Přednáška 3 – Diskrétní a logistický model

Diskrétní veličiny $y_t \in \{1, 2, \ldots, N\}$, $u_t \in \{1, 2, \ldots, K\}$

Diskrétní model

Podmíněná pf
$f(y_t|\psi_t, \Theta)$

Regresní vektor (data) $\psi_t = [u_t \;\; y_{t-1} \;\; u_{t-1} \;\; \ldots]'$
Neznámé parametry (tabulka pravděpodobností) Θ
Řád modelu – zpoždění výstupu
Závislost $y_t|\psi_t$ – konečný počet kombinací

Příklad:

Pozorovaný systém – úsek silnice

$y_t \in \{1, 2\}$ – stupeň dopravy
(1 = volný provoz, 2 = kolony)

$y_t|\psi_t$ – 4 kombinace:

1|1, 1|2, 2|1, 2|2

Model $f(y_t|y_{t-1}, \Theta)$, $\psi_t = y_{t-1}$

ψ_t \ y_t	1	2		
$y_{t-1} = 1$	$\Theta_{1\|1} = 0.7$	$\Theta_{2\|1} = 0.3$	$\Rightarrow$ 1	Alternativní rozdělení
$y_{t-1} = 2$	$\Theta_{1\|2} = 0.1$	$\Theta_{2\|2} = 0.9$	$\Rightarrow$ 1	

Dimenze diskrétního modelu

Vliv:

- počet možných realizací y_t
- počet veličin v ψ_t a jejich realizací

Příklad:

Model $f(y_t|y_{t-1}, v_t, \Theta)$, $\quad \psi_t = [y_{t-1} \quad v_t]'$
$y_t \in \{1, 2\}$ – stupeň dopravy (volný provoz, kolony)
$v_t \in \{1, 2\}$ – den v týdnu (všední den, víkend)

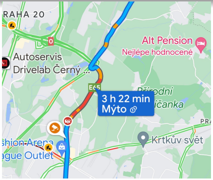
$\psi_t \backslash y_t$	1	2	
$y_{t-1} = 1, v_t = 1$	$\Theta_{1\|11}$	$\Theta_{2\|11}$	$\Rightarrow 1$
$y_{t-1} = 1, v_t = 2$	$\Theta_{1\|12}$	$\Theta_{2\|12}$	$\Rightarrow 1$
$y_{t-1} = 2, v_t = 1$	$\Theta_{1\|21}$	$\Theta_{2\|21}$	$\Rightarrow 1$
$y_{t-1} = 2, v_t = 2$	$\Theta_{1\|22}$	$\Theta_{2\|22}$	$\Rightarrow 1$

Příklad:

Model $f(y_t|y_{t-1}, v_t, \Theta)$, $\quad \psi_t = [y_{t-1} \quad v_t]'$

$y_t \in \{1, 2, 3\}$ – (plynulá jízda, menší kolony, větší kolony)

$v_t \in \{1, 2\}$ – den v týdnu (všední den, víkend)



$\psi_t \backslash y_t$	1	2	3	
$y_{t-1} = 1, v_t = 1$	$\Theta_{1\|11}$	$\Theta_{2\|11}$	$\Theta_{3\|11}$	$\Rightarrow 1$
$y_{t-1} = 1, v_t = 2$	$\Theta_{1\|12}$	$\Theta_{2\|12}$	$\Theta_{3\|12}$	$\Rightarrow 1$
$y_{t-1} = 2, v_t = 1$	$\Theta_{1\|21}$	$\Theta_{2\|21}$	$\Theta_{3\|21}$	$\Rightarrow 1$
$y_{t-1} = 2, v_t = 2$	$\Theta_{1\|22}$	$\Theta_{2\|22}$	$\Theta_{3\|22}$	$\Rightarrow 1$
$y_{t-1} = 3, v_t = 1$	$\Theta_{1\|31}$	$\Theta_{2\|31}$	$\Theta_{3\|31}$	$\Rightarrow 1$
$y_{t-1} = 3, v_t = 2$	$\Theta_{1\|32}$	$\Theta_{2\|32}$	$\Theta_{3\|32}$	$\Rightarrow 1$

- Pro $y_t \in \{1, 2, 3, 4, 5\}$, $v_t \in \{1, 2\}$ – dimenze 10×5
- Pro $y_t \in \{1, 2, 3, 4, 5\}$, $v_t \in \{1, 2, \dots, 7\}$?
- Pečlivý výběr veličin a hodnot – přínos vs. ztráta informace
- Kategorické rozdělení

Neurčitost v diskrétním modelu je dána parametrem Θ

$$f(y_t|y_{t-1}, u_t, \Theta)$$

Neurčitý diskrétní model

ψ_t \ y_t	1	2
$y_{t-1} = 1, u_t = 1$	0.45	0.55
$y_{t-1} = 1, u_t = 2$	0.5	0.5
$y_{t-1} = 2, u_t = 1$	0.4	0.6
$y_{t-1} = 2, u_t = 2$	0.7	0.3

Deterministický model – $\max\{y_{t-1}, u_t\}$

ψ_t \ y_t	1	2
$y_{t-1} = 1, u_t = 1$	1	0
$y_{t-1} = 1, u_t = 2$	0	1
$y_{t-1} = 2, u_t = 1$	0	1
$y_{t-1} = 2, u_t = 2$	0	1

Deterministický model – $y_t \neq y_{t-1}$

ψ_t \ y_t	1	2
$y_{t-1} = 1, u_t = 1$	0	1
$y_{t-1} = 1, u_t = 2$	0	1
$y_{t-1} = 2, u_t = 1$	1	0
$y_{t-1} = 2, u_t = 2$	1	0

Generování hodnot

Alternativní rozdělení $y_t \in \{1, 2\}$

Stačí jedna pravděpodobnost:

```
% pravděpodobnost hodnoty 1
  p=0.7;
% generuje skalár 1 nebo 2
  y=(rand(1,1)>p)+1
% nebo počet dat pro více dat
  nd=50;
% vektor nd hodnot 1 nebo 2
  y=(rand(1,nd)>p)+1
```

$$\underbrace{\underbrace{(\underbrace{\mathrm{rand}(1,1)}_{\text{hodnota mezi 0 a 1}} > 0.7)}_{\text{T nebo F}} + 1}_{\text{1 nebo 2 s pravděp. 0.7 pro 1}}$$

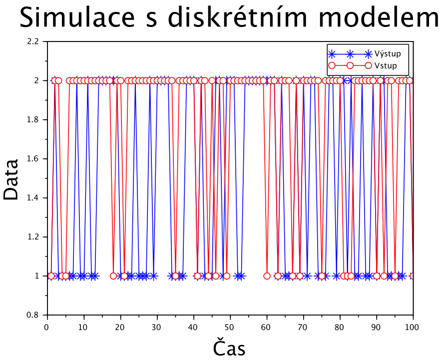
Kategorické rozdělení $y_t \in \{1, 2, 3\}$

Vektor pravděpodobností:

```
% pravděpodobnosti 3 hodnot
  p=[0.3  0.5  0.2];
% kumulativní součet [0.3  0.8 1]
  pp=cumsum(p);
% generuje skalár 1, 2 nebo 3
  y=sum(rand(1,1)>pp)+1
% nebo počet dat pro více dat
  nd=50;
for t=1:nd
% generuje vektor nd hodnot 1,2,3
  y(t)=sum(rand(1,1)>pp)+1
end
```

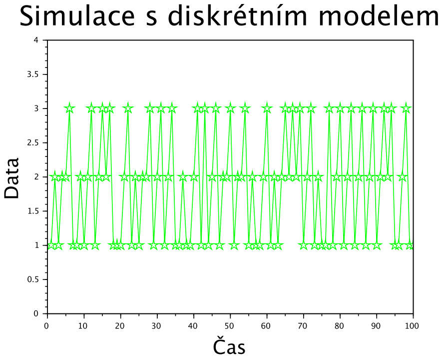
Simulace s alternativním modelem $f(y_t|y_{t-1}, u_t, \Theta)$

```
clear,clc,close
nd=100;  % počet dat
% tabulka pravděpodobností
th=[0.48 0.52;
    0.49 0.51;
    0.53 0.47;
    0.51 0.49];
y(1)=1;  % počáteční podmínky
% řídící vstup
u=(rand(1,nd)>0.2)+1;
for t=2:nd
  % číslo řádku
   j=2*(y(t-1)-1)+u(t);
  % pravděp. z 1.sloupce
  py=th(j,1);
   % generování výstupu
  y(t)=(rand(1,1)>py)+1
end
```



Simulace s kategorickým modelem $f(y_t|y_{t-1},\Theta)$

```
clear,clc,close
nd=100; % počet dat
% tabulka pravděpodobností
th=[0.2 0.7 0.1;
    0.2 0.1 0.7;
    0.7 0.2 0.1];
y(1)=1; % počáteční podmínky
for t=2:nd
  % číslo řádku
  j=y(t-1);
  % kumulativní součet
  pp=cumsum(th(j,:));
  % generování výstupu
  y(t)=sum(rand(1,1)>pp)+1;
end
```



Využití diskrétního modelu

Klasifikace diskrétních dat + vývoj dynamické veličiny

Logistický model (logistická regrese)

Modeluje diskrétní výstup $y_t \in \{0,1\}$
v závislosti na spojitých nebo spojitých+diskrétních veličinách

Obecný tvar pro $y_t \in \{0,1\}$

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

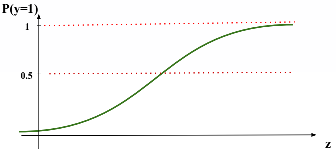
$$\swarrow \quad \searrow$$

$$\underbrace{\frac{\exp\{z_t\}}{1+\exp\{z_t\}}}_{P(y_t=1)} \qquad \underbrace{\frac{1}{1+\exp\{z_t\}}}_{P(y_t=0)}$$

$$\underbrace{z_t}_{\text{spojitá}} = \underbrace{\psi_t'\theta + e_t}_{\text{lineární regrese}}$$

Regresní vektor (data)
$\psi_t = [v_{1;t} \;\; v_{2;t} \;\; v_{3;t} \;\; \ldots \;\; 1]'$
Vektor regresních koeficientů
(neznámé parametry)
$\theta = [b_1 \;\; b_2 \;\; b_3 \;\; \ldots \;\; k]'$

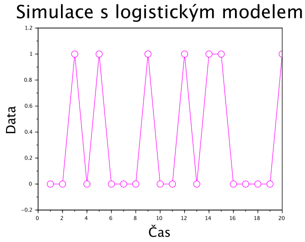
Příklad: Systém – úsek silnice
$y_t \in \{0,1\}$ – nehoda (NE, ANO)
$v_{1;t}$ – rychlost 1.auta (spojitá)
$v_{2;t}$ – rychlost 2.auta (spojitá)
$v_{3;t}$ – povrch vozovky (diskrétní)
$v_{4;t}$ – zkušenost řidiče (diskrétní)



$z_t \in \mathbb{R} \Rightarrow P \in (0,1)$

Simulace s logistickým modelem

```
clear,clc,close
nd=20; % počet dat
% regresní koeficienty a rozptyl
th=[1 3 -1 -4 0.5]; r=1;
% externí vstupy
v1=randn(1,nd); v2=randn(1,nd);
v3=randn(1,nd)
v4=(rand(1,nd)>0.5)+1
for t=1:nd
  e=sqrt(r)*rand(1,1,'n'); % šum
  % regresní vektor
  ps=[1 v1(t) v2(t) v3(t) v4(t)];
  z=th*ps'+e; % regrese z
  % pravděpodobnost 1
  p(t)=exp(z)/(1+exp(z));
  % generujeme výstup
  y(t)=double(rand(1,1)<p(t))
end
```



Využití

Klasifikace – statický model

Inverzní forma logistické regrese

Funkce logit

$$\text{logit}\,(p) = \ln \frac{p}{1-p}$$

Dosadíme

$$p = \frac{\exp\{z_t\}}{1+\exp\{z_t\}}$$

$$\ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{\frac{\exp(z_t)}{1+\exp(z_t)}}{1-\frac{\exp(z_t)}{1+\exp(z_t)}}\right) = \ln\left(\frac{\exp(z_t)}{1+\exp(z_t)}\right) - \ln\left(1-\frac{\exp(z_t)}{1+\exp(z_t)}\right)$$

$$= \ln(\exp(z_t)) - \ln(1+\exp(z_t)) - \ln\left(\frac{1+\exp(z_t)-\exp(z_t)}{1+\exp(z_t)}\right) =$$

$$= z_t - \ln(1+\exp(z_t)) - \ln(1) + \ln(1+\exp(z_t)) = z_t$$

Multinomiální logistická regrese – $y_t \in \{0, 1, 2, \cdots, N\}$

$$\ln \frac{p_1}{p_0} = z_{t(1)}, \quad \ln \frac{p_2}{p_0} = z_{t(2)}, \quad \ldots \quad , \ln \frac{p_N}{p_0} = z_{t(N)}$$