# 202C HW2

Zixi Zhang

October 21, 2022

1. **Jacobian.** Calculate the Jacobian for the following transformations.
(a) $Z = X^2$.
*Solution:*

$$\begin{aligned}
&Z = X^2 \\
&\sqrt{Z} = X \\
&\frac{dX}{dZ} = \frac{d}{dZ}(\sqrt{Z}) = \frac{1}{2\sqrt{Z}}
\end{aligned}$$

(b) $Z = \exp(X)$.
*Solution:*

$$\begin{aligned}
&\log(Z) = X \\
&\frac{dX}{dZ} = \frac{d}{dZ}(\log(Z)) = \frac{1}{Z}
\end{aligned}$$

(c) $Z = 1/X$.
*Solution:*

$$\begin{aligned}
&x = \frac{1}{Z} \\
&\frac{dX}{dZ} = \frac{d}{dz}(\frac{1}{z}) = -\frac{1}{z^2}
\end{aligned}$$

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

$$\begin{aligned}
&Z = \text{logit}(X) = \log(\frac{x}{1-x}) \\
&e^Z = \frac{X}{1-X} \\
&e^Z - Xe^Z = X \\
&e^Z = X + Xe^Z \\
&x = \frac{e^Z}{1+e^Z} \\
&\frac{dX}{dZ} = \frac{d}{dZ}(\frac{e^Z}{1+e^Z}) = \frac{e^Z}{(1+e^Z)^2}
\end{aligned}$$

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

$$\begin{aligned}
f(x) &= \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \\
f(z) &= \frac{b^a}{\Gamma(a)} (\frac{1}{x})^{a-1} e^{-\frac{b}{x}} |J| \\
&= \frac{b^a}{\Gamma(a)} (\frac{1}{x})^{a-1} e^{-\frac{b}{x}} |-\frac{1}{z^2}| \\
&= \frac{b^a}{\Gamma(a)} (\frac{1}{x})^{a+1} e^{-\frac{b}{x}} \\
&\Rightarrow Z \sim \text{Inverse Gamma}(a, b)
\end{aligned}$$

3. **Poisson.** Cont'd from HW 1, problem 2, same prior, sampling density, and data $Y$.
(a) Using the minus 2nd derivative log posterior evaluated at the posterior mode, evaluate the FIP fraction of posterior information coming from the prior.
*Solution:*

$$\lambda \mid \mathbf{Y} \sim \text{Gamma}(a + \Sigma y_i, b + n), i = 1, 2, ..., n$$

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

$$\lambda \sim \text{Gamma}(a, b), \text{mode} = \frac{a-1}{b}$$

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

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

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

$$\frac{\Sigma_{i=1}^n y_i + a}{n+b}$$

prior mean:

$$\frac{a}{b}$$

data mean:

$$\bar{y} = \frac{\Sigma_{i=1}^n y_i}{n} \Rightarrow \Sigma_{i=1}^n y_i = n\bar{y}$$

So we can write posterior mean as

$$\begin{aligned}
\frac{\Sigma_{i=1}^n y_i + a}{n+b} &= \frac{n\bar{y} + a}{n+b} \\
&= \frac{n\bar{y}}{n+b} + \frac{a}{n+b} \\
&= \frac{n}{n+b} \cdot \bar{y} + \frac{b}{n+b} \cdot \frac{a}{b}
\end{aligned}$$

where $\dfrac{n}{n+b}$ and $\dfrac{b}{n+b}$ are all non-negative and $\dfrac{n}{n+b}+\dfrac{b}{n+b}=1,$ so this is a convex combination.

(c) Interpret the prior parameters - what is the prior data mean, and prior sample size in this model?
*Solution:* prior parameter $\sim$ Gamma(a,b)
mean: $\frac{a}{b}$, sample size: $b$

4. **Power Distribution,** cont'd from HW 1, last problem. A colleague suggests transforming $z_i = -\log y_i$ before analyzing.
(a) What is the density of $z_i \mid \theta$ ? (Give name and give formula for the density.)
*Solution:*

$$
\begin{aligned}
z_i &= -\log y_i \\
y_i &= e^{-z_i} \\
\frac{dy}{dz} &= \frac{d}{dz}(e^{-z_i}) = -e^{-z_i} \\
\Rightarrow f(z_i|\theta) &= \theta(e^{-z_i})^{\theta-1} \cdot |-e^{-z_i}| \\
&= \theta e^{-z_i\theta}
\end{aligned}
$$

$$z_i|\theta \sim \exp(\theta)$$

5. **Power Distribution.** Continued from Homework 1. Use a $\theta \sim \text{Gamma}(a,b)$ prior.

(a) Calculate the posterior mean, variance, mode, and the negative 2nd derivative log posterior evaluated at the mode.
*Solution:*

$$
\begin{aligned}
f(\theta|y) &= \frac{f(y\mid\theta)f(\theta)}{f(y)} \\
&\propto L(y\mid\theta)f(\theta) \\
&\propto \theta^n \prod_{i=1}^n y_i^{\theta-1} \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \\
&\propto \theta^{n+a-1} \prod_{i=1}^n y_i^{\theta-1} e^{-b\theta} \\
&\propto \theta^{n+a-1} e^{\log(\prod_{i=1}^n y_i^{\theta-1})} e^{-b\theta} \\
&\propto \theta^{n+a-1} e^{(\theta-1)\Sigma_{i=1}^n \log y_i} e^{-b\theta} \\
&\propto \theta^{a+n-1} e^{-(b-\sum_{i=1}^n \log y_i)\theta}
\end{aligned}
$$

$$\theta \mid y \sim \text{Gamma}\left(a+n, b-\sum \log y_i\right)$$

Then we can get

$$\text{posterior mean} = \frac{a+n}{b-\sum \log y_i}$$
$$\text{posterior variance} = \frac{a+n}{(b-\sum \log y_i)^2}$$
$$\text{mode} = \frac{a+n-1}{b-\sum \log y_i}$$
$$\log f(\theta \mid y_i) = \log\left(\frac{(b-\sum \log y_i)^{(a+n)}}{\Gamma(a+n)}\theta^{a+n-1}e^{-(b-\sum \log y_i)\theta}\right)$$
$$= \log(\frac{(b-\sum \log y_i)^{(a+n)}}{\Gamma(a+n)}) + (a+n-1)\log\theta - (b-\sum \log y_i)\theta$$
$$\frac{d}{d\theta}\log f(\theta \mid y_i) = (a+n-1)\theta^{-1} - (b-\sum \log y_i)$$
$$\frac{d^2}{d\theta^2}\log f(\theta \mid y_i) = -(a+n-1)\theta^{-2}$$

evaluate at the mode

$$-\frac{d^2}{d^2\theta} = (a+n-1)\left(\frac{a+n-1}{b-\sum \log y_i}\right)^{-2} = \frac{(b-\sum \log y_i)^2}{a+n-1}$$

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

$$f(\lambda) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\lambda}$$
$$\log f(\lambda) = \log(\frac{b^a}{\Gamma(a)}) + (a-1)\log\theta - b\theta$$
$$\frac{d}{d\theta}[\log f(\theta)] = (a-1)\theta^{-1} - b$$
$$\frac{d^2}{d\theta^2}[\log f(\theta)] = (1-a)\theta^{-2}$$
$$-\frac{d^2}{d\theta^2}[\log f(\theta)]|_{\theta=\frac{a-1}{b}} = (a-1)(\frac{a+n-1}{b-\sum \log y_i})^{-2} = \frac{(a-1)(b-\sum \log y_i)^2}{(a+n-1)^2}$$
$$\text{FIP} = \frac{\frac{(a-1)(b-\sum \log y_i)^2}{(a+n-1)^2}}{\frac{(b-\sum \log y_i)^2}{a+n-1}} = \frac{a-1}{a+n-1}$$

(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.

*Solution:* The variance-related formula is better. Because the distributions are already known, so we can easily calculate the variances.

(d) Give two other names for the power distribution.
*Solution:* Pareto distribution and power law distribution.

6. **Normal approximations.**
(a) Construct two different (algebraic) normal approximations for the gamma posterior from HW 1 problem 2. (posterior mean-variance; posterior mode and 2nd derivative).
*Solution:*
posterior:

$$\lambda \mid \mathbf{Y} \sim \text{Gamma}(a+\Sigma y_i, b+n), i = 1, 2, ..., n$$

mean $= \frac{a+\Sigma y_i}{b+n}$ and variance $= \frac{a+\Sigma y_i}{(b+n)^2}$
Normal approximation(posterior mean-variance):

$$\lambda \mid \mathbf{Y} \sim \mathrm{N}(\frac{a+\Sigma y_i}{b+n}, \frac{a+\Sigma y_i}{(b+n)^2})$$

Normal approximation(posterior mode and 2nd derivative):

$$\begin{aligned}
&\lambda \mid \mathbf{Y} \sim \mathrm{N}(\tilde{\lambda}, \tilde{I}_p^{-1}) \\
&\text{where } \tilde{\lambda} = \frac{a+\Sigma y_i - 1}{b+n} \\
&\tilde{I}_\mathrm{p} = -\left.\frac{\partial^2 \log L(\lambda \mid Y)}{\partial \lambda^2}\right|_{\lambda=\hat{\lambda}} - \left.\frac{\partial^2 \log f(\lambda)}{\partial \lambda^2}\right|_{\lambda=\tilde{\lambda}} \\
&L(\lambda \mid Y) = \prod_{i=1}^{n} e^{-\lambda} \frac{1}{y_i!} \lambda^{y_i} \\
&\log L(\lambda \mid Y) = -n\lambda + \sum_{i=1}^{n} \log\left(\frac{1}{y_i!}\right) + \sum_{i=1}^{n} y_i \log(\lambda) \\
&\frac{\partial^2 \log L(\lambda \mid Y)}{\partial \lambda^2} \alpha \lambda^2 = -\sum_{i=1}^{n} y_i \lambda^{-2} \\
&\text{at mode } = -\sum_{i=1}^{n} y_i \left(\frac{a+\sum_{i=1}^{n} y_i - 1}{n+b}\right)^{-2} \\
&f(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\
&\log f(\lambda) = \log\left(\frac{b^a}{\Gamma(a)}\right) + (a-1)\log(\lambda) - b\lambda \\
&\frac{\partial^2 \log f(\lambda)}{\partial \lambda^2} = -(a-1)\lambda^{-2} \\
&\text{at } \hat{\lambda} = -(a-1)\left(\frac{a+\sum_{i=1}^{n} y_i - 1}{n+b}\right)^{-2} \\
\Rightarrow &\tilde{I}_\mathrm{p} = \left(a + \sum_{i=1}^{n} y_i - 1\right)\left(\frac{a+\sum_{i=1}^{n} y_i - 1}{n+b}\right)^{-2} \\
&\hat{I}_\mathrm{p}^{-1} = \frac{a+\sum_{i=1}^{n} y_i - 1}{(n+b)^2}
\end{aligned}$$

$$\lambda \mid \mathbf{Y} \sim \mathrm{N}(\frac{a+\Sigma y_i - 1}{b+n}, \frac{a+\Sigma y_i - 1}{(b+n)^2})$$

(b) Construct two data/prior examples. [That is, you pick $y_1$ (or $y_1, \ldots, y_n$ if you like), and also $a$ and $b$.]

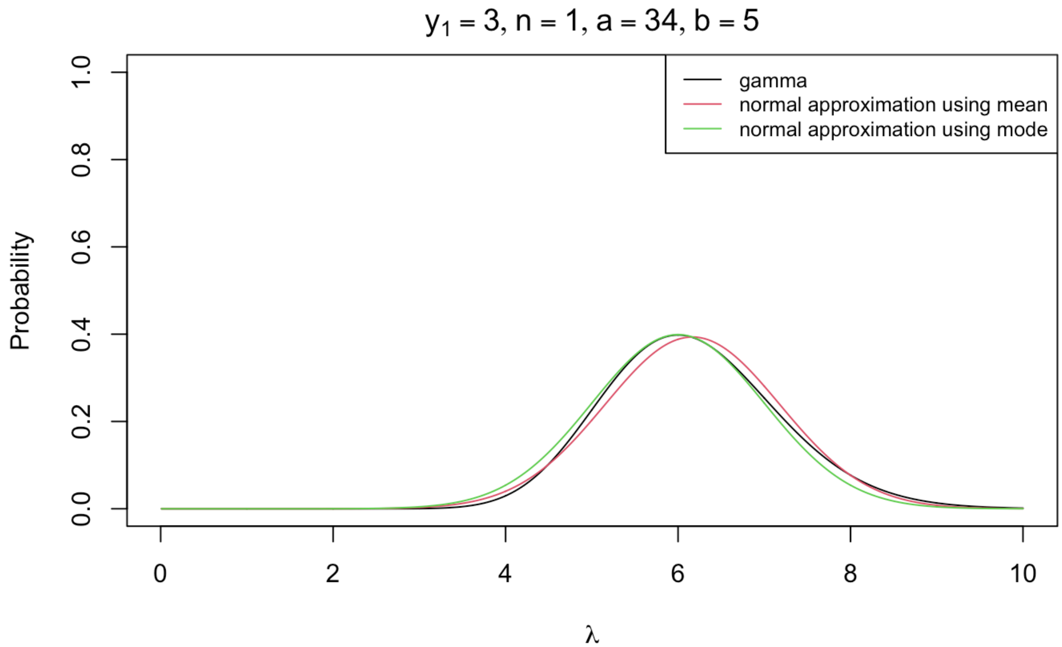
i. Make one example where the normal approximation(s) to the gamma posterior are good;
*Solution:*

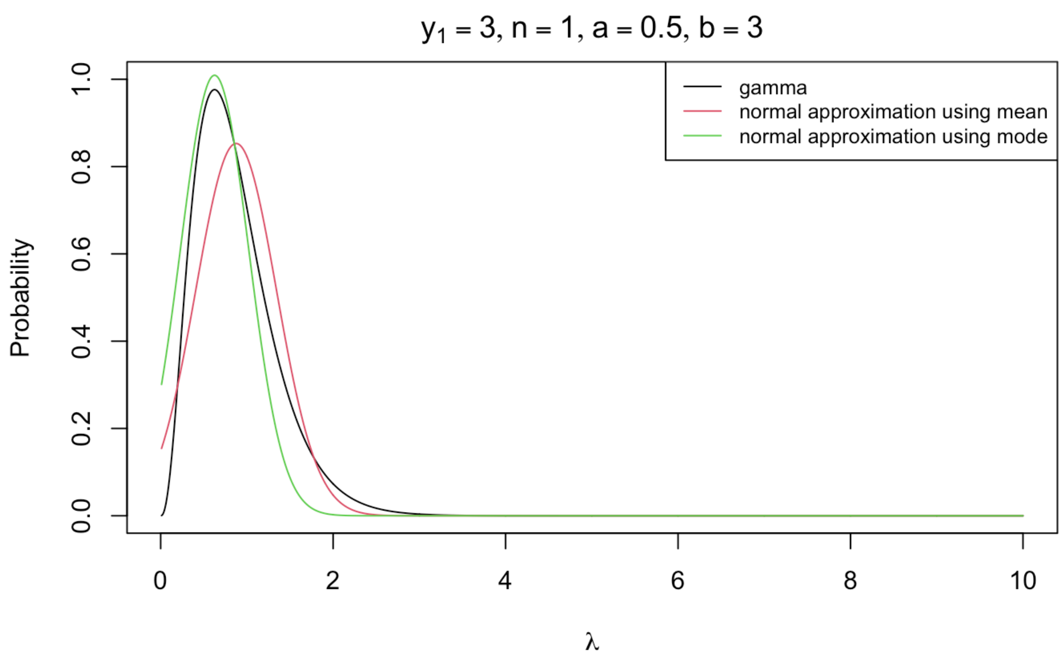
$$y_1 = 3, n = 1, a = 34, b = 5$$

ii. Make one example where the approximation(s) is (are) not very good. Specify your prior parameters and data for each combo.

$$y_1 = 3, n = 1, a = 0.5, b = 3$$

(c) For each data/prior combo, plot the posterior and your two normal approximations on a single graph. [So two plots total, each with three densities on it.]
*Solution:*




(d) Generalize: When will the normal approximation be good and when will it be poor? [1-2 sentences.]
*Solution:*
The normal approximation would be good when the original gamma distribution itself is close to a bell-shaped distribution. CLT tells us the error in a normal approximation to the gamma distribution is going to decrease as the shape $a$ grows larger. When $a$ is small, the normal approximation is poor.