Math 341 / 650 Spring 2018
Final Examination

Solutions

Professor Adam Kapelner

Wednesday, May 23, 2018

Full Name ______________________________

Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using an unauthorized cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

I acknowledge and agree to uphold this Code of Academic Integrity.

______________________________ signature ______________ date

Instructions

This exam is 120 minutes and closed-book. You are allowed **three** pages (front and back) of a "cheat sheet." You may use a graphing calculator of your choice. Please read the questions carefully. If the question reads "compute," this means the solution will be a number otherwise you can leave the answer in *any* widely accepted mathematical notation which could be resolved to an exact or approximate number with the use of a computer. I advise you to skip problems marked "[Extra Credit]" until you have finished the other questions on the exam, then loop back and plug in all the holes. I also advise you to use pencil. The exam is 100 points total plus extra credit. Partial credit will be granted for incomplete answers on most of the questions. $\boxed{\text{Box}}$ in your final answers. Good luck!

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)
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)
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 exam. 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 Imagine you draw n balls out of a bag with N total balls and θ of those balls are special (leaving $N-\theta$ not special balls). The number of special balls out of the n sampled is given by x. This is called the hypergeometric model; it is a discrete r.v. and it has the PMF:

$$X \sim \text{Hyper}(n, \theta, N) := \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}.$$

For the remainder of this problem, the total number of balls N is known and the number that you sample n is also known. The unknown is θ and that is the target of inference. Its support is $\{1, 2, \ldots, N-1\}$.

(a) [5 pt / 5 pts] What is the prior of indifference for θ?

$p(\theta) = \frac{1}{N-1}$

(b) [5 pt / 10 pts] Using the prior from (b), find the kernel of $\mathbb{P}(\theta \mid X)$. If you didn't get the answer to (a), assume $\mathbb{P}(\theta) \propto 1$ for the purposes of this problem.

$$P(\theta|x) \propto P(x|\theta)\,P(\theta) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}} \frac{1}{N-1} \propto \binom{\theta}{x}\binom{N-\theta}{n-x} = \frac{\theta!}{x!\,(\theta-x)!} \frac{(N-\theta)!}{(n-x)!\,(N-\theta-n+x)!}$$

$$\propto \frac{\theta!\,(N-\theta)!}{(\theta-x)!\,(N-\theta-n+x)!}$$

(c) [7 pt / 17 pts] [Extra Credit] Show that the posterior is beta-binomial and find its parameters.

Problem 2 Consider the normal likelihood model with a sample size of n where θ is known but σ^2 is unknown.

(a) [5 pt / 22 pts] What is the conjugate prior for σ^2 in this case?

$$\sigma^2 \sim \text{InvG}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$$

(b) [5 pt / 27 pts] What is the interpretation of the hyperparameters — what pseudodata do they represent?

n_0: # of previous samples

$n_0\sigma_0^2$: previous SSE

(c) [5 pt / 32 pts] After data is sampled, what is the posterior (i.e. given θ and $X_1, \ldots, X_n$) MMSE estimate for σ^2? Define any shorthand symbols explicitly.

$$P(\sigma^2 \mid \theta, X) = \text{InvG}\left(\frac{n_0+n}{2}, \frac{n_0\sigma_0^2 + n\hat{\sigma}^2_{MLE}}{2}\right) \Rightarrow \hat{\sigma}^2_{MMSE} = \frac{\frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{2}}{\frac{n_0+n}{2} - 1} = \frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{n_0+n-2}$$

where $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2$

(d) [7 pt / 39 pts] Prove this is a shrinkage estimator and find ρ.

$$\hat{\sigma}^2_{MMSE} = \frac{n_0\sigma_0^2}{n_0+n-2} + \frac{n\hat{\sigma}^2}{n_0+n-2} = \frac{n_0-2}{n_0+n-2}\underbrace{\frac{n_0\sigma_0^2}{n_0-2}}_{E[\sigma^2]} + \frac{n}{n_0+n-2}\underbrace{\hat{\sigma}^2_{MLE}}_{MLE} = \rho E[\sigma^2] + (1-\rho)\hat{\sigma}^2_{MLE}$$

$$E[\sigma^2] = \frac{\frac{n_0\sigma_0^2}{2}}{\frac{n_0}{2} - 1} = \frac{n_0\sigma_0^2}{n_0-2} \qquad \Rightarrow \rho = \frac{n_0-2}{n_0+n-2}$$

(e) [7 pt / 46 pts] Write an integral that when evaluated would find $\mathbb{P}(X^* \mid X_1, \ldots, X_n, \theta)$

$$P(X^* \mid X) = \int_{Supp[\sigma^2]} P(X^* \mid \theta, \sigma^2)\, P(\sigma^2 \mid \theta, X)\, d\sigma^2 = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X^*-\theta)^2} \frac{\left(\frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{2}\right)^{\frac{n_0+n}{2}}}{\Gamma\left(\frac{n_0+n}{2}\right)} (\sigma^2)^{-\frac{n_0+n}{2}-1} e^{-\frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{2\sigma^2}}\, d\sigma^2$$

Problem 3 Consider the independent Poisson model where the mean is a linear function of time but zero at inception i.e. $X_t \overset{ind}{\sim} \text{Poisson}(\theta t)$ with n independent samples. We employ the prior $\mathbb{P}(\theta) \propto 1$.

(a) [2 pt / 48 pts] Is the prior for θ proper? Yes / no.

(b) [5 pt / 53 pts] Try to find the posterior i.e. θ given $X_1, \ldots, X_n, t_1, t_2, \ldots, t_n$. Get as far as you can.

$$P(\theta|X_1,\ldots,X_n,t_1,\ldots,t_n) \propto P(X_1,\ldots,X_n|\theta,t_1,\ldots,t_n)P(\theta) = \prod_{i=1}^{n} \frac{(\theta t_i)^{x_i} e^{-\theta t_i}}{x_i!} \propto \prod_{i=1}^{n} (\theta t_i)^{x_i} e^{-\theta t_i}$$

$$= \prod \theta^{x_i} \prod t_i^{x_i} \prod e^{-\theta t_i} \propto \prod \theta^{x_i} \prod e^{-\theta t_i} = \theta^{\Sigma x_i} e^{-\theta \Sigma t_i} \propto \text{Gamma}(\Sigma x_i + 1, \Sigma t_i)$$

(c) [5 pt / 58 pts] Consider now the independent Poisson model where the mean is a linear function of time but one at inception i.e. $X_t \overset{ind}{\sim} \text{Poisson}(1 + \theta t)$. We employ the prior $\mathbb{P}(\theta) \propto 1$. Try to find the posterior; get as far as you can.

$$P(\theta|-) \propto \prod_{i=1}^{n} (1+\theta t_i)^{x_i} e^{-1+\theta t_i} = \prod_{i=1}^{n} (1+\theta t_i)^{x_i} \prod e^{-1+\theta t_i} = e^{-n} e^{-\theta \Sigma t_i} \prod_{i=1}^{n} (1+\theta t_i)^{x_i}$$

$$\propto e^{-\theta \Sigma t_i} \prod_{i=1}^{n} (1+\theta t_i)^{x_i} = k(\theta|-)$$

(d) [8 pt / 66 pts] Describe a means to test if $\theta > 0$ *approximately* using grid sampling. You can assume that we are fairly certain that $|\theta| \leq 100$. Full credit only given to answers that provide all details of the computation.

$H_0: \theta \leq 0,\ H_a: \theta > 0,\ \alpha = 5\%$

1. Pick a small $\Delta\theta$ and create a grid $G := \langle -100, -100+\Delta\theta, -100+2\Delta\theta$
2. Calculate $k(\theta|-)\ \forall \theta \in G$
3. Compute $c = 1 / \sum_{\theta \in G} k(\theta|-)$
4. compute $P(\theta|-) = c\, k(\theta|-)\ \forall \theta \in G$
5. compute $F(\theta|-)\ \forall \theta \in G$
6. Draw u from $U(0,1)$
7. Find $\theta_{samp} = \underset{\theta \in G}{\text{argmin}} \{ F(\theta|-) \geq u \}$
8. Repeat steps 6-7 for many times (T) to get $\{\theta_{samp,1}, \ldots, \theta_{samp,T}\}$
9. Calc approx $pval = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\theta_{samp,t} \leq 0}$
10. If $pval < \alpha = 5\% \Rightarrow$ Reject H_0 otherwise retain H_0.

(e) [5 pt / 71 pts] Will the test in (d) be nearly exact? Or will it suffer from the disadvantage of grid sampling? Explain.

It will be nearly exact. This is one-dimension and θ_{min} & θ_{max} are known. We can pick a very small $\Delta\theta$ and achieve very high resolution

Problem 4 Recall the mixture model of two normals:

$$X_1, \ldots, X_n \overset{iid}{\sim} \rho \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(x-\theta_1)^2\right) + (1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(x-\theta_2)^2\right)$$

We employed the uninformative priors $\mathbb{P}(\rho) \propto 1$, $\mathbb{P}(\theta_1) \propto 1$, $\mathbb{P}(\theta_2) \propto 1$, $\mathbb{P}(\sigma_1^2) \propto 1/\sigma_1^2$ and $\mathbb{P}(\sigma_2^2) \propto 1/\sigma_2^2$ but could not solve the problem because the kernel for the posterior did not reduce to a managable expression.

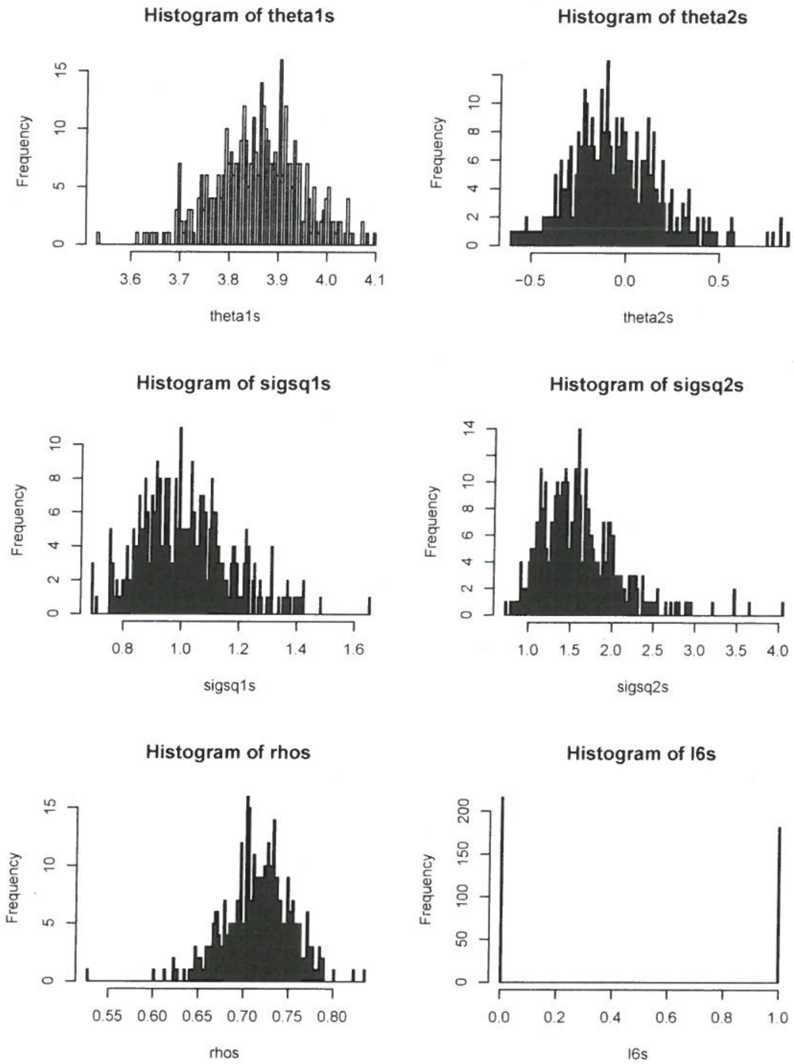
We the introduced "data augmentation" where we pretended that each x_i belonged to either the first normal distribution or the second based on the value of the indicator I_i which is 1 if from the first distribution. We then found the kernel of the posterior,

$$\mathbb{P}\left(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho, I_1, \ldots, I_n \mid X_1, \ldots, X_n\right) \propto$$
$$\frac{1}{\sigma_1^2\sigma_2^2} \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(x-\theta_1)^2\right)\right)^{I_i} \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(x-\theta_2)^2\right)\right)^{1-I_i} \rho^{I_i}(1-\rho)^{1-I_i}$$

We were then able to find all conditional distributions,

$$\begin{aligned}
&\mathbb{P}\left(\theta_1 \mid X_1, \ldots, X_n, \theta_2, \sigma_1^2, \sigma_2^2, \rho, I_1, \ldots, I_n\right) \\
&\mathbb{P}\left(\theta_2 \mid X_1, \ldots, X_n, \theta_1, \sigma_1^2, \sigma_2^2, \rho, I_1, \ldots, I_n\right) \\
&\mathbb{P}\left(\sigma_1^2 \mid X_1, \ldots, X_n, \theta_1, \theta_2, \sigma_2^2, \rho, I_1, \ldots, I_n\right) \\
&\mathbb{P}\left(\sigma_2^2 \mid X_1, \ldots, X_n, \theta_1, \theta_2, \sigma_1^2, \rho, I_1, \ldots, I_n\right) \\
&\mathbb{P}\left(I_1 \mid X_1, \ldots, X_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho, I_2, \ldots, I_n\right) \\
&\vdots \\
&\mathbb{P}\left(I_n \mid X_1, \ldots, X_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho, I_1, \ldots, I_{n-1}\right),
\end{aligned}$$

and all were easy to sample from. Hence we built a Gibbs sampler. After burning and thinning, we arrive at the following samples. The plot on the bottom right is the samples for I_6.



(a) [7 pt / 78 pts] Assume there were $n = 300$ samples. Create an approximate 95% credible region for the number of samples that were drawn from the first distribution, the normal with mean θ_1 and variance σ_1^2.

$CR_{\rho,95\%} \approx$ ~~0.69~~ $[0.63, 0.77]$ $\quad CR_{n \in 1^{st} distr, 95\%} \approx [189, 231]$

(b) [4 pt / 82 pts] Approximate the probability that the 6th observation came from the first distribution.

45%

(c) [3 pt / 85 pts] Find $\hat{I}_{6,MAP}$. 0

(d) [5 pt / 90 pts] Test $\sigma_1^2 > 2$. $H_0: \sigma_1^2 \le 2$, $H_a: \sigma_1^2 > 2$, $\alpha = 5\%$

$p_{val} = P(H_0|X) = P(\sigma_1^2 \le 2|X) \approx \frac{1}{T}\sum \mathbb{1}_{\sigma_{1,t}^2 \le 2} \approx 100\% \Rightarrow$ Retain H_0

(e) [7 pt / 97 pts] In finance, the Sharpe ratio is a useful indicator and it is defined as $\frac{\theta - r_f}{\sigma}$ where r_f is a constant. Explain how you would use this Gibbs chain to ~~find~~ approximate the posterior of the Sharpe ratio of the first distribution.

① Draw one vector $\begin{bmatrix} \theta_1 \\ \sigma_1^2 \end{bmatrix}$ from the burned and thinned chain

② Compute $S = \frac{\theta_1 - r_f}{\sigma_1}$

③ Repeat steps 1, 2 many times T to produce $\{S_1, \ldots, S_T\}$ which approximates the posterior of the Sharpe ratio for the "first distr."!

(f) [10 pt / 107 pts] Explain how you would draw one sample ($n^* = 1$) from $\mathbb{P}(X^* \mid X_1, \ldots, X_n)$ using the Gibbs chain.

① Draw one vector $\begin{bmatrix} \theta_1 \\ \sigma_1^2 \\ \theta_2 \\ \sigma_2^2 \\ \rho \end{bmatrix}$ from the burned and thinned chain.

② Draw Y_1 from $N(\theta_1, \sigma_1^2)$

③ Draw y_2 from $N(\theta_2, \sigma_2^2)$

④ Compute $X^* = \rho y_1 + (1-\rho) y_2$

(g) [3 pt / 110 pts] [Extra Credit] The alternative to data augmentation is to use the posterior directly and use 5 Metropolis steps in the sampler: one for θ_1, one for θ_2, one for σ_1^2, one for σ_2^2 and one for ρ. Why would this be worse than using the Gibbs sampler with the data augmentation?