1 Odhad spojitého modelu

Model je matematickým popisem vybraných veličin sledovaného procesu. Tyto veličiny popisujeme stochasticky (pomocí hustot pravděpodobnosti) v závislosti na jiných vybraných veličinách většinou jako lineární vazby pomocí diferenčních rovnic.

Při návrhu modelu řešíme dva kroky:

1. návrh struktury modelu, tj. výběr veličin a počtu kroků jejich zpoždění, na kterých modelovaná veličina závisí,
2. určení parametrů modelu, které vyjadřují konkrétní závislosti ve sledovaném procesu.

Zde se budeme zabývat druhým bodem - odhadem parametrů modelu.

1.1 Bayesovské odhadování

Parametry z hlediska bayesovství

V klasické statistice se odhadované parametry, ale i jiné odhadované veličiny, považují za náhodné veličiny. Jejich popisem je hustota pravděpodobnosti. Jestliže je pro konstrukci této hp použita jen předběžná (expertní) znalost, hovoříme o apriorním popisu. Jestliže jsou dále využita i měřená data, dostáváme aposteriorní popis.

V případě popisu parametrů se jedná o hp $f(\Theta)$, která se nazývá **apriorní hp** a která odráží prvotní znalosti o parametrech. V průběhu odhadování se měří data d_t v časech $t = 1, 2, \cdots, T$, kde T je horizont intervalu odhadování. Informace z měřených dat se postupně využívá pro zpřesnění popisu parametrů a původní apriorní hp se vyvíjí na **aposteriorní hp**

$$f(\Theta) \overset{d_1}{\rightarrow} f(\Theta|d(1)) \overset{d_2}{\rightarrow} f(\Theta|d(2)) \overset{d_3}{\rightarrow} \cdots \overset{d_T}{\rightarrow} f(\Theta|d(T))$$

Střední hodnota aposteriorní hp vypovídá o bodových odhadech parametrů (viz Přílohy **??**), kovarianční matice o nepřesnosti odhadů. Přepočty hp popisující parametry se provádí na základě **Bayesova vztahu**.

Bayesův vztah

Vývoj hp parametrů, tj. postupné upřesňování hp parametrů podle informace přicházející z měřených dat $d_1, d_2, \cdots, d_t$, se provádí podle Bayesova vzorce (viz Přílohy **??**)

$$f(\Theta|d(t)) \propto f(y_t|\psi_t, \Theta) f(\Theta|d(t-1)), \tag{1.1}$$

který se počítá pro $t = 1, 2, \cdots, T$ a startuje s tzv. apriorní hp $f(\Theta|d(0)) = f(\Theta)$, která odráží apriorní, expertní znalost.

Vztah (1.1) je možno zapsat také rovnou pro koncový čas T

$$f(\Theta|d(T)) = \prod_{t=1}^{T} f(y_t|\psi_t, \Theta) f(\Theta|d(0)) = L_T(\Theta) f(\Theta|d(0)),$$

kde

$$L_T(\Theta) = \prod_{t=1}^{T} f(y_t|\psi_t, \Theta) \tag{1.2}$$

je věrohodnostní funkce (likelihood).

Z uvedeného je patrné, že bayesovský odhad je vlastně klasický odhad maximální věrohodnosti korigovaný apriorní hp.

Reprodukovatelné parametrické vyjádření aposteriorní hp

Vztah (1.1) je rekurze pro funkce. Ta je prakticky nerealizovatelná, proto je třeba jednotlivé hp vyjádřit v konkrétním tvaru, který závisí na konečném počtu číselných charakteristik a tuto funkcionální rekurzi převést na rekurzi algebraickou pro charakteristiky rozdělení.[1]

Navíc je třeba parametrické vyjádření volit tak, aby při postupném odhadu v čase nevznikaly nové a nové charakteristiky, tj, aby se formální tvar hp parametrů reprodukoval. Jinak se tato hp velmi rychle stane tak složitou, že ji prakticky není možno počítat v rozumném čase.

Například bude-li mít model normální rozdělení, zvolíme apriorní hp také jako normální. Násobení normálních hp vede opět na normální rozdělení, které je určené svou střední hodnotou a rozptylem. Ty je možno počítat přímo ze středních hodnot a rozptylů modelu soustavy a apriorní hp. Dostáváme tak rekurzi na číslech, nikoli na funkcích. Pokud má model tuto vlastnost, řekneme, že při odhadu dostáváme **reprodukující se aposteriorní hp**.

Obecně tento problém vyřešit nelze, ale pro případy, kterými se zde budeme zabývat, tj. pro regresní spojitý model s normálním rozdělením a pro diskrétní model s multinomiálním rozdělením, takové tzv. adjungované apriorní distribuce na parametrech existují ve formě inverzního Gauss-Wishartova a Dirichletova rozdělení. Budeme o nich podrobněji hovořit v následujících kapitolách.

Příklad [odhadování bez reprodukce aposteriorní hp]

V tomto příkladu budeme ilustrovat situaci, kdy odhadujeme s modelem, který nevede na reprodukující se aposteriorní hp.

[1] Např. přepočítat funkci $f(x)$ na $g(x)$ lze jenom tak, že ji přepočítáme $f(x) \to g(x)$ pro každý bod $x \in R$. Jestliže ale je $f(x) = \exp\{ax\}$ a $g(x) = exp\{bx\}$, pak stačí přepočítat $a \to b$. Cela funkce už je dána svým předpisem.

Uvažujme model $f(y|a) = ay^2 - 2ay + \frac{4a+3}{6}$ pro $y \in (0,\, 2)$ a parametrem $a \in \left(-\frac{3}{4},\, \frac{3}{2}\right)$. Protože známe rozsah parametru a a nemáme o něm žádnou apriorní informaci, budeme apriorní hp volit jako rovnoměrnou, tj. $f(a) = \frac{4}{9}$ na intervalu $a \in \left(-\frac{3}{4},\, \frac{3}{2}\right)$.

V prvním kroku odhadu naměříme y_1 a dostaneme

$$f(a|y_1) \propto f(y_1|a)\, f(a) = \left(ay_1^2 - 2ay_1 + \frac{4a+3}{6}\right)\frac{4}{9}.$$

V druhém kroku

$$f(a|y_1, y_2) \propto f(y_2|a)\, f(a|y_1) = \left(ay_2^2 - 2ay_2 + \frac{4a+3}{6}\right)\left(ay_1^2 - 2ay_1 + \frac{4a+3}{6}\right)\frac{4}{9}$$

a tak dále. Z toho je dobře vidět, že formální tvar aposteriorní hp je stále složitější, protože je v něm třeba si pamatovat stále složitější výrazy. Po prvním kroku pracujeme s y_1^2 a y_1. Po druhém už to bude $y_1^2 y_2^2$, $y_1^2 y_2$, $y_1 y_2^2$, $y_1 y_2$ y_1^2, y_2^2, y_1, y_2 a tak dále.

Příklad [Odhadování s reprodukcí aposteriorní hp]

Zde ukážeme odhad s modelem, který vede na reprodukující se aposteriorní hp.

Vezmeme model $f(y_t|a) = \exp\{-ay\}$, $y \geq 0$, $a > 0$. Jako apriorní hp uvažujme $f(a) = \exp\{-ay_0\}$. Potom po prvním kroku odhadu máme

$$f(a|y_1) \propto f(y_t|a)\, f(a) = \exp\{-ay_1\}\exp\{-ay_0\} = \exp\{-a(y_1 + y_0)\},$$

po druhém

$$f(a|y_1, y_2) \propto \exp\{-a(y_2 + y_1 + y_0)\}$$

a obecně

$$f(a|y(k)) \propto \exp\left\{-a\sum_{i=0}^{k} y_i\right\}.$$

Je zřejmé, že když označíme statistiku odhadu $S_k = \sum_{i=0}^{k} y_i$, pak přepočet statistiky v čase k je

$$S_k = S_{k-1} + y_k.$$

Formální tvar hp parametrů se nemění a vzorec pro přepočet statistik má stále stejný tvar.

Výsledek odhadování

Výsledkem procedury odhadu je aposteriorní hp

$$f(\Theta|d(t)),$$

která dává úplný stochastický popis parametrů - tedy výčet všech možných hodnot a jejich pravděpodobnosti výskytu. Pokud je to možné (většinou z důvodů spočitatelnosti), mělo by se tam, kde máme neznámé parametry, počítat s touto celou distribucí.

Není-li možné využít v dalších výpočtech celou aposteriorní hp nebo potřebujeme-li jako odhady čísla, můžeme přejít k bodovým odhadům

$$\hat{\Theta}_t = E[\Theta|d(t)].$$

Obě tyto varianty využití procedury odhadování lze dobře ilustrovat v dalším odstavci Odhad výstupu systému.

Odhad výstupu

Předpověď výstupu je popsána prediktivní hp $f\left(y_t|\psi_t, d\left(t-1\right)\right)$. Protože parametry Θ obecně neznáme, nesmí se vyskytovat v podmínce.[2]

Prediktivní hp dostaneme tak, že modelujeme obě neznámé veličiny y_t i Θ pomocí sdružené hp, kterou rozložíme podle řetězového pravidla a integrujeme přes Θ, abychom dostali popis jen pro výstup y_t

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

$$= \int_{\Theta^*} f\left(y_t|\psi_t, \Theta\right) f\left(\Theta|d\left(t-1\right)\right) d\Theta, \tag{1.3}$$

kde chybějící veličiny v podmínkách vypadly z důvodu nezávislosti. Předpověď výstupu v čase t (jako náhodné veličiny) je tedy dána modelem $f\left(y_t|\psi_t, \Theta\right)$ a vyjádřením neznámého parametru pomocí aposteriorní hp parametrů $f\left(\Theta|d\left(t-1\right)\right)$.

Poznámka

Dobře si všimněte, jak bayesovství zachází s neznámou veličinou (tady parametrem). Každého by napadlo: mám model a potřebuji do něho parametr. Udělám odhad a ten tam dosadím. To ale není optimální. Správné je použití úplné pravděpodobnosti tak jako v (1.3). Do modelu postupně dosazuji všechny možné hodnoty parametrů a počítám vážený průměr - dosazení a sčítání dělá integrál, váhy jsou dány aposteriorní hp. ◁

Bodový odhad výstupu

Bodový odhad parametru zkonstruovaný s pomocí aposteriorní hp je (viz Příloha **??**) podmíněná střední hodnota parametru

$$\hat{\Theta}_t = E\left[\Theta|d\left(t\right)\right] = \int_{\Theta^*} \Theta f\left(\Theta|d\left(t\right)\right) d\Theta. \tag{1.4}$$

Tento bodový odhad už není úplný, ale jen částečný popis parametru (např. střední hodnota nic neříká o rozptylu). Pokud je ale aposteriorní hp dostatečně „štíhlá“, a to ona po správném odhadu je,[3] je možno ji nahradit Diracovým impulzem $\delta\left(\Theta - \hat{\Theta}_t\right)$, kde $\delta\left(0\right)$ je jedna a jinde nula, tedy

$$f\left(\Theta|d\left(t-1\right)\right) \to \delta\left(\Theta - \hat{\Theta}_{t-1}\right). \tag{1.5}$$

[2] Jinak bychom tuto hp nemohli přímo použít - neměli bychom za Θ co dosadit. Parametry ale potřebujeme pro model. Proto musíme odhad Θ zabudovat do konstrukce prediktivni hp.

[3] Při odhadu získáváme informaci, tím klesá neurčitost a s ní i rozptyl odhadu.

Dosadíme za aposteriorní hp a dostaneme

$$f\left(y_t|\psi_t, d\left(t-1\right)\right) = \int_{\Theta^*} f\left(y_t|\psi_t, \Theta\right) \delta\left(\Theta - \hat{\Theta}_{t-1}\right) d\Theta = f\left(y_t|\psi_t, \hat{\Theta}_{t-1}\right). \tag{1.6}$$

Výsledek je právě ten, který bychom čekali. Jestliže máme model s neznámými parametry a spočteme bodové odhady parametrů, pak tyto bodové odhady použijeme místo neznámých parametrů. Nedostáváme optimální řešení, ale pro „štíhlou aposteriorní hp“ je to řešení často přijatelné.

1.2 Odhad parametrů regresního modelu

Odhad parametrů normálního regresního modelu na základě apriorní informace a dat $d\left(t\right)$ měřených průběžně na systému provedeme na základě Bayesova vzorce (1.1) s modelem (**??**) a vhodně zvolenou apriorní hp $f\left(\Theta|d\left(0\right)\right) = f\left(\Theta\right)$.

Model

Regresní model s regresním vektorem ψ_t, jemu odpovídajícím vektorem regresních koeficientů θ a normálním šumem s rozptylem r má tvar

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

Pro účely odhadu je výhodné tuto hp (výraz v exponentu) upravit do následujícího tvaru (viz Příloha **??**, rovnice (**??**))

$$f\left(y_t|\psi_t, \Theta\right) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\}, \tag{1.7}$$

kde $D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix}\left[y_t,\, \psi_t'\right]$ je tzv. datová matice.

Dále v Příloze **??** v rovnici (**??**) je uveden tvar aposteriorní hp, který odpovídá normálnímu regresnímu modelu (jedná se o tzv. konjugovanou aposteriorní hp k normálnímu regresnímu modelu, tj. takový popis parametrů, jehož tvar se při odhadu podle Bayesova vzorce reprodukuje)

$$f\left(\Theta|d\left(\tau\right)\right) \propto r^{-0.5\kappa_\tau} \exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right] V_\tau \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\}, \tag{1.8}$$

kde V_τ a κ_τ jsou statistiky odhadu. Pro časový okamžik t vzorec s $\tau = t$ udává aposteriorní hp, pro $\tau = t-1$ je to apriorní hp.

Rekurze pro statistiku

Dosazením do Bayesova vzorce (1.1) dostaneme

$$\underbrace{r^{-0.5\kappa_t}\exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right]V_t\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}}_{\text{aposteriorní hp}}\propto$$

$$\propto\underbrace{r^{-0.5}\exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right]D_t\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}}_{\text{model}}\underbrace{r^{-0.5\kappa_{t-1}}\exp\left\{-\frac{1}{2r}\left[-1\,\theta'\right]V_{t-1}\begin{bmatrix}-1\\ \theta\end{bmatrix}\right\}}_{\text{apriorní hp}},$$

odkud porovnáním obou stran tohoto vztahu dostaneme rovnice pro přepočet statistik

$$V_t = V_{t-1} + D_t, \tag{1.9}$$
$$\kappa_t = \kappa_{t-1} + 1. \tag{1.10}$$

Uvedené statistiky se nazývají: V - rozšířená informační matice, κ - počítadlo vzorků a datová matice D_t. je

$$D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix}\begin{bmatrix} y_t & \psi_t' \end{bmatrix},$$

Algoritmus odhadu

Postup odhadu parametrů je následující:

1. Konstrukce apriorní hp $f\left(\Theta|d\left(0\right)\right)$. Obecně to není jednoduchá záležitost. Jedná se o převod předpokladů nebo požadavků týkajících se parametrů Θ (jako třeba Θ má nezáporné prvky, nebo menší než nějaká horní hranice apod.) do apriorních statistik konstruované apriorní hp.
   Výstupem jsou apriorní statistiky V_0 a κ_0 a z nich zkonstruovaná apriorní hp podle (1.8) s dosazenými apriorními statistikami.

2. Měření dat $d_1, d_2, \cdots$ kde $d = \{y,\, u\}$ a průběžný přepočet statistik podle vztahů (1.9) a (1.10) tj.
   $$V_t = V_{t-1} + D_t \;\text{ a }\; \kappa_t = \kappa_{t-1} + 1.$$
   Výstupem jsou aposteriorní statistiky V_t a κ_t.

3. Konstrukce aposteriorní hp $f\left(\Theta|d\left(t\right)\right)$ dosazením do vztahu (1.8) s $\tau = t$ a statistikami V_t, κ_t.
   Výstupem je aposteriorní hp $f\left(\Theta|d\left(t\right)\right)$.

4. Výpočet bodových odhadů $\hat{\Theta}_t$ (jsou-li potřeba). Bodové odhady jsou dány jako podmíněná střední hodnota (viz Příloha **??**)

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

V Příloze **??**, v rovnicích (**??**) a (**??**) je ukázáno, že platí

$$\hat{\Theta}_t = V_\psi^{-1} V_{y\psi}, \quad \hat{r}_t = \frac{V_y - V_{y\psi}' V_\psi^{-1} V_{y\psi}}{\kappa_t}. \tag{1.11}$$

Matice V_ψ a vektor $V_{y,\psi}$ dostaneme rozdělením informační matice V_t na submatice

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix}, \tag{1.12}$$

kde V_y je skalár, $V_{y\psi}$ je sloupcový vektor, $V_{y\psi}'$ je řádkový vektor a V_ψ je čtvercová matice a
κ_t je počítadlo datových vzorků, pro které platí $\kappa_\tau = \kappa_{\tau-1} + 1,\ \tau = 1, 2 \cdots t$ s počáteční hodnotou κ_0 (apriorní statistika).

Výsledkem uvedeného algoritmu je:

- konstrukce aposteriorní hp podle bodu 3,
- bodové odhady parametrů (1.11).

Bodový odhad výstupu s bodovým odhadem parametru

Střední hodnotu výstupu y_t (tedy jeho optimální odhad) dostaneme přímo z modelu takto

$$\hat{y}_t = E\left[y_t | \psi_t, d\left(t-1\right)\right] = E\left[\psi_t' \hat{\theta}_t + e_t\right] = \psi_t' \hat{\theta}_t.$$

To znamená, že do modelu bez šumu[4] dosadíme data a odhady parametrů a jednoduše spočteme výstup.

Odhadování parametrů regresního modelu a jeho výstupu budeme ilustrovat na příkladech.

Příklad [odhad regresního modelu]

Simulujte spojitý dynamický systém popsaný regresním modelem 2. řádu s regresním vektorem

$$\psi_t = \left[u_t, y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}, 1\right]',$$

[4] Odhadem je střední hodnota. Střední hodnota šumu je nula, takže predikce se provádí jakoby s modelem bez šumu.

jemu odpovídajícími regresními koeficienty

$$\theta = [1,\ 0.6,\ 0.5,\ -0.2,\ -0.3,\ 0.1]'$$

a rozptylem šumu $r = 0.01$. Odhadněte parametry $\Theta = \{\theta, r\}$ tohoto systému.

Simulace se provádí s pomocí modelu ($\sigma = \sqrt{r} = 0.1$)

$$y_t = \psi_t'\theta + \sigma e_t = u_t + 0.6y_{t-1} + 0.5u_{t-1} - 0.2y_{t-2} - 0.3u_{t-2} + 0.1 + 0.1e_t,$$

kde

$e_t \sim N_{e_t}(0,1)$ je realizace standardního normálního šumu, ψ_t je regresní vektor ze zadání úlohy, θ jsou regresní koeficienty a σ je směrodatná odchylka šumu.

Pro odhad vytváříme rozšířený regresní vektor

$$\Psi_t = \left[y_t,\ \psi_t'\right]' \tag{1.13}$$

pro $t = 1, 2, \cdots, n_t$, se kterým přepočítáváme informační matici a počítadlo

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

s počátečními (apriorními) hodnotami V_0 a κ_0.

Bodové odhady určíme rozkladem informační matice podle (??)

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix},$$

a dále podle vzorců

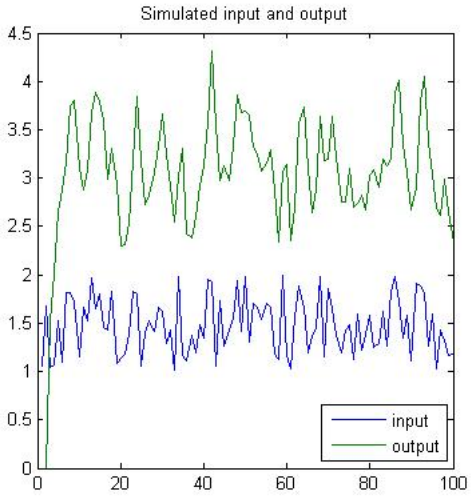
$$\hat{\theta}_t = V_\psi^{-1}V_{y\psi} \quad \text{a} \quad \hat{r}_t = \frac{V_y - V_{y\psi}'V_\psi^{-1}V_{y\psi}}{\kappa_t}. \tag{1.14}$$

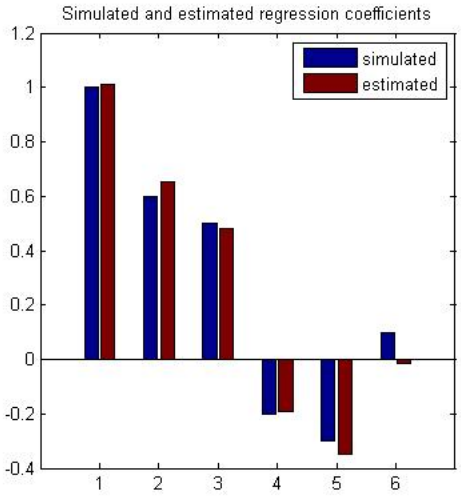
Podrobné řešení je uvedeno v následujícím programu

```
clc, clear all
// Simulation and estimation of second order
//    regression model

nt=100;                              // number of data

// SIMULATION
th=[1 .6 .5 -.2 -.3 .1]';            // regression coefficients
r=.01;                               // noise variance
s=sqrt(r);                           // standard deviation of noise
y=zeros(1,nt);                       // zero initial conditions + declar.
u=ones(1,nt)+rand(1,nt);             // input declaration
// time loop for simulation
for t=3:nt
  psi=[u(t) y(t-1) u(t-1) y(t-2) u(t-2) 1]';
```

```
  y(t)=psi'*th + s*randn;
end

// ESTIMATION
V=zeros(7);                          // initial statistics
// time loop for estimation
for t=3:nt
  Psi=[y(t) u(t) y(t-1) u(t-1) y(t-2) u(t-2) 1]';
  V=V+Psi*Psi';
end
Vy=V(1,1);
Vyp=V(2:end,1);
Vp=V(2:end,2:end);
Eth=inv(Vp)*Vyp;                     // point est. of regr. coeff.
Cth=(Vy-Vyp'*inv(Vp)*Vyp)/nt;        // point est. of noise var.

// print
Simulated_noise_variance=r
Estimated_noise_variance=Cth
```

Výsledky programu jsou na obrázcích





Příklad [bodový odhad regresního modelu]

Alternativní postup při odhadu parametrů normálního regresního modelu vycházející z metody nejmenších čtverců (který je ale s předchozím způsobem ekvivalentní) je následující.

Do rovnice regresního modelu (z předchozího příkladu)

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

postupně dosazujeme měřená data a rovnice pro $t = 1, 2, \cdots, N$ píšeme pod sebe

$$\begin{aligned} y_1 &= b_0 u_1 + a_1 y_0 + b_1 u_0 + a_2 y_{-1} + b_2 u_{-1} + k + e_1 \\ y_2 &= b_0 u_2 + a_1 y_1 + b_1 u_1 + a_2 y_0 + b_2 u_0 + k + e_2 \\ &\cdots \\ y_N &= b_0 u_N + a_1 y_{N-1} + b_1 u_{N-1} + a_2 y_{N-2} + b_2 u_{N-2} + k + e_N. \end{aligned}$$

Vytvořenou soustavu rovnic zapíšeme maticově

$$Y = \Psi\theta + E,$$

kde Y je vektor modelované veličiny y_t, Ψ je matice regresních vektorů v řádcích, θ je vektor parametrů v pořadí odpovídajícím pořadí veličiny v regresním vektoru a E je vektor šumů.

Pro uvažovaný model bude mít soustava tvar

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix} = \begin{bmatrix} u_1 & y_0 & u_0 & y_{-1} & u_{-1} & 1 \\ u_2 & y_1 & u_1 & y_0 & u_0 & 1 \\ & & \cdots & \cdots & & \\ u_N & y_{N-1} & u_{N-1} & y_{N-2} & u_{N-2} & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ a_1 \\ b_1 \\ a_2 \\ b_2 \\ k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_1 \\ \cdots \\ e_N \end{bmatrix}$$

Bodový odhad parametrů je

$$\hat{\theta} = \left(\Psi'\Psi\right)^{-1} \Psi Y,$$

odhad rozptylu šumu je

$$\hat{r} = Y' \left(Y - \hat{\theta}\Psi\right) = Y' E_p,$$

kde $E_p = Y - \hat{\theta}\Psi$ je vektor chyb predikce $Y_p = \hat{\theta}\Psi$.

Poznámka

Z porovnání obou uvedených příkladů plyne

$$Y'Y = V_y, \;\; Y'\Psi = V_{y\psi} \;\; \text{a} \;\; \Psi'\Psi = V_\psi.$$

◁

2 Odhad diskrétního a logistického modelu

2.1 Odhad parametrů diskrétního modelu

Diskrétní model popisuje systém, v němž jsou všechny veličiny diskrétní. Model je reprezentován tabulkou pravděpodobností. Model i úlohy s ním spojené jsou velice jednoduché, ale

- práce s tabulkami může být poněkud nezvyklá,
- pokud mají veličiny větší počet různých hodnot, bude tabulka modelu neúnosně veliká. V tom případě je lépe přejít na logistickou regresi.

Model a jeho součinový tvar

Model (**??**) obsahuje parametry $\Theta_{y|\psi}$ (pravděpodobnosti jednotlivých konfigurací rozšířeného regresního vektoru $\Psi_t = \left[y_t, \psi_t'\right]'$), které jsou v obecném případě neznámé, a tak je třeba je odhadovat z měřených dat. Pro odhad použijeme Bayesův vzorec (1.1), diskrétní model (**??**) a vhodně zvolený apriorní model parametrů $f\left(\Theta|d\left(0\right)\right) = f\left(\Theta\right)$. Dle odstavce 1.1 o analytickém tvaru hp parametrů a o reprodukovatelnosti její struktury je třeba tuto hp parametrů zvolit v analytickém tvaru, a to takovém, aby se při postupném násobení modelem soustavy jeho analytický tvar reprodukoval. Za tímto účelem přepíšeme model soustavy (**??**) formálně do tzv. **součinového tvaru**

$$f\left(y_t|\psi_t,\Theta\right) = \prod_{y|\psi\in\Psi^*} \Theta_{y|\psi}^{\delta(y|\psi,y_t|\psi_t)},$$

kde $y|\psi$ je multiindex (tj. vektorový index), který může nabývat všech možných konfigurací hodnot přípustných pro jednotlivé veličiny v něm obsažených; Ψ^* je množina všech takových konfigurací; symbol $\delta\left(y|\psi, y_t|\psi_t\right)$ je Kroneckerova funkce, která se rovná jedné, když platí $y|\psi = y_t|\psi_t$ (tj. $y = y_t$ a $\psi_1 = \psi_{1;t}$ až $\psi_{n_\psi} = \psi_{n_\psi;t}$), v ostatních případech je rovna nule.

Příklad [model falešné mince]

Pro model falešné koruny ($y = 1$ je líc, $y = 2$ je rub) lze zapsat jednotlivé pravděpodobnosti ve tvaru $f\left(y=1\right) = p_1$ a $f\left(y=2\right) = p_2$, kde $p_1, p_2 \geq 0$ a $p_1 + p_2 = 1$. Tento model můžeme zapsat následovně:

$$f\left(y\right) = p_1^{\delta(y,1)} p_2^{\delta(y,2)} = \prod_{i=1}^{2} p_i^{\delta(y,i)}.$$

Statistika

Přepis do součinového tvaru je čistě formální záležitost. Pro všechna $y|\psi \neq y_t|\psi_t$ dostáváme $\Theta^0 = 1$ a jen pro $y|\psi = y_t|\psi_t$ je $\Theta^1 = \Theta_{y_t|\psi_t}$. Nicméně je pro nás součinový tvar návodem, jak

volit apriorní model parametrů. Ten píšeme ve tvaru Dirichletova rozdělení - viz Přílohy **??**

$$f\left(\Theta|d\left(0\right)\right)=\prod_{y|\psi\in\Psi^{*}}\Theta_{y|\psi}^{\nu_{y|\psi;0}},$$

kde $\nu_{y|\psi;0}$ je apriorní statistika (pro čas $t=0$) pro odhad parametru Θ. Tato statistika je matice stejných rozměrů jako Θ. Podobně jako model ji můžeme reprezentovat pomocí tabulky. Počáteční statistika má tedy následující tvar

$$\nu_{y|\psi;0} \tag{2.1}$$

$[u_0, y_{-1}]$	$y_0=1$	$y_0=2$
1, 1	$\nu_{1\|11}$	$\nu_{2\|11}$
1, 2	$\nu_{1\|12}$	$\nu_{2\|12}$
2, 1	$\nu_{1\|21}$	$\nu_{2\|21}$
2, 2	$\nu_{1\|22}$	$\nu_{2\|22}$

Přepočet statistiky

Předchozí odvozené vztahy dosadíme do Bayesova vzorce (1.1) a dostaneme pro $t=1$

$$f\left(\Theta|d\left(1\right)\right)\propto\underbrace{\prod_{y|\psi\in\Psi^{*}}\Theta_{y|\psi}^{\delta(y|\psi,y_1|\psi_1)}}_{\text{model}}\underbrace{\prod_{y|\psi\in\Psi^{*}}\Theta_{y|\psi}^{\nu_{y|\psi;0}}}_{\text{apriorní}}=$$

$$=\underbrace{\prod_{y|\psi\in\Psi^{*}}\Theta_{y|\psi}^{\delta(y|\psi,y_1|\psi_1)+\nu_{y|\psi;0}}}_{\text{aposteriorní}}=\prod_{y|\psi\in\Psi^{*}}\Theta_{y|\psi}^{\overbrace{\nu_{y|\psi;1}}^{\text{nová statistika}}},$$

kde

$$\nu_{y|\psi;1}=\delta\left(y|\psi,y_1|\psi_1\right)+\nu_{y|\psi;0}$$

je statistika modelu parametrů $f\left(\Theta|d\left(1\right)\right)$ pro čas $t=1$.

Dále měříme data pro $t=2,3,\cdots,N$ a přepočítáváme statistiku ν podle obecného vzorce

$$\nu_{y|\psi;t}=\delta\left(y|\psi,y_t|\psi_t\right)+\nu_{y|\psi;t-1}. \tag{2.2}$$

Význam přepočtu statistiky ne následující: Po každém změření dat přičteme jedničku do toho políčka statistiky (viz (2.1)), které odpovídá dané konfiguraci hodnot rozšířeného regresního

vektoru $y_t|\psi_t$. Každé políčko tedy obsahuje počet, kolikrát příslušná konfigurace hodnot dosud nastala.

Aposteriorní hp parametrů se statistikou ν_t má tvar

$$f\left(\Theta|d\left(t\right)\right) \propto \prod_{y|\psi\in\Psi^*} \Theta_{y|\psi}^{\nu_{y|\psi;t}} \tag{2.3}$$

Bodový odhad parametrů

Bodový odhad parametru Θ je (viz Přílohy **??**)

$$\hat{\Theta}_{y|\psi;t} = \frac{\nu_{y|\psi;t}}{\sum_{y\in y^*} \nu_{y|\psi;t}}, \tag{2.4}$$

což zcela odpovídá statistické definici pravděpodobnosti jevu označeného indexem $y|\psi$. Pro každou konfiguraci regresního vektoru ψ je v čitateli počet případů, kolikrát nastala daná hodnota y v rámci tohoto regresního vektoru (počet příznivých pokusů) a ve jmenovateli je celkový počet pokusů.

Tabulka bodových odhadů má stejný tvar jako samotný parametr nebo statistika

$$\hat{\Theta}_t \tag{2.5}$$

$[u_t, y_{t-1}]$	$y_t = 1$	$y_t = 2$
1, 1	$\hat{\Theta}_{1\|11;t}$	$\hat{\Theta}_{2\|11;t}$
1, 2	$\hat{\Theta}_{1\|12;t}$	$\hat{\Theta}_{2\|12;t}$
2, 1	$\hat{\Theta}_{1\|21;t}$	$\hat{\Theta}_{2\|21;t}$
2, 2	$\hat{\Theta}_{1\|22;t}$	$\hat{\Theta}_{2\|22;t}$

a prvky této tabulky se počítají podle vzorce (2.4).

Bodový odhad výstupu

Podobně jako v případě spojitého regresního modelu i pro diskrétní model s neznámými parametry je možno neznámé parametry nahradit jejich bodovými odhady - z tabulky (viz (2.1)) vezmeme řádek odpovídající kombinaci hodnot v regresnímu vektoru ψ_t a v něm jako odhad $\hat{y}_t$ výstupu y_t vezmeme tu hodnotu, která má větší pravděpodobnost.

Příklad [odhad s diskrétním modelem]

V dopravní oblasti byly sledovány nehody a byly rozlišeny podle závažnosti na lehké (jen hmotná škoda) a těžké (vážné zranění nebo smrt). Nehody jako modelovanou veličinu označíme y_t (kde t nyní označuje pořadí nehody, nikoli čas) s hodnotami $y_t = 1$ - lehká nehoda a $y_t = 2$ - těžká nehoda. Předpokládáme, že na typ nehody mají hlavní vliv následující veličiny: ψ_1 - rychlost (1 normální, 2 velká), ψ_2 - počasí (1 sucho, 2 mokro) a ψ_3 - osvětlení (1 světlo, 2 tma).

Po dobu jednoho roku jsme měřili data a získali 18 následujících záznamů.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y_t	1	1	2	1	2	1	1	1	2	1	2	2	1	1	1	2	1	1
$\psi_{1;t}$	1	2	2	2	2	1	1	2	2	2	2	1	1	2	1	2	2	1
$\psi_{2;t}$	2	1	1	1	2	2	2	2	1	2	2	2	2	2	1	2	1	2
$\psi_{3;t}$	1	2	1	2	1	2	2	1	2	2	2	2	2	2	2	1	2	1

Model pro popis nehod podle (??) s využitím měřených dat bude mít tvar$f\left(y_t|v_t\right)$ s následující tabulkou

$$\Theta_{y|\psi}$$

$[v_{1;t},\, v_{2;t},\, v_{3;t}]$	$y_t = 1$	$y_t = 2$
$[1\,1\,1]$	$\Theta_{1\|111}$	$\Theta_{2\|111}$
$[1\,1\,2]$	$\Theta_{1\|112}$	$\Theta_{2\|112}$
$[1\,2\,1]$	$\Theta_{1\|121}$	$\Theta_{2\|121}$
$[1\,2\,2]$	$\Theta_{1\|122}$	$\Theta_{2\|122}$
$[2\,1\,1]$	$\Theta_{1\|211}$	$\Theta_{2\|211}$
$[2\,1\,2]$	$\Theta_{1\|212}$	$\Theta_{2\|212}$
$[2\,2\,1]$	$\Theta_{1\|221}$	$\Theta_{2\|221}$
$[2\,2\,2]$	$\Theta_{1\|222}$	$\Theta_{2\|222}$

Statistika odhadu podle (2.1) je reprezentována stejnou tabulkou. Aktualizace statistiky začíná s apriorní tabulkou a dále pokračuje podle vzorce (2.2), který říká: pro každé $t = 1, 2, \cdots, n_t$ podle hodnot příslušných veličin y_t a $\psi_t = [v_{1;t},\, v_{2;t},\, v_{3;t}]'$ najděte v tabulce odpovídající políčko a k němu přičtěte jedničku.

V našem příkladě zvolíme nulovou apriorní tabulku, která odpovídá nulové apriorní informací. S touto nulovou počáteční statistikou ν_0 dostaneme odhadovou statistiku

$$\nu_{y|\psi;0}$$

$[v_{1;t},\, v_{2;t},\, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	0	0
[1 1 2]	1	0
[1 2 1]	2	0
[1 2 2]	3	1
[2 1 1]	0	1
[2 1 2]	3	1
[2 2 1]	1	2
[2 2 2]	2	1

odkud (prostou normalizací řádků tabulky na součet jedna) dostaneme odhad parametrů modelu.

$$\hat{\Theta}_{y|\psi;n_t}$$

$[v_{1;t},\, v_{2;t},\, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	–	–
[1 1 2]	1	0
[1 2 1]	1	0
[1 2 2]	3/4	1/4
[2 1 1]	0	1
[2 1 2]	3/4	1/4
[2 2 1]	1/3	2/3
[2 2 2]	2/3	1/3

Výsledek je poměrně špatný. Parametry v prvním řádku nelze určit (zde statistika nebyla vůbec přepočtena), řádek 2,3 a 5 je deterministický (ve skutečnosti je to dáno tím, že do těchto řádků přišel jen jediný údaj) a ostatní řádky vypadají lépe, nicméně každý řádek odpovídá pokusu s hodem mincí. Umíme si představit, kolik hodů je třeba, abychom alespoň trochu objektivně posoudili pravděpodobnosti jejích stran. Podle hodnot statistiky vidíme, že do těchto řádků jsou započteny maximálně čtyři údaje. To je velmi málo a je zřejmé, že vzhledem k dimenzi tabulky statistiky máme neúnosně málo dat. Nedostatek dat v těchto případech je velmi častý a je třeba možnosti úlohy v takovém případě dobře zvážit.

Jednou z možností, jak řešit nedostatek dat, je uvažovat další informaci a tou je informace expertní. V krajním případě je dokonce možné postavit model na expertní (apriorní) informaci a změřená data použít jen k jakési korekci modelu. Tuto možnost ukážeme v další části příkladu.

Apriorní statistiku je možno sestavit takto: pro všechny řádky tabulky (tj. pro všechny možné regresní vektory)

1. určíme, jaké pravděpodobnosti bychom přiřadili hodnotám y pro konkrétní kombinaci hodnot ψ,
2. tyto pravděpodobnosti násobíme číslem, které vyjadřuje míru důvěry k našemu přiřazení.

Např.: první regresní vektor v tabulce je $\psi = [1,\, 1,\, 1]'$, tj. rychlost = normální, počasí = sucho, světlo = dobré. V této situaci rozhodujeme o typu nehody. Můžeme říci, že jsou to ideální podmínky, a tak nehoda může být jen lehká. Volíme proto první řádek statistiky

$$\nu_{:|111;0]} = [9,\, 1]\,.$$

To odpovídá 10 údajům o nehodách, z nichž 9 bylo lehkých a 1 těžká.

Druhý regresní vektor v tabulce je $\psi = [1,\, 1,\, 2]$ - rychlost = normální, počasí = sucho, světlo = šero. Za šera jsou někdy podmínky jízdy ošemetné zvláště pro chodce, kteří jsou špatně vidět. Proto volíme

$$\nu_{:|112;0]} = [1,\, 4]\,,$$

což odpovídá pěti záznamům, jeden s lehkou a čtyři s těžkou nehodou.

Stejně budeme pokračovat i pro další regresní vektory, a tak získáme následující tabulku pro apriorní statistiku.

ν_0

$[v_1,\, v_2,\, v_3]$	$y = 1$	$y = 2$
[1 1 1]	9	1
[1 1 2]	1	4
[1 2 1]	2	2
[1 2 2]	2	3
[2 1 1]	3	2
[2 1 2]	3	7
[2 2 1]	3	7
[2 2 2]	1	9

Odhadem z dat získáme parametry:

$$\hat{\Theta}_{y|\psi;n_t} \tag{2.6}$$

$[v_{1;t},\, v_{2;t},\, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	9/10	1/10
[1 1 2]	1/3	2/3
[1 2 1]	2/3	1/3
[1 2 2]	5/9	4/9
[2 1 1]	1/2	1/2
[2 1 2]	3/7	4/7
[2 2 1]	4/13	9/13
[2 2 2]	3/13	10/13

V řádku, kde nepřišla žádná data (např 1. řádek), zůstaly apriorní odhady. Tam, kam data přišla, je apriorní informace korigována daty.

2.2 Odhad parametrů logistického modelu

V případě, kdy modelovaná veličina je diskrétní a závisí jak na diskrétních tak i na spojitých veličinách, použijeme model logistické regrese. Tento model lze použít i v případě, kdy všechny veličiny jsou diskrétní ale mají velký počet různých hodnot, takže čistě diskrétní model by měl příliš vysokou dimenzi.

Zde se budeme zabývat odhadem modelu zavedeného podle (**??**, **??**), tedy modelem, který má tvar

$$logit\left(p_t\right) = \psi_t'\Theta + e_t,$$

kde $p_t = P\left(y_t = 1|\psi_t\right)$ a $y_t \in \{0,\, 1\}$, P je pravděpodobnost a y_t je dvouhodnotový výstup. ψ_t je regresní vektor externích veličin, např. "počasí" s hodnotami sucho, mokro, námraza nebo "den v týdnu" s hodnotami všední den, víkend. Parametry modelu jsou prvky vektoru Θ. Tyto parametry se odhadují na základě změřené množiny dat $d\left(t\right) = \{y_\tau, \psi_\tau\}_{\tau=1}^{t}$, kde y_τ jsou skaláry a $\psi_\tau = [1, x_{1;\tau}, x_{2;\tau}, \cdots x_{n;\tau}]'$ je sloupcový vektor hodnot nezávislé proměnné. Jednička na začátku je přidána pro odhad konstanty modelu.

Lze sestavit i algoritmus pro odhad logistické regrese s více hodnotami y_t. Ten je ale složitější a my se s ním zde nebudeme zabývat.

Odhad logistického modelu

Tato úloha nemá reprodukující se statistiku. Odhad se provádí jednorázově metodou ML (maximum likelihood) pro celý shromážděný datový vzorek. Za tímto účelem konstruujeme logaritmus věrohodnostní funkce

$$\ln L\left(\Theta\right) = \ln \prod_{\tau=1}^{t} f\left(y_\tau|\psi_\tau, \Theta\right)$$

jako součin modelů (**??**), kde platí (**??**). Po dosazení a drobných úpravách dostaneme

$$\ln L\left(\Theta\right) = \ln \prod_{\tau=1}^{t} \frac{\exp\left(y_\tau z_\tau\right)}{1 + \exp\left(z_\tau\right)} = \sum_{\tau=1}^{t} \left[y_\tau z_\tau - \ln\left(1 + \exp\left(z_\tau\right)\right)\right],$$

kde $z_t = \psi_t'\Theta$.

Bodové odhady leží v maximu logaritmu věrohodnostní funkce. Toto maximum lze výhodně nalézt Newtonovou metodou, protože jak gradient (první derivace), tak i Hessovu matici (druhá derivace) je možno vyjádřit v analytickém tvaru. Odvození je uvedeno v Příloze **??**.

Předpověď výstupu

Logistickou regresi nejspíše děláme proto, abychom byli schopní pro daný regresní vektor odhadnout (pro jednotnost s předchozím výkladem říkáme předpovědět) hodnotu odpovídajícího výstupu.

Toho lze dosáhnout tak, že bodové odhady $\hat{\Theta}_t$ dosadíme do modelu (**??**), (**??**) a pro libovolný regresní vektor ψ, dostáváme odhad[5] pravděpodobnosti hodnot y

$$f\left(y|\psi,\hat{\Theta}_t\right)=\frac{\exp\left(y\psi'\hat{\Theta}_t\right)}{1+\exp\left(\psi'\hat{\Theta}_t\right)}.$$

Bodovou predikci $\hat{y}$ pro regresní vektor ψ určíme jako podmíněnou střední hodnotu[6]

$$\hat{y}=E\left[y|\psi,d\left(t\right)\right]=\sum_{y=0}^{1}yf\left(y|\psi,\hat{\Theta}_t\right)=\frac{\exp\left(\psi'\hat{\Theta}_t\right)}{1+\exp\left(\psi'\hat{\Theta}_t\right)}.$$

Pokud požadujeme bodovou predikci pouze v přípustných hodnotách $\{0,1\}$, získanou hodnotu $\hat{y}$ zaokrouhlíme (pro $\hat{y}<0.5$ na hodnotu 0 a pro $\hat{y}\geq 0.5$ na hodnotu 1).

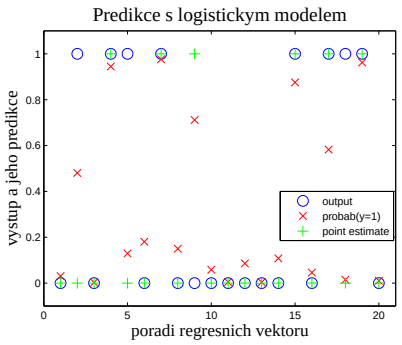
Příklad [odhad logistické regrese 1]

Uvažujme diskrétní veličinu y závislou na dvou veličinách regresního vektoru $\psi=[\psi_1,\psi_2]$. Tyto veličiny modelujeme jako normální náhodné veličiny $\psi_1\sim N\left(-0.1,\,0.25\right)$ a $\psi_2\sim N\left(0.3,\,1\right)$. Hodnoty y jsou přiřazeny následovně

$$y=\begin{cases}1 & \text{pro } 2\psi_1-\psi_2+e>1\\ 0 & \text{jinde}\end{cases}, \tag{2.7}$$

kde $e\sim N\left(0,\,1\right)$ je šum. Máme provést logistickou regresi těchto dat.

Řešení je dáno v programu T23LogRegM.m, který je možno nalézt v kapitole **??** Programy na str. **??** . Pro odhad byl použit vzorek 50 dat. Výsledek demonstrovaný na dvaceti následujících (testovacích) datech je na obrázku.



Z obrázku je vidět, že bodové odhady (+) většinou správně sedí na hodnotách výstupu (o), ale pravděpodobnosti predikcí (x) vykazují určitý stupeň neurčitosti. Ta je dána použitým heuristickým generátorem dat a šumem e, který jsme do dat přičetli při určování hodnot výstupu (2.7).

[5] Píšeme odhad, protože správné pravděpodobnosti bychom dostali nikoli dosazením bodových odhadů, ale násobením aposteriorní hp a integrací před parametry Θ. To je ale příliš složité.

[6] Z následujícího vzorce je patrné, že platí $\hat{y}=f\left(y|\psi,\hat{\Theta}_t\right)=P\left(y=1|\psi,\hat{\Theta}_t\right)$.

Příklad [odhad s logistickým modelem 2]

Uvažujme systém s výstupem $y \in \{1,2\}$ a dalšími veličinami, které tvoří regresní vektor $\psi = [\psi_1, \psi_2, \psi_3]$, $\psi_i \in \{1,2\}$. Dále jsme změřili data $d(t) = d(5)$, která jsou uvedena v následující tabulce.

data	ψ	y
1	$[2,2,2]$	1
2	$[1,2,2]$	2
3	$[1,1,1]$	1
4	$[1,1,1]$	1
5	$[1,1,1]$	2

Data jsou pod vlivem šumu. Je vidět, že poslední tři datové položky jsou ve sporu. Třetí a čtvrtá říká, že regresnímu vektoru $[1,1,1]$ odpovídá hodnota jedna, zatímco u páté položky je to dvojka.

Cílem příkladu je provést odhad koeficientů logistické regrese, určit predikce pro y a ukázat, jak je provedena klasifikace.

Pro odhad a predikci je opět možno použít program z m-souboru $\mathrm{T23LogRegM.m}$, který je uveden na str. ??. Odtud je možno získat výsledky:

Rovnice logistické regrese

$$logit\,(y) = b_0 + b_1\psi_1 + b_2\psi_2 + b_3\psi_3 + b_4\psi_4 \tag{2.8}$$

Parametry regrese

b_0	b_1	b_2	b_3
29.28	-63.22	16.49	16.76

V následující tabulce jsou uvedeny měřené hodnoty regresních vektorů ψ_t, výstupu y_t, predikované hodnoty výstupu y_t - pravděpodobnosti $P(y_t = 1)$ a bodové odhady $\hat{y}_t$ (zaokrouhlené hodnoty pravděpodobností).

t	ψ	y_t	$P(y_t = 1)$	$\hat{y}_t$
1	$[2,2,2]$	1	0	1
2	$[1,2,2]$	2	1	2
3	$[1,1,1]$	1	0.33	1
4	$[1,1,1]$	1	0.33	1
5	$[1,1,1]$	2	0.33	1

Z tabulky je patrné, že predikce pro hodnoty regresního vektoru, při kterých se vyskytovaly sporné hodnoty výstupu, nemají pravděpodobnost nula nebo jedna. Tyto predikce říkají, že jejich hodnoty jsou nejisté, ale přiklánějí se k hodnotě nula, protože nula odpovídala dvěma případům dat, zatímco jednička pouze jednomu. Ostatní predikce jsou správné a přesné.