DSC 383: Advanced Predictive Models for Complex Data

Section: Time Series Analysis > Subsection: State-Space and Hidden Markov Models

INSTRUCTOR:
Catherine (Kate) Calder
Department of Statistics & Data Sciences
University of Texas at Austin

PRACTICE: Bayesian Sequential Updating

Question:

The hematocrit level is the volume percentage of red blood cells in blood. Let θ_j be the true hematocrit level of a bone marrow transplant patient j days after surgery, and y_j be the measured hematocrit level j days after surgery. Hematocrit levels cannot be measure perfectly, but from laboratory quality control testing it is reasonable to assume that $y_j = \theta_j + \nu_j$, where $v_j \overset{iid}{\sim} N(0, 0.2)$. Based on her experience with other bone marrow transplant patients, the patient's doctor assumes that $\theta_1 \sim \text{N}(30, 2)$ before measuring the patient's hematocrit level the day after transplantation. Assume that θ_1 is independent of v_j for all j.

a) The predictive distribution of y_1 is normal with mean f_1 and variance Q_1. What are the numeric values of f_1 and Q_1?

b) The hematocrit level of the patient one day after surgery, y_1, is observed to be 32. The posterior distribution of θ_1 given $y_1 = 32$ is normal with mean m_1 and variance C_1? What are the numeric values of m_1 and C_1?

c) Assume that $\theta_j = \theta_{j-1} + \omega_j$, for $j = 2, \ldots,$ where $\omega_j \overset{iid}{\sim} \mathrm{N}(0, 0.2)$ and ω_j and ν_j are mutually independent. The predictive distribution of θ_2 given $y_1 = 32$ is normal with mean a_2 and variance R_2. What are the numeric values of a_2 and R_2?

LECTURE: State-Space Models

Definition: A state-space model consists of an unobserved $\mathbb{R}^p$-valued time series $\{\boldsymbol{\theta}_t : t = 0, 1, \ldots\}$ and a (partially) observable $\mathbb{R}^m$-valued times-series $\{\boldsymbol{y} : t = 1, 2, \ldots\}$ that satisfy the following

i) $\boldsymbol{\theta}_t$ is a Markov chain

ii) Conditional on $\{\boldsymbol{\theta}_t : t = 0, 1, \ldots\}$, the $\boldsymbol{y}_t$s are independent of each other and depend on $\boldsymbol{\theta}_t$ only

$$
\begin{tikzcd}
\theta_0 \arrow[r] & \theta_1 \arrow[r] \arrow[d] & \theta_2 \arrow[r] \arrow[d] & \cdots \arrow[r] & \theta_{t-1} \arrow[r] \arrow[d] & \theta_t \arrow[r] \arrow[d] & \theta_{t+1} \arrow[r] \arrow[d] & \cdots \\
 & Y_1 & Y_2 & & Y_{t-1} & Y_t & Y_{t+1} &
\end{tikzcd}
$$

OVERVIEW

- State-space models are a very general class of time-series models that include many special case of interest
- They were originally applied in problems in aerospace, but now are widely used across many applications
- They are particularly useful in the following situations:
  - time-series contains missing data
  - time-series observed with measurement error
  - time-series consisting of indirect measurement of a unobserved process of interest
  - massive data settings
  - forecasting using streaming data
- State-space models are very natural in a Bayesian setting

SPECIAL CASE: THE DYNAMIC LINEAR MODEL (DLM)

For $t = 1, 2, \ldots$

$$\boldsymbol{y}_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim \mathrm{N}_q(0, \boldsymbol{V}_t)$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + w_t, \quad \boldsymbol{w}_t \sim \mathrm{N}_p(0, \boldsymbol{W}_t)$$

$$\boldsymbol{\theta}_0 \sim \mathrm{N}_p(\boldsymbol{m}_0, \boldsymbol{C}_0)$$

where

- $\boldsymbol{G}_t$ is a *known* $p \times p$ matrix
- $\boldsymbol{F}_t$ is a *known* $q \times p$ matrix
- $\boldsymbol{\theta}_0$, $\boldsymbol{v}_t$, $\boldsymbol{w}_t$ are mutually independent

FILTERING

- **Notation:** Let $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t\}$
- **Kalman Filter**

  Let $\boldsymbol{\theta}_t | \boldsymbol{y}_{1:(t-1)} \sim N_p(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$.

  i) The one-step-ahead predictive distribution of $\boldsymbol{\theta}_t$ given $\boldsymbol{y}_{1:(t-1)}$ is Gaussian with parameters

ii) The one-step-ahead predictive distribution of $\boldsymbol{y}_t$ given $\boldsymbol{y}_{1:(t-1)}$ is Gaussian with parameters

iii) The filtering distribution of $\boldsymbol{\theta}_t$ given $\boldsymbol{y}_{1:t}$ is Gaussian with parameters

- Filtering with missing data

- **Forecasting**: Let $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$. Then, for $k \geq 1$

   i) the distribution of $\boldsymbol{\theta}_{t+k}$ given $\boldsymbol{y}_{1:t}$ is Gaussian with

   ii) the distribution of $\boldsymbol{y}_{t+k}$ given $\boldsymbol{y}_{1:t}$ is Gaussian with

SMOOTHING

- **Kalman Smoother**

  If $\boldsymbol{\theta}_{t+1}|\boldsymbol{y}_{1:T} \sim \mathrm{N}_p(\boldsymbol{s}_{t+1}, \boldsymbol{S}_{t+1})$, then for $t < T$

  $$\boldsymbol{\theta}_t|\boldsymbol{y}_{1:T} \sim \mathrm{N}_p(\boldsymbol{s}_{t+1}, \boldsymbol{S}_{t+1}),$$

  where

- Smoothing with missing data

PARAMETER ESTIMATION

- Unknown parameters in the DLM: $\boldsymbol{\psi} = \{\boldsymbol{V}_t,\ \boldsymbol{W}_t\, \boldsymbol{m}_0,\ \boldsymbol{C}_0\}$

  * Often, we assume

$$\boldsymbol{V}_t = \boldsymbol{V} = \mathrm{diag}(v_1, \ldots, v_m)$$

$$\boldsymbol{W}_t = \boldsymbol{W} = \mathrm{diag}(w_1, \ldots, w_m)$$

- Likelihood function:

PARAMETER ESTIMATION

- Maximum likelihood estimation of ψ

COMPUTING DEMO: Dynamic Linear Models

LECTURE: Regression DLMs

EXAMPLE: DYNAMIC CAPM

- The Capital Asset Pricing Model (CAPM) is a well known asset pricing tool that assumes that returns on an asset depend linearly on overall market returns
- **Notation:**

  r_t - returns at time t on asset of interest (e.g., an investment portfolio)

  r_t^M - returns at time t of the market (e.g., SP500 index)

  r_t^F - returns at time t on a risk free asset (e.g., treasury bonds)

  Excess returns of the asset: $y_t = r_t - r_t^F$

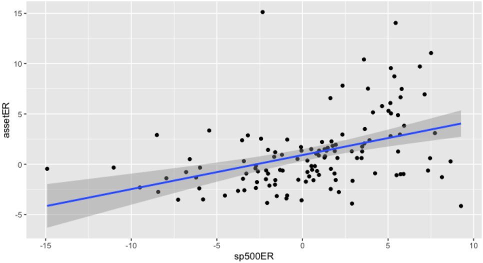
  Excess returns of the market: $x_t = r_t^M - r_t^F$

- The standard CAPM model assumes that

$$y_t = \alpha + \beta x_t + \nu_t, \quad \nu_t \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

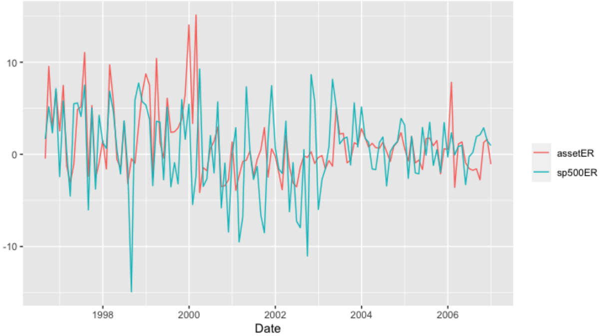
Interpretation of β:

- Consider the returns of a hypothetical asset manager, `HAM2`, in the R library `PerformanceAnalytics`.




```
library("PerformanceAnalytics")
assetER <- 100 * as.vector(managers[,"HAM2"] - managers[,"US 3m TR"])
sp500ER <- 100 * as.vector(managers[,"SP500 TR"] - managers[,"US 3m TR"])
CAPM_dat <- data.frame("assetER" = assetER,
                       "sp500ER" = sp500ER,
                       "date" = rownames(as.data.frame(managers)))
ggplot(data = CAPM_dat,
                     aes(y = assetER,
                         x = sp500ER)) +
  geom_point() +
  stat_smooth(method = lm)
```

- Time-varying relationship between assetER and sp500R?




```
CAPM_dat_long <- CAPM_dat[!is.na(assetER),]  %>%
  pivot_longer(cols = -date,
               names_to = "type",
               values_to = "y")
ggplot(data = CAPM_dat_long,
           aes(x = as.Date(date), y = y, color = as.factor(type))) +
  geom_line() +
  title("Excess Returns") + ylab("") + xlab("Date") + labs(color = "")
```

- Regression DLM CAPM:

$$y_t = \alpha_t + \beta_t x_t + \nu_t, \quad \nu_t \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

$$\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathrm{N}_2(0, \boldsymbol{V}), \quad \boldsymbol{V} = \mathrm{diag}([\sigma_\alpha^2, \sigma_\beta^2])$$

$$\begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} \sim \mathrm{N}_2(0, \mathrm{diag}([1 \times 10^7, 1 \times 10^7]))$$

Unobserved states:

Unknown parameters:

- Constructing the dlmModReg object and estimating the unknown parameters:

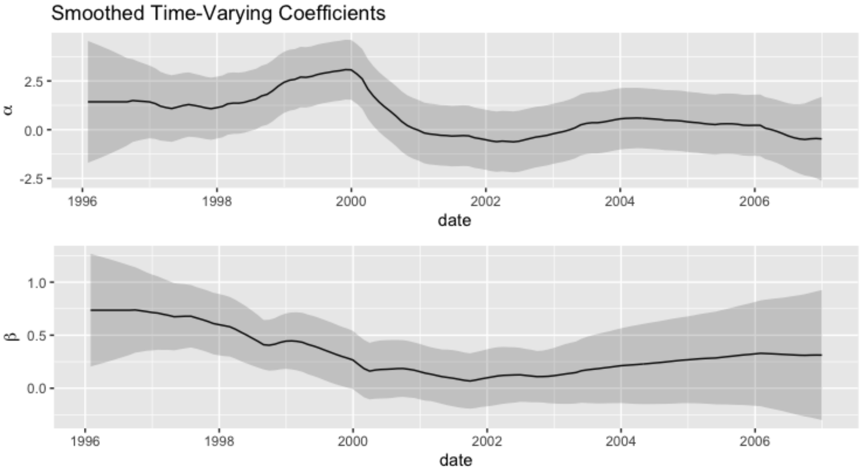
```
mod_regDLM <- dlmModReg(X = CAPM_dat$sp500ER,
                        addInt=TRUE,
                        dV=0.1,
                        dW=c(0.01, 0.01))
build_mod <- function(parm, x.mat){
   parm <- exp(parm)
   return(dlmModReg(X = x.mat,
                    dV = parm[1],
                    dW = c(parm[2], parm[3])))
}
modMLE <- dlmMLE(y = CAPM_dat$assetER,
                 parm = c(0, 0, 0),
                 x.mat = CAPM_dat$sp500ER,
                 build = build_mod,
                 hessian=T)
parm_est <- sqrt(exp(modMLE$par))
names(parm_est) = c("sigma", "sigma_alpha", "sigma_beta")
```

```
> sqrt(parm_est)
      sigma sigma_alpha  sigma_beta
  1.6980095   0.6486200   0.2498722
```

- Filtering and smoothing

```
fitted_dlm <- build_mod(modMLE$par, CAPM_dat$sp500ER)

CAPM_filtered <- dlmFilter(y = CAPM_dat$assetER,
                           mod = fitted_dlm)
CAPM_smoothed <- dlmSmooth(CAPM_filtered)

se_mat <- dropFirst(t(sapply(dlmSvd2var(CAPM_smoothed$U.S, CAPM_smoothed$D.S),
                   function(x) sqrt(diag(x)))))

a_dat <- data.frame(alpha_s = dropFirst(CAPM_smoothed$s[,1]),
                    a.u = dropFirst(CAPM_smoothed$s[,1]) + 1.96*se_mat[,1],
                    a.l = dropFirst(CAPM_smoothed$s[,1]) - 1.96*se_mat[,1],
                    date = as.Date(CAPM_dat$date))

b_dat <- data.frame(beta_s = dropFirst(CAPM_smoothed$s[,2]),
                    b.u = dropFirst(CAPM_smoothed$s[,2]) + 1.96*se_mat[,2],
                    b.l = dropFirst(CAPM_smoothed$s[,2]) - 1.96*se_mat[,2],
                    date = as.Date(CAPM_dat$date))
```

- Results



Interpretation?

```r
gg_a <- ggplot(data = a_dat,
       aes(y = alpha_s, x = date)) +
  geom_line() +
  geom_ribbon(aes(ymin = a.l,
                  ymax = a.u),
              alpha = .2) +
  labs(title = 'Smoothed Time-Varying Coefficients') +
  xlab("date") +
  ylab(expression(alpha))

gg_b <- ggplot(data = b_dat,
               aes(y = beta_s, x = date)) +
  geom_line() +
  geom_ribbon(aes(ymin = b.l,
                  ymax = b.u),
              alpha = .2) +
  xlab("date") +
  ylab(expression(beta))

grid.arrange(gg_a, gg_b, nrow = 2)
```

LECTURE: Hidden Markov Models

OVERVIEW

- Hidden Markov models (HMMs) originated in the signal processing literature, but are now widely applied in a various areas of application

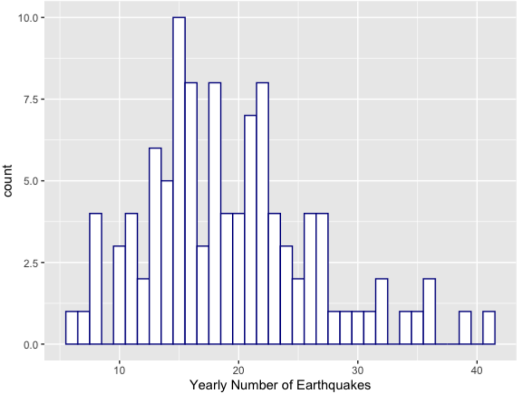
- Similar to state-space models, but the state/latent process is discrete

- Other names...

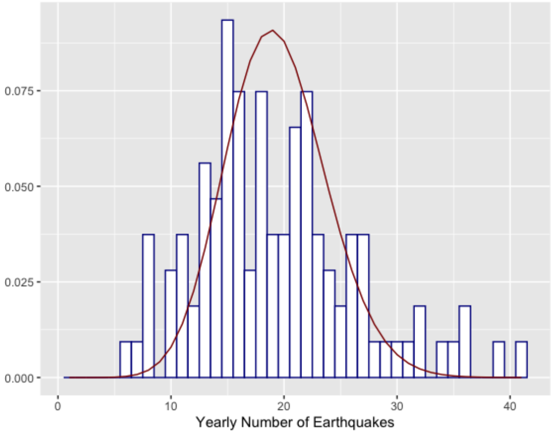
FINITE MIXTURE MODELS

- **Earthquake Example** (From Zucchini, MacDonald, and Langrock, 2016)

  Yearly count of earthquakes with magnitudes greater than 7 from 1900-2006



- Poisson Model: $y \overset{iid}{\sim} \text{Pois}(\hat{\lambda})$, where $\hat{\lambda} = \bar{y}$



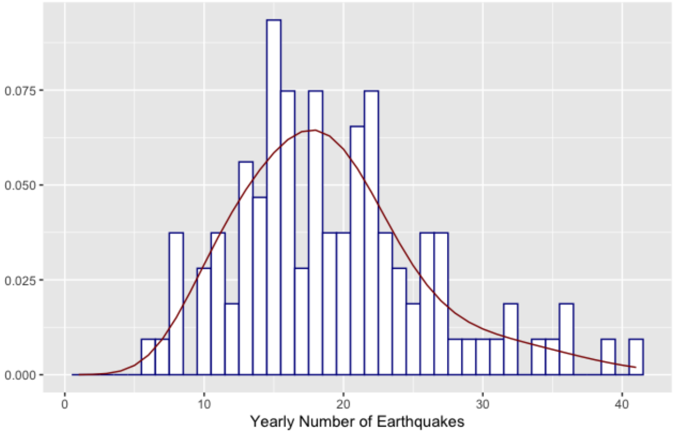
- A 2-component mixture model:

- An m-component mixture model:

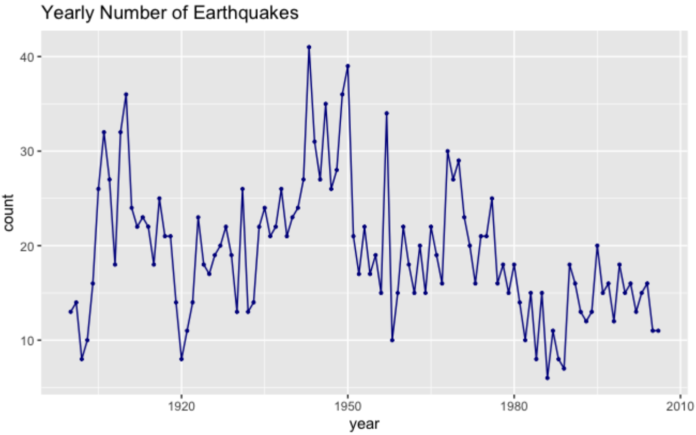
* Parameter estimation can be performed using maximum likelihood estimator or the EM algorithm (implemented in the `flexmix` R package)

- A 4-component mixture model for the earthquake data:

$$\hat{\delta} = (0.41, 0.11, 0.24, 0.23) \quad \hat{\lambda} = (18.66, 31.57, 12.44, 20.59)$$



- Serial dependence...



MARKOV CHAIN

- **Definition:** A discrete-time Markov chain is a sequence of random variables $\{C_t : t \in \mathbb{N}\}$ that satisfy the Markov property

$$P(C_{t+1}|C_{1:t}) = P(C_{t+1}|C_1, \ldots, C_t) = P(C_{t+1}|C_t)$$

  where $C_t \in \{1, \ldots, m\}$ for all t

- The transition probability of a Markov chain is $P(C_{s+t} = j|C_t = i)$

  $\rightarrow$ If the transition probability doesn't dependent on s than

$$\gamma_{ij}(t) = P(C_{s+t} = j|C_t = i)$$

  and the Markov chain is called homogeneous

- The matrix $\mathbf{\Gamma}(t)$ with (i,j) entry $\gamma_{ij}(t)$ is called the transition probability matrix

- The unconditional probability $P(C_t = j)$ is

$$\mathbf{u}(t) = (P(C_t = 1), \ldots, P(C_t = m))$$

and

$$\mathbf{u}(t+1) = \mathbf{u}(t)\mathbf{\Gamma}$$

- A Markov chain with transition probability matrix $\mathbf{\Gamma}$ is said to have stationary distribution $\boldsymbol{\delta}$ if

$$\boldsymbol{\delta} = \boldsymbol{\delta}\mathbf{\Gamma} \quad \text{and} \quad \boldsymbol{\delta}1^{\top} = 1$$

- A stationary homogeneous Markov chain has initial distribution $\boldsymbol{u}(1) = \boldsymbol{\delta}$

HIDDEN MARKOV MODEL

- A hidden Markov model (HMM) for $\{X_t : t \in \mathbb{N}\}$ is a type of dependent mixture model satisfying

$$P(X_t | X_{1:(t-1)}, C_{1:t}) = P(X_t | X_{t-1}, C_t)$$

$$P(C_t | C_{1:(t-1)}) = P(C_t | C_{t-1})$$