11. Умови максимізації ентропії. Визначення умовної ентропії H(X/Y). Ентропія складного експерименту при взаємозалежності подій. Визначення ентропії складного експерименту H(X,Y).

11.1. Умові максимізації ентропії.

Розглянемо тепер наступну задачу. Припустимо експеримент X має M наслідків $X_1, X_2, ..., X_M$. Який повинен бути розподіл ймовірностей наслідків $P(X_1)$, ..., $P(X_M)$, щоб ентропія експерименту

$$H(X) = -\sum_{i=1}^{M} P(X_i) \log P(X_i) = \max \tag{11.1}$$

тобто щоб вона була **максимальною**.

При цьому слід врахувати, що імовірність $\{P(X_i)\}$ не може вибиратися довільно.

Має місце природне обмеження

$$\sum_{i=1}^{M} P(X_i) = 1 \tag{11.2}$$

Сформульована задача, отже, є задачею на умовний екстремум.

Для її вирішення складемо функцію Лагранжа

$$F = -\sum_{i=1}^{M} P_i \log P_i + \lambda \sum_{i=1}^{M} P_i$$

Далі користуючись системою рівнянь

$$\left.\begin{array}{l} \dfrac{\partial F}{\partial P_1} = -\log P_1 - \dfrac{1}{\ln 2} + \lambda = 0 \\ \ldots.. \\ \dfrac{\partial F}{\partial P_M} = -\log P_M - \dfrac{1}{\ln 2} + \lambda = 0 \end{array}\right\} \tag{11.3}$$

знаходимо, що

$$\begin{array}{c} \sum_{i=1}^{M} P_i = 1 \\ P_1 = P_2 = \ldots = P_M \end{array} \tag{11.4}$$

тобто при даному числі наслідків M найбільшою ентропією володіє експеримент з рівноімовірними наслідками. Цей результат відповідає початковому положенню про те, що невизначеність тим менше, ніж більше шансів передбачити результат експерименту.

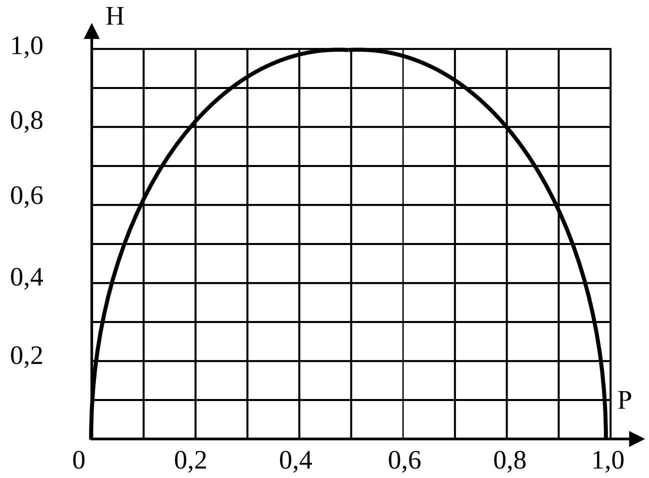
Так, наприклад, в простому випадку експеримент з двома наслідками при рівноімовірних наслідках $P_1 = P_2 = 0{,}5$

$$H = -2\frac{1}{2}\log\frac{1}{2} = 1 \text{дв.ед.}$$

Якщо наслідки не рівноімовірні, скажемо $P_1 = 0{,}99$, а $P_2 = 0{,}01$, то невизначеності майже немає.

Можна передбачити з великим ступенем вірності, що результатом експерименту буде X_1. Розрахункова формула попередньої лекції (10.13) при цьому дає

$$H = -0{,}99 \log 0{,}99 - 0{,}01 \log 0{,}01 \approx 0{,}08 \text{ дв. од.}$$



На рисунку зображений графік залежності ентропії H експерименту з двома наслідками від імовірності P_1 одного з наслідків.

У міру наближення закону розподілу до рівноімовірного ентропія зростає.

Ентропія досягає максимуму, якщо символи взаємонезалежні і рівноімовірні

$$H(A) = H_{\max} = \log M \ \textit{бит / симв}.$$

Величину H(A) називають **інформаційною місткістю алфавіту**. Приведемо ще приклад розрахунку ентропії джерела повідомлень, що створює текст російської мови.

Цікавитимемося ентропією, пов'язаною з незалежними появами однієї букви. У російському алфавіті є, включаючи пропуск між буквами, всього 32 букви. Якби поява всіх букв була рівноімовірна, то ентропія, пов'язана з появою однієї букви, була б рівна

$$H = \log 32 = 5 \text{ дв. од.}$$

Проте, імовірність появи різних букв не однакова. Для підрахунку безумовної імовірності появи різних букв потрібно підрахувати відносну їх частоту в типовому тексті.

Фахівці з математичної лінгвістики зробили такі підрахунки. Виявилося, що P(е)= 0,110, P(о)= 0,087, P(ф)= 0,002 і т.д.

Якщо тепер підставити імовірність появи різних букв у формулу попередньої лекції (4.14), то одержимо ентропію, пов'язану з незалежною появою однієї букви тексту:

$$H = -p(а) \log p(а) - \ldots -p(я) \log p(я)= 4{,}35 \text{ дв. од.}$$

Допущення про взаємну незалежність сусідніх букв тексту, використане в попередньому прикладі, викликає заперечення.

Граматика мови накладає певні зв'язки на сусідні букви. Щоб точніше оцінити ентропію джерела типового буквеного тексту, потрібно визначити ентропію складного експерименту, результатом якого є двовимірна (у загальному випадку n-мірна; $n \geq 2$) випадкова величина, не накладаючи умову незалежності окремих складових багатовимірної випадкової величини.

11.2. Визначення умовної ентропії H(X/Y).

Розглянемо складний експеримент XY, що складається в спільному проведенні експерименту X з наслідками $x_1, \ldots, x_M$ та експерименту Y з наслідками $y_1, \ldots, y_N$. Результатами експерименту XY є двовимірні дискретні випадкові величини x_iy_j(i=1, …, M; j=1, …, N), нас звичайно буде цікавити випадок N=M. Визначається заданим безумовний закон розподілу імовірності $P(y_j)$ компоненти y_j двовимірної величини x_iy_j й умовний закон розподілу імовірності $P(x_i/y_j)$ компоненти x_i за умови, що значення y_j задане. Помітимо, що задані закони розподілу дають вичерпний опис експерименту XY (чи випадкової величини x_iy_j). Зокрема, визначається спільний закон розподілу імовірності x_i й y_j .

$$P(x_i, y_j) = P(y_j)P(x_i/y_j), \tag{11.5}$$

Безумовний закон розподілу імовірності складової x_i

$$P(x_i) = \sum_{j=1}^{N} P(x_i, y_j), \tag{11.6}$$

і умовний закон розподілу ймовірностей складової y_j

$$P(y_j / x_i) = \frac{P(x_i, y_j)}{P(x_i)}, \tag{11.7}$$

Потрібно визначити:

1. Середню умовну ентропію H(X/Y) експерименту X за умови, що експеримент Y проведений і його результат відомий.
2. Ентропію складного експерименту XY (тобто сукупності випадкових величин XY) з урахуванням взаємозалежності складових x_i й y_j.

Сформульована задача носить загальний характер. Однак нас буде цікавити в основному два практичних додатки. По-перше, апостеріорна ентропія адресата, тобто невизначеність адресата у відношенні того, яке повідомлення було послано (у відношенні результату експерименту X) після того, як став відомий результат експерименту Y, тобто за умови, що відомо прийняте повідомлення. По-друге, ентропія, зв'язана з появою чергової букви тексту, чи чергового елементарного повідомлення, (експерименту X) за умови, що відома попередня буква тексту, чи попереднє елементарне повідомлення (експеримент Y).

Перейдемо до визначення ентропії складного експерименту при взаємозалежності подій.

5.3. Ентропія складного експерименту при взаємозалежності подій.

При визначенні ентропії і кількості інформації передбачалося, що елементи повідомлень статистично незалежні. Однак у реальних умовах незалежність елементів повідомлення - явище задоволене рідке. Наприклад, при передачі російського тексту імовірності появи окремих букв залежать від того, які букви їм передували. Наприклад, якщо передана буква "п", імовірність того, що наступної може бути "а", набагато більше, ніж імовірність того, що наступною буквою буде "р". Після букви "ъ" ніколи не очікується поява букви "н", "ы" і т.п. Така залежність між елементами утворилася історично в процесі тривалого формування сучасної російської мови. Очевидно, при визначенні ентропії й інформації в повідомленнях, елементи яких корельовано, не можна обмежуватися тільки безумовними ймовірностями елементів повідомлень. Необхідно враховувати й умовні імовірності появи елементів.

Кореляційні зв'язки між елементами повідомлень можуть поширюватися на різні групи елементів. Якщо мається корелятивний зв'язок тільки між двома сусідніми елементами, то імовірність передачі будь-якого елемента повідомлення буде залежати лише від того, який був попередній символ, тобто умовна імовірність передачі елемента x_j буде дорівнює $p(x_i/y_j)$. Сукупність умовних імовірностей $p(x_j/y_i)$ зручно записувати у вигляді матриці

$$P(X/Y)=\begin{vmatrix} p(x_1/y_1), & p(x_1/y_2), & \dots, & p(x_1/y_N) \\ p(x_2/y_1), & p(x_2/y_2), & \dots, & p(x_2/y_N) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ p(x_M/y_1), & p(x_M/y_2), & \dots, & p(x_M/y_N) \end{vmatrix}.$$

Якщо кореляtивні зв'язки охоплюють три елементи повідомлень, то умовна імовірність передачі елемента x_j буде дорівнює $p(x_j/y_i, z_k)$.

Більшість повідомлень у реальних умовах являють собою послідовність у який кореляtивні зв'язки поширюються на кінцеве число елементів. При достатній довжині такої послідовності з достатньою точністю можуть бути визначені імовірності й умовні імовірності появи окремих повідомлень. Мова являє типовий приклад такої послідовності. У будь-якій книзі даною мовою (крім вузькоспеціальних) частота повторення окремих букв і їхніх різних сполучень буде постійної, незалежно від змісту книги.

Нехай мається ***кореляtивний зв'язок*** тільки між сусідніми елементами. У цьому випадку ентропія елемента x_j буде визначатися умовною імовірністю $p(x_j/y_i)$.

Умовна ентропія $H(X/y_1)$ експерименту X за умови, що експеримент Y проведений і наступила подія y_1, дорівнює

$$H(X/y_1)=-\sum_{i=1}^{M}P(x_i/y_1)\log P(x_i/y_1), \tag{11.8}$$

Якщо результатом експерименту Y з'явилася подія y_2, то ентропія експерименту X визначається розподілом $P(x_i/y_2)$ і дорівнює

$$H(X/y_2)=-\sum_{i=1}^{M}P(x_i/y_2)\log P(x_i/y_2),$$

і таке інше. Ентропія $H(X/y_j)$ називається **приватною умовною ентропією**.

Умовна ентропія $H(X/y_1)$ визначає, наприклад, невизначеність, зв'язану з появою чергової букви тексту, за умови, що попередня буква була «а» (y_1=а).

Аналогічно $H(X/y_2)$ – ентропія, зв'язана з появою чергової букви за умови, що попередня буква була «б» (y_2=б) і т.д. Становить інтерес **середня умовна ентропія**, зв'язана з появою чергової букви за умови, що попередня буква відома. У загальному випадку для різних наслідків y_j експерименту Y, яким відповідають імовірності $P(y_j)$, величина умовної ентропії $H(X/y_j)$ експерименту X виходить різною. **Середнє значення величини** $H(X/y_j)$, отриманої шляхом осереднення її по всім можливим результатом y_j і називаємо **середньою умовної ентропією** експерименту X за умови, що відомий результат експерименту Y, дорівнює

$$H_2(X)=H(X/Y)=\sum_{j=1}^{N}P(y_j)H(X/y_j), \tag{11.9}$$

Таким чином, ми одержуємо відповідь на перше з поставлених питань:

$$H_2(X)=H(X/Y)=-\sum_{i=1}^{M}\sum_{j=1}^{N}P(y_j)P(x_i/y_j)\log P(x_i/y_j), \qquad (11.10)$$

або

$$H_2(X)=H(X/Y)=-\sum_{i=1}^{M}\sum_{j=1}^{N}P(x_i,y_j)\log P(x_i/y_j)$$

Ентропія $H_3(X)=H(X/YZ)$ визначаються, природно, умовним розподілом $P(x_k/y_iz_j)$ величини X при заданих значеннях $Y=y_j$ та $Z=z_i$:

$$H_3(X)=H(X/YZ)=-\sum_i\sum_j\sum_k P(x_k/y_j,z_i)P(y_i,z_j)\log P(x_k/y_j,z_i).$$

Якщо корелятивними зв'язками охоплене велике число елементів, то ентропія визначиться аналогічно.

При наявності корелятивних зв'язків між елементами ентропія повідомлень, а отже, і кількість переданої інформації *зменшуються*, причому це зменшення буде тим інтенсивніше, ніж сильніше корелятивні зв'язки і чим більше число елементів буде охоплено цими зв'язками.

Приклад. Джерело виробляє двійкові символи a_1 и a_2 з імовірностями $p(a_1)=0,2$; $p(a_2)=0,8$; $p(a_1/a_1)=0,4$; $p(a_2/a_1)=0,6$; $p(a_1/a_2)=0,15$; $p(a_2/a_2)=0,85$. Визначити умовну ентропію.

Ентропію даного джерела знаходимо, використовуючи формулу (11.6):

$$H_2(A)=-\sum_{i=1}^{2}p(a_i)\sum_{j=1}^{2}p(a_j/a_i)\log p(a_j/a_i)=$$
$$=-p(a_1)\left[p(a_1/a_1)\log p(a_1/a_1)+p(a_2/a_1)\log p(a_2/a_1)\right]-$$
$$-p(a_2)\left[p(a_1/a_2)\log p(a_1/a_2)+p(a_2/a_2)\log p(a_2/a_2)\right]=$$
$$=-0{,}2(0{,}4\log 0{,}4+0{,}6\log 0{,}6)-0{,}8(0{,}15\log 0{,}15+0{,}85\log 0{,}85)=0{,}682\text{бит}$$

Якщо би символи були б незалежними, то

$$H_1(A)=-\sum_{i=1}^{2}p(a_i)\log p(a_i)=-0{,}2\log 0{,}2-0{,}8\log 0{,}8=0{,}722\text{бит}\,.$$

Таким чином, якщо символи джерела залежні, то кожен з них несе в середньому меншу кількість інформації по порівнянню з незалежними. Фізично це зрозуміле, оскільки відомий попередній символ через наявність статистичних зв'язків зменшує невизначеність в появі подальшого символу, що і приводить до зниження його інформативності.

Знайдемо значення ентропії H(X/Y) для двох граничних випадків.

а) Експерименти X и Y статистично незалежні :

$$P(X_i/Y_j)=P(X_i), \qquad (11.11)$$

При цьому

$$H(X/Y)=\sum_j P(Y_j)\left[-\sum_i P(X_i)\log P(X_i)\right]=H(X), \qquad (11.12)$$

тобто середня умовна ентропія дорівнює безумовної. У силу незалежності X та Y значення Y не може зменшити невизначеність відносно X.

б) Результат експерименту цілком (функціонально) зв'язаний з результатом експерименту Y:

$$P(X_i/Y_j)=\begin{cases}1, i=j\\0, i\neq j\end{cases} \quad i,j=1,\ldots,M, \qquad (11.13)$$

При цьому, приймається в увагу, що $\lim\limits_{x\to 0}x\log x=0$ й $\log 1=0$

$$H(X/Y) = 0. \tag{11.14}$$

Результат експерименту Y однозначно визначає у випадку б) результат експерименту X и невизначеність відсутня.

Тому в загальному випадку

$$0 \le H(X/Y) \le H(X), \tag{11.15}$$

Для з'ясування впливу ступеня статистичного зв'язку між символами джерела на величину його ентропії розглянемо простий приклад. Хай символи двійкового джерела рівноімовірні $p(a_1)=p(a_2)=0,5$, а умовна вірогідність їх появи $p(a_1/a_1)=p(a_2/a_2)=0$; $p(a2/a1)=p(a_1/a_2)=1$. Такий розподіл імовірності означає, що на виході джерела послідовність символів має вигляд: ... $a_1a_2a_1a_2a_1a_2$... Не дивлячись на те, що джерело видає символи a_1 і a_2 з однаковою імовірністю (для незалежних символів – це максимум ентропії), наявність жорсткого зв'язку між ними (після a_1 – завжди a_2, а після a_2 – завжди a_1, і ніяк інакше) дозволяє припустити, що ентропія такого джерела рівна нулю. Це припущення підтверджується і розрахунками по формулі (11.6). Зменшення різниці між умовною імовірністю приводить до збільшення ентропії. Так, якщо $p(a_1/a_1)=p(a_2/a_2)=0,1$; $p(a_2/a_1)=p(a_1/a_2)=0,9$, то ентропія рівна 0,469 біт. При $p(a_1/a_1)=p(a_2/a_2)=0,3$; $p(a_2/a_1)=p(a_1/a_2)=0,7$ ентропія рівна 0,881 біт. І при $p(a_1/a_1)=p(a_2/a_2)=p(a_2/a_1)=p(a_1/a_2)=0,5$, коли рівноімовірні символи стають і незалежними, ентропія джерела максимальна і рівна 1 біту.

Таким чином, **величина умовної ентропії вкладена** в діапазоні між нулем і значенням безумовної ентропії. Перше має місце при жорсткому зв'язку між символами a_i і a_j. При цьому символ a_j не несе ніякої нової інформації, оскільки після появи символу a_i завжди можна однозначно передбачити, яким буде символ a_j. Друге має місце при незалежності символів a_i і a_j. В цьому випадку вся інформація в a_j є новою — в символі a_i немає навіть натяку про те, яким буде символ a_j.

11.4. Визначення ентропії складного експерименту H(X,Y).

При рішенні задач передачі інформації часто мають справу з декількома джерелами, що дають залежні повідомлення. *Сукупність повідомлень*, вироблюваних *декількома джерелами*, називають **складним повідомленням**.

Нехай мається два джерела повідомлень. Повідомлення першого джерела приймають значення x_1, x_2, ..., x_M з імовірностями $p(x_1)$, $p(x_2)$, ..., $p(x_M)$ і повідомлення другого джерела приймають значення y_1, y_2, ..., y_N з імовірностями $p(y_1)$, $p(y_2)$, ..., $p(y_N)$.

Спільну ентропію сукупності повідомлень X і Y (*ентропію об'єднання*) можна представити у виді

$$H(X,Y) = -\sum_{i=1}^{M}\sum_{j=1}^{N} P(x_i, y_j)\log P(x_i, y_j),$$

де $p(x_i, y_j)$ - імовірність спільної появи повідомлень x_i і y_j.

З огляду на, що спільна імовірність може бути представлена у виді

$$p(x_i, y_j) = p(x_i)\, p(y_i / x_i),$$

де $p(y_i / x_i)$ - умовна імовірність повідомлення y_j за умови, що надійшло повідомлення x_i, вираження для ентропії об'єднання шляхом математичних перетворень можна представити у виді

$$H(X,Y) = H(X) + \sum_{i=1}^{M} P(x_i) H(Y/x_i),$$

де $H(Y/x_i) = -\sum_{j=1}^{N} P(y_j/x_i) \log P(y_j/x_i),$ - приватна умовна ентропія, що виражає ентропію повідомлення Y за умови, що мало місце повідомлення X.

Вираження $\sum_{i=1}^{M} P(x_i) H(Y/x_i)$ являє собою усереднення $H(Y/x_i)$ по всім i=1 повідомленням x_1, x_2, ..., x_M і є середньою умовною ентропією H(Y/X) повідомлення Y за умови надходження повідомлення X. Тоді остаточно одержуємо

$$H(X,Y)= H(X) + H(Y/X).$$

Основний зміст умовної ентропії H(Y/X) полягає в тому, що вона показує, яку ентропію дають повідомлення Y, коли уже відома ентропія повідомлень X.

З очевидної рівності H(X,Y) = H(Y,X) одержимо:

$$H(X,Y) = H(Y) + H(X/Y).$$

Таким чином, спільна ентропія двох повідомлень дорівнює сумі безумовної ентропії одного з повідомлень і умовної ентропії другого повідомлення.

При деякій безлічі повідомлень X, Y, Z, ... спільна ентропія

$$H(X,Y,Z,...)=H(X)+H(Y,Z,.../X)=H(Y)+H(X,Z,.../Y)=H(Z)+H(X,Y,.../Z)=...$$

Розглянемо **основні властивості ентропії складних повідомлень** (ентропії об'єднання).

1. При *статично незалежних повідомленнях* X і Y спільна ентропія дорівнює сумі энтропій повідомлень

$$H(X,Y)= H(X) + H(Y),$$

тому що в цьому випадку H(Y/X) = H(Y).

2. При *повній статичній залежності повідомлень* X і Y спільна ентропія дорівнює безумовної ентропії одного з повідомлень, тому що в цьому випадку H(Y/X)=0 і H(X/Y)=0;

$$H(X,Y)= H(X) = H(Y).$$