Lecture 6: Exponential Families

MATH 667-01
Statistical Inference
University of Louisville

September 14, 2017
Last modified: 9/17/2017

Introduction

- We define an exponential family and state a few general properties which are described in Section 3.4 of Casella and Berger (2001)[1].
- We will also discuss a result from Section 5.2 which states that particular sum for a random sample from an exponential family also belongs to an exponential family.
- A sketch of a proof that derivatives can be computed under the integral for a special case of an exponential family from Section 2.7 of Lehmann (1959)[2] is presented at the end of the lecture.

[1] Casella, G. and Berger, R. (2001). Statistical Inference, second edition. Duxbury Press.

[2] Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley.

- *Definition L6.1* (p.111): A family of pdfs or pmfs is called an *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right).$$

Here $h(x) \geq 0$ and $t_1(x), \dots t_k(x)$ are real-valued functions of the observation x (they cannot depend on $\boldsymbol{\theta}$), and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ (they cannot depend on x).

- *Example L6.1*: Show that the normal distribution with mean μ and variance 1 can be expressed in the form of an exponential family.
- *Answer to Example L6.1*: Its pdf is

$$\begin{aligned} f(x|\mu) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2 + \mu x - \frac{1}{2}\mu^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right) e^{-\frac{1}{2}\mu^2} e^{\mu x} \\ &= h(x)c(\mu)e^{w_1(\mu)t_1(x)} \end{aligned}$$

where $h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $c(\mu) = e^{-\frac{1}{2}\mu^2}$, $w_1(\mu) = \mu$, and $t_1(x) = x$.

- *Example L6.2*: Show that the beta(α,β) distribution can be expressed in the form of an exponential family.
- *Answer to Example L6.2*: Its pdf is

$$
\begin{aligned}
f(x|\alpha,\beta) &= \frac{\Gamma(\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} I_{(0,1)}(x) \\
&= I_{(0,1)}(x)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{(\alpha-1)\ln x + (\beta-1)\ln(1-x)} \\
&= h(x)c(\alpha,\beta)e^{w_1(\alpha,\beta)t_1(x)+w_2(\alpha,\beta)t_2(x)}
\end{aligned}
$$

where $h(x) = I_{(0,1)}(x)$, $c(\alpha,\beta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $w_1(\alpha,\beta) = \alpha - 1$, $t_1(x) = \ln x$, $w_2(\alpha,\beta) = \beta - 1$, and $t_2(x) = \ln(1-x)$.

- *Example L6.3*: Consider the continuous distribution with density function

$$\begin{aligned} f(x|\theta) &= \frac{(\theta+1)x^{\theta}}{\theta^{\theta}}, 0 < x < \theta \\ &= \frac{(\theta+1)}{\theta^{\theta}} e^{\theta \ln x} \end{aligned}$$

  where $\theta > 0$. Is this an exponential family? Why or why not?

- *Answer to Example L6.3*: No, the support for the density cannot depend on θ.

- *Theorem L6.1* (Thm 3.4.2 on p.112): If X is a random variable with pdf or pmf of the form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right),$$

then

$$\mathsf{E}\left[\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right] = -\frac{\partial}{\partial \theta_j}\ln c(\boldsymbol{\theta})$$

and

$$\mathsf{Var}\left[\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right] = -\frac{\partial^2}{\partial \theta_j^2}\ln c(\boldsymbol{\theta}) - \mathsf{E}\left[\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial^2 \theta_j} t_i(X)\right].$$

- *Example L6.4*:
  (a) Assuming n is fixed, show that the binomial distribution with probability of success p based on n trials can be expressed in the form of an exponential family.
  (b) Use *Theorem L6.1* to show that $\mathsf{E}[X] = np$ and $\mathsf{Var}[X] = np(1-p)$.
- *Answer to Example L6.4*: (a) Its pmf is

$$
\begin{aligned}
f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\ldots,n\}}(x) \\
&= \left(\binom{n}{x} I_{\{0,1,\ldots,n\}}(x)\right)(1-p)^n \exp\left\{x \ln\left(\frac{p}{1-p}\right)\right\} \\
&= h(x)c(p)e^{w_1(p)t_1(x)}
\end{aligned}
$$

where $h(x) = \binom{n}{x} I_{\{0,1,\ldots,n\}}(x)$, $c(p) = (1-p)^n$,

$w_1(p) = \ln\left(\frac{p}{1-p}\right)$, and $t_1(x) = x$.

- *Answer to Example L6.4 continued*: Alternately, the pmf can be expressed as

$$f(x|p) = \tilde{h}(x)\tilde{c}(p)e^{w_1(p)t_1(x)}$$

where $\tilde{h}(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n I_{\{0,1,\dots,n\}}(x)$, $\tilde{c}(p) = 2^n(1-p)^n$, $w_1(p) = \ln\left(\frac{p}{1-p}\right)$, and $t_1(x) = x$ so that $\tilde{h}(x)$ is one of the pmf's in the family.

- (b) Directly applying the theorem to the first form, we see that

$$\mathsf{E}\left[\frac{1}{p(1-p)}X\right] = -\frac{-n}{1-p} \Rightarrow \mathsf{E}[X] = np.$$

$$\mathsf{Var}\left[\frac{1}{p(1-p)}X\right] = -\frac{-n}{(1-p)^2} - \frac{(2p-1)}{p^2(1-p)^2}np \Rightarrow \mathsf{Var}\left[X\right] = np(1-p)$$

- Sometimes, an exponential family is reparametrized in terms of the *natural parameter* $\boldsymbol{\eta}$ and *cumulant generating function* $\psi(\boldsymbol{\eta})$:

$$
\begin{aligned}
f(x|\boldsymbol{\eta}) &= h(x)c^*(\boldsymbol{\eta})\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right) \\
&= h(x)\exp\{-\psi(\boldsymbol{\eta})\}\exp\left\{\sum_{i=1}^{k}\eta_i t_i(x)\right\} \\
&= \exp\left\{\sum_{i=1}^{k}\eta_i t_i(x) - \psi(\boldsymbol{\eta})\right\}h(x) \\
&= \frac{\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)h(x)}{\int_{-\infty}^{\infty}\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)h(x)\ dx}
\end{aligned}
$$

where $\psi(\boldsymbol{\eta}) = \ln\left(\int_{-\infty}^{\infty}\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)h(x)\ dx\right)$.

- *Definition L6.2* (p.114): The set $\mathcal{H} = \left\{\eta = (\eta_1, \ldots, \eta_k) : \int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)\ dx < \infty\right\}$ is called the *natural parameter space* for the family. (The integral is replaced with an appropriate sum if the random variable is discrete.)
- Note that $e^{\psi(\boldsymbol{\eta})} = (c^*(\boldsymbol{\eta}))^{-1}$ is the moment generating function of $(t_1(X), \ldots, t_k(X))$ if X has pdf $h(x)$.
- Then the formulas for the first two central moments from *Theorem L6.1* reduce to

$$\mathsf{E}\left[t_j(X)\right] = -\frac{\partial}{\partial \eta_j} \ln c^*(\boldsymbol{\eta}) = \frac{\partial \psi(\boldsymbol{\eta})}{\partial \eta_j}$$

and

$$\mathsf{Var}\left[t_j(X)\right] = -\frac{\partial^2}{\partial \eta_j^2} \ln c^*(\boldsymbol{\eta}) = \frac{\partial^2 \psi(\boldsymbol{\eta})}{\partial \eta_j^2}.$$

- *Answer to Example L6.4 continued*: In terms of the natural parameterization $\eta = \ln(\frac{p}{1-p}) \Leftrightarrow p = \frac{e^\eta}{1+e^\eta}$, the pmf of the binomial distribution can be expressed as

$$f(x|\eta) = e^{\eta t_1(x) - \psi(\eta)} \tilde{h}(x)$$

where $\tilde{h}(x) = \binom{n}{x} \left(\frac{1}{2}\right)^n I_{\{0,1,\ldots,n\}}(x)$, $\psi(\eta) = n \ln\left(\frac{1+e^\eta}{2}\right)$ and $t_1(x) = x$.

- Then $\mathsf{E}[X] = \psi'(\eta) = n\frac{e^\eta}{1+e^\eta} = np$ and $\mathsf{Var}[X] = \psi''(\eta) = n\frac{e^\eta}{(1+e^\eta)^2} = np(1-p)$.

- *Definition L6.3* (p.115): A *full* exponential family is a family of pmf/pdf's for which the dimension of $\boldsymbol{\theta}$ is equal to k.
- *Definition L6.4* (p.115): A *curved* exponential family is a family of pmf/pdf's for which the dimension of $\boldsymbol{\theta}$ is less than k.

- *Example L6.5*: Show that the normal family of densities with mean μ and variance σ^2 can be expressed as an exponential family. What is its natural parameter space? Is it a full exponential or a curved exponential family?
- *Answer to Example L6.5*: Its pdf is

$$f(x|\boldsymbol{\eta}) = e^{\eta_1 t_1(x) + \eta_2 t_2(x) - \psi(\boldsymbol{\eta})} h(x)$$

where $h(x) = \frac{1}{\sqrt{2\pi}}$, $\psi(\boldsymbol{\eta}) = \frac{1}{2}\left(\frac{\eta_2^2}{2\eta_1} - \ln \eta_1\right)$, $t_1(x) = -\frac{x^2}{2}$, and $t_1(x) = x$ with $\eta_1 = 1/\sigma^2$ and $\eta_2 = \mu/\sigma^2$.

- The natural parameter space is

$$\{(\eta_1, \eta_2) : \eta_1 > 0, -\infty < \eta_2 < \infty\}$$

so this is a full exponential family.

- *Example L6.6*: Show that the normal family of densities with mean μ and standard deviation μ can be expressed as an exponential family. What is its natural parameter space? Is it a full exponential or a curved exponential family?
- *Answer to Example L6.6*: With $\sigma = \mu$, we obtain a one-dimensional curved exponential family

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\mu^2}} e^{-1/2} \exp\left(-\frac{x^2}{2\mu^2} + \frac{x}{\mu}\right)$$

with parameter space

$$\left\{(\mu, \mu^2) : \mu > 0\right\}.$$

Random Sample from an Exponential Family

- *Theorem L6.2* (Thm 5.2.11 on p.217): Suppose $X_1, \dots, X_n$ is a random sample from a pdf/pmf $f(x|\theta)$ where $f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$ is a member of an exponential family. Define statistics $T_1, \dots, T_k$ by

$$T_i(X_1, \dots, X_n) = \sum_{j=1}^{n} t_i(X_j), i = 1, \dots, k.$$

Suppose $\{(w_1(\theta), \dots, w_k(\theta)) : \theta \in \Theta\}$ contains an open subset of $\mathbb{R}^k$. Then the distribution of $(T_1, \dots, T_k)$ is an exponential family of the form

$$f_T(u_1, \dots, u_k|\theta) = H(u_1, \dots, u_k)\left[c(\theta)\right]^n \exp\left(\sum_{i=1}^{k} w_i(\theta)u_i\right).$$

- *Example L6.7*: Suppose that $X_1, \ldots, X_n$ are independent exponential random variables with pdf

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta} I_{(0,\infty)}(x)$$

where $\beta > 0$. Does the pdf of $Y = \sum_{i=1}^{n} X_i$ belong to an exponential family? If so, what is the function $H(y)$?

- *Answer to Example L6.7*: Since $f(x|\beta)$ is an exponential family with $h(x) = I_{(0,\infty)}(x)$, $c(\beta) = \frac{1}{\beta}$, $w(\beta) = -\frac{1}{\beta}$, and $t(x) = x$,
  Theorem L6.2 implies that the pdf of Y belongs to an exponential family with $C(\beta) = [c(\beta)]^n = -\frac{1}{\beta^n}$, $w(\beta) = \frac{1}{\beta}$, and $T(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ of the form

$$f_T(u) = H(u)\frac{1}{\beta^n}e^{-u/\beta}.$$

The pdf of a Gamma random variable is

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}I_{(0,\infty)}(x),$$

so $Y \sim \text{Gamma}(n,\beta)$ and $H(y) = \dfrac{1}{\Gamma(n)}y^{n-1}I_{(0,\infty)}(y)$.

Sketch of a Proof for Special Case of Thm L6.1

- *Theorem L6.3*: Suppose X is a random variable with pdf

$$f(x|\eta) = e^{\eta x - \psi(\eta)} h(x)\ dx$$

  where $\eta \in \mathcal{H} = \left\{\eta : \int e^{\eta x} h(x)\ dx < \infty\right\}$. If η is in the interior of $\mathcal{H}$, then $\mathsf{E}[X] = \psi'(\eta)$.
- *Proof of Theorem L6.3*: Since $\psi(\eta) = \ln \int e^{\eta x} h(x)\ dx$,

$$\begin{aligned}
\psi'(\eta) &= \frac{\frac{d}{d\eta}\int e^{\eta x} h(x)\ dx}{\int e^{\eta x} h(x)\ dx} \\
&= \frac{\int \frac{d}{d\eta} e^{\eta x} h(x)\ dx}{\int e^{\eta x} h(x)\ dx} \\
&= \frac{\int x e^{\eta x} h(x)\ dx}{\int e^{\eta x} h(x)\ dx} \\
&= \frac{\int x e^{\eta x} h(x)\ dx}{e^{\psi(\eta)}} = \int x e^{\eta x - \psi(\eta)} h(x)\ dx = \mathsf{E}[X].
\end{aligned}$$

Sketch of a Proof for Special Case of Thm L6.1

- We need to justify interchanging the order of the derivative and integral. This requires the following theorem.
- *Theorem L6.4* (Lebesgue Dominated Convergence Theorem): Let f_n be a sequence of measurable functions and let $f_n(x) \to f(x)$ for all x. If there exists an integrable function g such that $|f_n(x)| \leq g(x)$ for all n and x, then $\int f_n \, d\mu \to \int f \, d\mu$.

Sketch of a Proof for Special Case of Thm L6.1

- *Proof of Theorem L6.3 continued*: Let $M(\eta) = \int e^{\eta x} h(x)\ dx$.
- Then

$$\begin{aligned}
\frac{M(\eta_n) - M(\eta)}{\eta_n - \eta} &= \frac{\int e^{\eta_n x} h(x)\ dx - \int e^{\eta x} h(x)\ dx}{\eta_n - \eta} \\
&= \frac{\int \left(e^{\eta_n x} - e^{\eta x}\right) h(x)\ dx}{\eta_n - \eta} \\
&= \int e^{\eta x} \frac{\left(e^{(\eta_n - \eta)x} - 1\right)}{\eta_n - \eta} h(x)\ dx.
\end{aligned}$$

- For any $\delta > 0$, the following identity holds:

$$\left| \frac{e^{az} - 1}{z} \right| \leq \frac{e^{\delta |a|}}{\delta} \text{ when } |z| \leq \delta.$$

Sketch of a Proof for Special Case of Thm L6.1

- *Proof of Theorem L6.3 continued*: So we have

$$e^{\eta x}\left|\frac{\left(e^{(\eta_n-\eta)x}-1\right)}{\eta_n-\eta}\right| \le e^{\eta x}\frac{e^{\delta|x|}}{\delta} \le e^{\eta x}\frac{e^{\delta x}+e^{-\delta x}}{\delta} = \frac{1}{\delta}\left(e^{(\eta+\delta)x}+e^{(\eta-\delta)x}\right).$$

- Choose δ to be sufficiently small so that $\eta-\delta \in \mathcal{H}$ and $\eta+\delta \in \mathcal{H}$.
- Let $g(x) = \frac{1}{\delta}\left(e^{(\eta+\delta)x}+e^{(\eta-\delta)x}\right)$. Then

$$\int g(x)h(x)\,dx = \frac{1}{\delta}\left(\int e^{(\eta+\delta)x}h(x)\ dx + \int e^{(\eta-\delta)x}h(x)\ dx\right) < \infty.$$

Sketch of a Proof for Special Case of Thm L6.1

- *Proof of Theorem L6.3 continued*: So, the Lebesgue Dominated Convergence Theorem implies that

$$\lim_{\eta_n \to \eta} \frac{M(\eta_n) - M(\eta)}{\eta_n - \eta} = \lim_{\eta_n \to \eta} \int e^{\eta x} \frac{\left(e^{(\eta_n - \eta)x} - 1\right)}{\eta_n - \eta} h(x)\ dx$$
$$= \int e^{\eta x} \lim_{\eta_n \to \eta} \frac{\left(e^{(\eta_n - \eta)x} - 1\right)}{\eta_n - \eta} h(x)\ dx$$

- Computing the limits, we obtain

$$M'(\eta) = \int e^{\eta x} x h(x)\ dx$$
$$= \int x e^{\eta x} h(x)\ dx$$
$$= \int \frac{d}{d\eta} e^{\eta x} h(x)\ dx.$$