Bayesian Inference

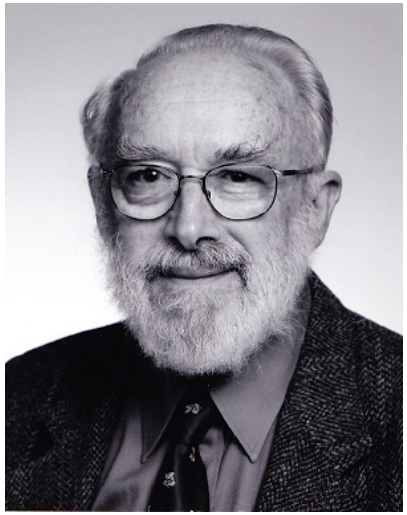
Chapter 4: Regression and Hierarchical Models

Conchi Ausín and Mike Wiper
Department of Statistics
Universidad Carlos III de Madrid

Master in Business Administration and Quantitative Methods
Master in Mathematical Engineering

Objective

AFM Smith



Dennis Lindley

We analyze the Bayesian approach to fitting normal and generalized linear models and introduce the Bayesian hierarchical modeling approach. Also, we study the modeling and forecasting of time series.

Contents

Normal linear models

A normal linear model is of the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is the observed data, $\mathbf{X}$ is a known $n \times k$ matrix, called the design matrix, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ is the parameter set and $\boldsymbol{\epsilon}$ follows a multivariate normal distribution. Usually, it is assumed that:

$$\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}_k, \frac{1}{\phi}\mathbf{I}_k\right).$$

- A simple example of normal linear model is the simple linear regression model where $\mathbf{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix}^T$ and $\boldsymbol{\theta} = (\alpha, \beta)^T$.

Normal linear models

Consider a normal linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$. A conjugate prior distribution is a normal-gamma distribution:

$$\boldsymbol{\theta} \mid \phi \sim \mathcal{N}\left(\mathbf{m}, \frac{1}{\phi}\mathbf{V}\right)$$
$$\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right).$$

Then, the posterior distribution given $\mathbf{y}$ is also a normal-gamma distribution with:

$$\begin{aligned}
\mathbf{m}^* &= \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}\left(\mathbf{X}^T\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right) \\
\mathbf{V}^* &= \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1} \\
a^* &= a + n \\
b^* &= b + \mathbf{y}^T\mathbf{y} + \mathbf{m}^T\mathbf{V}^{-1}\mathbf{m} - \mathbf{m}^{*T}\mathbf{V}^{*-1}\mathbf{m}^*
\end{aligned}$$

Normal linear models

The posterior mean is given by:

$$\begin{aligned} E[\theta \mid \mathbf{y}] &= \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}\left(\mathbf{X}^T\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right) \\ &= \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}\right) \\ &= \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}^{-1}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\theta}} + \mathbf{V}^{-1}\mathbf{m}\right) \end{aligned}$$

where $\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$ is the maximum likelihood estimator.

Thus, this expression may be interpreted as a weighted average of the prior estimator, $\mathbf{m}$, and the MLE, $\hat{\boldsymbol{\theta}}$, with weights proportional to precisions since, conditional on ϕ, the prior variance is $\frac{1}{\phi}\mathbf{V}$ and that the distribution of the MLE from the classical viewpoint is $\hat{\boldsymbol{\theta}} \mid \phi \sim \mathcal{N}\left(\boldsymbol{\theta}, \frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}\right)$

Normal linear models

Consider a normal linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, and assume the limiting prior distribution,

$$p(\boldsymbol{\theta}, \phi) \propto \frac{1}{\phi}.$$

Then, we have that,

$$\boldsymbol{\theta} \mid \mathbf{y}, \phi \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}, \frac{1}{\phi}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right),$$

$$\phi \mid \mathbf{y} \sim \mathcal{G}\left(\frac{n-k}{2}, \frac{\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\theta}}^T\left(\mathbf{X}^T\mathbf{X}\right)\hat{\boldsymbol{\theta}}}{2}\right).$$

Note that $\hat{\sigma}^2 = \frac{\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\theta}}^T\left(\mathbf{X}^T\mathbf{X}\right)\hat{\boldsymbol{\theta}}}{n-k}$ is the usual classical estimator of $\sigma^2 = \frac{1}{\phi}$.

In this case, Bayesian credible intervals, estimators etc. will coincide with their classical counterparts.

ANOVA model

The ANOVA model is an example of normal lineal model where:

$$y_{ij} = \theta_i + \epsilon_{ij},$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \frac{1}{\phi})$, for $i = 1, \ldots, k$, and $j = 1, \ldots, n_i$.

Thus, the parameters are $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, the observed data are $\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{k1}, \ldots, y_{kn_k})^T$, the design matrix is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 1_{n_1} & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1_{n_2} & & 0 \\ \vdots & & \vdots & \vdots \\ 0 & 0 & & 1 \end{pmatrix}$$

ANOVA model

- Assume conditionally independent normal priors, $\theta_i \sim \mathcal{N}\left(m_i, \frac{1}{\alpha_i \phi}\right)$, for $i = 1, \ldots, k$, and a gamma prior $\phi \sim \mathcal{G}(\frac{a}{2}, \frac{b}{2})$.
- This corresponds to a normal-gamma prior distribution for $(\boldsymbol{\theta}, \phi)$ where $\mathbf{m} = (m_1, \ldots, m_k)$ and $\mathbf{V} = diag\left(\frac{1}{\alpha_1}, \ldots, \frac{1}{\alpha_k}\right)$.
- Then, it is obtained that,

$$\boldsymbol{\theta} \mid \mathbf{y}, \phi \sim \mathcal{N}\left(\begin{pmatrix} \frac{n_1 \bar{y}_{1\cdot} + \alpha_1 m_1}{n_1 + \alpha_1} \\ \vdots \\ \frac{n_1 \bar{y}_{1\cdot} + \alpha_1 m_1}{n_1 + \alpha_1} \end{pmatrix}, \frac{1}{\phi} \begin{pmatrix} \frac{1}{\alpha_1 + n_1} & & \\ & \ddots & \\ & & \frac{1}{\alpha_k + n_k} \end{pmatrix}\right)$$

and

$$\phi \mid \mathbf{y} \sim \mathcal{G}\left(\frac{a+n}{2}, \frac{b + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 + \sum_{i=1}^{k} \frac{n_i}{n_i + \alpha_i} (\bar{y}_{i\cdot} - m_i)^2}{2}\right)$$

ANOVA model

- If we assume alternatively the reference prior, $p(\boldsymbol{\theta}, \phi) \propto \frac{1}{\phi}$, we have:

$$\boldsymbol{\theta} \mid \mathbf{y}, \phi \sim \mathcal{N}\left(\begin{pmatrix} \bar{y}_{1\cdot} \\ \vdots \\ \bar{y}_{k\cdot} \end{pmatrix}, \frac{1}{\phi}\begin{pmatrix} \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_k} \end{pmatrix}\right),$$

$$\phi \sim \mathcal{G}\left(\frac{n-k}{2}, \frac{(n-k)\,\hat{\sigma}^2}{2}\right),$$

  where $\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{k}\left(y_{ij} - \bar{y}_{i\cdot}\right)^2$ is the classical variance estimate for this problem.

- A 95% posterior interval for $\theta_1 - \theta_2$ is given by:

$$\bar{y}_{1\cdot} - \bar{y}_{2\cdot} \pm \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{n-k}\left(0.975\right),$$

  which is equal to the usual, classical interval.

Example: ANOVA model

- Suppose that an ecologist is interested in analysing how the masses of starlings (a type of birds) vary between four locations.
- A sample data of the weights of 10 starlings from each of the four locations can be downloaded from:
  `http://arcue.botany.unimelb.edu.au/bayescode.html`.
- Assume a Bayesian one-way ANOVA model for these data where a different mean is considered for each location and the variation in mass between different birds is described by a normal distribution with a common variance.
- Compare the results with those obtained with classical methods.

Simple linear regression model

Another example of normal linear model is the simple regression model:

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

for $i = 1, \ldots, n$, where $\epsilon_i \sim \mathcal{N}\left(0, \frac{1}{\phi}\right)$.

Suppose that we use the limiting prior:

$$p(\alpha, \beta, \phi) \propto \frac{1}{\phi}.$$

Simple linear regression model

Then, we have that:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \Big| \mathbf{y}, \phi \sim \mathcal{N}\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \frac{1}{\phi n s_x} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \right)$$

$$\phi \mid \mathbf{y} \sim \mathcal{G}\left(\frac{n-2}{2}, \frac{s_y\left(1-r^2\right)}{2} \right)$$

where:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \qquad \hat{\beta} = \frac{s_{xy}}{s_x},$$

$$s_x = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2, \qquad s_y = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2,$$

$$s_{xy} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right), \qquad r = \frac{s_{xy}}{\sqrt{s_x s_y}}, \qquad \hat{\sigma}^2 = \frac{s_y\left(1-r^2\right)}{n-2}.$$

Simple linear regression model

Thus, the marginal distributions of α and β are Student-t distributions:

$$\frac{\alpha - \hat{\alpha}}{\sqrt{\frac{\hat{\sigma}^2}{n}\frac{\sum_{i=1}^{n} x_i^2}{s_x}}} \mid \mathbf{y} \sim t_{n-2}$$

$$\frac{\beta - \hat{\beta}}{\sqrt{\frac{\hat{\sigma}^2}{s_x}}} \mid \mathbf{y} \sim t_{n-2}$$

Therefore, for example, a 95% credible interval for β is given by:

$$\hat{\beta} \pm \frac{\hat{\sigma}}{\sqrt{s_x}} t_{n-2}(0.975)$$

equal to the usual classical interval.

Simple linear regression model

Suppose now that we wish to predict a future observation:

$$y_{new} = \alpha + \beta x_{new} + \epsilon_{new}.$$

Note that,

$$\begin{aligned} E\left[y_{new} \mid \phi, \mathbf{y}\right] &= \hat{\alpha} + \hat{\beta} x_{new} \\ V\left[y_{new} \mid \phi, \mathbf{y}\right] &= \frac{1}{\phi}\left(\frac{\sum_{i=1}^{n} x_i^2 + n x_{new}^2 - 2n\bar{x}x_{new}}{n s_x} + 1\right) \\ &= \frac{1}{\phi}\left(\frac{s_x + n\bar{x}^2 + n x_{new}^2 - 2n\bar{x}x_{new}}{n s_x} + 1\right) \end{aligned}$$

Therefore,

$$y_{new} \mid \phi, \mathbf{y} \sim \mathcal{N}\left(\hat{\alpha} + \hat{\beta} x_{new}, \frac{1}{\phi}\left(\frac{\left(\bar{x} - x_{new}\right)^2}{s_x} + \frac{1}{n} + 1\right)\right)$$

Simple linear regression model

And then,

$$\frac{y_{new} - \hat{\alpha} + \hat{\beta} x_{new}}{\hat{\sigma}\sqrt{\left(\frac{(\bar{x}-x_{new})^2}{s_x} + \frac{1}{n} + 1\right)}} \mid \mathbf{y} \sim t_{n-2}$$

leading to the following 95% credible interval for y_{new} :

$$\hat{\alpha} + \hat{\beta} x_{new} \pm \hat{\sigma}\sqrt{\left(\frac{(\bar{x} - x_{new})^2}{s_x} + \frac{1}{n} + 1\right) t_{n-2}(0.975)},$$

which coincides with the usual, classical interval.

Example: Simple linear regression model

- Consider the data file `prostate.data` that can be downloaded from: `http://statweb.stanford.edu/~tibs/ElemStatLearn/`.
- This includes, among other clinical measures, the level of prostate specific antigen in logs (`lpsa`) and the log cancer volume (`lcavol`) in 97 men who were about to receive a radical prostatectomy.
- Use a Bayesian linear regression model to predict the `lpsa` in terms of the `lcavol`.
- Compare the results with a classical linear regression fit.

Generalized linear models

The generalized linear model generalizes the normal linear model by allowing the possibility of non-normal error distributions and by allowing for a non-linear relationship between **y** and **x**.

A generalized linear model is specified by two functions:

1. A conditional, exponential family density function of y given **x**, parameterized by a mean parameter, $\mu = \mu(\mathbf{x}) = E[Y \mid \mathbf{x}]$ and (possibly) a dispersion parameter, $\phi > 0$, that is independent of **x**.

2. A (one-to-one) link function, $g(\cdot)$, which relates the mean, $\mu = \mu(\mathbf{x})$ to the covariate vector, **x**, as $g(\mu) = \mathbf{x}\boldsymbol{\theta}$.

Generalized linear models

The following are generalized linear models with the canonical link function which is the natural parameterization to leave the exponential family distribution in canonical form.

- A logistic regression is often used for predicting the occurrence of an event given covariates:

$$Y_i \mid p_i \sim \mathcal{B}in(n_i, p_i)$$
$$\log \frac{p_i}{1-p_i} = \mathbf{x}_i\boldsymbol{\theta}$$

- A Poisson regression is used for predicting the number of events in a time period given covariates:

$$Y_i \mid p_i \sim \mathcal{P}(\lambda_i)$$
$$\log \lambda_i = \mathbf{x}_i\boldsymbol{\theta}$$

Generalized linear models

- The Bayesian specification of a GLM is completed by defining (typically normal or normal gamma) prior distributions $p(\boldsymbol{\theta}, \phi)$ over the unknown model parameters.
- As with standard linear models, when improper priors are used, it is then important to check that these lead to valid posterior distributions.
- Clearly, these models will not have conjugate posterior distributions, but, usually, they are easily handled by Gibbs sampling.
- In particular, the posterior distributions from these models are usually log concave and are thus easily sampled via adaptive rejection sampling.

Example: A logistic regression model

- The O-Ring data consist of 23 observations on Pre-Challenger Space Shuttle Launches
- On each launch, it is observed whether there is at least one O-ring failure, and the temperature at launch
- The goal is to model the probability of at least one O-ring failure as a function of temperature.
- Temperatures were 53, 57, 58, 63, 66, 67, 67, 67, 68, 69,70, 70, 70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 81
- Failures occurred at 53, 57, 58, 63, 70, 70, 75

Example: A logistic regression model

The table shows the relationship, for 64 infants, between gestational age of the infant (in weeks) at the time of birth (x) and whether the infant was breast feeding at the time of release from hospital (y).

x	28	29	30	31	32	33
$\#\{y=0\}$	4	3	2	2	4	1
$\#\{y=1\}$	2	2	7	7	16	14

Let x_i represent the gestational age and ni the number of infants with this age. Then we can model the probability that y_i infants were breast feeding at time of release from hospital via a standard binomial regression model.

Hierarchical models

- Suppose we have data, $\mathbf{x}$, and a likelihood function $f(\mathbf{x} \mid \boldsymbol{\theta})$ where the parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ are judged to be exchangeable, that is, any permutation of them has the same distribution.
- In this situation, it makes sense to consider a multilevel modeling assuming a prior distribution, $f(\boldsymbol{\theta} \mid \phi)$, which depends upon a further, unknown hyperparameter, ϕ, and use a hyperprior distribution, $f(\phi)$.
- In theory, this process could continue further, using hyperhyperprior distributions to estimate the hyperprior distributions. This is a method to elicit the optimal prior distributions.
- One alternative is to estimate the hyperparameter using classical methods, which is known as empirical Bayes. A point estimate $\hat{\phi}$ is then obtained to approximate the posterior distribution. However, the uncertainty in ϕ is ignored.

Hierarchical models

- In most hierarchical models, the joint posterior distributions will not be analytically tractable as it will be,

$$f(\boldsymbol{\theta}, \phi \mid \mathbf{x}) \propto f(\mathbf{x} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \phi) f(\phi)$$

- However, often a Gibbs sampling approach can be implemented by sampling from the conditional posterior distributions:

$$f(\boldsymbol{\theta} \mid \mathbf{x}, \phi) \propto f(\mathbf{x} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \phi)$$
$$f(\phi \mid \mathbf{x}, \boldsymbol{\theta}) \propto f(\boldsymbol{\theta} \mid \phi) f(\phi)$$

- It is important to check the propriety of the posterior distribution when improper hyperprior distributions are used. An alternative (as in for example Winbugs) is to use proper but high variance hyperprior distributions.

Hierarchical models

For example, a hierachical normal linear model is given by:

$$x_{ij} \mid \theta_i, \phi \sim \mathcal{N}\left(\theta_i, \frac{1}{\phi}\right), \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m.$$

Assuming that the means, θ_i, are exchangeable, we may consider the following prior distribution:

$$\theta_i \mid \mu, \psi \sim \mathcal{N}\left(\mu, \frac{1}{\psi}\right),$$

where the hyperparameters are μ y ψ.

Example: A hierarchical one-way ANOVA

Suppose that 5 individuals take 3 different IQ test developed by 3 different psychologists obtaining the following results:

	1	2	3	4	5
Test 1	106	121	159	95	78
Test 2	108	113	158	91	80
Test 3	98	115	169	93	77

Then, we can assume that:

$$X_{ij} \mid \theta_i, \phi \sim \mathcal{N}\left(\theta_i, \frac{1}{\phi}\right),$$

$$\theta_i \mid \mu, \psi \sim \mathcal{N}\left(\mu, \frac{1}{\psi}\right),$$

for $i = 1, \ldots, 5$, and $j = 1, 2, 3$, where θ_i represents the true IQ of subject i and μ the mean true IQ in the population.

Example: A hierarchical Poisson model

The number of failures, X_i at a power plant i is assumed to follow a Poisson distribution:

$$X_i \mid \lambda_i \sim \mathcal{P}(\lambda_i t_i), \qquad \text{para } i = 1, \ldots, 10,$$

where λ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in 1000s of hours). It seems natural to assume that the failure rates are exchangeable and thus we might assume:

$$\lambda_i \mid \gamma \sim \mathcal{E}(\gamma),$$

where γ is the prior hyperparameter. The observed data are:

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.5	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

Dynamic models

The univariate normal dynamic linear model (DLM) is:

$$y_t = \mathbf{F}_t \boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathcal{N}(0, V_t)$$
$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{W}_t).$$

These models are linear state space models, where $x_t = \mathbf{F}_t \boldsymbol{\theta}_t$ represents the signal, $\boldsymbol{\theta}_t$ is the state vector, $\mathbf{F}_t$ is a regression vector and G_t is a state matrix.

The usual features of a time series such as trend and seasonality can be modeled within this format.

If the matrices $\mathbf{F}_t$, $\mathbf{G}_t$, V_t and $\mathbf{W}_t$ are constants, the model is said to be time invariant.

Dynamic models

One of the simplest DLMs is the random walk plus noise model, also called first order polynomial model. It is used to model univariate observations and the state vector is unidimensional:

$$y_t = \theta_t + \nu_t, \qquad \nu_t \sim \mathcal{N}(0, V_t)$$
$$\theta_t = \theta_{t-1} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, W_t).$$

This is a slowly varying level model where the observations fluctuate around a mean which varies according to a random walk.

Assuming known variances, V_t and W_t, a straightforward Bayesian analysis can be carried out as follows.

Dynamic models

Suppose that the information at time $t-1$ is $\mathbf{y}^{t-1} = \{y_1, y_2, ..., y_{t-1}\}$ and assume that:

$$\theta_{t-1} \mid \mathbf{y}^{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1}).$$

Then, we have that:

- The prior distribution for θ_t is:

$$\theta_t \mid \mathbf{y}^{t-1} \sim \mathcal{N}(m_{t-1}, R_t)$$

  where $R_t = C_{t-1} + W_t$
- The one step ahead predictive distribution for y_t is:

$$y_t \mid \mathbf{y}^{t-1} \sim \mathcal{N}(m_{t-1}, Q_t)$$

  where $Q_t = R_t + V_t$.

Dynamic models

- The joint distribution of θ_t and y_t is:

$$\begin{pmatrix} \theta_t \\ y_t \end{pmatrix} \mid \mathbf{y}^{t-1} \sim N\left(\begin{pmatrix} m_{t-1} \\ m_{t-1} \end{pmatrix}, \begin{pmatrix} R_t & R_t \\ R_t & Q_t \end{pmatrix} \right)$$

- The posterior distribution for θ_t given $\mathbf{y}^t = \left\{\mathbf{y}^{t-1}, y_t\right\}$ is:

$$\begin{aligned} \theta_t \mid \mathbf{y}^t &\sim N(m_t, C_t), \qquad \text{where} \\ m_t &= m_{t-1} + A_t e_t, \\ A_t &= R_t / Q_t, \\ e_t &= y_t - m_{t-1}, \\ C_t &= R_t - A_t^2 Q_t. \end{aligned}$$

Note that e_t is simply a prediction error term. The posterior mean formula could also be written as:

$$m_t = (1 - A_t)\, m_{t-1} + A_t y_t.$$

Example: First order polynomial DLM

Assume a slowly varying level model for the water level in Lake Huron with known variances: $V_t = 1$ and $W_t = 1$.

1. Estimate the filtered values of the state vector based on the observations up to time t from $f(\theta_t \mid \mathbf{y}^t)$.
2. Estimate the predicted values of the state vector based on the observations up to time $t-1$ from $f(\theta_t \mid \mathbf{y}^{t-1})$.
3. Estimate the predicted values of the signal based on the observations up to time $t-1$ from $f(y_t \mid \mathbf{y}^{t-1})$.
4. Compare the results using e.g:
   - $V_t = 10$ and $W_t = 1$.
   - $V_t = 1$ and $W_t = 10$.

Dynamic models

- When the variances are not known, the Bayesian inference for the system is more complex.
- One possibility is the use of MCMC algorithms which are usually based on the so-called forward filtering backward sampling algorithm.
  1. The forward filtering step is the standard normal linear analysis to give $f(\theta_t|\mathbf{y}^t)$ at each t, for $t = 1, \ldots, T$.
  2. The backward sampling step uses the Markov property and samples θ_T^* from $f(\theta_T|\mathbf{y}^T)$ and then, for $t = T-1, \ldots, 1$, samples from $f(\theta_t \mid \mathbf{y}^t, \theta_{t+1}^*)$

  Thus, a sample from the posterior parameter structure is generated.
- However, MCMC may be computationally very expensive for on-line estimation. One possible alternative is the use of particle filters.

Dynamic models

Other examples of DLM are the following:

- A dynamic linear regression model is given by:

$$y_t = \mathbf{F}_t \boldsymbol{\theta}_t + \nu_t, \qquad \nu_t \sim \mathcal{N}(0, V_t)$$
$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{W}_t).$$

- The AR(p) model with time-varying coefficients takes the form:

$$y_t = \theta_{0t} + \theta_{1t} y_{t-1} + \ldots + \theta_{pt} y_{t-p} + \nu_t, \theta_{it} \qquad = \theta_{i,t-1} + \omega_{it},$$

  This model can be expressed in state space form by setting $\boldsymbol{\theta} = (\theta_{0t}, \ldots, \theta_{pt})$ and $F = (1, y_{t-1}, \ldots, y_t - p)$.

Dynamic models

The additive structure of the DLMs makes it easy to think of observed series as originating form the sum of different components,e.g.,

$$y_t = y_{1t} + \ldots, y_{h,t}$$

where y_{1t} might represent a trend component, y_{2t} a seasonal component, and so on. Then, each component, y_{it}, might be described by a different DLM:

$$\begin{aligned} y_t &= \mathbf{F}_{it}\boldsymbol{\theta}_{it} + \nu_{it}, \qquad & \nu_{it} &\sim \mathcal{N}(0, V_{it}) \\ \boldsymbol{\theta}_{it} &= \mathbf{G}_{it}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_{it}, \qquad & \boldsymbol{\omega}_{it} &\sim \mathcal{N}(0, \mathbf{W}_{it}). \end{aligned}$$

By the assumption of independence of the components, y_t is also a DLM described by:

$$\mathbf{F}_t = (\mathbf{F}_{1t} \,|\ldots|\, \mathbf{F}_{ht}), \quad V_t = V_{1t} + \ldots + V_{ht},$$

and

$$\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & & \\ & \ddots & \\ & & \mathbf{G}_{ht} \end{pmatrix}, \quad \mathbf{W}_t = \begin{pmatrix} \mathbf{W}_{1t} & & \\ & \ddots & \\ & & \mathbf{W}_{ht} \end{pmatrix}.$$