MATH 341 / 650.3 Spring 2020 Homework #5

Frank Palma Gomez

Sunday 29th March, 2020

Problem 1

These are questions about McGrayne's book, chapters 11–14.

(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?

(i) [easy] Write a one paragraph biography of John Tukey.

(j) [easy] Why did Alfred Kinsey's wife want to poison John Tukey?

(k) [easy] Tukey helped NBC with polling predictions for the presidential campaign. What was NBC's polling algorithm based on?

(l) [easy] Why is "objectivity an heirloom ... and ... a fallacy?"

(m) [easy] Why do you think Tukey called Bayes Rule by the name "borrowing strength?"

(n) [easy] Why is it that we don't know a lot of Bayes Rule's modern history?

(o) [easy] Generally speaking, how does Nate Silver predict elections?

(p) [easy] How many Bayesians of import were there in 1979?

(q) [easy] What advice did Chernoff give to Susan Holmes? (Note: Susan Holmes was my undergraduate advisor).

(r) [easy] How did Rasmussen's team estimate the probability of a nuclear plant core meltdown?

(s) [easy] How did the Three Mile Island accident vindicate Rasmussen's committee report?

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)
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)
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)
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 2

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

(a) [easy] Write the PDF of $\theta \sim \text{Gamma}(\alpha, \beta)$ which is the gamma distribution with the standard parameterization and notated with the hyperparameters we used in class.

$$\mathbb{P}(\theta) = \text{Gamma}(\alpha,\ \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(\beta\theta)$$

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

$$\text{Supp}[X] = (0, \infty), \alpha, \beta > 0$$

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

- $\mathbb{E}[Y] = \frac{\alpha}{\beta}$
- $\mathbb{V}\text{ar}[Y] = \frac{\alpha}{(\beta)^2}$
- $\text{Mode}[Y] = \frac{\alpha-1}{\beta}$ if $\alpha > 1$

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

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

$$\begin{aligned}
\mathbb{P}(\theta|X) &= \frac{\mathbb{P}(X|\theta)\,\mathbb{P}(\theta)}{\mathbb{P}(X)} \\
&\propto \mathbb{P}(X|\theta)\,\mathbb{P}(\theta) \\
&= \frac{\exp(-n\theta)\,\theta^{x}}{x!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(\beta\theta) \\
&= \exp(-n\theta)\,\theta^{x+\alpha-1}\exp(-\beta\alpha) \\
&= \theta^{x+\alpha-1}\exp(-(n+\beta)\theta) \\
&= \text{Gamma}(x+\alpha,\ n+\beta) \\
&= \text{Gamma}(x+\alpha,\ 1+\beta)
\end{aligned}$$

(f) [harder] Prove that the Poisson likelihood for n observations, i.e. $X_1, \ldots, X_n; \theta \overset{iid}{\sim} \text{Poisson}(\theta)$, with a gamma prior yields a gamma posterior and find its parameters.

$$\mathbb{P}(\theta|X) = \frac{\mathbb{P}(X|\theta)\,\mathbb{P}(\theta)}{\mathbb{P}(X)}$$

$$
\begin{aligned}
&\propto \mathbb{P}(X|\theta)\,\mathbb{P}(\theta) \\
&= \frac{\exp(-n\theta)\,\theta^{\sum x_i}}{\prod x_i!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(\beta\theta) \\
&= \exp(-n\theta)\,\theta^{\sum x_i+\alpha-1}\exp(-\beta\alpha) \\
&= \theta^{\sum x_i+\alpha-1}\exp(-(n+\beta)\theta) \\
&= \text{Gamma}\left(\sum x_i+\alpha,\, n+\beta\right)
\end{aligned}
$$

(g) [easy] Now that you see the posterior, provide a pseudodata interpretation for both hyperparameters.

$$\theta|X \sim \text{Gamma}\left(\sum x_i+\alpha,\, n+\beta\right)$$

- $\sum x_i$ is the number of total successes
- α is the number of psuedo successes
- n is the number of trials
- β is the number of pseudo trials

(h) [harder] Find the Bayesian point estimates as function of the data and prior's hyper-parameters (i.e. $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MMAE}}$ and $\hat{\theta}_{\text{MAP}}$).

- $\hat{\theta}_{\text{MMSE}} = \mathbb{E}[\theta|X] = \frac{\sum x_i+\alpha}{n+\beta}$
- $\hat{\theta}_{\text{MMAE}} = \text{median}[\theta|X] = qgamma(0.5, \sum x_i+\alpha, n+\beta)$
- $\hat{\theta}_{\text{MAP}} = \text{Mode}[\theta|X] = \frac{\sum x_i+\alpha-1}{n+\beta}$ if $\sum x_i+\alpha \geq 1$

(i) [harder] If $X_1,\ldots,X_n;\theta \overset{iid}{\sim} \text{Poisson}(\theta)$, find $\hat{\theta}_{\text{MLE}}$.

$$
\begin{aligned}
\mathcal{L}(\theta;X) &= \frac{\exp(-n\theta)\,\theta^{\sum x_i}}{\prod x_i!} \\
\ell'(\theta;X) &= -n+\frac{\sum x_i}{\theta}=0 \\
\hat{\theta}_{\text{MLE}} &= \frac{1}{n}\sum x_i = \bar{x}
\end{aligned}
$$

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

$$\begin{aligned}
\hat{\theta}_{\text{MMSE}} &= \frac{\sum x_i}{n+\beta}\frac{n}{n} + \frac{\alpha}{n+\beta}\frac{\beta}{\beta} \\
&= \frac{n}{n+\beta}\bar{x} + \frac{\beta}{n+\beta}\mathbb{E}\left[\theta\right] \\
&= (1-\rho)\bar{x} + \rho\mathbb{E}\left[\theta\right] \\
\rho &= \frac{\bar{x}}{\bar{x}+\mathbb{E}\left[\theta\right]}
\end{aligned}$$

(k) [harder] Demonstrate that $\mathbb{P}\left(\theta\right) \propto 1$ is improper.

If $\theta \sim U(0,\infty)$ then, $\mathbb{P}\left(\theta\right) \propto 1$ but $\mathbb{P}\left(\theta\right) = \frac{1}{\infty}$ therefore is not a valid pdf.

(l) [easy] [MA] Demonstrate that $\mathbb{P}\left(\theta\right) \propto 1$ 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).

(m) [harder] Find Jeffrey's prior for the Poisson likelihood model. Try to do it yourself.

$$\begin{aligned}
\ell'(\theta; X) &= -n + \frac{\sum x_i}{\theta} \\
\ell''(\theta; X) &= \frac{\sum x_i}{\theta} \\
\boldsymbol{I}(\theta) &= \mathbb{E}\left[\ell''(\theta; X)\right] \\
&= \frac{1}{(\theta)^2}\sum \mathbb{E}\left[x_i\right] \\
&= \frac{n}{\theta}
\end{aligned}$$

(n) [easy] What is the equivalent of the Haldane prior in the Binomial likelihood model for the Poisson likelihood model? Use an interpretation of pseudocounts to explain.

$\mathbb{P}\left(\theta|X\right) = \text{Gamma}\left(\sum x_i + \alpha,\, n+\beta\right)$ where α and β are 0. Therefore, $\text{Gamma}\left(\sum x_i,\, n\right)$

(o) [harder] Prove that posterior predictive distribution for the next Poisson realization (i.e. $n^* = 1$) given n observed Poisson realizations is negative binomially distributed and show its parameters are $p = \beta/(\beta+1)$ and $r = \alpha$ for $\alpha \in \mathbb{N}$.

$$\mathbb{P}\left(X_*|X\right) = \int \mathbb{P}\left(X_*|\theta\right)\mathbb{P}\left(\theta|X\right)\, d\theta$$

$$
\begin{aligned}
&= p^r(1-p)^{x_*}\frac{\Gamma\left(\sum x_i + x_* + \alpha\right)}{x_*!\Gamma\left(\sum x_i + \alpha\right)} \\
&= p^r(1-p)^{x_*}\frac{\Gamma\left(x_* + r\right)}{x_*!\Gamma\left(r\right)} \\
&= \binom{x_* + r - 1}{x_*} p^r(1-p)^{x_*} \\
&= \text{NegBin}\left(r,\, p\right)
\end{aligned}
$$

(p) [harder] If $\alpha \notin \mathbb{N}$, create an "extended negative binomial" r.v. and find its PMF. You can copy from Wikipedia.

(q) [harder] Why is the extended negative binomial r.v. also known as the gamma-Poisson mixture distribution? Why is it also called the "overdispersed Poisson"?

$$
\mathbb{P}(X; n, r, p) = \frac{\binom{x+r-1}{x} p^x}{(1-p)^{-r} - \sum_{i=0}^{n-1} \binom{i+r-1}{x} p^i}
$$

(r) [harder] If you observe $0, 3, 2, 4, 2, 6, 1, 0, 5$, give a 90% CR for θ. Pick an principled objective (uninformative) prior.

$$
\begin{aligned}
CR_{\theta, 95\%} &= [qgamma(0.05, 24, 9), qgamma(0.95, 24, 9)] \\
&= [1.839, 3.621]
\end{aligned}
$$

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

$$
\begin{aligned}
pval = \mathbb{P}\left(\theta < 2 | X\right) &= pgamma(2, 24, 9) - pgamma(0, 24, 9) \\
&= 0.101
\end{aligned}
$$

(t) [harder] Using the data and the prior from (s), find the probability the next observation will be a 7. Leave in exact form then use a calculator to compute it to the nearest two significat digits.

$$
\mathbb{P}\left(X_* = 7 | X\right) = \binom{30}{7} (0.9)^{24} (0.1)^7 = 0.02
$$

(u) [difficult] [MA] We talked about that the negative binomial is an "overdispersed" Poisson. Show that the negative binomial converges to a Poisson.

(v) [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.

Problem 3

We now discuss the theory of the normal-normal conjugate model. Assume

$$X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\theta, \sigma^2\right)$$

but where you see only "X", this is shorthand for all n samples.

(a) [easy] What is the kernel of $\theta \mid X,\ \sigma^2$?

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

(c) [easy] Find the Bayesian point estimates as function of the data and prior's hyperparameters (i.e. $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MMAE}}$ and $\hat{\theta}_{\text{MAP}}$).

- $\hat{\theta}_{\text{MMSE}} = \mathbb{E}\left[\theta|X\right] = \frac{\frac{n\bar{x}}{(\sigma)^2} + \frac{\mu_0}{(\tau)^2}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}}$
- $\hat{\theta}_{\text{MMAE}} = \text{median}\left[\theta|X\right] = \frac{\frac{n\bar{x}}{(\sigma)^2} + \frac{\mu_0}{(\tau)^2}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}}$
- $\hat{\theta}_{\text{MAP}} = \text{Mode}\left[\theta|X\right] = \frac{\frac{n\bar{x}}{(\sigma)^2} + \frac{\mu_0}{(\tau)^2}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}}$

(d) [harder] On a previous homework we showed that if $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\theta,\ \sigma^2\right)$ then $\hat{\theta}_{\text{MLE}} = \bar{x}$. Show that $\hat{\theta}_{\text{MMSE}}$ is a shrinkage estimator and find ρ.

$$\begin{aligned}
\hat{\theta}_{\text{MMSE}} &= \frac{\frac{n\bar{x}}{(\sigma)^2} + \frac{\mu_0}{(\tau)^2}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}} \\
&= \frac{\frac{n\bar{x}}{(\sigma)^2} + \hat{\theta}_{\text{MLE}}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}} \\
&= \frac{\frac{\mathbb{E}[\theta]}{(\tau)^2}}{\frac{n}{(\sigma)^2} + \frac{1}{(\tau)^2}} \\
&= (1-\rho)\hat{\theta}_{\text{MLE}} + \rho\mathbb{E}\left[\theta\right] \\
\rho &= \frac{\hat{\theta}_{\text{MMSE}} - \hat{\theta}_{\text{MLE}}}{\mathbb{E}\left[\theta\right] - \hat{\theta}_{\text{MLE}}}
\end{aligned}$$

(e) [harder] Setup the integral to find $\mathbb{P}(X_* \mid X)$ where $n_* = 1$ but don't solve.

$$
\begin{aligned}
\mathbb{P}(X_*|X) &= \int \mathbb{P}\left(X_*|\theta, (\sigma)^2\right) \mathbb{P}\left(\theta|X, (\sigma)^2\right) \, d\theta \\
&= \int \mathcal{N}\left(\theta,\, \sigma^2\right) \mathcal{N}\left(\theta_p,\, \sigma_p^2\right) \, d\theta \\
&= \int \exp\left(-\frac{1}{2\sigma^2}\left(x_* - \theta\right)^2\right) \exp\left(-\frac{1}{2\sigma_p^2}\left(x - \theta_p\right)^2\right) \, d\theta
\end{aligned}
$$