① Exponential family

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

natural parameter — sufficient statistics

$$\mathcal{N}(\bar{x}|\bar{\mu}, \Sigma) = g(\Sigma)\cdot e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma^{-1}(\bar{x}-\bar{\mu})}$$

$$-\frac{1}{2}\left(\bar{x}^T \Sigma^{-1} \bar{x} - 2\bar{x}^T \Sigma^{-1}\bar{\mu} + \bar{\mu}^T \Sigma^{-1}\bar{\mu}\right)$$

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

②

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

cumulant

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

$\nabla_{\bar{\theta}}$

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

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

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

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

mean parametrization

$$\frac{\partial^2 a}{\partial\theta_i \partial\theta_j} = \frac{\partial \mathbb{E}_p[t_i(\bar{x})]}{\partial\theta_j} = \frac{\partial}{\partial\theta_j} \int t_i(\bar{x})\, h(\bar{x})\, e^{\cdots}\, d\bar{x} =$$

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

$$= \int t_i(\bar{x}) t_j(\bar{x}) \, p(\bar{x}|\bar{\theta}) \, d\bar{x} - \frac{\partial a}{\partial \theta_j} \int t_i(\bar{x}) \cdot p(\bar{x}|\bar{\theta}) \, d\bar{x} =$$

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

$$\boxed{H(a) = \left(\frac{\partial^2 a}{\partial \theta_i \partial \theta_j} \right)^d_{i,j=1} = \mathrm{Var}[\bar{t}(\bar{x})]}$$

③ Maximum likelihood

$$p(\bar{x}_1 \ldots \bar{x}_N | \bar{\theta}) = \prod_{n=1}^{N} h(\bar{x}_n) \, e^{\bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x}_n) - a(\bar{\theta})}$$

$$\sum_n \bar{x}_n \quad \sum_n \bar{x}_n^2$$

$$\log p(\bar{x}_1 \ldots \bar{x}_N | \bar{\theta}) = \sum_n \log h(\bar{x}_n) + \bar{\eta}(\bar{\theta})^T \left[\sum_{n=1}^{N} \bar{t}(\bar{x}_n) \right] - N \cdot a(\bar{\theta})$$

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

Pitman - Koopman - Darmois thm

$$\nabla_\theta [\;\_\_\_\;] = \bar{0}$$

$$\boxed{N \cdot \nabla_{\bar{\theta}} a(\bar{\theta}) = \sum_{n=1}^{N} \bar{t}(\bar{x}_n)^T \, \nabla_\theta \bar{\eta}(\bar{\theta})}$$

Ecau $\bar{\eta}(\bar{\theta}) = id$, so: $\underbrace{\nabla_{\bar{\theta}} a(\bar{\theta})}_{\mathbb{E}[t(\bar{x})]} = \boxed{\frac{1}{N} \cdot \sum_n \bar{t}(\bar{x}_n) = \bar{\mu}_{ML}}$

④ Conjugate prior

$$p(\bar{\theta}|\bar{\chi}) \times p(\bar{x}_1 \ldots \bar{x}_N | \bar{\theta}) \propto p(\bar{\theta}|\bar{\chi}')$$

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

$$\log p(\bar{\theta} | \bar{x}_1 \ldots \bar{x}_N, \bar{\chi}_0, \nu_0) = \mathrm{Const} + \log p(\bar{\theta} | \bar{\chi}_0, \nu_0) + \log p(\bar{x}_1 \ldots \bar{x}_N | \bar{\theta}) =$$

$$= \mathrm{const} + \bar{\chi}_0^T \bar{\eta}(\bar{\theta}) - \nu_0 \cdot a(\bar{\theta}) + \bar{\eta}(\bar{\theta})^T \sum_n \bar{t}(\bar{x}_n) - N \cdot a(\bar{\theta})$$

$$p(\bar\theta \mid \bar x_1 - \bar x_N, \bar x_0, D_0) = p(\bar\theta \mid \bar x_N, D_N), \text{ ye}$$

$$D_N = D_0 + N, \quad \bar x_N = \bar x_0 + \sum_{n=1}^{N} \bar t(\bar x_n)$$

$$\mathcal N(x \mid \mu, \tau) = \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x-\mu)^2} = \frac{1}{\sqrt{2\pi}}\, e^{\overbrace{\begin{pmatrix} x^2 \\ x \end{pmatrix}^T}^{\bar t(x)} \overbrace{\begin{pmatrix} -\tau/2 \\ \tau\mu \end{pmatrix}}^{\eta(\bar\theta)} \overbrace{- \tau \cdot \frac{\mu^2}{2} + \frac{1}{2}\log\tau}^{a(\bar\theta)}}$$

$$p(\mu, \tau \mid \bar x_0, D_0) = f(x_0, D_0) \cdot e^{\begin{pmatrix} x_{0,2} \\ x_{0,1} \end{pmatrix}^T \begin{pmatrix} -\tau/2 \\ \tau\mu \end{pmatrix} - D_0 \tau \mu^2/2 + D_0 \cdot \log\tau/2} =$$

$$-x_{0,2} \cdot \tau/2 + x_{0,1} \tau \mu - D_0 \tau \mu^2/2 + \dots$$

$$-\frac{D_0 \tau}{2}\left(\mu - \frac{x_{0,1}}{D_0}\right)^2 + \frac{\tau x_{0,1}^2}{2D_0}$$

$$= f(x_0, D_0) \cdot \underbrace{e^{-\frac{D_0\tau}{2}\left(\mu - \frac{x_{0,1}}{D_0}\right)^2}}_{\mathcal N\left(\mu \,\middle|\, \underbrace{\frac{x_{01}}{D_0}}_{\mu_0},\, D_0\tau\right)} \cdot \underbrace{e^{\frac{D_0}{2}\log\tau - \frac{\tau}{2}\left(x_{0,2} - \frac{x_{0,1}^2}{D_0}\right)}}_{\mathrm{Gam}\left(\tau \,\middle|\, \underbrace{\frac{D_0}{2}}_{\alpha_0},\, \underbrace{\frac{1}{2}\left(x_{02} - \frac{x_{01}^2}{D_0}\right)}_{\beta_0}\right)}$$

$$p(\bar x \mid \bar\theta) = \mathcal N(\bar x \mid \bar\mu, \Lambda = \Sigma^{-1})$$

Wishart distribution

$$W(\Lambda \mid V, n, d) = \text{Const} \cdot \frac{1}{(\det V)^{n/2}} \cdot (\det \Lambda)^{\frac{n-d-1}{2}} \cdot e^{-\frac{1}{2}\mathrm{Tr}(V^{-1}\Lambda)}$$

(5) Predictive distribution

$$p(\bar x \mid D) = \int p(\bar x \mid \bar\theta)\, p(\bar\theta \mid D)\, d\bar\theta = \int p(\bar x \mid \bar\theta) \cdot p(\bar\theta \mid \bar x_N, D_N)\, d\bar\theta =$$

$$= \int h(\bar x)\, e^{\bar t(\bar x)^T \eta(\bar\theta) - a(\bar\theta)} \cdot \underline{f(x_N, D_N)\, e^{\bar x_N^T \eta(\bar\theta) - D_N \cdot a(\bar\theta)}}\, d\bar\theta =$$

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

$$= 1 / f(\bar x_N + \bar t(\bar x), D_N + 1)$$

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

6 Generalized linear models (GLM)

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

$h(c) = c$ – lin. regr

$h(c) = \sigma(c)$ – log. regr.

$c = \bar{w}^T\bar{x}$, $\mu = g^{-1}(c)$, где $g(\mu) = c$, g – link function

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

$p(y|\bar{x}, \bar{w}) = \quad \mu = g^{-1}(\bar{w}^T\bar{x})$ $\quad \theta = \bar{w}^T\bar{x}$

Overdispersed exp. family:

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

$t(y) = y/\sigma^2$, $a(\theta)/\sigma^2$, $E = \frac{\partial (a(\theta)/\sigma^2)}{\partial\theta}$

$$E[y] = \frac{\partial a(\theta)}{\partial\theta}$$

$$Var[y] = \sigma^2 \cdot \frac{\partial^2 a(\theta)}{\partial\theta^2}$$

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

$\theta = \mu = \bar{w}^T\bar{x}$, $g = g^{-1} = id$, $a(\theta) = -\frac{\mu^2}{2}$, $h(y, \sigma^2) = e^{\dots}$

2) Bernoulli

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

$$\mu = \sigma(\bar{w}^T\bar{x}) = \sigma(\theta) = g^{-1}(\theta)$$

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

$$= \log\left(1 - \frac{1}{1+e^{-\theta}}\right) = \log\left(1+e^{\theta}\right)$$

3) Poisson regression

$$p(y|\mu) = \frac{1}{y!}\mu^y e^{-\mu}, \quad \log p(y|\mu) = y\underbrace{\log\mu}_{\theta} - \mu - \log(y!)$$

$$\theta = \log\mu, \qquad \theta = \bar{\omega}^T\bar{x}$$

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