BIOSTAT 202C Homework 2

Lillian Chen (c.lillian@ucla.edu)

October 25, 2021

1. Jacobian.

Calculate the Jacobian for the following transformations.

a) $Z = X^2$.

$$X = Z^{1/2}$$
$$\frac{dX}{dZ} = \frac{1}{2Z^{1/2}}$$

b) $Z = \exp(X)$.

$$X = \ln Z$$
$$\frac{dX}{dZ} = \frac{1}{Z}$$

c) $Z = 1/X$.

$$X = \frac{1}{Z} = Z^{-1}$$
$$\frac{dX}{dZ} = -Z^{-2} = -\frac{1}{Z^2}$$

d) $Z = \text{logit}(X)$.

$$X = \text{logit}^{-1}(X) = \frac{1}{1+e^{-Z}} = \frac{e^Z}{1+e^Z}$$
$$\frac{dX}{dZ} = \frac{e^Z(1+e^Z) - e^Z(e^Z)}{(1+e^Z)^2} = \frac{e^Z}{(1+e^Z)^2}$$

2.

Starting from $X \sim \Gamma(a, b)$ density function, use the Jacobian for $Z = 1/X$, and derive the density of Z, which is distributed as an Inverse Gamma random variable.

$$\begin{aligned}
\frac{dX}{dZ} &= -\frac{1}{Z^2} \\
f(Z) &= f(X(Z))\left|\left|\frac{dX}{dZ}\right|\right| \\
&= \frac{b^a}{\Gamma(a)}\left(\frac{1}{Z}\right)^{a-1}\exp\left(-b\left(\frac{1}{Z}\right)\right)\left(\frac{1}{Z^2}\right) \\
&= \frac{b^a}{\Gamma(a)}Z^{-a+1-2}\exp\left(-\frac{b}{Z}\right) \\
&= \frac{b^a}{\Gamma(a)}Z^{-a-1}\exp\left(-\frac{b}{Z}\right) \\
Z &\sim IG(a,b), Z > 0
\end{aligned}$$

3. Poisson.

Data $Y = (y_1, ..., y_n)'$ is an n-vector of observations iid as $y_i \sim Poisson(\lambda)$, with $i = 1, ..., n$. The prior for λ is a Gamma distribution with known scalars $a > 0$ and $b > 0$ as parameters, $\lambda \sim Gamma(a, b)$.

Since this problem utilizes the same prior, sampling density, and data Y from HW 1 Problem 2, we know that the posterior distribution of λ given data Y can be described as follows:

$$\lambda|Y \sim Gamma(a + \Sigma y_i, b + n); i = 1, ..., n$$

a) Using the minus 2nd derivative log posterior evaluated at the posterior mode, evaluate the FIP fraction of posterior information coming from the prior.

$$\begin{aligned}
\log f(\lambda|y_i) &= \log\left(\frac{(b+n)^{(a+\Sigma y_i)}}{\Gamma(a+\Sigma y_i)}\lambda^{a+\Sigma y_i-1}e^{-(b+n)\lambda}\right) \\
&= \log\left[(b+n)^{(a+\Sigma y_i)}\right] - \log\left[\Gamma(a+\Sigma y_i)\right] + \log\left[\lambda^{a+\Sigma y_i-1}\right] + \log\left[e^{-(b+n)\lambda}\right] \\
&= (a+\Sigma y_i)\log(b+n) - \log\left[\Gamma(a+\Sigma y_i)\right] + (a+\Sigma y_i-1)\log\lambda - (b+n)\lambda \\
\frac{d}{d\lambda}\left[\log f(\lambda|y_i)\right] &= (a+\Sigma y_i-1)\lambda^{-1} - (b+n) \\
\hat{\lambda} = \text{posterior mode} &= \frac{a+\Sigma y_i-1}{b+n} \\
\frac{d^2}{d^2\lambda}\left[\log f(\lambda|y_i)\right] &= -(a+\Sigma y_i-1)\lambda^{-2} \\
-\frac{d^2}{d^2\lambda}|_{\lambda=\hat{\lambda}}\left[\log f(\lambda|y_i)\right] &= (a+\Sigma y_i-1)\left(\frac{a+\Sigma y_i-1}{b+n}\right)^{-2} = \frac{(b+n)^2}{a+\Sigma y_i-1} = \tilde{I}_p
\end{aligned}$$

$$
\begin{aligned}
\log f(\lambda) &= \log\left(\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}\right) \\
&= \log\left[b^a\right] - \log\left[\Gamma(a)\right] + \log\left[\lambda^{a-1}\right] + \log\left[e^{-b\lambda}\right] \\
&= \log\left[b^a\right] - \log\left[\Gamma(a)\right] + (a-1)\log\lambda - b\lambda \\
\frac{d}{d\lambda}\left[\log f(\lambda)\right] &= (a-1)\lambda^{-1} - b \\
\frac{d^2}{d^2\lambda}\left[\log f(\lambda\right] &= -(a-1)\lambda^{-2} \\
-\frac{d^2}{d^2\lambda}\left[\log f(\lambda\right] &= (a-1)\lambda^{-2} = \text{info from prior} \\
-\frac{d^2}{d^2\lambda}|_{\lambda=\hat{\lambda}}\left[\log f(\lambda\right] &= \frac{(a-1)(b+n)^2}{(a+\Sigma y_i-1)^2}
\end{aligned}
$$

$$
\begin{aligned}
\text{FIP} &= \frac{\frac{(a-1)(b+n)^2}{(a+\Sigma y_i-1)^2}}{\frac{(b+n)^2}{a+\Sigma y_i-1}} \\
&= \frac{a-1}{a+\Sigma y_i-1}
\end{aligned}
$$

(b) Write the posterior mean as a convex combination of the prior mean and the data mean.

$$
\begin{aligned}
\text{posterior mean} = \bar{\lambda} &= \frac{a+\Sigma y_i}{b+n} \\
\text{prior mean} = \lambda_0 &= \frac{a}{b} \\
\text{data mean} &= \bar{Y}
\end{aligned}
$$

We use the relationship between the posterior mean, prior mean, and data mean as a starting point to identifying the weights W and $1-W$ needed to write the posterior mean as a convex combination of the prior mean and data mean.

$$
\begin{aligned}
\frac{a+\Sigma y_i}{b+n} &= W\bar{Y} + (1-W)\frac{a}{b} \\
\frac{a/b + (\Sigma y_i)/b}{1+n/b} &= W\bar{Y} + (1-W)\frac{a}{b} \\
\left(\frac{1}{1+n/b}\right)\frac{a}{b} + \left(\frac{n\bar{Y}/b}{1+n/b}\right) &= (1-W)\frac{a}{b} + W\bar{Y} \\
\left(\frac{1}{1+n/b}\right)\frac{a}{b} + \left(\frac{n/b}{1+n/b}\right)\bar{Y} &= (1-W)\frac{a}{b} + W\bar{Y} \\
\bar{\lambda} &= \left(\frac{n/b}{1+n/b}\right)\bar{Y} + \left(\frac{1}{1+n/b}\right)\frac{a}{b}
\end{aligned}
$$

(c) Interpret the prior parameters - what is the prior data mean, and prior sample size in this model?

The prior data y_i is distributed $y_i \sim Poisson(\lambda)$, so the prior data mean is $lambda$ (the mean of a $Poisson(\lambda)$ is λ. The prior sample size, n, is the number of y_i's observed, which are iid.

4. Power Distribution.

We define the power distribution for $\theta > 0$ and for $0 < y < 1$ to have the density

$$f(y|\theta) \propto y^{\theta-1}\mathbf{1}\{0 < y < 1\}$$

a) We transform following $z_i = -\log y_i$ prior to analyzing. As a result, we get the relationship $y_i = e^{-z_i}$ and the Jacobian $\frac{d\,y_i}{d\,z_i} = -e^{-z_i}$.

$$\begin{aligned}
f(z_i|\theta) &= f(y_i|\theta(z_i)) \left|\left|\frac{d\,y_i}{d\,z_i}\right|\right| \\
&= \theta \exp(-z_i)^{\theta-1} \exp(-z_i) \\
f(z_i|\theta) &= \theta \exp(-\theta z_i) = \text{density for the exponential distribution}, 0 < z_i < \infty
\end{aligned}$$

5. Power Distribution.

We define the power distribution for $\theta > 0$ and for $0 < y < 1$ to have the density

$$f(y|\theta) \propto y^{\theta-1}\mathbf{1}\{0 < y < 1\}$$

and to have a prior

$$\theta \sim Gamma(a, b)$$

a) Calculate the posterior mean, variance, mode, and the negative 2nd derivative log posterior evaluated at the mode. We use the normalizing constant calculated from HW 1 to get the following likelihood:

$$f(y|\theta) = \theta y^{\theta-1}\mathbf{1}\{0 < y < 1\}$$

$$\begin{aligned}
f(\theta|y) &= \frac{f(y|\theta)f(\theta)}{f(y)} \\
&\propto f(y|\theta)f(\theta) \\
&\propto \left[\prod_{i=1}^{n} \theta y_i^{\theta-1}\right] \left[\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\right] \\
&\propto \theta^n \left[\prod_{i=1}^{n} y_i^{\theta-1}\right] \left[\theta^{a-1}e^{-b\theta}\right] \\
&\propto \theta^{a+n-1} \left[\prod_{i=1}^{n} y_i^{\theta-1}\right] e^{-b\theta} \\
&\propto \theta^{a+n-1} \exp\left[\log \prod_{i=1}^{n} y_i^{\theta-1}\right] \exp(-b\theta) \\
&\propto \theta^{a+n-1} \exp\left[(\theta-1)\sum_{i=1}^{n} \log y_i - b\theta\right] \\
&\propto \theta^{a+n-1} \exp\left[-(b - \sum_{i=1}^{n} \log y_i)\theta\right] \\
\theta|y &\sim Gamma(a+n, b - \sum \log y_i)
\end{aligned}$$

$$\text{posterior mean} = \bar{\theta} = \frac{a+n}{b - \sum \log y_i}$$

$$\text{posterior variance} = Var(\theta|y_i) = \frac{a+n}{(b - \sum \log y_i)^2}$$

$$\begin{aligned}
\log f(\theta|y_i) &= \log\left(\frac{(b-\sum\log y_i)^{(a+n)}}{\Gamma(a+n)}\theta^{a+n-1}e^{-(b-\sum\log y_i)\theta}\right)\\
&= \log\left[(b-\sum\log y_i)^{(a+n)}\right]-\log\left[\Gamma(a+n)\right]+\log\left[\theta^{a+n-1}\right]+\log\left[e^{-(b-\sum\log y_i)\theta}\right]\\
\frac{d}{d\theta}\left[\log f(\theta|y_i)\right] &= (a+n-1)\theta^{-1}-(b-\sum\log y_i)\\
\hat{\theta} = \text{posterior mode} &= \frac{a+n-1}{b-\sum\log y_i}\\
\frac{d^2}{d^2\theta}\left[\log f(\theta|y_i)\right] &= -(a+n-1)\theta^{-2}\\
-\frac{d^2}{d^2\theta}|_{\theta=\hat{\theta}}\left[\log f(\theta|y_i)\right] &= (a+n-1)\left(\frac{a+n-1}{b-\sum\log y_i}\right)^{-2}=\frac{(b-\sum\log y_i)^2}{a+n-1}=\tilde{I}_p
\end{aligned}$$

b) Using the 2nd derivative - log posterior evaluated at the posterior mode, evaluate the fraction of posterior information coming from the prior.

$$\begin{aligned}
\log f(\theta) &= \log\left(\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\right)\\
&= \log\left[b^a\right]-\log\left[\Gamma(a)\right]+\log\left[\theta^{a-1}\right]+\log\left[e^{-b\theta}\right]\\
&= \log\left[b^a\right]-\log\left[\Gamma(a)\right]+(a-1)\log\theta-b\theta\\
\frac{d}{d\theta}\left[\log f(\theta)\right] &= (a-1)\theta^{-1}-b\\
\frac{d^2}{d^2\theta}\left[\log f(\theta)\right] &= -(a-1)\theta^{-2}\\
-\frac{d^2}{d^2\theta}\left[\log f(\theta)\right] &= (a-1)\theta^{-2} = \text{info from prior}\\
-\frac{d^2}{d^2\theta}|_{\theta=\hat{\theta}}\left[\log f(\theta)\right] &= \frac{(a-1)(b-\sum\log y_i)^2}{(a+n-1)^2}
\end{aligned}$$

$$\begin{aligned}
\text{FIP} &= \frac{\frac{(a-1)(b-\sum\log y_i)^2}{(a+n-1)^2}}{\frac{(b-\sum\log y_i)^2}{a+n-1}}\\
&= \frac{a-1}{a+n-1}
\end{aligned}$$

https://www.overleaf.com/project/6172344ad9bdd3980f5dd940 c) (cont'd) Is it preferable to use the formula in the previous item or would it be easier/preferable to use the formula FIP = [1/prior variance]/[1/posterior variance]? One sentence: discuss.

It would be easier to use the FIP formula involving the inverse variances because we know that the variances exist (so there is no concern about the inverse variances existing) and the variances are fairly straightforward to calculate for a gamma distribution.

d) Two other names for the power distribution include the Pareto distribution and Zipf's law.

6. Normal Approximations.

a) The gamma posterior from HW 1, Problem 2, is as follows:

$$\lambda|y_i \sim Gamma(a+\Sigma y_i, b+n)$$

The normal approximation for the gamma posterior using the posterior mean and posterior variance is:

$$\lambda \sim N\left(\frac{a+n\bar{y}}{b+n}, \frac{a+n\bar{y}}{(b+n)^2}\right)$$

The normal approximation for the gamma posterior using the posterior mode and 2nd derivative is:

$$\lambda \sim N\left(\frac{a+n\bar{y}-1}{b+n}, \frac{a+n\bar{y}-1}{(b+n)^2}\right)$$

b) Let the data for each of the following examples consist of one datum y_1 ($n = 1$).

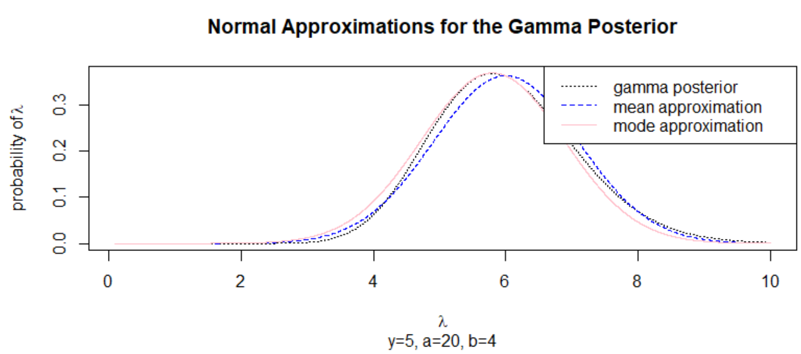
i) An example where the normal approximations to the gamma posterior are good is with the following:

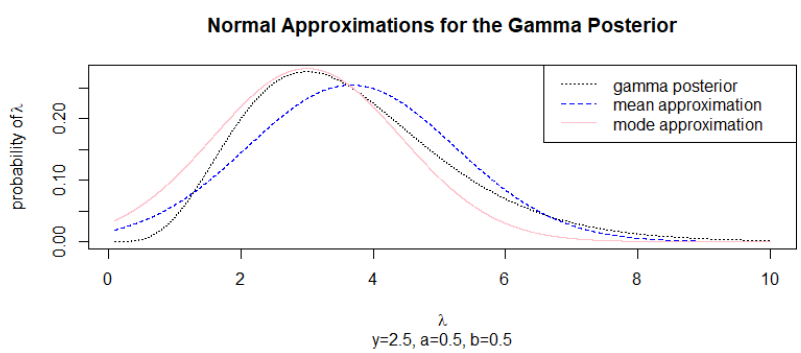
$$y_1 = 2.5, a = 23, b = 4$$

ii) An example where the normal approximations to the gamma posterior are bad is with the following:

$$y_1 = 2.5, a = 0.5, b = 0.5$$

c)





d) The gamma distribution is a two-parameter continuous distribution that can result in a long right-tailed distribution or a bell-curve shaped distribution. The normal approximation will be good when the shape parameter a is large since that tends towards a bell-curved shape, and the normal approximation will be poor when the gamma distribution is skewed, which occurs when shape parameter a is fairly small or close to 1 ($a = 1$ yields an exponential distribution).