

# Odhad - Problémy se sdruženým rozdělením pravděpodobnosti

20. listopadu 2013

V minulém materiálu jsme si ukázali, jak získat sdružené rozdělení pravděpodobnosti. Bylo to celkem jednoduché: Věrohodnostní funkci získám tak, že neustále násobím model a dosazuji do něj data. Sdružené rozdělení (aposteriorní) pak získám tak, že věrohodnostní funkci vynásobím případným apriorním sdruženým rozdělením.

Pokud to však zkusím v praxi, brzy narazím na několik problémů:

1. Graf sdruženého rozdělení mohu vykreslit jen pokud mám jednu nebo dvě neznámé. Pro více neznámých to není možné.
2. Při výpočtu zjistím, že hodnoty sdruženého rozdělení, které počítám, jsou tak malé, že je Matlab (či jiný software) zaokrouhlí na nulu a nic rozumného tedy nespočítám.
3. Když chci získané rozdělení napsat jako funkci (abych ji pak mohl např. integrovat), pak je někdy zápis tak dlouhý a složitý, že je to naprosto neúnosné.
4. Když chci spočítat sdružené rozdělení pro např. dvacet parametrů, z nichž každý budu počítat (byť jen) v deseti bodech, pak to znamená napočítat funkční hodnotu v  $10^{20}$  bodech, což obnáší miliardy Terabytů dat. To je pro počítač naprosto nezvládnutelné.

## 1 Grafické znázornění vícerozměrných rozdělení

Pokud opravdu chceme pro názornost vykreslit grafy, tak nám zde nám nezbyvá nic jiného, než udělat několik grafů, vždy pro dvě veličiny. Ostatní neznámé musíme ve funkci nahradit konstantami.

Jaké konstanty ale zvolit? Obvykle nás zajímá maximum funkce, abychom určili bodový odhad parametrů. Proto je vhodné za tyto konstanty zvolit právě bodové odhady parametrů.

Logika věci je následující: Necht' sdružené rozdělení nabývá maxima právě pro hodnoty parametrů:  $A = 1$ ,  $B = 3$ ,  $C = 5$  a  $D = 7$ . Pak můžeme vzít sdružené rozdělení, dosadit  $A = 1$  a  $B = 3$  a vykreslit graf pro neznámé  $C$  a  $D$ . Pak opravdu zjistíme, že nejvyšší bod rozdělení odpovídá hodnotám  $C = 5$  a  $D = 7$ .

Pokud bychom dosadili jiné  $A$  a  $B$ , vrchol grafu pro  $C$  a  $D$  by mohl ležet někde úplně jinde.

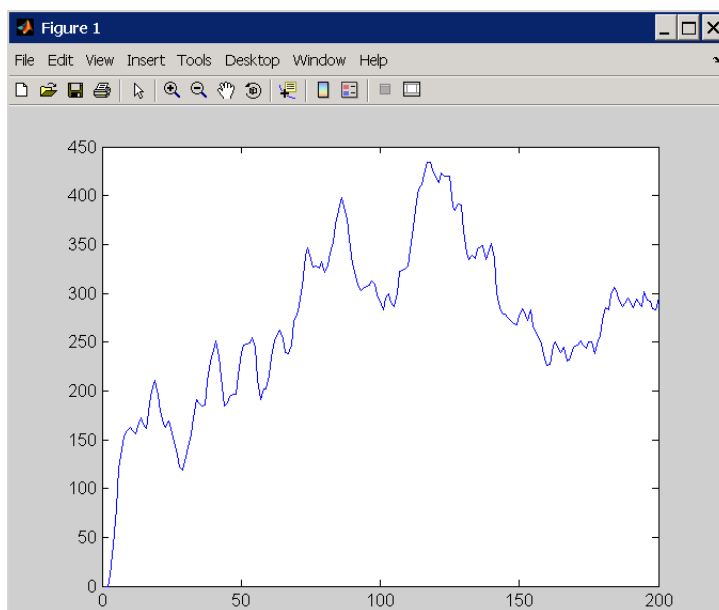
Jak získat bodové odhady, k tomu se dostaneme v dalším materiálu.

## Příklad

Načteme data ze souboru DataCviceni.mat, který je dostupný na našich stránkách. V tomto souboru je 200 hodnot  $y$  vygenerovaných modelem  $y_n = 1,5 \cdot y_{n-1} - 0,5 \cdot y_{n-2} + e_n$ , kde bílý šum  $e_n$  má normální rozdělení  $N(0, 100)$ .

Pokusíme se tato data modelovat modelem čtvrtého řádu  $y_n = A \cdot y_{n-1} + B \cdot y_{n-2} + C \cdot y_{n-3} + D \cdot y_{n-4} + e_n$  s normálním bílým šumem s neznámým rozptylem  $r$ . Pokusíme se také získat odhad pro hodnotu  $y_{201}$ .

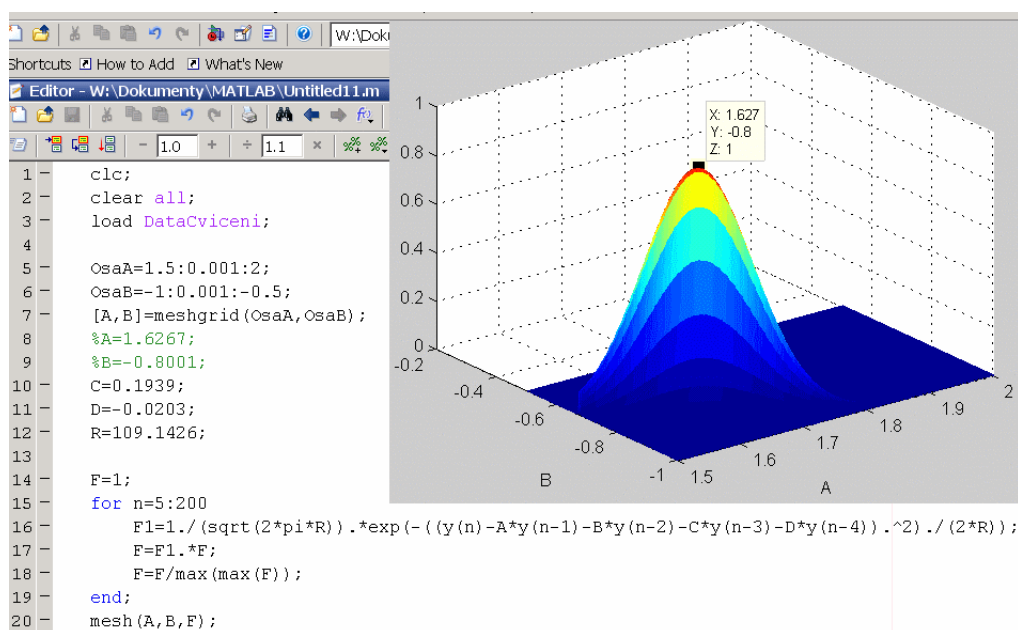
Příslušná data  $y$  zobrazuje následující graf:



Metodou, kterou si ukážeme v dalším materiálu, jsme získali bodové odhady parametrů:

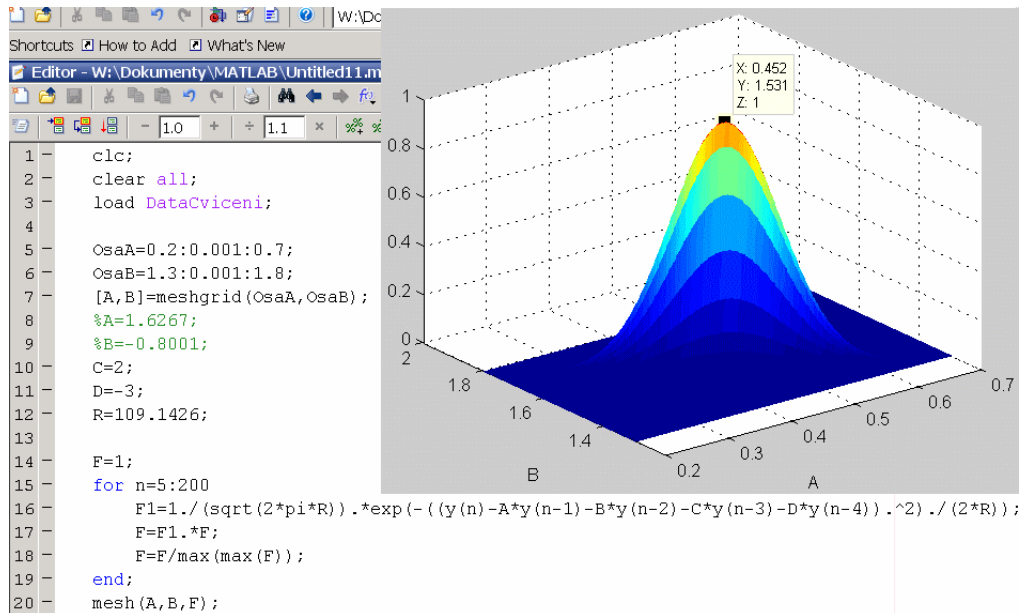
$A = 1.6267$ ,  $B = -0.8001$ ,  $C = 0.1939$ ,  $D = -0.0203$ ,  $R = 109.1426$ .

Vykreslíme grafy pro několik kombinací parametrů. Graf pro  $A$  a  $B$ :

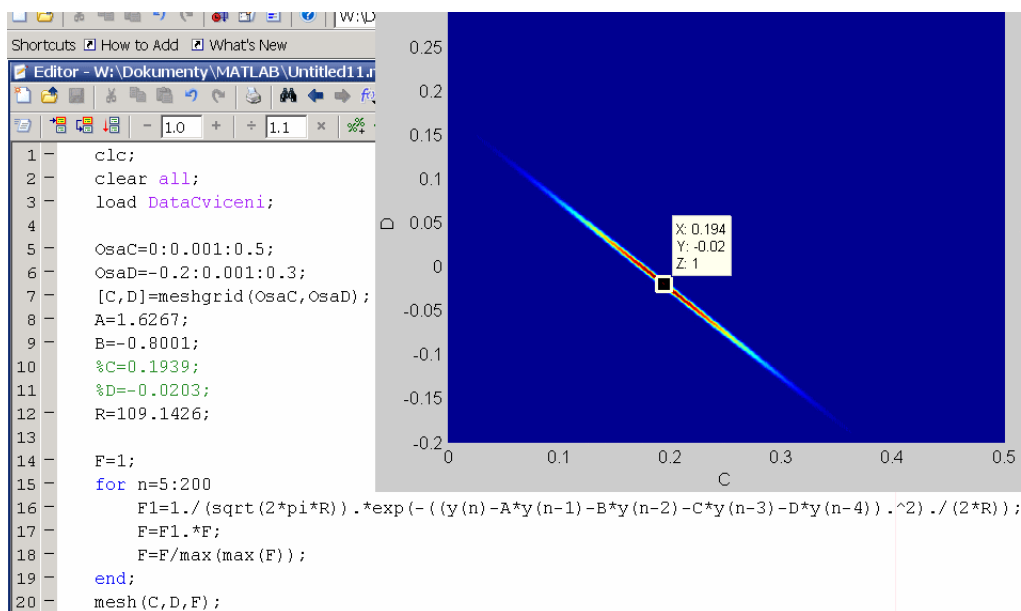


Vidíme, jak nechat parametry  $A$  a  $B$  jako proměnné a ostatní parametry dosadit.

Všimněte si, že pokud jsme zadali bodové odhady ostatních parametrů správně, graf nám opět správně ukáže i bodové odhady pro  $A$  a  $B$ . Pokud bychom ostatní parametry zvolili jinak, i pík pro  $A$  a  $B$  by vyšel jinde. To ostatně ukazuje následující graf, kde jsme za  $C$  a  $D$  zvolili hodnoty 2 a -3:



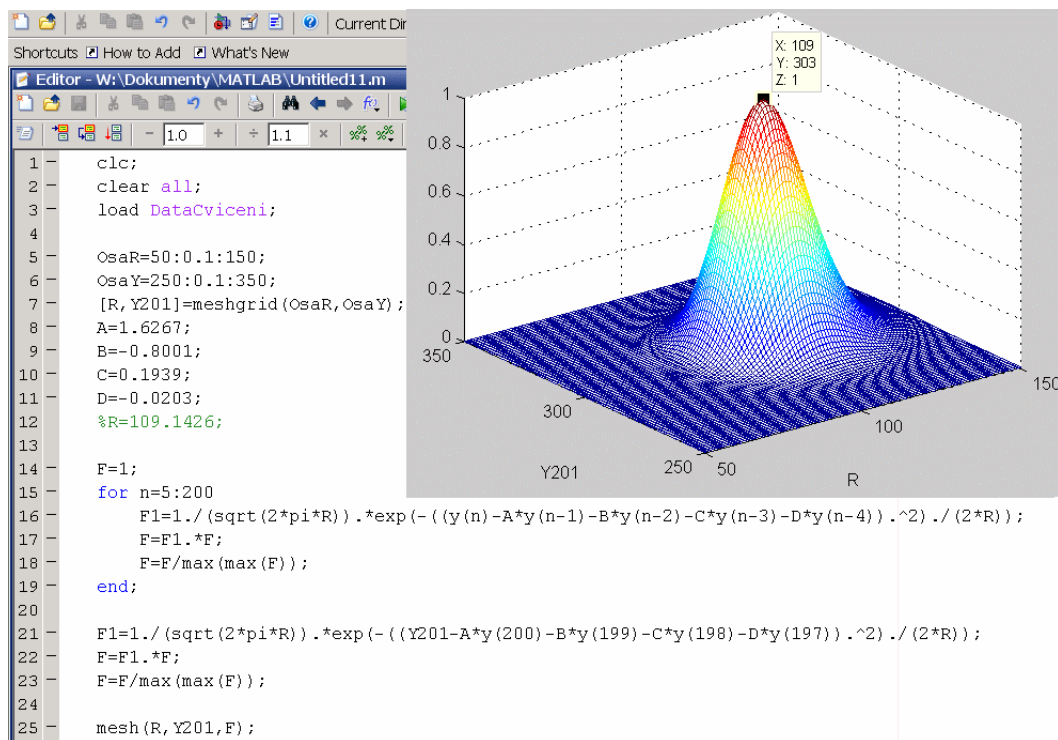
Pro úplnost si ukážeme ještě graf pro  $C$  a  $D$ . Vizualně je téměř stejný jako obrázek pro  $A$  a  $B$ , proto jsme pro zajímavost a pro názornost zvolili pohled shora. Vidíme, že parametry  $C$  a  $D$  jsou velmi silně závislé:



Na dalším grafu jsme vynesli na jednu osu rozptyl  $R$  a na druhou odhad pro 201. výstup  $Y_{201}$ .

Pro tento účel jsme ještě sdruženou hustotu pronásobili modelem pro neznámé  $y_{201}$ . Předtím jsme pracovali se sdruženou hustotou bez tohoto posledního modelu, protože jsme ho nepotřebovali a ostatně bychom ani nevěděli, jaký bodový odhad dosadit za  $y_{201}$ .

Z obrázku opět vyčteme bodový odhad pro  $R$  a i pro  $Y_{201}$ .



Teprve grafy pro všech šest neznámých veličin nám dávají jakousi představu o šestirozměrné sružené hustotě. V tomto případě to je poměrně jednoduché.

## 2 Napočítané hodnoty jsou příliš malé

Zde využijeme toho, že my vlastně uvedeným postupem nepočítáme přímo sružené rozdělení, ale pouze funkci úměrnou sruženému rozdělení. Můžeme tedy tuto funkci vynásobit libovolnou kladnou konstantou, aniž by se cokoli podstatného změnilo. Tvar funkce zůstane stejný. Tomuto násobení konstantou říkáme normalizace.

Sružené rozdělení můžeme v zásadě počítat dvěma způsoby. Jednak můžeme v nějakém for cyklu neustále do modelu dosazovat data a tímto modelem násobit, nebo můžeme mít už předpis pro celé sružené rozdělení hotový.

### Výpočet ve for cyklu

První způsob jsme použili např. u výše uvedeného příkladu na lineární regresní model. Normalizaci provádíme v každém cyklu a tím zachováme počítané hodnoty přiměřeně velké.

Normalizaci provádí řádek:

$$F=F/\max(\max(F));$$

Funkci dělíme jejím maximem, takže nové maximum bude vždy ve výšce jedna. To je ostatně vidět i na příslušném obrázku. Otázkou asi vyvolá dvojí použití funkce *max*. Funkce *max* aplikovaná na matici totiž nejprve udělá maxima z každého sloupce. A teprve druhé *max* udělá maximum i z těchto maxim, takže získá největší hodnotu z celé matice.

## Výpočet v jediném kroku

V minulém materiálu jsme v příkladu pro diskrétní model počítali rozdělení parametrů v jediném kroku. Bylo to snadné, neboť věrohodnostní funkce měla tvar součinu čtyř pravděpodobností s příslušnými mocninami. Pokud by mocniny byly příliš velké, snadno by se nám mohlo stát, že bychme se dostali do velmi malých čísel řádu  $10^{-1000}$  a podobně.

Řešení je snadné. Počítejme místo věrohodnostní funkce její logaritmus. Tím místo čísel řádu  $10^{-1000}$  dostaneme čísla řádu -1000, což již není problém. A nejvyšší bod věrohodnostní funkce bude samozřejmě odpovídat nejvyššímu bodu logaritmu.

Pokud chceme vykreslit graf rozdělení, může nám zde vadit, že logaritmus jej poněkud zdeformuje. Můžeme proto funkci ve tvaru logaritmu znormovat tak, aby maximum bylo v nule, a pak exponenciálou opět převést na původní tvar s maximem ve výšce jedna.

## Příklad

Zkusme vykreslit věrohodnostní funkci pro diskrétní model, pokud případ 1|1 (tedy jednička po jedničce) nastal 500 krát, případ 2|1 (dvojka po jedničce) 300 krát, případ 1|2 nastal 200 krát a případ 2|2 nastal 800 krát.

---

Tomu odpovídá věrohodnostní funkce:

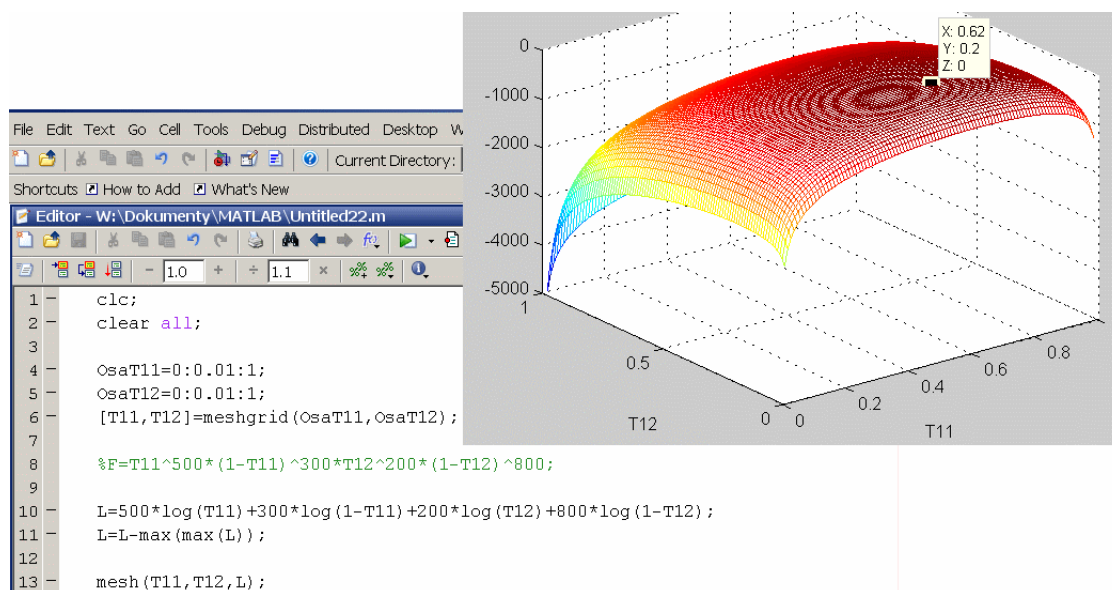
$$f(\theta_{1|1}, \theta_{1|2}) = \theta_{1|1}^{500} \cdot (1 - \theta_{1|1})^{300} \cdot \theta_{1|2}^{200} \cdot (1 - \theta_{1|2})^{800}.$$

Tuto funkci Matlab z výše uvedených důvodů nevykreslí. Vykresleme tedy přirozený logaritmus této funkce:

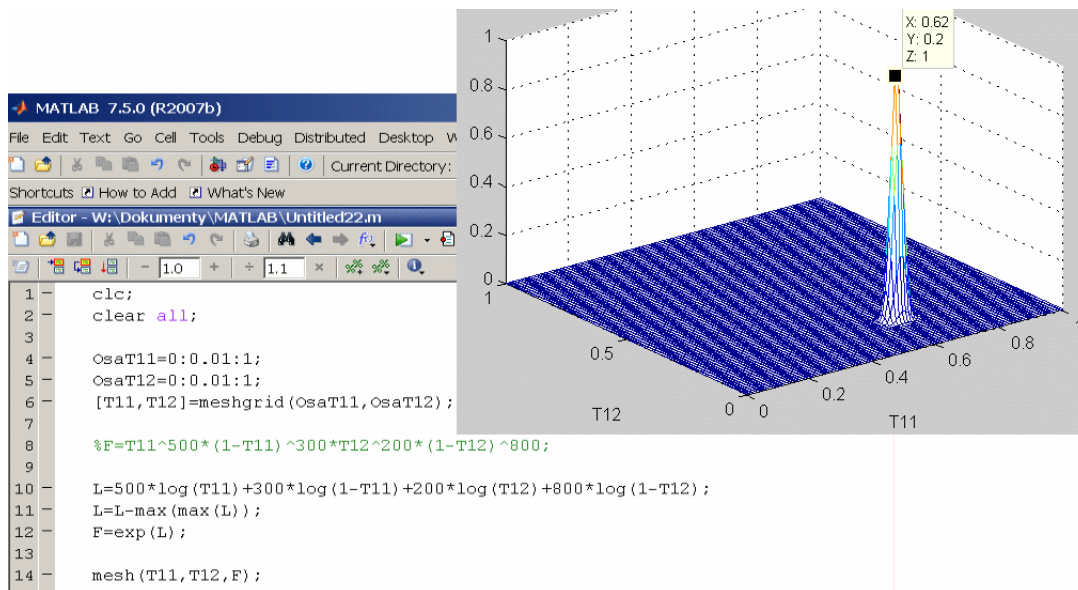
$$L(\theta_{1|1}, \theta_{1|2}) = 500 \cdot \ln \theta_{1|1} + 300 \cdot \ln(1 - \theta_{1|1}) + 200 \cdot \ln \theta_{1|2} + 800 \cdot \ln(1 - \theta_{1|2}).$$

Získáme následující graf, který nám pro určení bodového odhadu stačí.

Abychom nejvyšší bod funkce snadno našli, normujeme funkci na nulu. Tedy odečteme maximum funkce, takže nejvyšší bod bude na úrovni nula. Tvar funkce se normováním nezmění, jen se posune ve směru nahoru/dolů. Tuto normalizaci provádí řádek 11.



Pokud chceme graf pro původní věrohodnostní funkci, aplikujme na normovanou funkci  $L$  přirozenou exponenciálu:



### 3 Sdružené rozdělení, pokud je vyjádřím vzorcem, je příliš dlouhé a složité

Zde se nám někdy podaří najít ekvivalentní jednoduché vyjádření, které se s množstvím dat nezesložuje. Takovému vyjádření říkáme (samo)reprodukcující se rozdělení.

#### Diskrétní model

Příklad takového samoreprodukujícího se rozdělení jsme viděli v minulé kapitole pro diskrétní model:

$$f(\theta_{1|1}, \theta_{1|2}) = \theta_{1|1}^{500} \cdot (1 - \theta_{1|1})^{300} \cdot \theta_{1|2}^{200} \cdot (1 - \theta_{1|2})^{800}.$$

Pro libovolný počet dat má funkce tento jednoduchý tvar, jen se modifikují exponenty.

Podobný tvar má sdružená hustota pravděpodobnosti i pro jiné případy diskrétního modelu.

#### Lineární regresní model s normálním šumem

Také lineární regresní model s normálním šumem má samoreprodukující se tvar. Jednotlivé modely mají tvar gaussovek:

$$f(y_n | \theta, \psi) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(y_n - \theta \cdot \psi')^2}{2\sigma^2}}.$$

Ty se liší jen různými daty, která dosazujeme za  $y_n$  a do vektoru  $\psi$ . Při součinu těchto gaussovek se nám umocňuje úvodní zlomek a exponenty se sčítají. V součtu exponentů budou našťastí stále ty stejné členy. Pokud např. v čitateli exponentu máme:  $(y_n - Ay_{n-1} - Bx_n)^2$ , pak v součtu budou stále jen členy s  $A^2$ ,  $B^2$ ,  $AB$ ,  $A$ ,  $B$  a konstantou.

Elegantně to můžeme vyjádřit pomocí datové matice. Např. pro výše uvedený případ:

$$(y_n - Ay_{n-1} - Bx_n)^2 = (\Theta \cdot \Psi')^2 =$$

$$\begin{aligned}
&= \Theta \cdot \Psi' \cdot \Psi \cdot \Theta' = \\
&= \Theta \cdot (\Psi' \cdot \Psi) \cdot \Theta' = \\
&= \Theta \cdot D \cdot \Theta',
\end{aligned}$$

kde  $\Theta = (1, -A, -B)$  a  $\Psi = (y_n, y_{n-1}, x_n)$ . Datovou matici tedy spočteme:  $D = \Psi' \cdot \Psi$ . To je v našem případě matice 3 krát 3, v níž jsou součiny každého členu  $\Psi$  s každým.

Při součinu modelů se nám datové matice sčítají. Jejich součet nazveme informační maticí  $V = \sum_{i=1}^{\kappa} D_i$ , kde  $\kappa$  je počet modelů v součinu.

Lineární regresní model s normálním šumem má tedy v našem případě samoreprodukující se tvar:

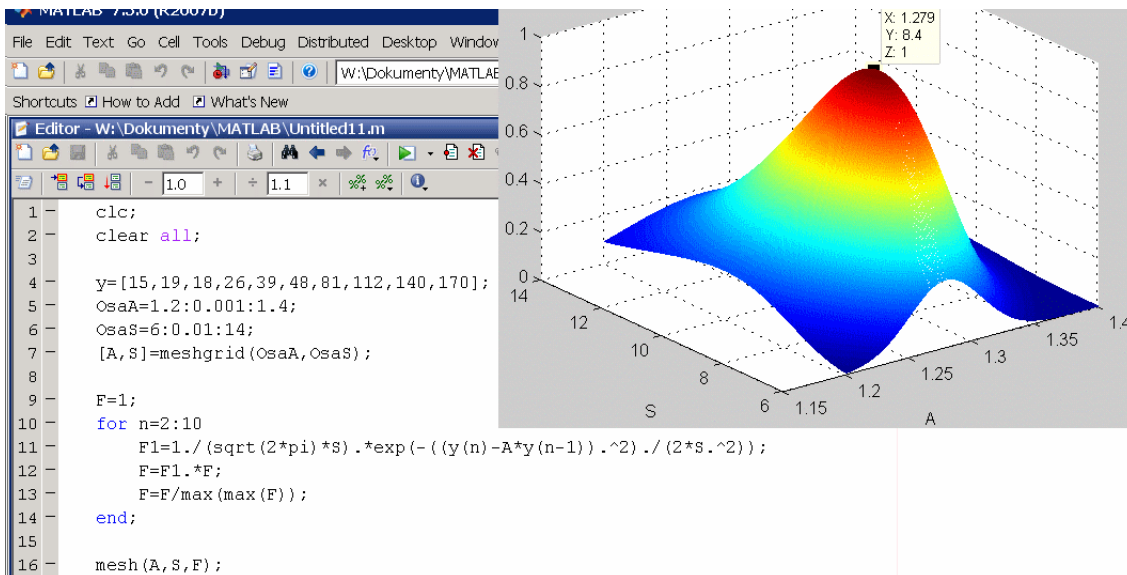
$$f(A, B, \sigma) = \left( \frac{1}{\sqrt{2\pi} \cdot \sigma} \right)^{\kappa} \cdot e^{-\frac{\Theta \cdot V \cdot \Theta'}{2\sigma^2}}.$$

Podobný tvar má sdružená hustota pravděpodobnosti i pro jiné případy lineárního regresního modelu s normálním šumem.

### Příklad

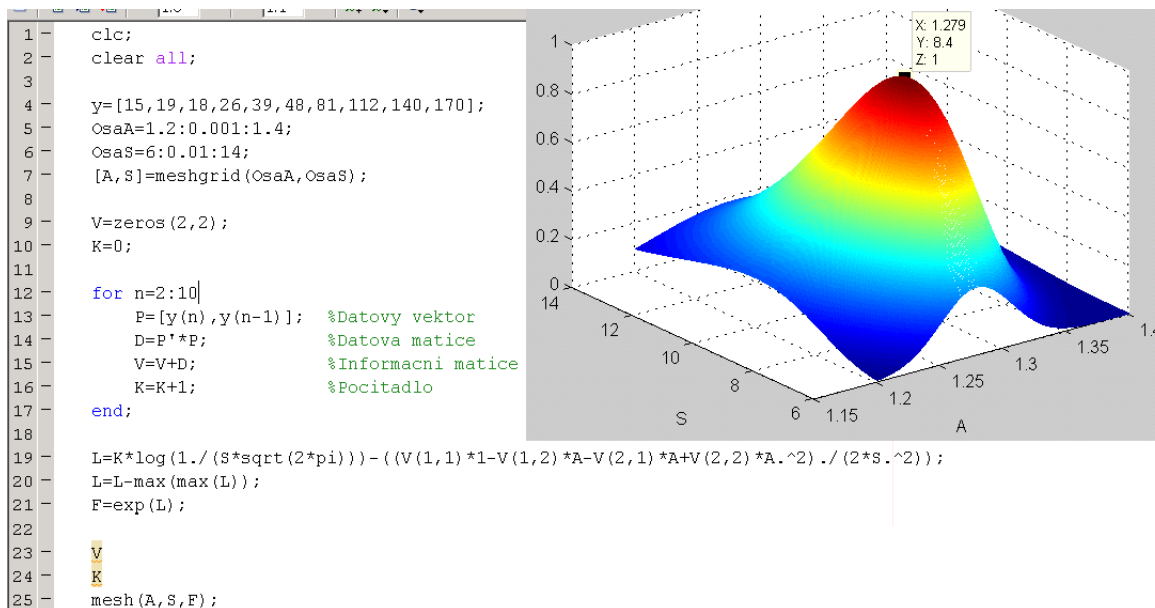
Máme lineární regresní model prvního řádu  $y_n = A \cdot y_{n-1} + e_n$  s normálním šumem  $e_n \sim N(0, \sigma^2)$ . Máme neznámé parametry  $A$  a  $\sigma$ . Vykreslete sdruženou hustotu pravděpodobnosti pro tyto parametry jednak postupným násobením modelů, jednak pomocí datové matice. Výsledky porovnejte. Data pro  $y$ : 15, 19, 18, 26, 39, 48, 81, 112, 140, 170

Postupným násobením modelů dostáváme tento výsledek:



Pomocí datové matice dostáváme tento výsledek:





Vidíme, že výsledky jsou naprosto stejné.

V tomto programu jsme na řádcích 19 až 21 použili trik s logaritmem. Nebyl v tomto případě ještě nutný, ale pro více dat by již nutný byl.

Exponent na řádku 19, jsme roznásobili, abychom ho mohli do Matlabu napsat. Tedy:

$$\begin{aligned}
 \Theta \cdot V \cdot \Theta' &= (1, -A) \cdot \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -A \end{pmatrix} = \\
 &= V_{11} - V_{12} \cdot A - V_{21} \cdot A + V_{22} \cdot A^2 = \\
 &= 72\,791 - 56\,421 \cdot A - 56\,421 \cdot A + 44\,116 \cdot A^2.
 \end{aligned}$$

Matici  $V$  a počítadlo  $K$  známe (řádek 23 a 24), proto můžeme výsledné sružené rozdělení přímo napsat:

$$f(A, S) = \left( \frac{1}{\sqrt{2\pi} \cdot S} \right)^9 \cdot e^{-\frac{72\,791 - 56\,421 \cdot A - 56\,421 \cdot A + 44\,116 \cdot A^2}{2S^2}}.$$

### Logistický model

Pro logistický model žádný úsporný samoreprodukcující se tvar neexistuje. Zde se můžeme jen pokusit aproximovat výsledné rozdělení nějakou podobnou funkcí s jednodušším zápisem.

## 4 Počet bodů, v kterých počítám sružené rozdělení pravděpodobnosti, je pro počítač nezvládnutelný

Toto se typicky týká mnohorozměrných rozdělení. Pokud chci nalézt maximum sruženého rozdělení a neznám lepší metodu, postupuji numericky.

Konkrétní metody jsou různé, ale princip je podobný. Začnu v bodě, který si jako maximum vytipuji. Zjistím, kterým směrem se hodnota funkce zvětšuje a tímto směrem popostoupím. To opakuji tak dlouho, až dospěji na vrcholek funkce.

Tyto metody fungují, pokud má rozdělení jen jedno (lokální) maximum. Pokud má lokálních maxim více, ledná se o poměrně náročnou úlohu.