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 $X$ [X not nec. in $\mathbb{R}^n$] of the form

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

where $T : X \to \mathbb{R}^s$ sufficient statistic

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

$\eta \in \Xi \subset \mathbb{R}^s$ natural parameter

$A : \mathbb{R}^s \to \mathbb{R}$ log-partition function

or normal/Big Const

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

$$\int_X P_\eta \, d\mu = 1 , \forall \eta,$$

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

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

Note $\Xi$, determined by $T, h, \mu$

We could take $\Xi \subset \Xi$, if we wanted

$A(\eta)$ is always convex

$$\Rightarrow \Xi, \text{is convex (HW1 Prob.2)}$$

---

Example: Poisson

$$X \sim \text{Pois} (\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x \in 0, 1, 2, ...$$

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

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

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

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

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

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

\[ \mathcal{E}_f = \{ \gamma \in \mathbb{R}^S : \int \text{Int} e^{\gamma \cdot T} \, h \, d\mu < \infty \} \]

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

\[ \Rightarrow \text{ on } \mathcal{E}_f, \ A(\gamma) \text{ has all partial derivatives} \]

Differentiate once:

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

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

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

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


Diff twice:

\[
\frac{\partial^2}{\partial \eta_i \partial \eta_k} e^{A(\eta)} = \frac{\partial^2}{\partial \eta_i \partial \eta_k} \int e^{\gamma^T \eta} \, h \, d\mu
\]

\[
e^{-A(\eta)} \left( \frac{\partial^2 A}{\partial \eta_i \partial \eta_k} + \frac{\partial A}{\partial \eta_i} \frac{\partial A}{\partial \eta_k} \right) = \sum_{T_j \neq T_k} \frac{E[T_j T_k]}{E[T_j]} e^{\gamma^T \eta - A(\eta)}
\]

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

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

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 \( e^{A(\xi)} \)

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

\[
M_{\tau}^{T(x)}(u) = E_{\tau}[e^{u'T(x)}] = \int e^{u'T} e^{\xi'T - A(\xi)} h(d\mu) \]
\[
= e^{A(\xi+u) - A(\xi)} \int e^{(\xi+u)'T - A(\xi+u)} h(d\mu) = 1
\]

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

Cumulant-generating function

\[
K_{\tau}^{T}(u) = \log M_{\tau}^{T}(u) = A(\xi+u) - A(\xi) \quad (A \text{ is sometimes called cgf})
\]
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** \( X \sim N(m, \sigma^2) \) \( m \in \mathbb{R} \) \( \sigma^2 > 0 \)

Let \( \Theta = (m, \sigma^2) \)

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

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

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

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

**Natural parameterization**

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

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

**Binomial**

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

\[ p_\theta(x) = \theta^x (1-\theta)^{n-x} \binom{n}{x} \quad x = 0, \ldots, n \]

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

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

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

**Beta**

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

\[ p_{\alpha,\beta}(x) = x^{\alpha-1} (1-x)^{\beta-1} / B(\alpha,\beta) \]

\[ = \exp \left\{ \alpha \log x + \beta \log (1-x) - \log B(\alpha,\beta) \right\} \frac{1}{x(1-x)} \]

\[ \eta = \left( \begin{array}{c} \alpha \\ \beta \end{array} \right) \quad T(x) = \left( \begin{array}{c} \log x \\ \log (1-x) \end{array} \right) \quad h(x) = \frac{1}{x(1-x)} \]

Practically everything else on wikipedia too:

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

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

1) Start with carrier \( h(x) \)

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

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

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

\( \mathcal{E}_1 \) = all tilts after which normalization is possible

\( \Rightarrow \) Decomposition into \( \xi, T, h, A \) very non-unique

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

2) Could absorb \( h \) into \( \mu \) \( (d\mu(x) = h d\nu(x)) \)

\((\text{wlog } h(x) \equiv 1 \text{ if we want})\)

3) Can add constant to \( T(x) \)
Repeated Sampling

Suppose \( X_1, \ldots, X_n \) iid \( p_{\gamma}^{(i)}(x) = e^{\gamma'T(x) - A(\gamma)h(x)} \)

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

\[
p_{\gamma}^{(i)}(x) = \prod_{i=1}^{\hat{\gamma}} e^{\gamma'T(x_i) - A(\gamma)h(x_i)}
= \exp \left\{ \gamma' \hat{\Sigma}_{i=1} T(x_i) - \gamma A(\gamma) \right\} \prod_{i=1}^{\hat{\gamma}} h(x_i)
\]

Important property!

This means \( \hat{\Sigma}_{i=1} T(X_i) \in \mathbb{R}^S \) 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.
Suppose \( X \sim p_\xi(x) = e^{\xi^T T(x) - A(\xi)} \) wrt \( \mu \) (wlog \( h \equiv 1 \))

Then, \( T(X) \sim q_\xi(t) = e^{\xi^t - A(\xi)} \) wrt \( \nu \),

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

\[
\nu(B) = \mu\left( \{ x : T(x) \in B \} \right).
\]

\[
\mathbb{P}_\xi(T(X) \in B) = \int_{B(T)} e^{\xi^T \pi(x) - A(\xi)} d\mu(x) = \int I_B(t) e^{\xi^t - A(\xi)} d\nu(t)
\]

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

\[
\mathbb{P}_\xi(T(X) = t) = \sum_{x : T(x) = t} e^{\xi^T \pi(x) - A(\xi)} h(x) \mu(\{x\}^c) = e^{\xi^t - A(\xi)} \sum_{x : T(x) = t} h(x) \mu(\{x\}^c) = \nu(\{t\}^c)
\]