Lecture 18: Exponential and location-scale families

Two important types of parametric families

Definition 2.2 (Exponential families). A parametric family $\{P_\theta : \theta \in \Theta\}$ dominated by a σ-finite measure ν on $(\Omega, \mathcal{F})$ is called an *exponential family* if and only if

$$\frac{dP_\theta}{d\nu}(\omega) = \exp\{[\eta(\theta)]^\tau T(\omega) - \xi(\theta)\}h(\omega), \quad \omega \in \Omega, \tag{1}$$

where $\exp\{x\} = e^x$, T is a random p-vector with a fixed positive integer p, η is a function from Θ to $\mathcal{R}^p$, h is a nonnegative Borel function on $(\Omega, \mathcal{F})$, and

$$\xi(\theta) = \log\left\{\int_\Omega \exp\{[\eta(\theta)]^\tau T(\omega)\}h(\omega)d\nu(\omega)\right\}.$$

In Definition 2.2, T and h are functions of ω only, whereas η and ξ are functions of θ only.
The representation (1) of an exponential family is not unique.
$\tilde{\eta}(\theta) = D\eta(\theta)$ with a $p \times p$ nonsingular matrix D gives another representation (with T replaced by $\tilde{T} = (D^\tau)^{-1}T$).
A change of the measure that dominates the family also changes the representation.
If we define $\lambda(A) = \int_A h d\nu$ for any $A \in \mathcal{F}$, then we obtain an exponential family with densities

$$\frac{dP_\theta}{d\lambda}(\omega) = \exp\{[\eta(\theta)]^\tau T(\omega) - \xi(\theta)\}. \tag{2}$$

In an exponential family, consider the reparameterization $\eta = \eta(\theta)$ and

$$f_\eta(\omega) = \exp\{\eta^\tau T(\omega) - \zeta(\eta)\}h(\omega), \quad \omega \in \Omega, \tag{3}$$

where $\zeta(\eta) = \log\left\{\int_\Omega \exp\{\eta^\tau T(\omega)\}h(\omega)d\nu(\omega)\right\}$.
This is the *canonical form* for the family (not unique).
The new parameter η is called the *natural parameter*.
The new parameter space $\Xi = \{\eta(\theta) : \theta \in \Theta\}$, a subset of $\mathcal{R}^p$, is called the *natural parameter space*.
An exponential family in canonical form is called a *natural exponential family*.
If there is an open set contained in the natural parameter space of an exponential family, then the family is said to be of *full rank*.

Example 2.6. The normal family $\{N(\mu, \sigma^2) : \mu \in \mathcal{R}, \sigma > 0\}$ is an exponential family, since the Lebesgue p.d.f. of $N(\mu, \sigma^2)$ can be written as

$$\frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \log\sigma\right\}.$$

Hence, $T(x) = (x, -x^2)$, $\eta(\theta) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$, $\theta = (\mu, \sigma^2)$, $\xi(\theta) = \frac{\mu^2}{2\sigma^2} + \log\sigma$, and $h(x) = 1/\sqrt{2\pi}$.
Let $\eta = (\eta_1, \eta_2) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$. Then $\Xi = \mathcal{R} \times (0, \infty)$ and we can obtain a natural exponential family of full rank with $\zeta(\eta) = \eta_1^2/(4\eta_2) + \log(1/\sqrt{2\eta_2})$.

A subfamily of the previous normal family, $\{N(\mu, \mu^2) : \mu \in \mathcal{R}, \mu \neq 0\}$, is also an exponential family with the natural parameter $\eta = (\frac{1}{\mu}, \frac{1}{2\mu^2})$ and natural parameter space $\Xi = \{(x, y) : y = 2x^2,\ x \in \mathcal{R},\ y > 0\}$. This exponential family is not of full rank.

For an exponential family, (2) implies that there is a nonzero measure λ such that

$$\frac{dP_\theta}{d\lambda}(\omega) > 0 \qquad \text{for all } \omega \text{ and } \theta. \tag{4}$$

We can use this fact to show that a family of distributions is not an exponential family.
Consider the family of uniform distributions, i.e., P_θ is $U(0, \theta)$ with an unknown $\theta \in (0, \infty)$.
If $\{P_\theta : \theta \in (0, \infty)\}$ is an exponential family, then (4) holds with a nonzero measure λ.
For any $t > 0$, there is a $\theta < t$ such that $P_\theta([t, \infty)) = 0$, which with (4) implies that $\lambda([t, \infty)) = 0$.
Also, for any $t \leq 0$, $P_\theta((-\infty, t]) = 0$, which with (4) implies that $\lambda((-\infty, t]) = 0$.
Since t is arbitrary, $\lambda \equiv 0$.
This contradiction implies that $\{P_\theta : \theta \in (0, \infty)\}$ cannot be an exponential family.

Which of the parametric families from Tables 1.1 and 1.2 are exponential families?

An important exponential family containing multivariate discrete distributions.

Example 2.7 (The multinomial family). Consider an experiment having $k + 1$ possible outcomes with p_i as the probability for the ith outcome, $i = 0, 1, ..., k$, $\sum_{i=0}^k p_i = 1$. In n independent trials of this experiment, let X_i be the number of trials resulting in the ith outcome, $i = 0, 1, ..., k$. Then the joint p.d.f. (w.r.t. counting measure) of $(X_0, X_1, ..., X_k)$ is

$$f_\theta(x_0, x_1, ..., x_k) = \frac{n!}{x_0!x_1! \cdots x_k!} p_0^{x_0} p_1^{x_1} \cdots p_k^{x_k} I_B(x_0, x_1, ..., x_k),$$

where $B = \{(x_0, x_1, ..., x_k) :\ x_i\text{'s are integers} \geq 0,\ \sum_{i=0}^k x_i = n\}$ and $\theta = (p_0, p_1, ..., p_k)$. The distribution of $(X_0, X_1, ..., X_k)$ is called the *multinomial* distribution, which is an extension of the binomial distribution. In fact, the marginal c.d.f. of each X_i is the binomial distribution $Bi(p_i, n)$.
$\{f_\theta :\ \theta \in \Theta\}$ is the multinomial family, where $\Theta = \{\theta \in \mathcal{R}^{k+1} :\ 0 < p_i < 1, \sum_{i=0}^k p_i = 1\}$.
Let $x = (x_0, x_1, ..., x_k)$, $\eta = (\log p_0, \log p_1, ..., \log p_k)$, and $h(x) = [n!/(x_0!x_1! \cdots x_k!)]I_B(x)$. Then

$$f_\theta(x_0, x_1, ..., x_k) = \exp\{\eta^\tau x\}\, h(x), \qquad x \in \mathcal{R}^{k+1}. \tag{5}$$

Hence, the multinomial family is a natural exponential family with natural parameter η.
However, representation (5) does not provide an exponential family of full rank, since there is no open set of $\mathcal{R}^{k+1}$ contained in the natural parameter space.
A reparameterization leads to an exponential family with full rank.
Using the fact that $\sum_{i=0}^k X_i = n$ and $\sum_{i=0}^k p_i = 1$, we obtain that

$$f_\theta(x_0, x_1, ..., x_k) = \exp\{\eta_*^\tau x_* - \zeta(\eta_*)\}\, h(x), \qquad x \in \mathcal{R}^{k+1}, \tag{6}$$

where $x_* = (x_1, ..., x_k)$, $\eta_* = (\log(p_1/p_0), ..., \log(p_k/p_0))$, and $\zeta(\eta_*) = -n \log p_0$.
The η_*-parameter space is $\mathcal{R}^k$.
Hence, the family of densities given by (6) is a natural exponential family of full rank.

If $X_1, ..., X_m$ are independent random vectors with p.d.f.'s in exponential families, then the p.d.f. of $(X_1, ..., X_m)$ is again in an exponential family.
The following result summarizes some other useful properties of exponential families.
Its proof can be found in Lehmann (1986).

Theorem 2.1. Let $\mathcal{P}$ be a natural exponential family given by (3).
(i) Let $T = (Y, U)$ and $\eta = (\vartheta, \varphi)$, where Y and ϑ have the same dimension.
Then, Y has the p.d.f.

$$f_\eta(y) = \exp\{\vartheta^\tau y - \zeta(\eta)\}$$

w.r.t. a σ-finite measure depending on φ.
In particular, T has a p.d.f. in a natural exponential family.
Furthermore, the conditional distribution of Y given $U = u$ has the p.d.f. (w.r.t. a σ-finite measure depending on u)

$$f_{\vartheta,u}(y) = \exp\{\vartheta^\tau y - \zeta_u(\vartheta)\},$$

which is in a natural exponential family indexed by ϑ.
(ii) If η_0 is an interior point of the natural parameter space, then the m.g.f. ψ_{η_0} of $P_{\eta_0} \circ T^{-1}$ is finite in a neighborhood of 0 and is given by

$$\psi_{\eta_0}(t) = \exp\{\zeta(\eta_0 + t) - \zeta(\eta_0)\}.$$

Furthermore, if f is a Borel function satisfying $\int |f| dP_{\eta_0} < \infty$, then the function

$$\int f(\omega) \exp\{\eta^\tau T(\omega)\} h(\omega) d\nu(\omega)$$

is infinitely often differentiable in a neighborhood of η_0, and the derivatives may be computed by differentiation under the integral sign.

Example 2.5. Let P_θ be the binomial distribution $Bi(\theta, n)$ with parameter θ, where n is a fixed positive integer. Then $\{P_\theta : \theta \in (0,1)\}$ is an exponential family, since the p.d.f. of P_θ w.r.t. the counting measure is

$$f_\theta(x) = \exp\left\{x \log \tfrac{\theta}{1-\theta} + n \log(1-\theta)\right\} \binom{n}{x} I_{\{0,1,\dots,n\}}(x)$$

($T(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$, $\xi(\theta) = -n \log(1-\theta)$, and $h(x) = \binom{n}{x} I_{\{0,1,\dots,n\}}(x)$). If we let $\eta = \log \frac{\theta}{1-\theta}$, then $\Xi = \mathcal{R}$ and the family with p.d.f.'s

$$f_\eta(x) = \exp\{x\eta - n \log(1 + e^\eta)\} \binom{n}{x} I_{\{0,1,\dots,n\}}(x)$$

is a natural exponential family of full rank.
Using Theorem 2.1(ii) and the result in Example 2.5, we obtain that the m.g.f. of the binomial distribution $Bi(\theta, n)$ is

$$\begin{aligned}\psi_\eta(t) &= \exp\{n \log(1 + e^{\eta+t}) - n \log(1 + e^\eta)\} \\ &= \left(\frac{1 + e^\eta e^t}{1 + e^\eta}\right)^n \\ &= (1 - \theta + \theta e^t)^n.\end{aligned}$$

Definition 2.3 (Location-scale families). Let P be a known probability measure on $(\mathcal{R}^k, \mathcal{B}^k)$, $\mathcal{V} \subset \mathcal{R}^k$, and $\mathcal{M}_k$ be a collection of $k \times k$ symmetric positive definite matrices. The family

$$\{P_{(\mu,\Sigma)} : \ \mu \in \mathcal{V}, \ \Sigma \in \mathcal{M}_k\} \tag{7}$$

is called a *location-scale family* (on $\mathcal{R}^k$), where

$$P_{(\mu,\Sigma)}(B) = P\left(\Sigma^{-1/2}(B-\mu)\right), \quad B \in \mathcal{B}^k,$$

$\Sigma^{-1/2}(B-\mu) = \{\Sigma^{-1/2}(x-\mu) : \ x \in B\} \subset \mathcal{R}^k$, and $\Sigma^{-1/2}$ is the inverse of the "square root" matrix $\Sigma^{1/2}$ satisfying $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. The parameters μ and $\Sigma^{1/2}$ are called the location and scale parameters, respectively.

The following are some important examples of location-scale families.
The family $\{P_{(\mu,I_k)} : \ \mu \in \mathcal{R}^k\}$ is a *location family*, where I_k is the $k \times k$ identity matrix.
The family $\{P_{(0,\Sigma)} : \ \Sigma \in \mathcal{M}_k\}$ is a *scale family*.
In some cases, we consider a location-scale family of the form $\{P_{(\mu,\sigma^2 I_k)} : \ \mu \in \mathcal{R}^k, \sigma > 0\}$.
If $X_1, ..., X_k$ are i.i.d. with a common distribution in the location-scale family $\{P_{(\mu,\sigma^2)} : \mu \in \mathcal{R}, \sigma > 0\}$, then the joint distribution of the vector $(X_1, ..., X_k)$ is in the location-scale family $\{P_{(\mu,\sigma^2 I_k)} : \ \mu \in \mathcal{V}, \sigma > 0\}$ with $\mathcal{V} = \{(x, ..., x) \in \mathcal{R}^k : \ x \in \mathcal{R}\}$.

A location-scale family can be generated as follows.
Let X be a random k-vector having a distribution P.
Then the distribution of $\Sigma^{1/2}X + \mu$ is $P_{(\mu,\Sigma)}$.

On the other hand, if X is a random k-vector whose distribution is in the location-scale family (7), then the distribution $DX + c$ is also in the same family, provided that $D\mu + c \in \mathcal{V}$ and $D\Sigma D^\tau \in \mathcal{M}_k$.

Let F be the c.d.f. of P.
Then the c.d.f. of $P_{(\mu,\Sigma)}$ is $F\left(\Sigma^{-1/2}(x-\mu)\right)$, $x \in \mathcal{R}^k$.
If F has a Lebesgue p.d.f. f, then the Lebesgue p.d.f. of $P_{(\mu,\Sigma)}$ is $\mathrm{Det}(\Sigma^{-1/2}) f\left(\Sigma^{-1/2}(x-\mu)\right)$, $x \in \mathcal{R}^k$ (Proposition 1.8).

Many families of distributions in Table 1.2 (§1.3.1) are location, scale, or location-scale families.
For example, the family of exponential distributions $E(a, \theta)$ is a location-scale family on $\mathcal{R}$ with location parameter a and scale parameter θ;
the family of uniform distributions $U(0, \theta)$ is a scale family on $\mathcal{R}$ with a scale parameter θ.
The k-dimensional normal family is a location-scale family on $\mathcal{R}^k$.