Přednáška 7 – Předpověď s regresním modelem

Předpověď modelované veličiny – obecně

Máme: model $f(y_t|\psi_t, \Theta)$
naměřená data $d(t)$

Potřebujeme: $\hat{y}_{t+k}$

k – počet kroků predikce

Rozlišujeme úlohu předpovědi:

počet kroků:
- jednokroková ($\hat{y}_{t+1}$), $k = 1$
- vícekroková ($\hat{y}_{t+k}$), $k > 1$

znalost Θ :
- známé parametry – Θ
- neznámé parametry – $\hat{\Theta}_t$

forma:
- úplný Bayesovský popis – celá prediktivní hp
- částečný popis (bodová predikce) – střední hodnota

model:
- spojitý
- diskrétní

Jednokroková předpověď s regresním modelem $\hat{y}_{t+1}$, $k = 1$

Bodový odhad výstupu = 0-kroková predikce = 1-kroková predikce

$d(t-1) \Rightarrow \hat{\Theta}_{t-1} \Rightarrow$ model $\Rightarrow \hat{y}_t$ $\qquad$ $d(t) \Rightarrow \hat{\Theta}_t \Rightarrow$ model $\Rightarrow \hat{y}_{t+1}$

Jednokroková předpověď se známými parametry – Θ

hp $f(y_{t+1}|y(t))$ nebo střední hodnota $\hat{y}_{t+1}$

Jednokroková předpověď s neznámými parametry – $\hat{\Theta}_t$

$$f(y_{t+1}|y(t)) = \int_{\Theta^*} \underbrace{f(y_{t+1}, \Theta|y(t))}_{\text{sdružená hp}} d\Theta = \int_{\Theta^*} \underbrace{f(y_{t+1}|\Theta, \overbrace{y(t)}^{\psi_t})}_{\text{model}} \underbrace{f(\Theta|y(t))}_{\text{GiW}} d\Theta$$

Nahradíme: GiW $= \delta(\Theta - \hat{\Theta}_t)$ $\qquad$ Diracův impulz $\qquad$ $\delta(0) = 1$, jinde $\delta = 0$

Předpověď: $f(y_{t+1}|\hat{\Theta}_t, y(t))$

Příklad: Model: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + e_t$

Předpověď: $\hat{y}_{t+1} = \underbrace{\hat{a}_1 y_t + \hat{a}_2 y_{t-1}}_{\text{střední hodnota}}$

Závěr:

Dosazujeme poslední data + odhady do modelu

Vícekroková předpověď s regresním modelem $\hat{y}_{t+k}$, $k > 1$

Pro $k = 2$: $$f(y_{t+2}|y(t)) = \int_{\Theta^*}\int_{{y^*}_{t+1}} \underbrace{f(y_{t+2}, y_{t+1}, \Theta|y(t))}_{\text{sdružená hp}} dy_{t+1} d\Theta =$$

$$= \int_{\Theta^*}\int_{{y^*}_{t+1}} \underbrace{f(y_{t+2}|\overbrace{y_{t+1}, y(t)}^{\tilde{\psi}_{t+1}}, \Theta)}_{\text{model}} \underbrace{f(y_{t+1}|\overbrace{y(t)}^{\psi_t}, \Theta)}_{\text{model}} \underbrace{f(\Theta|y(t))}_{\text{GIW}} dy_{t+1} d\Theta$$

$$= \int_{\Theta^*} f(y_{t+2}|\tilde{\psi}_{t+1}, \Theta)\delta(\Theta - \hat{\Theta}_t) d\Theta = f(y_{t+2}|\tilde{\psi}_{t+1}, \hat{\Theta}_t)$$

Zobecnění pro $k > 1$:

$f(y_{t+k}|y(t)) = \ldots k$ integrálů $\ldots = f(y_{t+k}|\tilde{\psi}_{t+k-1}, \hat{\Theta}_t)$

Příklad: Model $y_t = a_1 y_{t-1} + b_0 u_t + e_t$

1.krok $\hat{y}_t = a_1 y_{t-1} + b_0 u_t$

2.krok $\hat{y}_{t+1} = a_1 \hat{y}_t + b_0 u_{t+1}$

3.krok $\hat{y}_{t+2} = \underbrace{a_1 \hat{y}_{t+1} + b_0 u_{t+2}}_{\text{střední hodnota}}$

Závěr:

Dosazujeme předchozí předpovědi + odhady do modelu

Vícekroková $\neq k\cdot$jednokroková

Příklad – vícekroková predikce (částečný popis)

Systém: čerpací stanice v Praze

y_t – cena nafty v Kč

t – týdny v období 2009 – 2022

Model:

$\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + k + e_t$

Predikce

1.krok:

$\hat{y}_t = \hat{a}_1 y_{t-1} + \hat{a}_2 y_{t-2} + \hat{a}_3 y_{t-3} + \hat{k} = \mathbf{43.207}$

2.krok:

$\hat{y}_{t+1} = \hat{a}_1 \hat{y}_t + \hat{a}_2 y_{t-1} + \hat{a}_3 y_{t-2} + \hat{k} = \mathbf{43.095}$

3.krok:

$\hat{y}_{t+2} = \hat{a}_1 \hat{y}_{t+1} + \hat{a}_2 \hat{y}_t + \hat{a}_3 y_{t-1} + \hat{k} = \mathbf{42.845}$

4.krok:

$\hat{y}_{t+3} = \hat{a}_1 \hat{y}_{t+2} + \hat{a}_2 \hat{y}_{t+1} + \hat{a}_3 \hat{y}_t + \hat{k} = \mathbf{42.661}$

Odhady z dat:

$\hat{a}_1$=0.64,

$\hat{a}_2$=0.22,

$\hat{a}_3$=0.12,

$\hat{k}$=1.28

Ceny za poslední týdny:

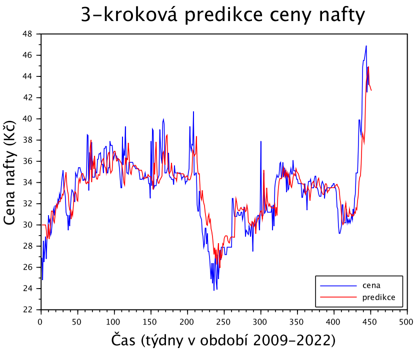
y_{t-1}=43.2 Kč

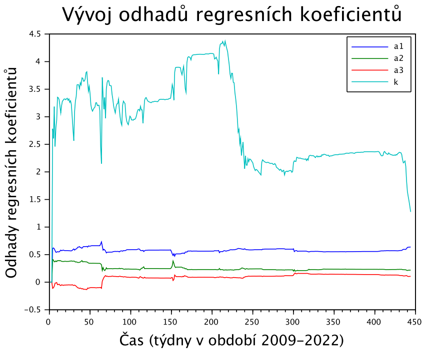
y_{t-2}=44.9 Kč

y_{t-3}=42.5 Kč

Chyba predikce (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T} (\hat{y}_t - y_t)^2}{T}}$$


3-kroková predikce ceny nafty
Cena nafty (Kč)
Čas (týdny v období 2009-2022)
cena
predikce


Vývoj odhadů regresních koeficientů
Odhady regresních koeficientů
Čas (týdny v období 2009-2022)
a1
a2
a3
k

Příklad – vícekroková predikce (úplný popis)

Systém: úsek silnice

y_t – intenzita dopravního proudu

t – minuta

$y_{t-1} = 8, y_{t-2} = 5, y_{t-3} = 7$

$\hat{a}_1 = 0.4, \hat{a}_2 = 0.35, \hat{a}_3 = 0.3, \hat{r}_t = 0.2$

Za předpokladu normality

Střední hodnota a rozptyl

Dosazujeme rovnici + šum

Model = 1.krok: $\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + e_t$

2.krok $\hat{y}_{t+1} = a_1 \underbrace{(\ldots)}_{\hat{y}_t} + a_2 y_{t-1} + a_3 y_{t-2} + e_{t+1}$

3.krok $\hat{y}_{t+2} = a_1 \underbrace{(\ldots)}_{\hat{y}_{t+1}} + a_2 \underbrace{(\ldots)}_{\hat{y}_t} + a_3 y_{t-1} + e_{t+2} =$

$$= \underbrace{\left(a_1^3 + a_1^2 a_3 + 2a_1 a_2\right) y_{t-1} + \left(a_1^2 a_2 + a_2^2 + a_1 a_3\right) y_{t-2} + a_2 a_3 y_{t-3}}_{\text{střední hodnota}=5.3635}$$

$+a_1^2 e_t + a_1 e_{t+1} + a_2 e_t + e_{t+2}$

var$[a_1^2 e_t + a_1 e_{t+1} + a_2 e_t + e_{t+2}] = (a_1^4 + a_1^2 + a_2^2 + 1) r = 0.262$ – rozptyl

Předpověď: $f(y_{t+2} | y(t-1)) \sim N(5.3635, 0.262)$

$\sigma = 0.51$, $3\sigma = 1.53$ Hodnoty predikce: od 3.8335 do 6.8935

Program:

```
% nahrajeme data
nd=size(y,2); % počet dat
np=3; % počet kroků predikce
% odhad
V=0.001*eye(5,5); % počáteční statistika V
k=0; % počáteční počítadlo
for t=4:nd
  Ps=[y(t) y(t-1) y(t-2) y(t-3) 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
```

```
% vícekroková predikce
for t=4:(nd-np)
  yy(1:(t-1))=y(1:(t-1)); % stará data
  for i=0:np
    % regresní vektor
    ps=[yy(t+i-1) yy(t+i-2) yy(t+i-3) 1];
    % generujeme střední hodnotu predikce
    yy(t+i)=ps*theta_odhad;
  end
  yp(t+np)=yy(t+np); % vícekroková predikce
end
% Chyba predikce
RMSE=sqrt(mean((y' - yp).^2));
```