

# Lecture 3

8/31/2023

## Outline

- 1) Review
- 2) Exponential families
- 3) Differential identities
- 4) MGF

# Exponential Families

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

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

$$T : \mathcal{X} \rightarrow \mathbb{R}^s$$

sufficient statistic

$$h : \mathcal{X} \rightarrow \mathbb{R}$$

carrier / base density

$$\eta \in \Xi \subseteq \mathbb{R}^s$$

natural parameter

$$A : \mathbb{R}^s \rightarrow \mathbb{R}$$

log-partition function

or normalizing const

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

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

The natural parameter space is the set of all  $\eta$  that give us normalizable  $p_\eta$

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

Note  $\Xi_1$  determined by  $T, h, \eta$

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

$A(\eta)$  is always convex

$\Rightarrow \Xi_1$  is convex (HW1 Prob. 2)

Example: Poisson

$$X \sim \text{Pois}(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x \in 0, 1, \dots$$

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

$$\eta(\lambda) = \log \lambda \quad T(x) = x$$

$$A(\lambda) = \lambda = e^\eta \quad h(x) = \frac{1}{x!}$$

# Differential Identities

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

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

Keener Thm 2.4 for  $f: \mathcal{X} \rightarrow \mathbb{R}$  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  $\eta \in \Xi_f^0$ . & we can get them by differentiating under the  $\int$  sign.

$\Rightarrow$  on  $\Xi_f^0$ ,  $A(\eta)$  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:

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

$$\cancel{e^{A(\eta)}} \left( \frac{\partial^2 A}{\partial \eta_i \partial \eta_k} + \underbrace{\frac{\partial A}{\partial \eta_i}}_{\mathbb{E}[T_i]} \underbrace{\frac{\partial A}{\partial \eta_k}}_{\mathbb{E}[T_k]} \right) = \underbrace{\int T_i T_k e^{\eta' T - A(\eta)} h d\mu}_{\mathbb{E}[T_i T_k]}$$

$$\frac{\partial^2 A}{\partial \eta_i \partial \eta_k}(\eta) = \text{Cov}_\eta(T_i, T_k)$$

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

Example: Poisson:  $T(x) = x$ ,  $A(\eta) = e^\eta (= \lambda)$

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

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

NB: We would get wrong answer by differentiating wrt  $\lambda$

## Moment-generating function

We can get  $k^{\text{th}}$  order moments of  $T(X)$  by

1) Differentiating (\*)  $k$  times, then

2) Dividing by  $e^{A(\eta)}$

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

$$\begin{aligned} M_{\eta}^{T(X)}(u) &= \mathbb{E}_{\eta} [e^{u'T(X)}] \\ &= \int e^{u'T} e^{\eta'T - A(\eta)} h d\mu \\ &= e^{A(\eta+u) - A(\eta)} \underbrace{\int e^{(\eta+u)'T - A(\eta+u)} h d\mu}_{=1} \end{aligned}$$

Useful for

- finding moments
- finding dist. of sums of indep. RVs

## Cumulant-generating function

$$K_{\eta}^T(u) = \log M_{\eta}^T(u) = A(\eta+u) - A(\eta) \quad (A \text{ is sometimes called cgf})$$

## Other Parameterizations

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

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

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

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

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

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

Natural parameterization

$$p_{\eta}(x) = e^{\eta' \begin{pmatrix} x \\ x^2 \end{pmatrix} - A(\eta)}$$

$$A(\eta) = \frac{-\eta_1^2}{4\eta_2} + \frac{1}{2} \log(-\pi/\eta_2)$$

## More examples

### Binomial

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

$$\begin{aligned} p_{\theta}(x) &= \theta^x (1-\theta)^{n-x} \binom{n}{x} & x = 0, \dots, n \\ &= \left(\frac{\theta}{1-\theta}\right)^x (1-\theta)^n \binom{n}{x} \\ &= \exp \left\{ \log\left(\frac{\theta}{1-\theta}\right) \cdot x + n \log(1-\theta) \right\} \binom{n}{x} \\ \eta(\theta) &= \log\left(\frac{\theta}{1-\theta}\right) \quad \text{"log odds ratio"} \end{aligned}$$

### Beta

$$X \sim \text{Beta}(\alpha, \beta)$$

$$\begin{aligned} p_{\alpha, \beta}(x) &= x^{\alpha-1} (1-x)^{\beta-1} / B(\alpha, \beta) & \leftarrow \text{Beta function} \\ &= \exp \left\{ \alpha \log x + \beta \log(1-x) - \log B(\alpha, \beta) \right\} \frac{1}{x(1-x)} \end{aligned}$$

$$\eta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad T(x) = \begin{pmatrix} \log x \\ \log(1-x) \end{pmatrix} \quad h(x) = \frac{1}{x(1-x)}$$

Practically everything else on wikipedia too:

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



## Interpretation: Exponential tilting

Can think of  $p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x)$  as  
an exponential tilt of the carrier  $h(x)$

1) Start with carrier  $h(x)$

2) Multiply by  $e^{\eta' T(x)}$

3) Re-normalize by  $e^{-A(\eta)}$

$T(x) = (T_1(x), \dots, T_s(x))$  gives linear space of directions  
in which we can tilt  $h(x)$

$\Xi_\eta =$  all tilts after which normalization is possible

$\Rightarrow$  Decomposition into  $\eta, T, h, A$  very non-unique

1) Only  $\text{span}(T_1, \dots, T_s)$  matters

2) Could absorb  $h$  into  $\mu$  ( $d\nu(x) = h d\mu(x)$ )  
(wlog  $h(x) \equiv 1$  if we want)

3) Can add constant to  $T(x)$

## Repeated Sampling

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

Then  $X = (X_1, \dots, X_n)$  comes from a closely related family

$$\begin{aligned} p_{\eta}^{(n)}(x) &= \prod_{i=1}^n e^{\eta' T(x_i) - A(\eta)} h(x_i) \\ &= \exp \left\{ \underbrace{\eta' \sum_{i=1}^n T(x_i)}_{\text{suff stat.}} - \underbrace{n A(\eta)}_{\text{log-part.fcn.}} \right\} \underbrace{\prod_{i=1}^n h(x_i)}_{\substack{\text{carrier density} \\ (\text{wrt } \mu^n \text{ on } \mathcal{X}^n)}} \end{aligned}$$

Important property!

This means  $\sum_{i=1}^n T(x_i) \in \mathbb{R}^S$  is an effective summary of a potentially very large sample  $X \in \mathcal{X}^n$

We will often analyze  $T(x)$  as a proxy for the whole data set.

## Distribution of $T(x)$

Suppose  $X \sim p_\eta(x) = e^{\eta' T(x) - A(\eta)}$  wrt  $\mu$   
(wlog  $h \equiv 1$ )

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

where  $\nu$  is the measure  $\mu$  "pushed forward" through  $T: \mathcal{X} \rightarrow \mathbb{R}^S$

$$\nu(B) \triangleq \mu(\{x: T(x) \in B\})$$

$$\begin{aligned} P_\eta(T(X) \in B) &= \int 1_B(T(x)) e^{\eta' T(x) - A(\eta)} d\mu(x) \\ &= \int 1_B(t) e^{\eta' t - A(\eta)} d\nu(t) \end{aligned}$$

Simplest in discrete case: (drop  $h \equiv 1$  assumption)

$$\begin{aligned} P_\eta(T(X) = t) &= \sum_{x: T(x) = t} e^{\eta' T(x) - A(\eta)} h(x) \mu(\{x\}) \\ &= e^{\eta' t - A(\eta)} \underbrace{\sum_{x: T(x) = t} h(x) \mu(\{x\})}_{\nu(\{t\})} \end{aligned}$$