

ЗАСТОСУВАННЯ ЕНТРОПІЇ ДО ОЦІНКИ ВАРТОСТІ ІНФОРМАЦІЇ

О. М. Витвицька

*Івано-Франківський національний технічний університет нафти і газу;
76000, м. Івано-Франківськ, вул. Карпатська, 15;
тел. +380 (342) 72-71-31; e-mail: okvitvitska@ukr.net*

У статті запропоновано принципово новий підхід до оцінки вартості інформації, який враховує ймовірнісний характер інформації і ґрунтується на використанні теорії інформації. Відповідно до цього підходу розроблено метод вартісної оцінки інформації із застосуванням ентропії. Розглянуто властивості ентропії. Запропонований підхід до оцінки вартості інформаційних ресурсів дозволяє визначити вплив повноти інформації на її вартість із врахуванням зміни ентропії.

***Ключові слова:** кількість інформації, оцінка вартості інформації, ентропія, інформаційна асиметрія, повнота інформації.*

Актуальність проблеми. Сучасний період цивілізаційного розвитку характеризується вступом суспільства в інформаційну епоху і, відповідно, виникненням нової соціально-економічної формації – інформаційного суспільства, заснованого на виробництві і споживанні інформаційних товарів і послуг, суспільства, в якому його виробнича основа складається з високотехнологічних засобів виробництва, контрольованих комп'ютерними технологіями, у якому відбувається віртуалізація виробничих і соціальних зв'язків, внаслідок чого створюються передумови для залучення широкого кола працюючих до процесів прийняття управлінських рішень. У таких умовах сфера безпосереднього спілкування людей розширюється до глобальних масштабів, і як наслідок, інформація, знання і творчість – нематеріальні ресурси – стають ключовими елементами у соціальній і економічній діяльності людей [1, с.75; 2, с.9]. Так, в результаті розвитку інформаційного менеджменту і створення інформаційних мереж, розвинуті країни змогли здійснити стрімкий технологічний стрибок. До прикладу, США на даний час вкладає понад 10% ВВП у розвиток інформаційних технологій та отримує за їх рахунок 25% ВВП [19, с.185], у вартості компанії British Petroleum нематеріальні активи становлять 71%, IBM – 83% [3, с.42]. Виключна роль інформації у життєдіяльності людського суспільства обумовила необхідність оцінки її вартості.

Аналіз досліджень і публікацій, у яких започатковано вирішення проблеми. На відміну від матеріальних об'єктів, вартісна

оцінка (*valuation*) яких регламентується Національним стандартом України № 1 та Міжнародними стандартами оцінки [5, 6], вартість нематеріальних активів, зокрема інформації, складно визначити за допомогою традиційного інструментарію, не існує єдиної загальноприйнятої методики оцінювання інформаційних продуктів у зв'язку з їх оригінальністю і неповторністю. У працях [4, 7, 8, 9] зазначалась актуальність цього важливого завдання і запропоновано принципові підходи до його вирішення. А саме, для оцінки вартості інформації найбільш прийнятними є підходи і методи, які використовуються при оцінці нематеріальних активів. Дуглас У. Хаббард в своїх дослідженнях використав той факт, що інформація знижує невизначеність по відношенню до показника, котрий впливає на рішення, які мають економічні наслідки і запропонував використовувати калібровані оцінки ймовірності для оцінки вартості інформації [10]. Зважаючи на ймовірнісний характер інформації, ми бачимо перспективним використання теорії інформації для вирішення поставленого завдання.

Результати дослідження. Будь-яка діяльність на сучасному етапі відбувається в умовах невизначеності, коли для описання стану досліджуваного об'єкта чи системи характерною є наявність альтернатив, причому кожен із станів може наступити з деякою ймовірністю. Тому зняття невизначеності, а відповідно, інформативність про стан досліджуваної системи має ймовірнісний характер. Власне, такий зв'язок понять ймовірності і невизначеності (або інформативності) став основою для вимірювання ступеня невизначеності в статистичній теорії інформації.

З другого боку, перехід системи з менш імовірного стану в більш імовірний характеризується зростанням ентропії системи. Тому ентропію можна розглядати як міру ймовірності перебування системи в даному стані. Вищенаведені твердження: “ймовірність – кількісна міра інформації” і “ентропія – кількісна міра ймовірності” дозволяють сформулювати логічний умовивід: “ентропія – кількісна міра інформації” [11, с. 219]. Ентропія характеризує ступінь невизначеності тієї чи іншої характеристики будь-якої системи чи явища, про яке маємо інформацію.

Ентропія $H(x)$ визначається за формулою згідно з теоремою К. Шеннона, на основі якої середня кількість інформації, що припадає на один символ, дорівнює [12, с. 178]:

$$H(x) = -\sum_{i=1}^n P_i(x_i) \ln P_i(x_i), \quad (1)$$

де $P_i(x_i)$ – ймовірність отримання кожного з n можливих значень показника, що характеризує різноманіття (невизначеність) стану системи, яка досліджується.

Ступінь невизначеності стану джерела інформації залежить від кількості можливих станів і від ймовірностей цих станів.

Виділимо основні властивості ентропії:

1. Ентропія є величиною дійсною і від'ємною.
2. Ентропія – величина обмежена.
3. Ентропія дорівнює нулю, якщо ймовірність одного із станів джерела інформації дорівнює одиниці – тим самим стан джерела визначений.
4. Ентропія джерела з двома станами x_1 і x_2 та відповідними їм ймовірностями p і $1-p$ визначається виразом:

$$H(x) = -(p \ln p + (1-p) \ln(1-p)). \quad (2)$$

5. Ентропія об'єднаних статистично незалежних джерел інформації дорівнює сумі їх ентропій.

6. Ентропія максимальна при рівних ймовірностях всіх станів джерела інформації:

$$H(x) = -\sum_{i=1}^n p \ln p = -n \cdot \frac{1}{n} \ln \frac{1}{n} = \ln n. \quad (3)$$

Ентропія є мірою безладу системи. Ентропія, взята з від'ємним знаком, є мірою впорядкованості системи. Меншому значенню ентропії (меншій імовірності системи) відповідає більше інформації, що фіксує цей стан. Якщо система переходить через стохастичні (імовірнісні) флуктуації з більш імовірнісного стану в менш імовірнісний стан, ентропія падає, а інформаційний зміст системи збільшується.

Ентропія може бути обчислена при будь-якій основі логарифма. При зміні основи логарифмів відбувається зміна масштабу виміру ентропії. При використанні у якості основи логарифма числа 2 одиницю виміру ентропії називають біт; числа e – ніт, числа 10 – хартлі. Здебільшого у якості одиниці виміру ентропії користуються нітом, оскільки ця одиниця виміру зручна тим, що математичні перетворення і формули, у які входить ентропія, отримують більш простий вигляд [13, с. 159].

На неперервні випадкові величини поняття ентропії, яке вираховується за формулою (1), не розповсюджується, так як кількість можливих значень неперервної величини нескінченне. Однак неперервну випадкову величину можна подати як дискретну, вважаючи рівними (співпадаючими) ті значення, які відрізняються на величину меншу, ніж величина похибки Δx її визначення. Це рівнозначне заміні плавної кривої розподілу $f(x)$ многокутником розподілу. При цьому кожна ділянка довжиною Δx є однією точкою. Імовірності потрапляння значень випадкової величини на ділянки Δx , очевидно будуть дорівнювати $f(x)$.

Тоді по аналогії з (1), ентропія неперервної випадкової величини X наближено буде дорівнювати

$$H(x) \approx -\sum_{i=1}^n f(x_i) \Delta x \ln[f(x_i) \Delta x]. \quad (4)$$

Ця формула використовується при обчисленні ентропії неперервних випадкових величин [13, с. 160].

Точні значення ймовірностей $P(x_i)$ досліджуваного явища, які відповідають різним значенням дискретного параметра або різним значенням неперервного параметра, нам ніколи не відомі. Тому точне значення ентропії $H(x)$ не може бути знайдене. Фактично ми завжди маємо не саме значення ентропії $H(x)$, а її оцінку $H_n(x)$ за деякою сукупністю спостережень, об'єм яких дорівнює n . Ця оцінка зв'язана з $H(x)$ співвідношенням $H(x) = \lim_{n \rightarrow \infty} H_n(x)$. Така оцінка називається вибірковою ентропією.

Якщо деякий параметр X економічного явища вимірюваний у n кількості спостережень і відомо, що похибка спостережень дорівнює Δx , то вибіркова ентропія $H_n(x)$ може бути обчислена за формулою

$$H_n(x) = -\sum_{i=1}^k \frac{n_i}{n} \ln \frac{n_i}{n}, \quad (5)$$

де n_i – число спостережень, що належить до i -го інтервалу; k – число інтервалів (кожний шириною Δx); $\omega = \frac{n_i}{n}$ – відносна частота (відношення частоти n_i у кожному з інтервалів зміни неперервної випадкової статистичної величини до загального числа спостережень n).

Саме ця формула найчастіше використовується на практиці для обчислення ентропії [13, с. 166].

Строго кажучи вибіркова ентропія $H_n(x)$ характеризує тільки різноманіття, що властиве наявній у нашому розпорядженні сукупності результатів спостережень за економічним явищем. Однак вона може слугувати оцінкою різноманіття, що притаманне досліджуваному явищу, тобто оцінкою генеральної сукупності значень параметра, якщо спостереження рівномірно виконані у всьому діапазоні зміни параметра, а також число спостережень достатньо велике. Очевидно, що у силу дії закону великих чисел, при рівномірному розміщенні достатньо великого числа спостережень у всьому їх діапазоні повинно виконуватись співвідношення

$$\sum_{i=1}^k \omega_i \ln \omega_i \cong \sum_{i=1}^k P_i \ln P_i, \quad (6)$$

яке вказує на те, що вибіркова ентропія $H_n(x)$ може слугувати наближеною оцінкою ентропії $H(x)$ генеральної сукупності. Мірою такого наближення є дисперсія вибіркової ентропії. При невеликому числі спостережень вибіркова ентропія виявляється сильно заниженою у порівнянні з ентропією генеральної сукупності, але з ростом числа спостережень різниця між $H(x)$ і $H_n(x)$ прямує до нуля [13, с. 166].

Для того, щоб сукупність значень спостережень $H_n(x)$ відображала різноманіття $H(x)$, яке притаманне досліджуваному явищу, різноманіття $H_n(x)$ повинно бути порівнюваним з різноманіттям $H(x)$. При числі спо-

спережень n максимальне різноманіття сукупності отриманих значень параметра досягається, якщо n спостережень приведуть до отримання значень параметра, що належить до різних інтервалів області змін параметра. У таких умовах ентропія сукупності отриманих значень параметра буде $H(x) = \ln n$. У всіх решти випадках $H_n(x)$ буде менша $\ln n$. Отже для вибіркової ентропії справедлива нерівність $H_n(x) \leq \ln n$ [13, с. 167].

Звідси зрозуміло, що сукупність значень параметра об'ємом n спостережень не буде відображати різноманіття $H(x)$, що притаманне досліджуваному явищу, до того часу, доки $\ln n$ не перевищить $H(x)$. Тому на початкових стадіях дослідження економічного явища, доки число спостережень за значеннями параметра X буде невелике, завжди повинна виконуватись нерівність $H_n(x) < H(x)$. З іншого боку, так як $H_n(x)$ є зміщеною оцінкою $H(x)$, яка має ту властивість, що із зростанням n зміщення зменшується, то при малих n вибіркова ентропія, залишаючись меншою $H(x)$, повинна до неї прямувати, тобто повинна систематично рости.

Очевидно, повинно існувати якесь граничне значення $n_{\text{гран}}$, (його можна назвати оптимальним числом спостережень) таке, що при $n < n_{\text{гран}}$ вибіркова ентропія $H_n(x)$ матиме тенденцію до зростання з ростом n , а при $n > n_{\text{гран}}$ вибіркова ентропія коливатиметься поблизу рівня, що відповідає ентропії $H(x)$ економічного явища. Його можна знайти, побудувавши графік залежності ентропії від кількості спостережень. Точка на такому графіку, яка відділяє ділянку систематичного росту вибіркової ентропії, від ділянки, де $H_n(x)$ вже не зростає, а коливається в межах точності спостережень поблизу прямої, паралельної осі абсцис, і буде граничним значенням вибіркової ентропії. Це граничне значення ентропії може слугувати у якості критерію про необхідний граничний обсяг спостережень за економічним явищем для отримання повної інформації про нього. Воно дозволяє також вирішити важливе практичне питання про граничний обсяг інформації, яка є корисною.

Як вже зазначалось вище, описані властивості ентропії використовують у теорії інформації для характеристики її кількості. Під кількістю інформації розуміють величину знятої невизначеності про стан системи [14]:

$$I(x) = H_0(x) - H(x), \quad (7)$$

де $I(x)$ – кількість інформації, отримана про систему у результаті її вивчення; $H_0(x)$ – міра ступеня невизначеності системи (ентропія) до початку спостережень; $H(x)$ – міра ступеня невизначеності системи (ентропія) по закінченні спостережень.

Якщо під $H(x)$ розуміти границю, до якої теоретично можна зменшувати ентропію системи, то $I(x)$ буде повною інформацією, тобто максимальною кількістю інформації, яку можна отримати про систему, яка вивчається. У ряді випадків $H(x)=0$, тобто у результаті спостережень

може бути виявлений той стан системи, у якому вона дійсно знаходиться. Тоді $I(x) = H_0(x)$, тобто кількість інформації дорівнює ентропії системи.

Процес досліджень іде таким чином, що із накопиченням результатів спостережень відбувається не зменшення невизначеності, а її збільшення. Це відбувається тому, що ми спочатку нічого не знаємо про явище, яке вивчаємо, і вимушені ототожнювати характеристики отриманої вибірки з характеристиками явища. Тому кількість повної інформації в цьому випадку може бути обчислено за формулою (6) тільки при умові врахування напрямку зміни ентропії. Врахування цього факту можна здійснити простою зміною знаку правої частини формули (7). Величина $H_0(x)$ у такому випадку дорівнює ентропії вибірки з одним можливим значенням параметру, тобто дорівнює нулю, а величина $H(x)$ дорівнює ентропії генеральної сукупності значень параметра. Отже, кількість повної інформації про параметри досліджуваного явища як складної системи буде чисельно дорівнювати ентропії $H(x)$ генеральної сукупності значень параметра і тоді $I(x) = H(x)$.

З точки зору оцінки, окрім величини повного обсягу інформації цікаво знати і ту її кількість, яка поступає з кожним новим спостереженням, тобто величину приросту інформації. Очевидно, що ця величина ΔI чисельно дорівнює приросту ентропії [13, с. 180]

$$\Delta I = -[H_j(x) - H_{j+1}(x)], \quad (8)$$

де $H_j(x)$ – ентропія до початку спостережень; $H_{j+1}(x)$ – ентропія після проведення спостережень.

З врахуванням викладеного оцінку вартості інформації можна здійснити за формулою:

$$V_i = \Delta I \cdot V_e - B_i = V_e(H(x)_{j+1} - H(x)_j) - B_i, \quad (9)$$

де V_i – вартість інформації, грн; ΔI – приріст інформації, нит; V_e – ціна одиниці ентропії, грн./нит; B_i – витрати на придбання або отримання інформації, грн.

Ціна одиниці ентропії може бути розрахована за формулою

$$V_e = \frac{\Delta E}{H_n^{span}(x)}, \quad (10)$$

де ΔE – величина економічної вигоди (попереджених збитків) від використання інформації, грн; $H_n^{span}(x)$ – граничне значення ентропії, яке може бути визначене тільки у разі наявності повної інформації про стан досліджуваної системи (явища), нит.

Необхідно зауважити, що при недостатніх обсягах інформації неможливо встановити $H_n^{span}(x)$. У такому випадку в якості $H_n^{span}(x)$ можна використати максимальне значення ентропії, але слід пам'ятати, що оцінка інформації при цьому буде дещо завищеною.

Проілюструємо викладене прикладом, наведеним у табл. 1, де подано інформацію про величини логістичних витрат на придбання одного з видів матеріалів (насосно-компресорних труб), які використовуються в процесі нафтогазовидобутку.

Таблиця 1. Ентропії величин логістичних витрат при різних обсягах інформації

№ спостереження	1	2	3	4	5	6	7	8	9	10
Величина логістичних витрат, тис. грн	19,0	24,6	20,9	13,9	17,6	16,2	18,6	19,6	21,5	20,1
Вибіркова ентропія, біт	0	0,69	1,1	1,39	1,61	1,56	1,55	1,5	1,52	1,5

Покажемо, як вираховується ентропія на прикладі H_{10} . Після отримання десятого значення інформації про величини логістичних витрат у сукупність спостережуваних значень H_k входили: 19,0; 24,6; 13,9; 17,6; 16,2; 18,6; 19,6; 21,5; 20,1 тис. грн. Розподіл цих значень за інтервалами величиною $\Delta H_k = 2$ тис. грн. показано у табл. 2.

Таблиця 2. Визначення ентропії для H_{10}

Інтервал, тис.грн	13-15	15-17	17-19	19-21	21-23	23-25
Число значень у інтервалі, n_i	1	1	3	3	1	1
Відносні частоти, ω_i	0,1	0,1	0,3	0,3	0,1	0,1

Обчислюємо ентропію за формулою (1):

$$H_{10} = -(0,1 \ln 0,1 + 0,1 \ln 0,1 + 0,3 \ln 0,3 + 0,3 \ln 0,3 + 0,1 \ln 0,1 + 0,1 \ln 0,1) =$$

$$= -(4(0,1 \ln 0,1) + 2(0,3 \ln 0,3)) = 4 \times 0,2303 + 2 \times 0,3612 = 1,6436 \text{ біт.}$$

Для даного прикладу відомо, що величина економічної вигоди від наявності повної інформації про логістичні витрати становить 30 тис. грн. Також на початковий момент у розпорядженні підприємства було тільки три значення величин логістичних витрат. Витрати підприємства на отримання одиниці інформації складають 500 грн.

З табл. 2 видно, що величина ентропії стабілізується вже після сьомого спостереження. Отже, ціна одиниці ентропії дорівнює

$$V_e = \frac{30}{14,5} = 20 \text{ тис. грн.}$$

Вартість інформації, обчислена за формулою (9), становить:

$$V_i = (1,5 - 1,1) 20 - 0,5 \times 7 = 8 - 3,5 = 4,5 \text{ тис. грн.}$$

Висновки та перспективи подальших наукових досліджень. Проблеми вибору в умовах невизначеності і ризику також тісно пов'язані з асиметрією інформації. В основних фундаментальних мікро-

економічних дослідженнях поведінка учасників ринку аналізується, виходячи з припущення, що всі агенти ринку володіють повною інформацією, необхідною для прийняття ними рішень. Реальні ж умови вкрай рідко відповідають припущенню про повноту і симетрію розподілу інформації. Навпаки, як правило, спостерігаємо нестачу і недоступність ринкової інформації, що перешкоджає прийняттю оптимальних рішень. Тоді має місце асиметрія інформації – нерівномірний розподіл між учасниками ринку інформації про умови здійснення ринкової угоди і наміри щодо іншої сторони [15]. Властивості ентропії можна використати для характеристики повноти інформації, тобто граничного значення тієї кількості інформації, яка є достатньою для прийняття рішення. Для вирішення проблем, пов'язаних із спотвореннями, породженими асиметрією і неповнотою інформації, доцільно використовувати ентропію, яка, зокрема, дозволяє встановити граничне значення кількості інформації, а відповідно – оптимальне число спостережень, вимірювань – яке впливає на вартість інформації і водночас є достатнім для вирішення поставленої задачі, що в свою чергу, сприятиме досягненню інформаційної симетрії.

Література

1. Социально-экономические проблемы информационного общества: монография / под ред. Л.Г. Мельника, М.В. Брюханова. – Вып.2. – Сумы: Университетская книга, 2010. – 896 с.
2. Чухно А.А. Сучасні економічні теорії / А.А. Чухно, П.І. Юхименко, П.М. Леоненко. – К.: Знання, 2007. – 878 с.
3. Мамонтова Н.А. Управління вартістю компаній нафтогазового комплексу в умовах інноваційного розвитку: монографія / Н.А. Мамонтова. – Львів: ПАІС, 2011. – 484 с.
4. Економіка підприємства: навч. посібник / [Й.М. Петрович, А.Ф. Кіт, Г.М. Захарчин та ін.]; за ред. Петровича Й. М. – Львів: Магнолія 2006, 2008. – 580 с.
5. Загальні засади майна і майнових прав. Постанова КМУ № 1440 від 10 вересня 2003 року [Інтернет-ресурс]. – Режим доступу: <http://www.akadem.kiev.ua> – (Національний стандарт № 1).
6. Международные стандарты оценки. Седьмое издание. 2005; [пер. с англ. И.Л. Артеменкова, Г.И. Микерина, Н.В. Павлова]. – М.: ООО “Российское общество оценщиков”, 2005. – 414 с. – (Международные стандарты оценки).
7. Шевчук О.Б. Інформаційний капітал: його сутність і види / О.Б. Шевчук // Економічна теорія. – 2005. – № 2. – С. 41-48.
8. Яремко І.Й. Оцінка капіталу підприємства: інформаційне забезпечення вартісно-орієнтованої концепції управління підприємством: монографія / І.Й. Яремко. – Львів: Новий світ – 2000, 2005. – 236 с.

9. Мендрул А.Г. Оценка стоимости нематериальных активов / А.Г. Мендрул, В.С. Ларцев. – К: ООО «Полиграф-Информ», 2004. – 264 с.
10. Дуглас У. Хаббард. Как измерить всё, что угодно. Оценка стоимости нематериального в бизнесе / Дуглас У. Хаббард / [Пер.с англ. Е. Пестеревой] – М.: ЗАО “Олимп-Бизнес”, 2009. – 320 с.
11. Основи стійкого розвитку: навчальний посібник / [Л.Г. Мельник, О.І. Карінцева, С.М. Шевченко та ін.]; за заг. ред. д. е. н., проф. Мельника Л. Г. – Суми: ВТД «Університетська книга», 2005. – 654 с.
12. Управление. Информация. Интеллект / под.ред А.И. Берга, Б.В. Бирюкова, Е.С. Геллера, Г.Н. Поварова. – М.: Мысль, 1976. – 383 с.
13. Дементьев Л.Ф. Применение математической статистики в нефтепромысловой геологии / Л.Ф. Дементьев, М.А. Жданов, А.Н. Кирсанов. – М.: Недра, 1977. – 255 с.
14. Яглом А.М. Вероятность и информация / А.М. Яглом, И.В. Яглом. – М.: Физматгиз, 1960. – 350 с.
15. Экономическая сущность асимметрии информации и её влияние на функционирование рынка [Интернет-ресурс]. – Режим доступа: <http://safbd.ru/magazine/article>

Стаття поступила в редакційну колегію 26.12.2014 р.

*Рекомендовано до друку к. ф-м. н., доцентом **Осипчуком М.М.**,
д.т.н., професором **Мойсишиним В. М.***

APPLICATION OF ENTROPY TO THE VALUE OF INFORMATION ASSESSMENT

O. M. Vytvytska

*Ivano-Frankivs'k National Technical University of Oil and Gas;
76019, Ivano-Frankivs'k, Carpathians str., 15;
ph. +380 (342) 72-71-31; e-mail: okvitvitska@ukr.net*

This article proposes brand-new approach to the value of information assessment that takes into account probabilistic nature of the information and based on application of information theory. Due to this approach the method of value of information assessment with application of entropy was developed. The properties of entropy were observed. Proposed approach to the value of information resources assessment enables to determine completeness impact of information on its value, taking into account changes in entropy.

Key words: *amount of information, value of information assessment, entropy, information asymmetry, completeness of information*