Families of Distributions

Andrew Nobel

September, 2024

Overview

Definition: A *distribution family* is an indexed collection $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ of densities or probability mass functions on a common sample space $\mathcal{X}$. Examples

- Location families
- Scale families
- Location-Scale families
- Exponential families
- Transformation families

Location and Scale Families

Location and Scale Families

Fact: Let V have density g. If $a \in \mathbb{R}$ and $b > 0$ then $X := bV + a$ has density

$$f(x) = \frac{1}{b}\, g\left(\frac{x-a}{b}\right)$$

Location Families

Definition: The *location family* generated by a density g is given by

$$\mathcal{P} = \{f(x|\theta) = g(x - \theta) : \theta \in \mathbb{R}\}$$

This is the set of densities of $X = V + \theta$ where $V \sim g$.

Ex: The location family generated by $\mathcal{N}(0, \sigma^2)$ is $\mathcal{P} = \{\mathcal{N}(\theta, \sigma^2) : \theta \in \mathbb{R}\}$

Ex: The location family generated by U$(0, 1)$ is $\mathcal{P} = \{$U$(\theta, \theta + 1) : \theta \in \mathbb{R}\}$

Scale Families

Definition: The *scale family* generated by a density g is given by

$$\mathcal{P} = \left\{ f(x|\theta) = \frac{1}{\theta}\, g\Big(\frac{x}{\theta}\Big) : \theta > 0 \right\}$$

This is the set of densities of $X = \theta V$ where $V \sim g$.

Ex: The scale family generated by $g = \mathcal{N}(0,1)$ is $\mathcal{P} = \{\mathcal{N}(0,\theta^2) : \theta > 0\}$

Ex: The scale family generated by $g = \mathsf{U}(0,1)$ is $\mathcal{P} = \{\mathsf{U}(0,\theta) : \theta > 0\}$

Location-Scale Families

Definition: The *location-scale* family generated by a density g is

$$\mathcal{P} = \left\{ f(x|\theta) = \frac{1}{b}\, g\left(\frac{x-a}{b}\right) : \theta = (a,b) \in \mathbb{R} \times (0,\infty) \right\}$$

This is the set of densities of $X = bV + a$ where $V \sim g$

Ex: The location-scale family generated by $\mathcal{N}(0,1)$ is the standard normal family

$$\mathcal{P} = \{\mathcal{N}(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$$

Ex: The location-scale family generated by $\mathsf{U}(0,1)$ is the standard uniform family

$$\mathcal{P} = \{\mathsf{U}(a,b) : -\infty < a < b < \infty\}$$

Product Families

Product Families

Definition: Given a family $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ of densities on $\mathcal{X}$ and $n \geq 1$ define the product family of densities on $\mathcal{X}^n$ by

$$\mathcal{P}^n = \left\{ h(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) : \theta \in \Theta \right\}$$

Here $x = (x_1, \ldots, x_n)$ denotes a point in $\mathcal{X}^n$

Note

- $\mathcal{P}^n$ is the family of densities associated with iid observations $X_1, \ldots, X_n$ having common distribution $f(\cdot|\theta) \in \mathcal{P}$
- $\mathcal{P}$ and $\mathcal{P}^n$ have different sample spaces, but the same parameter space

Exponential Families

Exponential Families

Family of probability density/mass functions with simple exponential structure

Attraction

- Flexible, includes many widely used families of distributions
- Nice theoretical properties, also useful in practice
- Close connections with physics, probability, and mathematics

Exponential Families

Setting: Sample space $\mathcal{X}$. Parameter space $\Theta \subseteq \mathbb{R}^d$

Definition: A family $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ of pmf's or pdf's on $\mathcal{X}$ is an *exponential family* if for each $x \in \mathcal{X}$ and $\theta \in \Theta$

$$f(x|\theta) = c(\theta)\, h(x) \exp\left\{ \sum_{j=1}^{k} \eta_j(\theta)\, T_j(x) \right\}$$

- Base function $h : \mathcal{X} \to [0, \infty)$ (depends only on data)
- Sufficient statistic $T : \mathcal{X} \to \mathbb{R}^k$, write as $T(x) = (T_1(x), \ldots, T_k(x))$
- Parametrization $\eta : \Theta \to \mathbb{R}^k$, write as $\eta(\theta) = (\eta_1(\theta), \ldots, \eta_k(\theta))$
- Normalization function $c : \Theta \to (0, \infty)$ (sometimes called a partition function)

Exponential Families: Features and Definitions

Recall: Exponential density $f(x|\theta) = c(\theta)\, h(x) \exp\left\{\sum_{j=1}^{k} \eta_j(\theta) T_j(x)\right\}$

- Note $f(x|\theta) > 0$ iff $h(x) > 0$ so all densities have same support
- Data x and parameter θ linked only through sum $\sum_{j=1}^{k} \eta_j(\theta) T_j(x)$
- Dimension k of functions η and T called the *order* of the family $\mathcal{P}$
- Assume $\mathcal{P}$ is *minimal*: no linear relationships between $\eta_1, \ldots, \eta_k$ or $T_1, \ldots, T_k$
- Recall $d =$ dimension of parameter space Θ. If $d = k$ family $\mathcal{P}$ said to be *full*. If $d < k$ family $\mathcal{P}$ said to be *curved*
- Natural parameter space consists of those $\theta \in \Theta$ such that $\int h(x) \exp\left\{\sum_{j=1}^{k} \eta_j(\theta) T_j(x)\right\} dx < \infty$ so that $c(\theta)$ is well defined

Exponential Families

Examples

- Binomial and Multinomial
- Geometric and Negative Binomial
- Poisson
- Normal
- Gamma, including Exponential and Chi-squared
- Beta

Non-Examples

- Uniform
- t-distributions

Identities for the Derivative of a Log-Likelihood

Given: Family $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ with $\Theta \subseteq \mathbb{R}^d$. Note $\int f(x|\theta)\, dx = 1$ for each θ, so

$$\frac{\partial}{\partial \theta_j} \int f(x|\theta)\, dx = 0 \quad \text{and} \quad \frac{\partial^2}{\partial \theta_j \partial \theta_l} \int f(x|\theta)\, dx = 0$$

Fact: If one can exchange the order of integration and differentiation in the first identity

$$\mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_j} \log f(X|\theta) \right] \;=\; 0$$

If one can exchange the order of integration and differentiation in both identities

$$\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f(X|\theta) \right] \;=\; -\mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_j} \log f(X|\theta) \cdot \frac{\partial}{\partial \theta_l} \log f(X|\theta) \right]$$

Moments and Covariances of Sufficient Statistics

Idea: For exponential families, we can find moments and covariances of the sufficient statistics $T_j(X)$ by differentiating the normalizing function $c(\theta)$

Fact: Let $\mathcal{P}$ be an exponential family with $d = k$. Then for $j = 1, \ldots, k$

$$\sum_{l=1}^{k} \frac{\partial \eta_l(\theta)}{\partial \theta_j} \mathbb{E}_\theta T_l(X) \;=\; -\frac{\partial \log c(\theta)}{\partial \theta_j}$$

In the special case $\eta_j(\theta) = \theta_j$ (called the natural parametrization) we have

$$\mathbb{E}_\theta T_j(X) \;=\; -\frac{\partial \log c(\theta)}{\partial \theta_j} \quad \text{and} \quad \mathrm{Cov}_\theta(T_j(X), T_l(X)) \;=\; -\frac{\partial^2 \log c(\theta)}{\partial \theta_j \partial \theta_l}$$

Joint and Conditional Distribution of Sufficient Statistics

Fact: Let $\mathcal{P}$ be an exponential family with sufficient statistics $T = (T_1, \ldots, T_k)$

- If $X \sim f(\cdot|\theta) \in \mathcal{P}$ the joint distribution of $T(X)$ has the exponential form

$$g(y|\theta) \;=\; \mathbb{P}_\theta\,(T(X) = y) \;=\; c(\theta)\,h'(y)\exp\left\{\sum_{j=1}^{k} \eta_j(\theta)\,y_j\right\}$$

  where $h'(y) = \int_{x:T(x)=y} h(x)\,dx$

- If $X \sim f(\cdot|\theta) \in \mathcal{P}$ and $S \subseteq \{1, \ldots, k\}$ the conditional distribution

$$\mathbb{P}_\theta(T_S(X) = y_S\,|\,T_{S^c}(X) = y_{S^c}) \;=\; c'(\theta)\,h'(y)\exp\left\{\sum_{j \in S} \eta_j(\theta)\,y_j\right\}$$

  where for $u \in \mathbb{R}^k$ and $A \subseteq \{1, \ldots, k\}$ we define $u_A = (u_j : j \in A)$ to be the restriction of u to the components in A

Independent Samples from an Exponential Family

Setting: Let $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ be an exponential family and let $X_1, \ldots, X_n \in \mathcal{X}$ be iid with $X_i \sim f(\cdot|\theta)$. The joint distribution of $X_1, \ldots, X_n$ at $x = (x_1, \ldots, x_n)$ is

$$g(x|\theta) \;=\; \prod_{i=1}^{n} f(x_i|\theta) \;=\; c(\theta)^n \prod_{i=1}^{n} h(x_i) \exp\left\{ \sum_{j=1}^{k} \eta_j(\theta) \left(\sum_{i=1}^{n} T_j(x_i) \right) \right\}$$

Upshot: Product family $\mathcal{P}^n$ is an exponential family $\tilde{\mathcal{P}}$ with

- Sample space $\tilde{\mathcal{X}} = \mathcal{X}^n$ and parameter space $\tilde{\Theta} = \Theta$
- Base function $\tilde{h}(x) = \prod_{i=1}^{n} h(x_i)$
- Sufficient statistics $\tilde{T}_j(x) = \sum_{i=1}^{n} T_j(x_i)$ for $j = 1, \ldots, k$
- Parametrization $\tilde{\eta}_j(\theta) = \eta_j(\theta)$ for $j = 1, \ldots, k$
- Normalization function $\tilde{c}(\theta) = c(\theta)^n$