Тема:
Дискретные ансамбли и источники

Дискретные ансамбли

- *Конечное множество X = {x_1 ,..., x_M} вместе с заданным на нем распределением вероятностей p(x) (т.е.* каждому элементу $x_i \in X$ сопоставлено число p (x_i), причем p(x_i)>0;   i= 1, 2, ..., M; $\sum_{i=1}^{M} p(x_i) = 1$*) называется дискретным вероятностным ансамблем или коротко — дискретным ансамблем (сообщений) и обозначается символом {X, p (x)}.*

Пусть XY есть произведение двух конечных множеств X и Y и на множестве XY задано совместное распределение вероятностей $p(x, y)$, которое каждой паре (x_i, y_j), $x_i \in X$, $y_j \in Y$, сопоставляет вероятность $p(x_i, y_j)$.

Если распределение вероятностей на произведении двух множеств X и Y удовлетворяет условию $p(x_i, y_j) = p_1(x_i)p_2(y_j)$ для всех $x_i \in X$, $y_j \in Y$, то ансамбли X и Y называются *статистически НЕзависимыми*.

В противном случае говорят, что эти ансамбли *статистически зависимы.*

- Пусть задан ансамбль *{XY, p (x, y)}*, предположим, что x_i - такой элемент множества X, для которого $p_i(x_i) \neq 0$. Число

$$p\left(y_j \middle| x_i\right) \equiv \frac{p(x_i, y_j)}{\sum_{y_{j \in Y}} p(x_i, y_j)}$$

называется условной вероятностью сообщения y_j при условии, что сообщение x_i известно (иногда это число называют условной вероятностью сообщения y_j относительно сообщения x_i).

- Задан ансамбль *{XY, p (x, y)}*,

пусть *A* — произвольное подмножество элементов из X с вероятностью $Pr_1(A)$, где распределение $p_1(x) \equiv \sum_{y_{j\in Y}} p\left(x_i, y_j\right)$.

Число $p\left(y_j \middle| A\right) \equiv \frac{1}{\mathrm{Pr_1(A)}} \cdot \sum_{x_i \in A} p\left(x_i, y_j\right)$

называется *условной вероятностью сообщения* y_j при условии, что сообщение x_i принадлежит множеству *A* (или условной вероятностью относительно множества *A*).

- Рассмотрим произведение n множеств $X_1 \ldots X_n$ и распределение вероятностей $p\ (x^{(1)}, \ldots, x^{(n)})$, $x^{(i)} \in X_i$, $i = 1, \ldots, n$, заданное на этом произведении.

Другими словами, рассмотрим вероятностный ансамбль $\{X_1 \ldots X_n, p(x^{(1)}, \ldots, x^{(n)})\}$. Пусть

$$
\begin{aligned}
p_1 &\equiv \sum_{X_2} \sum_{X_3} \ldots \sum_{X_n} p(x^{(1)}, \ldots, x^{(n)}) \\
p_2 &\equiv \sum_{X_1} \sum_{X_3} \ldots \sum_{X_n} p(x^{(1)}, \ldots, x^{(n)}) \qquad (1) \\
p_3 &\equiv \sum_{X_1} \sum_{X_2} \ldots \sum_{X_n} p(x^{(1)}, \ldots, x^{(n)})
\end{aligned}
$$

Соотношения (1) задают безусловные распределения вероятностей на множествах $X_1, X_2, \ldots, X_n$ соответственно. Если для любых $x^{(1)} \in X_1 ,\ldots, x^{(n)} \in X_n$ имеет место равенство

$$p(x^{(1)},\ldots,x^{(n)}) = p_1(x^{(1)}) \ldots p_n(x^{(n)})$$

то ансамбли $X_1 ,\ldots, X_n$ называют *статистически НЕзависимыми*.

Дискретный источник

Под *дискретным источником* понимают некоторое устройство, которое в каждую единицу времени (например, каждую секунду) выбирает одно из сообщений дискретного множества X.

Как правило, это множество одно и то же для каждого момента времени, хотя в некоторых случаях для каждого момента времени может быть свое множество сообщений.

С точки зрения теории информации источник считается заданным полностью, если имеется некоторая вероятностная схема (модель), позволяющая вычислить вероятность любого отрезка сообщений.

Например, НЕдостаточно сказать, что в качестве источника сообщений рассматривается телеграфный аппарат или передающее телевизионное устройство. НЕдостаточно также сказать, что известны вероятности букв, появляющихся на выходе телеграфного аппарата, или вероятности импульсов различной амплитуды на выходе телевизионного устройства. Для полного задания источника необходимо дать вероятностное описание процесса появлений сообщений на выходе источника. В примере с телеграфным аппаратом нужно дать такое описание, при котором можно вычислить вероятность любого буквенного сочетания, слова, предложения и т. д. в любой момент времени.

Таким образом, если источник задан, то для любых n и i, и любой последовательности сообщений ($x^{(i+1)}$, ..., $x^{(i+n)}$) определена вероятность p ($x^{(i+1)}$, ..., $x^{(i+n)}$) этой последовательности.

Верно и обратное, если для любых n и i определены вероятности всех последовательностей сообщений длины n, начинающихся с позиции (i+1), то говорят, что задан источник сообщений.

Отсюда следует, что все источники, которые могут иметь совершенно различную физическую природу, задаваемые одним и тем же набором вероятностей последовательностей сообщений, с позиции теории информации отождествляются.

- Пусть U_x — дискретный источник, выбирающий сообщения из множества X.

Будем говорить, что источник U_x задан, если для любых $n = 1, 2, \ldots$, любых $i=0, \pm1, \pm2, \ldots$

задано семейство распределений вероятностей $\{p(x^{(i+1)}, \ldots, x^{(i+n)})\}$, $x^{(j)} \in X$, $j=i+1, \ldots, i+n$, удовлетворяющих условию согласованности, состоящему в том, что распределение вероятностей

$p(x^{(i_1)}, \ldots, x^{(i_m)})$ для любого набора позиций $i_1 \ldots, i_m$ определено однозначным образом.

Стационарный дискретный источник

- Дискретный источник *Ux* называется стационарным, если сообщения на его выходе образуют стационарный случайный процесс.

В общем случае вероятность некоторого отрезка сообщений зависит как от самого отрезка, так и от его расположения на оси времени. Имеется, однако, важный класс источников, обладающих однородностью во времени или стационарностью.

Свойство стационарности состоит в том, что для любого целого числа j вероятности двух одинаковых последовательностей, одна из которых занимает временные позиции $i_1, \ldots, i_n$, а другая — временные позиции $i_1 + j, \ldots, i_n + j$, равны. Другими словами,

$$p\left(x^{(i_1+j)}, \ldots, x^{(i_n+j)}; i_1 + j, \ldots, i_n + j\right) = p(x^{(i_1)}, \ldots, x^{(i_n)}; i_1, \ldots, i_n)$$

(значение этой функции есть вероятность того, что в фиксированные моменты времени $i_1, \ldots, i_n$ источник породит сообщения (или процесс примет значения) $x^{(i_1)}, \ldots, x^{(i_n)}$ соответственно)

В случае стационарных источников все последовательности НЕ зависит от сдвига по оси времени и имеют одинаковые вероятности.

Дискретный источник без памяти

- *Дискретный источник называется источником без памяти*, *если для любых n = 1, 2 ..., любых i = 0, ±1, ±2, ... и любых последовательностей* $\left(x^{(i+1)}, \ldots, x^{(i+n)}\right)$, $x^{(j)} \in X$, *имеет место равенство*

$$p\left(x^{(i+1)}, \ldots, x^{(i+n)}\right) = \prod_{j=1}^{n} p_{i+j}\left(x^{(i+j)}\right)$$

Дискретный источник без памяти обладает тем свойством, что его выходной сигнал в любой момент времени не зависит от своей предыстории.

(В частности, это означает, что совместная вероятность n сигналов является просто произведением вероятностей соответствующих сигналов)

Дискретный источник с памятью

Дискретный источник называется источником с памятью, если его выходной сигнал в какой-либо момент времени зависит от своих значений, имевших место в несколько предшествующих моментов времени.

Если число этих моментов конечно, источник имеет конечный порядок, о противном случае - бесконечный.