Ewa Wędrowska

Miary entropii w statystyce i teorii informacji

Ekonomiczne Problemy Usług nr 67, 133-141

2011

ZESZYTY NAUKOWE UNIWERSYTETU SZCZECIŃSKIEGO

NR 650 EKONOMICZNE PROBLEMY USŁUG NR 67 2011

EWA WĘDROWSKA
Uniwersytet Warmińsko-Mazurski

MIARY ENTROPII W STATYSTYCE I TEORII INFORMACJI

Wprowadzenie

Rozwój telekomunikacji w początkach XX wieku zapoczątkował badania nad informacją, jej istotą, ilością oraz jakością. Przełomowa była praca C.E. Shannona, uważanego za twórcę matematycznej teorii informacji. Teoria informacji i teoria komunikacji, w ich matematycznym sformułowaniu podanym przez Shannona, traktują informację od strony ilościowej. Dotyczą one pomiaru ilości informacji, jaką charakteryzuje się każdy kanał informacyjny ze względu na stopień prawdopodobieństwa pojawienia się jednego z sygnałów. W przypadku najprostszego ze zbiorów, jakim jest zbiór dyskretny, modelem statystycznym źródła wiadomości dyskretnych jest zmienna losowa dyskretna.

Bardziej reprezentatywna dla procesów informacyjnych, jako procesów redukujących niepewność, jest oczekiwana ilość informacji rozumiana jako entropia źródła. Pojęcie entropii wykorzystywane było w badaniach systemów fizycznych, a zdefiniowane zostało przy okazji drugiej zasady termodynamiki. Zastosowanie termodynamiki w teorii informacji wprowadziło pojęcie entropii do systemów komunikowania się. Miara entropii zdefiniowana przez C.E. Shannona na gruncie teorii informacji znalazła w kolejnych latach zastosowanie w wielu dziedzinach nauki, między innymi w statystyce i informatyce. Obecnie teoria informacji nadal dotyczy głównie systemów łączności, pojawiają się jednak zastosowania pojęcia entropii w analizie zachowania się różnorodnych systemów, w tym systemów ekonomiczno-społecznych, a kolejne lata przyniosły liczne uogólnienia shannonowskiej miary entropii.

Celem artykułu jest przedstawienie miar entropii zmiennej losowej dyskretnej i ich własności umożliwiających zastosowanie entropii w badaniu dyskretnych rozkładów zmiennych losowych. Scharakteryzowano entropię Shannona wraz z jej uogólnieniami: entropią Rényiego oraz entropią Havrda–Charvát–Daróczy–Tsallisa, a także trygonometryczną postać entropii.

1. Entropia Shannona

W teorii informacji zdefiniowanie miary entropii zmiennej losowej *X* o rozkładzie dyskretnym $\{p(x_1), p(x_2),..., p(x_n)\}$ poprzedzone zostało sformułowaniem warunków stawianych funkcji entropii $H_S(X) = H_S(p(x_1), p(x_2),..., p(x_n))$. System uwarunkowań zaproponowany przez Shannona zakładał, że entropia powinna spełniać następujące warunki[1]:

1. Funkcja $H_S(X)$ powinna być ciągła względem wszystkich prawdopodobieństw $p(x_i)$ ($i = 1,2,..., n$), co oznacza, że niewielkim zmianom prawdopodobieństw powinna odpowiadać niewielka zmiana entropii.
2. Jeżeli wszystkie n zdarzeń zmiennej losowej X są jednakowo prawdopodobne $\left(p(x_1) = p(x_2) = ... = p(x_n) = \frac{1}{n} \right)$, to funkcja $H_S(X)$ powinna rosnąć monotonicznie wraz ze wzrostem n.
3. Funkcja $H_S(X)$ powinna być symetryczna, co oznacza, że wartość entropii jest niezmiennikiem permutacji prawdopodobieństw $p(x_1), p(x_2),..., p(x_n)$.
4. Funkcja $H_S(X)$ powinna być koherentna, co oznacza, 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[2], z dokładnością do stałej *k*, funkcja $H_S(X)$ *n*-zmiennych spełniająca powyższe warunki i jest ona określona wzorem:

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

gdzie $r > 1$, a prawdopodobieństwa $p(x_i)$ spełniają warunki unormowania oraz sumy jednostkowej: $0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1$. Stała *k* we wzorze (1) decyduje

[1] C.G. Chakrabarti, I. Chakrabarty: *Shannon entropy: axiomatic characterization and application*, „International Journal of Mathematics and Mathematical Sciences" 2005, vol. 17.

[2] E. Kuriata: *Teoria informacji i kodowania*, Oficyna Wydawnicza Politechniki Zielonogórskiej, Zielona Góra 2001.

o jednostce entropii. Jeżeli $k = \dfrac{1}{\log_r 2}$, jednostką entropii jest *bit*, a funkcja zapisana za pomocą wzoru (1) przyjmuje postać:

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

Jeśli z kolei $k = \dfrac{1}{\log_r e}$, jednostką entropii jest *nat* (*natural unit*), a formuła entropii staje się następująca:

$$H_S(X) = H_S(p(x_1), p(x_2),..., p(x_n)) = \sum_{i=1}^{n} p(x_i)\ln \frac{1}{p(x_i)}. \qquad (3)$$

Entropia $H_S(X)$ jest miarą niepewności związanej z rozkładem prawdopodobieństw $\{p(x_1), p(x_2),..., p(x_n)\}$, z jakimi zachodzą wartości $\{x_1, x_2,..., x_n\}$ dyskretnej zmiennej losowej *X*.

Probabilistyczna miara entropii $H_S(X)$ opisana formułą (2) posiada następujące własności:

- Entropia Shannona przyjmuje wartości nieujemne: $H_S(X) \geq 0$.
- Entropia Shannona przyjmuje wartość zero, gdy jedna z wartości $\{x_1, x_2,..., x_n\}$ dyskretnej zmiennej losowej X zachodzi z prawdopodobieństwem równym jedności, pozostałe zaś z prawdopodobieństwami równymi zeru.
- Entropia Shannona przyjmuje wartość największą równą $H_S(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa są sobie równe $\left(p(x_1) = p(x_2) = ... = p(x_n) = \dfrac{1}{n}\right)$.
- Entropia Shannona jest wklęsła.
- Entropia Shannona spełnia własność addytywności dla pary dyskretnych zmiennych losowych niezależnych X oraz Y: $H_S(X,Y) = H_S(X) + H_S(Y)$.

2. Entropia Rényiego

Zasłużony w dziedzinach kombinatoryki, teorii grafów, teorii liczb oraz teorii prawdopodobieństwa, węgierski matematyk A. Rényi wśród swoich licznych osiągnięć zaproponował uogólnienie miary entropii Shannona. Ostateczna postać formuły entropii Rényiego stopnia α ($\alpha > 0, \alpha \neq 1$) zmiennej losowej *X* o dyskretnym rozkładzie prawdopodobieństwa $\{p(x_1), p(x_2),..., p(x_n)\}$ jest następująca:

$$H_{R\alpha}(X) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p(x_i)^{\alpha}\right). \tag{4}$$

Entropia Rényiego $H_{R\alpha}(X)$ stopnia α ($\alpha > 0, \alpha \neq 1$) dyskretnej zmiennej losowej X o rozkładzie prawdopodobieństwa $\{p(x_i)\}$ ($0 \leq p(x_i) \leq 1, \sum_{i=1}^{n} p(x_i) = 1$) spełnia następujące własności:

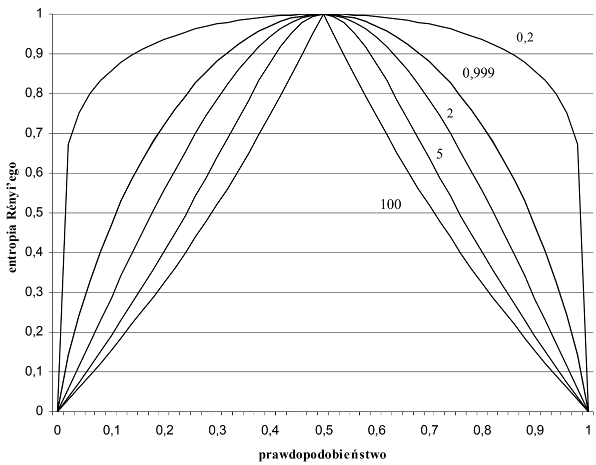
- Entropia Rényiego przyjmuje wartości nieujemne: $H_{R\alpha}(X) \geq 0.$
- Entropia Rényiego jest wklęsła dla każdego $\alpha \in (0,1)$ oraz wklęsła lub wypukła dla $\alpha > 1$[3].
- Entropia Rényiego przyjmuje wartość zero, gdy jedna z wartości $\{x_1, x_2, ..., x_n\}$ dyskretnej zmiennej losowej X zachodzi z prawdopodobieństwem równym jedności, pozostałe zaś z prawdopodobieństwami równymi zeru.
- Entropia Rényiego przyjmuje wartość największą równą $H_{R\alpha}(X) = \log_2 n$, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1,2,...,n.$
- Entropia Rényiego spełnia własność addytywności dla pary dyskretnych zmiennych losowych niezależnych X oraz Y: $H_{R\alpha}(X,Y) = H_{R\alpha}(X) + H_{R\alpha}(Y).$
- Entropia Shannona $H_S(X)$ jest granicą entropii Rényi'ego $H_{R\alpha}(X)$ dla $\alpha \to 1$:[4]

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

Własności entropii zmiennej losowej o dwupunktowym rozkładzie prawdopodobieństwa $\{p,(1-p)\}$ ilustruje rysunek 1, na którym przedstawiono wykres entropii Rényi'ego dla wybranych wartości stopnia α. Wykres entropii Rényiego jest przybliżony do wykresu entropii Shannona dla $\alpha = 0{,}999$. Dla każdej wartości α ($\alpha > 0, \alpha \neq 1$) entropia $H_{R\alpha}(X)$ osiąga wartość największą równą jedności, w przypadku gdy prawdopodobieństwa rozkładu są sobie równe, czyli dla $p = 1 - p = 0{,}5$.

[3] L.S. Hibbard: *Region segmentation using information divergence measures*, „Medical Image Analysis" 2004, no. 8, 233–244.

[4] E. Wędrowska: *Wykorzystanie entropii Shannona i jej uogólnień do badania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej*, „Przegląd Statystyczny" 2010, nr 4, s. 39–53.



Rys. 1. Entropia Rényiego zmiennej losowej o rozkładzie prawdopodobieństwa $\{p,(1-p)\}$ dla wartości α = 0,2; 0,999; 2; 5; 100

Źródło: opracowanie własne.

3. Entropia Havrda–Charvát–Daróczy–Tsallisa

Entropia określana mianem „entropii typu α" zaproponowana przez Tsallisa w 1988 roku na gruncie fizyki nieeksensywnej odpowiada dokładnie α-entropii zdefiniowanej wcześniej w teorii informacji przez Havrdę i Charváta w 1967 roku oraz Daróczego w 1970 roku. Obecnie w literaturze pojawiają się określenia „entropia Havrda–Charvát–Daróczy–Tsallisa" (HCDT) lub, w pracach z zakresu fizyki, entropia Tsallisa.

Entropia HCDT dyskretnej zmiennej losowej *X* o rozkładzie prawdopodobieństwa $\{p(x_i)\}$ $(0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1)$ określona jest następującą formułą dla $\alpha > 0, \alpha \neq 1$[5]:

$$H_{HCDT\alpha}(X) = \frac{\sum_{i=1}^{n} p(x_i)^{\alpha} - 1}{1-\alpha}, \tag{5}$$

Entropia HCDT typu α (dla $\alpha > 0, \alpha \neq 1$) dyskretnej zmiennej losowej *X* posiada poniższe własności:

– Entropia HCDT przyjmuje wartości nieujemne: $H_{HCDT\alpha}(X) \ge 0$.

[5] M. Masi: *A step beyond Tsallis and Rényi entropies*, „Physics Letters A" 2005, no. 338, s. 217–224.

- Entropia HCDT jest wklęsła dla każdego $\alpha > 0, \alpha \neq 1$.
- Entropia HCDT przyjmuje wartość największą, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1,2,...,n$.
- Entropia HCDT przyjmuje wartość zero, gdy jedna z wartości $\{x_1, x_2,..., x_n\}$ dyskretnej zmiennej losowej X zachodzi z prawdopodobieństwem równym jedności, pozostałe zaś z prawdopodobieństwami równymi zeru[6].
- Entropia $H_{HCDT\alpha}(X)$ dla stopnia $\alpha \to 1$ dąży do entropii Shannona[7]:
$$\lim_{\alpha \to 1} \frac{1}{1-\alpha} \left(\sum_{i=1}^{n} \omega_i^{\alpha} - 1 \right) = \ln 2 H_S(S^n).$$
- Entropia HCDT spełnia własność pseudoaddytywności (subaddytywności) dla zmiennych losowych niezależnych[8]:
$$H_{HCDT\alpha}(X,Y) = H_{HCDT\alpha}(X) + H_{HCDT\alpha}(Y) + (1-\alpha) H_{HCDT\alpha}(X) H_{HCDT\alpha}(Y).$$

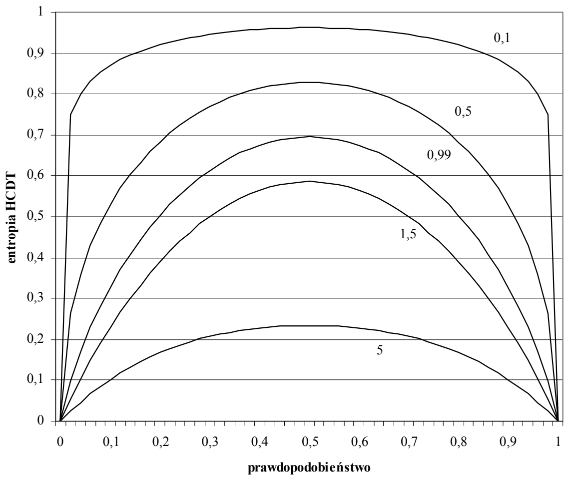
Entropia HCDT, w odróżnieniu do entropii Shannona oraz entropii Rényiego, nie spełnia dla pary zmiennych niezależnych własności addytywności, lecz jedynie tzw. własność pseudoaddytywności. Podobnie jak w przypadku entropii Shannona oraz Rényiego, entropia HCDT osiąga wartość największą dla równomiernego rozkładu prawdopodobieństwa. Jednak wartość największa entropii HCDT jest funkcją nie tylko wartości *n*, jak to było w przypadku entropii Shannona oraz Rényiego, ale i stopnia α. Wartość entropii $H_{HCDT\alpha}(X)$ rośnie wraz ze wzrostem wartości *n* dla danego stopnia α. Z kolei w przypadku zmiennej losowej dyskretnej przyjmującej *n* wartości $\{x_1, x_2,..., x_n\}$ wraz ze wzrostem stopnia α maleją wartości entropii $H_{HCDT\alpha}(X)$. Dla zmiennej losowej o dwupunktowym rozkładzie prawdopodobieństwa $\{p,(1-p)\}$ własność tę ilustruje rysunek 2.

---

[6] E. Wędrowska: *Wykorzystanie entropii Shannona…*, *op. cit.*, s. 39–53.

[7] P.K. Sahoo, G. Arora: *Image thresholding using two-dimensional Tsallis-Havrada-Charvát entropy*, „Pattern Recognition Letters” 2006, no. 27, s. 520–528.

[8] B.H. Lavenda: *Mean Entropies*, „Open Sys. Information Dyn.” 2004, no. 12, s. 289–302.



Rys. 2. Entropia HCDT zmiennej losowej o rozkładzie prawdopodobieństwa $\{p,(1-p)\}$ dla wartości $\alpha = 0{,}1;\ 0{,}5;\ 0{,}99;\ 1{,}5;\ 5$

Źródło: opracowanie własne.

4. Trygonometryczna postać entropii

Odmienna od poprzednich koncepcji jest trygonometryczna postać miary entropii, jaką zaproponował Lavenda[9], wskazując w swojej pracy na związki *Entropies of Mixing* (EOM) z funkcją logarytmiczną i wielomianami oraz własnościami trygonometrycznymi wielokątów. EOM zdefiniowana została następująco:

$$H_n^{EOM}(X) = \frac{1}{n}\sum_{i=1}^{n} \sin \pi\, p(x_i). \tag{6}$$

Własności entropii $H_n^{EOM}(X)$ danej wzorem (6) są następujące:

- Entropia EOM jest wielkością nieujemną: $H_n^{EOM}(X) \geq 0$.
- Entropia EOM przyjmuje wartość 0, gdy $p(x_i) = 1$ dla pewnego i $(i = 1,2,...,n)$.
- Entropia EOM przyjmuje wartość największą, gdy wszystkie prawdopodobieństwa $p(x_i)$ są sobie równe dla $i = 1,2,...,n$.
- Entropia EOM spełnia własność symetrii: $H_n^{EOM}(p(x_1), p(x_2),..., p(x_n)) = H_n^{EOM}(p(x_{(1)}), p(x_{(2)}),..., p(x_{(n)}))$.
- Entropia EOM jest wklęsła.

[9] B.H. Lavenda: *Geometric Entropies of Mixing (EOM)*, „Open Sys. Information Dyn.” 2006, no. 13, s. 91–101.

Wartość maksymalna entropii EOM wynosi $H_{\max}^{EOM} = \sin\frac{\pi}{n}$, zatem $H_n^{EOM}(X) \in \left[0, \sin\frac{\pi}{n}\right]$. Wraz ze wzrostem liczby n wartość H_n^{EOM} dąży do zera. Niespełnione jest więc założenie stawiane entropii Shannona i jej uogólnieniom, mówiące, że stopień nieokreśloności rozkładu, którego miarą jest entropia, rośnie wraz ze wzrostem liczby wartości będących realizacjami zmiennej losowej. Entropia EOM nie spełnia też własności addytywności dla pary dyskretnych zmiennych losowych niezależnych X oraz Y.

Podsumowanie

Wartości entropii EOM, tak jak entropii Shannona, Rényiego czy HCDT, zależą jedynie od prawdopodobieństw, jakie towarzyszą realizacji konkretnych wartości zmiennej X, a nie od tych wartości. Jednak opisane entropie przejawiają różne własności, co wynika z różnych postaci tych miar. Entropie Shannona i Rényiego mają postać funkcji logarytmicznych, entropia HCDT stanowiła pierwszą proponowaną w literaturze formułę nielogarytmiczną, natomiast entropia EOM przyjmuje postać funkcji trygonometrycznej.

Literatura

1. Chakrabarti C.G.: Chakrabarty I.: *Shannon entropy: axiomatic characterization and application*, „International Journal of Mathematics and Mathematical Sciences" 2005, vol. 17.
2. Hibbard L.S.: *Region segmentation using information divergence measures*, „Medical Image Analysis" 2004, no. 8.
3. Kuriata E.: *Teoria informacji i kodowania*, Oficyna Wydawnicza Politechniki Zielonogórskiej, Zielona Góra 2001.
4. Lavenda B.H.: *Mean Entropies*, „Open Sys. Information Dyn." 2004, no. 12.
5. Lavenda B.H.: *Geometric Entropies of Mixing (EOM)*, „Open Sys. Information Dyn." 2006, no. 13.
6. Sahoo P.K., Arora G.: *Image thresholding using two-dimensional Tsallis–Havrda–Charvát entropy*, „Pattern Recognition Letters" 2006, no. 27.
7. Wędrowska E.: *Wykorzystanie entropii Shannona i jej uogólnień do badania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej*, „Przegląd Statystyczny" 2010, nr 4.

MEASURES OF ENTROPY IN STATISTICS AND INFORMATION THEORY

Summary

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 and *Entropies of Mixing* were applied for studies on the properties of distributions in case of probabilities of the random variables.

Translated by Ewa Wędrowska