Exponential families

Outline

1) Exponential families
2) Differential identities
3) MGF

Exponential Families

An s-parameter exponential family is a family $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ with densities of the form

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

wrt base measure μ on sample space $\mathcal{X}$

Components of p_η:

- $\eta \in \Xi \subseteq \mathbb{R}^s$ called natural parameter
- $T(x)$ is s-dimensional sufficient statistic

  factorization theorem: $g_\eta(T(x)) = e^{\eta' T(x) - A(\eta)}$

- $h(x) \geq 0$ called base density or carrier density

  can be absorbed into base measure μ

- $A(\eta)$ called log-partition function or normalizing const.

  $A(\cdot)$ determined by T, h, μ:

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

The <u>natural parameter space</u> is the set of all η that give us normalizable p_η

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

<u>Note</u>: $\mathcal{P}$ can use strict subset ($\Xi \subsetneq \Xi_1$) if scientific considerations constrain η

$A(\eta)$ is always a convex function $\Rightarrow$ Ξ_1 convex set

If we absorb h into μ, log-densities almost linear

$$\log p_\eta(x) = \eta' T(x) - A(\eta) \qquad (\text{wrt } h \cdot d\mu)$$

Can think of $T(x)$ as basis

Very nice structure when we multiply densities

- Combining evidence from independent obs.
- Prior $\times$ likelihood in Bayesian calculations

or divide them

- Calculating conditional probabilities
- Likelihood ratios
- Relative densities

Examples

Poisson (single obs)

$$X \sim \mathrm{Pois}(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x \in 0, 1, \ldots$$

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

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

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

Poisson (n obs) $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Pois}(\lambda)$

$$p_\lambda(x) = \prod_{i=1}^{n} \exp\{(\log\lambda)x_i - \lambda\}\frac{1}{x_i!}$$

$$= \exp\{(\log\lambda)(\textstyle\sum_i x_i) - n\lambda\}\prod_i \frac{1}{x_i!}$$

$$\eta(\lambda) = \log\lambda \qquad T(x) = \sum_i x_i$$

$$A(\eta) = ne^\eta \qquad h(x) = \prod_i \frac{1}{x_i!}$$

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

$$p_\eta(x) = \prod_{i=1}^{n} \exp\{\eta' T^{(1)}(x_i) - A^{(1)}(\eta)\} h^{(1)}(x_i)$$

$$= \exp\{\eta' \underbrace{\left(\sum_i T^{(1)}(x_i)\right)}_{T(x)} - \underbrace{nA^{(1)}(\eta)}_{A(\eta)}\} \underbrace{\prod_i h^{(1)}(x_i)}_{h(x)}$$

$\Rightarrow$ Dimension of $T(x)$ doesn't grow with n

Differential Identities

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

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

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

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

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

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

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

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

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

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

NB: We would get wrong answer by differentiating wrt λ

Moment-generating function

We can get k^{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 $X \sim P_\eta$

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

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

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

Useful for

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

Cumulant-generating function

$$K_\eta^T(u) = \log M_\eta^T(u) = A(\eta+u) - A(\eta)$$

(A 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:

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

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

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

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

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

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

<u>Natural parameterization</u>

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

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

More examples

Binomial

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

$$p_\theta(x) = \theta^x (1-\theta)^{n-x} \binom{n}{x} \qquad x = 0, \ldots, n$$

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

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

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

Beta

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

$$p_{\alpha,\beta}(x) = x^{\alpha-1}(1-x)^{\beta-1} / B(\alpha,\beta)$$ Beta function

$$= \exp\{\alpha \log x + \beta \log(1-x) - \log B(\alpha,\beta)\} \frac{1}{x(1-x)}$$

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

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 <u>exponential tilt</u> of the carrier $h(x)$

1) Start with carrier $h(x)$
2) Multiply by $e^{\eta' T(x)}$
3) Re-normalize by $e^{-A(\eta)}$

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

Ξ_1 = all tilts after which normalization is possible

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

1) Only $\text{span}(T_1, \ldots, T_s)$ matters
2) Could absorb h into μ $\big(d\nu(x) = h \, d\mu(x)\big)$

   (wlog $h(x) \equiv 1$ if we want)
3) Can add constant to $T(x)$

⋮ many others

Distribution of $T(X)$

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

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

where ν is the measure μ "pushed forward" through $T : \mathcal{X} \to \mathbb{R}^s$

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

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

$$= \int 1_B(t) e^{\eta' t - A(\eta)} d\nu(t)$$

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

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

$$= e^{\eta' t - A(\eta)} \underbrace{\sum_{x : T(x) = t} h(x) \mu(\{x\})}_{\nu(\{t\})}$$

Canonical Form

The structure is most evident when:

- $T(x) = x$ (wlog: sufficiency reduction)
- $h(x) \equiv 1$ (wlog: absorb h into μ)
- $\theta = \eta$ (wlog: parameterize by η)

Then, we say the family is in <u>canonical form</u>:

$$p_\eta(x) = e^{\eta' x - A(\eta)}$$

Minimal form

Form of $p_\eta(x) = e^{\eta' T(x) - A(\eta)} h(x)$ minimal if

$\eta \in \Xi$ and $T(x)$ satisfy no linear constraints:

no $a \neq 0$, $b \in \mathbb{R}$ s.t. $\eta' a = b$ for all $\eta \in \Xi$

or $T(x)' a = b$ $\mathcal{P}$-a.s.

Otherwise we can represent $\mathcal{P}$ as an r-dim. ex. fam.

for some $r < s$

If p_η minimal, then $T(X)$ is minimal suff.

Need to show $\ell(\cdot\,; x) = \ell(\cdot\,; y) + c_{xy} \Rightarrow T(x) = T(y)$

($\Leftarrow$ holds by suff.)

$$\ell(\eta; x) - \ell(\eta; y) = \eta' \underbrace{(T(x) - T(y))}_{a}$$

Can find $\eta, \zeta \in \Xi$ s.t. $\eta' a \neq \zeta' a$ unless $a = 0$

$\Rightarrow T(x) = T(y)$