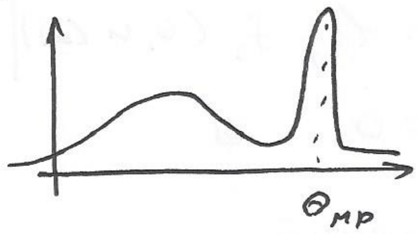
09.09.16 было (II)

$$p(\theta|X) = \frac{p(X|\theta)\cdot p(\theta)}{\int p(X|\theta)p(\theta)d\theta} = \frac{p(X,\theta)}{\int p(X,\theta)d\theta}$$

1) $\theta_{MP} = \arg\max_{\theta} p(\theta|X)$ точечная оценка

$\arg\min_{\hat\theta} l(\theta,\hat\theta)$

1) $l(\theta,\hat\theta) = \|\theta-\hat\theta\|^2$
2) $l(\theta,\hat\theta) = |\theta-\hat\theta|$
3) $l(\theta,\hat\theta) = I[\theta\neq\hat\theta]$



плохо

Сопряжённые распределения вероятностей

① $p(x|\theta) \sim \mathcal{A}(\theta)$

$p(\theta) \sim \mathcal{B}(\beta)$ сопряжённые $p(\theta|x) \sim \mathcal{B}(\beta')$

conjugate distribution

Тогда интеграл $\int p(x|\theta)p(\theta)d\theta$ берётся аналитически

① $p(X|\mu) \sim N(x|\mu,1)$

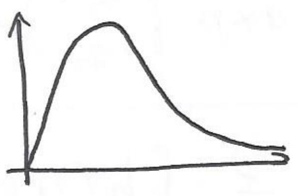
$$p(x|\mu) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}+x\mu-\frac{\mu^2}{2}\right)$$

$p(\mu) \simeq \exp(-a\mu^2+b\mu+c)$, $a>0$

$p(\mu) \sim N(\mu|m,s^2)$, $p(\mu|x)$?

② $p(X|\beta^{-1}) \sim N(x|\mu_0,\beta^{-1})$

$$p(x|\beta^{-1}) = \sqrt{\frac{\beta}{2\pi}}\exp\left(-\frac{x^2}{2}\beta\right)$$



$$p(\beta^{-1}) \simeq \exp(-\beta a)\cdot\beta^{c}\frac{1}{Z(a,c)}$$

$a>0$, $c>-1$, $\beta>0$

$$p(\beta^{-1}) \sim \Gamma(\beta|a,b) = \frac{b^a}{\Gamma(a)}\cdot\beta^{a-1}\exp(-b\beta)$$

$a,b\geqslant 0$, $\beta>0$

$p(\beta|x) \sim \Gamma(\beta|a',b')$

③ $p(x|\mu,\beta^{-1}) \sim N(x|\mu,\beta^{-1})$, $p(\mu,\beta)$?

$$p(x|\mu,\beta^{-1}) = \sqrt{\frac{\beta}{2\pi}}\exp\left(-\beta\frac{x^2}{2}+\beta\mu x-\beta\frac{\mu^2}{2}\right)$$ стр 1

$p(x|\mu,\beta^{-1})$; $p(\mu,\beta^{-1}) \neq p(\mu)\cdot p(\beta^{-1}) \Rightarrow p(\mu,\beta|x) \neq f(\mu)g(\beta)$

$p(\mu, \beta) \sim NG(\mu, \beta \mid m, \lambda, a, b) =$

$= N(\mu \mid m, (\lambda\beta)^{-1})\, \Gamma(\beta \mid a, b) = p(\mu \mid \vec{\beta})\, p(\vec{\beta})$

Т Критерий факторизации (Фишер)

$u(x)$ – достаточная статистика $\Longleftrightarrow$

$\exists\, p(x \mid \theta) = f_1(x)\, f_2(\theta)\, f_3(\theta, u(x))$

□ $\log p(x \mid \theta) = \log f_1(x) + \log f_2(\theta) + \log f_3(\theta, u(x)) \Big| \frac{\partial}{\partial \theta}, = 0$

$\frac{\partial}{\partial \theta} \log f_2(\theta) + \frac{\partial}{\partial \theta} \log f_3(\theta, u(x)) = 0$ ■

Экспоненциальный класс распределений.
Достаточные статистики.

О Параметрическое семейство $p(x \mid \theta)$ принадлежит экспоненциальному классу распределений, если

$$p(x \mid \theta) = \frac{h(x)}{g(\theta)} \exp(\theta^T u(x))$$

$h(x) \geq 0$, $u(x)$ – $\forall$ ф-ия, $g(\theta) = \int h(x) \exp(\theta^T u(x))\, dx$

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right)$$

$\theta_1 = -\frac{1}{2\sigma^2}$, $u_1(x) = x^2$, $h(x) = 1$,

$\theta_2 = \frac{\mu}{\sigma^2}$, $u_2(x) = x$, $g^{-1}(\theta) = \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma}$,

$\sigma^2 = -\frac{1}{2\theta_1}$, $\mu = \theta_2 \sigma^2 = -\frac{\theta_2}{2\theta_1}$, $\frac{\mu}{2\sigma^2} = \frac{\theta_2}{2}$

$g^{-1}(\theta) = \exp\left(\frac{\theta_2^2}{4\theta_1}\right) \cdot \frac{\sqrt{-2\theta_1}}{\sqrt{2\pi}}$; θ_1, θ_2 – естественные параметры

$$\frac{\partial g(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \int h(x) \exp(\theta^T u(x))\, dx = \int h(x)\, \theta^T u_j(x) \cdot \exp(\theta^T u(x))\, dx =$$

$$= g(\theta) \int \frac{h(x)\, u_j(x) \exp(\theta^T u(x))}{g(\theta)}\, dx = g(\theta) \int \frac{h(x) \exp(\theta^T u(x))}{g(\theta)} u_j(x)\, dx =$$

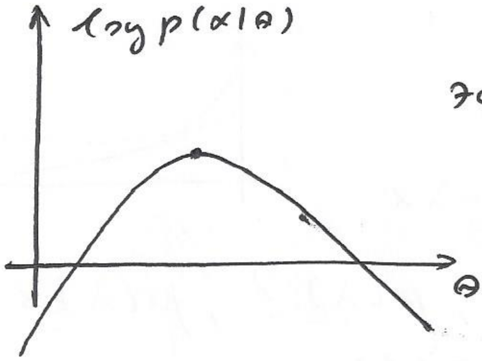
$= g(\theta)\, \mathbb{E}\, u_j(x)$, $\mathbb{E}\, u_j(x) = \frac{1}{g(\theta)} \frac{\partial g(\theta)}{\partial \theta_j}$

$$\mathbb{E}\, u_j(x) = \frac{\partial}{\partial \theta_j} \log g(\theta)$$

аналогично $\operatorname{cov}(u_j(x), u_i(x)) = \frac{\partial^2 \log g(\theta)}{\partial \theta_j \partial \theta_i}$

$\log g(\theta)$ – выпуклая ф-ия

$$\underbrace{\log p(x|\theta)}_{\text{вогнутая ф-ия}} \Leftarrow \underbrace{\log h(x)}_{\text{const}} - \underbrace{\log g(\theta)}_{\text{вогнутая}} + \underbrace{\theta^T u(x)}_{\text{линейная}}$$



эффективная численная оптимизация

$$X = (x_1 \dots x_n) \sim p(x|\theta)$$

$$\log p(X|\theta) = \sum_{i=1}^{n} \log p(x_i|\theta) = \sum_{i=1}^{n} \log h(x_i) - n \log g(\theta) + \theta^T \sum_{i=1}^{n} u_{i}(x_i) \to \max_\theta$$

$$-n \frac{\partial}{\partial \theta_j} \log g(\theta) + \sum_{i=1}^{n} u_j(x_i) = 0$$

$$\underbrace{\frac{\partial}{\partial \theta_j} \log g(\theta)}_{\mathbb{E} u_j(x)} = \frac{1}{n} \sum_{i=1}^{n} u_j(x_i) \approx \mathbb{E} u_j$$

$$\boxed{\underset{model}{\mathbb{E} u_j(x)} = \underset{data}{\mathbb{E} u_j(x)}}$$

сопряжённое распределение

$$p(x|\theta) = \frac{h(x)}{g(\theta)} \exp(\theta^T u(x))$$

$$p(\theta|\eta, \nu) = \exp(\theta^T \eta) \frac{1}{g^\nu(\theta)} \frac{1}{Z(\eta, \nu)}$$

$$p(\theta|\eta, \nu, X) \propto p(X|\theta) \cdot p(\theta|\eta, \nu) = \prod_{i=1}^{n} h(x_i) \cdot \frac{1}{g^n(\theta)} \cdot \exp\left(\theta^T \left(\sum_{i=1}^{n} u(x_i)\right)\right) \cdot \exp(\theta^T \eta) \cdot \frac{1}{g^\nu(\theta)} \cdot \frac{1}{Z(\eta, \nu)} =$$

$$= \exp\left(\theta^T \left(\sum_{i=1}^{n} u(x_i) + \eta\right)\right) \frac{1}{g^{n+\nu}(\theta)} \frac{\prod_{i=1}^{n} h(x_i)}{Z(\eta, \nu)}$$

$$\eta' = \eta + \sum_{i=1}^{n} u(x_i)$$

$$\nu' = \nu + n$$

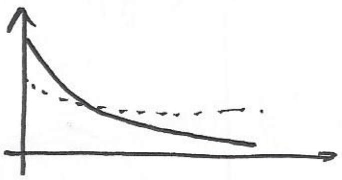
$$p(\theta|X, \eta, \nu) = p(\theta|\eta', \nu')$$

$p(\theta \mid X, \eta, \nu) = p(\theta \mid \eta', \nu')$ 09.09.16 бммо семинар

сопряжённые распределения

экспоненциальные семейства.

$$p(\theta \mid x) = \frac{p(x \mid \theta)\, p(\theta)}{\int p(x \mid \theta)\, p(\theta)\, d\theta}$$



① $x \sim \exp(\lambda)$, $p(x) = \lambda e^{-\lambda x}$

$x_1 \dots x_n$ – выборка, $\lambda_{МП}$?, $p(\lambda)$?, $p(\lambda \mid x)$?, $\mathbb{E}\, p(\lambda \mid x)$?

$$\mathcal{L}(x, \lambda) = \prod_{i=1}^{n} p(x_i \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$\log \mathcal{L}(x, \lambda) = \sum_{i=1}^{n} (\log \lambda - \lambda x_i)$$

$$\frac{\partial \log \mathcal{L}(x, \lambda)}{\partial \lambda} = -\sum_{i=1}^{n} x_i + n\lambda = 0, \quad \lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$p(\lambda) = G(\lambda \mid a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$p(x \mid \lambda) \cdot p(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} = \cancel{\lambda^{a-1} e^{-\lambda(x+a)} \frac{b^a}{n\Gamma(a)}}$$

$$= \frac{b^a}{\Gamma(a)} \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i} \cdot \lambda^{a-1} \cdot e^{-b\lambda} = \frac{b^a}{\Gamma(a)} \lambda^{n+a-1} e^{-\lambda\left(\sum_{i=1}^{n} x_i + b\right)} =$$

$$= G\left(\lambda \mid n+a, \sum_{i=1}^{n} x_i + b\right) = G(\lambda \mid a', b') = p(\lambda \mid x)$$

$$a' = n + a, \quad b' = b + \sum_{i=1}^{n} x_i$$

$$\mathbb{E}\, p(\lambda \mid x) = \frac{n+a}{\sum_{i=1}^{n} x_i + b} = \frac{a}{b} \frac{b}{\sum x_i + b} + \frac{n}{\sum x_i} \frac{\sum x_i}{\sum x_i + b} =$$

$$= \lambda_{pr} w + \lambda_{МП}(1 - w)$$

② $x \sim Bern(q)$

X	0	1
p(x)	$1-q$	q

$p(x) =$

$x_1 \dots x_n \sim Bern(q)$

1) $p(X \mid q) = \prod_{i=1}^{n} q^{\sum x_i} (1-q)^{N - \sum x_i}$

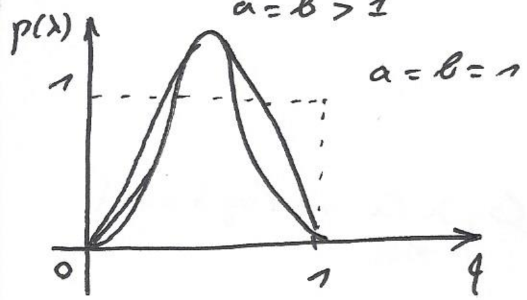
2) $p(k \mid N, q) = C_N^k q^k (1-q)^{N-k}$

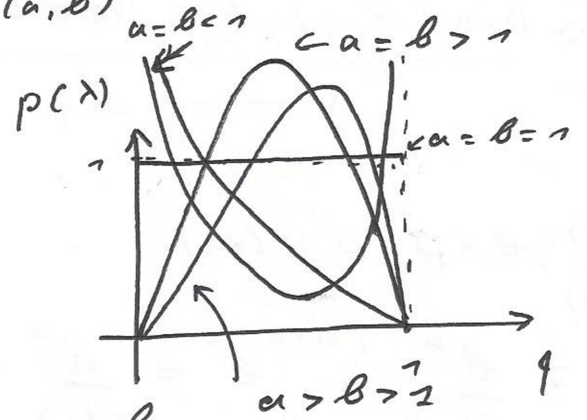
3) $p(k \mid q, r) = C_{k+r-1}^{k} q^k (1-q)^r$

Сопряжённый prior ?

$$p(q) = q^{a-1}(1-q)^{b-1} \cdot \frac{1}{B(a,b)}, \quad a>0,\ b>0$$

β-распределение





$$\mathbb{E}p = \frac{a}{a+b}, \quad \mathbb{D}p = \frac{ab}{(a+b)^2(a+b+1)}$$

$$\frac{1}{B(a,b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \quad C_N^k = \frac{N!}{k!(N-k)!}$$

$$p(q|x) = Beta(q|a', b')$$

$$q^{k+a-1}(1-q)^{b+N-k-1}, \quad a' = k+a,\ b' = b+N-k$$

$$\mathbb{E}p = \frac{k+a}{b+N-k}$$

$$\begin{cases} \dfrac{a}{a+b} = \mu \\ \dfrac{ab}{(a+b)^2(a+b+1)} = \sigma \end{cases}$$

$$\frac{a}{\mu} = a+b, \quad b = \frac{a}{\mu} - a$$

$$\frac{a^2(\frac{1}{\mu}-1)}{(a+a(\frac{1}{\mu}-1))(a+a(\frac{1}{\mu}-1)+1)} = \sigma$$

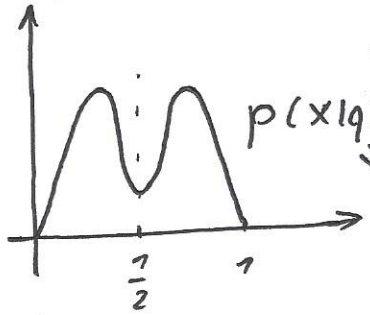
$$\frac{a\overbrace{(\frac{1}{\mu}-1)}^{\gamma}}{(1+(\frac{1}{\mu}-1))(a+a(\frac{1}{\mu}-1)+1)} = \sigma$$

$$\frac{a\gamma}{(1+\gamma)(a+a\gamma+1)} = \sigma$$

$$a\gamma = \sigma(a+a\gamma+1+a\gamma+a\gamma^2+\gamma)$$

$$a\gamma = \sigma a + 2a\gamma\sigma + \sigma + a\gamma^2\sigma + \sigma\gamma +$$

$$a(\sigma + 2\gamma\sigma + \gamma^2\sigma \cancel{\gamma} - \gamma) = -\sigma - \sigma\gamma$$



$$w_1 \, Beta(a_1, b_1) + w_2 \, Beta(a_2, b_2)$$

$$p(x|q) \approx \frac{Bin(N, q) \cdot (w_1 \, Beta(a_1, b_1) + w_2 \, Beta(a_2, b_2))}{\int -''- \, dq}$$

$$p(x|\Theta) = \frac{h(x)}{g(\Theta)} \exp(\Theta^T u(x))$$

④ $$G(\lambda | a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b} = \frac{b^a}{\Gamma(a)} e^{\cancel{\log \lambda} \, (a-1)\log\lambda - b\lambda}$$

$$\Theta = \begin{bmatrix} a-1 \\ -b \end{bmatrix}, \quad u(\lambda) = \begin{bmatrix} \log \lambda \\ \lambda \end{bmatrix}, \quad g(\Theta) = \frac{\Gamma(\Theta_1 + 1)}{\cancel{\Gamma(a)} (-\Theta_2)^{\Theta_1 + 1}}$$

$\cancel{\mathbb{E}\, u(\lambda) =}$

$$\mathbb{E} \, \cancel{\frac{\log \lambda}{g(\Theta)}} \log\lambda = \frac{\partial}{\partial \Theta_1} \left(\log \frac{\Gamma(\Theta_1 + 1)}{(-\Theta_2)^{\Theta_1 + 1}} \right) = \frac{\partial}{\partial \Theta_1} \Big(\log \Gamma(\Theta_1 + 1) -$$

$$- (\Theta_1 + 1) \log(-\Theta_2) \Big) = \psi(\Theta_1 + 1) - \log(-\Theta_2)$$

$\psi(x)$ - дигамма ф-ия, $\log(x - \frac{1}{2}) \le \psi(x) \le \log(x)$

$$\mathbb{E} \lambda = \frac{\partial}{\partial \Theta_2} \Big(\log \Gamma(\Theta_1 + 1) - (\Theta_1 + 1) \log(-\Theta_2) \Big) = \mp \frac{\Theta_1 + \cancel{\Theta_2}1}{\Theta_2} = \frac{a}{b}$$

⑤ $$p(x|q) = C_N^x q^x (1-q)^{N-x} = \cancel{\frac{N!}{x!(N-x)!}} C_N^x e^{x \log q + (N-x)\log(1-q)} =$$

$$= C_N^x \cdot e^{N \log(1-q)} \cdot e^{x \log q - x \log(1-q)} = C_N^x (1-q)^N e^{x \log \frac{q}{1-q}}$$

$$\Theta = \log \frac{q}{1-q}, \quad u(x) = x, \quad h(x) = C_N^x$$

$$e^\Theta = \frac{q}{1-q}, \quad (1-q) e^\Theta = q, \quad q(1 + e^\Theta) = 1, \quad q = \frac{1}{1 + e^\Theta}$$

$$g^{-1}(\Theta) = \left(1 - \frac{1}{1 + e^\Theta}\right)^N = \left(\frac{e^\Theta}{1 + e^\Theta}\right)^N$$