$$\mathcal{M}_1, \dots, \mathcal{M}_K \qquad p(\mathcal{M}_k|D) \propto p(\mathcal{M}_k)\, p(D|\mathcal{M}_k)$$

$$p(D|\mathcal{M}_k) = \int p(D|\bar\theta_k, \mathcal{M}_k)\, p(\bar\theta_k|\mathcal{M}_k)\, d\bar\theta_k$$

$$D = \{\bar x_1 \dots \bar x_N\} \qquad D \sim \underline{p_{data}(\bar x)} \approx p(\bar x|\bar\theta)$$

$$p(\underline{D}|\bar\theta) \to \max, \quad \bar\theta_{ML} \qquad \underline{p(\bar x|\bar\theta_{ML})}$$

min

$$KL\left(p_{data} \,\|\, p(\bar x|\bar\theta_{ML})\right) = \mathbb{E}_{p_{data}}\left[\log \frac{p_{data}(\bar x)}{p(\bar x|\bar\theta_{ML})}\right] =$$

$$= \mathbb{E}_{p_{data}(\bar x)}\left[\log p_{data}(\bar x)\right] - \underbrace{\mathbb{E}_{p_{data}}\left[\log p(\bar x|\bar\theta_{ML})\right]}_{\to \max}$$

$$\mathbb{E}_{p_{data}}\left[\log p(\bar x|\bar\theta_{ML})\right] \overset{D}{\approx} \underbrace{\frac{1}{N}\sum_{n=1}^{N} \log p(\bar x_n|\bar\theta_{ML})}$$

$$p_{data}(\bar x) = p(\bar x|\bar\theta_0) \qquad ? \qquad \approx \quad \frac{1}{N}\sum_n \log p(\bar x_n|\bar\theta) \to \max \quad \bar\theta_{ML}$$

$$\underline{\bar\theta_0} = \arg\max_{\bar\theta} \underbrace{\mathbb{E}_{p_{data}(\bar x)}\left[\log p(\bar x|\bar\theta)\right]}$$

---

1) $\bar\theta_{ML}(D) \xrightarrow[N\to\infty]{} \bar\theta_0$, где $D \sim p_{data}(\bar x) = p(\bar x|\bar\theta_0)$

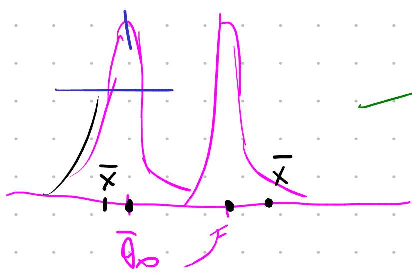
Для $\bar\theta_{ML}(D)$ верна асимпт. нормальность

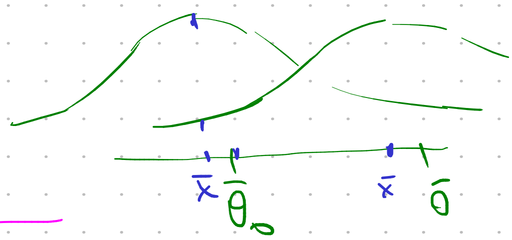
Fischer information matrix

$$\underbrace{\sqrt{N}\left(\bar\theta_{ML} - \bar\theta_0\right)} \xrightarrow{N\to\infty} \mathcal{N}\left(\,\cdot\,|\,\bar 0, I(\bar\theta_0)^{-1}\right)$$

$$\bar\theta_0 \qquad p(\bar x|\bar\theta_0) \sim \bar x \qquad \bar x \rightsquigarrow \bar\theta_0 ?$$

$$= p_{data}(\bar x) \qquad " \arg\max_{\bar\theta} p(\bar x|\bar\theta)$$





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

$\theta \in \mathbb{R}$

$$\mathbb{E}_{p(\bar{x}|\theta)}\left[\frac{\partial \log p(\bar{x}|\theta)}{\partial \theta}\right] = \int \cancel{p(\bar{x}|\theta)} \cdot \frac{\partial p / \partial \theta}{\cancel{p(\bar{x}|\theta)}} d\bar{x} =$$

$$= \frac{\partial}{\partial \theta} \int p(\bar{x}|\theta) d\bar{x} = \frac{\partial}{\partial \theta} 1 = 0$$

$$I(\theta) = \mathrm{Var}\left[\frac{\partial \log p(\bar{x}|\theta)}{\partial \theta}\right] = \mathbb{E}_{p(\bar{x}|\theta)}\left[\left(\frac{\partial \log p(\bar{x}|\theta)}{\partial \theta}\right)^2\right]$$

Fischer information

$$= \int \left(\frac{\partial \log p(\bar{x}|\theta)}{\partial \theta}\right)^2 p(\bar{x}|\theta) d\bar{x}$$

$$\frac{\partial^2 \log p(\bar{x}|\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{\frac{\partial p}{\partial \theta}}{p(\bar{x}|\theta)}\right) = \frac{\frac{\partial^2 p(\bar{x}|\theta)}{\partial \theta^2}}{p(\bar{x}|\theta)} - \frac{\left(\frac{\partial p(\bar{x}|\theta)}{\partial \theta}\right)^2}{p(\bar{x}|\theta)^2} =$$

$$= \underbrace{\frac{\partial^2 p / \partial \theta^2}{p(\bar{x}|\theta)}}_{=0} - \left(\frac{\partial \log p(\bar{x}|\theta)}{\partial \theta}\right)^2$$

$$\mathbb{E}_{p(\bar{x}|\theta)}\left[\frac{\partial^2 p / \partial \theta^2}{p}\right] = \int \cancel{p(\bar{x}|\theta)} \frac{\partial^2 p / \partial \theta^2}{\cancel{p(\bar{x}|\theta)}} d\bar{x} = \frac{\partial^2}{\partial \theta^2} \int p \, d\bar{x} = 0$$

$$I(\theta) = -\mathbb{E}_{p(\bar{x}|\theta)}\left[\frac{\partial^2 \log p(\bar{x}|\theta)}{\partial \theta^2}\right] = -\int \frac{\partial^2 \log p(\bar{x}|\theta)}{\partial \theta^2} p(\bar{x}|\theta) d\bar{x}$$

$\bar{\theta}$

$$I(\bar{\theta})_{ij} = \mathbb{E}_{p(\bar{x}|\bar{\theta})}\left[\frac{\partial \log p(\bar{x}|\bar{\theta})}{\partial \theta_i} \cdot \frac{\partial \log p(\bar{x}|\bar{\theta})}{\partial \theta_j}\right] =$$

$$= \underbrace{-\mathbb{E}_{p(\bar{x}|\bar{\theta})}\left[\frac{\partial^2 \log p(\bar{x}|\bar{\theta})}{\partial \theta_i \partial \theta_j}\right]}_{J}$$

2) $p_{data}(\bar{x}) \neq p(\bar{x}|\bar{\theta})$

$$\bar{\theta}_0 = \arg\max_{\bar{\theta}} \mathbb{E}_{p_{data}}\left[\log p(\bar{x}|\bar{\theta})\right]$$

$$\bar{\theta}_0 : \int p_{data}(\bar{x}) \frac{\partial \log p(\bar{x}|\bar{\theta}_0)}{\partial \bar{\theta}} d\bar{x} = \bar{0}$$

Если $\bar{x} \sim p_{data}(\bar{x})$, то:

$$\bar{\theta}_{ML}(D) \xrightarrow{N\to\infty} \bar{\theta}_0$$

$$N(\bar{\theta}_{ML}-\bar{\theta}_0)(\bar{\theta}_{ML}-\bar{\theta}_0)^T$$

$$\sqrt{N}(\bar{\theta}_{ML}-\bar{\theta}_0) \xrightarrow[N\to\infty]{} \mathcal{N}\left(\cdot \,\middle|\, \bar{0}, \; J^{-1}(\bar{\theta}_0)\, I(\bar{\theta}_0)\, J^{-1}(\bar{\theta}_0)\right)$$

$$I(\bar{\theta})_{ij} = \mathbb{E}_{p_{data}(\bar{x})}\left[\frac{\partial \log p(\bar{x}|\bar{\theta})}{\partial \theta_i} \cdot \frac{\partial \log p(\bar{x}|\bar{\theta})}{\partial \theta_j}\right]$$

$$J(\bar{\theta})_{ij} = -\mathbb{E}_{p_{data}(\bar{x})}\left[\frac{\partial^2 \log p(\bar{x}|\bar{\theta})}{\partial \theta_i \partial \theta_j}\right]$$

Если $p_{data} = p(\bar{x}|\bar{\theta}_0)$, то $I(\bar{\theta}_0) = J(\bar{\theta}_0)$

---

bias

$$b(p_{data}) = \mathbb{E}_{D\sim p_{data}}\left[\underline{\log p(D|\bar{\theta}_{ML}(D))} - N\cdot \mathbb{E}_{\bar{z}\sim p_{data}(\bar{z})}\left[\log p(\bar{z}|\bar{\theta}_{ML}(D))\right]\right]$$

$$\overbrace{\log p(D|w)}^{\bar{w}_{ML}} = \text{Const} - \frac{1}{2\sigma^2}\cdot\sum_n (y_n - w\cdot x_n)^2$$

$$\mathbb{E}_{D\sim p_{data}}\left[\log p(D|w, x)\right] = \text{Const} + \mathbb{E}\left[-\frac{1}{2\sigma^2}\sum_n (y_n - \bar{w}^T\bar{x}_n)^2\right] =$$

$$= -\frac{1}{2\sigma^2}\sum_n\left(\overbrace{\mathbb{E}[y_n^2]}^{=\mathbb{E}[y_n]^2+\sigma^2} - 2(\bar{w}^T\bar{x}_n)\cdot \overbrace{\mathbb{E}[y_n]}^{=\bar{w}_0^T\bar{x}_n} + (\bar{w}^T\bar{x}_n)^2\right) =$$

$$= -\frac{1}{2\sigma^2}\sum_n\left[(\bar{w}_0^T\bar{x}_n)^2 + \sigma^2 - 2(\bar{w}^T\bar{x}_n)(\bar{w}_0^T\bar{x}_n) + (\bar{w}^T\bar{x}_n)^2\right]$$

$$= -\frac{1}{2\sigma^2}\left(N\sigma^2 + \sum_n\left((\bar{w}-\bar{w}_0)^T\bar{x}_n\right)^2\right)$$

$$KL(p_{data} \| p(\bar{x}|\bar{\theta})) \longrightarrow \min$$

$$\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta})] \longrightarrow \max \qquad \bar{\theta}_0 = \arg\max_{\theta} \mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta})]$$

$$\underbrace{\mathbb{E}_{p_{data}}[\log p(\bar{x}|\bar{\theta}_{ML})]} = \log p(D|\bar{\theta}_{ML}) - b(p_{data})$$

$$\mathcal{M}_1, \dots, \mathcal{M}_K \qquad \mathcal{M}_k \quad p(\bar{x}|\bar{\theta}^{(k)}, \mathcal{M}_k)$$

$$\bar{\theta}^{(k)}_{ML} = \arg\max \log p(D|\bar{\theta}^{(k)}, \mathcal{M}_k)$$

$$p(\bar{x}|\bar{\theta}^{(k)}_{ML}, \mathcal{M}_k) \overset{?}{\approx} p_{data}(\bar{x})$$

$$KL(p_{data} \| p(\bar{x}|\bar{\theta}^{(k)}_{ML}, \mathcal{M}_k)) \xrightarrow[k]{} \min$$

$$\Updownarrow$$

$$\mathbb{E}_{p_{data}}[p(\bar{x}|\bar{\theta}^{(k)}_{ML}, \mathcal{M}_k)] \xrightarrow{k} \max$$

$$b(p_{data}) = \mathbb{E}_D[\log p(D|\bar{\theta}_{ML}) - N\,\mathbb{E}_{\bar{z}}[\log p(\bar{z}|\bar{\theta}_{ML})]] =$$

$$= \mathbb{E}_D[\log p(D|\bar{\theta}_{ML}) - \log p(D|\bar{\theta}_0)] \qquad B_1$$

$$+ \mathbb{E}_D[\log p(D|\bar{\theta}_0) - \underline{N\,\mathbb{E}_{\bar{z}}[\log p(\bar{z}|\bar{\theta}_0)]}] \qquad B_2$$

$$+ \mathbb{E}_D[N\,\underline{\mathbb{E}_{\bar{z}}[\log p(\bar{z}|\bar{\theta}_0)]} - N\,\underbrace{\mathbb{E}_{\bar{z}}[\log p(\bar{z}|\bar{\theta}_{ML})]}] \qquad B_3$$

$$\sum_n \log p(\bar{x}_n|\bar{\theta}_0)$$

$$B_2 = \mathbb{E}_D[\log p(D|\bar{\theta}_0)] - N \cdot \mathbb{E}_z[\log p(\bar{z}|\bar{\theta}_0)] = 0$$

B_3: $\quad \eta(\bar\theta) = \mathbb{E}_{\bar z}\left[\log p(\bar z|\bar\theta)\right] \qquad \bar\theta_0 = \arg\max \eta(\bar\theta)$

$$\hat\eta(\bar\theta) \approx \eta(\bar\theta_0) - \frac{1}{2}(\bar\theta - \bar\theta_0)^T \cdot J(\bar\theta_0)(\bar\theta - \bar\theta_0)$$

$$(\bar x - \bar x_0)^T A (\bar x - \bar x_0) = \mathrm{Tr}\left(A(\bar x - \bar x_0)(\bar x - \bar x_0)^T\right)$$

$$B_3 = N \cdot \mathbb{E}_D\left[\eta(\bar\theta_0) - \eta(\bar\theta_{ML})\right] \approx$$

$$\approx N \cdot \mathbb{E}_D\left[\frac{1}{2}(\bar\theta_{ML} - \bar\theta_0)^T J(\bar\theta_0)(\bar\theta_{ML} - \bar\theta_0)\right] =$$

$$= \frac{N}{2}\mathbb{E}_D\left[\mathrm{Tr}\left(J(\bar\theta_0)\cdot(\bar\theta_{ML} - \bar\theta_0)(\bar\theta_{ML} - \bar\theta_0)^T\right)\right] =$$

$$= \frac{N}{2}\mathrm{Tr}\left(J(\bar\theta_0)\cdot \mathbb{E}_D\left[(\bar\theta_{ML} - \bar\theta_0)(\bar\theta_{ML} - \bar\theta_0)^T\right]\right) \approx$$

$$\approx \frac{N}{2}\cdot\mathrm{Tr}\left(J(\bar\theta_0)\cdot\frac{1}{N}\cdot J(\bar\theta_0)^{-1} I(\bar\theta_0) J(\bar\theta_0)^{-1}\right)$$

$$B_3 = \frac{1}{2}\mathrm{Tr}\left(I(\bar\theta_0)\cdot J(\bar\theta_0)^{-1}\right)$$

$$B_1 = \mathbb{E}\left[\log p(D|\bar\theta_{ML}) - \log p(D|\bar\theta_0)\right]$$

$$\ell(\bar\theta) = \log p(D|\bar\theta)$$

$$\ell(\bar\theta) \approx \ell(\bar\theta_{ML}) - \frac{1}{2}(\bar\theta - \bar\theta_{ML})^T \cdot \left.\frac{\partial^2 \ell(\bar\theta)}{\partial\bar\theta\,\partial\bar\theta^T}\right|_{\bar\theta_{ML}} (\bar\theta - \bar\theta_{ML})$$

$N \to \infty$: $\quad \hat\theta_{ML} \to \bar\theta_0$

$$\frac{1}{N}\left.\frac{\partial^2 \ell(\bar\theta)}{\partial\bar\theta\,\partial\bar\theta^T}\right|_{\hat\theta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}\left.\frac{\partial^2 \log p(\bar x_n|\bar\theta)}{\partial\bar\theta\,\partial\bar\theta^T}\right|_{\bar\theta_{ML}} \longrightarrow J(\bar\theta_0)$$

$$\ell(\bar\theta_{ML}) - \ell(\bar\theta_0) \simeq \frac{N}{2}(\bar\theta_0-\bar\theta_{ML})^T J(\bar\theta_0)(\bar\theta_0-\bar\theta_{ML})$$

$$B_1 = \mathbb{E}_D\left[\ell(\bar\theta_{ML}) - \ell(\bar\theta_0)\right] \simeq \frac{N}{2}\mathbb{E}_D\left[Tr\left(J(\bar\theta_0)\cdot(\bar\theta_0-\bar\theta_{ML})(\bar\theta_0-\bar\theta_{ML})^T\right)\right]$$

$$= \frac{N}{2} Tr\left(J(\bar\theta_0)\cdot\frac{1}{N}J(\bar\theta_0)^{-1} I(\bar\theta_0) J(\bar\theta_0)^{-1}\right)$$

$$B_1 = \frac{1}{2} Tr\left(I(\bar\theta_0) J(\bar\theta_0)^{-1}\right)$$

$$b(p_{data}) = B_1 + B_2 + B_3 = Tr\left(I(\bar\theta_0)\cdot J(\bar\theta_0)^{-1}\right)$$

$$\mathbb{E}_{p_{data}}\left[\log p(D|\bar\theta_{ML})\right] \simeq \log p(D|\bar\theta_{ML}) - Tr\left(\underset{\simeq\hat I}{I(\bar\theta_0)}\underset{\simeq\hat J}{J(\bar\theta_0)}^{-1}\right)$$

Takeuchi inf. criterion

$$TIC = -2\log p(D|\bar\theta_{ML}) + 2\,\underline{Tr(\hat I \hat J^{-1})}$$

$$\hat I_{ij} = \frac{1}{N}\sum_n \frac{\partial \log p(\bar x_n|\bar\theta)}{\partial\theta_i}\cdot\frac{\partial\log p(\bar x_n|\bar\theta)}{\partial\theta_j}\Bigg|_{\bar\theta=\bar\theta_{ML}}$$

$$\hat J_{ij} = \frac{1}{N}\sum_n \frac{\partial^2 \log p(\bar x_n|\bar\theta)}{\partial\theta_i\partial\theta_j}\Bigg|_{\bar\theta=\bar\theta_{ML}}$$

Equal $p_{data}(\bar x) = p(\bar x|\bar\theta_0)$, so $I(\bar\theta_0) = J(\bar\theta_0)$

Akaike information criterion

$$AIC = -2\log p(D|\bar\theta_{ML}) + 2d$$

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

$$p(\theta|\alpha') \propto p(D|\theta) \times p(\theta|\alpha) \quad \text{conjugate priors}$$

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

Exponential family

норм. конст.

natural parameter

sufficient statistics

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

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

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

---

① Bernoulli / binomial

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

$$\frac{p}{1-p} = e^\theta, \quad p = \frac{e^\theta}{e^\theta + 1}$$

$$p = \sigma(\theta)$$

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

$$= \binom{n}{k} \cdot e^{k\log p + (n-k)\log(1-p)} = \binom{n}{k} e^{k\cdot\log\frac{p}{1-p} + n\log(1-p)}$$

$$h(k) = \binom{n}{k}, \quad t(k) = k, \quad \eta(p) = \log\frac{p}{1-p} = \theta, \quad a(p) = -n\log(1-p)$$

$$a(\theta) = -n\log(1-\sigma(\theta))$$

$$\mathrm{Bin}(k|n,\theta) = \binom{n}{k} e^{k\cdot\theta + n\log(1-\sigma(\theta))}$$

$$\nabla a(\theta) = +n\cdot\frac{\sigma(1-\sigma)}{1-\sigma} = +n\sigma$$

② $$\mathrm{Mult}(\bar{x}|n, p_1 \ldots p_k) = \frac{n!}{x_1! \ldots x_k!} \cdot p_1^{x_1} \ldots p_k^{x_k}, \quad \text{если } \sum x_i = n$$

$$\underbrace{\frac{n!}{x_1! \ldots x_k!}}_{h(\bar{x})} \cdot e^{\sum_i x_i \log p_i}$$

$$\Leftarrow \bar{\theta} = (\log p_1 \ldots \log p_k), \quad \sum e^{\theta_i} = 1$$

$$= (\log p_1 \ldots \log p_{k-1}, \log(1 - \sum p_i))$$

$$T(\bar{x}) = (x_1, \ldots, x_{n-1}, n - \sum x_i)$$

$$= h(\bar{x}) \cdot e^{\sum_{i=1}^{k-1} x_i \log p_i + (n - \sum x_i) \cdot \log(1 - \sum p_i)} =$$

$$= h(\bar{x}) \cdot e^{\sum_{i=1}^{k-1} x_i \cdot \log \frac{p_i}{1 - \sum p_i} + \underbrace{n \cdot \log(1 - \sum p_i)}_{= -a(\bar{\theta})}}$$

$\bar{t}(\bar{x}) = \bar{x}$ $\quad = \bar{\theta}$

$$\theta_i = \log \frac{p_i}{1 - \sum p_i} = \log \frac{p_i}{p_k}$$

$$\sum_{j=1}^{k-1} p_i = \frac{\sum_1^{k-1} e^{\theta_i}}{1 + \sum^{k-1} e^{\theta_i}}$$

$$p_i = p_k \cdot e^{\theta_i}$$

$$1 - p_k = \sum^{k-1} p_i = p_k \cdot \sum e^{\theta_i} \qquad p_k = \frac{1}{1 + \sum_1^{k-1} e^{\theta_i}}$$

$$\text{softmax}_i = \frac{e^{\theta_i}}{\sum_{j=1}^{k} e^{\theta_j}} = p_i$$

$$p_i = \frac{e^{\theta_i}}{1 + \sum_1^{k-1} e^{\theta_i}} \quad ; \quad \theta_k = \log \frac{p_k}{p_k} = 0$$

$$e^{\theta_k} =$$

---

③ Poisson

$h(x)$ $t(x)$ $\theta(\lambda)$ $a(\theta) = e^{\theta}$

$$p(x \mid \lambda) = \frac{1}{x!} e^{-\lambda} \lambda^x = \frac{1}{x!} e^{x \log \lambda - \lambda}$$

---

④ $$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x - \frac{\mu^2}{2\sigma^2} - \log\sigma} =$$

$$-\frac{\theta_2^2}{4\theta_1} - \frac{1}{2}\log(-2\theta_1) = \frac{\mu^2\tau}{2} - \frac{1}{2}\log\tau = a(\bar{\theta})$$

$h(x)$

$$= \frac{1}{\sqrt{2\pi}} e^{\begin{pmatrix} x^2 \\ x \end{pmatrix}^T \begin{pmatrix} -1/2\sigma^2 \\ \mu/\sigma^2 \end{pmatrix} - \left(\frac{\mu^2}{2\sigma^2} + \log\sigma\right)}$$

$= \bar{t}(x)$ $\quad = \bar{\theta}$

$$\bar{\theta} = \begin{pmatrix} -1/2\sigma^2 \\ \mu/\sigma^2 \end{pmatrix} = \begin{pmatrix} -\tau/2 \\ \mu\tau \end{pmatrix}$$

$$\frac{1}{\sqrt{2\pi \det \Sigma}} \cdot e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma^{-1} (\bar{x}-\bar{\mu})} =$$

$$= e^{\underbrace{-\frac{1}{2}\bar{x}^T \Sigma^{-1} \bar{x}}_{\sum_{ij} \Sigma^{-1}_{ij} x_i x_j} + \bar{x}^T(\Sigma^{-1}\bar{\mu}) - \overbrace{\left(\frac{1}{2}\bar{\mu}^T \underbrace{\Sigma^{-1}\bar{\mu}}_{\bar{\theta}_2} + \frac{1}{2}\log\det\Sigma\right)}^{a(\bar{\theta})}}$$

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

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

$$\bar{\theta} \longrightarrow \max$$

$$\log p(D|\bar{\theta}) = \sum_n \log h(\bar{x}_n) - N \cdot a(\bar{\theta}) + \sum_{n=1}^{N} \eta(\bar{\theta})^T t(\bar{x}_n) =$$

$$= \mathrm{Const} - N \cdot a(\bar{\theta}) + \eta(\bar{\theta})^T \left(\sum_n t(\bar{x}_n)\right)$$

$$\mathcal{N}(x|\mu, \sigma) \qquad \sum t(x_n) = \begin{pmatrix} \sum x_n^2 \\ \sum x_n \end{pmatrix}$$

$$\mu_{ML} = \frac{1}{N}\sum x_n, \quad \sigma_{ML} = \frac{1}{N}\sqrt{\sum (x_n - \bar{x})^2} = \frac{1}{N}\sqrt{\sum x_n^2 - \frac{2}{N}(\sum x_n)^2 + \frac{N}{N^2}(\sum x_n)^2}$$

$$\bar{x} = \frac{1}{N}\sum x_n$$

$$= \frac{1}{N}\sqrt{\sum x_n^2 - \frac{1}{N}(\sum x_n)^2}$$

$$\nabla_{\bar{\theta}} \log p(D|\bar{\theta}) = 0$$

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

$$\bar{\theta}_{ML}: \quad \nabla_{\bar{\theta}} a(\bar{\theta}) = \frac{1}{N} \cdot \nabla_{\bar{\theta}} \eta(\bar{\theta})^T \left(\sum_n t(\bar{x}_n)\right)$$

Ecan $\eta(\bar{\theta}) = \bar{\theta}$, то

$$\boxed{\nabla_{\bar{\theta}} a(\bar{\theta}) = \frac{1}{N} \cdot \sum_n t(\bar{x}_n)}$$

Bin: $x_i = t_i$

$$\frac{1}{N}\sum_i x_i = \bar{t}$$

$$n \cdot p_{ML} = \bar{t}$$