System

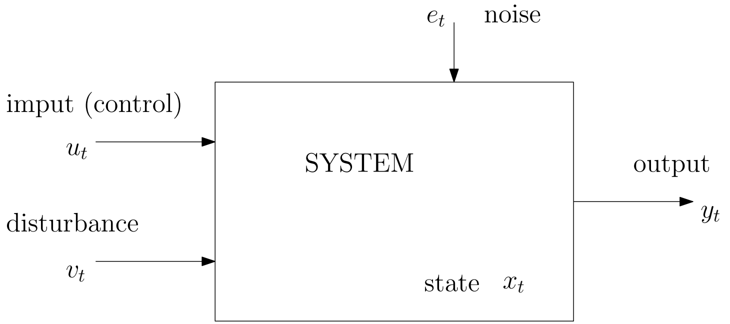
Process - System - Model

- Process - part of reality we are interested in.
- System - variables with their relations.
- Model - mathematical relation of the monitored variable and other explanatory variables.

Remark

If some delayed monitored variables are among the explanatory variables, the system is **dynamic**. Otherwise it is **static**.

Variables in the system



Output: Monitored variable.

Input: Manipulated variable - control.

Disturbance: Can be measured, cannot be manipulated.

State: Cannot be measured, is estimated from data.

Noise: Neither can be measured nor predicted.

Model

Bayesian view on model

Conditional probability density function (pdf)

$$f\left(y_t|\psi_t^{'},\Theta\right)$$

$\psi_t = [u_t, y_{t-1}, u_{t-1}, \cdots, y_{t-n}, u_{t-n}, 1]'$ - regression vector;

$\Theta = \{\theta, r\}$; $\theta = [b_0, a_1, b_1, \cdots, a_n, b_n, k]'$, θ - regression coefficients, r - noise variance.

It is a stochastic dependence of y_t on ψ_t with relations expressed by probability density function (pdf).

Regression model

The variables are continuous, ψ can have also some discrete ones.

The above pdf expression can be generated by the stochastic equation

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

$$= \psi_t' \theta + e_t$$

where e_t (noise) is i.i.d. (independent, identically distributed) random variable with zero expectation and variance r.

$E[y_t|\psi_t, \Theta] = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \cdots + a_n y_{t-n} + b_n u_{t-n} + k,$

$D[y_t] = D[e_t] = r$

Program: **T11simCont.sce** (page 83)

Discrete model

All variables are discrete (finite number of values)

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

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

$\sum_{i=1}^{2}\Theta_{i|jk}=1$ - conditional probabilities.

For given $[u_t, y_{t-1}]$ the output y_t is generated with the pdf $\left[\Theta_{1|u_t,y_{t-1}}\Theta_{2|u_t,y_{t-1}}\right]$.

Program: **T13simDisc.sce** (page 86)

Model of logistic regression

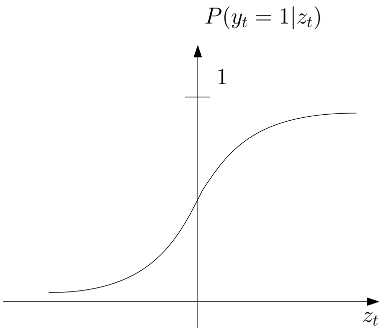
The output is discrete (0 or 1) and it depends on continuous variables.

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

where

$$z_t=\psi_t\Theta+e_t$$

The model has the following form - transformation from z to $p=f\left(y_t=1|z_t\right)$



State-space model

Describes the state variable x_t

– state model (state prediction)

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

– output model (state filtration)

$$y_t = Ax_t + Bu_t + v_t$$

M, N, A, B are known matrices,

w_t, v_t are noises with zero expectations and known covariances R_w, R_v

State form of regression model

For 2^{nd} order regression model

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + a_2 y_{t-2} + b_2 u_{t-2} + k + e_t$$

the state form is

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

Program: **T15simState.sce** (page 88)

Estimation

Bayesian estimation

Notation: d_t data at t, $d(t) = \{d_0, d_1, \cdots, d_t\}$ data up to t, d_0 prior.

$f(\Theta|d(t-1))$, $f(\Theta|d(t))$ description of parameters (prior, posterior)

Bayes rule

$$\underbrace{f(\Theta|d(t))}_{\textbf{posterior}} \propto \underbrace{f(y_t|\psi_t,\Theta)}_{\textbf{model}}\underbrace{f(\Theta|d(t-1))}_{\textbf{prior}}$$

– Natural conditions of control $\quad f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1))$

– Batch estimation

$$f(\Theta|d(t)) \propto \underbrace{\left[\prod_{\tau=1}^{t} f(y_\tau|\psi_\tau\Theta)\right]}_{\text{Likelihood}\, L_t(\Theta)} f(\Theta|d(0))$$

– Self reproducing prior $\quad f(\Theta|d(t-1)) \rightarrow f(\Theta|d(t))$ - the same form.

Results of estimation

- Posterior pdf $f\left(\Theta|d\left(t\right)\right)$ probabilities of parameter values
- Point estimate of parameter (expectation)

$$\hat{\Theta}_t = E\left[\Theta|d\left(t\right)\right] = \int_{-\infty}^{\infty} \Theta f\left(\Theta|d\left(t\right)\right) d\Theta$$

Estimation of regression model

Application of Bayes rule with regression model and prior/posterior in the form of Gauss-inverse-Wishart distribution

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

Statistics update

$$V_t = V_{t-1} + D_t$$
$$\kappa_t = \kappa_{t-1} + 1$$

where $D_t = \left[\begin{array}{c} y_t \\ \psi_t \end{array}\right]\left[y_t, \psi_t'\right]$ is data matrix, V_t is information matrix and κ_t is counter.

Programs: **T22estCont_B.sce; T22estCont_B2.sce; T22estCont_B3.sce;** (page 93 and further)
T22estCont_B4.sce (data from Strahov are on our web)

Point estimates of parameters

– division of information matrix

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

– estimates of regression coefficients

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

– estimate of noise variance

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

Batch estimation

For 2^{nd} order regression model

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + a_2 y_{t-2} + b_2 u_{t-2} + k + e_t$$

Construct

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} u_1 & y_0 & u_0 & y_{-1} & u_{-1} & 1 \\ u_2 & y_1 & u_1 & y_0 & u_0 & 1 \\ u_3 & y_2 & u_2 & y_1 & u_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ u_N & y_{N-1} & u_{N-1} & y_{N-2} & u_{N-2} & 1 \end{bmatrix}$$

Regression coefficients are $\hat{\theta}_N = (X'X)^{-1} X'Y$ in the order in which the rows of X are constructed.

Program: **T21estCont_LS.sce** (page 90)

Estimation of categorical model

The pdf of parameter has the Dirichlet form

$$f\left(\Theta|d\left(t\right)\right)\propto\prod_{y|\psi}\Theta_{y|\psi}^{\nu_{y|\psi;0}}$$

with the statistics update

$$\nu_{y_t|\psi_t;t}=\nu_{y_t|\psi_t;t-1}+1$$

The update runs as follows:

ν is a matrix with columns denoted by values of y_t and rows corresponding to configurations of values of ψ_t (the same as in model).

In the update we find the entry denoted by y_t and the row corresponding to the configuration of ψ_t and we increase it by one.

Point estimate of the parameter is given by ν normalized so that the sums of rows are equal to one.

Program: **T23estDisc.sce** (page 101)

Estimation of logistic model

It is not recursive - we must construct likelihood (for all measured data) and maximize it numerically.

Likelihood

$$L_t = \prod_{\tau=1}^{t} \frac{\exp\{y_\tau z_\tau\}}{1 + \exp\{z_\tau\}}$$

$$\ln L_t = \sum_{\tau=1}^{t} \left[y_\tau z_\tau - \ln\left(1 + \exp\{z_t\}\right)\right]$$

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

Prediction

Definition

Predictive pdf (k-step ahead)

$$f\left(y_{t+k}|y\left(t-1\right),u\left(t+k\right)\right)\rightarrow f\left(y_{t+k}|y\left(t-1\right)\right)$$

Point prediction

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

Zero step prediction

u_t given for all t needed.

Model $\quad f\left(y_t|y\left(t-1\right),\Theta\right)$

Predictive density

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

$$=\int_{\Theta^*}\underbrace{f\left(y_t|y\left(t-1\right),\Theta\right)}_{\text{model}}\;\underbrace{f\left(\Theta|y\left(t-1\right)\right)}_{\text{posterior from}\,t-1}\;d\Theta\rightarrow$$

$$\rightarrow\sum_{\theta_i\in\Theta}f\left(y_t|y\left(t-1\right),\theta_i\right)f\left(\theta_i|y\left(t-1\right)\right)$$

$\cdots$ average (expectation) of all possible models weighted by their probabilities.

One step prediction

u_t given for all t needed.

Model $\quad f\left(y_t | y\left(t-1\right), \Theta\right)$

Predictive density

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

$$= \int \int f\left(y_{t+1} | y\left(t\right), \Theta\right) f\left(y_t | y\left(t-1\right), \Theta\right) f\left(\Theta | y\left(t-1\right)\right) dy_t d\Theta = (*)$$

$$\int \int \left(\text{model}\left(y_{t+1}\right)\right) \left(\text{model}\left(y_t\right)\right) \left(\text{posterior}\left(t-1\right)\right) dy_t d\Theta$$

point estimates of parameters $\cdots$ $f(\Theta|y(t-1)) \doteq \delta\left(\Theta, \hat{\Theta}_{t-1}\right)$

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

$$\doteq \int f\left(y_{t+1}|y(t), \hat{\Theta}_{t-1}\right) f\left(y_t|y(t-1), \hat{\Theta}_{t-1}\right) dy_t = (**)$$

point estimates of outputs $\cdots$ $f\left(y_t|y(t-1), \hat{\Theta}_{t-1}\right) \doteq \delta(y_t, \hat{y}_t)$

$$(**) = \int f\left(y_{t+1}|y(t), \hat{\Theta}_{t-1}\right) \delta(y_t, \hat{y}_t)\, dy_t$$

$$= f\left(y_{t+1}|\hat{y}_t, y(t-1), \hat{\Theta}_{t-1}\right)$$

Point prediction

$$\hat{y}_{t+1} = E[y_{t+1}|y(t-1)] = \int y_{t+1} f(y_{t+1}|y(t-1))\, dy_{t+1}$$

$\cdots$ expectation conditioned by $y(t-1)$.

Prediction with regression model

Point prediction - repetitive substitution of model.

Example for model

$$y_t = ay_{t-1} + bu_t + e_t$$

Prediction

$$\begin{aligned}
\hat{y}_t &= ay_{t-1} + bu_t \\
\hat{y}_{t+1} &= a\hat{y}_t + bu_{t+1} \\
\hat{y}_{t+2} &= a\hat{y}_{t+1} + bu_{t+2} \\
& etc.
\end{aligned}$$

Programs: **T31preCont.sce; T32preCont_Adapt.sce; T32preCont_Adapt2.sce;** (page 104)
T32preCont_Adapt3.sce (with the data on web)

Full prediction for normal model

$$
\begin{aligned}
y_t &= a y_{t-1} + b u_t + e_t \\
y_{t+1} &= a y_t + b u_{t+1} + e_{t+1} = \\
&= a\left(a y_{t-1} + b u_t + e_t\right) + b u_{t+1} + e_{t+1} = \\
&= a^2 y_{t-1} + a b u_t + b u_{t+1} + a e_t + e_{t+1} \\
y_{t+2} &= a y_{t+1} + b u_{t+2} + e_{t+2} = \\
&= a^3 y_{t-1} + a^2 b u_t + a b u_{t+1} + b u_{t+2} + a^2 e_t + a e_{t+1} + e_{t+2}
\end{aligned}
$$

and predictive pdf is $N_{y_{t+2}}\left(\hat{\mu}, \hat{r}\right)$ where

$\hat{\mu} = E\left[y_{t+2} | y\left(t-1\right)\right] = a^3 y_{t-1} + a^2 b u_t + a b u_{t+1} + b u_{t+2}$

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

Prediction with discrete model

Predictive pdf is a row of the model matrix. Point prediction is generated from the predictive pdf.

Example: Model $f(y_t|u_t, y_{t-1})$; $y_t \in \{1,2,3\}$, $u_t \in \{1,2\}$

u_t, y_{t-1}	$y_t = 1$	$y_t = 2$	$y_t = 3$
1, 1	0.2	0.5	0.3
1, 2	0.1	0.3	0.6
1, 3	0.7	0.2	0.1
2, 1	0.3	0.3	0.4
2, 2	0.5	0.2	0.3
2, 3	0.6	0.1	0.3

For measured $u_t = 1$ and $y_{t-1} = 3$ the predictive pdf is

$f(y_t|u_t = 1, y_{t-1} = 3) \quad \rightarrow$

y_t	1	2	3
$f(y_t)$	0.7	0.2	0.1

Generation a prediction with discrete model

It is generated as a value from categorical distribution with the predictive pdf. The generation in Scilab can be done in the following way:

– model matrix

$$\Theta = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.6 \\ & \cdots & \\ 0.6 & 0.1 & 0.3 \end{bmatrix}$$

– find row r corresponding to u_t, y_{t-1} (u_t/y_{t-1} have n_u/n_y values)

$$r = n_y * (u_t - 1) + y_{t-1}$$

– generate from this row

$$y_t = (\text{sum}\,(\text{rand}(1,1,\text{'u'}) > \text{cumsum}\,(\Theta\,(\text{r},:))) + 1$$

Programs: **T33preCat_Off.sce; T34preCat_OffEst.sce; T35preCat_OnEst.sce** (page 114 and further)

Filtration

State-space model

– state model (state prediction)

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

– output model (state filtration)

$$y_t = Ax_t + Bu_t + v_t$$

M, N, A, B are known matrices,

w_t, v_t are noises with zero expectations and known covariances R_w, R_v

Filtration

State evolution: prediction $\rightarrow$ filtration

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

Prediction

$$f\left(x_t|d\left(t-1\right)\right) = \int_{x^*_{t-1}} f\left(x_t|x_{t-1},u_{t-1}\right) f\left(x_{t-1}|d\left(t-1\right)\right) dx_{t-1}$$

Filtration

$$f\left(\underbrace{x_t}_{\Theta}|d\left(t\right)\right) \propto \underbrace{f\left(y_t|x_t,u_t\right)}_{\text{model}} f\left(\underbrace{x_t}_{\Theta}|d\left(t-1\right)\right)$$

Kalman filter

For normal model and initial conditions we get Kalman filter

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

xt - state estimate (expectation)

Rx - state covariance matrix

yp - output prediction

yt, ut - output, input

M, N, F, A, B, G - state model parameters (F,G - constants)

Rw, Rv - model noise covariances

Program: **T46statEst_KF.sce; T47statEst_Noise.sce** (page 122 and further)

Nonlinear state estimation

Model

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

Model linearization (Taylor expansion)

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

where $\hat{x}$ is the last point estimate.

Result

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

where

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

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

Control

Control

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

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

Criterion can be minimized sequentially from the end. The recursion (Bellman equations) are

$\varphi_{N+1}^* = 0$

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

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

$$\varphi_t^* = \min_{u_t} \varphi_t \quad \text{minimization}$$

$$u_t^* = \arg\min \varphi_t \text{ control}$$

end

Control for regression model

It is performed for state form of the model.

$R_{N+1} = 0$

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

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

end

Here, the vectors S_t are computed and then they are use for control application (in time direction)

for $t = 1, 2, \cdots, N,\ u_t = u_t = -S_t x_{t-1};\ y_t = \text{gener}(u_t);$ **end**

Program: **T53ctrlX.sce; T54ctrlXEst.sce** (page 128 and further)

Remarks

1. If in criterion $(y_t - s_t)^2$ is used the output follows the setpoint s_t
2. If $J_t = y_t^2 + \lambda\,(u_t - u_{t-1})^2$ is used, steady-state deviation is avoided.
3. If the model parameters are not known, we must use sub-optimal control with receding horizon:
    (a) for existing parameter estimated design the control and use only the first step,
    (b) apply the computed control;
    (c) measure new output;
    (d) with new data recompute parameter estimates
    (e) go to (a).

Control with discrete model

It is performed exactly in the same way as continuous with the discrete model. However, the operations with tables are somewhat unusual. You can look at them into the text.

Program: **T52ctrlDisc.sce** (page 135)