4.3.5 Probit regression

$$egin{align} \Phi(a) &= \int_{-\infty}^a \mathcal{N}(heta \mid 0, 1) \mathrm{d} heta \ &= rac{1}{2} + \int_0^a \mathcal{N}(heta \mid 0, 1) d heta \ &= rac{1}{2} + rac{1}{\sqrt{2\pi}} \int_0^a \exp\left(- heta^2/2
ight) d heta \ \end{gathered}$$

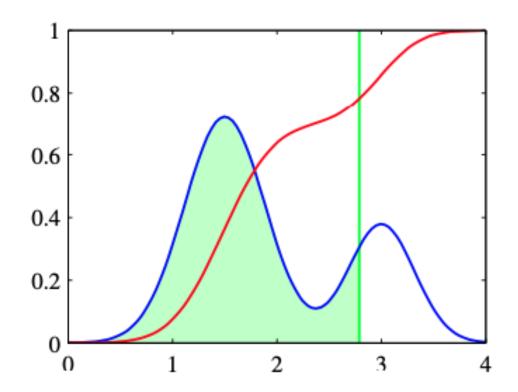
replace
$$heta=\sqrt{2} heta'$$

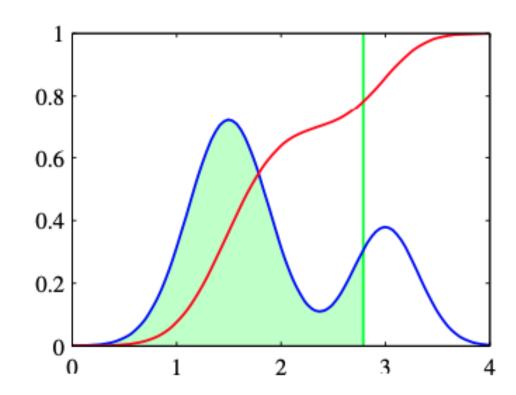
$$=rac{1}{2}+rac{1}{\sqrt{\pi}}\int_0^{a/\sqrt{2}}\exp\left(- heta'^2
ight)d heta'
onumber \ =rac{1}{2}\left\{1+ ext{erf}(rac{a}{\sqrt{2}})
ight\}$$

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

4.3.5 Probit regression

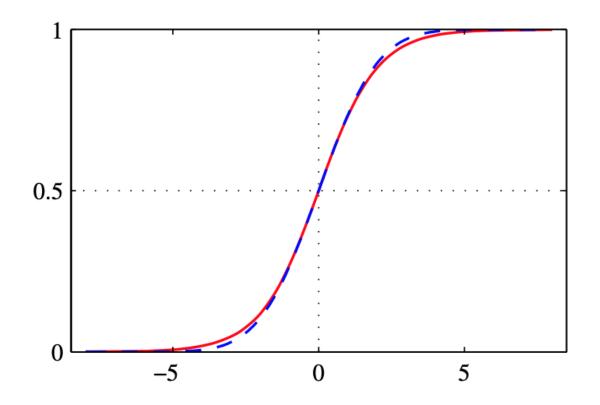
- aaa
- bbb





4.3.5 Probit regression

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



$$p(t \mid \mathbf{x}) = (1 - \epsilon)\sigma(\mathbf{x}) + \epsilon(1 - \sigma(\mathbf{x}))$$

= $\epsilon + (1 - 2\epsilon)\sigma(\mathbf{x})$

4.3.6 Canonical link functions

$$egin{aligned} p(t \mid \eta, s) &= rac{1}{s} h\left(rac{t}{s}
ight) g(\eta) \exp\left\{rac{\eta t}{s}
ight\} \ y &\equiv \mathbb{E}[t \mid \eta] = -srac{d}{d\eta} \ln g(\eta) \ y &= f\left(\mathbf{w}^{\mathrm{T}} oldsymbol{\phi}
ight) \ \ln p(\mathbf{t} \mid \eta, s) &= \sum_{n=1}^{N} \ln p\left(t_n \mid \eta, s
ight) = \sum_{n=1}^{N} \left\{ \ln g\left(\eta_n
ight) + rac{\eta_n t_n}{s}
ight\} + \mathrm{\,const.} \end{aligned}$$

$$egin{aligned}
abla_{\mathbf{w}} \ln p(\mathbf{t} \mid \eta, s) &= \sum_{n=1}^{N} \left\{ rac{d}{d\eta_n} \ln g\left(\eta_n
ight) + rac{t_n}{s}
ight\} rac{d\eta_n}{dy_n} rac{dy_n}{da_n}
abla a_n \end{aligned} \ &= \sum_{n=1}^{N} rac{1}{s} \left\{ t_n - y_n
ight\} \psi'\left(y_n
ight) f'\left(a_n
ight) oldsymbol{\phi}_n \end{aligned} \ f^{-1}(y) &= \psi(y) \end{aligned} \ egin{aligned}
abla \ln E(\mathbf{w}) &= rac{1}{s} \sum_{n=1}^{N} \left\{ y_n - t_n
ight\} \phi_n \end{aligned}$$

$$egin{aligned} p(z) &= rac{1}{Z} \ rac{df(z)}{dz}igg|_{z=z_0} &= 0 \ \ln f(z) &\simeq \ln f\left(z_0
ight) - rac{1}{2}A\left(z-z_0
ight)^2 \ A &= -\left.rac{d^2}{dz^2}\ln f(z)
ight|_{z=z_0} \ f(z) &\simeq f\left(z_0
ight) \exp\left\{-rac{A}{2}\left(z-z_0
ight)^2
ight\} \end{aligned}$$

$$egin{aligned} q(z) &= \left(rac{A}{2\pi}
ight)^{1/2} \exp\left\{-rac{A}{2}\left(z-z_0
ight)^2
ight\} \ & \ln f(\mathbf{z}) \simeq \ln f\left(\mathbf{z}_0
ight) - rac{1}{2}\left(\mathbf{z}-\mathbf{z}_0
ight)^{\mathrm{T}} \mathbf{A} \left(\mathbf{z}-\mathbf{z}_0
ight) \ & f(\mathbf{z}) \simeq f\left(\mathbf{z}_0
ight) \exp\left\{-rac{1}{2}\left(\mathbf{z}-\mathbf{z}_0
ight)^{\mathrm{T}} \mathbf{A} \left(\mathbf{z}-\mathbf{z}_0
ight)
ight\} \ & q(\mathbf{z}) = rac{|\mathbf{A}|^{1/2}}{(2\pi)^{M/2}} \exp\left\{-rac{1}{2}\left(\mathbf{z}-\mathbf{z}_0
ight)^{\mathrm{T}} \mathbf{A} \left(\mathbf{z}-\mathbf{z}_0
ight)
ight\} = \mathcal{N} \left(\mathbf{z} \mid \mathbf{z}_0, \mathbf{A}^{-1}
ight) \end{aligned}$$

4.4 The Laplace Approximation

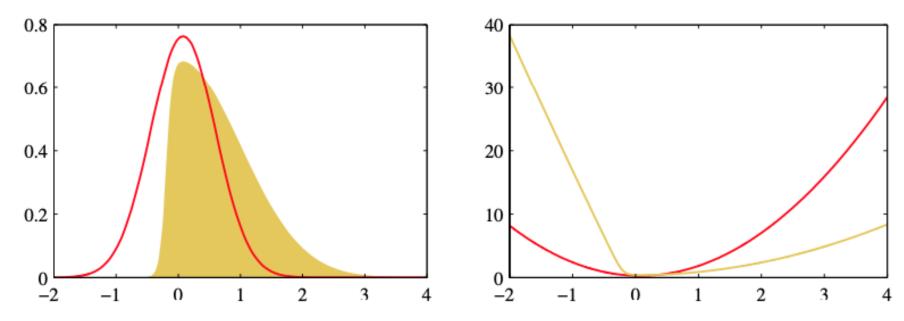


Figure 4.14 Illustration of the Laplace approximation applied to the distribution $p(z) \propto \exp(-z^2/2)\sigma(20z+4)$ where $\sigma(z)$ is the logistic sigmoid function defined by $\sigma(z) = (1+e^{-z})^{-1}$. The left plot shows the normalized distribution p(z) in yellow, together with the Laplace approximation centred on the mode z_0 of p(z) in red. The right plot shows the negative logarithms of the corresponding curves.

4.4.1 Model comparison and BIC

$$egin{aligned} Z &= \int f(\mathbf{z}) \mathrm{d}\mathbf{z} \ &\simeq f\left(\mathbf{z}_0
ight) \int \exp\left\{-rac{1}{2} \left(\mathbf{z} - \mathbf{z}_0
ight)^\mathrm{T} \mathbf{A} \left(\mathbf{z} - \mathbf{z}_0
ight)
ight\} \mathrm{d}\mathbf{z} \ &= f\left(\mathbf{z}_0
ight) rac{(2\pi)^{M/2}}{|\mathbf{A}|^{1/2}} \ &p(\mathcal{D}) = \int p(\mathcal{D} \mid oldsymbol{ heta}) p(oldsymbol{ heta}) \mathrm{d}oldsymbol{ heta} \ &\ln p(\mathcal{D}) \simeq \ln p\left(\mathcal{D} \mid oldsymbol{ heta}_\mathrm{MAP}
ight) + \ln p\left(oldsymbol{ heta}_\mathrm{MAP}
ight) + rac{M}{2} \ln(2\pi) - rac{1}{2} \ln |\mathbf{A}| \end{aligned}$$

$$egin{aligned} \mathbf{A} &= -
abla
abla \ln p \left(\mathcal{D} \mid oldsymbol{ heta}_{ ext{MAP}}
ight) p \left(oldsymbol{ heta}_{ ext{MAP}}
ight) = -
abla
abla \ln p \left(oldsymbol{ heta}_{ ext{MAP}} \mid \mathcal{D}
ight) \ & = \ln p \left(\mathcal{D} \mid oldsymbol{ heta}_{ ext{MAP}}
ight) - rac{1}{2} M \ln N \end{aligned}$$

Exercise 4.22

$$egin{aligned} p(D) &= \int p(D \mid oldsymbol{ heta}) p(oldsymbol{ heta}) p(oldsymbol{ heta}) doldsymbol{ heta} \ &= f\left(oldsymbol{ heta}_{MAP}
ight) rac{\left(2\pi
ight)^{M/2}}{|oldsymbol{ heta}|^{1/2}} \ &= p\left(D \mid oldsymbol{ heta}_{MAP}
ight) p\left(oldsymbol{ heta}_{MAP}
ight) rac{\left(2\pi
ight)^{M/2}}{|oldsymbol{ heta}|^{1/2}} \ &= p\left(D \mid oldsymbol{ heta}_{MAP}
ight) + \ln p\left(oldsymbol{ heta}_{MAP}
ight) + rac{M}{2} \ln 2\pi - rac{1}{2} \ln |oldsymbol{ heta}| \end{aligned}$$

Exercise 4.23

$$\begin{split} \ln p(D) &= \ln p\left(D \mid \boldsymbol{\theta}_{MAP}\right) + \ln p\left(\boldsymbol{\theta}_{MAP}\right) + \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{A}| \\ &= \ln p\left(D \mid \boldsymbol{\theta}_{MAP}\right) - \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{V}_{0}| - \frac{1}{2} \left(\boldsymbol{\theta}_{MAP} - \mathbf{m}\right)^{T} \mathbf{V}_{0}^{-1} \left(\boldsymbol{\theta}_{MAP} - \mathbf{m}\right) \\ &+ \frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{A}| \\ &= \ln p\left(D \mid \boldsymbol{\theta}_{MAP}\right) - \frac{1}{2} \ln |\mathbf{V}_{0}| - \frac{1}{2} \left(\boldsymbol{\theta}_{MAP} - \mathbf{m}\right)^{T} \mathbf{V}_{0}^{-1} \left(\boldsymbol{\theta}_{MAP} - \mathbf{m}\right) - \frac{1}{2} \ln |\mathbf{A}| \\ &\mathbf{A} = -\nabla \nabla \ln p\left(D \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{MAP}} \\ &= -\nabla \nabla \ln p\left(D \mid \boldsymbol{\theta}\right) \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{MAP}} - \nabla \nabla \ln p\left(\boldsymbol{\theta}\right) \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} - \nabla \nabla \left\{ -\frac{1}{2} \left(\boldsymbol{\theta} - \mathbf{m}\right)^{T} \mathbf{V}_{0}^{-1} \left(\boldsymbol{\theta} - \mathbf{m}\right) \right\} \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} + \nabla \left\{ \mathbf{V}_{0}^{-1} \left(\boldsymbol{\theta} - \mathbf{m}\right) \right\} \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{MAP}} \\ &= \mathbf{H} + \mathbf{V}_{0}^{-1} \end{split}$$

$$\ln \left\{ |\mathbf{V}_0| \cdot \left| \mathbf{H} + \mathbf{V}_0^{-1} \right| \right\}$$
 $\ln \left\{ |\mathbf{V}_0 \mathbf{H} + \mathbf{I}| \right\}$
 $\ln |\mathbf{V}_0| - \frac{1}{2} \ln |\mathbf{H}|$
 $\ln |\mathbf{H}| + \text{const}$

 $\mathbf{H} = \sum \mathbf{H_n} = N\hat{\mathbf{H}}$

n=1