Math 341 / 650 Fall 2021
Midterm Examination Two

Professor Adam Kapelner

Wednesday, April 28, 2021

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.

By taking this exam, you acknowledge and agree to uphold this Code of Academic Integrity.

Instructions

This exam is 70 minutes (variable time per question) and closed-book. You are allowed **one** page (front and back) of a "cheat sheet", blank scrap paper and a graphing calculator. Please read the questions carefully. No food is allowed, only drinks.

Problem 1 [7min] (and 7min will have elapsed) Let $\mathbb{P}(X \mid \theta) \sim \text{Poisson}(\theta)$, $\mathbb{P}(\theta) = \text{InvGamma}(\alpha, \beta)$ and $\alpha, \beta > 0$.

- [13 pt / 13 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) $\mathbb{P}(\theta)$ is the conjugate prior for the parametric model given by $\mathbb{P}(X \mid \theta)$

  (b) $\mathbb{P}(\theta)$ has the same support as the parameter space of $\mathbb{P}(X \mid \theta)$

  (c) $\mathbb{P}(\theta \mid X)$ is a mixture / compound distribution

  (d) $\mathbb{P}(X)$ is a mixture / compound distribution

  (e) The kernel of $\mathbb{P}(\theta \mid X)$ is $\theta^{x-\alpha-1} e^{-(\theta+\beta/\theta)}$

  (f) The kernel of $\mathbb{P}(\theta \mid X)$ is $\theta^x / x!$

  (g) The posterior is gamma-distributed

  (h) The normalization constant of $\mathbb{P}(\theta \mid X)$ is given by $\beta^\alpha / (x! \Gamma(\alpha))$

  (i) The posterior predictive distribution can be computed via $\int_0^\infty \mathbb{P}(X \mid \theta) \mathbb{P}(\theta) \, d\theta$

  (j) The posterior predictive distribution rv will have higher variance than the Poisson rv

  (k) The posterior predictive distribution rv will have one parameter: θ

  (l) The posterior predictive distribution rv will have two parameters: α, β

  (m) The posterior predictive distribution rv will have support $(0, \infty)$

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 2 [8min] (and 15min will have elapsed)

Let $\mathbb{P}(X \mid \theta) \sim \text{Poisson}(\theta)$ and then consider the reparameterization of θ called $\phi := t(\theta) = \ln(\theta)$ and $\theta = t^{-1}(\phi) = e^\phi$ so that $\mathbb{P}(X \mid \phi) = e^{x\phi} - e^\phi /x!$

- [8 pt / 21 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) To compute the Jeffrey's prior, the only information necessary is the likelihood function of $\mathcal{F}$

  (b) $\mathbb{P}_J(\theta) = \text{Gamma}(1/2,\, 0)$

  (c) $\ell(\phi; x) = x\phi - e^\phi - \ln(x!)$

  (d) $\ell'(\phi; x) = x - e^\phi$

  (e) $\mathbb{P}_J(\phi) \propto e^\phi$

  (f) $\mathbb{P}_J(\phi) \propto e^{\phi/2}$

  (g) If you parameterize by θ and use $\mathbb{P}_J(\theta)$ to derive the posterior, you will get the same point estimates $\hat{\theta}_{\text{MMSE}}, \hat{\theta}_{\text{MMAE}}, \hat{\theta}_{\text{MAP}}$ if you parameterize by ϕ and use $\mathbb{P}_J(\phi)$ to derive the posterior.

  (h) $f_1(\phi) := \mathbb{P}_J(\phi)$ can be derived from $f_2(\theta) := \mathbb{P}_J(\theta)$ using the change of variables equation, $f_1(\phi) = f_2(t^{-1}(\theta)) \left| \frac{d}{d\phi}[t^{-1}(\theta)] \right|$

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 3 [13min] (and 28min will have elapsed) We are looking to understand the annual risk of heart attacks in men with high cholesterol as defined by the mean number of annual heart attacks. In 2020, among $n = 20$ men with high cholesterol, one man had two heart attacks, four men had one heart attack and the other men had no heart attacks. An iid Poisson parametric model seems to be a reasonable choice for these annual heart attack events. We have absolutely no prior information so we are most comfortable employing a Haldane-type ignorance prior. All quantities below are rounded to the nearest two decimals.

- [14 pt / 35 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

    (a) $\hat{\theta}^{\text{MLE}} = 0.30$

    (b) $\mathbb{P}(\theta) = \text{Gamma}(1, 0)$

    (c) $\hat{\theta}_{\text{MMSE}} = 0.35$

    (d) Given quantities provided, you need a computing device to compute $\hat{\theta}_{\text{MMAE}}$

    (e) $\hat{\theta}_{\text{MMSE}} = 0.30$

    (f) $\hat{\theta}_{\text{MAP}} = 0.35$

    (g) $\hat{\theta}_{\text{MAP}} = 0.30$

    (h) Under this Haldane prior we chose, $\hat{\theta}_{\text{MMSE}}$ will shrink towards $x_0 = 0$

    (i) A 95% credible region for θ can be constructed by computing [qgamma(0.025, 6, 20), qgamma(0.975, 6, 20)]

    (j) To test the theory that high cholesterol induces higher heart attack risk than the average risk of 0.07, you can calculate pgamma(0.07, 6, 20) and compare it to your significance level (which is usually set to be 5%)

    (k) Using this data and prior, you can compute to the nearest two decimals the probability that a randomly selected man with high cholesterol having zero heart attacks in a year would be zero.

    (l) Using this data and prior, it is possible to compute to the nearest two decimals the probability that a randomly selected man with high cholesterol having one heart attack in a year would be θe^{θ}.

    (m) Using this data and prior, it is possible to compute to the nearest two decimals the probability that a randomly selected man with high cholesterol having one heart attack in a year would be $(20/21)^7$.

    (n) Using this data and prior, it is possible to compute to the nearest two decimals the probability that a randomly selected man with high cholesterol having one heart attack in a year would be $6 \times 20^6/21^7$.

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 4 [14min] (and 42min will have elapsed) We are looking to understand IQ levels in adults that were exposed to a toxin in youth. We sample $n = 7$ adults exposed to the toxin and measure IQ's of 79, 91, 103, 87, 86, 91 and 75 and thus $\bar{x} = 87.43$. The normal iid model is a reasonable model as the population IQ metric has been demonstrated to be normally distributed. It is also known that the population variance of IQ scores is $\sigma^2 = 15^2$ since the scale is designed pegged to that value (so it seems reasonable to assume the value of σ^2 is known in our setting). You have no strong prior belief on the IQ levels in these adults after being exposed to a toxin so you can employ any of the three principled objective priors we discussed found in the class lectures. All quantities are rounded to the nearest two decimals.

- [13 pt / 48 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) $\mathbb{P}(\theta) = \mathcal{N}(0, 0) \propto 1$

  (b) The kernel of the posterior can be expressed as $e^{a\theta - b\theta^2}$ where a, b are functions of $x_1, \ldots, x_n$, σ^2 and fundamental constants (but not functions of θ)

  (c) $\hat{\theta}_{\text{MMSE}} = 87.43$

  (d) $\hat{\theta}_{\text{MAP}} = \bar{x}$

  (e) Given quantities provided, you need a computing device to compute $\hat{\theta}_{\text{MMAE}}$

  (f) A 98% credible region can be constructed by computing [qnorm(0.01, $\bar{x}$, $\sqrt{15}$), qnorm(0.99, $\bar{x}$, $\sqrt{15}$)]

  (g) A 98% credible region can be constructed by computing [qnorm(0.01, $\bar{x}$, $\sqrt{15}/7$), qnorm(0.99, $\bar{x}$, $\sqrt{15}/7$)]

  (h) The mean IQ in the overall population is pegged to be 100. To test the theory that this toxin exposure in youth results in a below-mean IQ, you can calculate pnorm(100, θ_p, σ_p^2) where θ_p is the posterior mean and σ_p^2 is the posterior variance and compare it to your significance level (which is usually set to be 5%)

  (i) Under this prior we chose, $\hat{\theta}_{\text{MMSE}}$ will shrink towards $x_0 = 100$, the true population mean.

  (j) This prior we chose is equivalent to seeing $n_0 = 7$ pseudoobservations

  (k) Using this data and prior, you can compute the probability that a randomly selected man (exposed to the toxin in youth) will have an adult IQ less than 100 is pnorm(100, 87.43, 4.14)

  (l) A 95% posterior predictive interval can be computed as [qnorm(0.025, θ_p, $\sqrt{\sigma^2 + \sigma_p^2}$), qnorm(0.975, θ_p, $\sqrt{\sigma^2 + \sigma_p^2}$)]

  (m) A 95% posterior predictive interval can be computed as
  [qt.scaled(0.025, 7, θ_p, $\sqrt{\sigma^2 + \sigma_p^2}$), qt.scaled(0.975, 7, θ_p, $\sqrt{\sigma^2 + \sigma_p^2}$)]

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 5 [14min] (and 56min will have elapsed) We are looking to understand annual financial returns of a mutual fund for the NYC teachers' retirement system. The fund is guaranteed to produce a mean return of 7.5% over 40 years since inception but is allowed to vary its return year-year based on natural variation in the world economy. You unfortunately have only $n = 3$ years of data to use: you see returns 18%, -3% and 4%. Although it is probably not an appropriate parametric model for annual return, assume yearly returns are iid normal with known $\theta = 7.5\%$. For now, you have no strong prior belief on variance in these returns so you should employ the standard uninformative prior in this situation, the Jeffrey's prior. All quantities below are rounded to the nearest two digits.

- [13 pt / 61 pts]   Record the letter(s) of all the following that are **true** in general. At least one will be true.

    (a) $\mathbb{P}(\theta) \propto 1$

    (b) $\hat{\sigma}^2_{\text{MLE}} = 77.58$

    (c) $\hat{\sigma}^2_{\text{MAP}} = 77.58$

    (d) Given quantities provided, you need a computing device to compute $\hat{\sigma}^2_{\text{MMAE}}$

    (e) $\mathbb{P}(\sigma^2 \mid X, \theta) = \text{InvGamma}(1.50, 116.38)$

    (f) $\mathbb{P}(\sigma^2 \mid X, \theta) = \text{InvGamma}(1.50, 114.33)$

    (g) $\mathbb{P}(\sigma^2 \mid X, \theta) = \text{InvGamma}(3.00, 232.75)$

    (h) $\mathbb{P}(\sigma^2 \mid X, \theta) = \text{InvGamma}(0, 0)$

    (i) $\mathbb{P}(\sigma^2 \mid X, \theta) = k(\sigma^2 \mid X, \theta)$

    (j) A 98% credible region can be constructed by computing $[\text{qinvgamma}(0.01, n/2, ns^2/2), \text{qinvgamma}(0.99, n/2, ns^2/2)]$ where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

    (k) Using this data and prior, you can compute the probability that the variance of the returns is less than $5\%^2$ is $\text{pinvgamma}(5\%^2, n/2, n\hat{\sigma}^2_{\text{MLE}}/2)$

    (l) The probability that the annual return next year will be more than 10% can be exactly computed via $1 - \text{pt.scaled}(3, 10\%, 7.5\%, \sqrt{\hat{\sigma}^2_{\text{MLE}}})$

    (m) The probability that the annual return next year will be more than 10% can be approximated accurately via computing $1 - \text{pnorm}(10\%, 7.5\%, \sqrt{\hat{\sigma}^2_{\text{MLE}}})$

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.

Problem 6 [14min] (and 70min will have elapsed) We are looking to understand annual financial returns of a mutual fund for the NYC teachers' retirement system. The fund is guaranteed to produce a mean return of 7.5% over 40 years since inception but is allowed to vary its return year-year based on natural variation in the world economy. You unfortunately have only $n = 3$ years of data to use: you see returns 18%, -3% and 4%. Although it is probably not an appropriate parametric model for annual return, assume yearly returns are iid normal with known $\theta = 7.5\%$. Since we only have $n = 3$, inference for σ^2 will be very inaccurate as $\mathbb{Var}\left[\sigma^2 \mid X, \theta\right]$ from Problem 5 is too high to be useful. We wish to "borrow estimation strength" by examining other past financial products that are similar to this NYC teachers' retirement system mutual fund. Using the data from other financial products, we fit the following distribution to their lifetime variances: $\mathbb{P}\left(\sigma^2 \mid \theta\right) = \mathrm{InvGamma}\left(5.93, 18.17\right)$ and use this as our prior to perform our Bayesian statistical inference. All quantities below are rounded to the nearest two digits.

- [13 pt / 74 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) $\mathbb{P}\left(\sigma^2\right) \propto 1$

  (b) This prior is informative

  (c) Using this prior, the inference would constitute an example of "Empirical Bayes" in action

  (d) This prior's strength is approximately equivalent to observing 12 annual returns of the NYC teachers' retirement system mutual fund that we are investigating

  (e) To generate this prior, we looked at 12 similar past financial products

  (f) Under this prior, you expect σ^2 to be 3.69

  (g) If you employ $\hat{\sigma}^2_{\mathrm{MMSE}}$ for point estimation, your point estimate will be weighted more towards the prior's expectation than the maximum likelihood estimate from the data

  (h) $\hat{\sigma}^2_{\mathrm{MLE}} = 20.92$

  (i) $\hat{\sigma}^2_{\mathrm{MMSE}} = 20.92$

  (j) We estimate that the variance of returns for the NYC teachers' retirement system mutual fund is now smaller than previously when we employed the Jeffrey's prior

  (k) You need a computing device to compute $\hat{\sigma}^2_{\mathrm{MMAE}}$

  (l) $\mathbb{P}\left(X_* \mid X, \theta\right) = T_{n+n_0}\left(\theta, \hat{\sigma}^2_{\mathrm{MMSE}}\right)$

  (m) The probability that the annual return next year will be more than 10% can be exactly computed via
  1 - pt.scaled$(3, 10\%, 7.5\%, \sqrt{20.92})$

Your answer will consist of a lowercase string (e.g. `aebgd`) where the order of the letters does not matter.