Энтропия и KL-дивергенция

y	"1"	"2"	"3"	...
$\mathbb{P}\{y=y_i\}$	p_1	p_2	p_3	...

$$\boxed{H_p = \sum_i \mathbb{P}\{y=y_i\} \log_{\frac{1}{2}} \mathbb{P}\{y=y_i\}}$$

Смысл:
1) среднее число бинарных вопросов, к. нужно задать, чтобы завершить игру.
2) мера хаотичности и неопред. системы [информация]

y	"1"	"2"	"3"	...
$\mathbb{P}_A\{y=y_i\}$ истина	p_1	p_2	p_3	...
$\mathbb{P}_B\{y=y_i\}$ думаем	q_1	q_2	q_3	...

$$\boxed{CE(A\|B) = \sum_i \mathbb{P}_A\{y=y_i\} \log_{\frac{1}{2}} \mathbb{P}_B\{y=y_i\}}$$

Смысл: среднее число вопросов, к. нужно задать, чтобы завершить бин. игру при исп. оптим. стр., если страт. строим на основе распр. B, а ист. распр. A.

$$D_{KL}(a\|b) = CE(a\|b) - H(a)$$

Док-во: $x-1 \geq \ln x$

$$x = \frac{a_i}{b_i}: \quad \frac{a_i}{b_i} - 1 \geq \ln \frac{a_i}{b_i}$$

$$\underbrace{\sum_i a_i - b_i}_{=0\ (1-1)} \geq \sum_i b_i \ln a_i - b_i \ln b_i$$

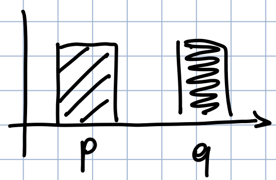
$$\sum_i b_i \ln a_i - b_i \ln b_i \leq 0$$

$$\left[\log_e a \cdot \log_{\frac{1}{2}} e = \log_{\frac{1}{2}} e^{\log_e a} = \log_{\frac{1}{2}} a\right]$$

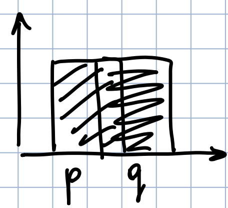
$$\sum_i b_i \log_{\frac{1}{2}} a_i - b_i \log_{\frac{1}{2}} b_i \geq 0$$

Замеч-я:

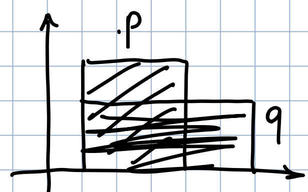
①



$D_{KL}(p\|q) = D_{KL}(q\|p) = \infty$



$D_{KL}(p\|q) = D_{KL}(q\|p) = \infty$



$D_{KL}(p\|q)$ – конечна

$D_{KL}(q\|p) = \infty$

② Непрерывный случай:
$H(X)$ может быть < 0.
(а в дискретном всегда $H(X) \geq 0$)

$D_{KL} \geq 0$

№1 Найти $H(X)$

а)

X	1	5	7
$\mathbb{P}\{X = x_i\}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$H(X) = \frac{1}{3}\log_{\frac{1}{2}}\frac{1}{3} + \frac{1}{3}\log_{\frac{1}{2}}\frac{1}{3} + \frac{1}{3}\log_{\frac{1}{2}}\frac{1}{3} = \log_{\frac{1}{2}}\frac{1}{3}.$$

б)
$$f_X(x) = \begin{cases} \frac{1}{a}, & \text{если } x \in [0;a] \\ 0, & \text{иначе} \end{cases}$$

$$H(X) = \int_0^a \frac{1}{a}\log_{\frac{1}{2}}\frac{1}{a}\,dx = \frac{1}{a}\log_{\frac{1}{2}}\frac{1}{a}(a-0) = \log_{\frac{1}{2}}\frac{1}{a} = \log_2 a.$$

в) $X \sim \mathcal{N}(\mu, \sigma^2)$

$$H(X) = -\int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \ln \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} =$$

$$= + \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left(+ \frac{1}{2}\ln 2\pi\sigma^2 + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right) =$$

$$= \frac{1}{2}\ln 2\pi\sigma^2 + \frac{1}{2\sigma^2}\int_{\mathbb{R}} \underbrace{\frac{dx}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{f_X(x)} \cdot \underbrace{(x-\mu)^2}_{g(X)} = \quad \text{(LOTUS)}$$

$$= \mathbb{E}\left((x-\mu)^2\right) = \sigma^2$$

$$= \frac{1}{2}\ln 2\pi\sigma^2 + \frac{1}{2} = \frac{1}{2}\ln 2\pi\sigma^2 e.$$

№2 Найти KL-дивергенцию.

а) $A = Bin(2, \frac{1}{3})$

$B =$ равнов. на $0, 1, 2$

X	0	1	2
A	$C_2^0\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^2 = \frac{4}{9}$	$C_2^1\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^1 = \frac{4}{9}$	$C_2^2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^0 = \frac{1}{9}$
B	$1/3$	$1/3$	$1/3$

$$CE(A\|B) = \frac{4}{9}\log_{\frac{1}{2}}\frac{1}{3} + \frac{4}{9}\log_{\frac{1}{2}}\frac{1}{3} + \frac{1}{9}\log_{\frac{1}{2}}\frac{1}{3} = \log_{\frac{1}{2}}\frac{1}{3} \approx 1.58$$

$$H(A) = \frac{4}{9}\log_{\frac{1}{2}}\frac{4}{9} + \frac{4}{9}\log_{\frac{1}{2}}\frac{4}{9} + \frac{1}{9}\log_{\frac{1}{2}}\frac{1}{9} \approx 1.04$$

$$D_{KL}(A\|B) = 1.58 - 1.04 = 0.54.$$

б) $B = Bin(2, \frac{1}{3})$

$A =$ равнов. на $0, 1, 2$

$$CE(A\|B) = \frac{2}{3}\log_{\frac{1}{2}}\frac{4}{9} + \frac{1}{3}\log_{\frac{1}{2}}\frac{1}{9} \approx 1.84$$

$$H(A) = \log_{\frac{1}{2}}\frac{1}{3} \approx 1.58$$

$$KL(A\|B) = 0.26$$

в) A − равнов. на $0, 1, 2, 3$ $\quad D_{KL}(A\|B) = \infty$

B − $Bin(2, \frac{1}{3})$

ММП

$X_1 \dots X_N$ – выборка случ. величин из какого-то распр.

$\Theta = (\Theta_1 \dots \Theta_k)$ – параметры распр.

Дискр:

$$L(X_1, \dots, X_N \mid \Theta) = \mathbb{P}\{X_1 = x_1 \dots X_n = x_n \mid \Theta\} \overset{iid}{=} \prod_{i=1}^{N} \mathbb{P}\{X_i = x_i \mid \Theta\}$$

Непр:

$$L(X_1, \dots, X_N \mid \Theta) = f_{X_1 \dots X_N \mid \Theta}(x_1, \dots, x_n \mid \Theta) \overset{iid}{=} \prod_{i=1}^{N} f(x_i \mid \Theta)$$

$\ell := \ln L$

Задача: $\hat{\Theta}_{ML} = \underset{\Theta}{\operatorname{argmax}}\ \ell(X \mid \Theta)$

№3 $\quad Y_i \sim iid$

Y	2	4	6
$\mathbb{P}\{Y = y_i\}$	$3a$	$1-5a$	$2a$

$Y_1 = 6,\ Y_2 = 4,\ Y_3 = 6$

$$L = \prod_{i=1}^{3} \mathbb{P}\{Y_i = y_i\} = 4a^2(1-5a)$$

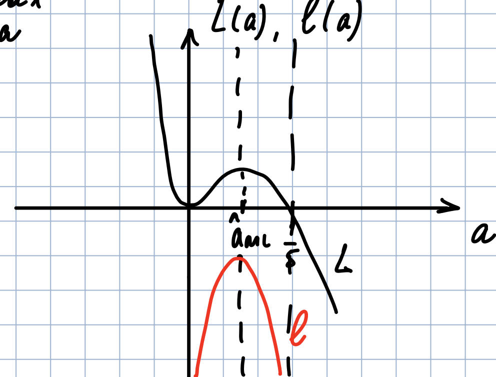
$$\ell = 2\ln 2a + \ln(1-5a) \to \max_a$$

$$\ell'_a = \frac{4}{2a} - \frac{5}{1-5a}$$

$$\frac{2}{\hat{a}} = \frac{5}{1-5\hat{a}}$$

$$2 - 10\hat{a} = 5\hat{a}$$

$$\hat{a}_{ML} = \frac{2}{15}$$



№4) $X_1 \dots X_n \sim N(\mu, \sigma^2)$

$\hat{\mu}_{ML} - ?$ $\quad \hat{\sigma}^2_{ML} - ?$

$$f_i(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$L = \prod_{i=1}^{N} f_i(x) \; ; \quad l = \sum_{i=1}^{N} \ln f_i(x)$$

$$l = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 \longrightarrow \max_{\mu,\sigma}$$

$$l'_\mu = +\frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) \Rightarrow \sum_{i=1}^{N} x_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{n} = \bar{X}.$$

$$l'_\sigma = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$n = \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{N}(x_i-\bar{X})^2$$

$$\hat{\sigma}^2 = \sum_{i=1}^{N}\frac{(x_i-\bar{X})^2}{n}$$ — смещ. оц-а дисп.

$$l''_{\mu\mu} = -\frac{n}{\sigma^2}\Big|_{\hat{\mu},\hat{\sigma}^2} = \frac{-n}{\hat{\sigma}^2}$$

$$l''_{\sigma\sigma} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu)^2\Big|_{\hat{\mu},\hat{\sigma}^2} = \frac{n}{\hat{\sigma}^2} - \frac{3\hat{\sigma}^2 \cdot n}{\hat{\sigma}^4} = \frac{-2n}{\hat{\sigma}^2}$$

$$l'_{\mu\sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{N}(x_i-\mu)\Big|_{\hat{\mu},\hat{\sigma}^2} = -\frac{2}{\hat{\sigma}^3}\sum_{i=1}^{N}(x_i-\bar{X}) = 0$$

$$H\Big|_{\hat{\mu},\hat{\sigma}} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{2n}{\hat{\sigma}^2} \end{pmatrix}$$ — отр. опред.

Св-а ML-оценок:

1. Неасимп-е:

$$\widehat{g(\theta_{ML})} = g(\hat{\theta}_{ML}) \quad \forall \text{ гладкой } g(\cdot)$$ (инвар. отн. гладк. преобр.)

т.е. $\widehat{(p_1 - p_2)}_{ML} = \hat{p}_{1ML} - \hat{p}_{2ML}$

2. Асимп-е:

если у L сущ-т единств. max

$\exists L'''$

Обл. знач. выб-и не зависит от парам-в, то при $n \to \infty$:

1. $\mathbb{E}(\hat{\theta}_{ML}) \overset{n\to\infty}{=} \theta$
2. $\operatorname{plim}\limits_{n\to\infty} \hat{\theta}_{ML} = \theta$
3. $\lim\limits_{n\to\infty} \mathbb{E}(\hat{\theta}' - \theta)^2 \overset{n\to\infty}{>} \mathbb{E}(\hat{\theta}_{ML} - \theta)^2 \quad \forall \hat{\theta}' \neq \hat{\theta}_{ML}$

   $\uparrow$ $\forall$ др. оценка
4. $\hat{\theta} \overset{n\to\infty}{\sim} N(\theta, Var(\hat{\theta}))$