Exponential Families

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

Surprisingly many of the distributions we use in statistics for random variables X taking value in some space $\mathcal{X}$ (often $\mathbb{R}$ or $\mathbb{N}_0$ but sometimes $\mathbb{R}^n$, $\mathbb{Z}$, or some other space), indexed by a parameter θ from some parameter set Θ, can be written in **exponential family** form, with pdf or pmf

$$f(x \mid \theta) = \exp\left[\eta(\theta)t(x) - B(\theta)\right]\ h(x)$$

for some **statistic** $t : \mathcal{X} \to \mathbb{R}$, **natural parameter** $\eta : \Theta \to \mathbb{R}$, and functions $B : \Theta \to \mathbb{R}$ and $h : \mathcal{X} \to \mathbb{R}_+$. The likelihood function for a random sample of size n from the exponential family is

$$f_n(\mathbf{x} \mid \theta) = \exp\left[\eta(\theta)\sum_{j=1}^{n} t(x_j) - nB(\theta)\right]\ \prod h(x_i),$$

which is actually of the same form with the same natural parameter $\eta(\cdot)$, but now with statistic $T_n(\mathbf{x}) = \sum t(x_j)$ and functions $B_n(\theta) = nB(\theta)$ and $h_n(\mathbf{x}) = \Pi h(x_j)$.

Examples

For example, the pmf for the binomial distribution $\mathsf{Bi}(m, p)$ can be written as

$$\binom{m}{x} p^x (1-p)^{m-x} = \exp\left[\left(\log\frac{p}{1-p}\right)x - m\log(1-p)\right]\binom{m}{x}$$

of Exponential Family form with $\eta(p) = \log\frac{p}{1-p}$ and natural sufficient statistic $t(x) = x$, and the Poisson

$$\frac{\theta^x}{x!}e^{-\theta} = \exp\left[(\log\theta)x - \theta\right]\ \frac{1}{x!}$$

with $\eta = \log\theta$ and again $t(x) = x$. The Beta distribution $\mathsf{Be}(\alpha,\beta)$ with either *one* of its two parmeters unknown can be written in EF form too:

$$\begin{aligned}
\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} &= \exp\left[\alpha\log x - \left(\log\frac{\Gamma(\alpha)}{\Gamma(\alpha+\beta)}\right)\right]\frac{(1-x)^{\beta}}{x(1-x)\Gamma(\beta)} \\
&= \exp\left[\beta\log(1{-}x) - \left(\log\frac{\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)\right]\frac{x^{\alpha}}{x(1{-}x)\Gamma(\alpha)}
\end{aligned}$$

with $t(x) = \log x$ or $\log(1{-}x)$ when $\eta = \alpha$ or $\eta = \beta$ is unknown, respectively. With *both* parameters unknown the beta distribution can be written as a *bivariate* Exponential Family with parameter $\theta = (\alpha,\beta) \in \mathbb{R}^2_+$:

$$f(x \mid \theta) = \exp\left[\eta(\theta)\cdot t(x) - B(\theta)\right] h(x) \tag{1}$$

with *vector* parameter $\eta = (\alpha,\beta)$ and statistic $t(x) = (\log x, \log 1{-}x)$ and scalar (one-dimensional) functions $B(\theta) = \log\Gamma(\alpha) + \log\Gamma(\beta) - \log(\alpha+\beta)$ and $h(x) = 1/x(1-x)$. Since this comes up often, we'll let η and T be q-dimensional below; usually in this course $q = 1$ or 2.

Natural Exponential Families

It is often convenient to reparametrize exponential families to the *natural parameter* $\eta = \eta(\theta) \in \mathbb{R}^q$, leading (with $A(\eta(\theta)) \equiv B(\theta)$) to

$$f(x \mid \eta) = e^{\eta\cdot t(x) - A(\eta)} h(x) \tag{2}$$

Since any pdf integrates to unity we have

$$e^{A(\eta)} = \int_{\mathcal{X}} e^{\eta\cdot t(x)} h(x)\, dx$$

and hence can calculate the moment generating function (MGF) for the **natural sufficient statistic** $t(x) = \{t_1(x), \cdots, t_q(x)\}$ as

$$\begin{aligned}
M_t(s) &= \mathsf{E}\left[e^{s\cdot t(X)}\right] \\
&= \int_{\mathcal{X}} e^{s\cdot t(x)}\, e^{\eta\cdot t(x) - A(\eta)} h(x)\, dx \\
&= e^{-A(\eta)} \int_{\mathcal{X}} e^{(\eta+s)\cdot t(x)} h(x)\, dx \\
&= e^{A(\eta+s) - A(\eta)},
\end{aligned}$$

so $\log M_t(s) = A(\eta + s) - A(\eta)$ and we can find moments for the natural sufficient statistic by

$$\begin{aligned} \mathsf{E}[t] &= \nabla \log M_t(0) &= \nabla A(\eta) \\ \mathsf{V}[t] &= \nabla^2 \log M_t(0) &= \nabla^2 A(\eta) \end{aligned}$$

provided that η is an interior point of the *natural parameter space*

$$\mathcal{E} \equiv \{\eta \in \mathbb{R}^q : 0 < \int_{\mathcal{X}} e^{\eta \cdot t(x)} h(x)\, dx < \infty\}$$

and that $A(\cdot)$ is twice-differentiable near η. For samples of size $n \in \mathbb{N}$ the sufficient statistic

$$T_n(\mathbf{x}) = \sum t(x_j)$$

is a sum of independent random variables, so by the Central Limit Theorem we have approximately

$$\sim \mathsf{No}\Big(n\nabla A(\eta),\ n\nabla^2 A(\eta)\Big).$$

Note that $\nabla^2 A(\eta) = -\nabla^2 \log f(\mathbf{x} \mid \theta)$ is both the observed and Fisher (expected) information (matrix) $I_n(\theta)$ for natural exponential families, and that the score statistic is $Z := \nabla \log f(\mathbf{x} \mid \theta) = \big[T_n(\mathbf{x}) - n\nabla A(\eta)\big]$.

Conjugate Priors

For hyper-parameters $\alpha \in \mathbb{R}^q$ and $\beta \in \mathbb{R}$ such that

$$c_{\alpha,\beta} := \int_\Theta e^{\eta(\theta)\cdot\alpha - \beta B(\theta)}\, d\theta < \infty,$$

we can define a prior density for θ by

$$\pi(\theta \mid \alpha, \beta) = c_{\alpha,\beta}^{-1} \int_\Theta e^{\eta(\theta)\cdot\alpha - \beta B(\theta)}\, d\theta.$$

With this prior and with data $\{X_i\} \overset{\text{iid}}{\sim} f(x \mid \theta)$ from the exponential family, the posterior is

$$\begin{aligned} \pi(\theta \mid \mathbf{x})) &\propto e^{\eta(\theta)\cdot\alpha - \beta B(\theta)} e^{\eta(\theta)\cdot T_n(\mathbf{x}) - nB(\theta)} \\ &\propto \pi(\theta \mid \alpha^* = \alpha + T_n(\mathbf{x}),\ \beta^* = \beta + n), \end{aligned}$$

again within the same family but now with parameters $\alpha^* = \alpha + T_n$ and $\beta^* = \beta + n$. For example, in the binomial example above this conjugate prior family is

$$\pi(\theta \mid \alpha, \beta) \propto \exp\left\{\alpha \log \frac{p}{1-p} - \beta \log(1-p)\right\} = p^{\alpha}(1-p)^{\beta-\alpha},$$

the Beta family, while for the Poisson example it is

$$\pi(\theta \mid \alpha, \beta) \propto \exp\{\alpha \log \theta - \beta\theta\} = \theta^{\alpha} e^{-\beta\theta},$$

the Gamma family. Conjugate families for every exponential family are available in the same way.

Note not *every* distribution we consider is from an exponential family. From (2), for exmple, it is clear set of points where the pdf or pmf is nonzero, the possible values a random variable X can take, is just

$$\{x \in \mathcal{X} : \ f(x \mid \theta) > 0\} = \{x \in \mathcal{X} : \ h(x) > 0\},$$

which does *not* depend on the parameter θ; thus any family of distributions where the "support" depends on the parameter (uniform distributions are important examples) can't be from an exponential family.

The next pages show several familiar (and some less familiar ones, like the Inverse Gaussian $\mathsf{IG}(\mu, \lambda)$ and Pareto $\mathsf{Pa}(\alpha, \beta)$) distributions in exponential family form. Some of the formulas involve the log gamma function $\gamma(z) = \log \Gamma(z)$ and its first and second derivatives, the "digamma" $\psi(z) = (d/dz)\gamma(z)$ and "trigamma" $\psi'(z) = (d^2/dz^2)\gamma(z)$, which are built into `R`, Mathematica, Maple, the `gsl` library in C, and such, but aren't on pocket calculators or most spreadsheets. In each case $\nabla^2 A(\eta)$ is the Information matrix in the natural parametrization, $I(\theta)$ in the usual parameterization.

1 Exponential Family Examples

$\mathsf{Be}(\alpha, \beta)$

$$\begin{aligned} f(x) &= \tfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1) \\ B(\alpha,\beta) &= \gamma(\alpha) + \gamma(\beta) - \gamma(\alpha+\beta) \\ A(\eta) &= \gamma(\eta_1) + \gamma(\eta_2) - \gamma(\eta_1+\eta_2) \\ \nabla A(\eta) &= \begin{bmatrix} \psi(\eta_1) - \psi(\eta_1+\eta_2) \\ \psi(\eta_2) - \psi(\eta_1+\eta_2) \end{bmatrix} \\ \nabla^2 A(\eta) &= \begin{pmatrix} \psi'(\eta_1) - c & -c \\ -c & \psi'(\eta_2) - c \end{pmatrix} \end{aligned}$$

$$\begin{aligned} T &= (\log x, \log 1{-}x) \\ \eta &= (\alpha, \beta) \\ \mathsf{E}T &= \begin{bmatrix} \psi(\alpha) - \psi(\alpha+\beta) \\ \psi(\beta) - \psi(\alpha+\beta) \end{bmatrix} \\ c &= \psi'(\eta_1+\eta_2) \end{aligned}$$

$\mathsf{Bi}(m, p)$

$$\begin{aligned} f(x) &= \tbinom{m}{x} p^x q^{(m-x)}, \quad x = 0...m \\ B(p) &= -m \log q \\ A(\eta) &= m \log(1+e^\eta) \\ \nabla A(\eta) &= \tfrac{me^\eta}{1+e^\eta} \\ \nabla^2 A(\eta) &= \tfrac{m\,e^\eta}{(1+e^\eta)^2} \end{aligned}$$

$$\begin{aligned} T &= x \\ \eta &= \log(p/q) \\ p &= e^\eta/(1+e^\eta) \\ \mathsf{E}T &= m\,p \\ I(p) &= m/pq \end{aligned}$$

$\mathsf{Ex}(\lambda)$

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x}, \quad x > 0 \\ B(\lambda) &= -\log \lambda \\ A(\eta) &= -\log(-\eta) \\ \nabla A(\eta) &= -1/\eta \\ \nabla^2 A(\eta) &= \eta^{-2} \end{aligned}$$

$$\begin{aligned} T &= x \\ \eta &= -\lambda \\ \mathsf{E}T &= 1/\lambda \\ I(\lambda) &= 1/\lambda^2 \end{aligned}$$

$\mathsf{Ga}(\alpha, \lambda)$

$$\begin{aligned} f(x) &= \tfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\lambda x}, \quad x > 0 \\ B(\alpha,\beta) &= \gamma(\alpha) - \alpha \log \lambda \\ A(\eta) &= \gamma(\eta_1) - \eta_1 \log(-\eta_2) \\ \nabla A(\eta) &= \begin{bmatrix} \psi(\eta_1) - \log(-\eta_2) \\ -\eta_1/\eta_2 \end{bmatrix} \\ \nabla^2 A(\eta) &= \begin{pmatrix} \psi'(\eta_1) & -1/\eta_2 \\ -1/\eta_2 & \eta_1/{\eta_2}^2 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} T &= (\log x, x) \\ \eta &= (\alpha, -\lambda) \\ \mathsf{E}T &= \begin{bmatrix} \psi(\alpha) - \log \lambda \\ \alpha/\lambda \end{bmatrix} \\ I(\alpha,\lambda) &= \begin{pmatrix} \psi'(\alpha) & -1/\lambda \\ -1/\lambda & \alpha/\lambda^2 \end{pmatrix} \end{aligned}$$

$\mathsf{Ge}(p)$

$$\begin{aligned} f(x) &= p\, q^x, \quad x = 0, 1, 2, ... \\ B(p) &= -\log p \\ A(\eta) &= -\log(1-e^\eta) \\ \nabla A(\eta) &= \tfrac{e^\eta}{1-e^\eta} \\ \nabla^2 A(\eta) &= \tfrac{e^\eta}{(1-e^\eta)^2} \end{aligned}$$

$$\begin{aligned} T &= x \\ \eta &= \log q \\ p &= 1 - e^\eta \\ \mathsf{E}T &= q/p \\ I(p) &= 1/p^2 q \end{aligned}$$

Exponential Family Examples (cont'd)

$\mathsf{IG}(a,b)$

$$\begin{aligned} f(x) &= a e^{-(a-bx)^2/2x}/\sqrt{2\pi x^3}, \quad x>0 & T &= (1/x, x) \\ B(a,b) &= -ab - \log a & \eta &= (-a^2/2, -b^2/2) \\ A(\eta) &= -2\sqrt{\eta_1\,\eta_2} - \tfrac{1}{2}\log(-2\eta_1) & a &= \sqrt{-2\eta_1}, \quad b = \sqrt{-2\eta_2} \\ \nabla A(\eta) &= \begin{bmatrix} \sqrt{\eta_2/\eta_1} - 1/2\eta_1 \\ \sqrt{\eta_1/\eta_2} \end{bmatrix} & \mathsf{E}T &= \begin{bmatrix} b/a + 1/a^2 \\ a/b \end{bmatrix} \\ \nabla^2 A(\eta) &= \tfrac{1}{2}\begin{pmatrix} \sqrt{\frac{\eta_2}{\eta_1{}^3}} + \frac{1}{\eta_1{}^2} & \frac{-1}{\sqrt{\eta_1\eta_2}} \\ \frac{-1}{\sqrt{\eta_1\eta_2}} & \sqrt{\frac{\eta_1}{\eta_2{}^3}} \end{pmatrix} & I(a,b) &= \begin{pmatrix} b/a + 2/a^2 & -1 \\ -1 & a/b \end{pmatrix} \end{aligned}$$

$\mathsf{NB}(\alpha,p)$

$$\begin{aligned} f(x) &= \tbinom{-\alpha}{x} p^\alpha (-q)^x, \quad x = 0,1,2,\ldots & T &= x \\ B(p) &= -\alpha\log p & \eta &= \log q \\ A(\eta) &= -\alpha\log(1-e^\eta) & p &= 1-e^\eta \\ \nabla A(\eta) &= \tfrac{\alpha e^\eta}{1-e^\eta} & \mathsf{E}T &= \alpha q/p \\ \nabla^2 A(\eta) &= \tfrac{\alpha e^\eta}{(1-e^\eta)^2} & I(p) &= \alpha/p^2 q \end{aligned}$$

$\mathsf{No}(\mu,\sigma^2)$

$$\begin{aligned} f(x) &= e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2} & T &= (x, x^2) \\ B(\mu,\sigma^2) &= \mu^2/2\sigma^2 + \tfrac{1}{2}\log\sigma^2 & \eta &= (\mu\sigma^{-2}, -\sigma^{-2}/2) \\ A(\eta) &= -\eta_1{}^2/4\eta_2 - \tfrac{1}{2}\log(-2\eta_2) & & \\ \nabla A(\eta) &= \begin{bmatrix} -\eta_1/2\eta_2 \\ \eta_1{}^2/4\eta_2{}^2 - 1/2\eta_2 \end{bmatrix} & \mathsf{E}T &= \begin{bmatrix} \mu \\ \mu^2 + \sigma^2 \end{bmatrix} \\ \nabla^2 A(\eta) &= \begin{pmatrix} -1/2\eta_2 & \eta_1/2\eta_2{}^2 \\ \eta_1/2\eta_2{}^2 & -\eta_1{}^2/2\eta_2{}^3 + 1/2\eta_2{}^2 \end{pmatrix} & I(a,b) &= \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4}/2 \end{pmatrix} \end{aligned}$$

$\mathsf{Po}(\lambda)$

$$\begin{aligned} f(x) &= \lambda^x e^{-\lambda}/x!, \quad x = 0,1,2,\ldots & T &= x \\ B(\lambda) &= \lambda & \eta &= \log\lambda \\ A(\eta) &= e^\eta & \lambda &= e^\eta \\ \nabla A(\eta) &= e^\eta & \mathsf{E}T &= \lambda \\ \nabla^2 A(\eta) &= e^\eta & I(\lambda) &= 1/\lambda \end{aligned}$$

$\mathsf{Pa}(\alpha,\beta)$

$$\begin{aligned} f(x) &= \beta\,\alpha^\beta/x^{\beta+1}, \quad x > \alpha & T &= \log x \\ B(\beta) &= -\log\beta - \beta\log\alpha & \eta &= -\beta \\ A(\eta) &= -\log(-\eta) + \eta\log\alpha & & \\ \nabla A(\eta) &= \log\alpha - 1/\eta & \mathsf{E}T &= \log\alpha + 1/\beta \\ \nabla^2 A(\eta) &= \eta^{-2} & I(\lambda) &= \beta^{-2} \end{aligned}$$