Lecture 5 – Exponential and Location-Scale Families

Last Modified: September 8, 2026

Outline:

5.1 Exponential families

In Lectures 3–4, we asked how much of a sample must be retained without losing information about an unknown parameter. In the examples, the answer varied greatly. For instance, $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{Ber}(p)$, has the one-dimensional minimal sufficient statistic $\sum_i X_i$, whereas the minimal sufficient statistic for a Cauchy location model is essentially the whole ordered sample.
This leads to a natural question:

When does a growing sample have a low-dimensional sufficient statistic?

We will see that this is the case for so-called *exponential families*.

Definition 1 (One-parameter exponential family)**.** Suppose $\Theta \subseteq \mathbb{R}$. A statistical model $\{f_\theta(x) : \theta \in \Theta\}$ is a **one-parameter exponential family** if its pmf or pdf can be written as

$$f_\theta(x) = g(x) \exp\big\{b(\theta)R(x) - c(\theta)\big\},$$

where $g(x) \geq 0$. We call $R(x)$ a **natural statistic**, and $b(\theta)$ a **natural parameter**.

The set on which $g(x) > 0$ is the support of every distribution in the family, so the support of distributions in an exponential family does not depend on θ.

Remark 1. An exponential *family* is a class of statistical models. It is not the same thing as the exponential *distribution*. The exponential distribution is one particular example of a distribution belonging to an exponential family.

Remark 2. Another way to write the density for a one-parameter exponential family is just

$$f_\theta(x) = g(x)h(\theta)e^{b(\theta)R(x)},$$

where $g(x) \geq 0$ for all x and $h(\theta) > 0$ for all θ. To bring it into the original form, simply define $c(\theta) = -\log h(\theta)$.

Remark 3.

- The function $R(x)$ has to actually depend on x. If it did not (i.e. $R(x) = R$ for all x), then the only part of the distribution depending on x would be $g(x)$. But then $e^{b(\theta)R-c(\theta)}$ would be forced to be the normalizing constant: $e^{b(\theta)R-c(\theta)} = 1/\int_{-\infty}^{\infty} g(x)dx$. So then $f_\theta(x) = g(x)/\int_{-\infty}^{\infty} g(x)dx$ would not actually depend on θ, and there would be no statistical inference to be done.

- By the same rationale, $b(\theta)$ has to actually depend on θ. If it did not (i.e. $b(\theta) = b$ for all θ), then we would have $f_\theta(x) = g(x)e^{bR(x)}e^{-c(\theta)}$, and $e^{-c(\theta)}$ would be forced to be a θ-independent normalizing constant. So again, no dependence on θ in f_θ, and no statistical inference.
- Once you have specified g, b, R, the $e^{-c(\theta)}$ part is forced: we must have

$$e^{c(\theta)} = \int g(x)e^{b(\theta)R(x)}\mathrm{d}x.$$

Example 1 (Bernoulli model)**.** Let $X \sim \mathsf{Ber}(p)$, where $0 < p < 1$. Since $x \in \{0, 1\}$,

$$\begin{aligned} f_p(x) &= p^x(1-p)^{1-x}\mathbb{1}_{\{0,1\}}(x) \\ &= \exp\left\{x\log\left(\frac{p}{1-p}\right) + \log(1-p)\right\}\mathbb{1}_{\{0,1\}}(x). \end{aligned}$$

Thus the Bernoulli model is an exponential family with

$$R(x) = x, \qquad b(p) = \log\left(\frac{p}{1-p}\right), \qquad c(p) = -\log(1-p),$$

and $g(x) = \mathbb{1}_{\{0,1\}}(x)$.

Example 2 (Poisson model)**.** Let $X \sim \mathsf{Poisson}(\lambda)$, where $\lambda > 0$. Its pmf is

$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}\mathbb{1}_{\{0,1,2,\ldots\}}(x) = \frac{1}{x!}\mathbb{1}_{\{0,1,2,\ldots\}}(x)e^{x\log(\lambda)-\lambda}.$$

Therefore the Poisson model is an exponential family with

$$R(x) = x, \qquad b(\lambda) = \log(\lambda), \qquad c(\lambda) = \lambda, \qquad g(x) = \frac{1}{x!}\mathbb{1}_{\{0,1,2,\ldots\}}(x).$$

Example 3 (Gamma model with fixed shape)**.** Let $X \sim \mathsf{Gamma}(\alpha, \beta)$, where the shape $\alpha > 0$ is known and the rate $\beta > 0$ is unknown. Then

$$\begin{aligned} f_\beta(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathbb{1}_{\{x>0\}} \\ &= \frac{x^{\alpha-1}}{\Gamma(\alpha)}\mathbb{1}_{\{x>0\}}\exp\{(-\beta)x + \alpha\log(\beta)\}. \end{aligned}$$

Thus this is an exponential family with

$$R(x) = x, \qquad b(\beta) = -\beta, \qquad c(\beta) = -\alpha\log(\beta).$$

When $\alpha = 1$, this gives the exponential distribution with rate β.

Example 4 (Normal model with known variance)**.** Let $X \sim \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Expanding the square gives

$$\begin{aligned} f_\mu(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\ &= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}}_{g(x)}\exp\left\{\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right\}. \end{aligned}$$

Therefore this is an exponential family with

$$R(x) = x, \qquad b(\mu) = \frac{\mu}{\sigma^2}, \qquad c(\mu) = \frac{\mu^2}{2\sigma^2}.$$

We will see later that this model is also a location family.

The multiparameter form

If θ is a vector in $\mathbb{R}^k$, the exponential-family has k corresponding natural parameters $b_1, \dots, b_k$, and k natural statistics $R_1, \dots, R_k$. The form of the pdf/pmf is then

$$f_\theta(x) = g(x) \exp\{b_1(\theta)R_1(x) + \cdots + b_k(\theta)R_k(\theta) - c(\theta)\}.$$

For example, when both parameters of $X \sim \mathcal{N}(\mu, \sigma^2)$ are unknown, the vector of unknowns is $\theta = (\mu, \sigma^2)$, and

$$f_{\mu,\sigma^2}(x) = \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}.$$

The natural statistics are therefore

$$R_1(x) = x, \qquad R_2(x) = x^2.$$

The corresponding natural parameters are

$$b_1(\mu, \sigma^2) = \frac{\mu}{\sigma^2}, \qquad b_2(\mu, \sigma^2) = -\frac{1}{2\sigma^2}.$$

5.2 Minimal sufficient statistics

In a one-parameter exponential family, there is a one-dimensional minimal sufficient statistic, as the next theorem shows.

Theorem 1. Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} f_\theta$, where f_θ belongs to the exponential family in Definition 1. Then

$$R_n(\underline{X}) = \sum_{i=1}^n R(X_i)$$

is a minimal sufficient statistic for θ.

Proof. Independence gives

$$\begin{aligned} f_{\underline{X},\theta}(\underline{x}) &= \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n [g(x_i)\exp\{b(\theta)R(x_i) - c(\theta)\}] \\ &= \left\{\prod_{i=1}^n g(x_i)\right\} \exp\left\{b(\theta)\sum_{i=1}^n R(x_i) - nc(\theta)\right\} = \left\{\prod_{i=1}^n g(x_i)\right\} e^{b(\theta)R_n(\underline{x}) - c(\theta)}. \end{aligned}$$

This satisfies the conditions of the Neyman–Fisher factorization theorem: there is a factor depending only on x, multiplied by a factor that is a function of θ and $R_n(\underline{x})$. Hence R_n is sufficient. To check minimal sufficiency, take two sample points $\underline{x}, \underline{y}$ such that $R_n(\underline{x}) \neq R_n(\underline{y})$. Their likelihood ratio is

$$\frac{L_{\underline{x}}(\theta)}{L_{\underline{y}}(\theta)} = \frac{\prod_{i=1}^n g(x_i)}{\prod_{i=1}^n g(y_i)} e^{b(\theta)(R_n(\underline{x}) - R_n(\underline{y}))} = c_1 e^{c_2 b(\theta)},$$

where c_1, c_2 depend on $\underline{x}, \underline{y}$ but are constant when viewing the likelihood ratio as a function of θ. Since $c_2 \neq 0$ and $b(\theta)$ varies with θ, the two likelihoods are not proportional. The likelihood criterion from Lecture 4 now shows that S_n is minimal sufficient. $\square$

Example 5 (Bernoulli sampling revisited)**.** For a Bernoulli observation, $R(x) = x$. Hence $R_n(\underline{x}) = \sum_{i=1}^{n} X_i$ is minimal sufficient for p. Thus an n-entry sequence of zeros and ones from Bernoulli(p) can be replaced, without losing information about p, by a single count.

The theorem tells us that if $f_\theta(x)$ belongs to the exponential family, then we have an automatic way to find a minimal sufficient statistic: it is $\sum_i R(x_i)$ where $R(x)$ is the natural statistic in the density. Also, the minimal sufficient statistic will be one-dimensional. Amazingly enough, a converse theorem holds!

Theorem 2 (Pitman-Koopman Lemma)**.** If a statistical model $\{f_\theta(x) \mid \theta \in \Theta\}$ is such that the support of $f_\theta(x)$ does not depend on θ and there exists a sufficient statistic of constant dimension (i.e. independent of n), then the model belongs to the exponential family.

The theorem tells us that within statistical models whose support does not depend on θ, only exponential families have low-dimensional sufficient statistics.

5.3 Finding moments from the exponential-family form

The normalizing term $c(\theta)$ and the natural parameter $b(\theta)$ contain useful information about the moments of the natural statistic.

Theorem 3 (Mean of the natural statistic)**.** Suppose f_θ belongs to the exponential family in Definition 1 and $b'(\theta) \neq 0$. Then

$$\mathbb{E}_\theta[R(X)] = \frac{c'(\theta)}{b'(\theta)}.$$

Proof. We prove it in the continuous case. Recall that we must have

$$e^{c(\theta)} = \int_{-\infty}^{\infty} g(x) e^{b(\theta)R(x)} dx,$$

since densities integrate to 1. Differentiating both sides gives

$$\begin{aligned} c'(\theta)e^{c(\theta)} &= \int_{-\infty}^{\infty} g(x) \frac{\partial}{\partial\theta} e^{b(\theta)R(x)} \mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x) R(x) b'(\theta) e^{b(\theta)R(x)} \mathrm{d}x \\ &= b'(\theta) \int_{-\infty}^{\infty} R(x) g(x) e^{b(\theta)R(x)} \mathrm{d}x. \end{aligned}$$

Now divide both sides by $e^{c(\theta)}b'(\theta)$ to get

$$\frac{c'(\theta)}{b'(\theta)} = \int_{-\infty}^{\infty} R(x) \Big\{ g(x) e^{b(\theta)R(x) - c(\theta)} \Big\} \mathrm{d}x = \int_{-\infty}^{\infty} R(x) f_\theta(x) \mathrm{d}x = \mathbb{E}_\theta[R(X)].$$

The discrete proof is identical with the integral replaced by a sum. $\square$

Example 6 (Recovering familiar expectations)**.** For $X \sim \mathsf{Ber}(p)$, Example 1 gives

$$b'(p) = \frac{1}{p(1-p)}, \qquad c'(p) = \frac{1}{1-p}.$$

Therefore

$$\mathbb{E}_p[X] = \frac{c'(p)}{b'(p)} = p.$$

For $X \sim \mathsf{Gamma}(\alpha, \beta)$ with fixed α, Example 3 gives

$$b'(\beta) = -1, \qquad c'(\beta) = -\frac{\alpha}{\beta}.$$

Therefore

$$\mathbb{E}_\beta[X] = \frac{c'(\beta)}{b'(\beta)} = \frac{\alpha}{\beta}.$$

In particular, an exponential random variable with rate β has mean $1/\beta$.

5.4 Ancillary statistics and location-scale families

Exponential-family structure helped us retain all the information about a parameter in a low-dimensional statistic. We now switch to a different construction: finding an observable statistic whose distribution does **not** involve the parameter.

Definition 2 (Ancillary statistic)**.** A statistic $V = V(\underline{X})$ is **ancillary** for a parameter θ if the distribution of V does not depend on θ.

In other words, an ancillary statistic has a fixed distribution, regardless of the value of the parameter. Because it is a statistic, it must be observable: it may depend on the data, but it cannot contain the unknown parameter.

Example 7 (A first ancillary statistic)**.** Suppose $X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$. Since the observations are independent,

$$X_1 - X_2 \sim \mathcal{N}(\mu - \mu, 1 + 1) = \mathcal{N}(0, 2).$$

The distribution does not depend on μ, so $X_1 - X_2$ is ancillary for μ. Equivalently, write

$$X_1 = \mu + Z_1, \qquad X_2 = \mu + Z_2, \qquad Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

Then $X_1 - X_2 = Z_1 - Z_2$: the location parameter cancels. This is a general feature of location families.

Location families

Definition 3 (Location family)**.** Let f_0 be a fixed pdf or pmf. Then a statistical model of the form $\{f_0(x - \mu) : \mu \in \mathbb{R}\}$ is called a *location family*, and μ is its *location parameter*.

Varying μ shifts the reference distribution without changing its shape.
We can construct a random variable X with pdf f_μ by drawing Z with pdf f_0 and setting $X = Z + \mu$. Indeed, we can check this using the cdf method. Let F_0 be the cdf corresponding to f_0. Then

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(Z \le x - \mu) = F_0(x - \mu).$$

Differentiating both sides gives $f_X(x) = f_0(x - \mu)$, which is precisely $f_\mu(x)$.

Ancillary statistics for location families: For a random sample from a location family, write $X_i = \mu + Z_i$, where $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} F_0$. Differences eliminate the location parameter:

$$X_i - X_j = Z_i - Z_j, \qquad X_{(n)} - X_{(1)} = Z_{(n)} - Z_{(1)}.$$

Consequently, $X_i - X_j$ and the sample range $X_{(n)} - X_{(1)}$ are ancillary for μ. Examples of location families include $\{\mathcal{N}(\mu, 1) : \mu \in \mathbb{R}\}$ and $\{\mathsf{Unif}(\mu, \mu + 1) : \mu \in \mathbb{R}\}$.

Scale families

Definition 4 (Scale family)**.** Let F_0 be a fixed cdf. The statistical model with cdfs

$$F_\sigma(x) = F_0\Big(\frac{x}{\sigma}\Big), \qquad \sigma > 0,$$

is a **scale family**, and σ is its **scale parameter**. If F_0 has pdf f_0, the corresponding densities are

$$f_\sigma(x) = \frac{1}{\sigma} f_0\Big(\frac{x}{\sigma}\Big).$$

We can construct a random variable X with cdf F_σ by drawing Z with cdf F_0 and setting $X = \sigma Z$. Indeed,

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\Big(Z \le \frac{x}{\sigma}\Big) = F_0\Big(\frac{x}{\sigma}\Big).$$

The righthand side is precisely $F_\sigma(x)$.

Ancillary statistics for scale families: For a random sample from a scale family, write $X_i = \sigma Z_i$. Ratios eliminate the scale parameter. Whenever the denominators are nonzero,

$$\frac{X_i}{X_j} = \frac{Z_i}{Z_j}, \qquad \frac{\sum_{j=1}^n X_j}{X_i} = \frac{\sum_{j=1}^n Z_j}{Z_i}.$$

These ratios are ancillary for σ.

Example 8 (Gamma and uniform scale families)**.** Let $Z \sim \mathsf{Gamma}(\alpha, 1)$, where α is fixed and the second parameter is the rate. If $X = Z/\beta$, then $X \sim \mathsf{Gamma}(\alpha, \beta)$. Hence the fixed-shape Gamma family is a scale family with scale parameter $\sigma = 1/\beta$. In particular, the exponential-distribution model is a scale family when parameterized by its scale.
If $Z \sim \mathsf{Unif}(0, 1)$ and $X = \theta Z$, then $X \sim \mathsf{Unif}(0, \theta)$. Thus $\{\mathsf{Unif}(0, \theta) : \theta > 0\}$ is also a scale family.

Location-scale families

Definition 5 (Location-scale family)**.** Let F_0 be a fixed cdf. The statistical model with cdfs

$$F_{\mu,\sigma}(x) = F_0\left(\frac{x-\mu}{\sigma}\right), \qquad \mu \in \mathbb{R}, \quad \sigma > 0,$$

is a **location-scale family**. If F_0 has pdf f_0, the corresponding densities are

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right).$$

Equivalently, if $Z \sim F_0$, then $X = \mu + \sigma Z$ belongs to the location-scale family generated by F_0.

Example 9 (Normal location-scale family)**.** Let $Z \sim \mathcal{N}(0, 1)$. Then $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$. Thus $\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$ is a location-scale family. Its density has the form

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}}_{f_0((x-\mu)/\sigma)}.$$

Ancillary statistics for location-scale families: For a random sample, write $X_i = \mu + \sigma Z_i$. Ratios of differences eliminate both parameters. Whenever the denominator is nonzero,

$$\frac{X_i - X_j}{X_k - X_r} = \frac{Z_i - Z_j}{Z_k - Z_r}.$$

The ratio is therefore ancillary for (μ, σ).

Another useful construction is based on standardized residuals. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean and $S_X^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ be the square of the sample standard deviation. Define $\overline{Z}$ and S_Z similarly. Then

$$\overline{X} = \mu + \sigma\overline{Z}, \qquad S_X = \sigma S_Z,$$

so

$$R_i = \frac{X_i - \overline{X}}{S_X} = \frac{Z_i - \overline{Z}}{S_Z}, \qquad i = 1, \ldots, n.$$

Thus the joint distribution of $(R_1, \ldots, R_n)$ depends on the reference shape F_0, but not on μ or σ. In other words, $R(\underline{X}) = (R_1, \ldots, R_n)$ is ancillary for (μ, σ).

Example 10 (Why ancillary statistics are useful: checking normal shape)**.** Suppose we observe $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$ and want to assess

$$H_0: \ P \in \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\} \qquad \text{versus} \qquad H_1: \ P \text{ is not normal.}$$

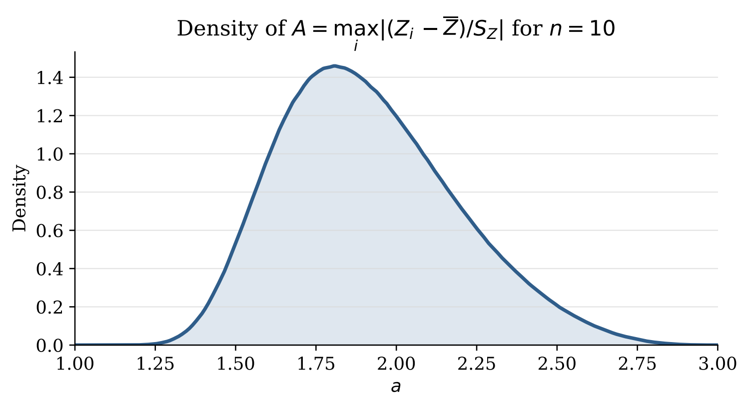
We want to check whether the data have a normal *shape*, without knowing their center or spread. A good way to do this is to use the ancillary statistic

$$A = \max_{1 \le i \le n} \frac{|X_i - \overline{X}|}{S_X}.$$

If the observations really are normal, then

$$A = \max_{1 \le i \le n} \left| \frac{Z_i - \overline{Z}}{S_Z} \right|, \qquad Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

For $n = 10$, the below figure shows the density of A. Suppose we observe say, $A > 5$ or $A < 0.5$.



This is extremely unlikely under the null hypothesis that the X_i are normally distributed, so we reject the null hypothesis.

5.5 Summary

This lecture considered two different questions.

retain information about the parameter	remove parameter dependence
Can the data be compressed without loss?	Can one reference distribution be used?
exponential-family structure	location-scale structure
$f_\theta(x) = g(x)e^{b(\theta)R(x)-c(\theta)}$	$X = \mu + \sigma Z$
$\sum_i R(X_i)$ is minimal sufficient	standardized statistics can be ancillary

- For exponential families, the main result is that $R_n(\underline{x}) = \sum_{i=1}^n R(x_i)$ is a one-dimensional minimal sufficient statistic, where $R(x)$ is the natural statistic in the pdf.
- For location-scale families, standardization removes location and scale:

$$X_i = \mu + \sigma Z_i \quad \Longrightarrow \quad \frac{X_i - \overline{X}}{S_X} = \frac{Z_i - \overline{Z}}{S_Z}.$$

  The resulting distribution depends on the reference shape F_0, but not on μ or σ. This makes it possible to check a whole location-scale family using a single reference distribution.

Some location/scale families are also exponential families, and some are not. The following table summarizes properties of some common distributions we have seen.

model	location/scale structure	exponential-family structure
$\mathcal{N}(\mu, \sigma^2)$	location-scale	yes
$\mathsf{Gamma}(\alpha, \beta),\ \alpha$ fixed	scale ($\sigma = 1/\beta$)	yes
$\mathsf{Unif}(0, \theta)$	scale	no: support depends on θ
$\mathsf{Ber}(p)$	neither	yes
$\mathsf{Poisson}(\lambda)$	neither	yes