

## 14. Metodologia Bayesowska

### 14.1. Podstawowe założenia

W metodologii Bayesowskiej zakłada się, że prawdopodobieństwa są odbiciem naszej subiektywnej wiedzy na temat rzeczywistości. Postawa ta różna jest od postawy klasycznej, która interpretuje prawdopodobieństwo jako miarę częstości z jaką zjawisko występuje.

W podejściu klasycznym, błędy są losowe ale parametry modelu są deterministyczne ale nieznanne. Nielosowość parametrów jest jednak jedynie użytecznym założeniem, ponieważ w praktyce obserwujemy jedynie estymatory parametrów, które są zmiennymi losowymi.

W podejściu bayesowskim wszystkie nieznanne elementy modelu są traktowane jako zmienne losowe. Błędy losowe i parametry są szacowane na podstawie danych, a niepełna wiedza na ich temat jest opisana za pomocą rozkładu prawdopodobieństwa.

Specjalną cechą metodologii Bayesowskiej jest używanie rozkładu *a priori* dla parametrów. W trakcie estymacji zakładamy, że dysponujemy pewną wstępną wiedzą na temat wielkości parametrów. Wiedza ta jest następnie modyfikowana po skonfrontowaniu z danymi. Na podstawie twierdzenia Bayesa, rozkładu *a priori* i modelu próbkowego możemy sformułować nowy rozkład parametrów, który uwzględni zarówno nasze pierwotne przekonania *a priori* jak i posiadane dane empiryczne. Taki zmodyfikowany na podstawie danych rozkład nazywamy rozkładem *a posteriori*.

### 14.2. Rozkłady

Oznaczmy gęstość rozkładu gamma o parametrach  $\alpha, \beta$  jako  $p_G(x|\alpha, \beta)$ . Można też analizować rozkład  $y = \frac{1}{x}$ , gdzie  $x$  pochodzi z rozkładu gamma. Rozkład ten nazywamy odwróconym rozkładem gamma i ma on postać (*inverted gamma*), którą można policzyć z twierdzenia 17.25

$$p_{IG}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha-1)} e^{-\beta y^{-1}} \left| -\frac{1}{y^2} \right| = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} e^{-\beta y^{-1}}$$

Jeśli  $y = \frac{\gamma}{x}$  i  $x$  ma rozkład gamma to gęstość  $y$  ma postać  $p_{IG}(y|\alpha, \gamma\beta)$ . Z definicji wynika, że  $p_{IG}(\sigma^2|\alpha, \beta) = p_G(\sigma^{-2}|\alpha, \beta)$ .

Gęstość  $k$ -wymiarowego rozkładu normalnego dla  $\mathbf{x} \in \mathbb{R}^k$ , wektorce wartości oczekiwanych  $\boldsymbol{\mu}$  i maciezy wariancji kowariancji  $\boldsymbol{\Sigma}$  oznaczmy jako  $p_N^k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Gęstość  $k$ -wymiarowego rozkładu rozkładu  $t$ -Studenta o  $r$  stopniach swobody, wektorce niecentralności  $\boldsymbol{\mu}$  i maciezy precyzji  $\mathbf{A} = \mathbf{V}^{-1}$  dodatnio określonej będziemy oznaczać jako  $p_S^k(\mathbf{x}|r, \boldsymbol{\mu}, \mathbf{V})$ . Rozkład normalny można zdekomponować w następujący sposób:

$$\begin{aligned}
 p_N^k(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) &= (2\pi\sigma^2)^{-\frac{1}{2}n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\
 &= (2\pi\sigma^2)^{-\frac{1}{2}n} \exp\left\{-\frac{1}{2\sigma^2}[e'e + (\boldsymbol{\beta} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \mathbf{b})]\right\} \\
 &= (2\pi\sigma^2)^{-\frac{1}{2}k} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \mathbf{b})\right] (2\pi\sigma^2)^{-\frac{1}{2}(n-k)} \exp\left\{-\frac{e'e}{2\sigma^2}\right\} \\
 &\propto p_N^k(\boldsymbol{\beta}|\mathbf{b}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) p_{IG}\left(\sigma^2\left|\frac{n-k-2}{2}, \frac{n-k}{2}s^2\right.\right) \quad (14.1)
 \end{aligned}$$

gdzie  $s^2 = \frac{e'e}{n-k}$ . W przypadku rozkładu gamma, odwróconego gamma i normalnego iloczynu gęstości dają funkcję gęstości z tej samej klasy:

$$p_G(x|\alpha_1, \beta_1) p_G(x|\alpha_2, \beta_2) = \frac{\beta_1^{\alpha_1}}{\Gamma(a_1)} \frac{\beta_2^{\alpha_2}}{\Gamma(a_2)} x^{\alpha_1+\alpha_2-2} e^{-(\beta_1+\beta_2)x} \propto p_G(\alpha_1 + \alpha_2 - 1, \beta_1 + \beta_2)$$

$$p_{IG}(x|\alpha_1, \beta_1) p_{IG}(x|\alpha_2, \beta_2) \propto p_{IG}(\alpha_1 + \alpha_2 - 1, \beta_1 + \beta_2) \quad (14.2)$$

$$\int_{\mathbb{R}^k} p_N^m(\mathbf{y}|\mathbf{Q}\mathbf{x} + \mathbf{q}, \boldsymbol{\Sigma}_q) p_N^n(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} = p_N^m(\mathbf{y}|\mathbf{Q}\boldsymbol{\mu} + \mathbf{q}, \boldsymbol{\Sigma}_q + \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}') \quad (14.3)$$

Iloczyn rozkładu normalnego i odwróconego gamma daje w rezultacie wielowymiarowy rozkład *t-studenta*

$$\begin{aligned} p_N^k(\mathbf{x} | \boldsymbol{\mu}, \sigma^2 \mathbf{V}) p_{IG} \left( \sigma^2 \mid \frac{a}{2}, \frac{b}{2} \right) \\ = p_{IG} \left( \sigma^2 \mid \frac{a+k}{2}, \frac{1}{2} [b + (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})] \right) p_S^k \left( \mathbf{x} \mid a, \boldsymbol{\mu}, \frac{b}{a} \mathbf{V} \right) \end{aligned}$$

Poza tym mamy następującą ważną tożsamość

$$\int_0^\infty p_N^k(\mathbf{x} | \boldsymbol{\mu}, \sigma^2 \mathbf{V}) p_{IG} \left( \sigma^2 \mid \frac{a}{2}, \frac{b}{2} \right) d\sigma^2 = p_S^k \left( \mathbf{x} \mid a, \boldsymbol{\mu}, \frac{b}{a} \mathbf{V} \right), \quad (14.4)$$

### 14.3. Twierdzenie Bayesa

łączną dystrybucję dla  $\boldsymbol{\theta} \in \Theta$  możemy zapisać jako

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

gdzie  $g(\boldsymbol{\theta} | \mathbf{y})$  jest rozkładem *a priori* parametrów funkcji wiarygodności  $f(\mathbf{y} | \boldsymbol{\theta}) = \ell(\boldsymbol{\theta})$ . Rozkład *a posteriori* parametrów spełnia więc równanie

$$f(\boldsymbol{\theta} | \mathbf{y}) = \frac{f(\mathbf{y} | \boldsymbol{\theta}) g(\boldsymbol{\theta})}{f(\mathbf{y})}, \quad (14.5)$$

ponieważ dystrybucja brzegowa  $f(\mathbf{y})$  nie zależy  $\boldsymbol{\theta}$ , więc  $g(\boldsymbol{\theta} | \mathbf{y})$  traktowana jako funkcja  $\boldsymbol{\theta}$  jest proporcjonalna do licznika ułamka po prawej stronie (14.5). Zapisujemy to

$$f(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}) g(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) g(\boldsymbol{\theta}),$$

gdzie znak  $\propto$  oznacza "jest proporcjonalne do". Stałą w tego typu wyrażeniach można znaleźć ze wzoru  $\int_\Theta g(\boldsymbol{\theta} | \mathbf{y}) = 1$

## 14.4. Rozkłady *a priori*

Rozkład *a priori* powinien odbijać wiedzę jaką posiadamy na temat parametrów a pochodzącą spoza próbki. Przykładowo możemy mieć pewną wiedzę na temat wielkości parametrów wynikającą ze znajomości wyników innych badań empirycznych. Określenie rozkładu *a priori* jest jednak z reguły sprawą trudną ponieważ wymaga sprecyzowania nie tylko wiedzy na temat wartości oczekiwanej parametrów ale i stopnia niepewności na ich temat oraz opisanie tej niepewności za pomocą rozkładu prawdopodobieństwa.

### • Informacyjne rozkłady *a priori*.

Rozkłady *a priori* dobiera się często w ten sposób, by uzyskany w rezultacie rozkład *a posteriori* był stosunkowo prosty. Obecnie jest już możliwe uzyskanie dobrego przybliżenia rozkładu *a posteriori* dla dowolnego rozkładu *a priori* na drodze obliczeń numerycznych.

Istnieje kilka problemów związanych z informacyjnymi rozkładami *a priori*.

Po pierwsze estymatory Bayesowskie będą lepsze od estymatorów klasycznych jedynie wtedy gdy wartość oczekiwana w parametrów w rozkładzie *a priori* będzie bliska prawdziwej wartości parametru. W razie przyjęcia błędnego rozkładu *a priori* estymator Bayesowski może być znacznie gorszy od klasycznego.

Jeśli chodzi o wybór stopnia niepewności (wariancji) w rozkładzie *a priori* mamy następujący problem: im mniejsza jest założona niepewność wiedzy *a priori* tym większe są zyski z jej uwzględnienia ale i możliwe straty w razie zaistnienia błędu.

Dodatkowy problem pojawia się, gdy zachodzi wyraźna sprzeczność między przyjętym rozkładem *a priori* a danymi. Czy w takim przypadku rzeczywiście powinniśmy się dalej upierać, że rozkład *a priori* jest prawidłowy?

Istnieje jeszcze inny problem związany z porównywalnością wyników wnioskowania statystycznego. W procedurze Bayesowskiej wynik estymacji i wnioskowania statystycznego uzależniony jest od postaci rozkładu *a priori*, który z definicji jest subiektywny i zależy od badacza. Jeśli jednak

tak jest to jak uzyskać porównywalność wyników uzyskanych przez różnych badaczy?

- **Nieinformacyjne rozkłady a priori.**

Najpopularniejszym rozwiązaniem subiektywizmu rozkładu *a priori* jest zastosowanie takiej jego postaci, która odbijałaby całkowity brak wiedzy na temat wielkości szacowanych parametrów. Okazuje się jednak, że znalezienie takiego rozkładu nie zawsze jest sprawą prostą. Najpopularniejszą regułą znajduwaną nieinformacyjnych rozkładów *a priori* jest reguła Jeffreysa, która mówi, że nieinformacyjny rozkład *a priori* parametru  $\theta$  powinien spełniać warunek

$$f(\theta) \propto |\mathbf{I}(\theta)|^{\frac{1}{2}}.$$

a więc być niewłaściwym rozkładem jednostajnym proporcjonalnym do pierwiastka wyznacznika macierzy informacyjnej Fishera.

## 14.5. Estymacja punktowa i przedziałowa

- Estymacja punktowa w przypadku metod bayesowskich polega najczęściej na wyznaczeniu wartości oczekiwanej  $\theta$  z rozkładu *a posteriori*

$$\hat{\theta} = E(\theta | \mathbf{y}) = \int_{\Theta} \theta f(\theta | \mathbf{y}) d\theta$$

- W pewnych szczególnych przypadkach za  $\hat{\theta}$  możemy przyjąć medianę lub dominantę  $f(\theta | \mathbf{y})$
- Estymacja przedziałowa dla skalarne  $\theta$  polega na znalezieniu takich  $\theta_1$  i  $\theta_2$ , że przy poziomie ufności  $\alpha$

$$\int_{\theta_1}^{\hat{\theta}} f(\theta | \mathbf{y}) d\theta = \int_{\hat{\theta}}^{\theta_2} f(\theta | \mathbf{y}) d\theta = \frac{1 - \alpha}{2}$$

## 14.6. Analiza Bayesowska w KMRL

### • Nieinformacyjny rozkład a priori (Jeffreys prior)

Często przyjmowanym nieinformacyjnym rozkładem *a priori* w *KMRL* jest rozkład

$$f(\boldsymbol{\beta}, \sigma) = f(\boldsymbol{\beta}) f(\sigma^2) \propto \sigma^{-2}$$

Taki rozkład *a priori* otrzymujemy przy założeniu, że  $g(\boldsymbol{\beta}) \propto \text{const}$  a  $g(\ln(\sigma^2)) = \text{const}$  dla  $-\infty < \ln \sigma < \infty$ . Oba te rozkłady są rozkładami niewłaściwymi, ponieważ  $\int_{-\infty}^{\infty} g(\beta_s) d\beta_s = \infty$  i  $\int_0^{\infty} \sigma^2 d\sigma = \infty$ . Jak wiemy z rozważań nad estymacją metodą największej wiarygodności *KMRL*

(wzór 10.14),  $\left| \frac{\partial^2 \ell^2(\boldsymbol{\beta}, \sigma^2)}{\partial^2 \sigma^2} \right|^{\frac{1}{2}} = \frac{\sqrt{n}}{2\sigma^2} \propto \sigma^{-2}$ . Podobnie  $\left| \frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right|^{\frac{1}{2}} = \left| \frac{\mathbf{X}' \mathbf{X}}{\sigma^2} \right|^{\frac{1}{2}} \propto \text{const}^1$ . Przyjęte rozkłady  $\sigma$  i  $\boldsymbol{\beta}$  są więc zgodne z regułą Jeffreysa.

Policzmy teraz rozkład *a posteriori* parametrów  $\boldsymbol{\beta}$  i  $\sigma$ . W modelu *KMRL* zakładamy, że funkcja wiarygodności pochodzi z rozkładu normalnego, tak więc

$$\ell(\boldsymbol{\beta}, \sigma^2) = p_N^k(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

a rozkład *a posteriori* ma postać

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) &\propto \ell(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) = p_N^n(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I}) \sigma^{-2} \\ &= p_N^k(\boldsymbol{\beta} | \mathbf{b}, \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}) p_{IG} \left( \sigma^2 \mid \frac{n-k-2}{2}, \frac{n-k}{2} s^2 \right) \sigma^{-2} \end{aligned}$$

$$p_{IG}(\sigma^2 | a, b) \sigma^{-2} = \frac{b^a}{\Gamma(a)} \sigma^{-2(a+1)} e^{-b\sigma^{-2}} \sigma^{-2} \propto \frac{b^{\bar{a}}}{\Gamma(\bar{a})} \sigma^{-2(\bar{a}+1)} e^{-b\sigma^{-2}} = p_{IG}(\sigma^2 | a+1, b)$$

$$f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) \propto p_N^k(\boldsymbol{\beta} | \mathbf{b}, \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}) p_{IG} \left( \sigma^2 \mid \frac{n-k}{2}, \frac{n-k}{2} s^2 \right)$$

---

<sup>1</sup>W rozumieniu, że nie zależy od  $\boldsymbol{\beta}$ .

wykorzystując wzór (14.4) otrzymujemy

$$f(\boldsymbol{\beta}|\mathbf{y}) = \int_0^\infty f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) d\sigma^2 = p_S^n \left( \boldsymbol{\beta} | n - k, \mathbf{b}, s^2 (\mathbf{X}'\mathbf{X})^{-1} \right)$$

Zauważmy, że uzyskany wynik jest całkowicie zgodny z analizą klasyczną. Nieobciążonym punktowym estymatorem wartości  $\boldsymbol{\beta}$  jest  $\mathbf{b}$  a  $\sigma^2$  jest  $s^2$  a wariancji  $\mathbf{b}$  jest  $s^2 (\mathbf{X}'\mathbf{X})^{-1}$ . Identycznie jak w przypadku klasycznego wnioskowania uzyskaliśmy rozkład  $\mathbf{b}$  przy nieznanym  $\sigma^2$  jako rozkład *t-studenta* a rozkład  $\frac{(n-k)s^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi_{n-k}^2$ .

- **Informacyjny rozkład a priori**
- **Prior naturalnie sprzężony**

Jak wiemy ze wzoru (14.1), rozkład normalny można zdekomponować do postaci proporcjonalnej do iloczynu rozkładu normalnego i gamma. W świetle wzorów (14.3), (14.2) narzuca się, że naturalnymi priorami naturalnie sprzężonymi będą w tym kontekście rozkłady normalny i gamma. Jeśli w świetle naszej wiedzy *a priori*

$$f(\boldsymbol{\beta}, \sigma^2) = f(\boldsymbol{\beta} | \sigma^2) f(\sigma^2)$$

gdzie

$$f(\boldsymbol{\beta} | \sigma^2) = p_N^k(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \sigma^2 \mathbf{V}^{-1})$$

a

$$f(\sigma^2) = p_{IG} \left( \frac{n_0 - k}{2}, \frac{n_0 - k}{2} s_0^2 \right)$$

to na mocy (14.4)

$$\begin{aligned} f(\boldsymbol{\beta}) &= \int_0^\infty p_N^k(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \sigma^2 \mathbf{V}^{-1}) p_{IG} \left( \sigma^2 \frac{n_0 - k}{2}, \frac{n_0 - k}{2} s_0^2 \right) d\sigma^2 \\ &= p_S^k(\boldsymbol{\beta} | n_0 - k, \boldsymbol{\beta}_0, s_0 \mathbf{V}^{-1}) \end{aligned}$$

Rozkład *a posteriori* będzie miał następującą postać:

$$f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) \propto \ell(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta} | \sigma^2) f(\sigma^2) = p_N^k(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) p_N^k(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \sigma^2 \mathbf{V}^{-1}) f(\sigma^2)$$

Możemy teraz znormalizować  $\boldsymbol{\beta}$  i  $\boldsymbol{\beta}_0$  za pomocą  $\mathbf{V}^{\frac{1}{2}}$  otrzymując

$$p_N^k(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \sigma_0^2 \mathbf{V}^{-1}) = p_N^k\left(\mathbf{V}^{\frac{1}{2}}\boldsymbol{\beta} \mid \mathbf{V}^{\frac{1}{2}}\boldsymbol{\beta}_0, \sigma_0^2 \mathbf{I}\right)$$

Z kolei iloczyn

$$p_N^n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) p_N^k\left(\mathbf{V}^{\frac{1}{2}}\boldsymbol{\beta}_0 \mid \mathbf{V}^{\frac{1}{2}}\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right) = p_N^{n+k}(\mathbf{Y} | \mathbf{W}\boldsymbol{\beta}, \sigma^2 \mathbf{I}),$$

gdzie  $\mathbf{Y} = \begin{bmatrix} \mathbf{y} \\ \mathbf{V}^{\frac{1}{2}}\boldsymbol{\beta}_0 \end{bmatrix}$  a  $\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{V}^{\frac{1}{2}} \end{bmatrix}$ . Korzystając teraz ze wzoru (14.1) otrzymujemy

$$p_N^{n+k}(\mathbf{Y} | \mathbf{W}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \propto p_N^k(\boldsymbol{\beta} | \bar{\mathbf{b}}, \sigma^2 (\mathbf{W}'\mathbf{W})^{-1}) p_{IG}\left(\sigma^2 \mid \frac{n}{2}, \frac{n-2}{2} s^2\right)$$

gdzie

$$\bar{\mathbf{b}} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'\mathbf{Y} = (\mathbf{X}'\mathbf{X} + \mathbf{V})^{-1} (\mathbf{X}'\mathbf{y} + \mathbf{V}\boldsymbol{\beta}_0) = (\mathbf{X}'\mathbf{X} + \mathbf{V})^{-1} (\mathbf{X}'\mathbf{X}\mathbf{b} + \mathbf{V}\boldsymbol{\beta}_0)$$

iloczyn

$$p_{IG}\left(\sigma^2 \mid \frac{n}{2}, \frac{n-2}{2} s^2\right) p_{IG}\left(\frac{n_0-k}{2}, \frac{n_0-k}{2} s_0^2\right) \propto p_{IG}\left(\sigma^2 \mid \frac{\bar{n}-k}{2}, \frac{(\bar{n}-k)}{2} \bar{s}^2\right)$$

gdzie  $\bar{n} = n + n_0$  a  $\bar{s}^2 = \frac{n-2}{\bar{n}-k} s^2 + \frac{n_0-k}{\bar{n}-k} s_0^2$ . W rezultacie

$$f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) \propto p_N^k(\boldsymbol{\beta} | \mathbf{b}, \sigma^2 (\mathbf{W}'\mathbf{W})^{-1}) p_{IG}\left(\sigma^2 \mid \frac{\bar{n}-k}{2}, \frac{(\bar{n}-k)}{2} \bar{s}^2\right)$$

aby znaleźć brzegowy rozkład  $\beta$  stosujemy ponownie wzór (14.4) uzyskując

$$f(\beta | \mathbf{y}) = \int_0^\infty f(\beta, \sigma^2 | \mathbf{y}) d\sigma^2 = p_S^k \left( \beta | \bar{n} - k, \bar{\mathbf{b}}, \bar{s}^2 (\mathbf{V} + \mathbf{X}'\mathbf{X})^{-1} \right)$$

- **g-prior**

W przypadku g-prioru zakładamy, że  $\mathbf{V} = g(\mathbf{X}'\mathbf{X})$ , gdzie  $g > 0$ . W rezultacie podane w powyżej wzory ulegną znacznemu uproszczeniu

$$\bar{\mathbf{b}} = (\mathbf{X}'\mathbf{X} + g(\mathbf{X}'\mathbf{X}))^{-1} (\mathbf{X}'\mathbf{y} + g(\mathbf{X}'\mathbf{X})\beta_0) = \frac{1}{1+g} \mathbf{b} + \frac{g}{1+g} \beta_0$$

a więc zwykłą średnią ważoną z estymatora *MNK* i wartości oczekiwanej *a priori* parametru  $\beta$ .

- **Analiza Bayesowska w KMRL (regresja grzbietowa)**

- W przypadku występowania problemu współliniowości zaproponowano estymator

$$\mathbf{b}_r = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{y}$$

- $\mathbf{D}$  jest macierzą diagonalną z elementami diagonalnymi macierzy  $\mathbf{X}'\mathbf{X}$  na przekątnej
- $r$  jest pewną stałą większą zera.
- Estymator ten uzyskamy jako estymator Bayesowski jeśli przyjmiemy, że

$$\beta_0 = \mathbf{0}$$

$$\mathbf{V} = r\mathbf{D}$$

## 14.7. Predykcja

## 14.8. Testowanie hipotez

- Mamy  $m$  konkurencyjnych modeli
- $p(M_i)$  prawdopodobieństwo *a priori* modelu  $M_i$
- Prawdopodobieństwo *a posteriori* modelu  $M_i$  z twierdzenia Bayesa

$$p(M_i | \mathbf{y}) = \frac{p(M_i) p(\mathbf{y} | M_i)}{\sum_{k=1}^m p(M_k) p(\mathbf{y} | M_k)}$$

gdzie

$$p(\mathbf{y} | M_i) = p_i(\mathbf{y}) = \int_{\Theta_i} p_i(\mathbf{y} | \boldsymbol{\theta}_i) p_i(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i$$

i  $\boldsymbol{\theta}_i$  jest wektorem parametrów dla modelu  $i$ .

- Zwykle wybieramy ten model, który ma największe prawdopodobieństwo *a posteriori*
- Można się do tego posłużyć iloczynem szans *a posteriori* (*posterior odds ratio*)

$$\frac{p(M_i | \mathbf{y})}{p(M_j | \mathbf{y})} = \frac{p(M_i) p(\mathbf{y} | M_i)}{p(M_j) p(\mathbf{y} | M_j)}$$

- Dwa najważniejsze przypadki szczególne:

– przyjmujemy, że  $p(M_1) = p(M_2) = \dots = p(M_m)$

$$\frac{p(M_i | \mathbf{y})}{p(M_j | \mathbf{y})} = \frac{p(\mathbf{y} | M_i)}{p(\mathbf{y} | M_j)}$$

– przyjmujemy, że mniej prawdopodobne modele o większej ilości parametrów np.

$$p(M_i) \propto 2^{-l_i}$$

- Niewłaściwe rozkłady *a priori* sprawiają problemy przy formułowaniu iloczynu szans.

## 14.9. Bayes empiryczny

Jednym z proponowanych rozwiązań problemu trudności ze znalezieniem hiperparametrów rozkładu *a priori* jest wyestymowanie ich na podstawie próby. W ten sposób można wyeliminować niebezpieczeństwo związane z przyjęciem złego rozkładu *a priori* oraz ewentualnego konfliktu między rozkładem *a priori* i danymi.

Problem z tym podejściem polega na tym, że rozkład *a posteriori* uzyskany został przy założeniu, że hiperparametry są znane a tak naprawdę są one estymowane. Wyznaczenie prawidłowego rozkładu *a posteriori* jest zazwyczaj bardzo trudne.

## 14.10. Bayes hierarchiczny

Bayes hierarchiczny polega na tym, że budujemy model Bayesowski dla hiperparametrów. W ten sposób rozwiązujemy problem arbitralności rozkładu *a priori* ponieważ parametry rozkładu *a priori* pochodzą także z modelu, który uwzględnia zarówno wiedzę *a priori* jak i dane. Niestety problem ten będzie z kolei dotyczył parametrów rozkładu *a priori* dla rozkładu *a priori*.

## 14.11. Metody numeryczne

- Rozkład *a posteriori*  $f(\boldsymbol{\theta} | \mathbf{y})$

1. Przyjmujemy arbitralnie wektor wartości początkowych  $\boldsymbol{\theta}^0 = [\theta_1^0, \theta_2^0, \dots, \theta_k^0]$
2. Wygenerowanie jednej realizacji  $\boldsymbol{\theta}^q$  w  $q$ -tym cyklu losowania Gibbsa polega na losowaniu z

warunkowych rozkładów *a posteriori*

$$\theta_1^q = p\left(\theta_1 | \theta_2^{q-1}, \theta_3^{q-1}, \dots, \theta_k^{q-1}\right)$$

$$\theta_2^q = p\left(\theta_2 | \theta_1^q, \theta_3^{q-1}, \dots, \theta_k^{q-1}\right)$$

...

$$\theta_k^q = p\left(\theta_k | \theta_1^q, \theta_2^{q-1}, \dots, \theta_{k-1}^{q-1}\right)$$

- Cykle powtarzane są aż do osiągnięcia zbieżności
- Rozkład *a posteriori*  $f(\theta | \mathbf{y})$  jest punktem stacjonarnym tego algorytmu