Notions and notation

June 3, 2016

Contents

1 Notation

1.1 Scilab notation

1.1.1 Dimensions

- $nd \rightarrow$ number of data
- $ni \rightarrow$ number of initial data
- $ns \rightarrow$ length of regression vector
- $nS \rightarrow$ length of extended regression vector
- $nL \rightarrow$ number of data for training (learning)
- $nT \rightarrow$ number of data for classification (testing)
- $np \rightarrow$ length of prediction
- $nh \rightarrow$ length of control interval
- $mc \rightarrow$ number of values of discrete variable
- $ky \rightarrow$ dimension of output
- $ku \rightarrow$ dimension of input
- $kx \rightarrow$ dimension of state

1.1.2 Variables

- $yt \rightarrow$ continuous output
- $ut \rightarrow$ continuous input
- $zt \rightarrow$discrete output
- $ut \rightarrow$ discrete input
- $ct \rightarrow$ pointer
- $yp \rightarrow$ prediction
- $ce \rightarrow$ pointer estimate
- $\psi \rightarrow$ continuous regression vector
- $\phi \rightarrow$ discrete regression vector

1.1.3 Objects

- $Sim \rightarrow$ object Simulation + data
- $Est \rightarrow$ object Estimation
- $Pre \rightarrow$ object Prediction
- $Con \rightarrow$ object Control
- $Fil \rightarrow$ object Filtration
- $Cla \rightarrow$ object Classification (can be also in Est)
- $Hyp \rightarrow$ object Hypotheses (can be also in Pre)

1.1.4 Component models

- $Cy \rightarrow$ component for continuous data
- $Cz \rightarrow$ component for discrete data
- $Cx \rightarrow$ state component
- $Ce \rightarrow$ exponential component
- $Cu \rightarrow$ uniform component
- $Cp \rightarrow$ component for pointer

1.1.5 Model requisites

- $ord \rightarrow$ model order
- $th \rightarrow$ parameters or their estimates (reg. coeffic. or table probs)
- tha, thb, $thk \rightarrow$ regression coefficients
- $cv \rightarrow$ noise covariance or its estimate
- $sd \rightarrow$ standard deviation or its estimate
- $psi \rightarrow$ regression vector
- $Psi \rightarrow$ extended regression vector
- $V \rightarrow$ statistics (either GiW or Di)
- $ka \rightarrow$ data counter
- rw, rv, $rx \rightarrow$ covariances in KF
- M, N, F, A, B, $G \rightarrow$ parameters of state model
- $stp \rightarrow$ set-point
- $Om \rightarrow$ penalization matrix $x' \cdot Om \cdot x$

- $om \rightarrow$ penalization of y-square
- $la \rightarrow$ penalization of u-square
- $T \rightarrow$ period of sampling
- $T0 \rightarrow$ basic period of sampling (measurements)

1.1.6 Structures

- **normal component**
  Sim.Cy(i).th → reg.coef.
  Sim.Cy(i).cv → covar.
  Mix.Cy(i).th → reg.coef.
  Mix.Cy(i).cv → covar.
  Mix.Cy(i).V → statistics
  Mix.Cy.ka → counter
  Mix.yp(:,t) → final store
- **categorical component**
  Mix.Cz(i).th → parameters
  Mix.Cz(i).V → statistics
- **state component**
  Sim.Cx(i).M
  Sim.Cx(i).N
  Sim.Cx(i).A
  Sim.Cx(i).B
  Sim.Cx(i).rw
  Sim.Cx(i).rv
  Mix.Cx(i).rx → current estimate
  Mix.Cx(i).x → current estimate
  Mix.Cx(i).yp → current estimate
  Mix.rx → merged covariance estimate (from components)
  Mix.x → merged state estimate (from components)
  Mix.yp(:,t) → final store of variable
  Mix.xt(:,t) → final store of variable
- **pointer model**
  Mix.Cp.th → alfa
  Mix.Cp.V → nu

1.2 Densities

1.2.1 Continuous distributions

For X, Y continuous random variables and $x\,y$ realizations we denote

$$f_{X|Y}(x|y) \equiv f(x|y),$$

the names of random variables are indicated by their realizations.

1.2.2 Discrete distributions

For I, J discrete random variables and i, j their realizations we denote

$$f_{I|J}(i|j) \equiv f(I = i|J = j)$$

1.2.3 Mixed distributions

For X, Y continuous and I, J discrete random variables and x, y, i, j their realizations we denote

$$f_{X,I|Y,J}(x, i|y, j) \equiv f(x,\, I = i|y,\, J = j)$$

1.2.4 Chain rule

$$f(A, B, C, \cdots, Z) = f(A|B, C, \cdots Z)\, f(B|C, \cdots, Z) \cdots f(Z)$$

1.2.5 Bayes rule

$$f(A|B, C) = \frac{f(B|A, C)\, f(A|C)}{f(B|C)}$$

2 Trifles

2.1 Formulas

2.1.1 Derivative of determinant

For symmetric matrix A the derivative of its determinant is

$$\frac{\partial}{\partial A}|A| = |A|A^{-1}$$

2.1.2 Derivative of matrix inversion

For symmetric matrix A the derivative of its inverse is given by the formula

$$\frac{\partial}{\partial A}A^{-1} = -A^{-2}$$

2.1.3 Derivative of square form

For x a vector and A symmetric matrix it holds

$$\frac{\partial}{\partial x}\mathrm{tr}\left(x'Ax\right) = 2Ax$$

$$\frac{\partial}{\partial A}x'A^{-1}x = -A^{-1}xx'A^{-1}$$

2.1.4 Inversion lemma

For matrices A, B, C, D of compatible dimensions, it holds

$$\left(A+BCD\right)^{-1} = A^{-1} - A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$$

2.1.5 Completion to squares

Scalar case

For scalar variables x and y and constants a, b, c it holds

$$ax^2+2bxy+cy^2 = a\left[x^2+2x\frac{b}{a}y+\left(\frac{b}{a}y\right)^2-\left(\frac{b}{a}y\right)^2\right]+cy^2 =$$

$$a\left(x+\frac{b}{a}y\right)^2+cy^2-\frac{b^2}{a}y^2 = a\left(x+\frac{b}{a}y\right)^2+\frac{ac-b^2}{a}y^2.$$

Vector case

For variables x and y in the form of column vectors and constant matrices A, B, C with corresponding dimensions, A a C sammetric, it holds

$$x'Ax+2x'By+y'Cy =$$

$$= x'Ax+2x'AA^{-1}By+\left(A^{-1}By\right)'AA^{-1}By-\left(A^{-1}By\right)'AA^{-1}By+y'Cy =$$

$$= \underbrace{\left(x+A^{-1}By\right)'A\left(x+A^{-1}By\right)}_{\text{square}}+\underbrace{y'\left(C-B'A^{-1}B\right)y}_{\text{reminder}}.$$

2.2 Functions

2.2.1 Kronecker function

The Kronecker function $\delta\left(i,j\right)$ is defined as follows

$$\delta\left(i,j\right) = \begin{cases} 1 & \text{for } i=j \\ 0 & \text{othervise} \end{cases} \tag{1}$$

If a and b are vectors of integers of the same length, the definition of the Kronecker function is analogous

$$\delta(a,b) = \begin{cases} 1 & \text{for } a_i = b_i,\ \forall i \\ 0 & \text{othervise} \end{cases} \tag{2}$$

2.2.2 Dirac function

The Dirac delta function $\delta(t-\tau)$ is defined as a functional acting on continuous functions $g(t)$ as follows

$$\int_{-\infty}^{\infty} g(\tau)\,\delta(t-\tau)\,d\tau = g(t) \tag{3}$$

It can also be characterized as a function which has almost all values equal to zero only in the origin, i.e. for its argument equal to zero, has a value equal to infinity so that it holds $\int_{-\infty}^{\infty}\delta(\tau)\,d\tau$=1.

2.2.3 Gamma function

$$\Gamma(u) = \int_0^{\infty} x^{u-1}\mathrm{e}^{-x}\,dx$$

Formulas

$$\Gamma(u+1) = u\Gamma(u)$$

2.2.4 Incomplete gamma function

$$\Gamma(u,s) = \int_0^{s} x^{u-1}\mathrm{e}^{-x}\,dx$$

2.2.5 Beta function

$$B(u,v) = \int_0^1 p^{u-1}(1-p)^{v-1}\,dp$$

Formulas

$$B(u,v) = \frac{\Gamma(u)\,\Gamma(v)}{\Gamma(u+v)}$$

$$B(u+1,v) = \frac{u}{u+v}B(u,v)$$

2.2.6 Multivariate beta function

$$B(u_1,u_1,\cdots,u_n) = \frac{\prod_i \Gamma(u_i)}{\Gamma\left(\sum_i u_i\right)}$$

for vector argument.

2.2.7 Generalized beta function

$$B(u) = \prod_j \frac{\prod_i \Gamma\left(u_{i|j}\right)}{\Gamma\left(\sum_i u_{i|j}\right)}$$

for matrix argument.

2.3 Distributions

2.3.1 Exponential family of distributions

A random variable is said to have a distribution belonging to the exponential family is its probability density function can be given the form

$$f(x;\Theta) = A(\Theta)\exp\{\langle Q(\Theta), D(x)\rangle\} \tag{4}$$

where Θ is a parameters of the distribution, A and Q are functions only of the parameter Θ (not data x), D is a function only of the data x (not the parameter Θ) and $\langle\cdot,\cdot\rangle$ denotes a scalar product, linear in Q, i.e. we demand so that it holds

$$\sum_i a_i \langle Q, D_i\rangle = \left\langle Q, \sum_i a_i D_i \right\rangle$$

2.3.2 Categorical distribution (static)

Probability function of the distribution is

$$f(y|p) = p_y$$

an it can be given a form of the table

y	1	2	$\cdots$	n_l
$f(y)$	p_1	p_2	$\cdots$	p_{n_l}

,

where p_i are probabilities, $p_i \geq 0$, $i = 1, 2, \cdots, n_l$ a $\sum_{i=1}^{n_l} p_i = 1$.

2.3.3 Categorical distribution (dynamic)

For ψ discrete regression vector the distribution is

$$f(y|\psi,\alpha) = \alpha_{y|\psi}.$$

and in the table, for $y \in \{1,2\}$ and $\psi = [u, v]'$, kde $u, v \in \{1,2\}$

$f(y|u,v)$

$[u,v]$	$y=1$	$y=2$
1, 1	$\alpha_{1\|11}$	$\alpha_{2\|11}$
1, 2	$\alpha_{1\|12}$	$\alpha_{2\|12}$
2, 1	$\alpha_{1\|21}$	$\alpha_{2\|21}$
2, 2	$\alpha_{1\|22}$	$\alpha_{2\|22}$

,

where $\alpha_{i|jk}$ conditional probabilities, $\sum_{i=1}^{2} \alpha_{i|jk} = 1$, $\forall j,k$.

2.3.4 Dirichlet distribution

Dirichlet distribution has pdf

$$f\left(\alpha|d\left(t\right)\right)=\frac{1}{B\left(\nu_t\right)}\prod_{i\in y^*}\prod_{\varphi\in\psi^*}\alpha_{i|\varphi}^{\nu_{i|\varphi;t}}, \tag{5}$$

kde

ν_t is a statistics of the distribution with the same structure as the parameters has,

$B\left(\nu\right)$ is generalized beta function (see 2.2.5)

2.3.5 Normal distribution

Normal distribution for **scalar** y has the form

$$f\left(y|m,r\right)=\frac{1}{\sqrt{2\pi}}r^{-0.5}\exp\left\{-\frac{1}{2r}\left(y_t-m\right)^2\right\}. \tag{6}$$

and in **multivariate** case

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

with expectation m and variance r.

Dynamic normal distribution

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

where ψ_t is a regression vector, $\Theta=\{\theta,r\}$ is a set of parameters {regression coefficients, noise variance}.

Expectation

$$E\left[y_t|\psi_t,\Theta\right]=\psi_t'\theta,$$

Variance

$$D\left[y_t|\psi_t,\Theta\right]=r.$$

Another form for normal distribution is

$$f\left(y_t|\psi_t,\Theta\right)=\frac{1}{\sqrt{2\pi}}r^{-0.5}\exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right]D_t\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}. \tag{7}$$

where $D_t=\begin{bmatrix}y_t\\ \psi_t\end{bmatrix}\left[y_t\,\psi_t\right]$ is called data matrix.

2.3.6 Gauss-inverse-Wishart (GiW) distribution

The distribution has the form

$$f\left(\Theta|d\left(t\right)\right) \propto r^{-0.5\kappa_t} \exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right] V_t \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\}, \tag{8}$$

where κ_t and V_t are statistic (κ_t is called counter and V_t is extended information matrix).

2.4 Principles

2.4.1 Natural conditions of control

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

as u_t is a function of $d\left(t-1\right)$ and nothing else. From it, using Bayes rule, we have

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

The assumption is based on the fact, that he, who estimates and who controls is one person.

2.4.2 Approximated prediction with regression model

If Θ is an unknown random variable, the Bayesian statistics treats it optimally through its posterior pdf $f\left(\Theta|d\left(t\right)\right)$. E.g. to express the data prediction $f\left(d_t|d\left(t-1\right)\right)$ using the model $f\left(d_t|\psi_t, \Theta\right)$ we proceed optimally as follows

$$f\left(d_t|d\left(t-1\right)\right) = \int f\left(d_t, \Theta|d\left(t-1\right)\right) d\Theta = \int f\left(d_t|\psi_t, \Theta\right) f\left(\Theta|d\left(t-1\right)\right) d\Theta$$

where $f\left(d_t|\psi_t, \Theta\right)$ is parametrized model and $f\left(\Theta|d\left(t-1\right)\right)$ is posterior pdf estimating Θ taken from the last step of estimation.

The approximate way, substantially simplifying computations, is lays in substituting the posterior pdf by a Dirac function (see 3)

$$f\left(\Theta|d\left(t-1\right)\right) \to \delta\left(\Theta, \hat{\Theta}_{t-1}\right)$$

where $\hat{\Theta}_{t-1}$ is point estimate of Θ for time $t-1$. We obtain

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

which is equivalent to substituting the point estimate $\hat{\Theta}_{t-1}$ for Θ instead of working with the full posterior pdf for Θ.

2.4.3 Quasi-Bayes approximation

In Quasi-Bayes we deal with the object $\delta\left(c_t, i\right)$ where c_t is random variable (pointer denoting active component) and i is integer (some realization of c_t).

If the activities of components in mixture model are known, $\delta\left(c_t, i\right)$ can be considered a distribution of c_t with all zeros and one only on the position of the active component.

If the activities are not known, i.e. we do not know the correct i, we must treat the function $\delta\left(c_t, i\right)$ only in its expected value.

2.4.4 Approximated likelihood

Continuous model

Likelihood $L_t(\Theta)$ for model $f(y_t|\psi_t,\Theta)$ is defined as

$$L_t(\Theta) = \prod_{\tau=1}^{t} f(y_\tau|\psi_\tau,\Theta) =$$

$$= f(y_t|\psi_t,\Theta)\prod_{\tau=1}^{t-1} f(y_\tau|\psi_\tau,\Theta) \propto f(y_t|\psi_t,\Theta)\, f(\Theta|d(t-1))$$

If we want to evaluate model behavior (model given by its parameters) with respect to the data or to compare several models (each given by its parameters) for the data sample, we can use likelihood of variances (v-likelihood)

$$L_t^v = \int_{\Theta^*} L_t(\Theta)\, d\Theta.$$

However, construction of the likelihood and even computation of the integral does not need to be simple. That is why we introduce approximated v-likelihood (a-likelihood) as

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

A motivation for the definition is as follows:

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

$$\text{and for } f(\Theta|d(t-1)) \to \delta\left(\Theta,\hat{\Theta}_{t-1}\right) \text{ we have}$$

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

where $\hat{\Theta}_{t-1}$ is point estimate of parameters and y_t, ψ_t is actual data sample.

Discrete model

Analogously to the continuous case, the a-likelihood for discrete model

$$f(y_t|\psi_t,\alpha) = \alpha_{y_t|\psi_t}$$

can be defined as a model with point estimates of parameters and actual data sample inserted

$$L_t^a = \hat{\alpha}_{y_t|\psi_t;t-1}$$

which cam be obtained from the model with $\hat{\alpha}_{t-1}$ inserted, taking the element of the table in the column given by the value of y_t and the row corresponding to the code of ψ_t.

3 Objects

3.1 Models

3.1.1 Dynamic pointer model

- pdf

$$f\left(c_t|c_{t-1},\alpha\right)$$

- parametric form

c_{t-1}	$c_t=1$	$c_t=2$	$c_t=3$
1	$\alpha_{1\|1}$	$\alpha_{2\|1}$	$\alpha_{3\|1}$
2	$\alpha_{1\|2}$	$\alpha_{2\|2}$	$\alpha_{3\|2}$
3	$\alpha_{1\|3}$	$\alpha_{2\|3}$	$\alpha_{3\|3}$

3.1.2 Static pointer model

- pdf

$$f\left(c_t|\alpha\right)$$

- parametric form

$c_t=1$	$c_t=2$	$c_t=3$
α_1	α_2	α_3

3.1.3 Regression component

- pdf of the c-th component

$$f_c\left(y_t|\psi_t,\Theta^{(c)}\right)$$

- parametric form

$$y_t=\psi_t^{'}\theta^{(c)}+e_t$$

3.1.4 Categorical component

- pdf of the c-th component

$$f_c\left(z_t|\phi_t, \Theta^{(c)}\right)$$

- parametric form

$[u_t\ z_{t-1}]$	$z_t = 1$	$z = 2$
1, 1	$\Theta^{(c)}_{1\|11}$	$\Theta^{(c)}_{2\|11}$
1, 2	$\Theta^{(c)}_{1\|12}$	$\Theta^{(c)}_{2\|12}$
2, 1	$\Theta^{(c)}_{1\|21}$	$\Theta^{(c)}_{2\|21}$
2, 2	$\Theta^{(c)}_{1\|22}$	$\Theta^{(c)}_{2\|22}$

3.1.5 State-space component

- pdfs of the c-th component

$$f_c\left(y_t|x_{t-1},\, u_t\right)$$
$$f_c\left(x_t|x_{t-1},\, u_t\right)$$

- parametric form

$$y_t = A^{(c)}x_{t-1} + B^{(c)}u_t + G^{(c)} + v_t$$
$$x_t = M^{(c)}x_{t-1} + N^{(c)}u_t + F^{(c)} + w_t$$

3.1.6 Mixed-data component

- pdf of continuous part of the c-th component

$$f_c\left(y_t|\psi_t, \Theta^{(c)}\right)$$

  where $\psi_t = [y_{t-1}, \cdots, z_t, \cdots, u_t, \cdots]$

- pdf of discrete part of the c-th component

$$f_c\left(z_t|\phi_t, \vartheta^{(c)}\right)$$

  where $\phi_t = [z_{t-1},\, z_{t-1}, \cdots]$ (no continuous variable present)

- parametric form of the continuous part

$$y_t = \psi_t^{'}\theta^{(c)} + e_t$$

- parametric form of the discrete part

z_{t-1}	$z_t = 1$	$z_t = 2$	$z_t = 3$
1	$\vartheta_{1\|1}$	$\vartheta_{2\|1}$	$\vartheta_{3\|1}$
2	$\vartheta_{1\|2}$	$\vartheta_{2\|2}$	$\vartheta_{3\|2}$
3	$\vartheta_{1\|3}$	$\vartheta_{2\|3}$	$\vartheta_{3\|3}$

3.1.7 Logistic model

- pdf of the c-th component

$$f_c\left(y_t = 0|x_t,\, b\right) \rightsquigarrow p$$
$$f_c\left(y_t = 1|x_t,\, b\right) \rightsquigarrow 1 - p$$

- parametric form

$$p = \frac{\exp\left\{x_t b\right\}}{1 + \exp\left\{x_t b\right\}}$$
$$1 - p = \frac{1}{1 + \exp\left\{x_t b\right\}}$$

3.2 Conjugated priors

3.2.1 Gauss-inverse-Wishart pdf

$$f\left(\Theta|d\left(t\right)\right) \propto r^{-0.5\kappa_t} \exp\left\{-\frac{1}{2r}\left[-1,\, \theta'\right] V_t \left[-1,\, \theta'\right]'\right\}$$

where $\Theta = \{\theta,\, r\}$ are model parameters: regression coefficients and noise variance, κ_t and V_t are statistics for estimation - see (2.3.6).

3.2.2 Dirichlet pdf

$$f\left(\alpha|d\left(t\right)\right) \propto \prod_{i|j} \alpha_{i|j}^{\nu_{i|j;t}}$$

where α is a matrix of parameters, ν_t is a matrix of statistics for estimation (with the same form as α) - see (2.3.4).

3.3 Statistics update

3.3.1 Update for continuous model

- regression model

$$\begin{aligned} V_t &= V_{t-1} + \Psi\Psi' \\ \kappa_t &= \kappa_{t-1} + 1 \end{aligned}$$

- regression component

$$\begin{aligned} V_t^{(c)} &= V_{t-1}^{(c)} + w_{c;t}\Psi\Psi' \\ \kappa_t^{(c)} &= \kappa_{t-1}^{(c)} + w_{c;t} \end{aligned}$$

3.3.2 Update for discrete model

- categorical model

$$\begin{aligned} \nu_{i;t} &= \nu_{i;t-1} + 1 \quad \text{static pointer} \\ \nu_{i|j;t} &= \nu_{i|j;t-1} + 1 \quad \text{dynamic pointer} \end{aligned}$$

- categorical component

$$\begin{aligned} \nu_{i;t} &= \nu_{i;t-1} + w_{i;t} \quad \text{static pointer} \\ \nu_{i|j;t} &= \nu_{i|j;t-1} + W_{ji;t} \quad \text{dynamic pointer} \end{aligned}$$

3.4 Point estimates

3.4.1 Point estimate of regression model parameters

- partitioning of information matrix

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix} = \begin{bmatrix} \cdot & \overline{} \\ | & \boxdot \end{bmatrix}$$

- point estimate of regression coefficients

$$\hat{\theta}_t = V_\psi^{-1} V_{y\psi}$$

- point estimate of model noise variance

$$\hat{r}_t = \frac{V_y - V_{y\psi} V_\psi^{-1} V_{y\psi}}{\kappa_t}$$

3.4.2 Point estimate of categorical model parameters

$$\hat{\alpha}_{i;t} = \nu_{i;t} / \sum_{j=1}^{n} \nu_{j;t} \quad \text{static model}$$

$$\hat{\alpha}_{i|j;t} = \nu_{i|j;t} / \sum_{k=1}^{n} \nu_{k|j;t} \quad \text{static model}$$

Point estimates of parameters are obtained by normalization of statistics (in rows in dynamic case).

3.4.3 Dynamic weights of components

$$W_{ji;t} = \underbrace{f_i\left(y_t|\psi_t, \hat{\Theta}_{t-1}\right)}_{\text{comp. model}} \underbrace{\hat{\alpha}_{i|j;t-1}}_{\text{ptr. model}} \underbrace{f\left(c_{t-1} = j|d\left(t-1\right)\right)}_{\text{ptr. prior}}$$

Implementation

- $\hat{m}$ - column vector of $f_i\left(y_t|\psi_t, \hat{\Theta}_{t-1}\right)$ for $i = 1, 2, \cdots, n_c = c^*$
- $\hat{\alpha}$- square matrix of $\hat{\alpha}_{i|j;t-1}$ for $i, j \in c^*$
- $\hat{f}$ - column vector of $f\left(c_{t-1} = j|d\left(t-1\right)\right)$ for $j \in c^*$

$$W_t = \left(\hat{f} \cdot \hat{m}\right) .* \hat{\alpha}$$

$$(| \quad —) \quad \boxdot$$

where $\cdot$ is matrix multiplication and $.*$ is entry-vise "dot product"

or

$$W_t = \text{diag}\left(\hat{m}\right) \cdot \hat{\alpha} \cdot \text{diag}\left(\hat{f}\right)$$

$$\boxdot \quad \boxdot \quad \boxdot$$

3.4.4 Static weights of components

$$w_{i;t} = \underbrace{f_i\left(y_t|\psi_t, \hat{\Theta}_{t-1}\right)}_{\text{comp. model}} \underbrace{f\left(c_{t-1} = i|d\left(t-1\right)\right)}_{\text{ptr. descripton}}$$

Implementation

- $\hat{m}$ - column vector of $f_i\left(y_t|\psi_t, \hat{\Theta}_{t-1}\right)$ for $i = 1, 2, \cdots, n_c = c^*$

- $\hat{f}$ - column vector of $f\left(c_{t-1}=i|d\left(t-1\right)\right)$ for $i \in c^*$

$$w_t = \hat{f} \cdot * \hat{m}$$

where $.*$ is entry-vise "dot product".

3.4.5 Kalman filter

$$f\left(x_{t-1}|d\left(t\right)\right) \propto f\left(y_t|x_{t-1}\right) f\left(x_{t-1}|d\left(t-1\right)\right) \quad \text{filtration}$$
$$f\left(x_t|d\left(t\right)\right) \propto f\left(x_t|x_{t-1},u_t\right) f\left(x_{t-1}|d\left(t\right)\right) \quad \text{prediction}$$

Point estimates for normal distribution

$$f\left(y_t|x_{t-1},u_t\right) = N_{y_t}\left(Ax_{t-1|t-1}+Bu_t+G,\ r_y\right)$$
$$f\left(x_t|x_{t-1},u_t\right) = N_{x_t}\left(Mx_{t-1|t}+Nu_t+F,\ r_x\right)$$

$$f\left(x_i|d\left(j\right)\right) = N_{x_i}\left(x_{i|j},\ R_{i|j}\right)$$
$$\text{for } i|j = t-1|t-1,\ t-1|t,\ t|t$$

Kalman filter

$y_p = Ax_{t-1|t-1}+Bu_t+G$ — output prediction

$R_p = r_y + AR_{t-1|t-1}A'$

$R_{t-1|t} = R_{t-1|t-1} - R_{t-1|t-1}A'R_p^{-1}AR_{t-1|t-1}$

$K = R_{t-1|t}A'r_y^{-1}$ — Kalman gain

$x_{t-1|t} = x_{t-1|t-1} + K\left(y_t - y_p\right)$ — state correction

$x_{t|t} = Mx_{t-1|t}+Nu_t+F$ — state prediction

$R_{t|t} = r_x + MR_{t-1|t}M'$

4 Derivations

4.1 Formula for beta function using Γ

The multivariate beta function of $x = [x_1, x_2, \cdots, x_n]$ is defined through the integral

$$B\left(x\right) = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} p_i^{x_i}\, dp_1 \cdots dp_n$$

where $p_i \geq 0 \ \forall i$ and $\sum_{i=1}^{n} p_i = 1$.
We are going to show, that it holds

$$B(x) = \frac{\prod_{i=1}^{n} \Gamma(x_i)}{\Gamma\left(\sum_{i=1}^{n} x_i\right)}$$

Proof: For $n = 3$ the integrand is

$$p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{x_3} = p_1^{x_1} p_2^{x_2} \left[(1 - p_1)\left(1 - \frac{p_2}{1 - p_1}\right)\right]^{x_3} = (*)$$

$$\text{we substitute } \tfrac{p_2}{1-p_1} = q_2 \rightarrow p_2 = q_2 (1 - p_1)$$

$$(*) = p_1^{x_1} q_2^{x_2} (1 - p_1)^{x_2} (1 - p_1)^{x_3} (1 - q_2)^{x_3} =$$
$$= p_1^{x_1} (1 - p_1)^{x_2 + x_3} q_2^{x_2} (1 - q_2)^{x_3}$$

After integration we obtain

$$B(x_1, x_2 + x_3) B(x_2, x_3) =$$
$$= \frac{\Gamma(x_1)\Gamma(x_2 + x_3)}{\Gamma(x_1 + x_2 + x_3)} \frac{\Gamma(x_2)\Gamma(x_3)}{\Gamma(x_2 + x_3)} = \frac{\Gamma(x_1)\Gamma(x_2)\Gamma(x_3)}{\Gamma(x_1 + x_2 + x_3)}$$

4.2 KL distance for categorical distributions

Kerridge inaccuracy between two discrete distributions f and g is

$$K = \sum_i f_i \ln \frac{1}{g_i} = -\sum_i f_i \ln g_i$$

4.3 KL distance for normal distributions

Kerridge inaccuracy between two normal distributions f and f_i is

$$K_i = \frac{1}{2}\left[\ln(2\pi r) + \frac{r_i}{r} + \frac{(m_i - m)^2}{r}\right]$$

Derivation

$$K_i = \int f_i(x) \ln \frac{1}{f(x)} dx =$$

$$= \int N_x(m_i, r_i) \ln \frac{1}{k r^{-0.5} \exp\left\{-\frac{1}{2r}(x - m)^2\right\}} dx =$$

$$= \int N_x(m_i, r_i) \left[0.5 \ln(2\pi r) + \frac{1}{2r}(x - m)^2\right] dx =$$

$$= 0.5 \ln(2\pi r) + \frac{1}{2r} \int (\{x - m_i\} + \{m_i - m\})^2 N_x(m_i, r_i)\, dx =$$

$$= 0.5 \left\{\ln(2\pi r) + \frac{1}{r}\left[r_i + (m_i - m)^2\right]\right\}$$