Math 153 - Lab 3

We are interested in the case where our data is assumed to be coming from a normal distribution parametrized by its mean μ and its variance σ^2. We start with the case where σ^2 is known, and thus $\theta = \mu$.

We have as the likelihood

$$L(x|\mu) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \propto \exp\{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\}.$$

a. Defining $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$, show

$$\sum(x_i - \mu)^2 = ns^2 + n(\bar{x} - \mu)^2.$$

A consequence of this is that

$$p(X|\mu) \propto \exp\{-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\} \propto \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

As we noticed with the binomial pmf, if we think about the likelihood being a function of p instead of x, we notice that it looks a lot like a Beta density. Thus, using a Beta prior, we got a Beta posterior out (conjugacy!). If we view the likelihood as a function of μ, it looks like a normal pdf as well. Let's see what happens if we use a normal prior.

In particular, we may try

$$f(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$$

We can now write down the posterior $f(\mu|x) \propto p(x|\mu)f(\mu)$ as usual, and completing the square, arrive at

$$f(\mu|x) \propto \exp\{-\frac{1}{2\sigma_n^2}(\mu - \mu_n)^2\}$$

with

$$\sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

and

$$\mu_n = \sigma_n^2(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}).$$

It's important to keep track of the various flavors of μ and σ^2 that are present here and what represents what.

b. Write some sentences making sense of these formulas, in particular relating this back to a representation we have of the beta posterior mean relating to the prior and the data.

You might notice that the formula for σ_n^2 is really unfortunate. It's *almost* pretty. To fix this, Bayesian's prefer not to talk about the variance, where large

numbers correspond to large uncertainty, but instead to **precision**, where large numbers correspond to less uncertainty.

Define $\lambda_i = \frac{1}{\sigma_i^2}$. The posterior becomes $f(\mu|x) = \mathcal{N}(\mu_n, \lambda_n)$ where

$$\lambda_n = \lambda_0 + n\lambda$$

and

$$\mu_n = \frac{\bar{x}n\lambda + \mu_0\lambda_0}{\lambda_n}.$$

That's better.

Alternatively, we might write $\sigma_0^2 = \frac{\sigma^2}{\kappa_0}$. This yields the interpretation that our prior knowledge is equivalent to κ_0 observations. This yields a posterior variance of $\frac{\sigma^2}{\kappa_n}$ where $\kappa_n = \kappa_0 + n$. I'm writing it in terms of σ^2 here, as that is likely more comfortable for you having just been introduced to precision, but we will use it with λ more often.

c. Similar to the last part of lab 2, create a plot that shows the prior being updated by data (just let $\bar{x}$ be a constant, based on increasing sample sizes).

A 'non-informative' prior can be arrived at by simply letting the prior variance equal to ∞, i.e. uniform over the reals. This is an *improper prior*, as it doesn't correspond to an actual density (one cannot define a uniform density over the reals), but still yields a proper posterior, namely

$$f(\mu|x) = \mathcal{N}(\bar{x}, \frac{\sigma^2}{n})$$

which looks like what frequentists expect to see.

While the math is nice here, I have yet to personally work with data for which μ is unknown and σ^2 (or equivalently λ) is known. Let's deal with the case where both are unknown. Writing the likelihood and seeing if anything looks familiar, we have

$$p(x|\mu, \lambda) \propto \lambda^{n/2} \exp\{-\frac{\lambda}{2}[n(\mu - \bar{x})^2 + ns^2]\}$$

As a function of μ only, this looks again like the kernel of a normal distribution. As a function of λ only, recall the gamma density:

$$f(x|\alpha, \beta) \propto x^{\alpha-1}e^{-\beta x}\mathbf{1}(x > 0)$$

The conjugate prior for (μ, λ) ends up being normal and gamma priors, respectively (thought the normal prior is a conditional density, conditioned on the value λ). So we need to specify the 4-tuple $(\mu_0, \kappa_0, \alpha_0, \beta_0)$ to specify our prior. We write

$$(\mu, \lambda) \sim \mathcal{NG}(\mu_0, \kappa_0, \alpha_0, \beta_0)$$

Fun fact: The marginal distribution of μ is a t-distribution.

d. Present some plots of various gamma distributions and discuss what they would say about your prior information if used as a prior for λ.

e. The mode is given as $\frac{\alpha-1}{\beta}$. Using *pbeta*, find a prior that encodes the following information coming from a scientist: my best guess is that $\sigma^2 = 4$ and I am 90 percent sure that it is no smaller than 2.

The posterior distribution is

$$
p(\mu, \lambda | x) = \mathcal{NG}(\frac{\kappa_0 \mu_0 + n\bar{x}}{\kappa_0 + n}, \kappa_0 + n, \alpha_0 + n/2, \beta_0 + \frac{n}{2}s^2 + \frac{\kappa_0 n(\bar{x} - \mu_0)^2}{2(\kappa_0 + n)})
$$

Let's stop doing calculus now. I can generate from one of these Normal-Gamma distributions via a two-step process.

- Generate λ^* from a $\Gamma(\alpha_n, \beta_n)$.

- Then generate μ^* from a $\mathcal{N}(\mu_n, \kappa_n \lambda^*)$

f. Write code that takes in a data set as well as the vector of hyperparameters, and returns the posterior mean for μ, a 95 percent credible interval for μ, and a 95 percent prediction interval for a new observation x. The predictive distribution is just a draw from a normal distribution with parameters μ^* and λ^*.

g. Think about fuel efficiency in cars. More specifically, think about what these numbers looked like in 1974. Specify a prior that you think is reasonable. Then use the *mtcars* data set in R (with a random car removed) to construct a prediction interval (using your code from e). Does it cover the missing observation? To visualize your prior, the easiest thing to do is just to sample from it and then create a scatter plot. You will likely need to limit the dimensions of your plot, using something like

```
plot(c(-2,2), c(0,4), t='n')
points(mu, lambda)  #these will have been generated first
```

This lab is in large parts based on `https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf` This document contains many more details than are given here.