§3 Exponential Families

§3.1 Introduction

3.1.1 **Definition.** A family of distributions $\{\mathbb{P}_\theta : \theta \in \Theta\}$ on a sample space $\mathcal{S}$ (free of θ) is an *exponential family* if its probability functions are of the form

$$f(\boldsymbol{x}|\theta) \propto \begin{cases} h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^k \pi_j(\theta) t_j(\boldsymbol{x})\right\}, & \boldsymbol{x} \in \mathcal{S}, \\ 0, & \text{otherwise.} \end{cases}$$

Without loss of generality, assume $h(\boldsymbol{x}) > 0$ for $\boldsymbol{x} \in \mathcal{S}$.

3.1.2 **Examples:**

(i) *Normal*, $N(\mu, \sigma^2)$:

$\mathcal{S} = \mathbb{R}, \quad \theta = (\mu, \sigma^2), \quad \Theta = (-\infty, \infty) \times (0, \infty)$

$f(x|\theta) = f(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-(x-\mu)^2/(2\sigma^2)\}$

$\pi_1(\theta) = -1/(2\sigma^2), \ \pi_2(\theta) = \mu/\sigma^2, \quad t_1(x) = x^2, \ t_2(x) = x, \quad h(x) = 1$

(ii) *Poisson* (λ):

$\mathcal{S} = \{0, 1, 2, \ldots\}, \quad \theta = \lambda, \quad \Theta = (0, \infty)$

$f(x|\lambda) = \exp(-\lambda)\lambda^x/x!$

$\pi_1(\lambda) = \ln\lambda, \quad t_1(x) = x, \quad h(x) = 1/x!$

(iii) The *Cauchy* pdf

$$f(x|\theta) = \frac{1}{\pi\{1 + (x-\theta)^2\}}$$

is **not** of exponential family form.

§3.2 Natural parameters and natural statistics

3.2.1 **Definitions.** Given exponential family $f(\boldsymbol{x}|\theta) \propto h(\boldsymbol{x}) \exp\{\sum_{j=1}^k \pi_j(\theta) t_j(\boldsymbol{x})\}\mathbf{1}\{\boldsymbol{x} \in \mathcal{S}\}$, $\theta \in \Theta$,

(i) **(natural parameter space)**

the set $\Pi = \{\boldsymbol{\pi} = (\pi_1(\theta), \pi_2(\theta), \ldots, \pi_k(\theta)) : \theta \in \Theta\} \subset \mathbb{R}^k$ is the *natural parameter space* for the exponential family;

(ii) **(natural parameter)**

the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k) \in \Pi$ is called the *natural parameter* of the exponential family;

(iii) **(natural statistic)**

for a random variable $\boldsymbol{X}$ drawn from $f(\boldsymbol{x}|\theta)$, the statistic $\big(t_1(\boldsymbol{X}),\ldots,t_k(\boldsymbol{X})\big)$ is called the *natural statistic*.

Note: Natural parameter and its associated natural statistic have the **same** dimension $(=k)$.

3.2.2 The exponential family of probability functions

$$f(\boldsymbol{x}|\theta) \propto h(\boldsymbol{x})\exp\Big\{\sum_{j=1}^{k}\pi_j(\theta)t_j(\boldsymbol{x})\Big\}\mathbf{1}\{\boldsymbol{x}\in\mathcal{S}\},\quad \theta\in\Theta,$$

is <u>equivalent</u> to the exponential family of probability functions (with slight abuse of notation)

$$f(\boldsymbol{x}|\boldsymbol{\pi}) \propto h(\boldsymbol{x})\exp\Big\{\sum_{j=1}^{k}\pi_j t_j(\boldsymbol{x})\Big\}\mathbf{1}\{\boldsymbol{x}\in\mathcal{S}\},\quad \boldsymbol{\pi}=(\pi_1,\ldots,\pi_k)\in\Pi.$$

Therefore we may **re-parameterize** an exponential family (indexed by θ) by its natural parameter $\boldsymbol{\pi}$.

3.2.3 Let $X_1,\ldots,X_n$ be i.i.d. under an exponential family

$$p(x|\boldsymbol{\pi}) \propto h(x)\exp\Big\{\sum_{j=1}^{k}\pi_j t_j(x)\Big\}\mathbf{1}\{x\in\mathcal{S}\}.$$

Joint probability function of $\boldsymbol{X}=(X_1,\ldots,X_n)$ is

$$\begin{aligned} f(\boldsymbol{x}|\boldsymbol{\pi}) = \prod_{i=1}^{n} p(x_i|\boldsymbol{\pi}) &\propto \prod_{i=1}^{n}\Big[h(x_i)\exp\Big\{\sum_{j=1}^{k}\pi_j t_j(x_i)\Big\}\mathbf{1}\{x_i\in\mathcal{S}\}\Big] \\ &= h(x_1)\cdots h(x_n)\exp\Big\{\sum_{j=1}^{k}\pi_j T_j(\boldsymbol{x})\Big\}\mathbf{1}\{\boldsymbol{x}\in\mathcal{S}^n\}, \end{aligned}$$

where $T_j(\boldsymbol{x})=\sum_{i=1}^{n} t_j(x_i)$.

Clearly the above joint probability function $f(\boldsymbol{x}|\boldsymbol{\pi})$ is also of exponential family form.

For this family, the natural parameter $\boldsymbol{\pi}$ and natural parameter space Π are the same as those for $p(x|\boldsymbol{\pi})$. The natural statistic is

$$\big(T_1(\boldsymbol{X}),\ldots,T_k(\boldsymbol{X})\big) = \Big(\sum_{i=1}^{n} t_1(X_i)\,,\ \ldots\ ,\ \sum_{i=1}^{n} t_k(X_i)\Big).$$

3.2.4 **Examples:**

(i) $N(\mu, \sigma^2)$ — (for $\mu \in \mathbb{R}$ and $\sigma > 0$)

Natural parameter: $\boldsymbol{\pi} = (\pi_1, \pi_2) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)$

Natural parameter space: $\Pi = (-\infty, 0) \times (-\infty, \infty)$

Natural statistic: $\left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ $\left[\text{for an i.i.d. sample } X_1, \ldots, X_n \sim N(\mu, \sigma^2)\right]$

(ii) Poisson (λ) — (for $\lambda > 0$)

Natural parameter: $\boldsymbol{\pi} = \pi_1 = \ln \lambda$

Natural parameter space: $\Pi = (-\infty, \infty)$

Natural statistic: $\sum_{i=1}^n X_i$ $\left[\text{for an i.i.d. sample } X_1, \ldots, X_n \sim \text{Poisson } (\lambda)\right]$

§3.3 Distributional properties of natural statistic

3.3.1 **Theorem.** Suppose $\boldsymbol{X}$ has probability function of exponential family form with natural statistic $\boldsymbol{T} = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ and natural parameter $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$. Then $\boldsymbol{T}$ has probability function also of exponential family form with the same natural parameter.

. . . .

Proof: [Assume $\boldsymbol{X}$ is discrete. A more general proof is outlined in Appendix A3.1.]

Let $\mathcal{S}(\boldsymbol{t}) = \mathcal{S}(t_1, \ldots, t_k) = \{\boldsymbol{x} \in \mathcal{S} : t_1(\boldsymbol{x}) = t_1, \ldots, t_k(\boldsymbol{x}) = t_k\}$.

Then joint probability function $g(\boldsymbol{t}|\boldsymbol{\pi})$ *of* $\boldsymbol{T}$ *is:*

$$g(\boldsymbol{t}|\boldsymbol{\pi}) = \sum_{\boldsymbol{x} \in \mathcal{S}(\boldsymbol{t})} f(\boldsymbol{x}|\boldsymbol{\pi}) \propto \exp\Big(\sum_{j=1}^k \pi_j t_j\Big) \sum_{\boldsymbol{x} \in \mathcal{S}(\boldsymbol{t})} h(\boldsymbol{x}) = h^*(\boldsymbol{t}) \exp\Big(\sum_{j=1}^k \pi_j t_j\Big),$$

which is also of exponential family form with the same natural parameter $\boldsymbol{\pi}$. ∎

3.3.2 **Theorem.** Suppose $\boldsymbol{X}$ has probability function of exponential family form with natural statistic $\boldsymbol{T} = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ and natural parameter $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$. Then any subset of components of $\boldsymbol{T}$ conditional on values of the rest has probability function also of exponential family form with natural parameter being the corresponding subvector of $\boldsymbol{\pi}$.

. . . .

Proof: W.l.o.g. we show that the probability function of $\boldsymbol{T}_1 = (t_1(\boldsymbol{X}), \dots, t_r(\boldsymbol{X}))$ *conditional on* $\boldsymbol{T}_2 = (t_{r+1}(\boldsymbol{X}), \dots, t_k(\boldsymbol{X}))$ *is of exponential family form with natural parameter* $\boldsymbol{\pi}^{(1)} = (\pi_1, \dots, \pi_r)$.

Note that $\boldsymbol{T}$ *has probability function* $g(\boldsymbol{t}|\boldsymbol{\pi}) \propto h^*(\boldsymbol{t}) \exp\left(\sum_{j=1}^k \pi_j t_j\right)$. *Let* $g_2(t_{r+1}, \dots, t_k|\boldsymbol{\pi})$ *be the marginal probability function of* $\boldsymbol{T}_2$. *Then* $\boldsymbol{T}_1$ *conditional on* $\boldsymbol{T}_2 = (t_{r+1}, \dots, t_k)$ *has probability function*

$$g(t_1, \dots, t_r|t_{r+1}, \dots, t_k, \boldsymbol{\pi}) = \frac{g(t_1, \dots, t_k|\boldsymbol{\pi})}{g_2(t_{r+1}, \dots, t_k|\boldsymbol{\pi})} \propto g(t_1, \dots, t_k|\boldsymbol{\pi}) \propto h^*(\boldsymbol{t}) \exp\Big(\sum_{j=1}^r \pi_j t_j\Big),$$

which has probability function of exponential family form with natural parameter $\boldsymbol{\pi}^{(1)}$. ▮

3.3.3 **Example.**

Two independent samples: $(X_1, \dots, X_m)$ i.i.d. $\sim$ Poisson (λ), $(Y_1, \dots, Y_n)$ i.i.d. $\sim$ Poisson (μ)

Joint mass function of the two samples:

$$\begin{aligned} f(x_1, \dots, x_m, y_1, \dots, y_n|\lambda, \mu) &= \left(\prod_{i=1}^m \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)\left(\prod_{j=1}^n \frac{e^{-\mu}\mu^{y_j}}{y_j!}\right) \\ &\propto \big(\prod_i x_i! \prod_j y_j!\big)^{-1} \exp\big\{(\ln\lambda)\sum_i x_i + (\ln\mu)\sum_j y_j\big\}. \end{aligned}$$

Theorem §3.3.1 implies that the natural statistic $\left(\sum_i X_i, \sum_j Y_j\right)$ *has a distribution of exponential family form, with natural parameter* $(\ln\lambda, \ln\mu)$.

Note: elementary probability arguments show that $\sum_i X_i$ and $\sum_j Y_j$ are independent Poisson $(m\lambda)$ and Poisson $(n\mu)$ random variables, respectively, so that $(T_1, T_2) \equiv \left(\sum_i X_i, \sum_j Y_j\right)$ has joint mass function

$$f(t_1, t_2|\lambda, \mu) = \left\{\frac{e^{-m\lambda}(m\lambda)^{t_1}}{t_1!}\right\}\left\{\frac{e^{-n\mu}(n\mu)^{t_2}}{t_2!}\right\} \propto \frac{m^{t_1}n^{t_2}}{t_1!\,t_2!} \exp\left(t_1 \ln\lambda + t_2 \ln\mu\right).$$

Suppose we wish to test $H_0 : \lambda = \mu$ vs $H_1 : \lambda > \mu$.

Joint mass function of the two samples:

$$\begin{aligned} f(x_1, \dots, x_m, y_1, \dots, y_n|\lambda, \mu) &\propto \big(\prod_i x_i! \prod_j y_j!\big)^{-1} \exp\big\{(\ln\lambda)\sum_i x_i + (\ln\mu)\sum_j y_j\big\} \\ &= h(x_1, \dots, x_m, y_1, \dots, y_n) \exp\Big\{\pi_1 \sum_i x_i + \pi_2\big(\sum_i x_i + \sum_j y_j\big)\Big\}, \end{aligned}$$

where $\pi_1 = \ln(\lambda/\mu)$*,* $\pi_2 = \ln \mu$*. It is of exponential family form. Rewrite* H_0, H_1 *in terms of* π_1, π_2 *—*

$$H_0 : \pi_1 = 0 \qquad vs \qquad H_1 : \pi_1 > 0.$$

We see that the hypotheses only involve π_1*, so that* π_2 *is a* nuisance parameter*.*

Theorem §3.3.2 says that the conditional distribution of $\sum_i X_i$ *given* $\sum_i X_i + \sum_j Y_j = t$ *depends on* π_1, π_2 *only through* π_1 *and hence we are rid of the* nuisance *parameter* π_2*.*

$$\left[\text{In fact, } \sum_i X_i \Big| \sum_i X_i + \sum_j Y_j = t \;\sim\; \text{binomial}\left(t, \frac{m\lambda}{m\lambda + n\mu}\right) \equiv \text{binomial}\left(t, \frac{me^{\pi_1}}{me^{\pi_1} + n}\right).\right]$$

Intuitively we would reject H_0 *in favour of* H_1 *if* $\sum_i X_i$ *is large. Under* H_0*, the conditional distribution of* $\sum_i X_i$ *given* $\sum_i X_i + \sum_j Y_j = t$ *is completely known, i.e. binomial* $\left(t, \dfrac{m}{m+n}\right)$*, based on which we can easily derive the desired critical value etc.*

Appendix

A3.1 A more formal treatment of §3.3.1...

Theorem. Suppose $\boldsymbol{X}$ has probability function of exponential family form with natural statistic $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X}) = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$. Then $\boldsymbol{T}$ has probability function also of exponential family form with the same natural parameter.

Outline of proof:

Fix $\boldsymbol{\pi}^* = (\pi_1^*, \ldots, \pi_k^*) \in$ *natural parameter space* Π.

Denote by $\boldsymbol{X}^*$ *a sample with probability function* $f(\boldsymbol{x}|\boldsymbol{\pi}^*) = C(\boldsymbol{\pi}^*)\, h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^k \pi_j^* t_j(\boldsymbol{x})\right\}$. *Let* $\boldsymbol{T}^* = (T_1^*, \ldots, T_k^*) = \boldsymbol{T}(\boldsymbol{X}^*)$ *and* $f_{\boldsymbol{T}^*}$ *be the probability function of* $\boldsymbol{T}^*$. *Note that* $f_{\boldsymbol{T}^*}$ *does not depend on* $\boldsymbol{\pi}$.

For any $\boldsymbol{T}$*-measurable set* $B \subset \mathbb{R}^k$, *consider*

$$\begin{aligned}
\mathbb{P}(\boldsymbol{T} \in B|\boldsymbol{\pi}) &= \int f(\boldsymbol{x}|\boldsymbol{\pi})\, \mathbf{1}\{\boldsymbol{T}(\boldsymbol{x}) \in B\}\, d\boldsymbol{x} \\
&= \int \frac{C(\boldsymbol{\pi})}{C(\boldsymbol{\pi}^*)} \exp\Big\{\sum_{j=1}^k (\pi_j - \pi_j^*)\, t_j(\boldsymbol{x})\Big\} f(\boldsymbol{x}|\boldsymbol{\pi}^*)\, \mathbf{1}\{\boldsymbol{T}(\boldsymbol{x}) \in B\}\, d\boldsymbol{x} \\
&= \mathbb{E}\Big[\frac{C(\boldsymbol{\pi})}{C(\boldsymbol{\pi}^*)} \exp\Big\{\sum_{j=1}^k (\pi_j - \pi_j^*)\, t_j(\boldsymbol{X}^*)\Big\}\, \mathbf{1}\{\boldsymbol{T}(\boldsymbol{X}^*) \in B\}\Big] \\
&= \mathbb{E}\Big[\frac{C(\boldsymbol{\pi})}{C(\boldsymbol{\pi}^*)} \exp\Big\{\sum_{j=1}^k (\pi_j - \pi_j^*)\, T_j^*\Big\}\, \mathbf{1}\{\boldsymbol{T}^* \in B\}\Big] \\
&= \int_B \frac{C(\boldsymbol{\pi})}{C(\boldsymbol{\pi}^*)} \exp\Big\{\sum_{j=1}^k (\pi_j - \pi_j^*)\, t_j\Big\}\, f_{\boldsymbol{T}^*}(\boldsymbol{t})\, d\boldsymbol{t}.
\end{aligned}$$

Thus $\boldsymbol{T}$ *has joint probability function*

$$\begin{aligned}
g(\boldsymbol{t}|\boldsymbol{\pi}) &= \frac{C(\boldsymbol{\pi})}{C(\boldsymbol{\pi}^*)} \exp\Big\{\sum_{j=1}^k (\pi_j - \pi_j^*)\, t_j\Big\}\, f_{\boldsymbol{T}^*}(\boldsymbol{t}) \\
&\propto \Big\{\exp\Big(-\sum_{j=1}^k \pi_j^* t_j\Big) f_{\boldsymbol{T}^*}(\boldsymbol{t})\Big\} \exp\Big(\sum_{j=1}^k \pi_j\, t_j\Big) \\
&= h^*(\boldsymbol{t}) \exp\Big(\sum_{j=1}^k \pi_j t_j\Big),
\end{aligned}$$

which is of exponential family form. ∎