Modelling Non-linear and Non-stationary Time Series

Chapter 7(extra):
Generalized State Space Models

Henrik Madsen

Lecture Notes

September 2016

Generalized State Space Models

- Consider the sequence of state variables:

$$\{X_t; t \geq 1\}$$

- And the sequence of observations:

$$\{Y_t; t \geq 1\}$$

- Only the one-dimensional case considered here.

Parameter-driven models

Here, the model evolve independently of the past history of the observation process.
Let the history of $\{X_t\}$ and $\{X_t\}$ be denoted $\mathcal{X}^{(t-1)}$ and $\mathcal{Y}^{(t-1)}$:

$$\mathcal{X}^{(t-1)} = (X_{t-1}, X_{t-2}, \ldots)$$
$$\mathcal{Y}^{(t-1)} = (Y_{t-1}, Y_{t-2}, \ldots)$$

Assumption: Y_t given $(X_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ is independent of $(\mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ with conditional probability density:

$$p\left(y_t|x_t\right) := p\left(y_t \mid x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}\right) \qquad \text{(obs. eq.)}$$

And similarly

$$p\left(x_{t+1}|x_t\right) := p\left(x_{t+1}|x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}\right) \qquad \text{(state eq.)}$$

The joint density of observation and state variables:

$$
\begin{aligned}
& p\left(y_1, \ldots, y_n, x_1, \ldots, x_n\right) \\
& \qquad = p\left(y_n | x_n, \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}\right) \cdot p\left(x_n, \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}\right) \\
& \qquad = p\left(y_n | x_n\right) \cdot p\left(x_n | \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}\right) \cdot p\left(\mathcal{Y}^{(n-1)}, \mathcal{X}^{(n-1)} \right. \\
& \qquad = \ldots \\
& \qquad = \prod_{j=1}^{n} p\left(y_j | x_j\right) \left(\prod_{j=2}^{n} p\left(x_j | x_{j-1}\right) \right) p\left(x_1\right)
\end{aligned}
$$

where $p\left(x_1\right)$ denotes the probability density of the initial state.

- Notice that $\{X_t\}$ is a first order Markov chain.
- It is also clearly seen that

$$p\left(y_1,\ldots,y_n|x_1,\ldots,x_n\right) = \prod_{j=1}^{n} p\left(y_i|x_j\right)$$

- that is, the dependence structure of $\{Y_t\}$ is inherited from that of $\{X_t\}$.
- The sequence $\{X_t\}$ is often referred to as the hidden or latent process.

Filtering and prediction

We want to determine $p\left(x_t|\mathcal{Y}^{(t)}\right)$ and $p\left(x_t|\mathcal{Y}^{(t-1)}\right)$. An application of Bayes' theorem yields

$$p\left(x_t|\mathcal{Y}^{(t)}\right) = \frac{p\left(y_t|x_t\right) p\left(x_t|\mathcal{Y}^{(t-1)}\right)}{p\left(y_t|\mathcal{Y}^{(t-1)}\right)}$$

and

$$p\left(x_{t+1}|\mathcal{Y}^{(t)}\right) = \int p\left(x_t|\mathcal{Y}^{(t)}\right) p\left(x_{t+1}|x_t\right)\ \mathrm{d}\mu(x_t)$$

where $\mathrm{d}\mu(x_t) = \mathrm{d}x$ in the continuous case.
In the discrete case, a summation is used.

Prediction of future observations

$$p\left(y_{t+1}|\mathcal{Y}^{(t)}\right) = \int p\left(y_{t+1}|x_{t+1}\right) p\left(x_{t+1}|\mathcal{Y}^{(t)}\right)\, \mathrm{d}\mu(x_{t+1})$$

Is built on a Bayesian interpretation of $\mathcal{X}^{(t)}$ being the parameter. In

$$P(y_1, \ldots, y_n, x_1, \ldots, x_n) = \prod_{j=1}^{n} p\left(y_j|x_j\right) \left(\prod_{j=2}^{n} p\left(x_j|x_{j-1}\right)\right) p\left(x_1\right)$$

the left side is the aposteriori distribution, and the right side is a multiplication of the likelihood and the prior.

A non-Gaussian example

- In general the recursions present serious computational problems.
- Consider a very simple example, with the state equation:

$$X_t = aX_{t-1} \qquad (*)$$

- The observation density is assumed to be

$$p\left(y_t|x_t\right) = \frac{(\pi x_t)^{y_t} e^{-\pi x_t}}{y_t!}, \quad y_t = 0, 1, 2, \ldots$$

  where $\pi \in]0, 1]$ is constant.
- (*) implies that

$$p(x_{t+1}|x_t) = \begin{cases} 1 & \text{if } x_{t+1} = ax_t, \\ 0 & \text{otherwise} \end{cases}$$

- Assume further that X_1 is gamma-distributed:

$$p_1(x_1) = g(x_1; \alpha, \lambda) = \frac{\lambda^\alpha x_1^{\alpha-1} e^{-\lambda x_1}}{\Gamma(\alpha)} \quad (x_1 > c)$$

- This could be a simple model for the evolution of the number, X_t, of individuals at time t infected with a rare disease in which X_t is treated as a continuous variable.
- The observation Y_t represents the number of infected individuals observed in a random sample consisting of a small fraction π of the population a time, t.
- Now

$$\begin{aligned}
p\left(x_1|y_1\right) &= p\left(y_1|x_1\right)p\left(x_1\right)/p\left(y_1\right) \\
&= \left(\frac{(\pi x_1)^{y_1}e^{-\pi x_1}}{y_1!}\right)\left(\frac{\lambda^\alpha x_1^{\alpha-1}e^{-\lambda x_1}}{\Gamma(\alpha)}\right)\frac{1}{p\left(y_1\right)} \\
&= c(y_1)x_1^{\alpha+y_1-1}e^{(\pi+\lambda)x_1} \quad (x_1 > 0) \\
&= g(x_1;\underbrace{\alpha_1}_{\alpha+y_1},\underbrace{\lambda_1}_{\pi+\lambda}) \quad (x_1 \text{ again gamma distributed})
\end{aligned}$$

- The prediction density becomes

$$p(x_2|\mathcal{Y}^{(1)}) = g(x_2;\alpha_1,\lambda_1/a)$$

Iterating the recursions, we find that

$$p\left(x_t|\mathcal{Y}^{(t)}\right) = g(x_t; \alpha_t, \lambda_t)$$

and

$$p\left(x_{t+1}|\mathcal{Y}^{(t)}\right) = g(x_{t+1}; \alpha_t, \lambda_t/a)$$

where

$$\alpha_t = \alpha_{t-1} + y_t = \alpha + y_1 + \ldots + y_t$$
$$\lambda_t = \lambda_{t-1}/a + \pi = \lambda a^{1-t} + \pi\frac{(1-a^{-t})}{(1-a^{-1})}$$

Observation-driven models

Assumption: (again)
Y_t given $(X_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ is independent of $(\mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$.
The model is specified by

$$p\left(y_t|x_t\right) = p\left(y_t|\mathcal{X}^{(t)}, \mathcal{Y}^{(t-1)}\right) \qquad \text{(obs. eqn.)}$$

$$p(x_{t+1}|\mathcal{Y}^{(t)}) = \text{depends on all obs.} \qquad \text{(state eqn.)}$$

The advantage of the observation-driven state equation is that Bayes' Theorem can be used directly without the use of the updating formula.

Note: The observation equation is the same for both types of models but the state vector of a parameter-driven model evolves independent of past observation, while the state vector of an observation-driven model depends on past observations.

Example – observation-driven model

An AR(1) process with i.i.d. noise can be expressed as an observation-driven model.
Suppose

$$Y_t = \Phi Y_{t-1} + Z_t$$

$\{Z_t\}$ is i.i.d. with mean 0 and probability density $f(x)$.
Then with

$$X_t := Y_{t-1}$$

we have

$$p\left(y_t | x_t\right) = f(y_t - \phi x_t)$$

$$p\left(x_{t+1} | \mathcal{Y}^{(t)}\right) = \begin{cases} 1, & \text{if } x_{t+1} = y_t \\ 0, & \text{otherwise} \end{cases}$$

The Poisson-gamma state-space model

- Useful for e.g. modeling the rain intensity measurements – see later
- Assume

$$p(y_t|X_t) = \frac{X_t^{y_t}}{y_t!} e^{-X_t} \ \sim \ p_0(x_t) \qquad \text{(obs. eqn.)}$$

where $y_t = 0, 1, 2, \ldots$ and $X_t > 0$.

$$p(x_t|\mathcal{Y}^{(t-1)}) = G(\alpha_{t|t-1}, \beta_{t|t-1}) \qquad \text{(state eqn.)}$$

- Such that

$$\alpha_{t|t-1} = \omega \alpha_{t-1}$$
$$\beta_{t|t-1} = \omega \beta_{t-1} \quad (0 < \omega \leq 1)$$

- Posterior is now (updating)

$$p(x_t|\mathcal{Y}^{(t)}) = G(\alpha_{t|t-1} + y_t, \beta_{t|t-1} + 1) = G(\alpha_t, \beta_t)$$

- Note that

$$\mathbb{E}\left[X_t|\mathcal{Y}^{(t-1)}\right] = \frac{\alpha_{t-1}}{\beta_{t-1}} \quad \left(= \frac{\alpha_{t-1|t-2} + y_{t-1}}{\beta_{t-1|t-2} + 1}\right)$$

$$V[X_t|\mathcal{Y}^{(t-1)}] = \frac{\alpha_{t-1}}{\beta_{t-1}^2}$$

- Whereas

$$\mathbb{E}\left[X_t|\mathcal{Y}^{(t-1)}\right] = \frac{\alpha_{t|t-1}}{\beta_{t|t-1}} = \frac{\omega\alpha_{t-1}}{\omega\beta_{t-1}} = \frac{\alpha_{t-1}}{\beta_{t-1}}$$

$$V[X_t|\mathcal{Y}^{(t-1)}] = \frac{\alpha_{t|t-1}}{\beta_{t|t-1}^2} = \frac{1}{\omega} V[X_{t-1}|\mathcal{Y}^{(t-1)}]$$

- Thus the mean of $X_t|\mathcal{Y}^{(t-1)}$ is the same as the mean of $X_{t-1}|\mathcal{Y}^{(t-1)}$, but the variance is increased.

The distribution of $Y_t|\mathcal{Y}^{(t-1)}$

$$
\begin{aligned}
p\left(y_t|Y^{(t-1)}\right) &= \int_0^\infty p\left(y_t|x_t\right)p\left(x_t|Y^{(t-1)}\right)\,\mathrm{d}x_t \\
&= \int_0^\infty \frac{x_t^{y_t}}{y_t!}e^{-x_t}\frac{x_t^{\alpha_{t|t-1}-1}}{\Gamma(\alpha_{t|t-1})\beta_{t|t-1}^{-\alpha_{t|t-1}}}e^{-\beta_{t|t-1}x_t}\,\mathrm{d}x_t \\
&= \frac{1}{y_t!\Gamma(\alpha_{t|t-1})\beta_{t|t-1}^{\alpha_{t|t-1}}}\int_0^\infty e^{-x_t}x_t^{y_t+\alpha_{t|t-1}-1}e^{-\beta_{t|t-1}x_t}\,\mathrm{d}x_t \\
&= \frac{1}{y_t!\Gamma(\alpha_{t|t-1})\beta_{t|t-1}^{-\alpha_{t|t-1}}}\int_0^\infty x_t^{y_t+\alpha_{t|t-1}-1}e^{-x_t(1+\beta_{t|t-1})}\,\mathrm{d}x_t \quad (\triangle)
\end{aligned}
$$

- Recall the definition of the Gamma function:

$$\int_0^\infty x^{U-1} e^{-xv}\, \mathrm{d}x = \frac{\Gamma}{v^U}$$

- Then $(\triangle)$ can be written as

$$p\left(y_t | x_{t-1}, y_{t-1}\right) = \frac{\Gamma(y_t + \alpha_{t|t-1})}{\Gamma(y_t+1)\Gamma(\alpha_{t|t-1})} \beta_{t|t-1}^{\alpha_{t|t-1}} (1 + \beta_{t|t-1})^{-1(y_t + \alpha_{t|t-1})}$$

which is the negative binomial density given by

$$p\left(x; k, p\right) = \frac{\Gamma(k+x)}{\Gamma(k+1)\Gamma(k)} p^k (1-p)^x$$

with

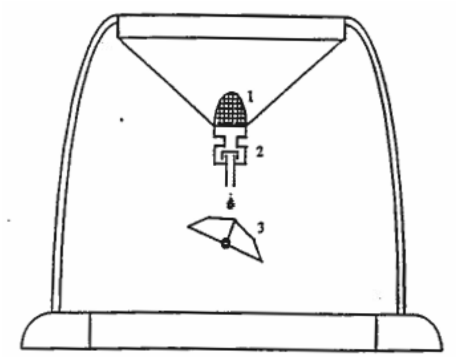
$$p = \frac{\beta_{t|t-1}}{1 + \beta_{t|t-1}}, \quad k = \alpha_{t|t-1} \quad \text{and} \quad x = y_t$$

The likelihood

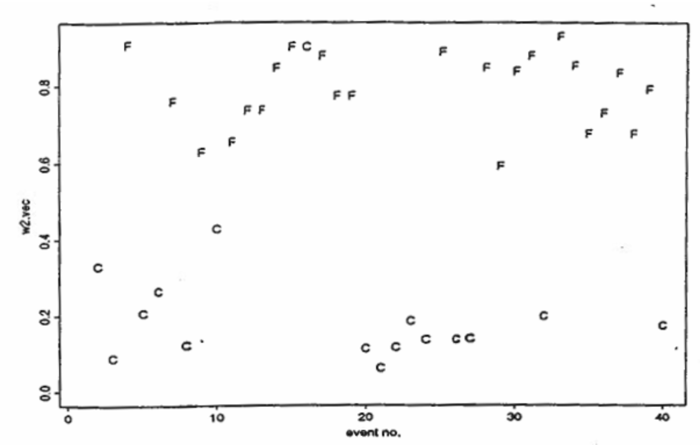
$$L(\omega, Y^{(n)}) = \prod_{t=1}^{n} p\left(y_t | \mathcal{Y}^{(t-1)}\right)$$
$$= \prod_{t=1}^{n} \frac{\Gamma(\alpha_{t|t-1} + y_t)}{\Gamma(y_t + 1)\Gamma(\alpha_{t|t-1})} \beta_{t|t-1}^{\alpha_{t|t-1}} (1 + \beta_{t|t-1})^{-(y_t + \alpha_{t|t-1})}$$

- The parameter ω can be regarded as a thinning parameter which indicates the strength of a carry-over effect.
- Example: The rainfall measurement process!

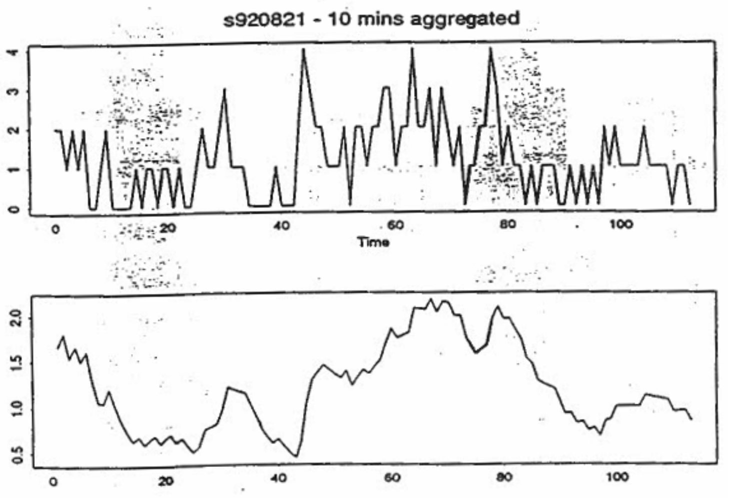
Tipping bucket rain gauge



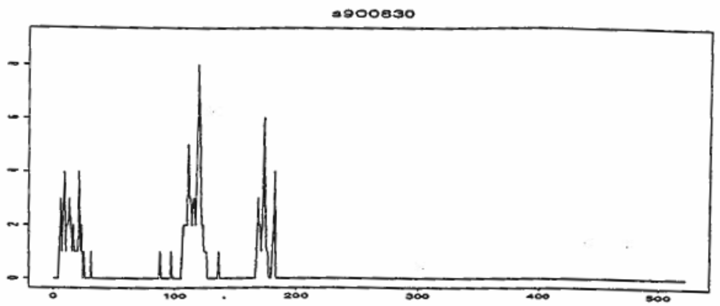
Parameter ω for 39 events



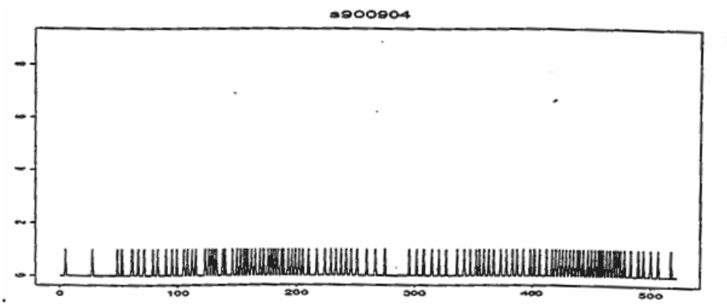
Original series and estimated latent process



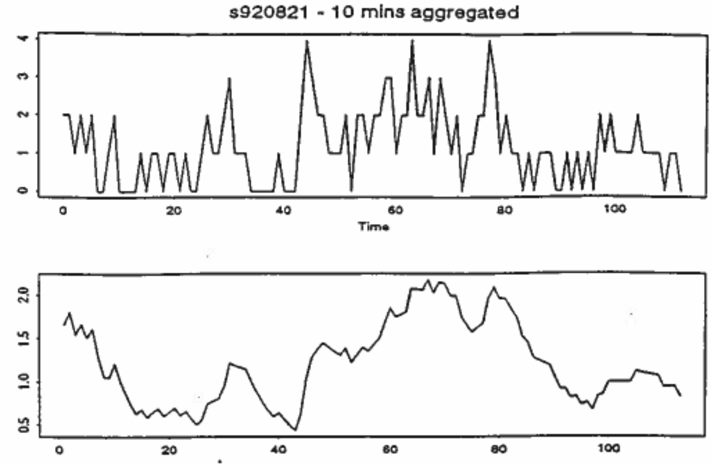
Convective rain event from August 30th 1990



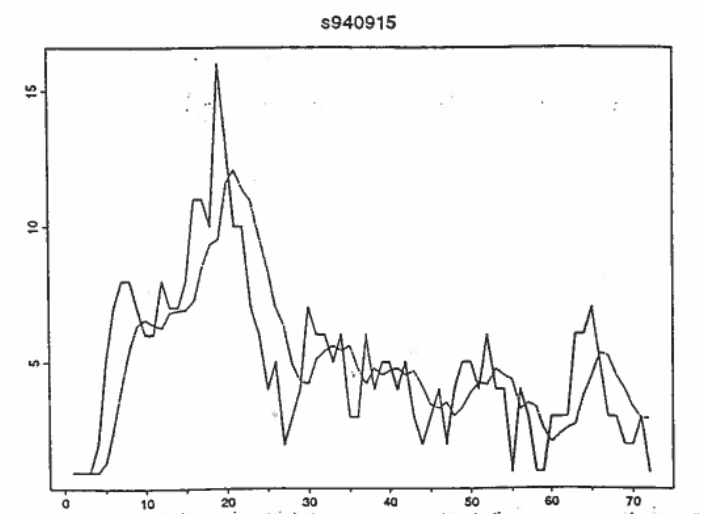
Frontal rain event from September 4th 1990



Original series and estimated latent process



Poisson-gamma predictions



Exponential family state space models

Defined as the model

$$
\begin{aligned}
\theta_t &= G_t \theta_{t-1} + \omega_t \quad ; \quad \omega_t \sim N_p(0, W_t) \\
\lambda_t &= F_t^T \theta_t \\
\eta_t &= v(\lambda_t) \\
p\left(y_t | \eta_t\right) &= \exp\left(Y_t^T \eta_t - b(\eta_t) + c(Y_t)\right)
\end{aligned}
$$

with initial prior $\theta_0 | D_o \sim N_p(m_0, C_0)$

- The relation between The mean

$$\mu_t = \tau(\eta_t)$$

and the linear predictor λ_t is specified through the response function, h:

$$\mu_t = h(\lambda_t)$$

- The inverse of the response function is the link function, g:

$$g(\mu_t) = \lambda_t$$

Extended Kalman filter and smoother

- Solving the state-filtering problem.

1. We know $(\theta_{t-1}|D_{t-1}) \sim N_p(m_{t-1}, C_{t-1})$ from previous slides.
2. Prior on the state:

$$(\theta_{t-1}|D_{t-1}) \sim N_p(\underbrace{G_t m_{t-1}}_{a_t}, \underbrace{G_t C_{t-1} G_t^T + W_t}_{R_t})$$

3. Determine point for Taylor expansion:
   If first pass:

$$\check{\lambda}_t = F_t^T a_t$$

   Otherwise use most recent run of Kalman smoother:

$$\check{\lambda}_t = F_t^T \tilde{m}_t$$

4. Determine:

$$\check{Y}_t = h'(\check{\lambda}_t)^{-1}\left\{Y_t - h(\check{\lambda}_t)\right\} + \check{\lambda}_t$$
$$\check{V}_t = \left\{h'(\check{\lambda}_t)\check{\Sigma}_t^{-1}h'(\check{\lambda}_t)\right\}^{-1}$$

5. One-step forecast:

$$(Y_t|D_{t-1}) \sim N_d\left(\underbrace{F_t^T a_t}_{f_t}, \underbrace{F_t^T R_t F_t + \check{V}_t}_{Q_t}\right)$$

6. Posterior for state ($A_t = R_t F_t Q_t^{-1}$)

$$(\theta_t|D_t)\ N_p\left(\underbrace{a_t + A_t(\check{Y}_t - f_t)}_{m_t}, \underbrace{R_t - A_t Q_t A_t^T}_{c_t}\right)$$