Homework 03: Due 9/27 (completion based)

Stat061-F23
Prof Amanda Luby

In HW1, we found the MOM and MLE estimates for each of the probability distributions below. For Q1-Q4, find (a) the posterior distribution for an iid sample of $X_1, X_2, \dots, X_n$ and (b) the posterior mean (our Bayesian estimate).

1. The parameter λ for a Poisson distribution where $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ for $k = 0, 1, 2, \dots$ and we assume the prior distribution for λ is Gamma(α, β). (This should be a named distribution; be sure to specify the parameters)

Solution:

(a) We need to solve for our likelihood:

$$L(\lambda) = \frac{\lambda^{\sum x_i}}{\prod x_i!}e^{-n\lambda}$$

Now, we can find our posterior distribution:

$$\begin{aligned} p(\lambda | X_1, X_2, \dots, X_n) &\propto \frac{1}{\beta^\alpha \Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda/\beta} \times \frac{\lambda^{\sum x_i}}{\prod x_i!}e^{-n\lambda} \\ &\propto \lambda^{\alpha + \sum x_i - 1}e^{-\lambda(1/\beta + n)} \end{aligned}$$

This is the kernel for a Gamma$(\alpha + \sum x_i, \frac{1}{1/\beta + n})$ distribution.

(b) We know the mean of a Γ distribution so our posterior mean is

$$E(\lambda | X_1, X_2, \dots, X_n) = \frac{\alpha + \sum x_i}{1/\beta + n}$$

2. The parameter p in the Geometric distribution where $P(X = k) = p(1-p)^{k-1}$ for $k = 1, 2, 3, \dots$ and we assume the prior distribution for p is Beta(a, b). (This should be a named distribution; be sure to specify the parameters)

Solution:

(a) We need to solve for our likelihood:

$$L(p) = (1-p)^{\sum x_i - n}p^n$$

Now, we can find our posterior distribution:

$$\begin{aligned} p(p | X_1, X_2, \dots, X_n) &\propto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1} \times (1-p)^{\sum x_i - n}p^n \\ &\propto p^{n+a-1}(1-p)^{\sum x_i - n + b - 1} \end{aligned}$$

This is the kernel for a Beta$(n + a + \sum x_i + b - n)$ distribution.

(b) We know the mean of a Beta distribution so our posterior mean is

$$E(p|X_1, X_2, \ldots, X_n) = \frac{n + a}{a + b + \sum x_i}$$

3. The parameter α in the distribution with pdf $f(x|\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}[x(1-x)]^{\alpha-1}$ where $x \in [0,1]$ and we assume the prior distribution for α is Unif(0,1). (You will probably not be able to solve the integral to determine the posterior mean, but write out the integral you would have to solve.)

Solution:

(a) We need to solve for our likelihood:

$$L(\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}^n \prod [x(1-x)]^{\alpha-1}$$

Now, we can find our posterior distribution:

$$\begin{aligned} p(\alpha|X_1, X_2, \ldots, X_n) &\propto 1 \times \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}^n \prod [x(1-x)]^{\alpha-1} \\ &\propto \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}^n \prod [x(1-x)]^{\alpha-1} \end{aligned}$$

This isn't a distribution I readily recognize.

(b) Since we don't know what the distribution is, we have so solve a complicated integral to find the posterior mean

$$E(\alpha|X_1, X_2, \ldots, X_n) = \int \alpha \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}^n \prod [x(1-x)]^{\alpha-1} d\alpha$$

Because the α is in the Γ functions, and the exponent and is a term on its own, how to solve this integral is not obvious to me. We'd likely have to solve numerically

4. The parameter β in the Pareto distribution with pdf $f(x|\beta) = \frac{\beta}{x^{\beta+1}}$ where $x > 1$ and we assume the prior distribution for β is Gamma(α, λ). (This should be a named distribution; be sure to specify the parameters)

Solution:

(a) We need to solve for our likelihood:

$$L(\beta) = \beta^n \prod x_i^{-\beta-1}$$

Now, we can find our posterior distribution:

$$\begin{aligned}
p(\beta|X_1, X_2, \ldots, X_n) &\propto \frac{1}{\lambda^\alpha \Gamma(\alpha)} \beta^{\alpha-1} e^{-\beta/\lambda} \times \beta^n \prod x_i^{-\beta-1} \\
&\propto \beta^{n+\alpha-1} e^{-\beta/\lambda} e^{-(\beta+1)\sum \ln x_i} \\
&\propto \beta^{n+\alpha-1} e^{-\beta(1/\lambda + \sum \ln x_i)}
\end{aligned}$$

This is the kernel for a Gamma$(n+\alpha, \frac{1}{1/\lambda+\sum \ln x_i})$ distribution.

Note that in the second step, we're using the fact that:

$$\prod x_i^{-\beta-1} = e^{\ln(\prod x_i^{-(\beta+1)})} = e^{-(\beta+1)\sum \ln x_i}$$

which is another Stats Trick™ we sometimes use.

(b) We know the mean of a Gamma distribution so our posterior mean is

$$E(\lambda|X_1, X_2, \ldots, X_n) = \frac{n+\alpha}{1/\lambda + \sum \ln x_i}$$

5. For a Binomial(n, π) observation y, consider the Bayes estimator of π using a Beta(α, β) prior distribution.

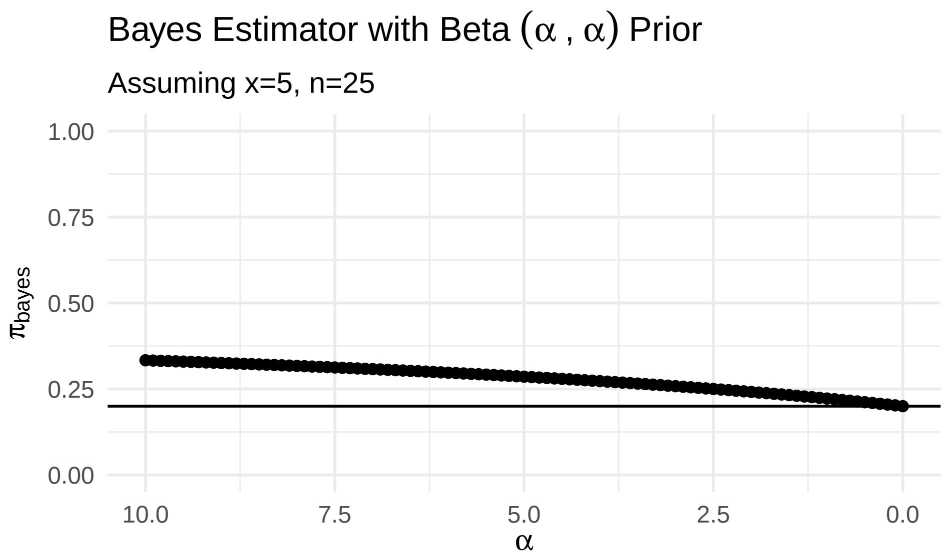
   (a) For large n, show that the posterior distribution of π has approximate mean $\hat{\pi} = \frac{y}{n}$ (it also has approximate variance $\frac{\hat{\pi}(1-\hat{\pi})}{n}$. Relate this result to classical estimation.

   (b) Show that the MLE estimator is a limit of Bayes estimators, for a certain sequence of $\alpha = \beta$ values.

Solution:

(a) In Notes03 in class, we showed that the Bayes estimator is $\hat{\pi} = \frac{x+\alpha}{n+\beta+\alpha}$ which can be written as $\frac{x}{n+\beta+\alpha} + \frac{\alpha}{n+\beta+\alpha}$. When n is large, α and β become negligible relative to n, and so the first term approaches $\frac{x}{n}$ and the second term approaches 0. $\hat{\pi}$ thus approaches $\bar{x}$, which is the MLE.

(b) In (a), we showed that the Bayes estimator approaches the MLE as $n \to \infty$. Here, we want to show that there is some sequence of α such that $\hat{\pi}$ approaches $\bar{x}$ for n held constant. Since $\hat{\pi} = \frac{x+\alpha}{n+\alpha+\alpha} = \frac{x+\alpha}{n+2\alpha}$, this occurs when $\alpha \to 0$. Note that because of the way the Gamma distribution is defined, α cannot actually be equal to zero. Below is one example of this sequence.



6. In class on Friday, we defined the posterior distribution for θ as:

$$f_{\theta|X}(\theta|x) = \frac{f_x(x|\theta)f_\theta(\theta)}{\int f_x(x|\theta)f_\theta(\theta)d\theta}$$

which is true if we observe 1 draw from the data model and have $X \sim f_x$.

If we have n IID observations $X_1, \ldots, X_n$, we replace $f_x(x|\theta)$ with the *joint pdf*:

$$f_{X^n}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_x(x_i|\theta) = L_n(\theta)$$

where x^n denotes the set of $(x_1, \ldots., x_n)$, and $L_n(\theta)$ is the same likelihood function that is so near and dear to our hearts.

Then, the posterior distribution is:

$$f(\theta|x^n) = \frac{f_{x^n}(x^n|\theta)f_\theta(\theta)}{\int f_{x^n}(x^n|\theta)f_\theta(\theta)d\theta} = \frac{L_n(\theta)f_\theta(\theta)}{c_n} \propto L_n(\theta)f_\theta(\theta)$$

(a) Why can we write the joint pdf as $\prod_{i=1}^{n} f_x(x_i|\theta)$?
(b) What is c_n and how do we know that it is a constant?
(c) Explain what the $\propto L_n f_\theta(\theta)$ means.
(d) Do you think the Bayes estimator will generally be more similar to the MLE or to the MoM? Why?

Solution:

(a) Since $X_1, \ldots, X_n$ are *independent*, we can multiple their individual pdfs to get the joint pdf.

Since they are *identical*, we know that their individual pdfs are all the same.

(b) $c_n = \int f_{x^n}(x^n|\theta) f_\theta(\theta) d\theta$. We know that it is constant (with respect to theta) since we are integrating over all possible theta. The integral will therefore not be a function of θ.

(c) $\propto$ means "proportional to", or "up to a normalizing constant", which means that the only difference between it and the previous step is being multipled by some constant c_n. We can always go back and figure out what c_n is, but it's generally not necessary for figuring out the posterior distribution $f(\theta|x^n)$.

(d) Remember that the MoM has no idea what the likelihood is, it only sees the moments. Since both the MLE and the Bayes estimator rely on the likelihood function, we might expect them to be more similar to one another than the MoM.

7. Wrap up lab activity
8. Review for quiz on Wednesday!