Odhady - Sdružené rozdělení pravděpodobnosti

14. listopadu 2013

Kdybych chtěl znát maximum informací o náhodné veličině, musel bych znát všechny hodnoty, které mohou padnout, a jejich pravděpodobnosti. Tedy musel bych znát rozdělení pravděpodobnosti. [1]

Pokud mám více náhodných veličin, pak bych maximum informací měl, pokud bych znal jejich sdružené rozdělení pravděpodobnosti.

Při tvorbě modelů typicky více náhodných veličin máme. Jsou to např. neznámé parametry nebo třeba jeden či více výstupů, které chceme předpovídat.

Ideální by tedy bylo mít sdružené rozdělení pravděpodobnosti všech těchto veličin. Z tohoto rozdělení pak můžeme získat různá marginální rozdělení, dělat intervalové nebo bodové odhady.

Jak získat sdružené rozdělení pravděpodobnosti?

Naštěstí je situace velmi jednoduchá a přehledná. Pomocí Bayesovy věty jsme si odvodili, že máme vzít model ve tvaru hustoty pravděpodobnosti (nebo ve tvaru pravděpodobnostní funkce pro diskrétní náhodnou veličinu) a dosazovat do něj data. Dosadíme všechna data (a případně všechny parametry, které známe) a ze všech těchto dosazených modelů uděláme součin. Pokud chceme odhadovat i několik budoucích výstupů, zahrneme do našeho součinu i modely pro tato virtuální data, která samozřejmě nedosadíme a tyto odhadované nedosazené veličiny se nám v našem součinu objeví jako nové neznámé.

Když si vzpomeneme na bodový odhad metodou maximální věrohodnosti, tak zjistíme, že náš součin modelů je vlastně věrohodnostní funkcí z této metody.

Naši věrohodnostní funkci ještě případně vynásobíme apriorním sdruženým rozdělením pravděpodobnosti, které nese tu informaci, která nepochází z našich dat. Tím získáme aposteriorní sdružené rozdělení pravděpodobnosti, tedy přesně to, co potřebujeme. [2]

Příklad

Máme jednoduchý exponenciální model s neznámým parametrem D:

$$f\left(y_n|D\right) = \frac{1}{D} \cdot e^{-\frac{y_n}{D}}.$$

[1] Pojem rozdělení pravděpodobnosti budeme používat pro diskrétní i spojité náhodné veličiny. Pro diskrétní pak mohou použít termín pravděpodobnostní funkce, pro spojité termín hustota pravděpodobnosti.

[2] Slovo apriorní znamená, že jsme PŘED naším měřením, slovo aposteriorní znamená, že jsme PO našem měření. Pokud řekneme, že aposteriorní rozdělení pravděpodobnosti je podmíněné na základě dat, tak tím zdůrazňujeme, že informace z těchto dat je v něm zahrnuta.

Skutečné aposteriorní rozdělení je pouze úměrné našemu výrazu přes nějakou normalizační konstantu. Tu (pokud ji budeme potřebovat) zvolíme tak, aby integrál (či suma) přes celý definiční obor byla 1.

Máme data pro y: 1, 2, 3.
V jisté vědecké práci určovali hodnotu D a vyšlo jim gaussovské rozdělení $N(2; 25)$.
Chceme odhadnout parametr D a příští dva výstupy.

---

1) Určíme apriorní hustotu pro D:

$$f(D) = \frac{1}{\sqrt{2\pi} \cdot 5} \cdot e^{-\frac{(D-2)^2}{50}}.$$

2) Do modelů dosadíme příslušná y. Dosadit můžeme už pro y_1. Protože chceme odhady i pro y_4 a y_5, uděláme modely i pro ně. Všechny tyto modely vynásobíme (včetně apriorní hustoty) a tím získáme aposteriorní hustotu pravděpodobnosti pro D, y_4 a y_5:

$$f(D, y_4, y_5 | y_1, y_2, y_3) = \frac{1}{\sqrt{2\pi} \cdot 5} \cdot e^{-\frac{(D-2)^2}{50}} \cdot \frac{1}{D} \cdot e^{-\frac{1}{D}} \cdot \frac{1}{D} \cdot e^{-\frac{2}{D}} \cdot \frac{1}{D} \cdot e^{-\frac{3}{D}} \cdot \frac{1}{D} \cdot e^{-\frac{y_4}{D}} \cdot \frac{1}{D} \cdot e^{-\frac{y_5}{D}}.$$

Tvorba modelu ve tvaru rozdělení pravděpodobnosti

Lineární regresní model s gaussovským šumem

Vyjdeme z modelu ve tvaru rovnice:

$$\begin{aligned} y_n &= b_0 \cdot x_n + a_1 \cdot y_{n-1} + b_1 \cdot x_{n-1} + \ldots + k + e_n = \\ &= \theta \cdot \psi' + e_n, \end{aligned}$$

kde e_n má rozdělení $e_n \sim N\left(0, \sigma^2\right)$ a θ je řádkový vektor parametrů $\theta = (b_0, a_1, b_1, \ldots k)$a ψ je řádkový vektor vstupů $\psi = (x_n, y_{n-1}, x_{n-1}, \ldots 1)$.

Z rovnice vyjádříme e_n. Druhá strana rovnice musí mít tudíž stejné rozdělení:

$$y_n - b_0 \cdot x_n - a_1 \cdot y_{n-1} - b_1 \cdot x_{n-1} - \ldots - k \sim N\left(0, \sigma^2\right).$$

Dosadíme do vzorce pro normální rozdělení:

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

Příklad

Máme lineární regresní model ve tvaru rovnice: $y_n = A \cdot y_{n-1} + K + e_n$, kde $e_n \sim N\left(0, \sigma^2\right)$. V tomto modelu jsou neznámé parametry A, K, σ. Napište model ve tvaru hustoty pravděpodobnosti.

Model ve tvaru hustoty pravděpodobnosti je tento:

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

Diskrétní model

Vyjdeme z modelu ve tvaru tabulky podmíněných pravděpodobností. Například tohoto:

y_{n-1}	$P(y_n = 1)$	$P(y_n = 2)$
1	$\theta_{1\|1}$	$\theta_{2\|1}$
2	$\theta_{1\|2}$	$\theta_{2\|2}$

Tento model má čtyři neznámé parametry $\theta_{i|j}$. Tyto parametry mají význam pravděpodobnosti, že $y_n = i$, za podmínky, že předchozí výstup byl $y_{n-1} = j$.

Součet pravděpodobností na každém řádku je 1, proto má tento model nezávislé parametry jen dva.

Nyní hledáme takovou funkci, která nám pro pravděpodobnost, že padne i, pokud předtím padla hodnota j, dá právě $\theta_{i|j}$. To můžeme jednoduše napsat takto:

$$f(y_n|\theta, \psi) = \theta_{y_n|y_{n-1}}$$

nebo o něco složitěji takto:

$$f(y_n|\theta, \psi) = \prod_{\forall i,j} \theta_{i|j}^{\delta(y_n|y_{n-1},\, i|j)}.$$

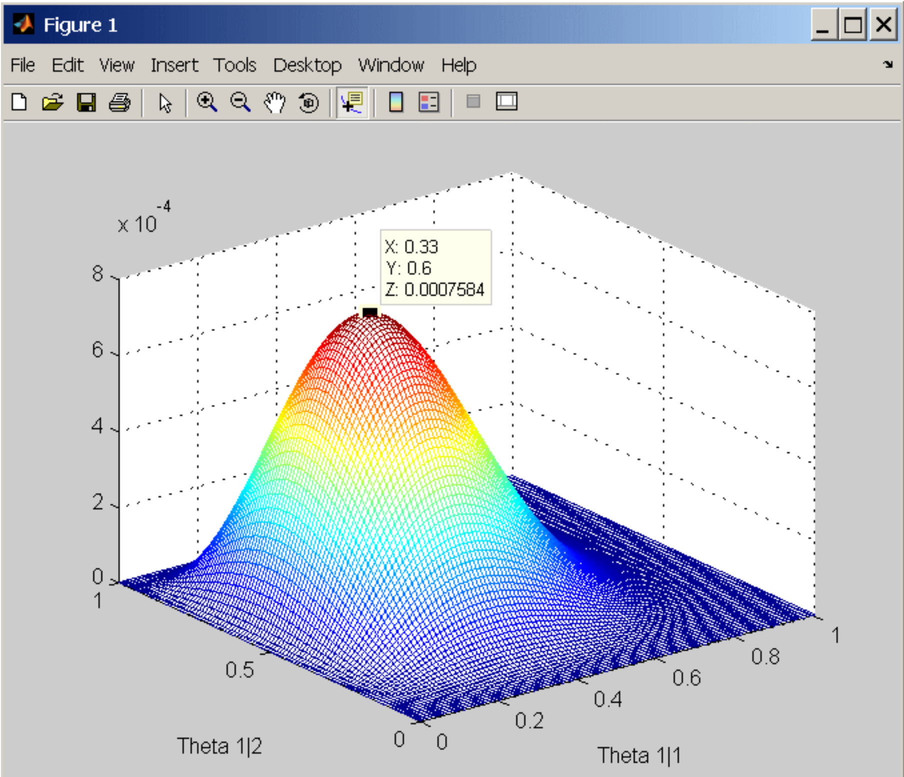
V druhém případě se jedná o součin všech $\theta_{i|j}$, ovšem s exponentem 1, pokud souslednost j, i nastala, nebo s exponentem 0, pokud souslednost j, i nenastala. Exponent se technicky vyřeší pomocí Kroneckerova delta $\delta(y_n|y_{n-1},\, i|j)$, které dává 1, pokud opravdu $y_n|y_{n-1} = i|j$, jinak dává nulu.

Příklad

Spočtěte věrohodnostní funkci pro výše uvedený diskrétní model, pokud jsme naměřili následující posloupnost y: 1, 1, 2, 1, 2, 2, 2, 1, 2, 1, 1, 2. Vykreslete věrohodnostní funkci pro nezávislé parametry $\theta_{1|1}$ a $\theta_{1|2}$.

$$\begin{aligned}
f\left(\theta_{1|1}, \theta_{1|2}\right) &= \prod_{\forall i,j} \theta_{i|j}^{\delta(1|1,\, i|j)} \cdot \prod_{\forall i,j} \theta_{i|j}^{\delta(2|1,\, i|j)} \cdot \prod_{\forall i,j} \theta_{i|j}^{\delta(1|2,\, i|j)} \cdot \ldots = \\
&= \theta_{1|1} \cdot \theta_{2|1} \cdot \theta_{1|2} \cdot \theta_{2|1} \cdot \theta_{2|2} \cdot \theta_{2|2} \cdot \theta_{1|2} \cdot \theta_{2|1} \cdot \theta_{1|2} \cdot \theta_{1|1} \cdot \theta_{2|1} = \\
&= \theta_{1|1}^2 \cdot \theta_{2|1}^4 \cdot \theta_{1|2}^3 \cdot \theta_{2|2}^2 = \\
&= \theta_{1|1}^2 \cdot \left(1 - \theta_{1|1}\right)^4 \cdot \theta_{1|2}^3 \cdot \left(1 - \theta_{1|2}\right)^2.
\end{aligned}$$

Graficky tato věrohodnostní funkce pro dvě neznámé $\theta_{1|1}$ a $\theta_{1|2}$ vypadá takto:



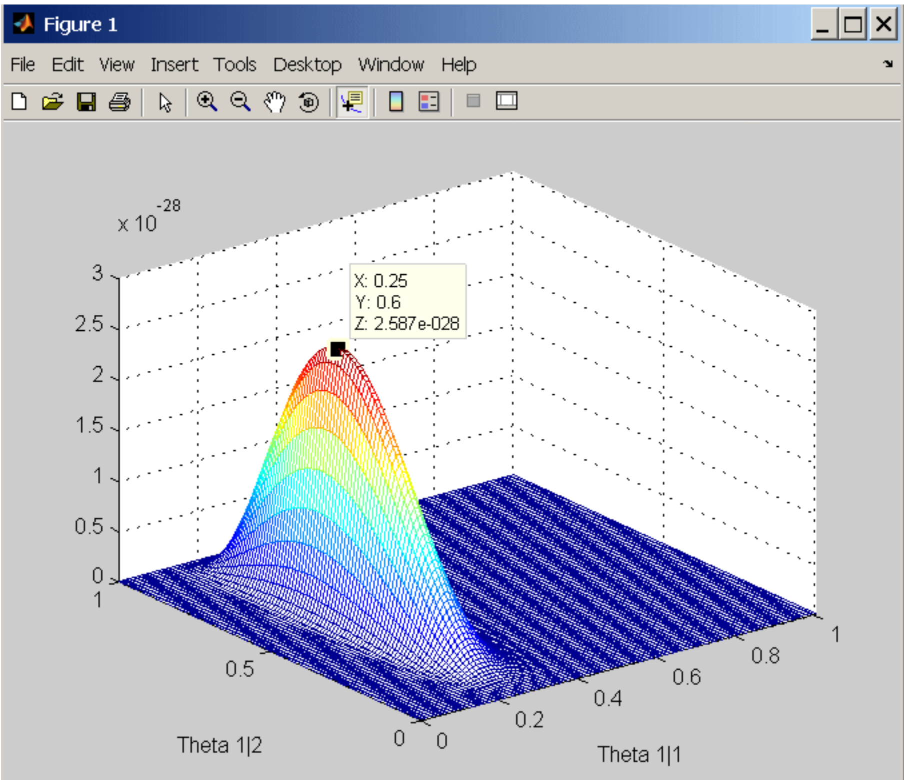
Komentář k příkladu

1. Vidíme, že pík grafu je hodně široký a správné hodnoty $\theta_{1|1}$ a $\theta_{1|2}$ jsou odhadovány velmi nepřesně. To souvisí s malým počtem dat.

2. Všimněte si, že exponenty ve věrohodnostní funkci odpovídají četnosti, kolikrát ten či onen případ nastal.

3. Všimněte si, že maximum věrohodnostní funkce je ve $\frac{2}{6}$ na první ose a $\frac{3}{5}$ na druhé ose a že to odpovídá poměru případů na jednotlivých řádcích tabulky.

4. Pokud nemáme žádnou apriorní informaci, můžeme věrohodnostní funkci považovat přímo za sdružené rozdělení pravděpodobnosti. Pokud naopak chceme apriorní informaci vložit, často se to dělá tak, že vytvořím apriorní rozdělení pravděpodobnosti, jako kdybych měl jakási virtuální data. Pokud jsem např. přesvědčen, že po 1 následuje jednička ve čtvrtině případů (tedy $\theta_{1|1} = \frac{1}{4}$) a toto mé přesvědčení má sílu, jako kdybych pozoroval stovku případů, vytvořím apriorní funkci takto: $f\left(\theta_{1|1}, \theta_{1|2}\right) = \theta_{1|1}^{25} \cdot \left(1 - \theta_{1|1}\right)^{75}$. Celkové rozdělení pravděpodobnosti bude tedy vypadat takto: $f\left(\theta_{1|1}, \theta_{1|2}\right) = \theta_{1|1}^{27} \cdot \left(1 - \theta_{1|1}\right)^{79} \cdot \theta_{1|2}^{3} \cdot \left(1 - \theta_{1|2}\right)^{2}$. To v grafu vypadá takto:



Vidíme, že odhad $\theta_{1|1}$, kterého se týkala apriorní informace, je mnohem přesnější, a že odpovídá oné apriorní $\frac{1}{4}$. Těch několik našich dat s tímto odhadem téměř nehlo. Odhad $\theta_{1|2}$, kterého se apriorní informace netýkala, zůstává stejně neurčitý jako v předchozím případě.

Logistický model

V logistickém modelu konstruujeme veličinu z, která bývá často lineární. Ve vzorci pro tuto veličinu vystupují neznámé parametry, které bychom rádi odhadli.

Pokud pro nějaké z nastane úspěch, pravděpodobnost tohoto úspěchu udává logistická funkce:

$$P\left(y_n = 1|\theta, \psi\right) = \frac{e^z}{1+e^z}.$$

Pravděpodobnost neúspěchu udává vzorec:

$$P\left(y_n = 0|\theta, \psi\right) = \frac{1}{1+e^z}.$$

Oba tyto vzorce mohu shrnout do:

$$f\left(y_n|\theta, \psi\right) = \frac{e^{z \cdot y_n}}{1+e^z}.$$

To je logistický model ve tvaru pravděpodobnostní funkce.

Příklad

Logistickým modelem modelujeme stihnutí/nestihnutí určité trasy v daném termínu. Odhadněte parametry tohoto modelu, pokud máte tato data:

čas k dispozici	15	20	21	23	25	27	29	32	35	40
stihnutí	0	0	0	1	0	1	0	1	1	1

Jakou zvolíme strukturu modelu? Doba t, kterou potřebujeme k ujetí trasy bude mít nějaké cca gaussovské rozdělení se střední hodnotou μ a rozptylem σ^2. Tedy Veličina $t-\mu$ bude mít nulovou střední hodnotu a po pronásobení nějakou konstantou dostanu i správný rozptyl pro logistické rozdělení (které použiji místo Gaussova). Tedy veličinu z volím: $z = k\cdot(t-\mu) = At+B$. Parametry A a B se pokusím odhadnout.

Nemám žádnou apriorní informaci, vyrobím tedy věrohodnostní funkci a to už bude mé aposteriorní sdružené rozdělení pravděpodobnosti pro parametry A a B.

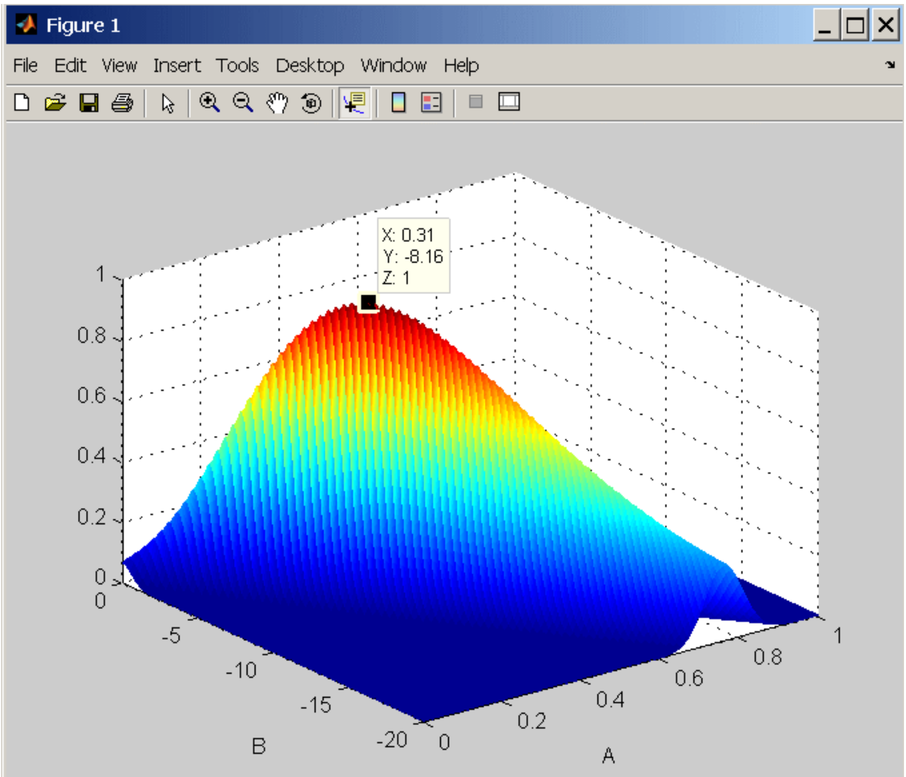
Model mám tento:

$$f\left(y_n|\theta,\psi\right) = \frac{e^{(At+B)\cdot y_n}}{1+e^{At+B}}.$$

Dosadím data do věrohodnostní funkce:

$$\begin{aligned} f\left(A,B\right) \quad = \quad & \frac{e^{(15A+B)\cdot 0}}{1+e^{15A+B}}\cdot\frac{e^{(20A+B)\cdot 0}}{1+e^{20A+B}}\cdot\frac{e^{(21A+B)\cdot 0}}{1+e^{21A+B}}\cdot\frac{e^{(23A+B)\cdot 1}}{1+e^{23A+B}}\cdot\frac{e^{(25A+B)\cdot 0}}{1+e^{25A+B}}\cdot \\ & \cdot\frac{e^{(27A+B)\cdot 1}}{1+e^{27A+B}}\cdot\frac{e^{(29A+B)\cdot 0}}{1+e^{29A+B}}\cdot\frac{e^{(32A+B)\cdot 1}}{1+e^{32A+B}}\cdot\frac{e^{(35A+B)\cdot 1}}{1+e^{35A+B}}\cdot\frac{e^{(40+B)\cdot 1}}{1+e^{40+B}}. \end{aligned}$$

Graf této funkce vypadá takto:



Vidíme, že kvůli malému počtu dat je odhad ještě značně neurčitý. Bodový odhad pro konstanty A a B získáváme $A = 0,31$ a $B = -8,16$.