Nonlinear and non-Gaussian state-space modelling by means of hidden Markov models

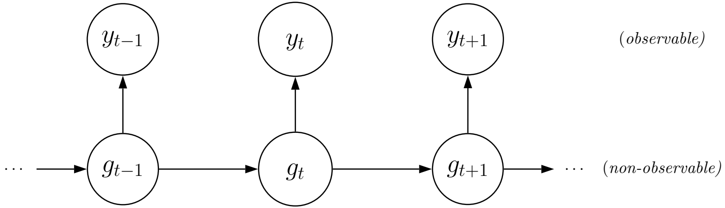
Roland Langrock
University of Göttingen
St Andrews, 13 December 2010

bla
bla
bla
bla

1. State-space models
    - Basics
    - Approximation via hidden Markov models

2. Application
    - Glacial varve thickness

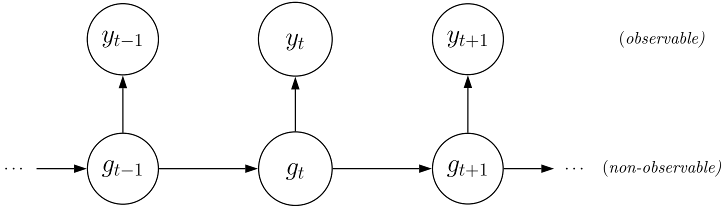
(General) state-space model (SSM):



$$y_t = a(g_t, \epsilon_t)$$
$$g_t = b(g_{t-1}, \eta_t)$$

- a, b: known functions (not necessarily linear)
- ϵ_t, η_t iid (not necessarily $\sim \mathcal{N}$)

(General) state-space model (SSM):



$$y_t = a(g_t, \epsilon_t)$$
$$g_t = b(g_{t-1}, \eta_t)$$

- a, b: known functions (not necessarily linear)
- ϵ_t, η_t iid (not necessarily $\sim \mathcal{N}$)

Example 1. Stochastic volatility model:

$$y_t = \epsilon_t \beta \exp(g_t/2)$$
$$g_t = \phi g_{t-1} + \sigma \eta_t$$

- $\epsilon_t \stackrel{iid}{\sim} t_\nu$ or $\mathcal{N}(0,1)$, $\eta_t \stackrel{iid}{\sim} \mathcal{N}(0,1)$
- g_t determines variance (volatility) of y_t

Example 2. Poisson autoregression:

$$y_t \sim \text{Poisson}\big(\beta \exp(g_t)\big)$$
$$g_t = \phi g_{t-1} + \sigma \eta_t$$

- $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1)$
- g_t determines mean (and variance) of y_t

Desired:

- parameter estimation
- state decoding
- forecasts
- model checking

SSM likelihood:

$$\mathcal{L}(\mathbf{y}) = \underbrace{\int \ldots \int}_{n-\text{fold}} f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

can not be evaluated directly...

(SSM *linear & Gaussian* → Kalman filter optimal)

Desired:

- parameter estimation
- state decoding
- forecasts
- model checking

SSM likelihood:

$$\mathcal{L}(\mathbf{y}) = \underbrace{\int \ldots \int}_{n-\text{fold}} f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

can not be evaluated directly...

(SSM *linear & Gaussian* $\rightarrow$ Kalman filter optimal)

Desired:

- parameter estimation
- state decoding
- forecasts
- model checking

SSM likelihood:

$$\mathcal{L}(\mathbf{y}) = \underbrace{\int \dots \int}_{n-\text{fold}} f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

can not be evaluated directly...

(SSM *linear & Gaussian* → Kalman filter optimal)

Parameter estimation in case of nonlinearity/non-Gaussianity:

- *Extended Kalman filter*
  - \+ simple implementation
  - – in general poor approximation
- *(Generalized) method of moments*
  - \+ simple implementation
  - – low efficiency, no state decoding
- *Monte Carlo methods*
  - \+ high efficiency
  - – computer-intensive
    nonstandard models require nontrivial modifications

Parameter estimation in case of nonlinearity/non-Gaussianity:

- *Extended Kalman filter*
  - \+ simple implementation
  - – in general poor approximation
- *(Generalized) method of moments*
  - \+ simple implementation
  - – low efficiency, no state decoding
- *Monte Carlo methods*
  - \+ high efficiency
  - – computer-intensive
    nonstandard models require nontrivial modifications

Parameter estimation in case of nonlinearity/non-Gaussianity:

- *Extended Kalman filter*
  - + simple implementation
  - − in general poor approximation
- *(Generalized) method of moments*
  - + simple implementation
  - − low efficiency, no state decoding
- *Monte Carlo methods*
  - + high efficiency
  - − computer-intensive
    nonstandard models require nontrivial modifications

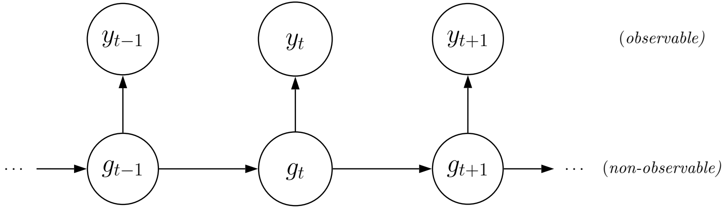
Parameter estimation in case of nonlinearity/non-Gaussianity:

- *Extended Kalman filter*
  - \+ simple implementation
  - – in general poor approximation
- *(Generalized) method of moments*
  - \+ simple implementation
  - – low efficiency, no state decoding
- *Monte Carlo methods*
  - \+ high efficiency
  - – computer-intensive
    nonstandard models require nontrivial modifications

1 State-space models
- Basics
- Approximation via hidden Markov models

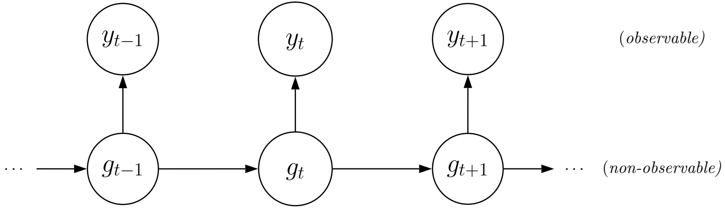
2 Application
- Glacial varve thickness

Hidden Markov model:



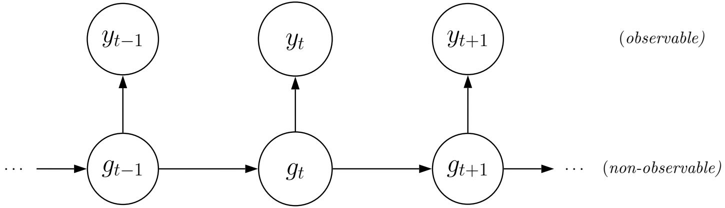
- Non-observable process: N-state Markov chain g_t
  - initial distribution $\delta_i = \mathbb{P}(g_1 = i)$
  - transition probabilities $\gamma_{ij} = \mathbb{P}(g_t = j \mid g_{t-1} = i)$
- Observable process: y_t
  - state-dependent density $f(y_t \mid g_t)$

Hidden Markov model:



- Non-observable process: N-state Markov chain g_t
  - initial distribution $\delta_i = \mathbb{P}(g_1 = i)$
  - transition probabilities $\gamma_{ij} = \mathbb{P}(g_t = j \,|\, g_{t-1} = i)$
- Observable process: y_t
  - state-dependent density $f(y_t \,|\, g_t)$

Hidden Markov model:



- Non-observable process: N-state Markov chain g_t
  - initial distribution $\delta_i = \mathbb{P}(g_1 = i)$
  - transition probabilities $\gamma_{ij} = \mathbb{P}(g_t = j \,|\, g_{t-1} = i)$
- Observable process: y_t
  - state-dependent density $f(y_t \,|\, g_t)$

Key idea:

- HMMs have the same two-process structure as SSMs
- in SSMs: g_t continuous-valued
- discretizing g_t yields approximation by HMM
- benefit: HMM methodology becomes applicable

- split essential range of g_t into m equidistant intervals

$$B_i := [b_{i-1}, b_i]$$

- b_i^*: midpoint of B_i

$$\Rightarrow \mathcal{L}(\mathbf{y}) = \int \ldots \int f(\mathbf{y}, \mathbf{g}) \, d\mathbf{g}$$

$$= \int f(g_1) f(y_1 | g_1) \prod_{t=2}^{n} \int f(g_t | g_{t-1}) f(y_t | g_t) \, dg_n \ldots dg_1$$

$$\approx \sum_{i_1=1}^{m} \mathbb{P}(g_1 \in B_{i_1}) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \sum_{i_t=1}^{m} \mathbb{P}(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*) f(y_t | g_t = b_{i_t}^*)$$

$$=: \mathcal{L}_{\text{approx}}(\mathbf{y})$$

- split essential range of g_t into m equidistant intervals

$$B_i := [b_{i-1}, b_i]$$

- b_i^*: midpoint of B_i

$$\Rightarrow \mathcal{L}(\mathbf{y}) = \int \ldots \int f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

$$= \int f(g_1) f(y_1 | g_1) \prod_{t=2}^{n} \int f(g_t | g_{t-1}) f(y_t | g_t) \, \mathrm{d}g_n \ldots \mathrm{d}g_1$$

$$\approx \sum_{i_1=1}^{m} \mathbb{P}(g_1 \in B_{i_1}) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \sum_{i_t=1}^{m} \mathbb{P}(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*) f(y_t | g_t = b_{i_t}^*)$$

$$=: \mathcal{L}_{\text{approx}}(\mathbf{y})$$

- split essential range of g_t into m equidistant intervals

$$B_i := [b_{i-1}, b_i]$$

- b_i^*: midpoint of B_i

$$\Rightarrow \mathcal{L}(\mathbf{y}) = \int \dots \int f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

$$= \int f(g_1) f(y_1 | g_1) \prod_{t=2}^{n} \int f(g_t | g_{t-1}) f(y_t | g_t) \, \mathrm{d}g_n \dots \mathrm{d}g_1$$

$$\approx \sum_{i_1=1}^{m} \mathbb{P}(g_1 \in B_{i_1}) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \sum_{i_t=1}^{m} \mathbb{P}(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*) f(y_t | g_t = b_{i_t}^*)$$

$$=: \mathcal{L}_{\text{approx}}(\mathbf{y})$$

- split essential range of g_t into m equidistant intervals

$$B_i := [b_{i-1}, b_i]$$

- b_i^*: midpoint of B_i

$$\Rightarrow \mathcal{L}(\mathbf{y}) = \int \ldots \int f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

$$= \int f(g_1) f(y_1 | g_1) \prod_{t=2}^{n} \int f(g_t | g_{t-1}) f(y_t | g_t) \, \mathrm{d}g_n \ldots \mathrm{d}g_1$$

$$\approx \sum_{i_1=1}^{m} \mathbb{P}(g_1 \in B_{i_1}) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \sum_{i_t=1}^{m} \mathbb{P}(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*) f(y_t | g_t = b_{i_t}^*)$$

$$=: \mathcal{L}_{\text{approx}}(\mathbf{y})$$

- split essential range of g_t into m equidistant intervals

$$B_i := [b_{i-1}, b_i]$$

- b_i^*: midpoint of B_i

$$\Rightarrow \mathcal{L}(\mathbf{y}) = \int \ldots \int f(\mathbf{y}, \mathbf{g}) \, \mathrm{d}\mathbf{g}$$

$$= \int f(g_1) f(y_1 | g_1) \prod_{t=2}^{n} \int f(g_t | g_{t-1}) f(y_t | g_t) \, \mathrm{d}g_n \ldots \mathrm{d}g_1$$

$$\approx \sum_{i_1=1}^{m} \mathbb{P}(g_1 \in B_{i_1}) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \sum_{i_t=1}^{m} \mathbb{P}(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*) f(y_t | g_t = b_{i_t}^*)$$

$$=: \mathcal{L}_{\text{approx}}(\mathbf{y})$$

Consider HMM with

- m-state MC (possible outcomes: midpoints b_i^*)
  - transition probabilities: $\gamma_{ij} := \mathbb{P}(g_t \in B_j | g_{t-1} = b_i^*)$
  - transition probability matrix: $\mathbf{\Gamma} = (\gamma_{ij})$
  - initial distribution: $\delta_i := \mathbb{P}(g_1 \in B_i)$
- observable process:
  - state-dependent density: $f(y_t | g_t = b_i^*)$
  - $\mathbf{P}(y_t)$: diag. matrix with ith entry $f(y_t \,|\, g_t = b_i^*)$

$$\Rightarrow \mathcal{L}_{\text{approx}}(\mathbf{y}) = \boldsymbol{\delta}\mathbf{P}(y_1)\mathbf{\Gamma}\mathbf{P}(y_2)\mathbf{\Gamma}\cdots\mathbf{\Gamma}\mathbf{P}(y_{n-1})\mathbf{\Gamma}\mathbf{P}(y_n)\mathbf{1}^t$$

$\rightarrow$ the HMM $(\delta, \mathbf{\Gamma}, f(y_t|\cdot))$ approximates the SSM

Consider HMM with

- m-state MC (possible outcomes: midpoints b_i^*)
  - transition probabilities: $\gamma_{ij} := \mathbb{P}(g_t \in B_j | g_{t-1} = b_i^*)$
  - transition probability matrix: $\mathbf{\Gamma} = (\gamma_{ij})$
  - initial distribution: $\delta_i := \mathbb{P}(g_1 \in B_i)$
- observable process:
  - state-dependent density: $f(y_t | g_t = b_i^*)$
  - $\mathbf{P}(y_t)$: diag. matrix with ith entry $f(y_t \,|\, g_t = b_i^*)$

$$\Rightarrow \mathcal{L}_{\text{approx}}(\mathbf{y}) = \boldsymbol{\delta}\mathbf{P}(y_1)\mathbf{\Gamma}\mathbf{P}(y_2)\mathbf{\Gamma}\cdots\mathbf{\Gamma}\mathbf{P}(y_{n-1})\mathbf{\Gamma}\mathbf{P}(y_n)\mathbf{1}^t$$

$\rightarrow$ the HMM $(\delta, \Gamma, f(y_t|\cdot))$ approximates the SSM

Consider HMM with

- m-state MC (possible outcomes: midpoints b_i^*)
  - transition probabilities: $\gamma_{ij} := \mathbb{P}(g_t \in B_j | g_{t-1} = b_i^*)$
  - transition probability matrix: $\mathbf{\Gamma} = (\gamma_{ij})$
  - initial distribution: $\delta_i := \mathbb{P}(g_1 \in B_i)$
- observable process:
  - state-dependent density: $f(y_t | g_t = b_i^*)$
  - $\mathbf{P}(y_t)$: diag. matrix with ith entry $f(y_t \,|\, g_t = b_i^*)$

$$\Rightarrow \mathcal{L}_{\text{approx}}(\mathbf{y}) = \boldsymbol{\delta}\mathbf{P}(y_1)\mathbf{\Gamma}\mathbf{P}(y_2)\mathbf{\Gamma}\cdots\mathbf{\Gamma}\mathbf{P}(y_{n-1})\mathbf{\Gamma}\mathbf{P}(y_n)\mathbf{1^t}$$

$\rightarrow$ the HMM $(\delta, \mathbf{\Gamma}, f(y_t|\cdot))$ approximates the SSM

Consider HMM with

- m-state MC (possible outcomes: midpoints b_i^*)
  - transition probabilities: $\gamma_{ij} := \mathbb{P}(g_t \in B_j | g_{t-1} = b_i^*)$
  - transition probability matrix: $\mathbf{\Gamma} = (\gamma_{ij})$
  - initial distribution: $\delta_i := \mathbb{P}(g_1 \in B_i)$
- observable process:
  - state-dependent density: $f(y_t | g_t = b_i^*)$
  - $\mathbf{P}(y_t)$: diag. matrix with ith entry $f(y_t \,|\, g_t = b_i^*)$

$$\Rightarrow \mathcal{L}_{\text{approx}}(\mathbf{y}) = \boldsymbol{\delta}\mathbf{P}(y_1)\mathbf{\Gamma}\mathbf{P}(y_2)\mathbf{\Gamma}\cdots\mathbf{\Gamma}\mathbf{P}(y_{n-1})\mathbf{\Gamma}\mathbf{P}(y_n)\mathbf{1}^{\mathbf{t}}$$

$\rightarrow$ the HMM $(\boldsymbol{\delta}, \mathbf{\Gamma}, f(y_t|\cdot))$ approximates the SSM

Pros and Cons (HMM method):

+ likelihood directly available
  extensions straightforward
  simple formulae for residuals, forecasts, decoding

− m and range of g_t have to be chosen
  only feasible for one-dimensional state spaces

Applications considered in Langrock (2010):

- stochastic volatility
- earthquake counts
- polio counts (seasonal)
- daily rainfall occurrence (seasonal)
- glacial varve thickness

Applications considered in Langrock (2010):

- stochastic volatility
- earthquake counts
- polio counts (seasonal)
- daily rainfall occurrence (seasonal)
- glacial varve thickness

- varves: layers of sediment deposited by melting glaciers
- can be useful for long-term climate research
- source: Shumway and Stoffer (*Time Series Analysis and Its Applications*, 2006)

- varves: layers of sediment deposited by melting glaciers
- can be useful for long-term climate research
- source: Shumway and Stoffer (*Time Series Analysis and Its Applications*, 2006)

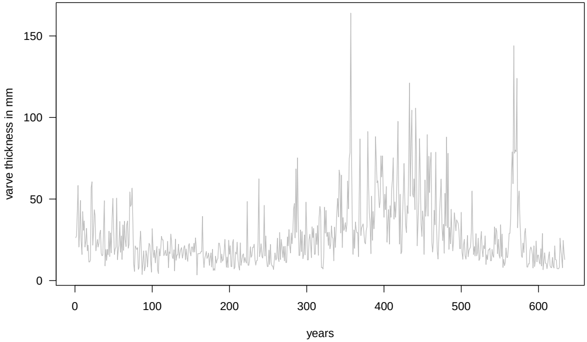


Figure: *Series of glacial varve thicknesses for a location in Massachusetts.*

$$y_t = \epsilon_t \beta \exp(g_t)$$
$$g_t = \phi g_{t-1} + \sigma \eta_t$$

$$\epsilon_t \sim \text{Gamma}\big(\text{shape} = c_v^{-2}, \text{scale} = c_v^2\big)$$

Properties:

- $\mathbb{E}(y_t | g_t) = \beta \exp(g_t)$
- (Conditional) coefficient of variation:

$$\frac{\text{sd}(y_t | g_t)}{\mathbb{E}(y_t | g_t)} = c_v$$

$$y_t = \epsilon_t \beta \exp(g_t)$$
$$g_t = \phi g_{t-1} + \sigma \eta_t$$

$$\epsilon_t \sim \text{Gamma}\big(\text{shape} = c_v^{-2}, \text{scale} = c_v^2\big)$$

Properties:

- $\mathbb{E}(y_t|g_t) = \beta \exp(g_t)$
- (Conditional) coefficient of variation:

$$\frac{\text{sd}(y_t|g_t)}{\mathbb{E}(y_t|g_t)} = c_v$$

$$y_t = \epsilon_t \beta \exp(g_t)$$
$$g_t = \phi g_{t-1} + \sigma \eta_t$$

$$\epsilon_t \sim \text{Gamma}\big(\text{shape} = c_v^{-2}, \text{scale} = c_v^2\big)$$

Properties:

- $\mathbb{E}(y_t | g_t) = \beta \exp(g_t)$
- (Conditional) coefficient of variation:

$$\frac{\text{sd}(y_t | g_t)}{\mathbb{E}(y_t | g_t)} = c_v$$

Table: *Estimated model parameters and bootstrap 95% confidence intervals (400 replications).*

para.	estimate	c.i.
ϕ	0.95	[0.90, 0.97]
σ	0.15	[0.11, 0.19]
β	24.42	[19.1, 31.1]
c_v	0.40	[0.37, 0.42]

$$\text{resolution: } m = 200$$

$$g_t - \text{range: } b_0 = -3,\ b_m = 3$$

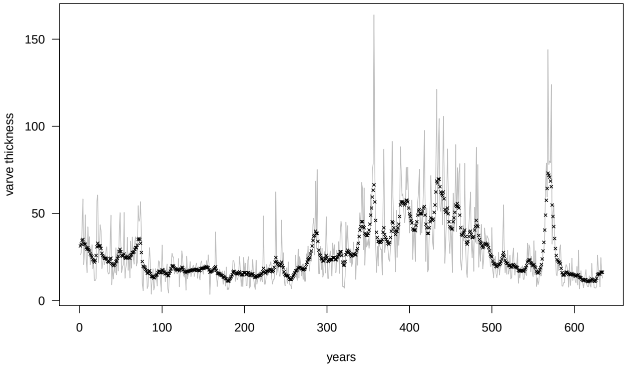


Figure: *Series of glacial varve thicknesses (solid grey line) and decoded mean sequence of the fitted gamma SSM (crosses).*

Summary

- HMM approximation convenient in SSM context
- whole HMM methodology applicable
- simple implementation of standard and nonstandard models

Langrock, R., MacDonald, I. M., Zucchini, W., 2010
Estimating standard and nonstandard stochastic volatility models using structured hidden Markov models. *(submitted)*

Langrock, R., 2010
Some applications of nonlinear and non-Gaussian state-space modeling by means of hidden Markov models. *(submitted)*