

Instytut Matematyki
Politechniki Wrocławskiej

Alicja Jokiel-Rokita

Rozprawa doktorska

**BAYESOWSKIE
I MINIMAKSOWE
PROCEDURY ESTYMACJI
I PREDYKCJI**

Promotor – prof. dr hab. Ryszard Magiera

Wrocław 1998

Spis treści

Wstęp	3
1 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego przy ustalonym rozmiarze próby	8
1.1 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego	9
1.2 Minimaksowa estymacja i predykcja dla wielowymiarowego rozkładu hipergeometrycznego	18
1.3 Przykłady	23
2 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego przy losowym rozmiarze próby	26
2.1 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego	27
2.2 Minimaksowa estymacja i predykcja dla wielowymiarowego rozkładu hipergeometrycznego	32
2.3 Przykłady	36
3 Estymacja parametru rozkładu wielomianowego na podstawie danych otrzymanych w chwilach losowych	40
3.1 Optymalne procedury sekwencyjne	42
3.1.1 Bayesowska procedura sekwencyjna	42
3.1.2 Minimaksowa procedura sekwencyjna	45
3.1.3 Bayesowska procedura sekwencyjna w przypadku, gdy F nie jest znane	46
3.2 Procedura adaptacyjna	47
3.3 Uwagi dodatkowe	50
4 Bayesowska predykcja mediany próbkowej z przesuniętego rozkładu wykładniczego przy ustalonym i losowym rozmiarze przyszłej próby	53
4.1 Predykcja mediany w przypadku ustalonego rozmiaru przyszłej próby . . .	54

4.2	Predykcja mediany w przypadku losowego rozmiaru przyszłej próby	57
4.3	Wnioski wynikające z obliczeń numerycznych	57

Literatura		61
-------------------	--	-----------

Wstęp

Praca poświęcona jest zagadnieniom bayesowskiej i minimaksowej estymacji i predykcji w sytuacjach, gdy rozmiar próby jest wielkością ustaloną z góry bądź realizacją pewnej zmiennej losowej.

Rozwój teorii estymacji minimaksowej został zapoczątkowany przez Walda (1950), Hogesa i Lehmana (1950, 1951). Narzędzia tej teorii zaczęto później wykorzystywać przy poszukiwaniu optymalnych predyktorów (Trybuła (1985), Wilczyński (1985), Trybuła (1987a, b)).

Większa część pracy (Rozdziały 1, 2, 3) poświęcona jest wyznaczaniu bayesowskich i minimaksowych estymatorów i predyktorów dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego. Rozdział 4 dotyczy bayesowskiej predykcji mediany próbkowej z przesuniętego rozkładu wykładniczego. Rozpatrywane modele odgrywają ważną rolę w teorii niezawodności i w statystycznej kontroli jakości.

Rozdziały 1 i 2 dotyczą klasy problemów, w których na podstawie obserwacji zmiennej losowej $\mathbf{X}$ należy podać prognozowaną wartość zmiennej losowej $\mathbf{Z}$, przy czym $\mathbf{X}$ i $\mathbf{Z}$ mają rozkłady zależne od tego samego parametru.

Niech $\mathbf{X}$ i $\mathbf{Y}$ oznaczają niezależne zmienne losowe o rozkładach wielomianowych, odpowiednio $\mathcal{W}(n, \mathbf{p})$ i $\mathcal{W}(m, \mathbf{p})$, o funkcji prawdopodobieństwa określonej wzorem (1.3), gdzie n i m są znane, $\mathbf{p} = (p_1, \dots, p_r)^T$ jest nieznanym parametrem. Na podstawie obserwacji zmiennej losowej $\mathbf{X}$ należy wyznaczyć postać minimaksowego predyktora kombinacji liniowej

$$\mathbf{Z} = a\mathbf{p} + b\mathbf{Y}, \quad (0.1)$$

gdzie a i b są liczbami rzeczywistymi. W Rozdziale 1 podano rozwiązanie tego problemu przy funkcji straty postaci

$$L(\mathbf{Z}, \mathbf{d}) = (\mathbf{d} - \mathbf{Z})^T C (\mathbf{d} - \mathbf{Z}) = \sum_{i,j=1}^r c_{ij} (d_i(\mathbf{X}) - Z_i)(d_j(\mathbf{X}) - Z_j), \quad (0.2)$$

gdzie $C = (c_{ij})$ jest macierzą nieujemnie określoną i $Z_i = ap_i + bY_i$, $i = 1, \dots, r$. W Rozdziale 1 rozpatruje się również model, w którym zmienne losowe $\mathbf{X}$ i $\mathbf{Y}$ są niezależne

o rozkładzie wielowymiarowym hipergeometrycznym z parametrami odpowiednio W , $\mathbf{W} = (W_1, \dots, W_r)^T$, n i W , $\mathbf{W} = (W_1, \dots, W_r)^T$, m o funkcji prawdopodobieństwa określonej wzorem (1.26), gdzie W , n i m są znane, $\mathbf{p} = (W_1/W, \dots, W_r/W)^T$ jest nieznanym parametrem. Podano również dla tego modelu rozwiązanie postawionego problemu predykcji przy funkcji straty (0.2) i stałym rozmiarze n obserwacji.

Zauważmy, że dla $b = 0$ rozważany problem predykcji staje się zagadnieniem estymacji.

Problemy rozpatrywane w Rozdziale 1 mają swoją genezę w pracach Hodgesa i Lehmana (1950) oraz Steinhausa (1957). Mianowicie, Hodges i Lehmann (1950) wyznaczyli minimaksowy estymator parametru rozkładu dwumianowego i hipergeometrycznego przy kwadratowej funkcji straty. Ich rezultaty zostały uogólnione przez Steinhausa (1957) w przypadku minimaksowej estymacji parametru $\mathbf{p}$ rozkładu wielomianowego przy funkcji straty postaci

$$L_1(\mathbf{p}, \mathbf{d}) = \sum_{i=1}^r (p_i - d_i)^2, \quad (0.3)$$

a następnie przez Trybułę (1958) w przypadku minimaksowej estymacji parametru $\mathbf{p}$ rozkładu wielomianowego i wielowymiarowego hipergeometrycznego przy funkcji straty postaci

$$L_2(\mathbf{p}, \mathbf{d}) = \sum_{i=1}^r c_i (p_i - d_i)^2. \quad (0.4)$$

Wilczyński (1985) wyprowadził minimaksowe estymatory parametru $\mathbf{p}$ rozkładu wielomianowego i wielowymiarowego hipergeometrycznego oraz minimaksowy predyktor zmiennej losowej o rozkładzie wielomianowym przy funkcji straty określonej wzorem (0.2). Wyniki uzyskane przez Hodgesa i Lehmana (1950), Steinhausa (1957), Trybułę (1958) oraz Wilczyńskiego (1985) stanowią szczególne przypadki rezultatów otrzymanych w Rozdziale 1. Na przykład, uzyskany przez Steinhausa (1957) przy funkcji straty $L_1(\mathbf{p}, \mathbf{d})$ określonej wzorem (0.3) minimaksowy estymator $\mathbf{d}_0(\mathbf{X}) = (d_1^0(\mathbf{X}), \dots, d_r^0(\mathbf{X}))^T$ parametru $\mathbf{p}$ rozkładu wielomianowego określony przez

$$d_i^0(\mathbf{X}) = \frac{X_i + \frac{\sqrt{n}}{r}}{n + \sqrt{n}},$$

$i = 1, \dots, r$, stanowi szczególny przypadek minimaksowego predyktora (1.18) kombinacji liniowej $\mathbf{Z}$, określonej wzorem (0.1), przy funkcji straty postaci (0.2), gdy $a = 1$, $b = 0$ i macierz C jest jednostkowa. Natomiast uzyskany przez Trybułę (1958) przy funkcji straty określonej wzorem (0.4) minimaksowy estymator $\mathbf{d}_0(\mathbf{X}) = (d_1^0(\mathbf{X}), \dots, d_r^0(\mathbf{X}))^T$ parametru

$\mathbf{W} = (W_1, \dots, W_r)^T$ rozkładu wielowymiarowego hipergeometrycznego określony przez

$$d_i(\mathbf{X}) = W \frac{X_i + s_i \sqrt{n \frac{W-n}{W-1}}}{n + \sqrt{n \frac{W-n}{W-1}}},$$

$i = 1, \dots, r$, gdzie $s_i, i = 1, \dots, r$, są stałymi zależnymi od $c_i, i = 1, \dots, r$, (w przypadku, gdy $c_i = 1, i = 1, \dots, r$, stałe $s_i = 1/r, i = 1, \dots, r$) stanowi szczególny przypadek minimaxowego predyktora (1.41) kombinacji liniowej $\mathbf{Z}$ przy funkcji straty postaci (0.2), gdy $a = W, b = 0$ i macierz C jest diagonalna.

W Rozdziale 2 rozważa się wyżej postawiony problem wyznaczania optymalnych predyktorów przyjmując, że rozmiar n próby $\mathbf{X}$ jest wielkością losową.

Rozpatrywanie takiego przypadku jest zasadne i ważne z praktycznego punktu widzenia, gdyż w wielu problemach nie jest możliwe przyjęcie ustalonego rozmiaru próby. Przykładem mogą być następujące dwie sytuacje: interesujące nas jednostki stanowią podgrupę w bazie danych, którą dysponujemy; wylosowana do próby jednostka odmówi nam udzielenia odpowiedzi lub będzie nieobecna w chwili przeprowadzania badania. Inne przykłady, które uzasadniają rozpatrywanie losowego rozmiaru próby można znaleźć w pracy Chaudhuri'ego i Stengera (1992).

W Rozdziale 2 wyprowadzono minimaxowe predyktory kombinacji liniowej $\mathbf{Z}$ określonej wzorem (0.1), przy funkcji straty postaci (0.2), w przypadku, gdy $\mathbf{X}$ i $\mathbf{Y}$ mają rozkład wielomianowy i wielowymiarowy hipergeometryczny, przy założeniu, że rozmiar n próby $\mathbf{X}$ jest realizacją zmiennej losowej o znanym rozkładzie niezależnym od nieznanego parametru $\mathbf{p}$, czyli jest statystyką dodatkową (Lehmann, (1986)).

W Rozdziale 2 udowodniono, że minimaxowy predyktor kombinacji liniowej $\mathbf{Z}$ przy funkcji straty określonej wzorem (0.2), w przypadku ustalonego rozmiaru próby, podany w Twierdzeniu 1.3 i w Twierdzeniu 1.4, nie jest minimaxowy, a nawet dopuszczalny, dla szerokiej klasy wartości współczynników a i b , parametrów n, m, W oraz rozkładów rozmiaru próby, w przypadku, gdy rozmiar próby jest statystyką dodatkową. Uzyskane w tym rozdziale wyniki są przykładem paradoksu, a mianowicie wpływu statystyki dodatkowej, którą jest tutaj rozmiar próby, na postać estymatora, czy predyktora. Po raz pierwszy podobny przykład podał Brown (1990), w którym dopuszczalność estymatora stałego czynnika w wielowymiarowej liniowej regresji zależy od rozkładu macierzy eksperymentu, która jest statystyką dodatkową. Następnie He (1990) pokazał, że w przypadku kwadratowej funkcji straty minimaxowy estymator parametru rozkładu wielomianowego przy ustalonym rozmiarze próby podany przez Steinhaus (1957) przestaje być estymatorem minimaxowym, a nawet dopuszczalnym, poza pewnymi trywialnymi przypadkami,

gdy rozmiar próby jest losowy. Amrhein (1995) wyprowadził postać minimaxowego estymatora parametru rozkładu wielowymiarowego hipergeometrycznego przy funkcji straty określonej wzorem (0.3) w przypadku, gdy rozmiar próby jest statystyką dodatkową i podobnie jak dla rozkładu wielomianowego różni się on od uzyskanego przez Trybułę (1958) estymatora w przypadku, gdy rozmiar próby jest ustalony. Uzyskane w Rozdziale 2 rezultaty stanowią uogólnienie wyników He (1990) i Amrheina (1995).

Otrzymane rezultaty dotyczące predykcji kombinacji liniowej $\mathbf{Z}$ określonej wzorem (1.1) zostały zilustrowane przykładami w Podrozdziałach 1.3 i 2.3.

Rozdział 3 poświęcony jest estymacji parametru $\mathbf{p} \in \mathcal{P} = \{(p_1, \dots, p_r)^T : p_i \geq 0, i = 1, \dots, r, p_1 + \dots + p_r = 1\}$ rozkładu wielomianowego o funkcji prawdopodobieństwa określonej wzorem (1.3) w przypadku, gdy dane otrzymywane są w chwilach losowych.

Rozpatruje się następujący model. Niech $\mathbf{Y}(i) = (Y_1(i), \dots, Y_r(i))^T, i = 1, \dots, n$, będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie wielomianowym $\mathcal{W}(1, \mathbf{p})$, to znaczy $P(\mathbf{Y}(i) = \mathbf{e}_j) = p_j, j = 1, \dots, r, i = 1, \dots, n$, gdzie $\mathbf{e}_1 = (1, 0, \dots, 0)^T, \dots, \mathbf{e}_r = (0, \dots, 0, 1)^T$. Zakłada się, że $\mathbf{Y}(i)$ jest obserwacją uzyskaną w chwili $t_i, i = 1, \dots, n$, gdzie $t_1, \dots, t_n$ są statystykami pozycyjnymi dodatnich zmiennych losowych $U_1, \dots, U_n$, które są niezależne od $\mathbf{Y}(1), \dots, \mathbf{Y}(n)$. Niech

$$k(t) = \sum_{i=1}^n I_{[0,t]}(U_i)$$

oznacza liczbę obserwacji uzyskanych do chwili $t \geq 0$, a $\mathcal{F}_t = \sigma\{k(s), s \leq t, \mathbf{Y}(1), \dots, \mathbf{Y}(k(t))\}$ będzie informacją dostępną w chwili t . Przyjmuje się, że strata, którą ponosimy, gdy estymujemy nieznany parametr $\mathbf{p}$ na podstawie obserwacji uzyskanych do chwili t jest postaci

$$L_t(\mathbf{p}, \mathbf{d}) = L(\mathbf{p}, \mathbf{d}) + c(t),$$

gdzie funkcja $L(\mathbf{p}, \mathbf{d})$ postaci

$$L(\mathbf{p}, \mathbf{d}) = \sum_{i=1}^r \frac{(d_i - p_i)^2}{p_i} \quad (0.5)$$

jest stratą związaną z błędem estymacji, gdy podejmiemy decyzję $\mathbf{d}$, a $\mathbf{p}$ jest prawdziwą wartością parametru. Funkcja $c(t)$ wyraża koszt obserwacji prowadzonych do chwili t . Zakłada się, że jest ona różniczkowalna, rosnąca, wypukła oraz taka, że $c(0) = 0$.

W Rozdziale 3 podane są przykłady, które uzasadniają rozpatrywanie powyższego modelu.

Rozpatruje się procedury estymacji sekwencyjnej postaci $\delta = (\tau, \mathbf{d}(\tau))$, gdzie τ jest czasem zatrzymania względem σ -algebry $\mathcal{F}_t, t \geq 0$, a $\mathbf{d}(\tau)$ jest $\mathcal{F}_\tau$ -mierzalną funkcją. W pierwszym podrozdziale wyprowadzono bayesowskie i minimaxowe procedury sekwencyjnej

estymacji nieznanego parametru $\mathbf{p}$ rozkładu wielomianowego na podstawie obserwacji $\mathbf{Y}(1), \dots, \mathbf{Y}(n)$ otrzymywanych w chwilach losowych. Problem estymacji rozwiązuje się przy założeniu, że rozkład $U_1, \dots, U_n$ należy do pewnej szerokiej klasy i jest znany, jak również przy założeniu, że jest rozkładem wykładniczym o nieznanym parametrze.

W Podrozdziale 3.2 zaproponowano procedurę adaptacyjną, która wymaga niewielkich założeń dotyczących rozkładu $U_1, \dots, U_n$.

W ostatnim podrozdziale prezentowane są analogiczne wyniki do tych uzyskanych w Podrozdziale 3.1 dla innych funkcji straty niż określona wzorem (0.5).

Problem estymacji na podstawie danych otrzymywanych w chwilach losowych rozpatrywany był przez Starra, Wardropa i Woodroofe'a (1976) w przypadku estymacji średniej z rozkładu normalnego o znanej wariancji. Niektóre z ich wyników zostały rozszerzone przez Magierę (1982) na jednoparametrową wykładniczą rodzinę rozkładów. W Rozdziale 3 wyprowadzono jawne postaci optymalnych procedur sekwencyjnych w przypadku estymacji parametru wielowymiarowego.

W Rozdziale 4 wyprowadzono bayesowski predyktor mediany z przyszłej próby o ustalonym i losowym rozmiarze na podstawie obserwacji obciętej II typu, przy założeniu kwadratowej funkcji straty, gdy obserwacje mają przesunięty rozkład wykładniczy o funkcji gęstości postaci

$$f(x, \lambda, \alpha) = \lambda \exp\{-\lambda(x - \alpha)\} I_{(\alpha, \infty)}(x), \quad (0.6)$$

$\alpha \geq 0$ i $\lambda > 0$. Jeśli α nie jest ograniczane do wartości nieujemnych, wówczas rozkład o gęstości (0.6) nazywany jest dwuparametrowym rozkładem wykładniczym. Wprowadzenie dwóch różnych terminów dla tych dwóch rozkładów jest zasadne ze względu na to, że tylko ten pierwszy, w którym $\alpha \geq 0$, stanowi właściwy model rozkładu czasu życia. Na podstawie danych pochodzących z pracy Dunsmore'a (1974) przedstawiono numeryczną ilustrację uzyskanych wyników i dokonano ich porównania w przypadku ustalonego i losowego rozmiaru próby. Problemy predykcji pewnych funkcji przyszłej obserwacji, szczególnie w przypadku, gdy obserwacje mają rozkład wykładniczy pojawiają się w teorii niezawodności, w wielu problemach biologicznych, rolniczych i kontroli jakości. Zagadnienia predykcji przedziałowej pewnych funkcji przyszłej obserwacji rozpatrywane były między innymi w pracach Lawlessa (1972, 1977), Chou'a i Owena (1986, 1989), Evansa i Nigma (1980).

Wyniki przedstawione w Rozdziale 1 zostały uzyskane w pracy Jokiel-Rokity (1997b) przyjętej do publikacji w "Applicationes Mathematicae". Rezultaty otrzymane w Rozdziałach 2, 3 i 4 zostały przeznaczone do opublikowania, odpowiednio w pracach Jokiel-Rokity (1998), Jokiel-Rokity i Magiery (1995) i Jokiel-Rokity (1997a).

Rozdział 1

Minimaksowa estymacja i predykcja dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego przy ustalonym rozmiarze próby

Rozdział ten poświęcony jest minimaksowej predykcji kombinacji liniowej

$$\mathbf{Z} = a\mathbf{p} + b\mathbf{Y} \quad (1.1)$$

nieznanego parametru $\mathbf{p}$ i zmiennej losowej $\mathbf{Y}$, na podstawie obserwacji $\mathbf{X}$, przy funkcji straty postaci

$$L(\mathbf{Z}, \mathbf{d}) = (\mathbf{d} - \mathbf{Z})^T C (\mathbf{d} - \mathbf{Z}) = \sum_{i,j=1}^r c_{ij} (d_i(\mathbf{X}) - Z_i)(d_j(\mathbf{X}) - Z_j) \quad (1.2)$$

w dwóch następujących modelach:

- zmienne losowe $\mathbf{X}$ i $\mathbf{Y}$ są niezależne o rozkładzie wielomianowym odpowiednio $\mathcal{W}(n, \mathbf{p})$ i $\mathcal{W}(m, \mathbf{p})$, o funkcji prawdopodobieństwa określonej wzorem (1.3), gdzie n i m są znane, $\mathbf{p} = (p_1, \dots, p_r)^T$ jest nieznanym parametrem;
- zmienne losowe $\mathbf{X}$ i $\mathbf{Y}$ są niezależne o rozkładzie wielowymiarowym hipergeometrycznym z parametrami odpowiednio W , $\mathbf{W} = (W_1, \dots, W_r)^T$, n i W , $\mathbf{W} =$

$(W_1, \dots, W_r)^T$, m o funkcji prawdopodobieństwa określonej wzorem (1.26), gdzie W , n i m są znane, $\mathbf{p} = (W_1/W, \dots, W_r/W)^T$ jest nieznanym parametrem.

Uzyskane w tym rozdziale wyniki stanowią uogólnienie rezultatów Steinhausa (1957), Trybuły (1958) oraz Wilczyńskiego (1985).

1.1 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego

Niech $\mathbf{X} = (X_1, \dots, X_r)^T$, $r \geq 2$, będzie zmienną losową o rozkładzie wielomianowym z parametrami $(n, \mathbf{p})$, gdzie n jest znane, to znaczy

$$P_{\mathbf{p}} \{ \mathbf{X} = \mathbf{x} = (x_1, \dots, x_r)^T \} = \begin{cases} \frac{n!}{x_1! \dots x_r!} p_1^{x_1} \dots p_r^{x_r}, & \text{gdzie } x_i \in \{0, 1, \dots, n\} \text{ i } x_1 + \dots + x_r = n, \\ 0, & \text{w przeciwnym wypadku,} \end{cases} \quad (1.3)$$

gdzie $\mathbf{p} \in \mathcal{P} = \{ \mathbf{p} = (p_1, \dots, p_r)^T : p_i \geq 0, i = 1, \dots, r; p_1 + \dots + p_r = 1 \}$. Obserwujemy wartość zmiennej losowej $\mathbf{X}$ i stosując ocenę $\mathbf{d}(\mathbf{X}) = (d_1(\mathbf{X}), \dots, d_r(\mathbf{X}))^T$ chcemy oszacować wartość kombinacji liniowej $\mathbf{Z}$, określonej wzorem (1.1), nieznanego parametru $\mathbf{p}$ i zmiennej losowej $\mathbf{Y} = (Y_1, \dots, Y_r)^T$, która ma rozkład wielomianowy z parametrami $(m, \mathbf{p})$, gdzie m jest znane. Zakładamy, że a i b ($a, b \in R$) są znane, $\mathbf{X}$ i $\mathbf{Y}$ są niezależne i funkcja straty związana z predyktorem $\mathbf{d}(\mathbf{X})$ jest postaci (1.2), gdzie C jest macierzą nieujemnie określoną.

Naszym celem jest znalezienie minimaksowego predyktora $\mathbf{d}_0(\mathbf{X}) = (d_1^0(\mathbf{X}), \dots, d_r^0(\mathbf{X}))^T$ kombinacji liniowej $\mathbf{Z} = (Z_1, \dots, Z_r)^T$, to znaczy predyktora $\mathbf{d}_0(\mathbf{X})$, dla którego

$$\sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{Z}, \mathbf{d}_0) = \inf_{\mathbf{d} \in \mathcal{D}} \sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{Z}, \mathbf{d}), \quad (1.4)$$

gdzie $R(\mathbf{Z}, \mathbf{d}) = E_{\mathbf{p}} [L(\mathbf{Z}, \mathbf{d})]$ jest funkcją ryzyka, $\mathcal{D}$ oznacza zbiór wszystkich mierzalnych po borelowsku odwzorowań $\mathbf{d} : (\mathbf{R}^r, \mathcal{B}_{\mathbf{R}^r}) \rightarrow (\mathbf{R}^r, \mathcal{B}_{\mathbf{R}^r})$, gdzie $\mathcal{B}_{\mathbf{R}^r}$ – σ -ciało podzbiorów borelowskich przestrzeni $\mathbf{R}^r$, $\mathbf{R} = (-\infty, +\infty)$ i $\mathcal{Z} = \{ \mathbf{Z} = (Z_1, \dots, Z_r)^T \in \mathbf{R}^r, Z_1 + \dots + Z_r = a + bm \}$.

Problem ten był rozważany przez Steinhausa (1957) w przypadku, gdy $a = 1$, $b = 0$ i C jest macierzą jednostkową, przez Trybułę (1958) w przypadku, gdy $a = 1$, $b = 0$ i C jest dowolną nieujemnie określoną macierzą diagonalną oraz przez Wilczyńskiego (1985), w przypadku $a = 1$, $b = 0$ i $a = 0$, $b = 1$, przy C dowolnej macierzy nieujemnie określonej.

Następne dwa twierdzenia i lemat będą wykorzystane w dowodach głównych wyników Rozdziałów 1 i 2, zamieszczonych w Twierdzeniach 1.3, 1.4, 2.1 i 2.2.

Twierdzenie 1.1 (zobacz Aubin (1979), Twierdzenie Siona) *Niech f będzie odwzorowaniem produktu kartezjańskiego $\mathcal{P} \times \mathcal{Q}$ w $\mathbf{R}$. Jeśli*

(a) $\mathcal{P}$ i $\mathcal{Q}$ są wypukłymi i zwartymi zbiorami,

(b) $\forall q \in \mathcal{Q} \ p \mapsto f(p, q)$ jest ciągła i wypukła,

(c) $\forall p \in \mathcal{P} \ q \mapsto f(p, q)$ jest ciągła i wklęsła,

to istnieje punkt siodłowy $(\bar{p}, \bar{q}) \in \mathcal{P} \times \mathcal{Q}$, to znaczy punkt $(\bar{p}, \bar{q})$, dla którego

$$\inf_{p \in \mathcal{P}} \sup_{q \in \mathcal{Q}} f(p, q) = \sup_{q \in \mathcal{Q}} f(\bar{p}, q) = f(\bar{p}, \bar{q}) = \inf_{p \in \mathcal{P}} f(p, \bar{q}) = \sup_{q \in \mathcal{Q}} \inf_{p \in \mathcal{P}} f(p, q).$$

□

Twierdzenie 1.2 (zobacz Karmanov (1986), Twierdzenie 3.5.4) *Niech $f : \mathcal{Q} \mapsto \mathbf{R}$ będzie wypukłą funkcją określoną na wypukłym zbiorze $\mathcal{Q} = \{q = (q_1, \dots, q_n)^T \in \mathbf{R}^n : a_i^T q - b_i \geq 0, i = 1, \dots, m\}$ dla pewnych $a_1, \dots, a_m \in \mathbf{R}^n$ i $b_1, \dots, b_m \in \mathbf{R}$. Punkt $\bar{q} \in \mathcal{Q}$ jest rozwiązaniem równania $\inf_{q \in \mathcal{Q}} f(q) = f(\bar{q})$ wtedy i tylko wtedy, gdy istnieją nieujemne liczby rzeczywiste $u_1, \dots, u_m$, $u_i \geq 0$, $i = 1, \dots, m$, dla których*

$$f'(\bar{q}) = \sum_{i=1}^m u_i a_i, \quad \sum_{i=1}^m (a_i^T \bar{q} - b_i) u_i = 0,$$

gdzie $f'(q)$ oznacza gradient funkcji f w punkcie q .

□

Lemat 1.1 (zobacz Ferguson (1967), Twierdzenie 2.11.1) *Niech π będzie rozkładem a priori parametru θ oraz $\bar{R}(\pi, d) = E_\pi R(\theta, d)$ bayesowskim ryzykiem estymatora d . Jeśli d_0 jest bayesowskim estymatorem parametru θ względem rozkładu a priori π_0 oraz*

$$\sup_{\theta \in \Theta} R(\theta, d_0) = \bar{R}(\pi_0, d_0),$$

to d_0 jest minimaksowy.

□

Oznaczmy $\mathbf{c} = (c_{11}, \dots, c_{rr})^T$. Niech $\pi_{\alpha, \beta}$ oznacza rozkład a priori Dirichleta $D(\alpha\beta_1, \alpha\beta_2, \dots, \alpha\beta_r)$ parametru $\mathbf{p}$ o funkcji gęstości

$$h(\mathbf{p}; \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha)}{\prod_{j \in A} \Gamma(\alpha\beta_j)} \prod_{j \in A} p_j^{\alpha\beta_j - 1}, & \text{gdzie } \sum_{i=1}^r p_i = \sum_{i \in A} p_i = 1, \\ p_i \geq 0, \ i = 1, \dots, r, \\ 0, & \text{w przeciwnym wypadku,} \end{cases}$$

gdzie $\alpha > 0$, $\beta = (\beta_1, \dots, \beta_r)^T \in \mathcal{P}$ i $A = \{i_1, \dots, i_k\} \subset \{1, \dots, r\}$, $k \geq 2$ (zobacz Uwaga 2), jest zbiorem takim, że $\beta_i > 0$ dla $i \in A$ oraz $\beta_i = 0$ dla $i \notin A$. Przez $\pi_{0,\beta}$ oznaczmy rozkład a priori parametru $\mathbf{p}$ postaci $P(\mathbf{p} = \mathbf{e}_j) = \beta_j$, gdzie $\mathbf{e}_1 = (1, 0, \dots, 0)^T$, $\mathbf{e}_2 = (0, 1, \dots, 0)^T, \dots, \mathbf{e}_r = (0, 0, \dots, 1)^T$ oraz przez $\pi_{\infty,\beta}$ rozkład a priori jednopunktowy parametru $\mathbf{p}$ postaci $P(\mathbf{p} = \beta) = 1$. Następujący lemat podaje postaci predyktorów bayesowskich kombinacji liniowej $\mathbf{Z}$, względem rozkładów a priori $\pi_{\alpha,\beta}$, $\pi_{0,\beta}$ i $\pi_{\infty,\beta}$ parametru $\mathbf{p}$, przy funkcji straty określonej wzorem (1.2).

Lemat 1.2 *Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określoną. Wówczas następujące predyktory*

$$\mathbf{d}^{\alpha,\beta}(\mathbf{X}) = \begin{cases} (a + bm)\beta, & \text{gdy } n = 0, \\ (a + bm) \left(\frac{1}{n + \alpha} \mathbf{X} + \frac{\alpha}{n + \alpha} \beta \right), & \text{gdy } n > 0, \end{cases} \quad (1.5)$$

$$\mathbf{d}^{0,\beta}(\mathbf{X}) = \begin{cases} (a + bm)\beta, & \text{gdy } n = 0, \\ \frac{a + bm}{n} \mathbf{X}, & \text{gdy } n > 0 \end{cases} \quad (1.6)$$

i

$$\mathbf{d}^{\infty,\beta}(\mathbf{X}) = (a + bm)\beta \quad (1.7)$$

kombinacji liniowej $\mathbf{Z}$ są bayesowskie względem rozkładów a priori parametru $\mathbf{p}$ odpowiednio $\pi_{\alpha,\beta}$, $\pi_{0,\beta}$ i $\pi_{\infty,\beta}$, a ich ryzyka bayesowskie są równe

$$\overline{R}(\pi_{\alpha,\beta}, \mathbf{d}^{\alpha,\beta}) = \begin{cases} \frac{1}{\alpha + 1} [(a + bm)^2 + \alpha b^2 m] (\mathbf{c}^T \beta - \beta^T C \beta), & \text{gdy } n = 0, \\ w_1 \left(\frac{\alpha}{\alpha + 1} \beta^T C \beta + \frac{1}{\alpha + 1} \mathbf{c}^T \beta \right) - w_2 \beta^T C \beta + w_3 \mathbf{c}^T \beta, & \text{gdy } n > 0, \end{cases} \quad (1.8)$$

$$\overline{R}(\pi_{0,\beta}, \mathbf{d}^{0,\beta}) = \begin{cases} (a + bm)^2 (\mathbf{c}^T \beta - \beta^T C \beta), & \text{gdy } n = 0, \\ 0, & \text{gdy } n > 0 \end{cases} \quad (1.9)$$

i

$$\overline{R}(\pi_{\infty,\beta}, \mathbf{d}^{\infty,\beta}) = b^2 m (\mathbf{c}^T \beta - \beta^T C \beta), \quad (1.10)$$

gdzie

$$w_1 = \left(\frac{a + bm}{n + \alpha} \right)^2 (\alpha^2 - n) - b^2 m = w_2 - w_3, \quad (1.11)$$

$$w_2 = \left[\frac{(a + bm)\alpha}{n + \alpha} \right]^2, \quad (1.12)$$

$$w_3 = \left(\frac{a + bm}{n + \alpha} \right)^2 n + b^2 m. \quad (1.13)$$

Dowód: Dla dowolnego predyktora $\mathbf{d}(\mathbf{X}) \in \mathcal{D}$, funkcję straty $L(\mathbf{Z}, \mathbf{d})$ możemy zapisać w postaci

$$\begin{aligned} L(\mathbf{Z}, \mathbf{d}) &= \sum_{i,j=1}^r c_{ij} (d_i - Z_i)(d_j - Z_j) \\ &= \sum_{i,j=1}^r c_{ij} (a + bm)^2 \left(\frac{d_i}{a + bm} - p_i \right) \left(\frac{d_j}{a + bm} - p_j \right) \\ &\quad - \sum_{i,j=1}^r c_{ij} b(a + bm) \left(\frac{d_i}{a + bm} - p_i \right) (Y_j - mp_j) \\ &\quad - \sum_{i,j=1}^r c_{ij} b(a + bm) (Y_i - mp_i) \left(\frac{d_j}{a + bm} - p_j \right) \\ &\quad + \sum_{i,j=1}^r c_{ij} b^2 (Y_i - mp_i)(Y_j - mp_j) \end{aligned}$$

i wobec tego funkcja ryzyka $R(\mathbf{Z}, \mathbf{d})$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}) &= E_{\mathbf{P}} \left[\sum_{i,j=1}^r c_{ij} (a + bm)^2 \left(\frac{d_i}{a + bm} - p_i \right) \left(\frac{d_j}{a + bm} - p_j \right) \right] \\ &\quad + E_{\mathbf{P}} \left[\sum_{i,j=1}^r c_{ij} b^2 (Y_i - mp_i)(Y_j - mp_j) \right]. \end{aligned} \quad (1.14)$$

Stąd, jeżeli chcemy minimalizować $\bar{R}(\pi, \mathbf{d}) = E_{\pi} R(\mathbf{Z}, \mathbf{d})$ po wszystkich $\mathbf{d} \in \mathcal{D}$ (gdzie π jest dowolnym rozkładem a priori parametru $\mathbf{p}$), to w związku z tym, że drugi składnik we wzorze (1.14) nie zależy od $\mathbf{d}(\mathbf{X})$, wystarczy znaleźć $\mathbf{d}^{\pi}(\mathbf{X})$, które minimalizuje wartość oczekiwaną

$$E_{\pi} \left\{ E_{\mathbf{P}} \left[\sum_{i,j=1}^r c_{ij} (a + bm)^2 \left(\frac{d_i}{a + bm} - p_i \right) \left(\frac{d_j}{a + bm} - p_j \right) \right] \right\}.$$

Stąd otrzymujemy, że szukany predyktor bayesowski kombinacji liniowej $\mathbf{Z}$ jest iloczynem $a + bm$ i estymatora bayesowskiego parametru $\mathbf{p}$. Łatwo sprawdzić, że jest on postaci (1.5), gdy π jest rozkładem Dirichleta $\pi_{\alpha, \beta}$; postaci (1.6), gdy π jest rozkładem $\pi_{0, \beta}$ oraz postaci (1.7), gdy π jest rozkładem jednopunktowym $\pi_{\infty, \beta}$. Uwzględniając, że dla $i = 1, \dots, r$ zachodzą równości

$$E_{\mathbf{P}} X_i = np_i \quad \text{ i } \quad E_{\mathbf{P}} (X_i - np_i)^2 = np_i(1 - p_i) \quad (1.15)$$

oraz dla $i \neq j$ zachodzi równość

$$E_{\mathbf{P}} [(X_i - np_i)(X_j - np_j)] = -np_i p_j, \quad (1.16)$$

funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\alpha, \beta})$ predyktora $\mathbf{d}^{\alpha, \beta}(\mathbf{X})$ jest postaci

$$R(\mathbf{Z}, \mathbf{d}^{\alpha, \beta}) = \begin{cases} [(a + bm)^2 - b^2m] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta^T C \beta \\ -2(a + bm)^2 \beta^T C \mathbf{p} + b^2 m \mathbf{c}^T \mathbf{p}, & \text{gd } n = 0, \\ w_1 \mathbf{p}^T C \mathbf{p} + w_2 \beta^T C \beta + w_3 \mathbf{c}^T \mathbf{p} - 2w_2 \beta^T C \mathbf{p}, & \text{gd } n > 0, \end{cases} \quad (1.17)$$

gdzie w_1 , w_2 i w_3 dane są odpowiednio wzorami (1.11), (1.12) i (1.13). Korzystając ze wzoru Liouville'a (Fichtenholz (1985), Tom 3) można pokazać, że dla $i = 1, \dots, r$ zachodzą równości

$$E_{\pi_{\alpha, \beta}}(p_i) = \beta_i,$$

$$E_{\pi_{\alpha, \beta}}(p_i^2) = \frac{\beta_i(\alpha \beta_i + 1)}{\alpha + 1}$$

oraz dla $i \neq j$ zachodzi równość

$$E_{\pi_{\alpha, \beta}}(p_i p_j) = \frac{\alpha \beta_i \beta_j}{\alpha + 1}.$$

Stąd ryzyko bayesowskie $\bar{R}(\pi_{\alpha, \beta}, \mathbf{d}^{\alpha, \beta})$ predyktora $\mathbf{d}^{\alpha, \beta}(\mathbf{X})$ jest postaci (1.8). Analogicznie, korzystając z równości (1.15) i (1.16), funkcje ryzyka predyktorów $\mathbf{d}^{0, \beta}(\mathbf{X})$ i $\mathbf{d}^{\infty, \beta}(\mathbf{X})$ są odpowiednio postaci

$$R(\mathbf{Z}, \mathbf{d}^{0, \beta}) = \begin{cases} [(a + bm)^2 - b^2m] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta^T C \beta \\ -2(a + bm)^2 \beta^T C \mathbf{p} + b^2 m \mathbf{c}^T \mathbf{p}, & \text{gd } n = 0, \\ \left[\frac{(a + bm)^2}{n} + b^2 m \right] (\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p}), & \text{gd } n > 0 \end{cases}$$

i

$$R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) = [(a + bm)^2 - b^2m] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta^T C \beta - 2(a + bm)^2 \beta^T C \mathbf{p} + b^2 m \mathbf{c}^T \mathbf{p}.$$

Tak więc ryzyka bayesowskie $\bar{R}(\pi_{0, \beta}, \mathbf{d}^{0, \beta})$ i $\bar{R}(\pi_{\infty, \beta}, \mathbf{d}^{\infty, \beta})$ są odpowiednio postaci (1.9) i (1.10). $\square$

Następujące twierdzenie podaje postać minimaxowego predyktora kombinacji liniowej $\mathbf{Z}$.

Twierdzenie 1.3 *Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określona. Wówczas predyktor $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ określony przez*

$$\mathbf{d}_0(\mathbf{X}) = \begin{cases} (a + bm) \left(\frac{1}{n + \alpha_0} \mathbf{X} + \frac{\alpha_0}{n + \alpha_0} \beta_0 \right), & \text{gd } (a + bm)^2 - b^2 m > 0 \text{ i } n > 0, \\ (a + bm) \beta_0, & \text{w przeciwnym wypadku,} \end{cases} \quad (1.18)$$

gdzie

$$\alpha_0 = \begin{cases} \frac{nb^2m + |a + bm| \sqrt{(a + bm)^2n + b^2mn(n-1)}}{(a + bm)^2 - b^2m}, & \text{gdy } n > 1, \\ \frac{(a + bm)^2 + b^2m}{(a + bm)^2 - b^2m}, & \text{gdy } n = 1 \end{cases} \quad (1.19)$$

i β_0 jest wektorem $(\beta_1^0, \dots, \beta_r^0)^T$ spełniającym warunek

$$\mathbf{c}^T \beta_0 - \beta_0^T C \beta_0 = \max_{\beta \in \mathcal{P}} (\mathbf{c}^T \beta - \beta^T C \beta), \quad (1.20)$$

jest minimaksowy.

Dowód: Rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\alpha, \beta}(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori $\pi_{\alpha, \beta}$ Dirichleta parametru $\mathbf{p}$, czyli predyktory postaci (1.5). Funkcja ryzyka predyktora określonego wzorem (1.5) jest postaci (1.17). Wobec tego, że w_1 określone wzorem (1.11) przyjmuje wartość zero wtedy i tylko wtedy, gdy

$$[(a + bm)^2 - b^2m] \alpha^2 - 2nb^2m\alpha - b^2mn^2 - (a + bm)^2n = 0,$$

otrzymujemy, że jeśli $(a + bm)^2 - b^2m > 0$, to istnieje $\alpha_0 > 0$, dla którego $w_1 = 0$. Łatwo można sprawdzić, że α_0 jest postaci (1.19). Oznaczmy $R_0(\beta, \mathbf{p}) = R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta})$. Wobec tego, że dla $\alpha = \alpha_0$, mamy $w_1 = 0$ i $w_2 = w_3$, więc funkcja ryzyka $R_0(\beta, \mathbf{p})$ sprowadza się do następującej postaci

$$R_0(\beta, \mathbf{p}) = w_2(\beta^T C \beta + \mathbf{c}^T \mathbf{p} - 2\beta^T C \mathbf{p}). \quad (1.21)$$

Zauważmy, że z nieujemnej określoności macierzy C wynika, że funkcja $R_0(\cdot, \mathbf{p}) : \mathcal{P} \mapsto \mathbf{R}$ jest wypukła względem zmiennej β , dla każdej wartości $\mathbf{p}$ (Karmanov (1986), Lemat 2.4.4). Ponadto odwzorowanie $R_0(\beta, \cdot) : \mathcal{P} \mapsto \mathbf{R}$ jest wklęsłe, a nawet liniowe, względem zmiennej $\mathbf{p}$. Oczywiście jest również to, że $R_0(\cdot, \cdot) : \mathcal{P} \times \mathcal{P} \mapsto \mathbf{R}$ jest ciągła względem $(\beta, \mathbf{p})$, a $\mathcal{P}$ jest domkniętym i ograniczonym podzbiorem wypukłym $\mathbf{R}^r$, a więc i zwartym. Wobec tego, z Twierdzenia 1.1 wynika, że istnieje punkt siodłowy $(\beta_0, \mathbf{p}_0) \in \mathcal{P} \times \mathcal{P}$ taki, że

$$\inf_{\beta \in \mathcal{P}} \sup_{\mathbf{p} \in \mathcal{P}} R_0(\beta, \mathbf{p}) = \sup_{\mathbf{p} \in \mathcal{P}} R_0(\beta_0, \mathbf{p}) = R_0(\beta_0, \mathbf{p}_0) = \inf_{\beta \in \mathcal{P}} R_0(\beta, \mathbf{p}_0) = \sup_{\mathbf{p} \in \mathcal{P}} \inf_{\beta \in \mathcal{P}} R_0(\beta, \mathbf{p}).$$

Jak wiadomo (Vorob'ev (1977), Uwaga 1.7.2) współrzędne punktu siodłowego $(\beta_0, \mathbf{p}_0)$ można wybrać niezależnie jako punkty, dla których osiągnięte są następujące równości:

$$\max_{\mathbf{p} \in \mathcal{P}} \inf_{\beta \in \mathcal{P}} R_0(\beta, \mathbf{p}) = \inf_{\beta \in \mathcal{P}} R_0(\beta, \mathbf{p}_0), \quad (1.22)$$

$$\min_{\beta \in \mathcal{P}} \sup_{\mathbf{p} \in \mathcal{P}} R_0(\beta, \mathbf{p}) = \sup_{\mathbf{p} \in \mathcal{P}} R_0(\beta_0, \mathbf{p}).$$

Zakładając chwilowo, że macierz C jest dodatnio określona, możemy stwierdzić, że punkt $\mathbf{p}_0 = (p_1^0, \dots, p_r^0)^T \in \mathcal{P}$, dla którego spełniony jest warunek

$$\mathbf{c}^T \mathbf{p}_0 - \mathbf{p}_0^T C \mathbf{p}_0 = \max_{\mathbf{p} \in \mathcal{P}} (\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p}) \quad (1.23)$$

jest jedynym rozwiązaniem (1.22). Powyższy fakt wynika stąd, że jeśli gradient funkcji wypukłej $R_0(\cdot, \mathbf{p})$ określonej na zbiorze wypukłym $\mathcal{P}$ przyjmuje wartość zero w punkcie należącym do $\mathcal{P}$, to w punkcie tym osiągnięte jest minimum globalne i jest ono jedyne, gdy funkcja $R_0(\cdot, \mathbf{p})$ jest ściśle wypukła (a jest ściśle wypukła, gdy macierz C jest dodatnio określona). W naszym przypadku gradient funkcji $R_0(\cdot, \mathbf{p})$ w punkcie β równa się $2w_2(C\beta - C\mathbf{p})$ i przyjmuje wartość zero jedynie dla $\beta = \mathbf{p}$ (ingeruje dodatnia określoność macierzy C) i wobec tego

$$\max_{\mathbf{p} \in \mathcal{P}} \inf_{\beta \in \mathcal{P}} R_0(\beta, \mathbf{p}) = \max_{\mathbf{p} \in \mathcal{P}} R_0(\mathbf{p}, \mathbf{p}) = \max_{\mathbf{p} \in \mathcal{P}} \{\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p}\}.$$

Z kolei rozpatrując ściśle wypukłą funkcję $R_0(\cdot, \mathbf{p}_0)$ i używając analogicznych argumentów wnioskujemy, że funkcja ta osiąga jedyne minimum w punkcie $\beta_0 = \mathbf{p}_0$. Punkt $(\mathbf{p}_0, \mathbf{p}_0)$ jest więc jedynym punktem siodłowym funkcji $R_0(\cdot, \cdot)$ na zbiorze $\mathcal{P} \times \mathcal{P}$. W przypadku, gdy macierz C jest nieujemnie określona, to dla $\beta = \mathbf{p}$, funkcja $R_0(\cdot, \mathbf{p})$ przyjmuje wartość minimalną na zbiorze $\mathcal{P}$. Wynika to stąd, że minima lokalne funkcji wypukłej, określonej na zbiorze wypukłym, są jej minimami globalnymi na danym zbiorze (Badach, Molisz i Seidler (1980)). Korzystając z Twierdzenia 1.2 znajdziemy teraz taki punkt $\mathbf{p}_0 \in \mathcal{P}$, dla którego spełniony jest warunek (1.23), przy założeniu, że macierz C jest nieujemnie określona. W naszym przypadku

$$\begin{aligned} a_1 &= (1, 0, \dots, 0)^T, \quad a_2 = (0, 1, \dots, 0)^T, \quad \dots, \quad a_r = (0, 0, \dots, 1)^T, \\ a_{r+1} &= a_1 + \dots + a_r, \quad a_{r+2} = -a_{r+1}, \\ b_i &= 0, i = 1, \dots, r, \quad b_{r+1} = -b_{r+2} = 1, \end{aligned}$$

$m = r + 2$, $n = r$. Punkt $\mathbf{p}_0$ spełnia więc warunek (1.23) (punkt taki oczywiście istnieje, gdyż $\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p}$ jest ciągłym odwzorowaniem zbioru zwartego $\mathcal{P}$) wtedy i tylko wtedy, gdy istnieje stała u_0 taka, że dla każdego $i \in \{1, 2, \dots, r\}$ zachodzą implikacje

$$\begin{aligned} \text{jeśli } p_i^0 &> 0, & \text{to } 2 \sum_{j=1}^r c_{ij} p_j^0 - c_{ii} &= u_0, \\ \text{jeśli } p_i^0 &= 0, & \text{to } 2 \sum_{j=1}^r c_{ij} p_j^0 - c_{ii} &\geq u_0. \end{aligned}$$

Z jednej strony otrzymaliśmy, że dla każdego $\beta \in \mathcal{P}$

$$R_0(\beta, \mathbf{p}_0) \geq R_0(\mathbf{p}_0, \mathbf{p}_0),$$

ponieważ dla $\beta = \mathbf{p}_0$, funkcja $R_0(\cdot, \mathbf{p}_0)$ przyjmuje wartość minimalną. Z drugiej strony

$$\begin{aligned} R_0(\mathbf{p}_0, \mathbf{p}) &= w_1 \left[\sum_{i,j=1}^r c_{ij} p_i^0 p_j^0 + \sum_{i=1}^r (c_{ii} - 2 \sum_{j=1}^r c_{ij} p_j^0) p_i \right] \\ &\leq w_1 \left[\sum_{i,j=1}^r c_{ij} p_i^0 p_j^0 - u_0 \right] = R_0(\mathbf{p}_0, \mathbf{p}_0), \end{aligned} \quad (1.24)$$

a więc $(\mathbf{p}_0, \mathbf{p}_0)$ jest punktem siodłowym funkcji $R_0(\cdot, \cdot)$ na zbiorze $\mathcal{P} \times \mathcal{P}$. Załóżmy, że $p_i^0 > 0$ dla $i \in A$ i $p_i^0 = 0$ dla $i \notin A$. Predyktor $\mathbf{d}^{\alpha_0, \mathbf{p}_0}(\mathbf{X})$ jest bayesowski względem rozkładu a priori $\pi_{\alpha_0, \mathbf{p}_0}$ Dirichleta. Korzystając z Lematu 1.2 otrzymujemy, że jego ryzyko bayesowskie $\bar{R}(\pi_{\alpha_0, \mathbf{p}_0}, \mathbf{d}^{\alpha_0, \mathbf{p}_0})$ jest równe

$$\bar{R}(\pi_{\alpha_0, \mathbf{p}_0}, \mathbf{d}^{\alpha_0, \mathbf{p}_0}) = R_0(\mathbf{p}_0, \mathbf{p}_0).$$

Następnie korzystając z nierówności (1.24) i Lematu 1.1 otrzymujemy, że w przypadku, gdy $(a + bm)^2 - b^2m > 0$ predyktor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\alpha_0, \mathbf{p}_0}(\mathbf{X})$ jest minimaksowy.

W przypadku, gdy $(a + bm)^2 - b^2m \leq 0$ rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\infty, \beta}(\mathbf{X}) = (a + bm)\beta$ względem rozkładu a priori $\pi_{\infty, \beta}$ parametru $\mathbf{p}$. Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$ rozpatrywanego predyktora $\mathbf{d}^{\infty, \beta}(\mathbf{X})$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) &= (a + bm)^2 \beta^T C \beta - 2(a + bm) \beta^T C \mathbf{p} \\ &\quad + [(a + bm)^2 - b^2m] \mathbf{p}^T C \mathbf{p} + b^2m \mathbf{c}^T \mathbf{p}, \end{aligned}$$

a więc, gdy tymczasowo założymy, że $a + bm \neq 0$, jest ona funkcją wypukłą względem zmiennej β dla każdej wartości $\mathbf{p}$ i wklęsłą względem zmiennej $\mathbf{p}$ dla każdej wartości β . Można pokazać, podobnie jak w przypadku, gdy $(a + bm)^2 - b^2m > 0$, że (β_0, β_0) , gdzie β_0 jest wektorem spełniającym warunek (1.20), jest punktem siodłowym funkcji $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$. Ponadto

$$\sup_{\mathbf{p} \in \mathcal{P}} R(\mathbf{Z}, \mathbf{d}^{\infty, \beta_0}) = b^2m(\mathbf{c}^T \beta_0 - \beta_0^T C \beta_0) = \bar{R}(\pi_{\infty, \beta_0}, \mathbf{d}^{\infty, \beta_0}).$$

Wobec tego, korzystając z Lematu 1.1 otrzymujemy, że predyktor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\infty, \beta_0}(\mathbf{X})$ jest minimaksowy.

W przypadku, gdy $a + bm = 0$ funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$ predyktora $\mathbf{d}^{\infty, \beta}$ jest postaci

$$R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) = b^2m(\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p})$$

i

$$\sup_{\mathbf{p} \in \mathcal{P}} R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) = \sup_{\mathbf{p} \in \mathcal{P}} R(\mathbf{Z}, \mathbf{d}^{\infty, \beta_0}) = b^2m(\mathbf{c}^T \beta_0 - \beta_0^T C \beta_0) = \bar{R}(\pi_{\infty, \beta_0}, \mathbf{d}^{\infty, \beta_0}),$$

gdzie β_0 jest wektorem spełniającym warunek (1.20), a więc na mocy Lematu 1.1 predyktor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\infty, \beta_0}(\mathbf{X})$ jest minimaksowy. $\square$

Uwaga 1. Otrzymany predyktor minimaksowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z} = a\mathbf{p} + b\mathbf{Y}$ jest kombinacją liniową minimaksowego estymatora parametru $\mathbf{p}$ i minimaksowego predyktora zmiennej losowej $\mathbf{Y}$, o współczynnikach zależnych od a, b, m i n .

Uwaga 2. Sformułowana definicja rozkładu a priori $\pi_{\alpha, \beta}$ jest poprawna wtedy, gdy $k \geq 2$. W przypadku, gdy $k = 1$ otrzymujemy, że

$$\max_{\mathbf{p} \in \mathcal{P}} (\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p}) = \mathbf{c}^T \mathbf{p}_0 - \mathbf{p}_0^T C \mathbf{p}_0 = 0.$$

Z drugiej strony $p_1 + \dots + p_r = 1$ dla każdego $\mathbf{p} \in \mathcal{P}$, więc

$$\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p} = \sum_{i=1}^r c_{ii} p_i - \sum_{i,j=1}^r c_{ij} p_i p_j = \frac{1}{2} \sum_{i,j=1}^r (c_{ii} + c_{jj} - 2c_{ij}) p_i p_j.$$

Z nieujemnej określoności macierzy C wynika, że dla każdego $i, j \in \{1, \dots, r\}$

$$c_{ii} + c_{jj} - 2c_{ij} \geq 0 \quad (1.25)$$

oraz

$$c_{ii} c_{jj} - c_{ij}^2 \geq 0.$$

Ponieważ $p_i \geq 0$, $i = 1, \dots, r$, więc na podstawie nierówności (1.25) wnioskujemy, że $k = 1$ wtedy i tylko wtedy, gdy dla każdego $i, j \in \{1, \dots, r\}$ zachodzi równość $c_{ii} + c_{jj} - 2c_{ij} = 0$. Tak więc

$$0 \leq c_{ii} c_{jj} - c_{ij}^2 = c_{ii} c_{jj} - \frac{1}{4} (c_{ii} + c_{jj})^2 = -\frac{1}{4} (c_{ii} - c_{jj})^2 \leq 0,$$

a to implikuje istnienie nieujemnej stałej c_0 takiej, że $c_{ij} = c_0$ dla każdego $i, j \in 1, \dots, r$. W tym przypadku funkcję ryzyka $R(\mathbf{Z}, \mathbf{d})$ można zapisać w postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}) &= E_{\mathbf{P}} \left[c_0 (a + bm)^2 \sum_{i,j=1}^r \left(\frac{d_i(\mathbf{X})}{a + bm} - p_i \right) \left(\frac{d_j(\mathbf{X})}{a + bm} - p_j \right) \right] \\ &\quad + E_{\mathbf{P}} \left[c_0 b^2 \sum_{i,j=1}^r (Y_i - m p_i)(Y_j - m p_j) \right] \\ &= E_{\mathbf{P}} \left\{ c_0 (a + bm)^2 \left[\sum_{i=1}^r \left(\frac{d_i(\mathbf{X})}{a + bm} - p_i \right) \right]^2 \right\} \\ &\quad + E_{\mathbf{P}} \left[c_0 b^2 \sum_{i,j=1}^r (Y_i - m p_i)(Y_j - m p_j) \right] \\ &= c_0 (a + bm)^2 E_{\mathbf{P}} \left[\frac{1}{(a + bm)} \sum_{i=1}^r d_i(\mathbf{X}) - 1 \right]^2 \\ &\quad + c_0 b^2 E_{\mathbf{P}} \left[\sum_{i,j=1}^r (Y_i - m p_i)(Y_j - m p_j) \right]. \end{aligned}$$

Ten przypadek nie jest jednak interesujący, gdyż

$$\inf_{\mathbf{d} \in \mathcal{D}} \sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{Z}, \mathbf{d}) = \sup_{\mathbf{p} \in \mathcal{P}} E_{\mathbf{p}} \left[c_0 b^2 \sum_{i,j=1}^r (Y_i - mp_i)(Y_j - mp_j) \right]$$

i wtedy każdy predyktor $\mathbf{d}(\mathbf{X}) = (d_1(\mathbf{X}), \dots, d_r(\mathbf{X}))^T$, dla którego $d_1(\mathbf{X}) + \dots + d_r(\mathbf{X}) = a + bm$ jest minimaksowy.

1.2 Minimaksowa estymacja i predykcja dla wielowymiarowego rozkładu hipergeometrycznego

Niech $\mathbf{X} = (X_1, \dots, X_r)^T$ będzie zmienną losową o rozkładzie wielowymiarowym hipergeometrycznym z parametrami $(W, \mathbf{W}, n)$, to znaczy,

$$P_{\mathbf{p}}\{\mathbf{X} = \mathbf{x} = (x_1, \dots, x_r)^T\} = \begin{cases} \frac{\binom{W_1}{x_1} \dots \binom{W_r}{x_r}}{\binom{W}{n}}, & \text{gdzie } x_i \in \{0, 1, \dots, W_i\} \text{ i } x_1 + \dots + x_r = n, \\ 0, & \text{w przeciwnym wypadku,} \end{cases} \quad (1.26)$$

gdzie $W_1 + \dots + W_r = W$, $W > 2$, $r \geq 2$, $0 \leq n \leq W$ są znane, $\mathbf{W} = (W_1, \dots, W_r)^T$ nie jest znane. Obserwujemy wartość zmiennej losowej $\mathbf{X}$ i stosując ocenę $\mathbf{d}(\mathbf{X}) = (d_1(\mathbf{X}), \dots, d_r(\mathbf{X}))^T$ chcemy oszacować wartość kombinacji liniowej

$$\mathbf{Z} = a\mathbf{p} + b\mathbf{Y} \quad (1.27)$$

nieznanego parametru $\mathbf{p} = \left(\frac{W_1}{W}, \dots, \frac{W_r}{W}\right)^T$ i zmiennej losowej $\mathbf{Y} = (Y_1, \dots, Y_r)^T$, która ma rozkład wielowymiarowy hipergeometryczny z parametrami $(W, \mathbf{W}, m)$, gdzie $0 < m \leq W$ jest znane. Zakładamy, że $\mathbf{X}$ i $\mathbf{Y}$ są niezależne, a i b ($a, b \in \mathcal{R}$) są znane i spełniają jeden z poniższych warunków:

$$(a + bm)^2(W - n)(W - n - 1) - b^2m(W - m)W \geq 0,$$

$$(a + bm)^2 - b^2m \frac{(W - m)}{W - 1} \leq 0.$$

Przyjmujemy, że strata związana z predyktorem $\mathbf{d}(\mathbf{X})$ określona jest wzorem (1.2), gdzie macierz $C = (c_{ij})$ jest nieujemnie określona. Naszym celem jest znalezienie minimaksowego predyktora $\mathbf{d}_0(\mathbf{X}) = (d_1^0(\mathbf{X}), \dots, d_r^0(\mathbf{X}))^T$ kombinacji liniowej $\mathbf{Z} = (Z_1, \dots, Z_r)^T$,

to znaczy predyktora $\mathbf{d}_0(\mathbf{X})$, który spełnia warunek (1.4), gdzie $R(\mathbf{Z}, \mathbf{d}) = E_{\mathbf{p}}[L(\mathbf{Z}, \mathbf{d})]$ jest funkcją ryzyka, $\mathcal{D}$ oznacza zbiór wszystkich mierzalnych po borelowsku odwzorowań $\mathbf{d} : (\mathbf{R}^r, \mathcal{B}_{\mathbf{R}^r}) \rightarrow (\mathbf{R}^r, \mathcal{B}_{\mathbf{R}^r})$, gdzie $\mathcal{B}_{\mathbf{R}^r}$ – σ -ciało podzbiorów borelowskich przestrzeni $\mathbf{R}^r$, $\mathbf{R} = (-\infty, +\infty)$ i $\mathcal{Z} = \{\mathbf{Z} = (Z_1, \dots, Z_r)^T \in \mathbf{R}^r, Z_1 + \dots + Z_r = a + bm\}$. Problem ten był rozważany przez Trybułę (1958) w przypadku, gdy $a = 1$, $b = 0$ i C jest macierzą diagonalną oraz przez Wilczyńskiego (1985) w przypadku, gdy $a = 1$, $b = 0$ i C jest dowolną macierzą nieujemnie określoną.

Następujący lemat podaje postaci bayesowskich predyktorów kombinacji liniowej $\mathbf{Z}$, przy funkcji straty określonej wzorem (1.2), względem następujących rozkładów a priori parametru $\mathbf{p}$: 1) Pòlya-Eggenbergera $\pi_{\alpha, \beta}$ z parametrami α, β ($\alpha > 0$, $\beta = (\beta_1, \dots, \beta_r)^T \in \mathcal{P}$) o funkcji prawdopodobieństwa postaci

$$P \left\{ \mathbf{p} = \left(\frac{W_1}{W}, \dots, \frac{W_r}{W} \right)^T \right\} = \begin{cases} \frac{W! \Gamma(\alpha)}{\Gamma(W + \alpha)} \prod_{j \in A} \frac{\Gamma(W_j + \alpha \beta_j)}{W_j! \Gamma(\alpha \beta_j)}, & \text{gdy } W_i \in \{0, \dots, W\}, i = 1, \dots, r, \\ & \sum_{i=1}^r W_i = \sum_{j \in A} W_j = W, \\ 0, & \text{w przeciwnym wypadku;} \end{cases}$$

2) wielomianowego $\pi_{\infty, \beta}$ o funkcji prawdopodobieństwa postaci

$$P \left\{ \mathbf{p} = \left(\frac{W_1}{W}, \dots, \frac{W_r}{W} \right)^T \right\} = \begin{cases} W! \prod_{j \in A} \frac{\beta_j^{W_j}}{W_j!}, & \text{gdy } W_i \in \{0, \dots, W\}, i = 1, \dots, r, \\ & \sum_{i=1}^r W_i = \sum_{j \in A} W_j = W, \\ 0, & \text{w przeciwnym wypadku;} \end{cases}$$

3) $\pi_{0, \beta}$, określonego przez $P(\mathbf{p} = \mathbf{e}_j) = \beta_j$;

4) jednopunktowego π_{β} , określonego przez $P(\mathbf{p} = \beta) = 1$, gdzie $A = \{i_1, \dots, i_k\} \subset \{1, \dots, r\}$, $k \geq 2$ (zobacz Uwaga 2, Rozdział 1.1), jest zbiorem takim, że $\beta_i > 0$ dla $i \in A$ oraz $\beta_i = 0$ dla $i \notin A$.

Lemat 1.3 Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określoną. Wówczas następujące predyktory

$$\mathbf{d}^{\alpha, \beta}(\mathbf{X}) = \begin{cases} (a + bm)\beta, & \text{gdy } n = 0, \\ (a + bm) \left[\frac{W + \alpha}{W(n + \alpha)} \mathbf{X} + \frac{\alpha(W - n)}{W(n + \alpha)} \beta \right], & \text{gdy } n > 0, \end{cases} \quad (1.28)$$

$$\mathbf{d}^{\infty, \beta}(\mathbf{X}) = \begin{cases} (a + bm)\beta, & \text{gdy } n = 0, \\ \frac{a + bm}{W} [\mathbf{X} + (W - n)\beta]; & \text{gdy } n > 0, \end{cases} \quad (1.29)$$

$$\mathbf{d}^{0,\beta}(\mathbf{X}) = \begin{cases} (a + bm)\beta, & \text{gdy } n = 0, \\ \frac{a + bm}{n}\mathbf{X}, & \text{gdy } n > 0 \end{cases} \quad (1.30)$$

i

$$\mathbf{d}^\beta(\mathbf{X}) = (a + bm)\beta \quad (1.31)$$

kombinacji liniowej $\mathbf{Z}$ są bayesowskie względem rozkładów a priori parametru $\mathbf{p}$ odpowiednio $\pi_{\alpha,\beta}$, $\pi_{\infty,\beta}$, $\pi_{0,\beta}$ i π_β , a ich ryzyka bayesowskie są równe

$$\overline{R}(\pi_{\alpha,\beta}, \mathbf{d}^{\alpha,\beta}) = \begin{cases} \left[(a + bm)^2 - b^2m \frac{W - m}{W - 1} \right] E_{\pi_{\alpha,\beta}}(\mathbf{p}^T C \mathbf{p}) \\ -(a + bm)^2 \beta^T C \beta + b^2m \frac{W - m}{W - 1} \mathbf{c}^T \beta, & \text{gdy } n = 0, \\ z_1 E_{\pi_{\alpha,\beta}}(\mathbf{p}^T C \mathbf{p}) - z_2 \beta^T C \beta + z_3 \mathbf{c}^T \beta, & \text{gdy } n > 0, \end{cases} \quad (1.32)$$

$$\overline{R}(\pi_{\infty,\beta}, \mathbf{d}^{\infty,\beta}) = \begin{cases} \frac{1}{W} [(a + bm)^2 + b^2m(W - m)] (\mathbf{c}^T \beta - \beta^T C \beta), & \text{gdy } n = 0, \\ \frac{(a + bm)^2(W - n)}{W^2} (\mathbf{c}^T \beta - \beta^T C \beta) \\ + \frac{b^2m(W - m)}{W} (\mathbf{c}^T \beta - \beta^T C \beta), & \text{gdy } n > 0, \end{cases} \quad (1.33)$$

$$\overline{R}(\pi_{0,\beta}, \mathbf{d}^{0,\beta}) = \begin{cases} (a + bm)^2 (\mathbf{c}^T \beta - \beta^T C \beta), & \text{gdy } n = 0, \\ 0, & \text{gdy } n > 0 \end{cases} \quad (1.34)$$

i

$$\overline{R}(\pi_\beta, \mathbf{d}^\beta) = b^2m \frac{W - m}{W - 1} (\mathbf{c}^T \beta - \beta^T C \beta), \quad (1.35)$$

gdzie

$$z_1 = (a + bm)^2 \frac{(W - n)^2(\alpha^2 - n) - (W - n)(n + \alpha)^2}{W(W - 1)(n + \alpha)^2} - b^2m \frac{W - m}{W - 1} = z_2 - z_3, \quad (1.36)$$

$$z_2 = \left[(a + bm) \frac{\alpha(W - n)}{W(n + \alpha)} \right]^2, \quad (1.37)$$

$$z_3 = (a + bm)^2 \frac{n(W - n)(W + \alpha)^2}{W^2(W - 1)(n + \alpha)^2} + b^2m \frac{W - m}{W - 1}. \quad (1.38)$$

Dowód: Wobec tego, że funkcja ryzyka dowolnego predyktora $\mathbf{d}(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$, przy funkcji straty określonej wzorem (1.2) wyraża się wzorem (1.14), analogicznie do przypadku rozkładu wielomianowego, otrzymujemy, że szukany predyktor bayesowski kombinacji liniowej $\mathbf{Z}$ jest iloczynem $a + bm$ i estymatora bayesowskiego parametru $\mathbf{p}$. Łatwo sprawdzić (zobacz Amrhein (1995), Lemat 1), że jest on postaci (1.28), jeśli π jest rozkładem Pólya-Eggenbergera $\pi_{\alpha,\beta}$; postaci (1.29), jeśli π jest rozkładem wielomianowym $\pi_{\infty,\beta}$; postaci (1.30), jeśli π jest rozkładem $\pi_{0,\beta}$ i postaci (1.31), jeśli π jest rozkładem jednopunktowym π_β . Uwzględniając, że dla $i = 1, \dots, r$ zachodzą równości

$$E_{\mathbf{p}} X_i = np_i \quad \text{ i } \quad E_{\mathbf{p}} (X_i - np_i)^2 = n \frac{W-n}{W-1} p_i (1-p_i)$$

oraz dla $i \neq j$ zachodzi równość

$$E_{\mathbf{p}} [(X_i - np_i)(X_j - np_j)] = -n \frac{W-n}{W-1} p_i p_j,$$

otrzymujemy, że dla $n > 0$ funkcje ryzyka predyktorów (1.28), (1.29), (1.30) i (1.31) są odpowiednio postaci

$$R(\mathbf{Z}, \mathbf{d}^{\alpha,\beta}) = z_1 \mathbf{p}^T C \mathbf{p} + z_2 \beta^T C \beta + z_3 \mathbf{c}^T \mathbf{p} - 2z_2 \beta^T C \mathbf{p}, \quad (1.39)$$

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\infty,\beta}) &= \left[(a + bm)^2 \frac{(W-n)(W-n-1)}{W(W-1)} - b^2 m \frac{W-m}{W-1} \right] \mathbf{p}^T C \mathbf{p} \\ &+ \frac{(a + bm)^2 (W-n)^2}{W^2} \beta^T C \beta - 2 \frac{(a + bm)^2 (W-n)^2}{W^2} \beta^T C \mathbf{p} \\ &+ \left[\frac{(a + bm)^2 (W-n)n}{W^2(W-1)} + \frac{b^2 m (W-m)}{W-1} \right] \mathbf{c}^T \mathbf{p}, \end{aligned}$$

$$R(\mathbf{Z}, \mathbf{d}^{0,\beta}) = \left[\frac{(a + bm)^2}{n} \frac{W-n}{W-1} + b^2 m \frac{W-m}{W-1} \right] (\mathbf{c}^T \mathbf{p} - \mathbf{p}^T C \mathbf{p})$$

i

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^\beta) &= \left[(a + bm)^2 - b^2 m \frac{W-m}{W-1} \right] \mathbf{p}^T C \mathbf{p} - 2(a + bm)^2 \beta^T C \mathbf{p} \\ &+ (a + bm)^2 \beta^T C \beta + b^2 m \frac{W-m}{W-1} \mathbf{c}^T \mathbf{p}, \end{aligned} \quad (1.40)$$

gdzie z_1 , z_2 , i z_3 określone są wzorami (1.36), (1.37) i (1.38). Natomiast dla $n = 0$ funkcje ryzyka predyktorów (1.28), (1.29), (1.30) i (1.31) są takiej samej postaci i są określone wzorem (1.40). Stąd łatwo można pokazać, że ryzyka bayesowskie predyktorów (1.28), (1.29), (1.30) i (1.31) są odpowiednio postaci (1.32), (1.33), (1.34) i (1.35). $\square$

Następujące twierdzenie uogólnia wyniki uzyskane przez Trybułę (1958) i Wilczyńskiego (1985).

Twierdzenie 1.4 Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określonej. Wówczas predyktor $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ określony przez

$$1^0 \quad \mathbf{d}_0(\mathbf{X}) = \begin{cases} (a + bm)\beta_0, & \text{gdy } n = 0, \\ (a + bm) \left[\frac{W + \alpha_0}{W(n + \alpha_0)} \mathbf{X} + \frac{\alpha_0(W - n)}{W(n + \alpha_0)} \beta_0 \right], & \text{gdy } n > 0, \end{cases} \quad (1.41)$$

$$\text{gdy } (a + bm)^2(W - n)(W - n - 1) - b^2m(W - m)W > 0, \quad (1.42)$$

$$2^0 \quad \mathbf{d}_0(\mathbf{X}) = \begin{cases} (a + bm)\beta_0, & \text{gdy } n = 0, \\ \frac{a + bm}{W} [\mathbf{X} + (W - n)\beta_0], & \text{gdy } n > 0, \end{cases} \quad (1.43)$$

$$\text{gdy } (a + bm)^2(W - n)(W - n - 1) - b^2m(W - m)W = 0 \quad (1.44)$$

$$3^0 \quad \mathbf{d}_0(\mathbf{X}) = (a + bm)\beta_0, \quad (1.45)$$

$$\text{gdy } (a + bm)^2 - b^2m \frac{W - m}{W - 1} \leq 0 \quad (1.46)$$

gdzie

$$\alpha_0 = \frac{n[(W - n)(a + bm)^2 + b^2mW(W - m)] + |a + bm|(W - n)\sqrt{\Delta}}{(a + bm)^2(W - n)(W - n - 1) - b^2mW(W - m)}, \quad (1.47)$$

$$\Delta = (a + bm)^2(W - n)(W - 1)n + b^2m(W - m)(n - 1)Wn$$

i β_0 jest wektorem spełniającym warunek (1.20), jest minimaksowy.

Dowód: Rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\alpha, \beta}(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori Pólya-Eggenbergera $\pi_{\alpha, \beta}$ parametru $\mathbf{p}$, czyli predyktory postaci (1.28). Funkcja ryzyka predyktora $\mathbf{d}^{\alpha, \beta}(\mathbf{X})$ określonego wzorem (1.28) jest postaci (1.39). Wobec tego, że z_1 określone wzorem (1.36) przyjmuje wartość zero wtedy i tylko wtedy, gdy

$$\begin{aligned} & [(a + bm)^2(W - n)(W - n - 1) - b^2m(W - m)W] \alpha^2 \\ & - 2[(a + bm)^2(W - n)n - b^2m(W - m)W] \alpha \\ & - (a + bm)^2(W - n)Wn - b^2m(W - m)Wn^2 = 0, \end{aligned} \quad (1.48)$$

otrzymujemy, że jeśli spełniony jest warunek (1.42), to istnieje $\alpha_0 > 0$, dla którego $z_1 = 0$. Łatwo można sprawdzić, że α_0 jest postaci (1.47). Oznaczmy przez $R_0(\beta, \mathbf{p}) =$

$R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta})$. Wobec tego, że dla $\alpha = \alpha_0$, mamy $z_1 = 0$ i $z_2 = z_3$, więc funkcja ryzyka $R_0(\beta, \mathbf{p})$ sprowadza się do następującej postaci

$$R_0(\beta, \mathbf{p}) = z_2(\beta^T C \beta - 2\beta^T C \mathbf{p} + \mathbf{c}^T \mathbf{p}). \quad (1.49)$$

Zauważmy, że funkcja $R_0(\cdot, \cdot) : \mathcal{P} \times \mathcal{P} \mapsto \mathbf{R}$ jest wypukła względem β , wklęsła względem $\mathbf{p}$, ciągła względem $(\beta, \mathbf{p})$, i $\mathcal{P}$ jest wypukłym, zwartym podzbiorem $\mathbf{R}^r$. Analogicznie do przypadku rozkładu wielomianowego można pokazać, że (β_0, β_0) , gdzie β_0 jest wektorem spełniającym warunek (1.20), jest punktem siodłowym funkcji $R_0(\beta, \mathbf{p})$. Ponadto,

$$R_0(\beta_0, \beta_0) = \bar{R}(\pi_{\alpha_0, \beta_0}, \mathbf{d}^{\alpha_0, \beta_0}),$$

tak więc w przypadku, gdy spełniony jest warunek (1.42) predyktor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\alpha_0, \beta_0}(\mathbf{X})$ określony wzorem (1.41) jest minimaksowy.

W przypadku, gdy spełniony jest warunek (1.44), rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\infty, \beta}(\mathbf{X})$ względem rozkładu a priori wielomianowego $\pi_{\infty, \beta}$ parametru $\mathbf{p}$. Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$ predyktora $\mathbf{d}^{\infty, \beta}(\mathbf{X})$ jest postaci

$$R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) = (a + bm)^2 \frac{(W - n)^2}{W^2} (\beta^T C \beta - 2\beta^T C \mathbf{p} + \mathbf{c}^T \mathbf{p}).$$

Analogicznie do poprzedniego przypadku, możemy pokazać, że predyktor określony wzorem (1.43) jest minimaksowy.

Podobnie, gdy spełniony jest warunek (1.46), funkcja ryzyka bayesowskiego predyktora $\mathbf{d}^\beta(\mathbf{X})$ względem rozkładu a priori π_β parametru $\mathbf{p}$, określona wzorem (1.40) jest wypukła względem β , wklęsła względem $\mathbf{p}$ oraz ciągła względem $(\beta, \mathbf{p})$. Łatwo można pokazać, że (β_0, β_0) , gdzie β_0 jest wektorem spełniającym warunek (1.20), jest punktem siodłowym funkcji $R(\mathbf{Z}, \mathbf{d}^\beta)$. Ponadto

$$\sup_{\mathbf{p} \in \mathcal{P}} R(\mathbf{Z}, \mathbf{d}^{\beta_0}) = \bar{R}(\pi_{\beta_0}, \mathbf{d}^{\beta_0}),$$

tak więc predyktor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\beta_0}(\mathbf{X})$ określony wzorem (1.45) jest minimaksowy. $\square$

Uwaga 1. Otrzymany predyktor minimaksowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z} = a\mathbf{p} + b\mathbf{Y}$ jest kombinacją liniową minimaksowego estymatora parametru $\mathbf{p}$ i minimaksowego predyktora zmiennej losowej $\mathbf{Y}$, o współczynnikach zależnych od W, a, b, m i n .

1.3 Przykłady

Przykład 1 (przypadek rozkładu wielomianowego). Załóżmy, że każde spośród n ($n \geq 1$) obserwowanych urządzeń, pracujących niezależnie, może ulec awarii z powodu jednej z r

możliwych przyczyn. Wówczas obserwowane wartości $X_i, i = 1, \dots, r$, reprezentują liczby urządzeń, które uległy awarii z powodu i -tej przyczyny. Celem naszym jest predykcja wartości $\mathbf{Z} = (Z_1, \dots, Z_r)^T$, $Z_i = ap_i + Y_i$, gdzie Y_i jest nieznaną liczbą awarii z i -tej przyczyny m pracujących urządzeń, natomiast ap_i jest wartością oczekiwaną liczby awarii z i -tej przyczyny a urządzeń, które będą podłączone do pracy. Korzystając z Twierdzenia 1.3 minimaksowy predyktor $\mathbf{d}_0(\mathbf{X})$ szukanej wielkości $\mathbf{Z}$, przy funkcji straty (1.2) z macierzą $C = I$, jest postaci

$$\mathbf{d}_0(\mathbf{X}) = (a + m) \left(\frac{1}{n + \alpha_0} \mathbf{X} + \frac{\alpha_0}{n + \alpha_0} \beta_0 \right),$$

gdzie

$$\alpha_0 = \frac{nm + |a + m| \sqrt{(a + m)^2 n + mn(n - 1)}}{(a + m)^2 - m}$$

$$\text{ i } \beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r} \right)^T.$$

Przykład 2 (przypadek rozkładu wielowymiarowego hipergeometrycznego). Partia W wyrobów podlega statystycznej kontroli jakości. Wartości $W_i, i = 1, \dots, r$, reprezentują nieznaną liczbę elementów i -tej jakości w danej partii. Na podstawie obserwacji $X_i, i = 1, \dots, r$, liczby wyrobów i -tej jakości w grupie n ($n \geq 1$) wyrobów wybranych bez zwracania, chcemy przewidzieć liczby wyrobów i -tej jakości w grupie m wyrobów, które zamierzamy wybrać z danej partii lub chcemy oszacować różnice tych liczb w porównaniu z wartością oczekiwaną liczby wyrobów i -tej jakości, gdybyśmy wybrali a wyrobów. W pierwszym przypadku chcemy znaleźć wartość predyktora $\mathbf{Z} = \mathbf{Y}$ ($a = 0, b = 1$; zwracamy uwagę, że nawet ten szczególny przypadek Twierdzenia 1.4 nie był dotychczas rozpatrywany), w drugim przypadku – wartość kombinacji liniowej $\mathbf{Z} = \mathbf{Y} - a\mathbf{p}$. Twierdzenie 1.4 możemy zastosować również wówczas, gdy chcemy przewidzieć wyniki wyborów lub różnice w wyniku wyborów, gdy do głosowania zgłosi się m uprawnionych a wartością oczekiwaną wyniku wyborów, gdyby głosowało a uprawnionych. Wartości $W_i, i = 1, \dots, r$, reprezentują w tym przypadku nieznaną liczbę osób popierających i -tego kandydata, W jest znaną liczbą osób uprawnionych do głosowania, m liczbą osób, które wezmą udział w wyborach. Korzystając z Twierdzenia 1.4 minimaksowy predyktor kombinacji liniowej $\mathbf{Z} = \mathbf{Y} - a\mathbf{p}$, przy funkcji straty (1.2) z macierzą $C = I$, jest postaci

$$1^0 \quad \mathbf{d}_0(\mathbf{X}) = (m - a) \left[\frac{W + \alpha_0}{W(n + \alpha_0)} \mathbf{X} + \frac{\alpha_0(W - n)}{W(n + \alpha_0)} \beta_0 \right],$$

gdzie $(m - a)^2(W - n)(W - n - 1) - m(W - m)W > 0$,

$$2^0 \quad \mathbf{d}_0(\mathbf{X}) = \frac{m-a}{W} [\mathbf{X} + (W-n)\beta_0],$$

$$\text{gdy } (m-a)^2(W-n)(W-n-1) - m(W-m)W = 0,$$

$$3^0 \quad \mathbf{d}_0(\mathbf{X}) = (m-a)\beta_0,$$

$$\text{gdy } (m-a)^2 - m\frac{W-m}{W-1} \leq 0,$$

gdzie

$$\alpha_0 = \frac{n[(W-n)(m-a)^2 + mW(W-m)] + |m-a|(W-n)\sqrt{\Delta}}{(m-a)^2(W-n)(W-n-1) - mW(W-m)},$$

$$\Delta = (m-a)^2(W-n)(W-1)n + m(W-m)(n-1)Wn$$

$$\text{i } \beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r}\right)^T.$$

Rozdział 2

Minimaksowa estymacja i predykcja dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego przy losowym rozmiarze próby

Rozdział ten poświęcony jest również minimaksowej predykcji kombinacji liniowej $\mathbf{Z}$ określonej wzorem (1.1), przy założeniu, że funkcja straty jest postaci (1.2), dla rozkładu wielomianowego i wielowymiarowego hipergeometrycznego, z tą jednak różnicą, że rozmiar próby jest zmienną losową o znanym rozkładzie niezależnym od nieznanego parametru. Rozpatrywanie przypadku losowego rozmiaru próby jest zasadne, gdyż w wielu praktycznych problemach nie jest możliwe przyjęcie ustalonego rozmiaru próby. Przykładem mogą być następujące dwie sytuacje: interesujące nas jednostki stanowią podgrupę w bazie danych, którą dysponujemy; wylosowana do próby jednostka odmówi nam udzielenia odpowiedzi lub będzie nieobecna w chwili przeprowadzania badania. Pokazuje się, że wyprowadzone w Rozdziale 1 predyktory kombinacji liniowej $\mathbf{Z}$ w przypadku ustalonego rozmiaru próby nie są minimaksowe, a nawet dopuszczalne, dla szerokiej klasy wartości współczynników a i b , parametrów n , m i W oraz rozkładów rozmiaru próby, przy losowym rozmiarze próby. Stanowi to przykład paradoksu, który po raz pierwszy zauważył Brown (1990) w przypadku dopuszczalności estymatora stałego czynnika w wielowymiarowej liniowej regresji. A mianowicie pokazał on, że ten estymator zależy od macierzy eksperymentu, która jest statystyką dodatkową (Lehmann (1986), Rozdział 10). Następnie He

(1990) pokazał, że w przypadku kwadratowej funkcji straty, estymator minimaksowy parametru rozkładu wielomianowego przy ustalonym rozmiarze próby, podany przez Steinhausa (1957), nie jest estymatorem minimaksowym, a nawet dopuszczalnym (poza pewnymi trywialnymi przypadkami), gdy rozmiar próby jest losowy. Również Amrhein (1995) wyprowadził postać minimaksowego estymatora parametru rozkładu wielowymiarowego hipergeometrycznego przy losowym rozmiarze próby i podobnie jak dla rozkładu wielomianowego różni się on od uzyskanego przez Trybułę (1958) estymatora, przy ustalonym rozmiarze próby (poza pewnymi trywialnymi przypadkami). Uzyskane w tym rozdziale wyniki stanowią uogólnienie rezultatów He (1990) i Amrhaina (1995).

2.1 Minimaksowa estymacja i predykcja dla rozkładu wielomianowego

Niech $\mathbf{X}$ będzie obserwacją o losowym rozmiarze N taką, że $\mathbf{X} \mid N = n$, $n = 0, 1, \dots$, ma rozkład wielomianowy o funkcji gęstości postaci (1.3) z parametrami n i $\mathbf{p}$, gdzie $\mathbf{p} = (p_1, \dots, p_r)^T$ nie jest znane. Zakładamy, że rozkład N jest znany i nie zależy od nieznanego parametru $\mathbf{p}$. Oznaczmy $f(n) = P(N = n)$. Następujące twierdzenie podaje postać minimaksowego predyktora $\mathbf{d}^*(\mathbf{X}, N) = (d_1^*(\mathbf{X}, N), \dots, d_r^*(\mathbf{X}, N))^T$ kombinacji liniowej $\mathbf{Z}$ określonej wzorem (1.1), na podstawie obserwacji $(\mathbf{X}, N)$, gdy funkcja straty jest postaci (1.2), to znaczy predyktora $\mathbf{d}^*(\mathbf{X}, N)$, dla którego

$$\sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{Z}, \mathbf{d}^*) = \inf_{\mathbf{d} \in \mathcal{D}} \sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{Z}, \mathbf{d}),$$

gdzie $R(\mathbf{Z}, \mathbf{d}) = E^N \{E_{\mathbf{p}} [L(\mathbf{Z}, \mathbf{d}) \mid N]\}$.

Twierdzenie 2.1 *Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określą. Wówczas predyktor $\mathbf{d}^*(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ określony przez*

$$1^0 \quad \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a + bm)\beta_0, & \text{gdy } N = 0, \\ (a + bm) \left(\frac{1}{N + \alpha^*} \mathbf{X} + \frac{\alpha^*}{N + \alpha^*} \beta_0 \right), & \text{gdy } N > 0, \end{cases}$$

$$\text{gdy } (a + bm)^2 \left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i} \right] - b^2 m < 0 \quad \text{ i } \quad (a + bm)^2 - b^2 m > 0, \quad (2.1)$$

$$2^0 \quad \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a + bm)\beta_0, & \text{gdy } N = 0, \\ \frac{a + bm}{N} \mathbf{X}, & \text{gdy } N > 0, \end{cases} \quad (2.2)$$

$$\text{gdy } (a + bm)^2 \left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i} \right] - b^2 m \geq 0, \quad (2.3)$$

$$\begin{aligned} 3^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = (a + bm)\beta_0, \\ & \text{gdy } (a + bm)^2 - b^2 m \leq 0, \end{aligned} \quad (2.4)$$

gdzie α^* jest rozwiązaniem równania

$$(a + bm)^2 E^N \left[\frac{\alpha^2 - N}{(N + \alpha)^2} \right] - b^2 m = 0$$

oraz β_0 jest wektorem spełniającym warunek (1.20), jest minimaksowy.

Dowód: Używając tych samych argumentów jak w pracy He (1990), można pokazać, że predyktory bayesowskie kombinacji liniowej $\mathbf{Z}$ postaci (1.5), (1.6) i (1.7) względem rozkładu a priori parametru $\mathbf{p}$ odpowiednio $\pi_{\alpha,\beta}$, $\pi_{0,\beta}$ i $\pi_{\infty,\beta}$ w przypadku ustalonego rozmiaru próby, pozostają predyktorami bayesowskimi kombinacji liniowej $\mathbf{Z}$, gdy rozmiar próby jest losowy.

W przypadku, gdy spełniony jest warunek (2.1) rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\alpha,\beta}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori $\pi_{\alpha,\beta}$ Dirichleta parametru $\mathbf{p}$, czyli predyktory określone wzorem (1.5). Funkcja ryzyka $R(\mathbf{d}^{\alpha,\beta}, \mathbf{Z})$ predyktora $\mathbf{d}^{\alpha,\beta}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\alpha,\beta}) &= E^N \{ E^{\mathbf{X}} [L(\mathbf{d}^{\alpha,\beta}, \mathbf{Z}) \mid N] \} \\ &= w_1 \mathbf{p}^T C \mathbf{p} + w_2 \beta^T C \beta + w_3 \mathbf{c}^T \mathbf{p} - 2w_2 \beta^T C \mathbf{p}, \end{aligned}$$

gdzie $\mathbf{c} = (c_{11}, \dots, c_{rr})^T$,

$$w_1 = (a + bm)^2 E^N \left[\frac{\alpha^2 - N}{(N + \alpha)^2} \right] - b^2 m = w_2 - w_3,$$

$$w_2 = (a + bm)^2 E^N \left[\frac{\alpha^2}{(N + \alpha)^2} \right],$$

$$w_3 = (a + bm)^2 E^N \left[\frac{N}{(N + \alpha)^2} \right] + b^2 m.$$

Zauważmy, że

$$\lim_{\alpha \rightarrow 0} w_1 = (a + bm)^2 \left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i} \right] - b^2 m$$

oraz

$$\lim_{\alpha \rightarrow \infty} w_1 = (a + bm)^2 - b^2m.$$

Jeżeli spełniony jest warunek (2.1), to pierwsza z granic jest ujemna, a druga dodatnia, czyli z twierdzenia o wartości średniej wynika istnienie $\alpha^* > 0$, dla którego $w_1 = 0$. Oznaczmy $R^*(\beta, \mathbf{p}) = R(\mathbf{Z}, \mathbf{d}^{\alpha^*, \beta})$. Ponieważ dla $\alpha = \alpha^*$, mamy $w_2 = w_3$, więc funkcja ryzyka $R^*(\beta, \mathbf{p})$ sprowadza się do następującej postaci

$$R^*(\beta, \mathbf{p}) = w_2(\beta^T C \beta + \mathbf{c}^T \mathbf{p} - 2\beta^T C \mathbf{p}).$$

Funkcja ta różni się od funkcji ryzyka $R_0(\beta, \mathbf{p})$ predyktora bayesowskiego $\mathbf{d}^{\alpha_0, \beta}(\mathbf{X})$ w przypadku ustalonego rozmiaru próby, określonej wzorem (1.21) tylko wartością współczynnika w_2 . Stąd dalszą część dowodu w tym przypadku można przeprowadzić analogicznie do odpowiedniej części dowodu Twierdzenia 1.3.

W przypadku, gdy spełniony jest warunek (2.3), rozpatrzmy predyktory bayesowskie $\mathbf{d}^{0, \beta}(\mathbf{X}, N)$ względem rozkładu a priori $\pi_{0, \beta}$ parametru $\mathbf{p}$, czyli predyktory określone wzorem (1.6). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{0, \beta})$ rozpatrywanego predyktora $\mathbf{d}^{0, \beta}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{0, \beta}) &= \left\{ (a + bm)^2 \left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i} \right] - b^2m \right\} \mathbf{p}^T C \mathbf{p} \\ &\quad + (a + bm)^2 f(0) \beta^T C \beta - 2(a + bm)^2 f(0) \beta^T C \mathbf{p} \\ &\quad + \left[(a + bm)^2 \sum_{i=1}^{\infty} \frac{f(i)}{i} + b^2m \right] \mathbf{c}^T \mathbf{p}. \end{aligned} \quad (2.5)$$

Oznaczmy $R_0(\beta, \mathbf{p}) = R(\mathbf{Z}, \mathbf{d}^{0, \beta})$. W przypadku, gdy spełniony jest warunek (2.3), funkcja $R_0(\beta, \mathbf{p})$ jest funkcją wypukłą ze względu na $\mathbf{p}$. Wobec tego funkcja $R_0(\beta, \cdot)$ przyjmuje wartość maksymalną na zbiorze wypukłym $\mathcal{P}$ w pewnych punktach $\mathbf{e}_j$, $j = 1, \dots, r$, gdzie $\mathbf{e}_1 = (1, 0, \dots, 0)^T$, $\mathbf{e}_2 = (0, 1, \dots, 0)^T$, $\dots$, $\mathbf{e}_r = (0, 0, \dots, 1)^T$. Niech $\beta = \beta_0$, gdzie β_0 jest wektorem spełniającym warunek (1.20). Z Twierdzenia 1.2 wynika istnienie stałej u_0 , dla której zachodzą następujące implikacje

$$\begin{aligned} \text{jeśli } \beta_i^0 &> 0, & \text{to } 2 \sum_{j=1}^r c_{ij} \beta_j^0 - c_{ii} &= u_0, \\ \text{jeśli } \beta_i^0 &= 0, & \text{to } 2 \sum_{j=1}^r c_{ij} \beta_j^0 - c_{ii} &\geq u_0. \end{aligned}$$

Ponieważ

$$R_0(\beta_0, \mathbf{e}_j) = (a + bm)^2 f(0) (c_{jj} - 2\beta_0^T \mathbf{c}_j + \beta_0^T C \beta_0),$$

gdzie $\mathbf{c}_j$ jest j -tą kolumną macierzy C , więc stąd otrzymujemy

$$R_0(\beta_0, \mathbf{e}_j) = R_0(\beta_0, \mathbf{e}_i) = (a + bm)^2 f(0) (\beta_0^T C \beta_0 - u_0) = \sup_{\mathbf{p} \in \mathcal{P}} R_0(\beta_0, \mathbf{p}),$$

gdzie $i, j \in A$ oraz

$$R_0(\beta_0, \mathbf{e}_j) \geq R_0(\beta_0, \mathbf{e}_i),$$

gdzie $j \in A$ i $i \notin A$, gdzie $A = \{i_1, \dots, i_k\} \subset \{1, \dots, r\}$ jest takim zbiorem, że $\beta_i^0 > 0$ dla $i \in A$ oraz $\beta_i^0 = 0$ dla $i \notin A$. Wobec tego zbiór tych $\mathbf{p} \in \mathcal{P}$, dla których $R_0(\beta_0, \mathbf{p})$ przyjmuje wartość największą na zbiorze $\mathcal{P}$ jest miary a priori jeden i z Lematu 1.1 wynika, że predyktor $\mathbf{d}^*(\mathbf{X}, N) := \mathbf{d}^{0, \beta_0}(\mathbf{X}, N)$ określony wzorem (2.2) jest minimaksowy.

W przypadku, gdy spełniony jest warunek (2.4), predyktor minimaksowy $\mathbf{d}^{\infty, \beta_0}(\mathbf{X}) = (a + bm)\beta_0$ przy ustalonym rozmiarze próby nie zależy od rozmiaru próby n . Stąd jest on również predyktorem minimaksowym przy losowym rozmiarze próby. $\square$

Uwaga 1. Predyktor minimaksowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ przy ustalonym rozmiarze próby określony wzorem (1.18) i predyktor minimaksowy $\mathbf{d}^*(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ przy losowym rozmiarze próby, podany w Twierdzeniu 2.1, mają taką samą postać wtedy i tylko wtedy, gdy $(a + bm)^2 - b^2m \leq 0$ lub $\max_n \{f(n)\} = 1$.

Uwaga 2. W przypadku, gdy macierz C jest dodatnio określona oraz, gdy spełnione są warunki

$$\begin{cases} \max_n \{f(n)\} < 1, \\ (a + bm)^2 - b^2m > 0, \end{cases} \quad (2.6)$$

predyktor minimaksowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ przy ustalonym rozmiarze próby, określony wzorem (1.18), nie jest predyktorem minimaksowym kombinacji liniowej $\mathbf{Z}$, gdy rozmiar próby jest losowy. Fakt ten wynika stąd, że predyktor minimaksowy $\mathbf{d}^*(\mathbf{X}, N)$ jest jedynym predyktorem bayesowskim względem odpowiedniego rozkładu a priori π parametru $\mathbf{p}$ takim, że $\sup_{\mathbf{p} \in \mathcal{P}} R(\mathbf{Z}, \mathbf{d}^*) = \overline{R}(\pi, \mathbf{d}^*)$, a więc jest jedynym predyktorem minimaksowym (zobacz Lehmann (1983), Twierdzenie 2.1). Ponadto, gdy spełniony jest warunek (2.6), predyktor $\mathbf{d}_0(\mathbf{X}) \neq \mathbf{d}^*(\mathbf{X}, N)$ (zobacz Uwaga 1).

Uwaga 3. W przypadku, gdy macierz C jest dodatnio określona, spełniona jest nierówność $(a + bm)^2 - b^2m > 0$ i nie zachodzi żaden z poniższych warunków

$$\max_n \{f(n)\} = 1 \quad (2.7)$$

$$\begin{cases} \max_n \{f(0) + f(n)\} = 1 \\ b^2 m = 0, \end{cases} \quad (2.8)$$

predyktor minimaxowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ przy ustalonym rozmiarze próby, określony wzorem (1.18), nie jest predyktorem dopuszczalnym kombinacji liniowej $\mathbf{Z}$ w przypadku losowego rozmiaru próby. Rozpatrzmy następujący predyktor bayesowski $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori π_{γ_0, β_0} parametru $\mathbf{p}$ postaci

$$\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N) = (a + bm) \frac{\mathbf{X} + \gamma_0 \beta_0}{N + \gamma_0},$$

gdzie $\gamma_0 > 0$ jest rozwiązaniem równania

$$(a + bm)^2 \sum_{n=1}^{\infty} \frac{\gamma^2 - n}{(n + \gamma)^2} f(n) - b^2 m \sum_{n=1}^{\infty} f(n) = 0,$$

(jeżeli $(a + bm)^2 - b^2 m > 0$, to rozwiązanie to istnieje), β_0 jest wektorem spełniającym warunek (1.20). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\gamma_0, \beta_0})$ predyktora $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\gamma_0, \beta_0}) &= f(0) \{ [(a + bm)^2 - b^2 m] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta_0^T C \beta_0 + b^2 m \mathbf{c}^T \mathbf{p} \\ &\quad - 2(a + bm)^2 \beta_0^T C \mathbf{p} \} \\ &\quad + (a + bm)^2 \sum_{n=1}^{\infty} \frac{\gamma_0^2}{(n + \gamma_0)^2} f(n) (\beta_0^T C \beta_0 + \mathbf{c}^T \mathbf{p} - 2\beta_0^T C \mathbf{p}), \end{aligned}$$

natomiast funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta_0})$ predyktora $\mathbf{d}_0(\mathbf{X}) = \mathbf{d}^{\alpha_0, \beta_0}(\mathbf{X})$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta_0}) &= f(0) \{ [(a + bm)^2 - b^2 m] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta_0^T C \beta_0 + b^2 m \mathbf{c}^T \mathbf{p} \\ &\quad - 2(a + bm)^2 \beta_0^T C \mathbf{p} \} \\ &\quad + (a + bm)^2 \sum_{n=1}^{\infty} \frac{\alpha_0^2}{(n + \alpha_0)^2} f(n) (\beta_0^T C \beta_0 + \mathbf{c}^T \mathbf{p} - 2\beta_0^T C \mathbf{p}). \end{aligned}$$

W związku z tym, że predyktor $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ jest dopuszczalny, jako jedyny predyktor bayesowski kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori π_{γ_0, β_0} parametru $\mathbf{p}$ (ingeruje tu założenie dodatniej określoności macierzy C), zachodzi następująca nierówność

$$\sum_{n=1}^{\infty} \frac{\gamma_0}{(n + \gamma_0)^2} f(n) \leq \sum_{n=1}^{\infty} \frac{\alpha_0}{(n + \alpha_0)^2} f(n). \quad (2.9)$$

W nierówności (2.9) równość zachodzi wtedy i tylko wtedy, gdy $\gamma_0 = \alpha_0$, czyli, gdy zachodzi jeden z warunków (2.7) lub (2.8). W przeciwnym razie istniałyby dwa predyktory o tej samej funkcji ryzyka, czyli ich kombinacja liniowa dominowałaby każdy z nich, wbrew dopuszczalności predyktora $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$.

2.2 Minimaksowa estymacja i predykcja dla wielowymiarowego rozkładu hipergeometrycznego

Niech $\mathbf{X}$ będzie obserwacją o losowym rozmiarze N , taką, że $\mathbf{X} \mid N = n, n = 0, 1, \dots, W$, ma rozkład wielowymiarowy hipergeometryczny o funkcji prawdopodobieństwa postaci (1.26) z parametrami $(W, \mathbf{W}, n)$, gdzie $\mathbf{W} = (W_1, \dots, W_r)^T$ nie jest znane. Zakładamy, że rozkład N jest znany i nie zależy od nieznanego parametru $\mathbf{W}$. Oznaczmy $f(n) = P(N = n)$. Następujące twierdzenie podaje postać minimaksowego predyktora kombinacji liniowej $\mathbf{Z}$ określonej wzorem (1.27), na podstawie obserwacji $(\mathbf{X}, N)$, gdy funkcja straty jest postaci (1.2).

Twierdzenie 2.2 *Niech funkcja straty będzie postaci (1.2) z macierzą C nieujemnie określą. Wówczas predyktor $\mathbf{d}^*(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ określony przez*

$$1^0 \quad \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a + bm)\beta_0, & \text{gdy } N = 0, \\ (a + bm) \left[\frac{W + \alpha^*}{W(N + \alpha^*)} \mathbf{X} + \frac{\alpha^*(W - N)}{W(N + \alpha^*)} \beta_0 \right], & \text{gdy } N > 0, \end{cases}$$

gdy

$$\begin{cases} (a + bm)^2 \left[f(0) - \sum_{i=1}^W \frac{f(i)(W - i)}{i(W - 1)} \right] - b^2 m \frac{W - m}{W - 1} < 0, \\ (a + bm)^2 \sum_{i=1}^W f(i) \frac{(W - i)(W - i - 1)}{W(W - 1)} - b^2 m \frac{(W - m)}{W - 1} > 0, \end{cases} \quad (2.10)$$

$$2^0 \quad \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a + bm)\beta_0, & \text{gdy } N = 0, \\ \frac{a + bm}{W} [\mathbf{X} + (W - N)\beta_0], & \text{gdy } N > 0, \end{cases} \quad (2.11)$$

gdy

$$(a + bm)^2 \sum_{i=1}^W f(i) \frac{(W - i)(W - i - 1)}{W(W - 1)} - b^2 m \frac{(W - m)}{W - 1} = 0, \quad (2.12)$$

$$3^0 \quad \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a + bm)\beta_0, & \text{gdy } N = 0, \\ \frac{a + bm}{N} \mathbf{X}, & \text{gdy } N > 0, \end{cases} \quad (2.13)$$

gdy

$$(a + bm)^2 \left[f(0) - \sum_{i=1}^W \frac{f(i)(W - i)}{i(W - 1)} \right] - b^2 m \frac{W - m}{W - 1} \geq 0, \quad (2.14)$$

$$4^0 \quad \mathbf{d}^*(\mathbf{X}, N) = (a + bm)\beta_0,$$

gdy

$$(a + bm)^2 - b^2 m \frac{W - m}{W - 1} \leq 0, \quad (2.15)$$

gdzie α_0 jest rozwiązaniem równania

$$(a + bm)^2 E^N \left[\frac{(W - N)^2(\alpha^2 - N) - (W - N)(N + \alpha)^2}{W(W - 1)(N + \alpha)^2} \right] - b^2 m \frac{W - m}{W - 1} = 0$$

oraz β_0 jest wektorem spełniającym warunek (1.20), jest minimaksowy.

Dowód: Używając tych samych argumentów jak w pracy He (1990), można pokazać, że predyktory bayesowskie kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori parametru $\mathbf{p}$ $\pi_{\alpha,\beta}$, $\pi_{\infty,\beta}$, $\pi_{0,\beta}$ i π_β , przy ustalonym rozmiarze próby, określone odpowiednio wzorami (1.28), (1.29), (1.30) i (1.31), pozostają predyktorami bayesowskimi kombinacji liniowej $\mathbf{Z}$, gdy rozmiar próby jest losowy.

W przypadku, gdy spełniony jest warunek (2.10) rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\alpha,\beta}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori $\pi_{\alpha,\beta}$ Pòlya-Eggenbergera parametru $\mathbf{p}$, czyli predyktory określone wzorem (1.28). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\alpha,\beta})$ predyktora $\mathbf{d}^{\alpha,\beta}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\alpha,\beta}) &= E^N \{ E^{\mathbf{X}} [L(\mathbf{d}^{\alpha,\beta}, \mathbf{Z}) \mid N] \} \\ &= z_1 \mathbf{p}^T C \mathbf{p} + z_2 \beta^T C \beta + z_3 \mathbf{c}^T \mathbf{p} - 2z_2 \beta^T C \mathbf{p}, \end{aligned}$$

gdzie $\mathbf{c} = (c_{11}, \dots, c_{rr})^T$,

$$\begin{aligned} z_1 &= (a + bm)^2 E^N \left[\frac{(W - N)^2(\alpha^2 - N) - (W - N)(N + \alpha)^2}{W(W - 1)(N + \alpha)^2} \right] - b^2 m \frac{W - m}{W - 1} \\ &= z_2 - z_3, \end{aligned}$$

$$z_2 = (a + bm)^2 E^N \left[\frac{\alpha(W - N)}{W(N + \alpha)} \right]^2,$$

$$z_3 = (a + bm)^2 E^N \left[\frac{N(W - N)(W + \alpha)^2}{W^2(W - 1)(N + \alpha)^2} \right] + b^2 m \frac{W - m}{W - 1}.$$

Zauważmy, że

$$\lim_{\alpha \rightarrow 0} z_1 = (a + bm)^2 \left[f(0) - \sum_{i=1}^W \frac{f(i)(W - i)}{i(W - 1)} \right] - b^2 m \frac{W - m}{W - 1}$$

oraz

$$\lim_{\alpha \rightarrow \infty} z_1 = (a + bm)^2 \sum_{i=1}^W f(i) \frac{(W - i)(W - i - 1)}{W(W - 1)} - b^2 m \frac{W - m}{W - 1}.$$

Jeżeli spełniony jest warunek (2.10), to pierwsza z granic jest ujemna, a druga dodatnia, czyli z twierdzenia o wartości średniej wynika istnienie $\alpha^* > 0$, dla którego $z_1 = 0$. Oznaczmy $R^*(\beta, \mathbf{p}) = R(\mathbf{Z}, \mathbf{d}^{\alpha^*, \beta})$. Ponieważ dla $\alpha = \alpha^*$, mamy $z_2 = z_3$, więc funkcja ryzyka $R^*(\beta, \mathbf{p})$ sprowadza się do następującej postaci

$$R^*(\beta, \mathbf{p}) = z_2(\beta^T C \beta + \mathbf{c}^T \mathbf{p} - 2\beta^T C \mathbf{p}).$$

Funkcja ta różni się od funkcji ryzyka $R_0(\beta, \mathbf{p})$ predyktora $\mathbf{d}^{\alpha_0, \beta}(\mathbf{X})$, określonej wzorem (1.49) tylko wartością współczynnika z_2 . Stąd dalszą część dowodu w tym przypadku można przeprowadzić analogicznie do odpowiedniej części dowodu Twierdzenia 1.4.

W przypadku, gdy spełniony jest warunek (2.12), rozpatrzmy predyktory bayesowskie $\mathbf{d}^{\infty, \beta}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori $\pi_{\infty, \beta}$ parametru $\mathbf{p}$, czyli predyktory określone wzorem (1.29). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$ rozpatrywanego predyktora $\mathbf{d}^{\infty, \beta}(\mathbf{X}, N)$ jest postaci

$$R(\mathbf{Z}, \mathbf{d}^{\infty, \beta}) = \frac{(a + bm)^2}{W^2} E^N [(W - N)^2] (\beta^T C \beta - 2\beta^T C \mathbf{p} + \mathbf{c}^T \mathbf{p}).$$

Analogicznie do poprzedniego przypadku można pokazać, że punkt (β_0, β_0) jest punktem siodłowym funkcji $R(\mathbf{Z}, \mathbf{d}^{\infty, \beta})$. Ponadto

$$R(\mathbf{Z}, \mathbf{d}^{\infty, \beta_0}) = \bar{R}(\pi_{\infty, \beta_0}, \mathbf{d}^{\infty, \beta_0}),$$

czyli na mocy Lematu 1.1 otrzymujemy, że predyktor $\mathbf{d}^*(\mathbf{X}, N) := \mathbf{d}^{\infty, \beta_0}(\mathbf{X}, N)$ określony wzorem (2.11) jest minimaksowy.

W przypadku, gdy spełniony jest warunek (2.14), rozpatrzmy predyktory bayesowskie $\mathbf{d}^{0, \beta}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori $\pi_{0, \beta}$ parametru $\mathbf{p}$, czyli predyktory postaci (1.30). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{0, \beta})$ rozpatrywanego predyktora $\mathbf{d}^{0, \beta}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{0, \beta}) = & \left\{ (a + bm)^2 \left[f(0) - \sum_{i=1}^W \frac{f(i)(W-i)}{i(W-1)} \right] - b^2 m \frac{W-m}{W-1} \right\} \mathbf{p}^T C \mathbf{p} \\ & + (a + bm)^2 f(0) \beta^T C \beta - 2(a + bm)^2 f(0) \beta^T C \mathbf{p} \\ & + \left[(a + bm)^2 \sum_{i=1}^W \frac{f(i)(W-i)}{i(W-1)} + b^2 m \frac{W-m}{W-1} \right] \mathbf{c}^T \mathbf{p}. \end{aligned}$$

Z uwagi na takie same własności powyższej funkcji ryzyka $R(\mathbf{Z}, \mathbf{d}^{0, \beta})$ i funkcji ryzyka określonej wzorem (2.5), dalszą część dowodu w tym przypadku można przeprowadzić analogicznie do odpowiedniej części dowodu Twierdzenia 2.1.

W przypadku, gdy spełniony jest warunek (2.15), minimaksowy predyktor $\mathbf{d}^{\infty, \beta_0}(\mathbf{X}) = (a + bm)\beta_0$ przy ustalonym rozmiarze próby nie zależy od rozmiaru próby n . Stąd jest on również predyktorem minimaksowym przy losowym rozmiarze próby. $\square$

Uwaga 1. W przypadku, gdy macierz C jest dodatnio określona, spełniona jest nierówność

$$(a + bm)^2 \sum_{n=1}^W \frac{(W-n)(W-n-1)}{W(W-1)} f(n) - b^2 m \frac{W-m}{W-1} \sum_{n=1}^W f(n) > 0$$

i nie zachodzi żaden z poniższych warunków

$$\max_n \{f(0) + f(n)\} = 1 \quad (2.16)$$

i

$$\begin{cases} \max_n \{f(0) + f(n) + f(W)\} = 1, \\ b(W-m) = 0, \end{cases} \quad (2.17)$$

predyktor minimaksowy $\mathbf{d}_0(\mathbf{X})$ kombinacji liniowej $\mathbf{Z}$ przy ustalonym rozmiarze próby, określony wzorem (1.18), nie jest predyktorem minimaksowym, a nawet dopuszczalnym, gdy rozmiar próby jest losowy. Rozpatrzmy następujący predyktor bayesowski $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori π_{γ_0, β_0} parametru $\mathbf{p}$ postaci

$$\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N) = (a + bm) \left[\frac{W + \gamma_0}{W(N + \gamma_0)} \mathbf{X} + \frac{\gamma_0(W - N)}{W(N + \gamma_0)} \beta_0 \right],$$

gdzie $\gamma_0 > 0$ jest rozwiązaniem równania

$$(a + bm)^2 \sum_{n=1}^W \frac{(W-n)^2(\gamma^2 - n) - (W-n)(n + \gamma)^2}{W(W-1)(n + \gamma)^2} f(n) - b^2 m \frac{W-m}{W-1} \sum_{n=1}^W f(n) = 0,$$

(jeśli $(a + bm)^2 \sum_{n=1}^W \frac{(W-n)(W-n-1)}{W(W-1)} f(n) - b^2 m \frac{W-m}{W-1} \sum_{n=1}^W f(n) > 0$, to rozwiązanie to istnieje) i β_0 jest wektorem spełniającym równanie (1.20). Funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\gamma_0, \beta_0})$ predyktora $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\gamma_0, \beta_0}) &= f(0) \left\{ \left[(a + bm)^2 - b^2 m \frac{W-m}{W-1} \right] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta_0^T C \beta_0 \right. \\ &\quad \left. + b^2 m \frac{W-m}{W-1} \mathbf{c}^T \mathbf{p} - 2(a + bm)^2 \beta_0^T C \mathbf{p} \right\} \\ &\quad + (a + bm)^2 \sum_{n=1}^W \frac{\gamma_0^2 (W-n)^2}{W^2 (n + \gamma_0)^2} f(n) (\beta_0^T C \beta_0 + \mathbf{c}^T \mathbf{p} - 2\beta_0^T C \mathbf{p}), \end{aligned}$$

natomiast funkcja ryzyka $R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta_0})$ predyktora $\mathbf{d}_0(\mathbf{X}) = \mathbf{d}^{\alpha_0, \beta_0}(\mathbf{X})$ jest postaci

$$\begin{aligned} R(\mathbf{Z}, \mathbf{d}^{\alpha_0, \beta_0}) &= f(0) \left\{ \left[(a + bm)^2 - b^2 m \frac{W-m}{W-1} \right] \mathbf{p}^T C \mathbf{p} + (a + bm)^2 \beta_0^T C \beta_0 \right. \\ &\quad \left. + b^2 m \frac{W-m}{W-1} \mathbf{c}^T \mathbf{p} - 2(a + bm)^2 \beta_0^T C \mathbf{p} \right\} \\ &\quad + (a + bm)^2 \sum_{n=1}^W \frac{\alpha_0^2 (W-n)^2}{W^2 (n + \alpha_0)^2} f(n) (\beta_0^T C \beta_0 + \mathbf{c}^T \mathbf{p} - 2\beta_0^T C \mathbf{p}). \end{aligned}$$

W związku z tym, że predyktor $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$ jest dopuszczalny, jako jedyny predyktor bayesowski kombinacji liniowej $\mathbf{Z}$ względem rozkładu a priori π_{γ_0, β_0} parametru $\mathbf{p}$, zachodzi następująca nierówność

$$\sum_{n=1}^W \frac{\gamma_0(W-n)^2}{W^2(n+\gamma_0)^2} f(n) \leq \sum_{n=1}^W \frac{\alpha_0(W-n)^2}{W^2(n+\alpha_0)^2} f(n). \quad (2.18)$$

W nierówności (2.18) równość zachodzi wtedy i tylko wtedy, gdy $\gamma_0 = \alpha_0$, czyli, gdy zachodzi jeden z warunków (2.16) lub (2.17). W przeciwnym razie istniałyby dwa predyktory o tej samej funkcji ryzyka, czyli ich kombinacja liniowa dominowałaby każdy z nich, wbrew dopuszczalności predyktora $\mathbf{d}^{\gamma_0, \beta_0}(\mathbf{X}, N)$.

2.3 Przykłady

Przykład 1 (przypadek rozkładu wielomianowego). Załóżmy, że w określonym czasie obserwujemy s niezależnie pracujących urządzeń, z których każde może ulec awarii ze znanym prawdopodobieństwem θ . Każdą awarię przypisujemy jednej z r możliwych przyczyn. Wówczas obserwowane wartości $X_i, i = 1, \dots, r$, reprezentują liczby urządzeń, które uległy awarii (w danym czasie) z powodu i -tej przyczyny, N oznacza liczbę tych urządzeń spośród s , które uległy awarii w określonym czasie. Wobec tego rozmiar N próby jest statystyką dodatkową o znanym rozkładzie dwumianowym $b(s, \theta)$ i $f(n) = \binom{s}{n} \theta^n (1-\theta)^{s-n}$. Celem naszym jest predykcja wartości $\mathbf{Z} = (Z_1, \dots, Z_r)^T$, $Z_i = ap_i + Y_i$, gdzie Y_i jest nieznaną liczbą awarii z i -tej przyczyny m pracujących urządzeń, a ap_i jest wartością oczekiwaną liczby awarii z i -tej przyczyny a urządzeń, które będą podłączone do pracy. Korzystając z Twierdzenia 2.1 minimaksowy predyktor szukanej wielkości $\mathbf{Z}$, przy funkcji straty wyrażonej wzorem (1.2) z macierzą $C = I$, jest postaci

$$\begin{aligned} 1^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = (a+m) \left(\frac{1}{N+\alpha^*} \mathbf{X} + \frac{\alpha^*}{N+\alpha^*} \beta_0 \right), \\ & \text{gd}y \\ & (a+m)^2 \left[(1-\theta)^s - \sum_{i=1}^s \frac{1}{i} \binom{s}{i} \theta^i (1-\theta)^{s-i} \right] - m < 0, \\ 2^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} (a+m)\beta_0, & \text{gd}y \ N = 0, \\ \frac{a+m}{N} \mathbf{X}, & \text{gd}y \ N > 0, \end{cases} \\ & \text{gd}y \\ & (a+m)^2 \left[(1-\theta)^s - \sum_{i=1}^s \frac{1}{i} \binom{s}{i} \theta^i (1-\theta)^{s-i} \right] - m \geq 0, \end{aligned}$$

gdzie α^* jest rozwiązaniem równania

$$(a + m)^2 \sum_{i=0}^s \left[\frac{\alpha^2 - i}{(i + \alpha)^2} \right] \binom{s}{i} \theta^i (1 - \theta)^{s-i} - m = 0$$

$$\text{ i } \beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r} \right)^T.$$

Przykład 2 (przypadek rozkładu wielowymiarowego hipergeometrycznego). Załóżmy, że każda spośród s ($s \leq W$) ankietowanych osób udzieli nam odpowiedzi na pytanie ankietowe (na przykład o preferencje wyborcze) ze znanym prawdopodobieństwem θ . Wówczas obserwowane wartości X_i , $i = 1, \dots, r$, reprezentują liczby osób, które wybrały i -tą opcję odpowiedzi (na przykład będą głosowały na i -tego kandydata), N oznacza liczbę tych osób spośród s , które udzieliły nam odpowiedzi. W tym przypadku, podobnie jak w poprzednim przykładzie, rozmiar N próby jest statystyką dodatkową o znanym rozkładzie dwumianowym $b(s, \theta)$ i $f(n) = \binom{s}{n} \theta^n (1 - \theta)^{s-n}$. Naszym celem jest predykcja zmiennej losowej $\mathbf{Y} = (Y_1, \dots, Y_r)^T$, gdzie Y_i , $i = 1, \dots, r$, jest liczbą tych osób spośród m , które wybiorą i -tą opcję odpowiedzi (na przykład Y_i może być liczbą głosów, które uzyska i -ty kandydat, gdy w wyborach weźmie udział m uprawnionych do głosowania). Korzystając z Twierdzenia 2.2 minimaksowy predyktor szukanej wielkości $\mathbf{Y}$, przy funkcji straty wyrażonej wzorem (1.2) z macierzą $C = I$, jest postaci

$$\begin{aligned} 1^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = m \left[\frac{W + \alpha^*}{W(N + \alpha^*)} \mathbf{X} + \frac{\alpha^*(W - N)}{W(N + \alpha^*)} \beta_0 \right], \\ \text{gdym} \quad & \begin{cases} m^2 \left[(1 - \theta)^s - \sum_{i=1}^s \binom{s}{i} \theta^i (1 - \theta)^{s-i} \frac{(W - i)}{i(W - 1)} \right] - m \frac{W - m}{W - 1} < 0 \\ m^2 \sum_{i=1}^s \binom{s}{i} \theta^i (1 - \theta)^{s-i} \frac{(W - i)(W - i - 1)}{W(W - 1)} - m \frac{W - m}{W - 1} > 0, \end{cases} \end{aligned}$$

$$\begin{aligned} 2^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = \frac{m}{W} [\mathbf{X} + (W - N) \beta_0], \\ \text{gdym} \quad & m^2 \sum_{i=1}^s \binom{s}{i} \theta^i (1 - \theta)^{s-i} \frac{(W - i)(W - i - 1)}{W(W - 1)} - m \frac{W - m}{W - 1} = 0, \end{aligned}$$

$$\begin{aligned} 3^0 \quad & \mathbf{d}^*(\mathbf{X}, N) = \begin{cases} m \beta_0, & \text{gdym } N = 0, \\ \frac{m}{N} \mathbf{X}, & \text{gdym } N > 0, \end{cases} \\ \text{gdym} \quad & m^2 \left[(1 - \theta)^s - \sum_{i=1}^s \binom{s}{i} \theta^i (1 - \theta)^{s-i} \frac{(W - i)}{i(W - 1)} \right] - m \frac{W - m}{W - 1} \geq 0, \end{aligned}$$

$$4^0 \quad \mathbf{d}^*(\mathbf{X}, N) = m\beta_0,$$

gdzie

$$m^2 - m \frac{W - m}{W - 1} \leq 0,$$

gdzie α^* jest rozwiązaniem równania

$$m^2 \sum_{i=0}^s \frac{(W - i)^2(\alpha^2 - i) - (W - i)(i + \alpha)^2}{W(W - 1)(i + \alpha)^2} \binom{s}{i} \theta^i (1 - \theta)^{s-i} - m \frac{W - m}{W - 1} = 0$$

oraz $\beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r}\right)^T$.

Przykład 3 (przypadek rozkładu wielomianowego). Rozpatrzmy grupę urządzeń ulegających awarii zgodnie z opisem w Przykładzie 1, Rozdział 1.3. Załóżmy, że w rozpatrywanym układzie pracuje W urządzeń typu A oraz $K - W$ urządzeń typu B . Interesuje nas awaryjność w podgrupie urządzeń typu A . Do statystycznej kontroli wybrano bez zwracania k urządzeń z całego układu. Wówczas rozmiar N próby związany z podgrupą urządzeń typu A jest zmienną losową o rozkładzie hipergeometrycznym, to znaczy zmienną losową o funkcji prawdopodobieństwa postaci

$$P(N = n) = \frac{\binom{W}{n} \binom{K - W}{k - n}}{\binom{K}{k}},$$

$n = 0, 1, \dots, \min\{k, W\}$. Korzystając z Twierdzenia 2.1 minimaksowy estymator $\mathbf{d}^*(\mathbf{X}, N)$ parametru $\mathbf{p}$, przy funkcji straty określonej wzorem (1.2) z macierzą $C = I$, jest postaci

$$\mathbf{d}^*(\mathbf{X}, N) = \frac{1}{N + \alpha^*} \mathbf{X} + \frac{\alpha^*}{N + \alpha^*} \beta_0,$$

gdzie α^* jest rozwiązaniem równania

$$E^N \left[\frac{\alpha^2 - N}{(N + \alpha)^2} \right] = 0$$

oraz $\beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r}\right)^T$.

Przykład 4 (przypadek rozkładu wielowymiarowego hipergeometrycznego). Rozpatrzmy populację K gospodarstw domowych, w której znajduje się W gospodarstw spełniających pewne warunki, na przykład gospodarstw emerytów i rencistów. Interesują nas frakcje gospodarstw danej podpopulacji, które możemy zaliczyć do jednej z r kategorii, czyli chcemy estymować parametr $\mathbf{p} = (W_1/W, \dots, W_r/W)^T$, gdzie $W_i, i = 1, \dots, r$,

stanowi nieznana liczbę gospodarstw danej podgrupy, które zaliczamy do i -tej kategorii. W celu oszacowania wartości nieznanego parametru $\mathbf{p}$ pobieramy bez zwracania k -elementową próbę gospodarstw z całej populacji. Wówczas rozmiar N próby związany z interesującą nas podgrupą gospodarstw domowych jest zmienną losową o rozkładzie hipergeometrycznym, to znaczy zmienną losową o funkcji prawdopodobieństwa postaci

$$P(N = n) = \frac{\binom{W}{n} \binom{K - W}{k - n}}{\binom{K}{k}},$$

$n = 0, 1, \dots, \min\{k, W\}$. Korzystając z Twierdzenia 2.2 minimaksowy estymator $\mathbf{d}^*(\mathbf{X}, N)$ parametru $\mathbf{p}$, przy funkcji straty określonej wzorem (1.2) z macierzą $C = I$, jest postaci

$$\mathbf{d}^*(\mathbf{X}, N) = \frac{W + \alpha^*}{W(N + \alpha^*)} \mathbf{X} + \frac{\alpha^*(W - N)}{W(N + \alpha^*)} \beta_0,$$

gdzie α^* jest rozwiązaniem równania

$$E^N \left[\frac{(W - N)^2(\alpha^2 - N) - (W - N)(N + \alpha)^2}{W(W - 1)(N + \alpha)^2} \right] = 0$$

oraz $\beta_0 = \left(\frac{1}{r}, \dots, \frac{1}{r} \right)^T$.

Rozdział 3

Estymacja parametru rozkładu wielomianowego na podstawie danych otrzymywanych w chwilach losowych

Rozdział ten poświęcony jest estymacji parametru $\mathbf{p} \in \mathcal{P} = \{(p_1, \dots, p_r)^T : p_i \geq 0, i = 1, \dots, r, p_1 + \dots + p_r = 1\}$ rozkładu wielomianowego o funkcji prawdopodobieństwa określonej wzorem (1.3) w przypadku, gdy dane otrzymywane są w chwilach losowych.

Rozważa się następujący model. Niech $\mathbf{Y}(i) = (Y_1(i), \dots, Y_r(i))^T$, $i = 1, \dots, n$, będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie wielomianowym $\mathcal{W}(1, \mathbf{p})$, to znaczy $P(\mathbf{Y}(i) = \mathbf{e}_j) = p_j$, $j = 1, \dots, r$, $i = 1, \dots, n$, gdzie $\mathbf{e}_1 = (1, 0, \dots, 0)^T, \dots, \mathbf{e}_r = (0, \dots, 0, 1)^T$. Zakłada się, że $\mathbf{Y}(i)$ jest obserwacją uzyskaną w chwili t_i , $i = 1, \dots, n$, gdzie $t_1, \dots, t_n$ są statystykami pozytywnymi dodatnich zmiennych losowych $U_1, \dots, U_n$, które są niezależne od $\mathbf{Y}(1), \dots, \mathbf{Y}(n)$.

Niech

$$k(t) = \sum_{i=1}^n I_{[0,t]}(U_i)$$

oznacza liczbę obserwacji uzyskanych do chwili $t \geq 0$, a $\mathcal{F}_t = \sigma\{k(s), s \leq t, \mathbf{Y}(1), \dots, \mathbf{Y}(k(t))\}$ będzie informacją dostępną w chwili t .

Jako przykład rozważa się grupę n urządzeń, które mogą ulec uszkodzeniu z powodu jednej z r możliwych przyczyn. Niech U_i ($i = 1, \dots, n$) oznacza czas efektywnego działania i -tego urządzenia i niech $k(t)$ będzie liczbą tych spośród danej liczby n urządzeń, które uległy uszkodzeniu przed czasem t . Obserwowana wartość $\mathbf{Y}(i) = \mathbf{e}_j$, $1 \leq j \leq r$, oznacza,

że w chwili $t_i = t_i(U_1, \dots, U_n)$ urządzenie uległo uszkodzeniu z powodu j -tej przyczyny. Inne przykłady można podać z dziedziny badań rynku, w przypadku, gdy chcemy ustalić udział konkurencyjnych firm w rynku lub zaplanować asortyment produkcji (klienci lub zamówienia przychodzą w chwilach losowych) lub z medycyny, kiedy interesuje nas częstość pojawiania się pewnych grup chorób (dane możliwe są do uzyskania w przypadku, gdy zgłosi się do nas pacjent).

Niech

$$L_t(\mathbf{p}, \mathbf{d}) = L(\mathbf{p}, \mathbf{d}) + c(t).$$

będzie stratą jaką ponosimy, gdy estymujemy nieznaną wartość parametru $\mathbf{p}$ na podstawie obserwacji uzyskanych do chwili t . Zakłada się, że $L(\mathbf{p}, \mathbf{d})$ – strata związana z błędem estymacji, gdy podejmiemy decyzję $\mathbf{d}$, a $\mathbf{p}$ jest prawdziwą wartością – jest postaci

$$L(\mathbf{p}, \mathbf{d}) = \sum_{i=1}^r \frac{(d_i - p_i)^2}{p_i}. \quad (3.1)$$

Funkcja $c(t)$ wyraża koszt obserwacji prowadzonych do chwili t . Zakładamy, że jest ona różniczkowalna, rosnąca, wypukła oraz taka, że $c(0) = 0$.

W problemie estymacji nieznanego parametru $\mathbf{p}$ rozkładu wielomianowego rozważa się sekwencyjne procedury postaci $\delta = (\tau, \mathbf{d}(\tau))$, gdzie τ jest czasem zatrzymania względem σ -algebry $\mathcal{F}_t$, $t \geq 0$, a $\mathbf{d}(\tau)$ jest $\mathcal{F}_\tau$ -mierzalną funkcją.

W pierwszym podrozdziale wprowadzono bayesowskie i minimaksowe procedury estymacji parametru $\mathbf{p}$ rozkładu wielomianowego na podstawie obserwacji $\mathbf{Y}(1), \dots, \mathbf{Y}(n)$ otrzymywanych w chwilach losowych. Problem estymacji rozwiązano przy założeniu, że rozkład $U_1, \dots, U_n$ należy do pewnej szerokiej klasy i jest znany, jak również przy założeniu, że jest rozkładem wykładniczym o nieznanym parametrze.

W Podrozdziale 3.2 zaproponowano procedurę adaptacyjną, która wymaga niewielkich założeń dotyczących rozkładu $U_1, \dots, U_n$.

W ostatnim podrozdziale prezentowane są analogiczne wyniki do tych uzyskanych w podrozdziale pierwszym dla pewnych modyfikacji funkcji straty określonej wzorem (3.1).

Problem estymacji na podstawie danych otrzymywanych w chwilach losowych rozpatrywany był przez Starra, Wardropa i Woodroofe'a (1976) w przypadku estymacji średniej z rozkładu normalnego o znanej wariancji. Niektóre z ich wyników zostały rozszerzone przez Magierę (1982) na jednoparametrową wykładniczą rodzinę rozkładów. W rozdziale niniejszym wprowadzono jawne postaci optymalnych procedur sekwencyjnych w przypadku modelu estymacji parametru wielowymiarowego.

3.1 Optymalne procedury sekwencyjne

Funkcję ryzyka procedury sekwencyjnej $\delta = (\tau, d(\tau))$ definiujemy przez

$$R(\mathbf{p}, \delta) = E_{\mathbf{p}} L_{\tau}(\mathbf{p}, \mathbf{d}) = E_{\mathbf{p}}[L(\mathbf{p}, \mathbf{d}(\tau)) + c(\tau)],$$

gdzie $E_{\mathbf{p}}$ oznacza wartość oczekiwaną ze względu na warunkowy rozkład, przy danym $\mathbf{p}$. Ryzyko bayesowskie $\bar{R}(\pi, \delta)$ procedury δ definiujemy natomiast przez

$$\bar{R}(\pi, \delta) = ER(\mathbf{p}, \delta) = \int_{\mathcal{P}} R(\mathbf{p}, \delta) \pi(d\mathbf{p}),$$

gdzie π jest rozkładem a priori parametru $\mathbf{p}$. Zakładamy, że π jest rozkładem Dirichleta $D(\alpha_1, \dots, \alpha_r)$ (naturalnym rozkładem sprzężonym dla prób z rozkładu wielomianowego) o funkcji gęstości postaci

$$g(\mathbf{p}; \mathbf{a}) = \begin{cases} \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_r)} p_1^{\alpha_1-1} \dots p_r^{\alpha_r-1}, & \text{gdzie } \sum_{i=1}^r p_i = 1, \\ & p_i \geq 0, i = 1, \dots, r, \\ 0, & \text{w przeciwnym wypadku,} \end{cases} \quad (3.2)$$

gdzie $\mathbf{a} \in \{(\alpha_1, \dots, \alpha_r)^T : \alpha_i > 0, i = 1, \dots, r\}$ i $\alpha = \sum_{i=1}^r \alpha_i$.

3.1.1 Bayesowska procedura sekwencyjna

Oznaczmy $X_i(k(\tau)) = \sum_{j=1}^{k(\tau)} Y_i(j)$. Następujący lemat podaje postać bayesowskiego estymatora parametru $\mathbf{p}$, względem rozkładu a priori Dirichleta $D(\alpha_1, \dots, \alpha_r)$, przy funkcji straty określonej wzorem (3.1).

Lemat 3.1 *Niech funkcja straty będzie postaci (3.1). Wówczas dla dowolnego czasu zatrzymania τ , bayesowski estymator parametru $\mathbf{p}$ względem rozkładu a priori π o funkcji gęstości określonej wzorem (3.2), gdy $\alpha_i > 1, i = 1, \dots, r$, $\alpha = \sum_{i=1}^r \alpha_i$, jest postaci $\mathbf{d}^*(\tau) = (d_1^*(\tau), \dots, d_r^*(\tau))^T$, gdzie*

$$d_i^*(\tau) = \frac{X_i(k(\tau)) + \alpha_i - 1}{k(\tau) + \alpha - 1} \quad (3.3)$$

i oczekiwana strata a posteriori jest równa

$$E[L(\mathbf{p}, \mathbf{d}^*(\tau)) \mid \mathcal{F}_{\tau}] = \frac{r - 1}{k(\tau) + \alpha - 1}.$$

Dowód: Funkcja wiarygodności procesu obserwowanego do czasu t jest postaci

$$\frac{k(t)!}{(X_1(k(t)))! \dots (X_r(k(t)))!} p_1^{X_1(k(t))} \dots p_r^{X_r(k(t))}.$$

Funkcja gęstości rozkładu a posteriori parametru $\mathbf{p}$ przy danym $\mathcal{F}_t$ określona jest następującym wzorem

$$g(\mathbf{p}; \mathbf{a} \mid \mathcal{F}_t) = \frac{\Gamma(k(t) + \alpha)}{\Gamma(X_1(k(t)) + \alpha_1) \dots \Gamma(X_r(k(t)) + \alpha_r)} p_1^{X_1(k(t)) + \alpha_1 - 1} \dots p_r^{X_r(k(t)) + \alpha_r - 1},$$

czyli jest rozkładem Dirichleta z parametrami $X_1(k(t)) + \alpha_1, \dots, X_r(k(t)) + \alpha_r$. Wobec tego ryzyko a posteriori jest postaci

$$E[L(\mathbf{p}, \mathbf{d}(t)) \mid \mathcal{F}_t] = \int_{\mathcal{P}} L(\mathbf{p}, \mathbf{d}) g(\mathbf{p}; \mathbf{a} \mid \mathcal{F}_t) d\mathbf{p} =$$

$$\frac{\Gamma(k(t) + \alpha)}{\prod_{i=1}^r \Gamma(X_i(k(t)) + \alpha_i)} \int \dots \int_{\mathcal{P}} \sum_{i=1}^r \frac{(d_i(t) - p_i)^2}{p_i} p_1^{X_1(k(t)) + \alpha_1 - 1} \dots p_r^{X_r(k(t)) + \alpha_r - 1} dp_1 \dots dp_{r-1}.$$

Funkcja ryzyka osiąga minimum, jeśli

$$\int \dots \int_{\mathcal{P}} \frac{d_i(t) - p_i}{p_i} p_1^{X_1(k(t)) + \alpha_1 - 1} \dots p_r^{X_r(k(t)) + \alpha_r - 1} dp_1 \dots dp_{r-1} = 0,$$

$i = 1, \dots, r$. Przy założeniu, że $\alpha_i > 1$ dla każdego $i = 1, \dots, r$, rozwiązaniem powyższego układu równań jest

$$d_i^*(t) = \frac{X_i(k(t)) + \alpha_i - 1}{k(t) + \alpha - 1},$$

$i = 1, \dots, r$. Oczekiwana strata a posteriori związana z rozpatrywanym rozkładem a priori π i estymatorem $\mathbf{d}^*(t)$ jest równa

$$\frac{r - 1}{k(t) + \alpha - 1}.$$

Przy założeniu, że τ jest losowym czasem obserwacji, lemat wynika z mocnej własności Markowa. $\square$

Z Lematu 3.1 wynika, że wybór optymalnej procedury sekwencyjnej może być utożsamiany z następującym problemem wyboru optymalnego czasu zatrzymania. Mianowicie całkowita strata (koszt) poniesiona do chwili τ , gdy zastosujemy regułę $\mathbf{d}^*(\tau)$, wyraża się następującym wzorem

$$\mathcal{L}(\tau) = \mathcal{L}(k(\tau), \tau) = \frac{r - 1}{k(\tau) + \alpha - 1} + c(\tau).$$

Problem polega na znalezieniu takiego czasu zatrzymania, który minimalizuje oczekiwaną stratę $E\mathcal{L}(\tau)$ po wszystkich czasach zatrzymania τ . Taki czas zatrzymania nazywamy *optymalnym czasem zatrzymania*, a odpowiadającą temu czasowi procedurę *bayesowską procedurą sekwencyjną*.

Założmy, że zmienne losowe $U_1, \dots, U_n$ są niezależne o jednakowym rozkładzie o dystrybuancie F . Założmy również, że $F(0) = 0$, $F(t) > 0$ dla $t > 0$, F jest absolutnie ciągła o funkcji gęstości f ; f jest prawostronną pochodną F na $(0, \infty)$. Oznaczmy klasę takich dystrybuant F przez $\mathcal{G}$.

Niech $\zeta = \sup\{t : F(t) < 1\}$ i $\rho(t) = f(t)[1 - F(t)]^{-1}$, $0 \leq t < \zeta$ oznacza intensywność awarii. Proces $k(t)$, $0 \leq t \leq \zeta$, jest niestacjonarnym łańcuchem Markowa ze względu na $\mathcal{F}_t$, $0 \leq t \leq \zeta$, i jego operator infinitesimalny jest postaci

$$\mathcal{A}_t h(k) = (n - k)\rho(t)[h(k + 1) - h(k)]$$

dla $k \in E_n = \{0, 1, \dots, n\}$ i wszystkich funkcji o wartościach rzeczywistych h , określonych na E_n (zobacz Starr, Wardrop, Woodroffe (1976)).

Niech h będzie daną funkcją o wartościach rzeczywistych, określoną na E_n i taką, że $0 \leq h(k) < \infty$ dla każdego $k \in E_n$ oraz niech

$$\mathcal{L}_h(t) = \mathcal{L}_h(k(t), t) = h(k(t)) + c(t),$$

$t \geq 0$, będzie ponoszona strata, jeśli proces jest zatrzymany w chwili t . Założmy, że $h(k) - h(k+1)$ jest funkcją nierosnącą dla $k \leq n-1$ oraz $F \in \mathcal{G}$ ma nierosnącą intensywność awarii. Modyfikując Twierdzenie 2.1 Starra, Wardropa i Woodroffe'a (1976) dla funkcji kosztu $c(t)$ zamiast funkcji ct możemy wywnioskować, że czas zatrzymania

$$\begin{aligned} \tau_h &= \inf\{t \geq 0 : \mathcal{A}_t h(k(t)) + c'(t) \geq 0\} \\ &= \inf\{t \geq 0 : [n - k(t)]\rho(t)[h(k(t)) - h(k(t) + 1)] \leq c'(t)\} \end{aligned}$$

jest optymalny. Jeśli spełnione są założenia dotyczące funkcji $\rho(t)$, $h(k)$ i $c(t)$, to zachodzi tak zwany przypadek monotoniczny, to znaczy, jeśli operator infinitesimalny straty $\mathcal{L}_h(t_0)$ jest nieujemny w punkcie t_0 , to pozostaje nieujemny dla każdego $t > t_0$. Optymalność τ_h , gdy dla $\mathcal{L}_h$ zachodzi przypadek monotoniczny wynika z tożsamości Dynkina (zobacz Breiman (1968)), która zastosowana do funkcji straty $\mathcal{L}_h(\tau)$ daje następującą równość

$$E\mathcal{L}_h(\tau) - h(0) = E \left\{ \int_0^\tau [\mathcal{A}_t h(k(t)) + c'(t)] dt \right\}$$

dla każdej chwili τ . Powyższa metoda była wykorzystana przy wyprowadzaniu jawnych procedur sekwencyjnych w innych modelach (zobacz, na przykład, Ross (1971), Chen i

Wardrop (1980), Shapiro i Wardrop (1980)). W szczególności, przyjmując za funkcję h funkcję postaci

$$h(k(t)) = \frac{r-1}{k(t) + \alpha - 1},$$

która związana jest z rozważanym modelem, otrzymujemy następujące twierdzenie.

Twierdzenie 3.1 *Niech $F \in \mathcal{G}$ ma nierosnącą intensywność awarii ρ . Wówczas bayesowska procedura sekwencyjna względem rozkładu a priori π parametru $\mathbf{p}$ o funkcji gęstości określonej wzorem (3.2), gdy $\alpha_i > 1, i = 1, \dots, r$, $\alpha = \sum_{i=1}^r \alpha_i$, jest postaci $\delta^\alpha = (\tau_\alpha, \mathbf{d}^*(\tau_\alpha))$, gdzie*

$$\tau_\alpha = \inf\{t \geq 0 : (r-1)[n - k(t)]\rho(t) \leq c'(t)[k(t) + \alpha - 1][k(t) + \alpha]\}$$

$$i \mathbf{d}^*(\tau_\alpha) = (d_1^*(\tau_\alpha), \dots, d_r^*(\tau_\alpha))^T \text{ z}$$

$$d_i^*(\tau_\alpha) = \frac{X_i(k(\tau_\alpha)) + \alpha_i - 1}{k(\tau_\alpha) + \alpha - 1},$$

$$i = 1, \dots, r.$$

3.1.2 Minimaksowa procedura sekwencyjna

W tym podrozdziale pokażemy, że czas zatrzymania τ_r określa minimaksową procedurę estymacji nieznanego parametru $\mathbf{p}$.

Twierdzenie 3.2 *Jeśli $E(t_1) < \infty$, to procedura sekwencyjna $\delta^r = (\tau_r, \mathbf{d}^0(\tau_r))$, gdzie*

$$\tau_r = \inf\{t \geq 0 : (r-1)[n - k(t)]\rho(t) \leq c'(t)[k(t) + r - 1][k(t) + r]\}$$

$$i \mathbf{d}^0(\tau_r) = (d_1^0(\tau_r), \dots, d_r^0(\tau_r))^T \text{ z}$$

$$d_i^0(\tau_r) = \frac{X_i(k(\tau_r))}{k(\tau_r) + r - 1},$$

$$i = 1, \dots, r, \text{ jest minimaksowa.}$$

Dowód: Jeśli π_ϵ jest sprzężonym rozkładem a priori parametru $\mathbf{p}$ z $\alpha_i = 1 + \epsilon, i = 1, \dots, r$, $\epsilon > 0$, wówczas z Lematu 3.1 wynika, że bayesowska procedura sekwencyjna redukuje się do problemu minimalizacji

$$\mathcal{V}(\epsilon, \tau) = E \left[\frac{r-1}{k(\tau) + r(1 + \epsilon) - 1} + c(\tau) \right].$$

ze względu na τ . Z Twierdzenia 3.1, procedura, która minimalizuje $\mathcal{V}(\epsilon, \tau)$ jest postaci $\tilde{\delta}^\epsilon := (\tilde{\tau}_\epsilon, \mathbf{d}^\epsilon(\tilde{\tau}_\epsilon))$, gdzie

$$\tilde{\tau}_\epsilon := \tau_{r(1+\epsilon)} = \inf\{t \geq 0 : (r-1)[n - k(t)]\rho(t) \leq c'(t)[k(t) + r(1+\epsilon) - 1][k(t) + r(1+\epsilon)]\}$$

i $\mathbf{d}^\epsilon(\tilde{\tau}_\epsilon) = (d_1^\epsilon(\tilde{\tau}_\epsilon), \dots, d_r^\epsilon(\tilde{\tau}_\epsilon))^T$ z

$$d_i^\epsilon(\tilde{\tau}_\epsilon) = \frac{X_i(k(\tilde{\tau}_\epsilon)) + \epsilon}{k(\tilde{\tau}_\epsilon) + r(1+\epsilon) - 1}.$$

Z Twierdzenia 2.1 Starra, Wardropa i Woodroofe'a (1976) wynika, że metoda minimalizacji $\mathcal{V}(\epsilon, \tau)$ jest niezależna od wybranego rozkładu a priori parametru $\mathbf{p}$ oraz $\mathcal{V}(\epsilon, \tilde{\tau}_\epsilon)$ jest niezależne od rozkładu π tak, że $\bar{R}(\pi_\epsilon, \tilde{\delta}^\epsilon) = \mathcal{V}(\epsilon, \tilde{\tau}_\epsilon)$. Łatwo można zauważyć, że funkcja ryzyka $R(\mathbf{p}, \tilde{\delta}^0) = \mathcal{V}(0, \tilde{\tau}_0)$ jest stałą (niezależną od $\mathbf{p}$). Używając argumentów Starra, Wardropa i Woodroofe'a (1976) (zobacz Twierdzenie 4.2) można pokazać, że $\mathcal{V}(\epsilon, \tilde{\tau}_\epsilon) \rightarrow \mathcal{V}(0, \tilde{\tau}_0)$, gdy $\epsilon \rightarrow 0$. Tak więc twierdzenie wynika ze znanych faktów z teorii podejmowania decyzji (zobacz Ferguson (1967), Twierdzenie 2, str. 90). $\square$

3.1.3 Bayesowska procedura sekwencyjna w przypadku, gdy F nie jest znane

Rozważmy problem estymacji parametru $\mathbf{p}$ w przypadku, gdy $U_1, \dots, U_n$ są warunkowo niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z nieznanym parametrem w , przy danym $W = w$, gdzie W jest zmienną losową o rozkładzie gamma z parametrami ν, β , to znaczy W ma rozkład o funkcji gęstości postaci

$$\Gamma(\nu)^{-1} \beta^\nu w^{\nu-1} e^{-\beta w} I_{(0,\infty)}(w). \quad (3.4)$$

Rozkład a posteriori W przy danym $\mathcal{F}_t$ jest rozkładem gamma z parametrami ν_t, β_t , gdzie

$$\nu_t = \nu + k(t) \quad \text{ i } \quad \beta_t = \beta + \sum_{j=1}^{k(t)} t_j + [n - k(t)]t.$$

Oznaczmy parametry rozkładu a priori o funkcji gęstości określonej wzorem (3.4) przez ν_0 i β_0 i weźmy pod uwagę proces $(\nu_t, \beta_t), t \geq 0$, o wartościach ze zbioru $\{\nu_0, \nu_0+1, \dots, \nu_0+n\} \times (0, \infty)$. Proces ten jest stacjonarnym procesem Markowa i jego operator infinitesimalny jest postaci

$$\mathcal{A}H(\nu, \beta) = [H(\nu+1, \beta) - H(\nu, \beta)](m - \nu)\nu\beta^{-1} + (m - \nu)H'_\beta(\nu, \beta),$$

gdzie $m = \nu_0 + n$ (Woodroofe (1976), Stadje (1990)). Dziedzina $\mathcal{A}$ zawiera wszystkie funkcje H , które są różniczkowalne w sposób ciągły ze względu na β dla każdego ν .

Niech całkowita strata związana z obserwowaniem procesu (ν_t, β_t) , $t \geq 0$, do chwili t będzie postaci

$$\mathcal{L}(t) = \mathcal{L}(\nu_t, \beta_t, t) = h(\nu_t) + c(t),$$

gdzie $h(\nu)$ jest funkcją określoną na zbiorze $\{\nu_0, \nu_0 + 1, \dots, \nu_0 + n\}$ oraz taką, że $0 \leq h(\nu) < \infty$ dla każdego $\nu \geq \nu_0$. Jeśli funkcja $[h(\nu) - h(\nu + 1)](m - \nu)\nu$ jest funkcją nierosnącą dla $\nu = \nu_0, \nu_0 + 1, \dots, \nu_0 + n - 1$, wówczas korzystając z metod z Podrozdziału 3.1.1 można pokazać, że czas zatrzymania

$$\begin{aligned} \tau_h &= \inf\{t \geq 0 : \mathcal{A}h(\nu_t) + c'(t) \geq 0\} \\ &= \inf\{t \geq 0 : [h(\nu_t) - h(\nu_t + 1)](m - \nu_t)\nu_t\beta_t^{-1} \leq c'(t)\} \end{aligned}$$

jest optymalny. W szczególności, dla

$$h(\nu_t) = \frac{r - 1}{\nu_t - \nu_0 + \alpha - 1} = \frac{r - 1}{k(t) + \alpha - 1}$$

zachodzi następujące twierdzenie.

Twierdzenie 3.3 *Niech rozkład $U_1, \dots, U_n$ będzie taki jak opisany powyżej. Wówczas bayesowska procedura sekwencyjna względem rozkładu a priori π parametru $\mathbf{p}$ o funkcji gęstości określonej wzorem (3.2), gdy $\alpha_i > 1$, $i = 1, \dots, r$, $\alpha = \sum_{i=1}^r \alpha_i$, jest postaci $(\tilde{\tau}_\alpha, \mathbf{d}^*(\tilde{\tau}_\alpha))$, gdzie*

$$\tilde{\tau}_\alpha = \inf\{t \geq 0 : (r - 1)(m - \nu_t)\nu_t \leq \beta_t c'(t)[k(t) + \alpha - 1][k(t) + \alpha]\}$$

i estymator $\mathbf{d}^(\tilde{\tau}_\alpha)$ jest określony wzorem (3.3).*

3.2 Procedura adaptacyjna

Procedury rozważane w podrozdziałach 3.1.1 3.1.2 wymagają znajomości dystrybucyjności F niezależnych zmiennych losowych $U_1, \dots, U_n$. W tym podrozdziale przedstawimy procedurę adaptacyjną, która nie wymaga znajomości ani n ani F i jest równie dobra, jak rozważane wcześniej procedury, gdy n jest dostatecznie duże, dla szerokiej klasy dystrybucyjności F . Niech jak poprzednio $\mathbf{Y}(i) = (Y_1(i), \dots, Y_r(i))^T$, $i = 1, 2, \dots, n$, będą niezależnymi zmiennymi losowymi, niezależnymi od $U_1, \dots, U_n$, i mają rozkład wielomianowy $\mathcal{W}(1, \mathbf{p})$ z nieznanym parametrem $\mathbf{p} = (p_1, \dots, p_r)^T \in \mathcal{P}$. Będziemy rozważać tylko takie procedury $\delta = (\tau, \mathbf{d}^*(\tau))$, dla których $\tau \geq t_1$ oraz $\mathbf{d}^*(\tau) = (d_1^*(\tau), \dots, d_r^*(\tau))^T$, gdzie

$$d_i^*(\tau) = \frac{X_i(k(\tau))}{k(\tau) + r - 1}, \quad i = 1, \dots, r.$$

Zakładamy, że funkcja kosztu obserwowania procesu do chwili t jest postaci $c(t) = ct$, gdzie c jest dodatnią stałą. Minimalne ryzyko takiej procedury jest wówczas równe

$$V(n, F) = \inf_{\tau} E\mathcal{L}(\tau) = \inf_{\tau} E \left\{ \frac{r-1}{k(\tau) + r - 1} + c\tau \right\},$$

gdzie infimum jest brane po wszystkich czasach zatrzymania τ (ze względu na $\mathcal{F}_t = \sigma\{k(s), s \leq t, \mathbf{Y}(1), \dots, \mathbf{Y}(k(t))\}$).

Twierdzenie 3.4 *Niech $F \in \mathcal{G}$ ma intensywność awarii ρ , dodatnią i ciągłą na $[0, \eta]$ dla pewnego $\eta > 0$. Wówczas*

$$\liminf n^{1/2} V(n, F) \geq 2 \left[\frac{(r-1)c}{\rho(0)} \right]^{1/2}$$

gdzie $n \rightarrow \infty$.

Dowód: W dowodzie wykorzystujemy argumenty Starra, Wardropa i Woodroofe'a (1976). Niech dane będzie $0 < \varepsilon < \rho(0)$ i niech $\delta > 0$ będzie tak małe, że $|\rho(t) - \rho(0)| < \varepsilon$ dla $0 \leq t \leq \delta$. Jeśli τ jest czasem zatrzymania i jeśli $\tau_\delta = \min\{\tau, \delta\}$, wówczas

$$E \left[\frac{r-1}{k(\tau) + r - 1} + c\tau \right] \geq E \left[\frac{r-1}{k(\tau_\delta) + r - 1} + c\tau_\delta \right] - E \left[\frac{r-1}{k(\delta) + r - 1} \right]. \quad (3.5)$$

Łatwo można sprawdzić, że

$$E \left[\frac{r-1}{k(\delta) + r - 1} \right] = O(n^{-1}), \quad (3.6)$$

gdzie $n \rightarrow \infty$. Oznaczmy

$$W(n, F) = \inf_{\tau \leq \delta} E \left[\frac{r-1}{k(\tau) + r - 1} + c\tau \right].$$

Korzystając z nierówności (3.5) oraz z równości (3.6), otrzymujemy, że

$$V(n, F) \geq W(n, F) - O(n^{-1}).$$

Niech $G(t) = 1 - \exp\{-(\rho(0) + \varepsilon)t\}$ dla $t \geq 0$. Wówczas dla $0 \leq t \leq \delta$ zachodzi nierówność $F(t) \leq G(t)$ (ponieważ $\rho(t) \leq \rho(0) + \varepsilon$ dla $0 \leq t \leq \delta$ i $F(t) = 1 - \exp\{-\int_0^t \rho(u)du\}$) oraz $W(n, F) \geq W(n, G) \geq V(n, G)$ (zobacz Starr, Wardrop i Woodroofe (1976), Lemat 2.2), tak, że

$$V(n, F) \geq V(n, G) - O(n^{-1}).$$

Z przykładu 2.1 Starra, Wardropa i Woodroofe'a (1976) wynika, że

$$V(n, G) = E \left[\frac{r-1}{k(\tau^*) + r - 1} + c\tau^* \right],$$

gdzie $\tau^* = t_k$ i k jest najmniejszą liczbą całkowitą l , dla której

$$(r-1)(n-l)(\rho_0 + \varepsilon) \leq c(r+l-1)(r+l)$$

oraz

$$V(n, G) = \frac{r-1}{k+r-1} + \sum_{i=1}^k \frac{c}{(\rho_0 + \varepsilon)(n-i+1)}$$

tak, że

$$\lim_{n \rightarrow \infty} n^{1/2} V(n, G) = 2 \left[\frac{(r-1)c}{\rho_0 + \varepsilon} \right]^{1/2}.$$

Twierdzenie wynika, gdy przejdziemy z $\varepsilon \rightarrow 0$. □

Rozważmy teraz strategię adhoc, w której czas zatrzymania jest postaci

$$\tau_0 = \inf\{t \geq 0 : r-1 \leq ct k(t)\}$$

i jako estymator parametru $\mathbf{p}$ przyjmijmy estymator $\mathbf{d}^0(\tau_0) = (d_1^0(\tau_0), \dots, d_r^0(\tau_0))^T$, gdzie $d_i^0(\tau_0) = X_i(k(\tau_0))/(k(\tau_0) + r - 1)$, $i = 1, \dots, r$. Zauważmy, że aby móc wykorzystać powyższą strategię nie musimy znać ani n ani F . Niech

$$V_0(n, F) = E \left[\frac{r-1}{k(\tau_0) + r - 1} + c\tau_0 \right]$$

będzie ryzykiem związanym z rozważaną strategią $\delta_0 = (\tau_0, \mathbf{d}^0(\tau_0))$. Pokażemy, że jeśli dla pewnego $m \geq 1$ spełniony jest dodatkowy warunek

$$\int_0^\infty [1 - F(t)]^m dt < \infty, \quad (3.7)$$

to $V_0(n, F) \sim V(n, F)$, gdy $n \rightarrow \infty$. Oczywiście, jeśli nie istnieje takie m , dla którego warunek (3.7) jest spełniony, to $V(n, F) = \infty$ dla każdego n .

Twierdzenie 3.5 *Niech spełnione będą założenia Twierdzenia 3.4 oraz zachodzi warunek (3.7). Wówczas*

$$\lim_{n \rightarrow \infty} n^{1/2} V_0(n, F) = 2 \left[\frac{(r-1)c}{\rho_0} \right]^{1/2}.$$

Dowód: Biorąc pod uwagę Twierdzenie 3.4, wystarczy pokazać, że $\limsup n^{1/2} V_0(n, F) = 2[(r-1)c/\rho_0]^{1/2}$. Zauważmy, że $c\tau_0 \geq (r-1)/(k(\tau_0) + r - 1)$, a więc $V_0(n, F) \leq 2cE\tau_0$. Niech $\tau_n = n^{1/2}\tau_0$ i $K_n(s) = k(s/n^{1/2})$ dla $s \geq 0$. Korzystając z tego, że

$$\tau_n > s \iff \tau_0 > s/\sqrt{n} \iff s/\sqrt{n}k(s/\sqrt{n}) < (r-1)/c \iff K_n(s) < \sqrt{n}(r-1)/cs$$

otrzymujemy, że

$$P(\tau_n > s) = P(K_n(s) < (r-1)n^{1/2}/cs). \quad (3.8)$$

Wobec tego, że dla każdego $s > 0$ zmienna losowa $K_n(s)$ ma rozkład dwumianowy z parametrami n i $F(s/n^{1/2})$, łatwo można pokazać, że $K_n(s)/n^{1/2} \rightarrow \rho_0 s$ względem prawdopodobieństwa, gdy $n \rightarrow \infty$. Stąd wynika, że prawa strona wyrażenia (3.8) dąży do 1, jeśli $s\rho_0 < (r-1)/cs$ oraz dąży do 0, jeśli $s\rho_0 > (r-1)/cs$. To znaczy

$$\tau_n \rightarrow \left[\frac{(r-1)c}{\rho(0)} \right]^{1/2},$$

względem prawdopodobieństwa. Tak więc wystarczy pokazać, że $\tau_n, n \geq 1$, jest jednostajnie całkowalne. W tym celu możemy użyć argumentów Starra, Wardropa i Woodroofe'a (1976). $\square$

Wniosek 1 *Jeśli spełnione są założenia Twierdzenia 3.5, to*

$$\lim n^{1/2}V(n, F) = 2 \left[\frac{(r-1)c}{\rho_0} \right]^{1/2},$$

gdy $n \rightarrow \infty$. $\square$

3.3 Uwagi dodatkowe

Jeśli zamiast funkcji straty określonej wzorem (3.1), weźmiemy pod uwagę następujące funkcje

$$L_1(\mathbf{p}, \mathbf{d}) = \sum_{i=1}^r \frac{(d_i - p_i)^2}{1 - p_i}, \quad (3.9)$$

$$L_2(\mathbf{p}, \mathbf{d}) = \left[\sum_{i=1}^r p_i(1 - p_i) \right]^{-1} \sum_{i=1}^r (d_i - p_i)^2, \quad (3.10)$$

to możemy uzyskać analogiczne wyniki do prezentowanych w Podrozdziałach 3.1 i 3.2. Rozważmy funkcję straty określoną wzorem (3.9). Dla dowolnego czasu zatrzymania τ , bayesowski estymator parametru $\mathbf{p}$ względem rozkładu a priori $\pi_1 := \pi$ Dirichleta o funkcji gęstości określonej wzorem (3.2), gdy $\alpha_i > 1$, $i = 1, \dots, r$, jest postaci $\mathbf{d}_1^*(\tau) = (d_{1,1}^*(\tau), \dots, d_{1,r}^*(\tau))^T$, gdzie

$$d_{1,i}^*(\tau) = \frac{X_i(k(\tau)) + \alpha_i}{k(\tau) + \alpha - 1}, \quad (3.11)$$

$i = 1, \dots, r$ i oczekiwana strata a posteriori jest postaci

$$E[L_1(\mathbf{p}, \mathbf{d}_1^*(\tau)) \mid \mathcal{F}_\tau] = \frac{1}{k(\tau) + \alpha - 1}.$$

Jeśli $F \in \mathcal{G}$ ma nierosnącą intensywność awarii ρ , wówczas bayesowską procedurą ze względu na rozkład a priori π_1 parametru $\mathbf{p}$ jest procedura $\delta_1^\alpha = (\tau_{1,\alpha}, \mathbf{d}_1^*(\tau_{1,\alpha}))$, gdzie

$$\tau_{1,\alpha} = \inf\{t \geq 0 : [n - k(t)]\rho(t) \leq c'(t)[k(t) + \alpha][k(t) + \alpha - 1]\}$$

i $\mathbf{d}_1^*(\tau_{1,\alpha}) = (d_{1,1}^*(\tau_{1,\alpha}), \dots, d_{1,r}^*(\tau_{1,\alpha}))^T$ jest zdefiniowana wzorem (3.11) dla $\tau = \tau_{1,\alpha}$. Procedura sekwencyjna $\delta_1^0 = (\tau_{1,0}, \mathbf{d}_1^0(\tau_{1,0}))$, gdzie

$$\tau_{1,0} = \inf\{t \geq 0 : [n - k(t)]\rho(t) \leq c'(t)[k(t) + r][k(t) + r - 1]\} \quad (3.12)$$

i $\mathbf{d}_1^0(\tau_{1,0}) = (d_{1,1}^0(\tau_{1,0}), \dots, d_{1,r}^0(\tau_{1,0}))^T$ z

$$d_{1,i}^0(\tau_{1,0}) = \frac{X_i(k(\tau_{1,0})) + 1}{k(\tau_{1,0}) + r - 1},$$

$i = 1, \dots, r$, jest minimaksowa.

W przypadku funkcji straty $L_2(\mathbf{d}, \mathbf{p})$ określonej wzorem (3.10) zdefiniujemy zmodyfikowaną rodzinę rozkładów a priori π_2 parametru $\mathbf{p}$ o funkcji gęstości postaci

$$g_2(\mathbf{p}; \mathbf{a}) = \frac{\Gamma(\alpha + 2)}{\sum_{i=1}^r [(\alpha - \alpha_i)\Gamma(\alpha_i + 1) \prod_{j=1, j \neq i}^r \Gamma(\alpha_j)]} \sum_{i=1}^r p_i(1 - p_i)p_1^{\alpha_1 - 1} \dots p_r^{\alpha_r - 1},$$

$\alpha_i > 0, i = 1, \dots, r, \alpha = \sum_{i=1}^r \alpha_i$. Dla dowolnego czasu zatrzymania τ , bayesowski estymator parametru $\mathbf{p}$ względem rozkładu a priori π_2 , przy funkcji straty określonej wzorem (3.10) jest postaci $\mathbf{d}_2^*(\tau) = (d_{2,1}^*(\tau), \dots, d_{2,r}^*(\tau))^T$, gdzie

$$d_{2,i}^*(\tau) = \frac{X_i(k(\tau)) + \alpha_i}{k(\tau) + \alpha}, \quad (3.13)$$

$i = 1, \dots, r$ i oczekiwana strata a posteriori jest postaci

$$E[L_2(\mathbf{p}, \mathbf{d}^*(\tau)) \mid \mathcal{F}_\tau] = \frac{1}{k(\tau) + \alpha}.$$

Jeśli $F \in \mathcal{G}$ ma nierosnącą intensywność awarii ρ , wówczas bayesowską procedurą ze względu na rozkład a priori π_2 parametru $\mathbf{p}$ jest procedura $\delta_2^\alpha = (\tau_{2,\alpha}, \mathbf{d}_2^*(\tau_{2,\alpha}))$, gdzie

$$\tau_{2,\alpha} = \inf\{t \geq 0 : [n - k(t)]\rho(t) \leq c'(t)[k(t) + \alpha][k(t) + \alpha + 1]\}$$

i $\mathbf{d}_2^*(\tau_{2,\alpha}) = (d_{2,1}^*(\tau_{2,\alpha}), \dots, d_{2,r}^*(\tau_{2,\alpha}))^T$ jest zdefiniowana wzorem (3.13) dla $\tau = \tau_{2,\alpha}$. Procedura sekwencyjna $\delta_2^0 = (\tau_{2,0}, \mathbf{d}_2^0(\tau_{2,0}))$, gdzie

$$\tau_{2,0} = \inf\{t \geq 0 : [n - k(t)]\rho(t) \leq c'(t)k(t)[k(t) + 1]\} \quad (3.14)$$

i $\mathbf{d}_2^0(\tau_{2,0}) = (d_{2,1}^0(\tau_{2,0}), \dots, d_{2,r}^0(\tau_{2,0}))^T$ z

$$d_{2,i}^0(\tau_{2,0}) = \frac{X_i(k(\tau_{2,0}))}{k(\tau_{2,0})},$$

$i = 1, \dots, r$, jest minimaksowa.

W szczególnym przypadku, gdy $r = 2$, to znaczy, gdy $\mathbf{Y}(1), \dots, \mathbf{Y}(n)$ mają rozkład dwumianowy określony przez $P(\mathbf{Y}(j) = (1, 0)^T) = p = 1 - P(\mathbf{Y}(j) = (0, 1)^T)$, $j = 1, \dots, n$, procedura sekwencyjna $\delta_2^0 = (\tau_{2,0}, \mathbf{d}_2^0(\tau_{2,0}))$, gdzie $\tau_{2,0}$ określone jest wzorem (3.14) i

$$d_{2,i}^0(\tau_{2,0}) = \frac{X_i(k(\tau_{2,0}))}{k(\tau_{2,0})},$$

$i = 1, \dots, r$, jest bayesowska względem rozkładu jednostajnego na odcinku $(0, 1)$ (granicznego rozkładu π_2 , w przypadku, gdy $\alpha_1 \rightarrow 0$ i $\alpha_2 \rightarrow 0$) i minimaksowa. Procedura sekwencyjna $\delta_2^0 = (\tau_{2,0}, \mathbf{d}_2^0(\tau_{2,0}))$ w przypadku rozkładu dwumianowego i przy założeniu, że funkcja kosztu obserwacji jest postaci $c(t) = ct$ podana była przez Starra, Wardropa i Woodroofe'a (1976).

Warto zauważyć, że minimaksowa procedura sekwencyjna przy funkcji straty określonej wzorem (3.10) nie zależy od r (wymiaru parametru), w przeciwieństwie do minimaksowych procedur przy funkcjach straty wyrażonych wzorami (3.1) i (3.9).

Rozdział 4

Bayesowska predykcja mediany próbkowej z przesuniętego rozkładu wykładniczego przy ustalonym i losowym rozmiarze przyszłej próby

Predykcja pewnych funkcji przyszłej obserwacji, szczególnie w przypadku, gdy obserwacje mają rozkład wykładniczy, ma szerokie zastosowanie w teorii niezawodności, w wielu problemach biologicznych, rolniczych i kontroli jakości. Lawless (1977) rozważał problem predykcji s -tej statystyki pozycyjnej oraz średniej z przyszłej próby ustalonego rozmiaru z rozkładu dwuparametrowego wykładniczego na podstawie obserwacji obciętej II typu. Wyprowadził wzory określające dolne i górne ograniczenie przedziału predykcji dla tych charakterystyk. Chou i Owen (1989) skonstruował jednoczesne predyktory dla statystyk pozycyjnych z l przyszłych prób (o ustalonym rozmiarze) również w przypadku dwuparametrowego rozkładu wykładniczego. Consul (1984) zwrócił uwagę, że w wielu problemach (szczególnie biologicznych i rolniczych) nie jest możliwe przyjęcie, że rozmiar przyszłej próby jest ustalony, gdyż część obserwacji z różnych powodów może być stracona. Wyprowadził on rozkłady statystyk pozycyjnych z rozkładu jednostajnego oraz rozstępu z rozkładu wykładniczego przy założeniu, że rozmiar próby ma uogólniony rozkład Poissona i uogólniony ujemny dwumianowy. Salem i El-Glad (1992, 1993) rozważał bayesowską predykcję r -tej statystyki pozycyjnej i rozstępu z dwuparametrowego rozkładu wykładniczego przy założeniu, że rozmiar próby ma uogólniony rozkład Poissona oraz ujemny dwumianowy. Ashour i El-Wakel (1994) wyprowadził bayesowski predyktor mediany z uogólnionego rozkładu Burra przy ustalonym i losowym rozmiarze próby.

Celem tego rozdziału jest wyprowadzenie bayesowskiego predyktora mediany z przyszłej próby, o ustalonym i losowym rozmiarze, przy założeniu kwadratowej funkcji straty, gdy obserwacje mają przesunięty rozkład wykładniczy o funkcji gęstości postaci

$$f(x, \lambda, \alpha) = \lambda \exp\{-\lambda(x - \alpha)\}, \quad (4.1)$$

gdzie $x > \alpha \geq 0$ i $\lambda > 0$, na podstawie obserwacji obciętej II typu. Jeśli α nie jest ograniczane do wartości nieujemnych, wówczas rozkład o gęstości (4.1) nazywany jest dwuparametrowym rozkładem wykładniczym. Wprowadzenie dwóch różnych terminów dla tych dwóch rozkładów jest zasadne ze względu na to, że tylko ten pierwszy ($\alpha \geq 0$) stanowi właściwy model rozkładu czasu życia. Na podstawie danych pochodzących z pracy Dunsmore'a (1974) przedstawiono numeryczną ilustrację uzyskanych wyników i dokonano ich porównania w przypadku ustalonego i losowego rozmiaru próby.

4.1 Predykcja mediany w przypadku ustalonego rozmiaru przyszłej próby

Na podstawie obserwacji n urządzeń do momentu awarii r spośród nich ($r \leq n$) chcemy oszacować wartość mediany z przyszłej próby o ustalonym rozmiarze. Niech $X_{(1)}, \dots, X_{(r)}$ będzie uzyskaną informacją, gdzie $X_{(i)}$ jest i -tą statystyką pozycyjną. Funkcja wiarygodności danej informacji próbkowej jest postaci

$$L(\lambda, \alpha) \sim \lambda^r \exp\{-\lambda[v + n(w - \alpha)]\}, \quad 0 \leq \alpha < w, \quad \lambda > 0,$$

gdzie

$$w = x_{(1)} \quad \text{ i } \quad v = \sum_{i=1}^r x_i + (n - r)x_r - nx_1.$$

Tak więc naturalną sprzężoną rodziną rozkładów a priori dla (λ, α) jest rodzina o funkcji gęstości postaci

$$f(\lambda, \alpha) \sim \lambda^g \exp\{-\lambda[h + c(b - \alpha)]\}, \quad 0 \leq \alpha < b, \quad \lambda > 0. \quad (4.2)$$

Aby funkcja określona wzorem (4.2) stanowiła właściwą funkcję gęstości należy położyć: $g > -1$, $h > 0$, $b > 0$, $c > 0$. Odpowiadająca gęstości a priori określonej wzorem (4.2) gęstość a posteriori dla (λ, α) jest postaci

$$f(\lambda, \alpha \mid \text{data}) = p\lambda^G \exp\{-\lambda[H + C(B - \alpha)]\}, \quad 0 \leq \alpha < B, \quad \lambda > 0, \quad (4.3)$$

gdzie

$$G = g + r, \quad B = \min(b, w), \quad C = c + n, \quad H = h + bc + v + nw - BC$$

i

$$p = \frac{C}{(G-1)! [H^{-G} - (H+BC)^{-G}]}$$

Niech $(Y_{(1)}, \dots, Y_{(m)})^T$ będzie wektorem statystyk pozycyjnych z przyszłej próby z rozkładu o funkcji gęstości określonej wzorem (4.1), gdzie m jest przyszłym ustalonym rozmiarem próby. Jeśli rozmiar próby jest nieparzysty, powiedzmy $m = 2k - 1$, wówczas rozkład mediany próbkowej $U = Y_{(k)}$ dla $u > \alpha$ ma funkcję gęstości postaci

$$f_{2k-1}(u \mid \lambda, \alpha) = K_1 \lambda \exp \{-\lambda k(u - \alpha)\} \sum_{i=1}^{k-1} \binom{k-1}{i} (-1)^i \exp \{-\lambda(u - \alpha)i\}, \quad (4.4)$$

gdzie $K_1 = (2k-1)! / [(k-1)!]^2$. Jeśli rozmiar próby jest parzysty, powiedzmy $m = 2k$, wówczas rozkład mediany próbkowej $U = (Y_{(k)} + Y_{(k+1)})/2$ dla $u > \alpha$ ma funkcję gęstości postaci

$$\begin{aligned} f_{2k}(u \mid \lambda, \alpha) = & K_2 \lambda \sum_{i=1}^{k-2} \binom{k-1}{i} \frac{(-1)^i}{k-1-i} [\exp \{-\lambda(k+1+i)(u - \alpha)\} \\ & - \exp \{-2\lambda k(u - \alpha)\}] + (-1)^{k-1} K_2 \lambda^2 (u - \alpha) \exp \{-2\lambda k(u - \alpha)\}, \end{aligned} \quad (4.5)$$

gdzie $K_2 = 2(2k)! / [(k-1)!]^2$. W przypadku, gdy m jest nieparzyste, powiedzmy $m = 2k-1$, to korzystając ze wzorów (4.3) i (4.4) bayesowski-przewidywany rozkład a posteriori mediany próbkowej U ma funkcję gęstości postaci

$$f_{2k-1}(u \mid data) = \int_0^\infty \int_0^{\min(u, B)} f_{2k-1}(u \mid \lambda, \alpha) f(\lambda, \alpha \mid data) d\alpha d\lambda.$$

Funkcja ta dla $u < B$ jest postaci

$$\begin{aligned} f_{2k-1}(u \mid data) = & A_1 \sum_{i=0}^{k-1} k_1(i) \left\{ [H + C(B - u)]^{-G-1} \right. \\ & \left. - [H + BC + (k+i)u]^{-G-1} \right\}, \end{aligned} \quad (4.6)$$

natomiast dla $u \geq B$ - postaci

$$\begin{aligned} f_{2k-1}(u \mid data) = & A_1 \sum_{i=0}^{k-1} k_1(i) \left\{ [H + (k+i)(u - B)]^{-G-1} \right. \\ & \left. - [H + BC + (k+i)u]^{-G-1} \right\}, \end{aligned} \quad (4.7)$$

gdzie $A_1 = K_1 p \Gamma(G+1)$, $k_1(i) = \binom{k-1}{i} (-1)^i / (k+i+C)$.

W przypadku, gdy m jest parzyste, powiedzmy $m = 2k$, to korzystając ze wzorów (4.3) i (4.5) bayesowski-przewidywany rozkład a posteriori mediany próbkowej U ma funkcję gęstości postaci

$$f_{2k}(u \mid data) = \int_0^\infty \int_0^{\min(u, B)} f_{2k}(u \mid \lambda, \alpha) f(\lambda, \alpha \mid data) d\alpha d\lambda.$$

Funkcja ta dla $u < B$ jest postaci

$$\begin{aligned} f_{2k}(u \mid data) = & A_2 \left\{ [H + C(B - u)]^{-G-1} \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{(-1)^i}{(k+1+i+C)(2k+C)} \right. \\ & + (2ku + H + CB)^{-G-1} \left[\sum_{i=0}^{k-2} \binom{k-1}{i} \frac{(-1)^i}{(k-1-i)(2k+C)} - \frac{(-1)^{k-1}}{(2k+C)^2} \right] \\ & - \sum_{i=0}^{k-2} \binom{k-1}{i} \frac{(-1)^i}{(k-1-i)(k+1+i+C)} [(k+1+i)u + H + BC]^{-G-1} \\ & \left. - (G+1) \frac{(-1)^{k-1} u}{2k+C} (2ku + H + CB)^{-G-2} \right\}, \end{aligned} \quad (4.8)$$

natomiast dla $u \geq B$ - postaci

$$\begin{aligned} f_{2k}(u \mid data) = & A_2 \left\{ (2ku + H + CB)^{-G-1} \left[\sum_{i=0}^{k-2} \frac{k_2(i)}{2k+C} - \frac{(-1)^{k-1}}{(2k+C)^2} \right] \right. \\ & - [2k(u - B) + H]^{-G-1} \left[\sum_{i=0}^{k-2} \frac{k_2(i)}{2k+C} - \frac{(-1)^{k-1}}{(2k+C)^2} \right] \\ & + \sum_{i=0}^{k-2} \frac{k_2(i)}{k+1+i+C} \left\{ [k_3(i)(u - B) + H]^{-G-1} - [k_3(i)u + H + CB]^{-G-1} \right\} \\ & + \frac{(-1)^{k-1}(G+1)u}{2k+C} \left\{ [2k(u - B) + H]^{-G-2} - (2ku + H + CB)^{-G-2} \right\} \\ & \left. - \frac{(-1)^{k-1}(G+1)B}{2k+C} [2k(u - B) + H]^{-G-2} \right\}, \end{aligned} \quad (4.9)$$

gdzie $A_2 = K_2 p \Gamma(G+1)$, $k_2(i) = \binom{k-1}{i} (-1)^i / (k-1-i)$, $k_3(i) = k+1+i$.

Wartość średnia i wariancja zmiennej losowej U o funkcji gęstości określonej wzorami (4.6) i (4.7) (gdy rozmiar próby jest nieparzysty) lub wzorami (4.8) i (4.9) (gdy rozmiar próby jest parzysty) jest odpowiednio bayesowskim predyktorem mediany próbkowej U oraz bayesowskim ryzykiem U (przy założeniu kwadratowej funkcji straty). Wielkości te mogą być obliczone metodami numerycznymi.

4.2 Predykcja mediany w przypadku losowego rozmiaru przyszłej próby

Niech $\mathbf{Y} = (Y_1, \dots, Y_M)^T$, gdzie Y_i , $i = 1, \dots, M$, ma przesunięty rozkład wykładniczy o funkcji gęstości określonej wzorem (4.1), będzie przyszłą próbą o losowym rozmiarze M . Na podstawie obserwacji obciętej II typu, $X_{(1)}, \dots, X_{(r)}$, wyprowadzimy bayesowski predyktor mediany próbkowej z przyszłej próby $\mathbf{Y}$. Niech $P(M = m)$ oznacza prawdopodobieństwo, że rozmiar próby M przyjmie wartość m . Wówczas funkcja gęstości mediany z przyszłej próby o losowym rozmiarze ma postać

$$f(u \mid data) = \frac{1}{P(M \geq 1)} \sum_{m=1}^{\infty} P(M = m) f_m(u \mid data).$$

Załóżmy, że M ma rozkład dwumianowy z parametrami s i θ , to znaczy rozkład o funkcji prawdopodobieństwa (dla $m = 0, 1, \dots, s$) postaci

$$P(M = m) = \binom{s}{m} \theta^m (1 - \theta)^{s-m}. \quad (4.10)$$

Biorąc pod uwagę rozkład rozmiaru próby, otrzymujemy następującą postać funkcji gęstości mediany próbkowej U

$$f(u \mid data) = \frac{1}{1 - (1 - \theta)^s} \sum_{m=1}^s f_m(u \mid data) \binom{s}{m} \theta^m (1 - \theta)^{s-m}. \quad (4.11)$$

Wartość średnia i wariancja zmiennej losowej U , o funkcji gęstości określonej wzorem (4.11) jest odpowiednio bayesowskim predyktorem mediany próbkowej U oraz bayesowskim ryzykiem U . Wielkości te mogą być obliczone metodami numerycznymi.

4.3 Wnioski wynikające z obliczeń numerycznych

Podrozdział ten poświęcony jest numerycznej ilustracji wyników uzyskanych w poprzednich podrozdziałach. Obliczenia zostały przeprowadzone na podstawie danych z przykładu z pracy Dunsmore'a (1974), w którym $n = 25$, $r = 10$, $b = 300$, $w = 276$, $v = 1164$. Przyjęto następujące wartości parametrów rozkładu a priori: $g = 0, 1, 2$, $c = 1, 2, 3$, $h = 1, 10, 100$ oraz $m = 3, 4, 5, 7$, gdy rozmiar próby jest ustalony. W przypadku, gdy rozmiar próby ma rozkład dwumianowy $b(s, \theta)$ przyjęto $s = 10$, $\theta = 3/10, 4/10, 5/10, 7/10$. Korzystając ze wzorów (4.6), (4.7), (4.8), (4.9) i (4.11) oraz pakietu "Maple V" obliczono predyktory mediany z przyszłej próby, jak również współczynniki wariancji jej rozkładu. Uzyskane wyniki zostały zestawione w Tabelach 1 i 2. Po analizie

tych wyników wyciągnięto następujące wnioski:

- (1) jeśli m i θ rośnie, to wartość predyktora maleje;
- (2) zmiana wartości parametru rozkładu a priori h nie ma dużego wpływu na zmianę wartości predyktora;
- (3) jeśli wartości parametrów rozkładu a priori c i h rosną, wartość predyktora również rośnie;
- (4) jeśli wartość parametru rozkładu a priori g rośnie, to wartość predyktora maleje;
- (5) współczynnik wariancji przy ustalonym rozmiarze próby jest mniejszy od odpowiadającemu mu współczynnikowi przy losowym rozmiarze.

Tabela 1

Predyktory mediany przy ustalonym i losowym rozmiarze przyszłej próby

Rozmiar próby		h	g=0			g=1			g=2		
			c=1	c=2	c=3	c=1	c=2	c=3	c=1	c=2	c=3
U s t a l o n y	m=3	1	381.01	383.32	385.63	370.51	372.59	374.67	361.92	363.81	365.70
		10	381.81	384.12	386.43	371.23	373.31	375.38	362.57	364.46	366.35
		100	389.75	392.08	394.40	378.38	380.47	382.56	369.07	370.98	372.87
	m=4	1	381.01	383.32	385.63	370.51	372.59	374.67	361.92	363.81	365.70
		10	381.81	384.12	386.43	371.23	373.31	375.38	362.57	364.46	366.35
		100	389.75	392.08	394.40	378.38	380.47	382.56	369.07	370.98	372.87
	m=5	1	374.31	376.58	378.76	364.57	366.53	368.48	356.51	358.30	360.07
		10	375.15	377.33	379.50	365.24	367.20	369.15	357.12	358.91	360.68
		100	382.60	384.79	386.98	371.94	373.91	375.88	363.22	365.01	366.80
	m=7	1	371.26	373.38	386.98	361.73	363.64	365.54	353.94	355.67	357.40
		10	371.98	374.10	375.48	362.38	364.29	366.19	354.53	356.26	357.99
		100	379.19	381.32	376.21	368.87	370.79	372.70	360.43	362.17	363.91
L o s o w y	$\theta = .3$	1	387.99	390.44	392.89	376.79	379.00	381.20	376.63	369.63	371.63
		10	388.84	391.29	393.74	377.55	379.76	381.96	368.32	370.33	372.33
		100	397.31	399.78	402.24	385.18	387.40	389.62	375.26	377.28	379.29
	$\theta = .4$	1	381.97	384.30	386.62	371.37	373.47	375.56	362.70	364.61	366.51
		10	382.77	385.10	387.43	372.09	374.19	376.29	363.36	365.27	367.17
		100	390.79	393.14	395.48	379.31	381.42	383.53	369.92	371.84	373.75
	$\theta = .5$	1	377.51	379.76	381.99	367.36	369.38	371.39	359.06	360.89	362.72
		10	378.28	380.53	382.76	368.05	370.07	372.09	359.69	361.52	363.35
		100	385.97	388.22	390.47	374.97	377.00	379.03	365.97	367.82	369.66
	$\theta = .7$	1	372.44	374.57	376.71	362.79	364.72	366.64	354.90	356.65	358.40
		10	373.17	375.31	377.44	363.50	365.38	367.30	355.50	357.25	359.00
		100	380.47	382.62	384.77	370.01	371.96	373.89	361.47	363.23	365.00

Tabela 2

Współczynniki wariancji rozkładu mediany próbkowej przy ustalonym i losowym rozmiarze przyszłej próby

Rozmiar próby		h	g=0			g=1			g=2		
			c=1	c=2	c=3	c=1	c=2	c=3	c=1	c=2	c=3
Ustalony	m=3	1	.2419	.2454	.2488	.2207	.2240	.2272	.2030	.2061	.2091
		10	.2433	.2467	.2501	.2220	.2252	.2284	.2042	.2072	.2102
		100	.2562	.2595	.2627	.2341	.2372	.2403	.2156	.2186	.2215
	m=4	1	.2040	.2069	.2098	.1855	.1882	.1909	.1701	.1727	.1752
		10	.2052	.2081	.2109	.1866	.1893	.1920	.1711	.1737	.1762
		100	.2161	.2188	.2216	.1968	.1994	.2020	.1807	.1832	.1856
	m=5	1	.1969	.1997	.2026	.1788	.1815	.1841	.1639	.1664	.1688
		10	.1980	.2008	.2036	.1798	.1825	.1851	.1648	.1673	.1698
		100	.2087	.2114	.2141	.1899	.1924	.1950	.1742	.1766	.1790
	m=7	1	.1729	.1755	.1780	.1566	.1589	.1613	.1430	.1452	.1474
		10	.1739	.1764	.1789	.1575	.1598	.1621	.1439	.1461	.1483
		100	.1834	.1858	.1882	.1663	.1686	.1708	.1522	.1543	.1564
Losowy	$\theta = .3$	1	.2584	.2620	.2655	.2361	.2395	.2428	.2175	.2207	.2238
		10	.2598	.2633	.2669	.2374	.2408	.2441	.2187	.2219	.2250
		100	.2733	.2767	.2801	.2502	.2534	.2566	.2308	.2338	.2369
	$\theta = .4$	1	.2280	.2313	.2345	.2078	.2109	.2139	.1910	.1938	.1967
		10	.2293	.2325	.2357	.2090	.2120	.2150	.1921	.1949	.1978
		100	.2414	.2445	.2476	.2204	.2233	.2262	.2028	.2056	.2083
	$\theta = .5$	1	.2039	.2068	.2097	.1854	.1881	.1908	.1700	.1726	.1751
		10	.2050	.2080	.2108	.1864	.1892	.1919	.1710	.1735	.1761
		100	.2160	.2188	.2216	.1967	.1993	.2020	.1807	.1831	.1856
	$\theta = .7$	1	.1734	.1759	.1785	.1570	.1593	.1617	.1434	.1457	.1478
		10	.1744	.1769	.1794	.1579	.1602	.1626	.1443	.1465	.1487
		100	.1839	.1863	.1887	.1667	.1690	.1712	.1526	.1547	.1568

Literatura

- [1] Amrhein, P. (1995). Minimax estimation of proportions under random sample size. *JASA* **90**, 1107–1111.
- [2] Ashour, S.K., El-Wakeel, M.A.A.H. (1994). Bayesian prediction of the median of the Burr distribution with fixed and random sample sizes. *Statistics*, **25** 113–122.
- [3] Aubin, J.P. (1979). Mathematical Methods of Game and Economic Theory. North-Holland Publishing Company.
- [4] Badasz, A., Molisz, W., Seidler, J. (1980). Metody Rozwiązywania Zadań Optymalizacji. WNT, Warszawa.
- [5] Breiman, L. (1968). Probability. Addison–Wesley, Reading.
- [6] Brown, L.D. (1990). An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.* **18**, 471–493.
- [7] Chaudhuri, A., Stenger, H. (1992). Survey Sampling–Theory and Methods. New York: Marcel Dekker.
- [8] Chen, E.E., Wardrop, R.L. (1980). Bayes sequential estimation in a life test and asymptotic properties. *Commun. Statist. - Theor. Meth.* **A 9(6)**, 659–672.
- [9] Chou, Y.M., Owen, D.B. (1986). One-sided distribution-free simultaneous prediction limits for p future samples. *Journal of Quality Technology* **18**, 96–98.
- [10] Chou, Y.M., Owen, D.B. (1989). Simultaneous one-sided prediction intervals for a two-parameter exponential distribution using complete or type II censored data. *Metrika* **36**, 279–290.
- [11] Consul, P.C. (1984). On the distribution of order statistics for a random sample size. *Statistica Neerlandica* **38**, 249–256.
- [12] Dunsmore, I.R. (1974). The Bayesian predictive distribution in life testing models. *Technometrics* **16**, 455–460.
- [13] Evans, I. G., Nigm, A. H. M. (1980). Bayesian prediction for the left truncated exponential distribution. *Technometrics* **22**, 201–204.
- [14] Ferguson, T. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York and London.

- [15] Fichtenholz, G.M. (1985). Rachunek różniczkowy i całkowy. PWN, Warszawa.
- [16] He, K. (1990). An ancillarity paradox in the estimation of multinomial probabilities. *JASA* **85**, 824–828.
- [17] Hodges, J. L., Lehmann, E. L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.* **21**, 182–196.
- [18] Hodges, J. L., Lehmann, E. L. (1951). Some applications of the Cramer–Rao inequality. *Second Berkeley Symp. Math. Statist. Prob.* **1**, 13–22.
- [19] Jokiel–Rokita, A., Magiera, R. (1995). Estimation with delayed observations for the multinomial distributions. Wysłano do opublikowania.
- [20] Jokiel–Rokita, A. (1997 a). Bayesian prediction of the median of the left truncated exponential distribution with fixed and random sample size. Wysłano do opublikowania.
- [21] Jokiel–Rokita, A. (1997 b). Minimax prediction for the multinomial and multivariate hypergeometric distributions. Przyjęto do publikacji w *Applicationes Mathematicae*.
- [22] Jokiel–Rokita, A. (1998). Minimax prediction for the multinomial and multivariate hypergeometric distributions with random sample size. Wysłano do opublikowania.
- [23] Karmanov, V.G. (1986). Matematicheskoye Programmirovaniye. Moskov.
- [24] Lawless, J.F. (1972). On prediction intervals for samples from the exponential distribution and prediction limits for system survival. *Sankhya-B* **34**, 725–730.
- [25] Lawless, J.F. (1977). Prediction intervals for the two parameter exponential distribution. *Technometrics* **19**, 469–472.
- [26] Lehmann, E.L (1983). Theory of Point Estimation. New York: John Wiley.
- [27] Lehmann, E.L (1986). Testing Statistical Hypotheses. New York: John Wiley.
- [28] Magiera, R. (1982). Estimation with delayed observations. *Applicationes Mathematicae* **17**, 249–258.
- [29] Ross, S.M. (1971). Infinitesimal look-ahead stopping rules. *Ann. Math. Statist.* **42**, 297–303.
- [30] Salem, S.A., El-glad, F.A. (1992). Bayesian prediction for the future r -th order statistics based on two parameter exponential distribution with random sample size. *The Egyptian Statistical Journal Cairo University* **36**, 255–273.
- [31] Salem, S.A., El-glad, F.A. (1993). Bayesian prediction for the range based on a two parameter exponential distributions with a random sample sizes. *Microelectron Reliab.* **33**, 623–632.
- [32] Shapiro, C.P., Wardrop, R.L. (1980). Dynkin’s identity applied to Bayes sequential estimation of a Poisson process rate. *Ann. Statist.* **8**, 171–182.
- [33] Stadje, W. (1990). A sequential estimation procedure for the parameter of an exponential distribution. *Statistics* **21**, 239–250.

- [34] Starr, N., Wardrop, R. and Woodroffe, M. (1976). Estimating a mean from delayed observations. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35**, 103–113.
- [35] Steinhaus, H. (1957). The problem of estimation. *Ann. Math. Statist.* **28**, 633–648.
- [36] Trybuła, S. (1958). Some problems of simultaneous minimax estimation. *Ann. Math. Statist.* **29**, 245–253.
- [37] Trybuła, S. (1985). Some investigations in minimax estimation theory. *Diss. Math.* **260**.
- [38] Trybuła, S. (1987 a). Minimax estimation and prediction for random variables with bounded sum. *Applicationes Mathematicae* **19**, 85–91.
- [39] Trybuła, S. (1987 b). Minimax prediction of the sum of processes. *Applicationes Mathematicae* **19**, 195–201.
- [40] Vorob'ev, N. N. (1977). Game Theory. Lectures for Economists and System Scientists. Springer-Verlag.
- [41] Wald, A. (1950). Statistical Decision Functions. New York: John Wiley.
- [42] Wilczyński, M. (1985). Minimax estimation for the multinomial and multivariate distributions. *Sankhyā* **47**, 128–132.

