MATH 341 / 650 Spring 2017 Homework #6

Professor Adam Kapelner

Due 2PM under my office door (KY604), Wednesday, April 6, 2017

(this document last updated Wednesday 29$^{\text{th}}$ March, 2017 at 4:23pm)

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, review the Normal-Normal model for mean estimation given the variance and variance estimation given the mean. Also, review the concepts of posterior predictive distributions, empirical Bayes prior designs, uninformative prior design, credible regions and Bayesian Hypothesis Tests.

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 341, 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: ____

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	`qbeta`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
betabinomial	`qbetabinom`(p, n, α, β)	`d-`(x, n, α, β)	`p-`(x, n, α, β)	`r-`(n, α, β)
betanegativebinomial	`qbeta_nbinom`(p, r, α, β)	`d-`(x, r, α, β)	`p-`(x, r, α, β)	`r-`(r, α, β)
binomial	`qbinom`(p, n, θ)	`d-`(x, n, θ)	`p-`(x, n, θ)	`r-`(n, θ)
exponential	`qexp`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
gamma	`qgamma`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
geometric	`qgeom`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
inversegamma	`qinvgamma`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
negative-binomial	`qnbinom`(p, r, θ)	`d-`(x, r, θ)	`p-`(x, r, θ)	`r-`(r, θ)
normal (univariate)	`qnorm`(p, θ, σ)	`d-`(x, θ, σ)	`p-`(x, θ, σ)	`r-`(θ, σ)
normal (multivariate)		`dmvnorm`($\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)		`r-`($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)
poisson	`qpois`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
T (standard)	`qt`(p, ν)	`d-`(x, ν)	`p-`(x, ν)	`r-`(ν)
T (nonstandard)	`qt.scaled`(p, ν, μ, σ)	`d-`(x, ν, μ, σ)	`p-`(x, ν, μ, σ)	`r-`(ν, μ, σ)
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 1

This question is about building models for the prices of cars.

The 2016 Honda Accord sells at many different dealerships in New York City but sell it for more and some for less. We'll assume that the final negotiated price is distributed normally because it's most likely the sum of many different negotiation factors.

Our goal here is to determine the mean price at a certain car dealership in Astoria that people have been saying is "too cheap" and if it's too cheap, Honda corporate may wish to investigate.

(a) [easy] Assume that each Accord's price at the Astoria dealership is normal and $\overset{iid}{\sim}$ given the parameters. Is this a good model? Why or why not? There is no "correct" answer here but I expect you to defend whatever answer you write using the concepts we discussed in class.

(b) [easy] Despite what you wrote in (b), assume $\stackrel{iid}{\sim}$ for the rest of the problem. The nationwide variance for a Honda Accord selling price we're going to assume is $\sigma^2 = \$1000^2$, an assumption we will relax later. Given a sample with average $\bar{x}$ and sample size n, what is the distribution of the mean price of a car from this shady Astoria dealership? Assume an uninformative prior of your choice but ensure to explicitly state it.

(c) [easy] You and your colleague go down to the Astoria dealership undercover and ask to buy a Honda. After much negotiation, they will sell it to you for \$19,000 and they will sell it to your colleague for \$18,200 but they sense something suspicious so you hesitate to send another one of your guys down there to do another faux negotiation. Unfortunately, we're going to have to estimate the mean with just $x_1 = 19000$ and $x_2 = 18200$. What is your best guess of the mean price of Honda Accords sold here? Assume your prior from (a). Compute explicitly as a number rounded to two decimals.

(d) [easy] What is the shrinkage value (which we have been denoting ρ) for this estimate? Compute explicitly as a number rounded to two decimals.

(e) [easy] Based on this data, we wish to test if this dealership is selling Honda Accords below the manufacturer sugested retail price (MSRP) of \$22,205 — if so, they would be subject to a fine. Calculate a p-value for this test below by using notation from Table 1 but do not solve numerically.

(f) [easy] What is the probability I get a really good deal — that I can buy a car from these Astoria people for under \$17,000? Use the notation from Table 1 but do not solve numerically.

(g) [easy] We will continue to not rely on the nationwide average of $\sigma = \$1000$. Here, instead of an uninformative prior, we use the Empirical Bayes concept to construct an informative prior (not uninformative).

Below are the sample average selling prices (in USD) of Honda Accords from 16 other car dealerships also in the NYC area that serve as a comparison:

22889.80 21159.16 23796.71 19132.65 23450.63 24088.28 19852.37 21306.45
24434.05 23150.34 21690.09 20640.79 21973.45 21984.48 22326.00 22239.98

Using this data *estimate* a conjugate prior for σ^2. Use the n_0 and μ_0 parameterization. You will still need σ^2 from above!

(h) [easy] Given the data in (d) which is $x_1 = 19000$ and $x_2 = 18200$, what is your best guess of the mean price of Honda Accords sold here? Assume the empirical Bayes conjugate prior. Round to the nearest cent.

Problem 2

We now continue questions on the normal-normal conjugate model.

(a) [difficult] [MA] Show that predictive distribution of $X^* \mid X,\ \sigma^2$ is normal if $\theta \sim \mathcal{N}(\mu_0,\ \sigma^2/n_0)$ by solving the integral and not using the convolution.

(b) [easy] If $X_1, \ldots, X_n \mid \theta,\ \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\ \sigma^2)$ and X represents all $X_1, \ldots, X_n$, in HW6 6(b) you found the kernel of $\sigma^2 \mid X,\ \theta$. Show that this is the kernel of an inverse gamma. Use the $\hat{\sigma}^2$ substituion we did in class.

(c) [harder] Why is using $\hat{\sigma}^2$ permitted in the setup in (a) but doesn't make sense in the ususal frequentist setup when the likelihood is normal? Hint: what is your target of estimation usually?

(d) [easy] In class we looked at $\sigma^2 \sim \text{InvGamma}(\alpha,\ \beta)$ but we used a different parameterization. Write the different parameterization below and explain why this was done i.e. interpret the meaning of the two new parameters.

(e) [harder] Show that $\sigma^2 \mid X,\ \theta$ is distributed as an inverse gamma with the prior from (d) and find its parameters.

(f) [easy] What is the Jeffrey's prior for σ^2 (look in the notes and write it down — no need to prove it). Is it proper?

(g) [easy] Show that the Jeffrey's prior for σ^2 is an improper inverse gamma distribution and find its parameters. Note these parameters are not in the parameter space of a proper inverse gamma distribution.

(h) [easy] Under the Jeffrey's prior for σ^2, what is the posterior?

(i) [harder] You are in a milk manufacturing plant producing 1 quart cartons of whole milk. You are willing to assume that the nozzle emits 1 qt on average. In your previous job, you remember inspecting 3 cartons of which you saw 1.02, 0.97, 1.03 quarts of milk inside. Create a prior based on what you've seen in your previous job. This forces you to understand (d).

(j) [difficult] The company wishes to test if there's too much variability i.e. that there is more than $\sigma = 0.1$ variability. You take a sample of 10 and see 1.153, 1.045, 1.268, 1.333, 0.799, 1.075, 1.27, 1.07, 1.192 and 1.079 quarts. Find the p value. You can write the answer below as a function of `rinvgamma`, `qinvgamma` or `pinvgamma` (i.e., expressions from Table 1). E.C. for computing it and testing this at $\alpha = 5\%$. You may want to use the `actuar` package (see here).

(k) [harder] Find $CR_{\sigma^2, 90\%}$ for the data above using expressions from Table 1.