Lecture 3

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 $\mathcal{X}$ such that

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

where

$T : \mathcal{X} \to \mathbb{R}^s$ is a sufficient statistic

$h : \mathcal{X} \to \mathbb{R}$ is the carrier/base density

$\eta \in \Xi \subseteq \mathbb{R}^s$ is the natural parameter

$A : \mathbb{R}^s \to \mathbb{R}$ is the log-partition function or normal/Bayes constant

Note: The function $A(\cdot)$ is totally determined by $h$ and $T$, since we must always have

$$ \int_{\mathcal{X}} p_\eta \, d\mu = 1, \quad \forall \eta. $$

$$ \Rightarrow A(\eta) = \log \left[ \int_{\mathcal{X}} e^{\eta' T(x)} h(x) \, d\mu(x) \right] \leq \infty $$
The natural parameter space is the set of all \( \eta \) that give us normalizable \( p_x \)

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

Note \( \Xi_1 \) determined by \( T, h, \mu \).

We could take \( \Xi \neq \Xi_1 \) if we wanted.

\( A(\eta) \) is always convex.

\[ \Xi_1 \] is convex (HW1 Prob. 2)

**Example: Poisson**

\[ X \sim \text{Pois}(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x \in \mathbb{N} \]

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

\[ \eta(\lambda) = \log \lambda \quad T(x) = x \]

\[ A(\lambda) = \lambda = e^\eta \quad h(x) = \frac{1}{x!} \]
Differential Identities

Write \( e^{A(\gamma)} = \int e^{\gamma^T T(x)} h(x) \, d\mu(x) \) (*)

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

Keener Thm 2.4 for \( f: X \rightarrow \mathbb{R} \) let

\[
\Xi_f = \{ \gamma \in \mathbb{R}^s : \int \text{exp}^{\gamma^T h} \, d\mu < \infty \}
\]

Then \( g(\gamma) = \int \text{exp}^{\gamma^T} \, h \, d\mu \) has cts partial derivatives of all orders for \( \gamma \in \Xi_f^0 \), & we can get them by differentiating under the \( \int \) sign.

\( \therefore \) on \( \Xi_f^0 \), \( A(\gamma) \) has all partial derivatives

Differentiate once:

\[
\frac{\partial}{\partial \gamma_j} e^{A(\gamma)} = \frac{\partial}{\partial \gamma_j} \int e^{\gamma^T T(x)} h(x) \, d\mu(x)
\]

\[
e^{A(\gamma)} \frac{\partial A}{\partial \gamma_j} (\gamma) = \int T_j(x) e^{\gamma^T T(x)} - A(\gamma)
\]

\( \Rightarrow \frac{\partial A}{\partial \gamma_j} (\gamma) = \mathbb{E}_\gamma[ T_j(x) ]
\]

\( \nabla A(\gamma) = \mathbb{E}_\gamma[ T(x) ] \)
Diff twice:

\[
\frac{\partial^2}{\partial \gamma_i \partial \gamma_k} \mathcal{A}(\eta) = \frac{\partial^2}{\partial \gamma_i \partial \gamma_k} \int e^{\eta^T T} \, h \, d\mu
\]

\[
e^{\frac{\partial^2 A}{\partial \gamma_i \partial \gamma_k} \bigg( \frac{\partial^2 A}{\partial \gamma_j \partial \gamma_k} + \frac{\partial A}{\partial \gamma_j} \frac{\partial A}{\partial \gamma_k} \bigg)} = \int \frac{T_j T_k}{E[T_j] E[T_k]} e^{\eta^T T - A(\eta)} \, h \, d\mu
\]

\[
\frac{\partial^2 A}{\partial \gamma_i \partial \gamma_k} (\eta) = \text{Cov}_{\eta} (T_j, T_k)
\]

\[
\nabla^2 A(\eta) = \text{Var}_{\eta} (T(x)) \in \mathbb{R}^{5 \times 5}
\]

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

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

\[
\text{Var}_{\eta} (X) = \frac{d^2}{d\eta^2} e^\eta = e^\eta = \lambda
\]

NB: We would get wrong answer by differentiating wrt \( \lambda \)
Moment-generating function

We can get \( k \)th order moments of \( T(X) \) by

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

2) Dividing by \( e^{A(x)} \)

That is because \( M^T_{\eta}(u) = e^{A(\eta + u)} - A(\eta) \)

is the moment-generating function (mgf) of \( T(X) \) when \( X \sim P_\eta \)

\[
M^T_{\eta}(u) = \mathbb{E}_\eta \left[ e^{u' T(X)} \right] 
= \int e^{u' T - A(\eta)} \, h \, d\mu 
= e^{A(\eta + u)} - A(\eta) \int e^{(\eta + u)' T - A(\eta + u)} \, h \, d\mu
\]

Useful for

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

Cumulant-generating function

\[
K^T_{\eta}(u) = \log M^T_{\eta}(u) = A(\eta + u) - A(\eta) \quad (A \text{ is sometimes called cgf})
\]
Other Parameterizations

Sometimes it is more convenient to use a different parameterization:

\[ p_\theta(x) = e^{\gamma(\theta)'T(x) - B(\theta)} h(x) \]
\[ B(\theta) = A(\gamma(\theta)) \]

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 \( \Theta = (\mu, \sigma^2) \)

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

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

\[ \gamma(\theta) = \left( \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right) \quad T(x) = (x^2) \quad h(x) = 1 \]

\[ B(\theta) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2} \log \left( 2\pi \sigma^2 \right) \]

**Natural parameterization**

\[ p_\gamma(x) = e^{\gamma'(x^2)} - A(\gamma) \]

\[ A(\gamma) = \frac{-\gamma^2}{4\gamma^2} + \frac{1}{2} \log \left( \frac{-\pi}{\gamma^2} \right) \]

**Ala II Elogtt**
More examples

**Binomial**

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

\[
\begin{align*}
\rho_\theta(x) &= \theta^x (1-\theta)^{n-x} \binom{n}{x} \\
&= \left(\frac{\theta}{1-\theta}\right)^x (1-\theta)^n \binom{n}{x} \\
&= \exp\left\{ \log\left(\frac{\theta}{1-\theta}\right) \cdot x + n \log(1-\theta) \right\} \binom{n}{x} \\
\eta(\theta) &= \log\left(\frac{\theta}{1-\theta}\right) \quad "\text{log odds ratio}" \\
\end{align*}
\]

**Beta**

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

\[
\begin{align*}
\rho_{\alpha,\beta}(x) &= x^{\alpha-1} (1-x)^{\beta-1} / B(\alpha,\beta) \\
&= \exp\left\{ \alpha \log x + \beta \log(1-x) - \log B(\alpha,\beta) \right\} \frac{1}{x(1-x)} \\
\eta(x) &= \left(\begin{array}{c}
\alpha \\
\beta
\end{array}\right) \\
T(x) &= \left(\begin{array}{c}
\log x \\
\log(1-x)
\end{array}\right) \\
h(x) &= \frac{1}{x(1-x)}
\end{align*}
\]

Practically everything else on wikipedia too:

Beta, Gamma, Multinom., Dirichlet, Pareto, Wishart...
Interpretation: Exponential tilting

Can think of $p_{\gamma}(x) = e^{\gamma T(x)} - A(\gamma) \cdot h(x)$ as an exponential tilt of the carrier $h(x)$.

1) Start with carrier $h(x)$
2) Multiply by $e^{\gamma T(x)}$
3) Re-normalize by $e^{-A(\gamma)}$

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

$\mathbb{E}_{\gamma_1}$ = all tilts after which normalization is possible

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

1) Only $\text{span}(T_1, \ldots, T_s)$ matters
2) Could absorb $h$ into $\mu$ (d$v(x) = h \cdot d\mu(x)$) ($\text{wlog} \ h(x) = 1$ if we want)
3) Can add constant to $T(x)$
Repeated Sampling

Suppose \( X_1, \ldots, X_n \sim \rho^{(i)}_\gamma (x) = e^{\gamma^T T(x) - A(\gamma) h(x)} \).

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

\[
\rho^{(n)}_\gamma (x) = \frac{\prod_{i=1}^n e^{\gamma^T T(x_i) - A(\gamma) h(x_i)}}{\exp \{ \gamma^T \sum_{i=1}^n T(x_i) - n A(\gamma) \} \prod_{i=1}^n h(x_i)}
\]

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 \( X \in \mathcal{X} \).

We will often analyze \( T(X) \) as a proxy for the whole data set.
Distribution of $T(X)$

Suppose $X \sim p(x) = e^{\tau^T T(x) - A(\tau)}$ wrt $\mu$  
($\omega \mu$ log $\lambda \equiv 1$)

Then $T(X) \sim q(t) = e^{\tau^T t - A(\tau)}$ wrt $\nu$,

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

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

$$
\begin{align*}
\mathbb{P}_\tau(T(X) \in B) &= \int_B (T(x)) e^{\tau^T T(x) - A(\tau)} d\mu(x) \\
&= \int_B (t) e^{\tau^T t - A(\tau)} d\nu(t)
\end{align*}
$$

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

$$
\begin{align*}
\mathbb{P}_\tau(T(X) = t) &= \sum_{x : T(x) = t} e^{\tau^T T(x) - A(\tau)} h(x) \mu(\{x\}) \\
&= e^{\tau^T t - A(\tau)} \sum_{x : T(x) = t} h(x) \mu(\{x\}) \\
&= e^{\tau^T t - A(\tau)} \nu(\{t\})
\end{align*}
$$