Exponential Families

An exponential family is a parameterized family of distributions

Single-Parameter Exponential Family

For some $X \subset \mathbb{R}$, a parameterized family of distributions over the alphabet Y is a **single parameter exponential family** if it can be expressed in the form:

$$p_y(y; x) = \exp\left(\lambda(x)t(y) - \alpha(x) + \beta y\right)$$

for some choice of functions λ, t, β

- λ: natural parameter
- t: natural statistic
- β: log base function

α is not part of the specification because it is for normalization

- A family may become no longer possible to be normalized for specific x

$$Z(x) = e^{\alpha(x)} = \sum_y \exp\left(\lambda(x)t(y) + \beta(y)\right)$$

- $Z(x)$ is known as the partition function
- $\alpha(x)$ is then known as the log-partition function

We denote that y follows an exponential family by:

$$y \sim p \in \mathcal{E}(\mathcal{X}, \mathcal{Y}, \lambda, t, \beta)$$

We note that if we instead factor out $\beta(y)$ to get:

$$p_y(y; x) \propto q(y)\exp\left(\lambda(x)t(y)\right)$$

- If we can get $q(y)$ to be a distribution (integrate to 1) then this is known as the **base distribution** of the family

Canonical Single-Parameter Exponential Family

A **canonical exponential family** is one with $\lambda(x) = x$

Natural Parameter Set

The **natural parameter space** is the collection of all possible x such that the corresponding form of the distribution can be normalized

- This is a convex set, which we will see shortly
- When $\mathcal{Y}$ is finite, it is an open set

Natural Exponential Families

Canonical exponential families with $\mathcal{Y} \subset \mathbb{R}$ and $t(y) = y$ are referred to as natural exponential families

Log-Partition Function Properties

Property 1

The log-partition function for a general exponential family satisfies:

$$\frac{d}{dx}\alpha(x) = \frac{d}{dx}\lambda(x) \cdot \mathbb{E}\left[t(y)\right]$$

Property 2

The log-partition function for a general exponential family satisfies:

$$\frac{d^2}{dx^2}\alpha(x) = \left(\frac{d}{dx}\lambda(x)\right)^2 \cdot \frac{d^2}{dx^2}\lambda(x) \cdot \mathrm{Var}\left[t(y)\right]$$

If we have a canonical exponential family, this simplies to just $\mathrm{Var}\left[t(y)\right]$

- This implies the log-partition function is convex since the variance is nonnegative
- This then verifies that the natural parameter space is convex for canonical exponential families

Property 3

The score function for a general exponential family satisfies:

$$S(y;x) = \frac{\partial}{\partial x}\ln p_y(y;x) = \frac{d}{dx}\lambda(x) \cdot (t(y) - \mathbb{E}\left[t(y)\right])$$

Property 4

The Fisher information for a general exponential family satisfies:

$$J_y(x) = \mathbb{E}\left[S(y;x)^2\right] = \left(\frac{d}{dx}\lambda(x)\right)^2 \cdot \mathrm{Var}\left[t(y)\right]$$

Property 5

For a general exponential family, we have:

$$\frac{d}{dx}\mathbb{E}\left[t(y)\right] = \frac{d}{dx}\lambda(x) \cdot \mathrm{Var}\left[t(y)\right]$$

This allows Property 4 to equivalently be expressed as:

$$J_y(x) = \frac{d}{dx}\lambda(x) \cdot \frac{d}{dx}\mathbb{E}\left[t(y)\right]$$

Exponential Family Constructions and Interpretations

Many familiar parameterized distributions can be expressed as exponential families:

- Bernoulli random variable
- Binomial random variable
- Guassian random variable
- Exponential random variable

The uniform random variable is **not** exponential

Geometric Mean of Distributions

Geometric Mean

COnsider two strictly positive probability distributions p_1 and p_2, the **weighted geometric mean of these distributions is defined for all $x \in [0,1]$:

$$p_y(y;x) = \frac{p_1(y)^x p_2(y)^{1-x}}{Z(x)}$$

This is an exponential family with:

- $\lambda(x) = x$
- $t(y) = \ln p_1(y) / p_2(y)
- $\beta(y) = \ln p_2(y)$
- $\alpha(x) = \ln Z(x)$

Any canonical exponential family over a finite alphabet $\mathcal{Y}$ can be expressed as the geometric mean of two distributions

- To see this, suppose we have a family specified by t and β
- Then choose $p_2(y) = c_1 e^{\beta(y)}$ and $p_1(y) = c_2 p_2(y) e^{t(y)}$
- This implies that given two distributions p_1 and p_2:
  - There is a single-parameter canonical exponential family that includes both as members
    - This is because we could weight the geometric mean to just get one or the other
  - That single-parameter canonical exponential family is unique
  - This is summarized in the following theorem

Geometric Mean Characterization of Canonical Exponential Families

Let $\mathcal{P}$ denote a single-dimensional family of distributions over $\mathcal{Y}$ for all $x \in \mathcal{X}$, where $\mathcal{X}$ is convex. Then $\mathcal{P}$ is a canonical exponential family iff for every $p_1, p_2, p_3 \in \mathcal{P}$ there is some $\lambda \in \mathbb{R}$ such that:

$$p_3(y) = \frac{p_1(y)^\lambda p_2(y)^{1-\lambda}}{Z(\lambda)}$$

Tilting Distributions

Given a base distribution q over an alphabet $\mathcal{Y} \subset \mathbb{R}$ we refer to $p_y(\cdot; x)$ as a tilted distribution where:

$$p_y(y; x) = \frac{q(y) e^{xt(y)}}{Z(x)}$$

It is a member of an exponential family with:

- $\lambda(x) = x$
- $t(y) = t(y)$
- $\beta(y) = \ln q(y)$
- $\alpha(x) = \ln Z(x)$

Any canonical exponential family with a finite alphabet $\mathcal{Y}$ can be expressed as a tilted distribution

Efficient Estimators

Exponential Fmaily Characterization of Efficient Estimators

An efficient estimator exists for estimating a nonrandom parameter x from observations y iff the model $p_y(y; x)$ is a member of an exponential family such that:

$$\frac{d}{dx}\lambda(x) = cJ_y(x)$$

for some constant $c > 0$.

When it exist, the efficient estimator takes the form for some constant b:

$$\hat{x} = ct(y) + b$$

which must also be the ML estimator (as shown previously)

Corollary

If the model is from a canonical exponential family, then an efficient estimator exists iff the Fisher information $J_y(x)$ is constant and does not depend on the parameter x