

## 4.3.5 Probit regression

$$\begin{aligned}\Phi(a) &= \int_{-\infty}^a \mathcal{N}(\theta \mid 0, 1) d\theta \\ &= \frac{1}{2} + \int_0^a \mathcal{N}(\theta \mid 0, 1) d\theta \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^a \exp(-\theta^2/2) d\theta\end{aligned}$$

replace  $\theta = \sqrt{2}\theta'$

$$\begin{aligned}&= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{a/\sqrt{2}} \exp(-\theta'^2) d\theta' \\ &= \frac{1}{2} \left\{ 1 + \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) \right\}\end{aligned}$$

$$\operatorname{erf}(a) = \frac{2}{\sqrt{\pi}} \int_0^a \exp(-\theta^2) d\theta$$

## 4.3.5 Probit regression

- aaa
- bbb



## 4.3.5 Probit regression

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

$$\begin{aligned} p(t \mid \mathbf{x}) &= (1 - \epsilon)\sigma(\mathbf{x}) + \epsilon(1 - \sigma(\mathbf{x})) \\ &= \epsilon + (1 - 2\epsilon)\sigma(\mathbf{x}) \end{aligned}$$

## 4.3.6 Canonical link functions

$$p(t \mid \eta, s) = \frac{1}{s} h\left(\frac{t}{s}\right) g(\eta) \exp\left\{\frac{\eta t}{s}\right\}$$

$$y \equiv \mathbb{E}[t \mid \eta] = -s \frac{d}{d\eta} \ln g(\eta)$$

$$y = f(\mathbf{w}^T \boldsymbol{\phi})$$

$$\ln p(\mathbf{t} \mid \eta, s) = \sum_{n=1}^N \ln p(t_n \mid \eta, s) = \sum_{n=1}^N \left\{ \ln g(\eta_n) + \frac{\eta_n t_n}{s} \right\} + \text{const}$$

$$\begin{aligned}
\nabla_{\mathbf{w}} \ln p(\mathbf{t} \mid \eta, s) &= \sum_{n=1}^N \left\{ \frac{d}{d\eta_n} \ln g(\eta_n) + \frac{t_n}{s} \right\} \frac{d\eta_n}{dy_n} \frac{dy_n}{da_n} \nabla a_n \\
&= \sum_{n=1}^N \frac{1}{s} \{t_n - y_n\} \psi'(y_n) f'(a_n) \phi_n
\end{aligned}$$

$$f^{-1}(y) = \psi(y)$$

$$\nabla \ln E(\mathbf{w}) = \frac{1}{s} \sum_{n=1}^N \{y_n - t_n\} \phi_n$$

$$p(z) = \frac{1}{Z}$$

$$\left.\frac{df(z)}{dz}\right|_{z=z_0}=0$$

$$\ln f(z) \simeq \ln f\left(z_0\right)-\frac{1}{2} A\left(z-z_0\right)^2$$

$$A=-\left.\frac{d^2}{dz^2}\ln f(z)\right|_{z=z_0}$$

$$f(z) \simeq f\left(z_0\right) \exp \left\{-\frac{A}{2}\left(z-z_0\right)^2\right\}$$



$$q(z) = \left(\frac{A}{2\pi}\right)^{1/2} \exp\left\{-\frac{A}{2} (z - z_0)^2\right\}$$

$$\ln f(\mathbf{z}) \simeq \ln f(\mathbf{z}_0) - \frac{1}{2} (\mathbf{z} - \mathbf{z}_0)^T \mathbf{A} (\mathbf{z} - \mathbf{z}_0)$$

$$f(\mathbf{z}) \simeq f(\mathbf{z}_0) \exp\left\{-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)^T \mathbf{A} (\mathbf{z} - \mathbf{z}_0)\right\}$$

$$q(\mathbf{z}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{M/2}} \exp\left\{-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)^T \mathbf{A} (\mathbf{z} - \mathbf{z}_0)\right\} = \mathcal{N}(\mathbf{z} \mid \mathbf{z}_0, \mathbf{A}^{-1})$$

## **4.4 The Laplace Approximation**

## 4.4.1 Model comparison and BIC

$$\begin{aligned} Z &= \int f(\mathbf{z}) d\mathbf{z} \\ &\simeq f(\mathbf{z}_0) \int \exp \left\{ -\frac{1}{2} (\mathbf{z} - \mathbf{z}_0)^T \mathbf{A} (\mathbf{z} - \mathbf{z}_0) \right\} d\mathbf{z} \\ &= f(\mathbf{z}_0) \frac{(2\pi)^{M/2}}{|\mathbf{A}|^{1/2}} \end{aligned}$$

$$p(\mathcal{D}) = \int p(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$\ln p(\mathcal{D}) \simeq \ln p(\mathcal{D} \mid \boldsymbol{\theta}_{\text{MAP}}) + \ln p(\boldsymbol{\theta}_{\text{MAP}}) + \frac{M}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{A}|$$

$$\mathbf{A} = -\nabla \nabla \ln p(\mathcal{D} \mid \boldsymbol{\theta}_{\text{MAP}}) p(\boldsymbol{\theta}_{\text{MAP}}) = -\nabla \nabla \ln p(\boldsymbol{\theta}_{\text{MAP}} \mid \mathcal{D})$$

$$\ln p(\mathcal{D}) \simeq \ln p(\mathcal{D} \mid \boldsymbol{\theta}_{\text{MAP}}) - \frac{1}{2} M \ln N$$

## Exercise 4.22

$$p(D) = \int p(D \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= f(\boldsymbol{\theta}_{MAP}) \frac{(2\pi)^{M/2}}{|\mathbf{A}|^{1/2}}$$

$$= p(D \mid \boldsymbol{\theta}_{MAP}) p(\boldsymbol{\theta}_{MAP}) \frac{(2\pi)^{M/2}}{|\mathbf{A}|^{1/2}}$$

$$\ln p(D) = \ln p(D \mid \boldsymbol{\theta}_{MAP}) + \ln p(\boldsymbol{\theta}_{MAP}) + \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{A}|$$

## Exercise 4.23

$$\begin{aligned}\ln p(D) &= \ln p(D \mid \boldsymbol{\theta}_{MAP}) + \ln p(\boldsymbol{\theta}_{MAP}) + \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{A}| \\ &= \ln p(D \mid \boldsymbol{\theta}_{MAP}) - \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{V}_0| - \frac{1}{2} (\boldsymbol{\theta}_{MAP} - \mathbf{m})^T \mathbf{V}_0^{-1} (\boldsymbol{\theta}_{MAP} - \mathbf{m}) \\ &\quad + \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{A}| \\ &= \ln p(D \mid \boldsymbol{\theta}_{MAP}) - \frac{1}{2} \ln |\mathbf{V}_0| - \frac{1}{2} (\boldsymbol{\theta}_{MAP} - \mathbf{m})^T \mathbf{V}_0^{-1} (\boldsymbol{\theta}_{MAP} - \mathbf{m}) - \frac{1}{2} \ln |\mathbf{A}|\end{aligned}$$

$$\begin{aligned}\mathbf{A} &= -\nabla \nabla \ln p(D \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \\ &= -\nabla \nabla \ln p(D \mid \boldsymbol{\theta}) \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} - \nabla \nabla \ln p(\boldsymbol{\theta}) \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} - \nabla \nabla \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \mathbf{m})^T \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}) \right\} \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} + \nabla \left\{ \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}) \right\} \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} + \mathbf{V}_0^{-1}\end{aligned}$$

$$\ln \left\{ \left| \mathbf{V}_0 \right| \cdot \left| \mathbf{H} + \mathbf{V}_0^{-1} \right| \right\}$$

$$\ln \left\{ \left| \mathbf{V}_0 \mathbf{H} + \mathbf{I} \right| \right\}$$

$$\ln \left| \mathbf{V}_0 \right| - \frac{1}{2} \ln \left| \mathbf{H} \right|$$

$$\ln \left| \mathbf{H} \right| + \text{const}$$

$$\mathbf{H} = \sum_{n=1}^N \mathbf{H}_{\mathbf{n}} = N \hat{\mathbf{H}}$$