MATH 341 / 650.3 Spring 2020 Homework #4

Frank Palma Gomez

Thursday 19th March, 2020

Problem 1

These are questions about McGrayne's book, chapters 8-10.

(a) [easy] When was experimentation introduced to medical science and who introduced it? Are you surprised that it was this recent?

Introduced in the 1950's by Jerome Cornfield

(b) [easy] Sir Ronald A. Fisher, the founder of modern experiments, did not believe cigarettes caused lung cancer. What were his two hypotheses for the cause of lung cancer?

Smoking causes lung cancer and that genetic factors can give people hereditary susceptibility for smoking and lung cancer

(c) [easy] Who invented, and what are Bayes Factors? (p116)

Bayes's Facotrs was inverted by Jerome Cornfield. Bayes factors is the probability of the observed data under a hypotheses divided by its probability under another hypotheses

(d) [easy] Trick question: who convinced Cornfield to stop smoking?

Cornfields own research convinced him that there were 14 studies showing high correletaion between lung cancer and smoking

(e) [easy] Why were frequentists at a loss to estimate the probability of a nuclear bomb being detonated by accident?

Because no atomic or hydrogen weapon had ever exploded by accident.

(f) [easy] What is Cromwell's Rule? And, when applying this principle to a Bayesian model what would it imply? (See the Wikipedia link and p123).

Cromwells rule is the ommision of prior probabilities with 0 or 1, except when applied to statements that are logically true or false. You should not use improper priors in Bayesian models.

(g) [easy] Did Bayesian Statistics prevent nuclear accidents? Discuss.

Bayesian statistics preventer nuclear accidents by calculating the number of accident opportunities based on the number of weapons, the number of longevity, the number of times they were aboard planes or handled in storage. With that, they were able to estimate when a nuclear accident was likely to occur.

(h) [easy] What is the main reason why there are so many variations of Bayesian interpretation? (p129)

The main reason why there was so many variations of Bayesian interpretation was because it subjective nature.

(i) [easy] What is a large practical drawback of Bayesian inference? (See mid-end of chapter 8).

Bayesian inference conclusion are largely based on frequintism.

Problem 2

Assume $\mathcal{F}$ = binomial with n fixed and $\mathbb{P}(\theta) = \text{Beta}(2.5, 2.5)$, $n = 100$ and $x = 39$, the prior and data from HW#3.

(a) [difficult] Find the posterior predictive distribution, $X_* \mid X$ where X_* denotes the random variable that counts the number of successes in n_* future trials. MA students — do this yourself. Other students — use my notes and justify each step. Remember, if $W \sim \text{Bernoulli}(p)$ then $\mathbb{P}(W = 1) = p$. Use that trick. Simplify.

$$\begin{aligned}\mathbb{P}(X_* \mid X) &= \int_0^1 \binom{n_*}{x} \theta^x (1-\theta)^{(n_*-x)} \frac{1}{B(\alpha+x, \beta+n-x)} \theta^{(\alpha+x)-1}(1-\theta)^{(\beta+n-x)-1}\, d\theta \\ &= \text{BetaBinomial}(n_*,\, \alpha+x,\, \beta+n_*-x)\end{aligned}$$

(b) [easy] If $n_* = 17$, what is the expectation and variance of $X_* \mid X$?

$$\mathbb{E}[X] = \frac{n\alpha}{\alpha+\beta} = \frac{17 * 2.5}{2.5+2.5} = 8.5$$

$$\mathbb{V}\text{ar}[X] = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{17 * 2.5 * 2.5 * (2.5+2.5+17)}{(2.5+2.5)^2(2.5+2.5+1)} = 15.58$$

(c) [harder] Plot the PMF of $X_* \mid X$ as best as you can. Mark critical points and label the axes.

(d) [easy] What is the probability of $x_* = 10$ given your data and prior? Write your answer in terms of the function(s) found in the Midterm's table 1.

$$pbetabinomial(10,\ 17,\ 2.5,\ 2.5)$$

(e) [harder] Answer the previous problem exactly and then round to two decimal places.

$$pbetabinomial(10,\ 17,\ 2.5,\ 2.5) = 0.67$$

Problem 3

Some quick question on mixture / compound distributions.

(a) [easy] Let X be $\mathcal{N}(0,\ 1^2)$ 1/3 of the time and $\text{Exp}(3)$ 2/3 of the time. What is its pdf?

$$\mathbb{P}(X) = \frac{1}{\sqrt{2\pi(1)^2}}\text{exp}\left(-\frac{1}{2(1)^2}(x-0)^2\right)(\frac{1}{3}) + 3\text{exp}(-3x)(\frac{2}{3})$$

(b) [difficult] Let's say $X \mid \beta \sim \text{Beta}(1,\ \beta)$ where $\beta \mid \lambda \sim \text{Exp}(\lambda)$. Write an integral expression which when solved, finds the compound / marginal density of X. DO NOT solve.

$$\int \frac{1}{B(\alpha,\beta)}(1-x)^{\beta-1}e^{\lambda\beta}d\beta$$

(c) [difficult] [MA] Let's say $X \mid \theta,\ \sigma^2 \sim \mathcal{N}(\theta,\ \sigma^2)$ where $\theta \mid \mu_0,\ \tau^2 \sim \mathcal{N}(\mu_0,\ \tau^2)$. Write an integral expression which when solved, finds the compound / marginal density of X. DO NOT solve.

Problem 4

These are questions about other vague priors: improper priors and Jeffreys priors.

(a) [easy] What is an improper prior?

A prior that is infenistesimal over an infinite range.

(b) [harder] Is $\theta \sim \text{Beta}(100, 0)$ improper? Yes / no and provide a proof.

(c) [easy] When are improper priors "legal"?

If $\sum_i^n x_i \neq 0$ and $\sum_i^n x_i \neq n$ or when it does not violate that parameter smapce for that particular prior.

(d) [easy] When are improper priors "illegal"?

If $\sum_i^n x_i = 0$ and $\sum_i^n x_i = n$ or the parameters violate the parameter space.

(e) [difficult] What does $I(\theta)$ tell you about the random variable with respect to its parameter θ?

The amount of information that X, the random variable, contains.

(f) [harder] If I compute a posterior on the θ scale and then measure the parameter on another scale, will I (generally) get the same posterior probability? Yes/no explain.

If using Jeffreys prior, yes since any parametarization provides the same prior.

(g) [easy] What is the Jeffrey's prior for θ under the binomial likelihood? Your answer must be a distribution.

$$\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

(h) [difficult] What is the Jeffrey's prior for $\theta = t^{-1}(r) = \frac{e^r}{1+e^r}$ (i.e. the log-odds reparameterization) under the binomial likelihood?

$$BetaPrime(\frac{1}{2}, \frac{1}{2})$$

(i) [difficult] Explain the advantage of Jeffrey's prior.

Jeffreys prior allows you to arrive to the same posterios with any parametarization by doing a change of variable. This is also known as the invariance principle. Jeffreys prior also reflects little knowledge about the data.

(j) [difficult] [MA] Prove Jeffrey's invariance principle i.e. prove that the Jeffrey's prior makes your prior probability immune to transformations. Use the second proof from class.

Recall:

$$\phi = \frac{\theta}{1-\theta} = \frac{\phi}{\phi+1} = \theta$$

$$X \sim \text{Binomial}\,(n,\, \theta) := \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

The log-likelihood functions:

$$\ell'(\theta; x) = \frac{x}{\phi} - \frac{n}{\phi+1}$$

$$\ell'' = \frac{x}{\phi^2} - \frac{n}{(\phi+1)^2}$$

$$\boldsymbol{I}(\phi) = \mathbb{E}\left[-\ell''\right] = n(\frac{1}{\phi(\phi+1)^2})$$

$$\mathbb{P}\left(\sqrt{I(\theta)}\right) = BetaPrime(\frac{1}{2}, \frac{1}{2})$$

Problem 5

This question is about estimation of "true career batting averages" in baseball.
Every hitter's *sample* batting average (BA) is defined as:

$$BA := \frac{\text{sample \# of hits}}{\text{sample \# of at bats}}$$

In this problem we care about estimating a hitter's *true* career batting average which we call θ. Each player has a different θ but we focus in this problem on one specific player. In order to estimate the player's true batting average, we make use of the sample batting average as defined above (with Bayesian modifications, of course).

We assume that each at bat (for any player) are conditionally $\overset{iid}{\sim}$ based on the players' true batting average, θ. So if a player has n at bats, then each successful hit in each at bat can be modeled via $X_1 \mid \theta,\ X_2 \mid \theta, \ldots,\ X_n \mid \theta \overset{iid}{\sim} \text{Bernoulli}\,(\theta)$ i.e. the standard assumption and thus the total number of hits out of n at bats is binomial.

Looking at the entire dataset for 6,061 batters who had 100 or more at bats, I fit the beta distribution PDF to the sample batting averages using the maximum likelihood approach and I'm estimating $\alpha = 42.3$ and $\beta = 127.7$. Consider building a prior from this estimate as $\theta \sim \text{Beta}\,(42.3,\, 127.7)$

(a) [easy] Is the prior "conjugate"? Yes / No.

Yes

(b) [easy] Is this prior "indifferent"? Yes / No.

No

(c) [easy] Is this prior "objective"? Yes / No.

No

(d) [easy] Is this prior "informative"? Yes / No.

Yes

(e) [easy] Using prior data to build the prior is called...

informative Prior. Imperical Bayes.

(f) [easy] This prior has the information contained in how many observations?

$\alpha + \beta = 170$

(g) [easy] We now observe four at bats for a new player and there were no hits. Find the $\hat{\theta}_{\text{MMSE}}$.

$$\hat{\theta}_{\text{MMSE}} = \mathbb{E}[\theta|X] = \frac{\alpha + x}{\alpha + \beta + n} = \frac{42.3}{42.3 + 127.7 + 4} = \frac{42.3}{174} = 0.243$$

(h) [easy] Why was your answer so far away from $\hat{\theta}_{\text{MLE}} = 0$? What is the shrinkage proportion in this estimation?

$$\rho = \frac{\alpha+\beta}{\alpha+\beta+n} = \frac{42.3. + 127.7}{42.3+127.7+4} = \frac{170}{174} = 0.977$$

(i) [harder] Why is it a good idea to shrink so hard here? Why do some consider this to be one of the beauties of Bayesian modeling?

Although $\sum x$ is 0, since you know the batting average then you can shrink your value towards the actual average despite having not a lot of data.

(j) [difficult] Write an exact expression for the batter getting 14 or more hits on the next 20 at bats. You can leave your answer in terms of the beta function. Do not compute explicitly.

$$X^*|X \sim \text{BetaBinomial}\left(n,\, \theta+x,\, \beta+n+\alpha\right) = \text{BetaBinomial}\left(20,\, 42.3,\, 147.7\right)$$

$$\mathbb{P}\left(X^*|X \geq 14\right) = \sum_{X^*=14}^{20} \binom{20}{X^*} \frac{B(42.3, 147.7)}{B(42.3, 127.7)}$$

(k) [harder] How many hits is the batter expected to get in the next 20 at bats?

Problem 6

We will now have lots of examples finding kernels from common distributions. Some of these questions are silly, but they will force you to think hard about what the kernel is under different situations. And... they're fun! Probabilistic pyromania!

(a) [easy] What is the kernel of $X \mid \theta,\ n \sim \text{Binomial}\left(n,\, \theta\right)$?

$$\mathbb{P}\left(X|\theta\right) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \propto \frac{1}{x!(n-x)!}\theta^x(1-\theta)^{n-x}$$

(b) [difficult] What is the kernel of $X, n \mid \theta \sim \text{Binomial}\left(n,\, \theta\right)$? Be careful...

$$\mathbb{P}\left(\theta|X\right) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^n(1-\theta)^{-x}$$

(c) [easy] What is the kernel of $X \mid \alpha,\ \beta \sim \text{BetaBinomial}\left(n,\ \alpha,\ \beta\right)$?

$$\mathbb{P}\left(X|\alpha,\beta\right) = \binom{n}{x}\frac{\text{Beta}\left(x+\alpha,\ n-x+\beta\right)}{\text{Beta}\left(\alpha,\ \beta\right)} \propto \text{Beta}\left(x+\alpha,\ n-x+\beta\right)$$

(d) [easy] What is the kernel of $X \mid \theta \sim \text{Poisson}\left(\theta\right) := \frac{e^{-\theta}\theta^{x}}{x!}$?

$$\mathbb{P}\left(X|\theta\right) = \frac{e^{-\theta}\theta^{x}}{x!} \propto \frac{\theta^{x}}{x!}$$

(e) [difficult] What is the kernel of $\theta \mid X \sim \text{Poisson}\left(\theta\right)$? Be careful...

$$\mathbb{P}\left(\theta|X\right) = \frac{\exp\left(-\theta\right)\theta^{x}}{x!} \propto \exp\left(-\theta\right)\theta^{x}$$

(f) [easy] What is the kernel of $X \mid \alpha,\ \beta \sim \text{Beta}\left(\alpha,\ \beta\right)$?

$$\begin{aligned}\mathbb{P}\left(X|\alpha,\beta\right) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} &\propto x^{\alpha-1}(1-x)^{\beta-1}\\ &\propto x^{\alpha-1}(1-x)^{\beta}(1-x)^{-1}\end{aligned}$$

(g) [easy] What is the kernel of $X \mid \theta \sim \text{Exp}\left(\theta\right) := \theta e^{-\theta x}$? This is the exponential distribution.

$$\mathbb{P}\left(X|\theta\right) = \theta e^{-\theta x} \propto e^{-\theta x}$$

(h) [easy] What is the kernel of $X \mid \theta,\ \sigma^2 \sim \mathcal{N}\left(\theta,\ \sigma^2\right)$?

$$\mathbb{P}\left(X|\theta,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right) \propto \exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$$

(i) [difficult] What is the kernel of $\theta,\ \sigma^2 \mid X \sim \mathcal{N}\left(\theta,\ \sigma^2\right)$? Be careful...

$$\mathbb{P}\left(X|\theta,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right)$$

(j) [easy] What is the kernel of $X \mid k \sim \chi^2_k := \dfrac{1}{2^{k/2}\Gamma(k/2)}\, x^{k/2-1}e^{-x/2}$, the chi-squared distribution with k degrees of freedom?

$$\mathbb{P}\left(X|k\right) = \frac{1}{2^{k/2}\Gamma(k/2)}\, x^{k/2-1}e^{-x/2} \;\propto x^{k/2-1}\text{exp}\left(\frac{-x}{2}\right)$$

(k) [harder] What is the kernel of

$$X \mid N,\ \theta,\ n \sim \text{Hypergeometric}\left(N,\, \theta,\, n\right) := \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}$$

where N is the number of total balls in the bag, θ is the number of success balls in the bag and n is the number drawn out of the bag?

$$\begin{aligned}\mathbb{P}\left(X|N,\theta,n\right) = \frac{\binom{\theta}{x}\binom{n-\theta}{n-x}}{\binom{N}{n}} &= \frac{\frac{\theta!}{x(\theta-x)}\frac{(N-\theta)!}{(n-x)!(N-\theta-n+x)!}}{\frac{N!}{n!(N-n)!}} \\ &= \frac{\theta!n!(n-\theta)!(N-n)!}{x!(\theta-x)!(n-x)!(N-\theta-n+x)!N!} \\ &\propto \frac{1}{x!(\theta-x)!(n-x)!(N-\theta-n+x)!}\end{aligned}$$

(l) [harder] [MA] If $X \sim F_{k_1,k_2}$, Snecedor's F-distribution, what is its kernel?

(m) [difficult] [MA] If $X \mid \theta,\ \sigma^2 \sim \mathcal{N}\left(\theta,\, \sigma^2\right)$ and $\theta \mid \mu_0,\ \tau^2 \sim \mathcal{N}\left(\mu_0,\, \tau^2\right)$, what is the kernel of $\theta \mid X,\ \sigma^2,\ \mu_0,\ \tau^2$?