MATH 390.03-02 / 650 Fall 2015 Homework #6

Professor Adam Kapelner

Due in class, Monday, March 28, 2016

(this document last updated Wednesday 23$^{\text{rd}}$ March, 2016 at 1:53pm)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read about the negative binomial-beta, exponential-beta, normal-normal models, mixture models and kernels of PMFs / PDFs. Also read ch11 and 12 in McGrayne.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

Problems marked "[MA]" are for the masters students only (those enrolled in the 650 course). For those in 390, doing these questions will count as extra credit.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 10 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: _____

Problem 1

These are questions about McGrayne's book, chapters 11 and 12.
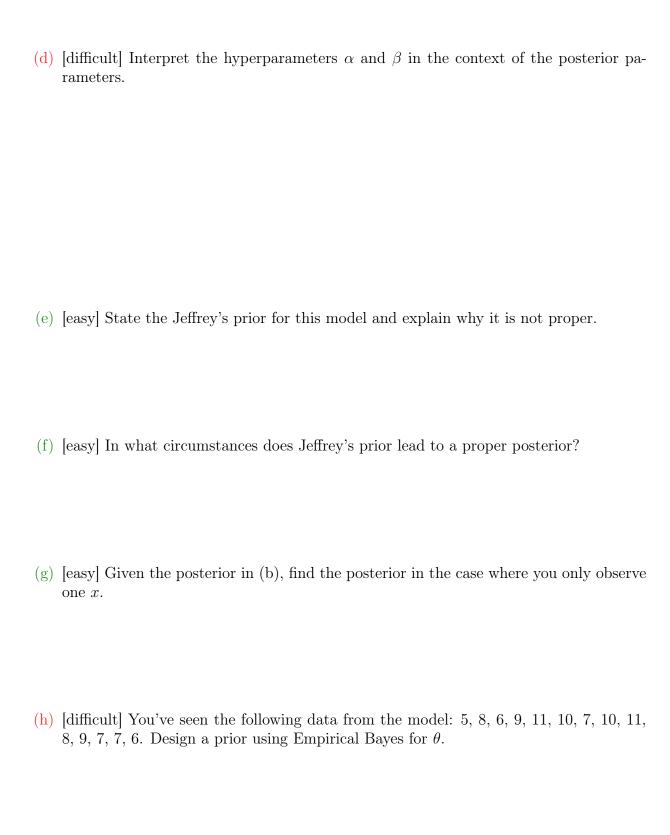
(a) [easy] Did Savage like Shlaifer? Yes / No and why?

(b) [easy] How did Neyman-Pearson approach statistical decision theory? What is the weakness to this approach? (p145)
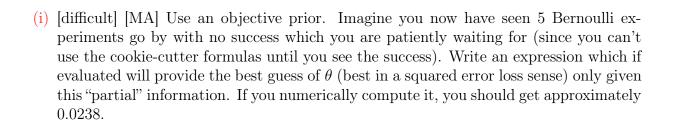
(c) [easy] Who popularized "probability trees" (and "tree flipping") similar to exercises we did in Math 241?

(d) [easy] Where are Bayesian methods taught more widely than any other discipline in academia?

(e) [easy] Despite the popularity of his Bayesian textbook on business decision theory, why didn't Schlaifer's Bayesianism catch on in the real world of business executives making decisions?

(f) [easy] Why did the pollsters fail (big time) to predict Harry Truman's victory in the 1948 presidential election?

(g) [easy] When does the diference between Bayesianism and Frequentism grow "immense"?

(h) [easy] How did Mosteller demonstrate that Madison wrote the 12 Federalist papers of unknown authorship?

Problem 2

We will ask some basic problems on Empirical Bayesianism.

(a) [easy] Explain the methodology of Empirical Bayes the best as you can.

(b) [difficult] In what situations do you think Empirical Bayes methods work the best?

Problem 3

We will again be looking at the beta-prior, negative-binomial-likelihood Bayesian model. But first consider the more basic case where $X_1, \ldots, X_n \overset{exch}{\sim} \text{Geometric}(\theta)$ and $\theta \sim \text{Beta}(\alpha, \beta)$.

(a) [easy] What is the likelihood model here? Write the PMF using the parameterization from class only.

(b) [harder] Demonstrate that the posterior in this case is Beta and find the posterior parameters.

(c) [easy] Give expressions for $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MAE}}$ and $\hat{\theta}_{\text{MAP}}$ (use the approximation for the median of the beta distribution given in the notes for $\hat{\theta}_{\text{MAE}}$). Are they all similar is n is large?

(d) [difficult] Interpret the hyperparameters α and β in the context of the posterior parameters.

(e) [easy] State the Jeffrey's prior for this model and explain why it is not proper.

(f) [easy] In what circumstances does Jeffrey's prior lead to a proper posterior?

(g) [easy] Given the posterior in (b), find the posterior in the case where you only observe one x.

(h) [difficult] You've seen the following data from the model: 5, 8, 6, 9, 11, 10, 7, 10, 11, 8, 9, 7, 7, 6. Design a prior using Empirical Bayes for θ.

(i) [difficult] [MA] Use an objective prior. Imagine you now have seen 5 Bernoulli experiments go by with no success which you are patiently waiting for (since you can't use the cookie-cutter formulas until you see the success). Write an expression which if evaluated will provide the best guess of θ (best in a squared error loss sense) only given this "partial" information. If you numerically compute it, you should get approximately 0.0238.

(j) [harder] Consider $X_1, \ldots, X_n \overset{exch}{\sim} \text{NegBin}(r, \theta)$ where r is considered known and $\theta \sim \text{Beta}(\alpha, \beta)$ Demonstrate that the posterior in this case is Beta and find the posterior parameters.

(k) [easy] Give expressions for $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MAE}}$ and $\hat{\theta}_{\text{MAP}}$ (use the approximation for the median of the beta distribution given in the notes for $\hat{\theta}_{\text{MAE}}$). Are they all similar is n is large?
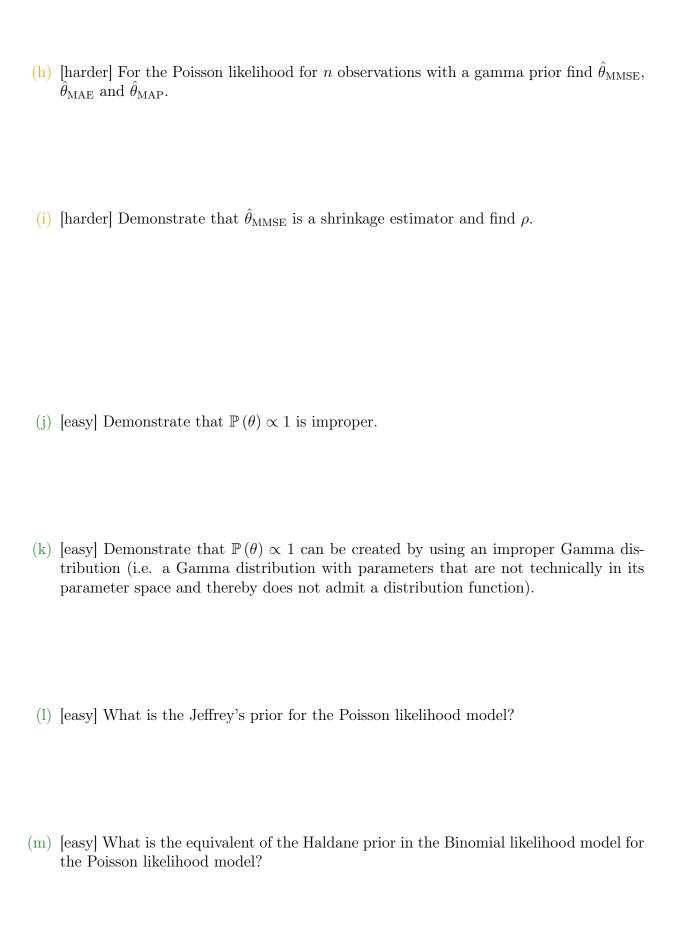
(l) [easy] [MA] Find the Jeffrey's prior for θ as a function of r. Look up the $I(\theta)$ on the Internet for the negative binomial given the parameterization we used in class. You do not need to do the derivation yourself.

(m) [difficult] [MA] Derive the posterior distribution PMF for one new negative binomial observation after seeing n observations. This is a lot of computation.

(n) [difficult] [MA] Write an integral expression for the joint posterior distribution for m new negative binomail observations. Extra credit if you can find the solution somewhere on the Internet.

(o) [harder] Derive the PMF of the BetaGeometric(α, β) distribution. Hint: the formula for the BetaNegativeBinomial(r, α, β) PMF was given in class. All you need to do is solve for the special case when $r = 1$. Leave in terms of the beta function.

(p) [harder] Find the kernel of the BetaGeometric distribution.

(q) [harder] Imagine $r = 3$ and you've seen the following data from the model: 5, 8, 6, 9, 11, 10, 7, 10, 11, 8, 9, 7, 7, 6. Find an expression for the probability the next observation will be 10.

(r) [easy] Evaluate the probability in (q).

Problem 4

We will ask some basic problems on the Gamma-Poisson conjugate model.

(a) [easy] If $X \sim \text{Poisson}(\theta)$, what is the kernel of $\theta \mid X$?

(b) [easy] Write the PDF of θ which is the gamma distribution with the standard hyper-parameters we used in class.

(c) [easy] What is the support and parameter space?

(d) [easy] What is the expectation and standard error and mode?

(e) [easy] Draw four different pictures of different hyperparameter combinations to demonstrate this model's flexibility

(f) [harder] Prove that the Poisson likelihood for $n = 1$ with a gamma prior yields a gamma posterior and find its parameters.

(g) [harder] Prove that the Poisson likelihood for n observations with a gamma prior yields a gamma posterior and find its parameters.

(h) [harder] For the Poisson likelihood for n observations with a gamma prior find $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MAE}}$ and $\hat{\theta}_{\text{MAP}}$.

(i) [harder] Demonstrate that $\hat{\theta}_{\text{MMSE}}$ is a shrinkage estimator and find ρ.

(j) [easy] Demonstrate that $\mathbb{P}(\theta) \propto 1$ is improper.

(k) [easy] Demonstrate that $\mathbb{P}(\theta) \propto 1$ can be created by using an improper Gamma distribution (i.e. a Gamma distribution with parameters that are not technically in its parameter space and thereby does not admit a distribution function).

(l) [easy] What is the Jeffrey's prior for the Poisson likelihood model?

(m) [easy] What is the equivalent of the Haldane prior in the Binomial likelihood model for the Poisson likelihood model?

(n) [harder] [MA] Prove that posterior predictive distribution for the next exchangeable Poisson realization given n observed Poisson realizations is negative binomially distributed and show its parameters are $p = \beta/(\beta + 1)$ and $r = \alpha$ for $\alpha \in \mathbb{N}$.
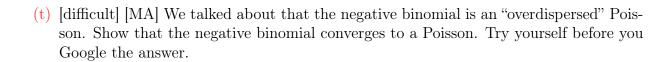
(o) [difficult] If $\alpha \notin \mathbb{N}$, create an "extended negative binomial" r.v. and find its PMF.

(p) [easy] Why is the extended negative binomial r.v. also known as the gamma-Poisson mixture distribution?

(q) [E.C.] [MA] Find the joint posterior predictive distribution for m future observations. I couldn't find the answer to this myself nor compute the integral.

(r) [harder] If you observe $0, 3, 2, 4, 2, 6, 1, 0, 5$, give a 95% CR for θ. Pick an objective prior.

(s) [harder] Using the data and the prior from (r), test if $\theta < 2$.

(t) [difficult] [MA] We talked about that the negative binomial is an "overdispersed" Poisson. Show that the negative binomial converges to a Poisson. Try yourself before you Google the answer.

Problem 5

We will ask some basic problems on the Gamma-Exponential conjugate model.

(a) [easy] If $X \sim \text{Exp}(\theta)$, what is the kernel of $\theta \mid X$?

(b) [harder] Prove that the Exponential likelihood for $n = 1$ with a gamma prior yields a gamma posterior and find its parameters.

(c) [harder] Prove that the exponential likelihood for n observations with a gamma prior yields a gamma posterior and find its parameters.

(d) [harder] For the exponential likelihood for n observations with a gamma prior find $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MAE}}$ and $\hat{\theta}_{\text{MAP}}$.

(e) [harder] Demonstrate that $\hat{\theta}_{\text{MMSE}}$ is a shrinkage estimator and find ρ.

(f) [harder] Use an uninformative prior like in the previous question. What is the posterior?

(g) [harder] Write the integral to solve for the posterior predictive distribution for a single observation given n observed data points.

(h) [difficult] [MA] Solve the integral.

(i) [E.C.] [MA] Find the joint posterior predictive distribution for m future observations.

(j) [difficult] [MA] What is the Jeffrey's prior for the exponential likelihood? Try yourself before you Google the answer.

We now begin the normal-normal conjugate model.

(a) [easy] If $X \sim \mathcal{N}(\theta, \sigma^2)$, what is the kernel of $\theta \mid X, \sigma^2$?

(b) [easy] If $X \sim \mathcal{N}(\theta, \sigma^2)$, what is the kernel of $\sigma^2 \mid X, \theta$?

(c) [easy] If $X \sim \mathcal{N}(\theta, \sigma^2)$, what is the kernel of $\theta, \sigma^2 \mid X$?

(d) [difficult] Show that posterior of $\theta \mid X, \sigma^2$ is normal if $\theta \sim \mathcal{N}(\mu_0, \tau^2)$. Try to do it yourself and only copy from the notes if you have to.

(e) [easy] What is the definition of the convolution of two r.v.'s X_1 and X_2?

(f) [difficult] Show that predictive distribution of $X^* \mid X$, σ^2 is normal if $\theta \sim \mathcal{N}\left(\mu_0,\ \tau^2\right)$ by solving the integral and not using the convolution.

(g) [difficult] Even though you solved this in (f), using the law of iterated expectation, find the expectation of the predictive distribution of $X^* \mid X$, σ^2.

(h) [difficult] Even though you solved this in (f), Using the law of total variance, find the variance of the predictive distribution of $X^* \mid X$, σ^2.

(i) [easy] In this problem we found the posterior, $\theta \mid X$, σ^2. What are all the other posteriors that could be of interest? Explain the inferential targets of each.