1 Exponential Family

Let

- μ be a measure on $\mathbb{R}^n$,
- $h : \mathbb{R}^n \mapsto [0, \infty)$ be a non-negative function,
- η_1, ..., η_s be measurable functions from $\mathbb{R}^p$ to $\mathbb{R}$,
- T_1, ..., T_s be measurable functions from $\mathbb{R}^n$ to $\mathbb{R}$,
- let $\theta \in \mathbb{R}^p$ be a $p \times 1$ column vector.

Definition 1 (Exponential family)**.** Suppose that

$$K(\theta) \quad \equiv \quad \log \int \exp\left\{\sum_{i=1}^{s} \eta_i(\theta) T_i(x)\right\} h(x)\, d\mu(x) < \infty.$$

The family of densities $\{p(\cdot \mid \theta) : \theta \in \Theta\}$ is called an s-parameter (or s-dimensional [maybe better]) exponential family , where

$$p(x \mid \theta) \quad = \quad \exp\left\{\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - K(\theta)\right\} h(x).$$

We can also do a reparametrization and define $\eta = \eta(\theta) \in \mathbb{R}^s$.

Definition 2 (Exponential family in Canonical Form)**.** Suppose that

$$A(\eta) \quad \equiv \quad \log \int \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x)\right\} h(x)\, d\mu(x) < \infty.$$

The family of densities $\{p(\cdot \mid \eta) : \eta \in \mathcal{N}\}$ is called an s-dimensional exponential family in canonical form, where

$$p(x \mid \eta) \quad = \quad \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right\} h(x),$$

η is called a natural parameter, and $\mathcal{N}$ is called the natural parameter space.

Example 1 (Examples of Exponential Family)**.**

1. Gaussian $N\left(\mu, \sigma^2\right)$ with $\theta = \left(\mu, \sigma^2\right)$, $\mu \in R$, $\sigma^2 \in (0, \infty)$:

$$p(x \mid \theta) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{\mu^2}{2\sigma^2}\right\},$$

   where $T_1 = x$, $\eta_1(\theta) = \frac{\mu}{\sigma^2}$, $T_2 = x^2$, $\eta_2(\theta) = -\frac{1}{2\sigma^2}$, $K(\theta) = \frac{1}{2}\log\left(2\pi\sigma^2\right) + \frac{\mu^2}{2\sigma^2}$, and $h(x) = 1$.

2. Bernoulli(θ) with $\theta \in (0, 1)$:

$$P(X = x \mid \theta) \quad = \quad \theta^x (1-\theta)^{1-x} = \exp\left\{x \log\theta + (1-x)\log(1-\theta)\right\},$$

   where $T_1 = x$, $\eta_1(\theta) = \log\theta$, $T_2 = 1 - x$, $\eta_2(\theta) = \log(1-\theta)$, $K(\theta) = 1$, and $h(x) = 1$.

3. If X_1, ..., X_n are iid $N\left(\mu, \sigma^2\right)$, their joint density is

$$p\left(x_1, ..., x_n \mid \theta\right) = \exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{n\mu^2}{2\sigma^2}\right\},$$

where $T_1 = \sum_{i=1}^{n} x_i$, $\eta_1\left(\theta\right) = \frac{\mu}{\sigma^2}$, $T_2 = \sum_{i=1}^{n} x_i^2$, $\eta_2\left(\theta\right) = -\frac{1}{2\sigma^2}$, $K\left(\theta\right) = \frac{1}{2}\log\left(2\pi\sigma^2\right) + \frac{\mu^2}{2\sigma^2}$, and $h\left(x\right) = 1$.

Several important properties of exponential family is in the following theorem.

Theorem 1. *Consider a k-parameter exponential family with natural parameter η. Let $\mathcal{Z}$ be an open subset of the natural parameter space and the function $\phi\left(x\right)$ to be integrable, i.e.,*

$$\int \left|\phi\left(x\right)\right| p\left(x \mid \eta\right) d\mu\left(x\right) < \infty,$$

for all $\eta \in \mathcal{Z}$. Then

$$\frac{\partial \int \phi\left(x\right) p\left(x \mid \eta\right) d\mu\left(x\right)}{\partial \eta} = \int \frac{\partial \phi\left(x\right) p\left(x \mid \eta\right)}{\partial \eta} d\mu\left(x\right),$$

that is, the derivative can be computed under integration sign.

An importance application of the theorem is to compute the moments of $T\left(x\right)$.

1. Let $\phi\left(x\right) \equiv 1$, we get

$$\mathrm{E}\left[T_i\left(x\right) \mid \eta\right] = \frac{\partial A\left(\eta\right)}{\partial \eta_i}.$$

2. If we take one more derivative, we get

$$\begin{aligned}\mathrm{cov}\left[T_i\left(X\right), T_j\left(X\right) \mid \eta\right] &= \frac{\partial^2 A\left(\eta\right)}{\partial \eta_i \partial \eta_j},\\ \mathrm{var}\left[T_i\left(X\right) \mid \eta\right] &= \frac{\partial^2 A\left(\eta\right)}{\partial \eta_i^2}.\end{aligned}$$

Example 2 (Examples of Moments of Exponential Family)**.**

We reparametrize Bernoulli $\left(\theta\right)$ with $\theta \in (0,1)$ to $\eta = \log\frac{\theta}{1-\theta} \in \mathbb{R}$, such that

$$P\left(X = x \mid \eta\right) = \exp\left\{x\eta - \log\left(1 + e^{\eta}\right)\right\}.$$

where $T\left(x\right) = x$, and $A\left(\eta\right) = \log\left(1 + e^{\eta}\right)$. Hence,

$$\begin{aligned}\mathrm{E}\left[X \mid \eta\right] &= \frac{\partial \log\left(1 + e^{\eta}\right)}{\partial \eta} = \frac{e^{\eta}}{1 + e^{\eta}} = \theta,\\ \mathrm{Var}\left[X \mid \eta\right] &= \frac{\partial^2 \log\left(1 + e^{\eta}\right)}{\partial \eta^2} = \frac{e^{\eta}}{\left(1 + e^{\eta}\right)^2} = \theta\left(1 - \theta\right).\end{aligned}$$

If we want the derivative with respect to θ, we can apply the chain rule to the natural parameter.

Theorem 2 (Convex set)**.** *The natural parameter space $\mathcal{N}$ is a convex set.*

Example 3 (Uniqueness in T and s)**.** We need to note that the functions $\{T\}$ and $\{\eta\}$ are not unique. The number s is not unique either. Bernoulli $\left(\theta\right)$ with $\theta \in (0,1)$:

$$P\left(X = x \mid \theta\right) = \theta^x \left(1 - \theta\right)^{1-x} = \exp\left\{x\log\theta + \left(1 - x\right)\log\left(1 - \theta\right)\right\},$$

where $T_1 = x$, $\eta_1\left(\theta\right) = \log\theta$, $T_2 = 1 - x$, $\eta_2\left(\theta\right) = \log\left(1 - \theta\right)$, $K\left(\theta\right) = 1$, and $h\left(x\right) = 1$, that is 2-dimensional exponential family. But x and $1 - x$ are not linearly independent.

1. We can write $P(X = x)$ as

$$P(X = x \mid \theta) \quad = \quad \exp\left\{x \log \frac{\theta}{1-\theta} + \log(1-\theta)\right\},$$

where $T(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$, $K(\theta) = -\log(1-\theta)$, and $h(x) = 1$. That is, 1-dimensional Exponential family, but not in canonical form.

2. If we reparametrize $\eta = \log \frac{\theta}{1-\theta} \in \mathbb{R}$, then

$$P(X = x \mid \eta) \quad = \quad \exp\{x\eta - \log(1 + e^{\eta})\}$$

where $T(x) = x$, and $A(\eta) = \log(1 + e^{\eta})$. That is, 1-dimensional Exponential family, in canonical form.

Definition 3 (Minimal or strictly s-dimensional exponential family)**.** If s in $\sum_{i=1}^{s} \eta_i(\theta) T_i(x)$ cannot be reduced, then we say the exponential family is minimal or it is a strictly s-dimensional exponential family.

Minimal exponential family essentially means that the functions 1, η_1, ..., η_s are linearly independent and the statistics T_1, ..., T_s are affine independent (i.e., a positive definite covariance matrix), i.e.,

$$\sum_{i=1}^{s} \lambda_i \eta_i(\theta) = \lambda_0 \quad \Rightarrow \quad \lambda_i = 0, \forall i,$$

$$\sum_{i=1}^{s} \lambda_i T_i(x) = \lambda_0 \quad \Rightarrow \quad \lambda_i = 0, \forall i.$$

There is no special reason to keep the exponential family non-minimal.

Definition 4 (Full rank and curved exponential family)**.** Consider a strictly s-dimensional exponential family in canonical form

$$p(x \mid \eta) \quad = \quad \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right\} h(x),$$

where $\eta \in \Delta$ is the natural parameter and Δ is the natural parameter space. If Δ contains an s-dimensional open set, then the family is said to be full rank. Otherwise, it is a curved exponential family.

Curved exponential family typically means the exponential family is minimal and $\{\eta_i\}$ are affine independent, but they are related in a nonlinear way. That is, we cannot reduce s, but the dimension of θ is lower than s.

Example 4 (Examples of Curved Exponential Family)**.**

Gaussian $N(\theta, \theta^2)$ with $\theta > 0$:

$$p(x \mid \theta) \quad = \quad \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{-\frac{(x-\theta)^2}{2\theta^2}\right\} = \exp\left\{\frac{1}{\theta}x - \frac{1}{2\theta^2}x^2 - \frac{1}{2}\log\left(2\pi\theta^2\right) - \frac{1}{2}\right\},$$

where $T_1 = x$, $\eta_1(\theta) = \frac{1}{\theta}$, $T_2 = x^2$, $\eta_2(\theta) = -\frac{1}{2\theta^2}$, $K(\theta) = \frac{1}{2}\log\left(2\pi\theta^2\right) + \frac{1}{2}$, and $h(x) = 1$. We only have one parameter θ, but we need $s = 2$ [cannot be reduced further].