ЛЕКЦІЯ 5

КОМПОЗИЦІЯ НЕЗАЛЕЖНИХ ВИПРОБУВАНЬ. ЕЛЕМЕНТАРНІ ФАКТИ З ТЕОРІЇ ІНФОРМАЦІЇ

5.1. Композиція двох випробувань

Композиція подій в теорії ймовірностей стосується операцій над подіями, які допомагають визначати нові події на основі вже існуючих.

Нехай (Ω_1, F_1, P_1) та (Ω_2, F_2, P_2) – ймовірнісні простори, що відповідають цим випробуванням, $\Omega_1 = \{E_1, E_2, \dots, E_{m_1}\}$, $\Omega_2 = \{Q_1, Q_2, \dots, Q_{m_2}\}$, F_k $(k = 1,2)$ складається з усіх підмножин множини Ω_k, а ймовірності складних подій $C_1 \subseteq \Omega_1$ і $C_2 \subseteq \Omega_2$ у випробуваннях

$$P_1(C_1) = \sum_{E_i \in C_1} P_1(E_i); P_2(C_2) = \sum_{Q_j \in C_2} P_2(Q_j);$$

задаються за допомогою ймовірностей елементарних подій $P_1(E_i)$, $i = 1, \dots, m_1$, та $P_2(Q_j)$, $j = 1, \dots, m_2$.

Зауважимо, що $P_1(\cdot)$ і $P_2(\cdot)$ – взагалі кажучи, різні ймовірнісні міри.

Композицією двох випробувань зветься складне випробування, що полягає у проведенні першого та другого випробування, елементарними подіями якого є всілякі можливі пари виду (E_i, Q_j), $i = 1, \dots, m_1$; $j = 1, \dots, m_2$. Таким чином, простір елементарних подій композицій двох випробувань є прямий добуток просторів Ω_1 і Ω_2 ,тобто

$$\Omega = \Omega_1 \text{x} \Omega_2.$$

Приклад 5.1. *Перше випробування полягає у підкиданні симетричної монети,* $\Omega_1 = \{\text{О}, \text{Р}\}$, *друге випробування – підкидання грального кубика,* $\Omega_2 = \{Q_1, Q_2, \dots, Q_6\}$. *Побудувати простір композиції цих випробувань.*

Розв'язання. Простір елементарних подій композиції цих двох випробувань –

$$\Omega = \{(\text{О}, Q_1), \dots, (\text{О}, Q_6), (\text{Р}, Q_1), \dots, (\text{Р}, Q_6)\}.$$

Як складні події для композиції розглядаються усі можливі підмножини множини Ω тобто усілякі теоретико-множинні суми пар виду (E_i, Q_j) (та неможлива подія $\varnothing$). Таким чином Ω і F для композиції побудовані.

Для повного завдання ймовірнісного простору композиції необхідно ще означити ймовірнісну міру $\text{P}(\cdot)$ на F. Але, взагалі кажучи, тільки позначенням $P_1(\cdot)$ і $P_2(\cdot)$ це зробити неможливо.

Відповідь. Композицію випробувань побудовано.

Розглянемо окремий випадок композиції.

Два випробування називаються *незалежними*, якщо для всіх пар ймовірність $P(\cdot)$ у композиційному просторі визначається співвідношенням:

$$P(E_i, Q_j) = P_1(E_i) \cdot P_2(Q_j). \tag{5.1}$$

Випробування *незалежні*, якщо неоднозначність результату кожного випробування визначена незв'язними між собою групами випадкових факторів, результат одного випробування не впливає на результат другого.

Ймовірність складної події C для композиції незалежних випробувань означимо як

$$P(C) = \sum_{(E_i,Q_j)\in C} P(E_i, Q_j) = \sum_{(E_i,Q_j)\in C} \left(P_1(E_i) \cdot P_2(Q_j)\right). \quad (5.2)$$

Ймовірність P, що так побудована, називається *прямим добутком* ймовірностей P_1та P_2 та позначається

$$P = P_1 \cdot P_2. \quad (5.3)$$

Загальна структура незалежних подій у композиційному просторі (Ω, F, P) має наступний вигляд .

Розглянемо подію $A \in F$

$$A = \bigcup_{l=1}^{s_1} \bigcup_{j=1}^{m_2} (E_l, Q_j).$$

Тобто внаслідок першого випробування відбувається одна з подій $E_{i_1}, E_{i_2}, \dots, E_{i_{s_1}}$, $s_1 \le m_1$ (подія $A_1 = \bigcup_{l=1}^{s_1} E_{i_l} \in F_1$), а внаслідок другого випробування відбувається усе що завгодно, тобто одна з подій $Q_1, Q_2, \dots, Q_{m_2}$ (ймовірна подія $\Omega_2 = \bigcup_{j=1}^{m_2} Q_j$).

Нехай складна подія $B \in F$

$$B = \bigcup_{i=1}^{m_1} \bigcup_{t=1}^{s_2} (E_i, Q_t),$$

тобто внаслідок першого випробування відбувається усе що завгодно, тобто одна з подій $E_1, E_2, \dots, E_{m_2}$ (вірогідна подія $\Omega_1 = \bigcup_{i=1}^{m_1} E_i$), а внаслідок другого випробування відбувається одна з подій $Q_{j_1}, Q_{j_2}, \dots, Q_{j_{s_2}}$, $s_2 \le m_2$ (подія $B_2 = \bigcup_{t=1}^{s_2} Q_{i_l} \in F_2$).

Можна коротко записати

$$A = A_1 \mathrm{x} \Omega_2, B = \Omega_1 \mathrm{x} B_2 \cdot$$

Переконаємось, що події A і B незалежні:

$$P(A \cap B) = P(\mathrm{A}) \cdot P(\mathrm{B}).$$

Дійсно,

$$P(A) = \sum_{l=1}^{s_1} \sum_{j=1}^{m_2} P(E_{i_l}, Q_j) = \sum_{l=1}^{s_1} \sum_{j=1}^{m_2} \left(P_1(E_{i_l}) \cdot P_2(Q_j)\right) = \sum_{l=1}^{s_1} P_1(E_{i_l}) \cdot$$

$$\cdot \sum_{j=1}^{m_2} P_2(Q_j) = \sum_{l=1}^{s_1} P_1(E_{i_l}) = P_1(A_1).$$

Аналогічно можна переконатись, що

$$P(B) = \sum_{l=1}^{s_1} P_2(Q_{j_t}) = P_2(B_2).$$

Очевидно,

$$A \cap B = \sum_{l=1}^{s_1} \sum_{t=1}^{s_2} (E_{i_l}, Q_{j_t}) = A_1 \mathrm{x} B_2.$$

Тоді

$$P(A\cap B)=\sum_{l=1}^{s_1}\sum_{t=1}^{s_2}P\left(E_{i_l},Q_{j_t}\right)=\sum_{l=1}^{s_1}\sum_{t=1}^{s_2}\left(P\left(E_{i_l}\right)\cdot P\left(Q_{j_t}\right)\right)=$$
$$\sum_{l=1}^{s_1}P\left(E_{i_l}\right)\cdot\sum_{t=1}^{s_2}P\left(Q_{j_t}\right)=P(A)\cdot P(B)=P_1(A_1)\cdot P_2(B_2),$$
що і треба було довести.

Застосування композиції подій

1. Визначення складних подій: Композиція допомагає створювати складніші події на основі простіших подій. Наприклад, у випадкових експериментах ми можемо цікавитися як комбінаціями успіхів, так і їхніми взаємними виключеннями.

2. Моделювання систем: У багатьох задачах ймовірності, наприклад в аналізі надійності систем, використовують композицію подій для визначення ймовірності відмови системи (коли відмовляє один чи більше компонентів).

3. Умовні ймовірності: Часто цікаво знати ймовірність однієї події за умови, що сталася інша подія. Це також можна розглядати як композицію подій, де друга подія є умовою для першої.

Застосування в реальному житті

1. Аналіз ризиків: Композиції подій використовуються для обчислення ймовірностей різних сценаріїв ризиків, наприклад, об'єднання подій може показати ймовірність виникнення катастрофічного випадку при настанні кількох ризиків одночасно.

2. Теорія ігор: Композиції подій застосовуються для моделювання стратегій та визначення ймовірностей різних результатів гри.

5.2. Композиція n незалежних випробувань

Зовсім аналогічно будується схема композиції n незалежних випробувань.

Нехай здійснюється n випробувань із скінченими просторами елементарних подій

$$\Omega_i=\left\{E_1^i,E_2^i,\dots,E_{m_i}^i\right\},i=\overline{1,n}.$$

Складними подіями у кожному i-му випробуванні є всілякі підмножини множини Ω_i, що визначає F_i. Задані на кожному F_i ймовірнісні міри $P_i(.)$. І, таким чином, маємо n ймовірнісних просторів $(\Omega_i,\mathrm{F}_i,P_i)$, $i=\overline{1,n}$.

Результат спільного проведення n випробувань вважаємо результатом проведення одного складного випробування (*композиції*).

Композиційний простір Ω будуємо як прямий добуток просторів Ω_i, $i=\overline{1,n}$,

$$\Omega=\Omega_1\mathrm{x}\Omega_2\mathrm{x}\dots\mathrm{x}\Omega_n,$$

тобто у якості елементарних подій композиції беруться всілякі n-ки виду

$$\left(E_{j_1}^1,E_{j_2}^2,\dots,E_{j_n}^n\right);\ j_1=\overline{1,m_1};\dots;j_n=\overline{1,m_n}.$$

Складними подіями в композиційному просторі є теоретико-множинні суми вказаних n-ок.

Випробування називаються *незалежними*, якщо ймовірність кожної n-ки визначається співвідношенням

$$P\left(E_{j_1}^1,E_{j_2}^2,\dots,E_{j_n}^n\right)=\prod_{i=1}^n P_i\left(E_{j_i}^i\right).\tag{5.4}$$

Ймовірності складних подій в композиційному просторі визначаються звичним для дискретних просторів способом – як сума ймовірностей n-ок, що належать даній складній події.

Побудована таким чином ймовірність P для подій композиційного простору Ω називається *прямим добутком ймовірностей* P_i і позначається

$$P = P_1 \cdot P_2 \cdot \ldots \cdot P_n. \qquad (5.5)$$

В побудованому просторі (Ω, F, P) виділимо клас подій A, що називаються *прямокутниками*. Нехай $A_i \in F_i$, $i = \overline{1,n}$. Прямокутник

$$A = A_1 \mathrm{x} A_2 \mathrm{x} \ldots \mathrm{x} A_n \qquad (5.6)$$

містить ті та тільки ті n-ки $E = \left(E_{j_1}^1, E_{j_2}^2, \ldots, E_{j_n}^n\right)$, у яких $E_{j_i}^i \in A_i$, $i = \overline{1,n}$.

З означення ймовірності для композиції незалежних випробувань виходить, що ймовірність прямокутника A дорівнює:

$$P(A) = \sum_{E \in A} P(E) = \sum_{E_{j_1}^1 \in A_1} \ldots \sum_{E_{j_n}^n \in A_n} \left(P_1\left(E_{j_1}^1\right) \cdot \ldots \cdot P_n\left(E_{j_n}^n\right)\right)$$

$$= \sum_{E_{j_1}^1 \in A_1} P_1\left(E_{j_1}^1\right) \cdot \ldots \cdot \sum_{E_{j_n}^n \in A_n} P_n\left(E_{j_n}^n\right) = P_1(A_1) \cdot P_2(A_2) \cdot \ldots \cdot P_n(A_n).$$

Загальна структура незалежних подій в композиційному просторі задається прямокутниками виду

$$A_i' = \Omega_1 \mathrm{x} \Omega_2 \mathrm{x} \ldots \mathrm{x} \Omega_{i-1} \mathrm{x} \mathrm{A}_i \mathrm{x} \Omega_{i+1} \mathrm{x} \ldots \mathrm{x} \Omega_n,$$

тобто подіями, які можна інтерпретувати наступним чином: в i-му випробуванні відбувається подія A_i, а в інших випробуваннях відбувається що завгодно.

З одержаного раніш співвідношення для ймовірності прямокутника маємо

$$P(A_i') = P_i(A_i).$$

Тому що $A = A_1 \mathrm{x} A_2 \mathrm{x} \ldots \mathrm{x} A_n = \bigcap_{i=1}^n A_i'$, то з того ж співвідношення виходить

$$P\left(\bigcap_{i=1}^n A_i'\right) = \prod_{i=1}^n P(A_i'),$$

тобто при різних і події A_i' незалежні.

Через те що у прямокутниках A_i' в якості A_i можуть фігурувати Ω_i, то ясно, що мова йде про незалежність у сукупності.

5.3. Елементарні факти з теорії інформації

Теорія інформації — це розділ математики, який досліджує процеси зберігання, перетворення і передачі інформації. Теорія інформації тісно пов'язана з такими розділами математики як теорія ймовірностей і математична статистика. Вона пов'язана з інформаційною ентропією, комунікаційними системами, криптографією, корекцією помилок і іншими важливими областями.

Основні розділи теорії інформації — кодування джерела (стискаюче кодування) і канальне (завадостійке) кодування. Теорія інформації тісно пов'язана з інформаційною ентропією, комунікаційними системами, криптографією і іншими суміжними дисциплінами.

Аксіоми теорії інформації

- Інформація є лише там, де функціонують пристрої керування.
- Інформація зберігається і передається тільки на матеріальному носії.
- Інформація має ідеальний характер.
- Інформація має різні форми.

Базові закони теорії інформації

Закон 1: на отримання інформації будь-яка кібернетична система витрачає не менше деякої мінімальної кількості енергії.

Закон 2: кількість інформації, яку отримує кібернетична система в процесі розпізнавання після прийняття певного сигналу, дорівнює логарифму при основі m від кількості варіантів вибору, що передували розпізнаванню.

Закон 3: що меншою є ймовірність завершення якогось випробування з певним результатом, то більше інформації для будь-якої кібернетичної системи несе саме цей результат, і навпаки.

Закон 4: будь-які сигнали, отримані кібернетичною системою, впливають на цю систему.

Властивості інформації

Адекватність — відповідність відображуваному об'єктові або реальному стану об'єктивної дійсності при відсутності прихованих помилок у такій інформації.

Повнота — властивість інформації, що дозволяє характеризувати об'єкт вичерпним для споживача засобом, що й надає можливість ухвалювати на основі такої інформації управлінські рішення.

Релевантність — відповідність потребам споживача, що характеризує, наскільки інформація сприяє досягненню поставлених перед споживачем цілей і завдань.

Доступність — можливість одержання будь-яким споживачем.

Актуальність — відповідність інформації теперішньому моменту часу.

Коректність — властивість, що полягає в такому її зображенні, щоб інформація однозначно сприймалася всіма її споживачами.

Захищеність — неможливість несанкціонованого доступу й цілеспрямованого спотворення інформації.

Ергономічність — достатність обсягу і форми інформації для даного споживача.

Власна інформація

Математична теорія інформації ґрунтується на теорії ймовірності й статистиці, і вимірює інформацію за допомогою декількох кількостей інформації (quantities of information). Застосовувану в наступних формулах одиницю інформаційної ентропії визначає вибір логарифмічної основи. Найзвичнішою одиницею інформації є *біт*, що ґрунтується на двійковому логарифмі. До інших одиниць належать *нат*, що ґрунтується на натуральному логарифмі, та *гартлі*, що ґрунтується на десятковому логарифмі. Надалі вираз $p \cdot log\,(p)$ $(p \cdot log_2(p); p \cdot lg(p)$ або $p \cdot ln\,(p))$, коли p є нулем вважається рівним нулю. Це є виправданим, оскільки для будь-якої логарифмічної основи $\lim_{p\to 0+}(p\cdot log\,(p))=0$.

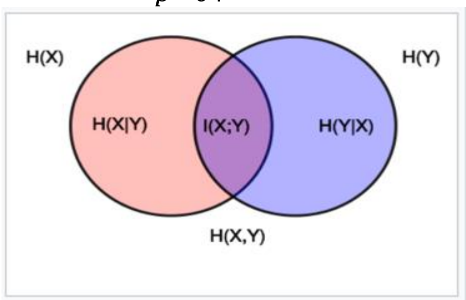

Рис. 5.1. Інформаційна діаграма

Міра інформаційного вмісту повідомлення, називається *власною інформацією* (*self-information*) або «*несподіваністю*»

Кількість власної інформації, яка міститься в ймовірнісній події, залежить лише від ймовірності цієї події: що меншою є її ймовірність, то більшою є власна інформація, пов'язана з отриманням інформації про те, що ця подія дійсно відбулася.

Міра власної інформації є додатною та адитивною. Якщо подія C є перетином двох незалежних подій A та B, то кількість інформації при оголошенні про те, що подія C сталася, дорівнює сумі кількостей інформації в оголошеннях про подію A та подію B відповідно:

$$I(A\cap B)=I(A\cdot B)=I(A)+I(B).$$

Із врахуванням цих властивостей, власною інформацією $I(x_n)$, пов'язаною з подією (повідомленням) x_n з ймовірністю $p(x_n)$, є

$$I(x_n)=log\left(\frac{1}{p(x_n)}\right)=-log(p(x_n)). \qquad (5.7)$$

Це визначення відповідає наведеним вище умовам. У формулі (5.4) не вказано основу логарифма: при застосуванні основи 2 одиницями $I(x_n)$ будуть біти, при застосуванні логарифма за основою e одиницею буде нат, длялогарифма за основою 10 одиницею буде гартлі.

Як швидке пояснення, кількістю інформації, пов'язаною з випадінням 4 аверсів (або будь-якого конкретного виходу) в 4 послідовних підкиданнях монети, буде 4 біти (ймовірність $\frac{1}{16}$), а кількістю інформації, пов'язаною з отриманням результату, відмінного від вказаного, буде 0,09 біт (ймовірність $\frac{15}{16}$).

Приклад 5.2. *При підкиданні монети шансом «реверсу» є* 0,5. *Обчислити кількість інформації.*

Розв'язання. Коли проголошується, що справді випав «реверс», то це дає кількість інформації

$$I(\text{«реверс»})=log_2\left(\frac{1}{0{,}5}\right)=log_2(2)=1 \text{ біт інформації.}$$

Відповідь. I(«реверс») $=1$ біт.

Приклад 5.3. *При викиданні правильного грального кубика ймовірність «четвірки» становить* $\frac{1}{6}$. *Обчислити кількість інформації.*

Розв'язання. Коли проголошується, що випала «четвірка», то кількістю власної інформації є

$$I(\text{«четвірка»}) = log_2\left(\frac{1}{\frac{1}{6}}\right) = log_2(6) = 2{,}585 \text{ бітів власної інформації.}$$

Відповідь. $I(\text{«четвірка»}) = 2{,}585$ бітів.

Приклад 5.4. *При незалежному викиданні двох гральних кубиків обчислити кількість інформації, пов'язаної з подією* $\{\text{викидання } 1 = \text{«два» і викидання } 2 = \text{«чотири»}\}$.

Розв'язання. При незалежному викиданні двох гральних кубиків кількість інформації, пов'язаної з $\{\text{викидання } 1 = \text{«два» і викидання } 2 = \text{«чотири»}\}$, дорівнює

$$I(\text{«викиданням 1 є два і викиданням 2 є чотири»}) =$$
$$log_2\left(\frac{1}{P(\text{викидання } 1=\text{«два» і викидання } 2=\text{«чотири»})}\right) = log_2\left(\frac{1}{\frac{1}{36}}\right) = log_2(36) =$$
$$5{,}170 \text{ бітів.}$$

Також шукана кількість інформації дорівнює сумі окремих кількостей власної інформації, пов'язаних із $\{\text{викидання } 1 = \text{«два»}\}$ і $\{\text{викидання } 2 = \text{«чотири»}\}$; а саме, $2{,}585 + 2{,}585 = 5{,}170$ бітів.

Відповідь. $(I\text{«викиданням 1 є два і викиданням 2 є чотири»}) = 5{,}170$ бітів.

Приклад 5.5. *При незалежному викиданні двох гральних кубиків обчислити кількість інформації, пов'язаної з подією* $\{\text{Сумою двох гральних кубиків є п'ять}\}$.

Розв'язання. В тій самій ситуації з двома гральними кубиками ми можемо розглядати інформацію, присутню в твердженні $\{\text{Сумою двох гральних кубиків є п'ять}\}$

$$I(\text{«Сумою викидів 1 та 2 є п'ять»}) = log_2\left(\frac{1}{P(\text{викиди 1 та 2 дають у сумі п'ять})}\right) =$$
$$log_2\left(\frac{1}{\frac{4}{36}}\right) = log_2(9) = 3{,}17 \text{ бітів.}$$

$P(\text{викиди 1 та 2 дають у сумі п'ять}) = \frac{4}{36}$, оскільки існує чотири варіанти з 36 можливих, щоби два кубики давали в сумі 5. Це показує, що складніші або неоднозначніші події теж можуть давати інформацію.

Відповідь. $I(\text{«Сумою викидів 1 та 2 є п'ять»}) = 3{,}17$ бітів.

Інформаційна ентропія

Розглянемо два ймовірнісних простори із скінченою кількістю елементарних подій (кожний задамо таблицею, у якої у верхньому рядку вказані елементарні події, у нижньому – ймовірності цих подій):

$$\begin{Bmatrix} x_1 & x_2 & x_3 & x_4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{Bmatrix}, \begin{Bmatrix} x_1' & x_2' & x_3' & x_4' \\ 0{,}01 & 0{,}02 & 0{,}96 & 0{,}01 \end{Bmatrix}.$$

Очевидно, що ступінь невизначеності результату випробування (до випробування) у першому ймовірнісному просторі вище, ніж у другому.

Дійсно, до випробування у першому просторі ні одній з подій не можна віддати перевагу, а у другому майже напевно відбудеться подія x_3'.

Виникає питання: чи можна якимось чином міряти ступінь невизначеності, що властива ймовірнісному простору. Першим, хто запропонував вираз для міри невизначеності, був К. Шеннон.

Нехай деякий ймовірнісний простір повідомлень Ω задано таблицею

$$\begin{Bmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p(x_1) & p(x_2) & p(x_3) & \dots & p(x_n) \end{Bmatrix},$$

де $x_i(i = \overline{1,n})$ — повідомлення з цього простору, $p(x_i)$ — ймовірність кожного можливого значення x_i, причому $\sum_{i=1}^{n} p(x_i) = 1$.

Таким чином, Ω може приймати лише окремі, чітко визначені значення, і кожне з цих значень має певну ймовірність, тому її надалі можна вважати *дискретною випадковою величиною*, яка полягає у отриманні якогось з повідомлень x_i $(i = \overline{1,n})$ або символів одного повідомлення з ймовірністю $p(x_i)$. Позначимо $\Omega = X$.

Для вимірювання ступеню невизначеності, що притаманна даному простору використовується інформаційна ентропія.

Інформаційна ентропія випадкової події — це математичне сподівання (середнє очікуване значення) її власної інформації.

$$H(X) = M(I(X)), \tag{5.8}$$

тобто, для дискретної випадкової величини X з ймовірностями $p(x_i)$ вона визначається як:

$$H(X) = -\sum_{i=1}^{n}\left(p(x_i) \cdot log_2\left(p(x_i)\right)\right). \tag{5.9}$$

Пояснення

1. Інформаційна ентропія враховує всі можливі значення випадкової величини та їх ймовірності. Чим більше значень і чим рівномірніше розподілені ймовірності, тим вища ентропія.
2. Використання логарифма в основі 2 означає, що ентропія вимірюється в бітах. Це зручно для цифрових систем і кодування інформації.
3. Від'ємний знак гарантує, що значення ентропії є додатним або нульовим, оскільки ймовірності лежать в інтервалі від 0 до 1.

Приклад 5.6. *Підкидається монета. Обчислити інформаційну ентропію.*

Розв'язання. Якщо підкидається монета (ймовірності орла і решки рівні 0,5) обчислити інформаційну ентропію, ентропія дорівнює 1 біту, оскільки обидва результати є однаково ймовірними:

$$H(X) = -(0{,}5 \cdot log_2(0{,}5) + 0{,}5 \cdot log_2(0{,}5)) = 1 \text{ біт}.$$

Відповідь. $H(X) = 1$ біт.

Приклад 5.7. *Підкидається шестигранний кубик. Обчислити інформаційну ентропію.*

Розв'язання. Якщо підкидається шестигранний кубик (ймовірність кожного результату дорівнює $\frac{1}{6}$), ентропія дорівнює:

$$H(X) = -6 \cdot \left(\frac{1}{6} \cdot log_2\left(\frac{1}{6}\right)\right) \approx 2{,}585 \text{ бітів}.$$

Відповідь. $H(X) = 2{,}585$ бітів.

Ентропії притаманні такі простіші **властивості**:

1. *Якщо ймовірнісний простір не має невизначеності, тобто якесь* $p(x_i) = 1$*, а інші дорівнюють нулеві, то ентропія дорівнює нулеві (очевидно).*

2. *Коли елементарні події рівноймовірні, тобто* $p(x_i) = \frac{1}{n}, i = \overline{1,n}$*, тоді ентропія приймає максимальне значення.*

Для **доведення** треба вирішити задачу максимізації $H(X)$ за умови $\sum_{i=1}^{n} p(x_i) = 1$. Для цього використаємо метод невизначених множників Лагранжа: будемо шукати екстремум функції

$$F = -\sum_{i=1}^{n}(p(x_i) \cdot log_2(p(x_i))) + \lambda \cdot (\sum_{i=1}^{n}(p(x_i)) - 1).$$

Диференцюємо цю функцію по p_i та прирівнюємо частинні похідні до нуля. Одержуємо систему рівнянь

$$-log_2(p(x_i)) - log_2(e) + \lambda = 0,\ i = \overline{1,n}.$$

або

$$log_2(p(x_i)) = \lambda - log_2(e),\ i = \overline{1,n}.$$

Звідси виходить, що екстремум (перевірте, що це максимум) досягається при рівних між собою $p(x_i)$. З умови нормування $\sum_{i=1}^{n} p(x_i) = 1$ маємо $p_i = \frac{1}{n}, i = \overline{1,n}$.

За одиницю невизначеності знову приймається біт.

<u>Застосування</u>

Інформаційна ентропія використовується в багатьох областях, включаючи:

Теорію інформації: Для вимірювання кількості інформації та оптимального кодування повідомлень.

Криптографію: Для оцінки безпеки шифрів та випадковості ключів.

Машинне навчання: Для оцінки невизначеності моделей та побудови дерев прийняття рішень.

Стиснення даних: Для визначення оптимальних алгоритмів стиснення.

Інформаційна ентропія є фундаментальною концепцією в розумінні передачі інформації та її ефективного кодування.

<u>Інформаційна ентропія композиції випробувань</u>

Розглянемо два ймовірнісних простори:

$$\Omega_1 = \begin{Bmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{Bmatrix}, \Omega_2 = \begin{Bmatrix} y_1 & y_2 & y_3 & \dots & y_m \\ q_1 & q_2 & q_3 & \dots & q_m \end{Bmatrix}.$$

Проведемо композицію двох випробувань. Для композиційного простору таблиця має наступний вигляд:

$$\Omega_1 \mathrm{x} \Omega_2 = \begin{Bmatrix} (x_i, y_j) \\ P(x_i, y_j) \end{Bmatrix}, i = \overline{1,n}; j = \overline{1,m}.$$

Максимальна ентропія композиційного ймовірнісного простору досягається тоді, коли випробування незалежні (хоча б на основі якісного аналізу). Обчислимо ентропію композиційного простору для випадку незалежних випробувань:

$$H(\Omega_1 \mathrm{x} \Omega_2) - \sum_{i=1}^{n}\sum_{j=1}^{m}\left(p_i \cdot q_j \cdot log_2\left(p_i \cdot q_j\right)\right) =$$

$$= -\sum_{i=1}^{s_1}\sum_{j=1}^{s_2}\left(p_i \cdot q_j \cdot log_2(p_i)\right) -$$

$$-\sum_{i=1}^{s_1}\sum_{j=1}^{s_2}\left(p_i \cdot q_j \cdot log_2\left(q_j\right)\right) = -\sum_{i=1}^{s_1}\left(p_i \cdot log_2(p_i)\right) -$$

$$-\sum_{j=1}^{s_2}\left(q_j \cdot log_2\left(q_j\right)\right) = H(\Omega_1) + H(\Omega_2).$$

Таким чином, для композиції двох випробувань маємо:

$$0 \le H(\Omega_1 \mathrm{x} \Omega_2) \le H(\Omega_1) + H(\Omega_2), \qquad (5.10)$$

або, враховуючи, що $\Omega_1 = X;\ \Omega_2 = Y$.

$$0 \le H(X, Y) \le H(X) + H(Y),$$

де $H(X, Y)$ – спільна ентропія.

Спільна ентропія

Спільну ентропію (joint entropy) двох дискретних випадкових змінних X та Y визначають як ентропію їхнього спільного розподілу:

$$H(X, Y) = M(-log_2(p(x, y))) = -\sum_{x,y}\left(p(x, y) \cdot log_2\left(p(x, y)\right)\right). \qquad (5.11)$$

Якщо X та Y є незалежними, то ця спільна ентропія є просто сумою їхніх окремих ентропій.

Умовна ентропія

За заданого конкретного значення випадкової змінної Y *умовну ентропію* X *за* $Y = y$ визначено як

$$H(X/y) = M(-log_2(p(x/y))) = -\sum_{x \in X}(p(x/y) \cdot log_2(p(x/y))). \qquad (5.12)$$

де $p(x/y) = \frac{p(x,y)}{p(y)}$ є умовною ймовірністю x за заданого y.

Умовну ентропію (conditional entropy) X за заданого Y, що також називають *ухильністю* (equivocation) X від Y, задають як

$$H(X/Y) = M(H(X/y)) = -\sum_{y \in Y}(p(y) \cdot \sum_{x \in X}(p(x/y) \cdot log_2(p(x/y)))) =$$

$$\sum_{x,y}\left(p(x, y) \cdot log_2\left(\frac{p(y)}{p(x,y)}\right)\right). \qquad (5.13)$$

Вона використовує умовне математичне сподівання з теорії ймовірності.

Базовою властивістю умовної ентропії є те, що

$$H(X/Y) = H(X, Y) - H(Y). \qquad (5.14)$$

Приклад 5.8. *Визначити інформаційну ентропію повідомлення, ймовірності появи символів якого такі:* $p(x_1) = p(x_2) = p(x_3) = p(x_4) = 0{,}01;\ p(x_5) = 0{,}96$.

Розв'язання. Складемо випадкову величину X появи символів повідомлення (табл. 5.1):

Таблиця 5.1. *Ряд розподілу випадкової величини X*

x_i	x_1	x_2	x_3	x_4	x_5
$p(x_i)$	0,01	0,01	0,01	0,01	0,96

Підставляємо значення з табл. 5.1 у формулу (5.9) для обчислення ентропії. Розрахункова таблиця представлена в табл. 5.2.

Таблиця 5.2. *Розрахункова таблиця до прикладу* 5.8

x_i	x_1	x_2	x_3	x_4	x_5	
$p(x_i)$	0,01	0,01	0,01	0,01	0,96	$H(X)$
$-(p(x_i) \cdot log_2(p(x_i)))$	0,066439	0,066439	0,066439	0,066439	0,056538	0,322292

Відповідь. $H(X) = 0{,}322292$ біта.

Приклад 5.9. *Визначити ентропію повідомлення «$aaaabbcddd$».*

Розв'язання. Складемо випадкову величину X появи символів повідомлення (табл. 5.3).

Всього літер – 10; a – 4; b – 2; c – 1; d – 3.

Тобто, ймовірності відповідно дорівнюють:

$$p(a) = 0{,}4;\ p(b) = 0{,}2;\ p(c) = 0{,}1;\ p(d) = 0{,}3.$$

Таблиця 5.3. *Ряд розподілу випадкової величини X*

x_i	a	b	c	d
$p(x_i)$	0,4	0,2	0,1	0,3

Підставляємо значення з табл. 5.3 у формулу (5.9) для обчислення ентропії. Розрахункова таблиця представлена в табл. 5.4.

Таблиця 5.4. *Розрахункова таблиця до прикладу* 5.9

x_i	a	b	c	d	
$p(x_i)$	0,4	0,2	0,1	0,3	$H(X)$
$-(p(x_i) \cdot log_2(p(x_i)))$	0,528771	0,464386	0,332193	0,52109	1,846439

Відповідь. $H(X) = 1{,}846439$ біта.

Приклад 5.10. *Визначити ентропію двійкового джерела інформації, якщо* $p(0) = 0{,}1$, *а* $p(1) = 0{,}9$.

Розв'язання. Джерело представляло собою булеву функцію від однієї змінної.

Складемо випадкову величину X появи символів повідомлення (табл. 5.5).

Таблиця 5.5. *Ряд розподілу випадкової величини X*

x_i	0	1
$p(x_i)$	0,1	0,9

Підставляємо значення з табл. 5.5 у формулу (5.9) для обчислення ентропії. Розрахункова таблиця представлена в табл. 5.6.

Таблиця 5.6. *Розрахункова таблиця до прикладу* 5.10

x_i	0	1	
$p(x_i)$	0,1	0,9	$H(X)$
$-(p(x_i) \cdot log_2(p(x_i)))$	0,332193	0,136803	0,468996

Відповідь. $H(X) = 0{,}468996$ біта.

Приклад 5.11. *Визначити ентропію джерела інформації, якщо* $p(00) = 0{,}01$; $p(01) = 0{,}09$; $p(10) = 0{,}09$; $p(11) = 0{,}81$.

Розв'язання. Дане джерело інформації математично представляє собою булеву функцію від двох змінних.

Складемо випадкову величину X появи символів повідомлення (табл. 5.7).

Таблиця 5.7. *Ряд розподілу випадкової величини X*

x_i	00	01	10	11
$p(x_i)$	0,01	0,09	0,09	0,81

Підставляємо значення з табл. 5.7 у формулу (5.9) для обчислення ентропії. Розрахункова таблиця представлена в табл. 5.8.

Таблиця 5.8. *Розрахункова таблиця до прикладу* 5.11

x_i	00	01	10	11	
$p(x_i)$	0,01	0,09	0,09	0,81	$H(X)$
$-(p(x_i) \cdot log_2(p(x_i)))$	0,066439	0,312654	0,312654	0,246245	0,937991

Відповідь. $H(X) = 0{,}937991$ біта.

Приклад 5.12. *Для двійкового джерела інформації визначити ймовірності появи «0» та «1», при яких ентропія максимальна.*

Розв'язання. Складемо випадкову величину X появи символів повідомлення (табл. 5.9).

Таблиця 5.9. *Ряд розподілу випадкової величини X*

x_i	0	1
$p(x_i)$	p	$1-p$

Підставляємо значення з табл. 5.9 у формулу (5.9) для обчислення ентропії.

$$H(X) = -p \cdot log_2(p) - (1-p) \cdot log_2(1-p).$$

Диференціюємо функцію по p.

$$\frac{dH(X)}{dp} = -log_2(p) - \frac{1}{ln(2)} + log_2(1-p) + \frac{1}{ln(2)}$$
$$= log_2(1-p) - log_2(p).$$

Прирівнюємо до нуля.

$$log_2(1-p) - log_2(p) = 0;$$
$$log_2(1-p) = log_2(p);$$
$$1 - p = p;$$
$$p = 0{,}5.$$

Перевірка:

$$\frac{d^2H(X)}{dp^2} = -\frac{1}{(1-p)\cdot ln(2)} - \frac{1}{p\cdot ln(2)}.$$

Підставляємо $p = 0{,}5$ і бачимо, що $\frac{d^2H(X)}{dp^2}$ =$-\frac{4}{ln(2)} < 0$, отже, при $p = 0{,}5$ ентропія максимальна.

Відповідь. $p = 0{,}5$.

Приклад 5.13. *Визначити ентропії $H(X)$, $H(Y)$, $H(X,Y)$, $H(X/Y)$, якщо задана матриця станів системи, яка об'єднує джерела X, Y:*

$$P(X \cdot Y) = \begin{pmatrix} 0{,}3 & 0{,}1 & 0 \\ 0 & 0{,}2 & 0{,}2 \\ 0 & 0 & 0{,}2 \end{pmatrix}.$$

Розв'язання. Складемо випадкову величину Y появи символів повідомлення (табл. 5.10).

Таблиця 5.10. *Ряд розподілу випадкової величини Y*

y_i	y_1	y_2	y_3
$p(y_i)$	0,3	0,3	0,4

Підставляємо значення з табл. 5.10 у формулу (5.9) для обчислення ентропії.

Таблиця 5.11. *Розрахункова таблиця* 1 *до прикладу* 5.13

y_i	y_1	y_2	y_3	
$p(y_i)$	0,3	0,3	0,4	$H(Y)$
$-(p(y_i) \cdot log_2(p(y_i)))$	0,5210897	0,5210897	0,5287712	1,57095059

Отже, $H(Y) = 1{,}57095059$ біта.

Аналогічно отримуємо

Таблиця 5.12. *Ряд розподілу випадкової величини X*

x_j	x_1	x_2	x_3
$p(x_j)$	0,4	0,4	0,2

Таблиця 5.13. *Розрахункова таблиця* 2 *до прикладу* 5.13

x_j	x_1	x_2	x_3	
$p(x_j)$	0,4	0,4	0,2	$H(X)$
$-\left(p(x_j) \cdot log_2\left(p(x_j)\right)\right)$	0,5287712	0,5287712	0,4643856	1,52192809

Отже, $H(X) = 1{,}52192809$ біта.

$H(X \cdot Y)$ обчислюємо за формулою (5.11).

Таблиця 5.14. *Розрахункова таблиця* 3 *до прикладу* 5.13

	y_1	y_2	y_2	$-\left(p(x_j, y_i) \cdot log_2\left(p(x_j, y_i)\right)\right)$			
x_1	0,3	0,1	0	0,5210897	0,3321928	0	
x_2	0	0,2	0,2	0	0,4643856	0,4643856	
x_3	0	0	0,2	0	0	0,4643856	$H(X,Y)$
				0,5210897	0,7965784	0,9287712	2,24643934

Отже, $H(X,Y) = 2{,}24643934$ біт.

З формули (5.14) отримаємо, що

$$H(X/Y) = 2{,}24643934 - 1{,}57095059 = 0{,}67558875 \text{ біта.}$$

Відповідь. $H(X) = 1{,}52192809$ біта; $H(Y) = 1{,}57095059$ біта; $H(X,Y) = 2{,}24643934$ біта; $H(X/Y) = 0{,}67558875$ біта.

5.4. Алгоритм формування комплексних послуг на основі визначення інформаційної ентропії

Для початку введемо основні означення.

Електронна публічна послуга – це послуга, яка надається громадянам, підприємствам та іншим організаціям через інформаційно-комунікаційні технології (ІКТ), такі як інтернет та інші електронні засоби зв'язку. Ці послуги забезпечують доступ до державних ресурсів, дозволяють подавати заяви, отримувати документи, здійснювати платежі тощо, без необхідності відвідувати державні установи особисто.

Комплексна електронна публічна послуга – електронна публічна послуга, в результаті якої суб’єкту звернення на підставі однієї заяви (звернення, запиту) на отримання декількох електронних публічних послуг одним або декількома суб’єктами надання публічних (електронних публічних) послуг або в автоматичному режимі програмними засобами інформаційно-телекомунікаційних систем (у тому числі з використанням Єдиного державного веб-порталу електронних послуг) надаються декілька електронних публічних послуг.

На теперішній момент формування комплексних послуг в Україні здійснюється за принципом об’єднання за життєвою подією, або ситуацією. На відміну від комплексних послуг, під *портфелем послуг* будемо розуміти набір послуг, які об’єднуються на основі іншого принципу (окрім життєвої події або ситуації).

Ідея алгоритму полягає у тому, що для кожної пари послуг визначається кількість інформації, яку одна послуга несе для іншої. Чим більше інформації, тим більше пов’язані послуги. Тобто вони можуть утворювати комплексну послугу.

1. Нехай є дві випадкові події:

   $X =$ {Користувач замовив послугу i}

   $Y =$ {Користувач замовив послугу j}.

2. Рахуємо коефіцієнт взаємозалежності Райського

$$K_R = \frac{I(X,Y)}{H(X,Y)}, \qquad (5.15)$$

де $I(X,Y)$ – взаємна інформація статистична функція двох випадкових величин, що описує кількість інформації, що міститься в одній випадковій величині відносно іншої.

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \qquad (5.16)$$

де $H(X)$, $H(Y)$, $H(X,Y)$ – інформаційні ентропії, що дорівнюють кількості інформації з послуг X, Y, та $X \cdot Y$ відповідно.

Ентропії та спільна ентропія обчислюються за формулами (5.9) та (5.11) відповідно.

В нашому випадку ймовірності $p(x_i), p(y_j), p(x_i, y_j)$ можна трактувати як відносну частоту замовлення клієнтами послуги i, j та одночасно обох послуг. Дана інформація повинна бути заздалегідь відома.

3. За властивостями ентропії видно, що коефіцієнт взаємозалежності Райського $1 \leq K_R < \infty$ (одиниці, якщо послуги залежні і прямує до нескінченості, якщо незалежні). Таким чином, чим ближче його значення до одиниці (не відрізняється від неї більше ніж на порядок), тим сильніше пов'язані між собою послуги i та j.

Вважатимемо, що якщо $K_R \leq \varepsilon$, то дві послуги i та j утворюють комплексну послугу, якщо більше, то не утворюють.

Взагалі кажучи, значення ε обирається експериментальним шляхом, в залежності від особливостей даних.

4. В разі, якщо маємо три послуги i, j та k, то користуємось правилом транзитивності:

В разі якщо послуги i та j є комплексними та послуги j та k також є комплексними, то послуги i та k теж є комплексними. Таким чином, i, j та k утворюють одну спільну комплексну послугу.

В разі якщо послуги i та j є комплексними та послуги i та k також є комплексними, то з цього не випливає, що послуги j та k теж є комплексними. Таким чином, i, j та k утворюють дві різні комплексні послуги.

Приклад 5.14. *Нехай є дві окремі послуги. Встановити, чи можна їх об'єднати в одну комплексну послугу.*

Розв'язання. Нехай є дві послуги X:

Таблиця 5.15. *Ряд розподілу випадкової величини* X

x_i	x_1	x_2	x_3
w_i	0,45	0,15	0,40

де x_i – документ, яку треба надати для отримання послуги X, w_i – відносна частота звернень по документ та послуга Y:

Таблиця 5.16. *Ряд розподілу випадкової величини* Y

y_k	y_1	y_2	y_3
w_k	0,35	0,35	0,30

де y_k – документ, яку треба надати для отримання послуги Y, w_k – відносна частота звернень по документ.

Одночасно послуга $X \cdot Y$ розглядається, як система двох випадкових величин:

Таблиця 5.17. *Ряд розподілу випадкової величини* $X \cdot Y$

	x_1	x_2	x_3
y_1	0,15	0,05	0,10
y_2	0,14	0,24	0,02
y_3	0,12	0,15	0,03

$$H(X) = -(0{,}45 \cdot log_2(0{,}45) + 0{,}15 \cdot log_2(0{,}15) + 0{,}40 \cdot log_2(0{,}40)) =$$
$$= 0{,}39062395106447 + 0{,}054805219444381 + 0{,}30258831894641 =$$
$$= 0{,}74801748945526.$$

$$H(Y) = -(0{,}35 \cdot log_2(0{,}35) + 0{,}35 \cdot log_2(0{,}35) + 0{,}30 \cdot log_2(0{,}30)) =$$
$$= 0{,}23108820773979 + 0.23108820773979 + 0{,}17271499274803 =$$
$$= 0{,}63489140822761.$$
$$H(X,Y) = -(0{,}15 \cdot log_2(0{,}15) + 0{,}05 \cdot log_2(0{,}05) + 0{,}10 \cdot log_2(0{,}10)$$
$$+ 0{,}14 \cdot log_2(0{,}14) + 0{,}24 \cdot log_2(0{,}24) + 0{,}02 \cdot$$
$$\cdot log_2(0{,}02) + 0{,}12 \cdot$$
$$\cdot log_2(0{,}12) + 0{,}15 \cdot log_2(0{,}15) + 0{,}03 \cdot log_2(0{,}03)) =$$
$$= 0{,}054805219444381 + 0{,}011568910657988 + 0{,}030102999566398 +$$
$$0{,}049356579386504 + 0{,}11656745623924 + 0{,}0035436764027112 +$$
$$0{,}039229869422866 + 0{,}054805219444381 + 0{,}0059301503142701 =$$
$$0{,}36591008087874.$$
$I(X,Y) = 0{,}74801748945526 + 0{,}63489140822761 - 0{,}36591008087874 = 1{,}0169988168041.$
$$K_R = \frac{1{,}0169988168041}{0{,}36591008087874} = 2{,}7793681288085.$$

Отже, оскільки значення коефіцієнту досить близьке до одиниці, можемо вважати, що послуги X, Y утворюють комплексну послугу.

Відповідь. Послуги X, Y утворюють комплексну послугу.