Bayesovské metódy
(seminár)

Doporučená literatúra:

Anděl, J., Matematická statistika, SNTL/ALFA, Praha, 1985.

Grendár, M., Bayesovská štatistika (http://www.savbb.sk/ grendar/pdf)

Hušková, M., BAYESOVSKÉ METODY, UNIVERZITA KARLOVA, Praha, 1985.
(http://www.karlin.mff.cuni.cz/ huskova)

Pázman, A., BAYESOVSKÁ ŠTATISTIKA, Univerzita Komenského, Bratislava, 2003.

Tento text v zásade sleduje knižku A.Pázmana.

1. Úvod a motivačný príklad

Definícia 1.1. Majme pravdepodobnostný priestor $(\Omega, \mathcal{A}, P)$. Náhodné javy $A_1, A_2, ... \in \mathcal{A}$ tvoria *úplný systém javov*, ak platí

$$A_i \cap A_j = \emptyset, \quad i \neq j, \quad \text{a} \quad \bigcup_{i=1}^{\infty} A_i = \Omega. \tag{1.1}$$

Poznámka. Úplný systém javov môže byť aj konečný.

Veta 1.1. (Vzorec pre úplnú pravdepodobnosť). Nech $A_1, A_2, ...$ je úplný systém javov v pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$ taký, že

$$P(A_i) > 0, \quad i = 1, 2, ... \tag{1.2}$$

Potom platí

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i). \tag{1.3}$$

Dôkaz:

$P(B) = P(B \cap \Omega) = P\left(B \cap \bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty}(B \cap A_i)\right) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$

♣

Veta 1.2. (1. Bayesov vzorec). Nech $A_1, A_2, ...$ je úplný systém javov v pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$ taký, že

$$P(A_i) > 0, \quad i = 1, 2, ...$$

Ak $P(B) > 0$, tak platí

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}, \quad j = 1, 2, ... \tag{1.4}$$

Dôkaz:

Pre ľubovoľné j je

$$P(A_j|B) = \frac{P(B \cap A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}. \quad \clubsuit$$

Veta 1.3. (2. Bayesov vzorec). Nech $A_1, A_2, ...$ je úplný systém javov v pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$ taký, že $P(A_i) > 0, \quad i = 1, 2, ...$. ďalej $A \in \mathcal{A}$, že $P(A) > 0$ a $B \in \mathcal{A}$. Platí

$$P(B|A) = \frac{\sum_{\{i:\ P(A \cap A_i) > 0\}} P(A_i)P(A|A_i)P(B|A \cap A_i)}{\sum_{i=1}^{\infty} P(A|A_i)P(A_i)}. \tag{1.5}$$

Dôkaz: spravte si sami.

Poznámka. Vety 1.1, 1.2 a 1.3 platia aj v prípade, že úplny systém javov je konečný.

Poznámka. $P(A_j)$ v Bayesových vzorcoch sú tzv. *apriórne pravdepodobnosti* a $P(A_j|B)$ *aposteriórne pravdepodobnosti* (po vykonaní pokusu s výsledkom B).

Poznámka. V prípade 1. Bayesovho vzorca ide o riešenie situácie, keď máme hypotézy $A_1, ...$, ktoré sa navzájom vylučujú, ale vyčerpávajú všetky možnosti. Poznáme ich (apriorné) pravdepodobnosti $P(A_i)$. Nastal jav A a poznáme pravdepodobnosti $P(A|A_i)$. Pýtame sa na (aposteriórne; nové, ktoré berú do úvahy skutočnosť, že mastal A) pravdepodobnosti $P(A_i|A)$

V prípade 2. Bayesovho vzorca ak nastal jav A, pýtame sa na pravdepodobnosť javu B.

Poznámka. Nie je vždy jednoduché voliť správny pravdepodobnostný model pre výpočet podmienených pravdepodobností.

Príklad 1.1. (Lekárska diagnostika). Vieme, že určitou (konkrétnou) chorobou *Ch* trpí 1% populácie. Choroba je diagnostikovaná na základe vyšetrenia, ktorého spoľahlivosť je

(i) 95% ak vyšetrovaná osoba trpí chorobou *Ch*

(ii) 70 % ak vyšetrovaná osoba netrpí chorobou *Ch*.

Vyšetrujeme náhodne zvolenú osobu. Určte pravdepodobnosť správnej diagnózy, ak výsledok vyšetrenia je

(a) pozitívny (podľa výsledku vyšetrenia je osoba chorá)

(b) negatívny (podľa výsledku vyšetrenia je osoba zdravá).

Riešenie:

Označme jav

A – vyšetrovaná osoba trpí chorobou *Ch* (je chorá)

B – výsledok vyšetrovania je pozitívny

Zo zadania vieme

$P(A) = 0.01$ (pravdepodobnosť, že vybraná osoba je chorá) Táto pravdepodobnosť sa volá <u>prevalencia</u> alebo tiež apriorná pravdepodobnosť choroby

Vyšetrenie (spoľahlivosť vyšetrenia) sa charakterizuje dvomi charakteristikami, a síce

pravdepodobnosťou $P(B|A) = 0.95$ tzv. <u>citlivosť</u> testu alebo aj <u>senzitivita</u> testu

pravdepodobnosťou $P(\overline{B}|\overline{A}) = 0.7$ tzv. <u>špecificita</u> testu.

(a) Máme určiť vlastne $P(A|B)$ (lebo v tomto prípade výsledok testu bol pozitívny, teda test hovorí, že vyšetrovaná osoba je chorá (diagnóza je, že pacient je chorý) a my máme určiť pravdepodobnosť správnej dignózy).

Zo zadania vieme, že $P(A) = 0.01$, $P(\overline{A}) = 0.99$, $P(B|A) = 0.95$ a $P(B|\overline{A}) = 1 - P(\overline{B}|\overline{A}) = 1 - 0.7 = 0.3$. Podľa Bayesovho vzorca ($A, \overline{A}$ sú hypotézy)

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})} = \frac{0.01 \cdot 0.95}{0.01 \cdot 0.95 + 0.99 \cdot 0.3} = 0.030995.$$

Je to aj aposteriórna pravdepodobnosť, že pacient je chorý, ak výsledok testu bol pozitívny. Je to prekvapivý výsledok. čakali by sme "omnoho lepší" výsledok.

Celkom máme 29 700 + 950 = 30 650 pozitívnych výsledkov, z toho správne pozitívnych je 950, čiže $P(A|B) = \dfrac{950}{30650} = 0.030995$.

(b) Analogicky (zase $A, \overline{A}$ sú hypotézy)

$$P(\overline{A}|\overline{B}) = \frac{P(\overline{A})P(\overline{B}|\overline{A})}{P(\overline{A})P(\overline{B}|\overline{A}) + P(A)P(\overline{B}|A)} = \frac{0.99 \cdot 0.7}{0.99 \cdot 0.7 + 0.01 \cdot 0.05} = 0.99928.$$

Je to aposteriórna pravdepodobnosť, že pacient nie je chorý, ak výsledok testu bol negatívny. Naozaj celkovo máme 69 300 + 50 = 69 350 negatívnych výsledkov, z toho správne negatívnych je 69 300 a teda pravdepodobnosť správnej diagnózy u negatívnych výsledkov testu je $P(\overline{A}|\overline{B}) = \dfrac{69300}{69350} = 0.99928$.

2. Bayesova veta

Uveďme si základné tvrdenia.

Veta 2.1. (O podmienenej hustote). Nech λ a ν sú $\sigma-$konečné miery a nech združená hustota náhodných vektorov $\mathbf{Y} \in \mathcal{R}^r$ a $\mathbf{Z} \in \mathcal{R}^{n-r}$ vzhľadom k $\mu = \lambda \times \nu$ je $p(\mathbf{y}, \mathbf{z})$. Nech

$$q(\mathbf{y}) = \int_{\mathcal{R}^{n-r}} p(\mathbf{y}, \mathbf{z}) d\nu(\mathbf{z}) \tag{2.1}$$

je marginálna hustota náhodného vektora $\mathbf{Y}$. Potom podmienená hustota vektora $\mathbf{Z}$ pri pevnom $\mathbf{Y} = \mathbf{y}$ je rovná

$$r(\mathbf{z}|\mathbf{y}) = \begin{cases} p(\mathbf{y}|\mathbf{z})/q(\mathbf{y}), & \text{ak } \ q(\mathbf{y}) \neq 0, \\ 0, & \text{ak } \ q(\mathbf{y}) = 0. \end{cases} \tag{2.2}$$

Dôkaz: nájdete v Andělovej knižke, str. 53.

Pri našich postupoch je kľúčová

Veta 2.2. (Bayesova). Nech $q(\mathbf{y})$ je marginálna hustota náhodného vektora $\mathbf{Y}$ a $r(\mathbf{z}|\mathbf{y})$ podmienená hustota náhodného vektora $\mathbf{Z}$ pri danom $\mathbf{Y} = \mathbf{y}$. Potom podmienená hustota $s(\mathbf{y}|\mathbf{z})$ náhodného vektora $\mathbf{Y}$ pri danom $\mathbf{Z} = \mathbf{z}$ je rovná

$$s(\mathbf{y}|\mathbf{z}) = \begin{cases} \dfrac{q(\mathbf{y})r(\mathbf{z}|\mathbf{y})}{\int_{\mathcal{R}^r} q(\mathbf{y})r(\mathbf{z}|\mathbf{y})d\lambda(\mathbf{y})}, & \text{ak } \ \int_{\mathcal{R}^r} q(\mathbf{y})r(\mathbf{z}|\mathbf{y})d\lambda(\mathbf{y}) \neq 0, \\ 0, & \text{inak.} \end{cases} \tag{2.3}$$

Dôkaz: nájdete v Andělovej knižke, str. 54.

Samozrejme Vety 2.1 a 2.2 platia aj v prípade, že Y a Z sú náhodné veličiny. V týchto textoch budeme uvažovať λ a ν alebo Lebesgueovu mieru, alebo sčítaciu mieru. Teda $q(\cdot), r(\cdot|\cdot\cdot)$ a $s(\cdot)$ sú hustoty vzhľadom k týmto $\sigma-$konečným mieram.

Inferencia (usudzovanie)

Aby sme videli, ako Bayesova veta slúži pre štatistickú inferenciu, predpokladajme, že máme dvojicu náhodných vektorov $(\boldsymbol{\Theta}', \mathbf{Y}')'$, pričom $P(\boldsymbol{\Theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m) = 1,\ P(\mathbf{Y} \in \mathcal{Y} \in \mathcal{B}_n) = 1$ ($\mathcal{B}_m$ je systém borelovských podmnožín priestoru $\mathcal{R}^m$). Poznáme

a) podmienenú hustotu pravdepodobnosti vektora $\mathbf{Y}$ vzhľadom k $\sigma-$konečnej miere $\lambda(\mathbf{y})$, za predpokladu, že je daná hodnota $\boldsymbol{\theta}$ náhodného vektora $\boldsymbol{\Theta}$

$$f(\mathbf{y}|\boldsymbol{\theta}), \tag{2.4}$$

b) hustotu pravdepodobnosti náhodného vektora $\boldsymbol{\Theta}$ vzhľadom k $\sigma-$konečnej miere $\nu(\boldsymbol{\theta})$

$$\pi(\boldsymbol{\theta}). \tag{2.5}$$

Podľa Vety 2.1 združená hustota pravdepodobnosti $(\boldsymbol{\Theta}', \mathbf{Y}')'$ vzhľadom k $\mu = \nu \times \lambda$ sa rovná

$$g(\boldsymbol{\theta}, \mathbf{y}) = \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}), \tag{2.6}$$

marginálna hustota pravdepodobnosti veličiny $\mathbf{Y}$ (predikčná hustota) je

$$f^*(\mathbf{y}) = \int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})$$

a podmienená hustota $\boldsymbol{\Theta}$, za predpokladu, že je známa hodnota $\mathbf{Y} = \mathbf{y}$, sa rovná

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{g(\boldsymbol{\theta}, \mathbf{y})}{f^*(\mathbf{y})} = \frac{\pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}. \tag{2.7}$$

Pre dáta $\mathbf{y}$ máme pravdepodobnostný model, t.j. predpokladáme, "že sa riadia" hustotou $f(\mathbf{y}|\boldsymbol{\theta})$ (v prípade spojitých údajov) alebo pravdepodobnostnou funkciou (v prípade diskrétnych údajov). Obyčajne hustotu zapisujeme ako $f(\mathbf{y}; \boldsymbol{\theta})$, ale v tomto prípade hustotu zapisujeme ako $f(\mathbf{y}|\boldsymbol{\theta})$, aby sme zdôraznili, že pravdepodobnostný model je hustotou pre dáta za predpokladu, že hodnota parametra je $\boldsymbol{\theta}$. Ďalej predpokladajme, že sme schopní vyjadriť "náš názor" (naše presvedčenie) týkajúce sa parametra $\boldsymbol{\theta}$ vo forme apriórnej hustoty $\pi(\boldsymbol{\theta})$. Túto hustotu musíme vedieť "skonštruovať" separátne (inou cestou, inak) než z nameraných dát.

Poznámka. Ak $\mathbf{Y}$ má (absolútne) spojité rozdelenie (λ je Lebesgueova miera) a $\boldsymbol{\Theta}$ tiež absolútne spojité rozdelenie, tak $\pi(\boldsymbol{\theta}),\ f(\mathbf{y}|\boldsymbol{\theta})$ a $\pi(\boldsymbol{\theta}|\mathbf{y})$ sú "obyčajné" hustoty.

Ak $\mathbf{Y}$ má spojité rozdelenie a $\boldsymbol{\Theta}$ diskrétne ($\nu(\boldsymbol{\theta})$ je sčítacia miera), $\boldsymbol{\Theta} \in \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$, tak $\pi(\boldsymbol{\theta})$ je pravdepodobnostná funkcia, $f(\mathbf{y}|\boldsymbol{\theta})$ je hustota a

$$\pi(\boldsymbol{\theta}_i|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta}_i) f(\mathbf{y}|\boldsymbol{\theta}_i)}{\sum_{k\geq 1} \pi(\boldsymbol{\theta}_k) f(\mathbf{y}|\boldsymbol{\theta}_k)}, \quad i = 1, 2, ...$$

je pravdepodobnostná funkcia.

Ak $\mathbf{Y}$ má diskrétne rozdelenie a $\boldsymbol{\Theta}$ spojité, tak $\pi(\boldsymbol{\theta})$ je hustota, $f(\mathbf{y}|\boldsymbol{\theta})$ je pravdepodobnstná funkcia a $\pi(\boldsymbol{\theta}|\mathbf{y})$ je hustota.

Ak $\mathbf{Y}$ je diskrétny náhodný vektor, $\mathbf{Y} \in \{y_1, y_2, ...\}$ a $\boldsymbol{\Theta}$ tiež diskrétny náhodný vektor, $\boldsymbol{\Theta} \in \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$, tak $\pi(\boldsymbol{\theta}),\ f(\mathbf{y}|\boldsymbol{\theta})$ aj

$$\pi(\boldsymbol{\theta}_i|\mathbf{y}_j) = \frac{\pi(\boldsymbol{\theta}_i) f(\mathbf{y}_j|\boldsymbol{\theta}_i)}{\sum_{k\geq 1} \pi(\boldsymbol{\theta}_k) f(\mathbf{y}_j|\boldsymbol{\theta}_k)}, \quad i = 1, 2, ...$$

sú pravdepodobnostné funkcie.

Ešte iný pohľad na Bayesov vzorec. V prípade diskrétnych náhodných premenných Θ a Y nech $A_1, ..., A_m$ je úplný systém borelovských množín na reálnej osi a $B_1, ..., B_n$ nejaké borelovské množiny. Nech sú známe podmienené pravdepodobnosti $P(B_i|A_k) = P(Y \in B_i|\Theta \in A_k)$ a tzv. apriórne pravdepodobnosti $P(A_k) = P(\Theta \in A_k)$. $P(A_k)$ vyjadrujú "očakávania" javov A_k ešte pred pozorovaním javov B_i. Vplyvom pozorovania javu B_i sa pravdepodobnosť "očakávania" $P(A_k)$ zmení na podmienenú pravdepodobnosť $P(A_k|B_i)$ určenú Bayesovým vzorcom (1.4) (alebo (2.7))

$$P(A_k|B_i) = \frac{P(B_i|A_k)P(A_k)}{\sum_{j=1}^{m} P(B_i|A_j)P(A_j)}.$$

Vzorce (2.4)–(2.7) samozrejme platia ak Θ a Y sú náhodné veličiny.

3. Príklad
(podľa Anděl, str. 279)

Výroba prebieha každý deň. pravdepodobnosť vyrobenia chybnej súčiastky je p. Predpokladajme, že táto pravdepodobnosť sa zo dňa na deň (mierne) mení a rozdelenie pravdepodobnosti veličiny p sa dá aproximovať beta rozdelením s hustotou

$$\pi(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1, \tag{3.1}$$

kde $a > 0$, $b > 0$ sú parametre, $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$ je beta funkcia. Keď výroba prebieha "ustálene", môžeme "dostatočne presne" odhadnúť hodnoty parametrov a tieto odhady považovať za skutočné hodnoty a a b.

Ak teraz spravíme pokus - náhodne vyberieme n výrobkov (s vrátením), pričom m z nich bude chybných, pravdepodobnosť vybratia práve m chybných súčiastok (ak je pravdepodobnosť vyrobenia chybnej súčiastky rovná p) je

$$f(m|p) = \binom{n}{m} p^m (1-p)^{n-m}, \quad m = 0, 1, ..., n. \tag{3.2}$$

Podľa (2.7) je aposteriórna hustota $\pi(p|m)$ náhodného parametra p pri zistenom počte m chybných súčiastok z n vybraných súčiastok rovná

$$\pi(p|m) = \frac{\pi(p)f(y|p)}{\int_{-\infty}^{\infty} \pi(p)f(y|p)dp} = \frac{\frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\binom{n}{m}p^m(1-p)^{n-m}}{\int_0^1 \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\binom{n}{m}p^m(1-p)^{n-m}dp} = \tag{3.3}$$

$$= \frac{1}{B(a+m, b+n-m)} p^{a+m-1}(1-p)^{b+n-m-1}, \quad 0 < p < 1.$$

4. Vlastnosti Bayesovho vzorca

Ak by sme vo vzorci (3.1) namiesto hustoty $\pi(p)$ uvažovali funkciu

$$\pi^*(p) = p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1,$$

(toto už nie je hustota), aposteriórna hustota (3.3) sa nezmení. Dokonca ak predpokladáme $a \to 0^+$, $b \to 0^+$, tak

$$\pi^*(p) \to \pi^{**}(p) = p^{-1}(1-p)^{-1}, \quad 0 < p < 1.$$

Funkcia $\pi^{**}(p)$ je nezáporná a platí pre ňu

$$\int_0^1 \pi^{**}(p)dp = \infty.$$

Neexistuje však žiadna normujúca konštanta k, aby $k\pi^{**}(p)$ bola hustotou. Ak ale platí $0 < m < n$, tak pre $a \to 0^+,\ b \to 0^+$ dostávame

$$\pi(p|m) \to \pi^{**}(p|m) = \frac{1}{B(m, n-m)} p^{a+m-1}(1-p)^{b+n-m-1}. \tag{4.1}$$

Tento výsledok by sme dostali aj priamo dosadením funkcie π^{**} za apriórnu hustotu. Ak ak teda platí $0 < m < n$, je $\pi^{**}(p|m)$ "obyčajná" hustota, ak ale $m = 0$ alebo $m = n$, nedostaneme konečnú limitu funkcie $\pi(p|m)$. Nezápornú merateľnú funkciu voláme *nevlastná hustota*.

Ďalšie vlastnosti Bayesovho vzorca.

Ak pre každé dve rôzne $\mathbf{y}_1, \mathbf{y}_2$ v (2.4) platí, že podiel $\dfrac{f(\mathbf{y}_1|\boldsymbol{\theta})}{f(\mathbf{y}_2|\boldsymbol{\theta})}$ nezávisí od $\boldsymbol{\theta}$, tak $\pi(\boldsymbol{\theta}|\mathbf{y}_1) = \pi(\boldsymbol{\theta}|\mathbf{y}_2)$ pre každé $\boldsymbol{\theta}$. Z bayesovského hľadiska sú výsledky experimentov $\mathbf{y}_1$ a $\mathbf{y}_2$ "informačne ekvivalentné".

V bayesovskom prístupe dôležitú úlohu má funkcia vierohodnosti

$$l_{\mathbf{y}}(\boldsymbol{\theta}) \propto f(\mathbf{y}|\boldsymbol{\theta}). \tag{4.2}$$

Je to trieda všetkých reálnych funkcií premennej $\boldsymbol{\theta}$, líšiacich sa od $f(\mathbf{y}|\boldsymbol{\theta})$ len o multiplikatívny člen nezávislý od $\boldsymbol{\theta}$. Z Bayesovho vzorca vyplýva, že štatistické postupy založené na aposteroórnej pravdepodobnosti spĺňajú tzv. *princíp vierohodnosti*:

Ak $\mathbf{y}, \mathbf{y}^*$ sú dva výsledky rôznych experimentov s tým istým parametrickým priestorom $\boldsymbol{\Omega}$ a ak $l_{\mathbf{y}}(\boldsymbol{\theta}) = l_{\mathbf{y}^*}(\theta)$, tak výsledky $\mathbf{y}$ a $\mathbf{y}^*$ vedú k tým istým štatistickým záverom o parametri $\boldsymbol{\theta}$.

5. Pravidlo reťazenia pre nezávislé pozorovania

Uvažujme dva nezávislé experimenty s tým istým parametrickým priestorom $\boldsymbol{\Omega}$. Experiment I je daný systémom podmienených hustôt

$$\{f(\mathbf{y}|\boldsymbol{\theta}) :\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\}, \tag{5.1}$$

v ktorom pozorujeme $\mathbf{y} \in \mathcal{Y}$ a experiment II nezávislý na experimente I, daný systémom podmienených hustôt

$$\{g(\mathbf{z}|\boldsymbol{\theta}) :\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\},$$

v ktorom pozorujeme $\mathbf{z} \in \mathcal{Z}$. V kombinovanom experimente II, v ktorom pozorujeme (nezávislú) dvojicu $(\mathbf{y}, \mathbf{z}) \in \mathcal{Y} \times \mathcal{Z}$, daným systémom podmienených hustôt

$$\{h(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{z}|\boldsymbol{\theta}) :\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\}$$

pri apriórnej hustote $\pi(\boldsymbol{\theta})$ dostávame z (2.7) aposteriórnu hustotu

$$\frac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{z}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{z}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}. \tag{5.2}$$

Teraz realizujme experiment II, ale za apriórnu hustotu neuvažujme $\pi(\boldsymbol{\theta})$, ale aposteriórnu hustotu z experimentu I, teda

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}.$$

Podľa (2.7) dostávame aposteriórnu hustotu

$$\text{(5.3)} \qquad \frac{[\pi(\boldsymbol{\theta}|\mathbf{y})]g(\mathbf{z}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} [\pi(\boldsymbol{\theta}|\mathbf{y})]\, g(\mathbf{z}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})} = \frac{\left[\dfrac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}\right] g(\mathbf{z}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \left[\dfrac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}\right] g(\mathbf{z}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{z}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{z}|\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}.$$

Hustoty (5.2) a (5.3) sú zhodné.

Toto pravidlo môžeme rozšíriť na ľubovolnú postupnosť nezávislých experimentov. *Aposteriórnu hustotu získanú z prvých k experimentov možno použiť ako apriórnu hustotu v ďalšom $(k+1)-$vom experimente.*

6. Bayesov vzorec a postačujúce štatistiky

Nech $\mathbf{Y} \in \mathcal{Y} \in \mathcal{B}_n$, $\boldsymbol{\Theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m$ a $\boldsymbol{\tau}$ je merateľné zobrazenie $\boldsymbol{\tau} : \mathcal{R}^n \to \mathcal{R}^k$. Takéto zobrazenie voláme *štatistika*. (Poznamenávame len, že ak $m = k$, štatistika je odhadom.) Štatistiku voláme *postačujúcou* ak podmienené hustoty pre $\mathbf{y}$, za podmienky, že $\boldsymbol{\tau}(\mathbf{y}) = \mathbf{t}$, nezávisia od $\boldsymbol{\theta}$. Z učebníc matematickej štatistiky (pozri napr. Anděl, str. 262) je známe, že štatistika $\boldsymbol{\tau}$ je postačujúca práve vtedy, ak existujú reálne (a merateľné) funkcie $h(\mathbf{y})$, $q(\mathbf{t}, \boldsymbol{\theta})$ také, že

$$\text{(6.1)} \qquad f(\mathbf{y}|\boldsymbol{\theta}) = h(\mathbf{y})q(\boldsymbol{\tau}(\mathbf{y}), \boldsymbol{\theta})$$

pre každé $\mathbf{y}, \boldsymbol{\theta}$ (veta o faktorizácii). Pripomíname len, že ak $\mathbf{Y}$ je spojitá náhodná veličina (alebo náhodný vektor), tak $f(\mathbf{y}|\boldsymbol{\theta})$ je hustota a ak $\mathbf{Y}$ je diskrétna náhodná veličina (alebo náhodný vektor), tak $f(\mathbf{y}|\boldsymbol{\theta})$ je pravdepodobnostná funkcia.

Označme $\pi^*(\boldsymbol{\theta}|\mathbf{t})_{\mathbf{t}=\boldsymbol{\tau}(\mathbf{y})}$ aposteriórnu hustotu, ak bola pozorovaná hodnota $\mathbf{t}$ štatistiky $\boldsymbol{\tau}$.

Veta 6.1. (Veta o postačujúcej štatistike). Ak $\boldsymbol{\tau}(\mathbf{y})$ je postačujúca štatistika, tak platí

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \pi^*(\boldsymbol{\theta}|\mathbf{t})_{\mathbf{t}=\boldsymbol{\tau}(\mathbf{y})}.$$

Dôkaz:

Ak $\boldsymbol{\Theta}$ aj $\mathbf{Y}$ sú diskrétne náhodné veličiny (alebo náhodné vektory), a $\boldsymbol{\tau}(\mathbf{y})$ je postačujúca štatistika, tak aposteriórna hustota (vlastne pravdepodobnostná funkcia)

$$\text{(6.2)} \qquad \pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{h(\mathbf{y})q(\boldsymbol{\tau}(\mathbf{y}), \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\sum_i h(\mathbf{y})q(\boldsymbol{\tau}(\mathbf{y}), \boldsymbol{\theta})\pi(\boldsymbol{\theta})} = \frac{q(\boldsymbol{\tau}(\mathbf{y}), \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\sum_i q(\boldsymbol{\tau}(\mathbf{y}), \boldsymbol{\theta})\pi(\boldsymbol{\theta})}.$$

Presnejšie v tomto prípade realizácie náhodnej veličiny (alebo náhodného vektora) $\boldsymbol{\Theta}$ sú $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ a realizácie náhodnej veličiny (alebo náhodného vektora) $\mathbf{Y}$ sú $\{\mathbf{y}_1, \mathbf{y}_2, ...\}$

Platí

$$\pi(\boldsymbol{\theta}_j) = P(\boldsymbol{\Theta} = \boldsymbol{\theta}_j), \quad f(\mathbf{y}_i|\boldsymbol{\theta}_j) = P(\mathbf{Y} = \mathbf{y}_i|\boldsymbol{\Theta} = \theta_j),$$

$$\pi(\boldsymbol{\theta}_j|\mathbf{y}_i) = \frac{h(\mathbf{y}_i)q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)}{\sum_{j\geq 1} h(\mathbf{y}_i)q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)} = \frac{q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)}{\sum_{j\geq 1} q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)}.$$

Pre diskrétne náhodné veličiny (resp. náhodné vektory) dostaneme analogicky ako vo vete 2.1, že združená pravdepodobnostná funkcia $g(\boldsymbol{\theta}_j, \mathbf{y}_i) = \pi(\boldsymbol{\theta}_j)f(\mathbf{y}_i|\boldsymbol{\theta}_j)$. Aposteriórna hustota náhodnej veličiny $\boldsymbol{\Theta}$ (vlastne pravdepodobnostná funkcia), ak bola pozorovaná hodnota $\mathbf{t}$ štatistiky $\boldsymbol{\tau}$ je

$$\pi^*(\boldsymbol{\theta}_j|\mathbf{t})_{\mathbf{t}=\boldsymbol{\tau}(\mathbf{y})} = P\{\boldsymbol{\Theta} = \boldsymbol{\theta}_j|\mathbf{Y} = \mathbf{y}_i :\ \boldsymbol{\tau}(\mathbf{y}_i) = \mathbf{t}\}, \quad j = 1, 2, \dots .$$

Ak označíme náhodné udalosti
$A_j - \{\boldsymbol{\Theta} = \boldsymbol{\theta}_j\}, \quad j = 1, 2, ...,$
$B_i - \{\mathbf{Y} = \mathbf{y}_i :\ \boldsymbol{\tau}(\mathbf{y}_i) = \mathbf{t}\}, \quad i = 1, 2, ...,$
tak

$$\pi^*(\boldsymbol{\theta}_j|\mathbf{t})_{\mathbf{t}=\boldsymbol{\tau}(\mathbf{y})} = P\{\boldsymbol{\Theta} = \boldsymbol{\theta}_j|\mathbf{Y} = \mathbf{y}_i :\ \boldsymbol{\tau}(\mathbf{y}_i) = \mathbf{t}\} = P(A_j| \cup B_i) = \frac{P(A_j \cap \cup B_i)}{P(\cup B_i)} = \frac{\sum_i P(A_j \cap B_i)}{\sum_i P(B_i)} =$$

$$= \frac{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}} P(\boldsymbol{\Theta} = \boldsymbol{\theta}_j,\ \mathbf{Y} = \mathbf{y}_i)}{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}} P(\mathbf{Y} = \mathbf{y}_i)} = \frac{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}[P(\boldsymbol{\Theta} = \boldsymbol{\theta}_j)P(\mathbf{Y} = \mathbf{y}_i|\boldsymbol{\Theta} = \boldsymbol{\theta}_j)]}{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}\{\sum_{s\geq 1} P(\boldsymbol{\Theta} = \boldsymbol{\theta}_s)P(\mathbf{Y} = \mathbf{y}_i|\boldsymbol{\Theta} = \boldsymbol{\theta}_s)\}} =$$

$$= \frac{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}[\pi(\boldsymbol{\theta}_j)f(\mathbf{y}_i|\boldsymbol{\theta}_j)]}{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}\{\sum_{s\geq 1} \pi(\boldsymbol{\theta}_s)f(\mathbf{y}_i|\boldsymbol{\theta}_s)\}}.$$

Ak $\boldsymbol{\tau}$ je postačujúca štatistika, tak podľa (6.1)

$$\pi^*(\boldsymbol{\theta}_j|\mathbf{t})_{\mathbf{t}=\boldsymbol{\tau}(\mathbf{y})} = \frac{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}[\pi(\boldsymbol{\theta}_j)h(\mathbf{y}_i)q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_j)]}{\sum_{\mathbf{y}_i:\ \boldsymbol{\tau}(\mathbf{y}_i)=\mathbf{t}}\{\sum_{s\geq 1} \pi(\boldsymbol{\theta}_s)h(\mathbf{y}_i)q(\boldsymbol{\tau}(\mathbf{y}_i), \boldsymbol{\theta}_s)\}} = \frac{\pi(\boldsymbol{\theta}_j)q(\mathbf{t}, \boldsymbol{\theta}_j)}{\sum_{s\geq 1} \pi(\boldsymbol{\theta}_s)q(\mathbf{t}, \boldsymbol{\theta}_s)}.$$

Zo vzťahu (6.2) dostávame, že ak $\boldsymbol{\tau}$ je postačujúca štatistika a $\mathbf{t}$ je pozorovaná hodnota tejto štatistiky, tak aposteriórna pravdepodobnostná funkcia je

$$\pi(\boldsymbol{\theta}_j|\mathbf{y}_i :\ \boldsymbol{\tau}(\mathbf{y}_i) = \mathbf{t}) = \frac{\pi(\boldsymbol{\theta}_j)q(\mathbf{t}, \boldsymbol{\theta}_j)}{\sum_i \pi(\boldsymbol{\theta}_i)q(\mathbf{t}, \boldsymbol{\theta}_i)},$$

teda vetu sme v prípade diskrétnych náhodných veličín (alebo vektorov) $\boldsymbol{\Theta}$ a $\mathbf{Y}$ dokázali. Vo všeobecnom prípade nájdeme dôkaz napr. v knihe A.Pázmana. ♣

V nasledujúcich kapitolách si zavedieme mieru množstva informácie a ukážeme, že vo všeobecnosti hustota $\pi^*(\boldsymbol{\theta}|\mathbf{t})$ obsahuje menej informácie o neznámom parametri $\boldsymbol{\theta}$ než hustota $\pi(\boldsymbol{\theta}|\mathbf{y})$.

7. I-divergencia a informácia získaná z experimentu

Bayesovské ponímanie znamená, že pred experimentom je všetka informácia o parametri $\boldsymbol{\Theta}$ zhrnutá v apriórnej hustote $\pi(\boldsymbol{\theta})$ a úplná informácia o parametri $\boldsymbol{\Theta}$ po experimente je zhrnutá v aposteriórnej hustote $\pi(\boldsymbol{\theta}|\mathbf{y})$.

Niekedy sa vyžaduje charakterizovať množstvo tejto informácie jediným číslom (podobne ako charakterizovať rozdelenie pravdepodobnosti jednou číselnou charakteristikou). Táto úloha samozrejme je diskutabilná a nie je jednoznačná. Jedna z možností je vyjadriť "ekonomicky užitočnú" informáciu podobne ako v kapitole 23. Cieľom môže byť aj vyjadrenie "presnosti", s ktorou hustota $\pi(\boldsymbol{\theta})$ alebo $\pi(\boldsymbol{\theta}|\mathbf{y})$ určujú hodnotu parametra $\boldsymbol{\theta}$ v danom experimente. Tu ide o miery variability rozdelenia pravdepodobnosti náhodnej premennej $\boldsymbol{\Theta}$. Môžeme merať aj nedostatok informácie obsiahnutej v $\pi(\boldsymbol{\theta})$ nejakou "mierou neurčitosti" apriórneho rozdelenia. Jedna z takýchto možných mier vznikla vo fyzike a neskôr sa osvedčila v rôznych vedných oblastiach.

Je to entropia. Budeme sa v ďalšom zaoberať aj vyjadrením stredného množstva informácie, ktoré možno získať z experimentu. Vyjadríme ho pomocou strednej I-divergencii. I-divergencia meria "odlišnosť" dvoch distribúcií.

8. Entropia

Definícia 8.1. Nech $\boldsymbol{\Theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m$ je diskrétna náhodná veličina, ktorá nadobúda hodnoty $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ s pravdepodobnosťami $\{p_1, p_2, ...\}$, teda $P(\boldsymbol{\Theta} = \boldsymbol{\theta}_i) = p_i$ $i = 1, 2, ...$ Pravdepodobnosná funkcia $\{\boldsymbol{\theta}_i, p_i\}_i$, $i = 1, 2, ...$ určuje rozdelenie pravdepodobnosti P. Entropia takto daného rozdelenia pravdepodobnosti P sa rovná

$$Ent(P) = -\sum_i p_i \ln p_i.$$

Ak množina hodnôt $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ je konečná (obsahuje n hodnôt), tak entropia je z uzavretého intervalu $\langle 0, \ln n \rangle$. Minimálnu hodnotu (rovnú 0) nadobúda, keď pre jeden index i_0 je $p_{i_0} = 1$ a pre ostatné indexy $i \neq i_0$ je $p_i = 0$. V tomto prípade s istotou nastáva jediná z hodnôt, a síce $\boldsymbol{\theta}_{i_0}$, a nie je neurčitosť. Maximálnu hodnotu (rovnú $\ln n$) nadobúda entropia, ak $p_i = \frac{1}{n}$ pre všetky i. Vtedy je neurčitosť, ktorá z hodnôt $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_n\}$ nastane, najväčšia. Kladieme $0 \ln 0 = 0$, lebo $\lim_{z \to 0^+}(z \ln z) = 0$.

Definícia 8.2. Nech $\boldsymbol{\Theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m$ je spojitá náhodná veličina, ktorej hustota (voči Lebesgueovej miere) je $f(\boldsymbol{\theta})$. Potom entropia (spojitého) rozdelenia pravdepodobnosti P určeného hustotou $f(\cdot)$ je

$$Ent(P) = -\int_{\boldsymbol{\Omega}} \ln f(\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} = -\mathcal{E}\left(\ln f(\boldsymbol{\Theta})\right).$$

Podotýkame len, že $ent(\boldsymbol{\Theta})$ má iné vlastnosti ako entropia diskrétneho rozdelenia $Ent(\boldsymbol{\Theta})$ (napr. nadobúda hodnoty z intervalu $(-\infty, \infty)$).

9. I-divergencia

Definícia 9.1. Nech $\boldsymbol{\Theta} \in \mathcal{R}^m$ je diskrétna náhodná veličina, ktorá nadobúda hodnoty $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ s pravdepodobnosťami $\{p_1, p_2, ...\}$, teda $P(\boldsymbol{\Theta} = \boldsymbol{\theta}_i) = p_i$ $i = 1, 2, ...$ a $\boldsymbol{\Theta}^*$ nadobúda tie isté hodnoty $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ s pravdepodobnosťami $\{q_1, q_2, ...\}$, teda $Q(\boldsymbol{\Theta}^* = \boldsymbol{\theta}_i) = q_i$ $i = 1, 2, ...$. P a Q sú teda dve rozdelenia pravdepodobnosti, pričom pre tieto dve rozdelenia platí, že ak $q_i > 0$, tak aj $p_i > 0$ (miera Q je absolútne spojitá vzhľadom k miere P). I-divergencia $I(P, Q)$ týchto mier je

$$I(P, Q) = \sum_i p_i \ln \frac{p_i}{q_i} \tag{9.1}$$

ak $\sum_i p_i \ln \frac{p_i}{q_i} < \infty$, inak položíme $I(P, Q) = \infty$. Kladieme $0 \ln \frac{0}{0} = 0$.

Definícia 9.2. Nech $\boldsymbol{\Theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m$ je spojitá náhodná veličina s hustotami $f_P(\boldsymbol{\theta})$ a $f_Q(\boldsymbol{\theta})$ (voči Lebesgueovej miere), pričom pravdepodobnostná miera Q je absolútne spojitá voči miere P (t.j. ak pre nejakú borelovskú množinu A platí $P(A) = 0$, tak aj $Q(A) = 0$). I-divergencia $I(P, Q)$ týchto mier je

$$I(P, Q) = \int_{\boldsymbol{\Omega}} \ln \frac{f_P(\boldsymbol{\theta})}{f_Q(\boldsymbol{\theta})} f_P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathcal{E}_P \left[\ln \frac{f_P(\boldsymbol{\Theta})}{f_Q(\boldsymbol{\Theta})}\right] \tag{9.2}$$

ak $\int_{\boldsymbol{\Omega}} \ln \frac{f_P(\boldsymbol{\theta})}{f_Q(\boldsymbol{\theta})} f_P(\boldsymbol{\theta}) d\boldsymbol{\theta}$ existuje, inak $I(P, Q) = \infty$.

Poznámka. Ńiekedy sa namiesto $I(P,Q)$ píše $I(f_P, f_Q)$. Platí tvrdenie, že vždy $I(P,Q) \geq 0$ a $I(P,Q) = 0$ práve vtedy ak $P = Q$ (pozri knihu A.Pázmana, str. 24). $I(P,Q)$ nie je metrika (vzdialenosť medzi rozdeleniami), lebo $I(P,Q) \neq I(Q,P)$. Miera P má vo výraze pre $I(P,Q)$ dominantné postavenie, lebo vzhľadom na ňu sa berie stredná hodnota.

Príklad 9.1. Majme $n-$rozmerný náhodný vektor $\mathbf{X}$ a jeho dve rôzne normálne rozdelené rozdelenia pravdepodobnosti P a Q dané hustotami

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} det^{\frac{1}{2}}(\boldsymbol{\Sigma}_i)} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}, \quad i \in \{P, Q\}$$

Spočítajte $I(f_P, f_Q)$.

Riešenie.

$$\ln\frac{f_P(\mathbf{x})}{f_Q(\mathbf{x})} = \frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)} - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_P^{-1}(\mathbf{x}-\boldsymbol{\mu}_P) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P-(\boldsymbol{\mu}_Q-\boldsymbol{\mu}_P))'\boldsymbol{\Sigma}_Q^{-1}(\mathbf{x}-\boldsymbol{\mu}_P-(\boldsymbol{\mu}_Q-\boldsymbol{\mu}_P)) =$$

$$= \frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)} - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_P^{-1}(\mathbf{x}-\boldsymbol{\mu}_P) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\mathbf{x}-\boldsymbol{\mu}_P) +$$

$$+(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q) + \frac{1}{2}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q).$$

Počítajme

$$I(f_P(\mathbf{x}), f_Q(\mathbf{x})) = \mathcal{E}_P\left[\ln\frac{f_P(\mathbf{X})}{f_Q(\mathbf{X})}\right] = \int_{\mathcal{R}^n}\ln\frac{f_P(\mathbf{x})}{f_Q(\mathbf{x})}f_P(\mathbf{x})d\mathbf{x} =$$

$$= \int_{\mathcal{R}^n}\frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)}f_P(\mathbf{x})d\mathbf{x} - \int_{\mathcal{R}^n}\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_P^{-1}(\mathbf{x}-\boldsymbol{\mu}_P)f_P(\mathbf{x})d\mathbf{x} + \int_{\mathcal{R}^n}\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\mathbf{x}-\boldsymbol{\mu}_P)f_P(\mathbf{x})d\mathbf{x} +$$

$$+\int_{\mathcal{R}^n}(\mathbf{x}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)f_P(\mathbf{x})d\mathbf{x} + \int_{\mathcal{R}^n}\frac{1}{2}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)f_P(\mathbf{x})d\mathbf{x} =$$

$$= \mathcal{E}_P\left\{\frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)}\right\} - \mathcal{E}_P\left\{\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_P^{-1}(\mathbf{X}-\boldsymbol{\mu}_P)\right\} + \mathcal{E}_P\left\{\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\mathbf{X}-\boldsymbol{\mu}_P)\right\} +$$

$$+\mathcal{E}_P\left\{(\mathbf{X}-\boldsymbol{\mu}_P)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)\right\} + \mathcal{E}_P\left\{\frac{1}{2}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)\right\} =$$

$$= \frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)} - tr\frac{1}{2}\{\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_P\} + tr\frac{1}{2}\{\boldsymbol{\Sigma}_Q^{-1}\boldsymbol{\Sigma}_P\} + \frac{1}{2}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q) =$$

$$\frac{1}{2}\ln\frac{\det(\boldsymbol{\Sigma}_Q)}{\det(\boldsymbol{\Sigma}_P)} - \frac{n}{2} + tr\frac{1}{2}\{\boldsymbol{\Sigma}_Q^{-1}\boldsymbol{\Sigma}_P\} + \frac{1}{2}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}_Q^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q).$$

V prípade, že $\boldsymbol{\Sigma}_P = \boldsymbol{\Sigma}_Q = \boldsymbol{\Sigma}$ (rozdelenia sa líšia len v strednej hodnote), dostávame

$$2I(P,Q) = (\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q).$$

Príklad 9.2. Majme diskrétnu náhodnú premennú $X \in \{0, 1, ..., n\}$ a jej dve rôzne binomické rozdelenia pravdepodobnosti P a Q dané hustotami (pravdepodobnostnými funkciami)

$$f_j(x) = \binom{n}{x}\theta_j^x(1-\theta_j)^{n-x}, \quad x = 0, 1, ..., n, \quad j \in \{P, Q\}$$

Spočítajte $I(P,Q)$.

Riešenie.

$$\ln\frac{f_P(x)}{f_Q(x)} = x\ln\frac{\theta_P}{\theta_Q} + (n-x)\ln\frac{1-\theta_P}{1-\theta_Q},$$

preto podľa (9.1)

$$I(P,Q) = \sum_{x=0}^{n} \left[x \ln \frac{\theta_P}{\theta_Q} + (n-x) \ln \frac{1-\theta_P}{1-\theta_Q} \right] \binom{n}{x} \theta_P^x (1-\theta_P)^{n-x} = n\theta_P \ln \frac{\theta_P}{\theta_Q} + n(1-\theta_P) \ln \frac{1-\theta_P}{1-\theta_Q}.$$

10. Súvis I-divergencie s Fisherovou informačnou maticou

(Spojitý prípad.) Majme štatistický experiment s výberovým priestorom $\mathbf{Y} \in \mathcal{Y} \in \mathcal{B}_n$ a hustotami

$$\{f(\cdot|\boldsymbol{\theta}) : \quad \boldsymbol{\theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m\}$$

(voči Lebesgueovej miere). Ak pre malé $\boldsymbol{\Delta} \in \boldsymbol{\Omega}$ sa hustoty $f(\cdot|\boldsymbol{\theta})$ a $f(\cdot|\boldsymbol{\theta}+\boldsymbol{\Delta})$ značne odlišujú, tak na základe experimentu budeme môcť dobre rozlišovať $\boldsymbol{\theta}$ od $\boldsymbol{\theta}+\boldsymbol{\Delta}$. Preto budeme môcť v tomto prípade "presne" odhadovať parameter $\boldsymbol{\theta}$. Odlišnosť $f(\cdot|\boldsymbol{\theta})$ od $f(\cdot|\boldsymbol{\theta}+\boldsymbol{\Delta})$ je vhodné merať pomocou I-divergencie

$$I_{\boldsymbol{\theta},\boldsymbol{\theta}+\boldsymbol{\Delta}} = \int_{\mathcal{Y}} \ln \frac{f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta}+\boldsymbol{\Delta})} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}). \tag{10.1}$$

(Diskrétny prípad.) Ak máme štatistický experiment s výberovým priestorom $\mathbf{Y} \in \{\mathbf{y}_1, \mathbf{y}_2, ...\}$ a hustotami (pravdepodobnostnými funkciami)

$$\{f(\cdot|\boldsymbol{\theta}) : \quad \boldsymbol{\theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m\},$$

tak odlišnosť $f(\cdot|\boldsymbol{\theta})$ od $f(\cdot|\boldsymbol{\theta}+\boldsymbol{\Delta})$ je vhodné merať pomocou I-divergencie

$$I_{\boldsymbol{\theta},\boldsymbol{\theta}+\boldsymbol{\Delta}} = \sum_i \ln \frac{f(\mathbf{y}_i|\boldsymbol{\theta})}{f(\mathbf{y}_i|\boldsymbol{\theta}+\boldsymbol{\Delta})} f(\mathbf{y}_i|\boldsymbol{\theta}). \tag{10.2}$$

Veta 10.1. (Súvis I-divergencie a Fisherovej informačnej matice.) Ak pre skoro všetky $\mathbf{y} \in \mathcal{Y} \in \mathcal{B}_n$ je $f(\mathbf{y}|\boldsymbol{\theta})$ dva razy spojite diferencovateľnou funkciou a ak možno zameniť poradie derivovania a integrovania vo výraze

$$\int_{\mathcal{Y}} \frac{\partial}{\partial \theta_i} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}),$$

potom platí

$$2I_{\boldsymbol{\theta},\boldsymbol{\theta}+\boldsymbol{\Delta}} = \boldsymbol{\Delta}' \mathbf{M}(\boldsymbol{\theta}) \boldsymbol{\Delta} + o(||\boldsymbol{\Delta}||^2),$$

kde $\mathbf{M}(\boldsymbol{\theta})$ je Fisherova informačná matica s prvkami

$$\mathbf{M}_{ij}(\boldsymbol{\theta}) = -\mathcal{E}_{\boldsymbol{\theta}} \left\{ \frac{\partial^2 \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}, \quad \boldsymbol{\Delta}' = (\Delta_1, \Delta_2, ..., \Delta_m)$$

a $o(||\boldsymbol{\Delta}||^2)$ je taká funkcia $||\boldsymbol{\Delta}||^2$, že $\lim_{\boldsymbol{\Delta}\to 0} \dfrac{o(||\boldsymbol{\Delta}||^2)}{||\boldsymbol{\Delta}||^2} = 0$.

Dôkaz: Dôkaz spravíme pre spojitý prípad. Pre diskrétnu náhodnú veličinu $\mathbf{Y}$ sa spraví analogicky. Pomocou Taylorovej vety dostávame

$$\ln f(\mathbf{y}|\boldsymbol{\theta}+\boldsymbol{\Delta}) = \ln f(\mathbf{y}|\boldsymbol{\theta}) + \frac{\partial \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \boldsymbol{\Delta} + \frac{1}{2} \boldsymbol{\Delta}' \frac{\partial^2 \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \boldsymbol{\Delta} + \tilde{o}(||\boldsymbol{\Delta}||^2). \tag{10.3}$$

Pretože

$$\int_{\mathcal{Y}} \left[\frac{\partial \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \boldsymbol{\Delta} \right] f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) = \int_{\mathcal{Y}} \sum_{i=1}^{n} \frac{\partial \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i} \Delta_i f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) = \tag{10.4}$$

$$= \sum_{i=1}^{n} \int_{\mathcal{Y}} \frac{\partial f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i} \Delta_i d\lambda(\mathbf{y}) = \sum_{i=1}^{n} \Delta_i \frac{\partial}{\partial \theta_i} \int_{\mathcal{Y}} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) = 0,$$

z (10.1) pomocou (10.3) a (10.4) dostávame

$$2I_{\boldsymbol{\theta},\boldsymbol{\theta}+\boldsymbol{\Delta}} = 2\int_{\mathcal{Y}} \ln \frac{f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta}+\boldsymbol{\Delta})} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) =$$

$$= -2\int_{\mathcal{Y}} \frac{\partial \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \boldsymbol{\Delta} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) - \int_{\mathcal{Y}} \boldsymbol{\Delta}' \frac{\partial^2 \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \boldsymbol{\Delta} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) - 2\tilde{o}(||\boldsymbol{\Delta}||^2) \int_{\mathcal{Y}} f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) =$$

$$= \boldsymbol{\Delta}' \mathbf{M}(\boldsymbol{\theta}) \boldsymbol{\Delta} + o(||\boldsymbol{\Delta}||^2). \qquad \clubsuit$$

11. Stredné množstvo informácie z experimentu

Bayesovská odpoveď na otázku, koľko informácie poskytuje daný experiment, vyplýva z porovnania apriórnej a aposteriórnej hustoty. V experimente danom výberovým priestorom $\mathcal{Y}$ a meraniami $\mathbf{Y}$, teda náhodnou veličinou (náhodným vektorom) $\mathbf{Y}$ a s hustotami (alebo pravdepodobnostnými funkciami)

$$\{f(\cdot|\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m\}$$

(voči Lebesgueovej miere alebo voči sčítacej miere), množstvo informácie o parametri $\boldsymbol{\theta}$, (keď výsledok experimentu bol $\mathbf{y}$), možno vyjadriť pomocou výrazu

$$I[\pi(\cdot|\mathbf{y}), \pi(\cdot)] = \int_{\boldsymbol{\Omega}} \left[\ln \frac{\pi(\boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta})} \right] \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \tag{11.1}$$

ak $\boldsymbol{\Theta}$ je (absolútne) spojitý (voči Lebesgueovej miere) náhodný vektor, alebo

$$I[\pi(\cdot|\mathbf{y}), \pi(\cdot)] = \sum_{i} \left[\ln \frac{\pi(\boldsymbol{\theta}_i|\mathbf{y})}{\pi(\boldsymbol{\theta}_i)} \right] \pi(\boldsymbol{\theta}_i|\mathbf{y}), \tag{11.2}$$

ak $\boldsymbol{\Theta}$ je diskrétny náhodný vektor s hodnotami $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$. Výraz (11.1) resp. (11.2) je vždy nezáporný, je nulový práve ak $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{y})$. Z Bayesovho vzorca (2.7) vyplýva, že v takomto prípade

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})},$$

čiže

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta}).$$

Pretože ľavá strana predchádzajúcej rovnosti závisí od $\boldsymbol{\theta}$ a pravá strana nie, musí byť funkcia vierohodnosti pre dané $\mathbf{y}$ konštantná. Vektor $\mathbf{y}$ neobsahuje žiadnu informáciu o parametri $\boldsymbol{\theta}$. Pri jednom výsledku $\mathbf{y}$ máme množstvo získanej informácie dané vzťahom (11.1) resp. (11.2). Prediktívna hustota náhodného vektora $\mathbf{Y}$ je $f^*(\mathbf{y})$ a preto stredné množstvo informácie, ktoré možno získať z experimentu sa rovná

$$I_{exp} = \mathcal{E}_{\mathbf{Y}}\{I[\pi(\cdot|\mathbf{y}), \pi(\cdot)]\} = \int_{\mathcal{Y}} I[\pi(\cdot|\mathbf{y}), \pi(\cdot)] f^*(\mathbf{y}) d\lambda(\mathbf{y}) \tag{11.3}$$

v spojitom prípade, alebo

$$I_{exp} = \sum_{i} I[\pi(\cdot|\mathbf{y}_i), \pi(\cdot)] f^*(\mathbf{y}_i), \tag{11.4}$$

ak $\mathbf{Y}$ je diskrétny náhodný vektor.

Pozrime sa na iné interpretácie I_{exp}.

Uvažujme spojitý prípad a predpokladajme, že môžeme zamieňať poradie integrácie. Pomocou (2.6) a (2.7) upravíme (11.3)

$$(11.5) \qquad I_{exp} = \mathcal{E}_{\mathbf{Y}}\{I[\pi(\cdot|\mathbf{y}), \pi(\cdot)]\} =$$

$$= \int_{\mathcal{Y}} I[\pi(\cdot|\mathbf{y}), \pi(\cdot)] f^*(\mathbf{y}) d\lambda(\mathbf{y}) = \int_{\mathcal{Y}} \left(\int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta}|\mathbf{y})]\, \pi(\boldsymbol{\theta}|\mathbf{y}) d\nu(\boldsymbol{\theta}) \right) f^*(\mathbf{y}) d\lambda(\mathbf{y}) =$$

$$= \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta}|\mathbf{y})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) - \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}).$$

Spočítajme teraz strednú I-divergenciu $f(\cdot|\boldsymbol{\theta})$ voči $f^*(\cdot)$

$$\int_{\mathbf{\Omega}} I[f(\cdot|\boldsymbol{\theta}), f^*(\cdot)] \pi(\boldsymbol{\theta}) d\nu(\boldsymbol{\theta}) =$$

$$= \int_{\mathbf{\Omega}} \left(\int_{\mathcal{Y}} [\ln f(\mathbf{y}|\boldsymbol{\theta})] f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}) \right) \pi(\boldsymbol{\theta}) d\nu(\boldsymbol{\theta}) - \int_{\mathbf{\Omega}} \int_{\mathcal{Y}} \left[\ln \frac{g(\boldsymbol{\theta}, \mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \right] f(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) =$$

$$= \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta}|\mathbf{y})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) - \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) = I_{exp}.$$

Stredné množstvo informácie, ktoré možno získať z experimentu sa rovná strednej I-divergencii $f(\cdot|\boldsymbol{\theta})$ voči $f^*(\cdot)$.

Spočítajme I-divergenciu združenej hustoty $g(\boldsymbol{\theta}, \mathbf{y})$ voči súčinu marginálnych hustôt $f^*(\mathbf{y})$ a $\pi(\boldsymbol{\theta})$. Pomocou (2.7) a (11.5) dostávame

$$\int_{\mathbf{\Omega} \times \mathcal{Y}} \left[\ln \frac{g(\boldsymbol{\theta}, \mathbf{y})}{f^*(\mathbf{y}) \pi(\boldsymbol{\theta})} \right] g(\boldsymbol{\theta}, \mathbf{y}) d(\nu \times \lambda)(\boldsymbol{\theta}, \mathbf{y}) = \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln g(\boldsymbol{\theta}, \mathbf{y})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) -$$

$$- \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} \left[\ln \frac{g(\boldsymbol{\theta}, \mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \right] g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) - \int_{\mathcal{Y}} \int_{\mathbf{\Omega}} [\ln \pi(\boldsymbol{\theta})]\, g(\boldsymbol{\theta}, \mathbf{y}) d\nu(\boldsymbol{\theta}) d\lambda(\mathbf{y}) = I_{exp}.$$

Stredné množstvo informácie, ktoré možno získať z experimentu sa rovná I-divergencii združenej hustoty $g(\boldsymbol{\theta}, \mathbf{y})$ voči súčinu marginálnych hustôt $f^*(\mathbf{y})$ a $\pi(\boldsymbol{\theta})$. Preto dostávame (pozri Poznámku pod Definíciou 9.2)

Veta 11.1. Stredné množstvo informácie získanej z experimentu I_{exp} je vždy nezáporné, pričom $I_{exp} = 0$ vtedy a len vtedy, keď pozorovaný vektor $\mathbf{Y}$ a parametre $\mathbf{\Theta}$ sú nezávislé náhodné vektory, t.j. keď v $\mathbf{Y}$ nie je obsiahnutá žiadna informácia o $\mathbf{\Theta}$.

Poznámka. Na str. 29 v knižke A.Pázmana je dokázané tvrdenie, že stredná informácia získaná z pozorovania $\boldsymbol{\tau}(\mathbf{y})$ ($\boldsymbol{\tau}(\cdot)$ je nejaká štatistika) je menšia alebo rovná strednej informácii získanej z pozorovania $\mathbf{y}$. Rovnosť nastáva práve vtedy, ak $\boldsymbol{\tau}(\mathbf{y})$ je postačujúca štatistika. "Náhrada" pozorovaného vektora $\mathbf{y}$ hodnotou $\boldsymbol{\tau}(\mathbf{y})$ nemôže zväčšiť množstvo informácie o $\mathbf{\Theta}$, nanajvýš ho zachovať. Toto nastane práve vtedy, ak je $\boldsymbol{\tau}(\cdot)$ postačujúca štatistika.

12. Využitie I-divergencie v asymptotike Bayesovho vzorca

Zopakujme si Zákon veľkých čísel - Chinčinovu vetu (pozri Anděl, str. 183).

Veta 12.1. (Chinčinova.) Nech $X_1, X_2, ...$ je postupnosť nezávislých náhodných veličín, ktoré majú rovnaké rozdelenie s konečnou strednou hodnotou μ. Potom pre $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$$

podľa pravdepodobnosti.

V kapitolke 5. "Pravidlo reťazenia pre nezávislé pozorovania" sme ukázali, že postupne pridávané nezávislé experimenty "akumulujú" informáciu. Majme zvolené nejaké apriórne rozdelenie (môže byť aj subjektívne). Nezávisle opakujeme ten istý experiment tak, aby skutočná hodnota parametra $\boldsymbol{\Theta}$ bola tá istá (ale neznáma). Označme ju $\boldsymbol{\theta}^*$. Ukážeme, že pri veľkom počte nezávislých opakovaní experimentu, výsledné aposteriórne rozdelenie sa v limte koncentruje do bodu $\boldsymbol{\theta}^*$.

Veta 12.2. Nech $\boldsymbol{\Theta}$ je diskrétny náhodný vektor s hodnotami $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$. Apriórne pravdepodobnosti $\pi(\boldsymbol{\theta}_i) > 0$ pre každé $i = 1, 2, ...$. Teda apriórna pravdepodobnostná funkcia diskrétnho náhodného vektora $\boldsymbol{\Theta}$ je $\{\pi(\boldsymbol{\theta}_i)\}_{i\geq 1}$. Nech $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ sú pozorované hodnoty v n nezávislých experimentoch s tou istou hustotou $f(\cdot|\boldsymbol{\theta})$, čiže $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ sú realizácie spojitých nezávislých rovnako rozdelených náhodných vektorov $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n$. Nech pre každé $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j$ platí $f(\cdot|\boldsymbol{\theta}_i) \neq f(\cdot|\boldsymbol{\theta}_j)$. Potom pre aposteriórnu pravdepodobnostnú funkciu platí

$$\lim_{n\to\infty} \pi(\boldsymbol{\theta}_i|\mathbf{y}^n) = \begin{cases} 1, & \text{ak } \ \boldsymbol{\theta}_i = \boldsymbol{\theta}^*, \\ 0, & \text{ak } \ \boldsymbol{\theta}_i \neq \boldsymbol{\theta}^*, \end{cases}$$

pričom $\mathbf{y}^n = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)$.

Dôkaz:

Platí

$$f(\mathbf{y}^n|\boldsymbol{\theta}_k) = \prod_{i=1}^{n} f(\mathbf{y}_i|\boldsymbol{\theta}_k), \quad k = 1, 2, ... \ .$$

Teda podľa Poznámky pod Vetou 2.2 (Bayesovou) platí

$$\pi(\boldsymbol{\theta}_i|\mathbf{y}^n) = \frac{\pi(\boldsymbol{\theta}_i) f(\mathbf{y}^n|\boldsymbol{\theta}_i)}{\sum_{k\geq 1} \pi(\boldsymbol{\theta}_k) f(\mathbf{y}^n|\boldsymbol{\theta}_k)} = \frac{\pi(\boldsymbol{\theta}_i)\dfrac{f(\mathbf{y}^n|\boldsymbol{\theta}_i)}{f(\mathbf{y}^n|\boldsymbol{\theta}^*)}}{\sum_{k\geq 1} \pi(\boldsymbol{\theta}_k)\dfrac{f(\mathbf{y}^n|\boldsymbol{\theta}_k)}{f(\mathbf{y}^n|\boldsymbol{\theta}^*)}} = \frac{\pi(\boldsymbol{\theta}_i)\exp\left\{nS_i^{(n)}\right\}}{\sum_{k\geq 1} \pi(\boldsymbol{\theta}_k)\exp\left\{nS_k^{(n)}\right\}},$$

kde

$$S_i^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \ln\frac{f(\mathbf{y}_k|\boldsymbol{\theta}_i)}{f(\mathbf{y}_k|\boldsymbol{\theta}^*)}.$$

Náhodné veličiny $\ln\dfrac{f(\mathbf{Y}_k|\boldsymbol{\theta}_i)}{f(\mathbf{Y}_k|\boldsymbol{\theta}^*)}, \quad k = 1, 2, ..., n$ majú strednú hodnotu

$$\mathcal{E}_{\boldsymbol{\theta}^*}\left[\ln\frac{f(\mathbf{Y}_k|\boldsymbol{\theta}_i)}{f(\mathbf{Y}_k|\boldsymbol{\theta}^*)}\right] = -I(f(\cdot|\boldsymbol{\theta}^*), f(\cdot|\boldsymbol{\theta}_i)) \leq 0, \quad k = 1, 2, ..., n,$$

pričom táto stredná hodnota je rovná 0 práve ak $f(\cdot|\boldsymbol{\theta}_i) = f(\cdot|\boldsymbol{\theta}^*)$, čo nastane práve vtedy ak $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*$. Teda ak $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*$, tak podľa pravdepodobnosti

$$\frac{1}{n}\sum_{k=1}^{n} \ln\frac{f(\mathbf{Y}_k|\boldsymbol{\theta}_i)}{f(\mathbf{Y}_k|\boldsymbol{\theta}^*)} \to 0.$$

Ak $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}^*$, tak podľa pravdepodobnosti

$$\frac{1}{n}\sum_{k=1}^{n} \ln \frac{f(\mathbf{Y}_k|\boldsymbol{\theta}_j)}{f(\mathbf{Y}_k|\boldsymbol{\theta}^*)} \to \alpha < 0$$

a

$$\pi(\boldsymbol{\theta}_j) \exp\left\{ n\frac{1}{n}\sum_{k=1}^{n} \ln \frac{f(\mathbf{Y}_k|\boldsymbol{\theta}_j)}{f(\mathbf{Y}_k|\boldsymbol{\theta}^*)} \right\} \to 0.$$

Konečne dostávame, že podľa pravdepodobnosti pre $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*$

$$\pi(\boldsymbol{\theta}_i|\mathbf{Y}^n) \to 1$$

a pre $\boldsymbol{\theta}_i \neq \theta^*$

$$\pi(\boldsymbol{\theta}_i|\mathbf{Y}^n) \to 0,$$

pričom $\mathbf{Y}^n = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n)'$. ♣

Dokázaná veta je len jednoduchším prípadom asymptotických tvrdení o aposteriórnej hustote. V knihe J.Anděla v kapitole o bayesovských metódach nájdete dôkaz, že aposteriórna hustota asymptoticky (pri veľkom počte pozorovaní) málo závisí od apriórnej hustoty. Podobný zmysel má aj veta 2.1 v skriptách M.Huškovej.

13. Princíp neurčitosti

Od vzniku bayesovskej štatistiky bola snaha stanoviť, aké apriórne rozdelenie treba zvoliť, ak nemáme žiadnu informáciu o $\boldsymbol{\Theta}$. Ak $\boldsymbol{\Theta}$ je absolútne spojitý (vzhľadom k Lebesgueovej miere), pričom $P(\boldsymbol{\Theta} \in \boldsymbol{\Omega}) = 1$, $\boldsymbol{\Omega} \in \mathcal{B}_m$ a Lebesgueova miera $\mu(\boldsymbol{\Omega}) > 0$, podľa princípu neurčitosti sa volí apriórne rozdelenie rovnomerné na $\boldsymbol{\Omega}$, teda $\pi(\boldsymbol{\theta}) = k > 0$, $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ (obyčajne sa volí $k = 1$). V prípade, že $\boldsymbol{\Theta}$ je diskrétny, jeho apriórne rozdelenie pravdepodobnosti volíme tiež rovnomerné. Často sa stáva, že $\mu(\boldsymbol{\Omega}) = \infty$ (alebo diskrétna náhodná veličina $\boldsymbol{\Theta}$ nadobúda spočítateľne veľa hodnôt). Potom apriórna hustota je nevlastná.

Príklad 13.1. Nech náhodná veličina Y má binomické rozdelenie pravdepodobnosti s pravdepodobnostnou fumkciou $f(y|\theta)$

$$f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, 1, 2, ..., n$$

(n je známe celé nezáporné číslo). O parametri $\Theta \in (0, 1)$ nemáme žiadnu apriórnu informáciu. Podľa princípu neurčitosti volíme apriórnu hustotu

$$\pi(\theta) = 1, \quad 0 < \theta < 1.$$

Podľa (2.7) je aposteriórna hustota parametra Θ rovná

$$\pi(\theta|y) = \frac{\theta^y(1-\theta)^{n-y}}{B(y+1, n-y+1)}, \quad 0 < \theta < 1.$$

Príklad 13.2. Nech $Y_1, Y_2, ..., Y_n$ je náhodný výber z normálneho rozdelenia $N(\Theta, \sigma^2)$, kde σ^2 je známe kladné číslo. O parametri $\Theta \in \Omega = \mathcal{R}$ nemáme žiadne apriórne informácie. V tomto prípade za apriórnu hustotu parametra Θ (podľa princípu neurčitosti) volíme

$$\pi(\theta) = 1, \quad \theta \in \mathcal{R}$$

(nevlastná hustota). Pretože

$$f(y_1, y_2, ..., y_n|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\theta)^2} = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\overline{y})^2}\ e^{-\frac{n}{2\sigma^2}(\overline{y}-\theta)^2},$$

aposteriórna hustota parametra Θ je podľa (2.7)

$$\pi(\theta|y_1, y_2, ..., y_n) = \pi(\theta|\mathbf{y}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\ e^{-\frac{n}{2\sigma^2}(\theta-\overline{y})^2}.$$

Aposteriórna hustota parametra Θ je $N\left(\overline{y}, \dfrac{\sigma^2}{n}\right)$.

Princíp neurčitosti (alebo princíp rovnakej pravdepodobnosti) vedie v niektorých prípadoch k protirečeniam. Napríklad ak máme experiment s hustotami $f(y|\theta)$, v ktorom $P(\Theta \in (0,1)) = 1$. Zaujíma nás ale "nový" parameter $\beta = \theta^2$. Tento tiež nadobúda hodnoty z $(0,1)$. Podľa princípu neurčitosti by oba parametre mali byť rozdelené rovnomerne, t.j. s apriornými hustotami

$$\pi(\theta) = 1, \qquad \pi(\beta) = 1. \tag{13.1}$$

Platí veta

Veta 13.1. (Anděl, str. 46.) Nech X má spojitú distribučnú funkciu $F(x)$. Predpokladajme, že $F'(x) = f(x)$ existuje všade s výnimkou najviac konečne veľa bodov. Nech t je rýdzo monotónna funkcia, ktorá má všade deriváciu. Položme $Y = t(X)$. Označme τ inverznú funkciu k t. Potom Y má hustotu

$$g(y) = f(\tau(y))\,|\tau'(y)|\,.$$

Označme $t(\theta) = \theta^2$, čiže $\tau(\beta) = \sqrt{\beta} = \theta$. Teda platí $\pi(\beta) = \pi(\sqrt{\beta})\left|\dfrac{d\sqrt{\beta}}{d\beta}\right| = \dfrac{1}{2\sqrt{\beta}}$. Toto je v spore s (13.1).

14. Jeffreysova apriórna hustota

Predchádzajúce úvahy viedli štatistikov k záveru, že bolo by vhodné namiesto princípu neurčitosti voliť apriórne rozdelenie tak, aby aposteriórne rozdelenie nezáviselo od parametrizácie modelu. V prípade, že $\mathbf{\Theta} \in \mathcal{R}^m$ je (absolútne) spojitý (vzhľadom na Lebesgueovu mieru μ), $P(\mathbf{\Theta} \in \mathbf{\Omega}) = 1$, pričom $\mu(\mathbf{\Omega}) > 0$, je riešenie tohto problému vo vete 14.2. Ešte pred jej sformulovaním si zopakujme niekoľko definícií a viet.

Definícia 14.1. Nech náhodný vektor $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ má hustotu $f(\mathbf{y}|\boldsymbol{\theta})$ (vzhľadom k $\sigma-$konečnej miere μ), pričom $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)'$. Predpokladajme, že platí:

(A) $\boldsymbol{\theta} \in \mathbf{\Omega}$, kde $\mathbf{\Omega}$ je neprázdna otvorená množina v $\mathcal{R}^m$.

(B) Množina $\mathbf{M} = \{\mathbf{y} :\ f(\mathbf{y}, \boldsymbol{\theta}) > 0\}$ nezávisí od $\boldsymbol{\theta}$.

(C) Pre skoro všetky $\mathbf{y} \in \mathbf{M}$ (vzhľadom k μ) existujú parciálne derivácie $f_i'(\mathbf{y}, \boldsymbol{\theta}) = \dfrac{\partial f(\mathbf{y}, \boldsymbol{\theta})}{\partial \theta_i}$, $i = 1, 2, ..., m$.

(D) Pre každé i a pre všetky $\boldsymbol{\theta} \in \mathbf{\Omega}$ platí $\int_{\mathbf{M}} f_i'(\mathbf{y}, \boldsymbol{\theta}) d\mu(\mathbf{y}) = 0$.

(E) Pre každú dvojicu (i, j) existuje konečný integrál

$$M_{ij}(\boldsymbol{\theta}) = \int_{\mathbf{M}} \frac{f_i'(\mathbf{y}, \boldsymbol{\theta}) f_j'(\mathbf{y}, \boldsymbol{\theta})}{f^2(\mathbf{y}, \boldsymbol{\theta})} f(\mathbf{y}, \boldsymbol{\theta}) d\mu(\mathbf{y}).$$

(F) Matica $\mathbf{M}(\boldsymbol{\theta})$ s prvkami $\{\mathbf{M}(\boldsymbol{\theta})\}_{ij} = M_{ij}(\boldsymbol{\theta})$ je pozitívne definitná pre každé $\boldsymbol{\theta} \in \mathbf{\Omega}$.

Ak sú splnené predpoklady (A) až (E), tak $\mathbf{M}(\boldsymbol{\theta})$ sa nazýva Fisherova informačná matica. Ak sú splnené predpoklady (A) až (F), tak tak hovoríme, že systém hustôt je *regulárny*.

Poznámka. Definícia 14.1 je platná aj keď náhodný vektor $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ má diskrétne rozdelenie pravdepodobnosti. Vtedy jeho "hustota" je jeho pravdepodobnostná funkcia. V bode (D) podmienka $\int_M f_i'(\mathbf{y}, \boldsymbol{\theta}) d\mu(\mathbf{y}) = 0$ je vlastne $\mathcal{E}\left[\dfrac{f_i'(\mathbf{Y}, \boldsymbol{\theta})}{f(\mathbf{Y}, \boldsymbol{\theta})}\right] = 0$ a Fisherova informačná matica má prvky $M_{ij}(\boldsymbol{\theta}) = \mathcal{E}_{\boldsymbol{\theta}}\left[\dfrac{f_i'(\mathbf{Y}|\boldsymbol{\theta}) f_j'(\mathbf{Y}|\boldsymbol{\theta})}{f^2(\mathbf{Y}|\boldsymbol{\theta})}\right]$.

Poznámka. V učebnici J. Anděla, str. 261 je dokázaná veta, ktorá tvrdí, že v prípade regulárneho systému hustôt $\{f(\mathbf{y}|\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\}$, ak existujú derivácie $f_{ij}'' = \dfrac{\partial^2 f(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}$, $i, j = 1, 2, ..., m$ a pre všetky $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ platí $\mathcal{E}_{\boldsymbol{\theta}}\left[\dfrac{f_{ij}''(\mathbf{Y}|\boldsymbol{\theta})}{f(\mathbf{Y}|\boldsymbol{\theta})}\right] = 0$, $i, j = 1, 2, ..., m$, tak $M_{ij}(\boldsymbol{\theta}) = -\mathcal{E}_{\boldsymbol{\theta}}\left[\dfrac{\partial^2 \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]$.

Definícia 14.2. Zobrazenie $\mathbf{f}$ z množiny A do množiny B nazývame *prostým*, ak platí

$$\{\mathbf{x}_1 \in A,\ \mathbf{x}_2 \in A,\ \mathbf{x}_1 \neq \mathbf{x}_2\} \implies \{\mathbf{f}(\mathbf{x}_1) \neq \mathbf{f}(\mathbf{x}_2)\}.$$

Definícia 14.3. Nech $\mathbf{f}$ je zobrazenie z $\mathcal{R}^r$ do $\mathcal{R}^r$. Ak je $\mathbf{y} = f(\mathbf{x})$, kde $\mathbf{y} = (y_1, ..., y_r)$ a $\mathbf{x} = (x_1, ..., x_r)$, položme $y_i = f_i(x_1, ..., x_r),\ i = 1, 2, ..., r$. Povieme, že zobrazenie $\mathbf{f}$ je *regulárne* v množine $M \subset \mathcal{R}^r$, ak platí:

1. M je otvorená.
2. Funkcie $f_1, ..., f_r$ majú parciálne derivácie prvého rádu spojité v M.
3. Pre každé $\mathbf{x} \in M$ platí, že *jakobián* $\mathbf{D_f}(\mathbf{x}) \neq 0$, pričom

$$\mathbf{D_f}(\mathbf{x}) = \det \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_r} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_r}{\partial x_1} & \cdots & \dfrac{\partial f_r}{\partial x_r} \end{pmatrix}.$$

Veta 14.1. (O substitúcii, Anděl, str. 47.) Nech $\mathbf{f}$ je zobrazenie otvorenej množiny $P \subset \mathcal{R}^r$ na $Q \subset \mathcal{R}^r$. Nech $\mathbf{f}$ je regulárne a prosté v P s jakobiánom $\mathbf{D_f}(\mathbf{x})$. Nech $M \subset Q$ je borelovská množina a F merateľná reálna funkcia. Potom platí

$$\int_M F(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{f}^{-1}(M)} F\left[\mathbf{f}(\mathbf{u})\right] |\mathbf{D_f}(\mathbf{u})| d\mathbf{u},$$

akonáhle jeden z integrálov existuje.

Veta 14.2. (Anděl, str. 291.) Nech náhodný vektor $\mathbf{Y}$ má pri danom parametri $\boldsymbol{\theta} \in \boldsymbol{\Omega} \in \mathcal{B}_m$ hustotu $f(\mathbf{y}|\boldsymbol{\theta})$ (alebo pravdepodobnostnú funkciu $\{f(\mathbf{y}_i|\boldsymbol{\theta})\}_{i\geq 1}$) . Predpokladajme, že systém hustôt $\{f(\mathbf{y}|\boldsymbol{\theta}) :\ \boldsymbol{\theta} \in \boldsymbol{\Omega}\}$ je regulárny a má Fisherovu informačnú maticu $\mathbf{M}(\boldsymbol{\theta})$, $\nu(\boldsymbol{\theta})$ je Lebesgueova miera. Nech platí

$$0 < \int_{\boldsymbol{\Omega}} f(\mathbf{y}|\boldsymbol{\theta}) |\mathbf{M}(\boldsymbol{\theta})|^{\frac{1}{2}} d\boldsymbol{\theta} < \infty. \tag{14.1}$$

Položme

$$c = \frac{1}{\int_{\boldsymbol{\Omega}} f(\mathbf{y}|\boldsymbol{\theta}) |\mathbf{M}(\boldsymbol{\theta})|^{\frac{1}{2}} d\boldsymbol{\theta}}. \tag{14.2}$$

Nech $\mathbf{H}$ je regulárne a prosté zobrazenie $\boldsymbol{\Omega}$ na $\boldsymbol{\Omega}^* \in \mathcal{B}_m$. Označme $\boldsymbol{\eta} = \mathbf{H}(\boldsymbol{\theta})$ a $f^*(\mathbf{y}|\boldsymbol{\eta}) = f(\mathbf{y}|\mathbf{H}^{-1}(\boldsymbol{\eta}))$. Potom je $\{f^*(\mathbf{y}|\boldsymbol{\eta}),\ \boldsymbol{\eta} \in \boldsymbol{\Omega}^*\}$ regulárny systém hustôt. Ak označíme $\mathbf{M}^*(\boldsymbol{\eta})$ jeho Fisherovu informačnú maticu, potom pre ľubovoľnú množinu $\mathbf{B} \in \mathcal{B}_m,\ \mathbf{B} \subset \boldsymbol{\Omega}$ platí

$$\int_{\mathbf{B}} c f(\mathbf{y}|\boldsymbol{\theta}) |\mathbf{M}(\boldsymbol{\theta})|^{\frac{1}{2}} d\boldsymbol{\theta} = \int_{\mathbf{H}(\mathbf{B})} c f^*(\mathbf{y}|\boldsymbol{\eta}) |\mathbf{M}^*(\boldsymbol{\eta})|^{\frac{1}{2}} d\boldsymbol{\eta}. \tag{14.3}$$

Dôkaz je uvedený v knižke J.Anděla, str.291.

Z Vety 14.2 vyplýva, že pri apriórnej hustote parametra $\mathbf{\Theta}$ rovnej funkcii $|\mathbf{M}(\boldsymbol{\theta})|^{\frac{1}{2}}$ (alebo ľubovolnému kladnému násobku tejto funkcie) je aposteriórna hustota parametra $\mathbf{\Theta}$ rovná $cf(\mathbf{y}|\boldsymbol{\theta})|\mathbf{M}(\boldsymbol{\theta})|^{\frac{1}{2}}$ a je to hustota. Vzorec (14.3) zaručuje, že aposteriórna pravdepodobnosť ľubovolnej "rozumnej" množiny $\mathbf{B}$ nezávisí od toho, či sa vychádza od parametra $\boldsymbol{\theta}$ alebo od nejakej jeho hladkej transformácie $\boldsymbol{\eta} = \mathbf{H}(\boldsymbol{\theta})$. Na ľavej strane (14.3) je totiž výraz $P(\boldsymbol{\theta} \in \mathbf{B}|\mathbf{y})$ a na pravej strane $P(\boldsymbol{\eta} \in \mathbf{H}(\mathbf{B})|\mathbf{y}) = P(\boldsymbol{\theta} \in \mathbf{B}|\mathbf{y})$. Pri Jeffreysovej voľbe apriórnej hustoty parametrov $\boldsymbol{\theta}$ a $\boldsymbol{\eta}$ sú obe tieto aposteriórne pravdepodobnosti rovnaké a nemožno dôjsť k paradoxnému výsledku ako v prípade princípu neurčitosti.

Jeffreys v r. 1946 [An invariant form for the prior probability in estimation problems. Proc. Roy. Soc. A, **186**, 453-461] navrhol "univerzálnu" apriórnu hustotu (vzhľadom na Lebesgueovu mieru)

$$\pi_J(\boldsymbol{\theta}) \propto [det(\mathbf{M}(\boldsymbol{\theta}))]^{\frac{1}{2}}.$$

Jeffreys navrhol svoju hustotu len pre jednorozmerný parameter. V takomto prípade sa jeho hustota osvedčila, lebo vtedy vždy možno zmeniť parametrizáciu modelu tak, aby informačná matica (vlastne Fisherova miera informácie - číslo) bola konštantná, t.j. nezávislá od hodnoty parametra. Toto je možné len výnimočne, ak parameter $\boldsymbol{\theta}$ je mnohorozmerný. Jeffreysova apriórna hustota býva často nevlastná.

Príklad 14.1. Nech náhodná veličina Y má binomické rozdelenie pravdepodobnosti s pravdepodobnostnou funkciou $f(y|\theta)$

$$f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y = 0, 1, 2, ..., n$$

(n je známe celé nezáporné číslo). Nájdite Jeffreysovu apriórnu hustotu.

Riešenie: Binomické rozdelenie spĺňa predpoklady Vety 14.2. Dostávame

$$\frac{\partial \ln f(y|\theta)}{\partial\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta} = \frac{y-n\theta}{\theta(1-\theta)}.$$

Fisherova miera informácie je

$$M(\theta) = \mathcal{E}_\theta\left[\frac{\partial \ln f(Y|\theta)}{\partial\theta}\right]^2 = \frac{\mathcal{E}(Y-n\theta)^2}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)},$$

lebo $\mathcal{E}(Y) = n\theta$ a disperzia $\mathcal{D}(Y) = \mathcal{E}(Y - \mathcal{E}(Y))^2 = n\theta(1-\theta)$. Jeffreysova apriórna hustota je rovná $k\dfrac{1}{\sqrt{\theta(1-\theta)}}$, kde $k > 0$.

Keď zvolíme apriórnu hustotu pre parameter θ

$$\pi(\theta) = \sqrt{\frac{M(\theta)}{n}} = \frac{1}{\sqrt{\theta(1-\theta)}}$$

dostávame z (2.7) aposteriórnu hustotu parametra θ

$$\pi(\theta|y) = \frac{\pi(\theta)f(y|\theta)}{\int_0^1 \pi(\theta)f(y|\theta)d\theta} = \frac{\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta} = \frac{1}{B\left(y+\frac{1}{2}, n-y+\frac{1}{2}\right)}\theta^{y-\frac{1}{2}}(1-\theta)^{n-y-\frac{1}{2}}.$$

Majme teraz záujem o parameter $\beta = \theta^2$. Podľa Vety 13.1 je aposteriórna hustota parametra β rovná

$$\pi(\beta|y) = \pi(\sqrt{\beta}|y)\left|\frac{1}{2\sqrt{\beta}}\right| = \frac{1}{2B\left(y+\frac{1}{2}, n-y+\frac{1}{2}\right)}\beta^{\frac{2y-3}{4}}(1-\beta^{\frac{1}{2}})^{n-y-\frac{1}{2}}. \tag{14.4}$$

Ak vyjdeme z parametra β, má Y z príkladu 14.1 pravdepodobnostnú funkciu

$$f(y|\beta) = \binom{n}{y}\beta^{\frac{y}{2}}(1-\beta^{\frac{1}{2}})^{n-y}, \quad y = 0, 1, 2, ..., n$$

(n je známe celé nezáporné číslo). Platí

$$\frac{\partial \ln f(y|\beta)}{\partial \beta} = \frac{y}{2\beta} - \frac{n-y}{2\beta^{\frac{1}{2}}(1-\beta^{\frac{1}{2}})}$$

a Fisherova miera informácie je

$$M^*(\beta) = \mathcal{E}_\beta\left[\frac{(Y-n\beta^{\frac{1}{2}})^2}{4\beta^2(1-\beta^{\frac{1}{2}})^2}\right] = \frac{n}{4}\beta^{-\frac{3}{2}}(1-\beta^{\frac{1}{2}})^{-1}.$$

Pri apriórnej hustote

$$\pi(\beta) = 2\frac{[M^*(\beta)]^{\frac{1}{2}}}{\sqrt{n}}$$

z (2.7) dostávame opäť aposteriórnu hustotu (14.4).

Príklad 14.2. Majme náhodný výber z rozdelenia $N(\theta, \sigma^2)$. Teda $\mathbf{Y} \sim N_n(\mathbf{1}\theta, \sigma^2\mathbf{I}_{n,n})$. Aká je Jeffreysova apriórna hustota v prípade, že

a) parameter σ poznáme,

b) poznáme parameter θ, ale nepoznáme parameter σ,

c) nepoznáme ani θ ani σ.

Riešenie:

Hustota náhodného vektora $\mathbf{Y}$ je $f(\mathbf{y}|\theta,\sigma) = f(y_1, y_2, ..., y_n|\theta,\sigma) = \dfrac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\theta)^2}$. Platí tiež, že

$$\ln f(\mathbf{y}|\theta,\sigma) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\theta)^2.$$

a) Ak parameter σ poznáme, tak Fisherova miera informácie je rovná (podľa Poznámky nad Definíciou 14.2)

$$M(\theta) = -\mathcal{E}_\theta\left[\frac{\partial^2 \ln f(\mathbf{Y}|\theta,\sigma)}{\partial\theta^2}\right] = \frac{n}{\sigma^2}.$$

Pri známom σ to je konštanta a preto Jeffreysova apriórna hustota parametra θ je

$$\pi(\theta) = 1, \quad \theta \in \mathcal{R}.$$

b) Ak parameter θ poznáme, ale nepoznáme parameter σ, tak Fisherova miera informácie je rovná (podľa Poznámky nad Definíciou 14.2)

$$M(\sigma) = -\mathcal{E}_\sigma\left[\frac{\partial^2 \ln f(\mathbf{Y}|\theta,\sigma)}{\partial\sigma^2}\right] = \frac{2n}{\sigma^2}.$$

Preto Jeffreysova apriórna hustota parametra σ je

$$\pi(\sigma) = \frac{1}{\sigma}, \quad \sigma > 0.$$

c) Ak nepoznáme ani parameter θ ani parameter σ, tak prvky Fisherovej informačnej matice sú (podľa Poznámky nad Definíciou 14.2)

$$M_{11}(\theta) = -\mathcal{E}_\theta\left[\frac{\partial^2 \ln f(\mathbf{Y}|\theta,\sigma)}{\partial\theta^2}\right] = \frac{n}{\sigma^2},$$

$$M_{12}(\theta) = -\mathcal{E}_\theta \left[\frac{\partial^2 \ln f(\mathbf{Y}|\theta,\sigma)}{\partial\theta\partial\sigma} \right] = 0,$$
$$M_{22}(\theta) = -\mathcal{E}_\theta \left[\frac{\partial^2 \ln f(\mathbf{Y}|\theta,\sigma)}{\partial\sigma^2} \right] = \frac{2n}{\sigma^2},$$
$$M_{21}(\theta) = M_{12}(\theta).$$

Jeffreysova apriórna hustota je

$$\pi(\theta,\sigma) = \frac{1}{\sigma^2}, \quad \theta \in \mathcal{R}, \quad \sigma > 0.$$

Poznámka. J. Anděl v učebnici na str. 294 poznamenáva, že väčšinou v prípade c) predchádzajúceho príkladu sa za apriórnu hustotu volí

$$\pi^*(\theta,\sigma) = \frac{1}{\sigma}, \quad \theta \in \mathcal{R}, \quad \sigma > 0.$$

Zdôvodňuje sa to tým, že sa dá mnohokrát dopredu predpokladať, že θ a σ sú nezávislé a apriórna hustota $\pi^*(\theta,\sigma)$ sa rovná súčinu apriórnych hustôt $\pi(\theta)\pi(\sigma)$ (ako keby $\pi(\theta)$ a $\pi(\sigma)$ boli "obyčajné" hustoty).

Príklad 14.3. (Poissonovský experiment.) Nech Y je diskrétna náhodná veličina s poissonovským rozdelením pravdepodobnosti a pravdepodobnostnou funkciou

$$f(y|\theta) = \frac{\theta^y}{y!}e^{-\theta}, \quad y = 0,1,... \quad \theta > 0.$$

Platí

$$\ln f(y|\theta) = -\ln y! - \theta + y\ln\theta.$$

Preto

$$M(\theta) = -\mathcal{E}_\theta \left[\frac{\partial^2 \ln f(y|\theta)}{\partial\theta^2} \right] = \mathcal{E}_\theta \left[\frac{Y}{\theta^2} \right] = \frac{1}{\theta}.$$

Preto Jeffreysova apriórna hustota je $\pi_J(\theta) \propto \dfrac{1}{\sqrt{\theta}}$.

15. Hierarchizácia apriórneho rozdelenia

V niektorých prípadoch možno apriórnu hustotu zapísať v tvare

$$\pi(\boldsymbol{\theta}) = \int_{\boldsymbol{\Gamma}} \pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})\pi_2(\boldsymbol{\gamma})d\mu(\boldsymbol{\gamma}), \tag{15.1}$$

kde $\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})$ je hustota, ktorej tvar je známy, ale neznáme sú parametre $\boldsymbol{\gamma} \in \mathcal{G}$ (realizácie náhodného vektora $\boldsymbol{\Gamma}$), ako aj neznáma je hustota $\pi_2(\boldsymbol{\gamma})$ vzhľadom na $\sigma-$konečnú mieru μ (my uvažujeme alebo Lebesgueovu mieru alebo sčítaciu mieru).

Poznámka. Upozorňujeme len, že "všeobecný" zápis (15.1) chápeme:

(i) Ak je $\boldsymbol{\Theta}$ spojitá aj $\boldsymbol{\Gamma}$ spojitá, tak $\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})$ je hustota známeho tvaru a $\pi_2(\boldsymbol{\gamma})$ je neznáma (apriorná) hustota náhodného vektora $\boldsymbol{\Gamma}$. Táto je nenulová na $\mathcal{G}$. Potom

$$\pi(\boldsymbol{\theta}) = \int_{\mathcal{G}} \pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})\pi_2(\boldsymbol{\gamma})d\boldsymbol{\gamma}. \tag{15.2}$$

(ii) Ak je $\boldsymbol{\Theta}$ spojitá a $\boldsymbol{\Gamma}$ diskrétna s hodnotami $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ...\}$, tak $\{\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma}_i)\}_{i\geq 1}$ sú hustoty známeho tvaru a $\{\pi_2(\boldsymbol{\gamma}_j)\}_{j\geq 1}$ je neznáma (apriorná) pravdepodobnostná funkcia parametra $\boldsymbol{\gamma} \in \{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ...\}$. Potom

$$\pi(\boldsymbol{\theta}) = \sum_{i\geq 1} \pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma}_i)\pi_2(\boldsymbol{\gamma}_i). \tag{15.3}$$

(iii) Ak je $\mathbf{\Theta}$ diskrétna s hodnotami $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ a $\mathbf{\Gamma}$ spojitá, tak $\{\pi_1(\boldsymbol{\theta}_j|\boldsymbol{\gamma})\}_{j\geq 1}$ je pravdepodobnostná funkcia známeho tvaru a $\pi_2(\boldsymbol{\gamma})$ je neznáma (apriorná) hustota náhodného vektora $\mathbf{\Gamma}$. Táto je nenulová na $\mathcal{G}$. Potom

$$\{\pi(\boldsymbol{\theta}_j)\}_{j\geq 1} = \left\{ \int_{\mathcal{G}} \pi_1(\boldsymbol{\theta}_j|\boldsymbol{\gamma})\pi_2(\boldsymbol{\gamma})d\boldsymbol{\gamma} \right\}_{j\geq 1}. \tag{15.4}$$

(iv) Ak je $\mathbf{\Theta}$ diskrétna s hodnotami $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ a $\mathbf{\Gamma}$ diskrétna s hodnotami $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ...\}$, tak $\{\pi_1(\boldsymbol{\theta}_j|\boldsymbol{\gamma}_i)\}_{j\geq 1}$ sú pravdepodobnostné funkcie známeho tvaru (pre každé $i \in \{1, 2, ...\}$) a $\{\pi_2(\boldsymbol{\gamma}_i)\}_{i\geq 1}$ je neznáma (apriorná) pravdepodobnostná funkcia parametra $\boldsymbol{\gamma} \in \{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ...\}$. Potom

$$\{\pi(\boldsymbol{\theta}_j)\}_{j\geq 1} = \left\{ \sum_{i\geq 1} \pi_1(\boldsymbol{\theta}_j|\boldsymbol{\gamma}_i)\pi_2(\boldsymbol{\gamma}_i) \right\}_{j\geq 1}. \tag{15.5}$$

Vzťah (15.1) znamená hierarchickú štrukturalizáciu. Hustota $\pi_2(\boldsymbol{\gamma})$ je primárna a hustota $\pi(\boldsymbol{\theta})$ je sekundárna v tejto chierarchizácii. Pôvodne sme mali experiment

$$\{f(\mathbf{y}|\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \mathbf{\Omega}\}$$

a teraz uvažujeme experiment

$$\{g(\mathbf{y}|\gamma) : \ \gamma \in \mathcal{G}\},$$

kde

$$g(\mathbf{y}|\boldsymbol{\gamma}) = \int_{\mathbf{\Omega}} f(\mathbf{y}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})d\nu(\boldsymbol{\theta}).$$

V tomto experimente je apriórna hustota $\pi_2(\boldsymbol{\gamma})$. Táto sa volí rovnomerná (t.j. podľa princípu neurčitosti), alebo akákoľvek iná. Jaj voľba nie je tak kritická, ako voľba apriórnej hustoty $\pi(\boldsymbol{\theta})$ v pôvodnom experimente $\{f(\mathbf{y}|\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \mathbf{\Omega}\}$.

Príklad 15.1. (Podľa knižky A. Pázmana, str. 9.) Novovyvinutý prístroj (alebo metóda) umožňuje detekovať chorobu, resp. neprítomnosť choroby s 95% pravdepodobnosťou. To znamená, že zo 100 pacientov, ktorým prístroj indikoval chorobu, približne 5 pacientov touto chorobou netrpí. Naopak, zo 100 pacientov, ktorým prístroj indikoval, že sú zdraví, v skutočnosti približne 5 je chorých. Náhodná veličina $Y-$indikácia choroby prístrojom má realizácie

y_1- prístroj chorobu indikoval

alebo

y_2- prístroj chorobu neindikoval.

Parameter θ nadobúda hodnoty

θ_1- vyšetrovaná osoba je chorá

alebo

θ_2- vyšetrovaná osoba je zdravá.

Teda $\pi(\theta)$ je pravdepodobnostná funkcia popisujúca apriórne rozdelenie pravdepodobnosti ochorenia v experimente $\{f(y|\theta) : \ \boldsymbol{\theta} \in \{\theta_1, \theta_2\}\}$. Zo zadania platí

$$f(y_1|\theta_1) = f(y_2|\theta_2) = 0,95 \quad \text{a} \quad f(y_2|\theta_1) = f(y_1|\theta_2) = 0,05.$$

Ľudia sú kategorizovaní podľa určitých predispozícií k chorobe na s kategórií $\gamma_1, ..., \gamma_s$ a podľa štatistík pravdepodobnostná funkcia ochorenia v jednotlivých kategóriách je

$$\{\pi_1(\theta_j|\gamma_i)\}_{j=1}^2, \quad i = 1, 2, ..., s,$$

Toto je apriórne rozdelenie pravdepodobnosti ochorenia v niektorej z kategórií. Nepoznáme však apriórnu primárnu hustotu $\pi_2(\gamma)$ na $\Gamma = \{1, 2, ..., s\}$. Môže sa zvoliť rovnomerná. Namiesto experimentu $\{f(y|\theta) : \Theta \in \{\theta_1, \theta_2\}\}$ s (neznámou) apriórnou pravdepodobnosťou $\pi(\theta)$ budeme teraz uvažovať experiment $\{g(y|\gamma) : \gamma \in \{1, 2, ..., s\}\}$, pričom

$$\{g(y_i|\gamma)\}_{i=1}^{2} = \left\{ \sum_{j=1}^{2} f(y_i|\theta_j)\pi_1(\theta_j|\gamma) \right\}_{i=1}^{2}.$$

Apriórna hustota v tomto experimente je $\pi_2(\lambda)$ (rovnomerná na $\{1, 2, ..., s\}$. Predchádzajúce úvahy zhrnieme vo vete

Veta 15.1. (Pázman, str. 47.) Ak platí (15.1), tak aj aposteriórna hustota má hierarchickú štruktúru

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \int_{\mathcal{G}} \pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma}, \mathbf{y})\pi_2(\boldsymbol{\gamma}|\mathbf{y})d\mu(\boldsymbol{\gamma}),$$

kde

$$\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma}, \mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})}{\int_{\boldsymbol{\Omega}} f(\mathbf{y}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta}|\boldsymbol{\gamma})d\nu(\boldsymbol{\theta})}$$

je aposteriórna hustota pre $\boldsymbol{\theta}$ pri daných parametroch $\boldsymbol{\gamma}$ a kde

$$\pi_2(\boldsymbol{\gamma}|\mathbf{y}) = \frac{g(\mathbf{y}|\boldsymbol{\gamma})\pi_2(\boldsymbol{\gamma})}{\int_{\mathcal{G}} g(\mathbf{y}|\boldsymbol{\gamma})\pi_2(\boldsymbol{\gamma})d\mu(\boldsymbol{\gamma})}.$$

16. Empirické metódy určenia apriórnej hustoty

Pre predikčnú hustotu platí (pozri časť Inferencia v 2. kapitole)

$$f^*(\mathbf{y}) = \int_{\boldsymbol{\Omega}} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta}). \tag{16.1}$$

Keby sme poznali $f^*(\mathbf{y})$, riešením integrálnej rovnice (16.1) by sme mohli dostať apriórnu hustotu $\pi(\boldsymbol{\theta})$. Z náhodného výberu $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n$ vieme len odhadnúť hustotu (resp. pravdepodobnostnú funkciu) $f^*(\mathbf{y})$ (napríklad jadrovými odhadmi hustoty). Potom môžeme riešiť integrálnu rovnicu (16.1) a takto získať odhad $\hat{\pi}(\boldsymbol{\theta})$ apriórnej hustoty $\pi(\boldsymbol{\theta})$. Pri tomto postupe odhadovanie predikčnej hustoty sa obyčajne nerobí bayesovskými metódami, preto tento postup nie je (obyčajne) čisto bayesovský.

V prípade, že v bayesovských metódach použijeme namiesto apriórnej hustoty $\pi(\boldsymbol{\theta})$ jej odhad $\hat{\pi}(\boldsymbol{\theta})$, hovoríme o *empirických bayesovských metódach*. Treba tu zdôrazniť dva fakty:

– Odhad $\hat{\pi}(\boldsymbol{\theta})$ treba získať z pozorovaní, ktoré nezávisia od pozorovaní v experimente, v ktorom používame bayesovské metódy.

– Odhad $\hat{\pi}(\boldsymbol{\theta})$ možno získať len z takých pozorovaní, v ktorých sa parameter $\boldsymbol{\theta}$ náhodne mení podľa $\pi(\boldsymbol{\theta})$. Nie je možné preto odhadovať subjektívnu apriórnu pravdepodobnosť.

Príklad 16.1. (Podľa knižky A. Pázmana, str. 48.) Nech $Y_1, Y_2, ..., Y_n$ je náhodný výber z Poissonovho rozdelenia s pravdepodobnostnou funkciou

$$f(y|\theta) = e^{-\theta}\frac{\theta^y}{y!}, \quad y = 0, 1, 2, ... \, , \tag{16.2}$$

pričom parameter Θ sa náhodne mení podľa hustoty $\pi(\theta)$ (vzhľadom k Lebesgueovej miere) a $P(\Theta \in (0, \infty)) = 1$. Ak aj pozorovanie Y_{n+1} je poissonovsky rozdelené s parametrom θ (realizáciou náhodnej premennej $\mathbf{\Theta}$) a nezávislé od $Y_1, Y_2, ..., Y_n$, aposteriórne rozdelenie (hustota) pri získanom pozorovaní y_{n+1}, teda $\pi(\theta|y_{n+1})$ je podľa (2.7) rovné

$$\pi(\theta|y_{n+1}) = \frac{\pi(\theta)f(y_{n+1}|\theta)}{\int_0^\infty \pi(\theta)f(y_{n+1}|\theta)d\theta}$$

a jeho stredná hodnota $\mathcal{E}\left[\pi(\theta|y_{n+1})\right]$ je

$$\mathcal{E}\left[\pi(\theta|y_{n+1})\right] = \frac{\int_0^\infty \theta\pi(\theta)f(y_{n+1}|\theta)d\theta}{\int_0^\infty \pi(\theta)f(y_{n+1}|\theta)d\theta}. \tag{16.3}$$

Táto stredná hodnota sa niekedy považuje za bayesovský odhad parametra θ. Po dosadení podľa (16.2) do (16.3) dostávame

$$\mathcal{E}\left[\pi(\theta|y_{n+1})\right] = \frac{(y_{n+1}+1)\int_0^\infty \pi(\theta)e^{-\theta}\dfrac{\theta^{y_{n+1}+1}}{(y_{n+1}+1)!}d\theta}{\int_0^\infty \pi(\theta)e^{-\theta}\dfrac{\theta^{y_{n+1}}}{y_{n+1}!}d\theta} = \frac{f^*(y_{n+1}+1)}{f^*(y_n)}\,(y_{n+1}+1). \tag{16.4}$$

Výraz (16.4) vieme spočítať len keď $\pi(\theta)$ poznáme. Ak $\pi(\theta)$ nepoznáme a n je dosť veľké číslo, postupujeme nasledovne:

Označme $n(y)$ počet výskytov hodnoty y v n nezávislých pozorovaniach $Y_1, Y_2, ..., Y_n$. $\dfrac{n(y)}{n}$ je relatívna početnosť výskytu hodnoty y v týchto n nezávislých pozorovaniach a je to frekvenčný odhad pravdepodobnosti $f^*(y)$. Teda podľa (16.4) empiricky získaná aposteriórna stredná hodnota $\mathcal{E}\left[\pi(\theta|y_{n+1})\right]$ (odhad tejto strednej hodnoty) je

$$\hat{\mathcal{E}}\left[\pi(\theta|y_{n+1})\right] = \frac{n(y_{n+1}+1)}{n(y_n+1)}\,(y_{n+1}+1).$$

17. Konjugované systémy apriórnych hustôt

Konjugovaným systémom apriórnych hustôt, priradeným k danému štatistickému modelu experimentu $\{f(\mathbf{y}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbf{\Omega}\}$ nazývame taký parametrizovaný systém $\mathcal{P}$ apriórnych hustôt na $\mathbf{\Omega}$, že

a) ak $\pi(\cdot) \in \mathcal{P}$, tak aj aposteriórna hustota $\pi(\cdot|\mathbf{y}) \in \mathcal{P}$ pre každé $\mathbf{y} \in \mathcal{Y}$,

b) prechod od $\pi(\cdot) \in \mathcal{P}$ k $\pi(\cdot|\mathbf{y})$ sa deje jednoduchou, ľahko vypočítateľnou zmenou nejakých parametrov.

17.1. Konjugované systémy pre hustoty exponenciálneho typu

V užšom zmysle konjugovaný systém je priraďovaný hustotám, ktoré sú exponenciálneho typu, t.j. ktoré sú tvaru

$$f(\mathbf{y}|\boldsymbol{\theta}) = \exp\left\{-\psi(\mathbf{y}) + \mathbf{t}'(\mathbf{y})\boldsymbol{\gamma}(\boldsymbol{\theta}) - \kappa(\boldsymbol{\theta})\right\}, \tag{17.1}$$

kde $\mathbf{Y} \in \mathcal{Y} \subset \mathcal{R}^n,\ \mathbf{\Theta} \in \mathbf{\Omega} \subset \mathcal{R}^m,\ \mathbf{t} :\ \mathcal{Y}\ \to \mathcal{R}^k,\ \boldsymbol{\gamma} :\ \mathbf{\Omega}\ \to \mathcal{R}^k,\ \psi :\ \mathcal{Y}\ \to \mathcal{R},\ \kappa :\ \mathbf{\Omega}\ \to \mathcal{R}$ sú dané funkcie. Konjugované apriórne hustoty majú tvar

$$\pi_{\mathbf{c},l}(\boldsymbol{\theta}) = K(\mathbf{c},l)\exp\left\{\mathbf{c}'\boldsymbol{\gamma}(\boldsymbol{\theta}) - l\kappa(\boldsymbol{\theta})\right\}, \tag{17.1 a}$$

kde $\mathbf{c} \in \mathcal{R}^k,\ l \in \mathcal{R}$ sú také, aby

$$\int_{\mathbf{\Omega}} e^{\mathbf{c}'\boldsymbol{\gamma}(\boldsymbol{\theta}) - l\kappa(\boldsymbol{\theta})} d\nu(\boldsymbol{\theta}) = [K(\mathbf{c},l)]^{-1} < \infty$$

a

$$[K(\mathbf{c}+\mathbf{t}(\mathbf{y}), l+1)]^{-1} < \infty$$

pre každé $\mathbf{y} \in \mathcal{Y}$. Po dosadení do Bayesovho vzorca sa presvedčte, že ak apriórna hustota je $\pi_{\mathbf{c},l}(\cdot)$, tak aposteriórna hustota je $\pi_{\mathbf{c}+\mathbf{t}(\mathbf{y}),l+1}(\cdot)$, teda v skutočnosti ide o veľmi jednoduchý prepočet parametrov

$$\mathbf{c} \to \mathbf{c}+\mathbf{t}(\mathbf{y}), \quad l \to l+1. \tag{17.2}$$

Problémom môže byť výpočet $K(\mathbf{c}, l)$. Metóda má význam hlavne ak rozdelenia $\pi_{\mathbf{c},l}(\cdot)$ patria k rozdeleniam známym z teórie pravdepodobnosti, kde normovací výraz $K(\mathbf{c}, l)$ už bol stanovený.

Systém konjugovaných apriórnych hustôt $\{\pi_{\mathbf{c},l}(\boldsymbol{\theta})\}$ priradený k hustotám exponenciálneho typu (17.1) možno podstatne rozšíriť, ak uvažujeme aj konvexné zmesi hustôt v tvare

$$\sum_{i=1}^{k} \omega_i \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}), \tag{17.3}$$

kde $\omega_i > 0$, $\sum_{i=1}^{k} \omega_i = 1$. Platí veta

Veta 17.1. (Pázman, str. 51.) Ak $\pi(\boldsymbol{\theta})$ má tvar (17.3), tak aposteriórna hustota je tiež konvexnou zmesou tvaru

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{k} \omega_i^*(\mathbf{y}) \pi_{\mathbf{c}_i+\mathbf{t}(\mathbf{y}),l_i+1}(\boldsymbol{\theta}),$$

kde koeficienty v tejto konvexnej kombinácii sú

$$\omega_i^*(\mathbf{y}) = \omega_i \frac{\dfrac{K(\mathbf{c}_i, l_i)}{K(\mathbf{c}_i+\mathbf{t}(\mathbf{y}), l_i+1)}}{\sum_{j=1}^{k} \omega_j \dfrac{K(\mathbf{c}_j, l_j)}{K(\mathbf{c}_j+\mathbf{t}(\mathbf{y}), l_j+1)}}.$$

Dôkaz:

Ak dosadíme (17.3) do Bayesovho vzorca (2.7), dostávame

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\sum_{i=1}^{k} \omega_i \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \sum_{j=1}^{k} \omega_j \pi_{\mathbf{c}_j,l_j}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})} = \sum_{i=1}^{k} \frac{\dfrac{\omega_i \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}}{\dfrac{\int_{\boldsymbol{\Omega}} \sum_{j=1}^{k} \omega_j \pi_{\mathbf{c}_j,l_j}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}} =$$

$$= \sum_{i=1}^{k} \frac{\omega_i \int_{\boldsymbol{\Omega}} \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}{\sum_{j=1}^{k} \omega_j \int_{\boldsymbol{\Omega}} \pi_{\mathbf{c}_j,l_j}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})} \; \frac{\pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi_{\mathbf{c}_i,l_i}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}.$$

Po dosadení (17.1) dostávame, že prvý zlomok je $\omega_i^*(\mathbf{y})$ a druhý je $\pi_{\mathbf{c}_i+\mathbf{t}(\mathbf{y}),l_i+1}(\boldsymbol{\theta})$. ♣

Pri "prepočte" parametrov $\mathbf{c}, l$ (17.2) má dôležitú úlohu postačujúca štatistika $\mathbf{t}(\mathbf{y})$. Metódu konštrukcie konjugovaného systému možno zovšeobecniť na prípad, keď hustoty $f(\mathbf{y}|\boldsymbol{\theta})$ nie sú exponenciálneho typu, ale existuje postačujúca štatistika (pozri v knižke J. Anděla, str. 283).

Príklad 17.1. (Binomický experiment, A. Pázman, str. 52.) Nech Y má binomické rozdelenie pravdepodobnosti s pravdepodobnostnou funkciou (hustotou vzhľadom k sčítacej miere)

$$f(y|\theta) = \binom{m}{y} \theta^y (1-\theta)^{m-y}, \quad y = 0, 1, ..., m\ ,$$

teda $\mathcal{Y} = \{0, 1, ..., m\}$, $\Theta \in \Omega = (0, 1)$, m je známe. Hustota $f(y|\theta)$ patrí do exponenciálnej triedy hustôt, lebo sa dá zapísať v tvare

$$f(y|\theta) = \exp\left\{\ln\binom{m}{y} + y\ln\frac{\theta}{1-\theta} + m\ln(1-\theta)\right\}.$$

Konjugovaná apriórna hustota (voči Lebesgueovej miere) na Ω má podľa (17.1 a) tvar

$$\pi_{c,l}(\theta) = K(c,l)\theta^c(1-\theta)^{lm-c}.$$

Pre beta funkciu platí $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx,\ a > 0,\ b > 0$ a $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Aby $\pi_{c,l}(\theta)$ bola hustota, musí platiť

$$[K(c,l)]^{-1} = \int_0^1 \theta^c(1-\theta)^{lm-c}d\theta = \int_0^1 \theta^{(c+1)-1}(1-\theta)^{(lm-c+1)-1}d\theta = B(c+1, lm-c+1) = \frac{\Gamma(c+1)\Gamma(lm-c+1)}{\Gamma(lm+2)}.$$

Pravidlá "prepočítavania" parametrov sú dané vzťahom (17.2), preto v tomto prípade

$$c \to c + y, \quad l \to l + 1.$$

a aposteriórna hustota je

$$\pi(\theta|y) = \pi_{c+y,l+1}(\theta) = K(c+y, l+1)\theta^{c+y}(1-\theta)^{(l+1)m-c-y} =$$
$$= \frac{\Gamma((l+1)m+2)}{\Gamma(c+y+1)\Gamma((l+1)m-c-y+1)}\ \theta^{c+y}(1-\theta)^{(l+1)m-c-y}.$$

Z tohoto vzťahu vieme napríklad spočítať aposteriórnu strednú hodnotu

$$\mathcal{E}[\pi(\theta|y)] = \frac{c+y+1}{(l+1)m+2},$$

čo je bayesovský odhad (vlastne jeho realizácia) parametra θ (pozri nižšie). V praxi je postup taký, že v súlade s kapitolou 16 nájdeme odhady neznámych parametrov apriornej hustoty $\pi(\theta)$ z konjugovaného systému hustôt (obyčajne nie bayesovskými postupmi) a neznáme parametre nahradíme ich odhadmi. V tomto príklade potom dostávame

$$\mathcal{E}[\widehat{\pi(\theta|y)}] = \frac{\hat{c}+y+1}{(\hat{l}+1)m+2}.$$

Príklad 17.2. (Poissonovský experiment, A. Pázman, str. 53.) Nech Y má Poissonovo rozdelenie pravdepodobnosti s pravdepodobnostnou funkciou (hustotou vzhľadom k sčítacej miere)

$$f(y|\theta) = e^{-\theta}\frac{\theta^y}{y!} \propto e^{-\theta+y\ln\theta}, \quad y = 0, 1, 2, \ldots,$$

teda $\mathcal{Y} = \{0, 1, 2, ..., \}$, $\Theta \in \Omega = (0, \infty)$. Hustota $f(y|\theta)$ patrí do exponenciálnej triedy hustôt. Konjugovaná apriórna hustota (voči Lebesgueovej miere) na Ω má podľa (17.1 a) tvar

$$\pi_{c,l}(\theta) = K(c,l)e^{-l\theta}\theta^c.$$

Pre gama funkciu platí $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx,\ a > 0$. Preto

$$[K(c,l)]^{-1} = \int_0^\infty e^{-l\theta}\theta^c d\theta = \frac{1}{l^{c+1}}\int_0^\infty e^{-u}u^{(c+1)-1}du = \frac{\Gamma(c+1)}{l^{c+1}}$$

pre $c > -1,\ l > 0$. Pravidlá "prepočítavania" parametrov sú dané vzťahom (17.2), preto v tomto prípade

$$c \to c + y, \qquad l \to l + 1.$$

a aposteriórna hustota je

$$\pi(\theta|y) = \pi_{c+y,l+1}(\theta) = K(c+y, l+1)e^{-(l+1)\theta}\theta^{c+y} = \frac{(l+1)^{c+y+1}}{\Gamma(c+y+1)}\ e^{-(l+1)\theta}\theta^{c+y}.$$

Z tohoto vzťahu vieme napríklad spočítať aposteriórnu strednú hodnotu

$$\mathcal{E}[\pi(\theta|y)] = \frac{c+y+1}{l+1},$$

čo je bayesovský odhad (vlastne jeho realizácia) parametra θ (pozri nižšie). V praxu opäť neznáme parametre c a l nahradíme ich (nie bayesovskými) odhadmi získanými z "iného" experimentu. Teda dostávame

$$\mathcal{E}[\widehat{\pi(\theta|y)}] = \frac{\hat{c}+y+1}{\hat{l}+1}.$$

Príklad 17.3. (A. Pázman, str. 54.) Nech $Y_1, Y_2, ..., Y_n$ je náhodný výber z rovnomerného rozdelenia na $(0, \theta)$, teda $\Theta \in \Omega = (0, \infty)$. $Y_i \in \mathcal{Y} = (0, \infty)$ má hustotu (vzhľadom k Lebesgueovej miere)

$$f(y_i|\theta) = \begin{cases} \theta^{-1}, & \text{ak } \ y_i \in (0, \theta), \\ 0, & \text{ak } \ y_i \in \langle\theta, \infty). \end{cases}$$

Hustota náhodného vektora $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ sa nedá zapísať v tvare (17.1), ale keď definujeme

$$t(\mathbf{y}) = \max_i y_i,$$

môžeme písať

$$f(\mathbf{y}|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \theta^{-n}\mathcal{I}_{(t(\mathbf{y}),\infty)}(\theta),$$

kde

$$\mathcal{I}_{(a,b)}(\theta) = \begin{cases} 1, & \text{ak } \ \theta \in (a, b), \\ 0, & \text{ak } \ \theta \notin (a, b). \end{cases}$$

Len podotýkame, že $t(\mathbf{Y})$ je postačujúca štatistika. Konjugovanú apriórnu hustotu zvolíme

$$\pi_{c,l}(\theta) = (l-1)K(c, l)c^{l-1}\theta^{-l}\mathcal{I}_{(c,\infty)}(\theta) \tag{17.4}$$

(presvedčte sa, že $\pi_{c,l}(\theta)$ je vskutku hustota). Preto

$$f(\mathbf{y}|\theta)\pi_{c,l}(\theta) \propto \theta^{-(n+l)}\mathcal{I}_{(\max\{c,t(\mathbf{y})\},\infty)}(\theta). \tag{17.5}$$

Zo vzťahov (17.4) a (17.5) vidíme pravidlá pre "prepočítavanie" parametrov

$$c \to \max\{c, t(\mathbf{y})\}, \qquad l \to l + n.$$

Príklad 17.4. Majme náhodný výber $Y_1, Y_2, ..., Y_n$ z rozdelenia $N(\theta, \sigma^2)$, pričom σ^2 poznáme. Parameter $\Theta \in \Omega = \mathcal{R}$. Podľa Príkladu 13.2 je hustota náhodného vektora $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ rovná

$$f(y_1, y_2, ..., y_n|\theta) = f(\mathbf{y}|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\overline{y})^2}\ e^{-\frac{n}{2\sigma^2}(\overline{y}-\theta)^2}.$$

Ako konjugovaný systém sa v tomto prípade v praxi berie množina všetkých rozdelení $N(a, b^2)$, kde $a \in \mathcal{R},\ b > 0$. Teda

$$\pi_{a,b}(\theta) = \frac{1}{\sqrt{2\pi}\ b} e^{-\frac{1}{2b^2}(\theta-a)^2}.$$

Preto hustota aposteriórneho rozdelenia parametra Θ je

$$\pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi_{a,b}(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{y}|\theta)\pi_{a,b}(\theta)d\theta} \propto e^{-\frac{n}{2\sigma^2}(\theta-\overline{y})^2} e^{-\frac{1}{2b^2}(\theta-a)^2}.$$

Úpravou dostaneme

$$-\frac{n}{2\sigma^2}(\theta-\overline{y})^2 - \frac{1}{2b^2}(\theta-a)^2 = -\frac{\left[\theta - \left(\frac{nb^2\overline{y}+a\sigma^2}{nb^2+\sigma^2}\right)\right]^2}{2\left(\frac{b^2\sigma^2}{nb^2+\sigma^2}\right)},$$

čiže

$$\pi(\theta|\mathbf{y}) \sim N\left(\frac{nb^2\overline{y}+a\sigma^2}{nb^2+\sigma^2}, \frac{b^2\sigma^2}{nb^2+\sigma^2}\right).$$

Stredná hodnota aposteriórneho rozdelenia je váženým priemerom parametra a (strednej hodnoty apriórneho rozdelenia, ktorú poznáme pred experimentom) a odhadu $\overline{y}$ založeného len na výberových hodnotách (na hodnotách získaných v experimente). Disperzia aposteriórneho rozdelenia je menšia než disperzia apriórneho rozdelenia b^2 (experimentom zmenšujeme rozptyl). Disperzia aposteriórneho rozdelenia je ale tiež menšia ako $\dfrac{\sigma^2}{n}$, čo je disperzia $\overline{Y}$. Keby sme čerpali informácie o θ len z experimentu, tak najlepší nestranný odhad parametra θ by bol $\overline{Y}$. Bayesovský odhad využíva apriórnu informáciu a je "presnejší" ako ten najlepší frekventistický odhad, ktorý apriórnu informáciu nevyužíva.

Ukážme si ešte príklady viacrozmerných konjugovaných systémov.

Príklad 17.5. (Normálne rozdelenie s neznámou strednou hodnotou a disperziou, A. Pázman, str. 55.) Nech $Y_1, ..., Y_n$ je náhodný výber z rozdelenia $N(\theta_1, \frac{1}{\theta_2})$, teda neznáma stredná hodnota je $\theta_1 \in \mathcal{R}$ a neznáma inverzná disperzia je $\theta_2 \in (0, \infty)$, $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)' \in \mathcal{R} \times (0, \infty)$. Hustota Y_i je

$$f(y_i|\theta_1, \theta_2) = f(y_i|\boldsymbol{\theta}) = \sqrt{\frac{\theta_2}{2\pi}} \exp\left\{-\frac{\theta_2}{2}(y_i-\theta_1)^2\right\}$$

a preto

$$f(y_1, y_2, ..., y_n|\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta}) = \frac{\theta_2^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^n e^{-\frac{\theta_2}{2}y_i^2 + \theta_1\theta_2 y_i - \frac{\theta_1^2\theta_2}{2}} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{(\sum_{i=1}^n y_i,\ \sum_{i=1}^n y_i^2)\binom{\theta_1\theta_2}{-\frac{\theta_2}{2}} - \frac{n}{2}(-\ln\theta_2 + \theta_1^2\theta_2)\right\}.$$

Pri označení (17.1) je $\mathbf{t}(\mathbf{y}) = (\sum_{i=1}^n y_i,\ \sum_{i=1}^n y_i^2)'$, $\boldsymbol{\gamma}(\boldsymbol{\theta}) = (\theta_1\theta_2,\ -\frac{\theta_2}{2})'$, $\kappa(\boldsymbol{\theta}) = \frac{n}{2}(-\ln\theta_2 + \theta_1^2\theta_2)$ a pravidlá pre "prepočítavanie" parametrov sú podľa (17.2)

$$c_1 \to c_1 + \sum_{i=1}^n y_i, \quad c_2 \to c_2 + \sum_{i=1}^n y_i^2, \quad l \to l+1.$$

Ešte spočítame $K(\mathbf{c}, l)$. Podľa formuly pod (17.1 a)

$$[K(\mathbf{c}, l)]^{-1} = \int\int_{(-\infty,\infty)\times(0,\infty)} \theta_2^{\frac{nl}{2}} e^{c_1\theta_1\theta_2 - \frac{nl}{2}\theta_1^2\theta_2 - c_2\frac{\theta_2}{2}} d\theta_1 d\theta_2 =$$

$$= \int_0^{\infty} \left[\int_{-\infty}^{\infty} \theta_2^{\frac{nl}{2}} e^{c_1\theta_1\theta_2 - \frac{nl}{2}\theta_1^2\theta_2 - c_2\frac{\theta_2}{2}} d\theta_1 \right] d\theta_2 =$$

$$= \int_0^{\infty} \left[\int_{-\infty}^{\infty} \theta_2^{\frac{nl}{2}} e^{-\frac{\theta_2}{2}\left[nl\left(\theta_1 - \frac{c_1}{nl}\right)^2 - \frac{c_1^2}{nl} + c_2\right]} d\theta_1 \right] d\theta_2 =$$

$$= \int_0^{\infty} \theta_2^{\frac{nl}{2}} e^{-\frac{\theta_2}{2}\left(c_2 - \frac{c_1^2}{nl}\right)} \left[\int_{-\infty}^{\infty} e^{-\frac{\theta_2 nl}{2}\left(\theta_1 - \frac{c_1}{nl}\right)^2} d\theta_1 \right] d\theta_2 =$$

$$= \int_0^{\infty} \theta_2^{\frac{nl}{2}} e^{-\frac{\theta_2}{2}\left(c_2 - \frac{c_1^2}{nl}\right)} \sqrt{\frac{2\pi}{\theta_2 nl}} \left[\int_{-\infty}^{\infty} \sqrt{\frac{\theta_2 nl}{2\pi}} e^{-\frac{\theta_2 nl}{2}\left(\theta_1 - \frac{c_1}{nl}\right)^2} d\theta_1 \right] d\theta_2 =$$

$$= \sqrt{\frac{2\pi}{nl}} \int_0^{\infty} \theta_2^{\frac{nl-1}{2}} e^{-\theta_2 \frac{c_2 nl - c_1^2}{2nl}} d\theta_2.$$

Z podmienok na parametre gama-hustoty dostávame

$$c_1 \in \mathcal{R}, c_2 > 0, l > 0, \qquad \text{alebo} \qquad c_1 \in \mathcal{R}, c_2 > \frac{c_1^2}{ln}, -\frac{1}{n} < l < 0.$$

Substitúciou

$$\frac{c_2 nl - c_1^2}{2nl}\,\theta_2 = u$$

dostávame z posledného integrálu

$$[K(\mathbf{c}, l)]^{-1} = \sqrt{\frac{2\pi}{nl}}\,\frac{\Gamma(p)}{a^p},$$

kde $a = \dfrac{c_2 - \frac{c_1^2}{nl}}{2}$ a $p = \dfrac{nl+1}{2}$.

Príklad 17.6. (Multinomický experiment, A. Pázman, str. 56.) Majme diskrétny náhodný vektor $\mathbf{Y} = (Y_1, Y_2, ..., Y_r)'$, ktorého zložky sú diskrétne náhodné veličiny. Môžu nadobúdať hodnoty $0, 1, ..., n$. Vektor $\mathbf{Y}$ má multinomické rozdelenie pravdepodobnosti s pravdepodobnostnou funkciou

$$f(y_1, ..., y_r|\theta_1, ..., \theta_r) = f(\mathbf{y}|\boldsymbol{\theta}) = P(Y_1 = y_1, ..., Y_r = y_r) = \frac{n!}{y_1!...y_r!}\theta_1^{y_1}...\theta_r^{y_r},$$

kde $y_i \in \{0, 1, ..., n\},\ \sum_{i=1}^r y_i = n,\ \theta_j \in (0, 1),\ j = 1, 2, ..., r,\ \sum_{j=1}^r \theta_j = 1$. Takouto hustotou (pravdepodobnostnou funkciou) je matematicky popísaný napríklad experiment, v ktorom robíme n nezávislých pokusov. Každý pokus može rezultovať v jednom z r výsledkov, pravdepodobnosť nastatia $j-$teho výsledku je θ_j. Náhodná veličina Y_i je počet nastatí $i-$teho výsledku v týchto n pokusoch. Konjugovaný systém hľadáme v tvare

$$\pi_{\mathbf{c}}(\boldsymbol{\theta}) = K(\mathbf{c}) \prod_{i=1}^r \theta_i^{c_i}$$

$(\mathbf{c} = (c_1, ..., c_r)',\ \mathcal{Y} = \{(y_1, y_2, ..., y_n) :\ y_i \in \{0, 1, ..., n\},\ \sum_{i=1}^r y_i = n\},\ \boldsymbol{\Omega} = \{\boldsymbol{\theta} = (\theta_1, ..., \theta_r) :\ \theta_j \in (0, 1), j = 1, 2, ..., r,\ \sum_{j=1}^r \theta_j = 1\}$. Pravidlá pre "prepočítavanie" parametrov sú

$$c_i \to c_i + y_i \ \ i = 1, 2, ..., r.$$

Rozdelenie tohoto tvaru je známe. Je to Dirichletovo rozdelenie s hustotou

$$\frac{\Gamma(\alpha_1 + ... + \alpha_r)}{\prod_{j=1}^r \Gamma(\alpha_j)} \prod_{i=1}^r \theta_i^{\alpha_i - 1}$$

(pozri napr. http://en.wikipedia.org/wiki/Dirichlet_distribution). Teraz už ľahko dostaneme vyjadrenie $K_{\mathbf{c}}$.

18. Aproximácie integrálov

Pri bayesovskej inferencii je kľúčové určenie aposteriórnej hustoty $\pi(\boldsymbol{\theta}|\mathbf{y})$. Túto dostávame zo vzorca (2.7), pričom musíme spočítať integrál $\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})$ (normovací faktor). V ďalšom budeme potrebovať spočítať strednú hodnotu aposteriórneho rozdelenia vektora $\boldsymbol{\Theta}$, čiže opäť len integrály $\int_{\boldsymbol{\Omega}} \theta_i \, \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})$, $i = 1, 2, ..., m$. Pokiaľ je dimenzia $\dim(\boldsymbol{\theta})$ malá, môžu sa použiť "bežné" numerické metódy približného výpočtu týchto integrálov. Načrtneme dve metódy vhodné pre $\dim(\boldsymbol{\theta}) = 1$. Obe metódy využívajú poznatok, že veľké počty pozorovaní je tvar funkcie vierohodnosti $f(\mathbf{y}|\boldsymbol{\theta})$ (ako funkcie $\boldsymbol{\theta}$ pri daných, nameraných $\mathbf{y}$) podobá tvaru hustoty normálneho rozdelenia. Teda uvažované integrály možno "rozumne" pretransformovať do tvaru

$$I = \int_{-\infty}^{\infty} \Phi(t) e^{-\frac{t^2}{2}} dt \tag{18.1}$$

s danou funkciou $\Phi(t)$.

18.1. Metóda kvadratúry (Hermiteova-Gaussova kvadratúra)

Táto metóda využíva ako pomocný nástroj systém Hermiteových polynómov $\{H_k(t)\}_{k=0}^{\infty}$, kde

$$H_k(t) = (-1)^k e^{t^2} \frac{d^k}{dt^k} e^{-t^2}.$$

Integrál (18.1) sa podľa tejto metódy rovná

$$I = \int_{-\infty}^{\infty} \Phi(t) e^{-\frac{t^2}{2}} dt = \sum_{j=1}^{k} H_j \Phi(t_j) + E,$$

pričom t_j, $j = 1, 2, ..., k$ sú korene Hermitovho polynómu $H_k(t)$ a H_j sú určené nasledovne

$$H_j = -\frac{2^{k+1} k! \sqrt{\pi}}{H_k'(t_j) H_{k+1}(t_j)}, \quad j = 1, 2, ..., k.$$

Chyba E sa rovná

$$E = \frac{k! \sqrt{\pi}}{2^k (2k)!} \Phi^{(2k)}(\eta).$$

Podrobnejšie pozri napr. v knižke Ralston, A., A first course in numerical analysis, MCGRAW-HILL, INC, 1965. Tam sú uvedené aj hodnoty t_j a H_j pre $j = 1, 2, 3, 4, 5$.

18.2. Metóda Laplaceovej transformácie

Metóda Laplaceovej transformácie využíva poznatok, že funkcia vierohodnosti $f(\mathbf{y}|\theta)$ (ako funkcia θ pri daných, nameraných $\mathbf{y}$) je pre veľké počty pozorovaní značne koncentrovaná okolo bodu

$$\hat{\theta} = \arg\max_{\theta} f(\mathbf{y}|\theta)$$

($\hat{\theta}$ je odhad získaný metódou maximálnej vierohodnosti). Touto metódou sa počíta integrál tvaru

$$I = \int_{-\infty}^{\infty} b(\theta) e^{-h(\theta)} d\theta, \tag{18.2}$$

kde $h(\theta) = -\ln f(\mathbf{y}|\theta)$. Funkcie $b(\theta)$ a $h(\theta)$ aproximujeme kvadratickými Taylorovými rozvojmi v okolí bodu $\hat{\theta}$:

$$b(\theta) = b(\hat{\theta}) + \left.\frac{db(\theta)}{d\theta}\right|_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \left.\frac{d^2 b(\theta)}{d\theta^2}\right|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 = b_0 + b_1(\theta - \hat{\theta}) + \frac{1}{2} b_2 (\theta - \hat{\theta})^2,$$

$$h(\theta) = h(\hat{\theta}) + \frac{1}{2} \left.\frac{d^2 h(\theta)}{d\theta^2}\right|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 = h_0 + \frac{1}{2} h_2 (\theta - \hat{\theta})^2.$$

Lineárny člen sme v druhom vzťahu vynechali, lebo $\left.\frac{dh(\theta)}{d\theta}\right|_{\theta=\hat{\theta}} = 0$. Integrál (18.2) aproximujeme integrálom

$$\hat{I} = e^{-h_0} \int_{-\infty}^{\infty} \left[b_0 + b_1(\theta - \hat{\theta}) + \frac{1}{2} b_2 (\theta - \hat{\theta})^2 \right] e^{-\frac{h_2}{2}(\theta - \hat{\theta})^2} d\theta =$$

$$= e^{-h_0} \sqrt{\frac{2\pi}{h_2}} \int_{-\infty}^{\infty} \left[b_0 + b_1(\theta - \hat{\theta}) + \frac{1}{2} b_2 (\theta - \hat{\theta})^2 \right] \sqrt{\frac{h_2}{2\pi}} e^{-\frac{h_2}{2}(\theta - \hat{\theta})^2} d\theta = \sqrt{\frac{2\pi}{h_2}} e^{-h_0} \left[b_0 + \frac{b_2}{2h_2} \right].$$

Metódu možno zovšeobecniť aproximáciou funkcie $b(\theta)$ rozvojom vyššieho rádu.

19. Niektoré simulačné metódy generovania nezávislých realizácií náhodnej veličiny

V mnohých prípadoch použitia bayesovských metód ide o určenie aposteriórnej hustoty, alebo o určenie strednej hodnoty aposteriórneho rozdelenia (teda o integrál). Základný princíp simulačných metód je generovanie postupnosti realizácií nezávislých náhodných veličín z nejakého (daného) rozdelenia. Hustotu potom aproximujeme histogramom, momenty aproximujeme výberovými priemermi. Tieto aproximácie sú tým lepšie, čím je väčší rozsah simulovaného súboru. Balíky počítačových programov obyčajne umožňujú generovať náhodný výber (resp. pseudonáhodný výber) z rovnomerného rozdelenia $R(0,1)$. Bližšie sa s mechanizmom činnosti generátora (pseudo)náhodných čísel a so spôsobmi testovania vytvorenej postupnosti môžeme oboznámiť napr. v skriptách Kalas, J., Pekár, J., Simulačné metódy, Bratislava: MFF UK, 1991. Tu uvedieme tri metódy. Viac o simulovaní náhodných veličín sa dočítate v knižke Antoch, J., Vorlíčková D., Vybrané metody statistické analýzy dat, ACADEMIA, Praha, 1992.

19.1 Metóda inverznej transformácie

Táto metóda umožňuje generovanie realizácií nezávislých (jednorozmerných) náhodných veličín $Y_1, Y_2, \ldots$ z rozdelenia aké má náhodná veličina Y s danou distribuňou funkciou $F_Y(y) = P(Y < y)$ a kvantilovu funkciou $F_Y^{-1}(u) = \inf\{y : \ F_Y(y) \geq u\},\ 0 < u < 1$. Toto rozdelenie môže byť diskrétne alebo aj spojité. Metóda je založená na nasledujúcom tvrdení:

Veta 19.1. Nech náhodná veličina U má rovnomerné rozdelenie na $(0,1)$. Nech $F_Y(y)$ je ľubovoľná distribučná funkcia. Potom náhodná veličina $Y = F_Y^{-1}(U)$ má rozdelenie s distribučnou funkciou $F_Y(y)$.

Dôkaz nájdeme v knižke J. Anděla, str. 6.

Postup pri generovaní je jednoduchý. Ak $x_1, x_2, \ldots$ je náhodný výber z rozdelenia $R(0,1)$, tak $F^{-1}(x_1)$, $F^{-1}(x_2), \ldots$ je náhodný výber z rozdelenia s (požadovanou) distribučnou funkciou $F_Y(y)$.

19.2 Zamietacia metóda

Metóda je založená na tvrdení:

Veta 19.2. Nech $a > 0$. Náhodný vektor $\mathbf{U} \in \mathcal{R}^n$ má hustotu (vzhľadom k Lebesgueovej miere) $f_{\mathbf{U}}(\mathbf{u})$ a súčasne podmienená hustota (inej) náhodnej veličiny (jednorozmernej) V, teda $f_{V|\mathbf{U}}(v|\mathbf{u})$ je rovnomerná na intervale $(0, af(\mathbf{u}))$

$$f_{V|\mathbf{U}}(v|\mathbf{u}) = \begin{cases} \dfrac{1}{af_{\mathbf{U}}(\mathbf{u})}, & \text{ak } \ v \in (0, af(\mathbf{u})), \\ 0, & \text{ak } \ v \notin (0, af(\mathbf{u})) \end{cases}$$

práve vtedy, keď náhodný vektor $(\mathbf{U}', V)' \in \mathcal{R}^{n+1}$ je rovnomerne rozdelený na množine

$$\left\{(\mathbf{u}', v)' \in \mathcal{R}^{n+1}: \ \mathbf{u} \in \mathcal{R}^n, \ f_{\mathbf{U}}(\mathbf{u}) > 0, \ v \in (0, af(\mathbf{u}))\right\}.$$

Dôkaz: nájdeme v knihe Devroye, L., Non-uniform random variate generation, Berlin, Springer, 1986.

Algoritmus generovania postupnosti vektorov $\mathbf{y}_1, \mathbf{y}_2, ... \in \mathcal{R}^n$ ktorá je náhodným výberom z rozdelenia s hustotou $f_{\mathbf{Y}}(\mathbf{y})$:

– Zvoľme nejakú hustotu $g_{\mathbf{U}}(\mathbf{u}), \ \mathbf{u} \in \mathcal{R}^n$, z ktorej vieme ľahko generovať náhodný výber, pričom táto hustota je taká, že existuje $a > 0$, že pre každé $\mathbf{u} \in \mathcal{R}^n$ platí $f_{\mathbf{Y}}(\mathbf{u}) \leq a g_{\mathbf{U}}(\mathbf{u})$.

– Generujme náhodný výber $\mathbf{u}_1, \mathbf{u}_2, ...$ zodpovedajúci hustote $g_{\mathbf{U}}(\mathbf{u})$.

– Na $i-$tom kroku generujeme x_i ako (jednorozmerný) jednobodový náhodný výber z rovnomerného rozdelenia $R(0, a g_{\mathbf{U}}(\mathbf{u}_i))$. Podľa Vety 19.2 to ale znamená, že body $(\mathbf{u}_i, x_i)$ sú nezávisle generované z rovnomerného rozdelenia na množine

$$G = \left\{(\mathbf{u}', x)' \in \mathcal{R}^{n+1}: \ 0 < x < ag(\mathbf{u})\right\}.$$

– Ak $x_i < f_{\mathbf{Y}}(\mathbf{u}_i)$, tak bod $\mathbf{u}_i$ zaradíme do hľadanej postupnosti $\{\mathbf{y}_j\}_{j \geq 1}$, ale ak $x_i \geq f_{\mathbf{Y}}(\mathbf{u}_i)$, tak bod $\mathbf{u}_i$ vyradíme z ďalšieho uvažovania.

– Je zrejmé, že takto vybrané vektory $(\mathbf{y}_i', x_i)'$ sú rovnomerne rozdelené na množine

$$F = \left\{(\mathbf{y}', x)' \in \mathcal{R}^{n+1}: \ f_{\mathbf{Y}}(\mathbf{y}) > 0, \ 0 < x < f_{\mathbf{Y}}(\mathbf{y})\right\},$$

pretože $F \subset G$, pričom pri danom $\mathbf{y}_j$ je bod x_j z rovnomerného rozdelenia $R(0, f_{\mathbf{Y}}(\mathbf{y}_j))$. Z Vety 19.2 vyplýva, že vybraná postupnosť vektorov $\mathbf{y}_1, \mathbf{y}_2, ...$ je náhodným výberom zodpovedajúcim hustote $f_{\mathbf{Y}}(\mathbf{y})$.

19.3 Kompozičná metóda

Táto metóda sa používa v prípade, že potrebujeme generovať nezávislé realizácie náhodného vektora s predikčnou hustotou $f^*(\mathbf{y}) = \int_{\mathbf{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})$ a vieme pomerne ľahko generovať body zodpovedajúce hustotám $f(\mathbf{y}|\boldsymbol{\theta})$ a $\pi(\boldsymbol{\theta})$. Algoritmus generovania postupnosti vektorov $\mathbf{y}_1, \mathbf{y}_2, ... \in \mathcal{R}^n$ ktorá je náhodným výberom z rozdelenia s hustotou $f^*(\mathbf{y})$:

– Generujeme náhodný výber $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...$ zodpovedajúci hustote $\pi(\boldsymbol{\theta})$.

– V $i-$tom kroku k už získanému $\boldsymbol{\theta}_i$ generujeme $\mathbf{y}_i$ ako jednoprvkový výber zodpovedajúci hustote $f(\mathbf{y}|\boldsymbol{\theta}_i)$.

– Takto získaná postupnosť $\mathbf{y}_1, \mathbf{y}_2, ...$ je náhodný výber zodpovedajúci hustote $f^*(\mathbf{y})$.

20. Monte-Carlo metóda integrovania

V horeuvadených, ale aj iných simulačných metódach generujeme nezávislé realizácie $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ toho istého náhodného vektora $\boldsymbol{\xi}$ (alebo, čo je správnejšie povedané - ale to isté, realizácie $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ nezávislých náhodných vektorov $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_n$ rovnako rozdelených ako náhodný vektor $\boldsymbol{\xi}$). Náhodný vektor $\boldsymbol{\xi}$ má hustotu $f(\mathbf{x})$. Ak potrebujeme vypočítať integrál typu

$$I = \int_{\mathbf{X}} \Phi(\mathbf{x}) f(\mathbf{x}) d\mu(\mathbf{x})$$

(obor hodnôt $\boldsymbol{\xi}$ je $\mathbf{X}$, hustota $f(\mathbf{x})$ je vzhľadom k $\sigma-$konečnej miere $\mu(\mathbf{x})$), tak tento aproximujeme sumou

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{x}_i).$$

Podľa silného zákona veľkých čísel s rastúcim n konverguje $\hat{I}_n$ k I. Presnejšie $\frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{X}_i)$ konverguje skoro iste k I.

21. MCMC metódy

Simulačné metódy z kapitoly 19.1 a 19.2 generujú nezávislé realizácie nejakej náhodnej premennej alebo náhodného vektora. Tieto metódy sa ukazujú ako nedostatočné, najmä ak dimenzia simulovaného vektora je veľká, alebo hustota simulovaného vektora je veľmi komplikovaná. Približne od roku 1990 sa v bayesovskej štatistike uplatňujú MCMC (Monte-Carlo Markov Chains) metódy. Ich základná myšlienka je v tom, že sa pomerne jednoduchými algoritmami vygeneruje ergodický markovovský proces s diskrétnym časom, ktorého stacionárne rozdelenie je práve to, ktoré potrebujeme simulovať. Teda negenerujú sa realizácie nezávislých náhodných vektorov, ale naopak, realizácie závislých náhodných vektorov.

V predchádzajúcom sme generovali realizácie $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...$ nezávislých náhodných vektorov $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, ...$, ktoré majú rovnaké rozdelenie s (aposteriórnou) hustotou $\pi(\boldsymbol{\theta}|\mathbf{y}) = \dfrac{\pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\nu(\boldsymbol{\theta})}$ a oborom hodnôt $\boldsymbol{\Omega}$, t.j. $P(\boldsymbol{\Theta}_i \in \boldsymbol{\Omega}) = 1$.

V ďalšom si označíme $\boldsymbol{\xi}(i)$ prvky náhodného vektorového procesu, $\mathbf{x}^{(i)}$ (alebo $\mathbf{y}^{(i)}$) realizácie náhodných vektorov $\boldsymbol{\xi}(i)$. Obor hodnôt generovaného náhodného vektora $\boldsymbol{\xi}$ označme $\mathbf{X}$. Jeho hustotu (v prípade, že má spojité rozdelenie) označme $f(\mathbf{x})$. Ak má diskrétne rozdelenie, jeho pravdepodobnostnú funkciu označme $p(\mathbf{x})$. Cieľom je generovať postupnosť vektorov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...$ patriacich do $\mathbf{X}$, ktoré sú realizáciami ergodického markovovského náhodného procesu $\boldsymbol{\xi}(1), \boldsymbol{\xi}(2), ...$ Stacionárne rozdelenie tohto náhodného procesu má hustotu $f(\mathbf{x})$ ak rozdelenie je spojité resp. pravdepodobnostnú funkciu $p(\mathbf{x})$ ak rozdelenie je diskrétne.

Voľne povedané, proces je ergodický, keď body generovanej postupnosti $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...$ sa po nejakom počte krokov priblížia ľubovoľne presne ku každému z bodov množiny $\{\mathbf{x} : \ f(\mathbf{x}) > 0\}$, ak stacionárne rozdelenie markovovského reťazca je spojté a $f(\cdot)$ je jeho hustota, alebo ku každému z bodov množiny $\{\mathbf{x} : \ p(\mathbf{x}) > 0\}$, ak stacionárne rozdelenie markovovského reťazca je diskrétne a $p(\cdot)$ je jeho pravdepodobnostná funkcia. Táto skutoňosť zaručí, že integrály tvaru $\int_{\mathbf{X}} \Phi(\mathbf{x}) f(\mathbf{x}) d\lambda(\mathbf{x})$ (resp. $\Phi(\mathbf{x}) \sum_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x})$) možno aproximovať sumou

$\frac{1}{n}\sum_{i=1}^{n}\Phi(\mathbf{x}^{(i)})$. Rozbor tohto problému patrí do teórie markovovských procesov (pozri napr. Robert, Ch., P. *Méthodes de Monte Carlo par Chaînnes de Markov*. Paris: Economica, 1996).

MCMC metódy možno rozdeliť do dvoch (hlavných) skupín:

a) metódy odvodené od algoritmu Hastinga a Metropolisa,

b) metódy odvodené od algoritmu Gibsa.

21.1. Algoritmus Hastinga a Metropolisa

Opis algoritmu. Zvolíme ľubovoľnú triedu podmienených hustôt (resp. podmienených pravdepodobnostných funkcií) na $\mathbf{X}$

$$\{q(\cdot|\mathbf{y}):\ \mathbf{y}\in\mathbf{X}\}$$

tak, aby platilo:

a) Ak pre $A\subset\mathbf{X}$ je $\int_A f(\mathbf{x})d\lambda(\mathbf{x})>0$ $\left(\sum_{\mathbf{x}\in A}p(\mathbf{x})>0\right)$, tak $\int_A q(\mathbf{x}|\mathbf{y})d\lambda(\mathbf{x})>0$ $\left(\sum_{\mathbf{x}\in A}q(\mathbf{x}|\mathbf{y})>0\right)$ pre každé $\mathbf{y}\in\mathbf{X}$.

b) Ľahko možno simulovať náhodné výbery zodpovedajúce hustote (pravdepodobnostnej funkcii) $q(\cdot|\mathbf{y})$.

c) existuje symetria v zmysle

$$q(\mathbf{x}|\mathbf{y})=q(\mathbf{y}|\mathbf{x});\quad \mathbf{x},\mathbf{y}\in\mathbf{X},$$

alebo je známy analytický výraz pre funkciu $(\mathbf{x},\mathbf{y})\in\mathbf{X}\times\mathbf{X}\to q(\mathbf{x}|\mathbf{y})$.

Hustoty (pravdepodobnostné funkcie) $q(\cdot|\mathbf{y})$ sú pomocné a nazývajú sa *inštrumentálne*. Ľubovôľa pri výbere inštrumentálnej hustoty je obmedzená požiadavkou, aby príslušný markovovský proces bol ergodický. Príkladom vhodnej hustoty je $q(\mathbf{x}|\mathbf{y})=g(\mathbf{y})$ nezávisle na $\mathbf{x}$ (pozri knižku A. Pázmana).

Nech $\mathbf{x}^{(n)}\in\mathbf{X}$ je vektor vygenerovaný algoritmom v $n-$tom kroku. Vektor $\mathbf{x}^{(n+1)}\in\mathbf{X}$ dostaneme v $(n+1)-$vom kroku takto:

1. Pri danom $\mathbf{x}^{(n)}$ generujeme vektor $\mathbf{y}^{(n)}\in\mathbf{X}$ zodpovedajúci hustote (pravdepodobnostnej funkcii) $q(\cdot|\mathbf{x}^{(n)})$.

2. Vypočítame výraz
$$R(\mathbf{x}^{(n)},\mathbf{y}^{(n)})=\min\left\{1,\frac{f(\mathbf{y}^{(n)})q(\mathbf{x}^{(n)}|\mathbf{y}^{(n)})}{f(\mathbf{x}^{(n)})q(\mathbf{y}^{(n)}|\mathbf{x}^{(n)})}\right\}.$$

3. S pravdepodobnosťou $R(\mathbf{x}^{(n)},\mathbf{y}^{(n)})$ zvolíme $\mathbf{x}^{(n+1)}=\mathbf{y}^{(n)}$ a s pravdepodobnosťou $1-R(\mathbf{x}^{(n)},\mathbf{y}^{(n)})$ zvolíme $\mathbf{x}^{(n+1)}=\mathbf{x}^{(n)}$.

Poznámka. Na stanovenie $R(\mathbf{x}^{(n)},\mathbf{y}^{(n)})$ nepotrebujeme vedieť normovací koeficient v hustote $f(\mathbf{x})$ (v pravdepodobnostnej funkcii $p(\mathbf{x})$). Toto využijeme ak požadovaná hustota (pravdepodobnostná funkcia) je aposteriórna, kde stačí použiť súčin $\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})$ namiesto $\pi(\boldsymbol{\theta}|\mathbf{y})$ (resp. $\pi(\boldsymbol{\theta}_i)f(\mathbf{y}|\boldsymbol{\theta}_i), i=1,2,...$ namiesto $\pi(\boldsymbol{\theta}_i|\mathbf{y}), i=1,2,...$) .

Výber bodu $\mathbf{x}^{(n+1)}$ v algoritme je náhodný a pravdepodobnosť, s ktorou je vyberaný, nezávisí od prdchádzajúcich $\mathbf{x}^{(i)}$, $i<n$. Dostaneme teda markovovský proces. Vzhľadom na značnú ľubovôľu pri výbere inštrumentálnych hustôt sa javí prekvapujúce, že práve hustota $f(\mathbf{x})$ zodpovedá stacionárnemu stavu tohto procesu. Práve táto ľubovôľa umožňuje vybrať inštrumentálne hustoty tak, aby sa stacionárny stav dosiahol čo najrýchlejšie, resp. výpočtovo najjednoduchšie. Označme $P(A|\mathbf{x})$ pravdepodobnosť, že proces sa (na ľubovoľnom) kroku dostane zo stavu $\mathbf{x}$ do stavu, ktorý patrí do množiny A. $P(A|\mathbf{x})$ voláme aj *prechodové jadro (transition probability kernel)*. Platí veta (Pázman, A., *Bayesovská štatistika*, str. 65)

Veta 21.1. Pri akejkoľvek voľbe inštrumentálnej hustoty (pravdepodobnostnej funkcie) spĺňajúcej podmienku a), algoritmus Hastinga a Metropolisa generuje realizáciu markovovského procesu a pre každú merateľnú množinu A platí

(i) ak $f(\mathbf{x})$ je hustota

$$\int_{\mathbf{X}} P(A|\mathbf{x})f(\mathbf{x})d\lambda(\mathbf{x}) = \int_A f(\mathbf{x})d\lambda(\mathbf{x}) \quad \text{pre každú merateľnú} \quad A \subset \mathbf{X}$$

($\lambda(\cdot)$ je Lebesgueova miera),

(ii) ak $p(\mathbf{x}_i),\ i = 1, 2, ...$ je pravdepodobnostná funkcia

$$\sum_{i \geq 1} P(A|\mathbf{x}_i)p(\mathbf{x}_i) = \sum_{\mathbf{x} \in A} p(\mathbf{x}) \quad \text{pre každú merateľnú} \quad A \subset \mathbf{X}.$$

Teda $f(\mathbf{x})$ je hustota stacionárneho rozdelenia ($p(\mathbf{x})$ je pravdepodobnostná funkcia stacionárneho rozdelenia) tohto procesu.

21.2. Algoritmus Gibsa

Tento algoritmus sa používa v mnohorozmernom prípade, keď:

– $\mathbf{X} = X_1 \times X_2 \times ... \times X_m$, pričom každé $\mathbf{x} \in \mathbf{X}$ možno rozložiť na m komponent $\mathbf{x} = (x_1, ..., x_m)'$ $x_i \in X_i,\ i = 1, 2, ..., m$. Predpokladáme, že $\lambda(A) = \lambda_1(A_1) \times ... \times \lambda_m(A_m)$, ak $A = A_1 \times ... \times A_m$, kde $A_i \subset X_i$. (Toto platí napríklad pre Lebesueovu mieru.)

– pre každé $i = 1, 2, ..., m$ sa dajú generovať náhodné výbery z množiny X_i zodpovedajúce podmieneným hustotám týchto komponentov, teda

$$f_i(\cdot|\{x_j\}_{j=1, j\neq i}^m).$$

Algoritmus predpisuje prechod od $\mathbf{x}^{(n)} = (x_1^{(n)}, ..., x_m^{(n)})'$ ku $\mathbf{x}^{(n+1)} = (x_1^{(n+1)}, ..., x_m^{(n+1)})'$ po jednotlivých komponentoch takto:

– generujeme výber $x_1^{(n+1)}$ z hustoty $f_1(\cdot|x_2^{(n)}, ..., x_m^{(n)})$,

– generujeme výber $x_2^{(n+1)}$ z hustoty $f_2(\cdot|x_1^{(n+1)}, x_3^{(n)}, ..., x_m^{(n)})$,

– generujeme výber $x_3(n+1)$ z hustoty $f_3(\cdot|x_1^{(n+1)}, x_2^{(n+1)}, x_4^{(n)} ..., x_m^{(n)})$,

atď,

– generujeme výber $x_m(n+1)$ z hustoty $f_m(\cdot|x_1^{(n+1)}, x_2^{(n+1)}, ..., x_{m-1}^{(n+1)})$.

Platí veta (Pázman, A., *Bayesovská štatistika*, str. 67)

Veta 21.1. Postupnosť $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...$ generovaná algoritmom Gibsa je realizáciou markovovského procesu a $f(\mathbf{x}) = f(x_1, ..., x_m)$ je hustota pravdepodobnosti stacionárneho rozdelenia tohto procesu. Dôkaz ergodičnosti príslušného markovovského procesu nájdeme napr. v knihe Robert, Ch., P. *Méthodes de Monte Carlo par Chaînnes de Markov*. Paris: Economica, 1996.

Poznámka. Poznamenávame len, že ak potrebujeme integrovať nejakú funkciu $h(x_i)$, ktorá závisí len od $i-$teho komponentu vektora $\mathbf{x}$

$$I = \int_{\mathbf{X}} h(x_i)f(\mathbf{x})d\lambda(\mathbf{x}) \quad (I = \sum_{\mathbf{x} \in \mathbf{X}} h(x_i)p(\mathbf{x}))$$

tak túto aproximujeme sumou

$$\hat{I} = \frac{1}{n}\sum_{k=1}^{n} h(x_i^{(k)}),$$

teda z postupnosti $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ používame iba príslušnú postupnosť $i-$tych komponentov: $x_i^{(1)}, x_i^{(2)}, \ldots$. Odtiaľ vyplýva **zovšeobecnený algoritmus Gibsa**: Vytvoríme nejaký priestor $\mathbf{Z}$ a nájdeme takú hustotu (pravdepodobostnú funkciu) $g(\mathbf{x}, \mathbf{z})$ na $\mathbf{X} \times \mathbf{Z}$, že

– hustota $f(\mathbf{x})$ je marginálna ku $g(\mathbf{x}, \mathbf{z})$,

– pomocou algoritmu Gibsa aplikovaného na $g(\mathbf{x}, \mathbf{z})$ generujeme postupnosť vektorov $(\mathbf{x}^{(1)}, \mathbf{z}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{z}^{(2)}), \ldots$,

– na aproximáciu integrálov typu

$$I = \int_{\mathbf{X}} \Phi(\mathbf{x}) f(\mathbf{x}) d\lambda(\mathbf{x}) \quad (I = \sum_{\mathbf{x} \in \mathbf{X}} \Phi(\mathbf{x}) p(\mathbf{x}))$$

použijeme sumu

$$\hat{I} = \frac{1}{n} \sum_{k=1}^{n} \Phi(x_i^{(k)})$$

a teda používame postupnosť $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$, ktorá zodpovedá hustote $f(\mathbf{x})$ (pravdepdobnostnej funkcii $p(\mathbf{x})$).

Takýto postup použijeme, keď generovanie podľa hustoty $g(\mathbf{x}, \mathbf{z})$ je jednoduchšie, než podľa hustoty $f(\mathbf{x})$ (pravdepodobnostnej funkcie $p(\mathbf{x})$). Je tu možnosť rôznej voľby $g(\mathbf{x}, \mathbf{z})$ a tým aj rôznych variantov algoritmu Gibsa.

22. Bayesovské bodové a intervalové odhady a testy pre jednorozmerný parameter

22.1. Bayesovské bodové a intervalové odhady pre jednorozmerný parameter

(Dve nasledujúce kapitoly sledujú text z knižky J.Anděla.) Nech $\dim \boldsymbol{\theta} = 1$. Za bodový bayesovský odhad parametra θ sa v niektorých prípadoch berie *stredná hodnota aposteriórneho rozdelenia* (presnejšie za realiźáciu bayesovského odhadu sa berie stredná hodnota aposteriórneho rozdelenia). V prípade príkladu z 3.kapitoly je hustota aposteriórneho rozdelenia

$$\pi(p|m) = \frac{1}{B(a+m, b+n-m)} p^{a+m-1}(1-p)^{b+n-m-1}, \quad 0 < p < 1$$

a stredná hodnota aposteriórneho rozdelenia

$$\int_0^1 \frac{1}{B(a+m, b+n-m)} p^{a+m-1}(1-p)^{b+n-m-1} p \, dp = \frac{B(a+m+1, b+n-m)}{B(a+m, b+n-m)} =$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a+m)\Gamma(b+n-m)} \frac{\Gamma(a+m+1)\Gamma(b+n-m)}{\Gamma(a+b+n+1)} = \frac{a+m}{a+b+n}.$$

Ak je n veľké (v zrovnaní s a, b), tak tento odhad je blízko hodnoty $\frac{m}{n}$. Toto je odhad parametra p bez ohľadu na apriórne rozdelenie). Je to v zhode so skúsenosťou: Ak n je veľké "veľa vieme" o p bez ohľadu na predchádzajúce vedomosti. Ak naopak nevieme nič o p, teda nerobíme žiaden výber - $m = 0, n = 0$, odhad je len z predchádzajúcich vedomostí, teda $\frac{a}{a+b}$ (stredná hodnota $B(a, b)$ rozdelenia).

Iná metóda je za bayesovský odhad zobrať *módus* aposteriórneho rozdelenia (presnejšie za realiźáciu bayesovského odhadu zobrať módus aposteriórneho rozdelenia). V tomto prípade

$$\max_p \frac{p^{a+m-1}(1-p)^{b+n-m-1}}{B(a+m, b+n-m)}.$$

Maximum spočítame štandardným spôsobom

$$\frac{d}{dp}\left(p^{a+m-1}(1-p)^{b+n-m-1}\right) = (a+m-1)p^{a+m-2}(1-p)^{b+n-m-1} - (b+n-m-1)p^{a+m-1}(1-p)^{b+n-m-2} =$$

$$= p^{a+m-2}(1-p)^{b+n-m-2}[(a+m-1)(1-p) - (b+n-m-1)p] = 0,$$

čiže

$$p(a+m-1+b+n-m-1) = a+m-1$$

a bayesovský odhad parametra p je v tomto prípade

$$\frac{a+m-1}{a+b+n-2}.$$

Na základe aposteriórnej hustoty $\pi(p|m)$ je možné zostrojiť aj bayesovský intervalový odhad parametra p. Nech čísla D a H $(0 \leq D < H \leq 1)$ spĺňajú podmienku

$$\int_D^H \pi(p|m)\ dp = 1 - \alpha, \tag{22.1}$$

kde α je dané číslo z intervalu $(0,1)$. Potom platí

$$P(D < p < H|m) = 1 - \alpha. \tag{22.2}$$

Interval (D, H) je *bayesovským intervalom spoľahlivosti* pre p. Samozrejme (22.1) neurčuje hranice D a H jednoznačne. Niekedy sa preto kladie požiadavka symetrie v tom zmysle, že má platiť

$$\int_0^D \pi(p|m)dp = \frac{\alpha}{2}, \qquad \int_H^1 \pi(p|m)dp = \frac{\alpha}{2}.$$

Ak existuje také p_0, že aposteriórna hustota je neklesajúca na $(0, p_0)$ a nerastúca na $(p_0, 1)$, volia sa niekedy D a H ako krajné body intervalu

$$I = \{p:\ \pi(p|m) > k\},$$

kde k je konštanta určená podmienkou

$$\int_I \pi(p|m)dp = 1 - \alpha.$$

22.2. Bayesovské testy v prípade jednorozmernáho parametra

Budeme pokračovať v úvahách z predchádzajúcej kapitoly. Ak už máme čísla D a H zvolené tak, aby spĺňovali podmienku (22.2), môžeme pristúpiť aj k bayesovskému testovaniu hypotéz o parametri p. Predpokladajme, že chceme testovať hypotézu

$$H_0:\ p \in A,$$

kde A je nejaký interval obsiahnutý v $(0,1)$. V prípade, že celý interval A padne mimo (D, H), hypotézu H_0 zamietame. V tomto prípade je zrejme aposteriórna pravdepodobnosť menšia alebo rovná α. Nebolo by však rozumné založiť test len na hodnote $P(A|m)$ (- pravdepodobnosť, že náhodná veličina s hustotou aposteriórneho rozdelenia $\pi(p|m)$ sa realizuje v intervale A) bez konfrontácie s intervalom (D, H). Keby bol totiž interval A veľmi krátky, mohla by byť pravdepodobnosť $P(A|m)$ veľmi malá dokonca aj v prípade, že by A obsahoval módus alebo strednú hodnotu aposteriórneho rozdelenia. To by viedlo k zamietnutiu H_0, hoci by poloha intervalu A vzhľadom k aposteriórnej hustote nebola nijako "podozrivá".

Niekedy sa hovorí o teste hypotézy

$$H_0: \; p = p_0,$$

kde $p_0 \in (0,1)$ je dané číslo. V tomto prípade A je uzavretý interval obsahujúci jediný bod p_0. Je jasné, že v tomto prípade je $P(A|m) = 0$, takže podmienka $P(A|m) \leq \alpha$ je splnená triviálne. O zamietnutí H_0 teda rozhoduje len poloha bodu p_0 vzhľadom k (D, H). Keby sme nemali nejakú podmienku vzťahujúcu sa k intervalu (D, H), hypotézu H_0 by sme vždy zamietli. Tu by sa dalo argumentovať, že hypotéza H_0 nie je rozumná, lebo dopredu vieme, že jej pravdepodobnosť je rovná nule. Keď však padne bod p_0 mimo $\langle D, H\rangle$, existuje také jeho okolie A_0, že je disjungtné s (D, H) a $P(A_0|m) = \int_{A_0} \pi(p|m)dp > 0$. Môže sa povedať, že ide vlastne o test hypotézy, že p je z nejakého okolia bodu p_0.

23. Štatistické rozhodovanie

(Kapitola sleduje knižku A. Pázmana: *Bayesovská štatistika.*) Teória štatistického rozhodovania je založená na ekonomickom princípe, podľa ktorého má rozhodovanie maximalizovať priemerný výnos, resp. minimalizovať priemernú stratu rozhodovania. Teória sa dá formulovať aj bez použitia apriórneho rozdelenia, teda nebayesovsky. Bayesovský prístup má tiež svoje opodstatnenie. Základy teórie sa dajú sformulovať dokonca aj bez pozorovaní (t.j. bez experimentu). Takýto prístup je síce značne obmedzujúci, v teórii však umožňuje elementárne definovať základné pojmy, ktoré potom možno ľahko preniesť na prípad rozhodovania s experimentom.

23.1. Základné pojmy rozhodovania bez experimentu

Základné pojmy si vysvetlime na elementárnom príklade: Treba rozhodnúť o spôsobe výstavby objektu v prípade, že máme pochybnosti o stave podložia, pričom nemáme možnosti ho preskúmať. Z pozície stavebníka sú dva možné stavy podložia:

θ_1: podložie je pevné, vhodné na stavbu,

θ_2: podložie je slabé a stavať sa môže len so zosilnenými základmi.

Rozhodovateľ (stavebník) má voliť medzi tromi rozhodnutiami:

a_1: nestavať vôbec,

a_2: stavať štandardným spôsobom,

a_3: stavať so zosilnenými základmi.

Označme $L(\theta, a)$ finančnú stratu, ktorú utrpí stavebník ak podložie je v stave θ a je prijaté rozhodnutie a. Hodnoty $L(\theta, a)$ sú v nasledujúcej tabuľke (záporná strata = zisk)

$a \setminus \theta$	θ_1	θ_2
a_1	5 000	0
a_2	−30 000	50 000
a_3	3 000	−20 000

Rozhodnutie a_3 je lepšie ako rozhodnutie a_1 pri akomkoľvek stave podložia. Rozhodnutia a_1 a a_2 sú neporovnateľné, kým nemáme apriórne (alebo nejaké experimentálne) informácie o stave podložia. Chápeme to v takom zmysle, že keď je podložie pevné (θ_1) lepšie je rozhodnúť sa pre a_2 (stavať štandardným spôsobom) a keď je podložie slabé (θ_2) lepšie je, keď sa rozhodneme pre a_1 (nestavať). Ak ale vieme s dosť veľkou apriórnou pravdepodobnosťou, že podložie môže byť v stave θ_2 (slabé), uprednostníme aj rozhodnutie a_3 pred a_2. V prípade, že nemáme apriórne informácie o stave podložia, rozhodnutie a_1 možno z ďalšieho rozhodovania vylúčiť.

Zovšeobecnime tieto úvahy: Nech $\mathbf{\Omega}$ je množina stavov prírody alebo (prozaickejšie) $\mathbf{\Omega}$ je **parametrický priestor**. Nech $\mathfrak{A}$ je **množina možných rozhodnutí**. Nech $L: \mathbf{\Omega} \times \mathfrak{A} \to \mathcal{R}$ je **stratová funkcia**, o ktorej budeme predpokladať, že je zdola ohraničená.

Rozhodnutie $\mathbf{a}_1$ považujeme za **ekvivalentné** s $\mathbf{a}_2$ (píšeme $\mathbf{a}_1 \sim \mathbf{a}_2$), ak pre každé $\boldsymbol{\theta} \in \mathbf{\Omega}$ platí $L(\boldsymbol{\theta}, \mathbf{a}_1) = L(\boldsymbol{\theta}, \mathbf{a}_2)$. Podobne $\mathbf{a}_1$ je **rovnomerne nie horšie** ako $\mathbf{a}_2$ (píšeme $\mathbf{a}_1 \preceq \mathbf{a}_2$), ak pre každé $\boldsymbol{\theta} \in \mathbf{\Omega}$ platí $L(\boldsymbol{\theta}, \mathbf{a}_1) \leq L(\boldsymbol{\theta}, \mathbf{a}_2)$. Ak však naviac platí, že existuje $\boldsymbol{\theta}^* \in \mathbf{\Omega}$ také, že $L(\boldsymbol{\theta}^*, \mathbf{a}_1) < L(\boldsymbol{\theta}^*, \mathbf{a}_2)$, potom $\mathbf{a}_1$ je **rovnomerne lepšie** ako $\mathbf{a}_2$ (píšeme $\mathbf{a}_1 \prec \mathbf{a}_2$). Zrejme relácia $\preceq$ definuje čistočné usporiadanie množiny rozhodnutí $\mathfrak{A}$. Nemusí existovať rovnomerne najlepšie rozhodnutie (pozri predchádzajúci príklad). Môžeme však použiť dôležité náhradné pojmy, a síce pojem prípustného rozhodnutia a pojem úplnej množiny rozhodnutí.

Definícia 23.1. Rozhodnutie $\mathbf{a} \in \mathfrak{A}$ je prípustné, ak neexistuje také $\mathbf{b} \in \mathfrak{A}$, že $\mathbf{b} \prec \mathbf{a}$. Množina rozhodnutí $C \subset \mathfrak{A}$ je úplná, ak ku každému rozhodnutiu $\mathbf{a} \in \mathfrak{A} - C$ existuje rozhodnutie $\mathbf{b} \in C$, že $\mathbf{b} \prec \mathbf{a}$.

Označme symbolom $\mathcal{P}$ množinu všetkých prípustných rozhodnutí. Rozhodnutia, ktoré nie sú prípustné, možno z rozhodovania vylúčiť. Podobne, ak sa podarí nájsť úplnú množinu C, možno vylúčiť rozhodnutia patriace do $\mathfrak{A} - C$. Množina $C_0 \subset \mathfrak{A}$ sa nazýva **minimálne úplná**, ak C_0 je úplná a je podmnožinou každej úplnej množiny.

Veta 23.1. Ak $\mathcal{P}$ je úplná, tak sa rovná minimálnej úplnej množine.

Dôkaz: Ak C je úplná a $\mathbf{a} \notin C$, tak $\mathbf{a} \notin \mathcal{P}$, pretože existuje $\mathbf{b} \in C$, že $\mathbf{b} \prec \mathbf{a}$. Teda $\mathcal{P} \subset C$. Odtiaľ vyplýva, že ak je $\mathcal{P}$ úplná, je aj minimálne úplná. ♣

Ak teda $\mathcal{P}$ je úplná možno množinu rozhodnutí $\mathfrak{A}$ redukovať na množinu $\mathcal{P}$ a ďalšia redukcia už nie je možná.

Ak máme k dispozícii apriórnu hustotu $\pi(\boldsymbol{\theta})$, **stredná strata** pri rozhodnutí $\mathbf{a}$ je

$$\int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{a}) \pi(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) \quad \text{resp.} \quad \sum_{i \geq 1} L(\boldsymbol{\theta}_i, \mathbf{a}) \pi(\boldsymbol{\theta}_i).$$

Rozhodnutie $\mathbf{a}_\pi \in \mathfrak{A}$ možno považovať za optimálne (vzhľadom na túto strednú stratu), ak

$$\mathbf{a}_\pi = \arg \min_{\mathbf{a} \in \mathfrak{A}} \int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{a}) \pi(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) \quad \text{resp.} \quad \mathbf{a}_\pi = \arg \min_{\mathbf{a} \in \mathfrak{A}} \sum_{i \geq 1} L(\boldsymbol{\theta}_i, \mathbf{a}) \pi(\boldsymbol{\theta}_i).$$

Takéto rozhodnutie sa volá **bayesovské**.

V súlade s tým, čo sme povedali o prípustných rozhodnutiach, platí, že ak množina $\mathcal{P}$ je úplná, tak rozhodnutie $\mathbf{a}_\pi$ je alebo prípustné, alebo existuje prípustné rozhodnutie $\mathbf{b}_\pi \prec \mathbf{a}_\pi$, pričom platí

$$\int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{a}_\pi) \pi(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) = \int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{b}_\pi) \pi(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta})$$

(a analogicky aj v diskrétnom prípade). Návod na dôkaz tohto tvrdenia je v knižke A. Pázmana, str. 74.

23.2. Rozhodovanie na základe experimentu

V príklade v kapitole 23.1 je rozumné, aby si stavebník, prv než sa rozhodne stavať, dal preskúmať podložie. To znamená, že príslušní odborníci vykonajú merania. Výsledkom týchto meraní je vektor údajov $\mathbf{y}$. Tento je samozrejme ovplyvnený náhodnými efektmi, ktoré v závislosti od stavu podložia $\boldsymbol{\theta}$ dajú vystihnúť hustotami pravdepodobnosti $f(\mathbf{y}|\boldsymbol{\theta})$ resp. príslušnými pravdepodobnostnými funkciami.

Prv než sa vykoná rozhodnutie v rozhodovacom probléme, je rozumné vykonať rôzne pozorovania, ktoré súhrnne nazývame experiment. Matematicky sme experiment charakterizovali triedou hustôt (alebo pravdepodobnostných funkcií)

$$\{f(\mathbf{y}|\boldsymbol{\theta}):\ \boldsymbol{\theta}\in\boldsymbol{\Omega}\}.$$

Úlhou štatistika je nájsť rozumné pravidlo, ako každému výsledku experimentu $\mathbf{y}\in\mathcal{Y}(\in\mathcal{B}_n)$ priradí niektoré rozhodnutie $\mathbf{a}\in\mathfrak{A}$, t.j. nájsť vhodné zobrazenie

$$\Delta:\ \mathbf{y}\in\mathcal{Y}\to\mathbf{a}\in\mathfrak{A}.$$

Toto zobrazenie sa nazýva **rozhodovacia funkcia**. Ak máme danú rozhodovaciu funkciu Δ, tak **rizikom**, alebo **rizikovou funkciou** nazývame funkciu

$$r:\quad (\boldsymbol{\theta},\Delta)\in\boldsymbol{\Omega}\times D\to r(\boldsymbol{\theta},\Delta)=\int_{\mathcal{Y}} L(\boldsymbol{\theta},\Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta})d\lambda(\mathbf{y}),$$

ktorá vyjadruje strednú stratu pri stave $\boldsymbol{\theta}$ a pri rozhodovaní podľa rozhodovacej funkcie Δ. Symbolom D tu značíme množinu rozhodovacích funkcií dovolených v danom rozhodovacom probléme. Základnou požiadavkou na každú $\Delta\in D$ je, aby existoval uvedený integrál.

Ak ten čo rozhoduje chce minimalizovať rizikovú funkciu, ocitá sa formálne v tej istej situácii, ako v prípade rozhodovania bez experimentu. Namiesto trojice

$$(\boldsymbol{\Omega},\mathfrak{A},L(\boldsymbol{\theta},\mathbf{a})),$$

ktorá sa uvažovala v kapitole 23.1, má teraz trojicu

$$(\boldsymbol{\Omega},D,r(\boldsymbol{\theta},\Delta)).$$

Mnohé pojmy možno teda priamo preniesť z kapitoly 23.1. Rozhodovacie funkcie sú vo vzťahu $\Delta_1\preceq\Delta_2$, ak pre každé $\boldsymbol{\theta}\in\boldsymbol{\Omega}$ platí $r(\boldsymbol{\theta},\Delta_1)\le r(\boldsymbol{\theta},\Delta_2)$, čím definujeme **rovnomerné usporiadanie** rozhodovacích funkcií. Rozhodovacia funkcia Δ_1 je **prípustná**, ak neexistuje iná rozhodovacia funkcia Δ_2, ktorá by bola rovnomerne lepšia, teda neexistuje Δ_2 pre ktorú platí $\Delta_2\prec\Delta_1$. Podobne ako v 23.1 definujeme **úplné** množiny rozhodovacích funkcií a platia presne tie isté všeobecné súvislosti medzi prípustnosťou a úplnosťou.

Pretože usporiadanie rozhodovacích funkcií reláciou $\preceq$ je čiastočné, vo všeobecnosti neexistuje rovnomerne najlepšia rozhodovacia funkcia. Ak však poznáme apriórnu hustotu $\pi(\boldsymbol{\theta})$, porovnávame rozhodovacie funkcie na základe stredného rizika (alebo strednej straty), teda na základe integrálu

$$R(\Delta)=\int_{\boldsymbol{\Omega}} r(\boldsymbol{\theta},\Delta)\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})=\int_{\boldsymbol{\Omega}}\left[\int_{\mathcal{Y}} L(\boldsymbol{\theta},\Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta})d\lambda(\mathbf{y})\right]\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta}). \tag{23.1}$$

Bayesovskou rozhodovacou funkciou voláme tú rozhodovaciu funkciu, pre ktorú platí

$$\Delta_\pi=\arg\ \min_{\Delta\in D} R(\Delta).$$

Tu však končí analógia s rozhodovaním bez experimentu. Pri rozhodovaní bez experimentu minimalizujeme integrál $\int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{a})\pi(\boldsymbol{\theta})d\lambda(\boldsymbol{\theta})$ resp. $\sum_{i\geq 1} L(\boldsymbol{\theta}_i, \mathbf{a})\pi(\boldsymbol{\theta}_i)$ vzhľadom na prvky množiny $\mathfrak{A}$, teraz potrebujeme minimalizovať dvojný integrál $\int_{\mathbf{\Omega}} \left[\int_{\mathcal{Y}} L(\boldsymbol{\theta}, \Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta})d\lambda(\mathbf{y})\right] \pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})$ vzhľadom na zobrazenia Δ, ktorých oborom hodnôt je mnžina $\mathfrak{A}$ (množina možných rozhodnutí), teda vzhľadom na podstatne komplikovanejšiu štruktúru. Ukazuje sa však, že za pomerne všeobecných podmienok táto minimalizačná úloha má elegantné riešenie, ktoré je dané v nasledujúcej vete.

Veta 23.2. (Základná veta o bayesovských rozhodovacích funkciách.) Nech stratová funkcia $L(\boldsymbol{\theta}, \mathbf{a})$ je merateľná a zdola ohraničená. Nech D je množina všetkých merateľných zobrazení z $\mathcal{Y}$ do $\mathfrak{A}$. Nech pre skoro všetky $\mathbf{y} \in \mathcal{Y}$ existuje riešenie úlohy

$$\Delta_\pi(\mathbf{y}) = \arg \min_{\mathbf{a}\in\mathfrak{A}} \int_{\mathbf{\Omega}} L(\boldsymbol{\theta}, \mathbf{a})\pi(\boldsymbol{\theta}|\mathbf{y})d\nu(\boldsymbol{\theta}), \tag{23.2}$$

ktoré je merateľnou funkciou $\mathbf{y}$. Potom toto riešenie je bayesovskou rozhodovacou funkciou pri apriórnej hustote $\pi(\boldsymbol{\theta})$.

Dôkaz nájdete v knižke A. Pázmana na str. 76.

Poznamenávame len, že v skutočnosti nie je ani potrebné určiť celú bayesovskú rozhodovaciu funkciu Δ_π. Ak zrealizujeme experiment a jeho výsledok je $\mathbf{y}$, stačí riešiť minimalizačnú úlohu (23.2) pre túto jedinú hodnotu $\mathbf{y}$.

Často sa popri bayesovských rozhodovacích funkciách uvažuje aj tzv. minimaxná rozhodovacia funkcia Δ_m definovaná vzťahom

$$\Delta_m(\mathbf{y}) = \arg \min_{\Delta\in D} \left\{ \max_{\boldsymbol{\theta}\in\mathbf{\Omega}} r(\boldsymbol{\theta}, \Delta) \right\},$$

ktorá minimalizuje riziko pri tom najnepriaznivejšom $\boldsymbol{\theta}$. Takáto stratégia rozhodovateľa vlastne nedôveruje žiadnemu apriórnemu rozdeleniu π a je teda veľmi opatrnícka. Pri istých dosť všeobecných predpokladoch sa dá dokázať, že existuje apriórna hustota π_0 taká, že $\Delta_m(\mathbf{y}) = \Delta_{\pi_0}(\mathbf{y})$. Hustotu π_0 nazývame **najnepriaznivejšou apriórnou hustotou**.

23.3. Aposteriórna hustota interpretovaná ako bayesovská rozhodovacia funkcia maximalizujúca informáciu

Aby sme sa vyhli topologickým komplikáciám, budeme v tejto kapitole predpokladať, že $\mathbf{\Theta}$ je diskrétny náhodný vektor s konečným oborom hodnôt $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_k\}$.

Nech množina rozhodnutí $\mathfrak{A}$ je množina všetkých (diskrétnych) rozdelení pravdepodobnosti (pravdepodobnstných funkcií) definovanćh na $\mathbf{\Omega}$. Pre každé $\boldsymbol{\theta} \in \mathbf{\Omega}$ a pre každé $P(\cdot) \in \mathfrak{A}$ definujeme stratu takto

$$L(\boldsymbol{\theta}, P(\cdot)) = -\ln P(\boldsymbol{\theta}),$$

teda $L(\boldsymbol{\theta}_i) = L_P(\boldsymbol{\theta}_i) = -\ln P(\boldsymbol{\theta}_i)$, $i = 1, 2, ..., k$. Takáto stratová funkcia má jasnú informačnú interpretáciu. Nech $\boldsymbol{\theta}^*$ je skutočná hodnota parametra $\boldsymbol{\theta}$. Ak $P(\boldsymbol{\theta}^*) = 1$, tak strata je nulová, pretože sme presne určili skutočnú hodnotu $\boldsymbol{\theta}$. Ak $P(\boldsymbol{\theta}^*)$ je malé číslo, tak strata je veľká, lebo sme pripísali malú pravdepodobnosť skutočnej hodnote parametra $\boldsymbol{\theta}$. V extrémnom prípade $P(\boldsymbol{\theta}^*) = 0$ je strata nekonečná. Minimalizovať takúto stratovú funkciu znamená teda maximalizovať informáciu získanú z experimentu. Ukážeme, že zobrazenie

$$\mathbf{y} \in \mathcal{Y} \to \pi(\cdot|\mathbf{y})$$

je bayesovskou rozhodovacou funkciou pri tejto stratovej funkcii.

Každá rozhodovacia funkcia v uvažovanom rozhodovacom probléme má tvar

$$\Delta(\mathbf{y}) = P_{\mathbf{y}}(\cdot),$$

t.j. pre každé $\mathbf{y} \in \mathcal{Y}$ je jej hodnota nejaké rozdelenie pravdepodobnosti (čiže nejaká pravdepodobnostná funkcia) na $\mathbf{\Omega}$, indexované vektorom $\mathbf{y}$. Teda platí

$$L(\boldsymbol{\theta}, \Delta(\mathbf{y})) = L(\boldsymbol{\theta}, P_{\mathbf{y}}(\cdot)) = -\ln P_{\mathbf{y}}(\boldsymbol{\theta})$$

a riziková funkcia je

$$r(\boldsymbol{\theta}, \Delta) = -\int_{\mathcal{Y}} \ln P_{\mathbf{y}}(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) d\lambda(\mathbf{y}).$$

Stredná strata pri apriórnom rozdelení π sa rovná

$$R(\Delta) = -\sum_{i\geq 1} \left[\int_{\mathcal{Y}} \ln P_{\mathbf{y}}(\boldsymbol{\theta}_i) f(\mathbf{y}|\boldsymbol{\theta}_i) d\lambda(\mathbf{y}) \right] \pi(\boldsymbol{\theta}_i).$$

Podľa Vety 23.2 bayesovská rozhodovacia funkcia pri danom $\mathbf{y}$ je

$$\Delta_\pi(\mathbf{y}) = \arg \min_{P(\cdot)} \left\{ -\sum_{i\geq 1} L(\boldsymbol{\theta}_i, P(\cdot))\pi(\boldsymbol{\theta}_i|\mathbf{y}) \right\} = \arg \min_{P(\cdot)} \left\{ -\sum_{i\geq 1} (\ln P(\boldsymbol{\theta}_i))\pi(\boldsymbol{\theta}_i|\mathbf{y}) \right\} =$$

$$= \arg \min_{P(\cdot)} \left\{ -\sum_{i\geq 1} (\ln P(\boldsymbol{\theta}_i))\pi(\boldsymbol{\theta}_i|\mathbf{y}) + \sum_{i\geq 1} (\ln \pi(\boldsymbol{\theta}_i|\mathbf{y}))\pi(\boldsymbol{\theta}_i|\mathbf{y}) \right\} = \arg \min_{P(\cdot)} I(\pi(\cdot|\mathbf{y}), P(\cdot)).$$

Minimalizujeme teda $I-$divergenciu. Podľa Poznámky pod Definíciou 9.2 toto minimum sa dosahuje ak $P(\cdot)$ a $\pi(\cdot|\mathbf{y})$ sa zhodujú. Teda bayesovská rozhodovacia funkcia sa v tomto prípade rovná

$$\Delta(\mathbf{y}) = \pi(\cdot|\mathbf{y}).$$

Tento výsledok je zaujímavým informačným zdôvodnením používania Bayesovho vzorca. Žiadne iné rozdelenie pravdepodobnosti na $\mathbf{\Omega}$ nepostihuje tak dobre informáciu získanú z experimentu ako práve aposteriórne rozdelenie.

24. Bayesovské odhady a testy z hľadiska teórie štatistického rozhodovania

Teória štatistického rozhodovania, ktorá vznikla ako ekonomicky motivovaná teória rozhodovania využívajúca experiment (pozorovania, merania), ovplyvnila myslenie v teoretickej štatistike, hlavne v teórii odhadu.

24.1. Bayesovské bodové odhady z hľadiska teórie štatistického rozhodovania

Ak v experimente

$$\{f(\mathbf{y}|\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \mathbf{\Omega}\}$$

zvolíme $\mathfrak{A} = \mathbf{\Omega}$ a ak stratovú funkciu $L(\boldsymbol{\theta}, \mathbf{a})$ zvolíme tak, aby vyjadrovala odchýlku odhadu od skutočnej hodnoty parametra, tak bayesovské rozhodovacie funkcie sú vlastne (bayesovskými) odhadmi parametra $\boldsymbol{\theta}$ (teda odhadovacími štatistikami). Takými stratovými funkciami sú napríklad

$$L(\boldsymbol{\theta}, \mathbf{a}) = ||\mathbf{a} - \boldsymbol{\theta}||^2$$

alebo

$$L(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{m} |a_i - \theta_i|.$$

Bayesovské odhady sa porovnávajú s "klasickými" odhadmi frekvenčnej štatistiky, hlavne s odhadom maxima vierohodnosti. Naopak, pri "klasických" odhadoch sa overujú vlastnosti formulované v teórii rozhodovania, ako sú prípustnosť (admisibilita), invariantnosť, optimálnosť vzhľadom na niektorú stratovú funkciu. Je vhodné špecializovať základnú vetu o rozhodovacích funkciách (Vetu 23.2) na uvedené stratové funkcie.

Veta 24.1. Nech $\mathbf{S}$ je symetrická pozitívne definitná matica a nech

$$L(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta} - \mathbf{a})'\mathbf{S}(\boldsymbol{\theta} - \mathbf{a}).$$

Nech $\pi(\boldsymbol{\theta})$ je apriórna hustota. Potom bayesovská rozhodovacia funkcia (t.j. bayesovský odhad) sa rovná podmienenej strednej hodnote $\boldsymbol{\Theta}$ pri danom $\mathbf{y}$

$$\Delta_\pi(\mathbf{y}) = \mathcal{E}(\boldsymbol{\Theta}|\mathbf{y}) = \begin{cases} \begin{pmatrix} \frac{\int_{\boldsymbol{\Omega}} \theta_1 f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})} \\ \frac{\int_{\boldsymbol{\Omega}} \theta_2 f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})} \\ \vdots \\ \frac{\int_{\boldsymbol{\Omega}} \theta_m f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} f(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta})d\nu(\boldsymbol{\theta})} \end{pmatrix}, & \text{ak menovateľ } \neq 0, \\ 0, & \text{ak menovateľ } = 0. \end{cases}$$

V prípade, že $\mathbf{S}$ je len pozitívne semidefinitná symetrická matica, zobrazenie $\mathbf{y} \in \mathcal{Y} \to \mathcal{E}(\boldsymbol{\Theta}|\mathbf{y})$ je jednou z bayesovských rozhodovacích funkcií. Stredná strata pri tejto rozhodovacej funkcii sa rovná

$$\mathcal{E}\left\{(\boldsymbol{\Theta} - \mathcal{E}(\boldsymbol{\Theta}|\mathbf{y}))'\mathbf{S}(\boldsymbol{\Theta} - \mathcal{E}(\boldsymbol{\Theta}|\mathbf{y}))\right\} = tr\left\{\mathbf{S}\ \mathcal{E}\left(cov(\boldsymbol{\Theta}|\mathbf{y})\right)\right\},$$

kde $cov(\boldsymbol{\Theta}|\mathbf{y})$ je kovariančná matica náhodného vektora $\boldsymbol{\Theta}|\mathbf{y}$.

Dôkaz: nájdete v knihe A. Pázmana, str. 82. Poznamenávame len, že ak vo vete 24.1 volíme $\mathbf{S} = \mathbf{I}$, tak $L(\boldsymbol{\theta}, \mathbf{a}) = ||\mathbf{a} - \boldsymbol{\theta}||^2$.

Veta 24.2. Nech $\nu(\cdot)$ je Lebesgueova miera na $\boldsymbol{\Omega} = \mathcal{R}^m$. Nech

$$L(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{m} |a_i - \theta_i|.$$

Nech $\pi_i(\theta_i|\mathbf{y})$ je marginálna aposteriórna hustota parametra θ_i

$$\pi_i(\theta_i|\mathbf{y}) = \int_{\mathcal{R}^{m-1}} \pi(\boldsymbol{\theta}|\mathbf{y})d\theta_1 ... d\theta_{i-1}d\theta_{i+1} ... d\theta_m$$

a nech $\mu_i(\mathbf{y})$ je medián tejto hustoty. Potom

$$\mathbf{y} \in \mathcal{Y} \to (\mu_1(\mathbf{y}), ..., \mu_m(\mathbf{y}))'$$

je bayesovská rozhodovacia funkcia (bayesovský odhad).

Dôkaz nájdete v knihe A. Pázmana na str. 83.

Používaným odhadom v bayesovskej štatistike je aj odhad

$$\widetilde{\boldsymbol{\Theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\Theta} \in \boldsymbol{\Omega}} \pi(\boldsymbol{\Theta}|\mathbf{y}),$$

ktorého realizácia je

$$\widetilde{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} \pi(\boldsymbol{\theta}|\mathbf{y}).$$

Z bayesovho vzorca resp. z (2.7) vyplýva, že ak hustota $\pi(\boldsymbol{\theta})$ je konštantná, tak odhad maxima aposteriórnej hustoty je totožný s odhadom maxima vierohodnosti

$$\widehat{\boldsymbol{\Theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\Theta}\in\boldsymbol{\Omega}} \ln f(\mathbf{y}|\boldsymbol{\Theta}),$$

ktorého realizácia je

$$\widehat{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} \ln f(\mathbf{y}|\boldsymbol{\theta}).$$

Odhad $\widetilde{\boldsymbol{\Theta}}(\mathbf{y})$ sa preto nazýva **bayesovským odhadom maxima vierohodnosti**. Podľa (2.7) možno písať

$$\widetilde{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} [\ln f(\mathbf{y}|\boldsymbol{\theta}) + \ln \pi(\boldsymbol{\theta})]$$

a preto sa tento odhad nazýva aj **penalizovaným odhadom maxima vierohodnosti**, pričom $\ln \pi(\boldsymbol{\theta})$ je penalizácia. Dá sa ukázať, že odhad $\widetilde{\boldsymbol{\Theta}}(\mathbf{y})$ nedostaneme zo žiadnej stratovej funkcie v zmysle kapitoly 23, môžemeho však vyjadriť ako limitu bayesovských rozhodovacích funkcií (pozri knihu A. Pázmana, str. 84).

24.2. Bayesovské intervalové odhady

Vo frekvenčnej štatistike sa ako intervalové odhady parametra $\boldsymbol{\theta}$ používajú oblasti spoľahlivosti. Sú to náhodné borelovské množiny, ktoré s predpísanou pravdepodobnosťou pokrývajú (pevný) parameter $\boldsymbol{\theta}$. Ich konštrukcia je spojená s konštrukciou optimálnych testov a môže byť veľmi komplikovaná. V bayesovskej štatistike je situácia jednoduchšia. Používajú sa oblasti ohraničené krivkou konštantnej aposteriórnej hustoty

$$\mathcal{O}(\mathbf{y}) = \{\boldsymbol{\theta} \in \boldsymbol{\Omega} : \ \pi(\boldsymbol{\theta}|\mathbf{y}) \geq c(\mathbf{y})\},$$

kde $c(\mathbf{y})$ je zvolené tak, aby platilo

$$\int_{\mathcal{O}(\mathbf{y})} \pi(\boldsymbol{\theta}|\mathbf{y}) d\nu(\boldsymbol{\theta}) = 1 - \alpha, \tag{24.1}$$

kde $1-\alpha$ je predpísaná spoľahlivosť. Teda aposteriórna pravdepodobnosť toho, že $\boldsymbol{\Theta} \in \mathcal{O}(\mathbf{y})$, je práve $1-\alpha$. V prípade, že sa rovnosť nedá realizovať, volíme za $c(\mathbf{y})$ supremum z tých čísel, pre ktoré platí

$$\int_{\mathcal{O}(\mathbf{y})} \pi(\boldsymbol{\theta}|\mathbf{y}) d\nu(\boldsymbol{\theta}) \geq 1 - \alpha.$$

Niekedy sa oblasť $\mathcal{O}(\mathbf{y})$ nazýva HDP oblasť (highest posterior density region). Zrejme platí nasledujúca veta

Veta 24.3. Nech $\mathbf{y} \in \mathcal{Y}$ je dané a $\pi(\boldsymbol{\theta}|\mathbf{y})$ je aposteriórna hustota vzhľadom na mieru $\nu(\boldsymbol{\theta})$ a nech pre $W \subset \boldsymbol{\Omega}$ platí

$$\int_W \pi(\boldsymbol{\theta}|\mathbf{y}) d\nu(\boldsymbol{\theta}) = 1 - \alpha.$$

Potom

$$\nu(\mathcal{O}(\mathbf{y})) \leq \nu(W).$$

24.3. Bayesovské testy

Vhodným výberom priestoru rozhodnutí $\mathfrak{A}$ a stratovej funkcie $L(\boldsymbol{\theta}, \mathbf{a})$ dostaneme rozhodovacie problémy podobné úlohám testovania hypotéz vo frekvenčnej štatistike. Podstatný rozdiel je v tom, že nemôžeme uvažovať hypotézy, ktorých apriórna pravdepodobnosť je nulová (lebo ich aposteriórna pravdepodobnosť je tiež nulová) a tiež v tom, že optimalita testov sa posudzuje úplne iným spôsobom než vo frekvenčnej štatistike.

24.3.1. Jednoduchá hypotéza a jednoduchá alternatíva

Nech $\boldsymbol{\Omega} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$. Priestor rozhodnutí bude dvojprvkový

$$\mathfrak{A} = \{\mathbf{a}_0, \mathbf{a}_1\},$$

kde $\mathbf{a}_0$ znamená, že prijímame hypotézu $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ a $\mathbf{a}_1$ znamená, že prijímame alternatívnu hypotézu $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. Stratová funkcia je $L(\boldsymbol{\theta}_0, \mathbf{a}_1) = w_0 > 0,\ L(\boldsymbol{\theta}_1, \mathbf{a}_0) = w_1 > 0\ L(\boldsymbol{\theta}_0, \mathbf{a}_0) = L(\boldsymbol{\theta}_1, \mathbf{a}_1) = 0$. Apriórne rozdelenie je diskrétne $\pi(\boldsymbol{\theta}_0) > 0,\ \pi(\boldsymbol{\theta}_1) = 1 - \pi(\boldsymbol{\theta}_0) > 0$. Experiment je daný hustotami $f(\mathbf{y}|\boldsymbol{\theta}_0),\ f(\mathbf{y}|\boldsymbol{\theta}_1)$ vzhľadom na ($\sigma-$konečnú) mieru $\lambda(\cdot)$. Môžu to byť "obyčajné" hustoty vzhľadom na Lebesgueovu mieru $\lambda(\cdot)$, alebo dve pravdepodobnostné funkcie $\{f(\mathbf{y}_i|\boldsymbol{\theta}_0)\}_{i\geq 1},\ \{f(\mathbf{y}_i|\boldsymbol{\theta}_1)\}_{i\geq 1}$ - vtedy je miera $\lambda(\cdot)$ sčítacia.

Bayesovská rozhodovacia funkcia (teda bayesovský test) sa v súlade s Vetou 23.2 rovná

$$\Delta(\mathbf{y}) = \arg \min_{\mathbf{a}\in\{\mathbf{a}_0, \mathbf{a}_1\}} [L(\boldsymbol{\theta}_0, \mathbf{a})\pi(\boldsymbol{\theta}_0|\mathbf{y}) + L(\boldsymbol{\theta}_1, \mathbf{a})\pi(\boldsymbol{\theta}_1|\mathbf{y})] =$$

$$= \begin{cases} \mathbf{a}_1, & \text{ak } \ w_0\pi(\boldsymbol{\theta}_0|\mathbf{y}) \leq w_1\pi(\boldsymbol{\theta}_1|\mathbf{y}), \\ \mathbf{a}_0, & \text{ak } \ w_1\pi(\boldsymbol{\theta}_1|\mathbf{y}) \leq w_0\pi(\boldsymbol{\theta}_0|\mathbf{y}). \end{cases}$$

Nulovú hypotézu zamietame, ak prijímame alternatívnu hypotézu, teda ak $\Delta(\mathbf{y}) = \mathbf{a}_1$ a to je práve vtedy ak

$$w_0\pi(\boldsymbol{\theta}_0|\mathbf{y}) \leq w_1\pi(\boldsymbol{\theta}_1|\mathbf{y}). \tag{24.2}$$

Podľa (2.7) je $\pi(\boldsymbol{\theta}|\mathbf{y}) = \dfrac{\pi(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{\sum_{i=1}^{2}\pi(\boldsymbol{\theta}_i)f(\mathbf{y}|\boldsymbol{\theta}_i)}$, teda (24.2) je splnené práve vtedy ak

$$\frac{w_0\pi(\boldsymbol{\theta}_0)}{w_1\pi(\boldsymbol{\theta}_1)} = c \leq \frac{f(\mathbf{y}|\boldsymbol{\theta}_1)}{f(\mathbf{y}|\boldsymbol{\theta}_0)}.$$

Porovnáme tento test s "klasickým" testom, preto napíšme oblasť zamietania nulovej hypotézy $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ ako

$$W(c) = \left\{\mathbf{y} \in \mathcal{Y} : \ \frac{f(\mathbf{y}|\boldsymbol{\theta}_1)}{f(\mathbf{y}|\boldsymbol{\theta}_0)} > c\right\},$$

kde

$$c = \frac{w_0\pi(\boldsymbol{\theta}_0)}{w_1\pi(\boldsymbol{\theta}_1)}. \tag{24.3}$$

Oblasť zamietania nulovej hypotézy je určená podielom vierohodností, podobne ako v Neymannovej-Pearsonovej leme. Zopakujme si ju

Veta 24.4. (Neymanova-Pearsonova lema.) Nech k danému $\alpha \in (0, 1)$ existuje také kladné číslo g, že pre množinu

$$W_g = \{\mathbf{y} : \ f(\mathbf{y}|\boldsymbol{\theta}_1) \geq g f(\mathbf{y}|\boldsymbol{\theta}_0)\}$$

platí

$$\int_{W_g} f(\mathbf{y}|\boldsymbol{\theta}_0) d\lambda(\mathbf{y}) = \alpha. \tag{24.4}$$

Potom pre ľubovoľnú množinu $W \in \mathcal{B}_n$ splňujúcu podmienku

$$\int_W f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}) = \alpha \tag{24.5}$$

platí

$$\int_{W_g} f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda(\mathbf{y}) \geq \int_W f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda\mathbf{y}).$$

Dôkaz nájdeme v knižke J. Anděla na str. 239. Treba si uvedomiť, že oblasťou zamietania nulovej hypotézy W_g je určený test s hladinou významnosti (pravdepodobnosťou chyby 1. druhu) α (t.j. pravdepodobnosť že test zamieta nulovú hypotézu, keď ona platí, je α). Pričom tento test je najsilnejší spomedzi všetkých testov s hladinou významnosti α. Všetky testy s hladinou významnosti α sú určené oblasťami zamietnutia W, pre ktoré platí (24.5). Chyba druhého druhu je taká, keď test nezamieta nulovú hypotézu, pričom platí alternatívna. Test určený oblasťou zamietania W má pravdepodobnosť chyby druhého druhu

$$\int_{\mathcal{Y}-W} f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda(\mathbf{y}) = 1 - \beta$$

a β je sila testu. Podľa Neymanovej-Pearsonovej lemy je oblasť zamietania pre najsilnejší test s hladinou významnosti α daný ako

$$W_g = \left\{\mathbf{y} : \frac{f(\mathbf{y}|\boldsymbol{\theta}_1)}{f(\mathbf{y}|\boldsymbol{\theta}_0)} \geq g\right\},$$

pričom konštantu g počítame zo vzorca (24.4).

Pri bayesovskom testovaní naproti tomu namiesto chýb 1. a 2. druhu máme aposteriórne pravdepodobnosti oboch hypotéz $\pi(\boldsymbol{\theta}_0|\mathbf{y})$, $\pi(\boldsymbol{\theta}_1|\mathbf{y})$ a strednú stratu (rizikovú funkciu, pozri (23.1)) rovnú

$$R(\Delta) = \pi(\boldsymbol{\theta}_0)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_0, \Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}) + \pi(\boldsymbol{\theta}_1)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_1, \Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda(\mathbf{y}).$$

Počítajme

$$\pi(\boldsymbol{\theta}_0)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_0, \Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}) =$$

$$= \begin{cases} \pi(\boldsymbol{\theta}_0)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_0, \mathbf{a}_1)f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}), & \text{ak } \ w_0\pi(\boldsymbol{\theta}_0|\mathbf{y}) \leq w_1\pi(\boldsymbol{\theta}_1|\mathbf{y}) \\ \pi(\boldsymbol{\theta}_0)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_0, \mathbf{a}_0)f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}), & \text{ak } \ w_1\pi(\boldsymbol{\theta}_1|\mathbf{y}) \leq w_0\pi(\boldsymbol{\theta}_0|\mathbf{y}) \end{cases}$$

$$= \begin{cases} \pi(\boldsymbol{\theta}_0)w_0\int_{\mathcal{Y}} f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}), & \text{ak } \ \mathbf{y} \in \mathcal{Y} - W(c) \\ 0, & \text{inokedy} \end{cases}$$

$$= \pi(\boldsymbol{\theta}_0)w_0\int_{\mathcal{Y}-W(c)} f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}) = \pi(\boldsymbol{\theta}_0)w_0\left[\int_{\mathcal{Y}} f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y}) - \int_{W(c)} f(\mathbf{y}|\boldsymbol{\theta}_0)d\lambda(\mathbf{y})\right] = \pi(\boldsymbol{\theta}_0)w_0(1-\alpha),$$

kde α počítame podľa (24.5). Podobne

$$\pi(\boldsymbol{\theta}_1)\int_{\mathcal{Y}} L(\boldsymbol{\theta}_1, \Delta(\mathbf{y}))f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda(\mathbf{y}) = \pi(\boldsymbol{\theta}_1)w_1\beta,$$

kde β je sila testu, t.j. $\int_{W(c)} f(\mathbf{y}|\boldsymbol{\theta}_1)d\lambda(\mathbf{y})$. Konštantu c počítame z (23.4). Pre strednú stratu vyššie uvedeného bayesovho testu dostávame

$$R(\Delta) = \pi(\boldsymbol{\theta}_0)w_0(1-\alpha) + \pi(\boldsymbol{\theta}_1)w_1\beta = \pi(\boldsymbol{\theta}_1)w_1(c(1-\alpha) + \beta).$$

24.3.2. Súčasné testovanie niekoľkých jednoduchých hypotéz

Na rozdiel od frekvenčnej štatistiky, v bayesovskej štatistike nerobí problémy rozhodovanie medzi niekoľkými hypotézami.

Nech $\mathcal{Y} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k\}$, $\mathfrak{A} = \{\mathbf{a}_1, ..., \mathbf{a}_k\}$, kde $\mathbf{a}_j$ znamená, že prijmeme hypotézu $\boldsymbol{\theta} = \boldsymbol{\theta}_j$. Predpokladáme, že $\pi(\boldsymbol{\theta}_j) > 0$ pre každé j. Ďalej

$$L(\boldsymbol{\theta}_i, \mathbf{a}_j) = \begin{cases} w_{ij} > 0, & \text{ak } \ i \neq j, \\ 0, & \text{ak } \ i = j. \end{cases}$$

Potom

$$\Delta(\mathbf{y}) = \arg \min_{\mathbf{a} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}} \sum_{i=1}^{k} L(\boldsymbol{\theta}_i, \mathbf{a}) \pi(\boldsymbol{\theta}_i | \mathbf{y}) =$$

$$= \mathbf{a}_j, \qquad \text{ak } \ \sum_{i=1}^{k} w_{ij} \pi(\boldsymbol{\theta}_i | \mathbf{y}) \leq \sum_{i=1}^{k} w_{il} \pi(\boldsymbol{\theta}_i | \mathbf{y}) \quad \text{pre každé } \ l.$$

24.3.3. Testovanie zloženej hypotézy proti zloženej alternatíve

Nech $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0 \cup \boldsymbol{\Omega}_1$, kde $\boldsymbol{\Omega}_0 \cap \boldsymbol{\Omega}_1 = \emptyset$. Nech $\mathfrak{A} = \{\mathbf{a}_0, \mathbf{a}_1\}$. Nech $\int_{\boldsymbol{\Omega}_i} \pi(\boldsymbol{\theta}) d\nu(\boldsymbol{\theta}) > 0$, $i = 0, 1$. Stratovú funkciu volíme v princípe tak, aby

$$L(\boldsymbol{\theta}, \mathbf{a}_j) = \begin{cases} 0, & \text{ak } \ \boldsymbol{\theta} \in \boldsymbol{\Omega}_j, \\ > 0, & \text{ak } \ \boldsymbol{\theta} \notin \boldsymbol{\Omega}_j. \end{cases}$$

Obvyklé voľby stratovej funkcie pre $\boldsymbol{\theta} \in \boldsymbol{\Omega}_i$ sú

$$L(\boldsymbol{\theta}, \mathbf{a}_j) = w_{ij} > 0, \quad \text{ak } \ i \neq j,$$

alebo

$$L(\boldsymbol{\theta}, \mathbf{a}_j) = K_j \min_{\boldsymbol{\theta}^* \in \ \boldsymbol{\Omega}_j} ||\boldsymbol{\theta} - \boldsymbol{\theta}^*||.$$

V zmysle Vety 23.2 hypotézu $\boldsymbol{\theta} \in \boldsymbol{\Omega}_1$ prijmeme (t.j. hypotézu $\boldsymbol{\theta} \in \boldsymbol{\Omega}_0$ zamietneme), keď

$$\int_{\boldsymbol{\Omega}_0} L(\boldsymbol{\theta}, \mathbf{a}_1) \pi(\boldsymbol{\theta} | \mathbf{y}) d\nu(\boldsymbol{\theta}) < \int_{\boldsymbol{\Omega}_1} L(\boldsymbol{\theta}, \mathbf{a}_0) \pi(\boldsymbol{\theta} | \mathbf{y}) d\nu(\boldsymbol{\theta}).$$

Takéto testovanie zloženej hypotézy proti zloženej alternatíve je typické pre diskriminačnú analýzu. (Podrobnejšie pozri v knižke J. Anděla str. 319-323.) Tvorba bayesovských testov môže byť jednoduchšia ako tvorba testov vo frekvenčnej štatistike.