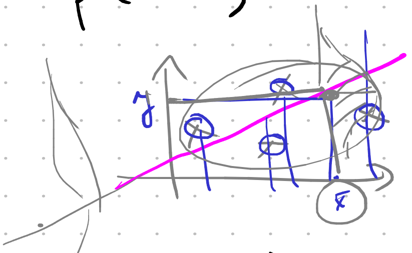
①

$$\overbrace{p(\underset{\text{params}}{\bar\theta} \mid \overset{\text{data}}{D})}^{\text{posterior distr.}} = \frac{\overbrace{p(\bar\theta)}^{\text{prior distr.}}\;\overbrace{p(D|\bar\theta)}^{\text{likelihood}}}{p(D)} \propto p(\bar\theta)\,p(D|\bar\theta)$$

(1) $p(D|\bar\theta) \xrightarrow{\bar\theta} \max$, $\bar\theta_{ML}$ – maximum likelihood

(2) $p(\bar\theta|D) \xrightarrow{\bar\theta} \max$, $\bar\theta_{MAP}$ – maximum a posteriori



(3) $p(\bar x|D)$ – predictive distribution

$\underline{p(y|\bar x, D)}$

$$p(\bar x|D) = \int p(\bar x, \bar\theta|D)\,d\bar\theta = \int \overbrace{p(\bar x|\boxed{\bar\theta}, D)}^{\text{likelihood}}\;\overbrace{p(\bar\theta|D)}^{\text{posterior}}\,d\bar\theta =$$

" $p(\bar x|\bar\theta)$

$$\underline{p(D|\bar\theta)} = \underline{\prod_{n=1}^{N} p(d_n|\bar\theta)}$$

$$= \underline{E_{p(\bar\theta|D)}[p(\bar x|\bar\theta)]}$$

$$\underline{p(y|\bar x, D)} = \int p(y|\bar x, \bar\theta)\,p(\bar\theta|D)\,d\bar\theta$$

② Bernoulli trials

$D = hhtth$, $\underline{\theta = p(heads)}$

$\underline{p(D|\theta) = \theta^m(1-\theta)^n}$, $m = \#$ heads, $n = \#$ tails

$\xrightarrow[\theta]{} \max$

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

$$\theta^{m-1}(1-\theta)^{n-1}\big(\underbrace{m(1-\theta) - n\theta}_{m-(m+n)\theta}\big) = 0$$

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

$$\theta = 0, 1, \frac{m}{m+n}$$

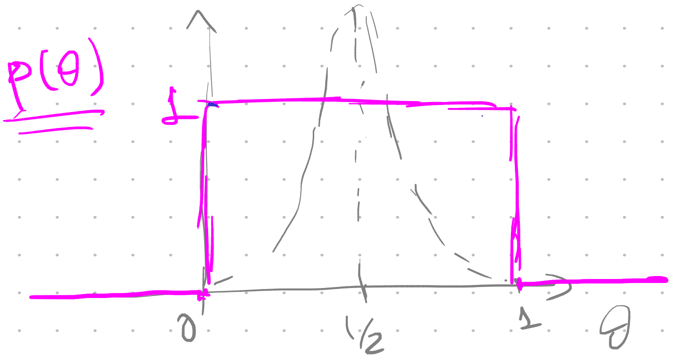
$D = t$, $\underline{\theta_{ML} = 0}$

$$\int_0^1 (1-\theta)\,p(\theta|x)\,d\theta$$

$p(D|\theta) = 1-\theta \to \max$, $\theta \in [0,1]$

$D = hhh$

$\theta_{ML} = \infty$




$$p(\theta) = \begin{cases} 1, & [0,1] \\ 0, & \text{—"—} \end{cases}$$

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

$p(y|x, D)$

$(w_0, w_1)$

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

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

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

$$p(\text{орёл}|D) = \int p(\text{орёл}|\theta)\, p(\theta|D)\, d\theta =$$

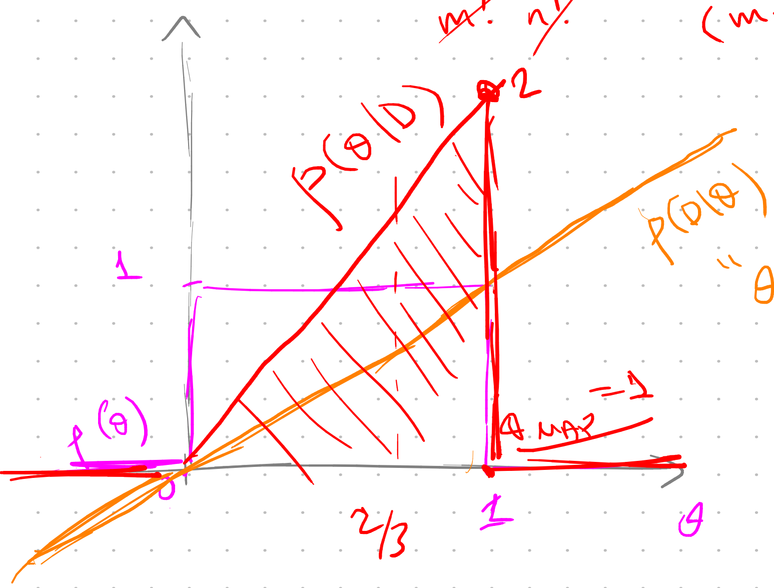
$$= \int_0^1 \theta \cdot \frac{(m+n+1)!}{m!\,n!}\theta^m(1-\theta)^n d\theta = \frac{(m+n+1)!}{m!\,n!}\int_0^1 \theta^{m+1}(1-\theta)^n d\theta =$$

$$= \frac{\cancel{(m+n+1)!}}{\cancel{m!}\,\cancel{n!}} \cdot \frac{(m+1)\cancel{!}\,\cancel{n!}}{(m+n+2)\cancel{!}} = \boxed{\frac{m+1}{m+n+2}}$$

правило Лапласа

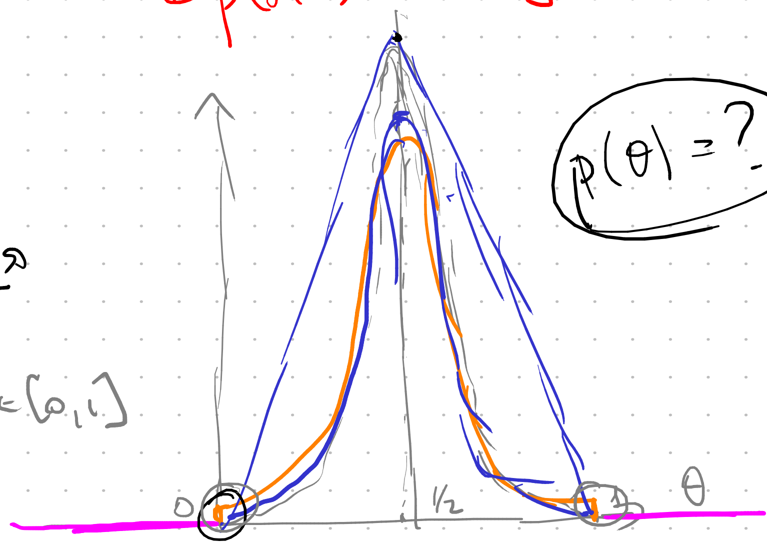
Laplace's Rule of Succession

Bayesian smoothing




$$E_{p(\theta|D)}[\theta]$$

$p(\theta) = ?$




## ③ Априорные распределения

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

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

$$e^{-\frac{1}{2\sigma^2}(\theta-\frac{1}{2})^2} \times \underbrace{\theta^m(1-\theta)^n} \propto \theta^m(1-\theta)^n \cdot e^{\cdot(\,)^2}$$

$$\theta^{\alpha-1}(1-\theta)^{\beta-1} \times \theta^m(1-\theta)^n \propto \theta^{m+\alpha-1}(1-\theta)^{n+\beta-1}$$

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

Conjugate priors — семейство распр. $p(\bar\theta \mid \bar\alpha)$, т.ч.:

для $p(D|\bar\theta)$

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

hyperparameter

$$Beta(\theta|\alpha,\beta) \times p\left(\begin{matrix} m \times heads \\ n \times tails\end{matrix} \middle| \theta\right) \propto Beta\left(\theta \middle| \begin{matrix}\alpha+m, \\ \beta+n\end{matrix}\right)$$

$$p(\bar\theta|\bar\alpha_0) \times p(\underset{\sim}{D}|\bar\theta) \propto p(\bar\theta|\bar\alpha_1) \times p(D'|\bar\theta) \propto p(\bar\theta|\bar\alpha_2)$$

$$p(D|\bar\theta) = \prod_{n=1}^{N} p(d_n|\bar\theta) \qquad \theta^{10}(1-\theta)^{10}$$

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

$$\boxed{p(\theta) = Beta(\theta \mid \alpha, \beta)} \iff$$ posterior w/ unif prior $+ \; D = (\alpha-1, \beta-1)$

equivalent sample size

$$Unif([0,1]) \times \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

"$p(\theta)$"

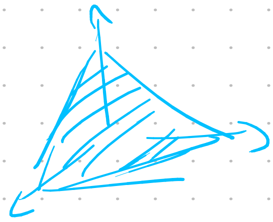
$$p(\theta) = \text{Beta}(\theta \mid \tfrac{1}{2}, \tfrac{1}{2}) = \theta^{-\frac{1}{2}}(1-\theta)^{-1/2} = \frac{1}{\sqrt{\theta(1-\theta)}}$$

## (4) Multinomial distribution

$$p(D \mid \bar{\theta}) = p(n_1, n_2, \dots n_k \mid \bar{\theta}) = \theta_1^{n_1}\, \theta_2^{n_2} \dots \underbrace{\theta_k^{n_k}}_{}$$

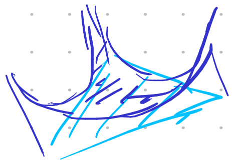
$$= 1-\theta_1-\theta_2-\dots-\theta_{k-1}$$

$$p(\bar{\theta}) = \text{Dir}(\bar{\theta} \mid \bar{\alpha}) = \frac{1}{\text{Dir}(\bar{\alpha})}\, \theta_1^{\alpha_1-1}\, \theta_2^{\alpha_2-1} \dots \theta_k^{\alpha_k-1}$$



$$p(\bar{\theta}) = \text{Dir}\left(\bar{\theta} \mid \bar{\alpha} = \left(\tfrac{1}{10}, \tfrac{1}{10}, \dots, \tfrac{1}{10}\right)\right)$$

Sparsity



LDA

SOK!

$\text{Dir}(\bar{\theta} \mid \bar{\alpha})$

$\bar{d}$ $\bar{t}$

$$p(d \mid t) = \frac{\overbrace{p(d)\,p(t \mid d)}^{\approx\, p(d,t)}}{\underbrace{p(d)\,p(t \mid d) + p(\bar{d})\,p(t \mid \bar{d})}_{p(t)}} \approx 1/6$$

$\sim 1/100$

$\sim 1/100$

$\sim 5/100$