Теория вероятности

Основные сведения из теории вероятности

Введение

Термин **информация** в курсе будет пониматься в узком научном смысле.

Теория информации – специальная математическая дисциплина. Её содержанием является абстрактно формулируемые теоремы и модели. ТИ имеет обширное применение к теории передачи сообщений, записывающих устройств, матлингвистике, компьютерной технике.

В самом общем виде теория информации понимается как теория передачи сигналов по линиям связи. Наиболее важное понятие ТИ –- сама информация. В нашей жизни большую роль играет информация и связанные с ней операции: передача, получение, обработка, хранение.

Информация имеет две стороны: количественную и качественную. Иногда важно получение общего количества информации (количественная сторона), иногда важно конкретное содержание самой ИИ. Отметим, что переработка ИИ является технически сложной процедурой, которая усложняет разработку общей теории информации.

Важнейшим этапом в открытии основных закономерностей ТИ были работы американского инженера-связиста, математика Клода Шеннона (1947-49гг).

Для вычисления количества информации была предложена т.н. **логарифмическая мера**. Понятие **количества информации** тесно связано с понятием энтропии как меры степени неопределённости. Приобретение информации сопровождается уменьшением неопределённости, следовательно, количество информации можно измерять количеством ''исчезнувшей неопределённости'' (энтропии).

Теория информации является математической теорией, использующей понятия и методы теории вероятности.

Вероятность. Случайные события и величины

Пусть производится серия из N опытов, причём некоторое событие A происходит в $N_a < N+1$. Тогда $h_n(A) = N_a/N$ называется частотой появления события A в серии из N опытов. Известный факт: с ростом $N \quad h_n(A) \to p$ (постоянная p - вероятность появления случайного события A).

Наука, изучающая свойства вероятности и применение этого понятия называется **теория вероятности**.

Событие, которое при выполнении некоторого комплекса условий обязательно выполняется называется **достоверным событием**.

Событие, которое при выполнении некоторого комплекса условий не выполняется называется **невозможным событием**.

Пример: Выпадение определённого числа очков на грани игральной кости - достоверное событие.

Выпадение семи очков на грани игральной кости - невозможное событие.

Случайное событие – событие, которое может произойти, а может и не произойти.

Задача: В урне 10 шаров : 5 белых, 3 чёрных и 2 красных. Найти вероятность выпадения шара определённого цвета (шары одинаковы).

Решение: Выписать случайные события:

A –- {вынутый шар белый} $P(A) = 5/10 = 1/2$;
B –- {вынутый шар чёрный} $P(B) = 3/10$;
C –- {вынутый шар красный} $P(C) = 2/10 = 1/5$.

Задача: Какова вероятность, что при бросании кости выпадет число очков, кратное 3?

Решение:

Кратны $3\ \{3, 6\}$.
N исходов $= 6$.
$P(A) = 2/6 = 1/3$.

Общий принцип решения задач сводится к понятию равновероятности или равновозможности. (например, все грани кости одинаковы, и вероятность выпадения той или иной грани равна $1/6$).

Классическое определение вероятности

Пусть из N возможных исходов опыта случайное событие A появляется M раз. Тогда вероятность случайного события A в модели с равновероятными исходами вычисляется по формуле $P(A) = M/N$

Каждому опыту отвечает своя таблица вероятности. К примеру, в задаче с урнами и шарами таблица вероятности имеет вид:

События	A	B	C
Вероятности	$P(A) = 1/2$	$P(B) = 3/10$	$P(C) = 1/5$

Можно сказать, что в опыте с бросанием кости число очков, выпадающих на грани является случайной величиной, которая может принимать одно из возможных 6 числовых значений в зависимости от случая.

Итак, случайная величина –- числовая функция, принимающая то или иное числовое значение в зависимости от случая.

Например, количество рождений в городе за год - случайная величина.

Свойства вероятности. Сложение и умножение случайных событий.

Несовместные и независимые случайные события.

Из определения вероятности вытекают основные свойства вероятности случайного события A:

- $0 \leqslant P(A) \leqslant 1$.

  $P(A) = 1$ – достоверное событие;

  $P(A) = 0$ – невозможное событие.

- Пусть опыт приводит к двум взаимоисключающим событиям или исходам A или B. В этом случае В называют противоположным A событием $(B = \bar{A})$

  Пусть $P(A) = \frac{m}{n}$;

  Тогда $P(\bar{A}) = \frac{(n-m)}{n} = 1 - \frac{m}{n} = 1 - P(A) \Rightarrow P(A) = 1 - P(\bar{A})$.

  Пусть случайное событие $A_1 \subset A$ влечёт появление события $A \Rightarrow P(A1) < P(A)$

- **Правило сложения вероятностей для двух событий:**

  - Пусть A и B – несовместны.
    Тогда $A \cap B = \varnothing$;
    $P(A) = \frac{m_1}{n}$;
    $P(B) = \frac{m_2}{n}$;
    $P(A + B) = \frac{(m_1 + m_2)}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$.
    Таким образом, $P(A + B) = P(A) + P(B)$.

    В примере с урной вероятность извлечь чёрный или белый шар равна $P(A + B) = P(A) + P(B) = \frac{1}{2} + \frac{3}{10} = \frac{4}{5}$;

    Замечание: Пусть некоторый опыт проиводит к появлению K различных (взаимоисключающих) исходов:

Исходы	A1	A2	…	An
Вероятности	P1	P2	…	Pn

    Заметим, что бывают случаи, когда

    $$\sum_{i=1}^{k} P(A_i) = P(A_1 + A_2 + \, \dots \, + A_n) = 1$$

    В этом случае говорят, что события $A_1, A_2, \, \dots \, , A_n$ составляют **полную группу** случайных событий, то есть $A_1, A_2, \, \dots \, , A_n$ попарно несовместны.

    $A_1, A_2, \, \dots \, , A_n \, : \, A_i \cap A_j = \varnothing \, \forall \, i, j \, : \, i \neq j$; если $A_1 + A_2 + \, \dots \, + A_n$ - достоверное событие.

  - Пусть A и B совместны. $P(A + B) = P(A) + P(B) - P(AB)$, где $P(AB)$ – вероятность одновременного происхождения двух случайных событий A и B.

Теорема сложения вероятности для совместных случайных событий.

(диаграмма Венна: $A = m_1; B = m_2, A \cap B = l$).

$P(AB) = \frac{l}{n}$

$P(A+B) = \frac{(m_1+m_2-l)}{n} =$ (в m_1 и m_2 входит l) $= \frac{m_1}{n} + \frac{m_2}{n} - \frac{l}{n}$

События A и B называются **независимыми**, если результат выполнения события A не связан с результатом события B. (извлечение двух чёрных шаров из разных урн – независимые события)

- **Теорема умножения вероятности для двух независимых событий:**

  Если A и B независимы, то $P(AB) = P(A) * P(B)$

Пример 1: Какова вероятность при двух бросках монеты оба раза выпадет орёл?

$P(AB) = ?$
A{орёл} $P(A) = \frac{1}{2}$;
B{решка} $P(B) = \frac{1}{2}$;
$P(AB) = \frac{1}{4}$.

Пример 2: В колоде 52 карты, 4 масти, 2 козыря. Какова вероятность того, что взятая наугад карта 2 является тузом или козырем?

$A\{$туз$\}$ $P(A) = 1/13$;
$B\{$козырь$\}$ $P(B) = 1/4$;
$P(AB) = 1/52$;
A и B совместны, независимы.
$P(A+B) = P(A) + P(B) - P(AB) = 1/13 + 1/4 - 1/52 = 4/13$.

Условная вероятность.

Рассмотрим пример: В урне M чёрных шаров и $N - M$ белых. Случайное событие
A {извлечение чёрного шара} и
B {извлечение чёрного шара из той-же урны после того, как из неё уже вынут один шар}

$$P(B|A) = \frac{(m-1)}{(n-1)}$$

Поскольку, если событие A имело место, то в урне осталось $M - 1$ чёрных шаров.

$$P(B|\bar{A}) = \frac{m}{(n-1)}$$

$\bar{A}$: $\{$первый вынутый шар - белый$\}$

Вероятность события B здесь разная. Вероятность, которую имеет событие B в том, случае, когда известно, что событие A имело место называется **условной вероятностью** события B при условии выполнения события A.

$$P(B/A) = P(B|A) = P_A(B)$$

Условные вероятности можно вычислять аналогично вычислению безусловных вероятностей.

В случае если A и B независимы, $P(A|B) = P(A) * P(B)$.

В случае зависимости $P(AB) = P(A) * P(B|A) = P(B) * P(A|B)$.

В обоих случаях мы имеем правило умножения вероятностей. В одном случае для независимых событий, в другом для зависимых. Последнее соотношение часто кладут в определение условной вероятности.

$$P(B|A) = \frac{P(AB)}{P(A)} \qquad P(A|B) = \frac{P(AB)}{P(B)}$$

Из предыдущей формулы можем составить пропорцию:

$$\frac{P(B|A)}{P(B)} = \frac{P(A|B)}{P(A)}$$

Из определения условной вероятности вытекают ее основные свойства:

1. $0 \leqslant P(B|A) \leqslant 1$, причём $P(B|A) = 1$ когда $A \subset B$; B - достоверное случайное событие.

   $P(B|A) = 0 \iff A, B$ несовместны, или известно, что B – невозможное событие.

2. Пусть $B_1 \subset B$ (появление B_1 вызывает событие B). $\quad P(B1|A) \leqslant P(B|A)$.

3. Если B и C несовместны $P(B+C|A) = P(B|A)+P(C|A)$ (теорема сложения вероятностей для несовместных событий)

4. $P(\bar{B}|A) = 1 - P(B|A)$

Замечание: Пусть имеется K (и только K) попарно несовместных исходов некоторого опыта $A_1, A_2, \dots, A_k$, называемых гипотезами. Пусть некоторое случайное событие B может произойти при выполнении одной из гипотез. Тогда очевидно, что $B = A_1B + A_2B + \dots + A_kB$ (все события A_iB несовместны, поэтому можно воспользоваться теоремой сложения вероятностей)

$$P(B) = P\left(\sum_{i=1}^{k} A_iB\right) = \sum_{i=1}^{k} P(A_iB) = \sum_{i=1}^{k} \big(P(A_i) * P(B|A_i)\big)$$

Формула носит название формулы полной вероятности

$$P(B) = \sum_{i=1}^{k} P(A_i) * P(B|A_i)$$

Задача: Имеется 5 урн : в двух по одному белому и пять чёрных шаров; в одной – 2 белых, 5 чёрных; в двух – 3 белых, 5 чёрных шаров. Наудачу выбирается одна урна. Из неё извлекается один шар. Какова вероятность того, что шар белый?

Решение: Выберем в качестве гипотез 3 способа

A_1 : {Выбрана урна с 1 б.ш}	$P(A_1) = 2/5$	$P(B|A_1) = 1/6$
A_2 : {Выбрана урна с 2 б.ш}	$P(A_2) = 1/5$	$P(B|A_2) = 2/7$
A_3 : {Выбрана урна с 3 б.ш}	$P(A_3) = 2/5$	$P(B|A_3) = 3/8$

B = извлечён белый шар

$P(B) = \frac{1}{6} * \frac{2}{5} + \frac{2}{7} * \frac{1}{5} + \frac{3}{8} * \frac{2}{5} = \frac{23}{84}$

Математическое ожидание случайной величины. Основные свойства математического ожидания

Введение.

Важнейшей числовой характеристикой ξ является её математическое ожидание или среднее значение, вычисляемое по правилу $M\xi = \sum\limits_{i=1}^{n} x_i p_i)$, где x_i – принимаемые ξ значения, p_i – вероятности их выпадения.

С помощью математического ожидания мы можем сравнивать между собой две случайные величины (например, из двух стрелков лучший тот, кто выбивает в среднем наибольшее число очков), однако встречаются задачи, в которых знание одного лишь $M\xi$ недостаточно.

Пример: Пушка ведёт прицельный огонь по мишени, удалённой от пушки на расстояние a. Обозначим дальность полёта снаряда через ξ километров; $M\xi = a$

Отклонение $M\xi$ от a свидетельствует о наличии систематической ошибки (производственный дефект, неправильный угол наклона). Ликвидация систематической ошибки достигается изменением угла наклона орудия.

Вместе с тем, отсутствие систематической ошибки ещё не гарантирует высокую точность стрельбы. Чтобы оценить точность надо знать, насколько близко ложатся снаряды к цели.

Как определить точность стрельбы и сравнить между собой качество стрельбы двух орудий?

Отклонение снаряда от цели - $\xi - a$

$M(\xi - a) = M\xi - a = a - a = 0$

В среднем, положительные и отрицательные значения $M\xi$ сокращаются. Поэтому принято характеризовать разброс значений случайной величины математическим ожиданием квадрата её отклонения от своего математического ожидания. Полученное таким образом число называется дисперсией случайной величины ξ.

$D\xi = M(\xi - a)^2 = M[\xi - M\xi]^2$

Ясно, что в случае орудий, ведущих стрельбу, лучшим следует считать орудие, у которого $D\xi$ будет наименьшей.

Пусть ξ характеризуется таблицей вероятностей

$x_i :$	x_1	x_2	$\dots$	x_n
$p_i :$	p_1	p_2	$\dots$	p_n

$$M\xi = \sum_{i=1}^{n} x_i p_i; \qquad D\xi = M(\xi - M\xi)^2 = \sum_{i=1}^{n} (x_i - M\xi)^2 * p_i$$

Определение математического ожидания

Пусть есть некоторое пространство, в котором имеется некоторое $\xi = \xi(\omega_i)$.

ω_i – неразделимое событие (пример: исходы броска монеты); $\omega_i : (i = 1, \bar{n})$.

Совокупность ω_i образует пространство элементарных событий $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

Математическим ожиданием случайной величины ξ называется число, обозначаемое $M\xi$ и равное

$$M\xi = \sum_{\omega_i \in \Omega} \{(\omega_i) * P(\omega_i)\} = \sum_{i=1}^{n} \xi(\omega_i) * p(\omega_i),$$ где p_i - элементарные вероятности.

Из определения математического ожидания вытекают следующие свойства:

1. Аддитивность. $M(\xi + \eta) = M\xi + M\eta$.

   Следствие $M\left(\sum\limits_{k=1}^{n} \xi_k\right) = \sum\limits_{k=1}^{n} (M\xi_k)$.

2. $\forall C = const : M(C * \xi) = C * M\xi$. Совокупность свойств 1 и 2 даёт нам свойство линейности математического ожидания:

$$M(C_1\xi_1 + C_2\xi_2 + \dots + C_n\xi_n) = C_1 M(\xi_1) + C_2 M(\xi_2) + \dots + C_n M(\xi_n)$$

3. Математическое ожидание индикатора случайного события равно вероятности этого случайного события.

   Индикатор $[\chi]$: $M\chi_A(\omega) = P(A)$ - случайная величина, принимающая 2 значения: $\chi_A(\omega) = \{1, \omega \in A \,|\, 0, \omega \notin A\}$

$$\sum_{\omega \in A} P(\omega) = P(A)$$

$$M\chi_A(\omega) = \sum_{\omega \in A} 1 * p(\omega) + \sum_{\omega \notin A} 0 * p(\omega) = \sum_{\omega \in A} 1 * p(\omega) = P(A).$$

4. Свойство монотонности $\xi \geqslant \eta \Rightarrow M\xi \geqslant M\eta$.

Докажем вначале, что имеет место следующее свойство $\xi \geqslant 0 \Rightarrow M\xi \geqslant 0$ (при разложении по определению неотрицательны).

$$M\xi = \sum_{\omega} \xi(\omega)p(\omega) \geqslant 0.$$

Применим полученное свойство:

$$\xi - \eta \geqslant 0 \Rightarrow M(\xi - \eta) \geqslant 0 \Rightarrow M\xi - M\eta \geqslant 0 \Rightarrow M\xi \geqslant M\eta.$$

Формулы вычисления математического ожидания

Пусть $x_1, x_2, \ldots, x_n$ –- значения случайной величины ξ, принимаемые с вероятностями $p_1, \ldots, p_i$. Тогда имеет место следующая формула для вычисления математического ожидания :

$$M\xi = \sum_{i=1}^{n} x_i * P(\xi = x_i)$$

Чтобы доказать формулу будем исходить из того, что ξ может быть представлена в виде линейной комбинации индикаторов случайных событий

$$\xi = \sum_{i=1}^{n} x_i * \chi_{A_i}(\omega)$$
$$A_i\{\omega_i : \xi = x_i\}$$

Левые и правые части соотношения совпадают. Применим к написанному равенству операцию математического ожидания:

$$M\left(\sum_{i=1}^{n} x_i\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} M\left(x_i\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} x_i M\left(\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} x_i P(\xi = x_i)$$

Рассуждая аналогично, нетрудно получить формулы вычисления математического ожидания от величин, представляющих собой функции случайных величин.

Пусть заданы $f(\xi), g(\xi, \eta)$.

В этом случае

$$M(f(\xi)) = \sum_{i=1}^{n} (f(x_i) * P(\xi = x_i))$$

$$M(g(\xi, \eta)) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} g(x_i, y_j) * P\left(\xi = x_i,\, \eta = y_i\right) \right)$$

Здесь $P(\xi, \eta)$ – совместная вероятность.

5 Мультипликативное свойство математического ожидания

Пусть ξ, η - независимые случайные величины, то $M(\xi, \eta) = M\xi * M\eta$

Доказательство:

$$M(\xi, \eta) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_i * y_j * P\left(\xi = x_i, \eta = y_j\right) \right)$$

Если ξ, η независимы, то для них применима теорема умножения вероятности.

$$P(\xi = x_i, \eta = y_j) = (\xi, \eta \text{ независимы}) = P(\xi = x_i) * P(\eta = y_j)$$

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_i * y_j * P(\xi = x_i) * P(\eta = y_j) \right) =$$

$$= \sum_{i=1}^{n} x_i * P(\xi = x_i) \, * \, \sum_{j=1}^{m} y_i * P(\eta = y_j) = M\xi * M\eta$$

Замечание: Все написанные формулы имеют место, если вероятностное пространство конечно, т.е. число элементарных событий конечно $\omega_i = (1, \bar{n})$.

В случае, если вероятностное пространство счётно, количество элементарных сообщений бесконечно, тогда для случайной величины $\xi(\omega), \omega \in$ (счетное вероятностное пространство) имеют место следующие формулы:

$$\omega_i, i = [\overline{1, \infty}]$$

$$M\xi = \sum_{i=1}^{\infty} (x_i * P(\xi = x_i))$$

$$Mf(\xi) = \sum_{i=1}^{\infty} (f(xi) * P(\xi = x_i))$$

В формулах справа стоят ряды. Чтобы математические ожидания существовали надо, чтобы эти ряды сходились. Ряд сходится, если он имеет конечную сумму.

Задача: Вычислить $M\xi$, распределённой по закону Пуассона. $P(\xi = k) = (a^k/k!)e^{-a}$, где $k = \{0, 1, 2, 3, 4, \ldots, \infty\};\ a > 0$ – заданный заранее характер распределения.

Решение:

$$M\xi = \sum_{k=0}^{\infty} k * \frac{(a^k * e^{-a})}{k!} = e^{-a} \sum_{k=0}^{\infty} k * \frac{(ka^k)}{k!} =$$

$$= e^{-a} \sum_{k=0}^{\infty} \frac{(k * a^{k-1} a)}{(k-1)!} = e^{-a} a \sum_{s=0}^{\infty} \frac{a^s}{s!} \text{ (формула Маклорена) } = e^{-a} a e^{a} = a$$

Математическое ожидание случайной величины, распределённой по закону Пуассона с параметром распределения a равно этому параметру распределения.

Если ξ непрерывна, её закон распределения определяется плотностью распределения $f_\xi(x) \geqslant 0 \Rightarrow M\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x) dx$. Если имеется функция $g(\xi),\ g(\xi, \eta)$, то математическое ожидание вычисляется по формулам:

$$M_{g_\xi} = \int\limits_{-\infty}^{\infty} g(x) f_\xi(x) dx$$

$$M(\xi, \eta) = \iint\limits_{-\infty}^{\infty} g(x, y) f_{\xi\eta}(x, y) dx\, dy$$

где $f(\xi, \eta)$ - плотность совместных случайных величин.

Эти математические ожидания существуют, если все написанные несобственные интегралы сходятся.

Пример: вычислить математическое ожидание ξ, равномерно распределённое $\sim$

Решение:

$$f_\xi(x) = \frac{1}{b-a}, a \leqslant x \leqslant b \mid 0, x \in \text{в остальных случаях}$$

$$M_\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x) dx = \frac{1}{b-a} \int\limits_{a}^{b} x dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{a+b}{2}$$

Пример 2: вычислить математическое ожидание случайной величины ξ, распределённой нормально (по закону распределения Гаусса)

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx +$$

$$+ \frac{a}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx \text{ (интеграл Лапласа)} = 0 + \frac{a}{\sqrt{2\pi}\sigma} \sqrt{2\pi}\sigma = a$$

Вывод: Распределение случайной величины, распределённой нормально, равно параметру распределения.

Дисперсия случайной величины и её основные свойства.

Дисперсия $D\xi$ - число, определяемое формулой $D\xi = M(\xi - M\xi)^2$ (1), т.е. дисперсия представляет собой квадрат разности случайной величины и её математического ожидания. Другое название - квадрат среднеквадратического отклонения.

Часто в прикладных задачах вместо D рассматривают величину $\sqrt{D}$, называемую среднеквадратическим отклонением

Формулу (1) можно продолжить, тогда мы получим

$$D_\xi = M\left(\xi^2 - 2\xi M\xi + (M\xi)^2\right) = M^2\xi - 2M\xi + 2M\xi + (M\xi)^2 = M^2\xi - (M_\xi)^2,$$

откуда $(2) \quad D\xi = M^2\xi + (M\xi)^2$

1. Пусть ξ - дискретная величина, принимающая значения $x_1, \ldots, x_n$ с вероятностями $p_1, \ldots, p_n$

$$D\xi = \sum_{k=1}^{n} (x_k M\xi)^2 * p_k = (2) = \sum_{k=1}^{n} (x_k^2 * p_k)^2 - (M\xi)^2$$

2. Пусть ξ - непрерывная случайная величина, значит может быть определена функция $f_\xi(x)$.

$$D\xi = (1) = \int\limits_{-\infty}^{\infty} x^2 f_\xi(x)dx - (M\xi)^2$$

Дадим механическую интерпретацию математического ожидания и дисперсии случайной величины. Будем представлять закон распределения вероятностей $p_k = P(\xi = x_k), \sum\limits_{k=1}^{n} p_k = 1$ случайной величины ξ, как закон распределения единичной массы на прямой: в точках x_k сосредоточены массы p_k :

$$- - -\frac{x_1}{p_1} - - - \frac{x_2}{p_2} - \cdots - \frac{x_n}{x_n} - -- > x$$

Тогда

$$M\xi = \sum_{k=1}^{n} x_k P(\xi = x_k) \text{ - центр тяжести СМАТ}$$

$$D\xi = \sum_{k=1}^{n} (x_k - M_\xi)^2 * p_k \text{ - момент инерции относительно начала координат}$$

Пример: $D\xi =?,\ f_\xi(x) = e^{-\frac{(x-a)^2}{2\sigma^2}}$

Решение:

$$D\xi = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 f_\xi(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

Произведём замену переменной по формуле $y = \frac{x-a}{\sigma},\ x - a = \sigma y,\ dx = \sigma dy$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} \sigma^2 e^{-\frac{y^2}{2}} \sigma dy = \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (-y) d e^{-\frac{y^2}{2}} == \frac{\sigma^2}{\sqrt{2\pi}} \left(-y e^{-\frac{y^2}{2}} \Big|_{-\infty}^{+\infty} + \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) = \sigma^2$$

Вывод: дисперсия нормального распределения случайной величины равна второму параметру распределения (σ^2): $M_\xi = a;\ D_\xi = \sigma^2$

Свойства дисперсии

1. Дисперсия неотрицательна: $D_\xi \geqslant 0$. $D_\xi = 0 \Longleftrightarrow \xi = const$.

   Доказательство: $D\xi = M(\xi - M\xi)^2 \geqslant 0-$ по свойству монотонности.

   Пусть $\xi = c = const$.

   Тогда $D_c = M(c - M_c)^2 = (c - c)^2 = 0$.

2. Если $a = const$, то дисперсия $D(a\xi) = a^2 D\xi$.

   Доказательство:

   $D(a\xi) = M(a\xi - Ma\xi)^2 = M(a\xi - aM\xi)^2 = M[a^2(\xi - M\xi)^2] = a^2 M(\xi - M\xi) = a^2 D\xi$

3. Если ξ, η независимы, то

$$D(\xi + \eta) = M\left(\xi + \eta - M(\xi + \eta)\right)^2 = M\left((\xi - M\xi) + (\eta - M\eta)\right)^2 =$$
$$= M(\xi - M\xi)^2 + 2(\xi - M\xi)(\eta - M\eta) + M(\eta - M\eta)^2 =$$
$$= M(\xi - M\xi)^2 + 2M\left((\xi - M\xi) * (\eta - M\eta)\right) + M(\eta - M\eta)^2 = D\xi + D_\eta$$

Энтропия и информация

Энтропия как мера неопределённости

Для практики важно уметь численно оценивать степень неопределённости самых разнообразных опытов, чтобы иметь возможность их сравнивать.

Начнём с рассмотрения опытов имеющих K равновероятных исходов. Степень неопределённости каждого такого опыта определяется числом K. При $K = 1$ исход опыта не является случайным. При большом значении K предсказание результата опыта становится затруднительным.

Таким образом, искомая численная характеристика степени неопределённости должна зависть от K, т.е быть функцией $f(k)$; $f(1) = 0$; при возрастании аргумента, функция должна возрастать. Для более полного определения функции $f(k)$ необходимо предъявить к ней дополнительные требования.

Рассмотрим сложный опыт $\alpha\beta$, состоящий в одновременном выполнении опытов α и β. Неопределённость выполнения сложного опыта больше неопределённости опыта α, т.к. к его неопределённости надо добавить неопределённость опыта β. Поэтому естественно считать, что **степень неопределённости** опыта $\alpha\beta$ равна сумме неопределённостей, характеризующих α и β.

Пусть $\alpha\beta$ имеет $k * l$ равновероятных исходов, $k\alpha,\ l\beta$. Приходим к следующему условию, которму должна удовлетворять функция $f(kl) = f(k) + f(l)$. Последнее условие наталкивает на мысль принять за меру неопределённости опыта, имеющего К равновероятных исходов число $\log k$: $\log(kl) = \log k + \log l$. Такое определение меры неопределённости согласуется с первоначальными условиями, что $f(1) = \log 1 = 0$; $f(k)$ - возрастающая функция. Можно доказать, что логарифмическая функция является единственной, удовлетворяющей этим условиям.

Замечание: отметим, что выбор основания логарифма большой роли не играет, поскольку в силу известной формулы перехода можем написать $\log_b a = \log_c a / \log_c b \Rightarrow \log_b k = \log_b a * log_a k$ сводится к домножению на константу, т.е. равносилен простому изменению **единицы измерения** степени неопределённости. Обычно за меру степени неопределённости берут логарифмы при основании 2: $log_2 k = log k$, причём основание 2 не фиксируют. Т.е. за единицу измерения степени неопределённости принимают неопределённость опыта, имеющего 2 равновероятных исхода: $\log_2 2 = 1$ бит. Везде далее будем пользоваться двоичными единицами измерения.

Таблица вероятности для опыта, имеющего K равновероятных исходов:

α				
Исходы	A_1	A_2	$\dots$	A_k
Вероятности	$\frac{1}{k}$	$\frac{1}{k}$	$\dots$	$\frac{1}{k}$

Поскольку при наших допущениях неопределённость равна $f(k) = \log k$. В этом случае каждый отдельный исход вносит неопределённость $\frac{1}{k}$. $\frac{\log k}{k} = \frac{1}{k} \log k = -\frac{1}{k} \log \frac{1}{k}$.

В самом общем случае опыт имеет следующую таблицу вероятности:

α				
Исходы	A_1	A_2	$\dots$	A_k
Вероятности	$P(A_1)$	$P(A_2)$	$\dots$	$P(A_k)$

Для опыта общая мера неопределённости равна $-p(A_1) \log p(A_1) - p(A_2) \log p(A_2) - \dots - p(A_k) \log p(A_k) = H(\alpha)$ - энтропия опыта α

Рассмотрим некоторые свойства энтропии $H(\alpha)$:

1. $H(\alpha) \geqslant 0$

   Доказательство:

   $-p(A) \log p(A) \geqslant 0$ (множители $\in$ промежутку $(0 \leqslant p(A) \leqslant 1)$)

   $-p(A) \log p(A) = 0 \Longleftrightarrow \{p = 0; p = 1\}$

   В случае, если опыт имеет K попарно несовместных исходов, то $H(\alpha) = 0$ равносильно тому, что один исход - достоверное событие, а все другие - невозможны, так как $\big(p(A_1) + \dots + p(A_k) = 1\big)$. Это обстоятельство хорошо согласуются с величиной $H(\alpha)$ - только в этом случае опыт вообще не содержит неопределённости.

2. Из всех опытов с K исходами самым неопределённым является опыт опыт с K равновероятными исходами. Можно показать, что имеет место неравенство

   $H(\alpha) = -p(A_1) \log p(A_1) - \dots - p(A_k) \log p(A_k) \leqslant H(\alpha_0)$

   $H(\alpha_0) = \log k = -\frac{1}{k} - \dots - \frac{1}{k}$.

   Равенство достигается при равных вероятностях $P(A_i);\ i = [\overline{1, k}]$

Пример: Имеется две урны с 20-ю шарами каждая. Первая - 10 белых, 5 чёрных, 5 красных. Вторая - 8 белых, 8 чёрных, 4 красных.

Из каждой урну вынимают по 1 шару. Исход какого из двух опытов следует считать более неопределённым?

Решение: Обозначим опыты как A1 и A2.

A1

Исходы	Бел	Чёр	Крас
Вероятности	1/2	1/4	1/4

A2

Исходы	Бел	Чёр	Крас
Вероятность	2/5	2/5	1/5

Энтропия опыта A1: $H(\alpha_1) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = -\frac{1}{2} * 1 - \frac{1}{2} * (-2) = -\frac{1}{2} + 1 = 1,5$бита.

Энтропия опыта A2: $H(\alpha_2) = -\frac{2}{5}\log\frac{2}{5} - \frac{2}{5}\log\frac{2}{5} - \frac{1}{5}\log\frac{1}{5} = -\frac{4}{5}(\log 2 - \log 5) - \frac{1}{5}(\log 1 - \log 5) = -0.8 + -\frac{4}{5}\log 5 + \frac{1}{5}\log 5 = -0.8 + \log 5 = 1,52$ бита.

Вывод: Если оценивать степень неопределённости опыта его энтропией, то исход второго опыта более неопределённый, нежели первого.

Историческая справка

Исторически первые шаги к введению понятия энтропии были сделаны в 1928 году американским инженером-связистом Хартли, предложившим характеризовать степень неопределённости опыта с К различными исходами числом $\log k$. Предложенная им мера степени неопределённости иногда бывает удобна в некоторых практических задачах, но часто оказывается малопоказательной, поскольку полностью игнорирует различие между характером имеющихся исходов. Поэтому почти невероятному исходу у Хартли придаётся такое-же значение, как и исходу весьма вероятному. Однако, он считал, что различия между отдельными исходами определяются в первую очередь "психологическими факторами" и должны учитываться лишь психологами, но не инженерами или математиками.

Ошибочность точки зрения Хартли была показана другим американским инженером - математиком К. Шенноном. Он предложил принять в качестве меры неопределённости опыта с K различными исходами $A_1, \ldots, A_k$ величину

$H(\alpha) = -p(A_1)\log p(A_1) - \ldots - p(A_k \log p(A_k).$

Иначе говоря, исходу A_i следует приписать неопределённость, равную $-\log p(A_i)$. В качестве неопределённости всего опыта $H(\alpha)$ принимается среднее значение случайной величины (математическое ожидание), равное $H(\alpha)\xi$,где ξ принимают значения $-\log p(A_i)$ с вероятностями $p(A_i)$.

Таким образом, загадочные "психологические факторы" учитываются с помощью использования понятия вероятности, имеющего чисто математический, а точнее статистический характер.

Использование величины $H(\alpha)$ в качестве меры неопределённости опыта A оказалось полезным во многих областях, а особенно в теории передачи сообщений по линиям связи.

Энтропия сложных событий. Условная энтропия

Условная энтропия. Пусть имеются два независимых опыта A, B с таблицей вероятностей $A_1, p(A_1);\ \ldots\ ; A_k, p(A_k); \quad B_1, p(B_1);\ \ldots\ ; B_l, p(B_l)$.

Рассмотрим сложный опыт $\alpha\beta$, когда осуществляются оба опыта одновременно, имеющий $k * l$ исходов ($A \times B$ - декартово произведение).

$A_1B_1 : \alpha = A_1; \beta = B_1$

Очевидно, что неопределённость опыта $\alpha\beta$ больше неопределённости каждого из опытов, из-за осуществления обоих опытов. Поэтому имеет место соотношение $H(\alpha\beta) = H(\alpha) + H(\beta)$. Написанное равенство называется правилом сложения энтропии для опытов α и β.

Для доказательства этого равенства рассмотрим выражение

$$H(\alpha\beta) = -p(A_1B_1)\log p(A_1B_1) - \ \ldots\ - p(A_kB_l)\log p(A_kB_l)$$

$$\alpha, \beta \text{ - независимы, следовательно } p(A_iB_j) = p(A_i) * p(B_j) \Rightarrow$$

$$\log p(A_iB_j) = \log p(A_i)p(B_j) = \log p(A_i) + \log p(B_j).$$

Предположим далее, что α и β - зависимые опыты (пример: α, β - последовательные извлечения двух шаров из одной урны.) Постараемся выяснить, чему равна энтропия сложного опыта $\alpha\beta$ в этом случае.

Здесь уже нельзя заменить $p(A_1B_1), p(A_1B_2), \ \ldots$ произведением вероятностей, а необходимо использовать условную вероятность $p(A_1B_1) = p(A_1) * p(B_1|A_1)$

В этом случае можно доказать следующую формулу:

$H(\alpha\beta) = H(\alpha) + [p(A_1) * H(\beta|A_1) + p(A_2) * H(\beta|A_2) + \ \ldots\ + p(A_k) * H(\beta|A_k)]\,(*)$, где $H(\beta|A_i)$ - условная энтропия опыта β при условии, что значение опыта α равно A_i.

$H(\beta|A_i) = -p(B_1|A_i)\log p(B_1|A_i) - p(B_2|A_i)\log p(B_2|A_i) - \ \ldots\ - p(B_e|A_i)\log p(B_e|A_i)(**)$

Это выражение представляет собой энтропию опыта β при условии, что имеет место событие A_i.

$$\begin{cases} H(\alpha,\beta)=H(\alpha)+H(\beta) & \text{(для независимых } \alpha,\beta) \\ H(\alpha,\beta)=H(\alpha)+[\ldots](*) & \text{(для зависимых } \alpha,\beta) \end{cases}$$

Первый член последнего выражения $(*)$ - энтропия опыта α. Что же касается второго - он есть математическое ожидание случайной величины, принимающей с вероятностями $p(A_1),\,\ldots\,,p(A_k)$ значения $H(\beta|A_1),\,\ldots\,,H(\beta|A_k)$, то есть значения, равные условной энтропии опыта β, при условии, что опыт α имеет исходы $\alpha:A_1,\,\ldots\,,A_k$. Это среднее значение естественно назвать **условной энтропией** выполнения опыта β при условии выполнения опыта α,

$$H(\beta|\alpha)=[\ldots](*)=p(A_1)H(\beta|A_1)+p(A_2)H(\beta|A_2)+\,\ldots\,+p(A_k)H(\beta|A_k)$$

Тогда соотношение $(*)$ переписывается как $H(\alpha\beta)\quad=\quad H(\alpha)\;+\;H(\beta|\alpha)(*)$; α,β - зависимы.

Это и есть общее правило для определения энтропии сложного опыта $\alpha\beta$. Его также можно назвать правилом сложения энтропии, для **зависимых** опытов $\alpha\beta$.

Укажем основные свойства условной энтропии:

1. $H(\beta|\alpha)\geqslant 0$.
2. $p(A_1),\,\ldots\,,p(A_k)\neq 0$ (опыт имеет к штук исходов).

   Тогда $H(\beta|\alpha)=0\iff H(\beta|A_1)=\,\ldots\,=H(\beta|A_k)=0$, т.е. при любом исходе опыта α результат опыта β полностью определён, и при этом имеем $H(\alpha\beta)=H(\alpha)$.

   Если α и β **независимы**, то тогда $H(\beta|\alpha)=H(\beta)$, и $H(\alpha\beta)=H(\alpha)+H(\beta)$.
3. Во всех случаях условная энтропия $H(\beta|\alpha)$ заключается между 0 и $H(\beta)$:

   $0\leqslant H(\beta|\alpha)\leqslant H(\beta)$.

   Таким образом случаи, когда исход β полостью предопределяется исходом α и когда опыты α и β независимы, являются в определённом смысле крайними.
4. Условная энтропия.

   $$H(\alpha\beta)=H(\beta\alpha)\Rightarrow H(\alpha)+H(\beta|\alpha)=H(\beta)+H(\alpha|\beta)\Rightarrow$$

   $$\Rightarrow H(\beta|\alpha)=H(\alpha|\beta)(.)+H(\beta)-H(\alpha)$$

   $H(\alpha|\beta)=0$ (исход опыта β полностью определяет опыта α

   $H(\beta|\alpha)=H(\beta)-H(\alpha)$

Задача: Задача о болезненной реакции.

Известно, что некоторой болезнью в среднем болеют 2 человека из 100. Для выявления больных используется определённая реакция, которая всегда оказывается положительной в том случае, когда человек болен. Если же человек здоров, то она столь же часто бывает положительной, как и отрицательной. Пусть опыт β состоит в определении того болен или здоров человек, а опыт α - в определении результата указанной реакции. Спрашивается, какова будет энтропия $H(\beta) =?$ опыта β и условная энтропия $H(\beta|\alpha) =?$.

Решение: Очевидно, что β имеет 2 исхода: $\beta : \{B_1$ - здоров; B_2 - болен$\}$.

$p(B_1) = 0.98; \quad p(B_2) = 0.02.$

$H(\beta) = -0.98 \log 0.98 - 0.02 \log 0.02 \approx 0.14$бит.

$H(\beta) \approx 0.14.$

Рассмотрим опыт α: $\alpha : A_1 -$ положительная реакция; $A_2 -$ отрицательная реакция

$p(A_1) = p(\frac{B_1}{2} + B_2) = p(\frac{B_1}{2}) + p(B_2) = 0.49 + 0.02 = 0.51.$

$p(A_2) = p(\frac{B_1}{2}) = 0.49.$

$\alpha = A_1 : p(B_1|A_1) = \frac{p(B_1 A_1)}{p(A_1)} = \frac{0.49}{0.51} = \frac{49}{51}.$

$\alpha = A_2 : p(B_2|A_1) = \frac{p(B_2 A_1)}{p(A_1)} = \frac{0.02}{0.51} = \frac{2}{51}.$

Пользуясь этими данными мы можем найти условную энтропию $H(\beta)$ при выполнении события A_1

$$H(\beta|A_1) = -\frac{49}{51} \log \frac{49}{51} - \frac{2}{51} \log \frac{2}{51} \approx 0.24 \text{ бит.}$$

При $\alpha = A_2 \Rightarrow \beta = B_1;\ H(\beta|A_2) = 0$, т.е. мы с уверенностью можем утверждать, что человек здоров, и опыт β имеет исход B_1.

Таким образом, условная энтропия β при условии осуществления α будет равна

$$H(\beta)\ = 0.14|sysH(\beta|A_1) \approx 0.045H(\beta|A_2) = 0 \sim\sim$$

$$H(\beta|\alpha) = p(A_1)H(\beta|A_1) + p(A_2)H(\beta|A_2) \approx 0.51 * 0.24 + 0.49 * 0 = 0.12 \text{ бит.}$$

Иначе говоря, выполнение опыта α уменьшает неопределённость опыта β на 0.002 бита.

Понятие об информации.

Вернёмся вновь к величине $H(\beta)$, характеризующей степень неопределённости опыта β. Равенство этой величины 0 означает, что исход опыта β заранее известен. Большее или меньшее значение числа $H(\beta)$ отвечает большей или меньшей проблематичности определения результата опыта β. Какое-либо измерение или наблюдение в виде опыта α, предшествующее β может ограничить количество возможных исходов опыта β, и тем самым уменьшить степень его неопределённости: так, к примеру степень неопределённости опыта, состоящего в нахождении самого тяжёлого из 3 грузов уменьшается после сравнения на весах двух из них.

Для того, чтобы результат измерения(наблюдения) α мог сказаться на последующем опыте β необходимо, чтобы α **не был известен заранее.** Поэтому, α можно рассматривать как вспомогательный опыт, также имеющий несколько допустимых исходов.

Тот факт, что осуществление α уменьшает степень неопределённости β отражается в неравенстве, где условная энтропия $H(\beta|\alpha) \leqslant H(\beta)$ первоначальной энтропии опыта β.

При этом, если опыт β не зависит от α, то осуществление α не уменьшает энтропии β. Это значит, что $H(\beta|\alpha) = H(\beta)$. Если же результат α полностью предопределяет исход опыта β, то энтропия β уменьшается до 0: $H(\beta|\alpha) = 0$. Таким образом, разность $\boxed{I(\beta,\alpha) = H(\beta) - H(\beta|\alpha)(*)}$.

Таким образом написанная разность указывает, насколько осуществление α уменьшает неопределённость β, т.е. как много мы узнаём об исходе опыта β, произведя измерение(наблюдение) в виде опыта α. Эта разность $(*)$ называют количеством информации относительно опыта β, содержащейся в опыте α. Таким образом, мы получаем возможность **численного измерения** информации. К примеру, в условиях задачи о болезненной реакции можно сказать, что используемая реакция в виде опыта α даёт информацию о заболевании в виде опыта β, равное $0.14 - 0.12 = 0.02$ бита. Эта цифра и оценивает пользу реакции.

Соотношение между понятиями энтропии и информации напоминает соотношение между физическими понятиями потенциала и разности потенциалов. Энтропия есть абстрактная мера неопределённости. Ценность этого понятия в значительной мере заключается в том, что оно позволяет оценить влияние на опыт β какого-либо другого опыта α как разность энтропий по формуле $(*)$.

Подчеркнём также, что информация относительно опыта β, содержащаяся в опыте α представляет собой среднее значение (математическое ожидание) случайной величины $H(\beta) - H(\beta|A_i)$, связанной с отдельными исходами A_i опыта α.

Пример: Задача о шарах и предварительной информации.

Пусть опыт β состоит в извлечении одного шара из урны:

β : 1 шар из 5 чёрных и 10 белых.

А опыт α_k состоит в предварительном извлечении (без возвращения обратно) K шаров:

α_k : K шаров извлечено.

$H(\beta) = ?$

$I(\beta, \alpha_1) = ?$

$I(\beta, \alpha_2) = ?$

$I(\beta, \alpha_{13}) = ?$

$I(\beta, \alpha_{14}) = ?$

Чему равна энтропия $H(\beta)$ и информация, содержащаяся в опыте α_1?

Решение:

$H(\beta) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} \approx 0.92$бита.

$$I(\beta, \alpha_1) = H(\beta) - H(\beta|\alpha_1)$$

$$H(\beta|\alpha_1) = \left[p(A_1^{\text{чёр}}) * H(\beta|A_1^{\text{чёр}})\right] + \left[p(A_1^{\text{бел}}) * H(\beta|A_1^{\text{бел}})\right] =$$

$$= -\frac{1}{3}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}})}_{4/14} * \log\frac{4}{14} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}})}_{5/7} * \log\frac{5}{7}\right] -$$

$$-\frac{2}{3}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{бел}})}_{5/11} * \log\frac{5}{11} + \underbrace{p(B^{\text{бел}}|A_1^{\text{бел}})}_{9/14} * \log\frac{9}{14}\right] \approx 0.004\text{бит.}$$

$$I(\beta|\alpha_2) = H(\beta) - H(\beta|\alpha_2)$$

$$H(\beta|\alpha_2) = p(A_1^{\text{ч}}A_2^{\text{ч}}) * H(\beta|A_1^{\text{ч}}A_2^{\text{ч}}) + p(A_1^{\text{ч}}A_2^{\text{б}}) * H(\beta|A_1^{\text{ч}}A_2^{\text{б}}) + p(A_1^{\text{б}}A_2^{\text{б}}) * H(\beta|A_1^{\text{б}}A_2^{\text{б}}) =$$

$$= -\left\{\frac{C_5^2}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}}A_2^{\text{чёр}})}_{3/13} * \log\frac{3}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}}A_1^{\text{чёр}})}_{10/13} * \log\frac{10}{13}\right] +\right.$$

$$+\frac{C_{10}^1 C_5^1}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}}A_2^{\text{бел}})}_{4/13} * \log\frac{4}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}}A_2^{\text{бел}})}_{9/13} * \log\frac{9}{13}\right] +$$

$$\left.+\frac{C_{10}^2}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{бел}}A_2^{\text{бел}})}_{5/13} * \log\frac{5}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{бел}}A_2^{\text{бел}})}_{8/13} * \log\frac{8}{13}\right]\right\} \approx 0.008\text{бит.}$$

$I(\beta, \alpha_{13}) = H(\beta) - H(\beta|\alpha_{13})$

$H(\beta|\alpha_{13})$ - здесь неопределённость только в оставшихся двух шарах: они должны быть двух цветов, значит, мы взяли 4 чёрных и 9 белых шаров.

$$H(\beta|\alpha_{13}) = \frac{C_5^4 C_{10}^9}{C_{15}^{13}}(-)\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2})\right) = -\frac{\frac{5!}{4!(5-4)!}\frac{10!}{9!(10-9)!}}{\frac{15!}{13!(15-13)!*(-1)}} = \frac{2*5*10}{14*15} \approx 0.44$$

$I(\beta, \alpha_{14}) = [H(\beta|\alpha_1 4) = 0] \approx 0.92.$