MATH 341 / 650.3 Spring 2018 Homework #1

Professor Adam Kapelner

Due *in class*, Friday 5PM, February 16, 2018

(this document last updated Thursday 8th February, 2018 at 12:18am)

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 Math 241 concerning random variables, support, parameter space, PMF's, PDF's, CDF's, Bayes Rule, read about parametric families and maximum likelihood estimators on the Internet, read the preface and ch 1 and 4 of Bolstad and read the preface and Ch1 of 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 650.3 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: ______________________________

Problem 1

These are questions about McGrayne's book, preface, chapter 1, 2 and 3.

(a) [easy] Explain Hume's problem of induction with the sun rising every day.

(b) [easy] Explain the "inverse probability problem."

(c) [easy] What is Bayes' billiard table problem?

(d) [difficult] [MA] How did Price use Bayes' idea to prove the existence of the deity?

(e) [easy] Why should Bayes Rule really be called "Laplace's Rule?"

(f) [difficult] Prove the version of Bayes Rule found on page 20. State your assumption(s) explicitly. Reference class notes as well.

(g) [easy] Give two scientific contexts where Laplace used inverse probability theory to solve major problems.

(h) [difficult] [MA] Why did Laplace turn into a frequentist later in life?

(i) [easy] State Laplace's version of Bayes Rule (p31).

(j) [easy] Why was Bayes Rule "damned" (pp36-37)?

(k) [easy] According to Edward Molina, what is the prior (p41)?

(l) [easy] What is the source of the "credibility" metric that insurance companies used in the 1920's?

(m) [easy] Can the principle of inverse probability work without priors? Yes/no.

(n) [difficult] In class we discussed the "principle of indifference" which is a term I borrowed from Donald Gillies' Philosophical Theories of Probability. On Wikipedia, it says that Jacob Bernoulli called it the "principle of insufficient reason". McGrayne in her research of original sources comes up with many names throughout history this principle was named. List all of them you can find here.

(o) [easy] Jeffreys seems to be the founding father of modern Bayesian Statistics. But why did the world turn frequentist in the 1920's? (p57)

Problem 2

These exercises will review the Bernoulli model.

(a) [easy] If $X \sim \text{Bernoulli}\,(\theta)$, find $\mathbb{E}\,[X]$, $\mathbb{V}\text{ar}\,[X]$, $\text{Supp}\,[X]$ and Θ. No need to derive from first principles, just find the formulas.

(b) [harder] If $X \sim \text{Bernoulli}\,(\theta)$, find $\text{median}\,[X]$.

(c) [harder] If $X \sim \text{Bernoulli}\,(\theta)$, write the "parametric statistical model" below using the notation we used in class only.

(d) [harder] Explain what the semicolon notation in the previous answer indicates. Hint: go back to precalc and think of the function $g(x; a) = ax^2$

(e) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(\theta)$, find the likelihood, $\mathcal{L}$, of θ.

(f) [difficult] Given the likelihood above, what would $\mathcal{L}$ be if the data was $< 0, 1, 0, 1, 3.7 >$? Why should this answer have to be?

(g) [easy] If $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, find the log-likelihood of θ, $\ell(\theta)$.

(h) [difficult] [MA] If $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$, explain why the log-likelihood of θ is normally distributed if n gets large.

(i) [easy] If $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, find the score function (i.e the derivative of the log-likelihood) of θ.

(j) [harder] If $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, find the maximum likelihood estimator for θ.

(k) [easy] If $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, find the maximum likelihood *estimate* for θ.

(l) [easy] Given the previous two questions, describe the difference between a random variable and a datum.

(m) [easy] If your data is $< 0, 1, 1, 0, 1, 1, 0, 1, 1, 1 >$, find the maximum likelihood estimate for θ.

(n) [easy] Given this data, find a 99% confidence interval for θ.

(o) [easy] Given this data, test $H_0 : \theta = 0.5$ versus $H_a : \theta \neq 0.5$.

(p) [easy] Write the PDF of $X \sim \mathcal{N}(\theta,\, 1^2)$.

(q) [difficult] Find the MLE for θ if $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta,\, 1^2)$.

(r) [difficult] [MA] Find the MLE for θ if $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu,\, \sigma^2)$. Solve the system of equations $\frac{\partial}{\partial \mu}\left[\ell(\theta)\right] = 0$ and $\frac{\partial}{\partial \sigma^2}\left[\ell(\theta)\right] = 0$ where $\ell(\theta)$ denotes the log likelihood. You can easily find this online. But try to do it yourself.

Problem 3

We will review the frequentist perspective here.

(a) [difficult] Why do frequentists have an insistence on θ being a fixed, immutable quantity? We didn't cover this in class explicitly but it is lurking behind the scenes. Use your reference resources.

(b) [easy] What are the three goals of inference? Give short explanations.

(c) [easy] What are the three reasons why *frequentists* (adherents to the frequentist perspective) use MLEs i.e. list three properties of MLEs that make them powerful.

(d) [difficult] [MA] Give the conditions for asymptotic normality of the MLE,

$$\frac{\hat{\theta}_{\text{MLE}} - \theta}{\mathbb{SE}\left[\hat{\theta}_{\text{MLE}}\right]} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\, 1\right).$$

You can find them online.

(e) [difficult] [MA] $\mathbb{SE}\left[\hat{\theta}_{\text{MLE}}\right]$ cannot be found without θ so we substituted $\hat{\theta}_{\text{MLE}}$ into $\mathbb{SE}\left[\hat{\theta}_{\text{MLE}}\right]$ and called it $\hat{\mathbb{SE}}\left[\hat{\theta}_{\text{MLE}}\right]$ (note the hat over the SE). Show that this too is asymptotically normal, *i.e.*

$$\frac{\hat{\theta}_{\text{MLE}} - \theta}{\hat{\mathbb{SE}}\left[\hat{\theta}_{\text{MLE}}\right]} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\, 1\right)$$

You need the continuous mapping theorem and Slutsky's theorem.

(f) [easy] [MA] Explain why the previous question allows us to build asymptotically valid confidence intervals using $\left[\hat{\theta}_{\text{MLE}} \pm z_{\alpha/2}\hat{\mathbb{S}\text{E}}\left[\hat{\theta}_{\text{MLE}}\right]\right]$.

(g) [harder] Why does all of frequentist inference break down if n isn't large?

(h) [easy] Write the most popular two frequentist interpretations of a confidence interval.

(i) [harder] Why are each of these unsatisfactory?

(j) [easy] What are the two possible outcomes of a hypothesis test?

(k) [difficult] [MA] What is the weakness of the interpretation of the p-val?

Problem 4

We review and build upon conditional probability here.

(a) [easy] Explain why $\mathbb{P}(B \mid A) \propto \mathbb{P}(A \mid B)$.

(b) [easy] If B represents the hypothesis or the putative cause and A represents evidence or data, explain what Bayesian Conditionalism is, going from which probability statement to which probability statement.