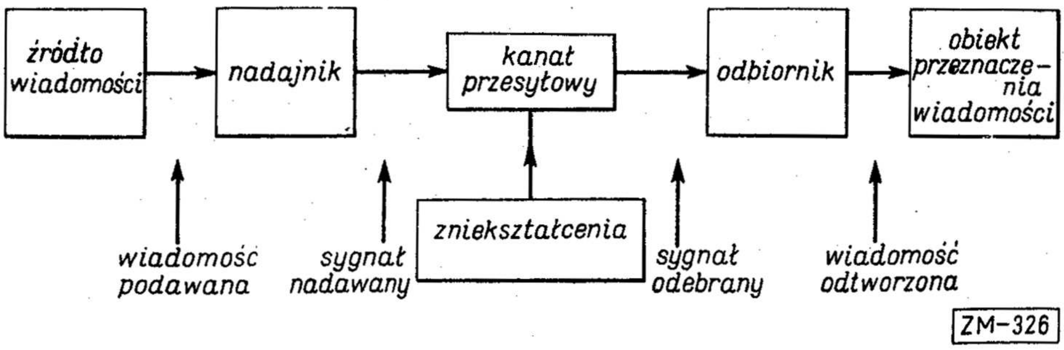
ZASTOSOWANIA MATEMATYKI
V (1960)

J. SEIDLER (Gdańsk)

ZWIĄZKI MIĘDZY TEORIĄ INFORMACJI A TEORIĄ FUNKCJI DECYZYJNYCH W ZASTOSOWANIU DO ZAGADNIEŃ TELEKOMUNIKACJI

I. Wstęp i sformułowanie zagadnienia

Zadaniem systemów telekomunikacyjnych jest przekazywanie na odległość wiadomości. Tymi wiadomościami mogą być np. teksty zbudowane z liter, dźwięki, obrazy lub przebiegi napięciowe uzyskane z przyrządów pomiarowych. Wiadomościom podawanym przez źródło wiadomości jest przyporządkowany w nadajniku sygnał, z zasady elektryczny, nadający się do przekazania na odległość. Sygnały po przebyciu kanału przesyłowego łączącego nadajnik z odbiornikiem są przekształcane i ostatecznie odtworzona wiadomość dochodzi do obiektu przeznaczenia wiado-



Rys. 1

mości (rys. 1). Jednakże w trakcie przesyłania sygnały ulegają zniekształceniom i wiadomość odtworzona na ogół różni się od wiadomości podanej przez źródło wiadomości.

Zniekształcenia są ostatecznym czynnikiem ograniczającym zasięg, w którym można przekazywać wiadomości. Z punktu widzenia urządzenia odbiorczego, można podzielić zniekształcenia na dwa zasadnicze rodzaje. Zniekształcenia *pierwszego rodzaju*, spowodowane przez czynniki w pełni znane, oraz zniekształcenia *drugiego rodzaju*, spowodowane przez czynniki przypadkowe. Przykładem zniekształceń pierwszego rodzaju jest znane osłabienie sygnału w trakcie przesyłania lub nałożenie się innego znanego sygnału, w szczególności tak zwany *przydźwięk sieciowy*. Przykładem zniekształceń drugiego rodzaju są zaniki, tj. wahania amplitudy sygnału spowodowane zmianami stanu jonosfery, oraz nałożenie się na sygnał szumu. Modele matematyczne niektórych typów zniekształceń przypadkowych można znaleźć w rozdziale V pracy [1] oraz w pracy [11].

Zniekształcenia pierwszego rodzaju dają się zawsze skorygować z wyjątkiem nierealnych przypadków. W podanych przykładach można to uczynić stosując wzmacniacz lub układ kompensacyjny. Dlatego też w dalszych rozważaniach nie będziemy brać pod uwagę zniekształceń tego typu.

Zniekształcenia drugiego rodzaju są w zasadzie nieuniknione. Zawsze bowiem obecne są efekty fluktuacji termicznych. W wielu zagadnieniach istotną rolę odgrywają znacznie silniejsze zakłócenia atmosferyczne, przemysłowe lub też pochodzące od innych systemów komunikacyjnych.

Patrząc od strony urządzenia odbiorczego można traktować jako wielkości przypadkowe nie tylko zakłócenia, ale również podawane wiadomości. W wielu przypadkach ich rozkłady prawdopodobieństwa można ustalić przez badanie własności źródła wiadomości bądź też określić eksperymentalnie (patrz np. [5]).

Tak więc odtwarzanie nadawanej wiadomości na podstawie odebranego sygnału jest zagadnieniem statystycznym, toteż metody statystyki matematycznej stosowane są szeroko w analizie problemów telekomunikacji (patrz np. [6], [12], [16]). Ponieważ najbardziej ogólne potraktowanie tych metod daje teoria funkcji decyzyjnych (patrz [20]), przeto i zagadnienia telekomunikacji można ująć w jednolity sposób za pomocą tej teorii.

Badaniem systemów telekomunikacyjnych zajmuje się również teoria podana przez C. E. Shannona [15], a później rozwinięta przez innych (patrz np. [2], [3]), zwana ogólnie *teorią informacji*, która nie wiąże się bezpośrednio ze schematem funkcji decyzyjnych.

Tematem niniejszej pracy jest ustalenie związków między zasadniczymi wielkościami występującymi w obu tych teoriach.

Celowość takich badań jest tym bardziej uzasadniona, że podczas gdy istnieje wyraźne powiązanie między praktyką a teorią funkcji decyzyjnych, to wyłaniają się istotne trudności w stosowaniu teorii informacji w konkretnych zagadnieniach telekomunikacji. Omówimy je pokrótce.

Podstawowe pojęcia tej teorii, a mianowicie entropia i miara ilości informacji, wiążą się organicznie z własnościami długich ciągów sygnałów i wynikającym z nich pojęciem optymalnego kodowania (patrz np.

[2], [3]). Tymczasem w praktyce większość systemów telekomunikacyjnych nie jest optymalnie zakodowana i nawet nie przewiduje się, że będą one kodowane w przyszłości. Dotyczy to w szczególności tak zwanych ciągłych (w sensie makroskopowym) systemów komunikacyjnych oraz systemów pomiarowych, np. mierzących odległość na zasadzie echa. Ujęcie Shannona nie wyjaśnia, jaki sens ma entropia i miara ilości informacji, gdy nie rozpatruje się długich ciągów sygnałów i gdy wprowadzenie pojęcia kodowania nie jest uzasadnione.

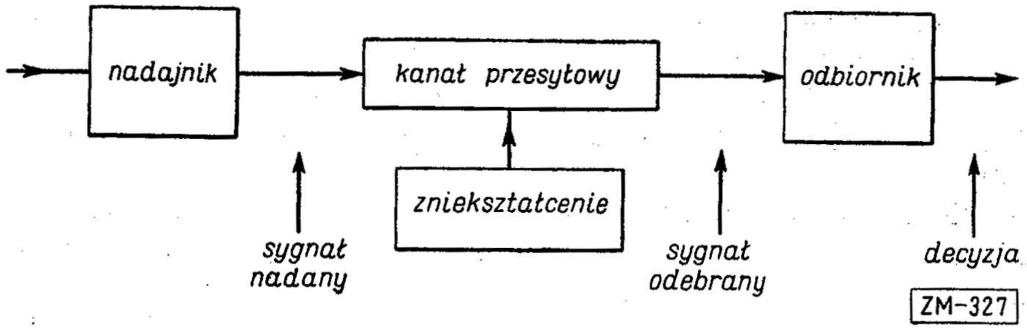
Omawianej trudności teorii Shannona usiłowało uniknąć wielu autorów (patrz np. [10], [14], [22]). Wszystkie te prace opierają się na *aksjomacie addytywności*, który postuluje, że jeżeli przesyła się dwa niezależne, elementarne sygnały, to ilość informacji, przesyłana przez sygnał złożony, będący parą sygnałów elementarnych, jest równa sumie ilości informacji przekazywanej przez każdy z elementarnych sygnałów z osobna. Przyjęcie tego aksjomatu jest uzasadnione, gdy sygnały nadawane mają strukturę ciągów zbudowanych z elementarnych jednostek i gdy interesuje nas wpływ zwiększenia ilości elementarnych sygnałów. Jednakże, taka struktura sygnałów nie jest bynajmniej konieczna do przesyłania wiadomości. Co więcej, z bliższej analizy cytowanych prac widać, że zarówno w rozważaniach ilościowych jak i przy wyprowadzaniu, nielicznych zresztą, wzorów, które mogą być użyteczne do badania systemów niekodowanych optymalnie, nie wykorzystuje się bezpośrednio wprowadzonego aksjomatu addytywności. Tak więc, mimo licznych prac, zagadnienie ogólnej interpretacji oraz zastosowań entropii i miary ilości informacji, zwłaszcza w wypadku systemów ciągłych, nie jest w pełni wyjaśnione. Wydaje się, że ustalenie związków między tymi wielkościami, a wielkościami występującymi w dobrze ugruntowanej teorii funkcji decyzyjnych, może być użyteczne.

II. Zadania urządzeń odbiorczych w świetle teorii funkcji decyzyjnych

Będziemy zakładali, że struktura nadajnika jest w pełni znana i że wiadomość jest związana z sygnałem nadanym w sposób jedno-jednoznaczny. Dlatego w dalszym ciągu będziemy się jedynie zajmowali sygnałem nadanym.

Przypadkowe zniekształcenia powodują, że znając sygnał odebrany nie można na ogół ustalić dokładnie, jaki sygnał został nadany. Można natomiast na podstawie odebranego sygnału powziąć decyzję o sygnale nadanym. Jeżeli sygnał nadawany jest określony przez jednowymiarowy parametr (np. wysokość impulsu), decyzja może być przybliżoną wartością parametru (estymatorem); również może ona być przedziałem, wewnątrz którego przypuszczalnie leży wartość parametru.

Zadaniem odbiornika jest podawanie decyzji; jest on jednoznacznie określony przez regułę przyporządkowującą wyjścia, tj. *decyzje*, wejściom, tj. *sygnałom odebranym*. Regułę tę nazywa się *funkcją decyzyjną*. Schemat systemu komunikacyjnego, którym będziemy się posługiwali w dalszym ciągu, jest pokazany na rysunku 2.



Rys. 2

Wskutek przypadkowych zakłóceń decyzje mogą być fałszywe lub niedokładne. Będziemy więc zakładali, że dana jest nieujemna funkcja, która mierzy straty, wynikające z błędnych decyzji. Będziemy ją krótko nazywali *stratą*.

W podanym jednowymiarowym przykładzie, gdy decyzja jest estymatorem, za stratę przyjmuje się często kwadrat błędu lub jego wartość bezwzględną.

Wartość średnia (oczekiwana) straty jest miarą jakości odbiornika; jej wartość wzięta z przeciwnym znakiem lub odwrotność nazywane bywają *dobrocią*.

Zasadniczy problem polega na określeniu funkcji decyzyjnej, innymi słowy, na zaprojektowaniu odbiornika tak, by przy narzuconych ograniczeniach technicznych był on optymalny, tj. by średnia strata była możliwie najmniejsza. Typ decyzji i związana z nim strata zależą na ogół od struktury nadawanych sygnałów oraz sposobu, w jaki są one wykorzystywane przez obiekt przeznaczenia wiadomości. Wydaje się jednak, że można wydzielić dwa podstawowe typy decyzji, obejmujące większość przypadków, występujących w telekomunikacji. Typy te stanowią naturalne uogólnienie podanego poprzednio jednowymiarowego przykładu. Opiszemy je nieco dokładniej dla lepszego zilustrowania wprowadzanych pojęć.

Wprowadźmy oznaczenia:

X — przestrzeń sygnałów nadawanych x,
Y — przestrzeń sygnałów odbieranych y,
$P_{X\times Y}$ — rozkład prawdopodobieństwa na $X\times Y$,

$P_{X|y}$ — wynikający z $P_{X\times Y}$ warunkowy rozkład prawdopodobieństwa na X przy ustalonym $y \in Y$,

P_Y — wynikający z $P_{X\times Y}$ brzegowy rozkład na Y.

1. Decyzje punktowe. Decyzja punktowa jest elementem przestrzeni X. Innymi słowy, funkcją decyzyjną jest w tym przypadku estymator $x^*(y)$ przyporządkowujący każdemu sygnałowi odebranemu $y \in Y$ jeden z sygnałów nadawanych: $x^*(y) \in X$. Dana też jest strata $R(x', x'')$ będąca nieujemną funkcją par sygnałów nadawanych $x' \in X$ i $x'' \in X$, którą należy rozumieć jako wyrażoną liczbowo szkodę z powzięcia decyzji, że nadany został sygnał x'', podczas gdy naprawdę nadano sygnał x'. Warunkową wartość oczekiwaną straty, pod warunkiem, że odebrany został sygnał y, będziemy nazywali *ryzykiem warunkowym* i oznaczali $r(x^*|y)$; a więc

$$r(x^*|y) = \int_X R[x, x^*(y)]dP_{X|y}. \tag{2.1}$$

Warunkowe ryzyko charakteryzuje dobroć odbioru przy ustalonym sygnale odebranym. Ogólną miarą jakości odbioru jest oczekiwana wartość straty rozciągnięta na całą przestrzeń $X\times Y$. Będziemy ją nazywali po prostu *ryzykiem* i oznaczali $r(x^*|Y)$. Tak więc

$$r(x^*|Y) = \int_{X\times Y} R[x, x^*(y)]dP_{X\times Y} = \int_Y r(x^*|y)dP_Y. \tag{2.2}$$

Czyni się przy tym założenie, zawsze spełnione w zastosowaniach, że strata $R(x', x'')$ i estymator $x^*(y)$ są takie, że wielkości określone przez (2.1) i (2.2) istnieją.

Przy decyzjach punktowych strata może być interpretowana jako uogólnienie miar zniekształceń stosowanych w teorii telekomunikacji. Jako przykład podajmy przypadek, gdy przestrzeń X jest przestrzenią $\tilde{L}_2(-\infty, \infty)$ (sygnały mają skończoną moc). Wówczas strata bywa na ogół określana jako

$$R(x', x'') = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x'(t) - x''(t)]^2 dt. \tag{2.3}$$

Wzory (2.1) oraz (2.2) dają w tym wypadku dobrze znany średni błąd kwadratowy. Pracę N. Wienera [21] można przytoczyć jako klasyczny przykład znalezienia optymalnej liniowej funkcji decyzyjnej, gdy przyjmuje się taką miarę jakości oraz zakłada, że sygnały nadawane i addytywny szum mają strukturę realizacji procesów stacjonarnych.

2. Decyzje obszarowe. Decyzja obszarowa jest podzbiorem przestrzeni sygnałów nadawanych, o którym przypuszcza się, że zawiera sygnał nadany. Innymi słowy, funkcja decyzyjna przyporządkowuje każdemu sygnałowi odebranemu $y \in Y$ podzbiór mierzalny $A|y \subset X$.

Stratę definiuje się wzorem

$$R(x, A) = \begin{cases} 0, & \text{jeśli} \quad x \in A, \\ 1, & \text{jeśli} \quad x \in X - A. \end{cases} \tag{2.4}$$

Znaczy to, że jednostkowa szkoda wynika tylko wtedy, gdy rzeczywiście nadany sygnał x nie jest zawarty w estymującym obszarze $A(y)$.

Z uwagi na zależność obszaru estymującego od sygnału odebranego strata określona wzorem (2.4) jest w gruncie rzeczy funkcją x i y. Analogicznie do przypadku poprzedniego definiuje się warunkowe ryzyko i ryzyko wzorami

$$r(A|y) = \int_X R[x, A(y)]dP_{X|y}, \tag{2.5}$$

$$r(A|Y) = \int_{X \times Y} R[x, A(y)]dP_{X\times Y} = \int_Y r(A|y)dP_Y. \tag{2.6}$$

Podkreślmy, że ryzyko jest funkcjonałem przyporządkowującym wartość liczbową f u n k c j i decyzyjnej $A(y)$ jako całości, a nie jej wartości w poszczególnym y.

Jest widoczne, że przy definicji (2.4) warunkowe ryzyko $r(A|y)$ jest równe prawdopodobieństwu $P_0(y)$ i że przy ustalonym sygnale odebranym y obszar estymujący $A(y)$ nie pokryje sygnału nadanego. Innymi słowy, zachodzi relacja

$$r(A|y) = P_{X|y}[X - A(y)] = P_0(y). \tag{2.7}$$

Jednakże samo prawdopodobieństwo $P_0(y)$ nie określa jakości systemu, gdyż jakość ta zależy od tego jaki „wielki" jest podzbiór $A(y)$. Dlatego założymy, iż w przestrzeni X dana jest dodatkowo miara γ określona na podzbiorach przestrzeni X.

Można ją interpretować jako miernik strat ponoszonych z powodu niedokładności odpowiedzi. Aby zasugerować techniczny sens tej funkcji podzbiorów, będziemy nazywali ją *objętością*.

W prostym jednowymiarowym przykładzie podanym poprzednio można w sposób naturalny zdefiniować miarę γ przedziałów jako ich długość. Gdy przestrzeń X składa się z przeliczalnej ilości elementów, miarę γ definiuje się z zasady jako liczebność podzbiorów.

Warunkową miarą jakości urządzenia podejmującego decyzje obszarowe, przy ustalonym y, jest więc para liczb: warunkowe prawdopodobieństwo błędu $P_0(y)$ i warunkowa dokładność estymacji

$$D(A|y) = \gamma[A(y)]. \tag{2.8}$$

Rolę ogólnej miary jakości urządzenia podejmującego decyzje obszarowe gra para liczb: ryzyko $r(A \mid Y)$ i średnia ze względu na y z warunkowych dokładności estymacji, tj.

$$D(A \mid Y) = \int_Y D(A \mid y) dP_Y = \int_Y \gamma[A(y)] dP_Y. \tag{2.9}$$

Zakładamy, że zbiory $A(y)$ oraz miara γ są takie, iż wielkości (2.8) i (2.9) istnieją.

Decyzje obszarowe są często spotykane w telekomunikacji, zwłaszcza gdy przestrzeń sygnałów nadawanych jest skończona lub przeliczalna. W szczególności z tego typu decyzjami wiąże się w istotny sposób praca Shannona [15].

Opisując decyzje punktowe i obszarowe zakładaliśmy, że sygnał odebrany jest ustalony, a sygnał nadawany traktowaliśmy jako przypadkowy (wzory (2.1), (2.5), (2.8)). Odpowiada to stanowisku obserwatora, który znajduje się po stronie odbiorczej kanału przesyłowego. W statystyce matematycznej taka metoda nazywa się *retrospektywną* [17] i związana jest z operowaniem prawdopodobieństwem a priori i twierdzeniem Bayesa.

W analogiczny sposób obserwator znajdujący się po stronie nadawczej będzie uważał sygnał nadany za ustalony, a sygnały odebrane, a zatem decyzje, za przypadkowe. Metoda ta bywa nazywana *prospektywną*; typowym jej przykładem jest metoda operująca pojęciem wiarogodności. Zasadniczą cechą omawianej metody jest obywanie się bez rozkładu a priori. Jednak, jak na to zwrócono uwagę w pracach [18], [19], stanowi to jedynie formalne ominięcie trudności, gdyż wiarogodność okazuje się równą prawdopodobieństwu a posteriori przy odpowiednio dobranym rozkładzie a priori. Biorąc to pod uwagę oraz to, że punkt widzenia strony odbiorczej jest bardziej naturalny, w pracy tej będziemy się wyłącznie posługiwali metodą retrospektywną.

III. Struktura urządzeń odbiorczych — pojęcie członu wstępnego

Konsekwentne zastosowanie w praktyce postępowania opisanego w poprzednim rozdziale jest możliwe tylko w nielicznych wypadkach. Na ogół bowiem rozwiązanie matematycznego problemu optymizacji nastręcza zbyt duże trudności rachunkowe, a gdy daje się uzyskać, techniczne urzeczywistnienie jest zbyt skomplikowane z praktycznego punktu widzenia. W takich sytuacjach dogodnie jest rozbić urządzenie odbiorcze na dwa człony: *człon wstępny*, dokonywający nad sygnałem odebranym wstępnych operacji upraszczających, i *człon drugi*, podejmujący decyzję na podstawie uproszczonych sygnałów. Potrzeba takiego rozbicia na człony wyłania się również w uniwersalnych systemach, gdzie ten sam sygnał odebrany steruje różnymi urządzeniami podejmującymi decyzje. Podobna sytuacja powstaje, gdy skalarne funkcje R i γ nie są dokładnie znane lub nie są zadowalającymi miernikami jakości i trzeba brać pod uwagę równocześnie różne postacie tych funkcji. Tak jest w szczególności wówczas, gdy zadania członu decyzyjnego spełnia człowiek. W takich przypadkach całą część elektronową odbiornika można traktować jako człon wstępny. W większości znanych urządzeń odbiorczych rozbicie na omawiane dwa człony jest bardzo wyraźne.

Analogiczna metoda stosowana bywa w statystyce matematycznej, gdzie w pewnych przypadkach możliwe są daleko idące uproszczenia próbek (sygnałów odebranych) bez utraty jakichkolwiek danych statystycznych o estymowanych parametrach (sygnałach nadanych). Operacje posiadające tę własność nazywamy *statystykami dostatecznymi.*

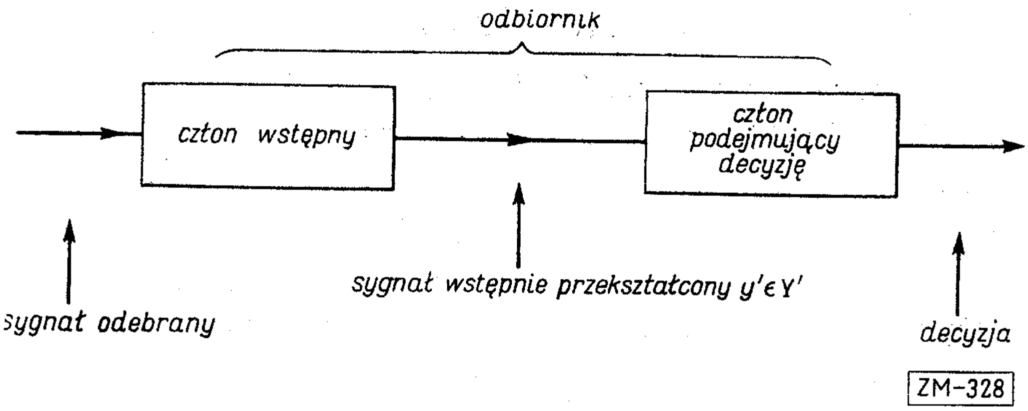
Obliczanie prawdopodobieństwa warunkowego $P_{X|y}$, gdzie y jest sygnałem odebranym, nie pociąga oczywiście żadnej straty danych o zbiorze sygnałów nadawanych X. Człon wstępny, dokonujący takiej operacji, będziemy nazywali *idealnym* członem wstępnym. Został on wprowadzony do teorii telekomunikacji przez P. M. Woodwarda [22] (rozdz. IV), który zresztą wbrew naszemu stanowisku nie rozważał w ogóle członów decyzyjnych.

Zasadniczą zaletą idealnego członu wstępnego jest to, że nie pogarsza on dobroci jakiegokolwiek optymalnego urządzenia podejmującego decyzje. Jednakże idealny człon wstępny jest faktycznie użyteczny tylko wówczas, gdy prawo rozkładu $P_{X|y}$ jest mniej skomplikowane niż sam sygnał odebrany y. Na ogół człon wstępny bywa określany przez czynniki ekonomiczne, technologiczne, lub wybiera się go w sposób heurystyczny. Z zasady nie jest on idealny i za cenę uzyskanych uproszczeń dobroć optymalnego członu podejmującego decyzje, sterowanego przez człon wstępny, jest mniejsza niż dobroć optymalnego urządzenia podejmującego decyzje bezpośrednio na podstawie odebranych sygnałów.

W dalszym ciągu będziemy przyjmowali, że odbiornik składa się, z dwu członów. Wejściami członu wstępnego są odebrane sygnały $y \in Y$, wyjściami — sygnały $y' = T(y)$, które należą do przestrzeni Y' i są wynikiem wstępnego upraszczającego przekształcenia T. Człon drugi jest członem podejmującym decyzje. Jego wejściami są sygnały $y' \in Y'$, wyjściami decyzje. W takiej sytuacji potrzebna jest miara jakości stopnia wstępnego ze względu na następujące po nim człony decyzyjne. Jest oczywiste, że wszystko, co wiemy o sygnale nadanym, gdy odebraliśmy sygnał y, jest zawarte w warunkowym rozkładzie prawdopodobieństwa $P_{X|y}$. Gdy dany jest człon wstępny opisany przez operację $y' = Ty$, to cała wiedza o sygnale nadanym, gdy znany jest tylko sygnał uproszczony y', tkwi w warunkowym rozkładzie prawdopodobieństwa $P_{X|Ty}$. Zatem miara jakości stopnia wstępnego powinna mieć sens „odległości" między rozkładami $P_{X|y}$ oraz $P_{X|Ty}$. W dalszym ciągu pokażemy, że dogodnie jest miarę jakości stopnia wstępnego zdefiniować za pomocą, dobrze znanego w teorii informacji, funkcjonału

$$\int_X \left[-\lg \frac{dP_{X|Ty}}{dP_{X|y}} \right] dP_{X|y}.$$

W rozdziale IV, posługując się tym funkcjonałem i wprowadzając w miejsce $P_{X|Ty}$ odpowiednio zdefiniowany rozkład prawdopodobieństwa



Rys. 3

uzależniony od decyzji, uzyskamy interesujące nas związki między entropią a ryzykiem. W rozdziale V pokażemy, w jaki sposób człon wstępny wpływa na te zależności. Wyłonią się przy tym powiązania między ilością informacji a ryzykiem. W rozdziale VI podamy pewne rozważania uzasadniające przyjętą tu definicję miary jakości stopnia wstępnego.

IV. Związki między entropią a ryzykiem

W rozdziale tym zachowamy wprowadzone poprzednio oznaczenia:

X — przestrzeń sygnałów nadawanych x,

Y — przestrzeń sygnałów odbieranych y,

$P_{X\times Y}, P_{X|y}, P_Y$ — rozkłady prawdopodobieństwa: na $X\times Y$, warunkowy na X przy ustalonym y i brzegowy na Y.

γ — pewna miara na X.

W dalszym ciągu przyjmiemy następujące założenie generalne:

(Z_1) Rozkłady $P_{X|y}$ są absolutnie ciągłe względem γ.

W związku z tym wprowadzamy oznaczenie:

$p(x|y) \stackrel{\mathrm{df}}{=} \dfrac{dP_{X|y}}{d\gamma}$ — gęstość prawdopodobieństwa rozkładu $P_{X|y}$ względem γ.

Istotną rolę odgrywać będzie lemat, udowodniony w pracy [9].

LEMAT 1. *Jeżeli $p_1(x)$ i $p_2(x)$ są gęstościami prawdopodobieństwa na X względem γ, to*

$$\int_X [-\lg p_2(x)]p_1(x)d\gamma(x) \geqslant \int_X [-\lg p_1(x)]p_1(x)d\gamma(x), \tag{4.1}$$

przy czym równość zachodzi wtedy i tylko wtedy, gdy

$$p_1(x) = p_2(x)$$

dla prawie wszystkich x w sensie miary γ.

Nierówność (4.1) można napisać także tak:

$$\int_X \left[-\lg \frac{p_2(x)}{p_1(x)}\right] p_1(x)d\gamma(x) \geqslant 0. \tag{4.1'}$$

Wprowadzimy pewne definicje: *entropia warunkowa*

$$H(X|y) = \int_X [-\lg p(x|y)]p(x|y)d\gamma(x) \tag{4.2}$$

i entropia

$$H(X|Y) = \int_Y H(X|y)dP_Y. \tag{4.3}$$

Zajmiemy się najpierw decyzjami punktowymi. Przypomnijmy oznaczenia wprowadzone w rozdziale II:

$x^*(y) \in X$ — estymator sygnału nadanego zdefiniowany dla $y \in Y$,

$R(x', x'')$, $x' \in X$, $x'' \in X$ — strata wynikła z tego, że nadano x', a postępuje się tak jak gdyby nadano x'',

$$r(x^*|y) = \int_X R(x, x^*(y))p(x|y)d\gamma(x) \quad \text{(ryzyko warunkowe)} \tag{4.4}$$

$$r(x^*|Y) = \int_Y r(x^*|y)dP_Y \quad \text{(ryzyko)}. \tag{4.5}$$

Zdefiniujmy pomocnicze rozkłady prawdopodobieństwa przyporządkowane sygnałom $x \in X$:

$$dP_{X|x',k_2}/d\gamma \stackrel{\mathrm{df}}{=} p^*(x|x', k_2) \stackrel{\mathrm{df}}{=} k_1 \exp[-k_2 R(x, x')], \tag{4.6}$$

gdzie

$$k_1 = \frac{1}{\int\limits_X \exp[-k_2 R(x, x')]\, d\gamma(x)}. \tag{4.7}$$

TWIERDZENIE 1. *Przy ustalonym estymatorze x^* i $y \in Y$ zachodzi równość*

$$r(x^*|y) = \frac{1}{k_2}\,[H(X|y) + \lg k_1 + \Delta_1(y, k_2)], \tag{4.8}$$

gdzie k_1 dane jest przez (4.7) *przy $x' = x^*(y)$, a*

$$\Delta_1(y, k_2) = \int\limits_X \left[-\lg \frac{p^*\big(x|x^*(y), k_2\big)}{p(x|y)}\right] p(x|y)\, d\gamma(x) \geqslant 0. \tag{4.9}$$

Dowód. Wobec lematu 1 tylko równość w (4.9) wymaga dowodu. Przekształcając (4.8) otrzymujemy

$$\Delta_1(y, k_2) = k_2 r(x^*|y) - H(X|y) - \lg k_1.$$

Wobec (4.4) mamy równość

$$k_2 r(x^*|y) - \lg k_1 = \int\limits_X \big[k_2 R\big(x, x^*(y)\big) - \lg k_1\big] p(x|y)\, d\gamma(x) =$$

$$= \int\limits_X \big[-\lg p^*\big(x|x^*(y), k_2\big)\big] p(x|y)\, d\gamma(x).$$

To wraz z (4.2) dowodzi równości w (4.9).

Twierdzenie 1 pozwala nam w niektórych przypadkach uzyskać oszacowanie warunkowego ryzyka niezależnie od sposobu estymacji. Z twierdzenia 1 można bowiem wyciągnąć bezpośredni następujący

WNIOSEK 1. *Przy założeniach twierdzenia* 1 *dla każdego $k_2 > 0$ zachodzi nierówność*

$$r(x^*|y) \geqslant \frac{1}{k_2}\,[H(X|y) + \lg k_1], \tag{4.10}$$

gdzie k_1 dane jest przez (4.7); *równość jest osiągnięta wtedy i tylko wtedy, gdy $p^*\big(x|x^*(y), k_2\big) = p(x|y)$ prawie wszędzie w sensie miary γ, gdzie $p^*\big(x|x^*(y), k_2\big)$ dane jest przez* (4.6).

W nierówności (4.10) k_1 na ogół zależy od sposobu szacowania, tj. od estymatora $x^*(y)$. W ważnych przypadkach szczególnych k_1 jest niezależne od x^*, a zależy tylko od k_2. Oto szczególnie interesujący przypadek:

TWIERDZENIE 2. *Jeśli w twierdzeniu* 1 *X jest przestrzenią euklidesową n-wymiarową o punktach $x = (x_1, x_2, \ldots, x_n)$, miara γ jest n-wymiarową* *miarą Lebesgue'a, a strata określona jest równością*

$$(4.11)\qquad R(x', x'') = \sum_{i,j=1}^{n} a_{ij}(x_i' - x_i'')(x_j' - x_j''),$$

gdzie macierz $\|a_{ij}\|$ *jest symetryczna i dodatnio określona, to przy ustalonym* $y \in Y$ *i dowolnym* $x^*(y)$ *dla każdego* $k_2 > 0$ *zachodzi nierówność*

$$(4.12)\qquad r(x^*|y) \geqslant \frac{1}{k_2}[H(X|y) + \lg k_1],$$

gdzie

$$(4.13)\qquad k_1 = k_2^{n/2}\pi^{-n/2}|a_{ij}|^{1/2},$$

a $|a_{ij}|$ *oznacza wyznacznik macierzy* $\|a_{ij}\|$.

Dowód. Twierdzenie 2 jest bezpośrednią konsekwencją wniosku 1. Wymagająca dowodu równość (4.13) jest bezpośrednią konsekwencją dobrze znanych własności n-wymiarowych rozkładów normalnych (por. Cramér [4], wzór (11.12.2)).

Udowodniona nierówność (4.12) zachodzi przy dowolnym $k_2 > 0$. Można więc nałożyć na k_2 dodatkowe warunki. Jednym z nich jest wybór tej wartości k_2, przy której prawa strona w (4.12) osiąga największą wartość. Otóż łatwym rachunkiem można udowodnić

LEMAT 2. *Dla dowolnego rzeczywistego H zachodzi relacja*

$$\max_{k_2>0}\left\{\frac{1}{k_2}\left(H + \lg k_2^{n/2}\pi^{-n/2}|a_{ij}|^{1/2}\right)\right\} = \frac{n}{2\pi e}|a_{ij}|^{1/n}e^{2H/n},$$

a maksimum to zachodzi dla

$$k_2 = \pi e|a_{ij}|^{-1/n}e^{-2H/n}.$$

Dowód. Pytamy, dla jakiego k_2 wyrażenie

$$(4.14)\qquad \frac{1}{k_2}\left[\left(H + \frac{1}{2}\lg|a_{ij}| - \frac{n}{2}\lg\pi\right) + \frac{n}{2}\lg k_2\right]$$

osiąga maksimum. Oznaczmy wyrazy w nawiasach ( ) przez u i obliczmy pochodną względem k_2 otrzymanego wyrażenia, a następnie wynik pomnóżmy przez k_2. Otrzymamy równanie

$$\frac{n}{2} - u - \frac{n}{2}\lg k_2 = 0.$$

Stąd wyznaczamy wartość k_2, a podstawiwszy ją do (4.14) wyznaczamy szukane maksimum.

Stosując lemat 2 do nierówności (4.12) otrzymujemy

TWIERDZENIE 3. *Przy założeniach twierdzenia 2 niezależnie od sposobu szacowania zachodzi nierówność*

$$r(x^*|y) \geqslant \frac{n}{2\pi e}|a_{ij}|^{1/n}\exp\frac{2H(X|y)}{n}. \tag{4.15}$$

Drugim naturalnym sposobem ustalania wartości k_2 jest taki dobór k_2, żeby oczekiwana strata obliczana względem rozkładu o gęstości $p^*(x|x^*(y), k_2)$ określonej przez (4.6) była równa ryzyku warunkowemu, to znaczy, żeby oprócz (4.7) spełniony był warunek

$$\int\limits_X R[x, x^*(y)]k_1\exp\{-k_2R[x, x^*(y)]\}d\gamma(x) = \tag{4.16}$$

$$= \int\limits_X R[x, x^*(y)]p(x|y)d\gamma(x).$$

Pokażemy, że dołączenie tego warunku do założeń twierdzenia 2 prowadzi również do oszacowania (4.15).

TWIERDZENIE 4. *Jeśli spełnione są założenia twierdzenia 2 oraz relacja* (4.16), *to*

$$r(x^*|y) \geqslant \frac{n}{2\pi e}|a_{ij}|^{1/n}\exp\frac{2H(X|y)}{n}. \tag{4.17}$$

Dowód. Obliczmy najpierw lewą stronę (4.16). Wobec założeń twierdzenia 2 mamy

$$L = \int\limits_X R[x, x^*(y)]k_1\exp\{-k_2R[x, x^*(y)]\}d\gamma(x) = \tag{4.18}$$

$$= \int\limits_X \sum_{i,j}^{n} a_{ij}[x_i - x_i^*(y)][x_j - x_j^*(y)]\times$$

$$\times k_1\exp\Big\{-\tfrac{1}{2}\sum_{i,j}^{n} 2k_2a_{ij}[x_i - x_i^*(y)][x_j - x_j^*(y)]\Big\}dx_1dx_2\ldots dx_n =$$

$$= \sum_{i,j}^{n} a_{ij}\int\limits_X [x_i - x_i^*(y)][x_j - x_j^*(y)]\times$$

$$\times k_1\exp\Big\{-\tfrac{1}{2}\sum_{i,j}^{n} 2k_2a_{ij}[x_i - x_i^*(y)][x_j - x_j^*(y)]\Big\}dx_1dx_2\ldots dx_n =$$

$$= \sum_{i,j}^{n} a_{ij}\lambda_{ij},$$

gdzie przez λ_{ij} oznaczyliśmy drugie momenty centralne występującego tu n-wymiarowego rozkładu normalnego.

Wiadomo, z własności rozkładu normalnego (patrz Cramér [4], § 24.1 i § 24.2), że macierz $\|\lambda_{ij}\|$ jest macierzą odwrotną do macierzy $2k_2\|a_{ij}\|$, a zatem

$$\|\lambda_{ij}\| = \frac{1}{2k_2}\|a_{ij}\|^{-1}. \tag{4.19}$$

Jak wiadomo,

$$\|a_{ij}\|^{-1} = \frac{1}{|a_{ij}|}\|A_{ij}\|,$$

gdzie przez A_{ij} oznaczamy algebraiczne dopełnienie elementu a_{ij} w macierzy $\|a_{ij}\|$. Ponieważ zaś dla każdego $j = 1, 2, \ldots, n$

$$\sum_{i=1}^{n} a_{ij}A_{ij} = |a_{ij}|,$$

przeto wobec (4.18) i (4.19) mamy

$$L = n/2k_2,$$

a zatem (4.16) jest spełnione dla $k_2 = n/2r(x^*|y)$. Podstawiając tę wartość do (4.12) i uwzględniając (4.13) otrzymujemy po prostych przekształceniach (4.17), co kończy dowód.

Zauważmy, że jeśli przy założeniach twierdzenia 2 podstawimy do (4.8) za k_2 wartość $n/2r(x^*|y)$ to otrzymamy równość

$$r(x^*|y) = \frac{n}{2\pi e}|a_{ij}|^{1/n}\exp\frac{2}{n}\left[H(X|y) + \Delta_1\left(y, \frac{n}{2r(x^*|y)}\right)\right]. \tag{4.20}$$

Występującą tu wielkość

$$\Delta_1[y, n/2r(x^*|y)] \tag{4.20a}$$

będziemy oznaczali krótko przez $\Delta_1(y)$.

W dotychczasowych twierdzeniach ustaliliśmy związki między warunkowym ryzykiem i warunkową entropią. A oto związek między ryzykiem i średnią entropią:

Twierdzenie 5. *Jeśli spełnione są założenia twierdzenia 2, to zachodzi nierówność*

$$r(x^*|Y) \geqslant \frac{n}{2\pi e}|a_{ij}|^{1/n}\exp\left\{\frac{2}{n}[H(X|Y) + \bar{\Delta}_1]\right\}, \tag{4.21}$$

gdzie

$$\bar{\Delta}_1 = \int_Y \Delta_1(y) dP_Y \geqslant 0. \tag{4.22}$$

Dowód. Całkując stronami (4.20) otrzymujemy

$$r(x^* \mid Y) = \frac{n}{2\pi e} |a_{ij}|^{1/n} \int_Y \exp\left\{\frac{2}{n}[H(X \mid y) + \Delta_1(y)]\right\} dP_Y. \tag{4.23}$$

Z nierówności Jensena (patrz np. [13], rozdział 3) wynika nierówność

$$\int_Y \exp\left\{\frac{2}{n}[H(X \mid y) + \Delta_1(y)]\right\} dP_Y \geqslant \exp\left\{\frac{2}{n}[H(X \mid Y) + \bar{\Delta}_1]\right\}. \tag{4.24}$$

To wraz z poprzednim daje nierówność (4.21). Nieujemność $\bar{\Delta}_1$ jest oczywista.

Przechodzimy do omówienia decyzji obszarowych. Oznaczenie

$$X,\ Y,\ P_{X\times Y},\ P_{X|y},\ P_Y,\ \gamma,\ p(x \mid y),\ H(X \mid y),\ H(X \mid Y)$$

zachowuje dawny sens. Przypomnijmy dalej oznaczenia wprowadzone w rozdziale II:

$$A(y) \quad \text{— obszar estymujący,}$$

$$R(x \mid A) = \begin{cases} 0, & x \in A, \\ 1, & x \in X - A, \end{cases} \quad \text{— strata,} \tag{4.25}$$

$$P_0(y) = P_{X|y}\big(X - A(y)\big) \quad \text{— ryzyko.} \tag{4.26}$$

A oto definicja pomocniczych rozkładów prawdopodobieństwa, przyporządkowanych obszarom estymującym:

$$dP^*_{X|A}/d\gamma = p^*(x \mid A) \stackrel{\text{df}}{=} \begin{cases} P_{X|y}(A)/\gamma(A) & \text{dla } x \in A, \\ P_{X|y}(X-A)/\gamma(X-A) & \text{dla } x \in X - A. \end{cases} \tag{4.27}$$

Zwracamy uwagę, że definicja (4.27) jest tak dobrana, że wynikają z niej równości

$$P^*_{X|A}(A) = P_{X|y}(A), \tag{4.28}$$

$$P^*_{X|A}(X-A) = P_{X|y}(X-A). \tag{4.29}$$

Ta ostatnia, wobec (4.26), jest analogonem relacji (4.16), a obie są równoważne.

Aby definicja (4.27) miała sens nie wymagający dodatkowych omówień, w rozważaniach dotyczących decyzji obszarowych przyjmujemy generalne założenie

$$\gamma(X) < \infty. \tag{Z_2}$$

Zdefiniujmy dalej wielkość analogiczną do $\Delta_1(y)$:

$$(4.30)\qquad \Delta_2(y) \overset{\mathrm{df}}{=} \int_X [-\lg p^*[x|A(y)]p(x|y)d\gamma(x) - \\ - \int_X [-\lg p(x|y)]p(x|y)d\gamma(x).$$

Wobec lematu 1 jest zawsze

$$(4.31)\qquad \Delta_2(y) \geqslant 0,$$

a równość zachodzi wtedy i tylko wtedy, gdy

$$(4.32)\qquad p^*[x|A(y)] = p(x|y)$$

prawie wszędzie według γ.

Oto analogon twierdzenia 1.

TWIERDZENIE 6. *Przy ustalonym obszarze estymującym $A(y)$ i $y \in Y$ dla każdego $k_2 > 0$ zachodzi równość*

$$(4.33)\qquad P_0(y) = \frac{1}{k_2}\left[H(X|y) - \lg\left(\gamma(A) + e^{-k_2}\gamma(X-A)\right) + \Delta_2(y, k_2)\right],$$

gdzie

$$(4.34)\qquad \Delta_2(y, k_2) = \int_X \left[-\lg \frac{p^*\left((x|A(y), k_2\right)}{p(x|y)}\right] p(x|y)d\gamma(x) \geqslant 0,$$

$$(4.35)\qquad dP^*_{X|A,k_2}/d\gamma \overset{\mathrm{df}}{=} p^*(x|A, k_2) \overset{\mathrm{df}}{=} k_1 e^{-k_2 R(x,A)},$$

$$(4.36)\qquad k_1 = \frac{1}{\int_X e^{-k_2 R(x,A)} d\gamma(x)}.$$

Dowód. Jest to wniosek z twierdzenia 1 z tym, że rolę $x^*(y)$ obejmuje $A(y)$, strata zdefiniowana jest przez (4.25), a pomocnicze rozkłady prawdopodobieństwa $P^*_{X|A,k_2}$ definiuje się analogicznie do (4.6) i (4.7) przez (4.35) i (4.36).

Wobec (4.25) można wyznaczyć k_1 explicite:

$$k_1 = [\gamma(A) + e^{-k_2}\gamma(X-A)]^{-1}.$$

Po podstawieniu k_1 do (4.8), (4.9) i uwzględnieniu (4.26) otrzymuje się twierdzenie 6.

Przez dobór k_2 możemy zapewnić spełnienie relacji (4.28). Łatwo sprawdzić, że dla

$$(4.37)\qquad k_2 = \lg \frac{[1-P_0(y)]\gamma(X-A)}{P_0(y)\gamma(A)}$$

uzyskujemy pokrycie się rozkładu $P^*_{X|A,k_2}$ z rozkładem $P^*_{X|A}$ zdefiniowanym przez (4.27). Podstawiając (4.37) do (4.33) otrzymujemy równość

$$(4.38)\quad P_0(y) = \frac{1}{\lg\dfrac{[1-P_0(y)]\gamma(X-A)}{P_0(y)\gamma(A)}}\cdot\{H(X|y)+\lg[1-P_0(y)]+\Delta_2(y)\},$$

która po dalszych przekształceniach daje

$$(4.39)\quad -P_0(y)\lg P_0(y)-[1-P_0(y)]\lg[1-P_0(y)]+P_0(y)\lg[1-\gamma(A)/\gamma(X)]+$$
$$+[1-P_0(y)]\lg\gamma(A)/\gamma(X) = H(X|y)-\lg\gamma(X)+\Delta_2(y).$$

Na mocy lematu 1 zastosowanego do lewej strony tej równości (X dwupunktowe, $p_1(x_1) = P_0(y)$, $p_1(x_2) = 1-P_0(y)$, $p_2(x_1) = 1-\gamma(A)/\gamma(x)$, $p_2(x_2) = \gamma(A)/\gamma(x)$ a rolę γ z lematu gra liczność zbioru) wnosimy, że jest ona niedodatnia. A zatem

$$(4.40)\quad H(X|y)-\lg\gamma(X)+\Delta_2(y) \leqslant 0$$

i równość zachodzi wtedy i tylko wtedy, gdy

$$(4.40\text{a})\quad P_0(y) = 1-\frac{\gamma(A)}{\gamma(X)}.$$

Rozpatrzmy teraz funkcję

$$(4.41)\quad M(P,z) = (1-P)\lg z+P\lg(1-z)-P\lg P-(1-P)\lg(1-P)$$
$$0\leqslant P\leqslant 1,\quad 0\leqslant z\leqslant 1.$$

Wiemy już, że $M(P,z)\leqslant 0$ i że $M(P,z) = 0$ dla $P = 1-z$. Zbadajmy przebieg funkcji $M(P,z)$ traktowanej przy ustalonym P jako funkcja z, a przy ustalonym z jako funkcja P. Mamy

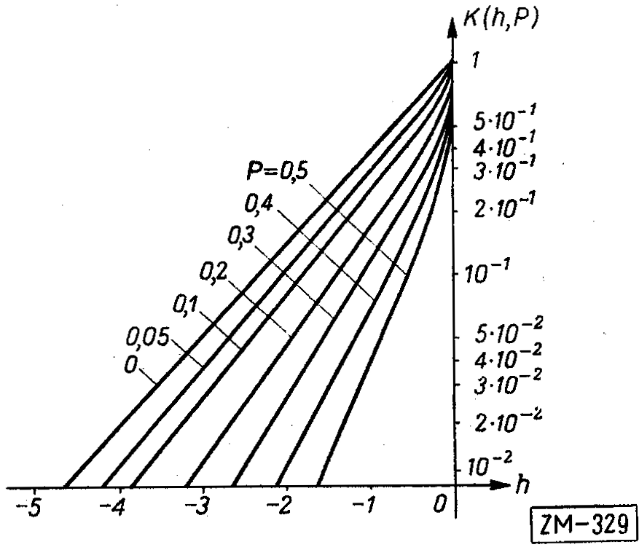
$$(4.42)\quad \frac{\partial M(P,z)}{\partial z} = \frac{1-z-P}{z(1-z)}.$$

Widzimy stąd, że przy ustalonym P

$$(4.43)\quad \frac{\partial M(P,z)}{\partial z}\begin{cases} >0 & \text{dla}\quad z<1-P,\\ =0 & \text{dla}\quad z=1-P,\\ <0 & \text{dla}\quad 1-P<z\end{cases}$$

i że wobec tego $M(P,z)$ traktowana jako funkcja z przy ustalonym P najpierw rośnie od $-\infty$ do 0, a potem maleje od 0 do $-\infty$.

Oznaczmy przez $K(h,P)$, $[0<P<1,\ -\infty<h\leqslant 0,\ 0<K(h,P)\leqslant 1-P]$, funkcję odwrotną do rosnącej gałęzi funkcji $M(P,z)$, $P=\text{const}$. Funkcja $K(h,P)$ jest więc przy ustalonym P funkcją określoną na półosi $-\infty<h\leqslant 0$, rosnącą od 0 do $1-P$. Wykresy tej funkcji pokazane są na rysunku 4:



Rys. 4

Mamy dalej

$$\frac{\partial M(P,z)}{\partial P}=\lg\frac{(1-P)(1-z)}{Pz}. \tag{4.44}$$

Widzimy stąd, że przy ustalonym z

$$\frac{\partial M(P,z)}{dP}\begin{cases}>0 & \text{dla}\quad P<1-z,\\ =0 & \text{dla}\quad P=1-z,\\ <0 & \text{dla}\quad 1-z<P,\end{cases} \tag{4.45}$$

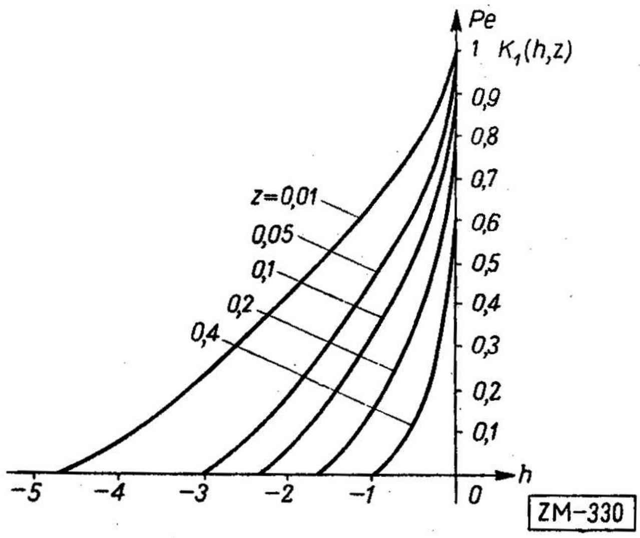
a więc $M(P,z)$ dla stałego z najpierw rośnie od $\lg z$ do 0, potem maleje od 0 do $\lg(1-z)$.

Oznaczmy przez $K_1(h,z)$, $[0<z<1,\ \lg z\leqslant 0,\ 0\leqslant K_1(h,z)\leqslant 1-z]$ funkcję odwrotną rosnącej gałęzi funkcji $M(P,z)$ dla stałego z. $K_1(h,z)$ jest więc, przy ustalonym z, określona w przedziale $\lg z\leqslant h\leqslant 0$ i rośnie w tym przedziale od 0 do $1-z$. Wykresy funkcji $K_1(h,z)$ pokazuje rysunek 5.

Zwróćmy uwagę, że $K(h,P)$ jest dla danego h i P mniejszym z dwu rozwiązań względem z równania

$$M(P,z)=h, \tag{4.46}$$

oraz że jest ono jednoznacznym rozwiązaniem równania (4.46), gdy $P+z < 1$. Podobnie $K_1(h, z)$ jest mniejszym z dwu rozwiązań tego równania względem P przy ustalonym z oraz jest jednoznacznym rozwiązaniem równania (4.46), gdy $P+z < 1$.



Rys. 5

Z tego, co dotychczas powiedzieliśmy o funkcjach $K(h, P)$ i $K_1(h, z)$, oraz z relacji (4.39) wynika następujące

TWIERDZENIE 7a. *Zachodzą następujące nierówności:*

$$\frac{\gamma(A)}{\gamma(X)} \geqslant K[H(X|y) - \lg\gamma(X) + \Delta_2(y), P_0'(y)], \tag{4.47}$$

$$P_0(y) \geqslant K_1\left[H(X|y) - \lg\gamma(X) + \Delta_2(y), \frac{\gamma(A)}{\gamma(X)}\right]. \tag{4.48}$$

Znak równości zachodzi, jeśli $\gamma(A)/\gamma(X)+P_0(y) \leqslant 1$.

Własności występujących tu funkcji $K(h, P)$ oraz $K_1(h, z)$ omówiliśmy poprzednio.

Udowodnimy ponadto

TWIERDZENIE 7b. *Zachodzą relacje*

$$K(h, P) > \exp\{[h - H(P)]/(1-P)\}, \tag{4.49}$$

gdzie

$$H(P) = -P\lg P - (1-P)\lg(1-P) \tag{4.49a}$$

oraz

$$\lim_{P\to 0}\frac{K(h,P)}{e^h} = 1. \tag{4.50}$$

Dowód. Ze wzoru (4.41) widać, że

$$M(P,z) < (1-P)\lg z + H(P),$$

a zatem dla funkcji odwrotnych do występujących w tym wzorze (a traktowanych jako funkcje zmiennej z) będzie zachodziła nierówność przeciwna, czyli (4.49).

Ze wzoru (4.41) widać również, że

$$M(0,z) = \lim_{P\to 0} M(P,z) = \lg z,$$

stąd zaś przechodząc do funkcji odwrotnych ze względu na z otrzymujemy relację

$$K(h,0) = \lim_{P\to 0} K(h,P) = e^h,$$

która dowodzi wzoru (4.50).

W twierdzeniach 6 i 7 była mowa o warunkowym prawdopodobieństwie $P_0(y)$ i warunkowej dokładności $\gamma[A(y)]/\gamma(X)$. Udowodnimy obecnie analogon twierdzenia 5 otrzymany przez wzięcie wartości oczekiwanych względem y. Oznaczmy

$$P_0 = r(A\,|\,Y) = \int_Y P_0(y)\,dP_Y, \tag{4.51}$$

$$\frac{\overline{\gamma A}}{\gamma X} = \frac{D(A\,|\,Y)}{\gamma X} = \int_Y \frac{\gamma[A(y)]}{\gamma X}\,dP_Y. \tag{4.51a}$$

W dalszym ciągu skorzystamy z następującego lematu:

LEMAT 3. *Jeśli $F(x,y)$ jest funkcją dwu zmiennych wypukłą w dół, to dla dowolnych zmiennych losowych ξ i η zachodzi nierówność*

$$\mathrm{E}F(\xi,\eta) \geqslant F(\mathrm{E}\xi, \mathrm{E}\eta). \tag{4.52}$$

Przez E *oznaczono operację uśrednienia.*

TWIERDZENIE 8. *Zachodzą następujące nierówności:*

$$\frac{\overline{\gamma A}}{\gamma X} \geqslant K[H(X\,|\,Y) + \bar{\Delta}_2 - \lg\gamma(X), P_0)], \tag{4.53}$$

$$P_0 \geqslant K_1\left[H(X\,|\,Y) + \bar{\Delta}_2 - \lg\gamma(X), \frac{\overline{\gamma A}}{\gamma X}\right], \tag{4.54}$$

gdzie

$$\bar{\Delta}_2 = \int \Delta_2(y)\,dP_Y. \tag{4.55}$$

Dowód. Ponieważ wartości argumentów funkcji K i K_1 występujące w (4.53) i (4.54) są oczekiwanymi wartościami odpowiednich argumentów występujących we wzorach (4.47) i (4.48), przeto twierdzenie 8 okaże się konsekwencją twierdzenia 7a i lematu 3, gdy wykażemy, że funkcje $K(h,P)$ i $K_1(h,z)$ są wypukłe w dół. Ponieważ funkcja $K(h,P)$ jest odwróceniem rosnących przy ustalonym P gałęzi funkcji $h = M(P,z)$ ze względu na z, $(z \leqslant 1-P)$, a $K_1(h,z)$ jest odwróceniem rosnących gałęzi tejże funkcji $h = M(P,z)$ ze względu na P, przeto wystarczy udowodnić własność wypukłości w górę dla funkcji $h = M(P,z)$. Wiadomo, że warunkiem dostatecznym wypukłości w górę jest niedodatniość drugiej pochodnej, a to jest równoważne z jednoczesnym spełnieniem nierówności

$$M_{PP} \leqslant 0 \quad \text{i} \quad M_{PP}M_{zz} - M_{Pz}^2 \geqslant 0, \tag{4.56}$$

gdzie M_{PP}, M_{Pz} i M_{zz} oznaczają drugie pochodne cząstkowe funkcji M względem wskazanych zmiennych. Wobec (4.42) i (4.44) mamy

$$M_{PP} = -\frac{1}{P} - \frac{1}{1-P},$$

$$M_{Pz} = -\frac{1}{z} - \frac{1}{1-z},$$

$$M_{zz} = -(1-P)\frac{1}{z^2} - P\frac{1}{(1-z)^2}.$$

Stąd widać, że pierwsza z nierówności (4.56) jest spełniona. Po przekształceniach otrzymujemy

$$M_{PP}M_{zz} - M_{Pz}^2 = \frac{(1-P-z)^2}{P(1-P)z^2(1-z)^2},$$

skąd widać, że także druga z nierówności (4.55) jest spełniona. To kończy dowód.

V. Wpływ stopnia wstępnego na stopnie decyzyjne

Posługiwać się będziemy oznaczeniami wprowadzonymi w rozdziale III:

Y — przestrzeń sygnałów odebranych,

Y' — przestrzeń sygnałów uproszczonych,

$T(y)$, $y \in Y$, $T(y) \in Y'$ — operacja upraszczająca,

$P_{X\times Y'}$ — rozkład prawdopodobieństwa w przestrzeni $X\times Y'$ wynikający z operacji T oraz rozkładu $P_{X\times Y}$,

$P_{X|y'}$ — warunkowy rozkład prawdopodobieństwa X przy ustalonym $y' \in Y$ wynikający z $P_{X \times Y'}$,
$p(x|y')$ — gęstość rozkładu $P_{X|y}$ względem γ,
$P_{Y'}$ — rozkład brzegowy na Y' wynikły z $P_{X \times Y'}$,

$$H(X|y') = \int_X [-\lg p(x|y')] p(x|y') d\gamma(x) \tag{5.1}$$

jest to entropia przy ustalonym y', a

$$H(X|Y') = \int_{Y'} H(X|y') dP_{Y'} \tag{5.2}$$

jest to średnia entropia.

Zgodnie z rozważaniami rozdziału III wprowadzimy jeszcze dwa oznaczenia:

$$\Delta_{\mathrm{sw}}(y) = \int_X \left[-\lg \frac{p(x|T(y))}{p(x|y)}\right] p(x|y) d\gamma(x) \tag{5.3}$$

jest to miara jakości stopnia wstępnego przy ustalonym $y \in Y$, a

$$\bar{\Delta}_{\mathrm{sw}} = \int_Y \Delta_{\mathrm{sw}}(y) dP_Y \tag{5.4}$$

jest to średnia miara jakości stopnia wstępnego.

TWIERDZENIE 9. *Zachodzą związki*

$$\bar{\Delta}_{\mathrm{sw}} = H(X|Y') - H(X|Y) \tag{5.5}$$

oraz

$$\bar{\Delta}_{\mathrm{sw}} \geqslant 0, \tag{5.6}$$

przy czym znak równości zachodzi wtedy i tylko wtedy, gdy

$$p(x|y') = p(x|y),$$

tj. gdy statystyka $y' = Ty$ jest dostateczna.

Dowód. Mamy

$$\bar{\Delta}_{\mathrm{sw}} = \int_Y \Delta_{\mathrm{sw}}(y) dP_Y = \int_Y \left\{ \int_X \left[-\lg \frac{p(x|T(y))}{p(x|y)}\right] p(x|y) d\gamma(x) \right\} dP_Y =$$

$$= \int_Y \left\{ \int_X [-\lg p(x|T(y)) - (-\lg p(x|y))] p(x|y) d\gamma(x) \right\} dP_Y =$$

$$= \int_Y \left\{ \int_X [-\lg p(x|T(y))] p(x|y) d\gamma(x) - \int_X [-\lg p(x|y)] p(x|y) d\gamma(x) \right\} dP_Y =$$

$$= \int\limits_{Y} \Big\{ \int\limits_{X} [-\lg p(x|T(y))] p(x|y) d\gamma(x) \Big\} dP_Y - \int\limits_{Y} H(X|y) dP_Y =$$

$$= \int\limits_{Y} \Big\{ \int\limits_{X} [-\lg p(x|T(y))] p(x|y) d\gamma(x) \Big\} dP_Y - H(X|Y) =$$

$$= \int\limits_{X\times Y} [-\lg p(x|T(y))] dP_{X\times Y} - H(X|Y),$$

czyli

$$\int \Delta_{sw}(y) dP_Y = \int\limits_{X\times Y} [-\lg p(x|T(y))] dP_{X\times Y} - H(X|Y). \tag{5.7}$$

Ale w myśl definicji $P_{X\times Y'}$ dla każdego $A \subset X\times Y'$ jest

$$P_{X\times Y}[T^{-1}(A)] = P_{X\times Y'}(A).$$

W szczególności dla każdego rzeczywistego α jest

$$P_{X\times Y}\big(\{(x, y): \lg p(x|T(y)) < \alpha\}\big) = P_{X\times Y'}\big(\{(x, y'): \lg p(x|y') < \alpha\}\big).$$

Wynika stąd, że

$$\int\limits_{X\times Y} [-\lg p(x|T(y))] dP_{X\times Y} = \int\limits_{X\times Y'} [-\lg p(x|y')] dP_{X\times Y'} =$$

$$= \int\limits_{Y'} \Big\{ \int\limits_{X} [-\lg p(x|y')] p(x|y') d\gamma(x) \Big\} dP_Y =$$

$$= \int\limits_{Y'} H(X|y') dP_{Y'} = H(X|Y').$$

Te związki wraz z (5.7) dowodzą wzoru (5.5). Na mocy lematu 1 dla każdego $y \in Y$ jest $\Delta_{sw}(y) \geqslant 0$, przy czym znak równości zachodzi wtedy i tylko wtedy, gdy $p(x|y) = p(x|y')$. Stąd, wobec definicji (5.4), otrzymujemy (5.6), co kończy dowód.

Wprowadzona tu średnia miara jakości stopnia wstępnego $\bar{\Delta}_{sw}$ jest ściśle związana z ilością informacji o przestrzeni X otrzymywaną na podstawie sygnałów odebranych $y \in Y$, zdefiniowaną przez Shannona [15] wzorem

$$I(X:Y) = H(X) - H(X|Y), \tag{5.8}$$

gdzie oznaczają

$$H(X) = \int \left[-\lg \frac{dP_X}{d\gamma} \right] dP_X$$

entropię a priori, a P_X rozkład a priori w przestrzeni X wynikający z rozkładu $P_{X\times Y}$.

Ilość informacji o przestrzeni X otrzymywana na podstawie sygnałów przekształconych $y' \in Y'$ wynosi

$$I(X:Y') = H(X) - H(X \mid Y'). \tag{5.9}$$

Zatem

$$I(X:Y) - I(X:Y') = H(X \mid Y') - H(X \mid Y) = \bar{\Delta}_{sw}. \tag{5.10}$$

Tak więc średnia miara jakości stopnia wstępnego równa się stracie informacji w sensie Shannona.

Podstawmy we wzorach (4.21), (4.53), (4.54), Y' zamiast Y oraz y' zamiast y. Oznaczmy przez $\bar{\Delta}_1'$ oraz $\bar{\Delta}_2'$ wielkości, które otrzymamy czyniąc podobne podstawienia we wzorach (4.22), (4.55) definiujących $\bar{\Delta}_1$ oraz $\bar{\Delta}_2$. Korzystając ze wzoru (5.5) mamy

$$r(x^* \mid Y') = \int\limits_{X\times Y'} R[x, x^*(y')]dP_{X\times Y'} \geqslant \tag{5.11}$$

$$\geqslant \frac{n}{2\pi e}\exp\left\{-\frac{2}{n}[H(X \mid Y) + \bar{\Delta}_{sw} + \bar{\Delta}_1']\right\},$$

$$\frac{D(A \mid Y')}{\gamma X} = \int\limits_{Y'} \frac{\gamma A(y')}{\gamma X} dP_{Y'} \geqslant K[H(X \mid Y) + \bar{\Delta}_{sw} + \bar{\Delta}_2', P_0], \tag{5.12}$$

$$r(A \mid Y') = \int\limits_{Y'} P_0(y')dP_{Y'} \geqslant K_1\left[H(X \mid Y) + \bar{\Delta}_{sw} + \bar{\Delta}_2', \frac{\gamma \bar{A}}{\gamma X}\right]. \tag{5.13}$$

Nierówność (5.11) jest analogiczna do nierówności Rao-Craméra (zobacz: [4], rozdz. XI). Jedna z różnic między tymi nierównościami polega na tym, że nierówności Rao-Craméra odpowiadają punktowi widzenia strony nadawczej, podczas gdy nierówność (5.8) odpowiada punktowi widzenia strony odbiorczej. Nierówności (5.12) i (5.13) stanowią analogon nierówności Rao-Craméra dla decyzji obszarowych.

Ze wzorów (5.11)-(5.13) widać, że wprowadzone przez nas wielkości $\bar{\Delta}_1$ oraz $\bar{\Delta}_2$ mają sens straty informacji wskutek decyzji. Ze wzorów tych widać też, że rozpatrywana tu średnia jakość stopnia wstępnego $\bar{\Delta}_{sw}$ ma duży wpływ na kresy dolne straty, gdy $\bar{\Delta}_{sw}$ jest duże w porównaniu z $\bar{\Delta}_1'$ lub $\bar{\Delta}_2'$.

Jak to wynika z rozdziału IV, dla pewnych rozkładów prawdopodobieństwa oraz straty, strata informacji wskutek decyzji może być równa zero. Zatem wielkość $\bar{\Delta}_{sw}$ stanowi rzeczywiście użyteczne kryterium jakości stopnia wstępnego w przypadku, gdy rozkłady prawdopodobieństwa i strata są zbliżone do takich specjalnych rozkładów i strat.

W przypadku decyzji punktowych $\Delta_1(y) = 0$, gdy strata zdefiniowana jest jako uogólniony kwadrat błędu (wzór (4.11)), a rozkłady są gaussowskie.

W przypadku decyzji obszarowych $\Delta_2(y) = 0$ dla rozkładów typu prostokątnego. Do rozkładów takiego typu zbliżają się rozkłady w zbiorze długich ciągów zbudowanych z elementarnych sygnałów, jeśli obowiązuje prawo wielkich liczb. Dla sformalizowania tego stwierdzenia wprowadzimy następujące oznaczenia:

V — sygnał elementarny będący zmienną losową mogącą przyjmować wartości v_k $(k = 1, 2, \ldots, K)$ z prawdopodobieństwami $P_k(y)$,

$x^{(M)} = (v^{(1)}, v^{(2)}, \ldots, v^{(M)})$,

v — sygnał nadawany, który jest ciągiem zbudowanym z M niezależnych zmiennych losowych typu V,

$A(y, M)$ — zbiór sygnałów $x^{(M)}$ takich, że ilość sygnałów elementarnych przyjmujących wartość v_k wynosi

$$M\big(P_k(y) + \varepsilon_k(M)\big),$$

$P_0(y, M)$ — prawdopodobieństwo, że sygnał x leży poza zbiorem $A(y, M)$ (wzór (2.7)),

γ_0 — miara w przestrzeni ciągów $x^{(M)}$ zdefiniowana jako liczebność,

$\Delta_2(y, M)$ — strata informacji zdefiniowana wzorem (4.30),

$$H(V|y) = \sum_{k=1}^{K} [-\ln P_k(y)] P_k(y) \tag{5.14}$$

— entropia zbioru sygnałów elementarnych.

Wydaje się, że można tak dobrać szybkość malenia liczb $\varepsilon_k(M)$ wraz z M, żeby były spełnione równocześnie dwie zależności

$$\lim_{M\to\infty} \frac{\Delta_2(y, M)}{M} = 0, \tag{5.15}$$

$$\lim_{M\to\infty} P_0(y, M) = 0. \tag{5.16}$$

Ponieważ $H(X^{(M)}|y) = MH(V|y)$, to korzystając ze wzorów (5.15) oraz (5.16), wzorów (4.47) oraz (4.50) i twierdzeń 7a oraz 7b otrzymalibyśmy podstawową w teorii Shannona zależność między entropią a licznością zbioru ciągów „prawdopodobnych" (patrz np. [2]):

$$\lim_{M\to\infty} \frac{\lg \gamma_0[A|y, M]}{M} = H(V|y). \tag{5.17}$$

Zaproponowana metoda prowadziłaby również do związków między $P_0(y, M)$, $\gamma_0 A(y, M)$ oraz $H(V|y)$ dla M skończonych.

Z przytoczonych tu rozważań widać, iż wielka rola ilości informacji w teorii Shannona wynika stąd, że wchodzą tam w grę rozkłady prawdopodobieństwa i decyzje, przy których strata informacji $\Delta_2(y)$ jest mała.

VI. Jednoznaczność definicji miary jakości stopnia wstępnego

Uzyskane w poprzednich rozdziałach rezultaty wynikają bezpośrednio z tego, że funkcjonał

$$\Delta(p_1, p_2) = \int\limits_X \left[-\lg \frac{p_2(x)}{p_1(x)}\right] p_1(x)\, d\gamma(x), \tag{6.1}$$

gdzie $p_1(x)$ i $p_2(x)$ są gęstościami prawdopodobieństwa, jest nieujemny i równa się zeru wtedy i tylko wtedy, gdy $p_1(x) = p_2(x)$ prawie wszędzie względem miary γ.

W rozumowaniu naszym przyjmowaliśmy punkt widzenia strony odbiorczej. Niektóre analogiczne wyniki, odpowiadające punktowi widzenia strony nadawczej, uzyskali S. Kullback i Leibler (patrz [7], [8], [9]). Łatwo zauważyć, że również i w tych pracach omawiana własność funkcjonału $\Delta(p_1, p_2)$ odgrywa zasadniczą rolę; w szczególności pokazano tam, iż wynikają z niej bezpośrednio nierówności Rao-Craméra.

Oznaczmy przez $\varphi(\xi)$ funkcję zmiennej rzeczywistej ξ. Zbadamy, czy oprócz funkcjonału $\Delta(p_1, p_2)$ istnieją funkcjonały

$$\Phi(p_1, p_2) = \int\limits_X \varphi(p_2(x)) p_1(x)\, d\gamma(x) - \int\limits_X \varphi(p_1(x)) p_1(x)\, d\gamma(x) \tag{6.2}$$

takie, że

$$\Phi(p_1, p_2) \geqslant 0,$$

a znak równości zachodzi wtedy i tylko wtedy, gdy $p_1(x) = p_2(x)$ prawie wszędzie względem γ. Dla takich funkcjonałów można by skonstruować nierówności analogiczne do nierówności dotyczących funkcjonału $\Delta(p_1, p_2)$ i można by stosować je jako miarę jakości stopnia wstępnego.

Zwróćmy uwagę, że zagadnienie sprowadza się do znalezienia funkcjonałów postaci

$$\Psi(p_1, p_2) = \int\limits_X \varphi(p_2(x)) p_1(x)\, d\gamma(x),$$

które przy każdym $p_1(x)$ osiągają infinimum ze względu na $p_2(x)$ dla $p_2(x) = p_1(x)$. Częściową odpowiedź daje twierdzenie, które, jak się wydaje, można uogólnić na znacznie szerszą klasę przestrzeni.

TWIERDZENIE 10. *Jeżeli funkcjonał*

$$\Psi(P_1, P_2) = \sum_{k=1}^{K} \varphi[P_2(x_k)]P_1(x_k), \tag{6.3}$$

gdzie dodatnie liczby $P_2(x_k)$ *oraz* $P_1(x_k)$ *spełniają zależności*

$$\sum_{k=1}^{K} P_1(x_k) = 1, \tag{6.4}$$

$$\sum_{k=1}^{K} P_2(x_k) = 1, \tag{6.5}$$

a $\varphi(\xi)$ *należy do klasy* C_1, *osiąga przy każdym dopuszczalnym zbiorze liczb* $P_1(x_k)$ *ekstremum ze względu na* $P_2(x_k)$ *dla* $\overline{P}_2(x_k) = P_1(x_k)$, *to*

$$\varphi(\xi) = \Lambda' \lg \xi + b, \tag{6.6}$$

gdzie Λ' *i* b *są stałymi.*

Dowód. Rozpatrzmy ustalony zbiór liczb dodatnich $P_1'(x_k)$, $k = 1, 2, \ldots, K$, spełniających warunek

$$\sum_{k=1}^{K} P_1'(x_k) = 1.$$

Warunkiem koniecznym, żeby funkcja

$$\Psi(P_1', P_2) = \sum_{k=1}^{K} \varphi[P_2(x_k)]P_1'(x_k) \tag{6.7}$$

osiągała ekstremum ze względu na K zmiennych $P_2(x_k)$, $k = 1, 2, \ldots, K$, spełniających dodatkowo warunek uboczny

$$\sum_{k=1}^{K} P_2(x_k) = 1, \tag{6.8}$$

jest, by wartości ekstremalizujące $\overline{P}_2(x_k)$ i mnożnik Lagrange'a Λ' spełniały $K+1$ równań

$$\left(\frac{d\varphi}{d\xi}\right)_{\xi=\overline{P}_2(x_k)} P_1'(x_k) + \Lambda' = 0, \quad k = 1, 2, \ldots, K, \tag{6.9}$$

$$\sum_{k=1}^{K} \overline{P}_2(x_k) = 1. \tag{6.10}$$

Zgodnie z założeniem ma zachodzić

$$\overline{P}_2(x_k) = P_1'(x_k), \quad k = 1, 2, \ldots, K.$$

Zatem funkcja $\varphi(\xi)$ musi spełniać K równań

$$\left(\frac{d\varphi}{d\xi}\right)_{\xi=P_1'(x_k)} P_1'(x_k)+\Lambda' = 0, \tag{6.11}$$

przy czym Λ' dla ustalonego zbioru liczb $P_1'(x_k)$ jest stałą.

Oznaczmy przez $P_1''(x_k)$ inny dopuszczalny zbiór, a przez Λ'' analogiczną stałą dla tego zbioru. Jeżeli przyjąć, że $P''(x_1) = P_1'(x_1)$, to $\Lambda' = \Lambda''$. Czyniąc dodatkowe założenie, że $K \geqslant 3$, i korzystając z założenia, że omawiana własność ekstremalna funkcji (6.7) ma być spełniona dla wszystkich dopuszczalnych zbiorów liczb $P_1''(x_k)$, widzimy, że dla $0 < \xi < 1$ funkcja $\varphi(\xi)$ musi spełniać równanie

$$\frac{d\varphi}{d\xi}\xi+\Lambda' = 0,$$

a zatem

$$\varphi(\xi) = -\Lambda'\lg\xi+b,$$

c. n. d.

Autor pragnie podziękować profesorowi S. Zubrzyckiemu za wiele cennych uwag i za pomoc przy przeredagowaniu pierwotnej wersji pracy.

Prace cytowane

[1] A. Blanc-Lapierre, R. Fortet, *Théorie des fonctions aléatoires*, Paris 1953.

[2] А. Я. Хинчин, *Понятие энтропий в теории вероятностей*, У. М. Н. VII, 3 (55) (1953), str. 3-20.

[3] А. Я. Хинчин, *Об основных теоремах теории информации*, У. М. Н. XI, 1 (67) (1956), str. 17-75.

[4] H. Cramér, *Mathematical methods of statistics*, Princeton 1946 (tłumaczenie polskie, Warszawa 1957).

[5] W. Davenport, *An experimental study of speech — wave probability distributions*, J. Acoust. Soc. Am. 24 (1952), str. 390-399.

[6] В. А. Котельников, *Теория потенциальной помехоустойчивости*, Москва 1956.

[7] S. Kullback, *An application of information theory to multivariate analysis*, Ann. Math. Stat. 23 (1952), str. 88-102.

[8] S. Kullback, *Certain inequalities in information theory and Cramér-Rao inequality*, Ann. Math. Stat. 26 (1955), str. 745-751.

[9] S. Kullback and R. A. Leibler, *On information and sufficiency*, Ann. Math. Stat. 22 (1951), str. 79-86.

[10] D. V. Lindley, *On a measure of information provided by an experiment*. Ann. Math. Stat. 27 (1956), str. 986-1005.

[11] D. Middleton, *On the theory of random noise. Phenomenological models of noise*, P. I, II, J. Appl. Phys. 22 (1951), str. 1143-1163.

[12] D. Middleton, *Optimum multialternative detection*, IRE Trans. on Inform. Theory, vol. IT1 No 2 (1955).

[13] И. П. Натансон, *Теория функции вещественной переменной*, Москва 1957.

[14] M. P. Schutzenberger, *On some measures of information used in statistics, III-rd symposium of information theory edited by C. Cherry*, London 1956.

[15] C. E. Shannon, *The mathematical theory of communication*, Bell Syst. Techn. Journ. 27 (1948), str. 379-423, 623-656.

[16] D. Slepian, *Estimation of signal parameters in the presence of noise*, IRE Trans. on Inform. Theory March 1954, str. 68-89.

[17] H. Steinhaus, *Podstawy kontroli statystycznej*, Zast. Mat. 1 (1953), str. 4-25.

[18] H. Steinhaus, *Prawdopodobieństwo, wiarogodność i możliwość*, Zast. Mat. 1 (1954), str. 149-172.

[19] H. Steinhaus i S. Zubrzycki, *O porównywaniu dwóch procesów produkcyjnych i zasadzie dualizmu*, Zast. Mat. 3 (1958), str. 229-257.

[20] A. Wald, *Statistical decision functions*, New York 1950.

[21] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series*, New York 1949.

[22] P. M. Woodward, *Probability and information theory with applications to radar*, London 1953 (tłumaczenie polskie, Warszawa 1959).

Praca wpłynęła w pierwszej wersji 28. 2. 1955 r., a w ostatecznej 1. 11. 1959 r.

J. SEIDLER (Gdańsk)

ZWIĄZKI MIĘDZY TEORIĄ INFORMACJI A TEORIĄ FUNKCJI DECYZYJNYCH W ZASTOSOWANIU DO ZAGADNIEŃ TELEKOMUNIKACJI

STRESZCZENIE

Celem pracy jest ustalenie związków między teorią informacji a teorią funkcji decyzyjnych, a w szczególności zależności między ryzykiem, entropią i miarą ilości informacji.

W pracy rozpatruje się dwa podstawowe typy decyzji: decyzje punktowe i decyzje obszarowe. Dla ich opisu wprowadzamy następujące oznaczenia:

X — przestrzeń sygnałów nadawanych,
Y — przestrzeń sygnałów odbieranych,
$P_{X\times Y}$ — prawdopodobieństwo w przestrzeni $X\times Y$,
$P_{X|y}$ — prawdopodobieństwo warunkowe w X przy ustalonym $y\in Y$,
P_Y — prawdopodobieństwo brzegowe w Y.

Decyzja punktowa jest sygnałem należącym do przestrzeni X, który jest przyporządkowany sygnałowi odebranemu $y\in Y$. Oznaczamy ją przez $x^*(y)\in X$. Dana jest funkcja straty $R(x', x'')$, $x'\in X$, $x''\in X$. Ryzyko warunkowe $r(x^*|y)$ zdefiniowane jest wzorem (2.1), a ryzyko $r(x^*|Y)$ wzorem (2.2).

Decyzja obszarowa jest zbiorem zawartym w przestrzeni X, który przyporządkowuje się sygnałom odbieranym. Oznacza się ją przez $A(y)$. Dana jest miara γ w X. Przy ustalonym $y\in Y$ jakość decyzji opisana jest przez dwie liczby: $P_0(y)$, $\gamma[A(y)]$ (wzory (2.7), (2.8)). Jakość średnia zdefiniowana jest wzorami (2.6), (2.9).

Następnie wprowadza się pojęcia stopnia wstępnego, który dokonuje operacji nad sygnałem odebranym, zanim zostanie powzięta decyzja. Człon ten opisany jest przez operację T odwzorowującą przestrzeń Y w przestrzeń Y'.

Związki między jakością decyzji a entropią $H(X|y)$ (wzór (4.2)) otrzymuje się definiując pomocnicze rozkłady prawdopodobieństwa zależne od decyzji.

Dla decyzji punktowych rozkład taki zdefiniowany jest wzorami (4.6) oraz (4.7), a związki dają wzory (4.8) oraz (4.9). Występująca w tych wzorach stała k_2 jest dowolna. Rozpatruje się różne metody wyboru k_2. W szczególności, gdy przestrzeń X jest euklidesowa a strata zdefiniowana wzorem (4.11), to k_2 można ustalić z warunku (4.16). Prowadzi to do zależności (4.20). Uogólnienie jej dla wartości średnich dają wzory (4.21) oraz (4.22).

Dla decyzji obszarowych pomocniczy rozkład definiuje się wzorem (4.27). Zależność między $P_0(y)$, $\gamma[A(y)]$ i $H(X|y)$, analogiczna do (4.20), dana jest wzorami (4.39), (4.47) oraz (4.48). Znak równości w dwu ostatnich wzorach występuje, jeśli

$$\frac{\gamma A}{\gamma X}+P_0(y)\leqslant 1.$$

Funkcje $K(h,P)$ oraz $K_1(h,z)$ są zdefiniowane jako funkcje odwrotne rosnącej gałęzi funkcji $M(P,z)$ danej wzorem (4.41). Zależność dla wartości średnich zdefiniowanych wzorem (4.51) dają wzory (4.53) oraz (4.54).

Następnie dowodzi się (wzór (5.5)), że średnia jakość stopnia wstępnego $\bar{\Delta}_{sw}$, zdefiniowana wzorami (5.3) oraz (5.4) jest równa różnicy entropii średnich (wzory (4.3), (5.2)) przy y lub $y' = Ty$ ustalonych. Korzystając z tego otrzymuje się oszacowanie jakości decyzji punktowych i obszarowych podejmowanych na podstawie sygnałów na wyjściu stopnia wstępnego — wzory (5.11) do (5.13). Są to nowe związki typu nierówności Rao-Craméra. Ze wzorów tych widać, że wprowadzone wielkości $\Delta_1(y)$ oraz $\Delta_2(y)$ mają sens straty informacji wskutek decyzji.

Rozpatruje się zakres zastosowań otrzymanych nierówności.

W ostatniej części pracy zwrócono uwagę na to, że otrzymane rezultaty wynikają bezpośrednio z tego, iż funkcjonał $\Delta(p_1,p_2)$ dany wzorem (6.1) jest nieujemny i równa się zeru wtedy i tylko wtedy, gdy $p_1(x) = p_2(x)$. Udowodniono następujące twierdzenie o jednoznaczności:

Jeżeli dla każdego zbioru liczb $P_1(x_k)$, $k = 1, 2, \ldots, K$, spełniających warunek (6.4) *funkcjonał*

$$\Psi(P_1,P_2) = \sum_{k=1}^{K} \varphi[P_2(x_k)]P_1(x_k)$$

osiąga minimum ze względu na $P_2(x_k)$ spełniające warunek (6.5) *dla $\bar{P}_2(x_k) = P_1(x_k)$, to*

$$\varphi(\xi) = A'\lg\xi+b,$$

gdzie A' i b są stałymi.

Й. СЕЙДЛЕР (Гданьск)

СВЯЗИ МЕЖДУ ТЕОРИЕЙ ИНФОРМАЦИЙ И ТЕОРИЕЙ РЕШАЮЩИХ ФУНКЦИЙ В ПРИЛОЖЕНИИ К ЗАДАЧАМ ТЕЛЕКОММУНИКАЦИИ

РЕЗЮМЕ

Целью работы является установление связей между теорией информаций и теорией решающих функций, а в особенности установление зависимости между: риском, энтропией и мерой количества информации.

В работе рассматривается два основных типа решений: решения точечные и решения областные. Для их описания вводятся следующие обозначения:

X — пространство переданных сигналов,

Y — пространство принятых сигналов,

$P_{X\times Y}$ — вероятность в пространстве $X\times Y$,

$P_{X|y}$ — условная вероятность в X при заданном $y\in Y$,

P_Y — граничная вероятность в Y.

Точечное решение — это сигнал принадлежащий пространству X — который ставится в соответствие принятому $y\in Y$. Обозначим её $x^*(y)\in X$. Задана функция потери $R(x', x'')$, $x'\in X$, $x''\in X$. Условные риски $r(x^*|y)$ определены формулой (2.1), риск $r(x^*|Y)$ формулой (2.2). Областное решение — это множество $A(y)\subset X$ содержимое в пространстве X, которое ставится в соответствие принятому $y\in Y$. Дана мера γ в X. При заданном $y\in Y$ качество решения определяется двумя числами: $P_0(y), \gamma[A(y)]$ (формулы (2.7), (2.8)). Среднее качество определено формулами (2.6), (2.9).

В дальнейшем вводится понятие предварительного звена которое преобразовывает принятый сигнал до совершения решения. Член этот описывается оператором T отображающим пространство Y в пространство Y'.

Связи между качеством решений и энтропией $H(X|y)$ (формула (4.2)) получаем, определяя вспомогательные распределения вероятностей которые зависят от решений.

Для точечных решений это разложение определяется формулами (4.6) и (4.7), а связи дают формулы (4.8) и (4.9). Постоянная k_2 в этих формулах является произвольной. Рассматриваются различные методы выбора k_2. В частности, когда X евклидово пространство, а потера определена формулой (4.11), то k_2 можно получить из условия (4.16). Приводит это к зависимости (4.20). Обобщение её для средних приводит к формулам (4.21) и (4.22).

Для областных решений вспомогательное распределение определяется формулой (4.27). Зависимость между $P_0(y), \gamma[A(y)]$ и $H(X|y)$, аналогична (4.20), определяется формулами (4.39), (4.47) и (4.48). Знак равенства в двух последних формулах имеет место, если

$$\frac{\gamma A}{\gamma X}+P_0(y)\leqslant 1.$$

Функции $K(h,P)$ и $K_1(h,z)$ определены как обратные функции возрастающей ветви функции $M(P,z)$, заданной формулой 4.41 при P и z, соответственно, зафиксированных.

Зависимости для средних значений, определённых формулой (4.51), дают формулы (4.53) и (4.54).

Потом доказывается (формула (5.5)), что среднее качество предварительного звена $\overline{\Delta}_{sw}$ определено формулами (5.3) и (5.4) и равно разнице средних энтропии (формулы (4.3), (5.2)) при y, $y' = Ty$ — зафиксированных. Используя это, получаем оценку качества точечных и областных решений, которые получается на базе выходов предварительного звена — формулы (5.11) до (5.13). Это новые связи типа неравенств Рао-Крамера. Из этих формул видно, что введённые величины $\Delta_1(y)$ и $\Delta_2(y)$ имеют смысл потери информации вследствии решения.

Рассматривается область применений полученных неравенств.

В последней части работы обращается внимание на то, что полученные результаты вытекают непосредственно из того, что функционал $\Delta(p_1, p_2)$, заданный формулой (6.1), неотрицателен и равен нулю тогда и только тогда, когда

$p_1(x) = p_2(x)$. Доказывается следующее предложение, касающееся однозначности:

Если для каждого множества чисел $P_1(x_k)$, $k = 1, 2, \ldots, K$, подчиняющегося условию (6.4) функционал

$$\Psi(P_1, P_2) = \sum_{k=1}^{K} \varphi[P_2(x_k)]P_1(x_k)$$

достигает условного минимума по переменным $P_2(X_k)$ при выполнению требования (6.5), то

$$\varphi(\xi) = \Lambda' \lg \xi + b,$$

где Λ' и b постоянные.

J. SEIDLER (Gdańsk)

RELATIONS BETWEEN THE THEORY OF INFORMATION AND THE THEORY OF DECISION FUNCTIONS AS APPLIED TO PROBLEMS IN TELECOMMUNICATION

SUMMARY

The aim of the paper is to establish the relations between the theory of information and the theory of decision functions, and in particular the relations between risk, entropy and amount of information.

Two fundamental types of decisions are considered: point decisions and domain decisions. The following notation is introduced for their description:

X — space of transmitted signals,
Y — space of received signals,
$P_{X\times Y}$ — probability in space $X\times Y$,
$P_{X|y}$ — conditional probability in X for $y \in Y$ fixed,
P_Y — boundary probability in Y.

A point decision is a signal belonging to space X which is associated with the received signal $y \in Y$. We denote it by $x^*(y) \in X$. We are given the loss function $R(x', x'')$, $x' \in X$, $x'' \in X$. The conditional risk $r(x^*|y)$ is defined by formula (2.1) and the risk $r(x^*|Y)$ by formula (2.2).

A domain decision is a set contained in space X which is associated with received signals. It is denoted by $A(y)$. We are given a measure γ in X. For a fixed $y \in Y$ the quality of the decision is described by two numbers: $P_0(y)$, $\gamma[A(y)]$ (formulas (2.7), (2.8)). The mean quality is defined by formulas (2.6), (2.9).

The author then introduces notions of the preliminary stage which performs operations on the received signal before the decision is made. That stage is described by operation T mapping space Y into space Y'.

The relations between the quality of the decision and entropy $H(X|y)$ (formula (4.2)) are obtained by defining auxiliary probability distributions dependent on the decision.

For point decisions such a distribution is defined by formulas (4.6) and (4.7) and the relations are given by formulas (4.8) and (4.9). The constant k_2 occurring in those formulas is arbitrary. Various methods of choosing k_2 are considered. In partic-

ular when space X is Euclidean and the loss is defined by formula (4.11), then k_2 can be determined by condition (4.16). This leads to relation (4.20). Its generalization for mean values is given by formulas (4.21) and (4.22).

For domain decisions the auxiliary distribution is defined by formula (4.27). The relation between $P_0(y)$, $\gamma[A(y)]$ and $H(X|y)$, analogous to (4.20), is given by formulas (4.39), (4.47) and (4.48). The sign of equality in the last two formulas holds if

$$\frac{\gamma A}{\gamma X}+P_0(y)\leqslant 1.$$

The functions $K(h,P)$ and $K_1(h,z)$ are defined as the inverse functions of the increasing branch of the function $M(P,z)$ given by formula (4.41). The relation for the mean values defined by formula (4.51), is given by formulas (4.53) and (4.54).

It is then proved (formula (5.5)) that the mean quality of the preliminary stage $\overline{\Delta}_{SW}$ defined by formulas (5.3) and (5.4) is equal to the difference of mean entropies (formulas (4.3), (5.2)) for fixed y or $y' = Ty$. Using this, we obtain an estimation of the quality of the point decisions and domain decisions made on the ground of signals at the output of the preliminary stage—formulas (5.11) to (5.13). They are new relations of the type of Rao-Cramér inequalities. Those formulas show that the quantities $\Delta_1(y)$ and $\Delta_2(y)$ which have been introduced have the meaning of information loss due to the decision.

The range of application of the inequalities obtained is then considered.

In the last section of the paper the author points to the fact that the results obtained follow immediately from the fact, that functional $\Delta(p_1, p_2)$ given by formula (6.1) is non-negative and equals zero if and only if $p_1(x) = p_2(x)$. The following unicity theorem is then proved: *if for every set of numbers* $P_1(x_k)$; $k = 1, 2, \ldots, K$ *satisfying condition* (6.4) *the functional*

$$\Psi(P_1, P_2) = \sum_{k=1}^{K} \varphi[P_2(x_k)]P_1(x_k)$$

has its minimum with respect to $P_2(x_k)$ *satisfying condition* (6.5), *for* $\overline{P}_2(x_k) = P_1(x_k)$, *then*

$$\varphi(\xi) = A'\lg\xi + b,$$

where A' *and* b *are constants.*