Решения задач из главы 2 книги Thomas M. Cover, Joy A. Thomas Elements of Information Theory, Second Edition

Урманов Максим Тимурович, ПМИ ФКН, группа 171-2

Июль 2019 г.

3'. Найти максимальное возможное значение энтропии для дискретного распределения с n значениями.

$\triangleright$ Пусть p_i — вероятность i-го значения. Тогда энтропия записывается в виде

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i, \text{ где } \sum_{i=1}^{n} p_i = 1.$$

Покажем, что максимум энтропии равен $\log n$ и достигается тогда и только тогда, когда все p_i равны $1/n$.

Рассмотрим произвольный набор неотрицательных значений p_i, где $\sum\limits_{i=1}^{n} p_i = 1$. Пусть не все p_i равны $1/n$.

Попробуем заменить часть значений p_i так, чтобы сумма всех p_i осталась прежней, а энтропия увеличилась. Так как не все p_i равны $1/n$, найдутся такие m и k, что $p_m < 1/n < p_k$. Заменим p_m и p_k на $1/n$ и $s - 1/n$ соответственно и покажем, что после такой замены энтропия увеличится. Для краткости обозначим $s = p_m + p_k$ и рассмотрим функцию

$$f(x) = -x \log x - (s - x) \log(s - x),$$

определённую на отрезке $[0, s]$. Очевидно, $f(p_m) = f(p_k) = p_m \log p_m + p_k \log p_k$ и нужно лишь доказать, что $f(1/n) > f(p_m)$. Для этого найдём производную функции f:

$$f'(x) = -x \cdot \frac{1}{x} - 1 \cdot \log x - (s - x) \cdot \frac{-1}{s - x} - (-1) \cdot \log(s - x) = \log(s - x) - \log x.$$

Отсюда видно, что f возрастает при $0 < x < \dfrac{s}{2}$ и убывает при $\dfrac{s}{2} < x < s$. Так как $p_m < \dfrac{s}{2} < p_k$, то f возрастает от p_m до $\dfrac{s}{2}$ и убывает от $\dfrac{s}{2}$ до p_k. Это значит, что на всём интервале (p_m, p_k) выполнено $f(x) > f(p_m) = f(p_k)$. Осталось заметить, что $p_m < \dfrac{1}{n} < p_k$, откуда $f(1/n) > f(p_m) = f(p_k)$, что и требовалось доказать.

Таким образом, в результате замены $p_m,\ p_k \to 1/n,\ s - 1/n$ энтропия увеличилась. Сделав так не более, чем $(n-1)$ замен, мы получим распределение, где все p_i равны $1/n$, причем при каждой замене энтропия строго увеличивалась. Значит, максимальное значение энтропии равно

$$\sum_{i=1}^{n} -\frac{1}{n} \log \frac{1}{n} = -\log \frac{1}{n} = \log n$$

и достигается лишь в случае, когда все p_i равны $1/n$. $\square$

29. *Неравенства.* Доказать неравенства:

(a) $H(X, Y|Z) \geqslant H(X|Z)$.

$\triangleright$ По цепному правилу для условной совместной энтропии, имеем

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \geqslant H(X|Z),$$

так как энтропия (в том числе условная) всегда неотрицательна. $\square$

(b) $I(X,Y;Z) \geqslant I(X;Z)$.

$\triangleright$ Применяя цепное правило для взаимной информации, имеем

$$I(X,Y;Z) = I(Y;Z|X) + I(X;Z) \geqslant I(X;Z).$$

Последний переход верен в силу того, что взаимная информация (в том числе условная) всегда неотрицательна. $\square$

(c) $H(X,Y,Z) - H(X,Y) \leqslant H(X,Z) - H(X)$.

$\triangleright$ Выведем это неравенство из предыдущего пункта. Для этого перепишем левую и правую части неравенства **(b)**, используя равенство (2.41) из книги «Elements of Information Theory, 2nd edition»:

$$I(X,Y;Z) = H(X,Y) + H(Z) - H(X,Y,Z),$$

$$I(X;Z) = H(X) + H(Z) - H(X,Z).$$

Тогда всё неравенство **(b)** можно переписать в виде

$$H(X,Y) + H(Z) - H(X,Y,Z) \geqslant H(X) + H(Z) - H(X,Z) \Longleftrightarrow H(X,Y,Z) - H(X,Y) \leqslant H(X,Z) - H(X),$$

что и требовалось доказать. $\square$

(d) $I(X;Z|Y) \geqslant I(Z;Y|X) - I(Z;Y) + I(X;Z)$.

$\triangleright$ Покажем, что это неравенство на самом деле всегда обращается в равенство. Перенесём $-I(Z;Y)$ из правой части в левую и заметим, что

$$I(X;Z|Y) + I(Z;Y) = I(X;Z|Y) + I(Y;Z) = |\text{по цепному правилу для информации}| = I(X,Y;Z),$$

$$I(Z;Y|X) + I(X;Z) = I(Y;Z|X) + I(X;Z) = |\text{по цепному правилу для информации}| = I(Y,X;Z).$$

Осталось заметить, что, очевидно, $I(X,Y;Z) = I(Y,X;Z)$. $\square$

29. *Неравенства.* Доказать неравенства:

(a) $H(X,Y|Z) \geqslant H(X|Z)$.

$\triangleright$ По цепному правилу для условной совместной энтропии, имеем

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z) \geqslant H(X|Z),$$

так как энтропия (в том числе условная) всегда неотрицательна. $\square$

(b) $I(X,Y;Z) \geqslant I(X;Z)$.

$\triangleright$ Применяя цепное правило для взаимной информации, имеем

$$I(X,Y;Z) = I(Y;Z|X) + I(X;Z) \geqslant I(X;Z).$$

Последний переход верен в силу того, что взаимная информация (в том числе условная) всегда неотрицательна. $\square$

(c) $H(X,Y,Z) - H(X,Y) \leqslant H(X,Z) - H(X)$.

$\triangleright$ Выведем это неравенство из предыдущего пункта. Для этого перепишем левую и правую части неравенства **(b)**, используя равенство (2.41) из книги «Elements of Information Theory, 2nd edition»:

$$I(X,Y;Z) = H(X,Y) + H(Z) - H(X,Y,Z),$$

$$I(X;Z) = H(X) + H(Z) - H(X,Z).$$

Тогда всё неравенство **(b)** можно переписать в виде

$$H(X,Y) + H(Z) - H(X,Y,Z) \geqslant H(X) + H(Z) - H(X,Z) \Longleftrightarrow H(X,Y,Z) - H(X,Y) \leqslant H(X,Z) - H(X),$$

что и требовалось доказать. $\square$

(d) $I(X;Z|Y) \geqslant I(Z;Y|X) - I(Z;Y) + I(X;Z)$.

$\triangleright$ Покажем, что это неравенство на самом деле всегда обращается в равенство. Перенесём $-I(Z;Y)$ из правой части в левую и заметим, что

$$I(X;Z|Y) + I(Z;Y) = I(X;Z|Y) + I(Y;Z) = |\text{по цепному правилу для информации}| = I(X,Y;Z),$$

$$I(Z;Y|X) + I(X;Z) = I(Y;Z|X) + I(X;Z) = |\text{по цепному правилу для информации}| = I(Y,X;Z).$$

Осталось заметить, что, очевидно, $I(X,Y;Z) = I(Y,X;Z)$. $\square$

23. *Условная взаимная информация.* Рассмотрим последовательность из n бинарных случайных величин $X_1.\dots.X_n$. Вероятность каждой последовательности с чётным числом единиц равна $1/2^{n-1}$, а вероятность каждой последовательности с нечётным числом единиц равна 0. Найти взаимные информации

$$I(X_1;X_2),\ I(X_2;X_3|X_1),\dots,\ I(X_{n-1};X_n|X_1,\dots,X_{n-2}).$$

$\triangleright$ Покажем, что случайные величины $X_1,\dots X_{n-1}$ независимы в совокупности. Для этого достаточно доказать, что для любого $k<n$ случайная величина X_k независима со случайным вектором $(X_1,\dots,X_{k-1})$. В свою очередь, это эквивалентно тому, что при всех значениях $\bar{a} = (a_1,\dots,a_{k-1}) \in \{0,1\}^{k-1}$ вектора $(X_1,\dots,X_{k-1})$ условные вероятности $\mathsf{P}[X_k = 0 \mid (X_1,\dots,X_{k-1}) = \bar{a}]$ и $\mathsf{P}[X_k = 1 \mid (X_1,\dots,X_{k-1}) = \bar{a}]$ равны $1/2$.

Будем доказывать последнее утверждение индукцией по k. База для $k=1$ будет доказываться так же, как и переход, с учётом того, что $1/2^0 = 1$. Поэтому сразу докажем переход.

Пусть $X_1,\dots X_{k-1}$ независимы в совокупности. Предположим, что X_{k+1} не независима с ними, то есть для некоторого бинарного вектора $\bar{a} = (a_1,\dots,a_{k-1})$ $\mathsf{P}[X_k = a_k \mid (X_1,\dots X_{k-1}) = \bar{a}] = p > 1/2$. Тогда рассмотрим такое значение $a_{k+1} \in \{0,1\}$ случайной величины X_{k+1}, что $\mathsf{P}[X_{k+1} = a_{k+1} \mid X_1 = a_1,\dots,\ X_k = a_k] \geqslant 1/2$. Далее рассмотрим аналогичное значение a_{k+2} для X_{k+2} (то есть такое, что $\mathsf{P}[X_{k+2} = a_{k+2} \mid X_1 = a_1,\dots,\ X_k = a_k] \geqslant 1/2$), потом аналогичное значение a_{k+3} для X_{k+3} и т. д. до X_{n-1}. Очевидно, такие значения a_{k+i} всегда найдутся, так как сумма соответствующих условных вероятностей для $a_{k+i} = 0$ и $a_{k+i} = 1$ равна 1. Тогда имеем

$$\mathsf{P}[(X_1,\dots,X_{n-1}) = (a_1,\dots,a_{n-1})] =$$

$$= \mathsf{P}[(X_1,\dots,X_{k-1}) = (a_1,\dots,a_{k-1})] \cdot \prod_{i=k}^{n-1} \mathsf{P}[X_i = a_i \mid (X_1,\dots,X_{i-1}) = (a_1,\dots,a_{i-1})] =$$

$$= |\text{из независимости } X_1, \ldots, X_{k-1} \text{ в совокупности}| =$$

$$= \left(\frac{1}{2}\right)^{k-1} \cdot\ p\ \cdot \prod_{i=k+1}^{n-1} \mathsf{P}[X_i = a_i \mid (X_1, \ldots, X_{i-1}) = (a_1, \ldots, a_{i-1})] \geqslant p \cdot \left(\frac{1}{2}\right)^{n-2}.$$

Заметим, что, поскольку вероятность вектора из нечётного числа единиц равна 0 (а из нечётного – не равна 0), то условные вероятности $\mathsf{P}[X_n = 0 \mid (X_1, \ldots, X_{n-1}) = (a_1, \ldots, a_{n-1})]$ и $\mathsf{P}[X_n = 1 \mid (X_1, \ldots, X_{n-1}) = (a_1, \ldots, a_{n-1})]$ равны 1 и 0 в каком-то порядке. Значит, выбирая a_n таким образом, чтобы соответствующая вероятность для a_n была равна 1, получаем

$$\mathsf{P}[(X_1, \ldots, X_n) = (a_1, \ldots, a_{n-1}, a_n)] =$$

$$= \mathsf{P}[(X_1, \ldots, X_{n-1}) = (a_1, \ldots, a_{n-1})] \cdot \mathsf{P}[X_n = a_n \mid (X_1, \ldots, X_{n-1}) = (a_1, \ldots, a_{n-1})] =$$

$$= \mathsf{P}[(X_1, \ldots, X_{n-1}) = (a_1, \ldots, a_{n-1})] \cdot 1 \geqslant p \cdot \left(\frac{1}{2}\right)^{n-2} > \left(\frac{1}{2}\right)^{n-1},$$

что противоречит условию задачи. Тогда предположение неверно и переход доказан.

Заметим, что мы на самом деле доказали, что любые $n-1$ случайных величин из $X_1, \ldots, X_n$ независимы в совокупности. Это следует из того, что все эти случайные величины равноправны.

Для решения задачи осталось заметить, что все условные информации вида $I(X_k; X_{k+1} | X_1, \ldots, X_{k-1})$ при $k < n-1$ равны 0, так как случайные величины $X_1, \ldots, X_{k+1}$ при $k < n-1$ независимы в совокупности.
В то же время, имеем

$$I(X_{n-1}; X_n | X_1, \ldots, X_{n-2}) = H(X_{n-1} | X_1, \ldots, X_{n-2}) - H(X_{n-1} | X_1, \ldots, X_{n-2}, X_n).$$

В получившейся разности уменьшаемое равно 1, так как условие никак не влияет на распределение и по сути у нас будет просто энтропия честной монетки (из доказательства независимости величин $X_1, \ldots, X_{n-1}$ в совокупности и самой независимости в совокупности следует, что для любого k $\mathsf{P}[X_k = 0] = \mathsf{P}[X_k = 1] = 1/2$). Вычитаемое же очевидно равно 0, так как X_{n-1} явно определяется через остальные $X_i,\ i \in \{1, \ldots, n-2\} \cup \{n\}$.

Значит, все взаимные информации, кроме $I(X_{n-1}; X_n | X_1, \ldots, X_{n-2})$, равны 0, а последняя равна 1. □

25. *Диаграммы Венна* (определение взаимной информации для трёх случайных величин).
Взаимная информация для трёх случайных величин определяется как

$$I(X; Y; Z) \stackrel{\text{def}}{=} I(X; Y) - I(X; Y | Z).$$

(a) Привести пример случайных величин $X,\ Y,\ Z$, таких что $I(X; Y; Z) < 0$.

▷ Это в точности задача 6b. Подойдёт пример $Z = X + Y$, где X и Y — независимые «честные монетки» $Bernoulli(1/2)$. □

(b) Доказать равенство $I(X; Y; Z) = H(X, Y, Z) - H(X) - H(Y) - H(Z) + I(X; Y) + I(Y; Z) + I(Z; X)$.

▷ По определению взаимной информации для трёх случайных величин имеем

$$I(X; Y; Z) = I(X; Y) - I(X; Y | Z) = I(X; Y) - H(X|Z) + H(X|Y, Z). \tag{1}$$

Используя цепное правило для энтропии, имеем

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

откуда

$$H(X|Y, Z) = H(X, Y, Z) - H(Y|Z) - H(Z). \tag{2}$$

Подставляя правую часть (2) вместо $H(X|Y, Z)$ в (1), получаем

$$I(X; Y; Z) = I(X; Y) - H(X|Z) + H(X, Y, Z) - H(Y|Z) - H(Z) =$$

$$= H(X,Y,Z) - H(Z) + I(X;Y) - H(X|Z) - H(Y|Z). \qquad (3)$$

Наконец, используя известное равенство для взаимной информации ((2.43) в книге)

$$I(A;B) = H(A) - H(A|B) \Longleftrightarrow -H(A|B) = I(A;B) - H(A)$$

и применяя его к $(A,B) = (X,Z)$ и $(A,B) = (Y,Z)$, а после подставляя в (3), получаем

$$I(X;Y;Z) = H(X,Y,Z) - H(Z) + I(X;Y) + \Big(I(X;Z) - H(X)\Big) + \Big(I(Y;Z) - H(Y)\Big) =$$

$$= H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X),$$

что и требовалось. С точки зрения диаграмм Венна это равенство можно интерпретировать как формулу включений-исключений, так как $I(X;Y)$ с точки зрения диаграмм означает «пересечение» $H(X)$ и $H(Y)$, а $I(X;Y;Z)$ по логике должно означать пересечение $H(X)$, $H(Y)$ и $H(Z)$. $\square$

(c) Доказать равенство $I(X;Y;Z) = H(X,Y,Z) - H(X,Y) - H(Y,Z) - H(Z,X) + H(X) + H(Y) + H(Z)$.

$\triangleright$ Выведем его из равенства **(b)**. Для этого вспомним известное равенство ((2.45) в книге)

$$I(A;B) = H(A) + H(B) - H(A,B).$$

Применяя его поочерёдно к $(A,B) = (X,Y)$, (Y,Z) и (Z,X) и подставляя в правую часть равенства **(b)** вместо $I(X;Y)$, $I(Y;Z)$ и $I(Z;X)$ соответственно, получаем

$$I(X;Y;Z) = H(X,Y,Z) - H(X) - H(Y) - H(Z) + \Big(H(X) + H(Y) - H(X,Y)\Big) +$$

$$+ \Big(H(Y) + H(Z) - H(Y,Z)\Big) + \Big(H(Z) + H(X) - H(Z,X)\Big) =$$

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

что и требовалось. $\square$

32. *Фано.* Случайные величины X и Y имеют следующее совместное распределение:

		Y		
		a	b	c
	1	1/6	1/12	1/12
X	2	1/12	1/6	1/12
	3	1/12	1/12	1/6

(a) Найти оценку $\hat{X}(Y)$ для X с минимальной вероятностью ошибки.

$\triangleright$ Очевидно, можно искать оценку $\hat{X}(Y)$ отдельно для каждого значения Y. Для примера, рассмотрим $Y = a$ (в силу симметричности совместного распределения оптимальные вероятности ошибки для $Y = b$ и $Y = c$ будут такими же). Любая оценка $\hat{X}(a)$ имеет вид

$$\hat{X}(a) = \begin{cases} 1 \text{ с вероятностью } p_1 \\ 2 \text{ с вероятностью } p_2 \\ 3 \text{ с вероятностью } p_3 = 1 - p_1 - p_2 \end{cases}$$

Найдём вероятность ошибки такой оценки. Но проще найти вероятность не ошибиться. Имеем

$$\mathsf{P}[X = \hat{X}(a) \mid Y = a] = \mathsf{P}[X = 1, \hat{X}(a) = 1 \mid Y = a] + \mathsf{P}[X = 2, \hat{X}(a) = 2 \mid Y = a] + \mathsf{P}[X = 3, \hat{X}(a) = 3 \mid Y = a] =$$

$$= p_1 \cdot \frac{1}{2} + p_2 \cdot \frac{1}{4} + (1 - p_1 - p_2) \cdot \frac{1}{4} = \frac{1}{4} + \frac{1}{4} p_1.$$

Таким образом, вероятность не ошибиться максимальна при $p_1 = 1$ и равна $1/2$, а значит, любая оценка $\hat{X}(a)$ ошибается с вероятностью не меньше, чем $1/2$. То же самое верно и для оценок $\hat{X}(b)$ и $\hat{X}(c)$, поэтому вероятность ошибки оценки $\hat{X}(Y)$ не меньше, чем $1/2$. Из рассуждения выше ясно, что вероятность ошибки $1/2$ достигается тогда и только тогда, когда $\hat{X}(a) = 1$, $\hat{X}(b) = 2$ и $\hat{X}(c) = 3$ с вероятностью 1. $\square$

(b) Записать неравенство Фано для пункта **(a)** и сравнить результаты.

$\triangleright$ Воспользуемся ослабленным неравенством Фано в форме

$$P_e \geqslant \frac{H(X|Y)-1}{\log|\chi|},$$

где χ — множество значений случайной величины X. Имеем $|\chi|=3$ и

$$H(X|Y) = H(X,Y) - H(Y) = -\frac{3}{6}\cdot\log\frac{1}{6} - \frac{6}{12}\cdot\log\frac{1}{12} + 3\cdot\frac{1}{3}\log\frac{1}{3} =$$
$$= \frac{1}{2}(1+\log 3) + \frac{1}{2}(2+\log 3) - \log 3 = \frac{3}{2} + \log 3 - \log 3 = \frac{3}{2}.$$

Подставляя полученные значения в неравенство Фано, получаем следующую оценку на вероятность ошибки:

$$P_e \geqslant \frac{3/2-1}{\log 3} = \frac{1}{2\log 3},$$

что заметно меньше, чем настоящая минимальная возможная вероятность ошибки. $\square$

37. *Относительная энтропия.* Пусть X, Y, Z — случайные величины с совместной функцией распределения $p(x,y,z)$. Относительная энтропия совместной функции вероятности к произведению маргинальных функций вероятности для этой тройки по определению равна

$$D\Big(P(x,y,z)||p(x)p(y)p(z)\Big) = \mathsf{E}\left[\log\frac{p(x,y,z)}{p(x)p(y)p(z)}\right].$$

Выразить это значение через энтропии. Когда оно равно нулю?

$\triangleright$ Имеем

$$D\Big(P(x,y,z)||p(x)p(y)p(z)\Big) = \mathsf{E}\left[\log\frac{p(x,y,z)}{p(x)p(y)p(z)}\right] = \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y,z)}{p(x)p(y)p(z)} =$$
$$= \sum_{x,y,z} p(x,y,z)\log p(x,y,z) - \sum_{x,y,z} p(x,y,z)\log p(x) - \sum_{x,y,z} p(x,y,z)\log p(y) - \sum_{x,y,z} p(x,y,z)\log p(z) =$$
$$= -H(X,Y,Z) - \sum_x p(x,y,z)\log p(x) - \sum_y p(y)\log p(y) - \sum_z p(z)\log p(z) = -H(X,Y,Z) + H(X) + H(Y) + H(Z).$$

Поймём, когда это выражение равно 0. По цепному правилу для энтропии, имеем

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

откуда

$$D\Big(P(x,y,z)||p(x)p(y)p(z)\Big) = H(X) + H(Y) + H(Z) - H(X) - H(Y|X) - H(Z|Y,X) =$$
$$= \Big(H(Y) - H(Y|X)\Big) + \Big(H(Z) - H(Z|Y,X)\Big).$$

Так как дополнительное условие может только уменьшить энтропию (неравенство (2.95) из книги), то оба слагаемых в последней сумме неотрицательны. Первое из них обращается в 0 тогда и только тогда, когда X и Y независимы, а второе — когда Z и случайный вектор (X,Y) независимы. Из этих двух условий и определения независимости в совокупности легко следует независимость X, Y и Z в совокупности. Обратное следствие очевидно. Таким образом, $D\Big(P(x,y,z)||p(x)p(y)p(z)\Big) = 0$ тогда и только тогда, когда X, Y и Z независимы в совокупности. $\square$

43. *Взаимная информация орлов и решек.*

(a) Рассмотрим подбрасывание честной монетки. Найти взаимную информацию верхней и нижней сторон монетки.

$\triangleright$ Всего есть два исхода с ненулевой вероятностью: (орёл сверху, решка снизу) и (решка сверху, орёл снизу). Так как вероятность каждого из них равна $1/2$, то взаимная информация равна

$$I = 2\cdot\frac{1}{2}\cdot\log\left(\frac{1/2}{(1/2)^2}\right) = \log 2 = 1.$$

$\square$

(b) Бросается честный шестигранный кубик. Найти взаимную информацию верхней и передней граней.

$\triangleright$ Здесь такая же логика, как для монетки. Нужно лишь заметить, что любая возможная пара из верхней и передней грани задаётся их общим ребром и одной из двух возможных ориентаций (какая грань верхняя, а какая — передняя). Все такие пары равновероятны, а их число равно удвоенному числу рёбер, то есть $12 \cdot 2 = 24$. При этом вероятность того, что фиксированная грань оказалась верхней (аналогично, передней), равна $1/6$, поэтому взаимная информация равна

$$I = 24 \cdot \frac{1}{24} \cdot \log\left(\frac{1/24}{(1/6)^2}\right) = \log\frac{36}{24} = \log\frac{3}{2} = \log 3 - 1.$$

$\square$

44. *Чистый рандом.* Пусть X — трёхсторонняя монетка с распределением

$$X = \begin{cases} A, \ p_A \\ B, \ p_B \\ C, \ p_C, \end{cases}$$

где p_A, p_B и p_C неизвестны.

(a) Используя два независимых подбрасывания, сгенерировать распределение $Bernoulli(1/2)$.

$\triangleright$ Заметим, что для каждого исхода из двух подбрасываний, когда выпадают разные результаты, перестановкой результатов получается исход с такой же вероятностью. Например, для исхода AB получится исход BA и эти исходы оба имеют вероятность $p_A p_B$. Тогда можно взять все исходы, когда выпадают разные результаты, разбить их на две равновероятные группы и считать, что при выпадении исхода из первой группы «выпадает орёл», а при выпадении исхода из второй группы — решка. Например, подойдёт разбиение на $\{AB,\ BC,\ CA\}$ и $\{BA,\ CB,\ AC\}$. В то же время, с исходами AA, BB и CC мы ничего не можем сделать, поэтому их мы будем просто игнорировать — считать, что если выпал один из них, то генерация не удалась. $\square$

(b) Какое максимальное ожидаемое число честных бит может быть так сгенерировано?

$\triangleright$ Честный бит будет сгенерирован если и только если два броска покажут разные результаты. Вероятность этого равна $1 - (p_A^2 + p_B^2 + p_C^2)$. Это и есть ожидаемое число сгенерированных честных бит. Так как $p_A + p_B + p_C = 1$, то по неравенству о средних имеем

$$\sqrt{\frac{p_A^2 + p_B^2 + p_C^2}{3}} \geqslant \frac{p_A + p_B + p_C}{3} = \frac{1}{3},$$

откуда

$$p_A^2 + p_B^2 + p_C^2 \geqslant 3 \cdot \left(\frac{1}{3}\right)^2 = \frac{1}{3} \Longleftrightarrow 1 - (p_A^2 + p_B^2 + p_C^2) \leqslant 1 - \frac{1}{3} = \frac{2}{3}$$

и равенство достигается тогда и только тогда, когда $p_A = p_B = p_C = 1/3$. $\square$

47. *Энтропия почти отсортированной колоды.* Дана отсортированная в возрастающем порядке колода из n карт с номерами от 1 до n. Из колоды равновероятно выбирается карта, вынимается, а потом вставляется в случайное место. Найти энтропию получающейся колоды.

$\triangleright$ Зафиксируем карту, которую вынули. Всего есть n мест для вставки, поэтому может получиться n разных колод, каждая с вероятностью $1/n$. В то же время, некоторые колоды, получающиеся в итоге при вынимании разных карт, могут совпасть. Поймём, когда это происходит. Ясно, что каждая колода — это просто перестановка на n элементах. Для произвольной нетождественной перестановки f определим *первое несоответствие* как

$$m(f) = f(\min\{i : f(i) \neq i\}).$$

Иными словами, это образ минимального элемента, такого что этот образ не равен самому этому элементу.

Легко понять, что для колоды, полученной после вынимания карты с номером i, первое несоответствие равно i, если карту вставили на позицию левее исходной, и $i+1$, если вставили на позицию правее исходной. Отсюда очевидно следует, что одинаковые колоды (не считая тождественной перестановки) могли получиться только при вынимании соседних карт (i и $i+1$), причём карта i должна была переместиться вправо, а карта $i+1$ — влево. Но заметим, что при перемещении карты i вправо карта $i-1$ по-прежнему останется левее, чем $i+1$ (если карты $i-1$ нет, работает такой же аргумент для перемещения карты $i+1$ и карт i и $i+2$). Поэтому одинаковые перестановки могли получиться, только если карта i поменялась местами с картой $i+1$, то есть транспозиция $(i,\ i+1)$ — единственная общая перестановка.

Теперь можно посчитать вероятности получения всех перестановок. Для тождественной перестановки она равна $n \cdot 1/n^2 = 1/n$. Для транспозиций вида $(i,\ i+1)$ вероятности равны $1/n^2 + 1/n^2 = 2/n^2$, а для всех остальных перестановок они равны $1/n^2$. Для подсчёта энтропии осталось найти число перестановок каждого типа. Тождественная перестановка единственна, а транспозиций вида $(i,\ i+1)$ всего $n-1$. Найдём число всех остальных возможных перестановок. При вытаскивании каждой карты получается n перестановок, из них одна тождественная. Кроме того, для карт с номерами 1 и n среди них есть одна транспозиция, а для остальных карт — две транспозиции. Итого получаем $2(n-2) + (n-2)(n-3) = n^2 - 3n + 2$ перестановок.

Наконец, можно найти энтропию:

$$H = -\frac{1}{n}\log\frac{1}{n} - (n-1)\cdot\frac{2}{n^2}\log\frac{2}{n^2} - (n^2-3n+2)\cdot\frac{1}{n^2}\log\frac{1}{n^2} =$$

$$= \frac{1}{n}\log n + \frac{2(n-1)(2\log n - 1)}{n^2} + \frac{(n^2-3n+2)\cdot 2\log n}{n^2} = \left(2 - \frac{1}{n}\right)\log n - \frac{2n-2}{n^2}.$$

□

48. *Длина последовательности.* Сколько информации длина последовательности даёт о её элементах? Дан случайный процесс $\{X_n\}_{n=1}^{\infty}$ с натуральным временем, для любого n $X_n \sim Bernoulli(1/2)$. Как только появляется первая единица, процесс останавливается. Пусть N — номер шага, на котором процесс завершился, а X^N — случайный вектор, состоящий из значений $X_1, \ldots, X_N$.

(a) Найти $I(N; X^N)$.

▷ Очевидно, каждому значению N соответствует ровно одно значение X^N, а именно, вектор из $(N-1)$ нулей и одной единицы. Вероятность того, что процесс продлится N шагов, равна $1/2^N$ поэтому взаимная информация равна

$$I(N; X^N) = \sum_{N=1}^{\infty} \frac{1}{2^N}\cdot\log\frac{1/2^N}{(1/2^N)^2} = \sum_{N=1}^{\infty}\frac{1}{2^N}\log(2^N) = \sum_{N=1}^{\infty}\frac{N}{2^N}.$$

Вспоминая, что $\displaystyle\sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}$, получаем $I(N, X^N) = \dfrac{1/2}{(1-1/2)^2} = 2$. □

(b) Найти $H(X^N|N)$.

▷ Так как X^N есть функция от N, то такая условная энтропия равна нулю. □

(c) Найти $H(X^N)$.

▷ Так как $I(X;Y) = H(X) - H(X|Y)$ (равенство (2.43) из книги), то имеем

$$H(X^N) = I(X^N; N) + H(X^N|N) = I(N; X^N) + 0 = 2.$$

□

Рассмотрим теперь другое время остановки. Будем останавливаться в момент $N = 6$ с вероятностью 1/3 и в момент $N = 12$ с вероятностью 2/3 независимо от значений членов последовательности $\{X_i\}_{i=1}^{12}$.

(d) Найти $I(N;X^N)$.

$\triangleright$ Есть два возможных значения N, это 6 и 12. Для $N=6$ есть 2^6 равновероятных возможных значений X^N, для каждого такого значения Y $p(N=6,\ X^6=Y)=1/3\cdot 1/2^6$.
С другой стороны, $p(N=6)=1/3,\ p(X^N=Y)=1/3\cdot 1/2^6$. Рассуждая аналогично для $N=12$, получаем

$$I(N;X^N)=2^6\cdot\frac{1}{3}\cdot\frac{1}{2^6}\log\frac{1/3\cdot 1/2^6}{1/3\cdot 1/3\cdot 1/2^6}+2^{12}\cdot\frac{2}{3}\cdot\frac{1}{2^{12}}\log\frac{2/3\cdot 1/2^{12}}{2/3\cdot 2/3\cdot 1/2^{12}}=$$

$$=\frac{1}{3}\log 3+\frac{2}{3}(\log 3-1)=\log 3-\frac{2}{3}.$$

$\square$

(e) Найти $H(X^N|N)$.

$\triangleright$ Для любого бинарного вектора Y длины 6 $\mathsf{P}[X^N=Y|N=6]=1/2^6$, аналогично для любого вектора длины 12 соответствующая условная вероятность равна $1/2^{12}$ Поэтому условная энтропия равна

$$H(X^N|N)=-2^6\cdot\frac{1}{3}\cdot\frac{1}{2^6}\log\frac{1}{2^6}-2^{12}\cdot\frac{2}{3}\cdot\frac{1}{2^{12}}\log\frac{1}{2^{12}}=2+8=10.$$

$\square$

(f) Найти $H(X^N)$.

$\triangleright$ Аналогично пункту **(c)**, имеем

$$H(X^N)=I(X^N;N)+H(X^N|N)=\log 3-\frac{2}{3}+10=\log 3+\frac{28}{3}.$$

$\square$