Lectures on

Stochastic systems

Contents

1 Introduction, probability, system **3**

1.1 Revision . . . 3

1.2 System and its variables . . . 4

2 Differential equations, regression model **5**

2.1 Differential equations . . . 5

2.2 Discretization . . . 6

2.3 Regression model . . . 7

2.4 Regression model in the state-space form . . . 8

3 Discrete and logistic models **9**

3.1 Discrete model . . . 9

3.2 Scilab generations . . . 10

3.3 Logistic model . . . 11

4 Estimation of regression model[1] **13**

5 Estimation of specific models **16**

5.1 Normal regression model . . . 16

5.2 Categorical model . . . 18

5.3 Model of logistic regression . . . 19

[1] two lectures

6 Prediction **21**

6.1 Output estimation (zero step prediction) 21

6.2 One step prediction 21

6.3 Multi-steps prediction 22

6.4 Prediction with discrete model 23

7 State-space model, state estimation **24**

7.1 Model 24

7.2 Estimation 24

8 Nonlinear state estimation **26**

8.1 Nonlinear model 26

8.2 Model with unknown parameters 28

9 Estimation of mixtures **29**

9.1 Mixture model 29

9.2 Estimation 29

10 Control with regression model **33**

10.1 Derivation in pdf 33

10.2 Derivation for regression model 35

11 Control with categorical model **37**

11.1 Optimization 37

11.2 Application 39

12 Appendix **40**

12.1 Elementary differential equations 40

12.2 Elementary difference equations 41

12.3 Discretization of a continuous model 43

1 Introduction, probability, system

www.fd.cvut.cz/personal/nagyivan + Stochastic systems

1.1 Revision

- **Variable** $\times$ **Random variable** (continuous, discrete)

  Remark: Variables are (i) continuous, (ii) discretized (ordinal), (iii) discrete (nominal) - can be ordered according to something (frequently money, some loss).

- **Distribution** (pf, pdf)

  - discrete: $f(x) \equiv P(X = x)$
  - continuous: $f(x) \equiv \lim P(O_x)/m(O_x)$ for $m(O_x) \to 0$, where $m(O_x)$ is a measure of the neighborhood O_x around the point x

- **Random vector**, joint; marginal; conditional distribution
  draw continuous and discrete uniform distribution for $X = [x_1, x_2]$

$$f(x_1, x_2) = f(x_1) f(x_2|x_1) = f(x_2) f(x_1|x_2)$$

Example

Discrete case

$f(x_1, x_2)$

$x_1 \backslash x_2$	1	2	$f(x_1)$
1	0.1	0.3	0.4
2	0.4	0.2	0.6
$f(x_2)$	0.5	0.5	

$f(x_2|x_1)$

$\frac{1}{4}$	$\frac{3}{4}$
$\frac{2}{3}$	$\frac{1}{3}$

$f(x_1|x_2)$

$\frac{1}{5}$	$\frac{3}{5}$
$\frac{4}{5}$	$\frac{2}{5}$

$f(x_1) f(x_2)$

0.2	0.2
0.3	0.3

Continuous case

$$f(x_1, x_2) = 6x_1^2 x_2, \quad x_1, x_2 \in (0,\, 1)$$

$$f(x_1) = \int_0^1 6x_1^2 x_2 dx_2 = 3x_1^2$$

$$f(x_2) = \int_0^1 6x_1^2 x_2 dx_1 = 2x_2$$

$$f(x_1|x_2) = \frac{6x_1^2 x_2}{2x_2} = 3x_1^2$$

$$f(x_2|x_1) = \frac{6x_1^2 x_2}{3x_1^2} = 2x_2$$

As it is $f(x_1, x_2) = f(x_1) f(x_2)$ the variables are independent.

- **Characteristics**

$$E[X] = \begin{bmatrix} E[x_1] \\ E[x_2] \end{bmatrix}, \quad C[X] = \begin{bmatrix} D[x_1] & \mathbf{cov}[x_1, x_2] \\ \mathbf{cov}[x_1, x_2] & D[x_2] \end{bmatrix}$$

- **Random process** is random variable indexed by time

time \ values	discrete	continuous
discrete	Markov chains	random sequences
continuous	queues	x

- **Categorical distribution**

x	1	2	$\cdots$	n
$f(x)$	p_1	p_2	$\cdots$	p_n

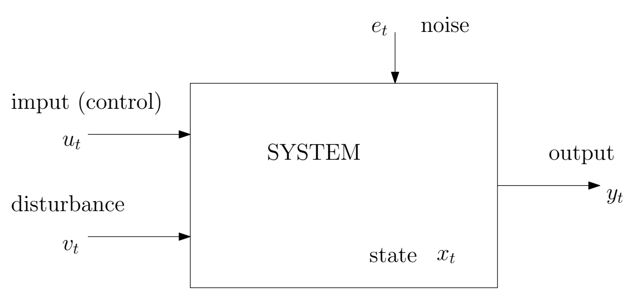
where $p_1 \geq 0$, $\sum p_i = 1$. Each realization has its probability.

- **Normal distribution**

$$f(X) = \frac{1}{\sqrt{(2\pi)^n |R|}} \exp\left\{-\frac{1}{2}(x-\mu)' R^{-1} (x-\mu)\right\}$$

1.2 System and its variables

System is a part of reality we are interested in, on which we measure data and which we want to learn about to be able to predict its behavior or influence it.



- output - the modeled variable, after application of the control it can be measured
- input - variable that influences the output and that can be fully manipulated by us
- disturbance - can be measured, cannot be influenced
- state - is influenced by input, influences output, cannot be measured
- noise - can be neither measured nor predicted

2 Differential equations, regression model

2.1 Differential equations

Dynamic process is described by differential equation

- **stationary** - constant coefficients
  - **homogeneous**: zero right-hand side (characteristic equation)
  - controlled: model with control variable (variation of constant)

First order

$$y' + ay = 0, \;\; y(0) = y_0$$

- Laplace
$$pY - y_0 + aY = 0$$
$$(p + a)\, Y = y_0$$
$$Y = y_0 \frac{1}{p + a} \to y(t) = y_0 \exp\{-at\}$$
- by guess
$$y = \alpha \exp\{\lambda t\}$$
characteristic equation
$$\lambda + a = 0 \;\; \to \;\; \lambda = -a$$
substitution
$$y = \alpha \exp\{-at\}$$
α according to initial condition
$$y(0) = \alpha = y_0$$
the solution is
$$y = y_0 \exp\{-at\}$$

Second order

$$y'' + a_1 y' + a_0 y = 0, \quad y(0) = y_0, \;\; y'(0) = d_0$$

Characteristic equation

$$\lambda^2 + a_1 \lambda + a_0 = 0$$

Solution

1. two real roots - two exponentials
2. one double root - exponential and polynomial
3. two complex roots - exponentials and sine, cosine

Stability - real parts of the roots must be in the left half-plane.

2.2 Discretization

Approximate

$$y' + ay = 0$$

$y'(t) \to \frac{y(t+T)-y(t)}{T} = \frac{y_{t+1}-y_t}{T}$ - T is step

$$\frac{y(t+T)-y(t)}{T} + ay(t) = 0$$
$$y_{t+1} - y_t + Tay_t = 0 \;\to\; y_{t+1} = \underbrace{(1-Ta)}_{\tilde{A}} y_t$$

Precise

Discretization: $\tau = t_0 + tT$ - t_0 is initial time, t discrete time, T period (t_0 often 0).
Notation: $y(tT) \equiv y_t$.
Equation

$$y'(\tau) + ay(\tau) = 0$$
$$\to y(\tau) = y_0 \exp\{-a\tau\} \quad \text{solution}$$

at time τ and $\tau + T$

$$y(tT) = y_0 \exp\{-atT\}$$
$$y(tT+T) = y_0 \exp\{-a(tT+T)\} = \underbrace{y_0 \exp\{-atT\}}_{y(tT)} \exp\{-aT\} = \exp\{-aT\} y(tT)$$

notation

$$y_{t+1} = \exp\{-aT\} y_t = Ay_t$$

Solution

$$\begin{aligned} y_1 &= Ay_0 \\ y_2 &= Ay_1 = A^2 y_0 \\ &\cdots \\ y_t &= A^t y_0 \end{aligned}$$

Stability: inside of unit circle in complex plane.

2.3 Regression model

$$y_t = \psi_t' \Theta + e_t$$

- y_t modeled variable (output) at time t
- ψ_t regression vector, containing samples of variables influencing the output
- Θ model parameters (regression coefficients θ and noise variance r)
- e_t noise, with zero expectation, constant variance, independent of variables in regression vector = sequence of independent and identically distributed r.v. = i.i.d.

$$\begin{aligned} \psi_t &= [u_t, y_{t-1}, u_{t-1} \cdots y_{t-n}, u_{t-n}, 1]' \\ \theta &= [b_0,\ a_1,\ b_1, \cdots\ a_n,\ b_n,\ k]', \end{aligned}$$

Model in detail

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 y_{t-1} + \cdots + a_n y_{t-n} + b_n u_{t-n} + k + e_t$$

Comments

1. Number of delayed y and u can be different. Number of delayed y is called **model order.**
2. The term $\psi_t' \theta$ is at time t known constant. Model represents a transformation of e_t to y_t according to the model equation.
3. If ψ_t contains no delayed outputs, the model is static. Otherwise, it is dynamic.
4. $y_t = \psi_t' \theta$ represents a difference equation.

A general description of the model as a tool, describing y_t as random variable is distribution

$$f\left(y_t | \psi_t, \Theta\right)$$

Moments of the model are

$$E\left[y_t | \psi_t, \Theta\right] = E\left[\psi_t' \theta + e_t\right] = \psi_t' \theta \equiv \hat{y}_t$$

$$D\left[y_t | \psi_t, \Theta\right] = D\left[\psi_t' \theta + e_t\right] = D\left[e_t\right] = r$$

Normal regression model

$$f\left(e_t\right)=\frac{1}{\sqrt{2\pi r}}\exp\left\{-\frac{1}{2r}e_t^2\right\}$$

transformation: $y_t=\hat{y}_t+e_t\rightarrow e_t=y_t-\hat{y}_t$, Jacobian is 1

$$f\left(y_t|\psi_t,\Theta\right)=\frac{1}{\sqrt{2\pi r}}\exp\left\{-\frac{1}{2r}\left(y_t-\psi_t^{'}\theta\right)^2\right\}$$

2.4 Regression model in the state-space form

The state model is

$$x_t=Mx_{t-1}+Nu_t+w_t.$$

We will demonstrate the transformation for the 2nd order model

$$y_t=b_0u_t+a_1y_{t-1}+b_1u_{t-1}+a_2y_{t-2}+b_2u_{t-2}+k+e_t$$

The state model is

$$\begin{bmatrix} y_t \\ u_t \\ y_{t-1} \\ u_{t-1} \\ 1 \end{bmatrix}=\begin{bmatrix} a_1 & b_1 & a_2 & b_2 & k \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_{t-1} \\ u_{t-1} \\ y_{t-2} \\ u_{t-2} \\ 1 \end{bmatrix}+\begin{bmatrix} b_0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}u_t+\begin{bmatrix} e_t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The first row is the regression model, the rest is only one-step time shift.

The advantage of the state-space model lies in recurrent computations. Its memory is only one.

Example: Compute y_3

$$y_1=b_0u_0+a_1y_0+a_2y_{-1}$$
$$y_2=b_0u_2+a_1\left(b_0u_0+a_1y_0+a_2y_{-1}\right)+a_2y_0$$
$$y_3=\cdots$$

$$x_1=Mx_0+Nu_1$$
$$x_2=M\left(Mx_0+Nu_1\right)+Nu_2=M^2x_0+MNu_1+Nu_2$$
$$x_3=M^3x_0+M^2Nu_1+MNu_2+Nu_3$$

In the state form we even can write a general recurrent formula

$$x_k=M^kx_0+\sum_{i=2}^{k}M^{k-i}u_i$$

3 Discrete and logistic models

3.1 Discrete model

All variables are discrete - there is a finite number of configurations of data vector $\Delta_t \equiv \left[y_t', \psi_t'\right]'$. In the model, each data configuration is assigned its own probability

$$f(y_t|\psi_t, \Theta) = \Theta_{y_t|\psi_t}$$

y_t - output, ψ_t - regression vector, Θ parameter.

For two-valued variables and $\psi_t = \left[u_t', y_{t-1}'\right]'$ the parameters are $\Theta_{y_t|u_t,y_{t-1}}$. The model can be given a form of a table

$[u_t, y_{t-1}]$	$y_t = 1$	$y_t = 2$
1, 1	$\Theta_{1\|11}$	$\Theta_{2\|11}$
1, 2	$\Theta_{1\|12}$	$\Theta_{2\|12}$
2, 1	$\Theta_{1\|21}$	$\Theta_{2\|21}$
2, 2	$\Theta_{1\|22}$	$\Theta_{2\|22}$

In the left, there are all configurations of the regression vector. The entries of the table denote all configurations of the data vector, each of them contains its parameter.

It holds:

$$\Theta_{i|jk} \geq 0, \ \sum_i \Theta_{i|jk} = 1, \ \forall jk$$

Remarks

1. The structure of the model is practically general. It is dynamic and possesses control variable.
2. The number of all data configurations is always finite. However, with increasing number of variables and number of values of the variables, its dimension rapidly grows.

Examples:

1. Coin

$y_t = 1$	$y_t = 2$
Θ_1	Θ_2

1. Coin with memory

$$f(y_t|y_{t-1}), \quad y \in \{1, 2\}$$

y_{t-1}	$y_t=1$	$y_t=2$
1	$\Theta_{1\|1}$	$\Theta_{2\|1}$
1	$\Theta_{1\|2}$	$\Theta_{2\|2}$

Uncertainty of the regression model is given by the noise variance. Here, it is given by Θ. If its entries are close to 0 or 1, the model is almost deterministic. If they are near to 0.5, the model is very uncertain. E.g.

$$\begin{bmatrix} 0.1, & 0.9 \\ 0.9, & 0.1 \end{bmatrix} \begin{bmatrix} 0.4, & 0.6 \\ 0.6, & 0.4 \end{bmatrix} \text{ or } \begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix} \begin{bmatrix} 0, & 1 \\ 1, & 0 \end{bmatrix}$$

1. Controlled coin
$$f\left(y_t|u_t\right), \quad y,u \in \{1,2\}$$
2. Controlled coin with memory
$$f\left(y_t|u_t,y_{t-1}\right), \quad y,u \in \{1,2\}$$

$[u_t, y_{t-1}]$	$y_t=1$	$y_t=2$
1, 1	0.8	0.2
1, 2	0.7	0.3
2, 1	0.25	0.75
2, 2	0.1	0.9

where y_t mostly obeys u_t

Other examples

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

First: y_t is the bigger from u_t and y_{t-1}, second: y_t is the opposite to u_t.

3.2 Scilab generations

- generate $y\in\{1,2\}$ so that $P\left(y=1\right)=0.3$

```
y=(rand(1,1,'u')>0.3)+1      (one value);

y=(rand(1,nd,'u')>0.3)+1     (nd values);
```

- generate $y\in\{1,2,\cdots,n\}$ so that $P\left(y=i\right)=p_i$; $p=[p_1\cdots p_n]$

```
pp=cumsum(p);
y=sum(rand(1,1,'u')>pp)+1;
```

- number of row i in the table for u_t, $y_{t-1} \in \{1,2\}$

```
i=2*(u(t)-1)+y(t-1);
```

- generate output y_t from the model $f(y_t|u_t, y_{t-1})$

```
i=2*(u(t)-1)+y(t-1);
pp=cumsum(th(i,:));
y(t)=sum(rand(1,1,'u')>pp)+1;
```

3.3 Logistic model

Output is discrete, regression vector contains at least one continuous variable.

Neither regression nor discrete model can be used!

For $y_t \in \{0,1\}$ the model is

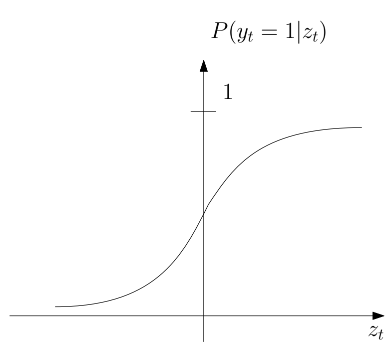
$$\begin{aligned} f(y_t|\psi_t,\Theta) &= \frac{\exp\{y_t z_t\}}{1+\exp\{z_t\}} \\ z_t &= \psi_t'\Theta + e_t \end{aligned}$$

ψ_t regression vector with continuous and possibly discrete variables,

$\psi_t \to z_t$ regression (z_t is continuous),

$P(y_t=1|z_t) = \frac{\exp\{z_t\}}{1+\exp\{z_t\}}$, and its complement $P(y_t=0|z_t) = 1 - \frac{\exp\{z_t\}}{1+\exp\{z_t\}} = \frac{1}{1+\exp\{z_t\}}$ transforms z_t to $(0,1)$.

The function has a form



and it holds

$$\begin{array}{lll} \text{if } z_t > 0 & \text{then } P(y_t = 1|z_t) > 0.5 & \text{estimate: } y_t = 1 \\ \text{if } z_t < 0 & \text{then } P(y_t = 1|z_t) < 0.5 & \text{estimate: } y_t = 0 \end{array}$$

Example:

y_t car accidents: 0 - just damage, 1 - injury or death

ψ_t (i) light: 1 - full, 2 - gloom, 3 - dark; (ii) weather: 1 dry, 2 - slippery; (iii) speed: continuous.

Output: discrete, its values denote system modes.

Regression vector: circumstances under which the output is measured.

Remarks

1. $f(y_t|\psi_t, \Theta) = [P(y_t = 0|z_t), P(y_t = 1|z_t)]$ where $P(y = 0|z_t) = 1/(1 + \exp\{z_t\})$ and $P(y = 1|z_t) = \exp\{z_t\}/(1 + \exp\{z_t\})$, $\quad P_1 + P_2 = 1$
2. Other form of logistic model is
$$logit\,(p_t) = \psi_t^{'}\Theta + e_t$$
where $logit\,(p) = \ln \frac{p}{1-P}$ and $p_t = P(y_t = 1|\psi_t, \Theta)$

   $z = logit\,(p)$ transforms $(0,1) \to R$, $p = logit^{-1}(z)$ transforms $R \to (0,1)$.
3. For $y_t \in \{0, 1, 2, \cdots n\}$ we have
$$\ln \frac{p_1}{p_0} = \psi^{'}\theta_1,\ \ln \frac{p_2}{p_0} = \psi^{'}\theta_2, \cdots,\ \ln \frac{p_n}{p_0} = \psi^{'}\theta_n$$
with the parameter $\Theta = [\theta_1, \theta_2, \cdots, \theta_n]$, where θ_i are columns.

4 Estimation of regression model[2]

Notation: y_t, $d_t = \{y_t, u_t\}$, $d(t) = \{d_0, d_1, d_2, \cdots, d_t\}$; d_0- prior, the rest are measurements.

Bayesian estimation

- classical statistics - parameters are unknown constants
  Bayesian statistics - parameters are random variables (their description is distribution)
- distributions
  $$\text{model description} \quad f(y_t|\psi_t, \Theta)$$
  $$\text{parameter description} \quad f(\Theta|d(t-1)), \; f(\Theta|d(t))$$
- evolution of parameter pdf
  $$f(\Theta|d(0)) \underbrace{\rightarrow}_{d_1=\{u_1,y_1\}} f(\Theta|d(1)) \underbrace{\rightarrow}_{d_2=\{u_2,y_2\}} \cdots \underbrace{\rightarrow}_{d_t=\{u_t,y_t\}} f(\Theta|d(t))$$
- The evolution is governed by the Bayes rule
  $$f(\Theta|d(\tau)) \propto f(y_\tau|\psi_\tau, \Theta) f(\Theta|d(\tau-1))$$
  starting from prior pdf $f(\Theta|d(0))$.

Remarks

1. Derivation of Bayes rule
   $$\begin{aligned} f(A, B|C) &= f(A|B, C) f(B|C) \\ &= f(B|A, C) f(A|C) \end{aligned}$$
   $\rightarrow$ $\qquad f(A|B, C) = \frac{f(B|A,C) f(A|C)}{f(B|C)}$
   where
   $$A \rightarrow \Theta, \;\; B \rightarrow d_t, \;\; C = d(t-1)$$
   and $\{B, C\} = \{d_t, \, d(t-1)\} = d(t)$.
2. Natural conditions of control: The person that estimates also controls. For both actions he uses only information from $d(t-1)$.

   $\rightarrow$

   $f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1))$ and conversely

   $f(u_t|d(t-1), \Theta) = f(u_t|d(t-1))$

[2]two lectures

It applies in estimation with controlled model

$$f(\Theta|d(t)) \propto f(y_t|\psi_t,\Theta)\, f(u_t|d(t-1),\cancel{\Theta})\, f(\Theta|d(t-1))$$

which means that $f(u_t|\cdots)$ goes to constant.

3. Self reproducing form of Bayes rule
B.r. is recursive for functions. To be able to manage functions it is necessary to parametrize the pdfs - e.g. normal distribution is given just by two numbers. Recursiveness requires so that the form of prior pdf (after multiplication by the model) is reproduced in the posterior pdf. E.g. normal pdf $\rightarrow$ normal pdf, with only statistics recomputed.

Example (not recursive)

$$f(y_t|a) = \frac{a}{1+a}\left(\frac{1}{y_t^2} + \exp\{-ay_t\}\right)$$

or

$$f(y_t|a) = \frac{1}{2+\pi a}(\sin(y_t) + a)$$

when computing product of models in measured y_t the number of different terms grows.

Example (recursive)

$$f(y_t|a) = a\exp\{-ay_t\}$$

Posterior

$$f(a|y_1,y_2,y_3) \propto a^3\exp\{-a(y_1+y_2+y_3)\} =$$
$$a^{\kappa_3}\exp\{-aS_3\}$$

where κ and S are statistics, evolving as follows

$$\begin{aligned}\kappa_t &= \kappa_{t-1}+1\\ S_t &= S_{t-1}+y_t\end{aligned}$$

with initial stats κ_0 and S_0 with the meaning:

- κ_0 is a virtual number of data samples, from which the prior statistics is constructed.
- $S_0 = \sum_{i=1}^{\kappa_0} y_i$ from which we have $\bar{y} = \frac{S_0}{\kappa_0}$ i.e. we say that average output is S_0/κ_0.

- Batch estimation

From Bayes rule if follows

$$f(\Theta|d(t)) \propto L_t(\Theta)\, f(\Theta)$$

where $L_t(\Theta) = \prod_{\tau=1}^{t} f(y_\tau|\psi_\tau,\Theta)$ is likelihood and $f(\Theta) \equiv f(\Theta|d(0))$ is the very prior pdf.

- Results of estimation

  (i) posterior pdf $f(\Theta|d(t))$ which brings full information and sometimes can be used as it is - e.g. in prediction

$$f(y_t|d(t-1)) = \int_{\Theta^*} f(y_t, \Theta|d(t-1))\, d\Theta = \int_{\Theta^*} f(y_t|\psi_t, \Theta)\, f(\Theta|d(t-1))\, d\Theta$$

  (ii) point estimates computed using posterior pdf

$$\hat{\Theta}_t = E[\Theta|d(t)] = \int_{\Theta^*} \Theta f(\Theta|d(t))\, d\Theta$$

$$\hat{y}_t = E[y_t|d(t-1)] = \int_{y^*} y_t\, f(y_t|d(t-1))\, dy_t =$$

$$\int_{y^*} y_t \left[\int_{\Theta^*} f(y_t|\psi_t, \Theta)\, f(\Theta|d(t-1))\, d\Theta\right] dy_t$$

- Point estimate with quadratic criterion

  E.g. for Θ and d - data

$$J = E\left[\left(\hat{\Theta} - \Theta\right)^2 |d(t)\right] \to \min$$

  We derive

$$\min_{\hat{\Theta}} E\left[\left(\hat{\Theta} - \Theta\right)^2 |d\right] = \min_{\hat{\Theta}} E\left[\hat{\Theta^2} - 2\hat{\Theta}\Theta + \Theta^2|d\right] =$$

$$= \min_{\hat{\Theta}} \left\{\hat{\Theta^2} - 2\hat{\Theta}E[\Theta|d] + E\left[\Theta^2|d\right]\right\} =$$

$$= \min_{\hat{\Theta}} \left\{\hat{\Theta^2} - 2\hat{\Theta}E[\Theta|d] + E[\Theta|d]^2 \underbrace{-E[\Theta|d]^2 + E\left[\Theta^2|d\right]}_{D[\Theta]}\right\} =$$

$$= \min_{\hat{\Theta}} \left\{\hat{\Theta^2} - 2\hat{\Theta}E[\Theta|d] + E[\Theta|d]^2\right\} + D[\Theta] =$$

$$= \min_{\hat{\Theta}} \left\{\left(\hat{\Theta} - E[\Theta|d]\right)\right\} + D[\Theta]$$

  $\to \hat{\Theta} = E[\Theta|d]$.

5 Estimation of specific models

5.1 Normal regression model

Model

$$f\left(y_t|\psi_t,\Theta\right)=\frac{1}{\sqrt{2\pi}}r^{-0.5}\exp\left\{-\frac{1}{2r}\left(y_t-\psi_t^{'}\theta\right)^2\right\}$$

For 1st order $y_t=bu_t+ay_{t-1}+e_t$ it is $\psi_t=\left[u_t,\,y_{t-1}\right]'$. The square in the exponent can be written

$$\left(y_t-bu_t-ay_{t-1}\right)\left(y_t-bu_t-ay_{t-1}\right)=$$

$$=(-1)\left[-1,\,b,\,a\right]\left[\begin{array}{c}y_t\\u_t\\y_{t-1}\end{array}\right](-1)\left[y_t,\,u_t,\,y_{t-1}\right]\left[\begin{array}{c}-1\\b\\a\end{array}\right]=$$

$$=\left[-1,\,\theta'\right]\underbrace{\left[\begin{array}{c}y_t\\\psi_t\end{array}\right]\left[y_t,\,\psi_t^{'}\right]}_{D_t}\left[\begin{array}{c}-1\\\theta\end{array}\right]$$

where D_t is data matrix.

Model (in modification)

$$f\left(y_t|\psi_t,\Theta\right)\propto r^{-0.5}\exp\left\{\left[-1,\,\theta'\right]D_t\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}$$

Prior pdf

In the same form as model

$$f\left(\Theta|d\left(0\right)\right)\propto r^{-0.5\kappa_0}\exp\left\{\left[-1,\,\theta'\right]V_0\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}$$

Bayes

$$f\left(\Theta|d\left(1\right)\right)\propto r^{-0.5}\exp\left\{\left[-1,\,\theta'\right]D_t\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}r^{-0.5\kappa_0}\exp\left\{\left[-1,\,\theta'\right]V_0\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}=$$

$$=r^{-0.5\kappa_1}\exp\left\{\left[-1,\,\theta'\right]V_1\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}$$

Posterior

$$f\left(\Theta|d\left(t\right)\right)\propto r^{-0.5\kappa_t}\exp\left\{\left[-1,\,\theta'\right]V_t\left[\begin{array}{c}-1\\\theta\end{array}\right]\right\}$$

Recursion

$$\begin{aligned}\kappa_t &= \kappa_{t-1}+1\\ V_t &= V_{t-1}+D_t\end{aligned}$$

with κ_0 and V_0 as prior statistics.

Result

(a) Posterior - GiW with statistics κ_t and V_t.

(b) Point estimates of parameters

$$V_t = \begin{bmatrix} V_y & V_{y\psi} \\ V_{y\psi} & V_\psi \end{bmatrix} \cdots \begin{bmatrix} \bullet & -- \\ | & \square \end{bmatrix}$$

$$\hat{\theta}_t = V_\psi^{-1} V_{y\psi} \quad \text{regression coefficients}$$

$$\hat{r}_t = \frac{V_y - V_{y\psi}' V_\psi^{-1} V_{y\psi}}{\kappa_t} \quad \text{noise variance}$$

Point estimate of output

$$\hat{y}_t = \psi_t \hat{\theta}_{t-1} \qquad (\theta \to \hat{\theta}_{t-1},\ e_t \to 0)$$

Batch estimation

$$y_t = b_0 u_t + \cdots a_n y_{t-n} + b_n u_{t-n} + k + e_t$$

for $t = 1, 2, \cdots, N$

$$\begin{gathered} y_1 = b_0 u_1 + \cdots a_n y_{1-n} + b_n u_{1-n} + k + e_1 \\ y_2 = b_0 u_2 + \cdots a_n y_{2-n} + b_n u_{2-n} + k + e_2 \\ \cdots \\ y_N = b_0 u_N + \cdots a_n y_{N-n} + b_n u_{N-n} + k + e_N \end{gathered}$$

$\rightarrow$ matrix form

$$Y = X\theta + E$$

optimization - least squares

$$J = \sum e_i^2 = E'E = (Y - X\theta)'(Y - X\theta) = Y'Y - 2\theta' X'Y + \theta' X'X\theta$$

$$\frac{\partial}{\partial \theta} J = -2X'Y + 2X'X\theta$$

$$X'X\theta = X'Y \quad \rightarrow \quad \hat{\theta}_t = \left(X'X\right)^{-1} X'Y$$

5.2 Categorical model

Product form of the model

$$f\left(y_t|\psi_t,\Theta\right)=\Theta_{y_t|\psi_t}=\prod_{y|\psi}\Theta_{y|\psi}^{\delta(y|\psi;\,y_t|\psi_t)}$$

i.e. product over all possible configurations of $y|\psi$; but only $y_t|\psi_t$ is chosen.

Posterior pdf

$$f\left(\Theta|d\left(t\right)\right)\propto\prod_{y|\psi}\Theta_{y|\psi}^{\nu_{y|\psi;t}}$$

where $\nu_{y|\psi;t}$ for all configurations of $y|\psi$ is statistics; $\nu_{y|\psi;0}$ is the prior one.

Statistics update

From Bayes rule

$$\nu_{y|\psi;t}=\nu_{y|\psi;t-1}+\delta\left(y|\psi;\,y_t|\psi_t\right)$$

for all configurations of $y|\psi$ (or $\nu_{y_t|\psi_t;t}=\nu_{y_t|\psi_t;t-1}+1$ for actual data)

Point estimate

$$\hat{\theta}_{y|\psi;t}=\frac{\nu_{y|\psi;t}}{\sum_i\nu_{i|\psi;t}}$$

which is normalization of the statistic matrix in rows.

Example (a coin)

Model

$$f\left(y|p\right)=p_y,\;\;y=1,2\,;\,p=\left[p_1,\,p_2\right]'$$

Product form

$$f\left(y|p\right)=p_1^{\delta(y,1)}p_2^{\delta(y,2)}$$

Posterior

$$f\left(p|d\left(t\right)\right)\propto p_1^{\nu_{1;t}}p_2^{\nu_{2;t}}$$

Statistics

$$\nu_t=\left[\nu_{1;t},\,\nu_{2;t}\right]$$

Update

– for $y=1$

$$\nu_{1;t}=\nu_{1;t-1}+1$$

– for $y=2$

$$\nu_{2;t}=\nu_{2;t-1}+1$$

For the data

t	1	2	3
y_t	1	1	2

and zero initial statistics

t	0	1	2	3
ν_1	0	1	2	2
ν_2	0	0	0	1
p_1	x	1	1	$\frac{2}{3}$
p_2	x	0	0	$\frac{1}{3}$

With initial statistics 10

t	0	1	2	3
ν_1	10	11	12	12
ν_2	10	10	10	11
p_1	x	0.524	0.546	0.522
p_2	x	0.476	0.454	0.478

The ratio $\frac{\nu_{1;0}}{\nu_{1;0}+\nu_{2;0}}$ expresses the value of p_1

The magnitude of ν expresses our belief in .our guess.

Output estimate

$$f\left(y_t|d\left(t-1\right)\right)=f\left(y_t|\psi_t,\Theta=\hat{\Theta}_{t-1}\right)$$

y_t	1	2	3	$\cdots$	n
$f\left(y_t\vert\psi_t\hat{\Theta}_{t-1}\right)$	$P\left(y_t=1\right)$	$P\left(y_t=2\right)$	$P\left(y_t=3\right)$		$P\left(y_t=n\right)$

Point estimate: e.g. the value with the biggest probability.

5.3 Model of logistic regression

For estimation, numerical maximization of log-likelihood is used.

For $y_t\in\{0,\,1\}$ the **model** is

$$f\left(y_t|z_t\right)=\frac{\exp\left\{y_tz_t\right\}}{1+\exp\left\{z_t\right\}},\quad z_t=\psi_t^{'}\Theta+e_t$$

Likelihood

$$L_t = \prod_{\tau=1}^{t} f\left(y_\tau | z_\tau\right) = \prod_{\tau=1}^{t} \frac{\exp\left\{y_\tau z_\tau\right\}}{1 + \exp\left\{z_\tau\right\}}$$

$$\ln L_t = \sum_{\tau=1}^{t} \left[y_t z_t - \ln\left(1 + \exp\left\{z_t\right\}\right)\right], \quad z_t = \psi_t' \Theta$$

$$\hat{\Theta}_t = \arg\min_{\Theta} \ln L_t$$

for minimization, Newton method can be used.

Output estimation

Substitute ψ_t into the model with parameter estimates. The value with the biggest probability can be selected.

Classification

The space of all possible ψ is divided into two subsets - one with $\hat{y} = 0$, the other with $\hat{y} = 1$.

6 Prediction

Estimation of the future output.

6.1 Output estimation (zero step prediction)

E.g. for 1st order regression model without control $f\left(y_t|y_{t-1},\Theta\right)$

$$f\left(y_t|y\left(t-1\right)\right)=\int_{\Theta^*}f\left(y_t,\Theta|y\left(t-1\right)\right)d\Theta=$$

$$\begin{aligned}(i)\quad &= \int_{\Theta^*}f\left(y_t|y_{t-1},\Theta\right)f\left(\Theta|y\left(t-1\right)\right)d\Theta \text{ posterior of } \Theta\\ (ii)\quad &\doteq f\left(y_t|y_{t-1},\hat{\Theta}_{t-1}\right) \text{ point estimate of } \Theta\end{aligned}$$

where (ii) is achieved by replacing $f\left(\Theta|y\left(t-1\right)\right)\to\delta\left(\Theta,\hat{\Theta}_{t-1}\right)$ and

$$\int_{\Theta^*}f\left(y_t|y_{t-1},\Theta\right)f\left(\Theta|y\left(t-1\right)\right)d\Theta\doteq$$

$$\doteq\int_{\Theta^*}f\left(y_t|y_{t-1},\Theta\right)\delta\left(\Theta,\hat{\Theta}_{t-1}\right)d\Theta=f\left(y_t|y_{t-1},\hat{\Theta}_{t-1}\right)$$

where $\hat{\Theta}_{t-1}=E\left[\Theta|y\left(t-1\right)\right]=\int_{\Theta^*}\Theta f\left(\Theta|y\left(t-1\right)\right)d\Theta$ is point estimate of Θ based on the data $y\left(t-1\right)$.

Remark: In $f\left(y_t|y\left(t-1\right)\right)$ the parameter Θ is missing. We need to supply it.

6.2 One step prediction

$$f\left(y_{t+1}|y\left(t-1\right)\right)=\int_{\Theta^*}\int_{y_t^*}f\left(y_{t+1},y_t,\Theta|y\left(t-1\right)\right)dy_td\Theta=$$

$$\begin{aligned}(i)\quad &= \int_{\Theta^*}\int_{y_t^*}f\left(y_{t+1}|y\left(t\right),\Theta\right)f\left(y_t|y_{t-1},\Theta\right)f\left(\Theta|y\left(t-1\right)\right)dy_td\Theta\\ (ii)\quad &\doteq f\left(y_{t+1}|\hat{y}_t,\hat{\Theta}_{t-1}\right)\end{aligned}$$

where for (ii) we lay $f\left(\Theta|y\left(t-1\right)\right)\to\delta\left(\Theta,\hat{\Theta}_{t-1}\right)$ and $f\left(y_t|y\left(t-1\right)\right)\to\delta\left(y_t,\hat{y}_t\right)$ with $\hat{\Theta}_{t-1}$ and $\hat{y}_t$ being point estimates.

Remark

- *Here, both Θ and y_t are missing. We must supply both.*
- *Comparing (i) and (ii) we can see the basic principle of Bayesian estimation. Basically, value of the missing unknown variable (Θ and y_t) is substituted (into the pdfs) and it is weighted by its probability (prior pdf + integration). In the second variant (ii) first point estimates are computed and then substituted for the unknown variables.*

6.3 Multi-steps prediction

Regression model with known parameters and point estimation

For a 1st order regression model $y_t = ay_{t-1} + bu_t + e_t$ with known parameters and point prediction we have

$$\begin{aligned} y_t &= ay_{t-1} + bu_t + e_t \\ \hat{y}_t &= ay_{t-1} + bu_t \\ \hat{y}_{t+1} &= a\hat{y}_t + bu_{t+1} = a\left(ay_{t-1} + bu_t\right) + bu_{t+1} = \\ &= a^2 y_{t-1} + abu_t + bu_{t+1} \\ \hat{y}_{t+2} &= a\hat{y}_{t+1} + bu_{t+2} = \\ &= a^3 y_{t-1} + a^2 bu_t + abu_{t+1} + bu_{t+2} \\ & etc. \end{aligned}$$

The point prediction can be achieved by a simple repetitive substitution of the model. For simulation, directly last estimates can be used.

Full prediction with regression model under condition of normality

Prediction with normal model with known parameters preserves normality. If e_t is normal, all predictions are normal, too.

$$\begin{aligned} y_t &= ay_{t-1} + bu_t + e_t \\ y_{t-1} &= ay_t + bu_{t+1} + e_{t+1} = \\ &= a\left(ay_{t-1} + bu_t + e_t\right) + bu_{t+1} + e_{t+1} = \\ &= a^2 y_{t-1} + abu_t + bu_{t+1} + ae_t + e_{t+1} \\ y_{t+2} &= ay_{t+1} + bu_{t+2} + e_{t+2} = \\ &= a^3 y_{t-1} + a^2 bu_t + abu_{t+1} + bu_{t+2} + a^2 e_t + ae_{t+1} + e_{t+2} \end{aligned}$$

$\rightarrow$

$E\left[y_{t+2} | y\left(t-1\right)\right] = a^3 y_{t-1} + a^2 bu_t + abu_{t+1} + bu_{t+2}$

$D\left[y_{t+2}|y\left(t-1\right)\right]=D\left[a^{2}e_{t}+ae_{t+1}+e_{t+2}\right]=\left(a^{4}+a^{2}+1\right)r$

Predictive pdf

$$f\left(y_{t+2}|y\left(t-1\right)\right)=N_{y_{t+2}}\left(E\left[y_{t+2}|y\left(t-1\right)\right],D\left[y_{t+2}|y\left(t-1\right)\right]\right)$$

(normal distribution is determined by its expectation and variance)

6.4 Prediction with discrete model

For model $f\left(y_{t}|y_{t-1},\Theta\right)$ we have

Zero step prediction

$$f\left(y_{t}|y_{t-1},\Theta\right)=\Theta_{y_{t}|y_{t-1}}$$

Multi-steps prediction

$$f\left(y_{t+k}|y\left(t-1\right)\right)=\left(\Theta^{k+1}\right)_{y_{t+k}|y_{t-1}}$$

Example

Two steps prediction

$$f\left(y_{t+2}|y\left(t-1\right)\right)=\sum_{y_{t+1}}\sum_{y_{t}}f\left(y_{t+2}|y_{t-1}\right)f\left(y_{t+1}|y_{t}\right)f\left(y_{t}|y_{t-1}\right)=$$
$$=\sum_{y_{t+1}}\Theta_{y_{t+2}|y_{t+1}}\sum_{y_{t}}\Theta_{y_{t+1}|y_{t}}\Theta_{y_{t}|y_{t-1}}=\left(\Theta^{3}\right)_{y_{t+2}|y_{t-1}}$$

For

$$\Theta=\left[\begin{array}{cc}0.4, & 0.6\\ 0.8, & 0.2\end{array}\right]$$

$$f\left(y_{t+2}|y\left(t-1\right)\right)=\left[\begin{array}{cc}0.4, & 0.6\\ 0.8, & 0.2\end{array}\right]^{3}=\left[\begin{array}{cc}0.544, & 0.456\\ 0.608, & 0.392\end{array}\right]$$

$\rightarrow$

for $y_{t-1}=1$ we have $f\left(y_{t+2}|1\right)=[0.544,\,0.456]$

for $y_{t-1}=2$ we have $f\left(y_{t+2}|2\right)=[0.608,\,0.392]$

Point prediction either MAP, or to generate from the distribution.

7 State-space model, state estimation

7.1 Model

$$\begin{array}{cc} f\left(x_t|x_{t-1},u_{t-1}\right) & \text{model of the state} \\ f\left(y_t|x_t,u_t\right) & \text{model of the output} \end{array}$$

is generated by the equations

$$\begin{aligned} x_t &= Mx_{t-1}+Nu_{t-1}+w_t \\ y_t &= Ax_t+Bu_t+v_t \end{aligned}$$

where M, N, A, B are matrices, w_t and v_t white noises with covariance matrices r_w and r_v.

7.2 Estimation

State description

$$f\left(x_{t-1}|d\left(t-1\right)\right) \underbrace{\rightarrow}_{\text{prediction}} f\left(x_t|d\left(t-1\right)\right) \underbrace{\rightarrow}_{\text{filtration}} f\left(x_t|d\left(t\right)\right)$$

Evolution

$$\begin{aligned} f\left(x_t|d\left(t-1\right)\right) &= \int_{x^*_{t-1}} f\left(x_t|x_{t-1},u_{t-1}\right) f\left(x_{t-1}|d\left(t-1\right)\right) \quad \text{prediction} \\ f\left(\underbrace{x_t}_{\Theta}|d\left(t\right)\right) &\propto \underbrace{f\left(y_t|x_t,u_t\right)}_{\text{model}} f\left(\underbrace{x_t}_{\Theta}|d\left(t-1\right)\right) \quad \text{Bayes} \end{aligned}$$

! In the above derivation Natural Conditions of Control are used !

Kalman filter

For normal model and normal prior state distribution the normality is preserved. Functional recursion becomes algebraic one for expectations and covariance matrices.

Notation

$$\begin{aligned} f\left(x_t|x_{t-1},u_t\right) &= N_{x_t}\left(Mx_{t-1}+Nu_t,\, r_w\right) \\ f\left(y_t|x_t,u_t\right) &= N_{y_t}\left(Ax_t+Bu_t,\, r_v\right) \end{aligned}$$

and

$$
\begin{aligned}
f\left(x_{t-1}|d\left(t-1\right)\right) &= N_{x_{t-1}}\left(x_{t-1|t-1},\, R_{t-1|t-1}\right)\\
f\left(x_{t}|d\left(t-1\right)\right) &= N_{x_{t}}\left(x_{t|t},\, R_{t|t}\right)\\
f\left(x_{t}|d\left(t\right)\right) &= N_{x_{t}}\left(x_{t|t},\, R_{t|t}\right)
\end{aligned}
$$

Substitution into the evolution equations gives Kalman filter (KF)

Kalman filter	
$x_{t\vert t-1}=Mx_{t-1\vert t-1}+Nu_t$	state prediction
$R_{t\vert t-1}=r_x+MR_{t-1\vert t-1}M'$	
$y_p=Ax_{t\vert t-1}+Bu_t$	output prediction
$R_p=r_y+AR_{t\vert t-1}A'$	
$R_{t\vert t}=R_{t\vert t-1}-R_{t\vert t-1}A'R_p^{-1}AR_{t\vert t-1}$	
$K=R_{t\vert t}A'r_y^{-1}$	Kalman gain
$x_{t\vert t}=x_{t\vert t-1}+K\left(y_t-y_p\right)$	state correction

The filter starts with prior $x_{0|0}$ and $R_{0|0}$, uses data y_t, u_t, $t=1,2,\cdots,N$ and currently computes $x_{t|t}$ and $R_{t|t}$. The result is either point state estimate $x_{t|t}$ or the full distribution of the state $f\left(x_t|u_t,d\left(t\right)\right)=N_{x_t}\left(x_{t|t},\,R_t\right)$.

Program with the task is in LecKalman.sce and Kalman.sci.

8 Nonlinear state estimation

8.1 Nonlinear model

$$x_t = g(x_{t-1}, u_t) + w_t$$
$$y_t = h(x_t, u_t) + v_t$$

EXAMPLE

For

$$x_t = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_t, \quad u_t, \quad y_t$$

the model is

$$\begin{aligned} x_{1;t} &= \exp\{-x_{1;t-1} - x_{2;t-1}\} + u_t + w_t \\ x_{2;t} &= x_{1;t-1} - 0.3u_t + w_{2;t} \\ y_t &= x_{2;t} + v_t \end{aligned}$$

Linearization

Is done using first two terms of Taylor expansion of nonlinear functions at the point of last point estimate. For the state equation it is $\hat{x}_{t-1}$ and for the output equation it is $\hat{x}_t$.

Generally, i.e. for a general value x the expansion reads

$$g(x, u_t) \doteq g(\hat{x}_{t-1}, u_t) + g'(\hat{x}_{t-1}, u_t)(x - \hat{x}_{t-1})$$

$$h(x, u_t) \doteq h(\hat{x}_t, u_t) + h'(\hat{x}_t, u_t)(x - \hat{x}_t)$$

REMARKS

1. x_t and x_{t-1} are random variables. x is their general value, $\hat{x}_t$ and $\hat{x}_{t-1}$ are special values: $\hat{x}_t$ is the point estimate of x_t and $\hat{x}_{t-1}$ is point estimate of x_{t-1}.
2. *Linearization can be applied only to nonlinear parts of the model. The linear parts can stay as they are.*

The derivatives g' and h' are

$$g'(\hat{x}_{t-1}, u_t) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ & & \cdots & \\ \frac{\partial g_n}{\partial x_1} & & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}_{|x=\hat{x}_{t-1}}, \quad h'(\hat{x}_t, u_t) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ & & \cdots & \\ \frac{\partial h_m}{\partial x_1} & & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}_{|x=\hat{x}_t}$$

After substitution the linearization into the model, we have

and for $x = \hat{x}_{t-1}$ in the case of the state equation and $x = \hat{x}_t$ for output equation we obtain the linearized model

$$\begin{aligned} x_t &= \bar{M} x_{t-1} + F + w_t \\ y_t &= \bar{A} x_t + G + v_t \end{aligned}$$

where

$$\bar{M} = g'\left(\hat{x}_{t-1}, u_t\right), \qquad F = g\left(\hat{x}_{t-1}, u_t\right) - g'\left(\hat{x}_{t-1}, u_t\right)\hat{x}_{t-1},$$

$$\bar{A} = h'\left(\hat{x}_t, u_t\right), \qquad G = h\left(\hat{x}_t, u_t\right) - h'\left(\hat{x}_t, u_t\right)\hat{x}_t.$$

Example (continuation) - $\cdots$ only first equation is nonlinear

$$g_1\left(x, u_t\right) = \exp\left\{-x_1 - x_2\right\} + u_t$$

$$g_1^{'}\left(x, u_t\right) = \left[\frac{\partial g_1}{\partial x_1}, \frac{\partial g_1}{\partial x_2}\right] = \left[-\exp\left\{-x_1 - x_2\right\}, -\exp\left\{-x_1 - x_2\right\}\right]$$

Fully linearized model is

$$\begin{aligned} x_{1;t} &= g_1^{'}\left(\hat{x}_{t-1}, u_t\right) x_{t-1} + g_1\left(\hat{x}_{t-1}, u_t\right) - g_1^{'}\left(\hat{x}_{t-1}, u_t\right)\hat{x}_{t-1} + w_t \\ x_{2;t} &= [1,\, 0]\, x_{t-1} - 0.3 u_t + w_{2;t} \\ y_t &= [0,\, 1]\, x_t + v_t \end{aligned}$$

where

$$\bar{M} = \begin{bmatrix} g_1^{'}\left(\hat{x}_{t-1}, u_t\right) \\ [1, 0] \end{bmatrix}, \quad F = \begin{bmatrix} g_1\left(\hat{x}_{t-1}, u_t\right) - g_1^{'}\left(\hat{x}_{t-1}, u_t\right)\hat{x}_{t-1} \\ -0.3 u_t \end{bmatrix},$$

$$N = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{A} = [0,\, 1], \quad G = 0, \quad B = 0.$$

With this, we can use subroutine Kalman

[xt,Rx,yp]=Kalman(xt,yt,ut,$\bar{M}$,N,F,$\bar{A}$,B,G,Rw,Rv,Rx)

8.2 Model with unknown parameters

The unknown parameters of the model are added to the state a and estimated. However, the model becomes nonlinear - model matrices contain state entries and they are multiplied by state. So, the technique of linearization must be used, again.

EXAMPLE

Model

$$\begin{aligned} x_t &= \exp\{-a x_{t-1}\} + b u_t + w_t \\ y_t &= x_t + v_t, \end{aligned}$$

where a and b are unknown.

We define new state

$$X_t = \left[x_t', \, a, \, b\right]', \;\; X_{t-1} = \left[x_{t-1}', \, a, \, b\right]'$$

and obtain new model

$$\begin{aligned} X_t &= \begin{bmatrix} \exp\{-X_{2;t-1}X_{1;t-1}\} + X_{3;t-1}u_t \\ X_{2;t-1} \\ X_{3;t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} w_t \\ \epsilon_{2;t} \\ \epsilon_{3;t} \end{bmatrix}}_{W_t} \\ y_t &= [1, \, 0, \, 0] \, X_t + v_t \end{aligned}$$

Only the first equation is nonlinear, however, we will treat the whole model as nonlinear (it is well possible)

$$g = \begin{bmatrix} \exp\{-X_{2;t-1}X_{1;t-1}\} + X_{3;t-1}u_t \\ X_{2;t-1} \\ X_{3;t-1} \end{bmatrix}_{X_{t-1}=\hat{X}_{t-1}}$$

$$g' = \begin{bmatrix} -X_{2;t-1}\exp\{-X_{2;t-1}X_{1;t-1}\}, & -X_{1;t-1}\exp\{-X_{2;t-1}X_{1;t-1}\}, & u_t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{X_{t-1}=\hat{X}_{t-1}}$$

model

$$\begin{aligned} X_t &= \underbrace{g'}_{\bar{M}} X_{t-1} + \underbrace{g - g'\hat{X}_{t-1}}_{F} + W_t \\ y_t &= \underbrace{[1, \, 0, \, 0]}_{\bar{A}} X_t + v_t \end{aligned}$$

and $N = [0, \, 0, \, 0]'$, $B = 0$, $G = 0$.

[x, Rx, yp] = Kalman(x, y, u, $\bar{M}$, N, F, $\bar{A}$, B, G, Rw, Rv, Rx)

9 Estimation of mixtures

9.1 Mixture model

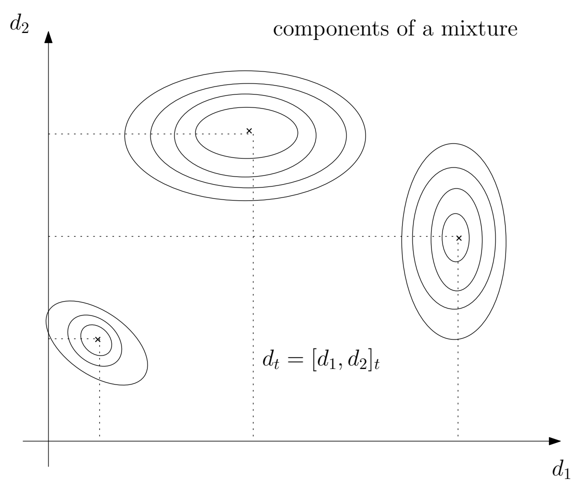
$$f\left(d_t|\Theta\right)=\sum_{k=1}^{n_c}\alpha_k f_k\left(d_t|\Theta_k\right)$$

where f_k denotes component; k is component index; Θ_k are parameters of k-th component; α_k stationary weights of components.

Example (for $n_c = 3$)

$f_1 :\quad d_t = m^{(1)} + e_{1;t} \quad e_{1;t} \sim N_{d_t}\left(0, r^{(1)}\right)$
$f_2 :\quad d_t = m^{(2)} + e_{2;t} \quad e_{2;t} \sim N_{d_t}\left(0, r^{(2)}\right)$
$f_3 :\quad d_t = m^{(3)} + e_{3;t} \quad e_{3;t} \sim N_{d_t}\left(0, r^{(3)}\right)$

$$d_t=\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}_t=\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}_t^{(k)}+\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}_t^{(k)},\ k=1,2,3$$



We cannot use d_t for updating all components - it would not respect multi-modality of the system. It is necessary to determine probabilities that the measured data item belongs to individual components (so called actual weights of components). This starts with assumption that the data belongs to a single component - called active component.

9.2 Estimation

Known active components

We introduce pointer $c_t \in \{1, 2, \cdots, n_c\}$ as a discrete process whose realizations at each time t point at the active component.

If the activities of the components are known, it holds

$$f\left(c_t\right)=\delta\left(c_t,\hat{c}_t\right)$$

where δ is Kronecker function, $\hat{c}_t$ is known number of active component at t.

In this case, the posterior is

$$f\left(\Theta_k|d\left(t\right)\right)\propto f_k\left(d_t|\Theta_k\right)f\left(\Theta_k|d\left(t-1\right)\right),\ \text{ for } k=\hat{c}_t,\ \forall t$$

$\cdots$ each component has its own prior and each time t only the prior of the active component is updated.

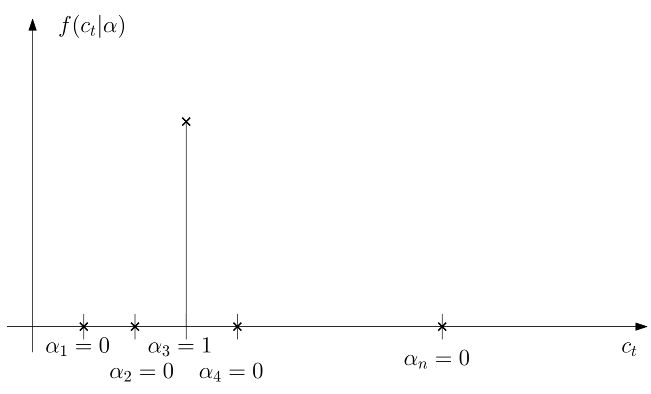
Unknown active components

If the active components are not known, we have to estimate them. We have two models - **active component model**

$$\begin{aligned}
f_{c_t}\left(d_t|\Theta_{c_t}\right) &= \prod_{k=1}^{n_c} f_{c_t}\left(d_t|\Theta_{c_t}\right)^{\delta(k;\,c_t)} \quad \text{(product form)}\\
&= \prod_{k=1}^{n_c} r^{0.5}\exp\left\{-\frac{1}{2r}\left(\left[\begin{array}{c} y_t\\ \psi_t\end{array}\right]' D_t\left[\begin{array}{c} y_t\\ \psi_t\end{array}\right]\right)\right\}^{\delta(k;\,c_t)} \quad \text{(normal model)}
\end{aligned}$$

and **pointer model** - with categorical description

$$f\left(c_t|\alpha\right)=\alpha_{c_t}=\prod_{k=1}^{n_c}\alpha_k^{\delta(k;\,c_t)} \quad \text{(product form)}$$



Bayes

$$f\left(c_t,\Theta,\alpha|d\left(t\right)\right)\underbrace{\propto}_{\text{Bayes}}f\left(d_t,c_t,\Theta,\alpha|d\left(t-1\right)\right)=$$

$$
\begin{aligned}
&= \underbrace{f\left(d_t|c_t,\Theta\right)}_{\text{component}}\ \underbrace{f\left(c_t|\alpha\right)}_{\text{pt. model}}\ \underbrace{f\left(\Theta,\alpha|d\left(t-1\right)\right)}_{\text{priors (independent)}} = \\
&= f_{c_t}\left(d_t|\Theta_{c_t}\right)\alpha_{c_t} f\left(\Theta|d\left(t-1\right)\right) f\left(\alpha|d\left(t-1\right)\right) = \\
&= \underbrace{f_{c_t}\left(d_t|\Theta_{c_t}\right) f\left(\Theta|d\left(t-1\right)\right)}_{\text{component part}} \times \underbrace{\alpha_{c_t} f\left(\alpha|d\left(t-1\right)\right)}_{\text{pointer part}} \propto (*)
\end{aligned}
$$

In product form with conjugated priors

$$
f\left(\Theta|d\left(t-1\right)\right) = \prod_{k=1}^{n_c} GiW_{\Theta_k}\left(V_k,\kappa_k\right) \propto \prod_{k=1}^{n_c} r^{0.5\kappa_{k;t-1}} \exp\left\{-\frac{1}{2r}\left(\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]' V_{k;t-1}\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]\right)\right\}
$$

$$
f\left(\alpha|d\left(t-1\right)\right) = \prod_{k=1}^{n_c} Di_\alpha\left(\nu_k\right) \propto \prod_{k=1}^{n_c} \alpha_k^{\nu_{k;t-1}}
$$

where GiW is Gauss-inverse-Wishart and Di is Dirichlet distributions, we have

$$
(*) \propto \prod_{k=1}^{n_c} r^{0.5} \exp\left\{-\frac{1}{2r}\left(\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]' D_t\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]\right)\right\}^{\delta(k;\,c_t)} \prod_{k=1}^{n_c} r^{0.5\kappa_{k;t-1}} \exp\left\{-\frac{1}{2r}\left(\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]' V_{k;t-1}\left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]\right)\right\} \times
$$

$$
\times \prod_{k=1}^{n_c}\prod_{k=1}^{n_c} \alpha_k^{\delta(k;\,c_t)} \prod_{k=1}^{n_c} \alpha_k^{\nu_{k;t-1}}
$$

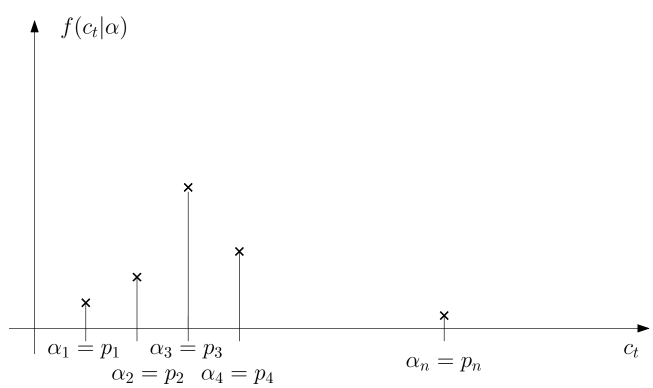
$\rightarrow$ statistics update

$$
\begin{aligned}
V_{k;t} &= V_{k;t-1} + \delta\left(k;\, c_t\right) D_t \\
\kappa_{k;t} &= \kappa_{k;t-1} + \delta\left(k;\, c_t\right) \\
\nu_{k;t} &= \nu_{k;\,t-1} + \delta\left(k;\, c_t\right)
\end{aligned}
$$

for all $k = 1, 2, \cdots, n_c$ (in practice, only statistics of the currently active component are updated).

Problem is that we do not know c_t. We will approximate $\delta\left(k;\, c_t\right)$ by its expectation

$$
E\left[\delta\left(k;\, c_t\right)|d\left(t\right)\right] = \sum_{k=1}^{n_c} \delta\left(k;\, c_t\right) f\left(c_t|d\left(t\right)\right) = P\left(c_t = k|d\left(t\right)\right) = w_{k;t}
$$



→ approximated statistics update

$$\begin{aligned} V_{k;t} &= V_{k;t-1} + w_{k;t} D_t \\ \kappa_{k;t} &= \kappa_{k;t-1} + w_{k;t} \\ \nu_{k;t} &= \nu_{k;\,t-1} + w_{k;t} \end{aligned}$$

i.e. only part of data is used for each component.

Computation of actual weights

Using posterior pdf

$$w_{k;t} = \int \int f\left(c_t = k, \Theta, \alpha | d\left(t\right)\right) d\alpha d\Theta =$$

$$= \underbrace{\int_{\Theta^*} f_{c_t}\left(d_t | \Theta_{c_t}\right) f\left(\Theta | d\left(t-1\right)\right) d\Theta}_{\text{data prediction } M_k} \times \underbrace{\int_{\alpha^*} \alpha_{c_t} f\left(\alpha | d\left(t-1\right)\right) d\alpha}_{\text{component prediction } \hat{\alpha}_{k;t-+}}$$

where M_k is a "distance" of d_t from k-th component - $f_k\left(d_t|\hat{\Theta}_{t-1}\right)$; $\hat{\alpha}_{k;t-1}$ is a probability of k-th component "historical" occurrence.

Algorithm

for $t = 1, 2, \cdots, n_c$

1. compute $\hat{\alpha}_k = \frac{\nu_{k;t-1}}{\sum_i \nu_{i;t-1}}$ and $\hat{\Theta}_{k;t-1}$ (LS)
2. measure new data d_t
3. compute $M_k = f_k\left(d_t|\hat{\Theta}_{k;t-1}\right)$
4. compute weights $w_k = M_k \hat{\alpha}_{k;t-1}$ and normalize $w = \frac{w}{\sum w}$
5. update component and pointer statistics

$$\begin{aligned} V_k &= V_k + w_k D_t \\ \kappa_k &= \kappa_k + w_k \\ \nu_k &= \nu_k + w_k \end{aligned}$$

6. compute point estimate of active component

$$\hat{c}_t = \arg\max(w_t)$$

10 Control with regression model

10.1 Derivation in pdf

Criterion

Optimal control needs criterion. We will use summation one

$$J = \sum_{t=1}^{N} J_t$$

where J_t is a penalization for time t. Mostly it is $J_t = y_t^2 + \omega u_t^2$.

We want to set $u_t,\ t = 1, 2, \cdots, N$ that minimizes J. But, J is a random variable, due to the output y_t. As random variable can take many different values it is not possible to speak about its minimization. So, we must minimize its estimate (which is expectation). So the minimized criterion is

$$E\left[J|d(0)\right] = E\left[\sum_{t=1}^{N} J_t|d(0)\right]$$

where in condition of the expectation is our preliminary knowledge - prior data.

Remark

For $N = 1$ we obtain one-step control. Here, we optimize control only for the next output. This control is dangerous, because the controller does not take into account future evolution of the system and to act best in one step it can generate too beg output. This can excite the system so much that it is not possible even to stabilize it in the future and the control fails.

Minimization

$$\min_{u_{1:N}} E\left[\varphi^*_{N+1} + \sum_{t=1}^{N} J_t|d(0)\right] =$$

$$= \min_{u_{1:(N-1)}} E\left[\min_{u_N} \underbrace{E\left[\varphi^*_{N+1} + J_N|u_N, d(N-1)\right]}_{\varphi^*_N} + \sum_{t=1}^{N-1} J_t \middle| d(0)\right] =$$

$$= \min_{u_{1:(N-1)}} E\left[\min_{u_N} \varphi_N + \sum_{t=1}^{N-1} J_t|d(0)\right] = \min_{u_{1:N}} E\left[\varphi^*_N + \sum_{t=1}^{N-1} J_t|d(0)\right]$$

which reproduces the initial form, only with $N \rightarrow N-1$ and where (due to the reproduction in general form)

Bellman equations

$$\varphi_t = E\left[\varphi^*_{t+1} + J_t | u_t, d(t-1)\right] \quad \text{expectation}$$
$$\varphi^*_t = \min_{u_t} \varphi_t \quad \text{minimization}$$

for $t = N$, $N-1$, $N-2$, $\cdots$, 1. Each minimization gives the formula for optimal control - it is $u_t = \arg\min \varphi_t(d(t-1))$. However, ti cannot be used immediately, because the data $d(t-1)$ is not known, yet. Only at time $t=1$ we need data $d(0)$ and the control can start to be generated.

Comments

1. The operator form of expectation is brief but not explicit. We will show its integral form:

$$\min_{u_{1:N}} E\left[\varphi^*_{N+1} + \sum_{t=1}^{N} J_t | d(0)\right] =$$

$$= \min_{u_{1:N}} \int \cdots \int \left(\varphi^*_{N+1} + \sum_{t=1}^{N} J_t\right) f(y(N), u(N) | d(0))\, dy(N)\, du(N) =$$

$$= \min_{u_{1:N}} \int \cdots \int \int \int \left(\left[\varphi^*_{N+1} + J_N\right] + \sum_{t=1}^{N-1} J_t\right) f(y_N | u_N, d(N-1))\, f(u_N | d(N-1)) \times$$
$$\times f(y(N-1), u(N-1) | d(0))\, dy(N)\, du(N) =$$

$$= \min_{u_{1:(N-1)}} \left\{ \int \cdots \int \min_{u_N} \int \underbrace{\int \left(\varphi^*_{N+1} + J_t\right) f(y_N | u_N, d(N-1))\, dy_N}_{\varphi_N(u_N, d(N-1))} f(u_N | d(N-1))\, du_N + \right.$$

$$\left. \sum_{t=1}^{N-1} J_t f(y(N-1), u(N-1) | d(0))\, dy(N-1)\, du(N-1) \right\}$$

Minimum over u_N

$$\min_{u_N} \int \underbrace{\int \left(\varphi^*_{N+1} + J_t\right) f(y_N | u_N, d(N-1))\, dy_N}_{\varphi_N(u_N, d(N-1))} f(u_N | d(N-1))\, du_N =$$

$$= \min_{u_N} \int \varphi_N(u_N, d(N-1))\, f(u_N | d(N-1))\, du_N$$

$\rightarrow u^*_N = \arg\min_{u_N} \varphi_N$ and $\quad f(u_N | d(N-1)) = \delta(u_N, u^*_N)$ - all u_t is concentrated into one point u^*_N.

10.2 Derivation for regression model

Regression model can be converted to state-space form (see lecture 2 - Regression model in state-space form).

$$x_t = Mx_{t-1} + Nu_t + w_t$$

where $x_t = [y_t, u_t, y_{t-1}, u_{t-1}, \cdots y_{t-n+1}, u_{t-n+1}]'$.

The penalty can be written as

$$y_t^2 + \omega u_t^2 = x_t' \Omega x_t \tag{10.1}$$

where Ω is a diagonal matrix

$$\Omega = \begin{bmatrix} 1 & & & & \\ & \omega & & & \\ & & 0 & & \\ & & & \cdots & \\ & & & & 0 \end{bmatrix}$$

Now the model and criterion is used in general Bellman equations, where we guess the form of $\varphi_{t+1}^* = x_t^{'} R_{t+1} x_t$

$$E\left[x_t^{'} R_{t+1} x_t + x_t^{'} \Omega x_t | u_t, d\,(t-1)\right] = E\left[x_t^{'} U x_t\right] =$$

$$= (Mx_{t-1} + Nu_t)'\, U\, (Mx_{t-1} + Nu_t) + \rho =$$

$$= x_{t-1}^{'} \underbrace{M'UM}_{C} x_{t-1} + 2u_t^{'} \underbrace{N'UM}_{B} x_{t-1} + u_t^{'} \underbrace{N'UN}_{A} u_t + \rho =$$

$$= u_t^{'} A u_t + 2u_t^{'} A \underbrace{A^{-1} B}_{S_t} x_{t-1} + x_{t-1}^{'} S_t^{'} A S_t x_{t-1} +$$

$$+ \underbrace{x_{t-1}^{'} C x_{t-1} - x_{t-1}^{'} S_t^{'} A S_t x_{t-1}}_{x_{t-1} R_t x_{t-1}} + \rho =$$

$$= (u_t + S_t x_{t-1})'\, A\, (u_t + S_t x_{t-1}) + x_{t-1}^{'} R_t x_{t-1} + \rho$$

Optimal $u_t = S_t x_{t-1}$.

Recursion

$R_{N+1} = 0$

for $t = N,\ N-1,\ \cdots,\ 1$

$U = R_{t+1} + \Omega$
$A = N'UN$
$B = N'UM$
$C = M'UM$
$S_t = A^{-1}B$
$R_t = C - S_t^{'}AS_t$
$u_t = S_t x_{t-1}.$

end

Remark

The penalty function (10.1) can be very easily extended to the following form

$$(y_t - s_t)^2 + \omega u_t^2 + \lambda (u_t - u_{t-1})^2$$

where the first term leads to the following the output y_t the prescribed set-point s_t and the last term introduces penalization of increments of the control variable. Penalizing the control increments calms control behavior and at the same time it does not result to steady-state deviation of the output and the set-point as it is when penalizing the whole control variable.

The solution how to introduce the above requirements for the control lies in construction of the penalization matrix as follows

$$\Omega = \begin{bmatrix} 1 & & & & & & -1 \\ & \omega+\lambda & & -\lambda & & & \\ & & 0 & & & & \\ & -\lambda & & \lambda & & & \\ & & & & \cdots & & \\ & & & & & 0 & \\ -1 & & & & & & 1 \end{bmatrix}$$

which is evident if we take into account that the criterion is

$$x_t^{'} \Omega x_t$$

and $x_t = [y_t,\, u_t,\, y_{t-1},\, u_{t-1},\, \cdots, 1]$.

11 Control with categorical model

Here, we are going to demonstrate synthesis for the controlled coin with memory. The model is introduced by the table below. The penalization matrix is of the same form. Each of its entries individually penalizes the corresponding configuration of values of actual data $[y_t,\, u_t,\, y_{t-1}]$.

$$\text{model} \quad f\left(y_t|u_t, y_{t-1}\right)$$
$$\text{penalty} \quad J_{y_t|u_t,y_{t-1}}$$

for three steps control, i.e. for $t = 1, 2, 3$ and the following model and penalization

model (Θ)

$u_3,\, y_2$	$y_3 = 1$	$y_3 = 2$
1, 1	0.7	0.3
1, 2	0.2	0.8
2, 1	0.9	0.1
2, 2	0.4	0.6

penalty (J)

$u_3,\, y_2$	$y_3 = 1$	$y_3 = 2$
1, 1	0	1
1, 2	1	0
2, 1	1	2
2, 2	2	1

11.1 Optimization

Optimization is a bit unusual, however, it again follows the Bellman equations (**??**) and (**??**). We will show the synthesis on the interval of the length 3. We begin at the end of it.

Step for $t = 3$: $\quad \varphi_4^* = 0$

Expectation

$$\varphi_3 = E\left[J|u_3, d\left(2\right)\right] = \sum_{y_3=1}^{2} J_{y_3|u_3,y_2}\Theta_{y_3|u_2,y_2} =$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix} .* \begin{bmatrix} 0.7 \\ 0.2 \\ 0.9 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} .* \begin{bmatrix} 0.3 \\ 0.8 \\ 0.1 \\ 0.6 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.2 \\ 1.1 \\ 1.4 \end{bmatrix} \begin{array}{l} \cdots \quad u_3 = 1,\, y_2 = 1 \\ \cdots \quad u_3 = 1,\, y_2 = 2 \\ \cdots \quad u_3 = 2,\, y_2 = 1 \\ \cdots \quad u_3 = 2,\, y_2 = 2 \end{array}$$

Minimization

$$\text{for}: \quad y_2 = 1 \rightarrow \min\left\{0.3,\, 1.1\right\} = 0.3 \ \text{ for } u_3 = 1$$
$$\text{for}: \quad y_2 = 2 \rightarrow \min\left\{0.2,\, 1.4\right\} = 0.2 \ \text{ for } u_3 = 1$$

$\rightarrow$

$$u_3 = \begin{cases} 1 & \text{for } y_2 = 1 \\ 1 & \text{for } y_2 = 2 \end{cases}$$

and reminder after minimization

$$\begin{array}{cc} y_2=1 & y_2=2 \\ \hline 0.3 & 0.2 \end{array} \forall u_2, y_1 \quad \rightarrow \quad \begin{array}{c|cc} u_2,\, y_1 & y_2=1 & y_2=2 \\ \hline 1,\, 1 & 0.3 & 0.2 \\ 1,\, 2 & 0.3 & 0.2 \\ 2,\, 1 & 0.3 & 0.2 \\ 2,\, 2 & 0.3 & 0.2 \end{array} = \varphi_3^*$$

Step for $t = 2$:

Expectation

$$\varphi_2 = E\left[J + \varphi_3^* | u_2, d\left(1\right)\right] = \sum_{y_2=1}^{2} \left(J_{y_2|u_2,y_1} + \varphi^*_{3;y_2|u_2,y_1}\right)\Theta_{y_2|u_2,y_1} =$$

$$= \left(\begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.3 \\ 0.3 \\ 0.3 \end{bmatrix}\right) .\varphi_2^* = 0 * \begin{bmatrix} 0.7 \\ 0.2 \\ 0.9 \\ 0.4 \end{bmatrix} + \left(\begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}\right) .* \begin{bmatrix} 0.3 \\ 0.8 \\ 0.1 \\ 0.6 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.8 \\ 0.7 \\ 1.6 \\ 1.9 \end{bmatrix} \begin{array}{l} \cdots \quad u_2 = 1,\, y_1 = 1 \\ \cdots \quad u_2 = 1,\, y_1 = 2 \\ \cdots \quad u_2 = 2,\, y_1 = 1 \\ \cdots \quad u_2 = 2,\, y_1 = 2 \end{array}$$

Minimization

$$\text{for}: \quad y_1 = 1 \rightarrow \min\{0.8,\, 1.6\} = 0.8 \;\; \text{for}\, u_2 = 1$$
$$\text{for}: \quad y_1 = 2 \rightarrow \min\{0.7,\, 1.9\} = 0.7 \;\; \text{for}\, u_2 = 1$$

$\rightarrow$

$$u_2 = \begin{cases} 1 & \text{for } y_1 = 1 \\ 1 & \text{for } y_1 = 2 \end{cases}$$

and reminder after minimization

$$\begin{array}{cc} y_1=1 & y_1=2 \\ \hline 0.8 & 0.7 \end{array} \forall u_1, y_0 \quad \rightarrow \quad \begin{array}{c|cc} u_1,\, y_0 & y_1=1 & y_1=2 \\ \hline 1,\, 1 & 0.8 & 0.7 \\ 1,\, 2 & 0.8 & 0.7 \\ 2,\, 1 & 0.8 & 0.7 \\ 2,\, 2 & 0.8 & 0.7 \end{array} = \varphi_2^*$$

Step for $t = 1$:

Expectation

$$\varphi_1 = E\left[J + \varphi_2^* | u_1, d\left(0\right)\right] = \sum_{y_1=1}^{2} \left(J_{y_1|u_1,y_0} + \varphi^*_{2;y_1|u_1,y_0}\right)\Theta_{y_1|u_1,y_0} =$$

$$= \left(\begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.8 \\ 0.8 \\ 0.8 \\ 0.8 \end{bmatrix} \right) .* \begin{bmatrix} 0.7 \\ 0.2 \\ 0.9 \\ 0.4 \end{bmatrix} + \left(\begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.7 \\ 0.7 \\ 0.7 \\ 0.7 \end{bmatrix} \right) .* \begin{bmatrix} 0.3 \\ 0.8 \\ 0.1 \\ 0.6 \end{bmatrix} =$$

$$= \begin{bmatrix} 1.8 \\ 1.7 \\ 2.6 \\ 2.9 \end{bmatrix} \begin{matrix} \cdots & u_1 = 1,\, y_0 = 1 \\ \cdots & u_1 = 1,\, y_0 = 2 \\ \cdots & u_1 = 2,\, y_0 = 1 \\ \cdots & u_1 = 2,\, y_0 = 2 \end{matrix}$$

Minimization

$$\text{for}: \quad y_0 = 1 \to \min\{1.8,\, 2.6\} = 1.8 \ \text{ for } u_1 = 1$$

$$\text{for}: \quad y_1 = 2 \to \min\{1.7,\, 2.9\} = 1.7 \ \text{ for } u_1 = 1$$

$\rightarrow$

$$u_1 = \begin{cases} 1 & \text{for } y_0 = 1 \\ 1 & \text{for } y_0 = 2 \end{cases}$$

and reminder after minimization

$y_0 = 1$	$y_0 = 2$
1.8	1.7

11.2 Application

For $t = 0$ let us have $y_0 = 2$.

For $y_0 = 2$ we have $u_1 = 1$; $\rightarrow$ [1, 2] $\Theta_{1,2}$ =[0.2, 0.8] let us obtain $y_1 = 2$

For $y_1 = 2$ we have $u_2 = 1$; $\rightarrow$ [1, 2] $\Theta_{1,2}$ =[0.2, 0.8] let us obtain $y_2 = 1$

For $y_2 = 1$ we have $u_3 = 1$; $\rightarrow$ [1, 1] $\Theta_{1,1}$ =[0.7, 0.3] let us obtain $y_3 = 2$

The final value of criterion is $J_{2|12} + J_{1|12} + J_{2|11} = 0 + 1 + 1 = 2$.

12 Appendix

12.1 Elementary differential equations

First order equations

The first order homogeneous differential equation with constant coefficients has the form

$$y' + a_0 y = 0, \qquad y(0) = \tilde{y}_0 \tag{12.1}$$

Characteristic equation is linear with unique solution

$$\lambda + a_0 = 0 \quad \rightarrow \quad \lambda = -a_0 \tag{12.2}$$

The solution to the differential equation (12.1) is

$$y(t) = \tilde{y}_0 \mathrm{e}^{\lambda t} \tag{12.3}$$

Second order equations

The second order homogeneous differential equation with constant coefficients has the form

$$y'' + a_1 y' + a_0 y = 0, \qquad y(0) = \tilde{y}_0,\ y'(0) = \tilde{y}_1 \tag{12.4}$$

Characteristic equation is quadratic

$$\lambda^2 + a_1 \lambda + a_0 = 0 \tag{12.5}$$

with the following types of solution

1. **Two different real roots** λ_1 and λ_2

   The equation (12.4) is

   $$y'' - (\lambda_1 + \lambda_2) y' + \lambda_1 \lambda_2 y = 0$$

   The solution is

   $$y(t) = c_1 \mathrm{e}^{\lambda_1 t} + c_2 \mathrm{e}^{\lambda_2 t}, \tag{12.6}$$

   where the coefficients c can be obtained as a solution of the set of linear equations

   $$\begin{aligned} c_1 + c_2 &= \tilde{y}_0 \\ \lambda_1 c_1 + \lambda_2 c_2 &= \tilde{y}_1 \end{aligned}$$

   which gives the solution

   $$c_1 = (\lambda_2 y_0 - y_0')/(\lambda_2 - \lambda_1) \text{ and } c_2 = (\lambda_1 y_0 - y_0')/(\lambda_1 - \lambda_2)$$

2. **One real double root λ**

   The equation (12.4) is
   $$y'' - 2\lambda y' + \lambda^2 y = 0$$
   The solution is
   $$y(t) = c_1 \mathrm{e}^{\lambda t} + c_2 t \mathrm{e}^{\lambda t}, \tag{12.7}$$
   where the coefficients c can be obtained as a solution of the set of linear equations
   $$\begin{aligned} c_1 &= \tilde{y}_0 \\ \lambda c_1 + c_2 &= \tilde{y}_1 \end{aligned}$$
   which gives the solution
   $$c_1 = \tilde{y}_0 \quad \text{and} \quad c_2 = \tilde{y}_1 - \lambda \tilde{y}_0$$

3. **Two complex roots** $\lambda_1 = \rho + \omega\mathrm{i}$ and $\lambda_2 = \rho - \omega\mathrm{i}$

   The equation (12.4) is
   $$y'' - 2\rho y' + \rho^2 + \omega^2 = 0$$
   The solution is
   $$y(t) = c_1 \mathrm{e}^{\rho t} \cos(\omega t) + c_2 \mathrm{e}^{\rho t} \sin(\omega t), \tag{12.8}$$
   where the coefficients c can be obtained as a solution of the set of linear equations
   $$\begin{aligned} c_1 &= \tilde{y}_0 \\ \rho c_1 + \omega c_2 &= \tilde{y}_1 \end{aligned}$$
   which gives the solution
   $$c_1 = \tilde{y}_0 \quad \text{and} \quad c_2 = (\tilde{y}_1 - \rho \tilde{y}_0)/\omega$$

12.2 Elementary difference equations

Here, we will consider discrete time k for which it holds $t = kT$, where t is continuous time and T is a fix period of sampling.

First order equations

The first order homogeneous difference equation with constant coefficients has the form
$$y_{k+1} + a_0 y_k = 0, \qquad y_0 = \tilde{y}_0 \tag{12.9}$$
Characteristic equation is linear with unique solution
$$\lambda + a_0 = 0 \quad \rightarrow \quad \lambda = -a_0 \tag{12.10}$$
The solution to the differential equation (12.9) is
$$y_k = \tilde{y}_0 . \lambda^k \tag{12.11}$$

Second order equations

The second order homogeneous difference equation with constant coefficients has the form

$$y_{k+2} + a_1 y_{k+1} + a_0 y = 0, \qquad y_0 = \tilde{y}_0, \; = y_1 = \tilde{y}_1 \tag{12.12}$$

Characteristic equation is quadratic

$$\lambda^2 + a_1\lambda + a_0 = 0 \tag{12.13}$$

with the following types of solution

1. **Two different real roots** λ_1 and λ_2

   The equation (12.12) is

   $$y_{k+2} - (\lambda_1 + \lambda_2) y_{k+1} + \lambda_1\lambda_2 y = 0$$

   The solution is

   $$y_k = c_1\lambda_1^k + c_2\lambda_2^k, \tag{12.14}$$

   where the coefficients c can be obtained as a solution of the set of linear equations

   $$\begin{aligned} c_1 + c_2 &= \tilde{y}_0 \\ \lambda_1 c_1 + \lambda_2 c_2 &= \tilde{y}_1 \end{aligned}$$

   which gives the solution

   $$c_1 = (\lambda_2\tilde{y}_0 - \tilde{y}_1)/(1-\lambda_1) \;\text{ and }\; c_2 = (\lambda_1\tilde{y}_0 - \tilde{y}_1)/(1-\lambda_2)$$

2. **One real double root** λ

   The equation (12.12) is

   $$y_{k+2} - 2\lambda y_{k+1} + \lambda^2 y = 0$$

   The solution is

   $$y_k = c_1\lambda^k + c_2 k\lambda^k, \tag{12.15}$$

   where the coefficients c can be obtained as a solution of the set of linear equations

   $$\begin{aligned} c_1 &= \tilde{y}_0 \\ \lambda c_1 + \lambda c_2 &= \tilde{y}_1 \end{aligned}$$

   which gives the solution

   $$c_1 = \tilde{y}_0 \;\text{ and }\; c_2 = \tilde{y}_1/\lambda - \tilde{y}_0$$

3. **Two complex roots** $\lambda_1 = \rho + \omega\mathrm{i}$ and $\lambda_2 = \rho - \omega\mathrm{i}$

   The equation (12.12) is
   $$y_{k+2} - 2\rho y_{k+1} + \rho^2 + \omega^2 = 0$$
   The solution is
   $$y_k = |c|^k [c_1 \cos(\omega k) + c_2 \sin(\omega k)], \tag{12.16}$$
   where the coefficients c can be obtained as a solution of the set of linear equations
   $$\begin{aligned} c_1 &= \tilde{y}_0 \\ c_1 |Re\lambda| + c_2 |Im\lambda| &= \tilde{y}_1 \end{aligned}$$
   which gives the solution
   $$c_1 = y_0 \;\; \text{and} \;\; c_2 = (\tilde{y}_1 - \tilde{y}_0 |Re\lambda|)/|Im\lambda|$$

12.3 Discretization of a continuous model

Our aim is to construct a discrete regression model whose output is a sampled output of the corresponding continuous one - homogeneous differential equation of 1st or 2nd order with constant coefficients. We will call this task **discretization**.

Let us denote the continuous time by t and the discrete time by k. It holds
$$t = kT, \qquad T \text{ is a period of sampling.}$$

First order equation

Consider a homogeneous differential equation with constant coefficient
$$y' + a_0 y = 0, \qquad y_0 = \tilde{y}_0. \tag{12.17}$$

Then the equivalent difference equation (whose response is the sampled response to the differential one) is
$$y_{k+1} = A_0 y_k, \qquad \text{where} \quad A_0 = \mathrm{e}^{-a_0 T}. \tag{12.18}$$

Solution: The solution to the differential equation is
$$y_t = \tilde{y}_0 \mathrm{e}^{-a_0 t}.$$

To get the discrete version of the solution, we set $t = k$ for actual sample and $t + T = k + 1$ for the shifted one. So, for the actual sample, the solution the same but the substitution k for t and for the shifted sample it holds
$$y_{k+1} = y_{t+T} = \tilde{y}_0 \mathrm{e}^{-a_0 (t+T)} = \tilde{y}_0 \mathrm{e}^{-a_0 t} \mathrm{e}^{-a_0 T} = \mathrm{e}^{-a_0 T} y_k$$
which proves (12.18). $\triangleleft$

Second order equation

- ***Two distinct real roots***

  Let us consider a homogeneous differential equation with constant coefficients

  $$y'' + a_1 y' + a_0 y = 0, \qquad y_0 = \tilde{y}_0,\, y_0' = \tilde{y}_1 \tag{12.19}$$

  whose characteristic equation $\lambda^2 + a_1\lambda + a_0 = 0$ has two different real roots λ_1 and λ_2. Then the equivalent difference equation (whose response is the sampled response to the differential one) is

  $$y_{k+2} = A_1 y_{k+1} + A_0 y_k, \tag{12.20}$$

  where

  $$A_1 = \mathrm{e}^{\lambda_1 T} + \mathrm{e}^{\lambda_2 T}, \qquad A_0 = -\mathrm{e}^{(\lambda_1+\lambda_2)T} \tag{12.21}$$

  Solution: A response to the considered continuous model is

  $$y_t = c_1 \mathrm{e}^{\lambda_1 t} + c_2 \mathrm{e}^{\lambda_2 t}.$$

  Sampling with $t = kT$ and the denotation $y_k = y_{kT}$ gives

  $$y_k = c_1 \mathrm{e}^{\lambda_1 kT} + c_2 \mathrm{e}^{\lambda_2 kT}$$

  This sampled response must obey the difference equation

  $$y_{k+2} = A_1 y_{k+1} + A_0 y_k.$$

  We express still the shifted responses

  $$\begin{aligned} y_{k+1} &= c_1 \mathrm{e}^{\lambda_1 kT} \mathrm{e}^{\lambda_1 T} + c_2 \mathrm{e}^{\lambda_2 kT} \mathrm{e}^{\lambda_2 T} \\ y_{k+2} &= c_1 \mathrm{e}^{\lambda_1 kT} \mathrm{e}^{\lambda_1 2T} + c_2 \mathrm{e}^{\lambda_2 kT} \mathrm{e}^{\lambda_2 2T} \end{aligned}$$

  and notice that they all are expressed in the basis with items $\mathrm{e}^{\lambda_1 kT}$ and $\mathrm{e}^{\lambda_2 kT}$. Thus we substitute into the difference equation and obtain

  $$c_1 \mathrm{e}^{\lambda_1 kT} \mathrm{e}^{\lambda_1 2T} + c_2 \mathrm{e}^{\lambda_2 kT} \mathrm{e}^{\lambda_2 2T} = A_1 (c_1 \mathrm{e}^{\lambda_1 kT} \mathrm{e}^{\lambda_1 T} + c_2 \mathrm{e}^{\lambda_2 kT} \mathrm{e}^{\lambda_2 T}) + A_0 (c_1 \mathrm{e}^{\lambda_1 kT} + c_2 \mathrm{e}^{\lambda_2 kT}).$$

  The coefficients B and A will be obtained by the comparison of items with the same basis element. We obtain the following system of equations

  $$\begin{aligned} c_1 \mathrm{e}^{\lambda_1 2T} &= A_1 c_1 \mathrm{e}^{\lambda_1 T} + A_0 c_1 \\ c_2 \mathrm{e}^{\lambda_2 2T} &= A_1 c_2 \mathrm{e}^{\lambda_1 T} + A_0 c_2. \end{aligned}$$

  The coefficients c get canceled (what is important) and the solution to this system is just what we want to prove. ◁

- ***One double root***

  Let the characteristic equation of (12.19) has one two-fold solution $\lambda = \lambda_1$.

  Then the equivalent difference equation (12.19) has the coefficients

$$A_1 = 2\mathrm{e}^{\lambda_1 T}, \qquad A_0 = -\mathrm{e}^{2\lambda_1 T} \tag{12.22}$$

  Solution: The proof is formally the same as for the two distinct roots, only with basis elements $\mathrm{e}^{\lambda_1 kT}$ and $kT\mathrm{e}^{\lambda_1 kT}$.

  The response to the continuous system is

$$y_k = c_1\mathrm{e}^{\lambda_1 t} + c_2 t\mathrm{e}^{\lambda_1 t}$$

  After expressing the sampled response and its shifted variants, substituting into (12.20) and comparing the terms at the individual basis items, we obtain the following set of equations

$$\begin{aligned} c_1\mathrm{e}^{2\lambda_1 T} + 2c_2 T\mathrm{e}^{2\lambda_1 T} &= A_1 c_1 \mathrm{e}^{\lambda_1 T} + A_1 T c_2 \mathrm{e}^{\lambda_1 T} + A_0 c_1 \\ c_2\mathrm{e}^{2\lambda_1 T} &= A_1 c_2 \mathrm{e}^{\lambda_1 T} + A_0 c_2 \end{aligned}$$

  Again, the solution is just what we wanted to proof. $\triangleleft$

- ***Two complex roots***

  Let the characteristic equation of (12.19) has two complex roots $\lambda_1 = \rho + \omega\mathrm{i}$ and $\lambda_2 = \rho - \omega\mathrm{i}$.

  Then the equivalent difference equation (12.19) has the coefficients

$$A_1 = 2\mathrm{e}^{\rho T}\cos(\omega T), \qquad A_0 = -\mathrm{e}^{2\rho T} \tag{12.23}$$

  Solution: Again, the proof is formally the same as for the previous cases, only with the basis elements $\mathrm{e}^{\rho kT \sin(\omega kT)}$ and $\mathrm{e}^{\rho kT}\cos(\omega kT)$.

  The response of the continuous system is

$$y_k = c_1\mathrm{e}^{\rho t}\cos(\omega t) + c_2\mathrm{e}^{\rho t}\sin(\omega t)$$

  After expressing the sampled response and its shifted variants, substituting into (12.20) and comparing the terms at the individual basis items, we obtain the following set of equations

$$\begin{aligned} -c_1\mathrm{e}^{2\rho T}\sin(2\omega T) + c_2\mathrm{e}^{2\rho T}\cos(2\omega T) &= -A_1 c_1\mathrm{e}^{\rho T}\sin(\omega T) + A_1 c_2 \mathrm{e}^{\rho T}\cos(\omega T) + A_0 c_2 \\ c_1\mathrm{e}^{2\rho T}\cos(2\omega T) + c_2\mathrm{e}^{2\rho T}\sin(2\omega T) &= A_1 c_1\mathrm{e}^{\rho T}\cos(\omega T) + A_1 c_2 \mathrm{e}^{\rho T}\sin(\omega T) + A_0 c_1 \end{aligned}$$

  Once more, the solution is just what we wanted to proof. $\triangleleft$