

# data specification, setting data to be predicted to NA
# ...

> prior.rw2 <-  c(1,0.00005)
> prior.iid <- c(1,0.005)

> country <- rep(c(1,2,3), each=AGE*PERIOD)

## model with correlated time effects and overdispersion
> model <- y~f(age, model="rw2", hyper=list(prec = list(param=prior.rw2)),
        group=country, constr=TRUE, rankdef=2) +
  f(period, model="rw2", hyper=list(prec = list(param=prior.rw2)),
        group=country, constr=TRUE, rankdef=2) +
  f(cohort, ...) +
  f(overdis, model="iid", hyper=list(prec = list(param=prior.iid)),
        group=country) + mu1 + mu2 + mu3 - 1

# with so many hyperparameters we have to increase the number of
# maximum function evaluations in the derivation of the posterior
# marginals for the hyperparameters
> results = inla(model, family="poisson", E=pop, data=data,
                control.predictor=list(compute=TRUE),
                control.inla=list(numint.maxfeval=80000000))
```

Projected rates for Norway 1900-1949 (80% CI)

Correlated multivariate APC (light blue shaded), univariate APC (dark blue), demographic model (orange).



Outline

Introduction

Applications

- Extrapolation of time trends in registry data
- Modelling seasonal patterns in multiple longitudinal profiles

Summary and outlook

Modelling seasonal variations

- Time-series of infectious disease counts are marked by occasional outbreaks, but additionally there are frequently seasonal variations.

- Typically, a superposition of sine and cosine functions is used.

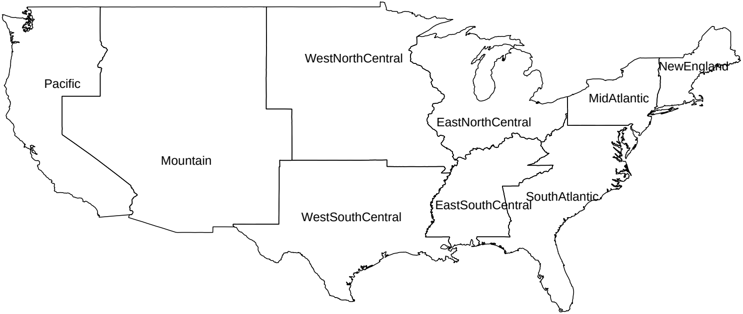
- However, in some cases this might be too simple for handling sharp peaks.

- Circular random walks (CRWs) are similar to periodic splines and represent a flexible alternative.

- To allow, for region-dependent disease onsets we use correlated CRWs in a multivariate setting.

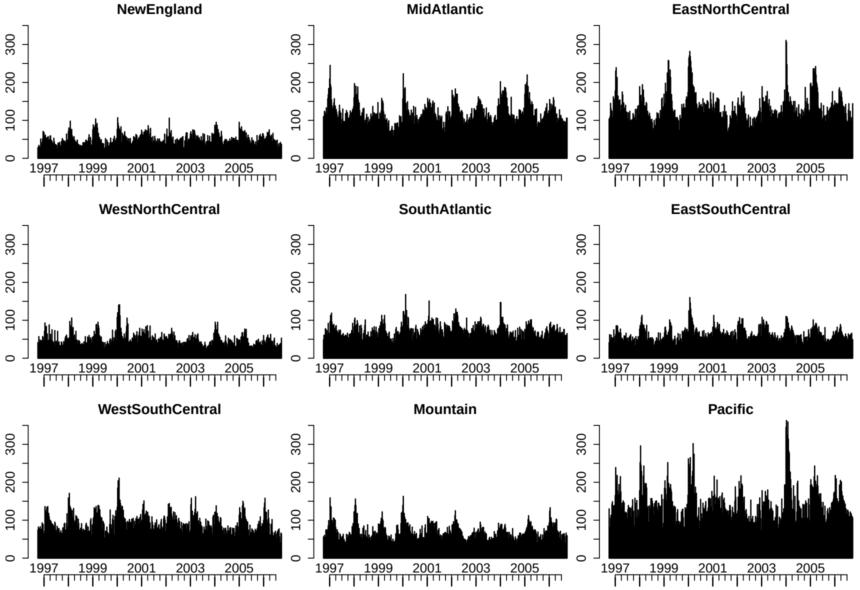
Death from influenza and pneumonia in the USA

– 9 major regions
– 520 weeks (40/1996 to 39/2006) $\Rightarrow$ Period $p = 52$



(Brownstein et al., 2006)

Weekly number of deaths from 40/1996 to 39/2006



Model details

y_{tr}: number of deaths in region r, $r = 1, \dots, 9$ at time t, $t = 1, \dots, 520$.
n_r: population size in region r (in the year 2000).

$$y_{tr} \sim \text{Poisson}(n_r \lambda_{tr}); \qquad \log(\lambda_{tr}) = \mu_r + \beta_{(t \bmod 52)r} + \alpha_{tr}$$

- Region specific intercepts μ_r.
- Seasonal effects $\beta_{(t \bmod 52)r}$ with period 52, modelled using a correlated CRW of second order (cCRW2).
- Time effects α_{tr} modelled using a correlated autoregressive process of first order (cAR1).

Model details

y_{tr}: number of deaths in region r, $r = 1, \dots, 9$ at time t, $t = 1, \dots, 520$.
n_r: population size in region r (in the year 2000).

$$y_{tr} \sim \text{Poisson}(n_r \lambda_{tr}); \qquad \log(\lambda_{tr}) = \mu_r + \beta_{(t \bmod 52)r} + \alpha_{tr}$$

- Region specific intercepts μ_r.
- Seasonal effects $\beta_{(t \bmod 52)r}$ with period 52, modelled using a correlated CRW of second order (cCRW2).
- Time effects α_{tr} modelled using a correlated autoregressive process of first order (cAR1).

Model details

y_{tr}: number of deaths in region r, $r = 1, \dots, 9$ at time t, $t = 1, \dots, 520$.
n_r: population size in region r (in the year 2000).

$$y_{tr} \sim \text{Poisson}(n_r \lambda_{tr}); \qquad \log(\lambda_{tr}) = \mu_r + \beta_{(t \bmod 52)r} + \alpha_{tr}$$

- Region specific intercepts μ_r.
- Seasonal effects $\beta_{(t \bmod 52)r}$ with period 52, modelled using a correlated CRW of second order (cCRW2).
- Time effects α_{tr} modelled using a correlated autoregressive process of first order (cAR1).

Model details

y_{tr}: number of deaths in region r, $r = 1, \dots, 9$ at time t, $t = 1, \dots, 520$.
n_r: population size in region r (in the year 2000).

$$y_{tr} \sim \text{Poisson}(n_r \lambda_{tr}); \qquad \log(\lambda_{tr}) = \mu_r + \beta_{(t \bmod 52)r} + \alpha_{tr}$$

- Region specific intercepts μ_r.
- Seasonal effects $\beta_{(t \bmod 52)r}$ with period 52, modelled using a correlated CRW of second order (cCRW2).
- Time effects α_{tr} modelled using a correlated autoregressive process of first order (cAR1).

The circular random walk of second order

Let $R = 1$, the circular random walk of second order for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{52})^\top$ is given by

$$f(\boldsymbol{\beta}|\kappa) \propto \kappa^{(52-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^\top \mathbf{R}^{CRW2} \boldsymbol{\beta}\right)$$

with precision matrix

$$\mathbf{R}^{CRW2} = \kappa \begin{pmatrix} 6 & -4 & 1 & 0 & \cdots & 0 & 1 & -4 \\ -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -4 & 6 & -4 \\ -4 & 1 & 0 & \cdots & 0 & 1 & -4 & 6 \end{pmatrix},$$

i.e. a circulant matrix with base $\mathbf{d} = \kappa \cdot (6, -4, 1, 0, \ldots, 0, 1, -4)^\top$ and unknown precision parameter κ. (Rue and Held, 2005, Sec. 2.6.1)

Correlated circular random walk of second order

The individual CRW2s $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_9$ can be correlated using the stacked vector $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_9^\top)^\top$:

$$f(\tilde{\boldsymbol{\beta}}|\mathbf{C}, \kappa) \propto |\kappa \mathbf{C}^{-1}|^{(52-2)/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\beta}}^\top \left\{\mathbf{C}^{-1} \otimes \mathbf{R}^{CRW2}\right\}\tilde{\boldsymbol{\beta}}\right).$$

where $\mathbf{C} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ denotes a 9×9 uniform correlation matrix with unknown correlation ρ.

(Riebler et al. (2012))

Hyperpriors

We need to assign hyperpriors to 5 hyperparameters:

- 1 autoregressive parameter:
  $\mathcal{N}(0, 0.2^{-1})$ for Fisher's z-transformed parameter.
- 2 precisions:
  Gamma$(1, 0.00005)$ for κ_{cCRW2} and Gamma$(0.1, 0.001)$ for κ_{cAR1}.
- 2 correlations:
  $\mathcal{N}(0, 0.2^{-1})$ for general Fisher's z-transformed parameters
  $\Rightarrow$ **C** is positive definite without any constraints.

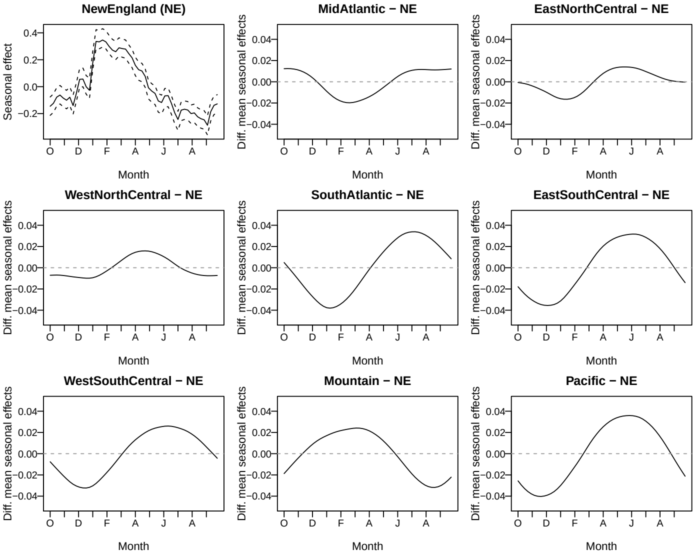
(Fisher, 1958, page 219).

INLA-call

```
> library(INLA)

# data specification (fluUSA)
# ...

> model <- count ~ offset(log(population)) + m1 + ... + m9 - 1 +
        f(week, model="rw2", hyper=list(prec=list(param=c(1, 0.00005))),
            cyclic=TRUE, group=region, constr=TRUE)  +
        f(time, model="ar1",
            hyper=list(prec=list(param=c(0.1,0.001)),
                       rho=list(param=c(0,0.2))), group=region)

> results <- inla(model, data = fluUSA, family = "poisson",
        control.compute=list(dic=TRUE))
```

Estimated seasonal effects

Correlation between seasonal patterns is 0.999 (95% CI:[0.998, 1]).



Model choice

Using DIC we compare the results of the proposed model to:

- a model assuming a common CRW2 for all regions.
- a model assuming independent CRW2s for each region.

	common CRW2	independent CRW2	correlated CRW2
DIC	36707	36716	36704

This is work in progress:

- Include dummies for known events, e.g. Christmas.
- Would like to account for spatial correlation, e.g. based on the degree of neighbourhood (?).
- Compare the results to those of models with a (co)sine function.

Model choice

Using DIC we compare the results of the proposed model to:

- a model assuming a common CRW2 for all regions.
- a model assuming independent CRW2s for each region.

	common CRW2	independent CRW2	correlated CRW2
DIC	36707	36716	36704

This is work in progress:

- Include dummies for known events, e.g. Christmas.
- Would like to account for spatial correlation, e.g. based on the degree of neighbourhood (?).
- Compare the results to those of models with a (co)sine function.

Model choice

Using DIC we compare the results of the proposed model to:

- a model assuming a common CRW2 for all regions.
- a model assuming independent CRW2s for each region.

	common CRW2	independent CRW2	correlated CRW2
DIC	36707	36716	36704

This is work in progress:

- Include dummies for known events, e.g. Christmas.
- Would like to account for spatial correlation, e.g. based on the degree of neighbourhood (?).
- Compare the results to those of models with a (co)sine function.

Model choice

Using DIC we compare the results of the proposed model to:

- a model assuming a common CRW2 for all regions.
- a model assuming independent CRW2s for each region.

	common CRW2	independent CRW2	correlated CRW2
DIC	36707	36716	36704

This is work in progress:

- Include dummies for known events, e.g. Christmas.
- Would like to account for spatial correlation, e.g. based on the degree of neighbourhood (?).
- Compare the results to those of models with a (co)sine function.

Summary

Experiences:

- Due to the increasing number of hyperparamters (?), I often had to increase `numint.maxfeval` in `control.inla` to avoid warnings.
- "Long" running times.

Summary and outlook:

- It works well.
- There are much more applications, e.g. invariant smoothing of multinomial data, ...
- A more general/flexible framework is desirable
  - to include, for example, spatial correlation depending on the distance between units/degree of neighbourhood.
  - to couple more than two models?
  - to let the correlation between units depend on associated covariate information of the units?

Summary

Experiences:

- Due to the increasing number of hyperparamters (?), I often had to increase `numint.maxfeval` in `control.inla` to avoid warnings.
- "Long" running times.

Summary and outlook:

- It works well.
- There are much more applications, e.g. invariant smoothing of multinomial data, ...
- A more general/flexible framework is desirable
  - to include, for example, spatial correlation depending on the distance between units/degree of neighbourhood.
  - to couple more than two models?
  - to let the correlation between units depend on associated covariate information of the units?

Thank you for your attention!

- Brownstein, J.S., Wolfe, C.J. and Mandl, K.D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine*, 3, e401. [Influenza data]
- Fisher, R.A. (1958). Statistical Methods for Research Workers. 13th (rev.) edn, Oliver & Boyd, Edinburgh. [general Fisher's z-transformation]
- Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction and estimation. *JASA*, 102, 359–378. [Scoring rules]
- Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics*, 11, 57–69. [MAPC]
- Riebler, A., Held, L. and Rue, H. (2012). Estimation and extrapolation of time trends in registry data–Borrowing strength from related populations. *Annals of Applied Statistics*, 6, 304–333. [correlated MAPC]