Exponential family and location-scale

Wednesday, June 12, 2024 09:57

$f(x;\boldsymbol\theta) = h(x)c(\boldsymbol\theta)\exp\{\sum_{i=1}^{k} w_i(\boldsymbol\theta)t_i(x)\}$

↳ θ is params of PDF

Simplified form: $f(y;\theta,\phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y;\phi)\}$

↳ θ is not the param of the PDF

↳ a function of the param of the PDF

θ is called canonical param

Binomial $f(x|p) = \binom{n}{x} p^{x}(1-p)^{n-x}$

$= \binom{n}{x}\exp(x\log p + (n-x)\log(1-p))$

$= \binom{n}{x}\exp\{x\log(\frac{p}{1-p}) + n\log(1-p)\}$

$= \underbrace{\binom{n}{x}}_{h(x)}\underbrace{(1-p)^n}_{c(p)}\exp\{\underbrace{x\cdot\log(\frac{p}{1-p})}_{t_1(x)\cdot w_1(p)}\}$

Normal $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$

$= \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x^2+\mu^2-2x\mu)\}$

$= \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x^2-2x\mu)\}\exp\{-\frac{1}{2\sigma^2}\mu^2\}$

$\Rightarrow h(x)=1 \quad c(\boldsymbol\theta) = \frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\{-\frac{1}{2\sigma^2}\mu^2\}$

$\Rightarrow w_1(\boldsymbol\theta) = -\frac{1}{2\sigma^2}$ $\frac{1}{\sigma^2}$; $\sigma>0$ $w_2(\boldsymbol\theta) = +\frac{1}{\cancel{2}\sigma^2}\cdot\cancel{x}\mu = \frac{\mu}{\sigma^2}$ $\sigma>0$

$t_1(x) = x^2$ $-\frac{x^2}{2}$; $t_2(x) = x$

W only defined over the range of the parameter

using indicator function.

$I_A(x) = \begin{cases}1, & x\in A\\ 0, & x\notin A\end{cases}$ or $I(x\in A)$

↳ a fcn of the data.

$\therefore f(x|\mu,\sigma^2) = h(x)c(\boldsymbol\theta)\exp\{\sum_{i=1}^{2} w_i(\boldsymbol\theta)t_i(x)\} I_{(-\infty,+\infty)}(x)$

$\therefore I_{(-\infty,+\infty)}(x)$ is a function of the data x, can be incorporated into $h(x)$. if $I(x)$ depends on $\boldsymbol\theta$, then not exponential family

e.g.

$f(x|\theta) = \theta^{-1}\exp\{1-(\frac{x}{\theta})\}\underline{I_{[\theta,\infty)}(x)}$ ↳ a fcn of both θ and x

Pois $\frac{e^{-\lambda}\lambda^x}{x!}$

$\therefore f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = \exp\{-\lambda + x\log\lambda - \log x!\} I_{(0,+\infty)}(x)$

$= \exp\{\underbrace{x}_{t_1(x)}\underbrace{\log\lambda}_{w_1(\lambda)}\}\cdot\underbrace{\exp\{-\lambda\}}_{c(\lambda)}\cdot\underbrace{\exp\{-\log x!\}I_{(0,+\infty)}(x)}_{h(x)}$

Nice properties of exponential family

① $E(\sum_{i=1}^{k}\frac{w_i(\boldsymbol\theta)}{\partial\theta_j}t_i(X)) = -\frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta)$

proof:

$f(x;\boldsymbol\theta) = h(x)c(\boldsymbol\theta)\exp\{\sum_{i=1}^{k} w_i(\boldsymbol\theta)t_i(x)\}$

$\log f(x;\boldsymbol\theta) = \log h(x) + \log c(\boldsymbol\theta) + \sum_{i=1}^{k} w_i(\boldsymbol\theta)t_i(x)$ ① 取 log

$\frac{\partial}{\partial\theta_j}\log f(x;\boldsymbol\theta) = \frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta) + \frac{\partial}{\partial\theta_j}w_i(\boldsymbol\theta)t_i(x)$ ② take derivative

→ 实际上,这是 score 的 Expectation, 它等于 0.

$E[\underbrace{\frac{\partial}{\partial\theta_j}\log f(x;\boldsymbol\theta)}_{=\frac{1}{f(x;\boldsymbol\theta)}\cdot\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)}] = E[\frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta)] + E[\frac{\partial}{\partial\theta_j}w_i(\boldsymbol\theta)t_i(x)]$ ③ take expectation:

$\int\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)\cdot\frac{1}{\cancel{f(x;\boldsymbol\theta)}}\cancel{f(x;\boldsymbol\theta)}\,dx = E[\frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta)] + E[\frac{\partial}{\partial\theta_j}w_i(\boldsymbol\theta)t_i(x)]$

swap $\int$ & $\frac{\partial}{\partial\theta_j}$

$\frac{\partial}{\partial\theta_j}\int f(x;\boldsymbol\theta)\,dx = \frac{\partial}{\partial\theta_j}1 = 0 = E[\frac{\partial}{\partial\theta_j}\log c(\boldsymbol\theta)] + E[\frac{\partial}{\partial\theta_j}w_i(\boldsymbol\theta)t_i(x)]$ ①

② $\mathrm{Var}(\sum_{i=1}^{k}\frac{w_i(\boldsymbol\theta)}{\partial\theta_j}t_i(X)) = -\frac{\partial^2}{\partial\theta_j^2}\log c(\boldsymbol\theta) - E(\sum_{i=1}^{k}\frac{\partial^2 w_i(\boldsymbol\theta)}{\partial\theta_j^2}t_i(X))$

$\frac{\partial^2}{\partial\theta_j^2}\log f(x;\boldsymbol\theta) = [\frac{\partial}{\partial\theta_j}\log f(x;\boldsymbol\theta)]' = [\frac{1}{f(x;\boldsymbol\theta)}\cdot\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)]'_{\theta_j}$ ② take derivative

$= \frac{\frac{\partial^2}{\partial\theta_j^2}f(x;\boldsymbol\theta)\cdot f(x;\boldsymbol\theta) - \cancel{\frac{\partial^2 f(x;\boldsymbol\theta)}{\partial\theta_j^2}}[\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)]^2}{[f(x;\boldsymbol\theta)]^2}$

$E(\frac{\partial^2}{\partial\theta_j^2}\log f(x;\boldsymbol\theta)) = \int\frac{\frac{\partial^2}{\partial\theta_j^2}f(x;\boldsymbol\theta)\cdot f(x;\boldsymbol\theta) - \cancel{\frac{\partial^2 f(x;\boldsymbol\theta)}{\partial\theta_j^2}}[\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)]^2}{[f(x;\boldsymbol\theta)]^2}\cdot f(x;\boldsymbol\theta)\,dx$ ③ take expectation

$(\frac{\frac{\partial}{\partial\theta_j}f(x;\boldsymbol\theta)}{f(x;\boldsymbol\theta)}) = \frac{\partial}{\partial\theta_j}\log f(x;\boldsymbol\theta)$

$= \int\frac{\partial^2}{\partial\theta_j^2}f(x;\boldsymbol\theta)\,dx - \int[\frac{\partial}{\partial\theta_j}\log f(x;\boldsymbol\theta)]^2 dx$

Location and scale family

location: $X\sim f(x-\mu)$ iff $X = Z+\mu$ with $Z\sim f(z)$

scale: $X\sim\frac{1}{\sigma}f(\frac{x}{\sigma})$ iff $X = \sigma Z$ with $Z\sim f(z)$

location-scale $X\sim\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$ iff $X = \sigma Z+\mu$. with $Z\sim f(z)$

- if $E(Z^2)<\infty$, 则 $E(X) = \sigma E(Z)+\mu$; $\mathrm{Var}(X) = \sigma^2\mathrm{Var}(Z)$
- standard in a location-scale family iff $\begin{cases}E(Z) = \int z f(z)\,dz = 0\\ \mathrm{Var}(Z) = \int z^2 f(z)\,dz = 1\end{cases}$

$\Rightarrow f(z)$ is the standard.

example: standard $f(z) = e^{-z}$ $z\geq 0$

$E(z) = \beta = 1$ $\mathrm{var}(z) = \beta^2 = 1$

$\therefore X = z+\mu. \Rightarrow z = x-\mu$

$f_X(x) = f_z(x-\mu) = e^{-(x-\mu)}$ $z = x-\mu\geq 0 \Rightarrow x\geq\mu.$