PML 13. State Space Models

Probabilistic Machine Learning Reading Group

Bruno Flores

April 29, 2026

Capgemini

Contents

- Introduction to SSMs
- Some types of SSMs
  - HMM
  - LDS and DLM
  - DGLM
- Applications
  - Forecasting aviation incidents
  - Supermarket Demand Forecasting
- Conclusions

Contents

State Space Models

State Space Models (SSM) are a type of partially observed Markov models.

Hidden states $\boldsymbol{z}_t$ generate observations $\boldsymbol{y}_t$ at each time step t.

Depending on the application, one of two main objectives:

- Infer hidden states $\boldsymbol{z}_t$
  - Predictive maintenance of machine parts
  - Identifying brain activity states from EEG data.
- Predict future observations $\boldsymbol{y}_t$
  - Sales forecasting from historical demand and other inputs
  - Weather forecasting

State Space Models. Definition

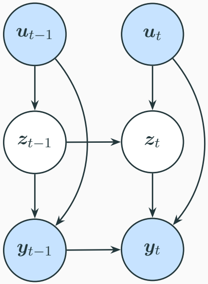
Transition model:

$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{u}_t) = p(\boldsymbol{z}_t|f(\boldsymbol{z}_{t-1}, \boldsymbol{u}_t))$$

Observation model:

$$p(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{u}_t, \boldsymbol{y}_{1:t-1}) = p(\boldsymbol{y}_t|h(\boldsymbol{z}_t, \boldsymbol{u}_t, \boldsymbol{y}_{1:t-1}))$$

where $\boldsymbol{u}_t$ are optional observed inputs, f the transition function and h the observation function.



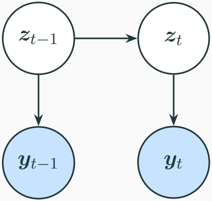
Contents

Contents

Hidden Markov Models

Hidden Markov Models (HMM) are SSM in which:

- Hidden states z_t are discrete, $z_t \in \{1, ..., K\}$
- Observations $\boldsymbol{y}_t$ may be discrete or continuous
- Hidden states are Markovian
- Observations are iid conditioned on the hidden states



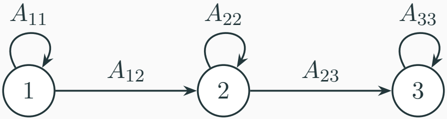
HMM. State Transition Model

Initial state distribution:

$$p(z_1 = j) = \pi_j$$

Transition model defined by row stochastic matrix $\boldsymbol{A}$:

$$p(z_t = j | z_{t-1} = i) = A_{ij}$$



State Transition Diagram

HMM. Observation Models

Observation model $p(\boldsymbol{y}_t|z_t = j)$ can be discrete or continuous

Examples:

- $p(\boldsymbol{y}_t|z_t = j) = \prod_{d=1}^{D} \mathsf{Cat}(y_{td}|\mathbf{p}_{d,j})$
- $p(y_t|z_t = j) = \mathsf{Poi}(y_t|\lambda_j)$
- $p(\boldsymbol{y}_t|z_t = j) = \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$
- $p(\boldsymbol{y}_t|\boldsymbol{y}_{t-L:t-1}, z_t = j, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{y}_t \,\middle|\, \sum_{\ell=1}^{L} \boldsymbol{W}_{j,\ell}\boldsymbol{y}_{t-\ell} + \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)$
- $p(\boldsymbol{y}_t|z_t = j, \boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(\boldsymbol{y}_t \mid z_t = j)$ Neural network likelihood

HMM. Parameter Learning

- Baum-Welch (EM) algorithm
- Stochastic Gradient Descent (SGD)
- Spectral Methods (standalone or as initial parameters for EM)
- Bayesian approach (variational Bayes or blocked Gibbs)

HMM. Parameter Learning. Baum-Welch Algorithm

$$\gamma_{n,t}(j) \triangleq p(z_t = j | \boldsymbol{y}_{n,1:T_n}, \boldsymbol{\theta}^{\text{old}}) \qquad \xi_{n,t}(j,k) \triangleq p(z_{t-1} = j, z_t = k | \boldsymbol{y}_{n,1:T_n}, \boldsymbol{\theta}^{\text{old}})$$

$$\hat{A}_{jk} = \frac{\mathbb{E}\left[N_{jk}\right]}{\sum_{k'} \mathbb{E}\left[N_{jk'}\right]}, \qquad \hat{\pi}_k = \frac{\mathbb{E}\left[N_k^1\right]}{N}, \qquad \hat{B}_{kv} = \frac{\mathbb{E}\left[M_{kv}\right]}{\mathbb{E}\left[N_k\right]}$$

Baum-Welch algorithm for HMMs

```
Initialize parameters θ
for each iteration do
    Initialize expected counts: E[N_k] = 0, E[N_jk] = 0, E[M_kv] = 0
    for each datacase n do
        Use forwards-backwards algorithm on y_n to compute γ_{n,t} and ξ_{n,t}
        E[N_k] := E[N_k] + Σ_{t=2}^{T_n} γ_{n,t}(k)
        E[N_jk] := E[N_jk] + Σ_{t=2}^{T_n} ξ_{n,t}(j,k)
        E[M_kv] := E[M_kv] + Σ_{t:x_{n,t}=v} γ_{n,t}(k)
    end for
    Compute new parameters θ = (A, B, π) using equations above
end for
```

Contents

LDS and DLM

A linear dynamic system (LDS) or linear-Gaussian state space model is defined as:

$$p(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{F}_t\boldsymbol{z}_t + \boldsymbol{D}_t\boldsymbol{u}_t, \boldsymbol{V}_t)$$
$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{G}_t\boldsymbol{z}_{t-1}, \boldsymbol{W}_t)$$

LDS and DLM

A linear dynamic system (LDS) or linear-Gaussian state space model is defined as:

$$p(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{F}_t\boldsymbol{z}_t + \boldsymbol{D}_t\boldsymbol{u}_t, \boldsymbol{V}_t)$$
$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{G}_t\boldsymbol{z}_{t-1}, \boldsymbol{W}_t)$$

and if we incorporate the inputs u_t into the block structured matrices $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ we get the common notation for what is also called a dynamic linear model (DLM)

$$p(\boldsymbol{y}_t|\boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{F}_t\boldsymbol{z}_t, \boldsymbol{V}_t)$$
$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) = \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{G}_t\boldsymbol{z}_{t-1}, \boldsymbol{W}_t), \qquad \boldsymbol{z}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$$

LDS and DLM

A linear dynamic system (LDS) or linear-Gaussian state space model is defined as:

$$p(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{F}_t\boldsymbol{z}_t + \boldsymbol{D}_t\boldsymbol{u}_t, \boldsymbol{V}_t)$$
$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{G}_t\boldsymbol{z}_{t-1}, \boldsymbol{W}_t)$$

If we incorporate the inputs u_t into the block structured matrices $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ we get the common notation for what is also called a dynamic linear model (DLM)

$$p(y_t|\boldsymbol{\theta}_t) = \mathcal{N}(y_t|\boldsymbol{F}_t\boldsymbol{\theta}_t, V_t) \qquad y_t \in R$$
$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) = \mathcal{N}(\boldsymbol{\theta}_t|\boldsymbol{G}_t\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_t), \qquad \boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$$

with observation matrix $\boldsymbol{F}_t \in \mathbb{R}^{m\times p}$ and state evolution matrix $\boldsymbol{G}_t \in \mathbb{R}^{p\times p}$.

DLM Building Blocks

Matrix with three model components $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$

$$\begin{pmatrix} A_{1,1} & A_{1,2} & 0 & 0 & 0 & 0 \\ A_{2,1} & A_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & B_{1,1} & B_{1,2} & B_{1,3} & 0 \\ 0 & 0 & B_{2,1} & B_{2,2} & B_{2,3} & 0 \\ 0 & 0 & B_{3,1} & B_{3,2} & B_{3,3} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{1,1} \end{pmatrix}$$

Building Blocks. Polynomial

- Local level model (1st-order polynomial)

$$\begin{aligned} y_t &= \mu_t + \epsilon_{y,t} \quad \epsilon_{y,t} \sim \mathcal{N}(0, \sigma_y^2) \\ \mu_t &= \mu_{t-1} + \epsilon_{\mu,t}, \quad \epsilon_{\mu,t} \sim \mathcal{N}(0, \sigma_\mu^2) \end{aligned} \quad \Rightarrow \quad \boldsymbol{F} = \boldsymbol{G} = 1$$

- Linear growth model (2nd-order polynomial)

$$\boldsymbol{F} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \qquad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

- nth-order polynomial model

$$\boldsymbol{F} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}}_{n}, \qquad \boldsymbol{G} = \underbrace{\begin{pmatrix} 1 & 1 & & \mathbf{0} \\ & 1 & \ddots & \\ & & \ddots & 1 \\ \mathbf{0} & & & 1 \end{pmatrix}}_{n} \Bigg\} n.$$

Building Blocks. Seasonality

When considering the presence of a seasonal effect of period s, one of the approaches available to model seasonality is through a seasonal factor component,

$$\boldsymbol{F} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}}_{s-1}, \qquad \boldsymbol{G} = \underbrace{\begin{pmatrix} -1 & -1 & \dots & -1 \\ 1 & 0 & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & & 1 & 0 \end{pmatrix}}_{s-1} \Bigg\} s-1.$$

with $\boldsymbol{F}$, $\boldsymbol{G}$ of dimensions $1 \times (s-1)$ and $(s-1) \times (s-1)$ respectively.

Building Blocks. Covariates

It is also straightforward to include covariates $\boldsymbol{u}_t \in \mathbb{R}^p$ into a DLM

$$\begin{aligned} y_t &= \boldsymbol{u}_t' \boldsymbol{\theta}_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_t) \end{aligned}$$

with $\boldsymbol{G}_t = \boldsymbol{1}_{1 \times p}$, and $\boldsymbol{F}_t = \boldsymbol{u}_t'$.

Note that when forecasting future values, u_{t+h}, must be known, have its own forecasts, or simply assume it stays the same, $u_{t+h} = u_t$.

Building Blocks

Observation and state evolution matrices of model with linear growth, seasonality of period 4 and one covariate u_t as components:

$$\boldsymbol{F}_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & u_t \end{pmatrix}$$

$$\boldsymbol{G}_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Update and forecast

For the univariate DLM, if we denote the available information at the beginning of period t as $D_t = \{D_{t-1}, y_{t-1}\}$, the sequential update and forecast procedure is given by the recursion:

- One-step ahead predictive distribution of $\boldsymbol{\theta}_t$, given D_t. It is $N(\boldsymbol{a}_t, \boldsymbol{R}_t)$, with $\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t$.

- One-step ahead predictive distribution of y_t, given D_t. It is $N(f_t, Q_t)$, with $f_t = \boldsymbol{F}_t \boldsymbol{a}_t$ and $Q_t = \boldsymbol{F}_t \boldsymbol{R}_t \boldsymbol{F}_t' + V_t$.

- Filtering or posterior distribution of $\boldsymbol{\theta}_t$, given D_t and y_t. It is $N(\boldsymbol{m}_t, \boldsymbol{C}_t)$, with $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t \boldsymbol{F}_t' Q_t^{-1}(y_t - f_t)$ and $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t \boldsymbol{F}_t' Q_t^{-1} \boldsymbol{F}_t \boldsymbol{R}_t$.

Discount Factors

Variance matrices $\boldsymbol{V}_t$, $\boldsymbol{W}_t$ are difficult to specify.

An option is to use MLE

Another more common option is to use $\boldsymbol{V}_t = \boldsymbol{I}_m$, and specify W_t via discount factors $\boldsymbol{\delta} \in (0, 1]$, which specify the increase in uncertainty (or loss of information) about state vector between t and $t+1$. It changes $\boldsymbol{R}_t$ in the previous slide to:

$$\boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t{}' / \delta$$

In practice, discount factors are usually assigned values between 0.8 and 0.99. It is possible to use different discount factors for each model component

Contents

DGLM

Dynamic generalized linear models (DGLM).

Observations y_t from a distribution in the exponential family:

$$p(y_t|\eta_t, V_t) = \exp\{V_t^{-1}[T(y_t)\eta_t - a(\eta_t)]\}b(y_t, V_t),$$

for some defining quantities η_t and V_t, and known functions $T(y_t)$, $a(\eta_t)$ and $b(y_t, V_t)$. The DGLM for the series y_t is defined through the components:

Observation model: $p(y_t|\eta_t)$ and $g(\eta_t) = \lambda_t = \boldsymbol{F}_t\boldsymbol{\theta}_t$,

State transition model: $\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t$ with $\boldsymbol{w}_t \sim (\boldsymbol{0}, \boldsymbol{W}_t)$,

Prior information: $\boldsymbol{\theta}_0 \sim (\boldsymbol{m}_0, \boldsymbol{C}_0)$,

Note that no particular distribution is assumed for $\boldsymbol{w}_t$, only its mean $\boldsymbol{0}$, and variance $\boldsymbol{W}_t$

Update and forecast

Similar to DLM. Uses moment matching technique.

At time $t > 0$, given the available information $D_t = \{D_{t-1}, y_{t-1}\}$:

- One-step ahead prior moments for $\boldsymbol{\theta}_t$, given D_t, are $\boldsymbol{\theta}_t | D_t \sim (\boldsymbol{a}_t, \boldsymbol{R}_t)$, with $\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t{}' + \boldsymbol{W}_t$.

- One-step ahead predictive distribution of y_t are based on the conjugacy-induced predictive distribution

$$p(\boldsymbol{y}_t | \boldsymbol{\alpha}_t, \beta_t) = b(\boldsymbol{y}_t, \phi_t) c(\boldsymbol{\alpha}_t, \beta_t) / c(\boldsymbol{\alpha}_t + \phi_t T(\boldsymbol{y}_t), \beta_t + \phi_t),$$

  with hyper-parameters $\{\boldsymbol{\alpha}_t, \beta_t\}$ estimated using moment matching, $E[g(\boldsymbol{\eta}_t) = \lambda_t | D_{t-1}] = f_t$ and $V[g(\boldsymbol{\eta}_t) = \lambda_t | D_{t-1}] = \boldsymbol{Q}_t$. Where $f_t = \boldsymbol{F}_t \boldsymbol{a}_t$ and $\boldsymbol{Q}_t = \boldsymbol{F}_t \boldsymbol{R}_t \boldsymbol{F}_t'$.

Update and forecast

- Posterior moments for $\boldsymbol{\theta}_t$ after observing $\boldsymbol{y}_t$, $\boldsymbol{\theta}_t | \mathcal{D}_t, y_t \sim (\boldsymbol{m}_t, \boldsymbol{C}_t)$

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t \boldsymbol{F}_t' \boldsymbol{Q}_t^{-1}(f_t^* - f_t), \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t \boldsymbol{F}_t' \boldsymbol{Q}_t^{-1}(\boldsymbol{I}_m - \boldsymbol{Q}_t^* \boldsymbol{Q}_t^{-1}) \boldsymbol{F}_t \boldsymbol{R}_t,$$

with $f_t^* = E[g(\eta_t)|D_t]$ and $\boldsymbol{Q}_t^* = V[g(\boldsymbol{\eta}_t)|D_t]$, given that the posterior of the natural parameter, $p(\boldsymbol{\eta}_t|D_t)$, is

$$c\Big(\boldsymbol{\alpha}_t + \phi_t T(\boldsymbol{y}_t), \beta_t + \phi_t\Big) \exp\left(\Big(\boldsymbol{\alpha}_t + \phi_t T(\boldsymbol{y}_t)\Big)\boldsymbol{\eta}_t - \Big(\beta_t + \phi_t\Big) a(\boldsymbol{\eta}_t)\right)$$

Contents

Contents

Forecasting aviation incidents. Data

Provided by the Spanish Aviation Safety Agency (AESA)

- 86 types of safety occurrences
- ICAO 5 severity levels
- Num. fatalities and injured (w/ severity)

Aviation safety risk management support: National Aviation Safety Plan

Basic Model. Poisson-Gamma

$$D_k = \{(\overset{\text{occur.}}{\overset{\downarrow}{x_i}}, \underset{\text{ops.}}{\underset{\uparrow}{n_i}})\}_{i=1}^{k}$$

$$x_k|\lambda, n_k \sim \mathit{Po}(\lambda n_k),$$
$$\lambda \sim \mathit{Ga}(a, p),$$

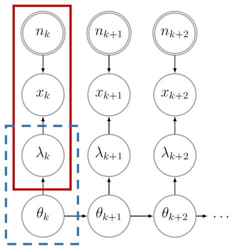
Effects

- Stress
- Seasonal
- Trend
- Group
- Dependence ops.
- Uncertain
- Underreporting
- Severities

Model	Types of occurrences
No Effect	25
Stress	1
Seasonal	1
Trend	44
Stress+Seasonal	3
Seasonal+Trend	12
	86

Variants. DLMs (trend, seasonality)

$$x_k | \lambda_k, n_k \sim \mathsf{Po}(\lambda_k n_k), \qquad \lambda_k = \exp(u_k),$$
$$u_k = \boldsymbol{F}_k \boldsymbol{\theta}_k + v_k, \qquad v_k \sim \mathcal{N}(0, V_k),$$
$$\boldsymbol{\theta}_k = \boldsymbol{G}_k \boldsymbol{\theta}_{k-1} + \boldsymbol{w}_k, \quad \boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W}_k),$$
$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{C}_0),$$



Influence diagram with log-rate as DLM

Variants. DLMs (trend, seasonality)

Particle filter:

- *Step 0. Forecast at beginning of k-th period, before observing x_k*

$$\left.\begin{array}{l} \theta_k^1 = \begin{pmatrix} \vdots \end{pmatrix} \to \overbrace{u_k^1}^{N(F_k\theta_k^1, V_k)} \to \lambda_k^1 \\ \vdots \\ \theta_k^N = \begin{pmatrix} \vdots \end{pmatrix} \to u_k^N \to \lambda_k^N \end{array}\right\} \approx \pi(x_k | n_k)$$

Variants. DLMs (trend, seasonality)

Particle filter:

- *Step 0. Forecast at beginning of k-th period, before observing x_k*
- *Step 1. Observation of x_k and update*

$$
\left.
\begin{array}{l}
\left.
\begin{array}{lll}
 & \nearrow & u_k^{1,1} \to \lambda_k^{1,1} \\
\theta_k^1 & \to & \vdots \\
 & \searrow & u_k^{1,N} \to \lambda_k^{1,N}
\end{array}
\right\} \approx \pi(x_k|\theta_k^1) \to \overline{\mathsf{w}}_k^1 = \mathsf{w}_{k-1}^1 \pi(x_k|\theta_k^1) \\
\vdots \\
\theta_k^N \to \quad \ldots \quad \} \quad \ldots \quad \to \overline{\mathsf{w}}_k^N
\end{array}
\right\} \mathsf{w}_k^i = \frac{\overline{\mathsf{w}}_k^i}{\sum_{j=1}^N \overline{\mathsf{w}}_k^j}
$$

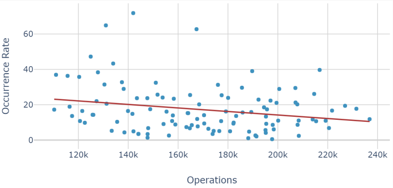
$$
N_{\text{eff}} = \underbrace{\frac{1}{\sum_{i=1}^N (\mathsf{w}_k^i)^2}}_{\in[1,N]} < \alpha
$$

Particle filter:

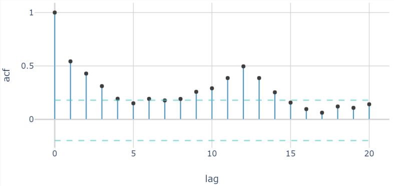
- *Step 0. Forecast at beginning of k-th period, before observing x_k*
- *Step 1. Observation of x_k and update*
- *Step 2. Propagation to period $(k+1)$*

$$\left.\begin{array}{l} \theta_k^1 \xrightarrow{\theta_k = G_k \theta_{k-1} + W_k} \theta_{k+1}^1 \\ \vdots \\ \theta_k^N \xrightarrow{\hspace{3cm}} \theta_{k+1}^N \end{array}\right\} \approx \pi(\theta_{k+1} | D_k)$$

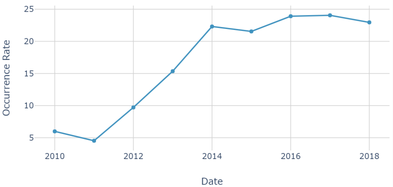
Example. Wind shear occurrences



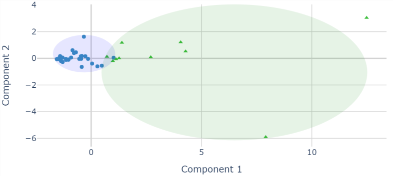
No stress effect



Seasonal effect



Linear trend



Group effect

Effect analysis for *wind shear*.

Example. Wind shear occurrences

Hierarchical model

$$
\begin{aligned}
&x_k^i|\lambda_k^i, n_k^i \sim Po(\lambda_k^i n_k^i), \quad \lambda_k^i = \exp(u_k^i), \qquad i = 1,2 \\
&u_k^i = \boldsymbol{F}\boldsymbol{\theta}_k^i + v^i, \quad v^i \sim \mathcal{N}(0, V^i), \\
&\boldsymbol{\theta}_k^i = \boldsymbol{G}\boldsymbol{\theta}_{k-1}^i + \boldsymbol{w}^i, \quad \boldsymbol{w}^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}^i), \\
&\boldsymbol{\theta}_0^i \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{C}_0),
\end{aligned}
$$

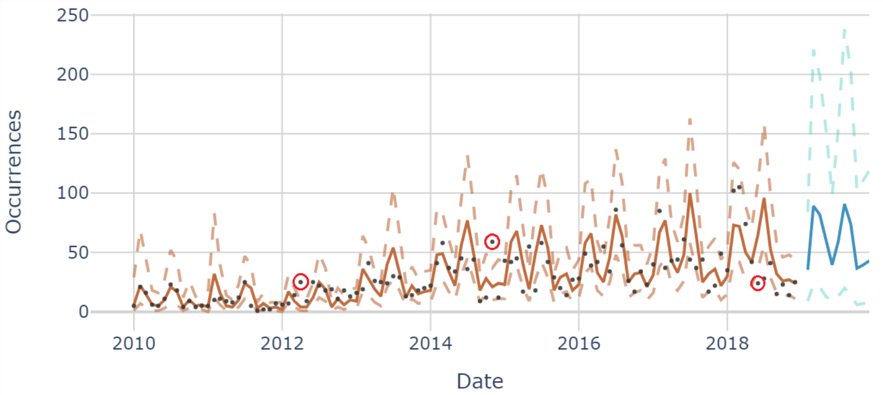
$\boldsymbol{F}$ and $\boldsymbol{G}$ have linear growth and period 12 seasonal components.

$\boldsymbol{W}^i$ and V^i estimated using maximum likelihood.

$\boldsymbol{C}_0 = \boldsymbol{I}_{13}$ and $\boldsymbol{m}_0$ initialized with first year of data (12 obs.)

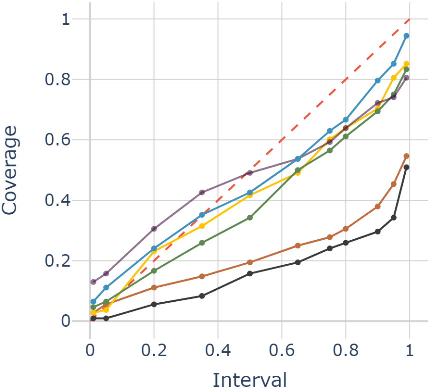
Example. Wind shear occurrences

Particle filter provides samples of predictive distribution $(x_k = x_k^1 + x_k^2 | D_k)$



Example. Wind shear occurrences

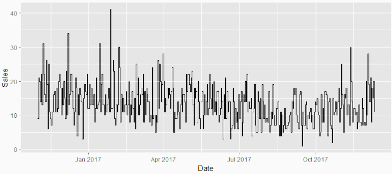
	This Model	GLARMA (Poi)	GLARMA (NB)	INGARCH (Poi)	INGARCH (NB)	DLM
MSE	296.74	350.33	392.35	425.51	471.22	379.7
MAE	11.96	13.35	14.03	14.81	15.37	14.53
MAPE	0.53	0.63	0.65	0.61	0.59	0.8
Theil's U	0.91	0.94	0.99	1.03	1.09	1.03



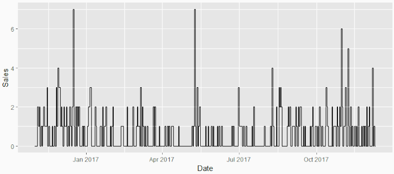
Coverage of credible intervals

Contents

Supermarket Demand Forecasting



High demand product.



Low demand product.

Large retail company

- Thousands of stores
- 4 countries
- Thousands of references per store
- No stock beyond shelves

Effects

- Trends (New items, vegan,...)
- Seasonality
- Overdispersion
- Price
- Promotions
- Substitute goods

DGLM Mixture

$\{\mathbf{y_t}\}$ original time series, $\{\mathbf{z_t}\}$ binary time series with $z_t = \mathbb{1}_{(y_t > 0)}$

$$z_t \sim Ber(\pi_t) \quad \text{and} \quad y_t | z_t = \begin{cases} 0, & \text{if } z_t = 0 \\ 1 + x_t, \quad x_t \sim \textit{Neg-Bin}(r_t, p_t) & \text{if } z_t = 1 \end{cases}$$

with λ_t for the Ber and Neg-Bin DGLMs: logit$(\pi_t) = F_t^{0\prime} \theta_t^0$,
$\log(p_t) = F_t^{+\prime} \theta_t^+$.

Fitting and forecasting procedure

- **One step ahead forecast for state** $\theta_t|\mathcal{D}_{t-1} \sim (a_t, R_t)$

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t' + W_t$$

- **One step ahead forecast for observations**: $(f_t = F_t' a_t,\ Q_t = F_t' R_t F_t)$

$$z_t|\mathcal{D}_{t-1} \sim Ber\left(\frac{\alpha_t^0}{\alpha_t^0 + \beta_t^0}\right) \quad \text{and} \quad x_t|\mathcal{D}_{t-1} \sim BNB(\beta_t^+ r_t + 1, \alpha_t^+, r_t)$$

with $\alpha_t^0, \alpha_t^+, \beta_t^0, \beta_t^+$ satisfying $f_t = E[\lambda_t|D_{t-1}]$, $Q_t = V[\lambda_t|D_{t-1}]$

- **Update after observing $\mathbf{y_t}$**, $\theta_t|\mathcal{D}_t \sim (m_t, C_t)$

$$m_t = a_t + R_t F_t Q_t^{-1}(\hat{f}_t - f_t) \qquad C_t = R_t - R_t F_t (Id - Q_t^{-1}\hat{Q}_t) Q_t^{-1} F_t' R_t$$

with $\hat{\alpha}_t = \alpha_t + \phi_t T(y_t)$, $\hat{\beta}_t = \beta_t + \phi_t$.

Resulting mixture distribution for observations:

$$p(y_t|D_t, \pi_t) = (1 - \pi_t)\delta_0(y_t) + \pi_t h_t(y_t - 1)$$

with $(\pi_t|D_t) \sim Be(\alpha_t^0, \beta_t^0)$, $h_t(y_t - 1) = BNB(y_t - 1|\beta_t^+ r_t + 1, \alpha_t^+, r_t)$.

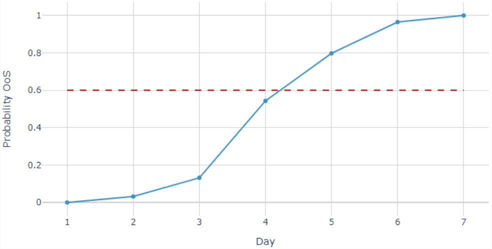
Out of Stock (OoS) events

Entail negative consequences.

$$stock_t = stock_{t-1} - sales_t + resupply_t$$

$$P(demand_t \geq stock_{t-1} + resupply_t)$$

$$p(OoS_i) = 1 - \sum_{j=0}^{stock_t} p(y_{t+1} + ... + y_{t+i} = j | D_t)$$



Out of Stock (OoS) probability for beer (SKU '182')

Out of Stock (OoS) events

Criteria

- Threshold α

$$A = \{i \in [1, ..., k] : P(OoS_i) \geq \alpha\}$$

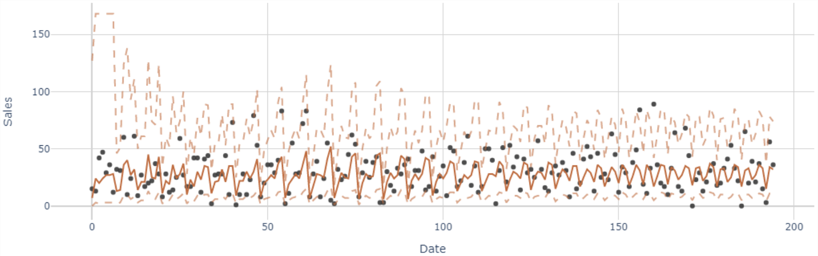
- Expected profit ($EProfit_i < 0$)

$$EProfit_i := (1 - P(OoS_i)) \times (\widehat{d}_i\, c) - P(OoS_i) \times (\widehat{d}_i - stock_t) f$$

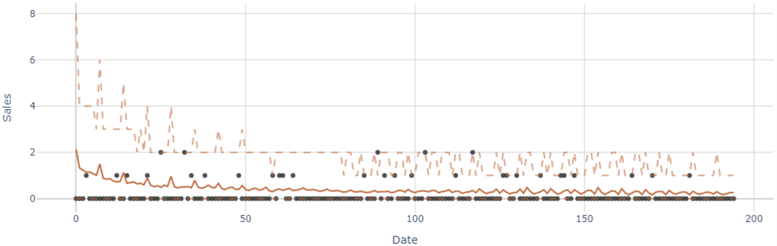
- Utility function with risk aversion ($EUtil_i < 0$)

$$EUtil_i := (1 - P(OoS_i)) \times E[u(d_i\, c)] - P(OoS_i) \times E[u((d_i - stock_t) f)]$$

Example



Beer 1



Shaving gel

Multivariate model

Similar to the univariate model:

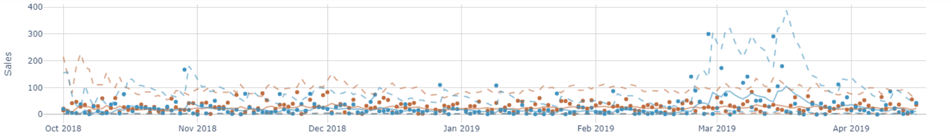
$y_t = \{\mathbf{y_{it}}\}_{i=1}^{m}$ original ts, $z_t = \{\mathbf{z_{it}}\}_{i=1}^{m}$ binary ts with $z_{it} = \mathbb{1}_{(y_{it}>0)}$

$$z_t \sim \textit{MBer}(p_t)$$
$$\textit{softmax}^{-1}(p_t) = F_t^{0\prime}{}_t\theta_t^0,$$
$$\theta_t = G_t^0\theta_{t-1}^0 + \omega_t^0$$

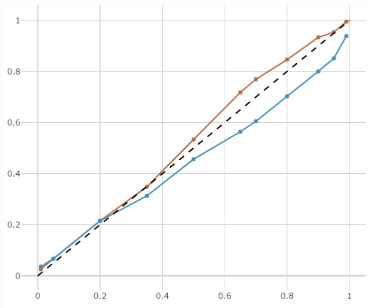
$$y_t \sim \textit{MNB}(r_t, \mu_t)$$
$$\log(\mu_t) = F_t^{+\prime}{}_t\theta_t^+,$$
$$\theta_t^+ = G_t^+\theta_{t-1}^+ + \omega_t^+$$

with MBer a multivariate Bernoulli and MNB a multivariate negative binomial

Fitting and forecasting procedure



Forecasts for Beer 1 and Beer 2

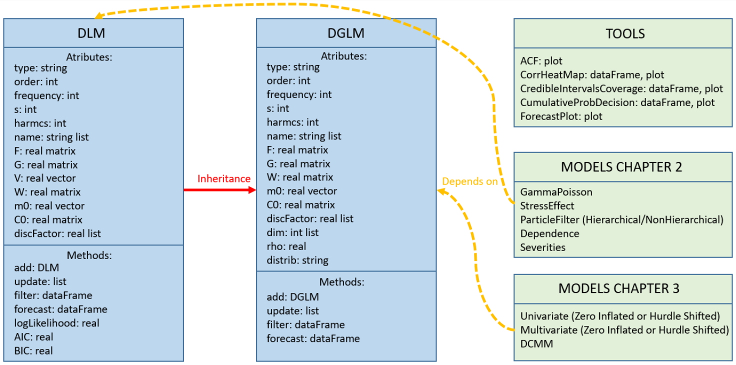


Coverage plot of credible intervals for Beer 1 and Beer 2

Python package

Object-Oriented Programming (OOP) paradigm

Competitors/Alternatives: `dlm` in R, and `pyDLM`, `pyBATS` in python.



UML diagram of two main classes in package (DLM and DGLM) in blue. Additional functions in green.

Contents

Conclusions

- SSMs provide a flexible framework for **inferring hidden states** (e.g., brain activity, machine condition) and **predicting future observations** (e.g., sales, weather)
- Multiple posibilities to **infer hidden states**. Baum-Welch (EM), SGD, spectral methods, blocked Gibbs.
- LDSs/DLMs allow **modular construction** of interpretable components (trend, seasonality, covariates) via block-diagonal matrices.
- DGLMs extend DLMs to **non-Gaussian observations** through exponential family distributions and moment matching
- Applications to aviation incidents and supermarket demand show SSMs are **competitive with state-of-the-art** models while offering richer probabilistic outputs.

Questions?

Thank you!

Questions?

bruno.fl.ba@gmail.com

Next Session

Decisions (Ch. 34)

May 13, 2026

José Manuel Camacho