

PGM Class Notes, Exponential Families

Karl Pichotta
pichotta@cs.utexas.edu

October 17, 2012

1 Exponential Distributions

$$P(x; \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\}$$

We have

$$\exp(A(\eta)) = \int h(x) \exp(\eta^T T(x)) dx$$

and

$$p(x; \eta) = \frac{h(x) \exp\{\eta^T T(x)\}}{\exp(A(\eta))}$$

2 Examples

1. **Bernoulli Dist.** Parametrized by π , bias of a coin. $X \in \{0, 1\}$.

$$p(x; \pi) = \pi^x (1 - \pi)^{1-x} \tag{1}$$

$$= \exp(x \log \pi + (1 - x) \log(1 - \pi)) \tag{2}$$

$$= \exp\left(x \log \frac{\pi}{1 - \pi} + \log(1 - \pi)\right) \tag{3}$$

$$\tag{4}$$

So to make it exponential, we have

$$\eta = -\log \frac{\pi}{1 - \pi} \Rightarrow \exp \eta = \pi / (1 - \pi) \Rightarrow 1 - \pi = (1 + \exp \eta)^{-1}$$

(might be missing a -1 sign)

$$T(x) = x$$

$$A(\eta) = -\log(1 - \pi)$$

$$= -\log((1 + \exp \eta)^{-1})$$

$$= \log(1 + \exp \eta)$$

2. Gaussian

$$p(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ \frac{1(x - \mu)^2}{2\sigma^2} \right\} \quad (5)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ \frac{-x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \log \sigma \right\} \quad (6)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ x^2 \frac{-1}{2\sigma^2} + x \frac{\mu}{\sigma^2} - \left(\frac{\mu^2}{2\sigma^2} + \log \sigma \right) \right\} \quad (7)$$

This is the “moment representation”.

So we have

$$\eta = \langle \mu/\sigma^2 \quad -1/2\sigma^2 \rangle$$

and

$$T(x) = \langle x \quad x^2 \rangle$$

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

(I might have copied that wrong; double check it.) And so we get our canonical representation

$$p(x; \eta) = (2\pi)^{-1/2} \exp \{ \eta_1 x + \eta_2 x^2 - A(\eta) \}$$

3. **Multinomial** $x = (x_1, \dots, x_p)$ parametrized by π_1, \dots, π_p s.t. $\pi_i \geq 0$ and $\sum \pi_i = 1$. Toss a die n times, let X_i be the number of times the die lands on face i .

$$p(x; \pi) = \frac{n!}{x_1! x_2! \dots x_p!} \pi_1^{x_1} \dots \pi_p^{x_p}$$

So we can write this as

$$p(x; \pi) = \frac{n!}{\prod_i x_i!} \exp \{ x_1 \log \pi_1 + \dots + x_p \log \pi_p \} \quad (8)$$

So we might think we can write $T(x) = \langle x_1, \dots, x_p \rangle$ and $\eta = \langle \log \pi_1, \dots, \log \pi_p \rangle$. This doesn't work though because there's a linear dependence between the x 's that we don't have here: $\sum_i x_i = n$. There's of course also one in the π_i 's. So we can express x_p in terms of x_1, \dots, x_{p-1} , ditto with the π 's. So we write

$$X_p = n - \sum_{i=1}^{p-1} X_i$$

and

$$\pi_p = 1 - \sum_{i=1}^{p-1} \pi_i$$

So our joint becomes

$$p(x; \pi) = \frac{n!}{\prod_i x_i!} \exp \left\{ x_1 \log \pi_1 + \cdots + \left(n - \sum_{i=1}^{p-1} X_i \right) \log \left(1 - \sum_{i=1}^{p-1} \pi_i \right) \right\} \quad (9)$$

$$= \frac{n!}{\prod_i x_i!} \exp \left\{ \sum_{i=1}^{p-1} x_i \log \frac{\pi_i}{1 - \sum_{i \neq p} \pi_i} + n \log \left(1 - \sum_{i \neq p} \pi_i \right) \right\} \quad (10)$$

giving us

$$T(x) = \langle x_1, \dots, x_p \rangle$$

and

$$\eta = \left\langle \log \frac{\pi_1}{1 - \sum_{i \neq p} \pi_i}, \dots, \log \frac{\pi_p}{1 - \sum_{i \neq p} \pi_i} \right\rangle$$

with our A term:

$$A(\eta) = -n \log \left(1 - \sum_{i \neq p} \pi_i \right)$$

Consider η_i :

$$\eta_i = \log \frac{\pi_i}{1 - \sum_{i \neq p} \pi_i} \quad (11)$$

$$\Leftrightarrow \exp \eta_i = \frac{\pi_i}{1 - \sum_{i \neq p} \pi_i} \quad (12)$$

$$\Leftrightarrow \sum_{i \neq p} \exp \eta_i = \frac{\sum_{i \neq p} \pi_i}{1 - \sum_{i \neq p} \pi_i} \quad (13)$$

$$\Leftrightarrow \sum_{i \neq p} \pi_i = \frac{\sum_{i \neq p} \exp \eta_i}{1 + \sum_{i \neq p} \exp \eta_i} \quad (14)$$

$$\Leftrightarrow 1 - \sum_{i \neq p} \pi_i = \frac{1}{1 + \sum_{i \neq p} \exp \eta_i} \quad (15)$$

(I don't get step (14), jeeez) and so

$$A(\eta) = n \log \left(1 + \sum_{i \neq p} \exp(\eta_i) \right)$$

Note that I think there's a minus sign somewhere that's flipped...

4. **Poisson** The pmf is

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (16)$$

$$= \frac{1}{x!} \exp\{x \log \lambda - \lambda\} \quad (17)$$

giving us $T(x) = x$, $\eta = \log \lambda$, and $A(\eta) = E^\eta$. And I can breathe a sigh of relief.

3 Moments and Canonical Parameters

What's the expectation of an exponential family?

$$\mathbb{E}[X] = \int_{\mathcal{X}} h(x) \exp\{\eta^T T(x) - A(\eta)\} dx$$

What is $A(\eta)$?

$$A(\eta) = \log \int h(x) \exp\{\eta^T T(x)\} dx$$

The grad is:

$$\nabla A(\eta) = \frac{\int T(x) h(x) \exp\{\eta^T T(x)\} dx}{\int h(x) \exp\{\eta^T T(x)\} dx}$$

The denominator is the normalization constant, So we have

$$\begin{aligned} \nabla A(\eta) &= \int T(x) p(x; \eta) dx \\ &= \mathbb{E}[T(x)] = \mu \end{aligned}$$

So if our sufficient statistics are just X , then we can take the gradient of our log partition to get the first moment. Hrm!

Note μ is a function of the parameters:

$$\mu = \nabla A(\eta)$$

if this is strictly convex, it is invertible:

$$\Rightarrow \eta = (\nabla A)^{-1}(\mu)$$

So this lets us go from the first moment to the canonical parameters.

What's the second moment like? Recall

$$A(\eta) = \log \int h(x) \exp(\eta^T T(x)) dx$$

So we have

$$A'(\eta) = \frac{\int T(x) h(x) \exp\{\eta^T T(x)\}}{\exp(A(\eta))}$$

Differentiating this again gives

$$\begin{aligned}
A''(\eta) &= \int T(x)h(x) \exp\{\eta^T T(x) - A(\eta)\} (T(x) - A'(\eta)) dx \\
&= \int T^2(x)p(x; \eta) dx - A'(\eta) \int T(x)p(x; \eta) dx \\
&= \mathbb{E}[T^2(x)] - \mathbb{E}[X]^2 \\
&= \text{Var}(T(X))
\end{aligned}$$

Rad.

3.1 Examples

For the Gaussian recall we had

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

with

$$\eta = \langle \mu/\sigma^2, -1/2\sigma^2 \rangle$$

Take the grad wrt one eta at a time:

$$\begin{aligned}
\frac{\partial A(\eta)}{\partial \eta_1} &= \frac{-2\eta_1}{4\eta_2} \\
&= \frac{\eta_1}{2\eta_2} \\
&= \mu \\
&= \mathbb{E}[X]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial A(\eta)}{\partial \eta_2} &= \frac{\eta_1^2}{4\eta_2^2} - \frac{2}{4\eta_2^2} \\
&= \sigma^2 + \mu^2 \\
&= \mathbb{E}[X^2] \\
&= \text{Var}(X) + \mathbb{E}[X]^2
\end{aligned}$$

The second derivative gives us the second centered moment:

$$\begin{aligned}
\frac{\partial^2 A(\eta)}{\partial \eta_2^2} &= \frac{1}{2\eta_2^2} \\
&= \sigma^2
\end{aligned}$$

So μ and η have a one-one mapping, so we can go from one to the other. So if we have

$$p(x; \eta) = \exp\{\eta^T T(x) - A(\eta)\}$$

we can express this as a function of μ , since η is just a function of μ :

$$p(x; \mu) = \exp\{\psi(\mu)^T T(x) - A(\psi(\mu))\}$$