STK4011 and STK9011
Autumn 2016

Exponential families

Covers section 3.4

Ørnulf Borgan
Department of Mathematics
University of Oslo

Many pdfs or pmfs may be expressed on the form

$$f(x\,|\,\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) \qquad (3.4.1)$$

where $h(x) \geq 0$ and $t_1(x),....,t_k(x)$ are real-valued functions of x and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}),...,w_k(\boldsymbol{\theta})$ are real valued functions of the possibly vector-valued parameter $\boldsymbol{\theta} = (\theta_1,\theta_2,...,\theta_d), \quad d \leq k$

We say that (3.4.1) defines an exponential family of distributions

Example: normal distribution

Consider the normal pdf

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

This may be written

$$f(x\,|\,\mu,\sigma^2) = \underbrace{1}_{h(x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{c(\mu,\sigma)} \exp\left\{\underbrace{\frac{1}{\sigma^2}}_{w_1(\mu,\sigma)}\underbrace{\left(-\frac{x^2}{2}\right)}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\mu,\sigma)}\,\underbrace{x}_{t_2(x)}\right\}$$

which is of the form (3.4.1)

Example: exponential distribution

Consider the exponential pdf

$$f(x\,|\,\beta) = \begin{cases} \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

For all x this may be written

$$f(x\,|\,\beta) = \underbrace{I_{\{x>0\}}(x)}_{h(x)}\underbrace{\frac{1}{\beta}}_{c(\beta)}\exp\left(\underbrace{-\frac{1}{\beta}}_{w_1(\beta)}\underbrace{x}_{t_1(x)}\right)$$

which is of the form (3.4.1)

Example: binomial distribution

Consider the binomial pmf

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,\dots,n$$

For all x this may be written

$$f(x \mid p) = \underbrace{I_{\{0,1,\dots,n\}}(x)\binom{n}{x}}_{h(x)}\underbrace{(1-p)^n}_{c(p)} \exp\left\{ \underbrace{\log\left(\frac{p}{1-p}\right)}_{w_1(p)} \underbrace{x}_{t_1(x)} \right\}$$

which is of the form (3.4.1)

In chapter 7 we will prove the following result (which may also be known from earlier courses):

If X is a random variable with pdf or pmf $f(x \mid \boldsymbol{\theta})$, and if $\int f(x \mid \boldsymbol{\theta})\,dx$ or $\sum f(x \mid \boldsymbol{\theta})$ may be differentiated twice with respect to θ_j by changing the order of integration/summation and differentiation, then we have

$$\mathrm{E}\left(\frac{\partial}{\partial \theta_j} \log f(X \mid \boldsymbol{\theta})\right) = 0$$

$$\mathrm{Var}\left(\frac{\partial}{\partial \theta_j} \log f(X \mid \boldsymbol{\theta})\right) = -\mathrm{E}\left(\frac{\partial^2}{\partial \theta_j^2} \log f(X \mid \boldsymbol{\theta})\right)$$

If we use this result for the exponential family of distributions, we obtain:

Theorem 3.4.2
If X is a random variable with pdf or pmf of the form (3.4.1), then

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

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

Example: exponential distribution

We have the pdf

$$f(x \mid \beta) = I_{\{x>0\}}(x) \frac{1}{\beta} \exp\left(-\frac{1}{\beta} x\right)$$

We have $c(\beta) = \frac{1}{\beta}, \quad w_1(\beta) = -\frac{1}{\beta}, \quad t_1(x) = x$

By Theorem 3.4.2 it follows that

$$\mathrm{E}\left(\frac{1}{\beta^2} X\right) = \frac{1}{\beta} \qquad \mathrm{Var}\left(\frac{1}{\beta^2} X\right) = -\frac{1}{\beta^2} - \mathrm{E}\left(-\frac{2}{\beta^3} X\right)$$

This gives

$$\mathrm{E}X = \beta \qquad \mathrm{Var}\, X = \beta^2$$

Example: binomial distribution

We have the pmf

$$f(x \mid p) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}(1-p)^n \exp\left\{\log\left(\frac{p}{1-p}\right)x\right\}$$

Note that

$$c(p) = (1-p)^n, \quad w_1(p) = \log\left(\frac{p}{1-p}\right), \quad t_1(x) = x$$

$$\frac{d}{dp}\log c(p) = \frac{-n}{1-p}, \qquad \frac{d}{dp}w_1(p) = \frac{1}{p(1-p)}$$

$$\frac{d^2}{dp^2}\log c(p) = \frac{-n}{(1-p)^2}, \qquad \frac{d^2}{dp^2}w_1(p) = \frac{-(1-2p)}{p^2(1-p)^2}$$

By Theorem 3.4.2 we have

$$\mathrm{E}\left(\frac{1}{p(1-p)}X\right) = \frac{n}{1-p}$$

$$\mathrm{Var}\left(\frac{1}{p(1-p)}X\right) = \frac{n}{(1-p)^2} - \mathrm{E}\left(\frac{-(1-2p)}{p^2(1-p)^2}X\right)$$

This gives

$$\mathrm{E}X = np$$

$$\mathrm{Var}\,X = np(1-p)$$

So far we have used the parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$

The parameter space for (3.4.1) is (usually) given as

$$\Theta = \left\{\boldsymbol{\theta}: \int_{-\infty}^{\infty} h(x)\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right)dx < \infty\right\} = \{\boldsymbol{\theta}: c(\boldsymbol{\theta}) > 0\}$$

(in the discrete case the integral is replaced by a sum)

Sometimes an exponential family is reparametrized as

$$f(x \mid \boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta})\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)$$

Here the $h(x)$ and $t_i(x)$ functions are as before, and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ is the natural parameter

The natural parameter space is given as

$$\mathcal{H} = \left\{\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k): \int_{-\infty}^{\infty} h(x)\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)dx < \infty\right\}$$

Note that

$$c^*(\boldsymbol{\eta}) = \left[\int_{-\infty}^{\infty} h(x)\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)dx\right]^{-1}$$

Example: binomial distribution

We have the pmf

$$f(x\,|\,p) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}(1-p)^n \exp\left\{\log\left(\frac{p}{1-p}\right)x\right\}$$

$$= I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}c(p)\exp\{w_1(p)\,x\}$$

where $c(p) = (1-p)^n$ and $w_1(p) = \log\big(p/(1-p)\big)$

The original parameter space is $\{p:\ 0 < p < 1\} = (0,1)$

The natural parameter is $\eta = \log\big(p/(1-p)\big) \in (-\infty, \infty)$

Note that $p = e^\eta/(1+e^\eta)$ and that the pmf becomes

$$f(x\,|\,\eta) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}\left(\frac{1}{1+e^\eta}\right)^n \exp\{\eta x\}$$

Example: normal distribution

The normal pdf may be written

$$f(x\,|\,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(\frac{1}{\sigma^2}\left(-\frac{x^2}{2}\right) + \frac{\mu}{\sigma^2}x\right)$$

Usually the original parameter space is given by
$\{(\mu,\sigma^2):\ -\infty < \mu < \infty,\ \sigma^2 > 0\}$

The natural parameters are $\eta_1 = 1/\sigma^2$ and $\eta_2 = \mu/\sigma^2$

The natural parameter space is
$\{(\eta_1,\eta_2):\ \eta_1 > 0,\ -\infty < \eta_2 < \infty\}$

The pdf may be written

$$f(x\,|\,\eta_1,\eta_2) = \frac{\sqrt{\eta_1}}{\sqrt{2\pi}}\exp\left(-\frac{\eta_2^2}{2\eta_1}\right)\exp\left(\eta_1\left(-\frac{x^2}{2}\right) + \eta_2 x\right)$$

In (3.4.1) it is usually the case that the dimension of the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ is equal to k, i.e. $d = k$

Then we have a full exponential family

If $d < k$, we have a curved exponential family

In the example on the previous slide, we have a full exponential family with parameter space
$\{(\mu,\sigma^2):\ -\infty < \mu < \infty,\ \sigma^2 > 0\}$

If we consider a normal model with $\sigma^2 = k\mu^2$, for a known constant k, we get a curved exponential family with parameter space $\{(\mu, k\mu^2):\ -\infty < \mu < \infty\}$