10. Кількість інформації та ентропія. Ентропія випадкового експерименту з рівноімовірними наслідками. Ентропія випадкового експерименту з нерівноімовірними наслідками.

10.1. Кількість інформації та ентропія.

Приступаючи до викладу елементів теорії інформації, необхідно, перш за все, ввести кількісну міру інформації, яка передається по каналу зв'язку. Нас цікавитиме деяка формалізована кількість інформації, що міститься в повідомленні, яке в першу чергу характеризує завантаження каналу зв'язку і трудність передачі даного повідомлення по каналу зв'язку.

Розгляд природно почати з простих повідомлень - дискретних. Останні доцільні також тому, що як ми покажемо надалі, про що говорилось раніше, <u>будь-які повідомлення можуть бути представлені у вигляді послідовності елементарних повідомлень</u>, які приймають одне з М наперед відомих значень, що надходять один за одним з інтервалом дискретності T_0.

Виділимо один інтервал дискретності $t \in (0,T_0)$. Джерело повідомлень на цьому інтервалі створює елементарне повідомлення, тут ми його позначимо через X, яке може прийняти одне з М значень: $x_1,x_2,...x_M$. Одержувач після обробки коливання, що приймається, ухвалює рішення про те, яке саме повідомлення з множини $\{x_i\}$ було послано. Відповідно до позначень, прийнятих в теорії інформації, це рішення (оцінку повідомлення, що приймається) позначимо Y. Рішень взагалі може бути М, тобто $Y \in \{y_1,y_2,...,y_M\}$, і y_1 означає рішення про те, що послано повідомлення x_1 і т.д.

Нагадаємо характеристику двох основних станів: **апріорного і апостеріорного**, в яких знаходиться одержувач в процесі прийому кожного елементарного повідомлення.

Відповідно до реальних умов роботи більшості систем зв'язку в теорії інформації приймається, що в **апріорному** стані одержувачу відомий набір можливих елементарних повідомлень $x_1,x_2,...x_M$ і імовірність посилки кожного повідомлення $P(X_1),...,P(X_M)$. Характерним є стан невизначеності у одержувача: він знає, що йому послано одне з М наперед відомих йому повідомлень, але яке саме послане, він не знає. Цю невизначеність одержувача в апріорному стані називають **<u>апріорною невизначеністю</u>**. Мабуть, кількісно ця невизначеність якось характеризується апріорним законом розподілу імовірності $P(x_i)$, $i=1,2,..,M$ посилки різних повідомлень.

Другий стан одержувача, зване **апостеріорним**, характерний тим, що на вхід приймальної системи (приймача) поступило коливання, що приймається, і по ньому сформовано рішення (оцінка) Y.

Покладемо одержувач, діючи по наказаному йому алгоритму обробки, ухвалив деяке конкретне рішення, скажімо, Y_j. Якщо перешкод в каналі немає, то одержувач в цьому випадку точно знає, що було послане повідомлення X_j. Ніякої невизначеності у нього не залишається. Проте випадок відсутності перешкод в каналі зв'язку може використовуватися тільки як перша ідеалізація при проведенні деяких досліджень. Основний інтерес представляє облік явищ, пов'язаних з обов'язковою наявністю різних перешкод в каналі зв'язку. Якщо перешкоди є, то після отримання оцінки Y_j у адресата ще залишається деяка невизначеність відносно того, яке повідомлення з множини $x_1, x_2, \ldots x_M$ насправді було послане.

Ця невизначеність, звана **апостеріорною невизначеністю адресата** (одержувача), характеризується апостеріорним розподілом $P(x_i/x_j)$, тобто розподілом імовірності посилки різних повідомлень x_i за умови, що прийняте повідомлення (оцінка) x_j. У окремому випадку, коли дію перешкод можна не враховувати, апостеріорний розподіл вироджується в розподіл вигляду $P(x_i/x_j)= P(x_i)$, що характеризує детермінованість результату.

Отримання інформації тотожне усуненню незнання чогось або усуненню невизначеності чогось. Тому, розумно визначити **кількість інформації**, яку одержує адресат або яку передають йому по каналу зв'язку, як **кількість невизначеності, що знімається у адресата при отриманні їм повідомлення**. Відповідно кількість інформації, що передана по каналу зв'язку, ми можемо визначити за допомогою наступної поки що словесною формулою:

кількість інформації, що передається по каналу зв'язку (одержуване адресатом)	=	апріорна невизначеність адресату	-	апостеріорна невизначеність адресату

У ідеалізованому каналі зв'язку без перешкод, апостеріорна невизначеність відсутня, тобто рівна нулю. При цьому

кількість інформації, що передається по каналу зв'язку (одержуване адресату)	=	апріорна невизначеність адресату

Таким чином, ми виразили кількість інформації через кількість невизначеності у адресата в апріорному і апостеріорному стані. Інакше кажучи, необхідність визначення кількісної міри інформації ми звели до необхідності визначення кількісної міри невизначеності, що, виявляється, є простішою задачею.

Апріорна і апостеріорна невизначеності адресата є окремим випадком загальнішої невизначеності - **невизначеності довільного апріорного експерименту**, або **невизначеності закону розподілу імовірності $P(x_i)$**, що описує довільний випадковий експеримент X з результатами $x_1, x_2, \ldots x_M$. Кількісно міру невизначеності випадкового експерименту в теорії інформації умовилися називати **ентропією** і позначати буквою H.

Термін "ентропія" запозичений з термодинаміки, де аналогічний вираз характеризує середню невизначеність стану системи молекул речовини.

Термін **ентропія**, таким чином, використовується надалі як синонім поняття **невизначеності** (кількості невизначеності), пов'язаний з реалізацією деякої апріорної величини X або з результатом деякого випадкового експерименту. Тому ми більш менш детально розглянемо кількісні співвідношення, що визначають ентропію довільних експериментів, після чого знову повернемося до поняття кількості інформації. Почнемо, природно з простіших випадків.

10.2. Ентропія випадкового експерименту з рівноімовірними наслідками.

Будь-який випадковий експеримент, зокрема робота джерел повідомлень, описується законом розподілу імовірності різних результатів і володіє деякою невизначеністю в тому значенні, що результат експерименту точно передбачити не можна. Мабуть, невизначеність тим більше, ніж менше шансів (чим менше імовірність) передбачити результат експерименту. І також, чим менша апріорна імовірність окремої події, тим більшу кількість інформації вона несе.

Це положення буде керівною ідеєю визначення кількісної міри невизначеності (ентропії) випадкового експерименту і кількості інформації, що що міститься в одному наслідку. Тому природно припустити, що цією кількісною мірою може бути величина, зворотна його апріорній імовірності $1/p(x_i)$. Однак при визначенні кількісної міри невизначеності ми повинні зажадати, щоб ентропія випадкового експерименту задовольняла наступним умовам:

а) Ентропія H випадкового експерименту з рівноімовірними наслідками монотонно зростає у міру збільшення числа результатів M.

Дійсно, при киданні монети M=2, імовірність передбачити результат експерименту рівна 1/2 (шанси передбачити результат рівні 50%).

При киданні гральної кістки M=6, імовірність передбачити результат експерименту рівна 1/6 і т.д.

До умови а) приєднаємо ще дві очевидні умови б) і в) яким повинна задовольняти кількісна міра невизначеності будь-яких випадкових експериментів (не тільки з рівноімовірними результатами).

б) Якщо число результатів експерименту рівне одиниці (M=1), то результат експерименту вирішений наперед і невизначеності немає: H=0.

в) Якщо проводиться два незалежні експерименти A і B, то повна невизначеність експерименту AB рівна сумі невизначеностей експерименту A і експерименту B (властивість адитивності):

$$H(AB)=H(A)+H(B) \tag{10.1}$$

Остання умова еквівалентна наступній очевидній вимозі, яку необхідно пред'явити до кількісної міри невизначеності.

Невизначеність H(AA) двічі незалежно повтореного експерименту A в два рази більше невизначеності H(A) одного експерименту H(AA)=2H(A). Якщо експеримент A незалежним чином повторюється n раз, то повна невизначеність n-кратного експерименту повинна бути рівна nH(A).

Виразимо тепер сформульовані умови, або вимоги до кількісної міри невизначеності H, в аналітичному вигляді. На підставі умови а) можна записати наступне.

Якщо експеримент має M рівноімовірних наслідків, то

$$H= f(M), \tag{10.2}$$

де f(•) - деяка монотонно зростаюча функція, яку належить визначити.

Умова б) дає $$f(1)=0 \tag{10.3}$$

Для того, щоб скористатися умовою в), покладемо, що експеримент A має M_A рівноможливих наслідків і відповідно ентропію $H(A)= f(M_A)$, а експеримент B має M_B рівноможливих наслідків і ентропію $H(B)= f(M_B)$.

Тоді експеримент AB, що є сукупністю експерименту A і експерименту B (які в даному випадку вважаються незалежними), має $M_A M_B$ рівноімовірних наслідків, і, отже його ентропія повинна виражатися формулою $H(A)=f(M_A M_B)$. На підставі умови в) $H(AB)=H(A)+H(B)$. отже, для функції $f(\bullet)$ повинна виконуватися рівність

$$f(M_A M_B)=f(M_A)+f(M_B) \tag{10.4}$$

Одержаним співвідношенням (10.3) і (10.4) при довільних значеннях M_A і M_B задовольняє єдина функція, а саме: логарифмічна функція. Отже,

$$f(\bullet)=\log_a(\bullet) \tag{10.5}$$

Вибір основи логарифма «а» принципового значення не має. Основа визначає величину одиниці вимірювання невизначеності. Нагадаємо, наприклад, що при вимірюванні довжини можуть використовуватися різні одиниці: фут і т.д. Вибір тієї або іншої одиниці є питанням зручності. Виявилося зручним прийняти підставу логарифма в (10.5) рівним 2.

Тоді **кількість інформації** $I(x_i)$, що міститься в символі x_i, вибираному з ансамблю $\{x_i\}$ (i=1, 2, 3,…, M, M- об'єм алфавіту) з імовірністю $P(x_i)$, причому $\sum^{K} P(\alpha)=1$, визначається як

$$I(x_i) = \log 1/P(x_i) = -\log P(x_i). \tag{10.6}$$

У випадку рівної імовірності повідомлень $P(x_1)=P(x_2)=\ldots=P(x_M)=1/M$ вираз (10.6) для кількості інформації приводиться до виду

$$I(x_i) = -\log P(x_i) = \log M. \tag{10.7}$$

Така міра кількості інформації була запропонована в 1928 році Р. Хартлі.

У (10.7) і скрізь в подальшому відповідно до символіки, прийнятої в теорії інформації, основа «а» логарифма не проставляється, якщо а=2.

Визначимо **ентропію випадкового експерименту з рівноімовірними наслідками**. Ентропія визначає середню кількість інформації I(X), що припадає на один символ. Тоді

$$\mathbf{H = \frac{\sum_{i=1}^{M} I(x_i)}{M} = \frac{-\sum_{i=1}^{M} \log P(x_i)}{M} = \frac{\sum_{i=1}^{M} \log M}{M} = \frac{M \log M}{M} = \log M} \tag{10.8}$$

Це значить, що <u>за одиницю невизначеності прийнята невизначеність, яку має випадковий процес з двома рівноімовірними наслідками</u> (M=2). Прикладами такого експерименту є кидання монети, посилка двійкового повідомлення 0 або 1 (при рівності імовірності 0 і 1).

Дійсно, підстановка в (10.6) M=2 дає H=1. Цю величину називають “двійкова одиниця” (дв. ед.), або “біт”, що є скороченням англійського терміну binary digit. Отже, в експерименті з двома рівноімовірними наслідками

$$H=\log_2 2=1 \text{ дв. ед. (біт)}$$

Одиниці вимірювання кількості інформації і ентропії залежать від вибору основи логарифма у формулах. При використовуванні десяткових логарифмів кількість інформації і ентропія визначаються в десяткових одиницях - дітах. У разі використовування двійкових логарифмів кількість інформації і ентропія визначаються в двійкових одиницях - бітах. Нарешті, при використовуванні натуральних логарифмів одиницею вимірювання є натуральних одиниця - ніт.

При аналізі інформаційних процесів в ЕОМ і інших пристроях, що функціонують на основі двійкової системи числення, зручно користуватися двійковими одиницями. У математичних викладеннях зручно користуватися натуральними одиницями.

Якщо джерело повідомлень створює одне з 32-х рівноімовірних повідомлень, то ентропія джерела повідомлень або апріорна ентропія адресата

$$H=\log_2 32=5 \text{ дв. ед.}$$

Кожну посилку при роботі такого джерела можна представити п'ятизначним двійковим числом і звести до послідовного проведення 5-ти простих незалежних експериментів з двома рівноімовірними наслідками $\{0,1\}$.

Не дивлячись на збіг залежностей (10.7) та (10.8), <u>ентропія</u> H(x) і <u>кількість інформації</u> I(x) **принципово різні**. H(x), що виражає середню невизначеність стану джерела повідомлень є об'єктивною характеристикою джерела повідомлень, і, якщо відома статистика повідомлень, може бути обчислена апріорно, тобто до отримання повідомлення, I(x) є апостеріорною характеристикою і визначає кількість інформації, одержувану з надходженням повідомлень. H(x) є міра недоліку інформації про стан окремої системи. З надходженням інформації про стан системи ентропія останньої знижується.

Збіг виразу для I(x) з виразом для H(x) свідчить лише про те, що кількість одержуваної інформації чисельно рівна ентропії, яка мала місце щодо джерела повідомлень.

10.3. Ентропія випадкового експерименту з нерівноімовірними наслідками.

Постараємося тепер перекласти ці результати на експеримент з нерівноімовірними наслідками.

Спосіб доказу, який буде використаний при цьому, вельми характерний для методів теорії інформації.

Покладемо, число різних наслідків випадкового експерименту знову рівної M: $X\in(x_1,\ldots,x_M)$. Але тепер імовірність настання різних наслідків x_i $P(x_i)=P_i$; $i=1,2,\ldots,M$ (наприклад, імовірність посилки різних повідомлень) вважатимемо різними. Вимагається визначити ентропію експерименту X.

Замість одноразово проведеного експерименту X розглядатимемо дуже довгу послідовність $X^{(1)},X^{(2)},\ldots,X^{(N)}$ що складається з N незалежно проведених експериментів X. Відповідно до закономірності (умова в)), ентропія H_N N - раз повтореного експерименту (N-кратного експерименту) при обумовленій нами статистичній незалежності окремих експериментів рівна ентропії одноразового експерименту, помноженої на N:

$$H_N=NH$$

і ентропія, що цікавить нас

$$H=\frac{1}{N}H_N \tag{10.9}$$

Розглянемо послідовність (ланцюжок) результатів N-кратного експерименту

$$\mathbf{(X^{(1)},X^{(2)},...,X^{(N)}\ ;\ X^{(i)} \in (x_1,...,x_M))} \qquad (10.10)$$

Якщо ланцюжок наслідків (10.2) достатньо довгий (теоретично, якщо $N\to\infty$), то на підставі цього ланцюжка можна підрахувати імовірність реалізації різних наслідків

$$P(x_1) = P_1 = \frac{N_1}{N},\ldots,P(x_M) = P_M = \frac{N_M}{N} \qquad (10.11)$$

де $N_1,...,N_M$ - число випадань в ланцюжку наслідків (10.10) наслідку $(x_1,...,x_M)$ відповідно. Навпаки, якщо імовірності різних наслідків $P_1,...,P_M$ задані, як це має місце в даному випадку, на підставі (10.11) можна передбачити наперед, яке число різних наслідків наступить в ланцюжку результатів (10.10) N-кратного експерименту. Якщо N достатньо велике, то практично обов'язково число перших наслідків x_1 буде рівне $N_1=P_1N$. На яких саме місцях ланцюжка з'являться ці P_1N перших наслідків x_1, не відомо. Число других наслідків x_2 буде $N_2=P_2N$ і т.д. Нарешті, M-й наслідок наступить $N_M=P_MN$ раз.

Отже, всі ланцюжки, що є різними реалізаціями N-кратного експерименту, матимуть один і той же набір елементів. Цей набір елементів або наслідків можна представити у вигляді рисунку 1.

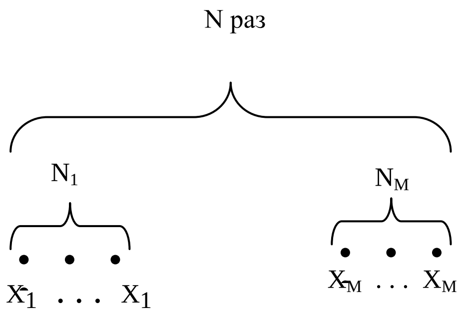

Рис. 10.1.

Різні реалізації відрізнятимуться один від одного тільки розташуванням елементів в ланцюжку наслідків. Невизначеність наслідку N-кратного експерименту полягає, власне, в невизначеності розташування наперед відомого набору наслідків.

Випадання кожного елементу ланцюжка (10.10) є незалежна подія. Тому імовірність P_Σ реалізації якого-небудь конкретного ланцюжка рівна добутку імовірностей реалізації окремих елементів і виходить одна і та же для всіх ланцюжків, що представляють різні реалізації N-кратного експерименту:

$$P_\Sigma = \underbrace{P_1 \cdots P_1}_{P_1N\ \text{раз}} \cdots \underbrace{P_M \cdots P_M}_{P_MN\ \text{раз}}$$

або

$$P_\Sigma = P_1^{P_1N} \cdots P_M^{P_MN} \qquad (10.12)$$

Таким чином, всі ланцюжки наслідків, які можуть реалізовуватися при N-кратному експерименті, при достатньо великому N (теоретично при $N \to \infty$) рівноімовірні, і ми приходимо до схеми експерименту з рівноімовірними наслідками.

Відповідно ентропія N-кратного експерименту HN повинна розраховуватися по формулі (10.8) і виходить рівною

$$H_N = -\log P_\Sigma = -N\,[P_1 \log P_1 + \ldots + P_M \log P_M].$$

Тепер неважко визначити ентропію H=H(X), що цікавить нас, експерименту X з нерівноімовірними наслідками $X_1, \ldots, X_M$.

Для цього слід тільки скористатися рівністю (10.9)

$$H = -\sum_{i=1}^{M} P_i \log P_i \qquad (10.13)$$

або в іншому часто використовуваному записі

$$H(X) = -\sum_{i=1}^{M} P(X_i) \log P(X_i) \qquad (10.14)$$

Одержана загальна формула для ентропії випадкового експерименту X (або випадкової величини X) є основною, на якій базуються всі подальші висновки теорії інформації. У окремому випадку, коли всі результати рівноімовірні, тобто коли $P_1 = P_2 = \ldots = P_M = 1/M$, формула (10.14) переходить в (10.8) і (10.7). В даному випадку ентропія визначає середню кількість інформації, що міститься в одному з символів (букв, цифр) алфавіту. Тоді кількість інформації повідомлення складається з до символів дорівнюватиме $I = k \cdot H$.

Помітимо, що саме за збіг формули (10.13) з формулою для ентропії у фізиці (термодинаміці) невизначеність в теорії інформації назвали ентропією. Деякі фахівці прагнуть зв'язати ентропію в теорії інформації з ентропією у фізиці. Проте до теперішнього часу змішення цих понять швидше приводило до плутанини, ніж до яких-небудь корисних узагальнень.