Normal models

David J. H. Shih

Intended learning outcomes

- Recognize and apply normal distributions in statistical models

Recall: Normal distribution

Definition

A random variable X has a normal (Gaussian) distribution if

$$f_X\left(x \mid \mu, \tau^{-1}\right) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(x-\mu)^2\right)$$

where μ is the mean parameter and $\tau > 0$ is the inverse-variance parameter.

Known mean and known variance

Given observed random variables $(X_i)_{i=1}^{N}$, known mean μ, and known inverse-variance τ, we can define

$$X_i \sim \mathsf{Normal}\left(\mu, \tau^{-1}\right).$$

Known mean and unknown variance

Given observed random variables $(X_i)_{i=1}^N$ and known mean μ, we can define

$$\tau \sim \mathsf{Gamma}\left(\alpha, \beta\right)$$
$$X_i \mid \tau \sim \mathsf{Normal}\left(\mu, \tau^{-1}\right),$$

where $\alpha > 0$ and $\beta > 0$ are fixed hyperparameters.

Using Bayes' theorem, we can derive the posterior:

$$p\left(\tau \mid \boldsymbol{x}\right) = \mathsf{Gamma}\left(\alpha + \frac{N}{2},\ \beta + \frac{r}{2}\right), \quad r = \sum_i (x_i - \mu)^2.$$

Unknown mean and known variance

Given observed random variables $(X_i)_{i=1}^{N}$ and known inverse-variance τ, we can define

$$\mu \sim \text{Normal}\left(\nu, \lambda^{-1}\right)$$
$$X_i \mid \mu \sim \text{Normal}\left(\mu, \tau^{-1}\right),$$

where ν and $\lambda > 0$ are fixed hyperparameter.

Using Bayes' theorem, we can derive the posterior:

$$p\left(\mu \mid \boldsymbol{x}\right) = \text{Normal}\left(\frac{\lambda\nu + \tau s}{\lambda + \tau N}, (\lambda + \tau N)^{-1}\right), \quad s = \sum_i x_i.$$

This involves completing the square.

Unknown mean and unknown variance

Given observed random variables $(X_i)_{i=1}^{N}$, we can define

$$\mu \mid \tau \sim \text{Normal}\left(\nu, (\lambda\tau)^{-1}\right)$$
$$\tau \sim \text{Gamma}\left(\alpha, \beta\right)$$
$$X_i \mid \mu, \tau \sim \text{Normal}\left(\mu, \tau^{-1}\right),$$

where ν, $\lambda > 0$, $\alpha > 0$, and $\beta > 0$ are fixed hyperparameters.

The joint prior on (μ, τ) is conjugate to the normal likelihood with unknown mean and variance:

$$p\left(\mu, \tau\right) = p\left(\mu \mid \tau\right) p\left(\tau\right) = \text{NormalGamma}\left(\nu, \lambda, \alpha, \beta\right).$$

Using Bayes' theorem, we can derive the posterior of

$$
\begin{aligned}
p\left(\mu, \tau \mid \boldsymbol{x}\right) &= \mathsf{NormalGamma}\left(\nu', \lambda', \alpha', \beta'\right) \\
\nu' &= \frac{\nu\lambda + s}{\lambda + N} \\
\lambda' &= \lambda + N \\
\alpha' &= \alpha + \frac{N}{2} \\
\beta' &= \beta + \frac{1}{2}\left(\lambda\nu^2 - \lambda'\nu'^2 + \sum_i x_i^2\right).
\end{aligned}
$$

Heteroscedastic normals

Given observed random variables $(X_i)_{i=1}^N$ and known mean μ, we can define

$$\tau_i \sim \mathsf{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

$$X_i \mid \mu, \tau_i \sim \mathsf{Normal}\left(\mu, \tau_i^{-1}\right).$$

Notice that X_i are independent but not identically distributed.

We can marginalize each τ_i out by integration:

$$p\left(X_i \mid \mu\right) = \int_0^\infty p\left(X_i \mid \mu, \tau_i\right) p\left(\tau_i\right) d\tau_i,$$

which will give

$$p\left(X_i \mid \mu\right) = \mathsf{StudentT}\left(\mu, \nu\right).$$

Student's t distribution

Definition

A random variable X has a Student's t distribution if

$$f_X(x \mid \mu, \nu) = (\pi\nu)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where μ is the mean parameter and $\nu > 0$ is the degree of freedom.

Exponential family

The exponential family is a set of distributions that follow the form:

$$f(x \mid \boldsymbol{\theta}) = h(\boldsymbol{x}) \, \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta})^\top \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\theta})\right),$$

where $\boldsymbol{\theta}$ is the parameter vector, $h(\boldsymbol{x})$, $\boldsymbol{\eta}(\boldsymbol{\theta})$, $\boldsymbol{T}(\boldsymbol{x})$, and $A(\theta)$ are functions for each distribution.

$\boldsymbol{\eta}(\boldsymbol{\theta})$ is a vector of natural parameters.

$\boldsymbol{T}(\boldsymbol{x})$ is a vector of sufficient statistics.

Normal distribution in exponential family form

$$\mathsf{Normal}(x \mid \boldsymbol{\theta}) = h(x)\,\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta})^\top \boldsymbol{T}(x) - A(\boldsymbol{\theta})\right)$$

$$h(x) = (2\pi)^{-\frac{1}{2}}, \quad A(\boldsymbol{\theta}) = \frac{1}{2}\mu^2\tau - \frac{1}{2}\log\tau$$

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \begin{bmatrix} \mu\tau \\ -\frac{1}{2}\tau \end{bmatrix}, \quad \boldsymbol{T}(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}.$$

Normal distribution with natural parameters

$$\begin{aligned}
\mathsf{Normal}(x \mid \boldsymbol{\eta}) &= h(x)\, \exp\left(\boldsymbol{\eta}^\top \boldsymbol{T}(x) - A(\boldsymbol{\eta})\right) \\
h(x) = (2\pi)^{-\frac{1}{2}}, &\quad A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \\
\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}, &\quad \boldsymbol{T}(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}, \quad \boldsymbol{\theta}(\boldsymbol{\eta}) = \begin{bmatrix} -\frac{\eta_1}{2\eta_2} \\ -\frac{1}{2\eta_2} \end{bmatrix}.
\end{aligned}$$

Summary

"The theory of probabilities is nothing more than common sense reduced to calculus." - Pierre-Simon Laplace

Intended learning outcomes

- Recognize and apply normal distributions in statistical models