1) Энтропия: условная, совместная

2) Dkc, KL

3) ф-ция правдоподобия и ABC

Опр. A - событие

münni векция

$$S \text{ (surprise)} = \text{неожиданность}$$
$$S(A) = \log_{\frac{1}{2}} P(A)$$

сигн. (рабо-тую)
послед из 0 и 1.
0010|111|0101...

$P(A) = \frac{1}{2}$ $S(A) = 1$ Чем ниже вероятн,
$P(A) = \frac{1}{8}$ $S(A) = 3$ тем выше неожид.

$$P(x_5 = 1, x_8 = 1, x_2 = 1) = \frac{1}{8}$$

Интуиция

бит → неожид.
кодировка

Земля.

	бит	кодировка
Зел. гел. $\frac{1}{8}$	3	000
Кр. гел. $\frac{1}{8}$	3	001
Фиол гел. $\frac{1}{4}$	2	01
Черн. гел. $\frac{1}{2}$	1	1

$1100\,/\sim Z\,Z\,K.$

KK КЗ ЗК ЗЗ

0,9 К 0,1З

к/З

KKK ЗК ЗК...

0 1

00 01 10 11
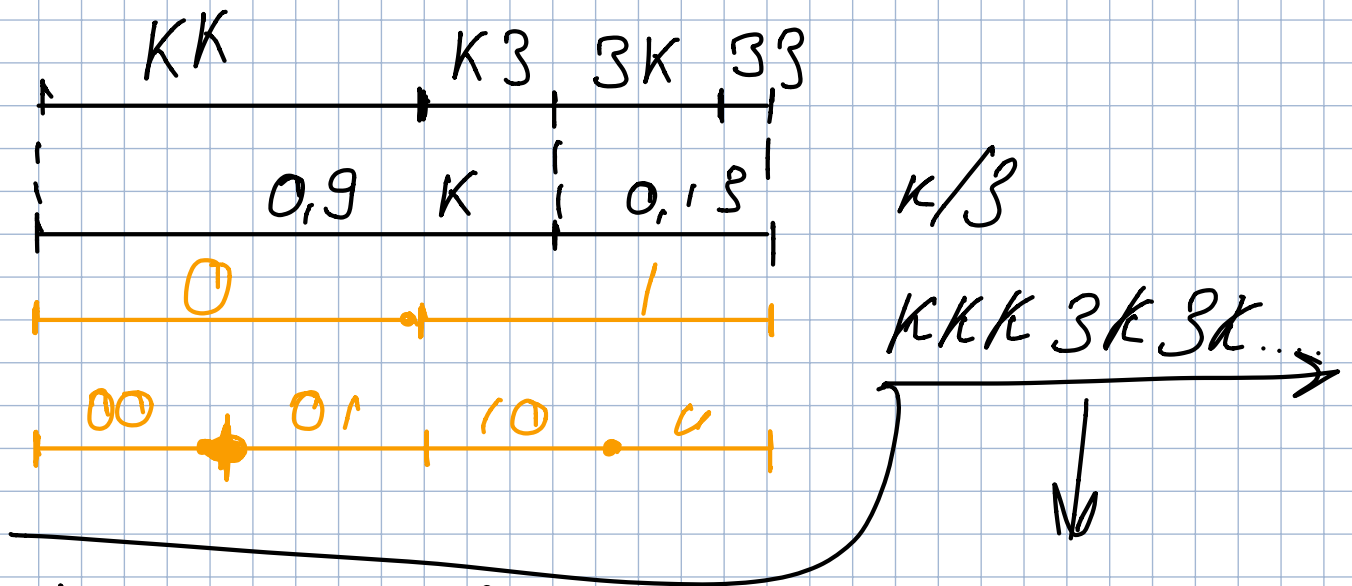
Ну а подобавляем....

Энтропия

с.в. X — дискретная с.в.

$H(X) =$ ожидаемая неожидан.ость от наблюдении X.

$$= \mathbb{E}(S_X) \quad S_\xi = \log_2 P(X = \xi)$$

ξ	5	6	7	8
$P(X = \xi)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$

$H(X) = \frac{1}{8} \cdot \log\left(\frac{1}{8}\right) +$
$+ \frac{1}{8} \log\left(\frac{1}{8}\right) + \frac{1}{2} \log\left(\frac{1}{2}\right) +$

$$+ \frac{1}{4} \log \left(\frac{1}{4} \right)$$

Зависит только от вероятн.

Поэтому можно заменить на функцию вероятностей. (Только диад. функции.)

$$H(X) = \frac{6}{8} + \frac{1}{2} + \frac{1}{2} = 1,75 \ (\text{бит})$$

Сколько в среднем уходит бит на передачу одного X. при оптимальном кодировании.

Данетки.

$H(X)$ — сколько вопросов в данетках нужно в среднем задать, чтобы угадать X.

Условная энтропия $H(X|Y)$

[смысл]: сколько (в среднем) бит информации нужно передать чтобы узнать X, если Y уже знаем.

$$H(X|Y=y) = \sum_{x} P(X=x|Y=y) \cdot \log_2 P(X=x|Y=y)$$

$$H(X|Y) = \sum_{y} P(Y=y) \cdot H(X|Y=y)$$

	$y=1$	$y=2$	$y=3$	Упр
$x=1$	0,1	0,2	0,3	$H(X|Y) - ?$
$x=2$	0,1	0,1	0,2	$H(Y|X) - ?$

Совместная энтропия $H(X,Y)$.

Сколько (в среднем) бит информации нужно узнать, чтобы узнать X и Y.

$$H(X,Y) = \sum_{x,y} P(X=x, Y=y) \cdot \log_2 P(X=x, Y=y)$$

Th.

① $H(X) + H(Y|X) = H(X,Y)$.

② $H(Y) + H(X|Y) = H(Y,X)$.   не симметр.

Опр. Взаимная информация $\begin{cases} (X|Y) \\ (X,Y) \\ \text{симметр.} \end{cases}$

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

Опр. Кросс-энтропия $CE(X||Y)$

сколько бит информации (в среднем) надо получить, если истинное распредел. X, а код-вка оптимальна под распред. Y.

Пример

		Б	Ф	К	З
ист.	p	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
предпол.	q	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$

кросс энтропия из p в q

$$CE(p||q) = \mathbb{E}_p\left(\log_{\frac{1}{2}} q\right) =$$

$$= \sum_w p(w) \cdot \log_{\frac{1}{2}} q(w)$$

Сколько затратим вопросов при условии, что наше распределение q ложное

Опр. Дивергенция Кульбака-Лейблера

$KL(X||Y)$ Потери в битах от неопт. кодирования.

$$KL(p||q) = CE(p||q) - H(p)$$