① Exponential family

sufficient statistics

$$p(\bar{x}|\bar{\theta}) = h(\bar{x})\, g(\bar{\theta})\, e^{\bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x})} = h(\bar{x})\, e^{\bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x}) - a(\bar{\theta})}$$

cumulant

$$\log p(\bar{x}|\bar{\theta}) = \log h(\bar{x}) + \bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x}) - a(\bar{\theta})$$

если $\bar{\eta}(\bar{\theta}) = \bar{\theta}$, то $\bar{\theta}$ - natural parameter

$$\mathrm{Binom}(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k} =$$

$$= \binom{n}{k} e^{k\log p + (n-k)\log(1-p)} = \underbrace{\binom{n}{k}}_{h(k)} e^{\overset{t(k)}{k}\log\frac{p}{1-p} + \overbrace{n\log(1-p)}^{-a(p)}}$$

$$\boxed{\theta = \log\frac{p}{1-p}}$$

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2 + x\cdot\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{\overbrace{\begin{pmatrix} x^2 \\ x \end{pmatrix}}^{\bar{t}(x)}{}^T \begin{pmatrix} -\frac{1}{2\sigma^2} \\ \mu/\sigma^2 \end{pmatrix} - \underbrace{\left(\frac{\mu^2}{2\sigma^2} + \log\sigma\right)}_{a(\bar{\theta})}}$$

$$\bar{\theta} = \begin{pmatrix} -\frac{\tau}{2} \\ \mu\cdot\tau \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$$\mathcal{N}(\bar{x}|\bar{\mu},\Sigma) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\left(\bar{x}^T\Sigma^{-1}\bar{x} - 2\bar{x}^T\Sigma^{-1}\bar{\mu} + \overbrace{\bar{\mu}^T\Sigma^{-1}\bar{\mu} + \log\det\Sigma}\right)}$$

$$\sum_{ij}(\Sigma^{-1})_{ij} x_i x_j = \mathrm{vec}(\bar{x}\bar{x}^T)^T \mathrm{vec}(\Sigma^{-1})$$

$$\bar{t}(\bar{x}) = \begin{pmatrix} \mathrm{vec}(\bar{x}\bar{x}^T) \\ \bar{x} \end{pmatrix}, \quad \bar{\theta}(\bar{\mu},\Sigma) = \begin{pmatrix} -\frac{1}{2}\mathrm{vec}(\Sigma^{-1}) \\ \Sigma^{-1}\bar{\mu} \end{pmatrix}$$

② Moments

natural param

$$\int p(\bar{x}|\bar{\theta})\,d\bar{x} = 1 \qquad \int e^{-a(\bar{\theta})} h(\bar{x})\, e^{\bar{\theta}^T \bar{t}(\bar{x})}\, d\bar{x} = 1$$

$$a(\bar{\theta}) = \log \int h(\bar{x}) e^{\bar{\theta}^T \bar{t}(\bar{x})}\, d\bar{x} \qquad / \ \nabla_{\bar{\theta}}$$

$$\nabla_{\bar{\theta}}\, a(\bar{\theta}) = \frac{\int h(\bar{x})\, \nabla_{\bar{\theta}}\, e^{\bar{\theta}^T \bar{t}(\bar{x})}\, d\bar{x}}{\int h(\bar{x}) e^{\bar{\theta}^T \bar{t}(\bar{x})}\, d\bar{x}} = \frac{\int h(\bar{x}) e^{\bar{\theta}^T \bar{t}(\bar{x})} \cdot \bar{t}(\bar{x})\, d\bar{x}}{e^{a(\bar{\theta})}} =$$

$$= \int \bar{t}(\bar{x}) \cdot \Big(\underbrace{h(\bar{x})\, e^{\bar{\theta}^T \bar{t}(\bar{x}) - a(\bar{\theta})}}_{p(\bar{x}|\bar{\theta})} \Big)\, d\bar{x}$$

$$\boxed{\mathbb{E}_{p(\bar{x}|\bar{\theta})}[\bar{t}(\bar{x})] = \nabla_{\bar{\theta}}\, a(\bar{\theta})}$$

Параметры средних

$$\bar{\mu} = \nabla_{\bar{\theta}}(a(\bar{\theta}))$$

$$\frac{\partial^2 a}{\partial\theta_i \partial\theta_j} = \frac{\partial\, \mathbb{E}[\bar{t}(\bar{x})]_i}{\partial\theta_j} = \int t_i(\bar{x})\, h(\bar{x})\, \frac{\partial}{\partial\theta_j} e^{\bar{\theta}^T \bar{t}(\bar{x}) - a(\bar{\theta})}\, d\bar{x}$$

$$= \int t_i(\bar{x})\, h(\bar{x})\, e^{\bar{\theta}^T \bar{t}(\bar{x}) - a(\bar{\theta})} \cdot \Big(t_j(\bar{x}) - \frac{\partial a}{\partial\theta_j}\Big)\, d\bar{x} =$$

$$= \int t_i(\bar{x}) \Big(t_j(\bar{x}) - \overbrace{\frac{\partial a}{\partial\theta_j}(\bar{\theta})}^{\text{"}\mathbb{E}[t_j(\bar{x})]} \Big) \cdot p(\bar{x}|\bar{\theta})\, d\bar{x}$$

$$= \mathbb{E}[t_i(\bar{x})\, t_j(\bar{x})] - \mathbb{E}[t_i(\bar{x})] \cdot \mathbb{E}[t_j(\bar{x})]$$

$$\boxed{H(a(\bar{\theta})) = \mathrm{Var}[\bar{t}(\bar{x})]}$$

3 Bayesian inference

$$p(\bar{x}_1, \dots, \bar{x}_N | \bar{\theta}) = \prod_n h(\bar{x}_n) e^{\bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x}_n) - a(\bar{\theta})}$$

$$\log p(D|\bar{\theta}) = \underbrace{\sum_n \log h(\bar{x}_n)}_{\not\ni \bar{\theta}} + \underbrace{\bar{\eta}(\bar{\theta})^T \left(\sum_n \bar{t}(\bar{x}_n)\right) - N \cdot a(\bar{\theta})}_{\bar{\theta} \to \max}$$

max

see max. likelihood

suff. stats $\underbrace{\sum_n \bar{t}(\bar{x}_n)}$

Hemp.: 6 $\mathcal{N}(x|\mu, \sigma^2)$ $\quad \bar{t}(x) = \begin{pmatrix} x^2 \\ x \end{pmatrix} \Rightarrow \begin{pmatrix} \sum_n x_n^2 \\ \sum_n x_n \end{pmatrix}$

Pitman-Koopman-Darmois theorem

$$\underline{\nabla_{\bar{\theta}} \bar{\eta}(\bar{\theta})^T} \sum_n \bar{t}(\bar{x}_n) - N \cdot \nabla_{\bar{\theta}} a(\bar{\theta}) = 0$$

natural param.: $\boxed{\nabla_{\bar{\theta}} a(\bar{\theta}) = \frac{1}{N} \sum_n \bar{t}(\bar{x}_n)} = \bar{\mu}$

conjugate prior:

$$p(\bar{\theta} | \bar{\chi}, \nu) = f(\bar{\chi}, \nu) \cdot e^{\bar{\chi}^T \bar{\eta}(\bar{\theta}) - \nu \cdot a(\bar{\theta})}$$

$$\log p(\bar{\theta} | \bar{x}_1 \dots \bar{x}_N, \bar{\chi}_0, \nu_0) = \log f(\bar{\chi}_0, \nu_0) + \bar{\chi}_0^T \bar{\eta}(\bar{\theta}) - \nu_0 a(\bar{\theta})$$
$$+ \sum_n \log h(\bar{x}_n) + \bar{\eta}(\bar{\theta})^T \left(\sum_n \bar{t}(\bar{x}_n)\right) - N a(\bar{\theta}) + \text{Const}$$

$$\nu_N = \nu_0 + N$$

$$\bar{\chi}_N = \bar{\chi}_0 + \sum_n \bar{t}(\bar{x}_n)$$

predictive distribution

$$p(\bar{x}|D) = \int p(\bar{x}|\bar{\theta})\,p(\bar{\theta}|D)\,d\bar{\theta} = \int p(\bar{x}|\bar{\theta})\,p(\bar{\theta}|\bar{\chi}_N, \nu_N)\,d\bar{\theta}$$

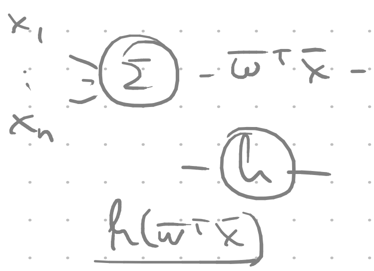
$$= \int h(\bar{x})\,e^{\bar{t}(\bar{x})^T \bar{\eta}(\bar{\theta}) - a(\bar{\theta})} \cdot f(\bar{\chi}_N, \nu_N)\,e^{\bar{\chi}_N^T \bar{\eta}(\bar{\theta}) - \nu_N a(\bar{\theta})}\,d\bar{\theta}$$

$$= f(\bar{\chi}_N, \nu_N)\,h(\bar{x}) \int e^{\bar{\eta}(\bar{\theta})^T (\bar{\chi}_N + \bar{t}(\bar{x})) - (\nu_N + 1) \cdot a(\bar{\theta})}\,d\bar{\theta}$$

$$\boxed{p(\bar{x}|D) = h(\bar{x}) \cdot \frac{f(\bar{\chi}_N, \nu_N)}{f(\bar{\chi}_N + \bar{t}(\bar{x}), \nu_N + 1)}}$$

4 Generalized linear models (GLM)

$\hat{y} = \bar{w}^T \bar{x}$ $\qquad\qquad$ $\hat{y} = \sigma(\bar{w}^T \bar{x})$



$$\boxed{\hat{y} = h(\bar{w}^T \bar{x})}$$

$h = id$ – lin. regr.

$h = \sigma$ – log. regr.

$$\boxed{\bar{w}^T \bar{x} = h^{-1}(\hat{y}) = g(\hat{y})}$$

g – link function

$p(y|\bar{x}, \bar{w})$ – среднее $\mu = h(\bar{w}^T \bar{x})$

μ – дост. статистика

$$p(y|\theta, \sigma^2) = h(y, \sigma^2)\,e^{\frac{y \cdot \theta - a(\theta)}{\sigma^2}}$$

overdispersed exponential family

dispersion parameter

$t(y) = \frac{y}{\sigma^2}$

$$\boxed{\theta = \bar{w}^T \bar{x}}$$

$$E\left[\frac{y}{\sigma^2}\right] = \frac{\partial\left(a(\theta)/\sigma^2\right)}{\partial\theta}, \quad \boxed{E[y] = \frac{\partial a}{\partial\theta}}$$

$$Var\left[\frac{y}{\sigma^2}\right] = \frac{1}{\sigma^4} Var[y] = \frac{\partial^2\left(a/\sigma^2\right)}{\partial\theta^2}, \quad \boxed{Var[y] = \sigma^2 \cdot \frac{\partial^2 a}{\partial\theta^2}}$$

$$h(\bar{w}^T\bar{x}) = \mu = \frac{\partial a}{\partial\theta} = a'(\bar{w}^T\bar{x})$$

dist. persp.

$$p(y|\mu,\sigma^2) = e^{-\frac{(y-\mu)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)} = e^{\frac{y\mu - \mu^2/2}{\sigma^2} - \left(\frac{y^2}{2\sigma^2} + \frac{1}{2}\log 2\pi\sigma^2\right)}$$

$$\mu = \bar{w}^T\bar{x}, \quad g^{-1} = id, \quad g = id, \quad a(\theta) = \frac{1}{2}\mu^2$$

$$h(y,\sigma^2) = e^{-\left(\frac{y^2}{2\sigma^2} + \frac{1}{2}\log 2\pi\sigma^2\right)}$$

dis. persp.

$$\mu = E[y|\bar{x}], \quad \mu = p(\text{heads})$$

$$p(y|\mu) = \mu^y(1-\mu)^{1-y} = e^{y\log\frac{\mu}{1-\mu} + \log(1-\mu)}$$

$$\boxed{\theta = \log\frac{\mu}{1-\mu}}, \quad a(\theta) = -\log(1-\mu)$$

$$\sigma^2 = 1 \qquad h(y,\sigma^2) = 1$$

$$\mu = g(\theta) = \frac{e^\theta}{1+e^\theta}, \quad g = \sigma^{-1}, \quad g^{-1} = \sigma$$

$$\log p(y_1 \ldots y_N | \bar{x}_1 \ldots \bar{x}_N, \bar{w}, \sigma^2) = \overbrace{\sum_n \log h(y_n,\sigma^2)}^{\text{const}} + \frac{1}{\sigma^2}\sum_n\left(y_n \cdot \bar{w}^T\bar{x}_n - a(\bar{w}^T\bar{x}_n)\right)$$

$$\nabla_{\bar{w}} \log p(D|\bar{w}, \sigma^2) = \frac{1}{\sigma^2} \sum_n \left(y_n \cdot \bar{x}_n - a'(\bar{w}^T \bar{x}_n) \cdot \bar{x}_n \right) = \underline{+ \nabla_{\bar{w}} \log p(\bar{w})}$$

$$= \frac{1}{\sigma^2} \sum_n \left(y_n - a'(\bar{w}^T \bar{x}_n) \right) \bar{x}_n =$$

$$= \frac{1}{\sigma^2} \sum_n (y_n - \mu_n) \bar{x}_n + \underline{\nabla_{\bar{w}} \log p(\bar{w})}$$

⑤ <u>Poisson regression</u>

$\mu = f(\bar{w}^T \bar{x})$

$y \sim \text{Pois}(\underline{\mu})$ $\qquad p(y|\mu) = \frac{1}{y!} \mu^y e^{-\mu}$

$$\log p(y|\mu) = y \log \mu - \mu - \log(y!)$$

$$\theta = \log \mu \quad , \quad \theta = \bar{w}^T \bar{x}$$

$$\log p(y|\bar{x}, \bar{w}) = y \cdot (\bar{w}^T \bar{x}) - e^{\bar{w}^T \bar{x}} - \log(y!)$$

$$\sigma^2 = 1, \quad \theta = \bar{w}^T \bar{x}, \quad a(\theta) = e^\theta = e^{\bar{w}^T \bar{x}}, \quad h(y, \sigma^2) = \frac{1}{y!}$$