STA 623 – Fall 2010 – Dr. Charnigo

Section 3.4: Exponential Families

Families of probability density functions. For simplicity I will focus on continuous random variables and probability density functions in this lecture. There are parallel developments for discrete random variables and probability mass functions.

To make explicit that a probability density function depends on a (scalar or vector) parameter θ, I can use the notation $f(x;\theta)$. For instance, the exponential distribution with (scalar) rate parameter θ has probability density function $f(x;\theta) = 1_{\{x\in(0,\infty)\}}\theta\exp[-\theta x]$.

Let Θ denote the parameter space (i.e., the set of all possible values for θ). If $f(x;\theta_1) \equiv f(x;\theta_2)$ implies that $\theta_1 = \theta_2$, then we say that θ is identifiable. In this case, we get a different $f(x;\theta)$ for each $\theta \in \Theta$, so we have not one probability density function (assuming that Θ is not singleton) but rather a whole family of probability density functions. We may refer to the family, rather than to an individual probability density function within the family, as $\{f(x;\theta) : \theta \in \Theta\}$.

Example (families of probability density functions). Put $\Theta := (0,\infty)$ and consider the family $\{1_{\{x\in(0,\infty)\}}\theta\exp[-\theta x] : \theta \in \Theta\}$. Is θ identifiable? Can you exhibit another family for which θ is not identifiable?

Exponential families. We say that $\{f(x;\theta) : \theta \in \Theta\}$ is an exponential family if each of its members has the form

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

Above, functions of x only are not allowed to depend on θ and vice versa.

Example (exponential families). Put $\Theta := (0, \infty)$ and consider the family $\{1_{\{x \in (0,\infty)\}} \theta \exp[-\theta x] : \theta \in \Theta\}$. To verify that this is an exponential family, we note that $k = 1$ and

$h(x) =$ $\qquad$ $c(\theta) =$

$w_1(\theta) =$ $\qquad$ $t_1(x) =$

Now put $\Theta := (0, \infty)$ and consider the family $\{1_{\{x \in [0,\theta]\}} \theta^{-1} : \theta \in \Theta\}$. This is not an exponential family because we cannot express $1_{\{x \in [0,\theta]\}}$ as a product of a function of x with a function of θ. This example provides a useful principle: if the support set depends on θ, then we do not have an exponential family.

For one more example, put $\Theta := (0, \infty)$ and consider the family $\{1_{\{x \in (0,\infty)\}} \theta / (x + 1)^{\theta + 1} : \theta \in \Theta\}$. To verify that this is an exponential family, we note that $k = 1$ and

$h(x) =$ $\qquad$ $c(\theta) =$

$w_1(\theta) =$ $\qquad$ $t_1(x) =$

Moment calculations. If the probability density function of X is a member of an exponential family, then

$$E\left[\sum_{i=1}^{k} \frac{\partial}{\partial \theta_j} w_i(\theta) t_i(X)\right] = -\frac{\partial}{\partial \theta_j} \log c(\theta) \quad \text{and}$$

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

where θ_j denotes the j^{th} component of a vector θ. If θ is a scalar, then replace each partial differentiation in θ_j with an ordinary differentiation in θ.

Example (moment calculations). To see where the expected value formula comes from, note that

$$\int_{\mathbb{R}} h(x)c(\theta) \exp\left[\sum_{i=1}^{k} w_i(\theta)t_i(x)\right] \, dx = 1$$

for all $\theta \in \Theta$. As such,

$$\int_{\mathbb{R}} h(x) \exp\left[\sum_{i=1}^{k} w_i(\theta)t_i(x)\right] \, dx = \exp[-\log c(\theta)].$$

Assuming that we can interchange the order of differentiation and integration (we can, but the justification is delicate), we have

$$\int_{\mathbb{R}} h(x) \exp\left[\sum_{i=1}^{k} w_i(\theta)t_i(x)\right] \sum_{i=1}^{k} \frac{\partial}{\partial\theta_j} w_i(\theta)t_i(x) \, dx = -\frac{\partial}{\partial\theta_j} \log c(\theta) \exp[-\log c(\theta)].$$

Multiplying both sides by $c(\theta)$ yields

$$\int_{\mathbb{R}} f(x;\theta) \sum_{i=1}^{k} \frac{\partial}{\partial\theta_j} w_i(\theta)t_i(x) \, dx = -\frac{\partial}{\partial\theta_j} \log c(\theta).$$

To illustrate the expected value and variance formulas, put $\Theta := (0, \infty)$ and consider the family $\{1_{\{x \in (0,\infty)\}}\theta/(x+1)^{\theta+1} : \theta \in \Theta\}$. We have

$E[\log(1+X)] =$ and

$Var[\log(1+X)] =$

Curved and full exponential families. An exponential family in which the dimension of θ equals k (the number of summands inside the exponential) is called a full exponential family. An exponential family in which the dimension of θ is less than k is called a curved exponential family. The distinction between curved and full exponential families becomes important in the search for best unbiased estimators, as you will see if you take STA 607.