Ryszard Magiera

STATYSTYCZNE FUNKCJE DECYZYJNE

Wydanie II rozszerzone

Oficyna Wydawnicza GiS
Wrocław 2018

ryszard.magiera@pwr.edu.pl

Autor projektu okładki *Ryszard Magiera*. Główny motyw obrazu tworzą wykresy powierzchni funkcji straty *SQUARED, BLINEX, GUMBEL, PRECAUTIONARY* i *LINEX*.

Printed in Poland

Skład komputerowy książki w systemie LaTeX wykonał autor. Do sporządzenia wykresów wykorzystano procedury pakietu komputerowego *Mathematica* 11 (licencja L4691-7788).

ISBN 978–83–62780–55–6

Wydanie II, Wrocław 2018
Oficyna Wydawnicza GiS, s.c.
Druk i oprawa: Drukarnia I-BiS sp. z o.o., A.Bieroński, P.Bieroński s.j.

Przedmowa

W książce przedstawiono metody wyznaczania optymalnych, przy różnych kryteriach, statystycznych funkcji decyzyjnych stosując różne podejścia do wnioskowań statystycznych dotyczących estymacji i testowania hipotez. W podejściach tych, statystyk może mierzyć wielkość poniesionych strat związanych z podjęciem błędnej decyzji przyjmując różne postaci funkcji straty, odpowiednie dla danego problemu statystycznego.

Najwięcej miejsca poświęcono podejściu bayesowskiemu, w którym zakłada się, że podlegający wnioskowaniu nieznany parametr modelu statystycznego jest realizacją obserwowanej zmiennej losowej. Podejście to dominuje ostatnio również w różnych naukach praktycznych, gdyż umożliwia wykorzystanie pewnej informacji a priori z przeszłości lub wiedzy eksperckiej. Opisano również podejście minimaksowe, przy którym statystyk podejmując decyzję kieruje się kryterium minimalizacji maksymalnej straty.

Książka zawiera również rozdział poświęcony metodom wnioskowania statystycznego opartego na próbie, której rozmiar nie jest z góry ustalony, ale jest zmienną losową zależną od uzyskiwanych wcześniej obserwacji. Takie podejście do wnioskowania statystycznego, zwane sekwencyjnym, ma szczególne znaczenie, gdy należy wziąć pod uwagę koszty związane z uzyskiwaniem danych w kolejnych krokach (chwilach), np. w badaniach niezawodności systemów lub badaniach medycznych. Rozdział kończy się szczegółową prezentacją przykładów procedur sekwencyjnych w modelach statystycznej kontroli jakości.

W przedstawionych metodach wyznaczania optymalnych decyzji statystyka, stosującego różne podejścia i kryteria, zwraca się jednocześnie uwagę na doniosłość stosowania przy wyborze decyzji dwóch podstawowych zasad statystyki matematycznej – zasady dostateczności i zasady niezmienniczości.

Książka adresowana jest przede wszystkim do środowiska akademickiego, w szczególności do studentów studiujących statystykę matematyczną. Może być przydatna również dla studentów ekonometrii, fizyki oraz różnych dziedzin nauk technicznych. Zakłada się, że Czytelnik zna podstawowe pojęcia i twierdzenia dotyczące dwóch głównych rodzajów wnioskowania statystycznego: estymacji i testowania hipotez.

Książka w dużej części zawiera materiał kursu *Statystyka Matematyczna*, który prowadziłem przez wiele lat na II stopniu kierunku *Matematyka* w Politechnice Wrocławskiej. Materiał zawarty w książce wykorzystywałem również na wykładzie *Teoria Statystycznych Funkcji Decyzyjnych* prowadzonym dla doktorantów.

Teorię ilustrują liczne przykłady, a dodatkowo książkę uzupełniają problemy i zadania zawarte w ostatnich podrozdziałach Rozdziałów 1 – 6. Do wielu z tych zadań można znaleźć rozwiązania w książce [21]. W spisie literatury zostały umieszczone pozycje, z których korzystałem przy opracowywaniu tematów zawartych w niniejszej książce.

W celu ułatwienia korzystania z literatury w języku angielskim oraz z angielskich wersji komputerowych pakietów statystycznych, polskie nazwy pojęć statystycznych uzupełniono w tekście terminologią angielską. Ponadto, oprócz skorowidza terminów polskich został dołączony skorowidz terminów angielskich.

W wydaniu II dokonano wiele zmian redakcyjnych polegających przede wszystkim na uzupełnieniu niektórych tematów, w szczególności w Rozdziałach 2 i 6. Dodano wiele przykładów, zmieniono graficzną prezentację większości wykresów oraz poprawiono zauważone błędy redakcyjne i typograficzne.

Wrocław, wrzesień 2018 *Ryszard Magiera*

Spis treści

Przedmowa **v**

Spis oznaczeń **x**

Spis tabel **xiv**

Spis wykresów **xvi**

1 Gra statystyczna i funkcje decyzyjne **1**
1.1 Gra dwuosobowa, gra decyzyjna i gra statystyczna 1
1.2 Randomizacja funkcji decyzyjnych i zrandomizowana gra statystyczna 5
1.3 Gra statystyczna jako podejście do wnioskowań statystycznych . 12
1.4 Funkcje straty w problemach estymacji 15
1.4.1 Funkcje straty w problemach estymacji parametrycznej . 15
1.4.2 Funkcje straty w problemach estymacji nieparametrycznej 24
1.5 Koncepcja dopuszczalności funkcji decyzyjnych 25
1.6 Nieobciążoność funkcji decyzyjnych 26
1.7 Zadania 28

2 Bayesowskie funkcje decyzyjne **37**
2.1 Podejście bayesowskie 37
2.1.1 Rozkład a priori 37
2.1.2 Rozkład a posteriori 38
2.1.3 Sprzężona rodzina rozkładów a priori 40
2.1.4 Wykładniczy model statystyczny 41
2.1.5 Sprzężona rodzina rozkładów a priori w modelu wykładniczym 43
2.1.6 Nieinformujące rozkłady a priori 49
2.2 Pojęcie bayesowskiej funkcji decyzyjnej 51
2.3 Ogólna metoda wyznaczania bayesowskich funkcji decyzyjnych . 52
2.4 Estymatory bayesowskie dla różnych postaci funkcji straty 55
2.5 Estymatory bayesowskie w wykładniczym modelu statystycznym 62

2.5.1 Estymatory bayesowskie przy kwadratowej funkcji straty . 62
2.5.2 Estymatory bayesowskie przy ważonej kwadratowej funkcji straty 66
2.5.3 Optymalny rozmiar próby 70
2.6 Testy bayesowskie 72
2.6.1 Test bayesowski oparty na funkcji straty 72
2.6.2 Test bayesowski oparty na ilorazie szans 74
2.6.3 Test bayesowski dla hipotez prostych 75
2.6.4 Porównania testów bayesowskich z testem klasycznym . . 77
2.7 Bayesowskie centralne twierdzenie graniczne 80
2.8 Bayesowska nierówność Craméra-Rao 84
2.9 Metody Monte Carlo we wnioskowaniu bayesowskim 87
2.9.1 Próbkowanie znaczące 87
2.9.2 Algorytm Metropolisa-Hastingsa 87
2.10 Zadania 92

3 Minimaksowe funkcje decyzyjne 103
3.1 Kryterium minimaksowe 103
3.2 Metody wyznaczania minimaksowych funkcji decyzyjnych 106
3.3 Twierdzenia minimaksowe 114
3.4 Zadania 115

4 Dopuszczalność statystycznych funkcji decyzyjnych 121
4.1 Klasy zupełne funkcji decyzyjnych 121
4.2 Funkcje decyzyjne oparte na statystyce dostatecznej 124
4.3 Dopuszczalność estymatorów w przestrzeniach regularnych 125
4.4 Dopuszczalność bayesowskich funkcji decyzyjnych 129
4.5 Dopuszczalność minimaksowych funkcji decyzyjnych 134
4.6 Zadania 136

5 Niezmienniczość funkcji decyzyjnych 139
5.1 Zasada niezmienniczości 139
5.2 Niezmienniczy model statystyczny 140
5.3 Rodziny rozkładów z parametrem położenia i parametrem skali . 141
5.4 Niezmienniczość funkcji decyzyjnych 143
5.5 Statystyki niezmiennicze 146
5.6 Estymatory ekwiwariantne 147
5.6.1 Estymatory ekwiwariantne parametru położenia 147
5.6.2 Estymatory ekwiwariantne parametru skali 151
5.7 Testy niezmiennicze 153
5.8 Nieinformujące rozkłady a priori w modelach niezmienniczych . . 156
5.9 Zadania 157

6 Sekwencyjne decyzje statystyczne **161**
6.1 Analiza sekwencyjna . . . 161
6.2 Procedura sekwencyjna . . . 163
6.3 Bayesowska procedura sekwencyjna . . . 164
6.4 Procedura bayesowska o stałym rozmiarze próby . . . 167
6.5 Procedura bayesowska obcięta . . . 170
6.6 Procedura o ustalonej liczbie kroków naprzód . . . 177
6.7 Podstawowa tożsamość analizy sekwencyjnej . . . 180
6.8 Sekwencyjny test ilorazowy . . . 185
6.8.1 Podstawy . . . 185
6.8.2 Modyfikacje sekwencyjnego testu ilorazowego . . . 191
6.9 Procedury sekwencyjne w prostych modelach statystycznej kontroli jakości . . . 194
6.9.1 Sekwencyjny test ilorazowy dla prawdopodobieństwa sukcesu . . . 194
6.9.2 Plan inspekcji skróconej . . . 197
6.10 Zadania . . . 201

Tablice rozkładów **208**

Literatura **216**

Skorowidz terminów polskich **221**

Skorowidz terminów angielskich **224**

1

Gra statystyczna i funkcje decyzyjne

Statistical Games and Statistical Decision Functions

1.1 Gra dwuosobowa, gra decyzyjna i gra statystyczna

Gra dwuosobowa

Podstawowe koncepcje teorii statystycznych funkcji decyzyjnych mają swe źródło w teorii gier dwuosobowych o sumie zerowej.

Grą dwuosobową [**two person game**] nazywamy grę, w której bierze udział dwóch graczy lub dwie stałe koalicje gry wieloosobowej.

Grą w postaci normalnej [**normal-form game**] nazywamy trójkę (X, Y, W), gdzie X i Y są ustalonymi (niepustymi) zbiorami strategii odpowiednio gracza I i II, a $W(x, y)$, $x \in X$, $y \in Y$, jest *funkcją wypłat* [**payoff function**], określoną dla każdej pary strategii x i y odpowiednio gracza I i II.

Grę dwuosobową nazywamy *grą o sumie zerowej* [**two person zero sum game**], jeżeli wygrana jednego gracza jest jednocześnie przegraną gracza drugiego.

Gra decyzyjna

Grą decyzyjną [**decision game**] nazywamy *grę dwuosobową o sumie zerowej*, w której rolę gracza pierwszego pełni natura, a rolę gracza drugiego – statystyk. W tej grze statystyka z naturą, zbiór strategii gracza I nazywamy zbiorem możliwych stanów natury i oznaczamy przez Θ, natomiast zbiór strategii gracza

II nazywamy zbiorem możliwych akcji statystyka i oznaczamy przez $\mathcal{A}$. Rolę *funkcji wypłaty* pełni tu *funkcja straty.*

DEFINICJA **1.1** *Funkcją straty* [**loss function**] nazywamy funkcję $\mathcal{L}(\vartheta, a) : \Theta \times \mathcal{A} \mapsto \mathbb{R}$, gdzie Θ jest zbiorem stanów natury, a $\mathcal{A}$ zbiorem wszystkich możliwych akcji statystyka. □

Funkcja $\mathcal{L}(\vartheta, a)$ określa stratę statystyka (lub zysk, gdy wartość jej jest ujemna) wynikłą z podjęcia akcji $a \in \mathcal{A}$, gdy prawdziwym stanem natury jest $\vartheta \in \Theta$. Zwykle zakłada się, że $\mathcal{L}(\vartheta, a) \geqslant -c > -\infty$.

W sensie matematycznym, gra decyzyjna wyznaczona jest przez trójkę (Θ, $\mathcal{A}$, $\mathcal{L}$). Przestrzeń mierzalną $(\mathcal{A}, \mathcal{G})$, gdzie $\mathcal{G}$ jest wyróżnionym σ-ciałem podzbiorów ze zbioru $\mathcal{A}$, nazywamy *przestrzenią akcji* [**action space**].

Istnieją ważne różnice między *grą dwuosobową* (X, Y, W) w postaci normalnej a grą decyzyjną $(\Theta, \mathcal{A}, \mathcal{L})$. Różnice te wynikają głównie z zachowań graczy w tych grach. Oto trzy najważniejsze różnice.

1. W grze (X, Y, W) obaj gracze dążą do uzyskania maksymalnej wygranej (czyli do zminimalizowania swoich strat), podczas gdy w grze decyzyjnej (Θ, $\mathcal{A}$, $\mathcal{L}$) natura wybiera swój stan $\vartheta \in \Theta$ bez takiego celu. Jest to zasadnicza różnica między grą (X, Y, W) a grą decyzyjną $(\Theta, \mathcal{A}, \mathcal{L})$, w której do wygranej dąży jedynie statystyk.

2. W grze (X, Y, W) każdy z graczy uzależnia swoje racjonalne zachowanie od inteligencji przeciwnika, podczas gdy w grze decyzyjnej kryterium racjonalnego zachowania się natury może nie istnieć, a nawet gdyby istniało, to statystyk może go nie znać.

PRZYKŁAD **1.1**

Niech $(\Theta, \mathcal{A}, \mathcal{L})$ będzie grą, w której $\Theta = \{\vartheta_1, \vartheta_2\}$, $\mathcal{A} = \{a_1, a_2\}$ i w której funkcja straty $\mathcal{L}$ określona jest w postaci podanej tablicy.

$\Theta \backslash \mathcal{A}$	a_1	a_2
ϑ_1	1	3
ϑ_2	0	-5
	$\mathcal{L}(\vartheta, a)$	

W grze dwuosobowej, w której zakłada się, że gracz I dysponujący strategiami ϑ_1 i ϑ_2 jest obdarzony inteligencją, jedynym racjonalnym jego postępowaniem jest wybór strategii ϑ_1, ponieważ bez względu na to jak postąpi przeciwnik, jego wygrana będzie większa, jeśli wybierze ϑ_1 zamiast ϑ_2. Wtedy statystyk powinien wybrać akcję a_1 zamiast a_2, bo wtedy straci tylko jedną jednostkę zamiast trzech. Jednak w grze decyzyjnej, natura może wybrać zarówno stan ϑ_1 jak ϑ_2 nie kierując się żadną rozsądną regułą. Jeśli wybierze ϑ_2, to wybór akcji a_2 przez statystyka daje mu większą korzyść. □

3. W grze decyzyjnej zakłada się, że natura wybiera swój stan raz na zawsze, a statystyk ma możliwość uzyskiwania informacji o wybranym przez naturę stanie przeprowadzając odpowiednie eksperymenty. Akcja statystyka zależy często od obserwacji pewnej zmiennej losowej lub pewnego procesu losowego, których rozkład zależy od stanu natury.

Gra statystyczna

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\})$ będzie *przestrzenią statystyczną* związaną z obserwacją zmiennej losowej X.

DEFINICJA **1.2** *Grą statystyczną* [**statistical game**] lub *statystycznym problemem decyzyjnym* [**statistical decision problem**] nazywamy *grę decyzyjną* $(\Theta, \mathcal{A}, \mathcal{L})$, która jest powiązana z eksperymentem umożliwiającym obserwowanie pewnej zmiennej losowej X, o wartościach (realizacjach) $x \in \mathcal{X}$, której rozkład P_ϑ zależy od stanu $\vartheta \in \Theta$ wybranego przez naturę. Przestrzeń mierzalną $(\mathcal{X}, \mathcal{B})$ nazywamy *przestrzenią prób* [**sample space**]. □

W statystycznym problemie decyzyjnym statystyk obserwuje zmienną losową X i po zaobserwowaniu $X = x$ podejmuje akcję $a = d(x) \in \mathcal{A}$.

DEFINICJA **1.3** *Niezrandomizowaną funkcją decyzyjną* [**nonrandomized decision function**] nazywamy funkcję mierzalną (statystykę) $d : (\mathcal{X}, \mathcal{B}) \mapsto (\mathcal{A}, \mathcal{G})$ z przestrzeni prób na przestrzeń akcji. □

Zbiór niezrandomizowanych funkcji decyzyjnych oznaczamy przez D.

W grze statystycznej, strata $\mathcal{L}(\vartheta, d(X))$ jest dla każdego ustalonego $\vartheta \in \Theta$ zmienną losową o rozkładzie wyznaczonym przez P_ϑ.

Grę decyzyjną $(\Theta,\ \mathcal{A},\ \mathcal{L})$ połączoną z obserwacją zmiennej losowej X o rozkładzie z rodziny $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ można zastąpić grą, w której zamiast zbioru akcji $\mathcal{A}$ rozpatruje się zbiór D *niezrandomizowanych funkcji decyzyjnych*, natomiast zamiast funkcji straty $\mathcal{L}(\vartheta, a)$ – *funkcję ryzyka* $R(\vartheta,\ d)$, będącą wartością oczekiwaną zmiennej losowej $\mathcal{L}(\vartheta, d(X))$.

DEFINICJA **1.4** *Funkcją ryzyka* $R(\vartheta, d)$ *niezrandomizowanej funkcji decyzyjnej* [**risk function of a nonrandomized decision function**] d nazywamy wartość oczekiwaną *funkcji straty* $\mathcal{L}(\vartheta,\ d(X))$, tzn.

$$R(\vartheta, d) = E_\vartheta[\mathcal{L}(\vartheta, d(X))] = \int_{\mathcal{X}} \mathcal{L}(\vartheta, d(x)) dP_\vartheta(x), \tag{1.1}$$

liczoną przy założeniu, że ϑ jest prawdziwą wartością stanu natury. □

Funkcja ryzyka $R(\vartheta, d)$ określa przeciętną stratę poniesioną przez statystyka, gdy d jest wybraną przez niego decyzją, a ϑ jest prawdziwym stanem natury.

DEFINICJA **1.5** *Niezrandomizowaną grą statystyczną* [**nonrandomized statistical game**] nazywamy trójkę (Θ, D, R), w której D jest zbiorem niezrandomizowanych funkcji decyzyjnych, a R oznacza funkcję ryzyka. □

Zrandomizowane funkcje decyzyjne i ich ryzyka określone będą w Podrozdziale 1.2.

PRZYKŁAD **1.2**
Niech $\Theta = \mathcal{A} = \{1, 2\}$ i niech funkcja straty $\mathcal{L}$ określona będzie w postaci podanej tablicy.
Zakłada się, że statystyk obserwuje zmienną losową X przyjmującą dwie wartości 1 i 2, przy czym

$\Theta\backslash\mathcal{A}$	1	2
1	0	1
2	1	-1
	$\mathcal{L}(\vartheta, a)$	

$$P_1(X=1) := P(X=1|\vartheta=1) = P_2(X=2) := P(X=2|\vartheta=2) = 2/3$$
$$P_2(X=1) := P(X=1|\vartheta=2) = P_1(X=2) := P(X=2|\vartheta=1) = 1/3,$$

tzn. obserwacja daje prawidłową odpowiedź z prawdopodobieństwem 2/3 i nieprawidłową z prawdopodobieństwem 1/3. Zatem $\mathcal{X} = \{1, 2\}$ i istnieją cztery funkcje decyzyjne $d(x) : \mathcal{X} \mapsto \mathcal{A}$; mianowicie

$$d_1(1) = 1, \quad d_2(1) = 1, \quad d_3(1) = 2, \quad d_4(1) = 2,$$
$$d_1(2) = 1, \quad d_2(2) = 2, \quad d_3(2) = 1, \quad d_4(2) = 2.$$

Wartości funkcji ryzyka obliczamy zgodnie ze wzorem (1.1), który w rozpatrywanym przykładzie sprowadza się do postaci

$$R(\vartheta, d) = \sum_i \mathcal{L}(\vartheta, d(x_i))P_\vartheta(X = x_i).$$

Mamy więc
$R(1, d_1) = \mathcal{L}(1, 1)P_1(X = 1) + \mathcal{L}(1, 1)P_1(X = 2) = \mathcal{L}(1, 1) = 0,$
$R(2, d_1) = \mathcal{L}(2, 1)P_2(X = 1) + \mathcal{L}(2, 1)P_2(X = 2) = \mathcal{L}(2, 1) = 1,$
$R(1, d_2) = \mathcal{L}(1, 1)P_1(X = 1) + \mathcal{L}(1, 2)P_1(X = 2) = (1/3)\mathcal{L}(1, 2) = 1/3,$
$R(2, d_2) = \mathcal{L}(2, 1)P_2(X = 1) + \mathcal{L}(2, 2)P_2(X = 2)$
$\qquad = (1/3)\mathcal{L}(2, 1) + (2/3)\mathcal{L}(2, 2) = -1/3,$
$R(1, d_3) = (2/3)\mathcal{L}(1, 2) = 2/3,$
$R(2, d_3) = (1/3)\mathcal{L}(2, 2) + (2/3)\mathcal{L}(2, 1) = 1/3,$
$R(1, d_4) = \mathcal{L}(1, 2) = 1,$
$R(2, d_4) = \mathcal{L}(2, 2) = -1.$

Zatem, wyjściowa gra decyzyjna $(\Theta, \mathcal{A}, \mathcal{L})$ została zastąpiona niezrandomizowaną grą statystyczną (Θ, D, R). Wartości funkcji ryzyka $R(\vartheta, d), d \in D = \{d_1, d_2, d_3, d_4\}$ można przedstawić w postaci podanej tablicy.

$\Theta\backslash D$	d_1	d_2	d_3	d_4
1	0	1/3	2/3	1
2	1	-1/3	1/3	-1
		$R(\vartheta, d)$		

□

PRZYKŁAD **1.3** Niech $\mathbf{X} = (X_1, \ldots, X_n)$ będzie próbą z rozkładu Bernoulliego $\mathcal{B}(1, \vartheta)$. Rozpatrzmy problem oszacowania prawdopodobieństwa sukcesu ϑ przy kwadratowej funkcji straty $\mathcal{L}(\vartheta, d) = (\vartheta - d)^2$.

Weźmy pod uwagę funkcję decyzyjną

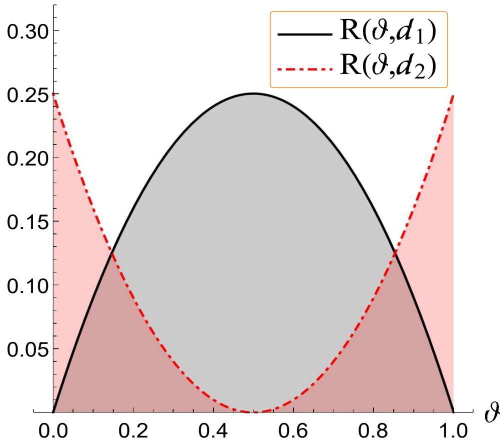
$$d = d_1(\mathbf{X}) = \overline{X},$$

gdzie $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ (średnia z próby $\mathbf{X}$), i funkcję decyzyjną

$$d = d_2(\mathbf{X}) = 1/2.$$

Odpowiednie funkcje ryzyka są postaci

$$R(\vartheta, d_1) = -\vartheta^2 + \vartheta,$$
$$R(\vartheta, d_2) = \vartheta^2 - \vartheta + 1/4.$$



Rys. 1.1: Wykresy funkcji ryzyka przy kwadratowej funkcji straty dla dwóch wybranych funkcji decyzyjnych.

□

1.2 Randomizacja funkcji decyzyjnych i zrandomizowana gra statystyczna

Niezrandomizowaną grę statystyczną (Θ, D, R) można rozszerzyć poprzez wprowadzenie zrandomizowanych decyzji statystycznych. Istnieją dwa sposoby randomizacji decyzji statystyka.

Jeden sposób randomizacji polega na wprowadzeniu miary probabilistycznej δ^* na przestrzeni $(D, \mathcal{F})$ funkcji decyzyjnych, gdzie $\mathcal{F}$ oznacza najmniejsze σ-ciało, względem którego *funkcje ryzyka* $R(\vartheta, \cdot)$ są mierzalne dla każdego $\vartheta \in \Theta$ oraz zawierające jako swoje elementy wszystkie jednoelementowe podzbiory zbioru D.

DEFINICJA **1.6** *Mieszaną funkcją decyzyjną* [**mixed decision function**] nazywamy miarę probabilistyczną δ^* określoną na przestrzeni mierzalnej $(D, \mathcal{F})$, dla której istnieje, dla każdego $\vartheta \in \Theta$, całka

$$R^*(\vartheta, \delta^*) = E^{\delta^*}[R(\vartheta, d)] = \int_D R(\vartheta, d) d\delta^*(d). \tag{1.2}$$

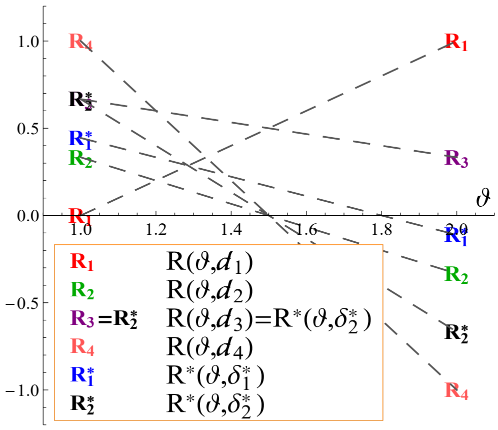
Funkcję $R^*(\vartheta, \delta^*), \vartheta \in \Theta$, nazywamy *funkcją ryzyka mieszanej funkcji decyzyjnej* [**risk function of a mixed decision function**] δ^*. □

Funkcja $R(\vartheta, d)$ we wzorze (1.2) oznacza funkcję ryzyka niezrandomizowanej funkcji decyzyjnej d. Zbiór wszystkich mieszanych funkcji decyzyjnych oznaczamy przez D^*.

Zbiór D można utożsamić z podzbiorem zbioru D^*, zawierającym rozkłady prawdopodobieństwa skupione w pojedynczych punktach zbioru D. W ten sposób gra $(\Theta,\ D,\ R)$ została rozszerzona do gry (Θ, D^*, R^*), w której strategiami statystyka są miary probabilistyczne określone na przestrzeni mierzalnej $(D, \mathcal{F})$. Grę (Θ, D^*, R^*) nazywamy *zrandomizowaną grą statystyczną* [**randomized statistical game**].

PRZYKŁAD **1.4** (cd. Przykładu 1.2) Weźmy pod uwagę dwie mieszane funkcje decyzyjne: $\delta_1^* = (\delta_{1,1}^*, \ldots, \delta_{1,4}^*) = (1/3, 1/3, 0, 1/3)$ i $\delta_2^* = (\delta_{2,1}^*, \ldots, \delta_{2,4}^*) = (0, 1/2, 0, 1/2)$. Biorąc pod uwagę wartości funkcji ryzyka $R(\vartheta, d_i)$ dla niezrandomizowanych funkcji decyzyjnych $d_i, i = 1, 2, 3, 4$, (z Przykładu 1.2), wyznaczone wartości funkcji ryzyka $R^*(\vartheta, \delta_1^*)$ i $R^*(\vartheta, \delta_2^*)$ dla mieszanych funkcji decyzyjnych δ_1^* i δ_2^* wyznaczone są poniżej:

$$
\begin{aligned}
R^*(\vartheta, \delta_1^*) &= \sum_{i=1}^{4} R(\vartheta, d_i)\delta_{1,i}^* \\
&= \begin{cases} 4/9, & \text{gdy } \vartheta = 1, \\ -1/9, & \text{gdy } \vartheta = 2, \end{cases} \\
R^*(\vartheta, \delta_2^*) &= \sum_{i=1}^{4} R(\vartheta, d_i)\delta_{2,i}^* \\
&= \begin{cases} 2/3, & \text{gdy } \vartheta = 1, \\ -2/3, & \text{gdy } \vartheta = 2. \end{cases}
\end{aligned}
$$



Rys. 1.2: Graficzne porównanie wartości funkcji ryzyka $R(\vartheta, d_i), i = 1, 2.3, 4$, z wartościami funkcji ryzyka $R^*(\vartheta, \delta_1^*)$ i $R^*(\vartheta, \delta_2^*)$.

□

Drugi sposób randomizacji polega na wprowadzeniu miar probabilistycznych Q na przestrzeni akcji $(\mathcal{A}, \mathcal{G})$ gry wyjściowej $(\Theta, \mathcal{A}, \mathcal{L})$, gdzie $\mathcal{G}$ oznacza najmniejsze σ-ciało, względem którego funkcje straty $\mathcal{L}(\vartheta, \cdot)$ są mierzalne dla każdego $\vartheta \in \Theta$ oraz zawierające jako swoje elementy wszystkie jednoelementowe podzbiory zbioru $\mathcal{A}$.

Niech $\mathcal{A}^*$ oznacza zbiór wszystkich miar probabilistycznych Q określonych na przestrzeni akcji $(\mathcal{A}, \mathcal{G})$, dla których istnieje, dla każdego $\vartheta \in \Theta$, całka

$$\mathcal{L}^*(\vartheta, Q) = E^Q[\mathcal{L}(\vartheta, a)] = \int_{\mathcal{A}} \mathcal{L}(\vartheta, a)dQ(a).$$

Funkcja $\mathcal{L}^*(\vartheta, Q), \vartheta \in \Theta$, nazywa się *funkcją straty akcji zrandomizowanej* [**loss function of a randomized action**]. Zbiór $\mathcal{A}$ możemy uważać za podzbiór zbioru $\mathcal{A}^*$, utożsamiając jego elementy z miarami probabilistycznymi skupionymi w pojedynczych punktach $a \in \mathcal{A}$. W ten sposób gra $(\Theta, \mathcal{A}, \mathcal{L})$ została rozszerzona do gry $(\Theta, \mathcal{A}^*, \mathcal{L}^*)$. Gra $(\Theta, \mathcal{A}^*, \mathcal{L}^*)$ nie jest powiązana z obserwacją zmiennej losowej X.

Niech X będzie obserwowalną zmienną losową o wartościach w $\mathcal{X}$ i o rozkładzie $P_\vartheta, \vartheta \in \Theta$.

DEFINICJA **1.7** Funkcję $\overline{\delta}$ określoną na $\mathcal{X} \times \mathcal{G}$, taką że dla każdego $A \in \mathcal{G}$, $\overline{\delta}(\cdot, A)$ jest mierzalna, a dla każdego $x \in \mathcal{X}$ jest miarą probabilistyczną $\overline{\delta}(x, \cdot)$ na $(\mathcal{A}, \mathcal{G})$ nazywamy *behawiorystyczną funkcją decyzyjną* [**behavioral decision function**]. □

Rozpatrywać będziemy tylko takie behawiorystyczne funkcje decyzyjne, dla których istnieje ryzyko

$$\begin{aligned}\overline{R}(\vartheta, \overline{\delta}) &= E_\vartheta[\mathcal{L}^*(\vartheta, \overline{\delta}(X, \cdot))] = E_\vartheta\left\{E^{\overline{\delta}(X,\cdot)}[\mathcal{L}(\vartheta, a)]\right\} \\ &= E_\vartheta\left[\int_{\mathcal{A}} \mathcal{L}(\vartheta, a) d\overline{\delta}(X, a)\right] = \int_{\mathcal{X}} \mathcal{L}^*(\vartheta, \overline{\delta}(x, \cdot)) dP_\vartheta(x) \qquad (1.3) \\ &= \int_{\mathcal{X}} \int_{\mathcal{A}} \mathcal{L}(\vartheta, a) d\overline{\delta}(x, a) dP_\vartheta(x).\end{aligned}$$

Podejmując decyzję behawiorystyczną, statystyk wybiera rozkład prawdopodobieństwa $\overline{\delta}(x, \cdot) \in \mathcal{A}^*$, gdy zaobserwowaną wartością zmiennej losowej X jest x; $\overline{\delta}(x, A)$ jest prawdopodobieństwem tego, że zostanie wybrana akcja ze zbioru A. Aby wybrać akcję w przestrzeni $\mathcal{A}$, należy wygenerować element pseudolosowy przestrzeni $\mathcal{A}$ zgodnie z rozkładem $\overline{\delta}(x, \cdot)$. Zatem, innym sposobem opisu zrandomizowanej funkcji decyzyjnej $\overline{\delta}$ jest podanie metody symulacji akcji z przestrzeni $\mathcal{A}$ dla każdej wartości $x \in \mathcal{X}$. Niezrandomizowaną funkcję decyzyjną $d(x)$ można traktować jako szczególny przypadek randomizacji $\overline{\delta}(x, A) = \mathbf{1}_A(d(x))$.

Funkcję $\overline{R}(\vartheta, \overline{\delta}), \vartheta \in \Theta$, nazywamy *funkcją ryzyka behawiorystycznej funkcji decyzyjnej* [**risk function of a behavioral decision function**]. Zbiór wszystkich behawiorystycznych funkcji decyzyjnych oznaczamy przez $\overline{D}$. Ponieważ zbiór $\overline{D}$ zawiera, jako swój podzbiór zbiór D funkcji z przestrzeni $\mathcal{X}$ w podzbiór $\mathcal{A}$ zbioru $\mathcal{A}^*$, więc w ten sposób gra (Θ, D, R) została rozszerzona do gry $(\Theta, \overline{D}, \overline{R})$, którą również nazywamy *zrandomizowaną grą statystyczną* [**randomized statistical game**].

PRZYKŁAD **1.5** (test zrandomizowany jako behawiorystyczna funkcja decyzyjna) Załóżmy, że statystyk obserwuje zmienną losową X o rozkładzie dwumianowym $\mathcal{B}(n, \vartheta)$. W problemie testowania hipotezy $H_0 : \vartheta \leqslant \vartheta_0$ przeciwko $H_1 : \vartheta > \vartheta_0$ niezrandomizowane testy mają postać $\varphi_B(x) = \mathbf{1}_B(x)$ z obszarem odrzucenia $B \subset \{0, 1, \ldots, n\} = \mathcal{X}$. Rozmiary tych testów

$$\beta_\varphi := \sup_{\vartheta \leqslant \vartheta_0} P_\vartheta(\varphi_B(X) = 1) = \sup_{\vartheta \leqslant \vartheta_0} P_\vartheta(X \in B)$$

mogą przyjmować tylko określoną liczbę wartości (zbiór $\mathcal{X}$ wartości zmiennej losowej X jest skończoną przestrzenią dyskretną i mamy skończoną liczbę podzbiorów B tej przestrzeni). W celu uzyskania testu o żądanej dowolnej ustalonej wartości rozmiaru α należy dokonać randomizacji. Na przykład, test zrandomizowany postaci

$$\varphi_{k,\gamma}(x) = \begin{cases} 1, & \text{jeżeli } x > k, \\ \gamma, & \text{jeżeli } x = k, \\ 0, & \text{jeżeli } x < k, \end{cases}$$

gdzie stałe k i γ wyznacza się z warunku

$$P_{\vartheta_0}(X > k) + \gamma P_{\vartheta_0}(X = k) = \alpha,$$

jest rozmiaru α. Na podstawie znanego z teorii testowania hipotez *twierdzenia Karlina-Rubina*, test tej postaci jest jednostajnie najmocniejszy na poziomie istotności α w testowaniu rozpatrywanych hipotez. Niech

$$\overline{\delta}(x, a_1) = \varphi(x) \quad \text{i} \quad \overline{\delta}(x, a_0) = 1 - \varphi(x),$$

gdzie a_i $(i = 0, 1)$ oznacza przyjęcie hipotezy H_i. Jeżeli zaobserwujemy $x > k$, to odrzucamy H_0 z prawdopodobieństwem 1; jeżeli zaobserwujemy $x = k$, to odrzucamy H_0 z prawdopodobieństwem γ i przyjmujemy H_0 z prawdopodobieństwem $1 - \gamma$; jeżeli zaobserwujemy $x < k$, to odrzucamy H_0 z prawdopodobieństwem zero. Opisany test zrandomizowany określa zatem behawiorystyczną funkcję decyzyjną w grze statystycznej z wyjściową dwuelementową przestrzenią akcji $\mathcal{A} = \{a_0, a_1\}$.

Przyjmując funkcję straty postaci

$$\mathcal{L}(\vartheta, a_0) = \begin{cases} 0, & \text{gdy } \vartheta \leqslant \vartheta_0, \\ 1, & \text{gdy } \vartheta > \vartheta_0, \end{cases} \qquad \mathcal{L}(\vartheta, a_1) = \begin{cases} 1, & \text{gdy } \vartheta \leqslant \vartheta_0, \\ 0, & \text{gdy } \vartheta > \vartheta_0, \end{cases}$$

funkcja straty akcji zrandomizowanej równa się

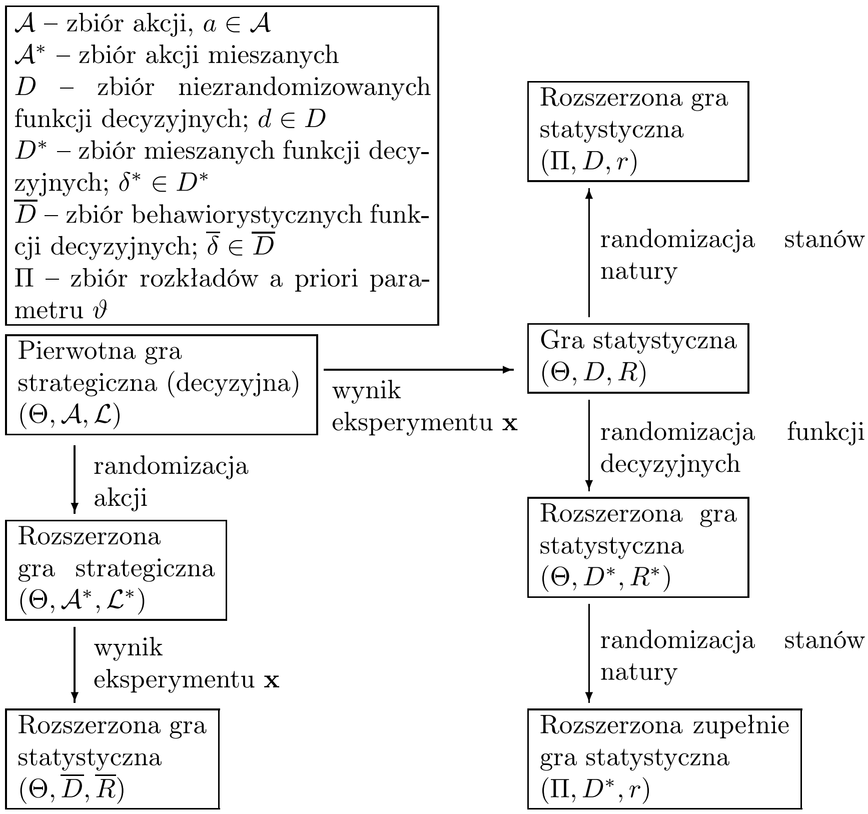
$$\begin{aligned} \mathcal{L}^*(\vartheta, \overline{\delta}(x, \cdot)) &= \int_{\mathcal{A}} \mathcal{L}(\vartheta, a) d\overline{\delta}(x, a) = \mathcal{L}(\vartheta, a_0)\overline{\delta}(x, a_0) + \mathcal{L}(\vartheta, a_1)\overline{\delta}(x, a_1) \\ &= \begin{cases} \varphi(x), & \text{gdy } \vartheta \leqslant \vartheta_0, \\ 1 - \varphi(x), & \text{gdy } \vartheta > \vartheta_0. \end{cases} \end{aligned}$$

Funkcja ryzyka behawiorystycznej funkcji decyzyjnej $\overline{\delta}$ jest postaci

$$\begin{aligned} \overline{R}(\vartheta, \overline{\delta}) &= E_\vartheta[\mathcal{L}^*(\vartheta, \overline{\delta}(X, \cdot))] = \begin{cases} E_\vartheta[\varphi(X)], & \text{gdy } \vartheta \leqslant \vartheta_0, \\ 1 - E_\vartheta[\varphi(X)], & \text{gdy } \vartheta > \vartheta_0, \end{cases} \\ &= \begin{cases} \beta_\varphi(\vartheta) = P_\vartheta(X > k) + \gamma P_\vartheta(X = k), & \text{gdy } \vartheta \leqslant \vartheta_0, \\ 1 - \beta_\varphi(\vartheta) = P_\vartheta(X < k) + (1 - \gamma)P_\vartheta(X = k), & \text{gdy } \vartheta > \vartheta_0, \end{cases} \end{aligned}$$

gdzie $\beta_\varphi(\vartheta) := E_\vartheta[\varphi(X)]$ jest *funkcją mocy testu zrandomizowanego* φ. Na przykład, dla $\vartheta = \vartheta_0$, $\overline{R}(\vartheta_0, \overline{\delta}) = \alpha$ (prawdopodobieństwo błędu I rodzaju), a dla $\vartheta = \vartheta_1 > \vartheta_0$, $\overline{R}(\vartheta_1, \overline{\delta}) = 1 - \beta_\varphi(\vartheta_1)$ (prawdopodobieństwo błędu II rodzaju). □

Tab. 1.1: Schemat graficzny rozszerzeń gry statystycznej



Przestrzeń $\overline{D}$ jest przestrzenią zrandomizowanych funkcji decyzyjnych $\overline{\delta}$, takich że dla każdego $x \in \mathcal{X}$, $\overline{\delta}(x, \cdot)$ jest rozkładem prawdopodobieństwa na $(\mathcal{A}, \mathcal{G})$ i ryzyko funkcji decyzyjnej $\overline{\delta}$ określone jest przez (1.3). Z poniższego lematu wynika, że nie jest konieczna inna randomizacja.

LEMAT **1.1** Niech Q będzie dowolną miarą probabilistyczną na przestrzeni akcji $(\mathcal{A}, \mathcal{G})$, a $\overline{\delta}(x, A)$, $x \in \mathcal{X}, A \in \mathcal{G}$, niech oznacza behawiorystyczną funkcję decyzyjną. Zdefiniujmy, dla każdego $x \in \mathcal{X}$, miarę probabilistyczną

$$\overline{\delta}_Q(x, A) = E^Q[\overline{\delta}(x, A)], A \in \mathcal{G}, \tag{1.4}$$

na $(\mathcal{A}, \mathcal{G})$. Wówczas miara $\overline{\delta}_Q$ jest równoważna Q, w tym sensie, że

$$\overline{R}(\vartheta, \overline{\delta}_Q) = E^Q[\overline{R}(\vartheta, \overline{\delta})].$$

Dowód. Zgodnie ze wzorem (1.3),

$$\overline{R}(\vartheta, \overline{\delta}_Q) = E_\vartheta\left[\mathcal{L}^*(\vartheta, \overline{\delta}_Q(X, \cdot))\right] = E_\vartheta\big\{E^{\overline{\delta}_Q(X,\cdot)}[\mathcal{L}(\vartheta, a)]\big\}.$$

Ponieważ, z definicji (1.4),

$$E^{\overline{\delta}_Q(x,\cdot)}[\mathcal{L}(\vartheta, a)] = E^Q E^{\overline{\delta}(x,\cdot)}[\mathcal{L}(\vartheta, a)],$$

więc

$$\overline{R}(\vartheta, \overline{\delta}_Q) = E_\vartheta E^Q \big\{ E^{\overline{\delta}(X,\cdot)}[\mathcal{L}(\vartheta, a)] \big\} = E^Q E_\vartheta \big\{ E^{\overline{\delta}(X,\cdot)}[\mathcal{L}(\vartheta, a)] \big\} = E^Q[\overline{R}(\vartheta, \overline{\delta})].$$

Zmiana kolejności całkowania w powyższych całkach uzasadniona jest przez twierdzenie Fubiniego, ponieważ z założenia $\mathcal{L}(\vartheta, a) \geqslant -c > -\infty$. ◻

Mieszana funkcja decyzyjna wskazuje przed dokonaniem eksperymentu rozkład prawdopodobieństwa, według którego należy wybrać funkcję przyporządkowującą wynikowi eksperymentu akcję, natomiast behawiorystyczna funkcja decyzyjna wskazuje po zaobserwowaniu wyniku eksperymentu, według jakiego rozkładu prawdopodobieństwa należy wybrać akcję. Gry (Θ, D^*, R^*) i $(\Theta, \overline{D}, \overline{R})$ różnią się zatem momentem dokonania losowego wyboru (randomizacji). Gry te jednak mogą być równoważne w następującym sensie: jeśli dla każdej mieszanej funkcji decyzyjnej $\delta^* \in D^*$ istnieje behawiorystyczna funkcja decyzyjna $\overline{\delta} \in \overline{D}$ taka, że

$$R^*(\vartheta, \delta^*) = \overline{R}(\vartheta, \overline{\delta}), \quad \text{dla każdego } \vartheta \in \Theta, \tag{1.5}$$

oraz, na odwrót, dla każdej behawiorystycznej funkcji decyzyjnej $\overline{\delta} \in \overline{D}$ istnieje mieszana funkcja decyzyjna $\delta^* \in D^*$ taka, że zachodzi (1.5), to gry (Θ, D^*, R^*) i $(\Theta, \overline{D}, \overline{R})$ nazywamy *równoważnymi* [**equivalent games**].

Następujące twierdzenie podaje warunek wystarczający równoważności gier $(\Theta,\ D^*,\ R^*)$ i $(\Theta, \overline{D}, \overline{R})$.

TWIERDZENIE **1.1** (*Walda-Wolfowitza*) [**Wald-Wolfowitz theorem**] Niech zbiór $\mathcal{A}$ będzie podzbiorem ośrodkowej przestrzeni metrycznej. Wówczas gry statystyczne (Θ, D^*, R^*) i $(\Theta, \overline{D}, \overline{R})$ są równoważne. ◻

PRZYKŁAD **1.6** Niech przestrzeń wartości zmiennej losowej X będzie skończona: $\mathcal{X} = \{x_1, x_2, \ldots, x_l\}$. Niech będą też skończone przestrzeń funkcji decyzyjnych $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$ i przestrzeń akcji $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$. Pokażemy, że dla każdej mieszanej funkcji decyzyjnej istnieje równoważna jej behawiorystyczna funkcja decyzyjna.

Niech $\delta^* = \{\delta^*(d_1), \delta^*(d_2), \ldots, \delta^*(d_n)\}$ oznacza funkcję decyzyjną mieszaną, tzn. miarę probabilistyczną na $(\mathcal{D}, \mathcal{F})$, gdzie $\delta^*(d_i)$ oznacza prawdopodobieństwo wyboru funkcji decyzyjnej d_i.

Definiujemy behawiorystyczną funkcję decyzyjną, tzn. funkcję $\overline{\delta}$ na $\mathcal{X} \times \mathcal{G}$ w następujący sposób:

$$\begin{aligned} \overline{\delta}(x,\cdot) &= \{\overline{\delta}(x_1,\cdot), \overline{\delta}(x_2,\cdot) \ldots, \overline{\delta}(x_l,\cdot)\} \\ &= \{\{\overline{\delta}(x_1,a_1), \ldots, \overline{\delta}(x_1,a_k)\}, \{\overline{\delta}(x_2,a_1), \ldots, \overline{\delta}(x_2,a_k)\}, \\ &\quad \ldots, \{\overline{\delta}(x_l,a_1), \ldots, \overline{\delta}(x_l,a_k)\}\}, \end{aligned}$$

gdzie

$$\overline{\delta}(x_r, a_s) = \sum_{i=1}^{n} \delta^*(d_i)\mathbf{1}\{d_i(x_r) = a_s\}, \quad r = 1, \ldots, l;\ s = 1, \ldots, k. \tag{1.6}$$

Pokażemy, że behawiorystyczna funkcja decyzyjna określona wzorem (1.6) jest równoważna mieszanej funkcji decyzyjnej w tym sensie, że jej funkcja ryzyka przyjmuje tę samą wartość co funkcja ryzyka mieszanej funkcji decyzyjnej.

Funkcja ryzyka mieszanej funkcji decyzyjnej wyraża się wzorem

$$\begin{aligned} R^*(\vartheta, \delta^*) &= \sum_{i=1}^{n} R(\vartheta, d_i)\delta^*(d_i) = \sum_{i=1}^{n} E_\vartheta[\mathcal{L}(\vartheta, d_i(X))]\delta^*(d_i) \\ &= \sum_{i=1}^{n} \left[\sum_{r=1}^{l} \mathcal{L}(\vartheta, d_i(x_r))P_\vartheta(X = x_r)\right] \delta^*(d_i) \\ &= \sum_{r=1}^{l} P_\vartheta(X = x_r) \sum_{i=1}^{n} \mathcal{L}(\vartheta, d_i(x_r))\delta^*(d_i). \end{aligned}$$

Funkcja ryzyka behawiorystycznej funkcji decyzyjnej ma następującą postać

$$\overline{R}(\vartheta, \overline{\delta}) = \sum_{r=1}^{l} \mathcal{L}^*(\vartheta, \overline{\delta}(x_r, \cdot))P_\vartheta(X = x_r).$$

Dla $\overline{\delta}(x_r, a_s)$ określonych wzorem (1.6) mamy

$$\begin{aligned} \mathcal{L}^*(\vartheta, \overline{\delta}(x_r, \cdot)) &= \sum_{s=1}^{k} \mathcal{L}(\vartheta, a_s)\overline{\delta}(x_r, a_s) = \sum_{s=1}^{k} \mathcal{L}(\vartheta, a_s) \sum_{i=1}^{n} \delta^*(d_i)\mathbf{1}\{d_i(x_r) = a_s\} \\ &= \sum_{i=1}^{n} \delta^*(d_i) \sum_{s=1}^{k} \mathcal{L}(\vartheta, a_s)\mathbf{1}\{d_i(x_r) = a_s\} = \sum_{i=1}^{n} \delta^*(d_i)\mathcal{L}(\vartheta, d_i(x_r)). \end{aligned}$$

Zatem

$$\overline{R}(\vartheta, \overline{\delta}) = \sum_{r=1}^{l} P_\vartheta(X = x_r) \sum_{i=1}^{n} \mathcal{L}(\vartheta, d_i(x_r))\delta^*(d_i) = R^*(\vartheta, \delta^*).$$

□

Mieszane oraz behawiorystyczne funkcje decyzyjne nazywamy ogólnie *zrandomizowanymi funkcjami decyzyjnymi* [**randomized decision functions**].

Ze względu na równoważność gier statystycznych (Θ, D^*, R^*) i $(\Theta, \overline{D}, \overline{R})$, spełniających w większości rozpatrywanych przypadków założenie Twierdzenia 1.1, zrandomizowane funkcje decyzyjne δ^* i $\overline{\delta}$ można oznaczać jednym symbolem δ, a ich zbiór przez $\mathcal{D}$.

1.3 Gra statystyczna jako podejście do wnioskowań statystycznych

Podstawowe rodzaje wnioskowań statystyki matematycznej mogą być przedstawione jako szczególne przypadki statystycznego problemu decyzyjnego. Opiszemy strukturę gry statystycznej w podstawowych wnioskowaniach statystycznych.

A. Estymacja punktowa.

W problemie *estymacji punktowej* zbiór akcji statystyka powinien zawierać zbiór możliwych wartości parametru ϑ, tzn. $\Theta \subset \mathcal{A}$. Funkcja straty uwzględnia fakt, że dla akcji a statystyka odległej od wartości ϑ poniesiona strata jest duża, natomiast dla wartości a bliskich ϑ strata jest mała i decyzja jest rozsądna. Funkcja decyzyjna $d : \mathcal{X} \mapsto \mathcal{A}$ pełni rolę estymatora parametru ϑ.

W estymacji punktowej stosuje się różne postaci funkcji straty w zależności od zastosowań rozważanego modelu statystycznego (patrz Podrozdział 1.4.1). Na przykład, dla kwadratowej funkcji straty $\mathcal{L}(\vartheta, a) = (a - \vartheta)^2$ funkcja ryzyka jest tzw. *błędem średniokwadratowym*:

$$R(\vartheta, d) = E_\vartheta[d(X) - \vartheta]^2 = \mathrm{Var}_\vartheta(d(X)) + b_\vartheta^2(d(X)),$$

gdzie $b_\vartheta(d(X)) = E_\vartheta(d(X)) - \vartheta$ jest obciążeniem estymatora $d(X)$. Dobry estymator powinien mieć małą wariancję i małe obciążenie.

B. Estymacja przedziałowa.

Przestrzeń akcji $\mathcal{A}$ w problemie *estymacji przedziałowej* składa się z podzbiorów $C(X)$ zbioru Θ, zależnych jedynie od obserwowalnej zmiennej losowej X. Oznaczenie C wiąże się ze słowem *confidence* (ang.). Zbiór ufności $C(X)$ (w szczególności przedział) identyfikujemy z akcją d oznaczającą wybór oceny przedziałowej "$\theta \in C$". Zatem funkcja decyzyjna $d : \mathcal{X} \mapsto \mathcal{A}$ określa, dla każdego $x \in \mathcal{X}$, który zbiór $C \in \mathcal{A}$ będzie wykorzystany jako ocena przedziałowa parametru ϑ po zaobserwowaniu $X = x$.

Funkcja straty w problemie *estymacji przedziałowej* zazwyczaj zawiera dwa składniki: miarę, która mierzy poprawność pokrycia przez ocenę przedziałową prawdziwej wartości parametru ϑ oraz miarę wielkości tej oceny przedziałowej. Najczęściej przyjmowaną miarą poprawności pokrycia prawdziwej wartości parametru ϑ przez zbiór $C(X)$ jest funkcja postaci

$$\mathbf{1}_{C(X)}(\vartheta) = \begin{cases} 1, & \text{gdy } \vartheta \in C(X), \\ 0, & \text{gdy } \vartheta \notin C(X). \end{cases}$$

Funkcja straty w tym problemie powinna przyjmować wartości małe dla dobrej decyzji $C(x)$, tzn. pokrywającej nieznany parametr ϑ, i małej miary (w przypadku przedziałów – małej długości) zbioru $C(x)$. Jedną z takich funkcji jest funkcja straty postaci

$$\mathcal{L}(\vartheta, C(x)) = b\mu(C(x)) - \mathbf{1}_{C(x)}(\vartheta), \tag{1.7}$$

gdzie b jest stałą dodatnią, a μ jest miarą zbioru $C(x)$. Jeżeli większą wagę przywiązujemy do poprawności pokrycia, to b powinno być małe, jeśli natomiast większą wagę przypisujemy do długości przedziału, to b powinno być duże.

Funkcja ryzyka związana z funkcją straty (1.7) jest postaci

$$\begin{aligned} R(\vartheta, C) &= bE_\vartheta[\mu(C(X))] - E_\vartheta[\mathbf{1}_{C(X)}(\vartheta)] \\ &= bE_\vartheta[\mu(C(X))] - P_\vartheta\left(\mathbf{1}_{C(X)}(\vartheta) = 1\right) \\ &= bE_\vartheta[\mu(C(X))] - P_\vartheta\left(\vartheta \in C(X)\right). \end{aligned}$$

Ryzyko ma dwie składowe: wartość oczekiwaną miary zbioru $C(X)$ (w przypadku przedziałów – wartość oczekiwaną długości przedziału) i prawdopodobieństwo pokrycia przez estymator przedziałowy. W podejściu klasycznym ustala się wartość poziomu ufności $1-\alpha$, tzn. infimum prawdopodobieństwa pokrycia, i dąży się do minimalizacji długości przedziału. W podejściu teoriodecyzyjnym, funkcja ryzyka uwzględnia fakt, że równocześnie zależy nam na tym by wartość oczekiwana długości była mała, a prawdopodobieństwo pokrycia było duże.

C. Testowanie hipotez.

Niech X będzie zmienną losową, której rozkład P_ϑ zależy od nieznanego parametru $\vartheta \in \Theta$. Rozpatrzmy problem testowania hipotezy $H_0 : \vartheta \in \Theta_0$ przeciwko hipotezie alternatywnej $H_1 : \vartheta \in \Theta_0^C$. Przestrzeń akcji $\mathcal{A}$ w tym problemie jest dwuelementowa: $\mathcal{A} = \{a_0, a_1\}$. Akcja a_0 oznacza przyjęcie weryfikowanej hipotezy H_0 (i równocześnie odrzucenie hipotezy H_1), natomiast akcja a_1 oznacza przyjęcie hipotezy H_1 (i równocześnie odrzucenie hipotezy H_0).

Akcję a_0 lub a_1 statystyk może podjąć w oparciu o zaobserwowaną wartość x zmiennej losowej X. Przestrzeń $\mathcal{X}$ wartości zmiennej X można podzielić na dwa rozłączne zbiory B oraz $B^C = \mathcal{X} \backslash B$ takie, że jeżeli $x \in B$, to podejmuje się akcję a_1 o odrzuceniu weryfikowanej hipotezy H_0. Zatem niezrandomizowana funkcja decyzyjna $d : \mathcal{X} \mapsto \mathcal{A}$ ma w problemie weryfikacji hipotez postać następującą:

$$d(x) = \begin{cases} a_1, & \text{jeżeli } x \in B, \\ a_0, & \text{jeżeli } x \in B^C. \end{cases} \tag{1.8}$$

Funkcja decyzyjna $d(x)$ określona wzorem (1.8) jest więc testem niezrandomizowanym z obszarem odrzucenia B. Funkcja mocy tego testu

$$\beta(\vartheta) = P_\vartheta(X \in B) = P_\vartheta(d(X) = a_1).$$

W teorii *testowania hipotez* przyjmuje się najczęściej następującą funkcję straty:

$$\mathcal{L}(\vartheta, a_0) = \begin{cases} 0, & \text{gdy } \vartheta \in \Theta_0, \\ c_0, & \text{gdy } \vartheta \in \Theta_0^C, \end{cases} \qquad \mathcal{L}(\vartheta, a_1) = \begin{cases} c_1, & \text{gdy } \vartheta \in \Theta_0, \\ 0, & \text{gdy } \vartheta \in \Theta_0^C, \end{cases} \tag{1.9}$$

gdzie $c_0, c_1 > 0$. Taka postać funkcji straty oznacza, że strata jest zerowa, gdy akcje a_0, i a_1 są prawidłowe, natomiast statystyk ponosi stratę w wysokości c_1, gdy popełnia *błąd I rodzaju* (odrzuca hipotezę H_0, gdy jest ona prawdziwa) oraz stratę w wysokości c_0, gdy popełnia *błąd II rodzaju* (przyjmuje hipotezę H_0, gdy jest ona fałszywa). Stałe c_0 i c_1 w funkcji straty $\mathcal{L}(\vartheta, a)$ można dobierać w zależności od konsekwencji i znaczenia błędów I i II rodzaju w konkretnych problemach.

Funkcja ryzyka dla tej funkcji straty jest postaci

$$R(\vartheta, d)$$
$$= \begin{cases} 0 \cdot P_\vartheta(d(X) = a_0) + c_1 P_\vartheta(d(X) = a_1) = c_1\beta(\vartheta), & \text{gdy } \vartheta \in \Theta_0, \\ c_0 P_\vartheta(d(X) = a_0) + 0 \cdot P_\vartheta(d(X) = a_1) = c_0[1 - \beta(\vartheta)], & \text{gdy } \vartheta \in \Theta_0^C. \end{cases}$$

Z powyższego wzoru wynika, że w teoriodecyzyjnym ujęciu problemu testowania hipotez funkcja ryzyka testu jest ściśle związana z funkcją mocy oraz podejście teoriodecyzyjne umożliwia uwzględnienie wag c_0 i c_1 przypisanym prawdopodobieństwom błędów I i II rodzaju.

Testy zrandomizowane są *behawiorystycznymi funkcjami decyzyjnymi* (Przykład 1.5).

D. Estymacja dystrybuanty.

Niech $\mathbf{X} = (X_1, \ldots, X_n)$ będzie próbą z rozkładu o nieznanej postaci dystrybuanty $F(t)$ pewnego rozkładu na $\mathbb{R}$. Problem polega na wyznaczeniu prawostronnie ciągłej funkcji $\widehat{F(t)} = d(t; \mathbf{X})$, $t \in \mathbb{R}$, jako oszacowania funkcji $F(t)$ w oparciu o obserwację $\mathbf{X}$. Nawet, jeśli wiadomo, że dystrybuanta $F(t)$ jest funkcją ciągłą, jej oszacowanie $\widehat{F(t)}$ nie musi być funkcją ciągłą. Przestrzeń Θ jest zbiorem dystrybuant na $\mathbb{R}$, o których ewentualnie wiadomo, że są typu ciągłego lub dyskretnego. Przestrzeń akcji jest zbiorem wszystkich dystrybuant na $\mathbb{R}$. Miarą rozbieżności między dystrybuantą F a jej oszacowaniem $\widehat{F}$ jest funkcja straty $\mathcal{L}(F(t), d(t, \mathbf{X}))$. W Podrozdziale 1.4.2 podano najczęściej używane postaci tej funkcji.

E. Problem klasyfikacji.

Niech $\Theta = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\}$ i $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$. Niech zmienna losowa X ma rozkład P_{ϑ_i}, jeżeli prawdziwym stanem natury jest $\vartheta_i, i = 1, 2, \ldots, k$. Podjęcie akcji a_j oznacza zaklasyfikowanie obserwacji $X = x$ jako otrzymanej z rozkładu P_{ϑ_j}. Jako funkcję straty można przyjąć funkcję postaci

$$\mathcal{L}(\vartheta_i, a_j) = \begin{cases} 1, & \text{gdy } i \neq j, \\ 0, & \text{gdy } i = j, \end{cases}$$

tzn. strata związana z decyzją niepoprawną równa jest 1, a z poprawną wynosi 0. W tym problemie rozsądne jest rozpatrywanie zrandomizowanych (behawiorystycznych) funkcji decyzyjnych. Rozpatrzymy miarę probabilistyczną, $\overline{\delta}(x, \cdot)$ na $(\mathcal{A}, \mathcal{G})$, która dla każdej zaobserwowanej wartości x jest miarą dyskretną skoncentrowaną w punktach a_j, $j = 1, \ldots, k$, z odpowiadającymi im masami $\overline{\delta}(x, a_j)$, tzn. $\overline{\delta}(x, A) = \int_A d\overline{\delta}(x, a) = \sum_{a_j \in A} \overline{\delta}(x, a_j)$, $A \subset \mathcal{A}$. Zgodnie ze wzorem (1.3), funkcja ryzyka funkcji decyzyjnej $\overline{\delta}(x,\cdot)$ ma następującą postać:

$$\begin{aligned}\overline{R}(\vartheta_i,\overline{\delta}) &= E_{\vartheta_i}[\mathcal{L}^*(\vartheta_i,\overline{\delta}(X,\cdot))] = E_{\vartheta_i}\Big[\sum_{j=1}^{k}\mathcal{L}(\vartheta_i,a_j)\overline{\delta}(X,a_j)\Big]\\ &= \sum_{j=1}^{k}\mathcal{L}(\vartheta_i,a_j)E_{\vartheta_i}[\overline{\delta}(X,a_j)] = \sum_{j\neq i}^{k}E_{\vartheta_i}[\overline{\delta}(X,a_j)] = 1 - E_{\vartheta_i}[\overline{\delta}(X,a_i)].\end{aligned}$$

F. Problem sekwencyjnego podejmowania decyzji.

Niech $\mathcal{A} = \{a_0, a_1,\ a_2,\ \ldots,\ a_k\}$, $k \geqslant 1$. Przykładem gry z takim zbiorem akcji może być wnioskowanie statystyczne (testowanie hipotez, estymacja czy klasyfikacja), w którym statystyk w każdym kroku może podjąć jedną z akcji ze zbioru $\mathcal{A}$, przy czym jedną z akcji w tym zbiorze (np. a_0) jest akcja oznaczająca pobranie dodatkowej (kolejnej) obserwacji. Po otrzymaniu nowych danych statystyk ponownie dokonuje wyboru akcji. Procedury tego typu rozpatrywane są w Rozdziale 6.

1.4 Funkcje straty w problemach estymacji

Funkcja straty powinna odzwierciedlać konsekwencje różnych błędów; jest jednak w dużym stopniu zdeterminowana względami wygody i możliwością uzyskania optymalnych, w pewnym sensie, decyzji statystycznych. Zależy również od rodzaju problemu statystycznego.

1.4.1 Funkcje straty w problemach estymacji parametrycznej

W problemach *parametrycznej estymacji punktowej* wiele ogólnych wyników teoretycznych nie wymaga szczegółowych założeń o funkcji straty i pozostaje w mocy dla dużej klasy takich funkcji, w szczególności dla funkcji straty $\mathcal{L}(\vartheta, a)$ wypukłych względem argumentu a dla każdego $\vartheta \in \Theta$. Wypukłe funkcje straty prowadzą do licznych uproszczeń problemu estymacji. Należy jednak rozważnie ocenić realistyczność takich funkcji. Jeżeli $\mathcal{L}(\vartheta, a)$ nie jest miarą niedokładności oszacowania, ale pewną realną (np. finansową) stratą, to można przypuszczać, że tego rodzaju straty są ograniczone: jeżeli ktoś już stracił wszystko, nie może stracić więcej. Z drugiej strony, jeżeli a może przyjmować wszystkie wartości z przedziału $(-\infty,\infty)$ lub z przedziału $(0,\infty)$, to żadna funkcja ograniczona, różna od stałej, nie może być wypukła. Ponadto, ograniczone funkcje straty z nieograniczonym argumentem a prowadzą do bezsensownych estymatorów. Można wtedy bowiem popełniać dowolnie duże błędy bez dodatkowych kosztów. Wypukłe funkcje straty prowadzą do bardziej sensownych estymatorów, gdyż wysokie straty, jakie przypisują dużym błędom, rekompensują nierealistyczne założenie o nieograniczoności a.