Assignment 1. Solutions

STAT314/STAT461

Set: Tue, July-27. Due: Fri Aug-06

Problem 1. Inverse Probability.

A total of 38 ancient manuscripts have so far been found in a certain area of England. Of those, 20 have been solidly attributed to the Historian A, 17 to the Historian B, and only 1 to the Historian C. All the chronicles pertain to the legendary King Arthur. Historian A tends to mention him on average 5 times per page, Historian B - 3 times per page, and Historian C is a fan and mentions King Arthur about 10 times per page.

Assume Poisson distribution for the number of mentions per page, so that you can use Poisson p.d.f. to evaluate the probability of a specific number of references to King Arthur on a page.

(a) If a single page is found from the same time period, and there is no mention of King Arthur, what is the probability that it was written by the historian C? (**1pt**)

(b) What assumptions have you made in the process of your analysis? List a couple; explain why you think they apply; and give counterexamples when they would not. (**1pt**)

Solution.

Using the information available so far, and assuming that the sample so far has been representative and that no other sources (Historians D, E etc.) exist. Then we can evaluate the respective probabilities that a manuscript has been written by a particular historian as

$$\begin{aligned}
Pr(\text{Hist=A}) &= 20/38, \\
Pr(\text{Hist=B}) &= 17/38, \\
Pr(\text{Hist=C}) &= 1/38.
\end{aligned}$$

Furthermore, if x denotes the number of mentions of King Arthur, then, assuming Poisson distribution,

$$Pr(x|Hist) = \frac{\lambda_{\text{Hist}}^x \exp(-\lambda_{\text{Hist}})}{x!},$$

where λ_{Hist} is the average frequency of mentions of King Arthur per page attributed to the respective Historian. In other words,$\lambda_A = 5, \lambda_B = 3, \lambda_C = 10$.

Specifically,

$$Pr(x = 0|Hist) = \frac{\lambda_{\text{Hist}}^0 \exp(-\lambda_{\text{Hist}})}{0!} = \exp(-\lambda_{\text{Hist}}).$$

Now, we are ready to use the Bayes' Formula:

$$\begin{aligned} Pr(Hist = C|x = 0) &= \frac{Pr(x = 0|Hist = C)Pr(Hist = C)}{Pr(x = 0|A)Pr(A) + Pr(x = 0|B)Pr(B) + Pr(x = 0|C)Pr(C)} \\ &= \frac{\exp(-10)1/38}{\exp(-5)20/38 + \exp(-3)17/38 + \exp(-10)1/38} \\ &\approx 0.000046. \end{aligned} \tag{1}$$

In other words, the probability that the newly found page was produced by Historian C is extremely small.

Some assumptions you might want to mention:

- one of only these three historians could have written the page (with an obvious counterexample of there being others, not before seen)
- Poisson distribution for the counts per page was assumed. A zero-inflated distribution is more likely (some pages may be about events which did not involve King Arthur at all).

Problem 2: Maximum Likelihood and Bayesian inference.

(a) To prove that the function $f(x) = \lambda \exp^{-\lambda x}$ is a valid p.d.f., we need to prove that it is non-negative everywhere and that it integrates to 1 over the domain of x.

Since the parameter λ is non-negative, and the exponent is always positive, the product will also be non-negative. I.e., $f(x) \geq 0$ for $\lambda > 0$, $x \geq 0$.

Let's look at the integral:

$$\int_0^\infty \lambda \exp^{-\lambda x} dx = -\exp^{-\lambda x} |_0^\infty = -(0-1) = 1.$$

QED.

(b) The joint likeihood can be found as:

$$L = \prod_i f(x_i) = \prod_i \left(\lambda \exp^{-\lambda x_i}\right) = \lambda^n \exp(-\lambda \sum_i x_i) \qquad \text{for } x_i \geq 0, i = 1, ..., n.$$

In order to find the maximum likelihood estimator for λ, we need to find a value $\hat{\lambda}$ that maximizes the above function.

It is easier to differentiate a sum rather than a product, and log() is a monotonically increasing function, so if $\hat{\lambda}$ maximizes $\log(L)$, then it will also maximize L.

$$\log L = n \log \lambda - \lambda \sum_i x_i.$$

Differentiating with respect to λ:

$$\frac{d \log L}{d\lambda} = n/\lambda - \sum_i x_i = 0.$$

when

$$\lambda = \frac{n}{\sum_i x_i} = \frac{1}{\bar{x}},$$

where $\bar{x}$ is the sample average.

Remember, that we need to confirm that this is indeed a global maximum. One way to do it is differentiate again:

$$\frac{d^2 \log L}{d\lambda^2} = \frac{d}{d\lambda}\left(n/\lambda - \sum_i x_i\right) = -\frac{1}{\lambda^2}.$$

The second derivative is negative for any positive value of λ. The function is thus strictly concave, and $\hat{\lambda} = \frac{1}{\bar{x}}$ is the MLE.

NB. Another way is to evaluate and compare the values of $\log L$ at the endpoints of the domain and at $1/\bar{x}$. This would involve the use of limits and is, perhaps, unnecessarily complicated, but is still a solution.

(c) The prior p.d.f. is

$$f(\lambda|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\lambda^{\alpha_0 - 1}\exp(-\beta_0\lambda)$$

Using Bayes' theorem and proportionality (i.e., concentrating on the numerator only, and discarding any terms which do not contain the parameter of interest), we get

$$\begin{aligned} f(\lambda|x, \alpha_0, \beta_0) &\propto \lambda^n \exp(-\lambda\sum_i x_i)\lambda^{\alpha_0 - 1}\exp(-\beta_0\lambda) \\ &\propto \lambda^{\alpha_0 + n - 1}\exp(-(\beta_0 + \sum_i x_i)\lambda) \end{aligned}$$

which is proportional to a Gamma density up to a constant. In other words,

$$\lambda|x, \alpha_0, \beta_0 \sim Gamma(\alpha_0 + n, \beta_0 + \sum_i x_i)$$

(d) The posterior mean of the above Gamma distribution is:

$$E(\lambda|x, \alpha_0, \beta_0) = \frac{\alpha_0 + n}{\beta_0 + \sum_i x_i} = \frac{\alpha_0/n + 1}{\beta_0/n + \bar{x}}$$

As $n \to \infty$, $E(\lambda|x, \alpha_0, \beta_0) \to \frac{1}{\bar{x}}$. I.e., the Bayesian estimate approaches the MLE.

Problem 3: Prior distribution.

Note, that the parameter λ is the rate parameter. I.e., how many buses per minute you expect on average. We thus want to find parameters α_0 and β_0 which would ensure that

$$\alpha_0/\beta_0 \approx 1/15$$

and that most of the distribution 1/16 and 1/14.

One way to do it is to set the variance to equal $((1/14 - 1/16)/4)^2 \approx 0.004^2$.

$$\alpha_0/\beta_0^2 = 0.004^2$$

Solving this yields $\beta_0 = 1/15/(0.004^2) \approx 4166.67$, and $\alpha_0 = \beta_0/15 = 277.78$.

To check, try making a plot or a simulation and see how well it incorporates your prior assumptions

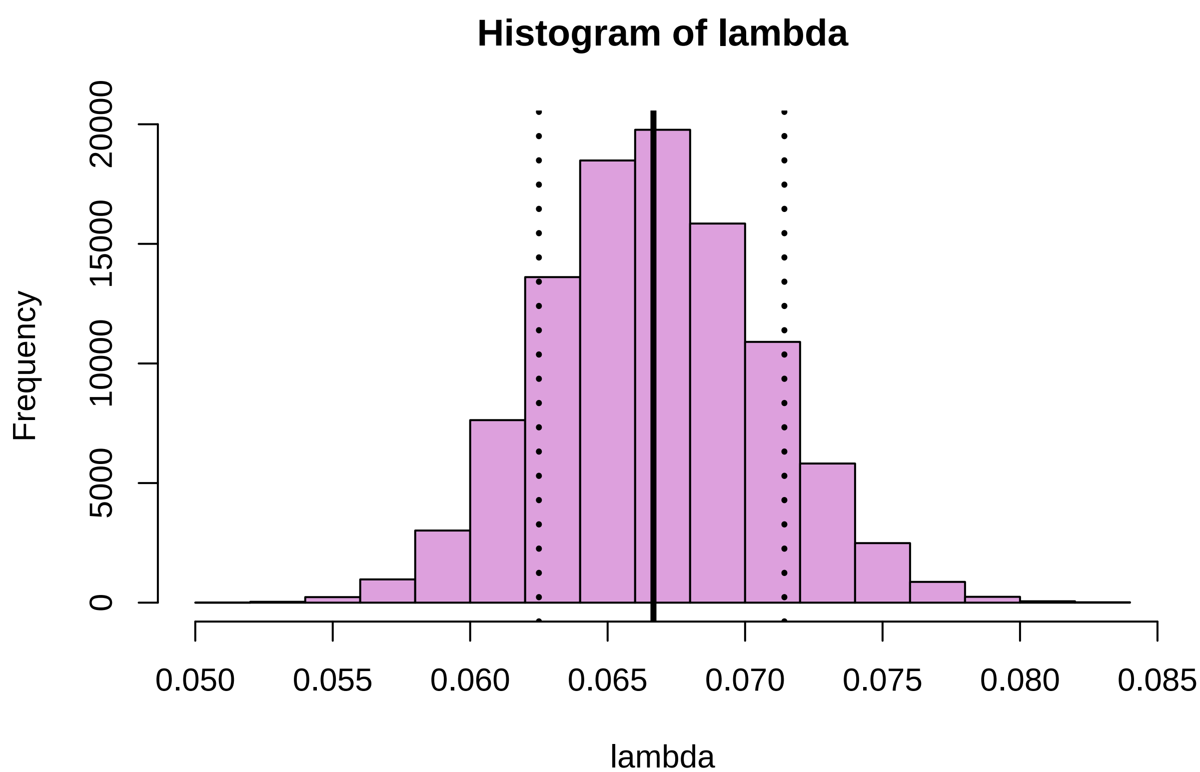
```r
alpha0 <- 277.78
beta0 <- 4166.67

lambda <- rgamma(10^5,alpha0,beta0)

mean(1/lambda)
```

```
## [1] 15.05029
```

```r
quantile(1/lambda,c(0.025,.975))
```

```
##     2.5%    97.5%
## 13.37896 16.92814
```

```r
hist(lambda,col='plum')
abline(v=c(1/14,1/15,1/16),lwd=3,lty=c(3,1,3))
```



Not too bad. Of course, the 4σ approximation only applies to the normal distribution density. Another way to incorporate this prior assumption would be to search for β_0 on a grid so that $\alpha_0/\beta_0 = 1/15$ and there is about 95% of the density mass between the 1/14 and 1/16.

Problem 4: Prior Predictive Distribution.

To generate a sample from the prior predictive distribution:

```r
alpha0 <- 277.78
beta0 <- 4166.67
```

```r
lambda <- rgamma(10^4,alpha0,beta0)
x <- rexp(10^4,lambda)

hist(x,col='plum',xlab='Time waiting for a bus',freq=F)
```

Based on this simulation, the probability that a random wait will be at least 20 minutes is

```r
mean(x>=20)
```

```
## [1] 0.2574
```

So, not particularly unusual. We can thus conclude that the data point and the model do not disagree.