BIOSTAT 202C Homework 1

Lillian Chen (c.lillian@ucla.edu)

October 18, 2021

1. Poisson-Gamma.

a) We are told that datum y is distributed $Poisson(\lambda)$ with λ as the mean of the Poisson distribution. The prior for this distribution is $\lambda \sim Gamma(a, b)$, with known scalars $a > 0$ and $b > 0$. To find the posterior distribution of λ given y, we use Bayes' theorem and multiply the likelihood function by the prior probability distribution and divide by the normalizing constant. Since the posterior distribution of λ given y is proportional to the likelihood multiplied by the prior, we can find the posterior distribution $f(\lambda|y)$ illustrated as follows:

$$\begin{aligned}
f(\lambda|y) &= \frac{f(y|\lambda)f(\lambda)}{f(y)} \\
&\propto f(y|\lambda)f(\lambda) \\
&\propto \frac{e^{-\lambda}\lambda^{y}}{y!}\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} \\
&\propto \frac{1}{y!}\frac{b^a}{\Gamma(a)}\lambda^{(a+y)-1}e^{-\lambda(b+1)} \\
&\propto \lambda^{(a+y)-1}e^{-\lambda(b+1)}
\end{aligned}$$

Since the gamma distribution is a conjugate prior for a Poisson likelihood, we know that our posterior distribution of λ will also belong to the gamma distribution. We can see that the last line above shows proportionality to a gamma probability distribution, where new scalars $a^* > 0, b^* > 0$ have the relationships $a^* = a + y$ and $b^* = b + 1$ with the known scalars a and b. Thus, the posterior distribution of λ given y can be described as follows:

$$\lambda|y \sim Gamma(a + y, b + 1)$$

b) We can calculate the normalizing constant $f(y) = \int_0^\infty f(y|\lambda)f(\lambda)d\lambda$ with the likelihood function and prior probability distribution used in part a) as follows:

$$\begin{aligned}
f(y) &= \int_0^\infty f(y|\lambda)f(\lambda)d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^{y}}{y!}\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}d\lambda \\
&= \int_0^\infty \frac{b^a}{y!\Gamma(a)}\lambda^{(a+y)-1}e^{-\lambda(b+1)}d\lambda
\end{aligned}$$

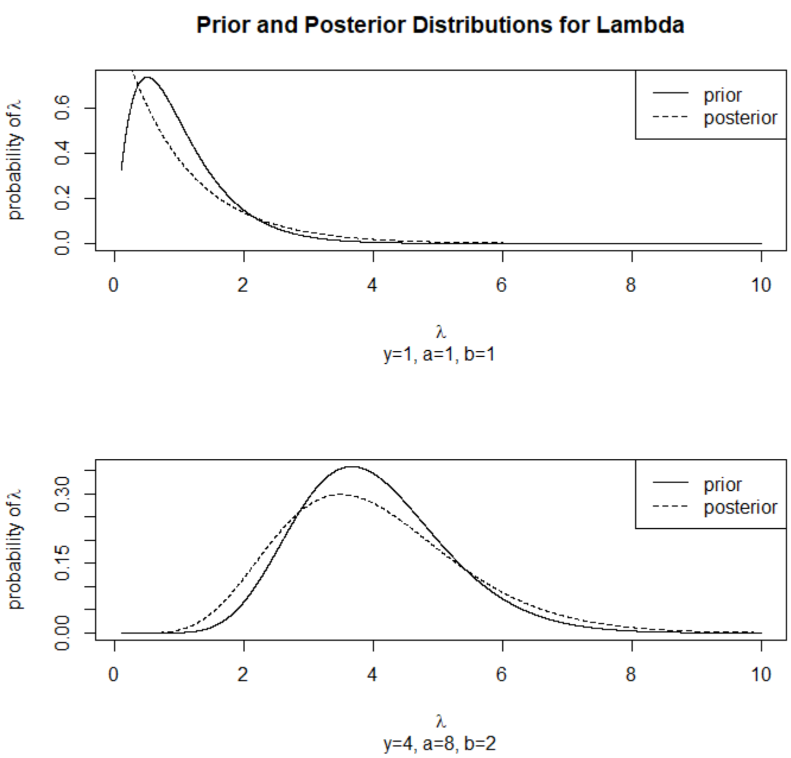
The last line shows proportionality to the form of a Gamma probability distribution, where new scalars $a^* > 0, b^* > 0$ have the relationships $a^* = a + y$ and $b^* = b + 1$ with the known scalars a and b. We can multiply by 1 to have the integral be equivalent to the pdf of a Gamma distribution as follows:

$$
\begin{aligned}
f(y) &= \frac{b^a}{y!\Gamma(a)} \int_0^\infty \lambda^{(a+y)-1} e^{-\lambda(b+1)} \frac{(b+1)^{a+y}}{\Gamma(a+y)} \frac{\Gamma(a+y)}{(b+1)^{a+y}} d\lambda \\
&= \frac{b^a \Gamma(a+y)}{y!\Gamma(a)(b+1)^{a+y}} \int_0^\infty \lambda^{(a+y)-1} e^{-\lambda(b+1)} \frac{(b+1)^{a+y}}{\Gamma(a+y)} d\lambda \\
&= \frac{b^a \Gamma(a+y)}{y!\Gamma(a)(b+1)^{a+y}} \times 1 \\
&= \frac{b^a \Gamma(a+y)}{y!\Gamma(a)(b+1)^{a+y}} \\
&= \text{normalizing constant}
\end{aligned}
$$

c) The distribution of $f(y|a,b)$ is negative binomial, also known as gamma-Poisson since we can view this distribution as a $Poisson(\lambda)$ distribution with λ as a random variable with parameters shape $= r$ and rate $B = 1 - p/p$. We write this as the following:

$$
\begin{aligned}
f(y|a,b) &= \int_0^\infty f_{Poisson(\lambda)}(y) * f_{Gamma(a,b)}(\lambda)\, d\lambda \\
&= \int_0^\infty \frac{\lambda^y}{y!} e^{-\lambda} \frac{\lambda^{a-1} e^{-\lambda(b)}}{(b)^a\, \Gamma(a)}\, d\lambda \\
&= \frac{\Gamma(a+y)}{y!\,\Gamma(a)} \left(\frac{1}{b+1}\right)^y \left(1 - \frac{1}{b+1}\right)^a
\end{aligned}
$$

d)



2. Poisson-Gamma, cont'd.

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, $\lambda \sim Gamma(a, b)$. a) Using Bayes' theorem and the fact that the posterior distribution is proportional to the likelihood function and the prior distribution, we can find the posterior distribution of λ given Y as follows:

$$
\begin{aligned}
f(\lambda|y) &= \frac{f(y|\lambda)f(\lambda)}{f(y)} \\
&\propto f(y|\lambda)f(\lambda) \\
&\propto \frac{e^{-n\lambda}\lambda^{\Sigma y_i}}{\Pi y_i}\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} \\
&\propto \lambda^{(a+\Sigma y_i)-1}e^{-\lambda(b+n)} \\
&\propto \lambda^{(a+\Sigma y_i)-1}e^{-\lambda(b+n)}
\end{aligned}
$$

We can see that the last line above shows proportionality to a gamma probability distribution, where new scalars $a^* > 0, b^* > 0$ have the relationships $a^* = a + y$ and $b^* = b + 1$ with the known scalars a and b. Thus, the posterior distribution of λ given data Y can be described as follows:

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

3. Half-normal distribution.

The half-normal distribution given parameter τ is defined with the density

$$f(y|\tau) = (\frac{2}{\pi\tau})^{1/2}\exp(-\frac{y^2}{2\tau})\mathbf{1}\{0 < y\}$$

and denoted as $y|\tau \sim \text{HN}(y|\tau)$. Let $y_i|\tau \sim \text{HN}(y_i|\tau)$ for $i = 1, ..., n$ with $Y = (y_1, ..., y_n)'$.

a) Finding the mean, median, mode, and variance of the half-normal distribution:

Mean - $E[y|\tau]$ is the integral from 0 to infinity of the cdf $f(y|\tau)$ multiplied by y with respect to y.

$$
\begin{aligned}
E[y|\tau] &= \int_0^\infty \left(\frac{2}{\pi\tau}\right)^{1/2} \exp\left(\frac{-y^2}{2\tau}\right) y\,dy \\
&= \left(\frac{2}{\pi\tau}\right)^{1/2} \int_0^\infty \exp\left(\frac{-y^2}{2\tau}\right) y\,dy \\
&= -\left(\frac{2}{\pi\tau}\right)^{1/2} \int_0^{-\infty} \exp(u)\tau\,du \\
&= -\left(\frac{2\tau}{\pi}\right)^{1/2} \int_0^{-\infty} \exp(u)\,du \\
&= -\left(\frac{2\tau}{\pi}\right)^{1/2} [e^u]_0^{-\infty} \\
&= -\left(\frac{2\tau}{\pi}\right)^{1/2} \left[e^{-\infty} - e^0\right] \\
&= -\left(\frac{2\tau}{\pi}\right)^{1/2} (0-1) \\
E[y|\tau] &= \left(\frac{2\tau}{\pi}\right)^{1/2}
\end{aligned}
$$

Median - The median of the half normal distribution is equivalent to the y value, $y_{0.5}$, such that the area under the half-normal density curve is equal to 0.5.

$$
\begin{aligned}
0.5 &= \int_0^{y_{0.5}} \left(\frac{2}{\pi\tau}\right)^{1/2} \exp\left(\frac{-s^2}{2\tau}\right) ds \\
&= \left(\frac{2}{\pi\tau}\right)^{1/2} \int_0^{y_{0.5}} \exp\left(\frac{-s^2}{2\tau}\right) ds \\
&= \left(\frac{2}{\pi\tau}\right)^{1/2} (2\tau)^{1/2} \int_0^{y_{0.5}/\sqrt{2\tau}} \exp(-u^2)\,du \\
&= \frac{2}{\pi^{1/2}} \int_0^{y_{0.5}/\sqrt{2\tau}} \exp(-u^2)\,du
\end{aligned}
$$

The integral on the right hand side is essentially the error function, defined by:

$$
\operatorname{erf} z = \frac{2}{\pi^{1/2}} \int_0^z \exp(-t^2)\,dt
$$

Rewriting the right hand side above, we get the following:

$$
\begin{aligned}
0.5 &= \operatorname{erf}\left(\frac{y_{0.5}}{\sqrt{2\tau}}\right) \\
\operatorname{erf}^{-1}(0.5) &= \frac{y_{0.5}}{\sqrt{2\tau}} \\
y_{0.5} &= \sqrt{2\tau}\operatorname{erf}^{-1}(0.5)
\end{aligned}
$$

Mode - The mode of the half-normal distribution is equivalent to the maximum of the half-normal density curve, and thus take the derivative as follows:

$$\begin{aligned}\frac{df}{dy} = 0 = \frac{d}{dy}&\left[\left(\frac{2}{\pi\tau}\right)^{1/2}\exp\left(\frac{-y^2}{2\tau}\right)\right]\\
&= \left(\frac{2}{\pi\tau}\right)^{1/2}\frac{d}{dy}\left[\exp\left(\frac{-y^2}{2\tau}\right)\right]\\
&= \left(\frac{2}{\pi\tau}\right)^{1/2}\left(\frac{-y}{\tau}\right)\exp\left(\frac{-y^2}{2\tau}\right)\end{aligned}$$

$\left(\frac{2}{\pi\tau}\right)^{1/2}$ is a constant and $\exp\left(\frac{-y^2}{2\tau}\right)$ is a nonzero component, so we set the derivative df/dy equal to 0 and equivalent to $-y/\tau$ as follows and get the mode as $y = 0$:

$$\begin{aligned}0 &= -y/\tau\\ y &= 0\end{aligned}$$

Variance - Let $y = |x|$, s.t. $x \sim \mathrm{N}(0, \tau)$. If $y|\tau \sim \mathrm{HN}(\tau)$ and $x \sim \mathrm{N}(0, \tau)$ then:

$$\begin{aligned}E[y^2] = E[|x|^2] &= E[x^2]\\
&= Var(x) + [E[x]]^2\\
E[y^2] &= \tau + 0 = \tau\\
Var(y) &= E[y^2] - [E[y]]^2\\
&= \tau - \left(\left(\frac{2\tau}{\pi}\right)^{1/2}\right)^2\\
&= \tau - \frac{2\tau}{\pi}\\
Var(y) &= \tau\left(1 - \frac{2}{\pi}\right)\end{aligned}$$

b) To find the sufficient statistic, we use the likelihood factorization theorem to isolate the factor that is both the function of the data and parameter of interest. The likelihood of the half normal distribution is given as follows:

$$\begin{aligned}L(\tau|y_i) &= \prod_{i=1}^{n}\left(\frac{2}{\pi\tau}\right)^{1/2}\exp\left(\frac{-y_i^2}{2\tau}\right), i = 1, ..., n\\
&= \left(\frac{2}{\pi\tau}\right)^{n/2}\exp\left(\frac{-\Sigma_{i=1}^{n} y_i^2}{2\tau}\right)\end{aligned}$$

The likelihood factorization theorem is given by the following:

$$u(x)v(T(x), \theta)$$

We can see that the part of the data that is the function of both the sample y_i and the parameter θ is

$$v(T(y_i, \tau)) = \exp\left(-\sum_{i=1}^{n}\frac{y_i^2}{2\tau}\right)$$

Thus, by the likelihood factorization theorem, the sufficient statistic is

$$T(y_i) = \sum_{i=1}^{n} y_i^2$$

c) Given that the conjugate prior for τ is the inverse gamma distribution $\tau \sim InverseGamma(a/2, b/2)$, we can derive the posterior of $\tau|Y$ as follows:

$$\begin{aligned}
f(\tau|Y) &= \frac{f(Y|\tau)f(\tau)}{f(y)} \\
&\propto f(Y|\tau)f(\tau) \\
&\propto \left(\frac{2}{\pi\tau}\right)^{n/2} \exp\left(\frac{-\sum_{i=1}^{n} y_i^2}{2\tau}\right) \frac{b/2}{\Gamma(a/2)} \tau^{-(a/2)-1} \exp\left(\frac{-b/2}{\tau}\right) \\
&\propto \left(\frac{2}{\pi}\right)^{n/2} \frac{b/2}{\Gamma(a/2)} \left(\frac{1}{\tau}\right)^{n/2} \exp\left(\frac{-\sum_{i=1}^{n} y_i^2 - b}{2\tau}\right) \tau^{-(a/2)-1} \\
&\propto \exp\left(\frac{-(\sum_{i=1}^{n} y_i^2 + b)/2}{\tau}\right) \tau^{-a/2-n/2-1} \\
&\propto \exp\left(\frac{-(\sum_{i=1}^{n} y_i^2 + b)/2}{\tau}\right) \tau^{-\left(\frac{a+n}{2}\right)-1} \\
a^* &= a + n \\
b^* &= \sum_{i=1}^{n} y_i^2 + b \\
\tau|y &\sim InverseGamma\left(\frac{a+n}{2}, \frac{\sum y_i^2 + b}{2}\right)
\end{aligned}$$

4. Half-normal distribution, cont'd.

a) Features of the data that would lead me to use a half-normal distribution as a sampling model would be if the data was continuous, non-negative, and concentrated around 0. The half-normal distribution can be used to model lifetime data and measurement data. Minimum daily temperatures in Fahrenheit in Fairbanks, Alaska would not be appropriately modeled by a half-normal density because there could be negative data points and the data would not be concentrated at 0 degrees Fahrenheit.

b) A specific example of data that might be modeled by the half-normal density would be in part inspection in an engineering setting – identifying the magnitude of the difference between the measurement of an engineering part and the acceptable measurement listed in the part specifications. There would ideally be magnitudes fairly close to 0 since the manufactured parts would be accurate to the specified lengths, and measurements are both continuous and non-negative.

c) Competitor distributions that might be used in place of the half-normal distribution to model non-negative data would be non-negative continuous distributions, which include distributions such as chi, gamma, and inverse gamma.

5. Half-normal distribution, cont'd.

a) If we parametrize the half-normal distribution in terms of the unknown standard deviation $\sigma = \tau^{1/2}$, we get that $y_i \sim \text{HN}(y_i|\sigma^2)$. This results in the sampling density of $y_i|\sigma$ written as follows:

$$\begin{aligned} f(y_i|\sigma) &= \prod_{i=1}^{n} \left[\left(\frac{2}{\pi\sigma^2} \right)^{1/2} \exp\left(\frac{-y_i^2}{2\sigma^2} \right) \right] \\ &= \left(\frac{2}{\pi\sigma^2} \right)^{n/2} \exp\left(\frac{-\sum_{i=1}^{n} y_i^2}{2\sigma^2} \right) \end{aligned}$$

b) No, because the sampling density of $y_i|\sigma$ is parameterized by σ^2, but the inverse gamma prior is being placed on σ in this scenario, which results in a posterior that does not take the form of an inverse gamma distribution since the exponent cannot be manipulated into a form that looks like the inverse gamma.

6. 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) To find the normalizing constant for $f(y|\theta)$ so that the right hand side is a density, we set the left hand side as an integral of the normalizing constant times the density with bounds for y, and set the right hand side equal to 1 since a density integrates to 1. We solve so that the normalizing constant c is equal to θ.

$$\begin{aligned} \int_0^1 cy^{\theta-1}dy &= 1 \\ \left[\frac{cy^\theta}{\theta} \right]_0^1 &= 1 \\ \frac{c}{\theta} - 0 &= 1 \\ c &= \theta \end{aligned}$$

b) The likelihood for a sample y_i of size n, $i = 1, ..., n$, with $Y = (y_1, ..., y_n)'$ is

$$\begin{aligned} L(y_i|\theta) &= \prod_{i=1}^{n} \theta y_i^{\theta-1} i = 1, ..., n \\ &= \theta^n \prod_{i=1}^{n} y_i^{\theta-1} \end{aligned}$$

Using the likelihood factorization theorem, we can isolate the sufficient statistic $T(x)$ by identifying the part of the likelihood that is a function of both the data Y and the parameter of interest θ. The likelihood factorization theorem is given by the following:

$$u(x)v(T(x), \theta)$$

We can see that the part of the data that is the function of both the sample y_i and the parameter θ is

$$v(T(y_i, \theta)) = \prod_{i=1}^{n} y_i^{\theta-1}$$

Thus, by the likelihood factorization theorem, the sufficient statistic is

$$T(y_i) = \prod_{i=1}^{n} y_i$$

The interpretable form of the sufficient statistic is the product of all data points in the sample.