Přednáška 5 – Odhad normálního regresního modelu

Bayesovo pravidlo: $\underbrace{f(\Theta|d(t))}_{\text{aposteriorní GiW}} \propto \underbrace{f(y_t|\psi_t,\Theta)}_{\text{model}}\ \underbrace{f(\Theta|d(t-1))}_{\text{apriorní GiW}}$

Algoritmus odhadu parametrů normálního regresního modelu:

1. Pro čas $t = 0$ nastavíme počáteční statistiky V_0, κ_0
2. Pro čas $t = 1, 2, \ldots$
   1. Měříme data $d_t = \{y_t, u_t\}$
   2. Datová matice $D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t \ \ \psi_t']$
   3. Update statistik $V_t = V_{t-1} + D_t, \quad \kappa_t = \kappa_{t-1} + 1$
   4. Jdeme na krok 2.1
3. Rozklad informační matice $V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix} = \begin{bmatrix} \cdot & \overline{\qquad\qquad} \\ | & \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \end{bmatrix} \end{bmatrix}$
4. Výpočet bodových odhadů parametrů $\hat{\theta}_t = V_\psi^{-1} V_{y\psi}$,

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

Příklad

Systém: řízené vozidlo

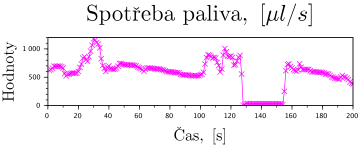
Reálná data:

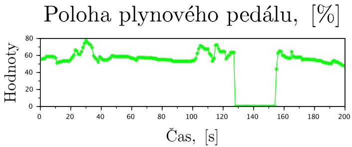
y_t – spotřeba paliva [μl/s]

u_t – poloha plynového pedálu [%]

čas t – sekundy (cca 4.5h)

y_t	622	635	661	695	692	...
u_t	55.6	56.4	56.4	58.8	58.8	...





Model 1.řádu:

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

$$\underbrace{\psi_t = [u_t \quad y_{t-1} \quad u_{t-1} \quad 1]'}_{\text{data}}$$

$$\underbrace{\theta = [b_0 \quad a_1 \quad b_1 \quad k]', \quad r}_{\text{neznámé parametry}}$$

Algoritmus:

1. $t = 0$, V_0 – dimenze informační matice

   V_0 – symetrická matice $m \times m$

   $m = dim(\psi_t) + 1 = 5$

   $dim(V_0) = 5 \times 5$

   zeros$(5, 5)$, $\quad 0.001 * \text{eye}(5, 5)$

   $\kappa_0 = 0$

Algoritmus – pokračování, kroky 2.1–2.4

2. $t = 1, 2, \ldots$
   1. Měříme data: $y_1 = 635$, $u_1 = 56.4$
   2. Datová matice: $D_1 = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t \ \ \psi'_t] = \begin{bmatrix} y_1 \\ u_1 \\ y_0 \\ u_0 \\ 1 \end{bmatrix} [y_1 \ u_1 \ y_0 \ u_0 \ 1]$

$$= \begin{bmatrix} 635 \\ 56.4 \\ 622 \\ 55.6 \\ 1 \end{bmatrix} [635 \ \ 56.4 \ \ 622 \ \ 55.6 \ \ 1]$$ – symetrická matice (5×5)

   3. Update statistik
   4. Jdeme na krok 2.1

$$V_1 = V_0 + D_1, \quad \kappa_1 = 0 + 1$$
$$\ldots\ldots\ldots$$
$$V_2 = V_1 + D_2, \quad \kappa_2 = 1 + 1$$
$$\ldots\ldots\ldots$$
$$V_3 = V_2 + D_3, \quad \kappa_3 = 2 + 1$$
$$\ldots\ldots\ldots$$
$$V_{100} = V_{99} + D_{100}, \quad \kappa_{100} = 99 + 1$$
$$\ldots\ldots\ldots$$

Algoritmus – pokračování, kroky 3–4

3. Rozklad informační matice

$$V_t = \begin{bmatrix} 46886009 & 3915576.2 & 46850676 & 3914765.2 & 66751 \\ 3915576.2 & 331361.79 & 3914315.9 & 331256.58 & 5704.7 \\ 46850676 & 3914315.9 & 46976957 & 3920783.4 & 66829 \\ 3914765.2 & 331256.58 & 3920783.4 & 331537.15 & 5706.3 \\ 66751 & 5704.7 & 66829 & 5706.3 & 99 \end{bmatrix}$$

$$= \begin{bmatrix} V_y & V'_{y\psi} \\ V_{y\psi} & V_\psi \end{bmatrix} = \begin{bmatrix} \cdot & \overline{} \\ | & \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \end{bmatrix} \end{bmatrix}, \quad \kappa_t = 100$$

$$f(\Theta | d(t)) \sim GiW(V_t, k_t)$$

4. Bodové odhady parametrů

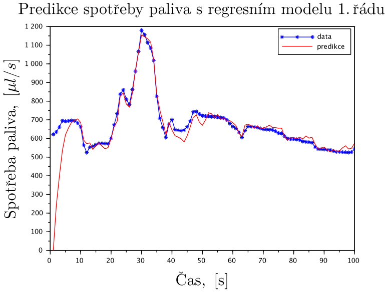
$$\hat{\theta}_t = V_\psi{}^{-1} V_{y\psi} = \begin{bmatrix} \hat{b}_0 \\ \hat{a}_1 \\ \hat{b}_1 \\ \hat{k} \end{bmatrix} = \begin{bmatrix} 18.774318 \\ 0.6098404 \\ -7.7999370 \\ -369.66761 \end{bmatrix}$$

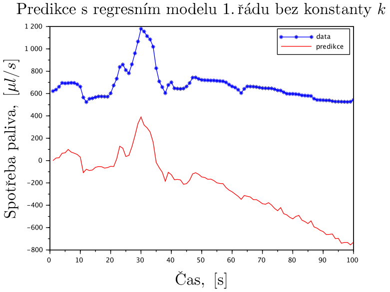
$$\hat{r}_t = \frac{V_y - V'_{y\psi} V_\psi{}^{-1} V_{y\psi}}{\kappa_t} = 130.377$$

Využití bodových odhadů

Dosadíme do modelu:

$$y_t = 18.774318u_t + 0.6098404y_{t-1} - 7.7999370u_{t-1} - 369.66761 + e_t$$





Poznámky

- Vliv koeficientu k
- Řád modelu pro odhad
- Nastavení počáteční statistiky V_0 z apriorních (expertních) dat
- Odhad z průběžně měřených dat = online odhad (real time)

Program

```
y=spotreba; u=plyn;
nd=100; % počet dat
m=4+1 % rozměr statistiky V
V=zeros(m,m); k=0;
for t=2:nd
% rozšířený regresní vektor
Ps=[y(t) u(t) y(t-1) u(t-1) 1];
V=V+Ps'*Ps% update statistik
k=k+1
end
% rozklad informační matice
Vy=V(1,1) % část Vy
Vyps=V(2:end,1) % část Vyps
Vps=V(2:end,2:end) % část Vps
% bodové odhady
theta_odhad=inv(Vps)*Vyps
r_odhad=(Vy-Vyps'*inv(Vps)*Vyps)/k
```

```
% predikce jako simulace s
odhadem parametrů
% počáteční podmínky
yp=zeros(1,nd);
for t=2:nd
% regresní vektor s predikcí
ps=[u(t) yp(t-1) u(t-1) 1];
% generování výstupu
yp(t)=ps*theta_odhad+...
sqrt(r_odhad)*randn(1,1);
end
```

Offline odhad – metoda nejmenších čtverců

Offline odhad ze všech dat najednou, není z ~~průběžně měřených~~ dat

Metoda nejmenších čtverců – stejný výsledek, jako s Bayesovským odhadováním

Postup:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{nd} \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} \psi_1' \\ \psi_2' \\ \dots \\ \psi_{nd}' \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} b_0 \\ a_1 \\ b_1 \\ \dots \\ k \end{bmatrix}}_{\theta} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_{nd} \end{bmatrix}$$

$\psi_t' = [u_t \; y_{t-1} \; u_{t-1} \dots 1]$ – regresní vektor

Bodové odhady parametrů:

$$\hat{\theta} = (X'X)^{-1}X'Y, \quad \hat{r} = \frac{Y'(Y - X\hat{\theta})}{nd}$$

Není index t

Jenom normální rozdělení