Exponential Family and Related Properties

A. One-Parameter Exponential Family (Natural Exponential Family):

One-parameter exponential family, if density function can be written as

\[ f(x \mid \theta) = h(x) \exp \{\eta(\theta)T(x) - B(\theta)\}, \]

where \( h(x) \) and \( B(\theta) \) are known functions, \( \eta(\theta) \) is a natural parameter.

Note:

(1) The support of the distribution (i.e., \( \{x : f(x \mid \theta) > 0\} \)) cannot depend on \( \theta \). If the support of \( f(x \mid \theta) \) is different for different \( \theta \), then the family is not an exponential family. For example, a uniform distribution does NOT belong to exponential family.

(2) The functions \( \eta(\theta) \) and \( T(x) \) are not unique. Since \( \eta \) can be multiplied by any nonzero constant \( c \), while \( T \) is multiplied by \( 1/c \) simultaneously.

Canonical Form for One-parameter Exponential Family:

Re-parameterize model in terms of \( \eta \) instead of \( \theta \), then

\[ f(x \mid \eta) = h(x) \exp \{\eta T(x) - A(\eta)\}. \]

We define the natural parameter space for \( \eta \) as

\[ \Omega_\eta = \{\eta : \int h(x) \exp \{\eta T(x)\} dx < \infty \}. \]

Some Properties (B&D, P. 49-52):

(1) In one-parameter exponential family, the random variable \( T(X) \) is sufficient for \( \theta \).

(2) The pdf of \( T(X) \) also belongs to one-parameter exponential family satisfying
\[ f(t; \theta) = h^*(t) \exp \{ \eta(\theta)t - B(\theta) \} . \]

(3) If \( \{X_i\} \) are iid r.v.’s from one-parameter exponential family, then the joint pdf for \( X = (X_1, ..., X_n) \) also belongs to one-parameter exponential family with the sufficient statistic \( T(X) = \sum_{i=1}^{n} T(X_i) \).

**Theorem (B&D, P. 52):** Based on the canonical form, the moment generating function of \( T(X) \) exists and is given by
\[
M(s) = \exp \{ A(s + \eta) - A(\eta) \},
\]
for \( s \) in some neighborhood of 0.
Moreover,
\[
E(T(X)) = A'(\eta), \ Var(T(X)) = A''(\eta).
\]

**Example:** Let \( \eta = \ln \lambda \), the Poisson family in canonical form is
\[
f(x | \eta) = \frac{1}{x!} \exp \{ \eta x - e^\eta \},
\]
Therefore, \( E(X) = e^\eta = \lambda, \ Var(X) = e^\eta = \lambda \).

**B. Multi-Parameter Exponential Family (i.e., \( \theta \) in \( R^k \)):**
\[
f(x | \theta) = h(x) \exp \left\{ \sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta) \right\}
\]

**Note:** The three properties of one-parameter exponential family still hold in multi-parameter exponential family.

(1) The vector \( T(X) = (T_1(X), ..., T_k(X))^\prime \) is sufficient.

(2) \( T(X) \) still form \( k \)-parameter exponential family.

(3) If \( X = (X_1, X_2, \cdots, X_n) \) and \( \{X_i\} \) are iid r.v. from a \( k \)-parameter exponential family, then \( T(X) = \left( \sum_{i=1}^{n} T_1(X_i), ..., \sum_{i=1}^{n} T_k(X_i) \right)^\prime \) is sufficient for \( \theta \).
 Canonical Form for a $k$-parameter Exponential Family:

$$f(x \mid \eta) = h(x) \exp \{T'(x) \eta - A(\eta)\},$$

and the natural parameter space is

$$\Omega_{\eta} = \{\eta = (\eta_1, \ldots, \eta_k)' : \int h(x) \exp \{T'(x) \eta\} \, dx < \infty\}.$$

Note: In general we assume that no linear constraint exists for $T$ and $\eta$, otherwise the dimension could be further reduced. (Consider the relation between minimal sufficient statistics and exponential family.)

Examples: $X = (X_1, \ldots, X_n)$ are iid r.v.’s from $N(\mu, \sigma^2)$.

1. $\sigma^2$ known to be $\sigma_0^2$,

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-x_i^2}{2\sigma_0^2} - \frac{\mu^2}{2\sigma_0^2} + \frac{\mu x_i}{\sigma_0^2} \right\},$$

The natural parameter and sufficient statistic is $\eta = \frac{\mu}{\sigma_0}$, $T(X) = \frac{1}{n} \sum X_i$.

Note: $\eta$ and $T$ also could be written as $\eta = \mu$, $T(X) = \frac{\sum X_i}{\sigma_0^2}$. (not unique!)

2. $\mu$ known to be $\mu_0$,

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-\mu_0^2}{2\sigma^2} - \frac{x_i^2}{2\sigma^2} + \frac{\mu_0 x_i}{\sigma^2} \right\}$$

Then $\eta = -\frac{1}{2\sigma_0}$, $T(X) = \sum (X_i - \mu_0)^2$.

Note: We could NOT set $\eta_i = -\frac{1}{2\sigma^2}$, $T_1 = \sum X_i^2$, and $\eta_2 = \frac{\mu_0}{\sigma^2}$, $T_2 = \sum X_i$, since $\eta_1$ and $\eta_2$ are linear dependent.

3. Both of $\mu$ and $\sigma^2$ are unknown,

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-\mu^2}{2\sigma^2} - \frac{x_i^2}{2\sigma^2} + \frac{\mu x_i}{\sigma^2} \right\},$$
Then, \( \eta = \left( \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right) \), \( T(X) = \left( \sum X_i, \sum X_i^2 \right) \).

Theorem (B&D, P. 59): For a canonical \( k \)-parameter exponential family with \( f(x | \eta) = h(x) \exp \{ T'(x) \eta - A(\eta) \} \), \( \eta \in \Omega_\eta \), we have

(a) \( \Omega_\eta \) is convex.

(b) \( A(): \Omega_\eta \to R \) is convex.

(c) If \( \eta_0 \in \Omega_\eta^0 \) for some nonempty interior \( \Omega_\eta \) in \( R^k \), then \( T(X) \) has a moment generating function satisfying

\[
M(s) = \exp \{ A(\eta_0 + s) - A(\eta_0) \},
\]

for all \( s \) such that \( \eta_0 + s \in \Omega_\eta \). Since \( \eta_0 \) is an interior point this set of \( s \) includes a ball about 0.

Corollary: Under the conditions of above theorem,

\[
E_\eta T(X) = \left( \frac{\partial A(\eta)}{\partial \eta_i}, \ldots, \frac{\partial A(\eta)}{\partial \eta_k} \right)_{\eta = \eta_0}, \quad Var_\eta T(X) = \left[ \frac{\partial^2 A}{\partial \eta_i \partial \eta_j}(\eta) \right]_{\eta = \eta_0}.
\]

Example: For a multinomial distribution with \( \eta_i = \log(p_i/p_0) \),

\[
E[T_j(x)] = \frac{\partial}{\partial \eta_j} n \log \left[ 1 + \sum_{i=1}^{k-1} e^{\eta_i} \right] = \frac{n e^{\eta_0}}{1 + \sum_{i=1}^{k-1} e^{\eta_i}} = \frac{n p_i}{1 + \sum_{i=1}^{k-1} p_i} = \frac{n p_i}{P_k} = n p_i,
\]

\[
Cov_\eta [T_i(x), T_j(x)] = \frac{\partial}{\partial \eta_j} \frac{\partial}{\partial \eta_k} n \log \left[ 1 + \sum_{i=1}^{k-1} e^{\eta_i} \right] = \frac{-n e^{\eta_0} e^{\eta_j}}{(1 + \sum_{i=1}^{k-1} e^{\eta_i})^2} = -n p_i p_j, \quad i \neq j,
\]

\[
Var_\eta [T_j(x)] = \frac{\partial^2}{\partial \eta_j^2} n \log \left[ 1 + \sum_{i=1}^{k-1} e^{\eta_i} \right] = np_i (1 - p_i).
\]
C. Curved Exponential Family:

\[ f(x \mid \theta) = h(x) \exp \left\{ \sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta) \right\}, \]

where \( \text{dim}(\theta) < k \) (i.e., there are some constraints in \( \theta \)).

**Example:** Let \( X_1, \ldots, X_n \) are iid r.v.’s from \( N(\theta, \theta^2) \):

\[
 f(x \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{ -\frac{(x_i - \theta)^2}{2\theta^2} \right\} = \left( \frac{1}{\sqrt{2\pi\theta}} \right)^n \exp\left\{ -\frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2 + \frac{1}{\theta} \sum_{i=1}^{n} x_i - \frac{n}{2} \right\}.
\]

This is an exponential family with \( \eta = \left( -\frac{1}{2\theta^2}, -\frac{1}{\theta} \right) \), and the parameter space is a curve.

![Parameter space for theta between 1 and 10](image)

D. Other Important Theorems and Properties:

**Theorem (Minimal sufficient statistic in the exponential family):**

Let \( X_1, \ldots, X_n \) be iid r.v.’s from an exponential family with the pdf

\[
 f(x \mid \theta) = h(x) \exp \left\{ \sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta) \right\}.
\]

If \( \{\eta_1(\theta), \ldots, \eta_k(\theta)\} \) is a linear independent set, then \( \left( \sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right) \) is a minimal sufficient statistic for \( \theta \).
Theorem (Complete statistics in the exponential family, C&B P.288): 

Let \( X_1, ..., X_n \) be iid r.v.’s from an exponential family with the pdf

\[
f(x \mid \boldsymbol{\theta}) = h(x) \exp \left\{ \sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) T_j(x) - B(\boldsymbol{\theta}) \right\}.
\]

If \( \{\eta_1(\boldsymbol{\theta}), ..., \eta_k(\boldsymbol{\theta})\} \) contains an open set in \( \mathbb{R}^k \), then \( \left( \sum_{i=1}^{n} T_1(X_i), ..., \sum_{i=1}^{n} T_k(X_i) \right) \) is a complete statistic for \( \boldsymbol{\theta} \).

Example: Assume \( \{X_i\} \) are iid\( N(\mu, \sigma^2) \) and \( X = (X_1, ..., X_n) \). Then

\[
f(x \mid \mu, \sigma) = h(x)c(\mu, \sigma)\exp\left\{ \frac{\mu}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 \right\}, \quad \text{where} \quad \eta_1 = \frac{\mu}{\sigma^2}, \quad \eta_2 = -\frac{1}{2\sigma^2},
\]

\( \Omega_\eta = (-\infty, \infty) \times (-\infty, 0) \) contains an open set in \( \mathbb{R}^2 \). Thus \( \left( \sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 \right) \) is complete.

Example: Assume \( \{X_i\} \) are iid\( N(\theta, \theta^2) \), \( \theta > 0 \) and \( X = (X_1, ..., X_n) \). Since

\[
\eta = \left(-\frac{1}{2\theta^2}, \frac{1}{\theta}\right)
\]

is a linear independent set, thus \( \left( \sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 \right) \) is minimal sufficient. But \( \Omega_\eta \) doesn’t contain an open set in \( \mathbb{R}^2 \), so we can’t apply the theorem.

However, we can take \( g(X) = \bar{X} - cS \) such that \( E_\theta(g(X)) = 0 \), where

\[
c = \frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}, \quad S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}.
\]

Thus \( \left( \sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 \right) \) is NOT complete.

Theorem (C&B P.289): 

If a minimal sufficient statistic exists, then any complete sufficient statistic is also a minimal sufficient statistic.