Lec 5. Distribution family

Zhang W.P.

Contents

5.1 Exponential family 1
5.1.1 Canonical Exponential Family 4
5.2 Location and Scale Families 11

Learning objectives

After this section, you should be able to ...

- understand the definition of exponential family and its properties
- list common distributions in exponential family
- know the location-scale distribution family

5.1 Exponential family

Exponential family comprises a set of flexible distribution ranging both continuous and discrete random variables. The members of this family have many important properties which merits discussing them in some general format. Many of the probability distributions are specific members of this family:

- Gaussian: $\mathbb{R}^p$
- Multinomial: categorical
- Binomial: counts of success/failure
- Poisson: $\mathbb{N}^+$
- Laplace: $\mathbb{R}^+$
- Exponential: $\mathbb{R}^+$
- Beta: (0,1)
- Weibull: $\mathbb{R}^+$
- Weishart: symmetric positive-definite matrices

A family $\{F_\theta, \theta \in \Theta\}$ of distributions forms an (k-dimensional) exponential family if the distributions F_θ have densities of the form:

$$f(x,\theta) = C(\theta) \exp\left[\sum_{i=1}^{k} Q_i(\theta) T_i(x)\right] h(x) \qquad (5.1)$$

Def.

where $C(\theta) > 0$, Q_i are functions which map θ to $\mathbb{R}$; $h(x) > 0$ and $T_i(x)$ are statistics.

- The support of exponential family,

$$G(x) = \{x : f(x,\theta) > 0\} = \{x : h(x) > 0\}$$

  is irrelevant with the parameter θ

The One-Parameter Case

Consider a sample: $X_1, \ldots, X_n$ iid $\sim f$, where $f \in \mathcal{F} = \{f(x,\theta), \theta \in \Theta \subset \mathbb{R}\}$ is a one-parameter exponential family distribution with density/pmf function

$$f(x,\theta) = h(x)C(\theta)\exp\{Q(\theta)T(x)\} \tag{5.2}$$

The sample $\mathbf{X} = (X_1, \ldots, X_n)$ is a random vector with density/pmf:

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \left(h\left(x_i\right) C(\theta) \exp\left[Q(\theta)T\left(x_i\right)\right]\right) \\ &= C^n(\theta)\left[\prod_{i=1}^{n} h\left(x_i\right)\right] \times \exp\left[Q(\theta)\sum_{i=1}^{n} T\left(x_i\right)\right] \\ &= h^*(x)C^*(\theta)\exp\left\{Q^*(\theta)T^*(\mathbf{x}) - B^*(\theta)\right\} \end{aligned}$$

where: $h^*(x) = \prod_{i=1}^{n} h\left(x_i\right), Q^*(\theta) = Q(\theta), T^*(\mathbf{x}) = \sum_{i=1}^{n} T\left(x_i\right), C^*(\theta) = C^n(\theta)$

5.1.1 Canonical Exponential Family

The One-Parameter Case

- Re-parametrize setting $\eta = Q(\theta)$ in (5.2)—the Natural Parameter
- The density has the form

$$f(x, \eta) = h(x)A(\eta)\exp\{\eta T(x)\}$$

- The function $A(\eta)$ replaces $C(\theta)$ and is defined as the normalization constant:

$$A^{-1}(\eta) = \int \cdots \int h(x)\exp\{\eta T(x)]dx$$

  if X is continuous or

$$A^{-1}(\eta) = \sum_{x \in \mathcal{X}} h(x)\exp\{\eta T(x)]$$

  if X is discrete

- The Natural Parameter Space $\mathcal{E} = \{\eta : 0 < A(\eta) < \infty\}$ is a convex set.

The One-Parameter Case

Theorem 1. *Suppose X is distributed according to a canonical exponential family, i.e., the density/pmf function is given by*

$$f(x,\eta) = h(x)\exp[\eta T(x) - B(\eta)], \text{ for } x \in \mathcal{X} \subset R^q$$

If η_0 is an interior point of $\mathcal{E}$, the natural parameter space, then

- *The **moment generating function** of $T(X)$ exists and is given by*

$$M_T(s) = E\left[e^{sT(X)}\right] = exp(B(s+\eta_0) - B(\eta_0))$$

  for s in some neighborhood of 0 .

- $E[T(X)] = B'(\eta_0)$
- $\text{Var}[T(X)] = B''(\eta_0)$

The Multi-parameter Case

- Let the model (5.1) be indexed by $\eta = (Q_1, \ldots, Q_k)^{\mathrm{T}}$ rather than θ. Thus, the canonical k-parameter exponential family generated by $\mathbf{T}$ and h is

$$f(x, \boldsymbol{\eta}) = h(x)A(\boldsymbol{\eta}) \exp\left\{\eta^{\mathrm{T}}\mathbf{T}(x)\right\}, x \in \mathcal{X} \subset R^q \tag{5.3}$$

  where $\mathbf{T}(x) = (T_1(x), \ldots, T_k(x))^{\mathrm{T}}$ and, in the continuous case,

$$A^{-1}(\eta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x) \exp\left\{\eta^{\mathrm{T}}\mathbf{T}(x)\right\} dx$$

  In the discrete case, $A(\eta)$ is defined in the same way except integrals over R^q are replaced by sums.

- The natural parameter space

$$\mathcal{E} = \left\{\boldsymbol{\eta} \in R^k : 0 < A(\boldsymbol{\eta}) < \infty\right\}$$

  is a convex set.

The Multi-parameter Case

Theorem 2. *If the distribution of X belongs to a canonical exponential family (5.3) and η is an interior point of $\mathcal{E}$, then for any bounded integrable function $g(x)$, the following function*

$$G(\eta) = \int \cdots \int g(x) \exp\left\{\eta^{\mathrm{T}} \mathbf{T}(x)\right\} h(x) dx$$

satisfy

$$\begin{aligned} \frac{\partial^m G(\eta)}{\partial \eta_1^{m_1} \cdots \partial \eta_k^{m_k}} &= \int \cdots \int \frac{\partial^m}{\partial \eta_1^{m_1} \cdots \partial \eta_k^{m_k}} \left[g(x) \exp\left\{\eta^{\mathrm{T}} \mathbf{T}(x)\right\} h(x)\right] dx \\ &= \int \cdots \int T_1^{m_1} \cdots T_k^{m_k} \left[g(x) \exp\left\{\eta^{\mathrm{T}} \mathbf{T}(x)\right\} h(x)\right] dx \end{aligned}$$

where $m_1 + \cdots + m_k = m$.

A curved exponential family is a family of densities of the form (5.1) for which the dimension of the vector θ is equal to $d < k$. If $d = k$, the family is a full exponential family. Def.

e.g. $N(\mu, \mu^2), N_2(0, 0, 1, 1, \rho)$ are curved exponential families.

Sufficiency and completeness

Theorem 3. *If a sample X is distributed according to a full-rank exponential family, then $T(x) = (T_1(x), \ldots, T_k(x))'$ is minimal sufficient.*

In an exponential family, it turns out that not only is the statistic T minimal sufficient, it is complete.

Theorem 4. *If a sample X is distributed according to a full-rank exponential family with minimal sufficient statistic $T(X) = (T_1(X), \ldots, T_s(X))'$, and the natural parameter space contains open subset of $\mathbb{R}^k$, then T is complete.*

5.2 Location and Scale Families

Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$, and any $\sigma > 0$, the family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the location-scale family with standard pdf $f(x)$; μ is called the location parameter and σ is called the scale parameter. Def.

if $Z \sim f(z)$, $X \sim 1/\sigma f((x-\mu)/\sigma)$,

- $X \stackrel{d}{=} \sigma Z + \mu$
- $EX = \sigma EZ + \mu, Var(X) = \sigma^2 Var(Z)$
- location family: $\{f(x-\mu), \mu \in \mathbb{R}\}$
- scale family: $\{f(x/\sigma)/\sigma, \sigma > 0\}$

(Location family ancillary statistic) Let $X_1, \ldots, X_n$ be iid observations from a location parameter family with cdf $F(x-\theta), -\infty < \theta < \infty$. Then $R = X_{(n)} - X_{(1)}$, is an ancillary statistic.

↑Exam

↓Exam

(Scale family ancillary statistic) Let $X_1, \ldots, X_n$ be iid observations from a scale parameter family with cdf $F(x/\sigma), \sigma > 0$. Then $X_1/X_n, \ldots, X_{n-1}/X_n$ is an ancillary statistic.

↑Exam

↓Exam