Lecture 3

9/3/2020

Outline

1) Review
2) Exponential families
3) Interpretation: exponential tilting
4) Demo
5) Examples

Exponential Families

An <u>s-parameter exponential family</u> is a family $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ with densities p_η wrt a common measure μ on $\mathcal{X}$ [$\mathcal{X}$ not nec. in $\mathbb{R}^n$] of the form

$$p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x) \quad , \quad \text{where}$$

$T : \mathcal{X} \to \mathbb{R}^s$ <u>sufficient statistic</u>

$h : \mathcal{X} \to \mathbb{R}$ <u>carrier / base density</u>

$\eta \in \Xi \subseteq \mathbb{R}^s$ <u>natural parameter</u>

$A : \mathbb{R}^s \to \mathbb{R}$ <u>cumulant-generating fcn</u> (<u>cgf</u>) or <u>normalizing const</u>

<u>Note</u> The cgf $A(\eta)$ is totally determined by h and T, since we must always have $\int_{\mathcal{X}} p_\eta \, d\mu = 1, \; \forall \eta.$

$$\Rightarrow A(\eta) = \log\left[\int_{\mathcal{X}} e^{\eta' T(x)} h(x)\, d\mu(x)\right]$$

p_η is normalizable iff $A(\eta) < \infty$

The natural parameter space is the set of all allowable (normalizable) η:

$$\Xi_1 = \{\eta : A(\eta) < \infty\}$$

We say $\mathcal{P}$ is in canonical form if $\Xi = \Xi_1$

Note Ξ_1 determined by T, h, μ

We could take $\Xi \subsetneq \Xi_1$ if we wanted

$A(\eta)$ is convex $\Rightarrow$ Ξ_1 is convex (HW 1 Prob. 2)

Sometimes it is more convenient to use a different parameterization:

$$p_\theta(x) = e^{\eta(\theta)' T(x) - B(\theta)} h(x)$$

$$B(\theta) = A(\eta(\theta))$$

Many, many examples, sometimes requires massaging to see that they are exp. fam.s:

Ex. 2.2: $X \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R}$ $\sigma^2 > 0$

Let $\theta = (\mu, \sigma^2)$

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\mu - x)^2/2\sigma^2}$$

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

$$\eta(\theta) = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix} \qquad T(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix} \qquad h(x) = 1$$

$$B(\theta) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)$$

In terms of η:

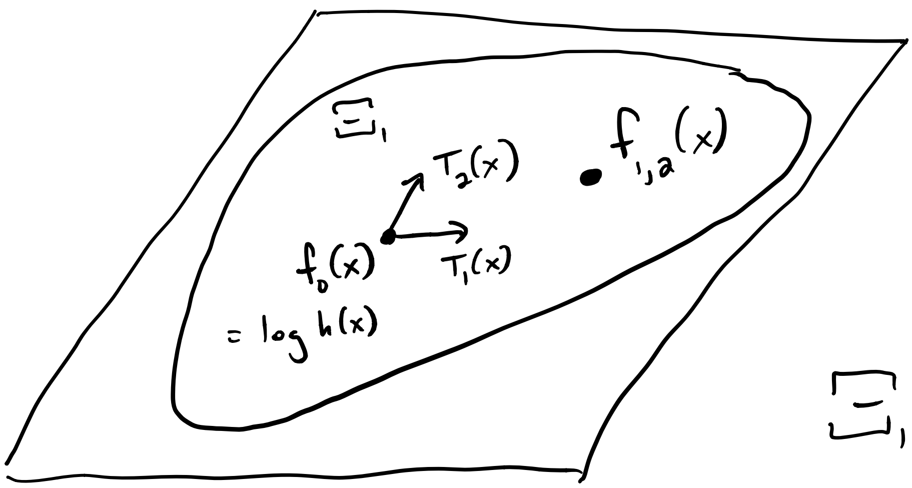
$$p_\eta(x) = e^{\eta'\binom{x}{x^2} - A(\eta)} \quad ; \quad A(\eta) = \frac{-\eta_1^2}{4\eta_2} + \frac{1}{2}\log(-\pi/\eta_2)$$

It can be useful to analogize $\{p_\eta : \eta \in \Xi\}$ to an s-dim. hyperplane in ∞-dimensional log-density space:

Assume wlog $h(x) > 0 \quad \forall x$

(o.w. we could truncate to $\tilde{\mathcal{X}} = \{x \in \mathcal{X} : h(x) > 0\}$)

Let $f_\eta(x) = \log h(x) + \eta' T(x)$



$e^{f_\eta(x)}$ is always a density, but not always normalizable.

$$\Xi_1 = \left\{\eta : \int e^{f_\eta} d\mu < \infty\right\}$$

Then
$$p_\eta(x) = e^{f_\eta(x)} \Big/ \int e^{f_\eta(x)} d\mu(x)$$
$$= e^{\eta' T(x)} h(x) \Big/ e^{A(\eta)}$$

for $\eta \in \Xi \subseteq \Xi_1 = \{\eta : e^{f_\eta} \text{ normiable}\}$

Functional form of p_η is in many ways not unique: for example we could always

a) Reformulate so $h(x) = 1$:

$$\mu \rightsquigarrow \tilde{\mu} \quad \text{with} \quad \frac{d\tilde{\mu}}{d\mu} = h$$
$$h \rightsquigarrow \tilde{h}(x) \equiv 1$$

b) Re-parameterize so $0 \in \Xi$:

Take some $\eta_0 \in \Xi$

$$\eta \rightsquigarrow \tilde{\eta} = \eta - \eta_0$$
$$h \rightsquigarrow \tilde{h}(x) = p_{\eta_0}(x)$$
$$A \rightsquigarrow \tilde{A}(\tilde{\eta}) = A(\eta_0 + \tilde{\eta}) - A(\eta_0)$$

c) For $c \in \mathbb{R}^s$, invertible $D \in \mathbb{R}^{s \times s}$,

change sufficient statistic to

$$T \rightsquigarrow \tilde{T}(x) = c + DT(x)$$
$$\eta \rightsquigarrow \tilde{\eta} = (D^{-1})' \eta$$
$$A \rightsquigarrow \tilde{A} \quad \text{as appropriate}$$

etc...

Interp:

- Start with a carrier density $h(x)$
- Apply exponential tilt:
  1) multiply by $e^{\eta' T(x)}$
  2) re-normalize to prob. density

Ex. fam in canonical form is all normalizable tilts of $h(x)$ using linear combos of $T_1(x), \dots, T_s(x)$

Ex. 2.3: $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$p_\theta(x) = \prod_{i=1}^n p_\theta^{(1)}(x_i)$$

$$= \exp\left\{ \sum_{i=1}^n \frac{\mu}{\sigma^2} x_i - \frac{1}{2\sigma^2} x_i^2 - \left(\frac{\mu}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)\right)\right\}$$

$$= \exp\left\{ \frac{\mu}{\sigma^2} \sum x_i - \frac{1}{2\sigma^2} \sum x_i^2 - n(\dots) \right\}$$

$$\eta(\theta) = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix} \qquad T(x) = \begin{pmatrix} \sum x_i \\ \sum x_i^2 \end{pmatrix}$$

$$B(\mu, \sigma^2) = n B^{(1)}(\mu, \sigma^2)$$

[No accident these came out so similar: an iid sample from an exp fam is also an exp fam]

Suppose $X_1, \dots, X_n \overset{iid}{\sim} p_\eta^{(1)}(x) = e^{\eta' T(x) - A(\eta)} h(x)$

Then $X \sim \prod_{i=1}^n e^{\eta' T(x_i) - A(\eta)} h(x_i)$

$$= \exp\{ \eta' \underbrace{\sum T(x_i)}_{\text{suff. stat.}} - \underbrace{n A(\eta)}_{\text{cgf}} \} \underbrace{\prod_i h(x_i)}_{\text{carrier}}$$

The sufficient statistic $T(X)$ follows a related exp fam too:

Suppose $X \in \mathcal{X}$, $T(X) \in \mathcal{T} = T(\mathcal{X})$,

μ base measure (wlog $h \equiv 1$)

For $B \subseteq \mathcal{T}$ define $\nu(B) = \mu(T^{-1}(B))$

(ν called "push-forward" measure)

Then $T(X) \sim q_\eta(t) = e^{\eta' t - A(\eta)}$ wrt ν

Discrete case:

$$\mathbb{P}_\eta(T(X) \in B) = \sum_{x:\, T(x) \in B} e^{\eta' T(x) - A(\eta)} \mu(\{x\})$$

$$= \sum_{t \in B} \sum_{x:\, T(x) = t} e^{\eta' t - A(\eta)} \mu(\{x\})$$

$$= \sum_{t \in B} e^{\eta' t - A(\eta)} \nu(\{t\})$$

More generally, ν satisfies

$$\int_{\mathcal{X}} f(T(x))\, d\mu(x) = \int_{\mathcal{T}} f(t)\, d\nu(t), \quad \forall \text{ ("nice") } f$$

$$\mathbb{P}_\eta(T(X) \in B) = \int_{\mathcal{X}} 1\{T(x) \in B\} e^{\eta' T(x) - A(\eta)} d\mu(x)$$

$$= \int_{\mathcal{T}} 1\{t \in B\} e^{\eta' t - A(\eta)} d\nu(t)$$

$\Rightarrow T \sim e^{\eta' t - A(\eta)}$ density wrt ν

More examples

Binomial $\quad X \sim \text{Binom}(n, \theta)$

$$p_\theta(x) = \theta^x (1-\theta)^{n-x} \binom{n}{x}$$

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

$$= \exp\left\{ \log\left(\frac{\theta}{1-\theta}\right) \cdot x + n \log(1-\theta) \right\} \binom{n}{x}$$

$$\eta(\theta) = \log\left(\frac{\theta}{1-\theta}\right) \quad \text{"log odds ratio"}$$

Poisson $\quad X \sim \text{Pois}(\theta) = \dfrac{\lambda^x e^{-\lambda}}{x!} \qquad x \in 0, 1, \ldots$

$$p_\lambda(x) = \exp\left\{ (\log \lambda)\, x - \lambda \right\} \frac{1}{x!}$$

$$\eta(\lambda) = \log \lambda$$

Practically everything else on wikipedia too:

Beta, Gamma, Multinom., Dirichlet, Pareto, Wishart...

Differential Identities

Write $e^{A(\eta)} = \int e^{\eta' T(x)} h(x)\, d\mu(x) \qquad (*)$

We can derive lots of useful identities by differentiating $(*)$ on both sides, pulling derivative inside $\int$ [not always allowed]

Keener Thm 2.4 for $f : \mathcal{X} \to \mathbb{R}$ let

$$\Xi_f = \{ \eta \in \mathbb{R}^s : \int |f| e^{\eta' T} h\, d\mu < \infty \}$$

Then $g(\eta) = \int f e^{\eta' T} h\, d\mu$ has cts partial derivatives of all orders for $\eta \in \Xi_f^\circ$. & we can get them by differentiating under the $\int$ sign.

$\Rightarrow$ on Ξ_1°, $A(\eta)$ has all partial derivatives

Differentiate once:

$$\frac{\partial}{\partial \eta_j} e^{A(\eta)} = \frac{\partial}{\partial \eta_j} \int e^{\eta' T(x)} h(x)\, d\mu(x)$$

$$\cancel{e^{A(\eta)}} \frac{\partial A}{\partial \eta_j}(\eta) = \int T_j(x)\, e^{\eta' T(x) - A(\eta)} h(x)\, d\mu(x)$$

$$\Rightarrow \frac{\partial A}{\partial \eta_j}(\eta) = \mathbb{E}_\eta[T_j(X)]$$

$$\nabla A(\eta) = \mathbb{E}_\eta[T(X)]$$

Diff twice:

$$\frac{\partial^2}{\partial \eta_j \partial \eta_k} e^{A(\eta)} = \frac{\partial^2}{\partial \eta_j \partial \eta_k} \int e^{\eta' T} h \, d\mu$$

$$\cancel{e^{A(\eta)}} \Big(\frac{\partial^2 A}{\partial \eta_j \partial \eta_k} + \underbrace{\frac{\partial A}{\partial \eta_j}}_{\mathbb{E}[T_j]} \underbrace{\frac{\partial A}{\partial \eta_k}}_{\mathbb{E}[T_k]} \Big) = \underbrace{\int T_j T_k \, e^{\eta' T - A(\eta)} h \, d\mu}_{\mathbb{E}[T_j T_k]}$$

$$\frac{\partial^2 A}{\partial \eta_j \partial \eta_k}(\eta) = \mathrm{Cov}_\eta(T_j, T_k)$$

$$\nabla^2 A(\eta) = \mathrm{Var}_\eta(T(x)) \in \mathbb{R}^{s \times s}$$

Example: Poisson : $T(x) = X$, $\eta(\lambda) = \log \lambda$

$$B(\lambda) = \lambda \Rightarrow A(\eta) = e^\eta$$

$$\mathbb{E}_\eta[X] = \frac{d}{d\eta} e^\eta = e^\eta = \lambda$$

$$\mathrm{Var}_\eta(X) = \frac{d^2}{d\eta^2} e^\eta = e^\eta = \lambda$$

Moment-generating function

Differentiating $(*)$ repeatedly we get

$$e^{-A(\eta)} \frac{\partial^{k_1+\cdots+k_s}}{\partial \eta_1^{k_1} \cdots \partial \eta_s^{k_s}} \left(e^{A(\eta)}\right) = \mathbb{E}_\eta\left[T_1^{k_1} \cdots T_s^{k_s}\right]$$

That is because $M_\eta^T(u) = e^{A(\eta+u) - A(\eta)}$ is the <u>moment-generating function</u> (mgf) of $T(X)$ when $X \sim p_\eta$

$$M_\eta^{T(X)}(u) = \mathbb{E}_\eta\left[e^{u'T(X)}\right]$$

$$= \int e^{u'T} e^{\eta'T - A(\eta)} h \, d\mu$$

$$= e^{A(\eta+u) - A(\eta)} \underbrace{\int e^{(\eta+u)'T - A(\eta+u)} h \, d\mu}_{=1}$$