Feynman Liang
STAT 210A
Lecture #2
8/30/2016

# 1 Exponential Family

See Keener Chapter 2.

**Definition 1.1.** An *s-parameter exponential family* is a family of PDFs $\{p_\eta : \eta \in \Xi\}$ wrt $\mu$ on $X$ of the form

$$p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x) \tag{1.1}$$

- $T : X \to \mathbb{R}^s$ is called the *sufficient statistic*
- $h : X \to \mathbb{R}_+$ is called the *carrier density*
- $\eta \in \Xi \subset \mathbb{R}^s$ is called the *natural parameters*
- $A(\eta) : \Xi \to \mathbb{R}$ is called the *cumulant generating function*

Because $\int p_\eta(x) d\mu(x) = 1$, $A(\eta) = \log \int e^{\eta' T(x)} h(x) d\mu(x)$ is a *normalizing constant*.

The *natural parameter space* $\Xi := \{\eta : A(\eta) < \infty\}$ is restricted such that $p_\eta$ is normalizable.

wlog $h(x) = p_0(x) = e^{-A(0)} h(x)$, can reparameterize such that $0 \in \Xi$. This motivates an affine decomposition: view $h(x)$ as a point and $\log \eta$ as a $s$-dimensional hyperplane in function space:

$$\mathcal{H} = \{\log p_\theta \perp \sum_{i=1}^{s} \eta_i \underbrace{T_i(x)}_{\text{the } s \text{ degrees of freedom}} : \eta \in \mathbb{R}^s\} \tag{1.2}$$

$p_\eta(x)$ is a "hyperplane of PDFs in function space."

[See exponential tilting demo.]

**Example 1.2** (Keener 2.2). $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

$$\begin{aligned} p_\theta(x) &= (2\pi\sigma^2)^{-1/2} \exp\left\{ \frac{(x-\mu)^2}{2\sigma^2} \right\} && (1.3) \\ &= \exp\left\{ \frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \left( \frac{\mu^2}{2\sigma^2} + \frac{1}{2} \log \sigma^2 \right) \right\} \frac{1}{\sqrt{2\pi}} && (1.4) \end{aligned}$$

Therefore

$$\left.\begin{aligned}
\theta &= (\mu, \sigma^2) \\
\eta(\theta) &= \left(\tfrac{\mu}{\sigma^2}, \tfrac{-1}{2\sigma^2}\right) \\
T(x) &= (x, x^2) \\
h(x) &= \tfrac{1}{\sqrt{2\pi}} \\
B(\theta) &= A(\eta(\theta)) = \tfrac{\mu^2}{2\sigma^2} + \tfrac{1}{2}\log\sigma^2
\end{aligned}\right\}\textit{canonical form}\text{ for exponential family} \tag{1.5}$$

The normal distribution is exponential family with constant carrier density$h(x)$ from Lebesgue measure and sufficient statistics $(x, x^2)$.

**Example 1.3** (Keener 2.3)**.** $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$.

$$p_\theta(X) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(\mu - x_i)^2}{2\sigma^2}\right\} \tag{1.6}$$

$$= (2\pi\sigma^2)^{-N/2} \exp\left\{\sum_{i=1}^{N}\left(\frac{\mu}{\sigma^2}x_i - \frac{1}{2\sigma^2}x_i^2\right) + B(\theta)\cdot N\right\} \tag{1.7}$$

$$T(X) = \begin{bmatrix} \sum_i x_i \\ \sum_i x_i^2 \end{bmatrix} = \sum_i T(x_i) \tag{1.8}$$

This shows that the sufficient statistics of IID exponential family data is the sum of $T(X)$. More generally

$$X_1, \cdots, X_n \sim p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x) \tag{1.9}$$

$$p_\eta(\{x_i\}_{i=1}^n) = \prod_i p_\eta(x_i) = \exp\left\{\eta' \sum_i T(x_i) - nA(\eta)\right\} \prod_i h(x_i) \tag{1.10}$$

**Example 1.4.** $X \sim \text{Binom}(n\theta)$.

$$p_\theta(x) = \theta^x (1-\theta)^{n-x} \binom{n}{x} = \left(\frac{\theta}{1-\theta}\right)^x (1-\theta)^n \binom{n}{x} \tag{1.11}$$

$$= \exp\left\{\underbrace{x}_{T(x)} \underbrace{\log\frac{\theta}{1-\theta}}_{\eta} + \underbrace{n\log(1-\theta)}_{A(\eta)}\right\} \underbrace{\binom{n}{x}}_{h(x)} \tag{1.12}$$

**Example 1.5** (Logistic Regression)**.** $(x_i, y_i)_{i=1}^n$, $y_i \overset{\text{indep}}{\sim} \text{Bern}(\theta_i)$, $x_i \in \mathbb{R}^s$ fixed.

Model $\log\frac{\theta_i}{1-\theta_i} = \beta' x_i$.

$$
\begin{aligned}
p_\theta(x) &= \prod_{i=1}^{n} \theta_i^{y_i}(1-\theta_i)^{1-y_i} && (1.13)\\
&= \prod_{i=1}^{n} \exp\{\beta' x_i y_i - \log(1+e^{-\beta' x_i})\} && (1.14)\\
&= \exp\{\beta' \underbrace{(\sum_i x_i y_i)}_{T(y)=\vec{\vec{X}}\vec{y}} - \sum_i \log(1+e^{\beta' x_i})\} && (1.15)
\end{aligned}
$$

**Example 1.6.** $X \sim U[0,\theta] \notin$ ExpF because $h(x)$ does not depend on $\theta$! Support of $X$ needs to be independent of $\eta$.

# 2 Differential Identities

$$
\begin{aligned}
e^{A(\eta)} &= \int e^{\eta' T(x)} h(x) d\mu(x) && (2.1)\\
\frac{\partial}{\partial \eta_j} e^{A(\eta)} &= \frac{\partial}{\partial \eta_j} \int e^{\eta' T(x)} h(x) d\mu(x) && (2.2)
\end{aligned}
$$

$$
\eta \in \Xi^\circ \implies \text{dominated conv.} \implies \frac{\partial}{\partial \eta_j} \int e^{\eta' T(x)} h(x) d\mu(x) = \int \frac{\partial}{\partial \eta_j} e^{\eta' T(x)} h(x) d\mu(x) \tag{2.3}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \eta_j} e^{A(\eta)} &= \int T_j(x) e^{\eta' T(x)} h(x) d\mu(x) && (2.4)\\
\frac{\partial}{\partial \eta_j} &= \int T_j(x) e^{\eta' T(x) - A(\eta)} h(x) d\mu(x) && (2.5)\\
\frac{\partial}{\partial \eta_j}(\eta) &= \mathbb{E}_\eta[T_j(x)] && (2.6)
\end{aligned}
$$

More generally

$$
\nabla \underbrace{A(\eta)}_{\text{“mean parameterization” } \mu(\eta)=A(\eta)} = \mathbb{E}_\eta[T(x)] \tag{2.7}
$$

$$
\frac{\partial^{k_1+\cdots+k_s}}{\partial \eta_1^{k_1} \cdots \partial \eta_s^{k_s}} e^{A(\eta)} = \mathbb{E}_\eta[T_1^{k_1}(x) \cdot \cdots \cdot T_s^{k_s}(x)] e^{A(\eta)} \tag{2.8}
$$

**Definition 2.1.** The *moment generating function (MGF)*

$$
\begin{aligned}
M_{T(x)} &= \mathbb{E}_\eta[e^{u' T(x)}] && (2.9)\\
&= \int e^{u' T(x)} e^{\eta' T(x) - A(\eta)} h(x) dx && (2.10)\\
&= e^{A(\eta+\mu)-A(\eta)} \cancelto{1}{\int e^{(\eta+\mu)' T(x) - A(\eta+\mu)} h(x) dx} && (2.11)
\end{aligned}
$$

$e^{A(\eta+u)-A(\eta)}$ is the MGF of $T(X)$ where $X \sim p_\eta$
$\nabla^2 A(\eta) = \mathrm{Var}_\eta(T())$.