Lectures 6-9

3 Exponential Family

[CB3.4, BD1.6]
Binomial and normal distributions have the property that the dimension of a sufficient statistic is independent of the sample size. We would like to identify and define a broad class of models that have this and other desirable properties.

Definition 1 *Let $\{f(x;\theta) : \theta \in \Theta\}$ be a family of pdf's (or pmf's). We assume that the set $\{\mathbf{x} : f(\mathbf{x};\theta) > 0\}$ is independent of θ, where $\mathbf{x} = (x_1, \cdots, x_n)$. We say that the family $\{f(x;\theta) : \theta \in \Theta\}$ is a k–parameter exponential family if there exist real–valued functions $Q_1(\theta), \cdots, Q_k(\theta)$ and $D(\theta)$ on Θ and $T_1(\mathbf{X}), \cdots, T_k(\mathbf{X})$ and $S(\mathbf{X})$ on $\mathbb{R}^n$ such that*

$$f(x;\theta) = exp(\sum_{i=1}^{k} Q_i(\theta)T_i(\mathbf{x}) + D(\theta) + S(\mathbf{x})).$$

We can express the k–parameter exponential family in canonical form for a natural $kx1$ parameter vector $\eta = (\eta_1, \cdots, \eta_k)'$ as

$$f(\mathbf{x};\eta) = h(\mathbf{x})c(\eta)exp(\sum_{i=1}^{k} \eta_i T_i(\mathbf{x})),$$

We define the natural parameter space as the set of points $\eta \in W \subset \mathbb{R}^k$ for which the integral $\int_{\mathbb{R}^n} exp(\sum_{i=1}^{k} \eta_i T_i(\mathbf{x}))h(\mathbf{x})d\mathbf{x}$ is finite.
We shall refer to T as a natural sufficient statistic.
Ex: Verify that Binomial and Normal belong to exponential family.
Uniform distribution U([0, θ]), $\theta \in \mathbb{R}^+$ does not belong to the exponential family, since its support depends on θ
If the probability distribution of X_1 belongs to an exponential family, the probability distribution of $(X_1, \cdots, X_n)$ also belongs to the same exponential family, where X_i are iid with distribution same as X_1.

Theorem 1 *Suppose $X_1, \cdots, X_n$ is a random sample from pdf or pmf $f_X(x|\theta)$ where $f_X(x|\theta) = h(x)d(\theta)exp(\sum_{i=1}^{k} w_i(\theta)t_i(x))$ is a member of an exponential family. Define a statistic $T(X)$ by $T(X) = (\sum_{j=1}^{n} t_1(X_j), \cdots, \sum_{j=1}^{n} t_k(X_j))$. The distribution of $T(X)$ is an exponential family of the form $f_T(u_1, \cdots, u_k|\theta) = H(u_1, \cdots, u_k)[d(\theta)]^n \exp(\sum_{i=1}^{k} w_i(\theta)u_i)$*

Theorem 2 (3.4.2 of CB) *If X is a random variable with pdf/pmf as in definition 1 then, for every j,*

$$\mathrm{E}(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(\mathbf{X})) = -\frac{\partial}{\partial \theta_j} D(\theta)$$

$$\text{Var}(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(\mathbf{X})) = -\frac{\partial^2}{\partial {\theta_j}^2} D(\theta) - \text{E}(\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial {\theta_j}^2} t_i(\mathbf{X}))$$

Ex: Use this to derive the mean and variance of the binomial and normal distributions.

Theorem 3 *If the distribution of X belongs to a canonical exponential family and η is an interior point of W, the mgf of Texists and is given by*

$$M(s) = c(\eta)/c(s+\eta)]$$

for s in some neighbourhood of 0.

Ex: Use this to derive the mean and variance of the natural sufficient statistic of Raleigh distribution

$$p(x,\theta) = (x/\theta^2)exp(-x^2/2\theta^2), x > 0, \theta > 0.$$

In an exponential family, if the dimension of Θ is k (there is an open set subset of $\mathbb{R}^k$ that is contained in Θ), then the family is a full exponential family.
Otherwise the family is a curved exponential family.
An example of a full exponential family is $\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0$.

Example 1 *An example of a curved exponential family is $\mathcal{N}(\mu, \mu^2), \mu \in \mathbb{R}$.*

Curved exponential families arise naturally in applications of CLT as approximation to binomial $\sigma^2 = p(1-p)/n$ or Poisson $\sigma^2 = \lambda/n$.

Theorem 4 *In the exponential family given by definition 1 and the set Θ contains an open subset of $\mathbb{R}^k$ then $(T_1(\mathbf{X}), \cdots, T_k(\mathbf{X}))$ is complete.*

Ex: In the curved exponential family of example 1, $k = 2$ and the set $Theta$ does not contain an open subset of $\mathbb{R}^2$. So we cannot apply the above theorem. Is it still true that $T(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is complete?
Ex: Show that the Cauchy family is not an exponential family.
Ex: Multinomial is a $(k-1)$ parameter exponential family.
Ex: Linear Regression model is 3 parameter exponential family.
Ex: Logistic regression model is 2-parameter exponential family.

Definition 2 *An exponential family is of rank k iff the natural sufficient statistic T is k-dimensional and $(1, T_1(X), \cdots, T_k(X))$ are linearly independent with positive probability. Formally, $P[\sum_{j=1}^k a_j T_j(X) = a_{k+1}] < 1$ unless all a_j are 0.*

Ex: multinomial is rank $k-1$.
Ex: Logistic with n=1 is rank 1 and θ_1 and θ_2 are not identifiable. For $n \geq 2$, the rank is 2.
The following theorem establishes the relation between rank and identifiability.

Theorem 5 *Suppose $\mathcal{P} = q(x, \eta); \eta \in W$ is a canonical exponential family generated by (T_{kxl}, h) with natural parameter space W such that W is open. Let $A(\eta) = -log(c(\eta))$. Then the following are equivalent.*

1. *$\mathcal{P}$ is of rank k.*
2. *η is a parameter (identifiable).*
3. *Var(T) is positive definite.*
4. *$\eta \to \dot{A}(\eta)$ is 1-1 on E*
5. *A is strictly convex in E.*

Ex: Multivariate normal. Show that this family is full rank and E is open.