CS6190: Probabilistic Modeling Homework 1
Exponential Families, Conjugate Priors

Gopal Menon

03 February, 2018

Written Part

1. **Expectation of sufficient statistics:** Consider a random variable X from a continuous exponential family with natural parameter $\eta = (\eta_1, \dots, \eta_n)$. Recall this means the pdf is of the form:

$$p(x) = h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right)$$

(a) Show that $E\left[T(x)|\eta\right] = \nabla A(\eta) = \left(\frac{\partial A}{\partial \eta_1}, \dots, \frac{\partial A}{\partial \eta_d}\right)$.

Hint: Start with the identity $\int p(x)dx = 1$, and take the derivative with respect to η.

$$\begin{aligned}
\int p(x)dx &= 1 \\
\Rightarrow \int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx &= 1 \\
\nabla \left(\int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx \right) &= 0 \\
\int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) . \left(T(x) - \nabla A(\eta)\right) dx &= 0 \\
\int T(x)h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx &= \nabla A(\eta) \int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx \\
\int T(x)p(x)dx &= \nabla A(\eta) \int p(x)dx \\
E\left[T(x)\right] &= \nabla A(\eta)
\end{aligned}$$

(b) Verify this formula works for the Gaussian distribution with unknown mean, μ, and known variance, σ^2.

Hint: Start by thinking about what the natural parameter η and the function $A(\eta)$ are, then verify that the expectation of the Gaussian is the same as $\nabla A(\eta)$.

The pdf for a Gaussian distribution with unknown mean, μ, and known variance, σ^2 is given by:

$$
\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\ln\left(\frac{1}{\sigma}\right)\right] \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\ln(\sigma) - \frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\ln(\sigma) - \frac{x^2}{2\sigma^2} + \frac{2x\mu}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\begin{bmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{bmatrix} \cdot \begin{bmatrix} x \\ x^2 \end{bmatrix} - \left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)\right]
\end{aligned}
$$

Based on the form of the exponential family pdf, we can see that:

$$
\begin{aligned}
A(\eta) &= \ln(\sigma) + \frac{\mu^2}{2\sigma^2}, \\
\eta &= \begin{bmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{bmatrix}, \text{ and} \\
T(x) &= \begin{bmatrix} x \\ x^2 \end{bmatrix}
\end{aligned}
$$

Based on the equation for $E[T(x)]$ above:

$$
\begin{aligned}
E[T(x)] &= \nabla A(\eta) \\
\Rightarrow E\left[\begin{bmatrix} x \\ x^2 \end{bmatrix}\right] &= \begin{bmatrix} \frac{\partial A}{\partial \eta_1} \\ \frac{\partial A}{\partial \eta_2} \end{bmatrix}, \text{ where } \eta_1 = \frac{\mu}{\sigma^2}, \text{ and } \eta_2 = \frac{-1}{2\sigma^2} \\
\begin{bmatrix} E[x] \\ E[x^2] \end{bmatrix} &= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\sigma^2}{2}\frac{\mu^2}{\sigma^4}\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln\left(\sqrt{\frac{-1}{2\eta_2}}\right) - \mu^2\eta_2\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\sigma^2}{2}\eta_1^2\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln\left(\frac{i}{\sqrt{2}\sqrt{\eta_2}}\right) - \mu^2\eta_2\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma^2}{2}.2\eta_1 \\ \frac{\sqrt{2}\sqrt{\eta_2}}{i}.\frac{i}{\sqrt{2}}.\frac{-1}{2}\eta_2^{\frac{-3}{2}} - \mu^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma^2}{2}.2\frac{\mu}{\sigma^2} \\ \frac{-1}{2\eta_2} - \mu^2 \end{bmatrix} \\
&= \begin{bmatrix} \mu \\ \sigma^2 - \mu^2 \end{bmatrix} \\
\Rightarrow E[x] &= \mu, \text{ and} \\
E[x^2] &= \sigma^2 - \mu^2, \text{ which is true since } \sigma^2 = E[x^2] - \mu^2
\end{aligned}
$$

2. **Noninformative priors for the Poisson distribution:** Let $X \sim Pois(\lambda)$. Recall that the pmf of the Poisson is

$$P(X = k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

(a) Rewrite the above pmf in exponential family form. What is the natural parameter? What is the sufficient statistic?

$$\begin{aligned} P(X = k; \lambda) &= \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \frac{1}{k!} \exp\left(\ln \lambda^k e^{-\lambda}\right) \\ &= \frac{1}{k!} \exp\left(k \ln \lambda - \lambda\right) \\ \eta &= \ln \lambda, \text{ natural parameter} \\ h(k) &= k, \text{ sufficient statistic} \end{aligned}$$

(b) Give at least two different options for noninformative priors for $p(\lambda)$.

i. **Uniform Prior:** In this case the prior is

$$p(\lambda) = c, \text{ for } 0 \leq x \leq \infty, \text{, where c is a constant}$$

ii. **Jeffrey's Prior:** In this case the prior is

$$
\begin{aligned}
\mathcal{I}(\lambda) &= E\left[\left(\frac{\partial}{\partial\lambda}\ln p(x|\lambda)\right)^2\right] \\
&= -E\left[\frac{\partial^2}{\partial\lambda^2}\ln p(x|\lambda)\right] \\
&= -E\left[\left(\frac{\partial^2}{\partial\lambda^2}\ln\frac{\lambda^x e^{-\lambda}}{x!}\right)^2\right] \\
&= -E\left[\left(\frac{\partial^2}{\partial\lambda^2}\left(x\ln\lambda-\lambda-\ln x!\right)\right)^2\right] \\
&= E\left[\frac{\partial}{\partial\lambda}\left(\frac{x}{\lambda}-1\right)^2\right] \\
&= -E\left[\frac{-x}{\lambda^2}\right] \\
&= E\left[\frac{x}{\lambda^2}\right] \\
&= \frac{1}{\lambda^2}E\left[x\right] \\
&= \frac{1}{\lambda^2}\sum_{i=1}^{\infty}xp(x) \\
&= \frac{1}{\lambda^2}\sum_{i=1}^{\infty}x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{e^{-\lambda}}{\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} \\
&= \frac{e^{-\lambda}}{\lambda}\left(\frac{\lambda^0}{0!}+\frac{\lambda^1}{1!}+\frac{\lambda^2}{2!}+\dots\right) \\
&= \frac{e^{-\lambda}}{\lambda}e^{\lambda} \\
&= \frac{1}{\lambda} \\
p(\lambda) &= \sqrt{\mathcal{I}(\lambda)} \\
&= \sqrt{\frac{1}{\lambda}}
\end{aligned}
$$

(c) What are the resulting posteriors for your two options? Are they proper (i.e., can they be normalized)?

i. **Uniform Prior:** The posterior is

$$
\begin{aligned}
p(\lambda|x) &\propto p(x|\lambda)p(\lambda) \\
&= \prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}.c, \text{based on the assumtion of i.i.d. samples} \\
&\propto \prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}
\end{aligned}
$$

Even though the prior cannot be normalized, since it will not add up to 1, the posterior shown above, where the constant c has been dropped, can be normalized. The reason is that the posterior is the product of distributions that can themselves be normalized.

ii. **Jeffrey's Prior:** The posterior is

$$\begin{aligned}
p(\lambda|x) &\propto p(x|\lambda)p(\lambda) \\
&= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \cdot \sqrt{\frac{1}{\lambda}}, \text{based on the assumtion of i.i.d. samples} \\
&= \prod_{i=1}^{n} \frac{\lambda^{x_i - \frac{1}{2}} e^{-\lambda}}{x_i!}
\end{aligned}$$

This can be normalized since it is the product of the likelihood, which is a product of Poisson distributions, and so can be normalized, and the prior, which can also be normalized.