Elementy statystyki bayesowskiej

Model statystyczny

\((X, \{P_{\theta}, \theta \in \Theta\})\)

Rozkład \textit{a priori}:
rozkład prawdopodobieństwa \(\Pi\) na zbiorze \(\Theta\)

Rozkład \(P_{\theta}\) oznaczać będziemy \(P(\cdot|\theta)\), a jego gęstość przez \(p(\cdot|\theta)\)

Niech \(\pi(\cdot)\) oznacza gęstość rozkładu \textit{a priori}

Rozkład \textit{a posteriori}:

\[
\pi(\theta|x) = \pi(\theta) \frac{p(x|\theta)}{p(x)}, \quad p(x) = \int_{\Theta} \pi(\theta)p(x|\theta)d\theta
\]
Estymacja bayesowska

Funkcja straty: $L(\hat{\theta}, \theta)$

Ryzyko bayesowskie:

$$\int_\Theta \left[ \int_X L(\hat{\theta}(x), \theta)p(x|\theta)dx \right] \pi(\theta) d\theta$$

$$= \int_X \left[ \int_\Theta L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta \right] p(x) dx$$

Estymator bayesowski minimalizuje

$$\int_\Theta L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta$$

Jeżeli $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, to estymatorem bayesowskim jest wartość oczekiwana rozkładu a posteriori
Bayesowskie testowanie hipotez

Zbiór decyzji

\[ D = \{d_1 = \{\theta = \theta_1\}, d_2 = \{\theta = \theta_2\}\} \]

Funkcja straty \( L(d, \theta) \)

<table>
<thead>
<tr>
<th></th>
<th>( \theta = \theta_1 )</th>
<th>( \theta = \theta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>( L_{11} )</td>
<td>( L_{12} )</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>( L_{21} )</td>
<td>( L_{22} )</td>
</tr>
</tbody>
</table>

Rozkład \textit{a priori}

\[ P\{\theta = \theta_1\} = \pi_1 \quad P\{\theta = \theta_2\} = \pi_2 \quad \pi_1 + \pi_2 = 1 \]

Oczekiwana strata \textit{a posteriori} reguły \( \delta \)

\[
\delta(x) = d_1 : \quad \pi_1 p(x|\theta_1) L_{11} + \pi_2 p(x|\theta_2) L_{12}
\]

\[
\delta(x) = d_2 : \quad \pi_1 p(x|\theta_1) L_{21} + \pi_2 p(x|\theta_2) L_{22}
\]

Reguła bayesowska: obszar przyjęć decyzji \( d_2 \)

\[
\left\{ x \in \mathcal{X} : \frac{p(x|\theta_2)}{p(x|\theta_1)} > \frac{\pi_1[L_{21} - L_{11}]}{\pi_2[L_{12} - L_{22}]} \right\}
\]
Przykład (Model dwumianowy).

Model pojedynczej obserwacji $X$:

$\{(0, 1), \{D(\theta), \theta \in [0, 1]\}\}$

Rozkład a priori $U(0, 1)$: $\pi(\theta) = 1$ ($\theta \in [0, 1]$)

Rozkład a posteriori ($n = 1$)

$$P(1) = \int_{0}^{1} P(1|\theta)\pi(\theta)d\theta = \int_{0}^{1} \theta d\theta = \frac{1}{2}$$

$$P(0) = \int_{0}^{1} P(0|\theta)\pi(\theta)d\theta = \int_{0}^{1} (1 - \theta) d\theta = \frac{1}{2}$$

$$\pi(\theta|1) = \frac{P(1|\theta)\pi(\theta)}{P(1)} = 2\theta$$

$$\pi(\theta|0) = \frac{P(0|\theta)\pi(\theta)}{P(0)} = 2(1 - \theta)$$
Rozkład a posteriori \((n = 2)\)

\[
P(2) = \int_0^1 P(1|\theta)^2 \pi(\theta) \, d\theta = \int_0^1 \theta^2 \, d\theta = \frac{1}{3}
\]

\[
P(1) = 2 \int_0^1 P(1|\theta) P(0|\theta) \pi(\theta) \, d\theta = 2 \int_0^1 \theta (1 - \theta) \, d\theta = \frac{1}{3}
\]

\[
P(0) = \int_0^1 P(0|\theta)^2 \pi(\theta) \, d\theta = \int_0^1 (1 - \theta)^2 \, d\theta = \frac{1}{3}
\]

\[
\pi(\theta|2) = \frac{P(1|\theta)^2 \pi(\theta)}{P(2)} = 3\theta^2
\]

\[
\pi(\theta|1) = \frac{P(1|\theta) P(0|\theta) \pi(\theta)}{P(1)} = 3\theta(1 - \theta)
\]

\[
\pi(\theta|0) = \frac{P(0|\theta)^2 \pi(\theta)}{P(0)} = 3(1 - \theta)^2
\]

***

\[
\int_0^1 x^{a-1}(1 - x)^{b-1} \, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}
\]

\[
\int_0^1 x^m(1 - x)^l \, dx = \frac{m!!}{(m + l + 1)!}
\]
Rozkład *a posteriori*

\[ P(k) = \binom{n}{k} \int_0^1 (1 - \theta)^{n-k} \pi(\theta) d\theta \]

\[ = \binom{n}{k} \int_0^1 \theta^k (1 - \theta)^{n-k} d\theta = \frac{1}{n+1} \]

\[ \pi(\theta|k) = \frac{\binom{n}{k} P(1|\theta)^k P(0|\theta)^{n-k} \pi(\theta)}{P(k)} \]

\[ = (n + 1) \binom{n}{k} \theta^k (1 - \theta)^{n-k} \]

Estymator bayesowski

\[ \hat{\theta}(k) = \int_0^1 \theta \pi(\theta|k) d\theta \]

\[ = (n + 1) \binom{n}{k} \int_0^1 \theta^{k+1} (1 - \theta)^{n-k} d\theta \]

\[ = (n + 1) \binom{n}{k} \frac{(k + 1)! (n - k)!}{(n + 2)!} \]

\[ = \frac{k + 1}{n + 2} \]
**Przykład (Model dwumianowy).**
Niech $\theta_1 < \theta_2 \in (0, 1)$. Model statystyczny
\[
\{\{0, 1, \ldots, n\}, \{B(n, \theta), \theta \in \Theta = \{\theta_1, \theta_2\}\}\}
\]

Zbiór decyzji
\[
\mathcal{D} = \{d_1 = \{\theta = \theta_1\}, d_2 = \{\theta = \theta_2\}\}
\]

Bayesowski obszar przyjęć decyzji $d_2$
\[
\left\{ x \in \mathcal{X} : \frac{p(x|\theta_2)}{p(x|\theta_1)} > \frac{\pi_1 [L_{21} - L_{11}]}{\pi_2 [L_{12} - L_{22}]} \right\}
\]
\[
\frac{\theta_2^k (1 - \theta_2)^{n-k}}{\theta_1^k (1 - \theta_1)^{n-k}} > \gamma
\]
\[
\left[ \frac{\theta_2 (1 - \theta_1)}{\theta_1 (1 - \theta_2)} \right]^k > \gamma \left[ \frac{1 - \theta_1}{1 - \theta_2} \right]^n
\]

Reguła bayesowska
\[
k > n \frac{\log \left[ \frac{1 - \theta_1}{1 - \theta_2} \right]}{\log \left[ \frac{\theta_2}{\theta_1} \right] + \log \left[ \frac{1 - \theta_1}{1 - \theta_2} \right]} + \log \gamma
\]
Niech $\pi_1 = \pi_2 = 0.5$. Funkcja straty $L(d, \theta)$

<table>
<thead>
<tr>
<th></th>
<th>$\theta = \theta_1$</th>
<th>$\theta = \theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$d_2$</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Reguła bayesowska

$$k > n \frac{\log \left[ \frac{1-\theta_1}{1-\theta_2} \right]}{\log \left[ \frac{\theta_2}{\theta_1} \right] + \log \left[ \frac{1-\theta_1}{1-\theta_2} \right]}$$

Ryzyko reguły: prawdopodobieństwo błędnjej decyzji

***

Niech $n = 100$, $\theta_1 = 0.05$, $\theta_2 = 0.15$
Decyzja $d_2$, jeżeli $k > 9.19$

.................................