Глава 1

СТАТИСТИЧЕСКАЯ МОДЕЛЬ ШЕННОНА–ВИНЕРА

Приводятся фундаментальные понятия и представления теории информации, основанной на статистической модели Шеннона–Винера. Обсуждаются логарифмические меры энтропии и информации различия для аддитивных случайных объектов и методы их получения.

Начало статистического подхода в теории информации связано с работами Р. Хартли [77], К. Шеннона [50], Н. Винера [5], В.А. Котельникова [30, 31], А.Н. Колмогорова [27–29] по теории передачи информации по системам связи и Р.А. Фишера [74, 75] по теории параметрического оценивания в математической статистике.

В отличие от традиционного изложения основ теории информации, представленного, например, в монографиях [5, 8, 31, 33, 35, 37, 38, 41, 42, 44, 46, 50, 76], в данной главе наряду с определением взвешенного среднего впервые вводится взвешенное среднее геометрическое для каждой случайной величины. Такой подход позволил статистическим, вариационным и групповым методами определить один тип аддитивной энтропии и информации различия, зависящих от средних геометрических для распределения.

1.1. Вероятность. Взвешенное среднее

Идея количественных мер в информационных представлениях состоит в том, что рассматривается вероятностно-статистическое описание объекта (процесса, системы или сигнала), который имеет дискретные или непрерывные случайные состояния.

Приведем основные определения.

Определение 1. Носителем информации являются два множества: множество всех состояний объекта, описываемых распределением вероятностей $p=\{p_1,\ldots,\ p_m\}$ $(p_i \geq 0,\ \ i=1,\ 2,\ \ldots,\ m)$ и множество случайных величин $T=\{T_1,\ldots,\ T_m\}$, характеризующих объект.

Определение 2. Взвешенное среднее каждой случайной величины T в состоянии с распределением p равно

$$\mathbf{E}(T)=\frac{\sum_i^m T_i p_i}{\sum_i^m p_i}, \tag{1.1.1}$$

где m – число возможных состояний объекта и

$$0<\sum_i^m p_i \leq 1. \tag{1.1.2}$$

Используется условие вероятностной нормировки распределения

$$\sum_i^m p_i = 1, \tag{1.1.3}$$

при выполнении которой взвешенное среднее примет следующий вид

$$\mathbf{E}(T)=\sum_i^m T_i p_i\,. \tag{1.1.4}$$

Значения p_i в выражениях (1.1.1) и (1.1.4) представляют собой так называемые веса значений T_i. В дальнейшем эти выражения будем называть просто средними, если веса не имеют своего обособленного представления.

Если $p_i=1$, то выражение (1.1.1) равняется обычному среднему арифметическому

$$\mathbf{E}(T)=\frac{1}{m}\sum_i^m T_i\,. \tag{1.1.5}$$

Из определения 2 вытекают основные свойства.

1. Однородность. При замене p на ap $(a>0)$ взвешенное среднее является однородным функционалом нулевой степени относительно p, то есть выполняется свойство однородности.

2. Нормированность. Для неслучайной постоянной величины C имеем равенство

$$\mathbf{E}(C) = C, \tag{1.1.6}$$

справедливое при $C = 0$ и $C = \infty$. В частности, при $C = 1$ из (1.1.6) следует $\mathbf{E}(1) = 1$, что означает нормированность взвешенного среднего на единицу.

3. Аддитивность. Пусть случайная величина $T_{12} = T_1 + T_2$ равняется сумме случайных величин двух независимых объектов. Совместное распределение вероятностей значений $T_{ij} = T_i + T_j$ мультипликативно $p_{ij} = p_i p_j$ (теорема умножения) и нормировано $\sum_i^m \sum_j^n p_{ij} = \sum_i^m p_i = \sum_j^n p_j = 1$.
Тогда получим равенство

$$\mathbf{E}(T_{12}) = \mathbf{E}(T_1) + \mathbf{E}(T_2). \tag{1.1.7}$$

Для случая одного объекта со случайной величиной $U = U_1 + U_2$ имеем

$$\mathbf{E}(U) = \mathbf{E}(U_1) + \mathbf{E}(U_2). \tag{1.1.8}$$

Здесь средние значения

$$\mathbf{E}(T_{12}) = \sum_i^m \sum_j^n T_{ij} p_i p_j, \quad \mathbf{E}(U) = \sum_i^m U_i p_i, \tag{1.1.9}$$

$$\mathbf{E}(T_1) = \sum_i^m T_i p_i, \quad \mathbf{E}(T_2) = \sum_j^n T_j p_j, \tag{1.1.10}$$

$$\mathbf{E}(U_1) = \sum_i^m U_{1i} p_i, \quad \mathbf{E}(U_2) = \sum_i^m U_{2i} p_i. \tag{1.1.11}$$

Равенства (1.1.7) и (1.1.8) означают аддитивность наблюдаемых макроскопических величин.

4. Мультипликативность. Среднее значение произведения независимых случайных величин $T_1 T_2$ для одного или двух объектов равно произведению средних значений

$$\mathbf{E}(T_1 T_2) = \mathbf{E}(T_1)\ \mathbf{E}(T_2). \tag{1.1.12}$$

5. f -взвешенное среднее. Функциональное обобщение взвешенного нормированного среднего запишется так

$$\mathbf{E}_f\left(T\right)=\frac{\sum_{i}^{m} T_i f\left(p_i\right)}{\sum_{i}^{m} f\left(p_i\right)}. \qquad (1.1.13)$$

Если $f\left(p_i\right)=p_i$, то f -взвешенное среднее совпадает с выражением (1.1.1)

Определение 3. Отклонение случайной величины T_i от среднего значения есть флуктуация

$$\Delta T_i = T_i - \mathbf{E}\left(T\right), \quad \sum_{i}^{m}\left(\Delta T_i\right) p_i = 0. \qquad (1.1.14)$$

Определение 4. Начальные и центральные моменты n-го порядка случайной величины T равны

$$\alpha_n = \sum_{i}^{m} T_i^n p_i, \qquad (1.1.15)$$

$$\mu_n = \sum_{i}^{m}\left[T_i - \mathbf{E}\left(T\right)\right]^n p_i. \qquad (1.1.16)$$

При $n=2$ получим дисперсию

$$\mu_2 = \mathbf{D}\left(T\right) = \sum_{i}^{m}\left[T_i - \mathbf{E}\left(T\right)\right]^2 p_i = \sum_{i}^{m} T_i^2 p_i - \left(\sum_{i}^{m} T_i p_i\right)^2. \qquad (1.1.17)$$

Статистическая зависимость двух величин характеризуется коэффициентом корреляции

$$r\left(T_1, T_2\right) = \frac{\mathbf{E}\left(\Delta T_1 \Delta T_2\right)}{\sigma\left(T_1\right)\sigma\left(T_2\right)}, \qquad (1.1.18)$$

где имеем квадратичные отклонения

$$\sigma\left(T_1\right) = \left[\mathbf{D}\left(T_1\right)\right]^{1/2}, \quad \sigma\left(T_2\right) = \left[\mathbf{D}\left(T_2\right)\right]^{1/2}. \qquad (1.1.19)$$

Отдельные моменты имеют существенное значение в статистической теории информации, поскольку являются наблюдаемыми макроскопическими величинами, содержащими количественные информационные свойства о случайном объекте.

Далее рассмотрим вероятностное пространство с непрерывной мерой

$$\Gamma(G)=\int_G p(X)dX, \quad 0<p(X)<\infty, \tag{1.1.20}$$

где $p=p(X)$ есть функция плотности распределения вероятностей по неравновозможным микросостояниям X в области G. В общем случае функция может зависеть от времени t и от некоторого параметра θ. Плотность распределения вероятностей (или, кратко, распределение) является так называемой производной Радона–Никодима $p=d\Gamma(X)/dX$ и всегда положительна [33].

Определение 5. Аналогом взвешенного среднего (1.1.1) каждой непрерывной случайной величины $T=T(X)$ в состоянии p является выражение

$$\mathbf{E}(T)=\frac{1}{\Gamma(G)}\int_G Td\Gamma=\frac{\int_G TpdX}{\int_G pdX}, \tag{1.1.21}$$

где

$$0<\int_G pdX<\infty. \tag{1.1.22}$$

При условии вероятностной нормировки для всего пространства из (1.1.21) вытекает следующее взвешенное среднее по вероятностной мере

$$\mathbf{E}(T)=\int Td\Gamma=\int TpdX, \quad \int pdX=1. \tag{1.1.23}$$

Весовой функцией случайной величины $T(X)$ является распределение $p(X)$.

1.2. Взвешенное среднее геометрическое

Продолжим рассмотрение основных определений.

Определение 6. Каждой случайной величине T соответствует взвешенное среднее геометрическое [4, 47]

$$N(T)=2^{\frac{\sum_i^m(\log_2 T_i)p_i}{\sum_i^m p_i}}=\left(\prod_i^m T_i^{p_i}\right)^{\frac{1}{\sum_i^m p_i}}. \tag{1.2.1}$$

Используется также выражение

$$\log_2 N(T) = \frac{\sum_i^m (\log_2 T_i) p_i}{\sum_i^m p_i}. \qquad (1.2.2)$$

В (1.2.1) и (1.2.2) взвешенное среднее определяется для случайной величины $\log_2 T_i$. Для нормированного распределения имеем среднее геометрическое и ее логарифм

$$N(T) = 2^{\sum_i^m (\log_2 T_i) p_i}, \qquad (1.2.3)$$

$$\log_2 N(T) = \sum_i^m (\log_2 T_i) p_i, \ \sum_i^m p_i = 1. \qquad (1.2.4)$$

Если $p_i = 1$, то из (1.2.1) вытекает обычное среднее геометрическое

$$N(T) = \sqrt[m]{\prod_i^m T_i}. \qquad (1.2.5)$$

Приведем основные свойства взвешенного среднего геометрического.

1. Однородность и вогнутость. Величина $N(T)$ является однородной относительно замены T на aT

$$N(aT) = aN(T), \ a = \text{const} \qquad (1.2.6)$$

и вогнутой, так как отрицательное значение среднего геометрического есть выпуклая функция.

Если T_1 и T_2 две положительные функции, то справедливо неравенство

$$N(T_1) + N(T_2) < N(T_1 + T_2), \qquad (1.2.7)$$

кроме случаев, когда $bT_1 = cT_2$, где b и c не равны нулю, или $N(T_1 + T_2) = 0$.

В общем случае (1.2.7) запишется так

$$N(T_1) + N(T_2) + \ldots + N(T_n) < N(T_1 + T_2 + \ldots + T_n). \qquad (1.2.8)$$

2. Нормированность. Для неслучайной постоянной величины C имеем равенство:

$$N(C) = 2^{\frac{\sum_i^m (\log_2 C) p_i}{\sum_i^m p_i}} = C, \qquad (1.2.9)$$

из которого вытекает свойство нормированности взвешенного среднего геометрического на единицу

$$N(1) = 1. \qquad (1.2.10)$$

3. Среднее геометрическое и взвешенное среднее. Для взаимосвязи средних геометрических приведем соотношения

$$N(T)\,N\left(T^{-1}\right) = 1, \qquad (1.2.11)$$

$$\frac{N(T_1)}{N(T_1+T_2)} = N\left(\frac{T_1}{T_1+T_2}\right) \le \mathbf{E}\left(\frac{T_1}{T_1+T_2}\right), \qquad (1.2.12)$$

$$N\left(T_1^{\alpha_1} T_2^{\alpha_2} \ldots T_n^{\alpha_n}\right) = \left[N(T_1)\right]^{\alpha_1}\left[N(T_2)\right]^{\alpha_2} \ldots \left[N(T_n)\right]^{\alpha_n}. \qquad (1.2.13)$$

Если взвешенное среднее случайной величины T конечно, то справедливо неравенство

$$2^{\frac{\sum_i^m (\log_2 T_i) p_i}{\sum_i^m p_i}} < \frac{\sum_i^m T_i p_i}{\sum_i^m p_i} < 2^{\frac{\sum_i^m (\log_2 T_i) T_i p_i}{\sum_i^m T_i p_i}}, \qquad (1.2.14)$$

кроме того случая, когда $T = C$, где C – постоянная.

Рассмотрим три случая зависимости T от распределений.

При $T = p$ взвешенное среднее геометрическое распределения является следующим функционалом

$$N(T) = 2^{\frac{\sum_i^m (\log_2 p_i) p_i}{\sum_i^m p_i}}, \qquad (1.2.15)$$

для которого выполняются свойства 1, 2 и 3 и следующие дополнительные свойства.

4. Мультипликативность. Взвешенное среднее геометрическое произведения $p_{ij} = p_i p_j$ распределений независимых объектов равно произведению их взвешенных средних геометрических

$$N\left(p_{12}\right) = N\left(p_1\right) N\left(p_2\right), \qquad (1.2.16)$$

где

$$N\left(p_{12}\right) = 2^{\frac{\sum_i^m \sum_j^n \left(\log_2 p_{ij}\right) p_{ij}}{\sum_i^m \sum_j^n p_{ij}}}, \qquad (1.2.17)$$

$$N\left(p_1\right) = 2^{\frac{\sum_i^m \left(\log_2 p_i\right) p_i}{\sum_i^m p_i}}, \quad N\left(p_2\right) = 2^{\frac{\sum_j^n \left(\log_2 p_j\right) p_j}{\sum_j^n p_j}}. \qquad (1.2.18)$$

5. Взвешенное среднее геометрическое равновероятного распределения. Подставим равновероятное распределение

$$p_i = \frac{1}{m} \qquad (1.2.19)$$

в определение (1.2.1) и получим среднее геометрическое

$$N\left(p\right) = \frac{1}{m}, \qquad (1.2.20)$$

обратно пропорциональное числу возможных состояний объекта.

Далее примем $T = u$ и $T = p/u$ и определим средние геометрические

$$N\left(u\right) = 2^{\frac{\sum_i^m \left(\log_2 u_i\right) p_i}{\sum_i^m p_i}}, \qquad (1.2.21)$$

$$N\left(\frac{p}{u}\right) = 2^{\frac{\sum_i^m \left(\log_2 \frac{p_i}{u_i}\right) p_i}{\sum_i^m p_i}}, \qquad (1.2.22)$$

которые имеют все приведенные свойства.

6. Взаимосвязь взвешенных средних геометрических распределений. Используем равенство (1.2.13) и получим следующее соотношение:

$$N\left(\frac{p}{u}\right)=\frac{N(p)}{N(u)}. \tag{1.2.23}$$

Подставив в (1.2.23) равновероятное распределение (1.2.19), получим

$$N\left(\frac{p}{u}\right)=mN(p). \tag{1.2.24}$$

Приведем непрерывные аналоги взвешенных средних геометрических распределений

$$N(p)=2^{\frac{\int_G (\log_2 p)pdX}{\int_G pdX}}, \quad N(u)=2^{\frac{\int_G (\log_2 u)pdX}{\int_G pdX}}, \tag{1.2.25}$$

$$N\left(\frac{p}{p_0}\right)=2^{\frac{\int_G \left(\log_2 \frac{p}{p_0}\right)pdX}{\int_G pdX}} \tag{1.2.26}$$

и их логарифмы

$$\log_2 N(p)=\frac{\int_G (\log_2 p)\,pdX}{\int_G pdX}, \quad \log_2 N(u)=\frac{\int_G (\log_2 u)\,pdX}{\int_G pdX}, \tag{1.2.27}$$

$$\log_2 N\left(\frac{p}{u}\right)=\frac{\int_G \left(\log_2 \frac{p}{u}\right)pdX}{\int_G pdX}. \tag{1.2.28}$$

Предопределяя дальнейшие результаты, отметим, что функционалы

$$H(p)=-\log_2 N(p), \tag{1.2.29}$$

$$H(p:u)=-\log_2 N(u), \tag{1.2.30}$$

$$I(p:u)=\log_2 N\left(\frac{p}{u}\right) \tag{1.2.31}$$

есть аддитивные выражения энтропии Шеннона–Винера, меры неточности Керриджа и информации различия Кульбака–Лейблера в статистической теории информации.

Взвешенное среднее геометрическое (1.2.1) можно записать в эквивалентном виде [47]

$$N(T)=\exp\left(\frac{\sum_{i}^{m}(\ln T_i)\,p_i}{\sum_{i}^{m}p_i}\right). \qquad (1.2.32)$$

Тогда функционалы (1.2.29) и (1.2.31) с натуральным логарифмом находят применение в статистической физике [19, 21].

1.3. Принцип минимума среднего геометрического распределения

Рассмотрим экстремум среднего геометрического распределения

$$N(p)=2^{\sum_{i}^{m}(\log_2 p_i)p_i} \qquad (1.3.1)$$

при сохранении нормировки

$$\sum_{i}^{m}p_i=1. \qquad (1.3.2)$$

Это позволит найти вероятное распределение, при котором достигается экстремальное значение среднего геометрического. Согласно вариационному принципу, вычислим безусловный экстремум функционала

$$L=2^{\sum_{i}^{m}(\log_2 p_i)p_i}-\alpha\sum_{i}^{m}p_i\,, \qquad (1.3.3)$$

где α есть множитель Лагранжа.

Приравнивая нулю первую вариацию

$$\delta L=2^{\sum_{i}^{m}(\log_2 p_i)p_i}\ln 2\cdot\sum_{i}^{m}\delta p_i\left(\log_2 p_i+\frac{1}{\ln 2}\right)-\alpha\sum_{i}^{m}\delta p_i=0\,, \qquad (1.3.4)$$

получим равенство

$$\gamma\left[(\ln 2)\log_2 p_i+1\right]-\alpha=0\,,\ \gamma=2^{\sum_{i}^{m}(\log_2 p_i)p_i}\,, \qquad (1.3.5)$$

из которого с учетом условия нормировки вытекает равновероят ное распределение

$$p_i = \frac{1}{m}. \qquad (1.3.6)$$

Подставим (1.3.6) в (1.3.1) и получим экстремальное значение среднего геометрического

$$N(p) = \frac{1}{m}. \qquad (1.3.7)$$

Далее рассмотрим вариационный принцип экстремума среднего геометрического распределения при дополнительных условиях заданности среднего значения случайной энтропии $h = \{h_1, \ldots, h_m\}$ и нормировки

$$H(p) = \sum_i^m h_i p_i, \quad \sum_i^m p_i = 1. \qquad (1.3.8)$$

Определим функционал

$$L = 2^{\sum_i^m (\log_2 p_i) p_i} + \tau \sum_i^m h_i p_i - \alpha \sum_i^m p_i, \qquad (1.3.9)$$

где α и τ есть множители Лагранжа. Используя равенство

$$\delta L = 2^{\sum_i^m (\log_2 p_i) p_i} \ln 2 \cdot \sum_i^m \delta p_i \left(\log_2 p_i + \frac{1}{\ln 2} \right) +$$
$$+ \tau \sum_i^m h_i \delta p_i - \alpha \sum_i^m \delta p_i = 0, \qquad (1.3.10)$$

получим

$$\gamma\left[(\ln 2) \log_2 p_i + 1 \right] + \tau h_i - \alpha = 0. \qquad (1.3.11)$$

Поскольку α и τ имеют произвольные значения, то примем $\gamma \ln 2 = \alpha = \tau$. Тогда из (1.3.11) следует распределение

$$p_i = 2^{-h_i} \qquad (1.3.12)$$

и значение случайной энтропии

$$h_i(p) = -\log_2 p_i. \qquad (1.3.13)$$

Усредняя (1.3.13), получим функционал:

$$H(p)=\sum_i^m h(p_i)\,p_i=-\sum_i^m(\log_2 p_i)\,p_i\,, \tag{1.3.14}$$

подстановка которого в (1.3.1) дает экстремальные выражения

$$N(p)=2^{-H(p)},\quad H(p)=-\log_2 N(p). \tag{1.3.15}$$

Наконец находим экстремум среднего геометрического от величины p/u

$$N\left(\frac{p}{u}\right)=2^{\sum_i^m\left(\log_2\frac{p_i}{u_i}\right)p_i} \tag{1.3.16}$$

при дополнительных условиях

$$I(p:u)=\sum_i^m I_i p_i,\quad \sum_i^m p_i=1. \tag{1.3.17}$$

Тогда задача сводится к нахождению безусловного экстремума следующего функционала

$$L=2^{\sum_i^m\left(\log_2\frac{p_i}{u_i}\right)p_i}-\tau\sum_i^m I_i p_i-\alpha\sum_i^m p_i \tag{1.3.18}$$

с множителями Лагранжа α и τ. Из условия $\delta L=0$ получим

$$\gamma\left[(\ln 2)\log_2\frac{p_i}{u_i}+1\right]-\tau I_i-\alpha=0 \tag{1.3.19}$$

и при $\gamma\ln 2=\alpha=\tau$ окончательно имеем выражения

$$p_i=u_i 2^{I_i}\,, \tag{1.3.20}$$

$$I_i=I(p_i:u_i)=\log_2\frac{p_i}{u_i}\,. \tag{1.3.21}$$

Усреднение случайной информации различия (1.3.21) приводит к функционалу

$$I(p:p_0)=\sum_i^m I_i(p:p_0)\,p_i=\sum_i^m\left(\log_2\frac{p_i}{p_{0i}}\right)p_i\,. \tag{1.3.22}$$

Для среднего геометрического получим экстремальное значение:

$$N\left(\frac{p}{p_0}\right)=2^{I(p:p_0)},\ I(p:u)=\log_2 N\left(\frac{p}{u}\right). \qquad (1.3.23)$$

Полученные экстремумы для средних геометрических соответствуют минимуму рассматриваемых функционалов, поскольку выполняются, соответственно, неравенства для второй вариации

$$\delta^2 L=\left[\delta N(p)\right]^2\left[N(p)\right]^{-1}+N(p)\sum_i^m\frac{(\delta p_i)^2}{p_i}\geq 0, \qquad (1.3.24)$$

$$\delta^2 L=\left[\delta N\left(\frac{p}{u}\right)\right]^2\left[N\left(\frac{p}{u}\right)\right]^{-1}+N\left(\frac{p}{u}\right)\sum_i^m\frac{(\delta p_i)^2}{p_i}\geq 0. \qquad (1.3.25)$$

Следовательно, в экстремуме энтропия Шеннона–Винера $H(p)$ и информация различия Кульбака–Лейблера $I(p:u)$ имеют максимум и минимум соответственно.

1.4. Тип аддитивной энтропии, меры неточности и информации различия

Рассмотрим случайный объект, который состоит их двух независимых объектов. Используем свойство мультипликативности среднего геометрического

$$N=N_1N_2, \qquad (1.4.1)$$

для произведения $p_{ij}=p_ip_j$ распределений независимых объектов, где $N=N(p_{12})$, $N_1=N(p_1)$ и $N_2=N(p_2)$ определяются формулами (1.2.17) и (1.2.18).

Далее положим, что выполняется свойство аддитивности для энтропий

$$H=H_1+H_2, \qquad (1.4.2)$$

которые зависят от соответствующих средних геометрических

$$H=H(N),\quad H_1=H_1(N_1),\quad H_2=H_2(N_2). \qquad (1.4.3)$$

Тогда, дифференцируя (1.4.1) и учитывая равенства

$$\frac{dN}{dH}=\frac{dN}{dN_1}\frac{dN_1}{dH_1}\frac{dH_1}{dH},\qquad \frac{dN}{dH}=\frac{dN}{dN_2}\frac{dN_2}{dH_2}\frac{dH_2}{dH} \qquad (1.4.4)$$

получим уравнение:

$$\frac{d\ln N}{dH}=\frac{d\ln N_1}{dH_1}=\frac{d\ln N_2}{dH_2}=-\lambda\,, \qquad (1.4.5)$$

где λ – произвольная постоянная.

Решением уравнения (1.4.5) является физическая безразмерная энтропия с точностью до коэффициента λ^{-1}

$$H^{phys}\left(p\right)=-\lambda^{-1}\ln N\left(p\right). \qquad (1.4.6)$$

Используем свойство нормированности энтропии на единицу в теории информации, и получим значение $\lambda=\ln 2$. Энтропия (1.4.6) принимает для всех объектов одинаковый вид

$$H\left(p\right)=-\log_2 N\left(p\right)=-\frac{\sum_i^m\left(\log_2 p_i\right)p_i}{\sum_i^m p_i}. \qquad (1.4.7)$$

Далее определим следующие геометрические средние:

$$N=N\left(u_{12}\right),\quad N_1=N\left(u_1\right),\quad N_2=N\left(u_2\right), \qquad (1.4.8)$$

$$N=N\left(\frac{p_{12}}{u_{12}}\right),\quad N_1=N\left(\frac{p_1}{u_1}\right),\quad N_2=N\left(\frac{p_2}{u_2}\right), \qquad (1.4.9)$$

где $p_{12}=p_1p_2$, $u_{12}=u_1u_2$. Аналогично используем свойства мультипликативности $N=N_1N_2$ для независимых объектов и аддитивности для мер неточности и информаций различия

$$H=H_1+H_2, \qquad (1.4.10)$$

$$I=I_1+I_2. \qquad (1.4.11)$$

В итоге получим уравнения

$$\frac{d\ln N}{dH}=\frac{d\ln N_1}{dH_1}=\frac{d\ln N_2}{dH_2}=-\lambda\,, \qquad (1.4.12)$$

$$\frac{d\ln N}{dI}=\frac{d\ln N_1}{dI_1}=\frac{d\ln N_2}{dI_2}=\lambda\,, \qquad (1.4.13)$$

решениями которых при $\lambda=\ln 2$ являются мера неточности и информация различия:

$$H(p:u) = -\log_2 N(u) = -\frac{\sum_i^m (\log_2 u_i)\, p_i}{\sum_i^m p_i}, \qquad (1.4.14)$$

$$I(p:u) = \log_2 N\left(\frac{p}{u}\right) = \frac{\sum_i^m \left(\log_2 \frac{p_i}{u_i}\right) p_i}{\sum_i^m p_i}. \qquad (1.4.15)$$

Таким образом, используя свойства мультипликативности, аддитивности и нормированности доказали, что существует только один тип энтропии, меры неточности и информации различия.

1.5. Энтропия Шеннона–Винера. Информация и мера Хартли

К статистическим свойствам случайного объекта относится мера информации как мера статистической неопределенности (или случайности) в состояниях, выражаемая некоторым функционалом от распределения. Для аддитивных объектов таким функционалом является энтропия Шеннона–Винера [5, 50]

$$H(p) = -\frac{\sum_i^m (\log_2 p_i)\, p_i}{\sum_i^m p_i}. \qquad (1.5.1)$$

При выполнении условия вероятностной нормировки

$$\sum_i^m p_i = 1 \qquad (1.5.2)$$

взвешенное среднее (1.5.1) случайной энтропии

$$h(p_i) = -\log_2 p_i \qquad (1.5.3)$$

принимает следующий вид:

$$H(p) = -\sum_i^m (\log_2 p_i)\, p_i. \qquad (1.5.4)$$

Приведем основные свойства энтропии Шеннона–Винера.

1. Положительность и выпуклость. Энтропия есть вещественный, неотрицательный и выпуклый функционал, то есть справедливы неравенства

$$H\left(p\right)\geq 0\,, \tag{1.5.5}$$

$$H\left(a_1 p_1 + a_2 p_2\right)\leq a_1 H\left(p_1\right)+a_2 H\left(p_2\right). \tag{1.5.6}$$

Здесь $a_1 + a_2 = 1$, $a_1 > 0$, $a_2 > 0$ и энтропии

$$H\left(p_1\right)=-\sum_i^m\left(\log_2 p_{1i}\right)p_{1i}\,,\;\; H\left(p_2\right)=-\sum_i^m\left(\log_2 p_{2i}\right)p_{2i} \tag{1.5.7}$$

с нормированными распределениями

$$\sum_i^m p_{1i} = \sum_i^m p_{2i} = 1\,. \tag{1.5.8}$$

2. Аддитивность для независимых объектов. Пусть состояние случайного объекта описывается совместным мультипликативным распределением $p_{ij} = p_i p_j$, а p_i и p_j относятся к разным независимым объектам. Общая энтропия запишется как

$$H\left(p_{12}\right)=-\sum_i^m\sum_j^n\left(\log_2 p_{ij}\right)p_{ij}\,, \tag{1.5.9}$$

где условия нормировки

$$\sum_i^m\sum_j^n p_{ij} = \sum_i^m p_i = \sum_j^n p_j = 1\,. \tag{1.5.10}$$

Тогда из (1.5.4) вытекает свойство аддитивности для энтропий независимых объектов

$$\begin{gathered} H\left(p_{12}\right)=H\left(p_1\right)+H\left(p_2\right), \\ H\left(p_1\right)=-\sum_i^m\left(\log_2 p_i\right)p_i,\quad H\left(p_2\right)=-\sum_j^n\left(\log_2 p_j\right)p_j. \end{gathered} \tag{1.5.11}$$

3. Аддитивность для зависимых объектов. В общем случае зависимых объектов для нормированных распределений имеем соотношения

$$p_{ij} = p_i p_{j|i} = p_j p_{i|j} \quad \text{(теорема умножения);} \tag{1.5.12}$$

$$p_i = \sum_j^n p_{ij},\quad p_j = \sum_i^m p_{ij} \quad \text{(теорема сложения);} \tag{1.5.13}$$

$$p_i = \sum_j^n p_{i|j} p_j, \quad p_j = \sum_i^m p_{j|i} p_i \quad \text{(теорема разложения).} \qquad (1.5.14)$$

Логарифмируя произведение распределений (1.5.12), получим аддитивность случайных энтропий

$$h(p_{ij}) = h(p_i) + h(p_{j|i}). \qquad (1.5.15)$$

Здесь $h(p_{j|i}) = -\log_2 p_{j|i}$ – случайная условная энтропия с условным распределением $p_{j|i}$. Учитывая условия нормировки

$$\sum_i^m \sum_j^n p_{ij} = \sum_i^m p_i = 1, \qquad (1.5.16)$$

получим свойство аддитивности для энтропий статистически зависимых объектов

$$H(p_{12}) = H(p_1) + H(p_2|p_1), \qquad (1.5.17)$$

где условная энтропия объекта с распределением $p_{2|1}$

$$H_i(p_{2|1}) = -\sum_j^n (\log_2 p_{j|i}) p_{j|i} \qquad (1.5.18)$$

и ее среднее значение

$$H(p_2|p_1) = \sum_i^m p_i H_i(p_{2|1}) \qquad (1.5.19)$$

вычисляются при условии реализации состояния с распределением p_1.

Если реализуется состояние с распределением p_2, то $p_{ij} = p_j p_{i|j}$ и тогда получаем энтропии

$$H(p_{12}) = H(p_2) + H(p_1|p_2), \qquad (1.5.20)$$

$$H_j(p_{1|2}) = -\sum_i^m (\log_2 p_{i|j}) p_{i|j}, \qquad (1.5.21)$$

$$H(p_1|p_2) = \sum_j^n p_j H_j(p_{1|2}). \qquad (1.5.22)$$

С учетом (1.5.17) и (1.5.20) имеем равенство:

$$H\left(p_1\right)+H\left(p_2\middle|p_1\right)=H\left(p_2\right)+H\left(p_1\middle|p_2\right). \qquad (1.5.23)$$

В частном случае независимости объектов соотношения (1.5.17) и (1.5.20) переходят в (1.5.11).

Для трех зависимых объектов имеем следующие распределения

$$p_i=\sum_j^n\sum_k^r p_{ijk}, \quad p_{ij}=\sum_k^r p_{ijk}, \quad p_{i|j}=\frac{\sum_k^r p_{ijk}}{\sum_i^m\sum_k^r p_{ijk}}, \qquad (1.5.24)$$

$$p_{i,j|k}=\frac{p_{ijk}}{\sum_i^m\sum_j^n p_{ijk}}, \qquad p_{i|j,k}=\frac{p_{ijk}}{\sum_i^m p_{ijk}}, \qquad (1.5.25)$$

для которых справедливы формулы

$$p_{ijk}=p_k p_{i,j|k}=p_{k|i,j}p_{ij}=p_i p_{j|i}p_{k|i,j}, \qquad (1.5.26)$$

$$p_{ij}=p_i p_{j|i}, \quad p_{i|j}=\sum_k^r p_{i,k|j}, \qquad (1.5.27)$$

$$p_{i,j|k}=p_{j|k}p_{i|j,k}=\frac{p_i p_{i|j} p_{i|j,k}}{\sum_i^m p_j p_{i|j}}. \qquad (1.5.28)$$

Вводим соответствующие энтропии

$$H\left(p_{123}\right)=-\sum_i^m\sum_j^n\sum_k^r\left(\log_2 p_{ijk}\right)p_{ijk}, \qquad (1.5.29)$$

$$H\left(p_{12}\right)=-\sum_i^m\sum_j^n\sum_k^r\left(\log_2 p_{ij}\right)p_{ijk}, \qquad (1.5.30)$$

$$H\left(p_1\right)=-\sum_i^m\sum_j^n\sum_k^r\left(\log_2 p_i\right)p_{ijk}, \qquad (1.5.31)$$

$$H\left(p_1\middle|p_2\right)=-\sum_i^m\sum_j^n\sum_k^r\left(\log_2 p_{i|j}\right)p_{ijk}, \qquad (1.5.32)$$

$$H\left(p_1, p_2 \middle| p_3\right) = -\sum_i^m \sum_j^n \sum_k^r \left(\log_2 p_{i,j|k}\right) p_{ijk}\,, \tag{1.5.33}$$

$$H\left(p_1 \middle| p_2, p_3\right) = -\sum_i^m \sum_j^n \sum_k^r \left(\log_2 p_{i|j,k}\right) p_{ijk} \tag{1.5.34}$$

и в итоге для них имеем свойство аддитивности в виде

$$\begin{gathered} H\left(p_{123}\right) = H\left(p_1\right) + H\left(p_2, p_3 \middle| p_1\right) = H\left(p_{12}\right) + \\ + H\left(p_3 \middle| p_1, p_2\right) = H\left(p_1\right) + H\left(p_2 \middle| p_1\right) + H\left(p_3 \middle| p_1, p_2\right). \end{gathered} \tag{1.5.35}$$

4. Флуктуация. Используем выражение флуктуации микроскопической энтропии

$$\Delta h\left(p_i\right) = h\left(p_i\right) - H\left(p\right) \tag{1.5.36}$$

и запишем распределение в виде

$$p_i = 2^{-\left[H(p) - \Delta h_i(p)\right]}. \tag{1.5.37}$$

Из условия нормировки получим формулы

$$p_i = \frac{2^{-\Delta h\left(p_i\right)}}{\sum_i^m 2^{-\Delta h\left(p_i\right)}}\,, \tag{1.5.38}$$

$$H\left(p\right) = \log_2 \sum_i^m 2^{-\Delta h\left(p_i\right)}\,, \tag{1.5.39}$$

позволяющие находить распределение и энтропию с использованием значения флуктуации $\Delta h\left(p_i\right)$.

5. Энтропия равновероятного состояния. Пусть в случайном объекте отсутствуют флуктуации микроскопической энтропии и $\Delta h_i\left(p\right) = 0$. Из (1.5.38) и (1.5.39) вытекает равновероятное распределение и соответствующая энтропия

$$p_i = \frac{1}{m}\,, \tag{1.5.40}$$

$$H\left(p\right) = \log_2 m\,. \tag{1.5.41}$$

6. Неравенства. Энтропия Шеннона–Винера также удовлетворяет основным неравенствам:

$$H\left(p_{12}\right) \leq H\left(p_1\right) + H\left(p_2\right), \tag{1.5.42}$$

$$H\left(p_2\right) \geq H\left(p_1 \middle| p_2\right), \quad H\left(p_1\right) \geq H\left(p_2 \middle| p_1\right), \tag{1.5.43}$$

$$H\left(p\right) \leq \log_2 m, \tag{1.5.44}$$

$$H\left(p_1 \middle| p_2, p_3\right) \leq H\left(p_1 \middle| p_3\right), \tag{1.5.45}$$

$$H\left(p_1, p_2 \middle| p_3\right) \leq H\left(p_1, p_2\right), \tag{1.5.46}$$

$$H\left(p_1 \middle| p_3\right) \leq H\left(p_1 \middle| p_2\right) + H\left(p_2 \middle| p_3\right), \tag{1.5.47}$$

$$\frac{H\left(p_1 \middle| p_3\right)}{H\left(p_1, p_3\right)} \leq \frac{H\left(p_1 \middle| p_2\right)}{H\left(p_1, p_2\right)} + \frac{H\left(p_2 \middle| p_3\right)}{H\left(p_2, p_3\right)}, \tag{1.5.48}$$

которые вытекают из свойства выпуклости информации различия.

7. Энтропийное расстояние. Геометрическая интерпретация энтропии в виде $\delta\left(p_1, p_2\right) = H\left(p_1 \middle| p_2\right) + H\left(p_2 \middle| p_1\right)$ соответствует расстоянию от p_1 до p_2. Выполняется неравенство треугольника

$$\delta\left(p_1, p_3\right) \leq \delta\left(p_1, p_2\right) + \delta\left(p_2, p_3\right), \tag{1.5.49}$$

где $\delta\left(p_1, p_2\right) \geq 0$, $\delta\left(p_1, p_1\right) = 0$ и $\delta\left(p_1, p_2\right) = \delta\left(p_2, p_1\right)$. Рассматривается также выражение

$$\delta'\left(p_1, p_2\right) = \frac{\delta\left(p_1, p_2\right)}{H\left(p_{12}\right)} = 1 - \frac{I\left(p_{12} : p_1 p_2\right)}{H\left(p_{12}\right)}, \tag{1.5.50}$$

которое равняется нулю при $H\left(p_{12}\right) = 0$. Подробно о статистических расстояниях можно ознакомиться в монографиях [71, 91, 125].

8. Нормированность и размерность. В результате эксперимента мера неопределенности состояний становится известной и является информацией объекта о самом объекте. За единицу измерения информации принимается неопределенность, при которой выполняется свойство нормированности энтропии

$$H\left(p\right) = -\sum_{i}^{m} \left(\log_2 p_i\right) p_i = 1. \tag{1.5.51}$$

Выбирая наименьшее число возможных состояний $m = 2$, получим из (1.5.51) значения $p_1 = p_2 = 1/2$. Таким образом, наименьшая единица измерения информации есть двоичная единица энтропии системы, которая имеет название один бит (от binary digit)

$$H\left(\frac{1}{2},\frac{1}{2}\right)=-\sum_{i}^{2}\left(\log_2 p_i\right)p_i=1. \qquad (1.5.52)$$

В случае десятичного и натурального логарифма в (1.5.4) единица измерения информации называется дитом или натом (от natural digit), соответственно. Для перехода между этими единицами измерения имеем равенство $1\text{дит}=\left(1/\lg 2\right)\text{бит}$ и $1\text{нат}=\left(1/\ln 2\right)$ бит. Для физической теории информации имеем энтропию

$$H^{phys}\left(p\right)=-\sum_{i}^{m}\left(\ln p_i\right)p_i\,. \qquad (1.5.53)$$

В статистической физике используется единица измерений, имеющая размерность постоянной Больцмана. Тогда получим размерную физическую энтропию Больцмана–Гиббса $H\left(p\right)=kH^{phys}\left(p\right)$ [1, 6].

Энтропия Шеннона–Винера есть отношение физической энтропии к ее значению при равновероятном состоянии с $m=2$, то есть выполняется равенство

$$H\left(p\right)=\frac{H^{phys}\left(p\right)}{H^{phys}\left(\frac{1}{2},\frac{1}{2}\right)},\quad H^{phys}\left(\frac{1}{2},\frac{1}{2}\right)=\ln 2\,. \qquad (1.5.54)$$

Рассмотрим случай, когда состояния аддитивного объекта являются равновероятными и статистически независимыми при n наблюдениях. Тогда справедливы следующие выражения

$$p_i=\prod_{k}^{n}p_{ik},\quad \sum_{i}^{m}p_i=\sum_{i}^{m}p_{ik}=1, \qquad (1.5.55)$$

где $k=1,\ldots,n$ и распределение

$$p_{ik}=\frac{1}{m}\,. \qquad (1.5.56)$$

В итоге из (1.5.4) получим меру информации Хартли [77]:

$$H\left(p\right)=-\sum_{i}^{m}\left(\log_2 p_i\right)p_i=n\log_2 m\,. \qquad (1.5.57)$$

Из (1.5.57) следует, что информация есть логарифмическая мера от числа равновероятных состояний случайного объекта, пропорциональная числу наблюдений n, и составляет $\log_2 m$ бит.

9. f-энтропия. Функциональные обобщения энтропии определяются выражениями

$$H_f(p)=\sum_i^m f(p_i)\,p_i\,, \tag{1.5.58}$$

$$H_f(p)=\sum_i^m f(p_i), \tag{1.5.59}$$

$$H_f(p)=f\left[N(p)\right], \tag{1.5.60}$$

где f есть выпуклая функция.

При $f=-\log_2 p_i$, $f=-(\log_2 p_i)\,p_i$ и $f=-\log_2 N(p)$ из (1.5.58), (1.5.59) и (1.5.60) следует традиционная мера информации Шеннона–Винера.

Функционал (1.5.58) является средним значением случайной энтропии, выражение (1.5.59) с информационной функцией $f(p)$ рассматривалось в работах [53, 58], а (1.5.60) есть функция от геометрического среднего распределения.

1.6. Аксиомы Хинчина. Аксиомы Фаддеева и метод информационной функции Дароши

Единственность функционала Шеннона–Винера, совместимого с приведенными свойствами, доказал впервые аксиоматическим подходом А.Я. Хинчин [48, 49]. Были сформулированы для дискретного случая распределения $p=\{p_1,\dots,p_m\}$ основополагающие аксиомы:

1. $H(p_1,p_2,\dots,p_m)$ непрерывна относительно $p_1,p_2,\dots,p_m$ в области $0\le p_i\le 1$ и $\sum_i^m p_i=1$.

2. $H(p_1,p_2,\dots,p_m)$ симметрична относительно $p_1,p_2,\dots,p_m$.

3. $H(p_1,p_2,\dots,p_m,0)=H(p_1,p_2,\dots,p_m)$. Это означает, что добавление к множеству состояний невозможного состояния не изменяет неопределенность.

4. $H\left(p_{11},\ldots,p_{1n};p_{21},\ldots,p_{2n};\ldots;p_{m1},\ldots,p_{mn}\right)=$

$$=H\left(p_1,p_2,\ldots,p_m\right)+\sum_i^m p_i H_i\left(\frac{p_{i1}}{p_i},\ldots,\frac{p_{in}}{p_i}\right), \tag{1.6.1}$$

где p_{ij} – нормированное совместное распределение статистически зависимых объектов с

$$\sum_i^m\sum_j^n p_{ij}=1,\quad p_i=\sum_j^n p_{ij}\quad \left(p_{ij}\ge 0\right). \tag{1.6.2}$$

5. $H\left(p_1,p_2,\ldots,p_m\right)\le H\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right)$. Это означает, что функционал имеет наибольшее значение при $p_1=p_2=\ldots=p_m=1/m$.

Определяя энтропию аксиомами Хинчина с точностью до постоянного положительного множителя λ^{-1}, получим логарифмическую меру [48, 49]

$$H\left(p_1,p_2,\ldots,p_m\right)=-\lambda^{-1}\sum_i^m\left(\log_a p_i\right)p_i. \tag{1.6.3}$$

При натуральном логарифме и $\lambda=1$ имеем физическую безразмерную энтропию $H^{phys}\left(p\right)=-\sum_i^m\left(\ln p_i\right)p_i$.

6. Условие нормированности. Выбирая двоичную единицу энтропии с $a=2$ и учитывая условие нормированности

$$H\left(\frac{1}{2},\frac{1}{2}\right)=1, \tag{1.6.4}$$

окончательно получим функционал Шеннона–Винера

$$H\left(p\right)=-\sum_i^m\left(\log_2 p_i\right)p_i. \tag{1.6.5}$$

Д.К. Фаддеев [43] упростил систему аксиом Хинчина и предложил следующие аксиомы:

1. $H\left(p_1,p_2\right)=H\left(1-p,p\right)$ непрерывна при $0\le p\le 1$ и положительна хотя бы в одной точке.

2. $H\left(p_1,p_2,\ldots,p_m\right)$ симметрична относительно $p_1,p_2,\ldots,p_m$.

3. При $m\ge 2$

$$H\left(p_1,\ldots,p_{m-1},p_m,p_{m+1}\right)=H\left(p_1,\ldots,p_{m-1},p_m+p_{m+1}\right)+$$
$$+\left(p_m+p_{m+1}\right)H\left(\frac{p_m}{p_m+p_{m+1}},\frac{p_{m+1}}{p_m+p_{m+1}}\right),\quad p_m+p_{m+1}>0. \qquad (1.6.6)$$

Аксиомами Фаддеева энтропия при $m=2$ определяется однозначно с точностью до коэффициента λ^{-1} и имеет вид

$$H\left(p,\ 1-p\right)=-\lambda^{-1}\left[p\log_a p+\left(1-p\right)\log_a\left(1-p\right)\right]. \qquad (1.6.7)$$

Учитывая размерность и условие нормированности (1.5.50), имеем $a=2,\quad \lambda^{-1}=1$, а (1.6.7) запишется так:

$$H\left(p,\ 1-p\right)=-\left[p\log_2 p+\left(1-p\right)\log_2\left(1-p\right)\right]. \qquad (1.6.8)$$

Переход к общему случаю с $m>2$ осуществляется методом математической индукции на основании аксиомы 3.

В работе [43] отмечается, что разница в этих двух системах аксиом заключается в следующем. Во-первых, аксиома 5 (экстремальность) в системе Хинчина заменяется требованием положительности энтропии в одной точке и, во-вторых, аксиомы 3 и 4 заменяются одной аксиомой 3 системы Фаддеева, очень естественной, если рассмотреть энтропию как меру неопределенности состояний случайного объекта.

Далее З. Дароши [52, 53, 72] представил аксиому 3 системы Фаддеева в общем виде с групповым усреднением

$$H\left(p_1,\ldots,p_m\right)=\sum_{k=2}^{m}\left(p_1+p_2+\ldots+p_k\right)f\left(\frac{p_k}{p_1+p_2+\ldots+p_k}\right). \qquad (1.6.9)$$

Здесь так называемая информационная функция $f\left(x\right)$ при $m=2$ удовлетворяет граничным условиям

$$f\left(0\right)=f\left(1\right),\quad f\left(\frac{1}{2}\right)=1 \qquad (1.6.10)$$

и функциональному уравнению

$$f\left(x\right)+\left(1-x\right)f\left(\frac{y}{1-x}\right)=f\left(y\right)+\left(1-y\right)f\left(\frac{x}{1-y}\right) \qquad (1.6.11)$$

для всех $\left(x,y\right)\in D$, где

$$D=\left\{\left(x,y\right):0\le x<1,\quad 0\le y<1,\quad x+y\le 1\right\}. \qquad (1.6.12)$$

Решением уравнения (1.6.11) является информационная функция

Дароши, равная энтропии Шеннона–Винера

$$H(p,\ 1-p) = f(p) = -p\log_2 p - (1-p)\log_2(1-p). \qquad (1.6.13)$$

Такой подход с информационной функцией при $m=2$, зависящей от двух переменных, позволяет получить информацию различия Кульбака–Лейблера

$$I(1-p, p:1-u, u) = f(p,u) = p\log_2\frac{p}{u} + (1-p)\log_2\frac{1-p}{1-u}, \qquad (1.6.14)$$

которая удовлетворяет условию нормированности $I\left(1,0:\frac{1}{2},\frac{1}{2}\right)=1$.

Подробные математические выкладки метода информационной функции приводятся в монографии [53].

1.7. Определение информации в каналах связи. Формула Шеннона

Рассмотрим определение информации как количество снятой меры неопределенности о состояниях случайного сигнала в каналах связи. Пусть переданный сигнал характеризуется распределением $p_i\ (i=1,\ldots,m)$. Количество информации определяется априорной энтропией случайного сигнала [50]

$$H(p_1) = -\sum_i^m (\log_2 p_i)\, p_i, \quad \sum_i^m p_i = 1. \qquad (1.7.1)$$

Если полученный сигнал не равен переданному, то остается мера неопределенности после приема сигнала, характеризуемая условным распределением $p_{i|j}\ (j=1,\ldots,n)$. Соответствующая условная энтропия равняется

$$H_j\left(p_{1|2}\right) = -\sum_i^m \left(\log_2 p_{i|j}\right) p_{i|j}, \quad \sum_i^m p_{i|j} = 1, \qquad (1.7.2)$$

а ее среднее значение [50]

$$H\left(p_1|p_2\right) = -\sum_j^n p_j H_j\left(p_{1|2}\right) = -\sum_i^m \sum_j^n \left(\log_2 p_{i|j}\right) p_{i|j} p_j \qquad (1.7.3)$$

вычисляется по распределению принимаемого сигнала p_j. Выражение (1.7.3) есть апостериорная энтропия случайного сигнала.

Таким образом, информация, полученная после приема сигнала, равняется следующей разности [50]:

$$I = H\left(p_1\right) - H\left(p_1 | p_2\right) = -\sum_i^m \left(\log_2 p_i\right) p_i + \sum_i^m \sum_j^n \left(\log_2 p_{i|j}\right) p_{i|j} p_j \qquad (1.7.4)$$

между априорной и апостериорной энтропиями и отражает количество снятой неопределенности о случайном сигнале.

Перепишем (1.7.4) в следующем виде

$$I\left(p_{12} : p_1 p_2\right) = -\sum_i^m \left(\log_2 p_i\right) p_i + \sum_i^m \sum_j^n \left(\log_2 p_{i|j}\right) p_{ij} =$$
$$= \sum_i^m \sum_j^n \left(\log_2 \frac{p_{i|j}}{p_i}\right) p_{ij} = \sum_i^m \sum_j^n \left(\log_2 \frac{p_{ij}}{p_i p_j}\right) p_{ij}, \qquad (1.7.5)$$

где $p_{ij} = p_j p_{i|j}$ есть совместное распределение переданного и полученного сигналов. Так как справедливо равенство $p_{ij} = p_i p_{j|i}$, то для полученной информации имеем также выражение

$$I\left(p_{12} : p_1 p_2\right) = H\left(p_2\right) - H\left(p_2 | p_1\right) =$$
$$= -\sum_j^n \left(\log_2 p_j\right) p_j + \sum_i^m \sum_j^n \left(\log_2 p_{j|i}\right) p_i p_{j|i} =$$
$$= \sum_i^m \sum_j^n \left(\log_2 \frac{p_{j|i}}{p_j}\right) p_{ij} = \sum_i^m \sum_j^n \left(\log_2 \frac{p_{ij}}{p_i p_j}\right) p_{ij}. \qquad (1.7.6)$$

В качестве примера из теории связи рассмотрим непрерывный канал с шумом. Источник посылает в канал сигналы с полосой частот $\Delta\nu$ за время t. Пусть x есть переданное сигнальное напряжение, а $x' = x + y$ определяет полученный сигнал при наличии шумового напряжения y. Величины x и y являются независимыми. Ограничимся гауссовыми сигналами и шумом. Тогда запишем непрерывный аналог информации (1.7.5) в натах

$$I = -\int p\left(x\right) \ln p\left(x\right) dx + \int p\left(x'\right) \left[\int p\left(x | x'\right) \ln p\left(x | x'\right) dx\right] dx' =$$
$$= \iint p\left(x, x'\right) \ln \frac{p\left(x, x'\right)}{p\left(x\right) p\left(x'\right)} dx dx'. \qquad (1.7.7)$$

Используя выражения совместного нормального распределения:

$$p(x,x')=\frac{1}{\left\{2\pi\mathbf{D}(x')\mathbf{D}(x)\left[1-r^2(x,x')\right]\right\}^{1/2}}\times$$

$$\times\exp\left\{-\frac{1}{2\left[1-r^2(x,x')\right]}\left[\frac{x^2}{\mathbf{D}(x)}-2r(x,x')\frac{xx'}{\mathbf{D}(x)\mathbf{D}(x')}+\frac{x'^2}{\mathbf{D}(x')}\right]\right\}, \quad (1.7.8)$$

и частных нормальных распределений

$$p(x)=\frac{1}{\left\{2\pi\mathbf{D}(x)\right\}^{1/2}}\exp\left[-\frac{x^2}{\mathbf{D}(x)}\right], \quad (1.7.9)$$

$$p(x')=\frac{1}{\left\{2\pi\mathbf{D}(x')\right\}^{1/2}}\exp\left[-\frac{x'^2}{\mathbf{D}(x')}\right], \quad (1.7.10)$$

из (1.7.7) вытекает количество информации

$$I=-\frac{1}{2}\ln\left[1-r^2(x,x')\right]. \quad (1.7.11)$$

Для расхождения имеем выражение

$$J=\frac{r^2(x,x')}{1-r^2(x,x')}. \quad (1.7.12)$$

Здесь дисперсии и коэффициент корреляции равняются

$$\mathbf{D}(x)=\mathbf{E}\left(x^2\right)=\int x^2p(x)dx,\quad \mathbf{D}(x')=\mathbf{E}\left(x'^2\right)=\int x'^2p(x')dx', \quad (1.7.13)$$

$$r(x,x')=\frac{1}{\left[\mathbf{D}(x)\mathbf{D}(x')\right]^{1/2}}\iint xx'p(x,x')dxdx'. \quad (1.7.14)$$

Поскольку x и x' зависимы, то принимается

$$p(x,x')=p(x)\,p\left(x'|x\right)=p(x)\,p(x'-x), \quad (1.7.15)$$

где условное нормальное распределение

$$p\left(x'|x\right)=\frac{1}{\left\{2\pi\mathbf{D}(x')\left[1-r^2(x,x')\right]\right\}^{1/2}}\times$$

$$\times\exp\left\{-\frac{1}{2\mathbf{D}(x')\left[1-r^2(x,x')\right]}\left[x'-xr(x,x')\left(\frac{\mathbf{D}(x')}{\mathbf{D}(x)}\right)^{1/2}\right]^2\right\}. \quad (1.7.16)$$

Учитывая (1.7.15), получим равенства

$$r(x,x')\left(\frac{\mathbf{D}(x')}{\mathbf{D}(x)}\right)^{1/2}=1,\quad r^2(x,x')=\frac{\mathbf{D}(x')}{\mathbf{D}(x)}=\frac{P}{P+N}, \qquad (1.7.17)$$

где $P=\mathbf{E}(x^2)$ и $N=\mathbf{E}(y^2)$ есть средние значения мощности переданного сигнала и шума. Подставим коэффициент корреляции в (1.7.11) и (1.7.12), в результате имеем выражение для количества информации в полученном сигнале и значение расхождения

$$I=\frac{1}{2}\ln\left(1+\frac{P}{N}\right), \qquad (1.7.18)$$

$$J=\frac{P}{N}. \qquad (1.7.19)$$

Для определения переданного сигнала необходимо $n=2\Delta\nu t$ независимых наблюдений [30, 31, 50]. Общее количество информации равняется nI и в итоге имеем знаменитую формулу Шеннона для пропускной способности канала

$$C=\frac{nI}{t}=\Delta\nu\ln\left(1+\frac{P}{N}\right). \qquad (1.7.20)$$

Из (1.7.20) следует, что уменьшение величины мощности шума позволяет увеличить количество переданной информации.

Общее количество расхождения равняется

$$nJ=2E/N_0, \qquad (1.7.21)$$

где $E=Pt$ есть полная энергия переданного сигнала, а $N_0=N/\Delta\nu$ – средняя мощность шума на единицу полосы частот.

В случае исчисления информации в битах необходимо разделить (1.7.18) и (1.7.20) на $\ln 2$, что в итоге дает

$$I=\frac{1}{2}\log_2\left(1+\frac{P}{N}\right), \qquad (1.7.22)$$

$$C=\Delta\nu\log_2\left(1+\frac{P}{N}\right). \qquad (1.7.23)$$

Перейдем к рассмотрению основных свойств информации различия.

1.8. Информация различия Кульбака–Лейблера

К статистическим свойствам объекта относится мера информации в одном состоянии относительно другого состояния, которая выражается функционалом Кульбака–Лейблера [33, 88]. В предыдущем параграфе приводился частный вид информации различия, имеющей важную роль в теории связи. Рассмотрим общий случай.

Пусть статистические наблюдения ведутся с распределением $p=\{p_1,\dots,p_m\}$ относительно состояния с $p_0=\{p_{01},\dots,p_{0m}\}$. Тогда наблюдения характеризуются случайной информацией различия в виде разности случайных энтропий

$$I\left(p_i:u_i\right)=-\left[h\left(p_i\right)-h\left(u_i\right)\right]=\log_2\frac{p_i}{u_i}. \qquad (1.8.1)$$

Взвешенное среднее значение есть информация различия Кульбака-Лейблера [33, 88]

$$I\left(p:u\right)=\frac{\sum_i^m\left(\log_2\frac{p_i}{u_i}\right)p_i}{\sum_i^m p_i} \qquad (1.8.2)$$

или просто различающая информация.

При выполнении условия вероятностной нормировки

$$\sum_i^m p_i=1 \qquad (1.8.3)$$

взвешенное среднее (1.8.2) принимает следующий вид

$$I\left(p:u\right)=\sum_i^m\left(\log_2\frac{p_i}{u_i}\right)p_i,\qquad \sum_i^m u_i=1. \qquad (1.8.4)$$

Приведем основные свойства различающей информации.

1. Положительность и выпуклость. Информация различия есть вещественный, неотрицательный и выпуклый функционал, то есть для произвольных p и u имеем неравенства

$$I\left(p:u\right)\ge 0, \qquad (1.8.5)$$

$$I\left(\left(a_1p_1+a_2p_2\right):\left(a_1u_1+a_2u_2\right)\right)\le a_1I\left(p_1:u_1\right)+a_2I\left(p_2:u_2\right). \qquad (1.8.6)$$

Здесь $a_1 + a_2 = 1,\quad a_1 > 0,\quad a_2 > 0$ и информации различия с нормированными распределениями

$$I\left(p_1 : u_1\right) = \sum_i^m \left(\log_2 \frac{p_{1i}}{u_{1i}}\right) p_{1i}, \qquad \sum_i^m p_{1i} = 1, \tag{1.8.7}$$

$$I\left(p_2 : u_2\right) = \sum_i^m \left(\log_2 \frac{p_{2i}}{u_{2i}}\right) p_{2i}, \qquad \sum_i^m p_{2i} = 1. \tag{1.8.8}$$

Равенство в (1.8.5) достигается тогда и только тогда, когда $p = p_0$. Аналогичные неравенства справедливы для функционала (1.8.2).

2. Аддитивность для независимых объектов. Пусть два состояния объекта описываются нормированными совместными распределениями p_{12} и u_{12}. Информация различия имеет вид

$$I\left(p_{12} : u_{12}\right) = \sum_i^m \sum_j^n \left(\log_2 \frac{p_{ij}}{u_{ij}}\right) p_{ij}, \tag{1.8.9}$$

$$\sum_i^m \sum_j^n p_{ij} = \sum_i^m \sum_j^n u_{ij} = 1. \tag{1.8.10}$$

В случае статистической независимости состояний имеем равенства $p_{ij} = p_i p_j$ и $u_{ij} = u_i u_j$. Из (1.8.9) следует свойство аддитивности для информации различия

$$I\left(p_{12} : u_{12}\right) = I\left(p_1 : u_1\right) + I\left(p_2 : u_2\right), \tag{1.8.11}$$

где

$$I\left(p_1 : q_1\right) = \sum_i^m \left(\log_2 \frac{p_i}{q_i}\right) p_i, \quad I\left(p_2 : q_2\right) = \sum_j^n \left(\log_2 \frac{p_j}{q_j}\right) p_j. \tag{1.8.12}$$

3. Аддитивность для зависимых объектов. Рассмотрим переход от состояния с p_{ij} к состоянию с $u_{ij} = p_i p_j$. Тогда свойства аддитивности определяются в терминах условной энтропии. Для этого определим, согласно теоремам сложения и разложения (1.5.10), нормированные распределения:

$$p_i = \sum_j^n p_{ij}, \quad p_j = \sum_i^m p_{ij} \tag{1.8.13}$$

и условные распределения

$$p_{i|j} = \frac{p_{ij}}{p_j}, \quad p_{j|i} = \frac{p_{ij}}{p_i}. \tag{1.8.14}$$

Согласно (1.8.9), получим свойство аддитивности для зависимых систем в виде

$$I\left(p_{12} : p_1 p_2\right) = \sum_i^m \sum_j^n \left(\log_2 \frac{p_{ij}}{p_i p_j}\right) p_{ij} =$$
$$= H\left(p_1\right) + H\left(p_2\right) - H\left(p_{12}\right) =$$
$$= H\left(p_1\right) - H\left(p_1 \middle| p_2\right) = H\left(p_2\right) - H\left(p_2 \middle| p_1\right), \tag{1.8.15}$$

где соответствующие энтропии

$$H\left(p_1\right) = -\sum_i^m \left(\log_2 p_i\right) p_i, \quad H\left(p_2\right) = -\sum_j^n \left(\log_2 p_j\right) p_j, \tag{1.8.16}$$

$$H_j\left(p_{1|2}\right) = -\sum_i^m \left(\log_2 p_{i|j}\right) p_{i|j}, \quad H_i\left(p_{2|1}\right) = -\sum_j^n \left(\log_2 p_{j|i}\right) p_{j|i}, \tag{1.8.17}$$

$$H\left(p_2 \middle| p_1\right) = \sum_i^m p_i H_i\left(p_{2|1}\right), \quad H\left(p_1 \middle| p_2\right) = \sum_j^n p_j H_j\left(p_{1|2}\right). \tag{1.8.18}$$

Таким образом, информация различия равняется разности энтропии Шеннона–Винера и средней условной энтропии. В теории связи, как было показано в предыдущем параграфе, эта информация различия характеризует пропускную способность канала. Остановимся подробнее на ее некоторых свойствах, в которых состояние с p_i и $p_{j|i}$ определяют, соответственно, как переданный и полученный сигналы.

Симметричность информации. Из определения функционала (1.8.15) следует равенство

$$I\left(p_{12} : p_1 p_2\right) = I\left(p_{21} : p_2 p_1\right) \geq 0, \tag{1.8.19}$$

которое означает, что положительная информация об объекте, содержащаяся в переданном и принимаемом сигналах, одинакова. Другими словами, неопределенность, снятая при посылке того, какой сигнал будет получен, равна неопределенности снятой при приеме сигнала того, какой сигнал был отправлен.

Информация о самом сигнале. Если $p_{ij} = p_i \delta_{ij}$ (где δ_{ij} – символ Кронекера), то из (1.8.15) вытекает количество информации

$$I = H(p) = -\sum_i^m (\log_2 p_i) p_i \tag{1.8.20}$$

о переданном сигнале, равное энтропии Шеннона–Винера.

В общем случае из определения положительной различающей информации вытекают известные неравенства для энтропий (1.5.42) и (1.5.43).

Уменьшение информации. Пусть новый переданный сигнал о системе характеризуется распределением $p' = \{p'_1, \ldots, p'_m\}$. Количество полученной информации не увеличивается, а становится меньше исходной. Доказательство приводится в монографии К. Шеннона [50].

Далее рассмотрим общий случай с совместными распределениями p_{12} и $u_{12} \neq p_1 p_2$. Тогда используя распределения

$$p_i = \sum_j^n p_{ij}, \quad u_i = \sum_j^n u_{ij}, \tag{1.8.21}$$

$$p_{j|i} = \frac{p_{ij}}{p_j}, \quad u_{j|i} = \frac{u_{ij}}{u_i}, \tag{1.8.22}$$

получим следующее свойство аддитивности

$$I(p_{12} : u_{12}) = I(p_1 : u_1) + I\left(p_{2|1} \middle| u_{2|1}\right). \tag{1.8.23}$$

Первое слагаемое в (1.8.23) есть информация различия в наблюдениях p_1 относительно u_1, а второе – среднее значение по p_1

$$I\left(p_{2|1} \middle| u_{2|1}\right) = \sum_i^m I_i\left(p_{2|1} \middle| u_{2|1}\right) p_i \tag{1.8.24}$$

условной различающей информации:

$$I_i\left(p_{2|1}\middle|u_{2|1}\right)=\sum_j^n\left(\log_2\frac{p_{j|i}}{u_{j|i}}\right)p_{j|i}\,. \qquad (1.8.25)$$

При $u_{ij}=p_i p_j$ из (1.8.23) вытекает выражение (1.8.15).

Из (1.8.9) имеем свойство аддитивности в виде

$$I\left(p_{12}:u_{12}\right)=I\left(p_2:u_2\right)+I\left(p_{1|2}\middle|u_{1|2}\right), \qquad (1.8.26)$$

где

$$I\left(p_{1|2}\middle|u_{1|2}\right)=\sum_j^n I_j\left(p_{1|2}\middle|u_{1|2}\right)p_j\,, \qquad (1.8.27)$$

$$I_j\left(p_{1|2}\middle|u_{1|2}\right)=\sum_i^m\left(\log_2\frac{p_{i|j}}{u_{i|j}}\right)p_{i|j}\,. \qquad (1.8.28)$$

Окончательно получим равенство

$$I\left(p_1:u_1\right)-I\left(p_2:u_2\right)=I\left(p_{1|2}\middle|u_{1|2}\right)-I\left(p_{2|1}\middle|u_{2|1}\right). \qquad (1.8.29)$$

Перейдем к случаю трех зависимых объектов. Учитывая формулы для распределений, энтропий и информаций различия

$$I\left(p_1:p_2\right)=H\left(p_1\right)-H\left(p_1\middle|p_2\right)=\sum_i^m\sum_j^n\left(\log_2\frac{p_{i|j}}{p_i}\right)p_{ij}\,, \qquad (1.8.30)$$

$$I\left(p_1:p_2\middle|p_3\right)=H\left(p_1\middle|p_3\right)-H\left(p_1\middle|p_2,p_3\right)=$$

$$=\sum_i^m\sum_j^n\sum_k^r\left(\log_2\frac{p_{i,j|k}}{p_{i|k}}\right)p_{ijk}, \qquad (1.8.31)$$

$$I\left(p_{12}:p_3\right)=H\left(p_{12}\right)-H\left(p_1,p_1\middle|p_3\right)=$$

$$=\sum_i^m\sum_j^n\sum_k^r\left(\log_2\frac{p_{i,j|k}}{p_{ij}}\right)p_{ijk}, \qquad (1.8.32)$$

получим свойства аддитивности в следующем виде:

$$H\left(p_1\right)+H\left(p_2\right)+H\left(p_1,p_2\middle|p_3\right)=I\left(p_1:p_2\right)+I\left(p_{12}:p_3\right), \qquad (1.8.33)$$

$$I\left(p_1:p_3\right)+I\left(p_1:p_2\middle|p_3\right)=I\left(p_1:p_2\right)+I\left(p_1:p_3\middle|p_2\right)=I\left(p_{12}:p_3\right). \quad (1.8.34)$$

4. Флуктуация. Рассмотрим флуктуацию случайной различающей информации

$$\Delta I\left(p_i:u_i\right)=I\left(p_i:u_i\right)-I\left(p:u\right) \quad (1.8.35)$$

и запишем распределение

$$p_i=u_i 2^{\left\{I\left(p_i:u_i\right)+\Delta I\left(p_i:u_i\right)\right\}}. \quad (1.8.36)$$

Используя условие нормировки (1.8.3), получим

$$p_i=\frac{u_i 2^{\Delta I\left(p_i:u_i\right)}}{\sum_i^m u_i 2^{\Delta I\left(p_i:u_i\right)}} \quad (1.8.37)$$

и информацию различия Кульбака–Лейблера

$$I\left(p:p_0\right)=\log_2\sum_i^m p_{0i}2^{\Delta I_i\left(p:p_0\right)}. \quad (1.8.38)$$

Введем флуктуации случайных энтропий

$$\Delta h\left(p_i\right)=-\log_2 p_i-H\left(p\right),\quad \Delta h\left(u_i\right)=-\log_2 u_i-H\left(u\right) \quad (1.8.39)$$

и информации различия

$$\Delta I_i\left(p:p_0\right)=-\left[\Delta h_i\left(p\right)-\Delta h_i\left(p_0\right)\right]+\mathbf{E}\left[\Delta h_i\left(p_0\right)\right], \quad (1.8.40)$$

где среднее значение

$$\mathbf{E}\left[\Delta h\left(u\right)\right]=\sum_i^m\left[\Delta h\left(u_i\right)\right]p_i. \quad (1.8.41)$$

Тогда выражения (1.8.37) и (1.8.38) примут следующий вид

$$p_i=\frac{p_{0i}2^{\left\{-\left[\Delta h_i\left(p\right)-\Delta h_i\left(p_0\right)\right]\right\}}}{\sum_i^m p_{0i}2^{\left\{-\left[\Delta h_i\left(p\right)-\Delta h_i\left(p_0\right)\right]\right\}}}, \quad (1.8.42)$$

$$I\left(p:u\right)=-\left[H\left(p\right)-H\left(u\right)\right]+\mathbf{E}\left[\Delta h\left(u\right)\right]=$$

$$= \log_2 \frac{\sum_i^m 2^{-\Delta h(u_i)}}{\sum_i^m 2^{-\{\Delta h(u_i) - \Delta I(p_i : u_i)\}}} = \log_2 \frac{\sum_i^m 2^{-\Delta h_i(p_0)}}{\sum_i^m 2^{-\{\Delta h_i(p) - \mathbf{E}(\Delta h_i(p_0))\}}}. \qquad (1.8.43)$$

Если отсутствует флуктуация случайной информации различия, то из (1.8.42) и (1.8.43) вытекают равенства

$$p_i = u_i, \quad I(p : u) = 0. \qquad (1.8.44)$$

Таким образом, задавая значения флуктуаций случайных энтропий и информации различия, можно находить распределение и среднее значение рассматриваемых случайных величин.

5. Информация различия с $u_i = 1/m$. Подставим в информацию различия (1.8.2) распределение равновероятного состояния

$$u_i = 1/m, \qquad (1.8.45)$$

где m – число состояний. Согласно (1.8.5) получим неравенство

$$I(p : u) = \frac{\sum_i^m \left(\log_2 \frac{p_i}{u_i} \right) p_i}{\sum_i^m p_i} = -\frac{\left[H(p) - \log_2 m \right]}{\sum_i^m p_i} \geq 0, \qquad (1.8.46)$$

из которого следует, что энтропия равновероятного состояния больше, чем энтропия Шеннона–Винера произвольного состояния.

6. Неравенства. Информация различия удовлетворяет неравенствам

$$I(p_{12} : u_{12}) \leq I(p_1 : u_1) + I(p_2 : u_2), \qquad (1.8.47)$$

$$I(p_{12} : p_3) \geq I(p_1 : p_3 | p_2), \quad I(p_{12} : p_3) > I(p_2 : p_3 | p_1), \qquad (1.8.48)$$

$$I(p_1 : p_3) \geq I(p_1 : p_2), \quad I(p_1 : p_3) > I(p_2 : p_3), \qquad (1.8.49)$$

$$\begin{gathered} I(p_1 : p_3) \geq I(p_1 : p_3 | p_2), \\ I(p_{12} : p_3) > I(p_2 : p_3), \quad I(p_{12} : p_3) \geq I(p_1 : p_3). \end{gathered} \qquad (1.8.50)$$

7. Негэнтропийный принцип Бриллюэна. Пусть среднее значение случайной энтропии $h(u_i)$ для распределений p_i и u_i одинаковы. Тогда справедливы равенства

$$\mathbf{E}\left[\Delta h(u)\right]=\sum_i^m\left[h(u_i)-H(u)\right]p_i=-\sum_i^m\left(\log_2 u_i\right)p_i+\sum_i^m\left(\log_2 u_i\right)\ u_i=0, \quad (1.8.51)$$

а в итоге получим значение различающей информации

$$I\left(p:p_0\right)=-\left[H(p)-H\left(p_0\right)\right]=\log_2\frac{\sum_i^m 2^{-\Delta h_i(p_0)}}{\sum_i^m 2^{-\Delta h_i(p)}}. \quad (1.8.52)$$

Из (1.8.52) следует соотношение

$$H(p)=H(u)-I(p:u), \quad (1.8.53)$$

где информация различия представлена в виде отрицательного вклада в энтропию и поэтому называется негэнтропией. Понятие негэнтропии, то есть изменение энтропии с обратным знаком, было предложено Э. Шредингером [51]. В общем случае выполняется негэнтропийный принцип Бриллюэна [2, 3]

$$I(p:u)+\left[H(p)-H(u)\right]\geq 0, \quad (1.8.54)$$

где знак неравенства соответствует необратимым процессам, происходящим в случайном объекте.

Из (1.8.54) следует, что увеличение энтропии $H(p)$ до значения $H\left(p_0\right)$ происходит совместно с потерей информации различия $I(p:u)$. Н. Винер [5] также записал частный вид информации для равновероятных распределений с $p_i=1/m$ и $u_j=1/n$ в виде

$$I(p:u)=-\left[H(p)-H(u)\right]=$$
$$=\left[\sum_i^m\left(\log_2 p_i\right)p_i-\sum_j^n\left(\log_2 u_j\right)u_j\right]=\log_2\frac{m}{n} \quad (1.8.55)$$

логарифмической меры отношения чисел возможных состояний случайного объекта.

8. Расхождение. Определим количественную меру информации различия в наблюдениях u относительно p [33, 80]:

$$I(u:p)=\sum_i^m\left(\log_2\frac{u_i}{p_i}\right)u_i,\quad \sum_i^m u_i=\sum_i^m p_i=1 \tag{1.8.56}$$

и запишем следующий функционал

$$J(p:p_0)=I(p:p_0)+I(p_0:p)=\sum_i^m\left(\log_2\frac{p_i}{p_{0i}}\right)(p_i-p_{0i})\geq 0. \tag{1.8.57}$$

Величина $J(p:u)$, введенная в работе [80], называется расхождением и является логарифмической мерой трудности различия в наблюдениях p и u. Отметим некоторые свойства расхождения.

Расхождение обладает свойством симметричности относительно p и u, то есть $J(p:u)=J(u:p)$. Для информаций различия это свойство не выполняется $I(p:u)\neq I(u:p)$ и поэтому $I(p:u)$ и $I(u:p)$ можно рассматривать как направленные расхождения. Также функционал (1.8.57) является выпуклым и аддитивным для независимых объектов.

9. Мера неточности. К информационным свойствам объекта относится мера статистической неточности определения одного состояния случайного объекта относительно другого, которая выражается функционалом при аддитивности мер

$$H(p:u)=H(p)+I(p:u). \tag{1.8.58}$$

Из (1.8.58) следует, что мера неточности Керриджа [87]

$$H(p:u)=-\sum_i^m(\log_2 u_i)\,p_i \tag{1.8.59}$$

является средним значением случайной энтропии $h(u_i)=-\log_2 u_i$. Информация различия представляет собой отрицательный вклад в меру неточности и является, таким образом, информацией о снятой мере неточности.

При $u=p$ функционал (1.8.58) совпадает с энтропией Шеннона–Винера, то есть $H(p:p)=H(p)$.

На рис.1.1 представлены зависимости энтропии $H(p)$, информации различия $I(p:u)$ и меры неточности $H(p:u)$ от распределения при значениях $m=2$, $p_1=p$; а) $-u_1=1/3$, б) $-u_1=1/2$.

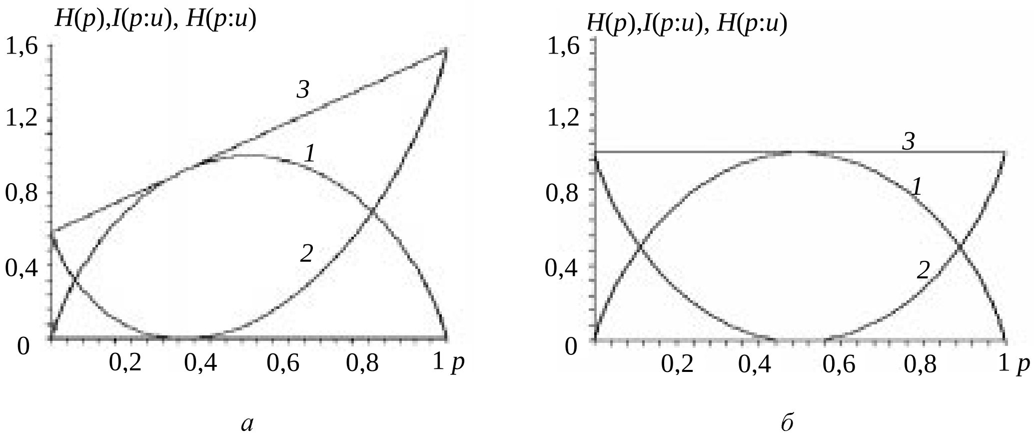

Рис.1.1. Зависимости функционалов модели Шеннона–Винера от распределения:
1 – энтропия $H(p)$, *2* – информация различия $I(p:u)$,
3 – мера неточности $H(p:u)$

Мера неточности изображается касательной к функции энтропии в точке u_1.

10. Информационный радиус. Геометрическая интерпретация неравенства (1.5.6) при $a_1 = a_2 = 1/2$ дает определение информационного радиуса [108]

$$R(p:u) = H\left(\frac{p+u}{2}\right) - \frac{1}{2}\left[H(p) + H(u)\right] =$$

$$= \sum_i^m \left\{\frac{1}{2}\left[\left(\log_2 p_i\right) p_i + \left(\log_2 u_i\right) u_i\right] - \frac{p_i + u_i}{2} \log_2 \frac{p_i + u_i}{2}\right\} \geq 0. \qquad (1.8.60)$$

Используя функционал (1.8.60), перепишем расхождение в виде

$$J(p:u) = 4\left[R(p:u) + T(p:u)\right], \qquad (1.8.61)$$

где второе слагаемое

$$T(p:u) = \sum_i^m \frac{p_i + u_i}{2} \log_2 \frac{p_i + u_i}{2 p_i^{1/2} u_i^{1/2}} \geq 0 \qquad (1.8.62)$$

введено и изучается в работе [121]. Функционалы (1.8.60) и (1.8.62) задаются также, как полусуммы информаций различия Кульбака–Лейблера [121]:

$$R(p:u)=\frac{1}{2}\left[I\left(p:\frac{p+u}{2}\right)+I\left(u:\frac{p+u}{2}\right)\right], \tag{1.8.63}$$

$$T(p:u)=\frac{1}{2}\left[I\left(\frac{p+u}{2}:p\right)+I\left(\frac{p+u}{2}:u\right)\right]. \tag{1.8.64}$$

11. Информационный коэффициент корреляции. Статистическая зависимость случайных объектов определяется информационным коэффициентом корреляции [92]

$$R_{12}=\sqrt{1-2^{-2I(p_{12}:p_1p_2)}}, \tag{1.8.65}$$

где информация различия $I(p_{12}:p_1p_2)$ выражается формулой (1.8.15). Информационный коэффициент корреляции имеет значение $R_{12}=0$ тогда и только тогда, когда объекты независимы. В этом случае совместное распределение равняется $p_{ij}=p_ip_j$, что приводит к равенству $I(p_{12}:p_1p_2)=0$. В теории передачи информации в системах связи информационный коэффициент корреляции совпадает с коэффициентом корреляции для переданного и полученного сигналов, то есть $R_{12}=r(x,x')$ (см. разд. 1.7).

Информационный коэффициент корреляции также записывается в эквивалентном виде

$$R_{12}=\sqrt{1-N^2\left(\frac{p_{12}}{p_1p_2}\right)} \tag{1.8.66}$$

или, разрешая (1.8.66) относительно среднего геометрического, получим

$$N\left(\frac{p_{12}}{p_1p_2}\right)=\sqrt{1-R_{12}^2}. \tag{1.8.67}$$

12. Информационное расстояние. Геометрическая интерпретация информации различия в виде $\delta(p,u)=\frac{1}{2}I^2(p:u)$ соответствует половине несимметричного расстояния от p до u. Тогда неравенство треугольника не выполняется и имеем несимметричный аналог теоремы Пифагора:

$$I(p:u)=I(p:w)+I(w:u). \tag{1.8.68}$$

Равенство в (1.8.68) достигается тогда и только тогда, когда справедливо условие

$$\sum_{i}^{m}\left(\log_2\frac{w_i}{u_i}\right)p_i=\sum_{i}^{m}\left(\log_2\frac{w_i}{u_i}\right)w_i\,. \tag{1.8.69}$$

В рассматриваемой интерпретации имеет место тождество параллелограмма

$$I(p:u)+I(w:u)=I\left(p:\frac{p+w}{2}\right)+I\left(w:\frac{p+w}{2}\right)+2I\left(\frac{p+w}{2}:u\right). \tag{1.8.70}$$

Вопрос об информационных расстояниях является частным случаем общей проблемы статистических расстояний [71, 91, 125].

13. Нормированность и размерность. Нормированность информации различия определяется равенством

$$I\left(1,0:\frac{1}{2},\frac{1}{2}\right)=\sum_{i}^{2}\left(\log_2\frac{p_i}{u_i}\right)p_i=1, \tag{1.8.71}$$

которое выполняется тождественно. Информация различия есть отношение физической различающей информации к значению физической энтропии $H^{phys}(u)$ при равновероятном состоянии с $m=2$

$$I(p:u)=\frac{I^{phys}(p:u)}{H^{phys}\left(\frac{1}{2},\frac{1}{2}\right)}, \tag{1.8.72}$$

где

$$I^{phys}(p:u)=\sum_{i}^{m}\left(\ln\frac{p_i}{u_i}\right)p_i\,,\ H^{phys}(u)=-\sum_{i}^{2}(\ln u_i)u_i\,, \tag{1.8.73}$$

$$H^{phys}\left(\frac{1}{2},\frac{1}{2}\right)=I^{phys}\left(1,0:\frac{1}{2},\frac{1}{2}\right)=\ln 2\,. \tag{1.8.74}$$

В статистической физике используется размерная физическая информация различия $I(p:u)=kI^{phys}(p:u)$ [19, 21].

14. f-информация различия. Функциональные обобщения информации различия определяются выражениями:

$$I_f\left(p:u\right)=\sum_i^m f\left(\frac{u_i}{p_i}\right)p_i\,, \tag{1.8.75}$$

$$I_f\left(p:u\right)=\sum_i^m f\left(p_i,u_i\right), \tag{1.8.76}$$

где вводится выпуклая функция f. В частности, в теории передачи информации по каналам связи из (1.8.75) имеем следующую f-информацию различия

$$I_f\left(p_{12}:p_1p_2\right)=\sum_i^m\sum_j^n f\left(\frac{p_{ij}}{p_ip_j}\right)p_{ij}\,. \tag{1.8.77}$$

Для функции $f=-\left[h\left(p_i\right)-h\left(u_i\right)\right]$, являющейся случайной информацией различия, из (1.8.75) вытекает известная мера Кульбака–Лейблера. Функционал (1.8.75) представляет собой среднее значение от выпуклой функции, а функция $f\left(p,u\right)$ в (1.8.76) есть информационная функция. Свойства информаций различия (1.8.75) и (1.8.77) детально исследованы в работах [69 – 71, 125].

В случае трех распределений функционал (1.8.75) запишется так [122]:

$$I_f\left(p:w:u\right)=\sum_i^m f\left(\frac{w_i}{u_i}\right)p_i\,. \tag{1.8.78}$$

Если $f=\log_2\left(w_i/u_i\right)$ есть случайная информация различия, то обобщенное выражение [122]

$$I\left(p:w:u\right)=\sum_i^m\left(\log_2\frac{w_i}{u_i}\right)p_i=I\left(p:w\right)-I\left(p:u\right)=$$
$$=-\left[H\left(p:w\right)-H\left(p:u\right)\right] \tag{1.8.79}$$

представляет собой разность информаций различия либо разность мер неточности.

Определим также f-информации различия:

$$I_f\left(p:u\right)=f\left[N\left(\frac{p}{u}\right)\right],\ I_f\left(p:w:u\right)=f\left[N\left(\frac{w}{u}\right)\right], \tag{1.8.80}$$

где среднее геометрическое распределение

$$N\left(\frac{w}{u}\right) = 2^{\sum_i^m \left(\log_2 \frac{w_i}{u_i}\right) p_i} . \qquad (1.8.81)$$

Пусть f есть логарифмическая функция, тогда из (1.8.80) следуют выражения (1.8.4) и (1.8.79). При $w = p$ из (1.8.79) и (1.8.80) вытекают, соответственно, информация различия Кульбака–Лейблера и f-информация различия (1.8.75).

1.9. Экстремальные свойства мер информации

Приведем экстремальные свойства мер информации в состояниях объекта, которые позволяют находить наиболее вероятное распределение. Принцип максимума энтропии был сформулирован впервые в статистической механике [6], а принцип минимума различающей информации – в математической статистике [33]. Оба принципа широко используются в теории информации, а автор применил их в теории самоорганизации открытых физических систем [14 – 24, 130, 131].

Рассмотрим случай экстремума энтропии Шеннона–Винера

$$H(p) = -\sum_i^m \left(\log_2 p_i\right) p_i \qquad (1.9.1)$$

при сохранении нормировки

$$\sum_i^m p_i = 1. \qquad (1.9.2)$$

Согласно вариационному методу, вычислим безусловный экстремум функционала

$$L = -\sum_i^m \left(\log_2 p_i\right) p_i - \alpha \sum_i^m p_i , \qquad (1.9.3)$$

где α – неопределенный множитель Лагранжа.

Из условия равенства нулю первой вариации функционала

$$\delta L = -\sum_i^m \delta p_i \left(\log_2 p_i + \frac{1}{\ln 2} \right) - \alpha \sum_i^m \delta p_i = 0 \qquad (1.9.4)$$

следует $p_i = \text{const}$. Учитывая условие нормировки, находим равновероятное распределение и соответствующее значение энтропии Хартли:

$$p_i = \frac{1}{m}, \quad H(p) = \log_2 m\,. \tag{1.9.5}$$

Пусть в макроскопическом опыте наблюдается среднее значение

$$\mathbf{E}(T) = \sum_i^m T_i p_i \tag{1.9.6}$$

случайной величины $T = \{T_1, \ldots, T_m\}$. Тогда имеем дополнительное ограничение на распределение p. Определение этого распределения соответствует задаче на экстремум энтропии при дополнительных условиях нормировки (1.9.2) и равенстве (1.9.6). Находим безусловный экстремум функционала

$$L = -\sum_i^m \left(\log_2 p_i\right) p_i + \tau \sum_i^m T_i p_i - \alpha \sum_i^m p_i\,, \tag{1.9.7}$$

где α и τ есть множители Лагранжа.

Из условия

$$\delta L = -\sum_i^m \delta p_i \left(\log_2 p_i + \frac{1}{\ln 2} - \tau T_i + \alpha\right) = 0 \tag{1.9.8}$$

вытекает распределение

$$p_i = 2^{\tau T_i} \mathit{\Gamma}^{-1}(\tau),\ \mathit{\Gamma}(\tau) = \sum_i^m 2^{\tau T_i} \tag{1.9.9}$$

с параметром τ, который меняется в пределах допустимых значений. Функция $\mathit{\Gamma}(\tau)$ имеет следующие свойства

$$\frac{\partial \log_2 \mathit{\Gamma}(\tau)}{\partial \tau} = \mathbf{E}(T), \quad \frac{\partial}{\partial (1/\tau)}\left[\frac{1}{\tau}\log_2 \mathit{\Gamma}(\tau)\right] = H(p). \tag{1.9.10}$$

Согласно (1.9.9) и (1.9.10), получим энтропию Шеннона–Винера в экстремуме

$$H(p) = -\sum_i^m \left(\log_2 p_i\right) p_i = -\tau \mathbf{E}(T) + \log_2 \mathit{\Gamma}(\tau) \tag{1.9.11}$$

и соответствующее дифференциальное соотношение:

$$dH(p) = -\tau d\mathbf{E}(T) \tag{1.9.12}$$

взаимосвязи со средним значением случайной величины T.

Для флуктуации случайной величины имеем формулу

$$-\frac{\partial h(p_i)}{\partial \tau} = T_i - \mathbf{E}(T),\; h(p_i) = -\log_2 p_i, \tag{1.9.13}$$

использование которой позволяет находить произвольные центральные моменты.

Далее рассмотрим экстремум информации различия Кульбака–Лейблера

$$I(p:u) = \sum_i^m \left(\log_2 \frac{p_i}{u_i}\right) p_i \tag{1.9.14}$$

для двух выше рассмотренных случаев и получим, соответственно, распределения

$$p_i = u_i, \tag{1.9.15}$$

$$p_i = u_i 2^{\tau T_i} \Gamma^{-1}(\tau),\quad \Gamma(\tau) = \sum_i^m 2^{\tau T_i} u_i \tag{1.9.16}$$

и свойства функции $\Gamma(\tau)$

$$\frac{\partial \log_2 \Gamma(\tau)}{\partial \tau} = \mathbf{E}(T),\quad -\frac{\partial}{\partial(1/\tau)}\left[\frac{1}{\tau}\log_2 \Gamma(\tau)\right] = I(p:u). \tag{1.9.17}$$

Для производной случайной информации различия имеем соотношение

$$\frac{\partial I(p_i : u_i)}{\partial \tau} = T_i - \mathbf{E}(T). \tag{1.9.18}$$

Распределение (1.9.16) дает экстремальные значения информации различия и расхождения

$$I(p:u) = \sum_i^m \left(\log_2 \frac{p_i}{u_i}\right) p_i = \tau \mathbf{E}(T) - \log_2 \Gamma(\tau), \tag{1.9.19}$$

$$I(u:p) = \sum_i^m \left(\log_2 \frac{u_i}{p_i}\right) u_i = -\tau \mathbf{E}_0(T) + \log_2 \Gamma(\tau), \tag{1.9.20}$$

$$J(p:u) = I(p:u) + I(u:p) = \tau\left[\mathbf{E}(T) - \mathbf{E}_0(T)\right]. \tag{1.9.21}$$

Пусть распределение p произвольное, а распределение u принадлежит параметрическому семейству распределений

$$u_i = 2^{\tau_0 T_i} \Gamma^{-1}(\tau_0), \quad \Gamma(\tau_0) = \sum_i^m 2^{\tau_0 T_i} . \tag{1.9.22}$$

Тогда, используем значения энтропии Шеннона–Винера и среднего значения величины T при усреднении с распределением u

$$H(u) = -\sum_i^m (\log_2 u_i) u_i \quad , \quad \mathbf{E}_0(T) = \sum_i^m T_i u_i \tag{1.9.23}$$

и перепишем информацию различия (1.9.19) в следующем виде:

$$I(p:u) = -\left[H(p) - H(u)\right] - \tau_0 \left[\mathbf{E}(T) - \mathbf{E}_0(T)\right]. \tag{1.9.24}$$

В дифференциальной форме (1.9.24) запишется так:

$$dI(p:u) = -dH(p) - \tau_0 d\mathbf{E}(T). \tag{1.9.25}$$

Отметим некоторые свойства функции $I(p:u)$ и среднего значения $\mathbf{E}(T)$, зависящих от τ.

1. Функция $I(p:u)$ является выпуклой, то есть $I(p:u) \geq 0$. Равенство достигается тогда, когда $\tau = 0$.

2. Функция $I(p:u)$ монотонно возрастает при $\mathbf{E}(T) \geq \mathbf{E}(T)_{\tau=0}$ и монотонно убывает при $\mathbf{E}(T) \leq \mathbf{E}(T)_{\tau=0}$.

Наконец, рассмотрим вторую вариацию функционалов и получим соответствующие неравенства

$$\delta^2 L = -\frac{1}{\ln 2} \sum_i^m \frac{(\delta p_i)^2}{p_i} \leq 0, \tag{1.9.26}$$

$$\delta^2 L = \frac{1}{\ln 2} \sum_i^m \frac{(\delta p_i)^2}{p_{0i} p_i} \geq 0. \tag{1.9.27}$$

Из (1.9.26) следует, что экстремум энтропии Шеннона–Винера достигается в ее максимуме. В случае информации различия Кульбака–Лейблера вторая вариация (1.9.27) является положительной, что указывает на ее минимум. Аналогично минимум достигается и для расхождения.

1.10. Информация Фишера. Неравенство Рао–Крамера

Рассмотрим семейство распределений одного и того же функционального вида, которые различаются значениями одномерного или многомерного параметра θ, характеризующего наблюдение. Пространство параметров есть пространство всех допустимых параметрических точек. Для двух точек $\theta = \{\theta_1, \ldots, \theta_n\}$ и $\theta' = \{\theta'_1, \ldots, \theta'_n\}$ в n-мерном пространстве имеем распределения $p_i(\theta)$ и $p_i(\theta')$, соответственно.

Мера информации в состоянии $p_i(\theta)$ относительно состояния $p_i(\theta')$ выражается информацией различия Кульбака–Лейблера [33]

$$I(\theta:\theta') = \sum_i^m \left(\log_2 \frac{p_i(\theta)}{p_i(\theta')} \right) p_i(\theta). \qquad (1.10.1)$$

При малом отклонении точек θ и $\theta' = \theta + \delta\theta$ пространства имеем распределения $p_i(\theta)$ и $p_i(\theta + \delta\theta)$. Тогда случайная информация различия принимает следующий вид

$$\begin{aligned} I_i(\theta:\theta+\delta\theta) &= -\left[s_i(\theta) - s_i(\theta+\delta\theta) \right] = \\ &= -\sum_k^n \delta\theta_k \frac{\partial \log_2 p_i(\theta)}{\partial \theta_k} - \frac{1}{2} \sum_k^n \sum_\ell^n \delta\theta_k \delta\theta_\ell \frac{\partial^2 \log_2 p_i(\theta)}{\partial\theta_k \partial\theta_\ell} \end{aligned} \qquad (1.10.2)$$

при разложении в ряд Тейлора по θ. Пренебрегая членами выше второго порядка в (1.10.2) и учитывая равенства

$$\frac{\partial \log_2 p_i(\theta)}{\partial \theta_k} = \frac{1}{(\ln 2)\, p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_k}, \qquad (1.10.3)$$

$$\frac{\partial^2 \log_2 p_i(\theta)}{\partial\theta_k \partial\theta_\ell} = \frac{1}{\ln 2} \left[\frac{1}{p_i(\theta)} \frac{\partial^2 p_i(\theta)}{\partial\theta_k \partial\theta_\ell} - \frac{1}{p_i^2(\theta)} \frac{\partial p_i(\theta)}{\partial\theta_k} \frac{\partial p_i(\theta)}{\partial\theta_\ell} \right], \qquad (1.10.4)$$

получим значение информации различия

$$I(\theta:\theta+\delta\theta) = \frac{1}{2\ln 2} \sum_k^n \sum_\ell^n \Gamma_{k\ell} \delta\theta_k \delta\theta_\ell, \qquad (1.10.5)$$

где

$$\Gamma_{k\ell} = \sum_{i}^{m} \left(\frac{\partial \ln p_i(\theta)}{\partial \theta_k} \frac{\partial \ln p_i(\theta)}{\partial \theta_\ell} \right) p_i(\theta) \qquad (1.10.6)$$

есть положительно определенная корреляционная матрица. Для расхождения имеем значения

$$J(\theta : \theta + \delta\theta) = 2I(\theta, \theta + \delta\theta) = \frac{1}{\ln 2} \sum_{k}^{n} \sum_{\ell}^{n} \Gamma_{k\ell} \delta\theta_k \delta\theta_\ell . \qquad (1.10.7)$$

Матрица $\Gamma_{k\ell}$ является фундаментальной информационной матрицей Фишера в математической статистике [32, 33, 39, 74, 75].

Рассмотрим случай изменения одномерного параметра. По определению параметр не испытывает флуктуаций и, следовательно, не имеет статистических характеристик. Однако можно поставить задачу об оценке нефлуктуирующего параметра по результатам наблюдений, подверженных статистическим отклонениям. При этом используются идеи теории параметрического оценивания.

Пусть случайная величина θ_i^*, которая не зависит от θ, выражается через известные дискретные величины, характеризующие случайный объект. Такая величина называется оценкой истинного значения нефлуктуирующего параметра θ. При выборе оценки естественно предположить, что ее среднее значение $\mathbf{E}(\theta^*)$ было близко к параметру θ. В общем случае оценка имеет смещение $b(\theta)$, зависящее от θ, и выражается так:

$$\mathbf{E}(\theta^*) = \sum_{i}^{m} \theta_i^* p_i(\theta) = \theta + b(\theta). \qquad (1.10.8)$$

При равенстве $\mathbf{E}(\theta^*) = \theta$ данная оценка представляет собой несмещенную оценку, при которой отсутствует систематическая ошибка наблюдений, обусловленная величиной $b(\theta)$. Для дисперсии оценки справедливо фундаментальное информационное неравенство Рао–Крамера [32, 39]:

$$\mathbf{D}\left(\theta^*\right) \geq \mathbf{D}\left(\theta^*\right)_{\min} = \frac{\left[1 + \frac{\partial b(\theta)}{\partial\theta}\right]^2}{\sum\limits_i^m \left[\frac{\partial \ln p_i(\theta)}{\partial\theta}\right]^2 p_i(\theta)}, \tag{1.10.9}$$

дающее ее нижнюю грань. Некоторые обобщения для (1.10.9) приводятся в [9-13]. При $b(\theta) = 0$ неравенство (1.10.9) обращается в

$$\mathbf{D}\left(\theta^*\right) \geq \mathbf{D}\left(\theta^*\right)_{\min} = \Gamma_{\theta\theta}^{-1}. \tag{1.10.10}$$

Величина $\Gamma_{\theta\theta}$ соответствует количеству информации Фишера о величине θ, которая содержится в оценке θ^*. Чем выше информация, тем меньше дисперсия и, соответственно, выше точность наблюдения. Это связано с тем, что дисперсия определяет меру точности статистических наблюдений в случайном объекте.

Мера информации Фишера

$$\Gamma_{\theta\theta} = \sum_i^m \left[\frac{\partial \ln p_i(\theta)}{\partial\theta}\right]^2 p_i(\theta) \tag{1.10.11}$$

характеризует предельное значение информации различия

$$I(\theta : \theta + \delta\theta) = \frac{1}{2\ln 2}\Gamma_{\theta\theta}(\delta\theta)^2 = \frac{(\delta\theta)^2}{2}\sum_i^m \frac{\partial^2 h\left[p_i(\theta)\right]}{\partial\theta^2} p_i(\theta). \tag{1.10.12}$$

1.11. Физические приложения

Пусть случайный объект представляет собой частицу в статистической физике, который имеет состояния с распределением вероятностей $p = \{p_1, \ldots, p_m\}$. Максимальное значение физической размерной энтропии Больцмана–Гиббса [6]

$$H(p) = -k\sum_i^m \left(\ln p_i\right) p_i = \beta(E - F) \tag{1.11.1}$$

достигается, согласно (1.9.9), при равновесном каноническом распределении Гиббса:

$$p_i = \exp\left\{-k^{-1}\beta\left(H_i - F\right)\right\}. \tag{1.11.2}$$

Здесь $\beta = -k\,\tau/\ln 2 = 1/T$ есть обратная абсолютная температура, H_i – дискретные значения энергии частицы, $E = \mathbf{E}\left(H\right)$ – средняя энергия частицы. Свободная энергия

$$F = -kT\ln\sum_i^m \exp\left\{-k^{-1}\beta H_i\right\} \tag{1.11.3}$$

определяется из условия нормировки распределения.

Дифференциальные соотношения равновесной статистической термодинамики замкнутых систем имеют, согласно (1.9.10) и (1.9.12), следующий вид

$$TdH\left(p\right) = dE\,, \quad dF = -H\left(p\right)dT\,. \tag{1.11.4}$$

Для открытых систем, находящихся в окружении с температурой T_0, физическая размерная информация различия имеет, согласно (1.9.24), значение [12]

$$I\left(p : p_0\right) = k\sum_i^m \left(\ln\frac{p_i}{p_{0i}}\right)p_i = -\delta H + \frac{1}{T_0}\delta E =$$
$$= -\left[H\left(p\right) - H\left(p_0\right)\right] + \frac{1}{T_0}\left(E - E_0\right) \geq 0 \tag{1.11.5}$$

с равенством тогда и только тогда, когда $p_i = p_{0i}$. Здесь для неравновесной и равновесной энтропий и, соответственно, средних энергий имеем

$$H\left(p\right) = -k\sum_i^m\left(\ln p_i\right)p_i\,,\ H\left(p_0\right) = -k\sum_i^m\left(\ln p_{0i}\right)p_{0i}\,, \tag{1.11.6}$$

$$E = \sum_i^m H_i p_i\,,\ E_0 = \sum_i^m H_i p_{0i}\,. \tag{1.11.7}$$

Состояние полного равновесия с окружением описывается каноническим распределением

$$p_{0i} = \exp\left\{-k^{-1}\beta_0\left(H_i - F_0\right)\right\} \tag{1.11.8}$$

с обратной температурой $\beta_0 = 1/T_0$ и свободной энергией F_0.

Использование диаграммы энтропия-энергия придает простой и наглядный смысл информации различия. На рис.1.2 приведено геометрическое представление термодинамических величин [12].

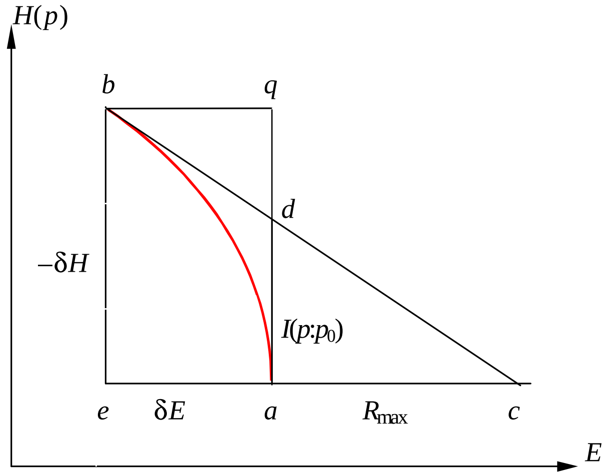

Рис.1.2. Энтропия-энергия для физических систем

Кривая ab изображает изменение функции $H(E)$ от произвольного состояния, соответствующего точке $a=(H,E)$, к полному равновесному в точке $b=(S_0,E_0)$. Из точки b проводится касательная прямая bc. Тогда, согласно (1.11.5), отрезок ad, параллельный оси $H(p)$, есть информация различия Кульбака–Лейблера I. Отрезок ac, параллельный оси E, изображает максимальную работу $R_{\max}$, выполняемую системой над окружением. Минимальная работа $R_{\min}=-R_{\max}$ связана с изменением полной энтропии замкнутой системы (система+окружение) от своего наибольшего значения, если система неравновесна, выражением $\Delta H^{\text{п}}=-R_{\min}/T_0$.

В первом случае сравним значения энтропий произвольного и полного равновесного состояний системы при одинаковых средних значениях энергии, что соответствует условию Гиббса

$$\sum_i^m H_i p_i=\sum_i^m H_i p_{0i}\,. \tag{1.11.9}$$

Перепишем (1.11.5) в следующем виде

$$I(p:p_0)=-\left[H(p)-H(p_0)\right]\geq 0\,. \tag{1.11.10}$$

Так как информация различия есть знакоопределенный функционал, то из (1.11.10) следует теорема Гиббса о максимуме энтропии равновесного состояния [6]

$$H(p) \geq H(p_0). \tag{1.11.11}$$

Таким образом, не прибегая к рассмотрению временной эволюции энтропии, можно определить изменения ее при переходах между различными состояниями системы. Причем энтропия приобретает свойства знакоопределенного функционала только при выполнении дополнительного условия постоянства средней энергии (1.11.9)

Принцип максимума энтропии показывает, что при спонтанном переходе от произвольного неравновесного состояния к равновесному степень неопределенности (разупорядоченности) системы увеличивается и при равновесии достигает максимального значения.

Для необратимых информационных физических процессов к равенству (1.11.5) добавляется знак неравенства и в дифференциальном виде имеем [12]

$$dI(p:p_0) \geq -dH(p) + \frac{1}{T_0} dE. \tag{1.11.12}$$

Рассматриваемый случай соответствует такой физической ситуации в термодинамике незамкнутых систем, когда в системе не происходит изменения энергии $(dE = 0)$. Из (1.11.12) вытекает выражение негэнтропийного принципа Бриллюэна [2, 3]

$$dI(p:p_0) + dH(p) \geq 0, \tag{1.11.13}$$

обобщающего принцип Карно–Клаузиуса о возрастании энтропии для замкнутых систем

$$dH(p) \geq 0. \tag{1.11.14}$$

При отсутствии изменения работы $(\delta R = 0)$ выполняется общий принцип уменьшения информации различия [12]

$$dI \leq 0. \tag{1.11.15}$$

Таким образом, взаимное изменение энтропии и физической информации различия сопровождается потерей информации (упорядоченности) состояний системы. Из (1.11.13) следует, что при необратимых явлениях увеличение энтропии $dH(p) > -dI(p:p_0)$ больше, чем уменьшение различающей информации. Следовательно, происходит самораспад неравновесной системы и спонтанные переходы приведут в конце концов к полному равновесному состоянию незамкнутой системы с максимальной неопределенностью (разупорядоченностью) в состояниях.

Согласно Бриллюэну [2, 3], изменение работы для рассматриваемого случая на величину kT_0 соответствует одной натуральной единице измерения количества информации (1 нит) в тепловой среде.

Во втором случае сравниваем значения средних энергий при одинаковых энтропиях

$$H(p) = H(p_0). \tag{1.11.16}$$

Тогда имеем выражение принципа минимума средней энергии в равновесном состоянии

$$R_{\max} = (E - E_0) \geq 0\,. \tag{1.11.17}$$

Диаграмма энтропия-энергия позволяет найти геометрическое представление коэффициенту полезного действия (КПД) преобразования энергии для обратимых процессов. Известно, что КПД преобразования энергии выражается отношением совершенной максимальной работы открытой системы к затраченной энергии в окружении

$$\eta = \frac{R_{\max}}{R_{\max} + \delta E}\,. \tag{1.11.18}$$

Данное равенство представляет собой отношение отрезков ac/ec, которое можно записать в виде $\eta = ad/eb$. Тогда КПД преобразования энергии выражается как отношение информации различия к изменению энтропии при переходе от произвольного состояния к равновесному [12]

$$\eta = \frac{I}{-\delta H} = \frac{I}{H_0 - S} \leq 1\,. \tag{1.11.19}$$

С другой стороны, интерпретация неравенства (1.11.19) на микроскопическом уровне означает, что степень упорядоченности состояний открытой системы при спонтанном переходе системы меньше, чем изменение степени разупорядоченности.

Информация различия Кульбака является знакоопределенной функцией Ляпунова. Поэтому, чтобы состояние полного равновесия было устойчивым, необходимо выполнение следующего неравенства для производной:

$$\frac{dI\left(p:p_0\right)}{dt}=-\frac{d\left(H\left(p\right)-H\left(p_0\right)\right)}{dt}\le 0\,. \qquad (1.11.20)$$

Из (1.11.20) следует закон временной эволюции энтропии (H –теорема Больцмана [1])

$$\frac{dH\left(p\right)}{dt}\ge 0 \qquad (1.11.21)$$

при приближении к состоянию полного равновесия. Происходит самораспад макроскопической системы при спонтанных переходах.

Пусть система находится в локальном (частичном) равновесии с температурой T. Подставляя (1.11.4) в основное уравнение (1.11.12), получим

$$dI\left(p:p_0\right)\ge-\left(\frac{1}{T}-\frac{1}{T_0}\right)dE,, \qquad (1.11.22)$$

где знак неравенства соответствует необратимым информационным процессам при контакте локально равновесной системы с окружением.