BML. 1 Intro

1.2 Means. Normal-Normal model

DataLab CSIC

Schedule

- Week 0. Props,
- Week 1. Means, Regression
- Week 2. DLMs, Recap.

Lab1-1

Expo-Gamma (Reliability)

Beta-binomial (Fraud detection)

Normal-normal (Ecology)

Normal-normal model. A typical example

Consider detecting fraud in fishing activities. Only sardines of size bigger than 5.5 cm are allowed to be captured by an EU regulation.

Inspectors catch the boat *Ayuso* right at the entrance of Cedeira port. We would like to assess whether the *Ayuso* crew has respected the EU regulation.

Numerical example in Lab. Here we just go through the concepts and methods.

Normal-normal model. A typical example

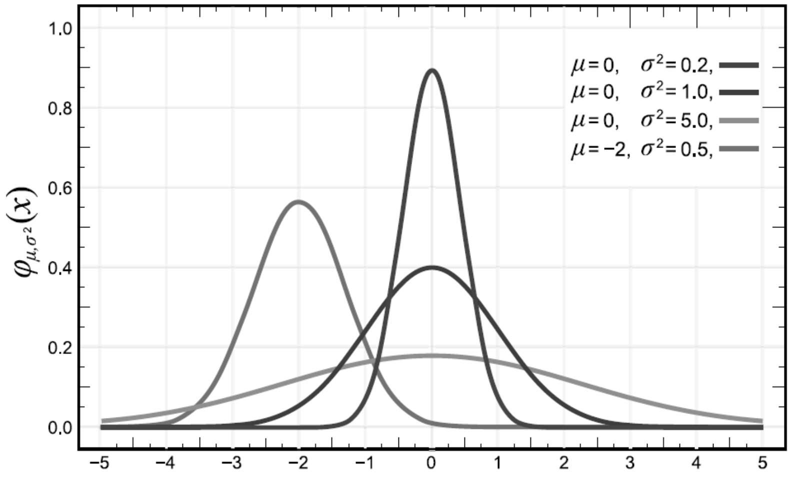
Consider detecting fraud in fishing activities. Only sardines of size bigger than 5.5 cm are allowed to be captured by an EU regulation.

Inspectors catch the boat *Ayuso* right at the entrance of Cedeira port. We would like to assess whether the *Ayuso* crew has respected the EU regulation.

Numerical example in Lab. Here we just go through the concepts and methods.

Let's start with the model

Check
https://en.wikipedia.org/wiki/Normal_distribution



Bayesian inference with the normal-normal model

Parameter θ

Model $X|\theta \sim N(\theta,\sigma^2)$ $\quad f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right)$ $\quad E(X)=\theta$, $Var(X)=\sigma^2 \longrightarrow$ KNOWN!!

Data $x_1, x_2, \dots, x_n$

Likelihood $f(x_1,x_2,\dots,x_n|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x_i-\theta}{\sigma}\right)^2\right) \propto \exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)\right) = \ell(\theta|\underline{x})$

(MLE) $h(\theta) = -\log \ell(\theta|x) = \frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)$ $\quad h'(\theta)=0 \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$

Bayesian inference with the normal-normal model

Likelihood

$$\ell(\theta) \propto \exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)\right)$$

Non-informative Prior

$$f(\theta) = 1 \longrightarrow \text{IMPROPER PRIOR !!!}$$

$$f(\theta|x_1,\dots,x_n) \propto \exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)\right) \times 1$$

$$= \exp\left(-\frac{1}{2}\left(\theta^2\frac{n}{\sigma^2} - 2\theta\frac{\sum x_i}{\sigma^2}\right)\right)$$

Posterior

$$\theta \sim 1 \longrightarrow N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

Sequential update

$$\mu_{n+1} = \frac{n\mu_n + x_{n+1}}{n+1} \qquad \sigma_{n+1} = \frac{n\sigma_n}{n+1}$$

Bayesian inference with the normal-normal model

Likelihood

$$l(\theta) \propto \exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)\right)$$

Prior

$$\theta \sim N(\theta_0, \sigma_0^2) \qquad f(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma_0}\right)^2\right) \propto \exp\left(-\frac{1}{2\sigma_0^2}\left(\theta^2 - 2\theta_0\theta\right)\right)$$

Posterior

$$f(\theta \mid x_1,\dots,x_n) = \frac{f(x_1,\dots,x_n \mid \theta)\, f(\theta)}{\int f(x_1,\dots,x_n \mid \theta)\, f(\theta)\, d\theta} \propto f(x_1,\dots,x_n \mid \theta)\, f(\theta)$$

$$\propto \exp\left(-\frac{1}{2\sigma_0^2}\left(\theta^2 - 2\theta_0\theta\right)\right)\exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum x_i\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\theta^2\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) - 2\theta\left(\frac{\sum x_i}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}\right)\right)\right)$$

Sequential update

$$\theta \sim N(\theta_0, \sigma_0^2) \longrightarrow \theta \mid x_1,\dots,x_n \sim N\left(\frac{\frac{\sum x_i}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}\right)$$

Bayesian inference with the normal-normal model (analytic and simulation approaches)

Point estimation

Posterior mean

$$\frac{\frac{\sum x_i}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} \frac{\sum x_i}{n} + \frac{\frac{1}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} \theta_0$$

Mix of prior and data

What if n grows??

$$n \to \infty \quad \approx \quad \bar{x}$$

Posterior median

SAME

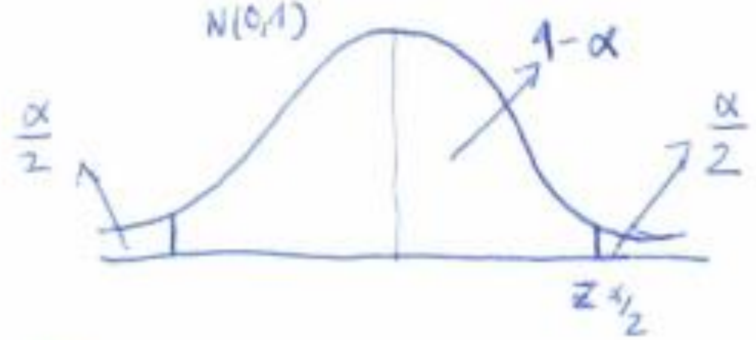
Posterior mode

SAME

Bayesian inference with the normal-normal model

Credible interval

Symmetric interval



$$\left[\frac{\frac{\sum x_i}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}}\right]$$

HPD

SAME

Bayesian inference with the normal-normal model

Hypothesis testing

$H_0: \theta \leq a$

$H_1: \theta > a$

EVIDENCE SUPPORTS H_0 IF

$$Pr(\theta \leq a \mid \underline{x}) \geq Pr(\theta > a \mid x)$$

$$\Updownarrow$$

$$Pr(\theta \leq a \mid x) \geq \frac{1}{2}$$

Testing two lateral hypothesis

		IS	
		$\theta \leq a$	$\theta > a$
SAYS	$\theta \leq a$	1	0
	$\theta > a$	c	1

What if utility is not 0-1??

$\theta \leq a \longrightarrow Pr(\theta \leq a \mid x)$

$\theta > a \longrightarrow c\, Pr(\theta \leq a \mid x) + Pr(\theta > a \mid x)$

SAY $\theta \leq a$ IFF $Pr(\theta \leq a \mid x) \geq \frac{1}{2-c}$

Bayesian inference with the normal-normal model

$$y \mid \theta \sim N(\theta, \sigma^2) \quad \theta \mid \underline{x} \sim N(\mu_1, \sigma_1^2)$$

Forecasting. The predictive distribution

$$f(y \mid \underline{x}) = \int f(y \mid \theta)\, f(\theta \mid \underline{x})\, d\theta$$

$$y \mid \underline{x} \sim N\left(\mu_1, \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}}\right)$$

Summarising the predictive distribution

$$E(y \mid \underline{x}) = \mu_1$$

$$Var(y \mid \underline{x}) = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}}$$

Pending issues

What if variance unknown?

Multivariate case?

Try them as exercises   after seeing 1.3

BML 1. Intro

1.3 Regression models

DataLab CSIC

In this deck

Linear regression

(Low dimension) Logistic regression

Lab 1.2

Linear regression (ornitology)

Logistic regression (bioassay)

Dynamic linear models (hydrology)

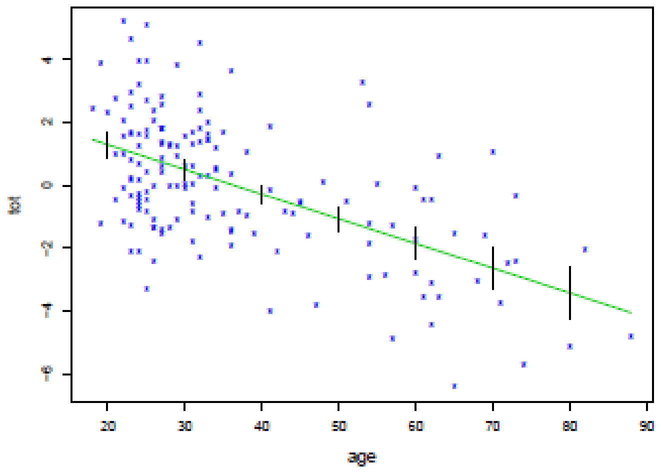
Linear regression model. A typical example

Consider a study of kidney function. The data represent (x=age of person, y=tot, a composite measure of the overall function). Kidney function declines with age and we need to provide additional information concerning decline rate. This is important in managing kidney transplant.

Numerical example in the Lab. Here we just go through the concepts and methods.

Let’s start with the model

Check
https://en.wikipedia.org/wiki/Simple_linear_regression



Linear regression model

Data structure. Response

$$y$$

Explanatory variables

$$x = (x_1, \dots, x_n)$$

Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad , \quad i = 1, \dots, n$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad \text{INDEPENDENT}$$

Likelihood

$$\theta = (\beta_0, \beta_1, \dots, \beta_k, \sigma)$$

$$p(\theta \mid \underline{x}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n \prod_{i=1}^{n} \exp\left(-\frac{1}{2} \left(\frac{y_i - \beta x_i}{\sigma} \right)^2 \right)$$

(MLE)

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$s^2 = \frac{1}{n-k} (y - X\hat{\beta})^T (y - X\hat{\beta})$$

Bayesian inference with linear regression model

Model

$$y \mid \beta, \sigma, X \sim N(X\beta, \sigma^2 I)$$

Standard noninformative prior

$$p(\beta, \sigma^2) \propto \sigma^{-2}$$

Posterior

$$p(\beta, \sigma^2 \mid y) = p(\beta \mid \sigma^2, y)\; p(\sigma^2 \mid y)$$

- $\beta \mid \sigma, y \sim N(\hat{\beta}, V_\beta \sigma^2)$ $\quad \hat{\beta} = V_\beta X^T y$, $\quad V_\beta = (X^T X)^{-1}$
- $p(\sigma^2 \mid y) = \dfrac{p(\beta, \sigma^2 \mid y)}{p(\beta \mid \sigma^2, y)} \sim \text{Inv-}\chi^2(n-k, s^2)$ $\quad s^2 = \dfrac{1}{(n-k)} (y - X\hat{\beta})^T (y - X\beta)$

Bayesian inference with linear regression model

Sampling from the posterior

COMPUTE V_β, $\hat{\beta}$, s^2

DRAW $\sigma^2 \sim \text{INV-}\chi^2(n-K, s^2)$

DRAW $\beta|\sigma \sim N(\hat{\beta}, V_\beta \sigma^2)$

COMPUTE $X = QR$ $\quad Q_{n \times K}$ ORTH. COLUMNS , $R_{K \times K}$ UPPER TRIANGULAR

COMPUTE $R^{-1} \rightarrow V_\beta = R^{-1}(R^{-1})^T$

COMPUTE $\hat{\beta}$: $R\hat{\beta} = Q^T y$

Bayesian inference with linear regression model

Hypothesis testing for coefficients

$\beta_i = 0$??

$$\beta_i \mid \sigma, y \sim N(\hat{\beta}_i, V_{\beta ii}\,\sigma^2)$$

$$\sigma^2 \mid y \sim \text{INV-}\chi^2(n-K, s^2)$$

```
FOR I = 1, 1000
    GENERATE σ_I^2 ~ σ^2 | y
    GENERATE β_I ~ β_i | σ_I, y
```

$\beta_{(I)} \leftarrow \text{SORT}(\beta_I)$

Check whether $0 \in [\beta_{(50)}, \beta_{(950)}]$

Bayesian inference with the linear regression model

Forecasting. The predictive distribution

$y \quad . \quad (\tilde{X}, \tilde{y}) \qquad f(\tilde{y} \mid \tilde{X}, y) = \iint f(\tilde{y} \mid \beta, \sigma, \tilde{X})\, f(\beta, \sigma \mid y)\, d\sigma\, d\beta$

GIVEN σ, y $\quad \tilde{y} \sim N\left(\tilde{X}\hat{\beta}, (I + \tilde{X} V_\beta \tilde{X})\sigma^2\right)$

GIVEN y $\quad \tilde{y} \sim t_{n-k}\left(\tilde{X}\hat{\beta}, s^2(I + \tilde{X} V_\beta \tilde{X})\right)$

Sampling from the predictive

GENERATE $\quad \beta, \sigma \sim f(\beta \mid \sigma, y)\, f(\sigma \mid y)$

GENERATE $\quad \tilde{y} \sim f(\tilde{y} \mid \beta, \sigma, \tilde{X})$

Bayesian inference with linear regression model

Model checking (Recall discussion in IntroML)

Examining plots of residuals againts explanatory variables,… interpretable as posterior predictive checks

Advantage: Using simulation, compute the posterior predictive distribution for any data summary

Example in Lab

Pending issues

Hierarchical models

Nonlinear regression (including NNs)

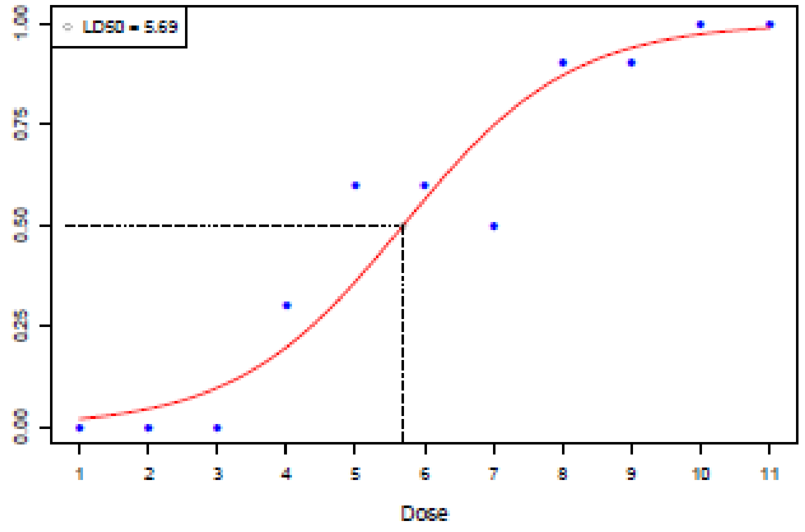
To be seen

Logistic regression. A typical example

A new anti-cancer drug is being developed. Before human testing can begin, animal studies are needed to determine safe dosages. A bioassay or dose-response experiment is carried out: 11 groups of 10 mice are treated with an increasing dose of drug and the proportion of deaths are observed.

Numerical example in the Lab. Here we just go through the concepts and methods.

Let's start with the model



Check
https://en.wikipedia.org/wiki/Logistic_regression

Bayesian inference with logitic regression

Data

$$\{x_i, n_i, y_i\} \quad i=1,\dots,k$$

x_i, EXP. VARIABLE

n_i, TRIALS

y_i, SUCCESSES.

Model

$$y_i \mid \theta_i \sim \mathrm{Bin}(n_i, \theta_i) \qquad \mathrm{logit}(\theta_i) = \alpha + \beta x_i \qquad \mathrm{logit}(\theta_i) = \log \frac{\theta_i}{1-\theta_i} \qquad \theta_i = \mathrm{logit}^{-1}(\alpha + \beta x_i)$$

Likelihood

$$\ell(\alpha, \beta \mid x, n, y) = \prod_{i=1}^{k} \left[\binom{n_i}{y_i} \mathrm{logit}^{-1}(\alpha+\beta x_i)^{y_i} \left(1 - \mathrm{logit}^{-1}(\alpha+\beta x_i)\right)^{n_i - y_i} \right]$$

(MLE)

$$\max_{\alpha, \beta} \sum_{i=1}^{k} y_i \left[\log\left(\mathrm{logit}^{-1}(\alpha+\beta x_i)\right) + (n_i - y_i) \log\left(1 - \mathrm{logit}^{-1}(\alpha+\beta x_i)\right) \right]$$

$$\longrightarrow \hat{\alpha}, \hat{\beta}$$

Bayesian inference with logistic regression

Likelihood

Generic prior

$$p(\alpha, \beta \mid y, n, \mathbf{x}) \propto p(\alpha, \beta) \prod_{i=1}^{k} \left[\left(\text{logit}^{-1}(\alpha + \beta x_i) \right)^{y_i} \left(1 - \text{logit}^{-1}(\alpha + \beta x_i) \right)^{n_i - y_i} \right]$$

Generic posterior

Bayesian inference with logistic regression

Sampling from the joint posterior. Piecewise constant approximation

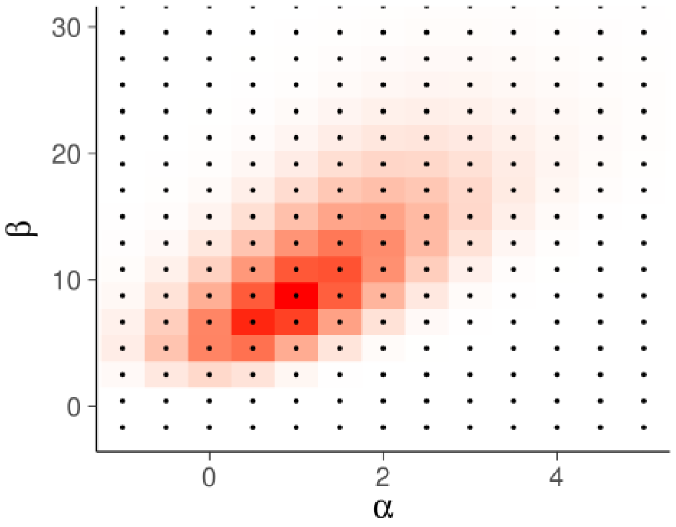
- Get some finite region where posterior is concentrated
- Choose grid over region
- Evaluate density over grid
- (Normalised) Density x Cell area approximates cell prob. in each cell

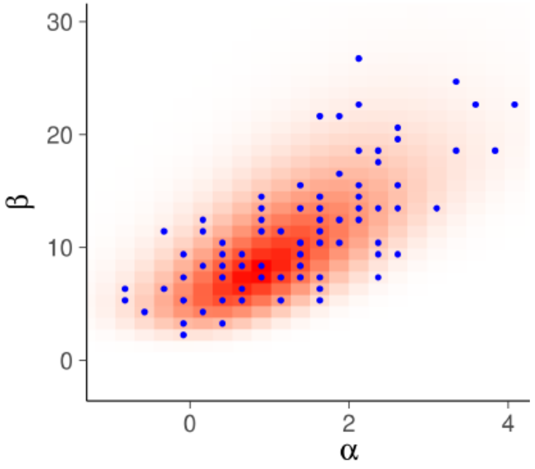
Sample according to grid cell probabilities

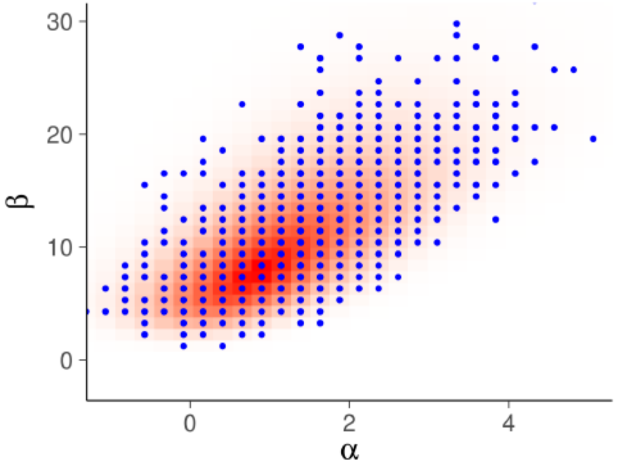
Jitter for visualisation

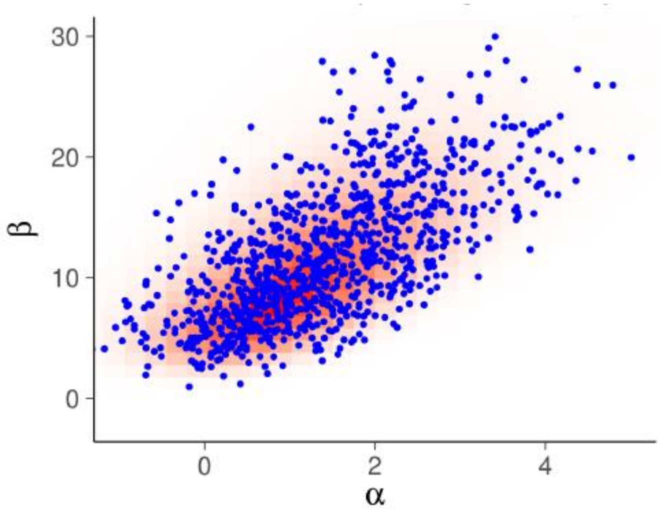
Bayesian inference with logistic regression

Sampling from the joint posterior









Bayesian inference with logistic regression

Posteriors of derived quantities. The case of LD50, dose level at which death probability is 50%

$$E\left(\frac{y}{n}\right) = \theta = \mathrm{logit}^{-1}(\alpha + \beta x) = 0.5$$

$$\alpha + \beta x = \mathrm{logit}(0.5) = 0$$

$$x_{LD50} = -\frac{\alpha}{\beta}$$

$$E(x_{LD50} \mid y, n, x) \simeq \frac{1}{S} \sum_{s=1}^{S} \frac{\alpha^{(s)}}{\beta^{(s)}}$$

$$(\alpha^{(s)}, \beta^{(s)})_{s=1}^{S} \sim p(\alpha, \beta \mid y, n, x)$$

Pending issues

Logistic multiple regression

Hierarchical models

Other GLMs (e.g., Probit)

To be seen

BML. 1 Intro

1.4 Dynamic Linear Models

DataLab CSIC

In this deck

Dynamic Linear Models

Lab 1.2

Linear regression (ornitology)

Logistic regression (bioassay)

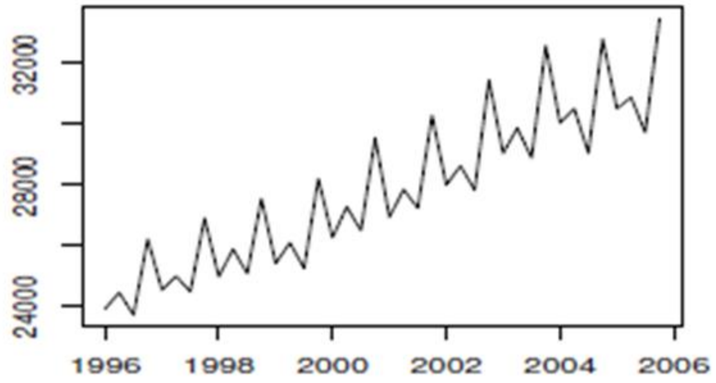
Dynamic linear models (hydrology)

Dynamic linear models. A typical example

We have available quarterly consumptions of gas in Spain. We need to forecast consumption over the next year to plan production.

Numerical example in the Lab. Here we just go through the concepts and methods.

Check e.g.
https://www.math.unm.edu/~ghuerta/tseries/dlmch2.pdf



Time series: three approaches

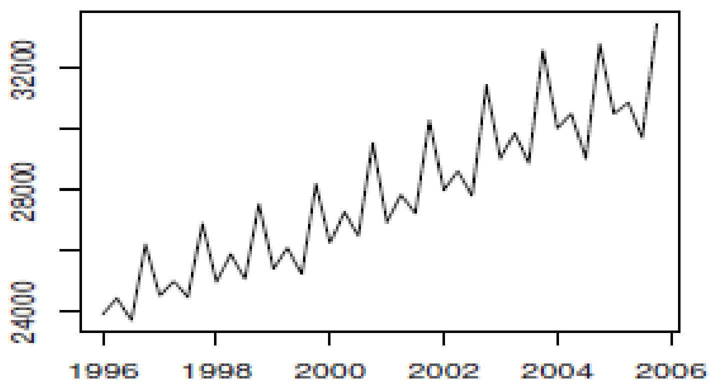
- Models in time domain: ARIMA. Box Jenkins
- Models in frequency domain. Spectral analysis.
- State space models. Series as output of dynamic systems subject to random perturbations.

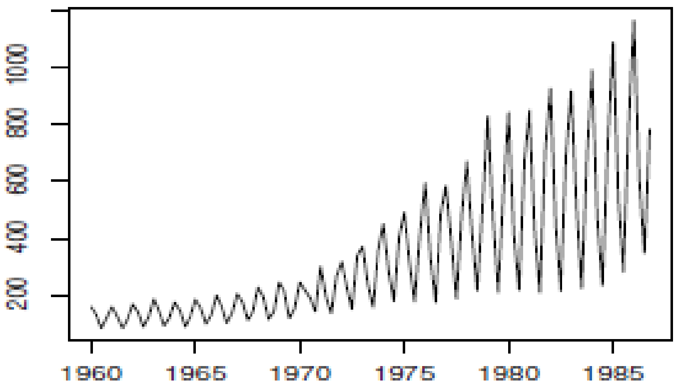
- Recurrent NNs, LSTMs, GRUs, Transformers,…

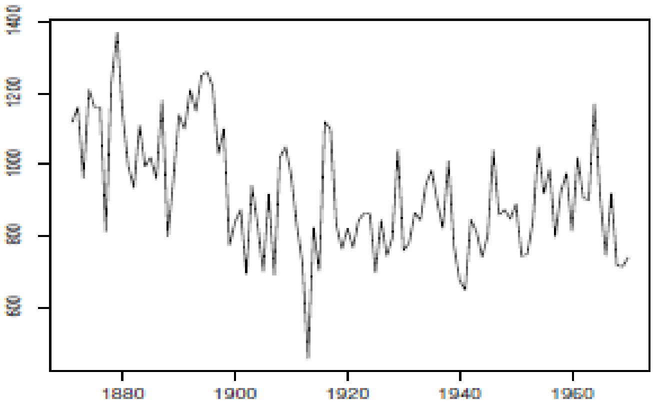
State space models: Advantages

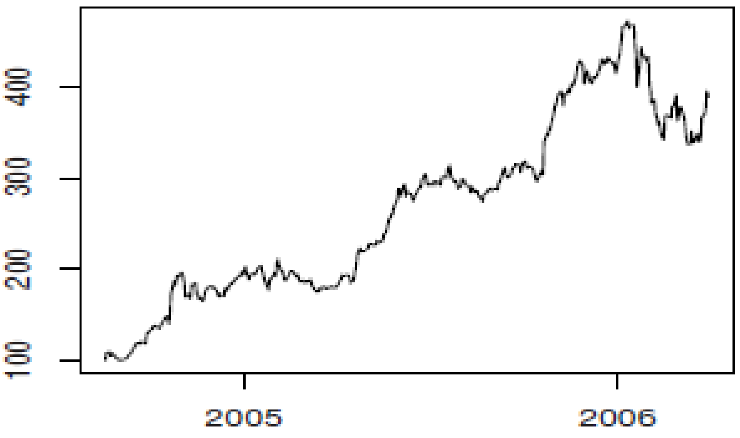
- Natural interpretation through combination of components: trend, seasonal, regression, autoregression
- Powerful probabilistic structure facilitating flexible modelling in numerous domains
- Recursive computation
- Bayesian treatment naturally
- Uni and multivariate
- Non stationarity, structural changes, irregular patterns naturally
- Linear (Kalman filter) and non-linear (MCMC).
- Very useful in applications

Time series features









State space models: simple example SLAM

Object position from measurements subject to random errors

$$\theta \qquad\qquad (Y_t : t = 1, 2, \ldots)$$

$$Y_t = \theta + \epsilon_t, \qquad \epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

$$Y_1, Y_2, \ldots | \theta \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2). \qquad\qquad \theta \sim \mathcal{N}(m_0, C_0)$$

$$\theta | y_{1:n} \sim \mathcal{N}(m_n, C_n),$$

$$m_n = \mathrm{E}(\theta | y_{1:n}) = \frac{C_0}{C_0 + \sigma^2/n} \bar{y} + \frac{\sigma^2/n}{C_0 + \sigma^2/n} m_0$$

$$C_n = \mathrm{Var}(\theta | y_{1:n}) = \left(\frac{n}{\sigma^2} + \frac{1}{C_0}\right)^{-1} = \frac{\sigma^2 C_0}{\sigma^2 + nC_0}.$$

Re
cu
rsi
ve
ly

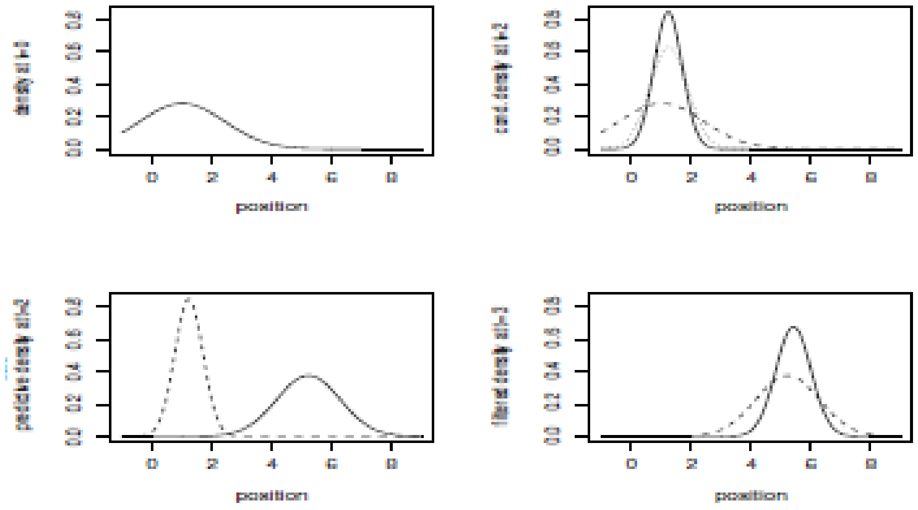
$$m_n = \frac{C_{n-1}}{C_{n-1} + \sigma^2} y_n + \left(1 - \frac{C_{n-1}}{C_{n-1} + \sigma^2}\right) m_{n-1}$$

$$= m_{n-1} + \frac{C_{n-1}}{C_{n-1} + \sigma^2} (y_n - m_{n-1})$$

$$C_n = \left(\frac{1}{\sigma^2} + \frac{1}{C_{n-1}}\right)^{-1} = \frac{\sigma^2 C_{n-1}}{\sigma^2 + C_{n-1}}.$$

State space models: Simple example

Time	Observation	Mean	Variance
0	-	1	2
1	1.3	1.24	0.4
2	1.2	1.222	0.222
3	……..	………….	…………..



Simple example

- Introduce dynamic component. From t=2, object adopts a velocity

With in $\theta_t = \theta_{t-1} + \nu + w_t, \qquad w_t \sim \mathcal{N}(0, \sigma_w^2)$ $\qquad Y_t = \theta_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$

- Initial step:

$$\theta_2 | y_{1:2} \sim \mathcal{N}(m_2 = 1.222, C_2 = 0.222)$$

- Prediction:

$$\theta_3 | y_{1:2} \sim \mathcal{N}(a_3, R_3) \qquad Y_3 | y_{1:2} \sim \mathcal{N}(f_3, Q_3)$$

$$a_3 = \mathrm{E}(\theta_2 + \nu + w_3 | y_{1:2}) = m_2 + \nu = 5.722 \qquad f_3 = \mathrm{E}(\theta_3 + \epsilon_3 | y_{1:2}) = a_3 = 5.722$$

$$R_3 = \mathrm{Var}(\theta_2 + \nu + w_3 | y_{1:2}) = C_2 + \sigma_w^2 = 1.122 \qquad Q_3 = \mathrm{Var}(\theta_3 + \epsilon_3 | y_{1:2}) = R_3 + \sigma^2 = 1.622$$

Simple example

- Initial step:

$$\theta_2|y_{1:2} \sim \mathcal{N}(m_2 = 1.222, C_2 = 0.222)$$

- Forecast:

$$\theta_3|y_{1:2} \sim \mathcal{N}(a_3, R_3) \qquad Y_3|y_{1:2} \sim \mathcal{N}(f_3, Q_3)$$

$$a_3 = \mathrm{E}(\theta_2 + \nu + w_3|y_{1:2}) = m_2 + \nu = 5.722 \qquad f_3 = \mathrm{E}(\theta_3 + \epsilon_3|y_{1:2}) = a_3 = 5.722$$

$$R_3 = \mathrm{Var}(\theta_2 + \nu + w_3|y_{1:2}) = C_2 + \sigma_w^2 = 1.122 \qquad Q_3 = \mathrm{Var}(\theta_3 + \epsilon_3|y_{1:2}) = R_3 + \sigma^2 = 1.622$$

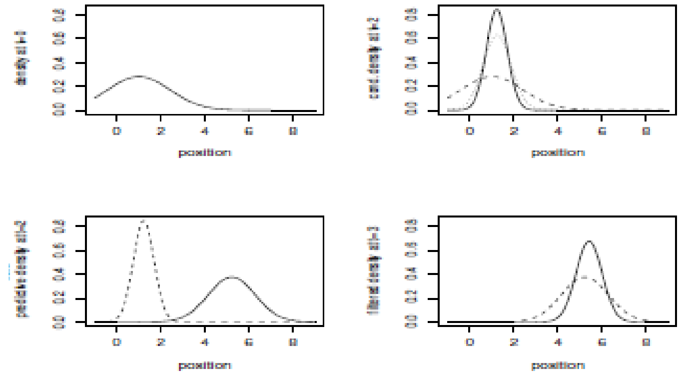
- Filter:

$$Y_3 = 5 \qquad e_t = y_t - f_t = -0.722$$

$$\theta_3|y_1, y_2, y_3 \sim \mathcal{N}(m_3, C_3)$$

$$m_3 = a_3 + \frac{R_3}{R_3 + \sigma^2}(y_3 - f_3) = 5.568$$

$$C_3 = \frac{\sigma^2 R_3}{\sigma^2 + R_3} = R_3 - \frac{R_3}{R_3 + \sigma^2} R_3 = 0.346$$



Lessons from introductory example

- Observable process determined by latent process (except for normal error)
- Latent process depends on previous state linearly (except for normal error)
- Forecast and estimation sequentially as data arrives

- Linearity and normality define dynamic linear models (DLMs)
- Time dependence structure, define state space models

State space models. Definition

- p-variate and m-variate time series (θ_t) Y_t such that
  - (θ_t) is Markov chain
  - Given (θ_t) Y_t are independent and just depends on θ_t

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \cdot \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j)$$

Dynamic linear models. Definition

- State space model with
  - Observation equation

$$Y_t = F_t\theta_t + v_t, \qquad v_t \sim \mathcal{N}_m(0, V_t)$$

  - State equation

$$\theta_t = G_t\theta_{t-1} + w_t, \qquad w_t \sim \mathcal{N}_p(0, W_t)$$

  - Prior

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0)$$

DLM: Random walk with noise AKA: Local level model

- State space with

$$
\begin{aligned}
Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\
\mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W)
\end{aligned}
$$

DLM: Linear growth model

- State space with

$$
\begin{aligned}
Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V), \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1}, & w_{t,1} &\sim \mathcal{N}(0, \sigma_\mu^2) \\
\beta_t &= \beta_{t-1} + w_{t,2}, & w_{t,2} &\sim \mathcal{N}(0, \sigma_\beta^2)
\end{aligned}
$$

Problems

Target $\pi(\theta_s | y_{1:t})$

Filtering $s = t$

Forecasting $s > t$

Smooth $s < t$

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

Estimation and Forecasting

- Predictive density for state one step ahead

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})\,\mathrm{d}\theta_{t-1}$$

- Predictive density

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})\,\mathrm{d}\theta_t$$

- Filtering density

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})}$$

Estimation and forecasting with DLM (Kalman filter)

- Currently, $\theta_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$
- State predictive density one step ahead, normal

$$a_t = \mathrm{E}(\theta_t|y_{1:t-1}) = G_t m_{t-1},$$
$$R_t = \mathrm{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t$$

- Predictive density, normal

$$f_t = \mathrm{E}(Y_t|y_{1:t-1}) = F_t a_t,$$
$$Q_t = \mathrm{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t$$

- Filtering density

$$m_t = \mathrm{E}(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t,$$
$$C_t = \mathrm{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t$$

$$e_t = Y_t - f_t$$

K steps ahead forecasts

- Looking ahead in the future

$$\begin{array}{ccccccc} \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots & \longrightarrow & \theta_{t+k} \\ | & & & & & & | \\ Y_{1:t} & & & & & & Y_{t+k} \end{array}$$

- State k steps ahead

$$\pi(\theta_{t+k}|y_{1:t}) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|y_{1:t})\,\mathrm{d}\theta_{t+k-1}$$

- Observation k steps ahead

$$\pi(y_{t+k}|y_{1:t}) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|y_{1:t})\,\mathrm{d}\theta_{t+k}$$

K steps ahead prediction with DLM

- State, normal with mean and variance

$$a_t(k) = G_{t+k} a_{t,k-1},$$
$$R_t(k) = G_{t+k} R_{t,k-1} G'_{t+k} + W_{t+k}$$

- Observation, normal with mean and variance

$$f_t(k) = F_{t+k} a_t(k),$$
$$Q_t(k) = F_{t+k} R_t(k) F'_{t+k} + V_t$$

Model validation

- Prediction errors

$$e_t = Y_t - \mathrm{E}(Y_t|y_{1:t-1}) = Y_t - f_t$$

$$\begin{aligned} e_t &= Y_t - F_t a_t = F_t\theta_t + v_t - F_t a_t \\ &= F_t(\theta_t - a_t) + v_t. \end{aligned}$$

- Properties facilitating validation
  - Expected value is 0
  - Observations from gaussian process
  - White noise gaussian process

$$\tilde{e}_t = e_t/\sqrt{Q_t}$$

Model specification. Superposition principle

- The sum of independent DLMs is a DLM

$$Y_t = Y_{1,t} + \cdots + Y_{h,t}$$

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t}, & v_{i,t} &\sim \mathcal{N}(0, V_{i,t}), \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + w_{i,t}, & w_{i,t} &\sim \mathcal{N}(0, W_{i,t}) \end{aligned}$$

results in

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, & v_t &\sim \mathcal{N}(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t), \end{aligned}$$

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{bmatrix} \qquad F_t = [F_{1,t}|\cdots|F_{h,t}] \qquad G_t = \begin{bmatrix} G_{1,t} & & \\ & \ddots & \\ & & G_{h,t} \end{bmatrix} \quad W_t = \begin{bmatrix} W_{1,t} & & \\ & \ddots & \\ & & W_{h,t} \end{bmatrix} \quad V_t = \sum_{i=1}^{j} V_{i,t}$$

Model building strategy based on blocks

Trend models (polynomial)

- Order n

$$\begin{cases} Y_t = \theta_{t,1} + v_t \\ \theta_{t,j} = \theta_{t-1,j} + \theta_{t-1,j+1} + w_{t,j} \quad j = 1, \ldots, n-1 \\ \theta_{t,n} = \theta_{t-1,n} + w_{t,n}. \end{cases}$$

$$F = (1, 0, \ldots, 0)$$

$$G = \begin{bmatrix} 1 & 1 & 0 & & \ldots & 0 \\ 0 & 1 & 1 & 0 & \ldots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & & \ldots & 0 & 1 & 1 \\ 0 & & \ldots & & 0 & 1 \end{bmatrix}$$

$$W = \mathrm{diag}(W_1, \ldots, W_n).$$

- Predictive function

$$f_t(k) = E(Y_{t+k}|y_{1:t}) = a_{t,0} + a_{t,1}k + \cdots + a_{t,n-1}k^{n-1}$$

- Consider cases n=1,2,3

Seasonal models

- State vector dimension s-1

$$F = (1, 0, \ldots, 0)$$

$$G = \begin{bmatrix} -1 & -1 & \ldots & -1 & -1 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 1 & 0 \end{bmatrix}.$$

(Dynamic) regression models

- From the standard regression model

$$Y_t = \beta_1 + \beta_2 x_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- To the dynamic regression model

$$Y_t = \beta_{1,t} + \beta_{2,t} x_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\begin{aligned} Y_t &= x_t' \theta_t + v_t, \qquad & v_t &\sim \mathcal{N}(0, \sigma_t^2) \\ \theta_t &= G_t \theta_{t-1} + w_t, \qquad & w_t &\sim \mathcal{N}_p(0, W_t) \end{aligned}$$

Modelling strategy

- Trend + Seasonal + Regression term+ Low order AR term (1)

θ_t

```
dlmModPoly(2) + dlmModSeas(4)
```

Some pending issues

- What if V, W unknown?

- What if non-normal?

- What if non-linear?

Some discussion in Lab1-2. A lot of discussion in later chapters