Metody bayesowskie w statystyce

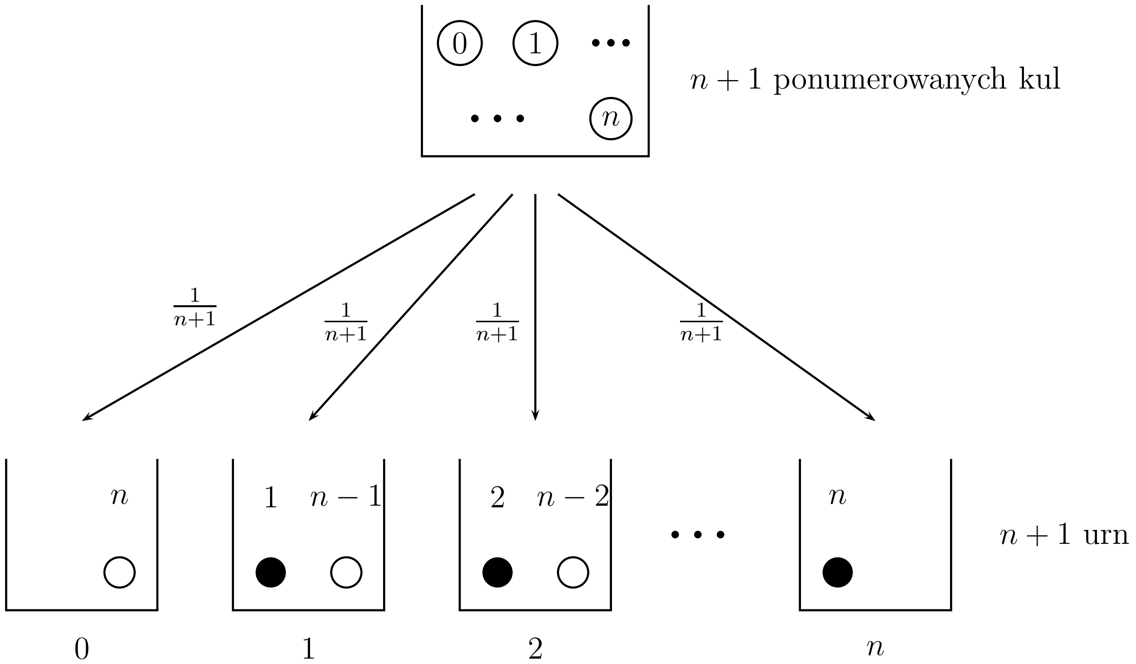
prof. dr hab. W. Niemiro

Spis treści

1. Klasyczny i bayesowski punkt widzenia w statystyce. Przykłady.

05.10.2006r.

Przykład 1 (LAPLACE'A).



1. Wybieramy jedną z urn z jednakowym prawdopodobieństwem $\frac{1}{n+1}$.

2. Losujemy m razy ze zwracaniem (z/z) z wybranej uprzednio urny.

3. Losujemy jeszcze raz, $(n+1)$-szy raz z tejże urny.

Wiadomo, że w m losowaniach wybraliśmy same kule białe.
Jakie jest prawdopodobieństwo, że w kolejnym $(n+1)$-szym losowaniu też wyjdzie biała?

U – numer urny wylosowanej w 1-szym etapie

$$\mathbb{P}(U = i) = \tfrac{1}{n+1} \quad i = 0, 1, \ldots n \quad \text{(rozkład } \textit{a priori} \text{)}$$

S – liczba kul białych w m losowaniach 2-go etapu

$$\mathbb{P}(S = k|U = i) = \tbinom{m}{k}\left(\tfrac{n-i}{n}\right)^k\left(\tfrac{i}{n}\right)^{n-k}$$

X – liczba kul białych w $(m+1)$-szym losowaniu 3-go etapu

$$\mathbb{P}(X = 1|U = i) = \tfrac{n-i}{n} \qquad \mathbb{P}(X = 0|U = i) = \tfrac{i}{n}$$

$$\mathbb{P}(X = 1|S = m) = \frac{\mathbb{P}(X = 1, S = m)}{\mathbb{P}(S = m)} \quad \text{(Twierdzenie Bayesa)}$$

$$\mathbb{P}(S=m|U=i) = \left(1-\frac{i}{n}\right)^m$$
$$\mathbb{P}(S=m) = \sum_{i=0}^{n} \mathbb{P}(S=m|U=i)\mathbb{P}(U=i) = \sum_{i=0}^{n}\left(1-\frac{i}{n}\right)^m \cdot \frac{1}{n+1}$$
$$\mathbb{P}(S=m, X=1) = \sum_{i=0}^{n}\left(1-\frac{i}{n}\right)^{m+1} \cdot \frac{1}{n+1}$$

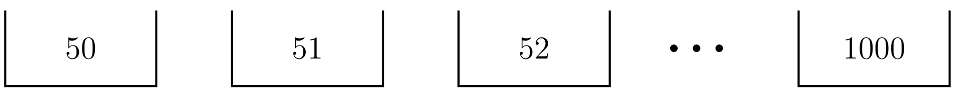
Zatem:

$$\mathbb{P}(X=1|S=m) = \frac{\sum_{i=0}^{n}\left(1-\frac{i}{n}\right)^{m+1} \cdot \frac{1}{n+1}}{\sum_{i=0}^{n}\left(1-\frac{i}{n}\right)^{m} \frac{1}{n+1}} = \frac{\frac{1}{n}\sum_{j=0}^{n}\left(\frac{j}{n}\right)^{m+1}}{\frac{1}{n}\sum_{j=0}^{n}\left(\frac{j}{n}\right)^{m}} =$$
$$= \frac{\frac{1}{n}\sum_{j=0}^{n} f\left(\frac{j}{n}\right)}{\frac{1}{n}\sum_{j=0}^{n} f\left(\frac{j}{n}\right)} \longrightarrow \frac{\int_0^1 x^{m+1}dx}{\int_0^1 x^m dx} = \frac{m+1}{m+2}$$

- Rozkład prawdopodobieństwa definiuje się przy pomocy prawdopodobieństwa warunkowego.
- Subiektywna interpretacja prawdopodobieństwa na 1-szym etapie.
- Twierdzenie Bayesa.
- Nacisk na predykcję.

Przykład 2 (Kapsle Tymbark).

60 kapsli	50 różnych napisów
200 dalszych kapsli	$X = ?$ różnych napisów



$\frac{1}{951}$ – prawdopodobieństwo wyboru urny

$S = 50$

$X = ?$

Zadanie domowe 1.
$\mathbb{P}(X = 190|U = 1000)$

Algorytm 1.

U	S	X
ukryte	znamy	chcemy przewidzieć

1. Losujemy z rozkładu U, $\quad \mathbb{P}(U=i)=\frac{1}{951}$ $\quad$ (z rozkładu a priori)
2. Znając U losujemy z rozkładu $S|U$

rozkład a posteriori $\mathbb{P}(U=i|S=50)$

Jeśli $S \neq 50$, odrzuć (przejdź do 1)

Jeśli $S=50$, wylosuj $X|U$, zapamiętaj

Przykład 3.
Są 2 typy kierowców ostrożni – $\mathcal{O}$ i ryzykanci – $\mathcal{R}$. Na jednego ostrożnego przypada 4 ryzykantów. Jeśli kierowca jest ostrożny, to powoduje szkodę w ciągu roku z prawdopodobieństwem 0.1. Jeśli kierowca jest ryzykantem, to powoduje szkodę w ciągu roku z prawdopodobieństwem 0.4.

1. Ubezpieczamy nowego kierowcę. Jakie jest prawdopodobieństwo, że zgłosi szkodę?
2. Kierowca zgłosił szkodę w 1-szym roku ubezpieczenia. Jakie jest prawdopodobieństwo, że jest typu $\mathcal{O}$?
3. Kierowca zgłosił szkodę w 1-szym roku. Jakie jest prawdopodobieństwo, że zgłosi szkodę i w 2-gim roku?
4. Kierowca zgłosił szkody w dwóch kolejnych latach. Jakie jest prawdopodobieństwo, że zgłosi szkodę i w trzecim roku?

2. Podstawy probabilistyczne: warunkowe rozkłady prawdopodobieństwa, warunkowe wartości oczekiwane, twierdzenie Bayesa.

19.10.2006r.

Warunkowe rozkłady prawdopodobieństwa i wartości oczekiwane.

Gęstość rozkładu prawdopodobieństwa:

$$(\mathscr{X}, \mathcal{A}, \mu)$$

przestrzeń $\qquad$ σ-ciało $\qquad$ miara σ-skończona

$$\mathcal{A} \subseteq 2^{\mathscr{X}}, \quad \mu : \mathcal{A} \longrightarrow \mathbb{R}_+$$

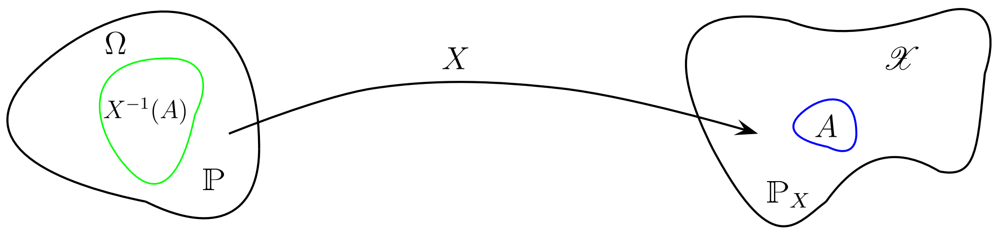
$(\Omega, \mathcal{F}, \mathbb{P})$ – przestrzeń probabilistyczna

$X : \Omega \longrightarrow \mathscr{X}$ – zmienna losowa o wartościach w $\mathscr{X}$

$$\forall_{A\in\mathcal{A}} \qquad X^{-1}(A) = \{\omega \in \Omega;\ X(\omega) \in A\} = \{X \in A\} \in \mathcal{F}$$

Rozkład prawdopodobieństwa zmiennej losowej X

$$\mathcal{A} \ni A \longmapsto \mathbb{P}(X \in A) = \mathbb{P}_X(A)$$
$$\mathbb{P}_X : \mathcal{A} \longrightarrow [0,1]$$
$$\mathbb{P} : \mathcal{F} \to [0,1]$$



Definicja 1.
Rozkład prawdopodobieństwa zmiennej losowej X ma gęstość $f = f_X : \mathscr{X} \longrightarrow \mathbb{R}_+$, jeśli:

$$\forall_{A \in \mathcal{A}} \qquad \mathbb{P}(X \in A) = \int_A f(x)\mu(dx) = \int_A f d\mu$$

Przykład 4.
$\mathscr{X} = \mathbb{R}^{\mathrm{d}}$, $\quad \mu$ – miara Lebesgue'a, $\quad \mathcal{A}$ – σ - ciało borelowskie

$$\int_A f(x)\mu(dx) = \int_A f(x)dx$$

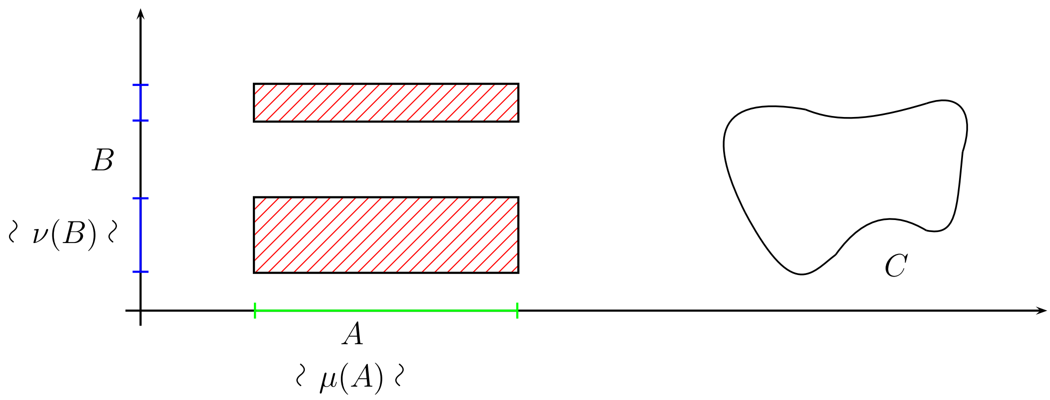
Przykład 5.

$$\left.\begin{array}{l} \mathscr{X} \text{ – przeliczalny, lub skończony,} \qquad \mathcal{A} = 2^{\mathscr{X}} \\ \mu \text{ – miara licząca, tzn.:} \qquad \mu(A) = \begin{cases} \#A & \text{jeśli A jest skończony} \\ \infty & \text{w p.p.} \end{cases} \\ \qquad\qquad \displaystyle\int_A f(x)\mu(dx) = \sum_{x \in A} f(x) \end{array}\right\} \text{przestrzeń dyskretna}$$

$$\left.\begin{array}{l} (\mathscr{X}, \mathcal{A}, \mu) \\ (\mathscr{Y}, \mathcal{B}, \nu) \end{array}\right\} \text{przestrzeń z miarami } \sigma \text{ - skończonymi}$$

$$(\mathscr{X} \times \mathscr{Y}, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$$

$$\mathcal{A} \otimes \mathcal{B} = \sigma(\{A \times B; \quad A \in \mathcal{A}, \quad B \in \mathcal{B}\})$$



$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B)$$

Umowa 1.

$f : \mathscr{X} \times \mathscr{Y} \longrightarrow \mathbb{R}$

$$\iint\limits_{\mathscr{X}\times\mathscr{Y}} f(x,y)\mu \otimes \nu(dx,dy) \overset{tw.\ Fubiniego}{=} \int_{\mathscr{Y}}\int_{\mathscr{X}} f(x,y)\mu(dx)\nu(dy) = \int_{\mathscr{Y}}\int_{\mathscr{X}} f(x,y)dxdy$$

$(X,Y) : \Omega \longrightarrow \mathscr{X} \times \mathscr{Y}$ – zmienna losowa, czyli odwzorowanie mierzalne względem σ - ciała produktowego

Definicja 2.
Rozkład prawdopodobieństwa zmiennej losowej (X,Y) ma gęstość łączną $f = f_{X,Y}$, $\mathscr{X} \times \mathscr{Y} \longrightarrow \mathbb{R}_+$, jeśli:

$$\forall_{C\in\mathcal{A}\otimes\mathcal{B}} \qquad \iint\limits_{C} f(x,y)\mu(dx)\nu(dy) = \mathbb{P}\Big((X,Y) \in C\Big)$$

Uwaga 1.
Wystarczy jeśli

$$\int_B\int_A f(x,y)\mu(dx)\nu(dy) = \mathbb{P}(X \in A, Y \in B) \quad \forall_{A\in\mathcal{A}} \quad \forall_{B\in\mathcal{B}}$$

Uwaga 2.
Rozkład brzegowy zmiennej losowej X ma gęstość

$$f_X(x) = \int_Y f(x,y)\nu(dy)$$

$$\mathbb{P}(X \in A) \overset{tw.\ Fubiniego}{=} \int_A f_X(x)\mu(dx)$$

Uwaga 3.
Jeżeli $\mathscr{X}$ i $\mathscr{Y}$ – dyskretne, to

$$f_X(x) = \mathbb{P}(X = x) = \sum_y f(x,y)$$
$$f(x,y) = \mathbb{P}(X = x, Y = y)$$

Definicja 3.
Załóżmy, że: $(X,Y):\Omega \longrightarrow \mathscr{X}\times\mathscr{Y}$ ma gęstość łączną f.
Gęstość warunkową określamy wzorem:

$$f(y|x)=f_{Y|X}(y|x)=\begin{cases}\frac{f(x,y)}{f_X(x)} & \text{jeśli } f_X(x)>0\\ 0 & \text{w p.p.}\end{cases}$$

Uwaga 4.
Jeżeli $\mathscr{X}$ i $\mathscr{Y}$ dyskretne, to:

$$f(y|x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)}=\mathbb{P}(Y=y|X=x)$$

Definicja 4.
Prawdopodobieństwo warunkowe

$$\mathbb{P}(Y\in B|X=x)=\int_B f(y|x)\nu(dy)\overset{ozn.}{=}p(B,x)$$

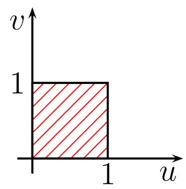
1. Dla dowolnego x, z wyjątkiem zbioru $N=\{x; f(x)=0\}$
   $p(\cdot,x)$ jest miarą probabilistyczną na $(\mathscr{Y},\mathcal{B})$
   Jeżeli $f_X(x)>0$, to $\int_{\mathscr{Y}} f(y|x)\nu(dy)=1$
2. Dla dowolnego $B\in\mathcal{B}$
   $p(\mathcal{B},\cdot):(\mathscr{X},\mathcal{A})\longrightarrow\mathbb{R}$ – jest mierzalne

Uwaga 5.
Jeżeli $\mathscr{X}$ jest przestrzenią dyskretną, to:

$$\mathbb{P}(Y\in B|X=x)=\frac{\mathbb{P}(Y\in B,X=x)}{\mathbb{P}(X=x)}$$

Przykład 6.
U, V – niezależne o rozkładzie $\mathcal{U}(0,1)$

$f_{U,V}(u,v)=1$ dla $(u,v)\in[0,1]^2$



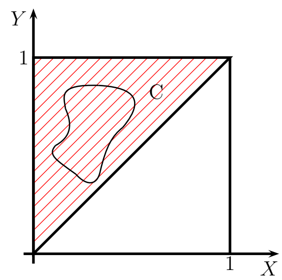
Przykład 7.
Niech: $X=\min(U,V)$ i $Y=\max(U,V)$

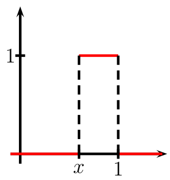
$$f_{X,Y}(x,y)=f(x,y)=\begin{cases}2 & 0\leqslant x\leqslant y\leqslant 1\\ 0 & \text{w p.p.}\end{cases}$$

$X,Y\in C\Leftrightarrow (U,V)\in C$ lub $(U,V)\in C$

$$f_X(x)=\int_x^1 f(x,y)dy=\int_x^1 2dy=2(1-x)$$

$$f(y|x)=\frac{f(x,y)}{f_X(x)}=\begin{cases}\frac{1}{1-x} & 0\leqslant x\leqslant y\leqslant 1\\ 0 & \text{w p.p.}\end{cases}$$





$f(y|x)$ jest gęstością rozkładu $\mathcal{U}(x,1)$

$$\mathbb{P}\Big(Y > \tfrac{9}{10}|X=x\Big) = \int_{\frac{9}{10}}^{1} f(y,x)dy = \begin{cases} 1 & \text{dla } x > \frac{9}{10} \\ \frac{1}{10}\cdot\frac{1}{1-x} & \text{dla } x \geqslant \frac{9}{10} \end{cases}$$

Definicja 5.
Warunkowa wartość oczekiwana

$$\mathbb{E}(Y|X=x) = \int_{\mathscr{Y}} yf(y|x)dy$$

Uwaga 6.
Bardziej ogólnie, jeśli $h : \mathscr{Y} \longrightarrow \mathbb{R} \quad \text{lub} \quad \mathbb{R}^{\mathrm{d}}$

$$\mathbb{E}\Big(h(Y)|X=x\Big) = \int_{\mathscr{Y}} h(y)f(y|x)dy$$

Uwaga 7.
Rozpatrujemy warunkową wartość oczekiwaną jako funkcję x

$$\mathbb{E}\Big(h(Y)|X=x\Big) = r(x) \qquad r : \mathscr{X} \longrightarrow \mathbb{R} \quad \text{lub} \quad \mathbb{R}^{\mathrm{d}}$$

25.10.2006r.

Prawdopodobieństwo warunkowe i warunkowa wartość oczekiwana.

$X, Y : \Omega \longrightarrow \mathscr{X} \times \mathscr{Y}$

$f(x,y)$ – gęstość względem $\mu \otimes \nu$, $\Big($ μ, ν – albo miary Lebesqea albo miary liczące konwencja: $\mu(dx) = dx, \quad \nu(dy) = dy$ $\Big)$

$f(y|x) = \frac{f(x,y)}{f_X(x)}$

$$\mathbb{P}(Y \in B|X=x) \stackrel{df}{=} \int_B f(y|x)dy$$

Przykład 8.
U_1, U_2 – niezależne $\sim \mathcal{U}(0,1)$
$X = \min(U_1,U_2)$, $\quad Y = \max(U_1,U_2)$

$$f(x,y) = \begin{cases} 2\,, & 0 \leqslant x \leqslant y \leqslant 1 \\ 0\,, & \text{w p. p.} \end{cases}$$

Dla ustalonego $x \in (0,1)$

$$f(y|x) = \begin{cases} \frac{1}{1-x}\,, & y \in (x,1) \\ 0\,, & \text{w p. p.} \end{cases}$$

Rozkład warunkowy przy danym $X = x$ jest $\mathcal{U}(x,1)$

Wzór na prawdopodobieństwo całkowite

$$\mathbb{P}(Y \in B) = \int_{\mathscr{X}} \mathbb{P}(Y \in B|X = x) f_X(x)dx$$

Dowód.

$$\mathbb{P}(Y \in B) = \int_B f_Y(y)dy = \int_B \int_{\mathscr{X}} f(x,y)dxdy = \int_B \int_{\mathscr{X}} f(y|x)f_X(x)dxdy$$

$$= \int_{\mathscr{X}} \underbrace{\int_B f(y|x)dy}\, f_X(x)dx$$

□

Wzór Bayesa

$$f(x|y) = \frac{f(y|x)f_X(x)}{f_Y(y)}$$

Dowód.

$f(x|y) = \dfrac{f(x,y)}{f_Y(y)}$

□

Jeżeli μ i ν – miary liczące

$f(x,y) = \mathbb{P}(X = x, Y = y)\,,\ f_X(x) = \mathbb{P}(X = x)\,,\ f(y|x) = \mathbb{P}(Y = y|X = x)$

$$\mathbb{P}(Y \in B) = \sum_x \mathbb{P}(Y \in B|X = x)\mathbb{P}(X = x)$$

Wzór Bayesa dla miar liczących

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(Y = y|X = x)\mathbb{P}(X = x)}{\mathbb{P}(Y = y)}$$

Definicja 6.

$$\mathbb{E}\,(Y|X = x) = \int_{\mathscr{Y}} y f(y|x)dy = r(y)\,, \qquad r : \mathscr{X} \longrightarrow \mathscr{Y}$$

Wzór na prawdopodobieństwo całkowite

$$\mathbb{E}\,Y = \int_{\mathscr{X}} \underbrace{\mathbb{E}\,(Y|X = x)}\, f_X(x)dx$$

Dowód.

$$\mathbb{E}\,Y = \int_{\mathscr{Y}} y f_Y(y)dy = \int_{\mathscr{Y}} y \int_{\mathscr{X}} f(x,y)dxdy = \int_{\mathscr{Y}} y \int_{\mathscr{X}} f(y|x)f_X(x)dxdy$$

$$= \int_{\mathscr{X}} \underbrace{\int_{\mathscr{Y}} y f(y|x)dy}\, f_X(x)dx$$

$$\mathbb{E}\,Y = \int_{\mathscr{X}} \mathbb{E}\,(Y|X = x)f_X(x)dx = \int_{\mathscr{X}} r(x)f_X(x)dx = \mathbb{E}\,r(X)$$

□

Definicja 7.
Zmienną losową $r(X)$ oznaczamy $\mathbb{E}\,(Y|X)$

$$\Omega \xrightarrow{\text{X}} \mathscr{X} \xrightarrow{\text{r}} \mathscr{Y}$$

$$r(X) = r \circ X$$
$$r(X)(\omega) = r(X(\omega))$$

Wniosek 1 (Wzór na prawdopodobieństwo całkowite)**.**

$$\mathbb{E}\,Y = \mathbb{E}\,\mathbb{E}\,(Y|X)$$

Własności warunkowej wartości oczekiwanej – $\mathbb{E}\,(Y|X)$

0. $\mathbb{E}\,(Y|X)$ jest funkcją zmiennej losowej X
1. $\mathbb{E}\,(Y_1 + Y_2|X) = \mathbb{E}\,(Y_1|X) + \mathbb{E}\,(Y_2|X)$
2. $\mathbb{E}\,(aY|X) = a\mathbb{E}\,(Y|X)$
3. Jeżeli X i Y są niezależne, to $\mathbb{E}\,(Y|X) = \mathbb{E}\,Y$

$$f(x,y) = f_X(x)f_Y(y)$$
$$f(y|x) = f_Y(y)$$
$$\mathbb{E}\,(Y|X=x) = \mathbb{E}\,Y$$

4. $\mathbb{E}\,(h(Y)|X=x) = \int_{\mathscr{Y}} h(y)f(y|x)dy\,, \qquad h : \mathscr{Y} \longrightarrow R$
5. $\mathbb{E}\,(h(Y,X)|X=x) = \int_{\mathscr{Y}} h(y,x)f(y|x)dy\,, \qquad h : \mathscr{Y} \times \mathscr{X} \longrightarrow \mathbb{R}$
6. $\mathbb{E}\,(h(Y)g(X)|X=x) = g(x)\mathbb{E}\,(h(Y)|X=x)$

   $\mathbb{E}\,(h(Y)g(X)|X) = g(X)\mathbb{E}\,(h(Y)|X)$
7. $\mathbb{E}\,\mathbb{E}\,(Y|X) = \mathbb{E}\,Y$

Definicja 8.

$$Var(Y|X=x) = \int_{\mathscr{Y}} (y - r(x))^2 f(y|x)dy \overset{ozn.}{=} v(x)$$

Definicja 9.
Zmienną losową $v(X)$ ozn. $Var(Y|X)$

$$Var(Y|X) = \mathbb{E}\,[(Y - \mathbb{E}\,(Y|X))^2|X] = \mathbb{E}\,(Y^2|X) - \mathbb{E}\,(Y|X)^2$$

Ponieważ:

$$\begin{aligned} Var(Y|X) &= \mathbb{E}\,[(Y - \mathbb{E}\,(Y|X))^2|X] = \mathbb{E}\,[Y^2 - 2Y\mathbb{E}\,(Y|X) + \mathbb{E}\,(Y|X)^2|X] \\ &= \mathbb{E}\,(Y^2|X) - 2\mathbb{E}\,(Y|X)\mathbb{E}\,(Y|X) + \mathbb{E}\,(Y|X)^2 = \mathbb{E}\,(Y^2|X) - \mathbb{E}\,(Y|X)^2 \end{aligned}$$

Twierdzenie 1.

$$Var(Y) = \mathbb{E}\,Var(Y|X) + Var\mathbb{E}\,(Y|X)$$

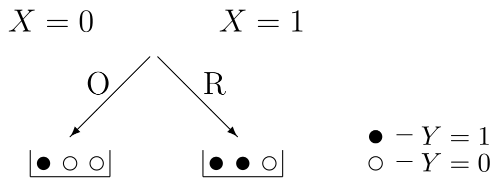
Dowód. $Var(Y) = \mathbb{E}\, v(X) + Var\, r(X)\,, \qquad m = \mathbb{E}\, Y = \mathbb{E}\, r(X)$

$$
\begin{aligned}
Var(Y) &= \int_{\mathscr{Y}} (y-m)^2 f_Y(y)dy = \int_{\mathscr{X}}\int_{\mathscr{Y}} (y-m)^2 f(y|x) f_X(x)dydx \\
&= \int_{\mathscr{X}}\int_{\mathscr{Y}} (y - r(x) + r(x) - m)^2 f(y|x)dy f_X(x)dx \\
&= \int_{\mathscr{X}}\underbrace{\int_{\mathscr{Y}} (y-r(x))^2 f(y|x)dy}_{v(x)} f_X(x)dx + \int_{\mathscr{X}} (r(x)-m)^2 \underbrace{\int_{\mathscr{Y}} f(y|x)dy}_{1} f_X(x)dx \\
&= \int_{\mathscr{X}} v(x) f_X(x)dx + Var\, r(X) = \mathbb{E}\, v(X) + Var\, r(X)
\end{aligned}
$$

$$
\int_X\int_{\mathscr{Y}} (y-r(x))(r(x)-m) f(y|x)dy f_X(x)dx = \int_{\mathscr{X}} (r(x)-m)\underbrace{\int_{\mathscr{Y}} (y-r(x)) f(y|x)dy f_X(x)dx}_{0}
$$

□

Przykład 9.



$Var(Y|X=0) = \frac{2}{9} \qquad Var(Y|X) = \frac{2}{9}$

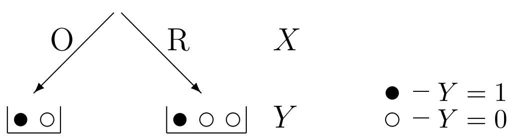
$Var(Y|X=1) = \frac{2}{9} \qquad \mathbb{E}\, Var(Y|X) = \frac{2}{9}$

$\mathbb{E}\,(Y|X=0) = \frac{1}{3}$

$\mathbb{E}\,(Y|X=1) = \frac{2}{3} \qquad Var\mathbb{E}\,(Y|X) = \frac{1}{4\cdot 9}$

26.10.2006r.

Specyfikacja rozkładu łącznego przez rozkłady warunkowe.



$\mathbb{P}(X=0) = \frac{1}{2} = \mathbb{P}(X=R)$

$\mathbb{P}(Y=1|X=0) = \frac{1}{2}$

$\mathbb{P}(Y\neq 1|X=R) = \frac{1}{3}$

Przykład 10.

N – liczba rozszerzeń. $N \sim Poiss(\lambda)$. Każde roszczenie jest uwzględniane z prawdopodobieństwem p lub odrzucane z prawdopodobieństwem $q = 1-p$, niezależnie dla każdego roszczenia.

K – liczba roszczeń uwzględnionych. $\mathbb{P}(K=k) =$ „?”

$\mathbb{P}(K=k|N=n) = \binom{n}{k} p^k q^{n-k}$

$$\mathbb{P}(K=k) = \sum_{n=k}^{\infty} \mathbb{P}(K=k|N=n)\mathbb{P}(N=n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}$$
$$= e^{-\lambda} p^k \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} q^{n-k} \frac{\lambda^n}{n!} = e^{-\lambda} \frac{p^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}\lambda^k}{(n-k)!} q^{n-k}$$
$$\overset{i=n-k}{=} e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{i=0}^{\infty} \frac{(\lambda q)^i}{i!} = e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda q} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}, \qquad K \sim Poiss(\lambda p)$$

3. Budowa statystycznych modeli bayesowskich. rozkłady *a priori* i *a posteriori* , warunkowa niezależność, statystyki dostateczne.

Definicja 10.
Załóżmy, że X, Y_1, Y_2 mają łączną gęstość na $\mathscr{X} \times \mathscr{Y}_1 \times \mathscr{Y}_2$. Mówimy, że Y_1 i Y_2 są WARUNKOWO NIEZALEŻNE przy danym X jeśli

$$\forall_{x,y_1,y_2} \qquad f(y_1, y_2|x) = f(y_1|x)f(y_2|x) \qquad \text{ozn. } Y_1 Y_2 \perp\!\!\!\perp X$$

Wniosek 2.
Jeśli $Y_1 Y_2 \perp\!\!\!\perp X$, to $f(y_2|x, y_1) = f(y_2|x)$

Dowód.

$$f(y_2|x, y_1) = \frac{f(x, y_1, y_2)}{f(x, y_1)} = \frac{f(y_1, y_2|x)f(x)}{f(y_1|x)f(x)} = \frac{f(y_1|x)f(y_2|x)f(x)}{f(y_1|x)f(x)} = f(y_2|x)$$

□

Uogólnienie: $(Y_1, \ldots, Y_n) \perp\!\!\!\perp X$

$$f(y_1, \ldots, y_n|x) = \prod_{i=1}^{n} f(y_i|x)$$

I Rzucamy monetę.

II Losujemy kolejno z tej samej, wybranej urny z/z kule $Y_1, \ldots, Y_n$

Podstawowy model statystyki bayesowskiej.

$\mathscr{Y}$ – przestrzeń obserwacji
Y – zmienna losowa interpretowana jako „obserwacja”

Rodzina rozkładów prawdopodobieństwa na przestrzeni $\mathscr{Y}$ o gęstościach $f_\theta(y)$ (względem ustalonej miary ν na σ - ciele $\mathcal{A}$)

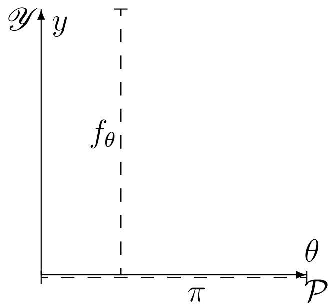
θ – parametr rozkładu prawdopodobieństwa, $\theta \in \mathcal{P}$ – przestrzeń parametrów

$$\Big(\mathscr{Y}, \mathcal{A}, \{f_\theta \nu : \theta \in \mathcal{P}\}\Big) \quad \text{– przestrzeń statystyczna}$$

gdzie $f_\theta \nu$ oznacza miarę probabilistyczną o gęstości f_θ względem ν

Jeżeli na zbiorze $\mathcal{P}$ mamy σ - ciało $\mathcal{C}$ i miara probabilistyczna $\pi\mu$ (o gęstości π względem miary μ na $\mathcal{P}$)

$$\Big(\mathscr{Y}, \mathcal{A}, \{f_\theta \nu : \theta \in \mathcal{P}\}, \mathcal{C}, \pi\mu\Big) = \Big(\mathscr{Y}, \{f_\theta : \theta \in \mathcal{P}\}, \pi\Big) \quad \text{– bayesowska przestrzeń statystyczna}$$



Łączny rozkład prawdopodobieństwa na $\mathcal{P} \times \mathscr{Y}$ ma gęstość

$$f(\theta, y) \stackrel{df}{=} \pi(\theta) f_\theta(y)$$

Możemy traktować parametr jako zmienną losową Θ o rozkładzie prawdopodobieństwa π

Gęstość warunkowa

$$f(y|\theta) = \frac{f(\theta, y)}{\pi(\theta)} = f_\theta(y)$$

Terminologia

π – rozkład *a priori*

$f(y|\theta) = f_\theta(y)$ – wiarogodność

Rozkład *a posteriori*

$$\pi_y(\theta) = f(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)}, \qquad \text{gdzie } f(y) = \int_{\mathcal{P}} f(y|\theta)\pi(\theta) d\theta$$

Przykład 11.

$\mathscr{Y} = \{0, 1, \ldots, n\}$ – przestrzeń obserwacji

$\mathcal{P} = [0, 1]$ – przestrzeń parametrów

Y – liczba sukcesów w schemacie Bernoulliego z prawdopodobieństwem sukcesu θ i liczbą doświadczeń n

$$\mathbb{P}(Y = k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k} = f(k|\theta)$$

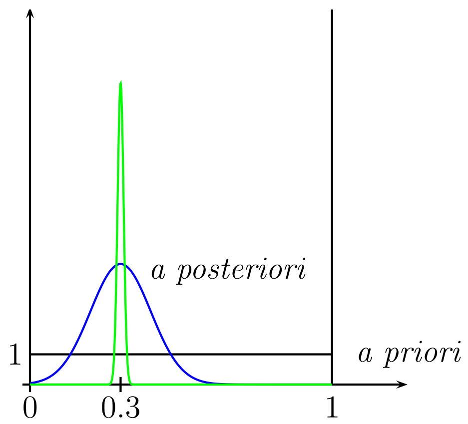
π – rozkład jednostajny $\mathcal{U}(0, 1)$

$\pi(\theta) \equiv 1$

$$f(k) = \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta = \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} d\theta = \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!}$$

Stąd

$$\pi_k(\theta) = \frac{f(k|\theta)\pi(\theta)}{f(k)} = \frac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{\binom{n}{k}\frac{k!(n-k)!}{(n+1)!}} = \frac{(n+1)!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$
$$= (n+1)\binom{n}{k}\theta^k(1-\theta)^{n-k} = const \cdot \theta^k(1-\theta)^{n-k}$$



$$\left.\begin{array}{l} n = 100 \\ k = 30 \end{array}\right\} \quad const\ \theta^{30}(1-\theta)^{70}$$

$$\left.\begin{array}{l} n = 10 \\ k = 3 \end{array}\right\} \quad const\ \theta^{3}(1-\theta)^{7}$$

Szukamy max gęstości

$$\begin{aligned}
&\log \pi_k(\theta) = const + k\log\theta + (n-k)\log(1-\theta) \\
&\big(\log \pi_k(\theta)\big)' = k\cdot\frac{1}{\theta} + (n-k)\cdot\frac{1}{1-\theta}\cdot(-1) \\
&k\cdot\frac{1}{\theta} + (n-k)\cdot\frac{1}{1-\theta}\cdot(-1) = 0 \\
&k\cdot\frac{1}{\theta} = (n-k)\cdot\frac{1}{1-\theta} \\
&k(1-\theta) = (n-k)\theta \\
&(n-k+k)\theta = k \\
&\theta = \frac{k}{n}
\end{aligned}$$

Definicja 11.
Mówimy, że $\Theta \sim \mathcal{B}eta(\alpha,\beta)$, jeśli ma gęstość

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,, \qquad \text{dla } 0<\theta<1\,, \quad \alpha>0\,, \quad \beta>0$$

$\mathcal{B}eta(1,1) \sim U(0,1)$
$\mathcal{B}eta(2,1) \sim 2\theta$
$\mathcal{B}eta(1,2) \sim 2(1-\theta)$

02.11.2006r.

Podstawowy model bayesowski.

Przestrzeń $\begin{array}{ccc} \mathcal{P} & \times & \mathcal{Y} \\ \mu & & \nu \end{array}$

$\Theta \quad Y$

$f(\theta; y) \stackrel{df}{=} \pi(\theta) f(y|\theta)$

gdzie $\pi(\theta)$ – rozkład *a priori* , (gęstość względem miary μ)
$f(y|\theta)$ – wiarogodność, (gęstość względem miary ν)

Zwykle (niekoniecznie) $Y = (Y_1, \ldots, Y_n)$, $\quad Y_1, \ldots, Y_n$ – są warunkowo niezależne przy danym $\Theta = \theta$

$$f(\theta; y_1, \ldots, y_n) = \pi(\theta) \prod_i f(y_i|\theta)$$

$\mathscr{Y} = \mathscr{Y}_1^n$

Rozkład brzegowy obserwacji

$$f(y) = \int_{\mathcal{P}} f(\theta, y) d\theta$$
$$f(y_1, \ldots, y_n) = \int_{\mathcal{P}} f(\theta, y_1, \ldots, y_n) d\theta$$

Rozkład *a posteriori*

$$\pi_y(\theta) = f(\theta|y) = \frac{\pi(\theta) f(y|\theta)}{f(y)} \propto \pi(\theta) f(\theta)$$

Definicja 12.

$$g_1(\theta) \propto g_2(\theta) \quad \Leftrightarrow \quad \exists_{c \neq 0} \quad g_1(\theta) = c g_2(\theta)$$

Czasami $\quad \mathscr{Y} = \mathscr{Y}_1^{n+1}$, Θ, $Y_1, \ldots, Y_n$, Y_{n+1}

gdzie Θ – nieobserwowana zmienna losowa; („wymyślona”)
$Y_1, \ldots, Y_n$ – obserwacje
Y_{n+1} – nieobserwowana zmienna losowa; (nie znamy jej, ale poznamy)

$f(y) = \int f(y|\theta) \pi(\theta) d\theta$

$f(y_2|y_1) = \int f(y_2|\theta, y_1) f(\theta|y_1) d\theta$

Rozkład predykcyjny

$$f(y_{n+1}|y_1, \ldots, y_n) = \int f(y_{n+1}|y_1, \ldots, y_n, \theta) f(\theta|y_1, \ldots, y_n) d\theta = \int f(y_{n+1}|\theta) \pi_{y_1, \ldots, y_n}(\theta) d\theta$$

Przykład 12.
Bernoulli/Beta $\quad \mathcal{P} = [0, 1]$, $\quad \mathscr{Y} = \{0, 1\}^{n+1}$

Schemat Bernoulliego z nieznanym prawdopodobieństwem sukcesu θ
$Y_1, \ldots, Y_n, Y_{n+1}$

$$\begin{cases} \mathbb{P}_\theta(Y = 1) = \theta = \mathbb{P}(Y = 1|\theta) = f(1|\theta) \\ \mathbb{P}_\theta(Y = 0) = 1 - \theta = \mathbb{P}(Y = 0|\theta) = f(0|\theta) \end{cases}$$

$f(y|\theta) = \theta^y(1-\theta)^{1-y}$

$f(y_1,\ldots,y_n|\theta) = \mathbb{P}_\theta(Y_1 = y_1,\cdots,Y_n = y_n) = \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i}$ – rozkład Bernoulliego

Rozkład *a priori* $\Theta \sim \mathcal{B}eta(\alpha,\beta)$

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Rozkład *a posteriori*

$$\pi_{y_1,\ldots,y_n}(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} = \theta^{\alpha+\Sigma Y_i-1}(1-\theta)^{\beta+n-\Sigma y_i-1} \sim \mathcal{B}eta(\alpha+\Sigma y_i,\beta+n-\Sigma y_i)$$

$$\mathbb{E}\left(\Theta|y_1,\ldots,y_n\right) = \frac{\alpha+\Sigma y_i}{\alpha+\beta+n} = \frac{\alpha}{\alpha+\beta}\cdot\frac{\alpha+\beta}{\alpha+\beta+n} + \frac{\Sigma y_i}{n}\cdot\frac{n}{\alpha+\beta+n} = (1-z)\frac{\alpha}{\alpha+\beta} + \overline{y}\cdot z$$

gdzie $z = \dfrac{n}{\alpha+\beta+n}$

dla $n\to\infty \quad z\to 1$

$\dfrac{\alpha}{\alpha+\beta}$ – średnia *a priori*

$\overline{y}$ – średnia z obserwacji

Rozkład brzegowy

$$\begin{aligned}
f(y_1,\ldots,y_n) &= \mathbb{P}(Y_1 = y_1,\ldots,Y_n = y_n) = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+\Sigma y_i-1}(1-\theta)^{\beta+n-\Sigma y_i-1}d\theta \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+\Sigma y_i)\Gamma(\beta+n-\Sigma y_i)}{\Gamma(\alpha+\beta+n)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+s)\Gamma(\beta+n-s)}{\Gamma(\alpha+\beta+n)} \\
&= \frac{\alpha(\alpha+1)(\alpha+2)\ldots(\alpha+s-1)\beta(\beta+1)(\beta+2)\ldots(\beta+n-s-1)}{(\alpha+\beta)(\alpha+\beta+1)\ldots(\alpha+\beta+n-1)} \\
&= \frac{-\alpha(-\alpha-1)(-\alpha-2)\ldots(-\alpha-s+1)(-\beta)(-\beta-1)(-\beta-2)\ldots(-\beta-n+s+1)}{(-\alpha-\beta)(-\alpha-\beta-1)\ldots(-\alpha-\beta-n+1)} \\
&= \frac{\binom{-\alpha}{s}\binom{-\beta}{n-s}}{\binom{-\alpha-\beta}{n}}\cdot\frac{s!(n-s)!}{n!} = \frac{\binom{-\alpha}{s}\binom{-\beta}{n-s}}{\binom{-\alpha-\beta}{n}}\cdot\frac{1}{\binom{n}{s}} = \mathbb{P}(Y_1 = y_1,\ldots,Y_n = y_n)
\end{aligned}$$

$$\mathbb{P}\Big(\sum Y_i = s\Big) = \frac{\binom{-\alpha}{s}\binom{-\beta}{n-s}}{\binom{-\alpha-\beta}{n}}$$

Rozkład predykcyjny

$$\begin{aligned}
f(1|y_1,\ldots,y_n) = \mathbb{P}(Y_{n+1} = 1|y_1,\ldots,y_n) &= \int_0^1 \mathbb{P}(Y_{n+1} = 10|\theta)\pi_{y_1,\ldots,y_n}(\theta)d\theta \\
&= \int_0^1 \theta\pi_{y_1,\ldots,y_n}(\theta)d\theta = \mathbb{E}\left(\Theta|y_1,\ldots,y_n\right) = \frac{\alpha}{\alpha+\beta}(1-z) + \overline{y}\cdot z
\end{aligned}$$

Przykład Laplace'a

$Y_1,\ldots,Y_n,Y_{n+1}$

$$\mathbb{P}(Y_1 = y_1,\ldots,Y_n = y_n|\theta) = \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i}, \qquad \text{gdzie} \quad \theta = \frac{k}{m}$$

Rozkład *a priori*

$$\pi(\frac{k}{m}) = \mathbb{P}(\Theta = \frac{k}{m}) = \frac{1}{m+1}$$

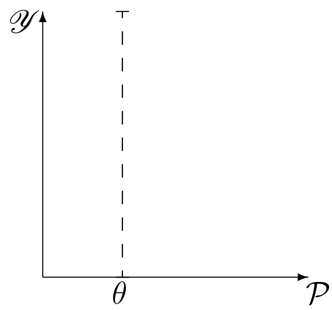
$$z = \frac{n}{n+\alpha+\beta} = \frac{n}{n+2}$$

$$\mathbb{P}(Y_{n+1} = 1|Y_1 = \ldots = Y_n = 1) = \tfrac{1}{2} \cdot \frac{2}{n+2} + 1 \cdot \frac{n}{n+2} = \frac{n+1}{n+2} \qquad \text{w granicy}$$

09.11.2006r.

Dostateczność.

$\mathscr{Y}$ – przestrzeń obserwacji
$\mathcal{P}$ – przestrzeń parametrów
$\{f_\theta\}$ – rodzina rozkładów prawdopodobieństwa na $\mathscr{Y}$
π – rozkład *a priori* na $\mathcal{P}$



$f_\theta(y) = f(y|\theta)$

$f(\theta, y) \stackrel{df}{=} f_\theta(y)\pi(\theta)$

Definicja 13.
Statystyka jest to funkcja mierzalna

$$T : \mathscr{Y} \longrightarrow \mathscr{T}$$

Przykład 13.
$\mathscr{Y} = \mathbb{R}^n\,, \quad Y = (Y_1, \ldots, Y_n)$

$\overline{Y} = \frac{1}{n}\sum_i Y_i$

$T : \mathbb{R}^n \longrightarrow \mathbb{R}$

Definicja 14.
Statystyka T jest dostateczna, jeśli dla każdego θ rozkład warunkowy zmiennej Y przy danym $T(Y) = t$ jest taki sam (nie zależy od θ).

$\mathscr{Y}$ i $\mathscr{T}$ – przestrzenie dyskretne (założenie upraszczające)

Definicja 15.
T jest dostateczne, jeśli $\mathbb{P}_\theta(Y = y|T(Y) = t)$ – nie zależy od θ

Uwaga 8 (DOST)**.**
$\mathbb{P}_\theta(Y = y|T(Y) = t) = 0$, jeśli $T(y) \neq t$
rozkład warunkowy jest skupiony na $\{y; T(y) = t\}$

Uwaga 9.
W modelu bayesowskim dostateczność można zapisać

$$\mathbb{P}(Y=y|T(Y)=t,\Theta=\theta)=\mathbb{P}(Y=y|T(Y)=t)$$

Definicja 16 (FAKT). Kryterium faktoryzacji
T jest dostateczna, jeśli

$$f_\theta(y)=\mathbb{P}_\theta(Y=y)=\mathbb{P}(Y=y|\Theta=\theta)$$

daje się przedstawić jako $\quad h_\theta\big(T(y)\big)g(y)$ (nie zależy od θ)

Definicja 17 (WN).
Warunkowa niezależność

$$\mathbb{P}\Big(Y=y,\Theta=\theta|T(Y)=t\Big)=\mathbb{P}\Big(Y=y|T(Y)=t\Big)\mathbb{P}\Big(\Theta=\theta|T(Y)=t\Big)$$

Umowa: $\quad T\overset{df}{=}T(Y)$

Definicja 18 (BDOST). Bayesowska dostateczność

$$\mathbb{P}\Big(\Theta=\theta|Y=y\Big)=\mathbb{P}\Big(\Theta=\theta|T(Y)=T(y)\Big)$$

Twierdzenie 2.
DOST $\Leftrightarrow$ FAKT $\Leftrightarrow$ WN $\Leftrightarrow$ BDOST

Dowód. DOST $\Rightarrow$ FAKT

$$\begin{aligned}f_\theta(y)&=\mathbb{P}(Y=y|\Theta=\theta)=\mathbb{P}(Y=y,T=T(y)|\Theta=\theta)\\&=\mathbb{P}(Y=y|T=T(y),\Theta=\theta)\mathbb{P}(T=T(y)|\Theta=\theta)\\&\overset{DOST}{=}\mathbb{P}(Y=y|T=T(y))\mathbb{P}(T=T(y)|\Theta=\theta)\end{aligned}$$

FAKT $\Rightarrow$ BDOST

$$\begin{aligned}\mathbb{P}(\Theta=\theta|Y=y)&=\frac{\mathbb{P}(Y=y|\Theta=\theta)\mathbb{P}(\Theta=\theta)}{\mathbb{P}(Y=y)}=\frac{h_\theta\big(T(y)\big)g(y)\mathbb{P}(\Theta=\theta)}{\mathbb{P}(Y=y)}\\&=\frac{h_\theta\big(T(y)\big)g(y)\mathbb{P}(\Theta=\theta)}{\sum\limits_{\theta'}h_{\theta'}\big(T(y)\big)g(y)\mathbb{P}(\Theta=\theta')}=\mathbb{P}(\Theta=\theta|T=T(y))\end{aligned}$$

$\mathbb{P}(\Theta=\theta|Y=y,T=T(y))=\mathbb{P}(\Theta=\theta|T=T(y))$

WN $\Leftrightarrow$ DOST $\Leftrightarrow$ BDOST

$\square$

Przykład 14.
Model Bin/Beta

$Y_1, \ldots, Y_n$

$f(1|\theta) = \mathbb{P}(Y_i = 1|\theta) = \theta\,, \qquad \mathbb{P}(Y_i = 0|\theta) = 1 - \theta = f(0|\theta)$

$Y_1, \ldots, Y_n$ – warunkowo niezależne $|\theta$

$\pi(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$

Kryterium faktoryzacji

$$f(y_1, \ldots, y_n|\theta) \overset{WN}{=} f(y_1|\theta)\ldots f(y_n|\theta) = \prod_{i=1}^{n}\Big(\theta^{y_i}(1-\theta)^{1-y_i}\Big) = \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i}\cdot 1$$

$S(y) = \sum y_i$ – jest statystyką dostateczną

Rozkład warunkowy

$(Y_1, \ldots, Y_n)$ przy danym $S(Y_1, \ldots, Y_n) = \sum Y_i = S$

$$\mathbb{P}\Big(Y_1 = y_1, \ldots, Y_n = y_n|\sum Y_i = s\Big) = \begin{cases} 0 & \text{jeśli } \sum y_i \neq s \\ \frac{\mathbb{P}(Y_1=y_1,\ldots,Y_n=y_n,S=s)}{\mathbb{P}(S=s)} & \end{cases}$$

$$= \frac{\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n)}{\mathbb{P}(S=s)} = \frac{\theta^s(1-\theta)^{n-s}}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} = \frac{1}{\binom{n}{s}}$$

BDOST

$$\pi_y(\theta) = f(\theta|y_1, \ldots, y_n) = \frac{f(y_1, \ldots, y_n|\theta)\pi(\theta)}{f(y_1, \ldots, y_n)} \propto \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$f(\theta|y_1, \ldots, y_n) = f(\theta|S=s) \quad$ gdzie $s = \sum y_i$

Jeśli $f(y|\theta)$ i $\pi(\theta)$ są gęstościami względem dowolnych miar σ - skończonych, to

FAKT $\quad f(y|\theta) = h_\theta\big(T(y)\big)g(y)$

WN $\quad f(y,\theta|t) = f(y|t)f(\theta|t)$

BDOST $\quad f(\theta|y) = f(\theta|t) \quad$ gdzie $t = T(y)$

15.11.2006r.

Wykładnicze rodziny rozkładów prawdopodobieństwa.

$\mathscr{X}$ – przestrzeń obserwacji
X – zmienna losowa (obserwacja) o gęstości $f_\theta(x)\,, \quad x \in \mathscr{X}\,, \quad \theta \in \Theta$

Definicja 19.
Rodzina rozkładów prawdopodobieństwa na $\mathscr{X}$ jest rodziną wykładniczą jeśli gęstości są postaci:

$$(*) \qquad f_\theta(x) = \exp\left(\sum_{j=1}^{k} T_j(x)g_j(\theta) + g_0(\theta)\right)h(x)\,, \qquad \theta \in \Theta$$

Przykład 15.
Rodzina rozkładów $U(0,\theta)$, $\theta > 0$ nie jest rozkładem, wykładniczym (ponieważ zbiór na którym sie zeruje ta gęstość zależy od θ)

$$f_\theta(x) = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x)$$

Przykład 16.
$\{Ex(\theta) : \theta > 0\}$ jest rodziną wykładniczą, $\quad \mathscr{X} = [0,\infty)$

$$f_\theta(x) = \theta e^{-\theta x} = \exp(-\theta x + \log\theta)$$

Przykład 17.
$\{Poiss(\theta) : \theta > 0\}$ jest rodziną wykładniczą, $\quad \mathscr{X} = \{0,1,2,\ldots\}$

$$f_\theta(x) = e^{-\theta}\frac{\theta^x}{x!} = \exp\Big(-\theta + x\log\theta\Big)\frac{1}{x!}$$

Przykład 18.
$\{Gamma(\alpha,\lambda) : \alpha > 0, \lambda > 0\}$ jest rodziną wykładniczą, $\quad \mathscr{X} = [0,\infty), \quad \theta = (\alpha,\lambda)$

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} = \exp\left(-\lambda x + (\alpha-1)\log x + \log\frac{\lambda^\alpha}{\Gamma(\alpha)}\right)$$

Jeżeli $X_1,\ldots,X_n$ jest próbką (iid) $\sim f_\theta$, to

$$f_\theta(x_1,\ldots,x_n) = \prod_{i=1}^n f_\theta(x_i) = \exp\left\{\sum_{j=1}^k\left(\sum_{i=1}^n T_j(x_i)\right)g_j(\theta) + ng_0(\theta)\right\}\prod_{i=1}^n h(x_i)$$

Z kryterium faktoryzacji wynika, że

$$\left(\sum_{i=1}^n T_1(X_i),\ldots,\sum_{i=1}^n T_k(X_i)\right)$$ – jest k - wymiarową statystyką dostateczną

gdzie n – rozmiar próbki
k – ilość składników rodziny wykładniczej

Naturalna parametryzacja rodziny wykładniczej

$$f_\theta(x) = \exp\left\{\sum_{j=1}^k T_j(x)\theta_j + \ell(\theta)\right\}h(x)$$

$$\ell(\theta) = -\log\int_{\mathscr{X}}\exp\left\{\sum_{j=1}^k T_j(x)\theta_j\right\}h(x)dx$$

$\ell(\theta) = \ell(\theta_1,\ldots,\theta_k)$

Stwierdzenie 1.

$$\mathbb{E}_\theta T_j(X) = -\frac{\partial\ell(\theta)}{\partial\theta_j}$$

Dowód.

$$0 = \frac{\partial}{\partial \theta_j} \int_{\mathscr{X}} f_\theta(x)dx = \int_{\mathscr{X}} \frac{\partial}{\partial \theta_j} f_\theta(x)dx = \int_{\mathscr{X}} \left(T_j(x) + \frac{\partial \ell}{\partial \theta_j} \right) \exp\left(\sum_{r=1}^k T_r(x)\theta_r + \ell(\theta) \right) h(x)dx$$
$$= \int_{\mathscr{X}} T_j(x) f_\theta(x)dx + \frac{\partial \ell}{\partial \theta_j} \int_{\mathscr{X}} f_\theta(x)dx = \mathbb{E}_\theta T_j(X) + \frac{\partial \ell}{\partial \theta_j}$$

□

Przykład 19.
Rozkład dwumianowy, $\mathcal{B}in(p,n)$, $p \in (0,1)$

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \exp\left\{ x \log p + (n-x)\log(1-p) \right\} \binom{n}{x}$$
$$= \exp\left\{ x \log \frac{p}{1-p} + n \log(1-p) \right\} \binom{n}{x}$$

$\theta = \log \dfrac{p}{1-p}$

$\ell(\theta) = n \log(1-p) = n \log \dfrac{1}{1+e^\theta}$

$\dfrac{\partial \ell}{\partial \theta} = -n \cdot \dfrac{e^\theta}{1+e^\theta} = -n \cdot p$

Definicja 20.
Dla rodziny wykładniczej postaci $(*)$ rodzina gęstości

$$\pi(\theta) = \pi_\alpha(\theta) \propto \exp\left\{ \sum_{j=1}^k \alpha_j g_j(\theta) + \alpha_0 g_0(\theta) \right\}$$

nazywa się sprzężoną rodziną rozkładów *a priori* , $\big(\ \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_k)\ \big)$

Załóżmy, że $X_1, \ldots, X_n$ są warunkowo *iid* (przy danym θ) o gęstości f_θ

Stwierdzenie 2.
Rozkład *a posteriori* ma postać

$$\pi(\theta|x_1) \propto f_\theta(x_1)\pi(\theta) \propto \exp\left\{ \sum_{j=1}^k T_j(x) g_j(\theta) + g_0(\theta) \right\} \exp\left\{ \sum_{j=1}^k \alpha_j g_j(\theta) + \alpha_0 g_0(\theta) \right\}$$
$$= \exp\left\{ \sum_{j=1}^k \Big(\alpha_j + T_j(x) \Big) g_j(\theta) + (\alpha_0 + 1) g_0(\theta) \right\}$$

$$\pi(\theta|x_1, \ldots, x_n) \propto f_\theta(x_1, \ldots, x_n)\pi(\theta) \propto \exp\left\{ \sum_{j=1}^k \left(\alpha_j + \sum_{i=1}^n T_j(x_i) \right) g_j(\theta) + (\alpha_0 + n) g_0(\theta) \right\} = \pi_\alpha(\theta)$$

gdzie $\alpha_1' = \alpha_1 + \sum_{i=1}^n T_1(x_i)$

$$\begin{aligned}
&\vdots\\
&\alpha_k' = \alpha_k + \sum_{i=1}^{n} T_k(x_i)\\
&\alpha_0' = \alpha_0 + n
\end{aligned}$$

Przykład 20.
$\{Ex(\theta) : \theta > 0\}$

$$f_\theta(x) = \theta e^{-\theta x} = \exp\Big\{ -\theta x + \log\theta \Big\}$$

$$f_\theta(x_1, \ldots, x_n) = \exp\left\{ -\theta \sum_{i=1}^{n} x_i + n\log\theta \right\}$$

sprzężona rodzina rozkładów *a priori*

$$\pi_\alpha(\theta) = \pi_{\alpha_0,\alpha_1}(\theta) \propto \exp\Big\{ -\theta\alpha_1 + \alpha_0 \log\theta \Big\} = \theta^{\alpha_0} \exp\Big\{ -\alpha_1 \cdot \theta \Big\} = \mathcal{G}amma(\alpha_0 + 1, \alpha_1)$$

Rozkład *a posteriori* $\quad \mathcal{G}amma(\alpha_0 + n + 1, \alpha_1 + \Sigma x_i)$

16.11.2006r.

Typowe modele bayesowskie ze sprzężonymi rodzinami rozkładów.

Bin/Beta

$X_1, \ldots, X_n$
$\mathbb{P}(X_i = 1|\theta) = \theta = f(1|\theta) \qquad \theta \in (0, 1)$
$\mathbb{P}(X_i = 0|\theta) = 1 - \theta = f(0|\theta)$

rozkład *a priori*

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$\mathcal{U}(0, 1) = \mathcal{B}eta(1, 1)$

rozkład *a posteriori*

$$\begin{aligned}
f(\theta|x_1, \ldots, x_n) =& \frac{f(x_1, \ldots, x_n|\theta)\pi(\theta)}{f(x_1, \ldots, x_n)} \propto \left(\prod_{i=1}^{n} f(x_i|\theta) \right) \pi(\theta)\\
=& \left(\prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \right) \pi(\theta) = \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i} \theta^{\alpha-1}(1-\theta)^{\beta-1}\\
=& \theta^{\Sigma x_i + \alpha - 1}(1-\theta)^{n - \Sigma x_i + \beta - 1} \sim \mathcal{B}eta(\Sigma x_i + \alpha, n - \Sigma x_i + \beta)
\end{aligned}$$

2 sukcesy i 3 porażki $\qquad \Rightarrow \qquad \left.\begin{matrix} \mathcal{B}eta(\alpha+2, 3+\beta) \\ \alpha, \beta = 1 \end{matrix}\right\} \quad \mathcal{B}eta(3, 4)$

$\widehat{\theta}_B = \mathbb{E}\left(\theta|x_1, \ldots, x_n\right) = \dfrac{s+\alpha}{n+\alpha+\beta}$ – estymator bayesowski parametru θ

$\widehat{\theta} = \frac{s}{n}$

$$f_\theta(x) = \theta^x(1-\theta)^{1-x} = \exp\left\{x\log\theta + (1-x)\log(1-\theta)\right\} = \exp\left\{x\log\frac{\theta}{1-\theta} + \log(1-\theta)\right\}$$

$\sum x_i$ – statystyka dostateczna

Sprzężona rodzina rozkładu *a priori*

$$\begin{aligned}\pi(\theta) &\propto \exp\left\{\alpha_1\log\frac{\theta}{1-\theta} + \alpha_0\log(1-\theta)\right\}\\ &\propto \left(\frac{\theta}{1-\theta}\right)^{\alpha_1}(1-\theta)^{\alpha_0} = \theta^{\alpha_1}(1-\theta)^{\alpha_0-\alpha_1} \sim \mathcal{B}eta(\alpha_1+1, \alpha_0-\alpha_1+1)\end{aligned}$$

Poisson/Gamma

$f(k|\theta) = \mathbb{P}(X=k) = e^{-\theta}\frac{\theta^k}{k!}, \qquad X \sim Poiss(\theta)$

rozkład *a priori*

$\pi(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta} \propto \theta^{\alpha-1}e^{-\lambda\theta}$ rozkład *a posteriori*

$$\begin{aligned}f(\theta|x_1,\ldots,x_n) &= \frac{f(x_1,\ldots,x_n)|\theta)\pi(\theta)}{f(x_1,\ldots,x_n)} \propto \left(\prod_{i=1}^n f(x_i|\theta)\right)\pi(\theta)\\ &= \left(\prod_{i=1}^n \frac{\theta^{x_i}}{x_i!}e^{-\theta}\right)\theta^{\alpha-1}e^{-\lambda\theta} \propto \theta^{\Sigma x_i}\theta^{\alpha-1}e^{-\lambda\theta}e^{-n\theta} = \theta^{\Sigma x_i+\alpha-1}e^{-(\lambda+n)\theta}\end{aligned}$$

$\widehat{\theta}_B = \mathbb{E}\left(\theta|X_1,\ldots,X_n\right) = \frac{s+\alpha}{\lambda+n} \sim \mathcal{G}amma(s+\alpha, \lambda+n)\,, \qquad$ gdzie $s = \Sigma x_i$

4. Funkcje straty, estymacja i predykcja bayesowska.

23.11.2006r.

Teoria decyzji statystycznych.

$\mathscr{X}$ – przestrzeń obserwacji
$\mathcal{P}$ – przestrzeń parametrów
$\{\mathbb{P}_\theta; \theta\in\Theta\}$ – rodzina rozkładów prawdopodobieństwa na $\mathscr{X}$
X – obserwacja, $X \sim \mathbb{P}_\theta$
$\mathcal{A}$ – przestrzeń akcji (decyzji), $a \in \mathcal{A}$

Definicja 21.
Reguła decyzyjna

$$\delta : \mathscr{X} \longrightarrow \mathcal{A}$$

„na podstawie (losowej) obserwacji X podejmujemy akcję (decyzję) $\delta(X)$"

Definicja 22.
Funkcja strat
$$L : \mathcal{P} \times \mathcal{A} \longrightarrow \mathbb{R} \qquad (\mathbb{R}_+)$$
$L(\theta, a)$ – strata jaką ponosi statystyk, jeśli podejmie akcję a, a wartością parametru jest θ

GRA

I gracz – Natura	II gracz – Statystyk
wybiera $\theta \in \Theta$	wybiera $a \in \mathcal{A}$
losujemy $X \sim \mathbb{P}_\theta$	$a = \delta(X)$

1940 Abraham Wald
$\mathbb{P}_\theta$ – ma gęstość f_θ

Definicja 23.
Funkcja ryzyka
$$R(\theta, \delta) = \mathbb{E}_\theta L(\theta, \delta(X)) = \int_{\mathscr{X}} L(\theta, \delta(X)) f_\theta(x) dx$$

Przykład 21 (estymacja).
$\mathcal{P} \subseteq \mathbb{R}, \quad \mathcal{A} = \mathbb{R}$
$L(\theta, a) = (\theta - a)^2$ – kwadratowa funkcja straty
$R(\theta, \delta) = \mathbb{E}_\theta(\theta - \delta(X))^2 \qquad \delta$ – estymator, $\delta(X) \approx \hat{\theta}(X)$

Przykład 22 (estymacja).
$L(\theta, a) = |\theta - a|$ – błąd absolutny

Przykład 23 (estymacja).
$\mathcal{P} = \mathbb{R}_+ = (0, \infty)$
$$L(\theta, a) = \frac{|\theta - a|}{\theta}$$
$$L(\theta, a) = \frac{(\theta - a)^2}{\theta^2}$$

Przykład 24 (testowanie hipotez).
$\mathcal{P} = \{\theta_0, \theta_1\}$
H_0 – hipoteza zerowa
H_1 – hipoteza alternatywna
$X \sim \mathcal{P}_\theta \qquad \theta = \theta_0 \qquad$ lub $\qquad \theta = \theta_1$
$\mathcal{A} = \{a_0, a_1\}$
a_0 – brak podstaw do odrzucenia H_0

a_1 – odrzucamy H_0 na rzecz H_1

Rzut kostką
$\mathscr{X} = \{1, 2, \ldots, 6\}$

$\mathbb{P}_{\theta_0}$

1	2	3	4	5	6
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

„dobra kostka”

$\mathbb{P}_{\theta_1}$

1	2	3	4	5	6
$\frac{1}{6} - \frac{1}{100}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6} + \frac{1}{100}$

„oszukana kostka”

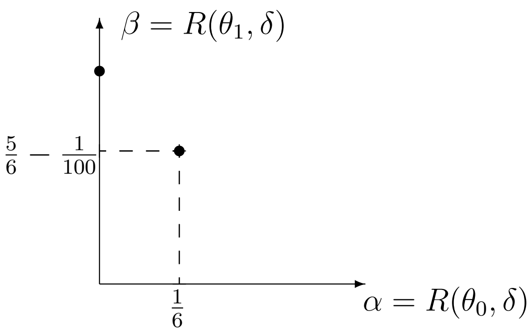
$L(\theta_0, a_0) = L(\theta_1, a_1) = 0$
$L(\theta_1, a_0) = L(\theta_0, a_1) = 1$

$$\delta(x) = \begin{cases} a_1\,; & \text{jeśli} \quad x = 6 \\ a_0\,; & \text{w p.p.} \end{cases}$$

$$\underbrace{\begin{array}{c|c|c|c|c} 1 & 2 & 3 & 4 & 5 \\ \hline \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array}}_{a_0} \underbrace{\begin{array}{|c} 6 \\ \hline \frac{1}{6} \end{array}}_{a_1}$$ – obszar krytyczny

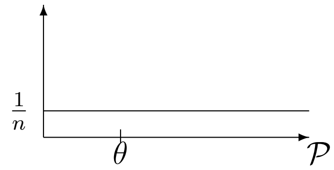
$$R(\theta_0, \delta) = \mathbb{E}_{\theta_0} L(\theta_0, \delta(X)) = \mathbb{P}_{\theta_0}(\delta(X) = a_1) = \frac{1}{6} \qquad \text{błąd I - go rodzaju}$$
$$R(\theta_1, \delta) = \mathbb{E}_{\theta_1} L(\theta_1, \delta(X)) = \mathbb{P}_{\theta_1}(\delta(X) = a_0) = \frac{5}{6} - \frac{1}{100} \qquad \text{błąd II - go rodzaju}$$

decyzje statystyka / stan natury	a_0	a_1
$H_0 \ : \theta_0$	O.K.	błąd I - szego rodzaju ozn. α
$H_1 \ : \theta_1$	błąd II - go rodzaju ozn. β	O.K.

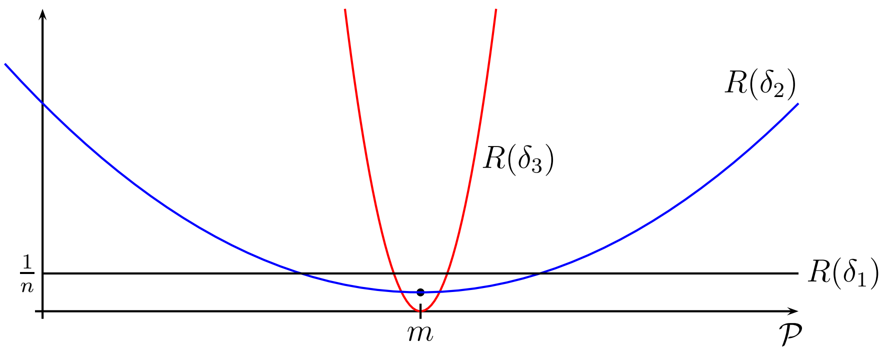


Zadanie domowe 2.
Znaleźć funkcję ryzyka dla wszystkich reguł decyzyjnych.

Przykład 25.
$X = (X_1, \ldots X_n) \underset{iid}{\sim} \mathcal{N}(\theta, 1)$
$L(\theta, a) = (\theta - a)^2$
$\overline{X} = \delta_1(X)$
$R(\theta, \delta_1) = \mathbb{E}_\theta(\theta - \overline{X})^2 = Var_\theta(\overline{X}) = \frac{1}{n}$



Estymator bayesowski dla rozkładu *a priori* $\theta \sim \mathcal{N}(m, a^2)$
$\widehat{\theta}_B = z\overline{X} + (1 - z)m$
$$z = \frac{na^2}{na^2 + \delta^2} = \frac{na^2}{na^2 + 1}$$
$\sigma^2 = 1$



$\delta_3(X) = m$
$R(\theta, \delta_3) = \mathbb{E}_\theta(\theta - m)^2 = (\theta - m)^2$

Definicja 24.
Ryzyko bayesowskie (dla rozkładu *a priori* π)

$$r(\delta) = r(\pi, \delta) = \mathbb{E}\, L(\Theta, \delta(X))$$
$$= \int_{\mathcal{P}} \int_{\mathscr{X}} L(\theta, \delta(x)) f_\theta(x) dx \pi(\theta) d\theta = \int_{\mathcal{P}} R(\theta, \delta) \pi(\theta) d\theta = \int_{\mathcal{P}} \mathbb{E}\, \Big(L(\Theta, \delta(X)) | \Theta = \theta\Big) \pi(\theta) d\theta$$

$$r(\delta) = \mathbb{E}\,\mathbb{E}\, \Big(L(\Theta, \delta(X)|X)\Big)$$
$$= \int_{\mathscr{X}} \underbrace{\mathbb{E}\, \Big(L(\Theta, \delta(X)|X = x)\Big)}_{\text{ryzyko } a \text{ posteriori}} f(x) dx = \int_{\mathscr{X}} \underbrace{\int_{\mathcal{P}} L(\theta, \delta(x)) f(\theta|x) d\theta}_{\text{ryzyko } a \text{ posteriori}} f(x) dx$$

gdzie $\quad f(\theta|x) = \dfrac{f_\theta(x)\pi(\theta)}{f(x)}$

Jak zminimalizować $r(\delta)$?

Dla każdego $x \in \mathscr{X}$ znajdujemy $\delta(x)$ dla którego

$$\int_{\mathcal{P}} L(\theta,a)f(\theta|x)d\theta \longrightarrow \min_a$$

Przykład 26.
$X = (X_1, \dots X_n) \underset{iid}{\sim} \mathcal{N}(\theta, \sigma^2), \qquad \Theta \sim \mathcal{N}(m, a^2)$
$L(\theta, a) = (\theta - a)^2$

Rozkład *a posteriori*

$$\Theta \mid X_1, \dots X_n \sim \mathcal{N}\Big(z\overline{X} + (1-z)m, \tfrac{a^2\sigma^2}{na^2+\sigma^2}\Big)$$

Ryzyko *a posteriori*

$$\mathbb{E}\left[(\Theta - d)^2 | X_1, \dots X_n\right]$$

$\mathbb{E}\,(\Theta - d)^2$ jest najmniejsze dla $d_* = z\overline{X} + (1-z)m, \quad r_* = \frac{a^2\sigma^2}{na^2+\sigma^2}$

30.11.2006r.

Problemy klasyfikacji/dyskryminacji.

R. Fisher 1930

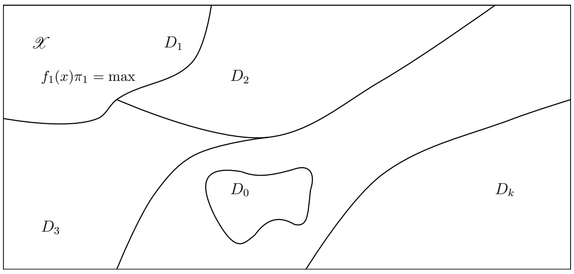
Obiekt opisany „wektorem cech” $X = (X_1, \dots, X_d)$ należy do jednej z k klas. Na podstawie obserwacji X mamy zdecydować, do której klasy zaliczyć obiekt.
$\mathcal{P} = \{1, 2, \dots, k\}$ – identyfikator klasy (przestrzeń parametrów)
$\mathscr{X} \subseteq \mathbb{R}^d$ – przestrzeń obserwacji
$\mathcal{A} = \{1, 2, \dots, k\}$ lub $\mathcal{A} = \{0, 1, \dots, k\}$ – przestrzeń decyzji, (0 – zawieszenie decyzji)

Rozkłady prawdopodobieństwa wektora X przy danym $\theta \in \mathcal{P}$ są znane.
(gęstość f_θ) $f_1, \dots, f_k$ – gęstości w klasach
$\pi(\theta) = \pi_\theta$ – prawdopodobieństwo *a priori* pojawienia się obiektu z klasy θ
$\pi_1 + \ldots + \pi_k = 1$

Funkcja strat $L(\theta, a)$ – macierz $k \times (k+1)$
$L(\theta, \theta) = 0$
$L(\theta, a) \geqslant 0$
$L(\theta, 0) \leqslant L(\theta, a)$ dla dowolnego $a \neq \theta$

Reguła decyzyjna $\delta : \mathscr{X} \longrightarrow \mathcal{A}$

Obszary decyzyjne $D_j = \{x \in \mathscr{X}; \delta(x) = j\}$ ozn. $\theta = i$
$a = j$



Ryzyko

$$\begin{aligned} R_i =& \mathbb{E}\left[L(i,\delta(X))|\theta=i\right] = \int_{\mathscr{X}} L(i,\delta(x))f_i(x)dx \\ =& \sum_{j=0}^{k}\int_{D_j} L(i,j)f_i(x)dx = \sum_{j=0}^{k} L(i,j)\underbrace{\mathbb{P}(X\in D_j|\theta=i)}_{\int_{D_j} f_j(x)dx} \end{aligned}$$

Ryzyko bayesowskie

$$r = \sum_{i=1}^{k}\pi_i R_i = \mathbb{E}\, L\Big(\theta,\delta(X)\Big) = \sum_{i=1}^{k}\sum_{j=0}^{k}\pi_i L(i,j)\mathbb{P}(X\in D_j|\theta=i)$$

$$r = \int_{\mathscr{X}} \underbrace{\mathbb{E}\left[L\Big(\Theta,\delta(X)|X=x\Big)\right]}_{\text{ryzyko } a\ posteriori} f(x)dx \qquad \text{gdzie} \quad f(x) = \sum_{i=1}^{k}\pi_i f_i(x)$$

$$r_x = \sum_{i=1}^{k} L\Big(i,\delta(x)\Big)\underbrace{\mathbb{P}(\theta=i|X=x)}_{\text{rozkład } a\ posteriori} = \sum_{i=1}^{k} L\Big(i,\delta(x)\Big)\pi(i|x) \qquad \text{gdzie} \quad \pi(i|x) = \frac{f_i(x)\pi_i}{f(x)}$$

Reguła bayesowska $\delta(x)$ dla ustalonego x $\delta^*(x) = j^*$, jeśli

$$\sum_i L(i,j^*)\pi(i,x) \leqslant \sum_i L(i,j)\pi(i,x) \qquad \forall_j$$

$$\Updownarrow$$

$$\sum_i L(i,j^*)f_i(x)\pi_i \leqslant \sum_i L(i,j)f_i(x)\pi_i \qquad \forall_j$$

Przykład 27.

$L(i,j) = \begin{cases} 0\,; & j=i \\ 1\,; & j\neq i \end{cases} \qquad \mathcal{A} = \mathcal{P} = \{1,\ldots,k\}$

$r = \mathbb{P}(\theta \neq \delta(X))$ – prawdopodobieństwo błędnej klasyfikacji

Reguła bayesowska

$$\sum_i L(i,j)f_i(x)\pi_i = \sum_{i\neq j} f_i(x)\pi_i = f(x) - f_j(x)\pi_j \longrightarrow \min_j$$

$\delta^*(x) = j^*$ dla takiego j^*, że $\quad f_{j^*}(x)\pi_{j^*} \geqslant f_j(x)\pi_j \quad \forall_j$

$\pi(j^*|x) \geqslant \pi(j|x) \quad \forall_j$

Przykład 28.

$$L(i,j) = \begin{cases} 0\,; & j = i \\ \lambda\,; & j = 0 \text{– zawieszenie decyzji} \\ 1\,; & j \neq i \quad j \neq 0 \text{– błędna klasyfikacja} \end{cases} \qquad \lambda \in (0,1)$$

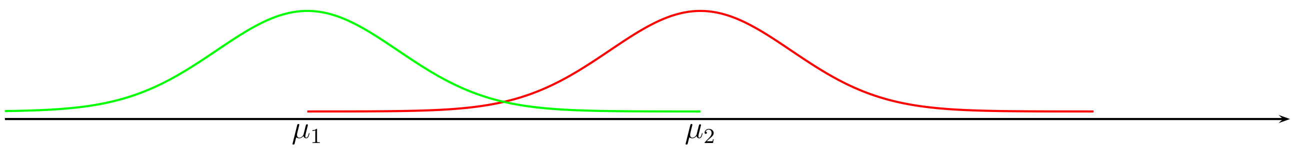
$$\textstyle\sum_i L(i,j)\pi(i|x) = \begin{cases} 1 - \pi(j|x)\,; & \text{jeśli} \quad j \neq 0 \\ \lambda\,; & \text{jeśli} \quad j = 0 \end{cases}$$

Reguła bayesowska
Niech $\pi(j^*|x) \geqslant \pi(j,x) \qquad \forall_j$

$$\delta^*(x) = \begin{cases} j^*\,; & \text{jeśli} \quad 1 - \pi(j^*|x) \leqslant \lambda \qquad \pi(j^*) \geqslant 1-\lambda \\ 0\,; & \text{jeśli} \quad 1 - \pi(j^*|x) \geqslant \lambda \qquad \pi(j^*) \leqslant 1-\lambda \end{cases}$$

Przykład 29.
Dwa rozkłady normalne, d=1

$$f_i(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\tfrac{1}{2\sigma^2}(x-\mu_i)^2\Big) \qquad i = 1,2$$



Funkcja strat

$$\begin{array}{cc} & \begin{array}{ccc} \text{a=1} & \text{a=2} & \text{a=0} \end{array} \\ \begin{array}{c} \theta = 1 \\ \theta = 2 \end{array} & \begin{bmatrix} 0 & L(1,2) & L(1,0) \\ L(2,1) & 0 & L(2,0) \end{bmatrix} \end{array}$$

$$r_x = \textstyle\sum_i L(i,j)\pi(i|x) = \begin{cases} L(2,1)\pi(2|x)\,; & j = 1 \\ L(1,2)\pi(1|x)\,; & j = 2 \\ L(1,0)\pi(1|x) + L(2,0)\pi(2|x)\,; & j = 0 \end{cases}$$

$1 \succ 0\;:$

$$L(2,1)\pi(2|x) \leqslant L(1,0)\pi(1|x) + L(2,0)$$

$$L(2,1)f_2(x)\pi_2 \leqslant L(1,0)f_1(x)\pi_1 + L(2,0)f_2(x)\pi_2$$

$$L(2,1)\frac{f_2}{f_1}\pi_2 \leqslant L(1,0)\pi_1 + L(2,0)\frac{f_2}{f_1}\pi_2$$

$$\frac{f_2}{f_1}\Big[L(2,1) - L(2,0)\Big]\pi_2 \leqslant L(1,0)\pi_1$$

$$\delta^*(x) = 1 \qquad \Leftrightarrow \qquad \frac{f_2(x)}{f_1(x)} \leqslant \frac{\pi_1 L(1,0)}{\pi_2\Big[L(2,1) - L(2,0)\Big]} = c_1$$

$$\delta^*(x) = 2 \qquad \Leftrightarrow \qquad \frac{f_2(x)}{f_1(x)} \geqslant \frac{\pi_1\Big[L(1,2) - L(1,0)\Big]}{\pi_2 L(2,0)} = c_2$$

$$\delta^*(x) = 0 \qquad \Leftrightarrow \qquad c_1 \leqslant \frac{f_2(x)}{f_1(x)} \leqslant c_2$$

14.12.2006r.

Zmienne losowe w $\mathbb{R}^{\mathbf{d}}$.

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d$

$f(x) = f(x_1, \ldots, x_d)$ – gęstość rozkładu prawdopodobieństwa

$X = (X_1, \ldots, X_d)^T = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$

$\mathbb{E}\, X = (\mathbb{E}\, X_1, \ldots, \mathbb{E}\, X_d)^T = \mu$

$$Var\ X = \Big(Cov(X_i, X_j)\Big)_{i,j=1}^d = \mathbb{E}\,(X-\mu)(X-\mu)^T$$
$$= \Big(\mathbb{E}\,(X_i - \mu_i)(X_j - \mu_j)\Big)_{i,j=1}^d$$

Twierdzenie 3.
$X \sim f_X$, $\quad h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ – dyfeomorfizm, $\quad Y \stackrel{df}{=} h(X)$, $\quad Dh^{-1} = \left(\frac{\partial h_i^{-1}(y)}{\partial y_j}\right)$
Wtedy Y ma gęstość f_Y, gdzie

$$f_Y(y) = f_X\Big(h^{-1}(y)\Big)\Big|det Dh^{-1}(y)\Big|$$

5. Wielowymiarowe rozkłady normalne.

$Z = (Z_1, \ldots, Z_d)^T$, $\qquad$ gdzie $\quad Z_i \underset{iid}{\sim} \mathcal{N}(0,1)$
$\mathbb{E}\, Z = 0$
$Var\ Z = I$
$Z \sim \mathcal{N}(0, I)$

$$f_Z(z) = \prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}} \exp\left(\tfrac{1}{2} z_i^2\right)\right) = (2\pi)^{-\frac{d}{2}} \exp\left(\frac{1}{2}\sum_{i=1}^d z_i^2\right) = (2\pi)^{-\frac{d}{2}} \exp\left(\tfrac{1}{2} z^T z\right)$$

$x = h(z) = Rz$ – przekształcenie liniowe, $\quad h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $\quad h^{-1}(x) = R^{-1}x$, $\quad \det R \neq 0$
nieosobliwe

$$f_X(x) = f_Z(z) \det R^{-1} = (2\pi)^{-\frac{d}{2}} \det R^{-1} \exp\Big(-\tfrac{1}{2} x^T (R^{-1})^T (R^{-1}) x\Big)$$

Ale

$$Var\ X = \mathbb{E}\, X X^T = \mathbb{E}\, R Z Z^T R^T = R(\mathbb{E}\, Z Z^T) R^T = R R^T = \Sigma\,, \qquad \det \Sigma = (\det R)^2$$

Stąd
$$f_X(x) = (2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp\Big(-\tfrac{1}{2}x^T\Sigma^{-1}x\Big)$$

gdzie: $\Sigma = Var\,X$, $\quad X \sim \mathcal{N}(0,\Sigma)$

$Y \sim \mathcal{N}(\mu,\Sigma)$, $\quad Y = X + \mu$, $\quad X \sim \mathcal{N}(0,\Sigma)$

$\mathbb{E}\,Y = \mu$, $\quad Var\,Y = \Sigma$

$$f_Y(y) = (2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp\Big(-\tfrac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\Big)$$

Przykład 30.
Dyskryminacja dwóch rozkładów normalnych $\pi_1,\ \pi_2 = 1-\pi_1$ gdzie π_1 – to prawdopodobieństwo, że obiekt pochodzi z pierwszej klasy

Jeśli $\ i=1$, to $\ X \sim \mathcal{N}(\mu_1,\Sigma_1)$
$\quad\quad i=2$, to $\ X \sim \mathcal{N}(\mu_2,\Sigma_2)$

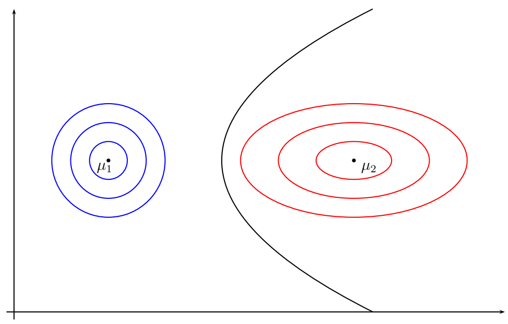
Funkcja straty $\ L_{12}\,, L_{21}$
$\quad L_{11} = L_{22} = 0$

Reguła bayesowska
$$\delta(x) = \begin{cases} 1\,; & \frac{f_2(x)}{f_1(x)} \leqslant \frac{\pi_1 L_{12}}{\pi_2 L_{21}} \qquad \log\frac{f_2(x)}{f_1(x)} \leqslant \log\frac{\pi_1 L_{12}}{\pi_2 L_{21}} \\ 2\,; & \frac{f_2(x)}{f_1(x)} > \frac{\pi_1 L_{12}}{\pi_2 L_{21}} \qquad \log\frac{f_2(x)}{f_1(x)} > \log\frac{\pi_1 L_{12}}{\pi_2 L_{21}} \end{cases}$$

$$\frac{f_2(x)}{f_1(x)} = \frac{1}{2}\Big(\log\tfrac{\det\Sigma_1}{\det\Sigma_2} + (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\Big)$$ – funkcja kwadratowa

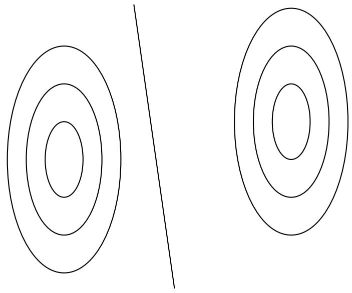
Obszary decyzyjne

$D_i = \{x;\ \delta(x) = i\}$



Przykład 31.
Dwa rozkłady normalne $\mathcal{N}(\mu_1,\Sigma)$, $\mathcal{N}(\mu_2,\Sigma)$ z tą samą macierzą kowariancji.

$$\log\tfrac{f_2(x)}{f_1(x)} = \tfrac{1}{2}\Big((x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\Big)$$



$\pi_1 = \pi_2$, $L_{12} = L_{21}$ (minimalizacja prawdopodobieństwa błędnej klasyfikacji)

$\Sigma = I$

$$\delta(x) = \begin{cases} 1\,; & \left(x - \frac{\mu_1+\mu_2}{2}\right)^T(\mu_2 - \mu_1) < 0 \quad \Leftrightarrow \quad \|x-\mu_1\|^2 < \|x-\mu_2\|^2 \\ 2\,; & \|x-\mu_1\| > \|x-\mu_2\| \end{cases}$$

Definicja 25.
Odległość Mahalanobisa: $\|x-y\|^2_\Sigma = (x-y)^T\Sigma^{-1}(x-y)$

$$\delta(x) = \begin{cases} 1\,; & \|x-\mu_1\|_\Sigma < \|x-\mu_2\|_\Sigma \\ 2\,; & \|x-\mu_1\|_\Sigma > \|x-\mu_2\|_\Sigma \end{cases}$$

Przykład 32.
Dyskryminacja k rozkładów normalnych $\mathcal{N}(\mu_1,\Sigma),\ldots,\mathcal{N}(\mu_k,\Sigma)$ ze wspólną macierzą kowariancji $L_{i,j} = \begin{cases} 0\,; & i=j \\ 1\,; & i\neq j \end{cases}$, $\pi_i = \frac{1}{k}$

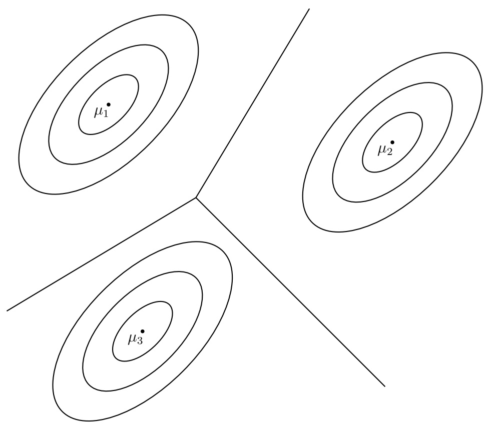
$\delta(x) = i^*$, jeśli $f_{i^*}(x) \geqslant f_i(x)$ dla wszystkich i

$$\pi(i|x) = \frac{f_i(x)}{f(x)}$$

$$f(x) = \frac{1}{k}\sum_j f_j(x)$$

$$f_i(x) = (2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp\Big(-\tfrac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)\Big)$$

$\delta(x) = i^*$ jeśli $\|x-\mu_{i^*}\|_\Sigma \leqslant \|x-\mu_i\|_\Sigma$



21.12.2006r.

Definicja 26.
Niech $(Y_1, \ldots, Y_d)$ będzie zmienną losową o wartościach w $\mathbb{R}^d$ taką, że $\sum_{i=1}^d Y_i = 1$, zaś $(Y_1, \ldots, Y_{d-1})$ mają gęstość

$$f_{\alpha_1,\ldots,\alpha_d}(y_1, \ldots, y_{d-1}) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_d)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_d)} y_1^{\alpha_1 - 1} \ldots y_{d-1}^{\alpha_{d-1}-1}(1 - y_1 - \ldots - y_{d-1})^{\alpha_d - 1}$$

$\left(\text{dla } 0 \leqslant y_i, \quad \sum_{i=1}^{d-1} y_i \leqslant 1\right)$

Mówimy, że $(Y_1, \ldots, Y_d)$ ma ROZKŁAD DIRICHLETA $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$, $\quad \alpha_i > 0$

Definicja 27.
$N_1, \ldots, N_d$ ma ROZKŁAD WIELOMIANOWY $\mathcal{M}(n, \theta_1, \ldots, \theta_d)$, jeśli

$$\mathbb{P}(N_1 = n_1, \ldots, N_d = n_d) = \begin{cases} \frac{n!}{n_1!\ldots n_d!}\theta_1^{n_1}\ldots\theta_d^{n_d} & \text{jeśli} \quad n_1 + \ldots + n_d = n \\ 0 & \text{w p. p.} \end{cases}$$

Uwaga 10.

$$(N_1, N_2) \sim \mathcal{M}(n, \theta_1, \theta_2) \equiv N_1 \sim \mathcal{B}in(n, \theta_1)$$

Uwaga 11.

$$(Y_1, Y_2) \sim \mathcal{D}(\alpha_1, \alpha_2) \equiv Y_1 \sim \mathcal{B}e(\alpha_1, \alpha_2)$$

Stwierdzenie 3.
Jeśli $(N_1, \ldots, N_d) \sim \mathcal{M}(n, \theta_1, \ldots, \theta_d)|\theta_1, \ldots, \theta_d$ warunkowo i $(\theta_1, \ldots, \theta_d)$ ma rozkład *a priori* $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$, to rozkład *a posteriori* jest $\mathcal{D}(\alpha_1 + n_1, \ldots, \alpha_d + n_d)$

Dowód.

$$\begin{aligned} \pi(\theta|n) =& \pi(\theta_1, \ldots, \theta_{d-1}|n_1, \ldots, n_d) \propto f(n|\theta)\pi(\theta) = f(n_1, \ldots, n_d|\theta_1, \ldots, \theta_{d-1})\pi(\theta_1, \ldots, \theta_{d-1}) \\ =& \theta_1^{n_1}\ldots\theta_d^{n_d}\theta_1^{\alpha_1-1}\ldots\theta_d^{\alpha_d-1} = \theta_1^{\alpha_1+n_1-1}\ldots\theta_d^{\alpha_d+n_d-1}, \quad \text{gdzie} \quad \theta_d = (1 - \theta_1 - \ldots - \theta_{d-1}) \end{aligned}$$

□

ESTYMATORY BAYESOWSKIE PARAMETRÓW $\theta_1, \ldots, \theta_d$

$\mathbb{E}\,(\theta_i) = \dfrac{\alpha_i}{\alpha_1 + \ldots + \alpha_d}$ – *a priori*

$\mathbb{E}\,(\theta_i|n_1, \ldots, n_d) = \dfrac{\alpha_i + n_i}{\alpha_1 + n_1 + \ldots + \alpha_d + n_d}$ – *a posteriori*
(α_i – pseudo-zliczenia)

NORMALNY/NORMALNY

$X_1, \ldots, X_n|\theta \underset{iid}{\sim} \mathcal{N}(\theta, \sigma^2) \qquad \theta \sim \mathcal{N}(m, a^2)$

σ^2 – znane

NORMALNY/$\mathcal{IG}$

$X_1, \ldots, X_n|\nu \underset{iid}{\sim} \mathcal{N}(\mu, \frac{1}{\nu})$ $\quad \nu \sim \mathcal{G}amma(\alpha, \lambda)$
μ – znane $\quad \sigma^2 \sim \mathcal{IG}(\alpha, \lambda)$

$\nu = \frac{1}{\sigma^2}$ – „precyzja”

a posteriori : $\sigma^2 \sim \mathcal{IG}\Big(\frac{n}{2} + \alpha, \lambda + \frac{1}{2}\sum(x_i - \mu)^2\Big)$

$$f(x_1, \ldots, x_n|\nu) = \left(\frac{\nu^{\frac{1}{2}}}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\nu}{2}\sum_{i=1}^n (x_i - \mu)^2\right\} \propto \nu^{\frac{n}{2}} \exp\left\{-\frac{\nu}{2}\sum_{i=1}^n (x_i - \mu)^2\right\}$$

gęstość *a priori*

$$\pi(\nu) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\nu^{\alpha-1}e^{-\lambda\nu} \propto \nu^{\alpha-1}e^{-\lambda\nu}$$

gęstość *a posteriori*

$$\pi(\nu|x_1, \ldots, x_n) \propto \nu^{\frac{n}{2}+\alpha-1} \exp\left\{-\nu\Big[\lambda + \tfrac{1}{2}\sum_{i=1}^n (x_i - \mu)^2\Big]\right\}$$

Definicja 28.

$$Y \sim \mathcal{IG}(\alpha, \lambda) \equiv \frac{1}{Y} \sim \mathcal{G}amma(\alpha, \lambda)$$

Różne funkcje straty w zadaniach estymacji.

$\theta \in \mathcal{P} \subseteq \mathbb{R}$

Reguła decyzyjna $\delta : \mathscr{X} \longrightarrow \mathbb{R}$, $\quad \delta(x)$ – estymator θ

1. Funkcja kwadratowa $L(\theta, a) = (\theta - a)^2$
   $\min\limits_a \mathbb{E}\,(\theta - a)^2$, $\quad a_* = \mathbb{E}\,\theta$, $\quad \mathbb{E}\,(\theta - a_*)^2 = Var(\theta)$

   Estymator bayesowski
   $\min\limits_\delta \mathbb{E}\,[(\theta - \delta(X))^2|X]$, $\quad \delta^*(X) = \mathbb{E}\,(\theta|X)$
   $\mathbb{E}\,[(\theta - \delta^*(X))^2|X] = Var(\theta|X)$ – ryzyko *a posteriori*
   Ryzyko bayesowskie $\mathbb{E}\,Var(\theta|X)$

2. Wartość bezwzględna błędu $L(\theta, a) = |\theta - a|$
   $\min\limits_a \mathbb{E}\,|\theta - a|$, $\quad a_* = med(\theta)$
   $f(a) = |a|$, $\quad f'(a) = sign(a)$

   $\dfrac{\partial}{\partial a}\mathbb{E}\,|\theta - a| = \mathbb{E}\,\dfrac{\partial}{\partial a}|\theta - a| = \mathbb{E}\,\dfrac{\partial}{\partial a}|a - \theta| = \mathbb{E}\,sign(a - \theta) = -\mathbb{P}(\theta > a) + \mathbb{P}(\theta < a) = 0$

   $\mathbb{P}(\theta > a) = \mathbb{P}(\theta < a)$

Lemat 1.

$$\mathbb{E}\,(\theta - a)_+ = \int_a^\infty \mathbb{P}(\theta > x)dx$$

Lemat 2.

$$\mathbb{E}\,|\theta - a| = \mathbb{E}\,(\theta - a)_+ + \mathbb{E}\,(a-\theta)_+ = \int_a^\infty \mathbb{P}(\theta > x)dx + \int_{-\infty}^a \mathbb{P}(\theta - x)dx$$

$$Q(a) = \int_a^\infty [1 - F(x)]dx + \int_{-\infty}^a F(x)dx$$

Jeżeli F jest ciągła w a, to

$$\frac{\partial}{\partial a} Q(a) = F(a) - [1 - F(a)] = 2F(a) - 1$$

Jeżeli $F(a-) < F(a)$, to

$$\frac{\partial}{\partial a-} Q(a) = 2F(a-) - 1$$
$$\frac{\partial}{\partial a+} Q(a) = 2F(a) - 1$$

$$\frac{\partial}{\partial a-} Q(a) \leqslant 0 \leqslant \frac{\partial}{\partial a} Q(a)$$

04.01.2007r.

Testowanie hipotez statystycznych.

Przypadek dwóch hipotez prostych.

$\mathscr{X}$ – przestrzeń obserwacji
$\mathcal{P} = \{0, 1\}$ – przestrzeń parametrów

$H_0 : \quad \theta = 0 \qquad f(x|0) = f_0(x) \qquad X \sim f_0$

$H_1 : \quad \theta = 1 \qquad f(x|1) = f_1(x) \qquad X \sim f_1$

δ – reguła decyzyjna (test), $\qquad \delta : \mathscr{X} \longrightarrow \mathcal{A} = \{0, 1\}$
0 – nie mamy podstaw do odrzucenia H_0
1 – odrzucamy H_0 na rzecz H_1

Funkcja straty $\qquad L(0, 1) = L(1, 0) = 1$
$L(0, 0) = L(1, 1) = 0$

Rozkład *a priori* $\pi_0, \pi_1 \qquad \pi_0 + \pi_1 = 1, \quad \pi_i = \mathbb{P}(\Theta = i) \quad$ dla $i = 1, 2$

Błąd I rodzaju $\qquad \mathbb{P}(\delta(X) = 1|\Theta = 0) = \mathbb{P}_0(\delta(X) = 1) = \alpha(\delta) = \mathbb{E}\Big(L\big(0, \delta(X)\big)|\Theta = 0\Big)$

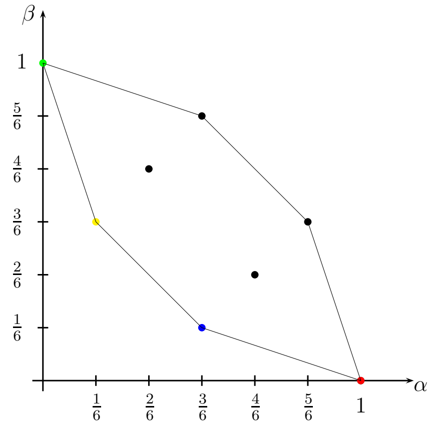
II rodzaju $\qquad \mathbb{P}(\delta(X) = 0|\Theta = 1) = \mathbb{P}_1(\delta(X) = 0) = \beta(\delta) = \mathbb{E}\Big(L\big(1, \delta(X)\big)|\Theta = 1\Big)$

Zbiór ryzyka

$$\mathcal{R} = \Big\{\big(\alpha(\delta), \beta(\delta)\big) : \delta \text{ jest testem}\Big\}$$

Przykład 33.
$\mathscr{X} = \{1, 2, 3\}$

x	1	2	3
f_1	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$
f_0	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
$\frac{f_1}{f_0}$	$\frac{1}{3}$	1	3



Reguły bayesowskie przy wszystkich możliwych rozkładach *a priori*

$$\delta(x) = \begin{cases} 1\,, & \text{jeśli } \frac{f_1(x)}{f_0(x)} > \frac{\pi_0}{\pi_1} = c \\ 1 \text{ lub } 0 & \text{jeśli } \frac{f_1(x)}{f_0(x)} = \frac{\pi_0}{\pi_1} \\ 0\,, & \text{jeśli } \frac{f_1(x)}{f_0(x)} < \frac{\pi_0}{\pi_1} = c \end{cases} \qquad \text{test ilorazu wiarogodności}$$

$\delta' = 1 - \delta$

$\alpha(\delta') = 1 - \alpha(\delta)$

$\beta(\delta') = 1 - \beta(\delta)$

Definicja 29.
δ^* jest najmocniejszym testem na poziomie istotności α^*, jeśli

1. $\alpha(\delta^*) \leqslant \alpha^*$
2. Jeśli $\alpha(\delta^*) \leqslant \alpha^*$, to $\beta(\delta^*) \leqslant \beta(\delta)$

Lemat 3 (Neymana – Pearsona)**.** Jeżeli δ^* jest najmocniejszym testem na poziomie istotności $\alpha^* \in (0, 1)$ i $\alpha(\delta^*) = \alpha^*$, to istnieje $c \in [0, \infty)$ takie, że

$$\delta^*(x) = \begin{cases} 1\,, & \text{jeśli } \frac{f_1(x)}{f_0(x)} > c \\ 0\,, & \text{jeśli } \frac{f_1(x)}{f_0(x)} < c \end{cases} \qquad (*)$$

Jeżeli δ^* spełnia $(*)$ i $\alpha(\delta^*) = \alpha^*$, to δ^* jest najmocniejszy na poziomie istotności α^*.

Przykład 34.

$\mathscr{X} = [0,1] \qquad f_0(x) = 1 \qquad f_1(x) = 2x \; \dfrac{f_1(x)}{f_0(x)} = 2x$

$$\alpha(\delta) = \mathbb{P}_0(2X > c) = \mathbb{P}_0(X > \tfrac{c}{a}) = 1 - \frac{c}{2}$$

$$\beta(\delta) = \mathbb{P}_1(2X \leqslant c) = \mathbb{P}_1(X \leqslant \tfrac{c}{2}) = \left(\frac{c}{2}\right)^2$$

$\beta(\alpha) = (1-\alpha)^2$

Najmocniejszy test na poziomie istotności $\alpha^* = 0.005$

$\alpha = 0.005$
$\beta = (1-\alpha)^2 = (0.95)^2$

Test zrandomizowany $\qquad \delta : \mathscr{X} \longrightarrow [0,1]$

$\delta(x)$ – prawdopodobieństwo podjęcia decyzji 1, (odrzucenie hipotezy H_0)

Zbiór ryzyka dla reguł zrandomizowanych jest wypukły.

$\delta_1 \qquad \big(\alpha(\delta_1), \beta(\delta_1)\big)$

$\delta_2 \qquad \big(\alpha(\delta_2), \beta(\delta_2)\big)$

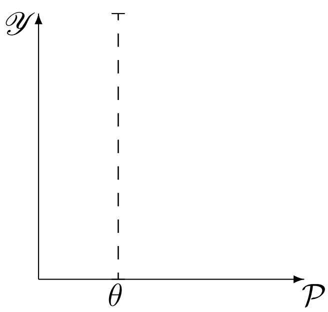
$\delta = \lambda\delta_1 + (1-\lambda)\delta_2$

$$\alpha(\delta) = \int_{\mathscr{X}} \delta(x) f_0(x) dx$$

$$\beta(\delta) = \int_{\mathscr{X}} \big(1 - \delta(x)\big) f_1(x) dx$$

Przykład (33 c. d.**).** Najmocniejszy test niezrandomizowany na poziomie istotności $\alpha^* = 0.05$ jest $\delta(x) \equiv 0$ (nigdy nie odrzucamy).
Najmocniejszy test zrandomizowany na poziomie istotności $\alpha^* = 0.05$ jest $\delta(x) \equiv 0$.

$$\delta(x) = \begin{cases} 0\,, & \text{jeśli } x = 1 \text{ lub } x = 2 \\ \frac{\frac{1}{20}}{\frac{1}{6}} = \frac{6}{20} = \frac{3}{10}\,, & \text{jeśli } x = 3 \end{cases}$$



11.01.2007r.

Krzywa $(\alpha - \beta)$ Neymana – Pearsona.

$H_0: \quad \theta = 0 \qquad f_0$

$$h = \tfrac{f_1}{f_0}$$

$H_1: \quad \theta = 1 \qquad f_1$

Testy ilorazu wiarogodności:

$$\delta(x) = \begin{cases} 1\,; & \text{jeśli} \quad h(x) > c \\ \varrho\,; & \text{jeśli} \quad h(x) = c \\ 0\,; & \text{jeśli} \quad h(x) < c \end{cases}$$

dokładniej, jeśli $h(x) = c$, to bierzemy

$$\delta(x) = \begin{cases} 1\,; & \text{z prawdopodobieństwem } \varrho \\ 0\,; & \text{z prawdopodobieństwem } 1-\varrho \end{cases}$$

$$\alpha(c,\varrho) = \int_{\{h>c\}} f_0 + \varrho \int_{\{h=c\}} f_0$$

$$\beta(c,\varrho) = \int_{\{h<c\}} f_1 + (1-\varrho) \int_{\{h=c\}} f_1$$

$$F_0(c) = \mathbb{P}_0(h \leqslant c) = \int_{\{h\leqslant c\}} f_0$$

$$F_1(c) = \mathbb{P}_1(h \leqslant c) = \int_{\{h\leqslant c\}} f_1$$

$$\alpha(c,\varrho) = 1 - F_0(c) + \varrho[F_0(c) - F_0(c-)] = \mathbb{E}_0 \frac{f_1}{f_0}\, \mathbb{1}(h \leqslant c)$$

$$\beta(c,\varrho) = F_1(c-) + (1-\varrho)[F_1(c) - F_1(c-)] = \mathbb{E}_0 h\, \mathbb{1}(h \leqslant c)$$

Testowanie hipotez złożonych w ujęciu bayesowskim.

$\mathcal{P}$ – przestrzeń parametrów

$\mathscr{X}$ – przestrzeń obserwacji

$f(x|\theta)$ – wiarogodność

$\pi(\theta)$ – *a priori*

$\pi(\theta|x)$ – *a posteriori*

$H_0: \quad \theta \in \mathcal{P}_0 \subset \mathcal{P} \qquad \mathcal{P}_0 \cap \mathcal{P}_1 = \varnothing$

$H_1: \quad \theta \in \mathcal{P}_1 \subset \mathcal{P} \qquad \mathcal{P}_1 = \mathcal{P} \setminus \mathcal{P}_0$

$$\mathbb{P}(H_0|X=x) = \mathbb{P}(\theta \in \mathcal{P}_0|X=x) = \int_{\mathcal{P}_0} \pi(\theta|x)d\theta$$

$$\mathbb{P}(H_1|X=x) = \mathbb{P}(\theta \in \mathcal{P}_1|X=x) = \int_{\mathcal{P}_1} \pi(\theta|x)d\theta$$

Przykład 35.

$X_1, \ldots, X_n \underset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ $\qquad$ $H_0: \quad \theta \leqslant \theta_0$

$H_1: \quad \theta > \theta_0$

a priori $\qquad \theta \sim \mathcal{N}(m, a^2)$

a posteriori $\qquad \theta \sim \mathcal{N}\Big(z\overline{x} + (1-z)m, \frac{a^2\sigma^2}{na^2+\sigma^2}\Big) \qquad z = \frac{na^2}{na^2+\sigma^2}$

$$\begin{aligned}
\mathbb{P}(H_0|x_1, \ldots, x_n) &= \int_{-\infty}^{\theta_0} \pi(\theta|x_1, \ldots, x_n)d\theta = \mathbb{P}(\theta \leqslant \theta_0|x_1, \ldots, x_n) \\
&= \mathbb{P}\left(\frac{\theta - 2\overline{x} - (1-z)m}{a\sigma}\sqrt{na^2+\sigma^2} \leqslant \frac{\theta_0 - 2\overline{x} - (1-z)m}{a\sigma}\sqrt{na^2+\sigma^2}\right) \\
&= \Phi\left(\frac{\theta_0 - 2\overline{x} - (1-z)m}{a\sigma}\sqrt{na^2+\sigma^2}\right)
\end{aligned}$$

$H_0: \quad \theta = \theta_0$

$H_1: \quad \theta \neq \theta_0$

Rozkład *a priori*

$\mathbb{P}(\theta = \theta_0) = c, \qquad c \in (0, 1)$

$$\begin{tikzcd}
& {} \arrow[dl, "c"'] \arrow[dr, "1-c"] & \\
\theta = \theta_0 & & \theta \sim \mathcal{N}(m, a^2)
\end{tikzcd}$$

Zbiory (przedziały ufności) w ujęciu bayesowskim.

$\mathcal{P} \qquad \mathscr{X}$

$\pi \qquad f$

Definicja 30.
Przyporządkowanie $x \longmapsto C(x) \subset \mathcal{P}$

$$\mathbb{P}\Big(\theta \in C(x)|X\Big) = 1 - \alpha$$

nazywamy PRZEDZIAŁEM UFNOŚCI na poziomie $1 - \alpha$

Uwaga 12.
W statystyce klasycznej

$$\mathbb{P}\Big(\theta \in C(x)|\theta\Big) = 1 - \alpha$$

Stwierdzenie 4.
Załóżmy, że π jest gęstością względem miary Lebesque'a.
$\pi(\cdot|x)$ też jest gęstością względem miary Lebesque'a.

Jeżeli istnieje takie $c \in (0, \infty)$, że

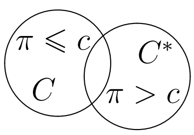
$$\mathbb{P}(\pi(\theta|x) > c|x) = \int_{\{\theta;\ \pi(\theta|x)>c\}} \pi(\theta|x)d\theta = 1 - \alpha$$

to $C^*(x) = \{\theta;\ \pi(\theta|x) > c\}$ jest najmniejszym w sensie miary Lebesgue'a przedziałem ufności na poziomie $1 - \alpha$

Dowód. $x \in \mathscr{X}$ – ustalone
Jeżeli $\int_{C(x)} \pi(x|\theta)d\theta = 1 - \alpha$, to $|C(x)| \geqslant |C^*(x)|$

$C(x) = C$
$C^*(x) = C^*$
$\pi(\theta|x) = \pi(\theta)$
$\int_C \pi = 1 - \alpha = \int_C^* \pi$ na C^* mamy $\pi > c$
poza C^* mamy $\pi \leqslant c$

$$c\,|C \setminus C^*| \geqslant \int_{C\setminus C^*} \pi = \int_{C^*\setminus C} \pi > \int_{C^*\setminus C} c = c\,|C^* \setminus C|$$



$$|C \setminus C^*| > |C^* \setminus C|$$

$$|C| > |C^*| \qquad \text{jeśli} \quad |C \setminus C^*| > 0$$

□

Predykcja liniowa i liniowe modele bayesowskie.

θ – jednowymiarowa zmienna losowa, X

Znaleźć funkcję $\delta : \mathscr{X} \longrightarrow \mathbb{R}$ taką, żeby $\mathbb{E}\left(\delta(X) - \theta\right)^2$ była minimalna

$$\delta^*(X) = \mathbb{E}\,(\theta|X) \quad - \quad \text{rozwiązanie}$$

Twierdzenie 4.
Jeżeli $\theta, X_1, \ldots, X_n$ są zmiennymi losowymi $X = (X_1, \ldots, X_n)^T$, to funkcja liniowa

$$\delta(x) = \delta(X_1, \ldots, X_n) = c_0 + \sum_{i=1}^{n} c_i X_i = c_0 + c^T X$$

minimalizująca $\mathbb{E}\left(\theta - \delta(X)\right)^2$ jest dana wzorami

$$c_* = Var(X)^{-1}\, Cov(X, \theta)$$
$$c_0 = \mathbb{E}\ \theta - c_*^T\ \mathbb{E}\,X$$

gdzie

$$
\begin{aligned}
Var(X) &= \Big(Cov(X_i, X_j);\, i, j = 1, \dots, n\Big) \\
Cov(X, \theta) &= \Big(Cov(X_i, \theta);\, i = 1, \dots, n\Big)
\end{aligned}
$$

18.01.2007r.

Empiryczne metody bayesowskie i hierarchiczne modele bayesowskie.

Przykład 36 (NORMALNY/NORMALNY).
$X_1, \dots, X_n \underset{iid}{\sim} \mathcal{N}(\theta, s^2), \qquad \theta \sim \mathcal{N}(m, a^2)$

a posteriori $\quad \theta \sim \mathcal{N}\Big(z\overline{X} + (1-z)m, \frac{a^2 s^2}{na^2+s^2}\Big)$, gdzie: $z = \frac{na^2}{na^2+s^2}$

$\widehat{\theta}_B = z\overline{X} + (1-z)m$

INTERPRETACJA UBEZPIECZENIOWA.

$\theta; X_1, \dots, X_n$ – sumaryczne szkody zgłoszone przez <u>klienta</u> w poszczególnych latach $1, 2, \dots, n$

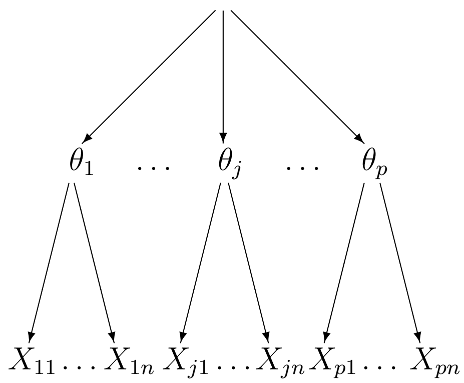
$\theta = \mathbb{E}\,(X_i|\theta)$ – średnia wartość szkód dla naszego klienta

$s^2 = Var(X_i|\theta)$ – zmienność pomiędzy latami

m – średnia szkoda w całej populacji klientów

a^2 – zmienność w populacji klientów

MODEL UWZGLĘDNIAJĄCY PRÓBKĘ Z POPULACJI KLIENTÓW



p – klientów

X_{ji} – szkody j-tego klienta w i-tym roku, $j = 1, \dots, p \quad i = 1, \dots, n$

$X_{ji} \sim \mathcal{N}(\theta_j, s^2), \qquad \theta_1, \dots, \theta_p \underset{iid}{\sim} \mathcal{N}(m, a^2)$

Łączny rozkład prawdopodobieństwa

$$
f\Big((\theta_j), (x_{ji})\Big) = \prod_{j=1}^{p} \pi(\theta_j) \prod_{i=1}^{n} f(x_{ji}|\theta_j)
$$

EMPIRYCZNE PODEJŚCIE BAYESOWSKIE

Estymujemy rozkład *a priori* parametru θ na podstawie próbek odpowiadających różnym „realizacjom” zmiennej losowej θ.

$$\begin{array}{lcllll} \theta_1 & ; & X_{11} & \dots & X_{1n} & \\ \vdots & & & & & \\ \theta_j & ; & X_{j1} & \dots & X_{jn} & \longleftarrow \text{Dane} \\ \vdots & & & & & \\ \theta_p & ; & X_{p1} & \dots & X_{pn} & \end{array}$$

1. Na podstawie całej tablicy estymujemy parametry rozkładu *a priori* m, a^2 (i parametr s^2).

2. W rozwiązaniach bayesowskich wstawiamy wyestymowany rozkład *a priori* .

Ad. 1.

$$\left.\begin{aligned} \widehat{m} &= \overline{X} = \frac{1}{np}\sum_{j=1}^{p}\sum_{i=1}^{n} X_{ji} = \frac{1}{p}\sum_{j=1}^{p}\overline{X}_j\,; \qquad \text{gdzie} \quad \overline{X}_j = \frac{1}{n}\sum X_{ji} \\ \widehat{s}^2 &= \frac{1}{p}\sum_{j=1}^{p}\frac{1}{n-1}\sum_{i=1}^{n}(X_{ji}-\overline{X}_j)^2\,; \qquad \text{gdzie} \quad \overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ji} \\ \widehat{a}^2 &= \frac{1}{p-1}\sum_{j=1}^{p}(\overline{X}_j-\overline{X})^2 - \frac{1}{n}\widehat{s}^2 \end{aligned}\right\}\text{estymatory nieobciążone}$$

Ad. 2. Estymator bayesowski parametru θ_j

$$\widehat{\theta}_j^B = z\overline{X}_j + (1-z)m$$

Estymator empiryczny bayesowski

$$\widehat{\theta}_j^{EB} = \widehat{z}\overline{X}_j + (1-\widehat{z})\widehat{m} \qquad \text{gdzie} \quad \widehat{z} = \frac{n\widehat{a}^2}{n\widehat{a}^2 + \widehat{s}^2}$$

HIERARCHICZNY MODEL BAYESOWSKI.

$X_{ji} \underset{iid}{\sim} \mathcal{N}(\theta_j, s^2)$ $\qquad$ $\theta_j \underset{iid}{\sim} \mathcal{N}(m, a^2)$

$s^2 \sim \mathcal{IG}(p, c)$ $\qquad$ $m \sim \mathcal{N}(\nu, b^2)$

$a^2 \sim \mathcal{IG}(q, d)$

$\tau = \frac{1}{s^2} \sim \mathcal{G}amma(p, c)$

$\zeta = \frac{1}{a^2} \sim \mathcal{G}amma(q, d)$

Znane: ν, b^2, p, c, q, d

Parametry losowe: $s^2, a^2, m, \theta_1, \dots, \theta_q$

Łączny rozkład

$$f\Big((x_{ji}), (\theta_j), m, \tau, \zeta\Big) = h(\zeta)g(\tau)\varphi(m)\prod_j \pi(\theta_j|m,\zeta)\prod_i f(x_{ji}|\theta_j,\tau)$$

Cel: estymacja bayesowska θ_j

Próbnik Gibsa.
Rozważmy zmienną losową $\theta = (\theta_1, \dots, \theta_k)$ o wartośiach w przestrzeni $\mathscr{X}^k$ i gęstości $\pi(\theta_1, \dots, \theta_k)$

Umiemy <u>losować</u> zmienne losowe jednowymiarowe z gęstości warunkowych

$$\pi(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$$

Generujemy ciąg wektorów losowych $\theta(0), \theta(1), \dots, \theta(n), \dots$

Jeżeli $\theta(n) = (\theta_1, \dots, \theta_k) \stackrel{ozn}{=} \theta$, to $\theta(n+1) = (\theta'_1, \dots, \theta'_k) = \theta'$ losujemy w k krokach

1. $\theta'_1 \sim \pi(\cdot | \theta_2, \dots, \theta_k)$
2. $\theta'_2 \sim \pi(\cdot | \theta'_1, \theta_3, \dots, \theta_k)$

$\vdots$

k-1. $\theta'_{k-1} \sim \pi(\cdot | \theta'_1, \dots, \theta'_{k-2}, \theta_k)$

k. $\theta'_k \sim \pi(\cdot | \theta'_1, \dots, \theta'_{k-1})$

Przy pewnych założeniach na rozkład π

$$\theta(n) \xrightarrow[n \to \infty]{} \pi$$

$$\frac{1}{n} \sum_{t=1}^{n} h\Big(\theta(t)\Big) \xrightarrow[n \to \infty]{} \int_{\mathscr{X}^n} h(\theta) \pi(\theta) d\theta = \mathbb{E}_{\pi} h(\theta)$$