MATH 341 / 650.3 Spring 2020 Homework #7

Frank Palma Gomez

Friday 8th May, 2020

Problem 1

These are questions about McGrayne's book, chapter 16.

(a) [easy] What was the main problem facing Bayesian Statistics in the early 1980's?

The main problem that Bayesian Statistic was facing in the early 1980's is the "curse of dimensionality."

(b) [harder] What is the "curse of dimensionality?"

The curese of dimensionality is when data dispersion occurs and dimension increases.

(c) [easy] How did Bayesian Statistics help sociologists?

It helped them by allowing them to test hypothesis accurately given the data within their model.

(d) [easy] How did Gibbs sampling come to be?

Gibbs sampling came to be when the German brothers were trying to reduce the noise and increase the resolution of unfocused images.

(e) [easy] Were the Geman brothers the first to discover the Gibbs sampler?

No

(f) [easy] Who officially discovered the expectation-maximization (EM) algorithm? And who *really* discovered it?

Aurthur Dempster and Nan Laird first published the algorithm despite Wong discovering it.

(g) [harder] How did Bayesians "break" the curse of dimensionality?

They broke it by reducing the dimension.

(h) [harder] Consider the integrals we use in class to find expectations or to approximate PDF's / PMF's — how can they be replaced?

We can use Gibbs sampling.

(i) [easy] What did physicists call "Markov Chain Monte Carlo" (MCMC)? (p222)

They called it statistical sampling.

(j) [easy] Why is sampling called "Monte Carlo" and who named it that?

Named after Stanislaw Ulams gambling uncle by Nicolas Metropolis

(k) [easy] The Metropolis-Hastings (MH) Algorithm is world famous and used in myriad applications. Why didn't Hastings get any credit?

When Hasting published the paper, powerful computers were not around.

(l) [easy] The combination of Bayesian Statistics + MCMC has been called ... (p224)

"Most powerful mechanism ever created for processing data and knowledge."

(m) [E.C.] p225 talks about Thomas Kuhn's ideas of "paradigm shifts." What is a "paradigm shift" and does Bayesian Statistics + MCMC qualify?

(n) [easy] How did the BUGS software change the world?

Made MCMC available to everyone.

(o) [easy] Lindley said that Bayesian Statistics would win out over Frequentist Statistics because it was more logical. What in reality was the reason for the eventual victory of Bayes?

It gained influence from multiple desciplines that used it.

(p) [E.C.] One of my PhD advisors, Ed George at Wharton told me that "Bayesian Statistics is really 'knowledge engineering."' Is this true? Explain.

(q) [E.C.] Take a look at the software Stan. What kind of potential does it have to change the world? Note: I had an opportunity to work on Stan as a postdoc (right after I finished his PhD) but chose to come to QC instead.

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	`qbeta`$(p,\ \alpha,\ \beta)$	`d-`$(x,\ \alpha,\ \beta)$	`p-`$(x,\ \alpha,\ \beta)$	`r-`$(\alpha,\ \beta)$
betabinomial	`qbetabinom`$(p,\ n,\ \alpha,\ \beta)$	`d-`$(x,\ n,\ \alpha,\ \beta)$	`p-`$(x,\ n,\ \alpha,\ \beta)$	`r-`$(n,\ \alpha,\ \beta)$
binomial	`qbinom`$(p,\ n,\ \theta)$	`d-`$(x,\ n,\ \theta)$	`p-`$(x,\ n,\ \theta)$	`r-`$(n,\ \theta)$
exponential	`qexp`$(p,\ \theta)$	`d-`$(x,\ \theta)$	`p-`$(x,\ \theta)$	`r-`(θ)
gamma	`qgamma`$(p,\ \alpha,\ \beta)$	`d-`$(x,\ \alpha,\ \beta)$	`p-`$(x,\ \alpha,\ \beta)$	`r-`$(\alpha,\ \beta)$
inversegamma	`qinvgamma`$(p,\ \alpha,\ \beta)$	`d-`$(x,\ \alpha,\ \beta)$	`p-`$(x,\ \alpha,\ \beta)$	`r-`$(\alpha,\ \beta)$
negative-binomial	`qnbinom`$(p,\ r,\ \theta)$	`d-`$(x,\ r,\ \theta)$	`p-`$(x,\ r,\ \theta)$	`r-`$(r,\ \theta)$
normal (univariate)	`qnorm`$(p,\ \theta,\ \sigma)$	`d-`$(x,\ \theta,\ \sigma)$	`p-`$(x,\ \theta,\ \sigma)$	`r-`$(\theta,\ \sigma)$
poisson	`qpois`$(p,\ \theta)$	`d-`$(x,\ \theta)$	`p-`$(x,\ \theta)$	`r-`(θ)
T (standard)	`qt`$(p,\ \nu)$	`d-`$(x,\ \nu)$	`p-`$(x,\ \nu)$	`r-`(ν)
T (nonstandard)	`qt.scaled`$(p,\ \nu,\ \mu,\ \sigma)$	`d-`$(x,\ \nu,\ \mu,\ \sigma)$	`p-`$(x,\ \nu,\ \mu,\ \sigma)$	`r-`$(\nu,\ \mu,\ \sigma)$
uniform	`qunif`$(p,\ a,\ b)$	`d-`$(x,\ a,\ b)$	`p-`$(x,\ a,\ b)$	`r-`$(a,\ b)$

Table 1: Functions from `R` (in alphabetical order) that can be used on this assignment and exams. The hyphen in colums 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 2

Now we will move to the Bayesian normal-normal model for estimating both the mean and variance and demonstrate similarities with the classical results.

(a) [harder] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\, \sigma^2)$ and X represents all $X_1, \ldots, X_n$, Find the kernel of $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ if $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$. Use the substitution that we made in class:

$$\sum_{i=1}^{n}(x_i - \theta)^2 = (n-1)s^2 + n(\bar{x} - \theta)^2$$

where $s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. We do this here because this substitution is important for what comes next.

$$\mathbb{P}\left(\theta,\ \sigma^2 \mid X\right) \propto \mathbb{P}\left(X \mid \theta,\ \sigma^2\right)\mathbb{P}\left(\theta,\ \sigma^2\right)$$

$$\propto (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n}(x_i-\theta)^2\right)(\frac{1}{\sigma^2})$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{(n-1)s^2/2}{\sigma^2}\right) \exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)$$

$$\propto \text{NormInvGamma}(\bar{x},\ n,\ \frac{n}{2},\ \frac{(n-1)s^2}{2})$$

(b) [harder] Using Bayes Rule, break up $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ into two pieces. How are those two pieces distributed?

$$\text{Recall: } \mathbb{P}\left(\theta,\ \sigma^2 \mid X\right) \propto \mathbb{P}\left(X \mid \theta,\ \sigma^2\right)\mathbb{P}\left(\ \sigma^2 \mid X\right)$$

$$\mathbb{P}\left(X \mid \theta,\ \sigma^2\right) \propto \mathcal{N}\left(\bar{x},\ \frac{\sigma^2}{n}\right)$$

$$\mathbb{P}\left(\ \sigma^2 \mid X\right) \propto \text{InvGamma}(\frac{n-1}{2}, \frac{(n-1)s^2}{2})$$

(c) [harder] Using your answer from (b), explain in English how you can create samples from the distribution $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ that look like $\{[\theta_1, \sigma_1^2], [\theta_2, \sigma_2^2], \ldots, [\theta_S, \sigma_S^2]\}$.

(a) Sample σ^2_{samp} from $\mathbb{P}(\ \sigma^2 \mid X)$

(b) Sample θ_{samp} from $\mathcal{N}\left(\bar{x},\ \frac{\sigma^2}{n}\right)$

(c) Return $[\theta_{samp}, \sigma^2_{samp}]$

(d) Repeat previous steps s times

(d) [difficult] Using these samples, how would you estimate $\mathbb{E}[\theta \mid X]$ and $\mathbb{E}[\sigma^2 \mid X]$? Why is $\mathbb{E}[\theta \mid X]$ of paramount importance?

(a) $\mathbb{E}[\theta \mid X] \approx \frac{\theta_{samp,1}+\cdots+\theta_{samp,s}}{s}$

(b) $\mathbb{E}[\sigma^2 \mid X] \approx \frac{\sigma^2_{samp,1}+\cdots+\sigma^2_{samp,s}}{s}$

Because $\mathbb{E}[\theta \mid X]$ is the best guess for θ.

(e) [difficult] Using these samples, how would you estimate a 95% CR for θ?

$$CR_{\theta,95\%} = [\texttt{samplequantile}(\theta_j, 0.025), \texttt{samplequantile}(\theta_j, 0.975)]$$

(f) [difficult] Using these samples, how would you obtain a p-val for testing if $\sigma^2 > 1.364$?

$$p_{val} = \frac{1}{N}\sum_{l=1} \mathbb{1}\sigma^2_{l,j} \leq 1.364$$

(g) [difficult] [MA] Using these samples, how would you estimate $\text{Corr}\left[\theta \mid X,\ \sigma^2 \mid X\right]$ i.e. the correlation between the posterior distributions of the two parameters?

(h) [easy] Find $\mathbb{P}\left(\theta \mid X,\ \sigma^2\right)$ by using the full posterior kernel from (a) and then conditioning on σ^2. You should get the same answer as we did before the midterm.

$$\mathbb{P}\left(\theta \mid X,\ \sigma^2\right) = \mathcal{N}\left(\bar{x},\ \frac{\sigma^2}{n}\right)$$

(i) [easy] Find $\mathbb{P}\left(\sigma^2 \mid X,\ \theta\right)$ by using the full posterior kernel from (a) and then conditioning on θ. You should get the same answer as we did before the midterm.

$$\mathbb{P}\left(\sigma^2 \mid X,\ \theta\right) = \text{InvGamma}\left(\frac{n - n_0}{2},\ \frac{n\sigma^2_{MLE} + n_0\sigma_0^2}{2}\right)$$

(j) [difficult] Show that $\mathbb{P}\left(\theta \mid X\right)$ is a non-standard T distribution and find its parameters. Assume the prior $\mathbb{P}\left(\theta,\ \sigma^2\right) \propto \frac{1}{\sigma^2}$. The answer is in the notes, but try to do it yourself.

$$\begin{aligned}
\mathbb{P}\left(\theta \mid X\right) &= \int_0^\infty \mathbb{P}\left(\theta,\ \sigma^2 \mid X\right) d\sigma^2 \\
&\propto \int_0^\infty \left(\sigma^2\right)^{-\frac{n}{2}-1} \exp\left(-\frac{(n-1)s^2/2}{\sigma^2}\right) \exp\left(-\frac{n}{2\sigma^2}(\theta - \bar{x})^2\right) d\sigma^2 \\
&\propto \Gamma\left(\frac{n}{2}\right)\left(\frac{(n-1)s^2 + n(\theta - \bar{x})^2}{2}\right)^{-\frac{n}{2}} \\
&\propto \left(\frac{(n-1)s^2 + n(\theta - \bar{x})^2}{2}\right)^{-\frac{n}{2}} \\
&\propto \left(\frac{(n-1)s^2}{2}\right)^{-\frac{n}{2}}\left(1 + \frac{n(\theta - \bar{x})^2}{(n-1)s^2}\right)^{\frac{-(n-1)-1}{2}} \\
&\propto \left(1 + \frac{1}{n-1}\left(\frac{(\theta - \bar{x})^2}{s^2/2}\right)\right)^{\frac{-(n-1)-1}{2}} \\
&\propto T_{n-1}\left(\bar{x},\ \frac{s}{\sqrt{n}}\right)
\end{aligned}$$

(k) [difficult] Show that $\mathbb{P}\left(\sigma^2 \mid X\right)$ is an inverse gamma and find its parameters. Assume the prior $\mathbb{P}\left(\theta,\ \sigma^2\right) \propto \frac{1}{\sigma^2}$. The answer is in the notes, but try to do it yourself.

(l) [easy] Write down the distribution of $\mathbb{P}(X^* \mid X)$ assuming the prior $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$. This is in the notes.

$$T_{n-1}(\bar{x}, s\ \sqrt{\frac{n+1}{n}})$$

(m) [E.C.] Prove what you wrote in the previous question: $\mathbb{P}(X^* \mid X)$ is the non-standard T distribution and find its parameters.

(n) [harder] Explain how to sample from the distribution of $\mathbb{P}(X^* \mid X)$. Also in the notes.

(a) Sample σ^2_{samp} from InvGamma $\left(\frac{n-1}{2},\ \frac{(n-1)s^2}{2}\right)$

(b) Sample θ_{samp} from $\mathcal{N}\left(\bar{x},\ \frac{\sigma^2}{n}\right)$ where $\theta = \theta_{samp}$

(c) Sample X^*_{samp} from $\mathcal{N}(\theta,\ \sigma^2)$ where $\theta = \theta_{samp}$ and $\sigma^2 = \sigma^2_{samp}$

(d) Return X^*_{samp}

(e) Repeat previous steps s times

(o) [harder] Now consider the informative conjugate prior of $\mathbb{P}(\theta,\ \sigma^2) = \mathbb{P}(\theta \mid \sigma^2)\,\mathbb{P}(\sigma^2)$ where $\mathbb{P}(\theta \mid \sigma^2) = \mathcal{N}\left(\mu_0,\ \frac{\sigma^2}{m}\right)$ and $\mathbb{P}(\sigma^2) = \text{InvGamma}\left(\frac{n_0}{2},\ \frac{n_0\sigma_0^2}{2}\right)$ i.e. the general normal-inverse-gamma. What is its kernel? Collect common terms and be neat.

$$\begin{aligned}
\mathbb{P}\left(\theta,\ \sigma^2\right) &= \mathcal{N}\left(\mu_0,\ \frac{\sigma^2}{m}\right) \text{InvGamma}\left(\frac{n_0}{2},\ \frac{n_0\sigma_0^2}{2}\right) \\
&\propto \exp\left(\frac{\mu_0 m}{\sigma^2}\theta - \frac{m}{2\sigma^2}\theta^2 (\sigma^2)^{-\frac{n_0}{2}-1}\right) \exp\left(-\frac{n_0\sigma_0^2}{2\sigma^2}\right) \\
&\propto \exp\left(\frac{\mu_0 m}{\sigma^2}\theta - \frac{m}{2\sigma^2}\theta^2 - \frac{n_0\sigma_0^2}{2\sigma^2}\right) (\sigma^2)^{-\frac{n_0}{2}-1}
\end{aligned}$$

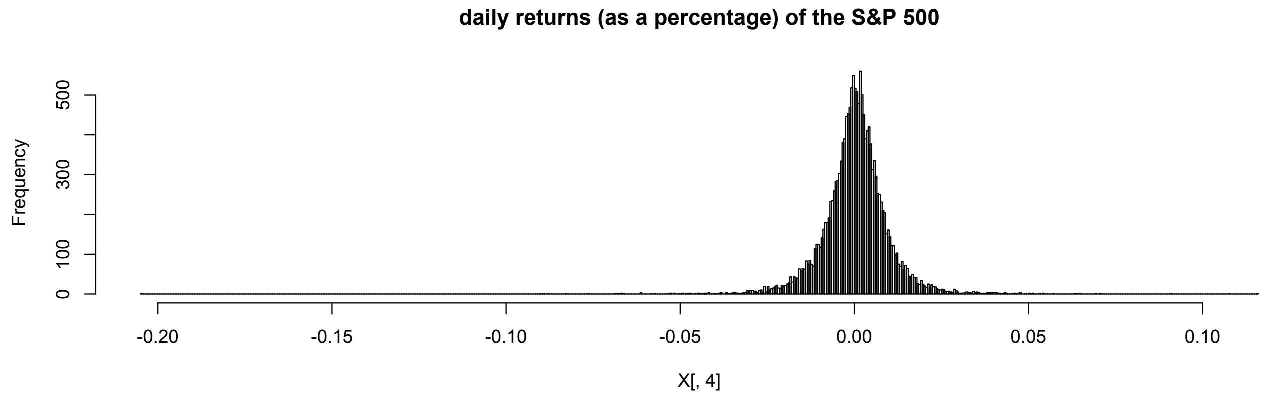
(p) [difficult] [MA] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\ \sigma^2)$ and given the general prior above, find the posterior and demonstrate it that the normal-inverse gamma is conjugate for the normal likelihood with both mean and variance unknown. This is what I did *not* do in class.

Problem 3

We model the returns of S&P 500 here.

(a) [easy] Below are the 16,428 daily returns (as a percentage) of the S&P 500 dating back to January 4, 1950 and the code used to generate it. Does the data look normal? Yes/no

Yes.



(b) [harder] Do you think the data is $\stackrel{iid}{\sim}$? Explain.

No because previous days can influence the days in the future.

(c) [harder] Assume $\stackrel{iid}{\sim}$ normal data regardless of what you wrote in (a) and (b). The sample average is $\bar{x} = 0.0003415$ and the sample standard deviation is $s = 0.0096$. Under an objective prior, give a 95% credible region for the true mean daily return.

$$\begin{aligned} CR_{\theta,95\%} &= [\texttt{qnorm}(0.025, 0.0003415, \frac{0.0096^2}{16428}), \texttt{qnorm}(0.975, 0.003415, \frac{0.0096^2}{16428})] \\ &= [0.00341489, 0.00341511] \end{aligned}$$

(d) [difficult] Give a 95% credible region for *tomorrow's* return using functions in Table 1.

$$\begin{aligned} CR_{95\%} &= [\texttt{qnorm}(0.025, 0.003415, (0.0096)^2), \texttt{qnorm}(0.975, 0.003415, (0.0096)^2)] \\ &= [0.0001608697, 0.0005221303] \end{aligned}$$