① ML как Байесовский вывод

параметры → $\bar\theta$, данные → D

$$\underbrace{p(\bar\theta \mid D)}_{\text{posterior}} = \frac{\overbrace{p(\bar\theta)}^{\text{prior}}\ \overbrace{p(D \mid \bar\theta)}^{\text{likelihood}}}{p(D)}$$

$p(D)$ — evidence probability

$$p(\bar\theta \mid D) \propto p(D \mid \bar\theta) \cdot p(\bar\theta)$$

1) $p(D \mid \bar\theta) \xrightarrow[\bar\theta]{} \max$, $\quad \bar\theta_{ML} = \arg\max_{\bar\theta} p(D \mid \bar\theta)$

2) $p(\bar\theta \mid D) \xrightarrow[\bar\theta]{} \max$, $\quad \bar\theta_{MAP} = \arg\max_{\bar\theta} p(\bar\theta \mid D)$

3) $p(\bar x \mid D) = \int p(\bar x, \bar\theta \mid D)\, d\bar\theta = \int p(\bar x \mid \boxed{\bar\theta}, D)\, p(\bar\theta \mid D)\, d\bar\theta$

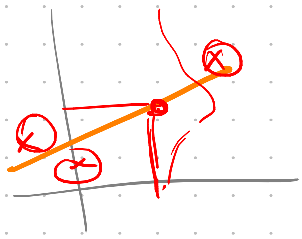
predictive distribution

$$= \int \underbrace{p(\bar x \mid \bar\theta)}_{\text{likelihood of } \bar x}\ \underbrace{p(\bar\theta \mid D)}_{\text{posterior}}\, d\bar\theta$$

Conditional independence

x и y усл. нез. при усл. z, если

$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z)$$

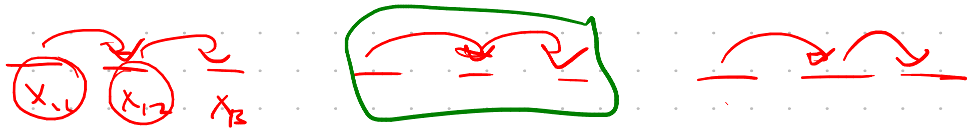


Assumption №1

$D = \{\bar x_1, \dots, \bar x_n\}$

$$p(\bar x_1, \dots, \bar x_n \mid \bar\theta) = p(\bar x_1 \mid \bar\theta)\, p(\bar x_2 \mid \bar\theta) \dots p(\bar x_n \mid \bar\theta)$$

$$p(D \mid \bar\theta) = p(\bar x_1 \dots \bar x_N \mid \bar\theta) = \prod_{n=1}^{N} p(\bar x_n \mid \bar\theta)$$



$$\log p(D \mid \bar\theta) = \sum_{n=1}^{N} \log p(\bar x_n \mid \bar\theta)$$

$$\log p(\bar\theta \mid D) = \log p(\bar\theta) + \sum_{n=1}^{N} \log p(\bar x_n \mid \bar\theta) + \overbrace{\text{Const}}^{-\log p(D)}$$

② Bernoulli trials   heads / tails

$D = hhthttht$ $\quad x_1 = h,\ x_2 = h, \dots, x_6 = t, \dots$

$p(x_1|\theta) = \theta \qquad p(x_6|\theta) = 1-\theta$

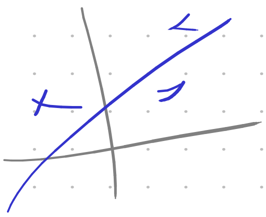
$\theta = p(\text{heads}), \quad 1-\theta = p(\text{tails})$

$$p(D|\theta) = \prod_n p(x_n|\theta) = \theta^4 (1-\theta)^5$$

$D = \{ n \times \text{heads},\ m \times \text{tails} \}$ $\qquad p(D|\theta) = \theta^n (1-\theta)^m$

1) $\theta_{ML} = \underset{\theta}{\operatorname{argmax}}\ p(D|\theta)$

$$\frac{\partial p(D|\theta)}{\partial \theta} = n \cdot \theta^{n-1} \cdot (1-\theta)^m - m(1-\theta)^{m-1} \cdot \theta^n = 0$$

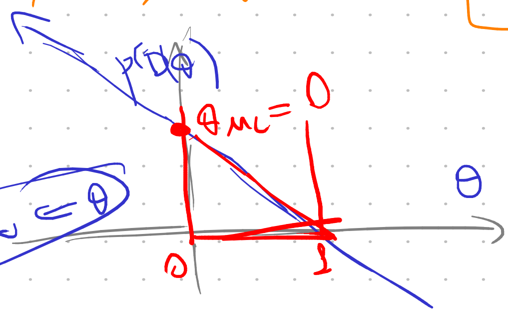


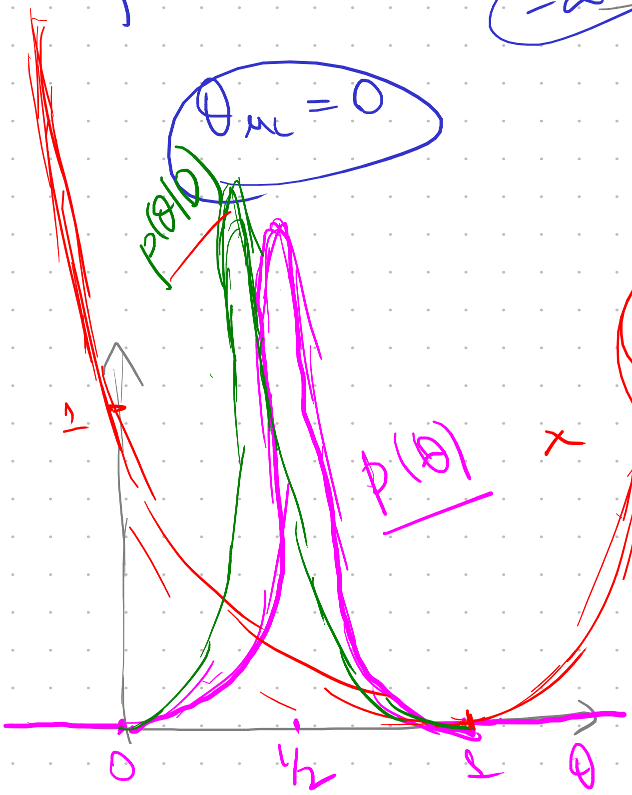
$$\theta^{n-1}(1-\theta)^{m-1}\left(n(1-\theta) - m \cdot \theta\right) = 0$$

$$\theta = \cancel{0},\ \cancel{1},\ \frac{n}{n+m} \qquad \boxed{\theta_{ML} = \frac{n}{n+m}}$$

$D = t$

$p(D|\theta) = 1-\theta \qquad -\infty < \theta$

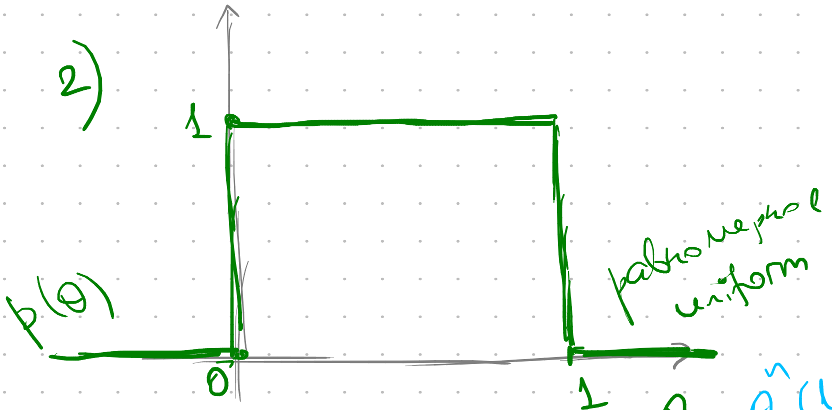




$D = tt$

$p(D|\theta) = (1-\theta)^2$

2)



$$p(\theta)=\begin{cases}1, & \theta\in[0,1]\\ 0, & \theta\notin[0,1]\end{cases}$$

$\theta^n(1-\theta)^m$

$$p(\theta|D)=\frac{p(\theta)\,p(D|\theta)}{p(D)}=\begin{cases}\frac{1}{p(D)}\theta^n(1-\theta)^m, & \theta\in[0,1]\\ 0, & —''—\end{cases}$$

$\theta \to \max$

$$\theta_{MAP}=\frac{n}{n+m}$$

$$p(D)=\int p(\theta)\,p(D|\theta)\,d\theta=\int_0^1\theta^n(1-\theta)^m\,d\theta$$

Бета – функция

$$B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$\Gamma(n)=(n-1)!$$

$$\int_0^1\theta^n(1-\theta)^m\,d\theta=\frac{\Gamma(n+1)\Gamma(m+1)}{\Gamma(n+m+2)}=\frac{n!\,m!}{(n+m+1)!}$$

$$p(\theta|D)=\begin{cases}\frac{(n+m+1)!}{n!\,m!}\cdot\theta^n(1-\theta)^m, & \theta\in[0,1]\\ 0, & —''—\end{cases}$$

3) Predictive distribution

$$p(\text{heads}\mid D) = \int \overbrace{p(\text{heads}\mid\theta)}^{\theta}\ \underbrace{p(\theta\mid D)}_{\text{posterior}}\, d\theta =$$

$$= \int_0^1 \theta \cdot \frac{(n+m+1)!}{n!\,m!}\,\theta^n (1-\theta)^m \, d\theta =$$

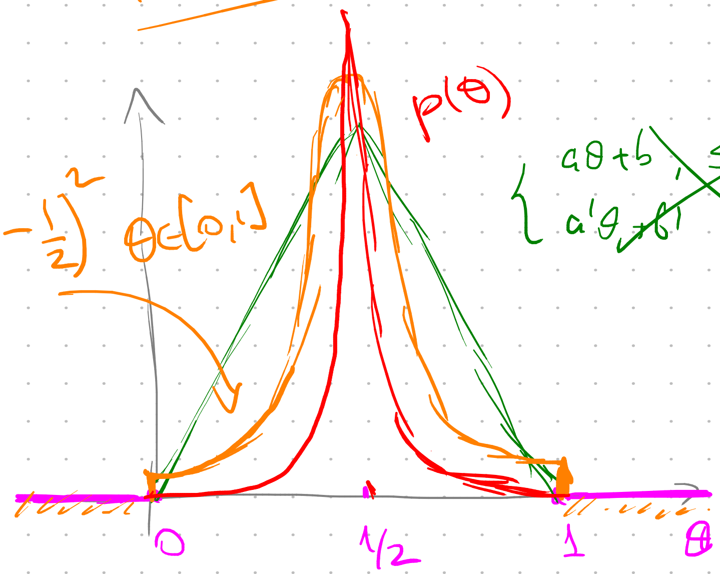
$$= \frac{(n+m+1)!}{n!\,m!} \int_0^1 \theta^{n+1} (1-\theta)^m \, d\theta = \frac{\cancel{(n+m+1)!}}{\cancel{n!}\,\cancel{m!}} \cdot \frac{(n+1)!\ \cancel{m!}}{(n+m+2)!}$$

$$\boxed{p(\text{heads}\mid D) = \frac{n+1}{n+m+2}}$$

– Laplace's Rule of Succession

Bayesian smoothing

③ Априорные распределения



$$\left\{\begin{array}{ll} \cancel{a\theta+b}, & \cancel{\le 1/2} \\ \cancel{a'\theta+b'}, & \cancel{\ge 1/2} \end{array}\right.$$

$$\mathcal{N}(\theta\mid \tfrac{1}{2}, \sigma^2) = \begin{cases} \frac{1}{Z}\cdot e^{-\frac{1}{2\sigma^2}(\theta-\frac{1}{2})^2}, & \theta\in[0,1] \\ 0, & \text{—"—} \end{cases}$$

$$'' \quad p(\theta)$$

$$p(\theta) \times p(D\mid\theta) = \begin{cases} \frac{1}{Z}\cdot e^{-\frac{1}{2\sigma^2}(\theta-1/2)^2}\cdot\theta^n(1-\theta)^m, & [0,1] \\ 0 & \text{—"—} \end{cases}$$

$$\underbrace{\theta^{a}(1-\theta)^{b}}_{} \qquad\qquad\qquad\qquad \theta^{a+n}(1-\theta)^{b+m}$$

$$p(\theta) \times p(D\mid\theta) \propto p(\theta\mid D)$$

$$? \quad \times \quad \theta^n(1-\theta)^m \quad \propto \quad ?$$

Conjugate priors (сем-во сопр. апр. распр.) - ~~770~~

$p(\bar\theta|\bar\alpha)$, $\bar\alpha$ - hyperparameters;

$$p(\bar\theta|\bar\alpha)\cdot p(D|\bar\theta) \propto p(\bar\theta|\bar\alpha')$$

$$p(\theta) = Beta(\theta|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1},\quad \theta\in[0,1]$$

$D = \{n, m\}$

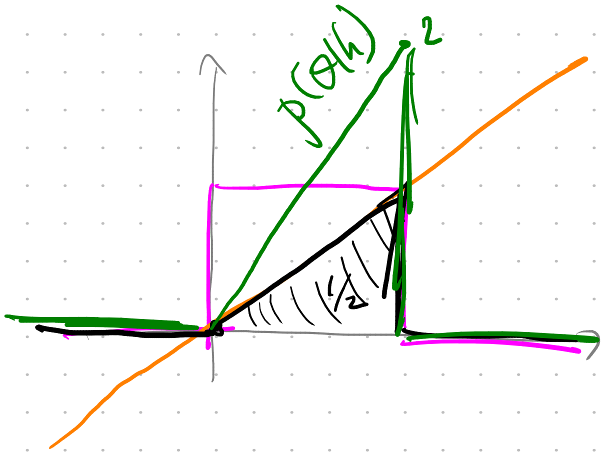
$$p(\theta|D) \propto p(\theta)\,p(D|\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\cdot\theta^{n}(1-\theta)^{m} =$$

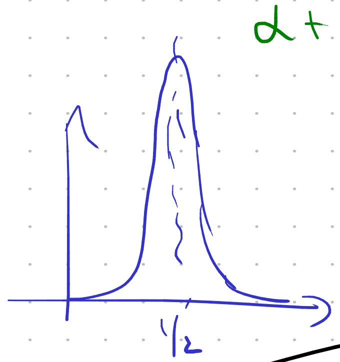
$$= \theta^{(\alpha+n)-1}(1-\theta)^{(\beta+m)-1},\quad \theta\in[0,1]$$

$$p(\theta|D) = Beta(\theta|\alpha+n,\beta+m)$$

$D' = \{n'\times h,\ m'\times t\}$

$$p(\theta|D)\cdot p(D'|\theta) \propto p(\theta|D,D') = Beta(\theta|\alpha+n+n',\beta+m+m')$$





equivalent sample size

$\alpha = 5t,\ \beta = 5T$

$$p(\theta|\alpha,\beta) = \frac{1}{Beta(\alpha,\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta|\tfrac{1}{2},\tfrac{1}{2}) = \frac{1}{Beta(\frac{1}{2},\frac{1}{2})}\cdot\frac{1}{\sqrt{\theta(1-\theta)}}$$

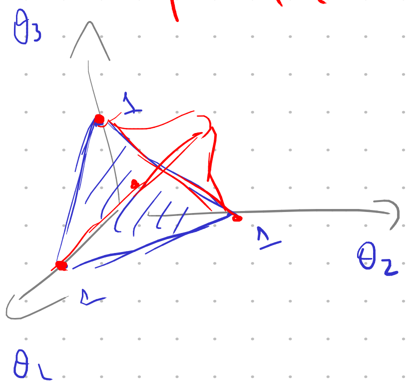
K: $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_K)$, $\sum \theta_k = 1$

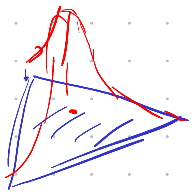
$\theta_K = 1 - \sum \theta_k$

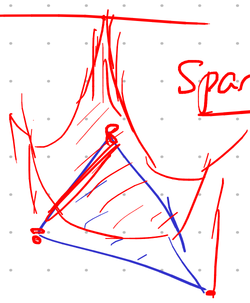
$D = \{n_1, n_2, \dots, n_K\}$

$$p(D|\bar{\theta}) = \theta_1^{n_1} \theta_2^{n_2} \dots \theta_K^{n_K}$$

$$p(\bar{\theta}|\alpha_1 \dots \alpha_K) = \frac{1}{Dir(\bar{\alpha})} \cdot \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \dots \theta_K^{\alpha_K - 1}$$







Sparsity