PRZEGLĄD STATYSTYCZNY
R. LVII – ZESZYT 4 – 2010

EWA WĘDROWSKA

WYKORZYSTANIE ENTROPII SHANNONA I JEJ UOGÓLNIEŃ DO BADANIA ROZKŁADU PRAWDOPODOBIEŃSTWA ZMIENNEJ LOSOWEJ DYSKRETNEJ

1. WPROWADZENIE

Pojęcie entropii, które wywodzi się z termodynamiki znalazło zastosowanie w wielu dziedzinach nauki, począwszy od teorii informacji poprzez mechanikę statystyczną, teorię chaosu do sieci neuronowych, lingwistyki matematycznej czy też taksonomii. Możliwość wykorzystania entropii do określania stopnia nieokreśloności badanych stanów czy też rozkładów zmiennych losowych dało przyczynek do określania różnych postaci entropii. Do teorii informacji pojęcie entropii wprowadził Shannon (1948), po czym w drugiej połowie ubiegłego wieku pojawiło się wiele uogólnień probabilistycznej miary shannonowskiej entropii.

W artykule przedstawione zostało pojęcie entropii Shannona wraz z jej wybranymi uogólnieniami: entropią Rényiego (1965) oraz entropią Havrda-Charvát-Daróczy-Tsallisa (1970). Entropia Rényiego jest wynikiem uogólnienia średniej Kołmogorowa-Nagumo, entropia Tsallisa jest natomiast pewną funkcją entropii Rényiego.

Własności entropii rozkładu zmiennej losowej typu skokowego wskazują na liczne związki miar koncentracji z entropią Shannona ([2], [3], [7], [9], [10], [16], [17], [23], [24]). Dlatego też w artykule zaproponowano wykorzystanie entropii Rényiego, będącej uogólnieniem entropii Shannona, do oceny stopnia nierównomierności dyskretnego rozkładu prawdopodobieństwa.

Celem teoretycznym artykułu jest przedstawienie pojęć i własności entropii zmiennej losowej dyskretnej umożliwiające zastosowanie jej w badaniu własności dyskretnych rozkładów zmiennych losowych. Celem aplikacyjnym natomiast jest wskazanie na możliwość wykorzystania miar entropii do określania stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa.

2. ENTROPIA SHANNONA

Niech X będzie zmienną losową dyskretną o rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \dots, p(x_n)\}$ z prawdopodobieństwami $\{p(x_i)\}$ spełniającymi warunek $0 \leq p(x_i) \leq 1$, $\sum_{i=1}^{n} p(x_i) = 1$. W teorii informacji zdefiniowanie entropii poprzedzone zostało sformułowaniem następujących postulatów, jakie powinna spełniać poszukiwana funkcja entropii $H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n))$ ([4],[12]):

P1. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być ciągła względem wszystkich argumentów $p(x_i)$ ($i = 1, 2, \dots, n$), co oznacza, że niewielkim zmianom prawdopodobieństw powinna odpowiadać niewielka zmiana entropii.

P2. Jeżeli wszystkie n zdarzeń zmiennej losowej X są jednakowo prawdopodobne, to funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna rosnąć monotonicznie wraz ze wzrostem n, gdzie $p(x_1) = p(x_2) = \dots = p(x_n) = \frac{1}{n}$.

P3. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być symetryczna, co oznacza, że wartość funkcji entropii jest niezmiennikiem permutacji prawdopodobieństw $p(x_1), p(x_2), \dots, p(x_n)$.

P4. Funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ powinna być koherentna, co oznacza w tym przypadku, że jeżeli realizacja zdarzeń odbywa się w dwóch kolejno następujących po sobie etapach, to entropia początkowa powinna być sumą ważoną entropii poszczególnych etapów.

Istnieje dokładnie jedna, z dokładnością do stałej k [4], funkcja $H_S(p(x_1), p(x_2), \dots, p(x_n))$ n-zmiennych spełniająca powyższe warunki i jest ona określona wzorem:

$$H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n)) = k\sum_{i=1}^{n} p(x_i)\log\frac{1}{p(x_i)}. \tag{1}$$

Dowód tego twierdzenia przedstawiono między innymi w pracach ([4], [12]).

Stała k we wzorze (1) stanowi wybór jednostki entropii. Jeżeli $k = \frac{1}{\log 2}$ jednostką entropii jest bit, a funkcja zapisana wzorem (1) przyjmuje postać:

$$H_S(X) = H_S(p(x_1), p(x_2), \dots, p(x_n)) = \sum_{i=1}^{n} p(x_i)\log_2\frac{1}{p(x_i)}. \tag{2}$$

Entropia Shannona H_S zmiennej losowej X, której wartości występują z prawdopodobieństwami $\{p(x_1), p(x_2), \dots, p(x_n)\}$ jest określona wzorem (2) i jest traktowana jako miara niepewności związanej z rozkładem dyskretnym o prawdopodobieństwach $\{p(x_1), p(x_2), \dots, p(x_n)\}$ [25].

Określenie shannonowskiej entropii pozwala ustalić szereg jej istotnych własności, przedstawionych między innymi w pracy [20]:

1) Entropia Shannona jest wielkością nieujemną, $\forall p(x_i) \in [0,1]\ H_S(X) \geq 0$.

2) Entropia H_S przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \dots, n$).

3) Entropia H_S przyjmuje wartość największą równą $H_S(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \dots, n$.

4) Entropia H_S jest wklęsła.

5) Entropia Shannona spełnia własność addytywności dla dyskretnych zmiennych losowych niezależnych.

Własność 5 dotyczy pary zmiennych losowych niezależnych. Jeśli zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_j)\}$ dla $i = 1, 2, \ldots, n$ oraz $j = 1, 2, \ldots, m$ są niezależne, to rozkład prawdopodobieństwa zmiennej losowej dwuwymiarowej (X, Y) zadany jest iloczynem $\{p(x_i)q(y_j)\}$, gdzie $\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)q(y_j) = 1$. Wtedy entropia Shannona spełnia własność addytywności, a zatem prawdziwa jest równość $H_S(X,Y) = H_S(X) + H_S(Y)$. Prawdziwość podanej równości wynika z własności funkcji logarytmicznych [20].

3. ENTROPIA RÉNYIEGO

Węgierski matematyk Alfred Rényi wśród swoich licznych osiągnięć z dziedziny teorii prawdopodobieństwa zaproponował uogólnienie pojęcia entropii Shannona. Wyprowadzenie uogólnionej postaci entropii wymaga zastosowania średniej Kołmogorowa-Nagumo (KN) [15].

Niech $\varphi(x_i)$ będzie funkcją ciągłą i ściśle monotoniczną. Średnia Kołmogorowa-Nagumo związana z funkcją $\varphi(x_i)$ zmiennej losowej dyskretnej X przyjmującej wartości $\{x_1, x_2, \ldots, x_n\}$, definiowana jest następująco [13]:

$$\langle X \rangle_\varphi = \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\varphi(x_i)\right). \tag{3}$$

Średnia arytmetyczna jest średnią Kołmogorowa-Nagumo związaną z funkcją liniową $\varphi(x_i) = x_i$, uwzględniając bowiem w formule (3) zamiast $\varphi(x_i)$ wartość x_i uzyskana zostanie znana postać wartości oczekiwanej.

Średnie Kołmogorowa-Nagumo są niezmiennicze ze względu na transformacje afiniczne $\varphi(x_i) \to \gamma(x_i) = a\varphi(x_i) + b$, gdzie a, b są liczbami. Stąd funkcją odwrotną do $\varphi(x_i)$ jest $\gamma^{-1}(x_i) = \varphi^{-1}\left(\frac{x_i - b}{a}\right)$, a średnia KN związana w funkcją $\varphi(x_i)$ zmiennej losowej X, czyli $\langle X\rangle_\varphi$ równa jest średniej KN $\langle X\rangle_\gamma$ związanej z funkcją $\gamma(x_i)$.

Rozważmy dwie niezależne zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_i)\}$ dla $i = 1, 2, \ldots, n$. Poszukiwanie funkcji $\varphi(x_i)$ spełniającej własność addytywności średnich $\langle X + Y\rangle_\varphi = \langle X\rangle_\varphi + \langle Y\rangle_\varphi$ daje, oprócz funkcji typu $ax_i + b$, jedynie funkcję postaci $\varphi(x_i) = a\mu^{\nu x_i} + b$, gdzie a, b są liczbami rzeczywistymi. Warunek addytywności spełniają jedynie średnie KN związane z dwiema funkcjami, pierwsza z nich to funkcja liniowa $\varphi(x_i) = x_i$, druga to funkcja wykładnicza $\varphi(x_i) = a\mu^{\nu x_i} + b$ [15]. Rozpatrzmy dalej funkcję $\varphi(x_i) = \mu^{\nu x_i}\,(a = 1, b = 0)$, przyjmując $\mu = 2$, $\nu = 1 - \alpha$. Wtedy funkcja $\varphi(x_i)$ przyjmuje postać $2^{(1-\alpha)x_i}$, a funkcja odwrotna $\varphi^{-1}(x_i) = \frac{1}{1-\alpha}\log_2 x_i$.

Średnia ilość informacji w sensie Kołmogorowa-Nagumo związana z funkcją $\varphi(x_i) = 2^{(1-\alpha)x_i}$ i rozkładem prawdopodobieństwa $\{p(x_i)\}$ zmiennej losowej X jest entropią Rényiego. W teorii informacji ilość informacji dostarczaną przez wiadomość o przyjęciu przez dyskretną zmienną losową X wartości x_i wyraża się miarą $I(x_i) = -\log_2 p(x_i)$ [25]. Stąd uwzględniając przyjęte powyżej założenia, jawną postać entropii Rényiego otrzymuje się drogą poniższych przekształceń ([5], [13]):

$$\begin{aligned} H_{R\alpha}(X) &= \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\varphi(I(x_i))\right) \\ &= \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\varphi(-\log_2 p(x_i))\right). \end{aligned} \tag{4}$$

Uwzględniwszy, że przyjęto powyżej funkcję $\varphi(x_i)$ postaci $2^{(1-\alpha)x_i}$ wraz z funkcją do niej odwrotną $\varphi^{-1}(x_i) = \frac{1}{1-\alpha}\log_2 x_i$, wyrażenie (4) można wyrazić następująco:

$$\begin{aligned} &\frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)\cdot 2^{-(1-\alpha)\log_2 p(x_i)}\right) \\ &= \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)\cdot 2^{\log_2 p(x_i)^{(\alpha-1)}}\right) \\ &= \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)p(x_i)^{(\alpha-1)}\right) \\ &= \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)^{\alpha}\right). \end{aligned} \tag{5}$$

Postać $\frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)^{\alpha}\right)$ powyższej formuły jest podawana najczęściej w literaturze jako definicja entropii Rényiego, a jej związek ze średnią Kołmogorowa-Nagumo jest mniej znany. Zwracają na to uwagę Czachor i Naudts w pracy [5] oraz Lavenda [13]. Entropia $H_{R\alpha}(X)$ określona wzorem (5) nazywana jest najczęściej entropią Rényiego stopnia α lub α-entropią Rényiego dla $\alpha > 0$, $\alpha \neq 1$.

Entropia Rényiego stopnia α dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_i)\}$ $\left(0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1\right)$ spełnia następujące własności:

I. Jest wielkością nieujemną $\forall\alpha \in (0,1) \cup (1,+\infty)$, $H_{R\alpha}(X) \geq 0$.

II. Jest wklęsła $\forall\,\alpha \in (0,1)$ [8].

III. Jest wklęsła lub wypukła dla $\alpha > 1$ [8].

IV. Przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \ldots, n$).

V. Przyjmuje wartość największą równą $H_{R\alpha}(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \ldots, n$.

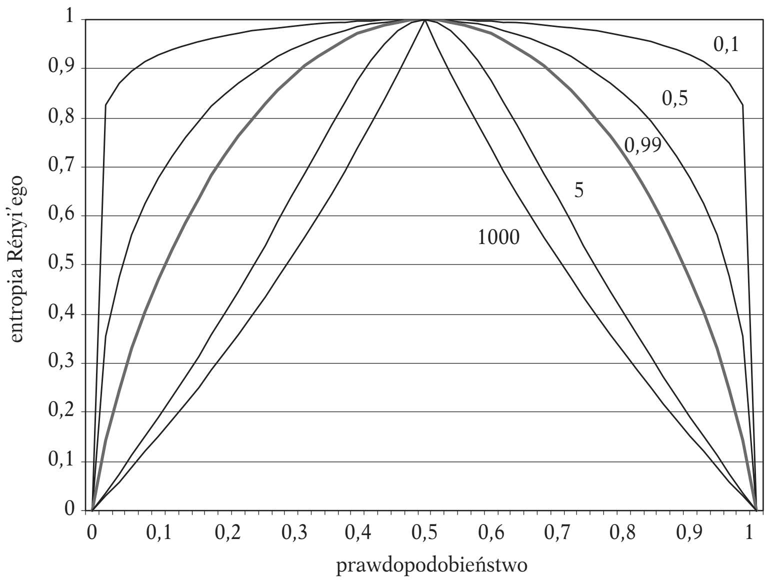
VI. Entropia Rényiego spełnia własność addytywności dla zmiennych losowych niezależnych.

VII. Entropie Shannona H_S oraz Rényiego $H_{R\alpha}$ spełniają relację $H_{R\alpha_1} \geq H_S \geq H_{R\alpha_2}$ jeżeli $0 < \alpha_1 < 1$ oraz $\alpha_2 > 1$ [19].

VIII. Entropia Shannona H_S jest granicą entropii Rényiego $H_{R\alpha}$ dla $\alpha \to 1$,

$$\lim_{\alpha \to 1} \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{n} p(x_i)^{\alpha} \right) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i).$$

Z własności V wynika, że dla zmiennej losowej o rozkładzie równomiernym $(p(x_1) = p(x_2) = \ldots = p(x_n))$ wartość największa entropii Rényiego równa jest entropii Shannona dla tego rozkładu. Zatem w nierówności $H_{R\alpha_1} \geq H_S \geq H_{R\alpha_2}$ podanej we własności VII zachodzi równość dla rozkładu równomiernego. Wartości entropii Rényiego rosną monotonicznie wraz ze wzrostem n, co jest spełnieniem jednego z postulatów stawianych funkcji entropii przez Shannona.



Rysunek 1. Entropia Rényiego zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1-p)\}$

Źródło: opracowanie własne.

Własność VI dotyczy pary zmiennych losowych niezależnych. Jeśli zmienne losowe dyskretne X i Y o rozkładach prawdopodobieństwa odpowiednio $\{p(x_i)\}$ oraz $\{q(y_j)\}$ dla $i = 1, 2, \ldots, n$ oraz $j = 1, 2, \ldots, m$ są niezależne, prawdziwa jest równość $H_{R\alpha}(X,Y) = H_{R\alpha}(X) + H_{R\alpha}(Y)$. Prawdziwość podanej równości wynika, podobnie jak w przypadku równości dla entropii Shannona, z własności funkcji logarytmicznych.

Entropia Rényiego dla zmiennej losowej o dwumianowym rozkładzie prawdopodobieństwa $\{p,(1-p)\}$ przyjmuje wartość największą równą jeden dla $p = 1 - p$, czyli dla rozkładu o prawdopodobieństwach {0,5; 0,5}. Na rysunku 1 przedstawiono wykres entropii Rényiego dla wybranych wartości stopnia α rozkładu o prawdopodobieństwach $\{p,(1-p)\}$. Dla $\alpha = 0{,}99$ wykres entropii Rényiego jest przybliżony do wykresu entropii Shannona.

4. ENTROPIA TSALLISA

Entropię określaną mianem „entropii typu α" zaproponowali niezależnie w swoich pracach z roku 1970 Havrda i Charvát oraz Daróczy, a następnie Tsallis w roku 1988 przy okazji tzw. termodynamiki nieekstensywnej [5]. Stąd często w literaturze pojawia się określenie entropia Havrda-Charvát-Daróczy-Tsallis (HCDT) lub w pracach z zakresu fizyki, entropia Tsallisa. Entropia HCDT została zaproponowana jako pierwsza nielogarytmiczna postać entropii.

Do wyprowadzenia ogólnej postaci entropii HCDT wykorzystane zostaną ponownie własności średniej Kolmogorowa-Nagumo dotyczące niezmienniczości ze względu na transformacje afiniczne $\varphi(x_i) \to \gamma(x_i) = a\varphi(x_i) + b$. Z rodziny funkcji typu $\varphi(x_i) = a\mu^{vx_i} + b$ wybrana zostanie funkcja $\varphi(x_i) = \frac{1}{1-\alpha} 2^{(1-\alpha)x_i} - \frac{1}{1-\alpha} = \frac{2^{(1-\alpha)x_i} - 1}{1-\alpha}$, gdzie $\alpha \in (0,1) \cup (1, +\infty)$.

Postać entropii HCDT uzyskana jest drogą modyfikacji średniej Kołmogorowa-Nagumo związanej z funkcją $\varphi(x_i) = \frac{1}{1-\alpha} 2^{(1-\alpha)x_i} - \frac{1}{1-\alpha}$:

$$\begin{aligned} H_{R\alpha} &= \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\varphi(-\log_2 p(x_i))\right) \\ &= \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\frac{2^{(1-\alpha)(-\log_2 p(x_i))} - 1}{1-\alpha}\right) \\ &= \varphi^{-1}\left(\sum_{i=1}^{n} p(x_i)\frac{2^{\log_2\left(\frac{1}{p(x_i)}\right)^{1-\alpha}} - 1}{1-\alpha}\right). \end{aligned} \tag{6}$$

Wykorzystując własności funkcji wykładniczych i logarytmicznych wyrażenie $2^{\log_2\left(\frac{1}{p(x_i)}\right)^{1-\alpha}}$ można wyrazić w prostszej postaci $\left(\frac{1}{p(x_i)}\right)^{1-\alpha}$. Stąd wyrażenie (6) przyjmuje postać:

$$\begin{aligned} &\varphi^{-1}\left(\frac{\sum_{i=1}^{n} p(x_i)\left(\frac{1}{p(x_i)}\right)^{1-\alpha} - 1}{1-\alpha}\right) \\ &= \varphi^{-1}\left(\frac{\sum_{i=1}^{n} p(x_i)^{\alpha} - 1}{1-\alpha}\right). \end{aligned} \tag{7}$$

Wyrażenie uzyskane w drodze przekształceń prowadzących do otrzymania entropii $H_{R\alpha}(X)$ postaci:

$$H_{T\alpha}(X) = \frac{\sum_{i=1}^{n} p(x_i)^{\alpha} - 1}{1-\alpha} \tag{8}$$

nosi miano entropii Tsallisa lub entropii Havrda-Charvát-Daróczy-Tsallisa.

Entropia Tsallisa typu α dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_i)\}\left(0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1\right)$ spełnia następujące własności:

a) Jest wielkością nieujemną $\forall \alpha \in (0,1) \cup (1, +\infty)$, $H_{T\alpha}(X) \ge 0$.

b) Jest wklęsła $\forall \alpha \in (0,1) \cup (1, +\infty)$

c) Przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i ($i = 1, 2, \ldots, n$).

d) Przyjmuje wartość największą, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1, 2, \ldots, n$.

e) Dla $\forall \alpha_1, \alpha_2 \in (0,1) \cup (1, +\infty)$ jeżeli $\alpha_1 < \alpha_2$ zachodzi relacja $H_{T\alpha_1}(X) > H_{T\alpha_2}(X)$.

f) Iloczyn entropii Shannona H_S i $\ln 2$ jest granicą entropii Tsallisa $H_{T\alpha}$ dla $\alpha \to 1$;

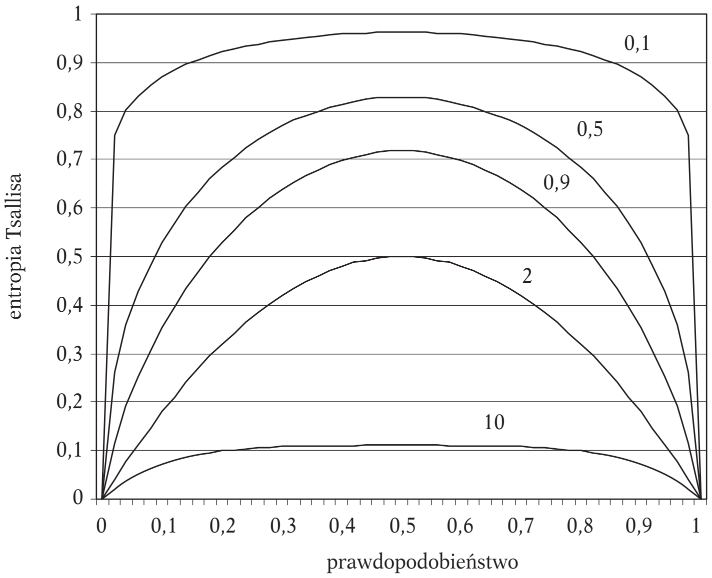
$$\lim_{\alpha \to 1} \frac{\sum_{i=1}^{n} p(x_i)^{\alpha} - 1}{1-\alpha} = -\ln 2 \sum_{i=1}^{n} p(x_i) \log p(x_i) = \ln 2 H_S(X).$$

g) Spełnia tzw. własność pseudo-addytywności (subaddytywności) dla zmiennych losowych niezależnych [13]:

$$H_{T\alpha}(X,Y) = H_{T\alpha}(X) + H_{T\alpha}(Y) + (1-\alpha) H_{T\alpha}(X) H_{T\alpha}(Y).$$

Z własności pseudo-addytywności wynika, że dla $\alpha \in (0,1)$ entropia $H_{T\alpha}(X, Y)$ jest większa od sumy $H_{T\alpha}(X) + H_{T\alpha}(Y)$, natomiast dla $\alpha \in (1, +\infty)$ zachodzi relacja $H_{T\alpha}(X, Y) < H_{T\alpha}(X) + H_{T\alpha}(Y)$. Własność pseudo-addytywności w fizyce nazywana jest często nieekstensywnością. Własność pseudo-addytywności odróżnia entropię Tsallisa od entropii Shannona i Rényiego, które spełniały własność addytywności dla dyskretnych zmiennych losowych niezależnych X i Y.

Entropia HCDT przyjmuje wartość największą dla zmiennej losowej o rozkładzie $\{p, (1-p)\}$ dla $p = 0{,}5$. Wartość ta jest funkcją stopnia α oraz jest mniejsza dla każdego $\alpha \in (0,1) \cup (1, +\infty)$ od wartości maksymalnej entropii Shannona i Rényiego dla tego rozkładu $\{0{,}5; 0{,}5\}$. Na rysunku 2 przedstawiono wykres entropii Tsallisa dla wybranych wartości stopnia α rozkładu o prawdopodobieństwach $\{p, (1-p)\}$.



Rysunek 2. Entropia Tsallisa zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1-p)\}$

Źródło: opracowanie własne.

Entropia Tsallisa osiąga wartość największą dla równomiernego rozkładu prawdopodobieństwa $(p(x_1) = p(x_2) = \ldots = p(x_n))$. Wartość największa entropii HCDT jest funkcją nie tylko wartości n, jak to było w przypadku entropii Shannona oraz Rényiego, ale i stopnia α. Wraz ze wzrostem n rośnie wartość entropii $H_{T\alpha}$ dla danego stopnia α. Wartości największe omawianych w artykule entropii zamieszczono w tabeli 1.

Tabela 1

Wartości entropii rozkładu równomiernego zmiennej losowej dyskretnej

n	Entropia					
	Shannona	Rényiego $\alpha \in (0,1) \cup (1,+\infty)$	Tsallisa			
			$\alpha = 0,1$	$\alpha = 0,5$	$\alpha = 2$	$\alpha = 10$
2	1,00000	1,00000	0,96230	0,82843	0,50000	0,11089
4	2,00000	2,00000	2,75800	2,00000	0,75000	0,11111
5	2,32193	2,32193	3,61856	2,47214	0,80000	0,11111
10	3,32193	3,32193	7,71476	4,32456	0,90000	0,11111
50	5,64386	5,64386	36,45796	12,14214	0,98000	0,11111
100	6,64386	6,64386	68,99526	18,00000	0,99000	0,11111
200	7,64386	7,64386	129,71200	26,28427	0,99500	0,11111

Źródło: opracowanie własne.

5. MIARY KONCENTRACJI BAZUJĄCE NA ENTROPII

Związek pomiędzy miarami koncentracji, czy też miarami dekoncentracji, a entropią wydaje się być oczywisty. Entropia jest bowiem miarą nieokreśloności i niejednorodności rozkładu dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$ z prawdopodobieństwami $\{p(x_i)\}$ spełniającymi warunek $0 \le p(x_i) \le 1$, $\sum_{i=1}^{n} p(x_i) = 1$. Jednym z przykładów związku miar koncentracji z entropią jest indeks Giniego-Simpsona. Indeks ten jest miarą równomierności dyskretnego rozkładu prawdopodobieństwa i określony jest formułą [17]:

$$GS = 1 - \sum_{i=1}^{n} p(x_i)^2. \tag{9}$$

Miernik ten jest szczególnym przypadkiem entropii Tsalisa $H_{T\alpha}(X)$. Jeśli bowiem w formule (8) uwzględniony zostanie stopień α równy dwa, entropia HCDT przyjmie następującą postać:

$$H_{T2}(X) = \frac{1}{\alpha - 1}\left[1 - \sum_{i=1}^{n} p(x_i)^{\alpha}\right] = \left[1 - \sum_{i=1}^{n} p(x_i)^2\right] = GS.$$

Miernik ten nazywany jest także w literaturze entropią kwadratową Vajdy (zob. [16]). Warto zauważyć, że suma kwadratów prawdopodobieństw $\sum_{i=1}^{n} p(x_i)^2$ występująca w formule indeksu Giniego-Simpsona jest znanym w literaturze miernikiem koncentracji, zwanym indeksem Herfindahla-Hirschmana (*HHI*).

Konstrukcja indeksu Herfindahla-Hirschmana wskazuje, że wybór wykładnika potęgi prawdopodobieństw $p(x_i)$ implikuje relatywny wpływ na pomiar koncentracji wartości x_i, zachodzących odpowiednio z małymi bądź znacznymi prawdopodobieństwami. Dobór wyższych wykładników potęgi sprawia, że większe znaczenie przypisywane jest tym wartościom zmiennej losowej X, które występują z dużym prawdopodobieństwem. Stąd w literaturze pojawiły się modyfikacje i uogólnienia indeksu *HHI*. Hannah i Kay zaproponowali w 1977 roku swobodny dobór wykładnika, podając następującą formułę miernika koncentracji [2]:

$$HKI_{(\alpha)} = \begin{cases} \left[\sum_{i=1}^{n} p(x_i)^{\alpha}\right]^{\frac{1}{\alpha-1}} & \text{dla } \alpha > 0,\ \alpha \neq 1 \\ \exp\left[\sum_{i=1}^{n} p(x_i)\ln p(x_i)\right] & \text{dla } \alpha = 1 \end{cases}. \tag{10}$$

Indeks Hannah-Kaya dla $\alpha = 1$ ($HKI_{(1)}$) jest granicą indeksu $HKI_{(\alpha)}$ dla $\alpha \to 1$, zaś w przypadku przyjęcia w formule (10) $\alpha = 2$ indeks Hannah-Kaya staje się indeksem Herfindahla-Hirschmana. Związki indeksu Hannah-Kaya z miarami entropii określają Bajo i Salas (2002) w pracy [2].

Odrębną klasę mierników koncentracji stanowią miary wykorzystujące definicję entropii Shannona. Entropia Shannona $H_S(X)$ zdefiniowana formułą (2) jest odwrotnością stopnia koncentracji prawdopodobieństw $p(x_i)$ w rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$. Entropia $H_S(X)$ równa jest bowiem zeru, gdy jedna z wartości x_i ($i = 1, 2, \ldots, n$) osiągana jest z prawdopodobieństwem równym jedności. W przypadku gdy wszystkie n wartości dyskretnej zmiennej losowej X zachodzą z jednakowymi prawdopodobieństwami, entropia $H_S(X)$ osiąga wartość największą równą $H_S^{\max}(X) = \log_2 n$. W wielu pracach entropia Shannona wykorzystywana jest bezpośrednio do oceny stopnia niejednorodności rozkładu badanej cechy [3] bądź do analizy stopnia dywersyfikacji działalności przedsiębiorstw na poziomie korporacyjnym (zob. [9], [10]) czy też na poziomie regionalnym (zob. [1], [7]).

Znacznie bardziej popularne w naukach ekonomicznych są mierniki będące funkcją wystandaryzowanej postaci entropii. Mierniki te stanowią miarę koncentracji, tak jak miara

$$h = 1 - \frac{H_S(X)}{H_S^{\max}(X)} \tag{11}$$

przedstawiona przez Roeske-Słomka (1994) w pracy [21] bądź też miarę dekoncentracji postaci:

$$e = \frac{H_S(X)}{H_S^{\max}(X)} \tag{12}$$

proponowaną miedzy innymi w pracach ([17], [23], [24]).

Z konstrukcji miary (11) wynika, że w pomiarze stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa największy wpływ na otrzymane wyniki mają wartości zmiennej losowej X realizowane z dużym prawdopodobieństwem. Powoduje to sytuację, że zdecydowana większości obszaru zmienności miary h przeznaczona jest dla rozkładów bliskich pełnej koncentracji.

W niniejszym artykule skonstruowana zostanie miara koncentracji dyskretnego rozkładu prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$ bazująca na entropii Rényiego $H_{R\alpha}(X)$. Proponowana miara jest jednocześnie uogólnieniem miary (11) i zdefiniowana zostanie następującą formułą:

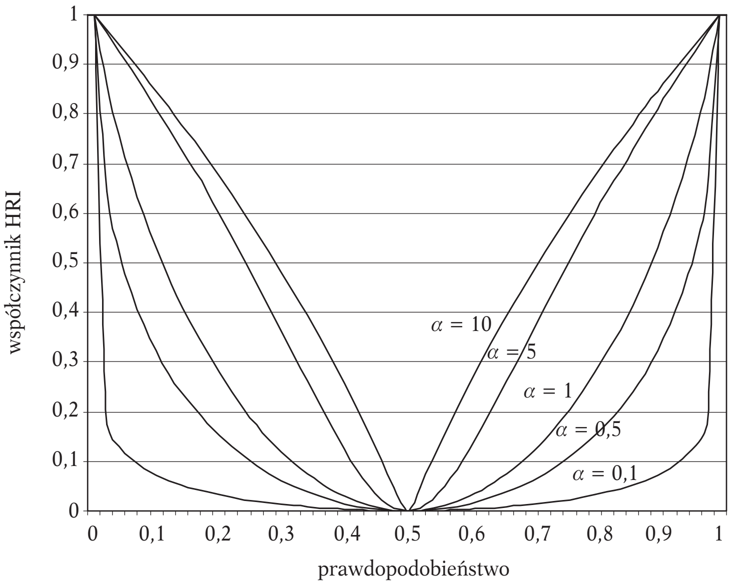
$$HRI_\alpha(X) = \begin{cases} 1 - \dfrac{\log_2\left(\sum\limits_{i=1}^{n} p(x_i)^\alpha\right)}{(1-\alpha)\log_2 n}, \text{ dla } \alpha > 0 \ i \ \alpha \neq 1 \\ 1 - \dfrac{\sum\limits_{i=1}^{n} p(x_i)\log_2 \dfrac{1}{p(x_i)}}{\log_2 n}, \text{ dla } \alpha = 1 \end{cases}. \tag{13}$$

Widać, że proponowaną miarę można następująco wyrazić jako funkcję entropii Rényiego:

$$HRI_\alpha(X) = \begin{cases} 1 - \dfrac{H_{R\alpha}(X)}{H_{R\alpha}^{\max}(X)}, \text{ dla } \alpha > 0 \ i \ \alpha \neq 1 \\ 1 - \dfrac{H_S(X)}{H_S^{\max}(X)}, \text{ dla } \alpha = 1 \end{cases}. \tag{14}$$

Własności entropii Rényiego determinują własności miary HRI_α. Miara koncentracji HRI_α jest symetryczna, to oznacza, że jest niezależna od permutacji prawdopodobieństw $p(x_i)$ $(i = 1, 2, \ldots, n)$ w dyskretnym rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$. Spełnia także kryterium unormowania w przedziale [0,1] i jest jednoznacznie określona dla krańcowych sytuacji. W przypadku równomiernego rozkładu prawdopodobieństw $p(x_i)$ miernik HRI_α przyjmuje wartość zero, natomiast w sytuacji pełnej koncentracji osiąga wartość jeden. Uogólnienie miary (11) bazującej na entropii Shannona daje możliwość swobodnego wyboru potęgi prawdopodobieństw $p(x_i)$. Dobór stopnia α entropii Rényiego $H_{R\alpha}(X)$ umożliwia relatywny wpływ na ocenę stopnia koncentracji wartości x_i zmiennej losowej X występujących z odpowiednio niskimi bądź wysokimi prawdopodobieństwami. Dla $\alpha \in (0,1]$ zdecydowanie większa część przedziału zmienności miary HRI_α przeznaczona jest dla rozkładów prawdopodobieństw bliskich pełnej koncentracji. Występuje znikoma reakcja miernika przy odbieganiu rozkładu $\{p(x_1), p(x_2), \ldots, p(x_n)\}$ od rozkładu równomiernego o prawdopodobieństwach $p(x_1) = p(x_2) = \ldots = p(x_n) = \frac{1}{n}$. Dla rozkładów bliskich pełnej koncentracji miernik HRI_α przyjmuje wartości z dolnego krańca przedziału [0,1]. Zatem omawiany miernik dla stopnia α z przedziału (0,1] charakteryzuje się niewielką awersją do niejednorodnego rozkładu prawdopodobieństwa.

Proponowana miara HRI_α jest wypukła dla każdego stopnia $\alpha \in (0,1]$. Na rysunku 3 przedstawiono wartości miary HRI_α dla rozkładu o prawdopodobieństwach $\{p, (1-p)\}$, przyjmując wybrane wartości stopnia α.



Rysunek 3. Wartości miary HRI_α dla zmiennej losowej o rozkładzie prawdopodobieństwa $\{p, (1-p)\}$

Źródło: opracowanie własne.

Miara koncentracji HRI_α dla α większego od jedności jest wypukła bądź wklęsła. Charakteryzuje się większą wrażliwością (bezwzględną i względną) na zmiany w rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$. Wraz ze wzrostem α zwiększa się awersja miernika do nierównomiernego rozkładu prawdopodobieństw, z jakim zachodzą wartości x_i dyskretnej zmiennej losowej X.

6. PODSUMOWANIE

Możliwość wykorzystania entropii do określania stopnia nieokreśloności badanych stanów czy też rozkładów prawdopodobieństw przyczyniło się do określania różnych postaci entropii. Zastosowanie pojęcia entropii w teorii informacji przez Shannona dało początek licznym uogólnieniom miary entropii. W artykule wskazano jedynie wybrane uogólnienia entropii Shannona, mianowicie entropię Rényiego oraz Tsallisa. Entropia Rényiego stopnia α doczekała się kolejnych uogólnień, czego przykładem jest chociażby entropia stopni (α, β) zaproponowana przez Kapura (1967). Inne probabilistyczne miary entropii proponują miedzy innymi Lavenda [14], Sharma i Taneja [22], Chakrabarti [4], Garbaczewski [6], Nanda i Paul [18], Kumar i in. [11].

Entropia Shannona jest granicą entropii Rényiego oraz Tsallisa dla stopnia α dążącego do jedności. Entropie Shannona i Rényiego mają postać funkcji logarytmicznych, natomiast entropia Tsallisa jest pierwszą proponowaną w literaturze nie-logarytmiczną entropią. Warto podkreślić, że entropie Shannona i Rényiego są addytywne dla dyskretnych zmiennych losowych X i Y, natomiast entropia Tsallisa nie spełnia własności addytywności, a jedynie tzw. własność pseudo-addytywności.

Wartości entropii Shannona, Rényiego oraz Tsallisa zależą jedynie od prawdopodobieństw, jakie towarzyszą realizacji wartości $\{x_1, x_2,\ldots, x_n\}$ dyskretnej zmiennej losowej X, a nie od tych wartości. Wartość największą każda ze wskazanych entropii osiąga dla rozkładu równomiernego o prawdopodobieństwach $p(x_1) = p(x_2) = \ldots = p(x_n) = \frac{1}{n}$, przy czym monotoniczny wzrost entropii wraz ze wzrostem n. Entropia jest odwrotnością stopnia koncentracji prawdopodobieństw $p(x_i)$ w dyskretnym rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2), \ldots, p(x_n)\}$. Stąd znajomość entropii może być przydatna do rozpoznania stopnia koncentracji rozkładu prawdopodobieństwa zmiennej losowej dyskretnej. W artykule zaproponowano miarę umożliwiającą ocenę stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa będącą funkcją entropii Rényiego. Istotną zaletą proponowanej miary jest swobodny dobór potęgi prawdopodobieństw $p(x_i)$. Dobór stopnia α umożliwia bowiem relatywny wpływ na ocenę stopnia koncentracji rozkładu wartości x_i zmiennej losowej X występujących z odpowiednio niskimi bądź wysokimi prawdopodobieństwami.

Uniwersytet Warmińsko-Mazurski w Olsztynie

LITERATURA

[1] Attaran M., [1985], *Industrial diversity and economic performance in U.S. areas*, The Annals of Regional Science 20, s. 44-54.

[2] Bajo O., Salas R., [2002], *Inequality foundations of concentration measures: An application to the Hannah-Kay indices*, Spanish Economic Review 4, s. 311-316.

[3] Campos D., Isaza J.F., [2009], *A geometrical index for measuring species diversity*, Ecological Indicators 9, s. 651-658.

[4] Chakrabarti C.G., Chakrabarty I., [2005], *Shannon entropy: axiomatic characterization and application*, International Journal of Mathematics and Mathematical Sciences, Vol. 17, s. 2847-2854.

[5] Czachor M., Naudts J., [2002], *Thermostatistics based on Kolmogorov-Nagumo averages: unifying framework for extensive and nonextensive generalizations*, Physics Letters A 298, s. 369-374.
[6] Garbaczewski P., [2006], *Differential Entropy and Dynamics of Uncertainty*, „Journal of Statistical Physics", Vol. 123, No. 2, s. 315-355.
[7] Hackbart M.W., Anderson D.A., [1975], *On measuring economic diversification*, Land Economics 51, s. 374-378.
[8] Hibbard L.S., [2004], *Region segmentation using information divergence measures*, Medical Image Analysis 8, s. 233-244.
[9] Jacquemin A.P., Berry C.H., [1979], *Entropy measure of diversification and corporate growth*, „Journal of Industrial Economics" 27 (4), s. 359-369.
[10] Koen F., [2007], *Entropy and information theory*, [w:] Hanusch H., Pyka A. (red.), The Elgar Companion to Neo-Schumpeterian Economics.
[11] Kumar R., Kumar S., Kumar A., [2010], *A New Measure of Probabilistic Entropy and its Properties*, Applied Mathematical Sciences, Vol. 4, s. 1387-1394.
[12] Kuriata E., [2001], *Teoria informacji i kodowania*, Oficyna Wydawnicza Politechniki Zielonogórskiej.
[13] Lavenda B.H., [2004], *Mean Entropies*, Open Sys. Information Dyn. 12, s. 289-302.
[14] Lavenda B.H., [2006], *Geometric Entropies of Mixing (EOM)*, Open Sys. Information Dyn. 13, s. 91-101.
[15] Masi M., [2005], *A step beyond Tsallis and Rényi entropies*, Physics Letters A 338, s. 217-224.
[16] Mayoral M.M., [1998], *Renyi's entropy as an index of diversity in simple-stage cluster sampling*, „Journal of Information Sciences" 105, s. 101-114.
[17] McBratney A., Minasny B., [2007], *On measuring pedodiversity*, Geoderma 141, s. 149-154.
[18] Nanda A.K., Paul P., [2006], *Some results on generalized residual entropy*, Information Sciences 176 (1), s. 27-47.
[19] Principe J.C., Xu D., Fisher III J.W., [????], *Information-theoretic learning*, in: Haykin S. (ed.), Unsupervised Adaptive Filtering. Wiley, New York, s. 265-319.
[20] Przybyszewski R., Wędrowska E., [2005], *Algorytmiczna teoria entropii*, „Przegląd Statystyczny" nr 2, tom 52 Warszawa, s. 85-102.
[21] Roeske-Słomka I., [1994], *Klasyczne a entropijne miary koncentracji i asymetrii w badaniu rozmieszczenia ludności*, Zesz. Nauk. AE Pozn. Seria I, nr 219, s. 48-59.
[22] Sharma B.D., Taneja I.J., [1975], Entropy of type (α, β) and other generalized measures in information theory, Metrika 22, s. 205-215.
[23] Troutt M. D., Acar W., [2005], *A Lorenz-Pareto measures of pure diversification*, „European Journal of Operational Research" 167, s. 543-549.
[24] Wędrowska E., [2003], *Datalogiczna miara ilości informacji strukturalnej jako instrument zarządzania zasobami informacji statystycznej*, [w:] „Pozyskiwanie wiedzy i zarządzanie wiedzą", Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 975 Wrocław, s. 447-459.
[25] Woodward P.M., [1959], *Wstęp do teorii informacji*, PWN, Warszawa.

Praca wpłynęła do redakcji kwiecień 2010 r.

WYKORZYSTANIE ENTROPII SHANNONA I JEJ UOGÓLNIEŃ DO BADANIA ROZKŁADU PRAWDOPODOBIEŃSTWA ZMIENNEJ LOSOWEJ DYSKRETNEJ

Streszczenie

O entropii można mówić wszędzie tam, gdzie istnieje potrzeba rozpoznania i zmierzenia nieokreśloności badanych rozkładów zmiennych losowych. Potrzeba zmierzenia stopnia koncentracji występuje w badaniu wielu systemów, procesów czy zjawisk, w szczególności w badaniu zjawisk społeczno-ekonomicznych. Właśnie dlatego w badaniu tychże zjawisk coraz częściej wykorzystuje się miary zdefiniowane na gruncie teorii informacji. W artykule zaprezentowano pojęcia i własności entropii zmiennej losowej dyskretnej. Wykorzystano entropię Shannona wraz z jej uogólnieniami (entropią Rényiego i Tsallisa) do badania własności rozkładów prawdopodobieństw dyskretnych zmiennych losowych. Wskazano ponadto na liczne związki miar koncentracji dyskretnego rozkładu prawdopodobieństwa z entropią. W artykule zaproponowano miarę umożliwiającą ocenę stopnia koncentracji dyskretnego rozkładu prawdopodobieństwa będącą funkcją entropii Rényiego.

Słowa kluczowe: entropia Shannona, entropia Rényiego, entropia Havrda-Charvát-Daróczy-Tsallisa, miary koncentarcji

APPLICATION OF SHANNON'S ENTROPY AND ITS GENERALIZATIONS FOR STUDIES ON THE DISCRETE RANDOM VARIABLE PROBABILITY DISTRIBUTION

Summary

Everywhere where there is need for identification and measurement of indeterminacy of the distributions studied we can talk about entropy. The need to measure the degree of diversity occurs in studies on numerous systems, processes and phenomena, in particular in studies on socioeconomic phenomena. That is why in studies on those phenomena the measures or models defined on grounds of the information theory are used increasingly often. The paper presents categorization of notions and characteristics of the entropy of a discrete random variable. In addition to Shannon's entropy, the Rényi's and Tsallis entropies were applied for studies on the properties of distributions in case of probabilities of the random variables. The notion of entropy stemming from thermodynamics found application in many fields of sciences. Shannon's entropy was defined on the grounds of the information theory, the Rényi's entropy is the result of generalization of the Kolmogorov-Nagumo average while the Tsallis entropy is a certain function of Rényi's entropy. The present paper unifies these approaches by presenting one general model of concentration measure that applies Rényi's entropy.

Key words: Shannon's entropy, Rényi's entropy, Tsallis entropy, concentration measure