STK4150 - Environmental and spatial statistics
Statistical preliminaries

Odd Kolbjørnsen, Geir Storvik

16 January 2017

Statistical preliminaries

- Multivariate Gaussian distribution
- Conditional probabilities and hierarchical modelling
- Bayesian hierarchical modelling
- Inference
- Prediction
- Computation
- Graphical representation of statistical models

Multivariate Gaussian distribution

- Central in modelling dependence
- Definition: A random vector $\mathbf{Y} = (Y_1, \cdots Y_m)^T$ is said to have the multivariate normal distribution if it satisfies the following condition: Every linear combination of its components

  $$V = a_1 Y_1 + \cdots + a_m Y_m$$

  is normally distributed. That is, for any constant vector $\mathbf{a} \in \mathcal{R}^n$, the random variable $V = \mathbf{a}^T \mathbf{Y}$ has a univariate normal distribution.
- Density

  $$[\mathbf{Y}] = (2\pi)^{-m/2} |\mathbf{\Sigma}|^{-1/2} \exp[-(1/2)(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})]$$
- Applications:
  - $\mathbf{Z}$ actual observations
  - $\mathbf{Y}$ latent variables which is multivariate Gaussian, data

  $$[\mathbf{Z}|\mathbf{Y}] = \prod_i [Z_i | Y_i]$$

  $$[\mathbf{Z}] = \int_{\mathbf{Y}} [\mathbf{Z}|\mathbf{Y}][\mathbf{Y}] d\mathbf{Y}$$

  Allow for MANY complicated multivariate distributions

Conditional probabilities and hierarchical modelling

	Variable	Densities	Notation in book
Data model:	$\mathbf{Z}$	$p(\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta})$	$[\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta}]$
Process model:	$\mathbf{Y}$	$p(\mathbf{Y}\|\boldsymbol{\theta})$	$[\mathbf{Y}\|\boldsymbol{\theta}]$
Parameter model:	$\boldsymbol{\theta}$	$p(\boldsymbol{\theta})$	$[\boldsymbol{\theta}]$

$$[\boldsymbol{\theta},\mathbf{Y},\mathbf{Z}] = [\boldsymbol{\theta}] \times [\mathbf{Y}|\boldsymbol{\theta}] \times [\mathbf{Z}|\mathbf{Y},\boldsymbol{\theta}]$$

- Model built up by several conditional probabilities/densitities
- Easier than specifying a multivariate distribution on $(\boldsymbol{\theta},\mathbf{Y},\mathbf{Z})$ directly.
- Model for **Y** often build up by physical/biological knowledge!

Example: Repeated counts

Seizure counts in a randomised trial of anti-convulsant therapy in epilepsy. Table below: successive seizure counts for 59 patients.

- treatment (0,1)
- 8-week baseline seizure counts
- age in years

Patient	y_1	y_2	y_3	y_4	Trt	Base	Age
1	5	3	3	3	0	11	31
2	3	5	3	3	0	11	30
3	2	4	0	5	0	6	25
4	4	4	1	4	0	8	36
....							
8	40	20	21	12	0	52	42
9	5	6	6	5	0	12	37
....							
59	1	4	3	2	1	12	37

Possible model

$$\log \mu_{jk} = Y_{jk} = \mathbf{x}_j\boldsymbol{\beta} + b_{jk} + \varepsilon_{jk}, \quad \mathrm{Cov}[b_{jk}, b_{jl}] = \sigma^2 a^{|k-l|}$$

$$Z_{jk}|\mu_{jk} \sim \mathrm{Poisson}(\mu_{jk})$$

Multivariate Poisson distribution!

Bayesian approach

- Assume $\boldsymbol{\theta}$ is *stochastic*
- Prior distribution $[\boldsymbol{\theta}]$ specifies our prior knowledge about $\boldsymbol{\theta}$
- Knowledge *updated* when data $\mathbf{Z}$ is obtained (Bayes theorem)

$$[\boldsymbol{\theta}|\mathbf{Z}] = \frac{[\boldsymbol{\theta},\mathbf{Z}]}{[\mathbf{Z}]} \quad \left(= \frac{\int_{\mathbf{Y}}[\boldsymbol{\theta},\mathbf{Y},\mathbf{Z}]d\mathbf{Y}}{\int_{\boldsymbol{\theta},\mathbf{Y}}[\boldsymbol{\theta},\mathbf{Y},\mathbf{Z}]d\mathbf{Y}d\boldsymbol{\theta}}\right)$$

- Can handle uncertainty in parameters coherently

$$\begin{aligned} E[\mathbf{Z}^*|\mathbf{Z}] =& E^{\boldsymbol{\theta}}[E^{\mathbf{Z}^*}[\mathbf{Z}^*|\boldsymbol{\theta},\mathbf{Z}]] \\ \text{Var}[\mathbf{Z}^*|\mathbf{Z}] =& E^{\boldsymbol{\theta}}[\text{Var}^{\mathbf{Z}^*}[\mathbf{Z}^*|\boldsymbol{\theta},\mathbf{Z}]] + \text{Var}^{\boldsymbol{\theta}}[E^{\mathbf{Z}^*}[\mathbf{Z}^*|\boldsymbol{\theta},\mathbf{Z}]] \end{aligned}$$

Frequentist approach

$$\begin{aligned} E[\mathbf{Z}^*|\mathbf{Z}] \approx& E[\mathbf{Z}^*|\hat{\boldsymbol{\theta}},\mathbf{Z}] \\ \text{Var}[\mathbf{Z}^*|\mathbf{Z}] \approx& \text{Var}[\mathbf{Z}^*|\hat{\boldsymbol{\theta}},\mathbf{Z}] \end{aligned}$$

Multiple data sources

Assume several data sets $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$, all giving information about $\mathbf{Y}$, the process of interest.
$\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$ may have a complex dependence structure due to their common relation to $\mathbf{Y}$.
Hierarchical modelling:

$$[\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3 | \mathbf{Y}, \boldsymbol{\theta}_D] = [\mathbf{Z}_1 | \mathbf{Y}, \boldsymbol{\theta}_{D,1}] \times [\mathbf{Z}_2 | \mathbf{Y}, \boldsymbol{\theta}_{D,2}] \times [\mathbf{Z}_3 | \mathbf{Y}, \boldsymbol{\theta}_{D,3}]$$

i.e. *conditional independence*

Process model as a hierarchical model

- One time series: $Y_1, ..., Y_T$

  $$[Y_1, ..., Y_T] = [Y_1][Y_2|Y_1][Y_3|Y_2, Y_1] \cdots [Y_T|Y_{T-1}, ..., Y_1]$$

  Common assumption:

  $$[Y_t|Y_{t-1}, ..., Y_1] = [Y_t|Y_{t-1}]$$

  First order Markov assumption

  Why question: Do Y_{t-1} influence Y_t?

Process model as a hierarchical model

- One time series: $Y_1, ..., Y_T$
  $$[Y_1, ..., Y_T] = [Y_1][Y_2|Y_1][Y_3|Y_2, Y_1]\cdots[Y_T|Y_{T-1}, ..., Y_1]$$
  Common assumption:
  $$[Y_t|Y_{t-1}, ..., Y_1] = [Y_t|Y_{t-1}]$$
  First order Markov assumption
  Why question: Do Y_{t-1} influence Y_t?
- Two processes $\mathbf{X}, \mathbf{Y}$. Assumption:
  $$[\mathbf{X}] = \prod_{t=1}^{T}[X_t|X_{t-1}]$$
  $$[\mathbf{Y}|\mathbf{X}] = \prod_{t=1}^{T}[Y_t|Y_{t-1}, X_t, X_{t-1}]$$
  Why question: Do X_t and/or X_{t-1} influence Y_t?
  Easier to answer than: Do $\mathbf{X}$ influence $\mathbf{Y}$?

Model building - challenges

- Critic: Prior $[\boldsymbol{\theta}]$ is subjective
  Subjectivity in all kind of model specification
  If no prior knowledge: Non-informative/non-sensitive priors can be used

Model building - challenges

- Critic: Prior $[\theta]$ is subjective
  Subjectivity in all kind of model specification
  If no prior knowledge: Non-informative/non-sensitive priors can be used
- Computational demanding

Model building - challenges

- Critic: Prior $[\theta]$ is subjective
  Subjectivity in all kind of model specification
  If no prior knowledge: Non-informative/non-sensitive priors can be used
- Computational demanding
- Hierarchical modelling makes it possible to construct VERY complex models
  How to evaluate these models?
  - Model selection criteria: AIC, BIC, DIC ...
    - Predictability, parsimony
  - *Scientific interpretation*

Optimal prediction

Typical aim: Predict Y from Z (assumed now scalar)
Available: $[Y|\boldsymbol{\theta}], [Z|\boldsymbol{\theta}, Y]$ and observations Z.
Decision theory

- Assume $a(Z)$ is a prediction of Y
- Loss in prediction: $L(a(Z), Y)$
  - Example: $L(a(Z), Y) = (a(Z) - Y)^2$
- Aim: Find $a^*(Z)$ such that

$$E[L(a^*(Z), Y)|Z] \leq E[L(a(Z), Y)|Z] \text{ for all possible } a(\cdot)$$

  - Example: $a^*(Z) = E[Y|Z]$

Optimal prediction and binary Y

For $Y \in \{0,1\}$, $E[Y|Z] \in [0,1]$
Alternative:

- $a(Y) \in \{0,1\}$
- $L(a(Y), Y) = I(a(Y) = Y)$
- $a^*(Y) = max_{Y\in\{0,1\}}[Y|Z]$

Computation

Inference for space time-processes difficult due to

- Complex models
- Latent processes (likelihood not directly specified)
- Huge amounts of data

Computation

Inference for space time-processes difficult due to

- Complex models
- Latent processes (likelihood not directly specified)
- Huge amounts of data

Possibilities

- Use available software
  - Typically for specific models
- Monte Carlo metods
  - Preferred method in the book, described in sec 2.3
- Using *integrated nested Laplace approximation* (INLA)
  - Flexible software for latent Gaussian processes
  - Will be used throughout the course
- Computation within **R**