Biostats-202C-hw1

Jiahao Tian

2022-10-06

```
library(tidyr)
library(ggplot2)
```

“““ **Question 1**

a)

$$
\begin{gathered}
y \sim pois(\lambda), \quad \lambda \sim gamma(a, b) \\
f(y|\lambda) = \frac{(\lambda^y e^{-\lambda})}{y!} \quad L(y|\lambda) = \frac{\lambda^{\sum(yi)} e^{-ny}}{\Pi y!} \\
p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b} \\
p(\lambda|y) \propto \lambda^{y+a-1} e^{-\lambda(b+1)}, \lambda > 0 \\
posterior\ follows\ gamma\ distribution\ \sim Gamma(y+a, b+1)
\end{gathered}
$$

b)

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

c)

From part b,

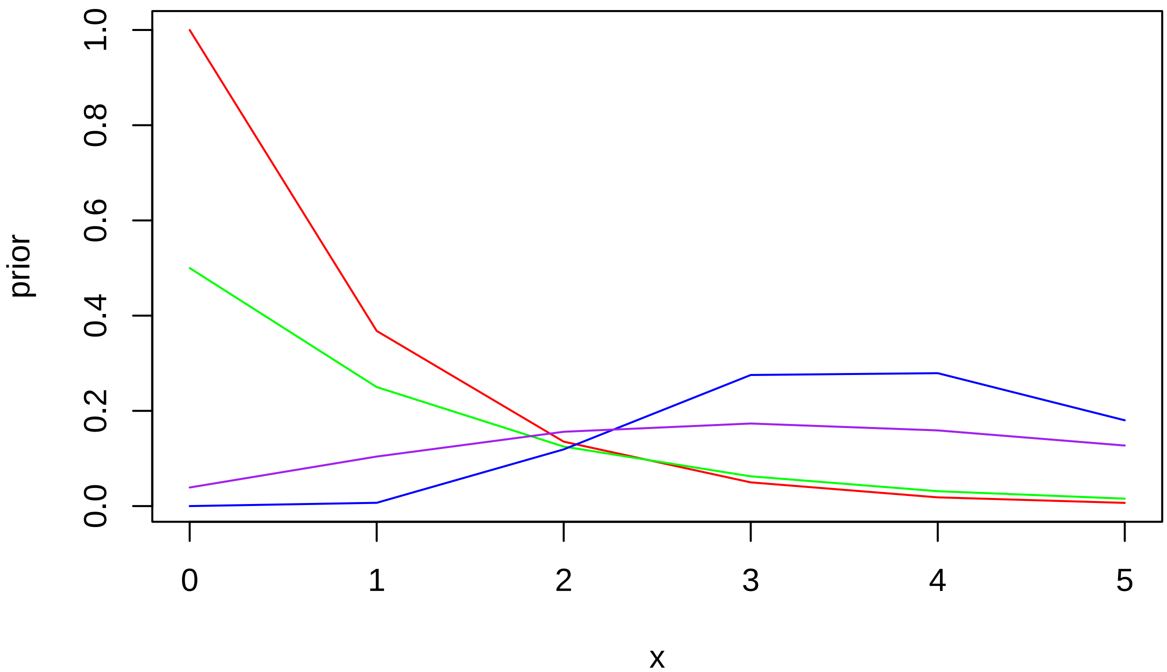
$$
\begin{aligned}
f(y|a, b) &= \frac{b^a}{\Gamma(a) y!} \frac{\Gamma(y+a)}{(b+1)^{y+a}} \\
&= \frac{(y+a-1)!}{(a-1)! y!} \frac{b^a}{(b+1)^{y+a}} \\
&= \binom{y+a-1}{y} b^a (b+1)^{-y-a} \\
&= \binom{y+a-1}{y} \left(\frac{b}{1+b}\right)^a \left(\frac{1}{1+b}\right)^y
\end{aligned}
$$

$$Which \quad f(y|a,b) \sim NB(a, \frac{b}{1+b})$$

d)

```
x <- seq(from = 0, to = 5, by = 1)
prior <- dgamma(x, 1, 1)
posterior <- dnbinom(x, 1, 1/2)
prior2 <- dgamma(x, 8, 2)
posterior2 <- dnbinom(x, 8, 2/3)

plot(x, prior, type = "l", col = "red")
lines(x, posterior, type = "l", col = "green")
lines(x, prior2, type = "l", col = "blue")
lines(x, posterior2, type = "l", col = "purple")
```



Question 2

a)

$$L(y|\lambda) = \frac{\lambda^{\sum(yi)} e^{-ny}}{\Pi y!} \quad p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b}$$

$$p(\lambda|y) \propto \lambda^{\sum(yi)+a-1} e^{-\lambda(b+n)}, \lambda > 0$$

$$posterior \ follows \ gamma \ distribution \ \sim Gamma(\sum(yi) + a, b+n)$$

Question 3

a)

$$
\begin{aligned}
mean; \quad E(y|\tau) &= \int_0^\infty y(\frac{2}{\pi\tau})^{1/2} e^{\frac{-(y)^2}{2\tau}} dy \\
&= (\frac{2}{\pi\tau})^{1/2}[-\tau e^{\frac{-y^2}{2\tau}}|_0^\infty] \\
&= (\frac{2\tau}{\pi})^{1/2}
\end{aligned}
$$

$$
variance; \quad \tau(1-\frac{2}{\pi}) \qquad mode; \quad 0 \qquad median; \quad \sqrt{2\tau} erf^{-1}(1/2)
$$

b)

$$
\begin{aligned}
indicator, \quad I &= \{^{1}_{0} \quad {}^{0\,<\,yi}_{otherwise} \\
Lf(y|\tau) &= (\frac{2}{\pi\tau})^{n/2} e^{-\frac{(\sum yi)^2}{2\tau}} I \\
logLf(y|\tau) &= \frac{n}{2} log\frac{2}{\pi\tau} - \frac{\sum yi^2}{2\tau} I \\
\triangledown logLf(y|\tau) &= -\frac{n}{2\tau^3} + \frac{\sum yi^2}{2\tau^2} \text{ , set equal to } 0 \\
we\ get\ , \hat{\tau} &= \frac{n}{\sum yi^2}, \quad mle
\end{aligned}
$$

$$
\begin{aligned}
Lf(y|\tau) &= (\frac{2}{\pi\tau})^{n/2} e^{-\frac{(\sum yi)^2}{2\tau}} I \\
T &= \sum yi^2 \text{ ,is the sufficient statistics for } \tau
\end{aligned}
$$

c)

$$
indicator, \quad I = \{^{1}_{0} \quad {}^{0\,<\,yi}_{otherwise}
$$

$$
\begin{aligned}
Lf(y|\tau) &= (\frac{2}{\pi\tau})^{n/2} e^{-\frac{(\sum yi)^2}{2\tau}} I \\
f(\tau) &= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \tau^{-\frac{a}{2}-1} e^{-\frac{b}{2\tau}} \\
posterior; \quad f(\tau|y) &\propto \tau^{-n/2-a/2-1} e^{\frac{-(\sum yi^2+b)}{2\tau}}, \ yi > 0 \\
posterior\ follows &\sim inver.\Gamma(\frac{n}{2}+\frac{a}{2}, \frac{(\sum yi^2+b)}{2})
\end{aligned}
$$

Question 4

a)

When data is non-negative, also collected at the individual level (binary response) or group level (proportion) might use half-normal distribution. When data shows binary response and non-negative values will lead to use half-normal distribution. And my answer is NO for Alaska temperature question.

b)

Example: Half-normal can be use to model distance at which wildlife observe on a boat detect an animal when surveying, such as whales in the ocean.

c)

The competitor distribution I might use is “Poisson distribution”.

Question 5

a)

$$f(y|\sigma) = (\frac{2}{\pi\sigma^2})^{1/2} e^{-\frac{(y)^2}{2\sigma^2}},\ y > 0$$

b)

No, because because the prior distribution is not conjugate prior. It is about σ not about σ^2

c)

Yes, I thinks that it can be inverse gamma.

Question 6

a)

$$f(y|\theta) \sim beta(\theta, 1)$$
$$so\ the\ normalizing\ constant\ is;\ \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)} = \theta$$

b)

$$Lf(y|\theta) = \left(\frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)}\right)^n \Pi_{i=1}^n yi^{\theta-1} I_{0<yi<1}$$

$$\begin{aligned}
&By\ Factorization \\
h(y) &= \Pi_{i=1}^n yi^{-1} I_{0<yi<1} \\
k(g(y)|\theta) &= \theta^n \Pi_{i=1}^n yi^{\theta} \\
&= \theta^n exp(\theta log(\sum_{i=1}^n yi) \\
T &= log(\sum_{i=1}^n yi), is\ the\ sufficient\ statistics \\
&So\ we\ can\ say\ that\ T = y^{\frac{1}{n}}\ \ is\ also\ the\ sufficient\ statistics\ which\ follows\ Geometric\ mean.
\end{aligned}$$