Math 341 / 650 Spring 2021
Final Examination

Professor Adam Kapelner

Monday, May 24, 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 110 minutes (variable time per question) and closed-book. You are allowed **three** pages (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 [17min] (and 17min will have elapsed) Consider the case of the beta-binomial conjugate model with fixed n under the prior of indifference. Let X_* be the number of successes in n_* future observations from the same process.

- [26 pt / 26 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

(a) $\mathbb{P}(\theta) \propto \text{Beta}(0,\, 0)$

(b) $\mathbb{P}(\theta) = 1$

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

(d) $\mathbb{P}(\theta) \propto \binom{n}{x}\theta^x (1-\theta)^{n-x}$

(e) $\mathbb{P}(\theta) \propto \theta^x (1-\theta)^{n-x}$

(f) $\mathbb{P}(\theta) \propto (x!(n-x)!)^{-1} (\theta/(1-\theta))^x$

(g) $\mathbb{P}(X \mid \theta) = 1$

(h) $\mathbb{P}(X \mid \theta) \propto 1$

(i) $\mathbb{P}(X \mid \theta) \propto \binom{n}{x}\theta^x (1-\theta)^{n-x}$

(j) $\mathbb{P}(X \mid \theta) \propto \theta^x (1-\theta)^{n-x}$

(k) $\mathbb{P}(X \mid \theta) \propto (x!(n-x)!)^{-1} (\theta/(1-\theta))^x$

(l) $\mathbb{P}(\theta \mid X) = 1$

(m) $\mathbb{P}(\theta \mid X) \propto 1$

(n) $\mathbb{P}(\theta \mid X) \propto \binom{n}{x}\theta^x (1-\theta)^{n-x}$

(o) $\mathbb{P}(\theta \mid X) \propto \theta^x (1-\theta)^{n-x}$

(p) $\mathbb{P}(\theta \mid X) \propto (x!(n-x)!)^{-1} (\theta/(1-\theta))^x$

(q) $\mathbb{P}(X) = 1$

(r) $\mathbb{P}(X) \propto 1$

(s) $\mathbb{P}(X) \propto \binom{n}{x}\theta^x (1-\theta)^{n-x}$

Be careful about X and X_* in the following problems:

(t) $\mathbb{P}(X_* \mid \theta) \propto \binom{n}{x_*}\theta^{x_*} (1-\theta)^{n-x_*}$

(u) $\mathbb{P}(X_* \mid \theta) \propto \theta^{x_*} (1-\theta)^{n-x_*}$

(v) $\mathbb{P}(X_* \mid \theta) \propto (x_*!(n-x_*)!)^{-1} (\theta/(1-\theta))^{x_*}$

(w) $\mathbb{P}(X_* \mid X) \propto \binom{n}{x_*}\theta^{x_*} (1-\theta)^{n-x_*}$

(x) $\mathbb{P}(X_* \mid X) \propto \theta^x (1-\theta)^{n-x}$

(y) $\mathbb{P}(X_* \mid X) \propto \theta^{x_*} (1-\theta)^{n-x_*}$

(z) $\mathbb{P}(X_* \mid X) \propto B(1+x+x_*, 1+n-x+n_*-x_*)$

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

Problem 2 [8min] (and 25min will have elapsed) Recall the baseball batter example from class. Here, $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$ where each X_i is either a hit or not a hit. We are trying to infer θ, his true probability of getting a hit (or his true "batting average"). We employed the prior $\mathbb{P}(\theta) = \text{Beta}(\alpha = 78.7, \beta = 224.8)$. Let $X = \sum_{i=1}^{n} X_i$. We collected the following data: $x = 11$ hits out of $n = 25$ at bats.

- [10 pt / 36 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) θ is considered fixed for every one of the n at bats

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

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

  (d) $\mathbb{P}(\theta)$ is a principled uninformative prior even though it is not one of the three we explicitly studied in class

  (e) $\mathbb{P}(\theta)$ has hyperparameters that were fit using the principle of indifference

  (f) With a different value of n, $\mathbb{P}(\theta)$ could be considered uninformative

  (g) $\mathbb{P}(\theta)$ can be sampled from using the function `pbeta`

  (h) $\mathbb{P}(\theta)$ can be sampled from using the grid sampling algorithm

  (i) $\hat{\theta}_{\text{MMSE}}$ is numerically closer to $\mathbb{E}[\theta]$ than $\bar{x} = 11/25$

  (j) A 95% posterior predictive interval for the value of the next at bat could be computed by $[\texttt{qbeta}(2.5\%, \alpha + x, \beta + n - x), \texttt{qbeta}(97.5\%, \alpha + x, \beta + n - x)]$

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

Problem 3 [10min] (and 35min will have elapsed) We will continue with our discussion of the batter. But now we have a slightly different situation. There is reason to believe the player's θ changes during the course of his n at bats (where n is fixed). We denote this change point as m. Thus, for the first m at bats, he bats with probability of getting a hit θ_0 and for the remaining $n - m$ at bats, he bats with probability of getting a hit θ_1 where $\theta_0 \neq \theta_1$. All at bats are considered independent. Since we are no longer interested in his lifetime career batting average, we have no need for the empirical Bayes prior of the previous problem. We now use Laplace's prior for θ_0 and θ_1. Since we have no strong feeling of where the change point is, we use Laplace's prior as well there. We are interested in inference for all unknown parameters using the data $x := [x_1, \ldots, x_n]$ where each value is 1 if there is a hit at time t or 0 if not.

- [14 pt / 50 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

(a) $X_1, \ldots, X_m \overset{iid}{\sim} \text{Binomial}\,(m,\, \theta_0)$

(b) The unknown parameters are $\theta_0, \theta_1, m, n, t$

(c) The unknown parameters are θ_0, θ_1, m, n

(d) The unknown parameters are θ_0, θ_1, m

(e) $\theta_0 + \theta_1 = 1$

(f) The shrinkage ρ for posterior expectation of θ_0 can be computed as $m\theta_0 + (n-m)\theta_1$

(g) $\mathbb{P}\,(\theta_0) = 1$

(h) $\mathbb{P}\,(\theta_1) = 1$

(i) $\mathbb{P}\,(m) = 1$

(j) The likelihood of the data is the same if $m = 0$ as if $m = n$

(k) The posterior is the same value if calculated at $m = 0$ and $m = n$

(l) If X_* denotes the random variable (rv) of only the next at bat, then X_* is a bernoulli random variable

(m) If X_* denotes the rv of the next $n_* > 1$ at bats, then X_* is a binomial random variable if the data and θ_1 is known

(n) If X_* denotes the rv of the next $n_* > 1$ at bats, then X_* is a betabinomial random variable if the data and m is known

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

Problem 4 [11min] (and 46min will have elapsed) We will continue with our discussion of the batter. But now we have a slightly different situation. There is reason to believe the player's θ changes during the course of his n at bats (where n is fixed). We denote this change point as m. Thus, for the first m at bats, he bats with probability of getting a hit θ_0 and for the remaining $n-m$ at bats, he bats with probability of getting a hit θ_1 where $\theta_0 \neq \theta_1$. All at bats are considered independent. Since we are no longer interested in his lifetime career batting average, we have no need for the empirical Bayes prior of the previous problem. We now use Laplace's prior for θ_0 and θ_1. Since we have no strong feeling of where the change point is, we use Laplace's prior as well there. We are interested in inference for all unknown parameters using the data $x := [x_1, \ldots, x_n]$ where each value is 1 if there is a hit at time t or 0 if not. Letting $a := \sum_{i=1}^{m} x_i$ and $b = \sum_{i=m+1}^{n} x_i$ (where a is defined to be zero if $m=0$ and b is defined to be zero if $m=n$), the likelihood of the data can be written as:

$$\mathbb{P}\left(X \mid \theta_0, \theta_1, m\right) = \binom{m}{a} \theta_0^a (1-\theta_0)^{m-a} \binom{n-m}{b} \theta_1^b (1-\theta_1)^{n-m-b}.$$

- [8 pt / 58 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) The likelihood of the data can be decomposed as $\mathbb{P}\left(X \mid \theta_0, \theta_1, m\right) = \mathbb{P}\left(X_1, \ldots, X_m \mid \theta_0, m\right) \mathbb{P}\left(X_{m+1}, \ldots, X_n \mid \theta_1, m\right)$

  (b) The likelihood of the data can be decomposed as $\mathbb{P}\left(X \mid \theta_0, \theta_1, m\right) = \mathbb{P}\left(X_1, \ldots, X_m \mid \theta_0\right) \mathbb{P}\left(X_{m+1}, \ldots, X_n \mid \theta_1\right) \mathbb{P}\left(X \mid m\right)$

  (c) $\mathbb{P}\left(\theta_0, \theta_1, m \mid X\right) \propto \mathbb{P}\left(X \mid \theta_0, \theta_1, m\right)$

  (d) $k(\theta_0, \theta_1, m \mid X) = \mathbb{P}\left(X \mid \theta_0, \theta_1, m\right)$ where k denotes the irreducible kernel of the distribution

  (e) $\mathbb{P}\left(\theta_0 \mid X, \theta_1, m\right) = \text{Beta}\left(a+1,\, m-a+1\right)$

  (f) $\mathbb{P}\left(\theta_1 \mid X, \theta_0, m\right) = \text{Beta}\left(b+1,\, m-b+1\right)$

  (g) $\mathbb{P}\left(m \mid X, \theta_0, \theta_1\right) = \text{BetaBinomial}\left(n,\, \theta_0,\, \theta_1\right)$

  (h) $k(m \mid X, \theta_0, \theta_1) = \mathbb{P}\left(m \mid X, \theta_0, \theta_1\right)$ where k denotes the irreducible kernel of the distribution

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

Problem 5 [6min] (and 52min will have elapsed) We will continue with our discussion of the batter. But now we have a slightly different situation. There is reason to believe the player's θ changes during the course of his n at bats (where n is fixed). We denote this change point as m. Thus, for the first m at bats, he bats with probability of getting a hit θ_0 and for the remaining $n-m$ at bats, he bats with probability of getting a hit θ_1 where $\theta_0 \neq \theta_1$. All at bats are considered independent. Since we are no longer interested in his lifetime career batting average, we have no need for the empirical Bayes prior of the previous problem. We now use Laplace's prior for θ_0 and θ_1. Since we have no strong feeling of where the change point is, we use Laplace's prior as well there. We are interested in inference for all unknown parameters using the data $x := [x_1, \ldots, x_n]$ where each value is 1 if there is a hit at time t or 0 if not.Letting $a := \sum_{i=1}^{m} x_i$ and $b = \sum_{i=m+1}^{n} x_i$ (where a is defined to be zero if $m = 0$ and b is defined to be zero if $m = n$), we can show that:

$$
\begin{aligned}
\mathbb{P}(\theta_0 \mid X, \theta_1, m) &= \text{Beta}(a+1,\, m-a+1), \\
\mathbb{P}(\theta_1 \mid X, \theta_0, m) &= \text{Beta}(b+1,\, n-m-b+1), \\
k(m \mid X, \theta_0, \theta_1) &\propto \binom{m}{a}\theta_0^a(1-\theta_0)^{m-a}\binom{n-m}{b}\theta_1^b(1-\theta_1)^{-m-b}.
\end{aligned}
$$

- [13 pt / 71 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) $\mathbb{P}(m \mid X, \theta_0, \theta_1)$ can be solved for in closed form and thus can be computed exactly

  (b) Regardless of whether (a) is true, $\mathbb{P}(m \mid X, \theta_0, \theta_1)$ can be sampled from using a grid sampler

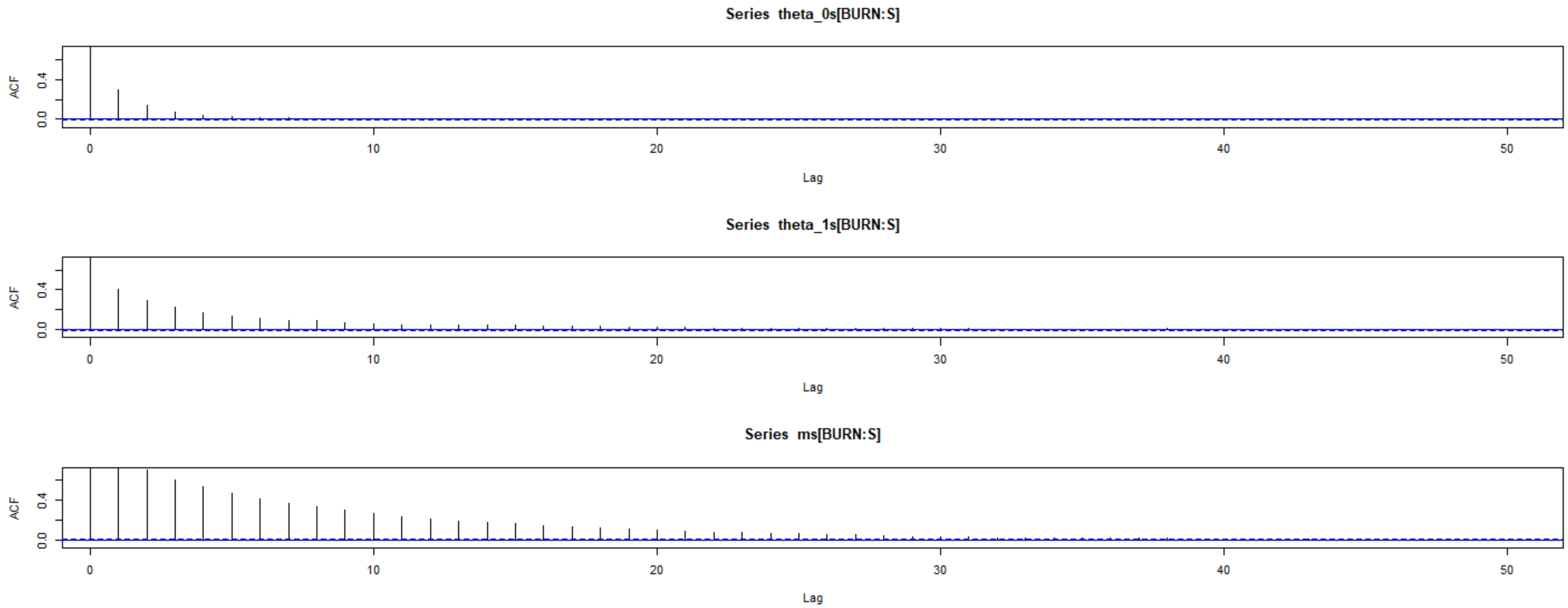
  (c) Regardless of whether (b) is true, a grid sampler for $\mathbb{P}(m \mid X, \theta_0, \theta_1)$ will suffer since we do not know how to set the minimum and maximum sizes of m nor the grid resolution Δ

  (d) Using the information above, you can sample from $\mathbb{P}(\theta_0, \theta_1, m \mid X)$ using a systematic sweep Gibbs sampler which will eventually converge

  (e) Assume the Gibbs sampler will converge, starting at $\theta_0 = 0$, $\theta_1 = 0$ and $m = 0$ will allow for convergence

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

Problem 6 [4min] (and 56min will have elapsed) In our dataset, $n = 30$. Using the three conditional distributions from the previous problem within a systematic sweep Gibbs sampler running for $S = 100,000$ iterations, we find that the Gibbs sampler burns in almost immediately, so we set $B = 10$. We now wish to assess sample dependence. Here are the autocorrelation plots for θ_0 (top), θ_1 (middle) and m (bottom):



- [4 pt / 75 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) The samples of θ_0 become independent of each other more quickly than the samples of θ_1 which become more quickly independent than the samples of m

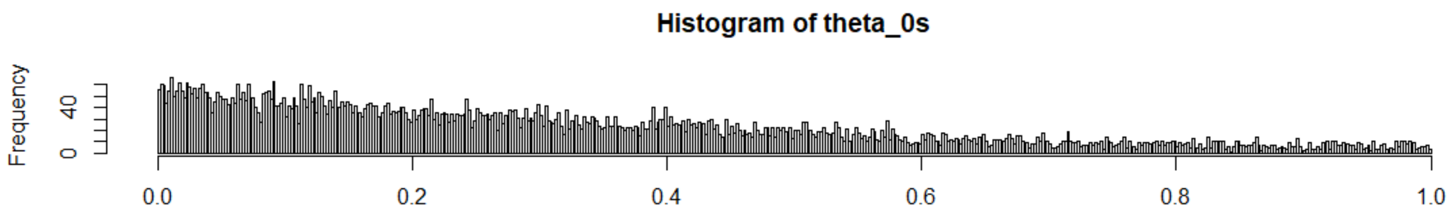
  (b) We can thin the Gibbs chain every 10 iterations to arrive at iid samples from the burned and thinned chain

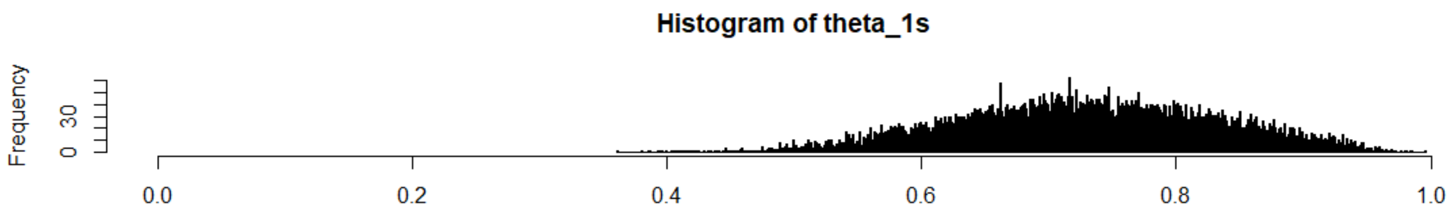
  (c) We can thin the Gibbs chain every 40 iterations to arrive at iid samples from the burned and thinned chain

  (d) We cannot tell when to thin the Gibbs chain from these plots

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

Problem 7 [6min] (and 62min will have elapsed) In our dataset, $n = 30$. Using the three conditional distributions from the previous problem within a systematic sweep Gibbs sampler running for $S = 100,000$ iterations, we find that the Gibbs sampler burns in almost immediately, so we set $B = 10$. We thinned every 40 iterations. Below is a histogram of the burned and thinned samples of θ_0 (top) and θ_1 (bottom):





- [6 pt / 81 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) There are 11,110 samples for θ_0

  (b) There are a different number of samples for θ_0 and θ_1

  (c) It is likely that $\mathbb{V}\text{ar}[\theta_0 \mid X] > \mathbb{V}\text{ar}[\theta_1 \mid X]$

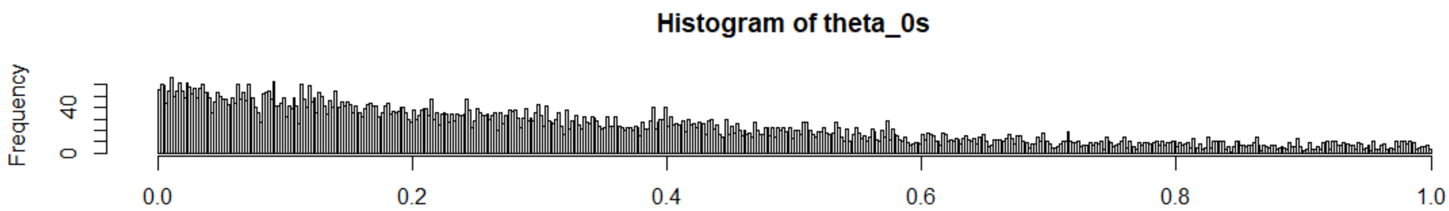
  (d) It is likely that m is small relative to $n - m$

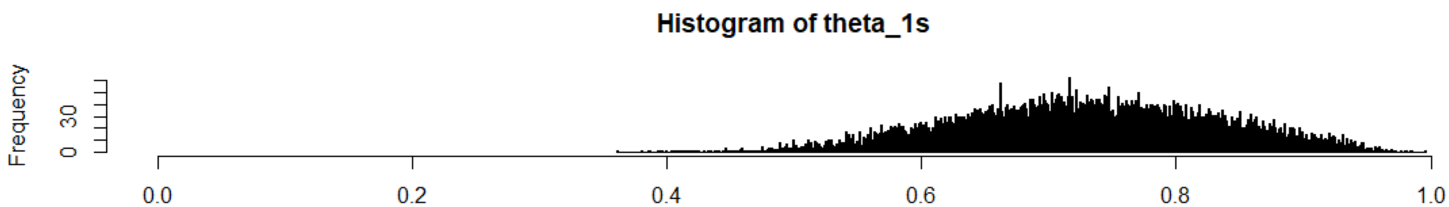
  (e) It is not possible to approiximate a point estimate for θ_0 given this illustration

  (f) It is likely that $\theta_1 > \theta_0$

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

Problem 8 [6min] (and 68min will have elapsed) In our dataset, $n = 30$. Using the three conditional distributions from the previous problem within a systematic sweep Gibbs sampler running for $S = 100,000$ iterations, we find that the Gibbs sampler burns in almost immediately, so we set $B = 10$. We thinned every 40 iterations. Below is a histogram of the burned and thinned samples of θ_0 (top) and θ_1 (bottom):





- [5 pt / 86 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) The MMSE estimate for θ_0 is near 0.3

  (b) The MAP estimate for θ_0 is definitely zero

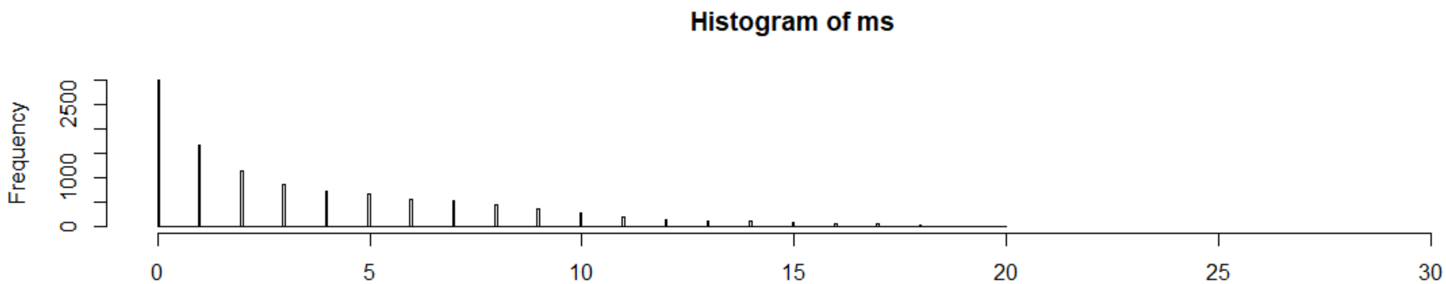
  (c) The MMSE estimate for θ_1 is slightly higher than 0.7

  (d) The MMAE estimate for θ_1 is slightly higher than 0.7

  (e) The MAP estimate for θ_1 is likely near the MMSE estimate for θ_1

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

Problem 9 [6min] (and 74min will have elapsed) In our dataset, $n = 30$. Using the three conditional distributions from the previous problem within a systematic sweep Gibbs sampler running for $S = 100,000$ iterations, we find that the Gibbs sampler burns in almost immediately, so we set $B = 10$. We thinned every 40 iterations. Below is a histogram of the burned and thinned samples of m:



- [7 pt / 93 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) The MMSE estimate for m must be a value in the set $\{0, 1, \ldots, n\}$

  (b) The MMSE estimate for m is near 4

  (c) The MAP estimate for m is definitely zero

  (d) If we were to test that $m > 15$, the null hypothesis would likely be rejected

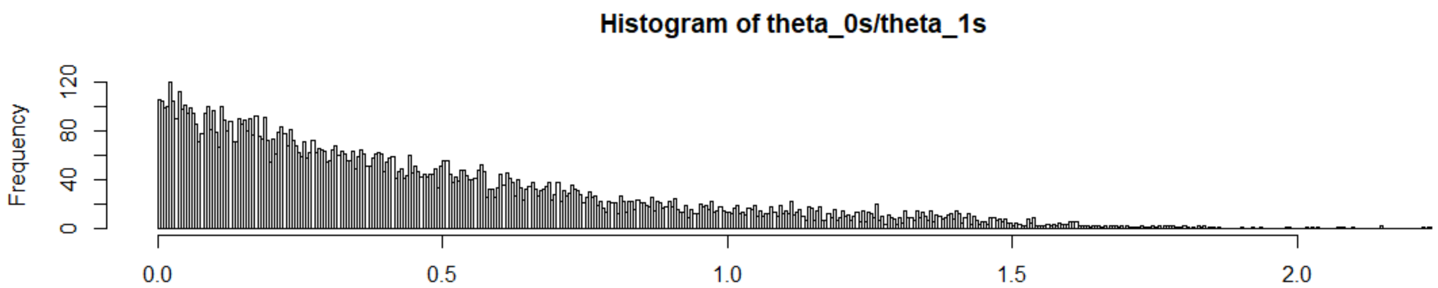
  (e) The 95% credible region for m is $\{0, 1, \ldots, 14\}$ and this means the true value of m is one of the values in the set $\{0, 1, \ldots, 14\}$

  (f) Since there is no mass in $\mathbb{P}(m \mid X)$ above $m = 20$, this means it is likely we burned too soon in our Gibbs chain

  (g) Since there is no mass in $\mathbb{P}(m \mid X)$ above $m = 20$, this means it is likely we did not thin the Gibbs chain appropriately

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

Problem 10 [5min] (and 79min will have elapsed) In our dataset, $n = 30$. Using the three conditional distributions from the previous problem within a systematic sweep Gibbs sampler running for $S = 100,000$ iterations, we find that the Gibbs sampler burns in almost immediately, so we set $B = 10$. We thinned every 40 iterations. We are interested in the difference in probability of getting a hit in the assumed two periods of the players batting. Below is a histogram of the samples of the ratio of θ_0/θ_1



- [3 pt / 96 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) If we were to test $H_a : \theta_0 \neq \theta_1$, the null hypothesis would be rejected since there is little mass in vicinity of 1.0 in the plot above

  (b) If we were to test $H_a : \theta_0 \neq \theta_1$, the null hypothesis would be rejected since there is little mass in vicinity of 0.0 in the plot above

  (c) If we were to test $H_a : \theta_0 \neq \theta_1$, the null hypothesis would be rejected since there is little mass in the right tail

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

Problem 11 [17min] (and 96min will have elapsed) Consider the iid normal model where both the mean and variance are unknown. After a long derivation we find that $\mathbb{P}(\theta, \sigma^2 \mid X) = \text{NormInvGamma}(\mu = \bar{x}, \lambda = n, \alpha = n/2, \beta = (n-1)s^2/2)$.

- [17 pt / 113 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

(a) We used Jeffrey's priors for both the prior on θ and the prior on σ^2

(b) The prior on θ and the prior on σ^2 are the same distribution

(c) The prior we used is an informative prior

(d) It is possible to arrive at this posterior if $\mathbb{P}(\theta) = \mathcal{N}(\mu_0, \tau^2)$ for specific values of the constants μ_0 and τ^2

(e) $\mathbb{P}(\theta, \sigma^2 \mid X) = \mathbb{P}(\theta \mid X, \sigma^2)\,\mathbb{P}(\sigma^2 \mid X)$ where $\mathbb{P}(\sigma^2 \mid X) = \text{InvGamma}(\alpha, \beta)$

(f) $\mathbb{P}(\theta, \sigma^2 \mid X) = \mathbb{P}(\theta \mid X, \sigma^2)\,\mathbb{P}(\sigma^2 \mid X)$ where $\mathbb{P}(\theta \mid X, \sigma^2) = \mathcal{N}(\mu, \lambda)$

(g) $\mathbb{P}(\theta, \sigma^2 \mid X) = \mathbb{P}(\sigma^2 \mid X, \theta)\,\mathbb{P}(\theta \mid X)$ where $\mathbb{P}(\sigma^2 \mid X, \theta)$ has an inverse gamma distribution

(h) $\mathbb{P}(\theta, \sigma^2 \mid X) = \mathbb{P}(\sigma^2 \mid X, \theta)\,\mathbb{P}(\theta \mid X)$ where $\mathbb{P}(\theta \mid X)$ has a normal distribution

(i) You can sample from $\mathbb{P}(\theta, \sigma^2 \mid X)$ exactly by first sampling from $\mathbb{P}(\sigma^2 \mid X)$ to obtain σ^2_{samp} and then sampling θ_{samp} from $\mathbb{P}\left(\theta \mid X, \sigma^2 = \sigma^2_{samp}\right)$ and returning the pair $\left\langle \theta_{samp}, \sigma^2_{samp} \right\rangle$

(j) You can sample from $\mathbb{P}(\theta, \sigma^2 \mid X)$ exactly by first sampling from $\mathbb{P}(\sigma^2 \mid X, \theta = \theta_0)$ to obtain σ^2_{samp} and then sampling θ_{samp} from $\mathbb{P}\left(\theta \mid X, \sigma^2 = \sigma^2_{samp}\right)$ and returning the pair $\left\langle \theta_{samp}, \sigma^2_{samp} \right\rangle$ as long as θ_0 is in the parameter space of the parametric model which in our case is any real number

(k) You can sample from $\mathbb{P}(\theta, \sigma^2 \mid X)$ exactly by first sampling from $\mathbb{P}(\theta \mid X)$ to obtain θ_{samp} and then sampling σ^2_{samp} from $\mathbb{P}(\sigma^2 \mid X, \theta = \theta_{samp})$ and returning the pair $\left\langle \theta_{samp}, \sigma^2_{samp} \right\rangle$

(l) You can sample from $\mathbb{P}(\theta, \sigma^2 \mid X)$ exactly by first sampling from $\mathbb{P}(\theta \mid X, \theta = \sigma^2_0)$ to obtain θ_{samp} and then sampling σ^2_{samp} from $\mathbb{P}(\sigma^2 \mid X, \theta = \theta_{samp})$ and returning the pair $\left\langle \theta_{samp}, \sigma^2_{samp} \right\rangle$ as long as σ^2_0 is in the parameter space of the parametric model which in our case is any positive real number

(m) $\int_0^\infty \mathbb{P}(\theta, \sigma^2 \mid X)\, d\sigma^2$ is a known distribution

(n) $\int_{\mathbb{R}} \int_0^\infty \mathbb{P}(\theta, \sigma^2 \mid X)\, d\sigma^2 d\theta$ is a known distribution

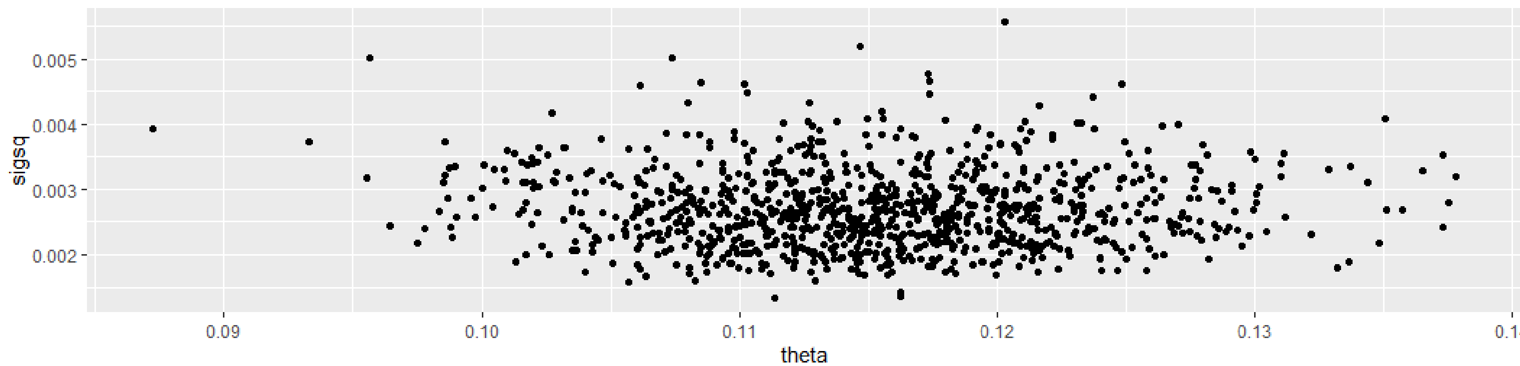
(o) The estimate $\mathbb{E}[\theta \mid X]$ has no shrinkage to the prior expectation

(p) $\mathbb{E}[\theta \mid X] = \mathbb{E}[\theta \mid X, \sigma^2]$ for any value of n and any value of σ^2 in its parameter space

(q) $\mathbb{E}[\sigma^2 \mid X] = \mathbb{E}[\sigma^2 \mid X, \theta]$ for any value of n and any value of θ in its parameter space

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

Problem 12 [9min] (and 105min will have elapsed) Consider the iid normal model where both the mean and variance are unknown. After a long derivation we find that $\mathbb{P}(\theta, \sigma^2 \mid X) = \text{NormInvGamma}(\mu = \bar{x}, \lambda = n, \alpha = n/2, \beta = (n-1)s^2/2)$. We sample $S = 10,000$ realizations from this posterior $\mathcal{S} := \{\langle \theta_1, \sigma_1^2 \rangle, \langle \theta_2, \sigma_2^2 \rangle, \ldots, \langle \theta_S, \sigma_S^2 \rangle\}$. A sample of 1,000 points $\in \mathcal{S}$ is displayed below where θ is on the x-axis and σ^2 is on the y-axis.



- [9 pt / 122 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) A 95% HDR region would be rectangular in shape

  (b) $\hat{\sigma}^2_{\text{MMSE}}$ can be approximated from this plot

  (c) $\hat{\sigma}^2_{\text{MMSE}}$ can be approximated with $\frac{1}{S}\sum_{t=1}^{S} \sigma_t^2$

  (d) A 95% CR for $\mathbb{P}(\sigma^2 \mid X)$ can be approximated from this plot

  (e) A 95% CR for $\mathbb{P}(\sigma^2 \mid X, \theta)$ can be approximated from this plot for every value of θ in the parameter space of θ

  (f) During a test of $\theta > 0.90$, the null hypothesis would likely be rejected

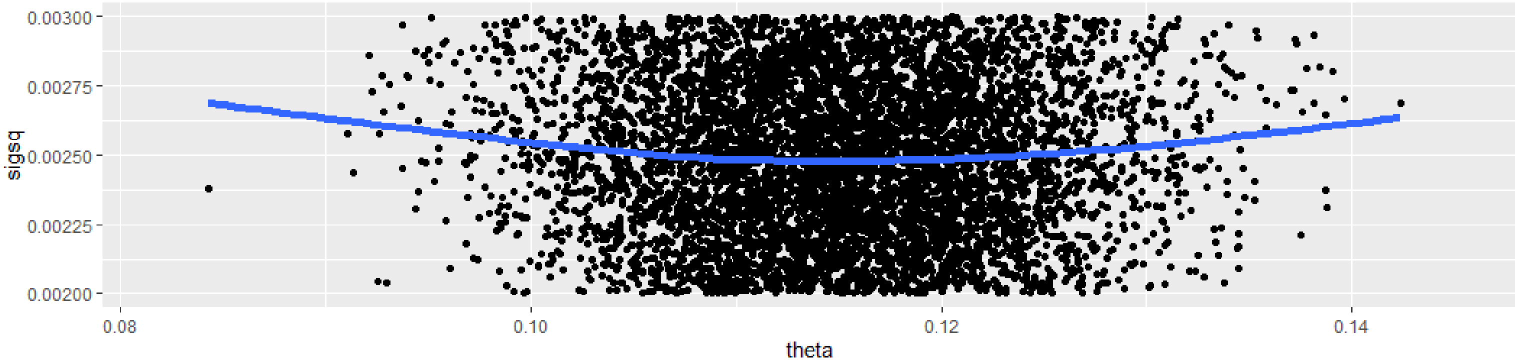
  (g) During a test of $\sigma > 0.045$, the null hypothesis would likely be rejected

  (h) Sampling a future observation x_* given the data can be done as follows: first draw a random sample $\langle \theta_s, \sigma_s^2 \rangle$ from the set $\mathcal{S}$ and then draw x_* via $\texttt{rnorm}(\theta_s, \sqrt{\sigma_s^2})$

  (i) Sampling a future observation x_* given the data can be done as follows: first draw a random sample $\langle \theta_s, \sigma_s^2 \rangle$ from the set $\mathcal{S}$ and then draw x_* via $\texttt{rt.scaled}(n-1, \theta_s, \sqrt{\sigma_s^2})$

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

Problem 13 [5min] (and 110min will have elapsed) Consider the iid normal model where both the mean and variance are unknown. After a long derivation we find that $\mathbb{P}(\theta, \sigma^2 \mid X) = \text{NormInvGamma}(\mu = \bar{x}, \lambda = n, \alpha = n/2, \beta = (n-1)s^2/2)$. We sample $S = 10,000$ realizations from this posterior $\mathcal{S} := \{\langle \theta_1, \sigma_1^2 \rangle, \langle \theta_2, \sigma_2^2 \rangle, \ldots, \langle \theta_S, \sigma_S^2 \rangle\}$. We now also plot a best fit line in blue which tries to approximate $\mathbb{E}[\sigma^2 \mid X, \theta]$ over all values of θ on the plot. To see the line clearer, we zoom in to display σ^2 values between 0.002 and 0.003 (which is different from the previous plot that showed σ^2 values between 0.002 and 0.006):



- [3 pt / 125 pts] Record the letter(s) of all the following that are **true** in general. At least one will be true.

  (a) This line should be flat (i.e. slope zero) but it is curved only due to random variation (if we had more S samples, then we would see that it is truly flat)

  (b) The line should be curved but the curve should've been cupped downwards instead of cupped upwards (if we had more S samples, then we would see the true relationship)

  (c) The line should be curved and the curve has a minimum point at $\theta = \bar{x}$

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