

# CS6190: Probabilistic Modeling Homework 1

## Exponential Families, Conjugate Priors

*Gopal Menon*

*04 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:

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

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

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

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

- (b) Verify this formula works for the Gaussian distribution with unknown mean,  $\mu$ , and known variance,  $\sigma^2$ .

**Hint:** Start by thinking about what the natural parameter  $\eta$  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,  $\mu$ , and known variance,  $\sigma^2$  is given by:

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

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:

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

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?

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

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

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

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

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

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

(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

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

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

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

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.

3. **Non-conjugate priors:** Let  $X_i$  be from a Gaussian with known variance  $\sigma^2$  and mean  $\mu$  with uniform prior, i.e.,

$$\begin{aligned} \mu &\sim \text{Unif}(a, b) \\ X_i &\sim N(\mu, \sigma^2) \end{aligned}$$

What is the posterior pdf,  $p(\mu|x_1, \dots, x_n; \sigma^2, a, b)$ ?

**Hint:** There will be an integral that you won't be able to analytically solve (just leave it in integral form).

$$\begin{aligned} p(\mu|x_1, \dots, x_n; \sigma^2, a, b) &= p(x_1, \dots, x_n|\mu; \sigma^2)p(\mu; a, b) \\ &= \int_{-\infty}^{\infty} p(x|\mu; \sigma^2)p(\mu; a, b)dx \\ &= \int_a^b p(x|\mu; \sigma^2) \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

## R Coding Part

```
# Density function for Inverse Gamma: conjugate prior for Gaussian sigma^2
density_inverse_gamma <- function(x_coordinates, alpha, beta) {

  return (beta^alpha / gamma(alpha) * x_coordinates^(-alpha - 1) * exp(-beta / x_coordinates))

}

# Density function for Normal-Inverse-Gamma with parameters mu0, lambda, alpha, beta
density_normal_inverse_gamma <-
  function(distribution_of_mean, distribution_of_variance, mu0 = 0, lambda = 1, alpha = 8, beta = 16)
  {
    return (dnorm(distribution_of_mean, mean = mu0, sd = sqrt(distribution_of_variance / lambda)) *
            density_inverse_gamma(distribution_of_variance, alpha, beta))
  }

# Read data for hippocampus volume and patient details
hippocampus_volume <- read.csv(file="oasis_hippocampus.csv", header=TRUE, sep=",")
patient_details <- read.csv(file="oasis_cross-sectional.csv", header=TRUE, sep=",")
```

```

# Join both frames on identifier
patient_hippocampus_volume <- merge(hippocampus_volume, patient_details, by="ID")

# Select right hippocampal volume for control group
control_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR==0.0, select=RightHippoVol)

# Select right hippocampal volume for mild dementia group
mild_dementia_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR!=0.0 & CDR!=2.0 & !is.null(CDR),
    select=RightHippoVol)

# Select right hippocampal volume for dementia group
dementia_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR==2.0, select=RightHippoVol)

# Model each group as a normal variable with its own mean and variance
library(fitdistrplus)

```

```
## Loading required package: MASS
```

```
## Loading required package: survival
```

```

control_group_rt_hippo_vol_normal_model <-
  fitdistr(control_group_rt_hippo_vol[[1]], "normal")
mild_dementia_group_rt_hippo_vol_normal_model <-
  fitdistr(mild_dementia_group_rt_hippo_vol[[1]], "normal")
dementia_group_rt_hippo_vol_normal_model <-
  fitdistr(dementia_group_rt_hippo_vol[[1]], "normal")

```

TODO: Explain joint posterior density and marginal posterior density.

```

# Find marginal posterior distribution of variance. Create a range for the variance that
# is 20% on both sides of the sample variance. Add up the posterior density for mean value
# that is three standard deviations on both sides.

```