University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Exponential families

A probability density function or probability mass function is called an exponential family if it can be expressed as

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

Note: $h(x), t_1(x), \ldots, t_k(x)$ do not depend on $\boldsymbol{\theta}$ and $c(\boldsymbol{\theta})$ does not depend of x.

Example:
Consider $X \sim b(n,p)$ with n fixed. Show that $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$ can be expressed in the exponential family form.

$$\begin{aligned}
p(x) &= \binom{n}{x}p^x(1-p)^{n-x} \\
&= \binom{n}{x}\left(\frac{p}{1-p}\right)^x (1-p)^n \\
&= \binom{n}{x}(1-p)^n e^{log(\frac{p}{1-p})^x} \\
&= \binom{n}{x}(1-p)^n e^{xlog(\frac{p}{1-p})}
\end{aligned}$$

Therefore this pmf is an exponential family with
$h(x) = \binom{n}{x}, c(p) = (1-p)^n, t_1(x) = x, w_1(p) = log\frac{p}{1-p}$.

Theorem:
Suppose a random variable X has a pdf or pmf that can be expressed in the form of exponential family. Then,

$$\text{(a)} \qquad E\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j}t_i(x)\right) = -\frac{\partial}{\partial \theta_j}logc(\boldsymbol{\theta}).$$

and

$$\text{(b)} \qquad var\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j}t_i(x)\right) = -\frac{\partial^2}{\partial \theta_j^2}logc(\boldsymbol{\theta}) - E\left(\sum_{i=1}^{k}\frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2}t_i(x)\right).$$

Note: Here log is the natural logarithm.

Proof of (a):

$$\begin{aligned} \int_x f(x|\boldsymbol{\theta})dx &= 1 \\ \int_x h(x)c(\boldsymbol{\theta})exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx &= 1 \end{aligned}$$

Differentiate both sides w.r.t. θ_j:

$$\begin{aligned} &\int_x h(x)\frac{\partial c(\boldsymbol{\theta})}{\partial\theta_j}exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx \\ + &\int_x h(x)c(\boldsymbol{\theta})\sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(x)exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx = 0 \end{aligned}$$

Multiply the first integral by $\frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\theta})}$ and note that $\frac{\partial logc(\boldsymbol{\theta})}{\partial\theta_j} = \frac{\partial c(\boldsymbol{\theta})}{\partial\theta_j}\frac{1}{c(\boldsymbol{\theta})}$.

$$\begin{aligned} &\int_x h(x)\frac{\partial c(\boldsymbol{\theta})}{\partial\theta_j}exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)\frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\theta})}dx \\ + &\int_x h(x)c(\boldsymbol{\theta})\sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(x)exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx = 0 \end{aligned}$$

After rearranging we get

$$\begin{aligned} &\int_x \sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(x)h(x)c(\boldsymbol{\theta})exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx = \\ &\qquad -\frac{\partial logc(\boldsymbol{\theta})}{\partial\theta_j}\int_x h(x)c(\boldsymbol{\theta})exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)\right)dx \end{aligned}$$

Or

$$E\left(\sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(x)\right) = -\frac{\partial}{\partial\theta_j}logc(\boldsymbol{\theta}).$$

To prove statement (b) of the theorem differentiate a second time and rearrange.