Exponential Families

Definition. $\{P_\theta, \theta \in \Theta\}$ with density (pdf or pmf) is a single-parameter exponential family if there exist real valued functions $c(\theta)$, $d(\theta)$ on Θ and $T(\mathbf{x})$ and $S(\mathbf{x})$ on $\mathcal{R}^n$ and a set $A \subset \mathcal{R}^n$ such that

$$f(\mathbf{x}|\theta) = \exp\left[c(\theta)T(\mathbf{x}) + d(\theta) + S(\mathbf{x})\right] I_A(\mathbf{x}),$$

where A must not depend on θ.

Example. $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Then $f(x|\lambda) = \exp(-\lambda)\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$. Take $A = \{0, 1, 2, \ldots\}$ and write the density as

$$f(x|\lambda) = \exp\left(x\log(\lambda) - \lambda - \log(x!)\right) I_A(x).$$

Choosing $c(\lambda) = \log(\lambda)$, $d(\lambda) = -\lambda$, $T(x) = x$ and $S(x) = -\log(x!)$ shows that $\text{Poisson}(\lambda)$, $\lambda > 0$ is a single-parameter exponential family of distributions.

Example. $X \sim U(0, \theta)$, $\theta > 0$. Then

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Is $U(0, \theta)$, $\theta > 0$ exponential family? Note that $f(x|\theta) = \exp(-\log(\theta))I_A(x)$, where the support of the density, $A = (0, \theta)$ depends on θ. It is not possible to express the density in the required exponential form on a common support of the entire family, so $U(0, \theta)$, $\theta > 0$ is not exponential family.

Example. $X \sim N(\theta, 1)$. Then

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(x-\theta)^2)I_A(x), \quad A = (-\infty, \infty),$$

which can be written as

$$\begin{aligned} f(x|\theta) &= \exp\left(-\frac{1}{2}[x^2 + \theta^2 - 2\theta x] - \log(\sqrt{2\pi})\right) I_A(x) \\ &= \exp\left(\theta x - \frac{\theta^2}{2} - [\frac{x^2}{2} + \log(\sqrt{2\pi})]\right) I_A(x). \end{aligned}$$

Choosing $c(\theta) = \theta$, $d(\theta) = -\theta^2/2$, $T(x) = x$ and $S(x) = -(x^2/2 + \log(\sqrt{2\pi}))$ we can show that $N(\theta, 1)$, $-\infty < \theta < \infty$ is a single-parameter exponential family.

Consider $X_1, \ldots, X_m$ i.i.d P_θ with density $f(x|\theta)$. Suppose $\{P_\theta, \theta \in \Theta\}$ is exponential family. i.e., $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))I_A(x)$. Then

$$\begin{aligned} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m|\theta) &= \prod_{i=1}^m \exp(c(\theta)T(x_i) + d(\theta) + S(x_i))I_A(x_i) \\ &= \exp(c(\theta)\sum_{i=1}^m T(x_i) + md(\theta) + \sum_{i=1}^m S(x_i))I_{A^m}(x_1, \ldots, x_m). \end{aligned}$$

Therefore, $(X_1, \ldots, X_m)$ has distribution belonging to a single-parameter exponential family.

Result. If $\{P_\theta, \theta \in \Theta\}$ is a single-parameter exponential family with density $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))I_A(x)$, then $T(X)$ is sufficient for θ. (Actually minimal sufficient if Θ contains an open interval, as shown later.)

Simply note that

$$f(x|\theta) = \exp(c(\theta)T(x) + d(\theta)) \exp(S(x))I_A(x).$$

Thus we have $g(t, \theta) = \exp(c(\theta)t + d(\theta))$ and $h(x) = \exp(S(x))I_A(x)$. Combining this result with above we get the following.

Corollary. If $X_1, \ldots, X_m$ are i.i.d P_θ with density $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))I_A(x)$, the $\sum_{i=1}^m T(X_i)$ is sufficient for θ.

Example. $X \sim$ Bernoulli(θ). Then

$$\begin{aligned} f(x|\theta) &= \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x) \\ &= \exp\left(x \log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta)\right) I_{\{0,1\}}(x), \end{aligned}$$

so $T(X) = X$ is sufficient for θ. Now consider $X_1, \ldots, X_n$ i.i.d Bernoulli(θ). Then using the Corollary (or otherwise) observe that $\sum_{i=1}^n T(X_i) = \sum_{i=1}^n X_i$ is sufficient for θ.

Theorem. Let $\{P_\theta, \theta \in \Theta\}$ be a one-parameter exponential family with density $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))I_A(x)$. Suppose that either P_θ is discrete, or $T(X)$ has a continuous distribution. Then the family of distributions $\{Q_\theta\}$ for $T(X)$ is also a one-parameter exponential family, and has density $q(t|\theta) = \exp(c(\theta)t + d(\theta) + S^*(t))I_{A^*}(t)$.

Proof. Discrete case:

$$\begin{aligned} q(t|\theta) &= P_\theta\left(T(X)=t\right) = \sum_{\{x:T(x)=t\}} f(x|\theta) \\ &= \sum_{\{x:T(x)=t\}} \exp(c(\theta)T(x)+d(\theta)+S(x))I_A(x) \\ &= \exp(c(\theta)t+d(\theta))\left\{\sum_{\{x\in A:T(x)=t\}} \exp(S(x))\right\} I_{A^*}(t), \end{aligned}$$

where $A^* = \{t : t = T(x), x \in A\}$. Now define

$$S^*(t) = \begin{cases} \log\sum_{\{x\in A:T(x)=t\}}\exp(S(x)) & \text{if } t\in A^*; \\ 0 & \text{otherwise.} \end{cases}$$

The continuous case is similar.