Matematika III – 14. týden
Bayesovská interpretace

Jan Slovák

Masarykova univerzita
Fakulta informatiky

15. – 19.12. 2014

Obsah přednášky

Plán přednášky

Kde je dobré číst?

- Karel Zvára, Josef Štěpán, Pravděpodobnost a matematická pravděpodobnost statistika, Matfyzpress, 2006, 230pp.
- J. Slovák, M. Panák, M. Bulant, Matematika drsně a svižně, Muni Press, Brno 2013, v+773 s., elektronická edice `www.math.muni.cz/Matematika_drsne_svizne`
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, Teorie pravděpodobnosti a matematická statistika (sbírka příkladů), Masarykova univerzita, 3. vydání, 2004, 117 stran, ISBN 80-210-3313-4.
- Marie Budíková, Tomáš Lerch, Štěpán Mikoláš, Základní statistické metody, Masarykova univerzita, 2005, 170 stran, ISBN 80-210-3886-1.
- Riley, K.F., Hobson, M.P., Bence, S.J. Mathematical Methods for Physics and Engineering, second edition, Cambridge University Press, Cambridge 2004, ISBN 0 521 89067 5, xxiii + 1232 pp.

Plán přednášky

Bayesův vzorec pro podmíněnou pravděpodobnost (tzv. inverzní pravděpodobnost):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Bayesův vzorec pro podmíněnou pravděpodobnost (tzv. inverzní pravděpodobnost):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Na úrovni hustot (nebo pravděpodobnostních funkcí) náhodných veličin: máli vektor (X, Θ) hustotu $f(x|\theta)$, pak podmíněná pravděpodobnost komponenty Θ za podmínky $X = x$ hustotu $g(\theta|x)$ danou

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}.$$

Bayesův vzorec pro podmíněnou pravděpodobnost (tzv. inverzní pravděpodobnost):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Na úrovni hustot (nebo pravděpodobnostních funkcí) náhodných veličin: máli vektor (X, Θ) hustotu $f(x|\theta)$, pak podmíněná pravděpodobnost komponenty Θ za podmínky $X = x$ hustotu $g(\theta|x)$ danou

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}.$$

Mluvíme o **apriorní hustotě** $g(\theta)$ a **aposteriorní hustotě** $g(\theta|x)$.

Bayesův vzorec pro podmíněnou pravděpodobnost (tzv. inverzní pravděpodobnost):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Na úrovni hustot (nebo pravděpodobnostních funkcí) náhodných veličin: máli vektor (X, Θ) hustotu $f(x|\theta)$, pak podmíněná pravděpodobnost komponenty Θ za podmínky $X = x$ hustotu $g(\theta|x)$ danou

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}.$$

Mluvíme o **apriorní hustotě** $g(\theta)$ a **aposteriorní hustotě** $g(\theta|x)$. Protože předem víme, že $g(\theta|x)$ je hustota pravděpodobnosti, nemusí nás konstanta $f(x)$ vůbec zajímat — počítáme prostě až na násobek konstantou.

Předpokládejme, že na univerzitě je spokojenost studentů v jednotlivých předmětech náhodná veličina $X \sim \mathrm{N}(\theta, \sigma^2)$, zatímco parametr θ dosahovaný jednotlivými učiteli je náhodná veličina $\theta \sim \mathrm{N}(a, b)$.
Můžeme tedy počítat (pořád až na konstatní násobky, tj. ignorujeme součinitele, ve kterých nevystupuje θ) a dostaneme

$$\theta \sim \mathrm{N}\left(\frac{b^2}{b^2+\sigma^2}x + \frac{\sigma^2}{b^2+\sigma^2}a, \frac{b^2\sigma^2}{b^2+\sigma^2}\right).$$

Předpokládejme, že na univerzitě je spokojenost studentů v jednotlivých předmětech náhodná veličina $X \sim \mathrm{N}(\theta, \sigma^2)$, zatímco parametr θ dosahovaný jednotlivými učiteli je náhodná veličina $\theta \sim \mathrm{N}(a, b)$.
Můžeme tedy počítat (pořád až na konstatní násobky, tj. ignorujeme součinitele, ve kterých nevystupuje θ) a dostaneme

$$\theta \sim \mathrm{N}\left(\frac{b^2}{b^2+\sigma^2}x + \frac{\sigma^2}{b^2+\sigma^2}a, \frac{b^2\sigma^2}{b^2+\sigma^2}\right).$$

Když tedy z dlouhodobého vyhodnocování anket známe parametry a, b, σ, můžeme po vyjádření nějakého studenta upřesnit apriorní představu o parametrech pro jeden konkrétní předmět. Ve výsledném odhadu rozložení je pak střední hodnota dána váženým průměrem zjištěné hodnoty x a apriorně předpokládané střední hodnoty a, v závislosti na rozptylech σ a b.

Bayesovská interpretace?

Pro $\sigma \to 0$ je váha jediného názoru stále rostoucí a tomu odpovídá 100% váha u x v případdě $\sigma = 0$. Je to plně v souladu s interpretací, že Bayesovská statistika je pravděpodobnostní rozšíření standardní diskrétní matematické logiky.

Bayesovská interpretace?

Pro $\sigma \to 0$ je váha jediného názoru stále rostoucí a tomu odpovídá 100% váha u x v případdě $\sigma = 0$. Je to plně v souladu s interpretací, že Bayesovská statistika je pravděpodobnostní rozšíření standardní diskrétní matematické logiky.
Místo jednoho studenta použijeme výběrový průměr $\bar{X}$ výsledku šetření. Opět jde o normální rozdělení, jen budeme místo σ^2 dosazovat σ^2/n. Pišme

$$c_n = \frac{nb^2}{nb^2 + \sigma^2}$$

a aposteriorní odhad pro θ je

$$\theta \sim \mathrm{N}(c_n \bar{X} + (1 - c_n)a, c_n \sigma^2/n).$$

Bayesovská interpretace?

Pro $\sigma \to 0$ je váha jediného názoru stále rostoucí a tomu odpovídá 100% váha u x v připaddě $\sigma = 0$. Je to plně v souladu s interpretací, že Bayesovská statistika je pravděpodobnostní rozšíření standardní diskrétní matematické logiky.
Místo jednoho studenta použijeme výběrový průměr $\bar{X}$ výsledku šetření. Opět jde o normální rozdělení, jen budeme místo σ^2 dosazovat σ^2/n. Pišme

$$c_n = \frac{nb^2}{nb^2 + \sigma^2}$$

a aposteriorní odhad pro θ je

$$\theta \sim \mathrm{N}(c_n\bar{X} + (1 - c_n)a, c_n\sigma^2/n).$$

Pro rostoucí n se bude střední hodnota našeho rozdělení pro θ stále více blížit výběrovému průměru a jeho rozptyl půjde k nule. Čím je tedy n větší, tím více se blížíme bodovému odhadu z frekventistického přístupu.

Přínosem Bayesovského přístupu je, že s použitím odhadnutého rozdělení můžeme odpovídat na dotazy typu „s jakou pravděpodobností je nový vyučující horší než předchozí?“ Potřebujeme k tomu apriorní údaje.
Předpokládejme, že máme docela dobře hodnocené učitele: $a = 7{,}5$, $b = 2{,}5$ a a ponecháme směrodatnou odchylku $\sigma = 2$. Pro $n = 15$ a výběrový průměr 5,133 dostaneme aposteriorní odhad pro rozdělení $\theta \sim \mathrm{N}(5{,}230,\ 0{,}256)$.

Přínosem Bayesovského přístupu je, že s použitím odhadnutého rozdělení můžeme odpovídat na dotazy typu „s jakou pravděpodobností je nový vyučující horší než předchozí?“
Potřebujeme k tomu apriorní údaje.
Předpokládejme, že máme docela dobře hodnocené učitele: $a = 7{,}5$, $b = 2{,}5$ a a ponecháme směrodatnou odchylku $\sigma = 2$. Pro $n = 15$ a výběrový průměr 5,133 dostaneme aposteriorní odhad pro rozdělení $\theta \sim \mathrm{N}(5{,}230,\ 0{,}256)$.
Zajímá nás $P(\theta < 6)$. Odpověď získáme dotazem na hodnotu distribuční funkce příslušného normálního rozdělení pro argument 6 – odpověď je cca $93,6\%$. Je tedy podobná, jako jsme viděli v frekventistickém přístupu.