# Math 390 / 650 Spring 2022 Final Examination

Solutions

Professor Adam Kapelner

Monday, May 23

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.

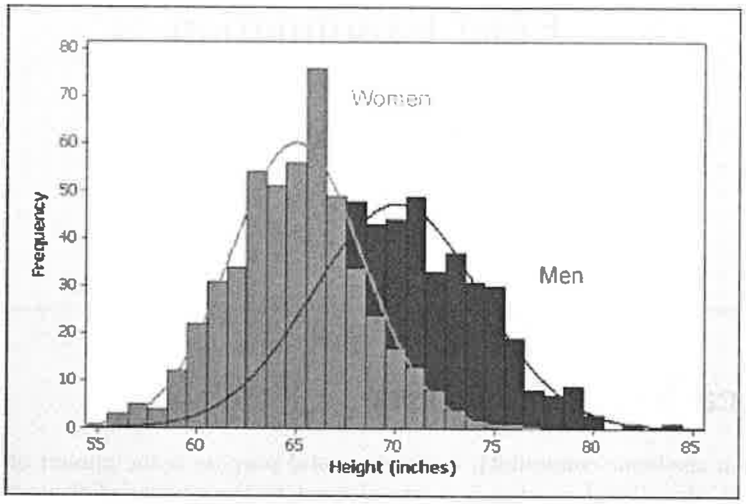
I acknowledge and agree to uphold this Code of Academic Integrity.

______________________________ signature ______________ date

## Instructions

This exam is seventy five 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. 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. NO FOOD but drinks okay. Good luck!

Problem 1 You take a sample of $n = 200$ American people at random and measure their heights. According to this approximation, male height is normally distributed with mean $\theta_M = 70$ inches and $\sigma_M^2 = 4^2$ squared-inches and female height is normally distributed with mean $\theta_F = 65$ inches and $\sigma_F^2 = 3.5^2$ squared-inches.




We can assume this approximation is the truth i.e. $\theta_M, \theta_F, \sigma_M^2, \sigma_F^2$ are known and male heights are $\stackrel{iid}{\sim} \mathcal{N}(\theta_M, \sigma_M^2)$ and female heights are $\stackrel{iid}{\sim} \mathcal{N}(\theta_F, \sigma_F^2)$. We don't know the *proportion* of males in our sample of $n$ people and we'll denote this proportion $\rho$ which is our main target of inference with $\mathbb{S}\text{upp}[\rho] = [0, 1]$. Let $X_1, X_2, \ldots, X_n$ denote the measured heights in the sample.

(a) [3 pt / 3 pts] Write out the explicit PDF of the likelihood of the data given all the parameters: $\theta_M, \theta_F, \sigma_M^2, \sigma_F^2, \rho$.

$$p(X \mid \theta_m, \theta_F, \sigma_m^2, \sigma_F^2, \rho) = \prod_{i=1}^{n} \rho \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{1}{2\sigma_m^2}(x_i-\theta_m)^2} + (1-\rho)\frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{1}{2\sigma_F^2}(x_i-\theta_F)^2}$$

(b) [3 pt / 6 pts] Why would it be difficult to find a closed-form solution for $\hat{\rho}_{\text{MLE}}$ given the likelihood you found in (a)? Write a couple sentences in English to answer.

The likelihood is of the form $\prod_{i=1}^{n} a_i + b_i$ which is difficult to compute a derivative wrt $\rho$ due to the product of sums.

We now "augment the data" by introducing the parameters $I_1, I_2, \ldots, I_n$

$$I_i := \begin{cases} 1 & \text{if the } i\text{th data point is male, coming from the } \mathcal{N}(\theta_M, \sigma_M^2) \text{ distribution} \\ 0 & \text{if the } i\text{th data point is female, coming from the } \mathcal{N}(\theta_F, \sigma_F^2) \text{ distribution} \end{cases}$$

(c) [4 pt / 10 pts] Write out the explicit PDF of the likelihood of the data given all the parameters and the parameters under data augmentation $\theta_M, \theta_F, \sigma_M^2, \sigma_F^2, \rho, I_1, I_2, \ldots, I_n$. Simplify as much as possible.

$$p(X \mid \theta_m, \theta_F, \sigma_m^2, \sigma_F^2, \rho, I_1, \ldots, I_n) = \prod_{i=1}^{n} \left(\rho \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{1}{2\sigma_m^2}(X_i - \theta_m)^2}\right)^{I_i} \left((1-\rho)\frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{1}{2\sigma_F^2}(X_i - \theta_F)^2}\right)^{1-I_i}$$

$$= (2\pi)^{-n/2} \rho^{\Sigma I_i} (1-\rho)^{n - \Sigma I_i} e^{-\frac{1}{2\sigma_m^2} \Sigma I_i (X_i - \theta_m)^2} e^{-\frac{1}{2\sigma_F^2} \Sigma (1 - I_i)(X_i - \theta_F)^2} (\sigma_m^2)^{-\frac{\Sigma I_i}{2}} (\sigma_F^2)^{-\frac{n - \Sigma I_i}{2}}$$

(d) [3 pt / 13 pts] Since $\theta_M, \theta_F, \sigma_M^2, \sigma_F^2$ are assumed known constants, we do not need to specify a prior for them. Specify the Laplace prior for $\rho$ explicitly. Do not simply write $\mathbb{P}(\rho) \propto 1$. You need to write $\mathbb{P}(\rho) =$ a legal distribution.

$$p(\rho) = \text{Beta}(1,1) = U(0,1)$$

(e) [3 pt / 16 pts] Specify the Laplace prior for all the $I_i$'s explicitly. Do not simply write $\mathbb{P}(I_i) \propto 1$. You need to write $\forall i \;\; \mathbb{P}(I_i) =$ a legal distribution.

$$\forall i \quad p(I_i) = \text{Bern}\left(\tfrac{1}{2}\right)$$

(f) [4 pt / 20 pts] Regardless of what you wrote for the previous two questions, you can now assume that $\mathbb{P}(\rho, I_1, I_2, \ldots, I_n) \propto 1$. Find the kernel of the posterior as best as possible $k(\rho, I_1, I_2, \ldots, I_n \mid X_1, X_2, \ldots, X_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2)$.

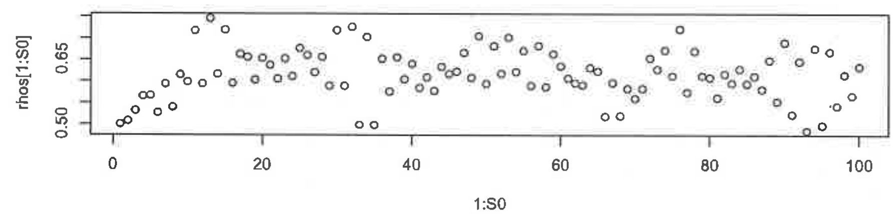
$$p(\rho, I_1, \ldots, I_n \mid \theta_m, \theta_F, \sigma_m^2, \sigma_F^2, X) \propto p(X \mid \text{---}) \propto \rho^{\Sigma I_i}(1-\rho)^{n-\Sigma I_i} e^{-\frac{1}{2\sigma_m^2}\Sigma I_i (X_i - \theta_m)^2} e^{-\frac{1}{2\sigma_F^2}\Sigma(1-I_i)(X_i-\theta_F)^2} (\sigma_m^2)^{-\frac{\Sigma I_i}{2}} (\sigma_F^2)^{-\frac{n-\Sigma I_i}{2}}$$

$$\propto \rho^{\Sigma I_i}(1-\rho)^{n-\Sigma I_i} e^{-\frac{1}{2\sigma_m^2}\Sigma I_i(X_i-\theta_m)^2 + \frac{1}{2\sigma_F^2}\Sigma I_i (X_i - \theta_F)^2} (\sigma_m^2)^{-\frac{\Sigma I_i}{2}} \sigma_F^{2\,\frac{\Sigma I_i}{2}}$$

$$= k(\rho, I_1, \ldots, I_n \mid \theta_m, \theta_F, \sigma_m^2, \sigma_F^2, X)$$

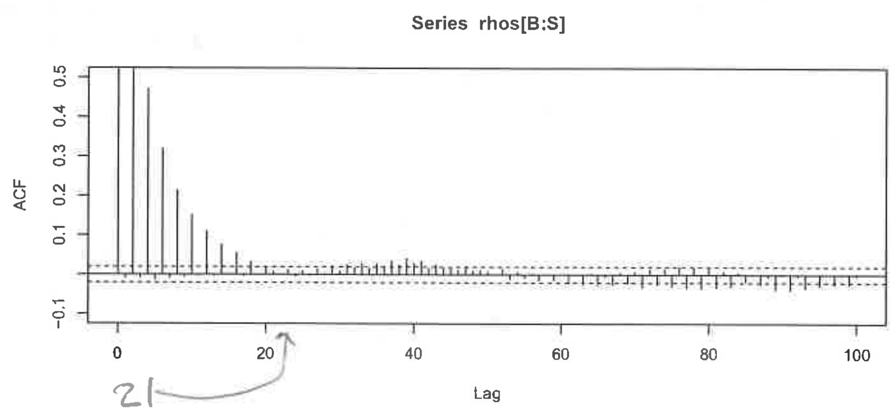
(g) [4 pt / 24 pts] The kernel you found in the previous example is not any known distribution that you know how to sample from. Thus we will employ a Gibbs sampler. Find the conditional distribution $\mathbb{P}(\rho \mid I_1, I_2, \ldots, I_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2, X_1, X_2, \ldots, X_n)$. It will be a known distribution. Compute its parameters.

$$k(\rho \mid I_1 \ldots I_n, \theta_m, \theta_F, \sigma_m^2, \sigma_F^2, X) = \rho^{\sum I_i}(1-\rho)^{n-\sum I_i} \propto \text{Beta}\left(\sum I_i + 1,\ n - \sum I_i + 1\right)$$




(h) [1 pt / 25 pts] Above is the first 100 samples from the Gibbs sampler's conditional distribution $\mathbb{P}(\rho \mid I_1, I_2, \ldots, I_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2, X_1, X_2, \ldots, X_n)$. At what approximate iteration number would you burn?

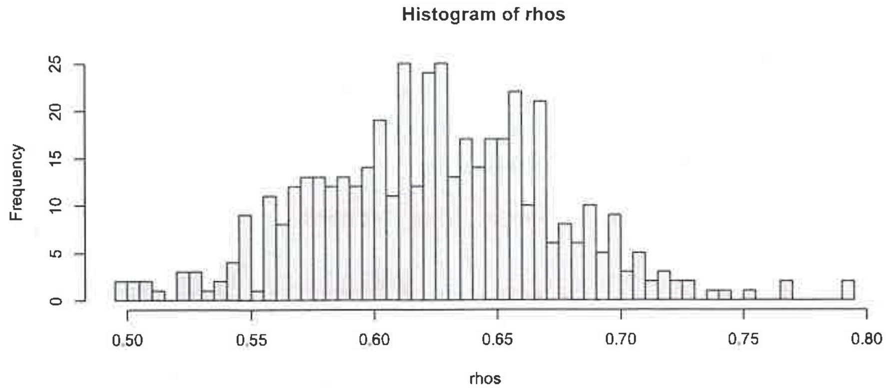
15-20




(i) [2 pt / 27 pts] Above is an autocorrelation plot of post-burned samples from the Gibbs sampler's conditional distribution $\mathbb{P}(\rho \mid I_1, I_2, \ldots, I_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2, X_1, X_2, \ldots, X_n)$. At what approximate iteration number would you thin?

21 or 22

but not higher!




(j) [2 pt / 29 pts] Above is the post-burned and thinned samples from the Gibbs sampler's conditional distribution $\mathbb{P}(\rho \mid I_1, I_2, \ldots, I_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2, X_1, X_2, \ldots, X_n)$. Estimate $\hat{\hat{\rho}}_{\text{MMAE}}$.

$0.625$

(k) [2 pt / 31 pts] Provide an estimated $CR_{\rho,95\%}$.

$[0.53, 0.74]$

(l) [5 pt / 36 pts] Test the theory that this sample has an equal number of men and women. Show all work and be explicit about your assumptions. Write a concluding statement.

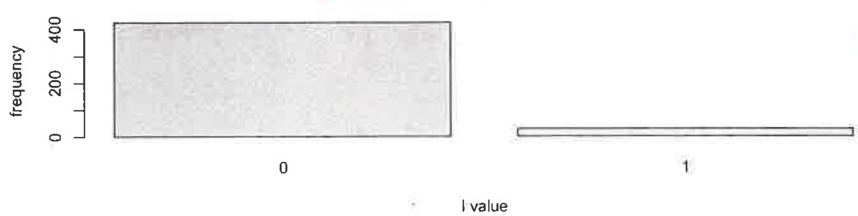
Let $\delta = 0.02 \Rightarrow H_0: \rho \in [0.5 \pm \delta] = [0.48, 0.52]$,
$H_a: \rho \notin [0.5 \pm \delta] = [0, 0.48) \cup (0.52, 1]$

Let $\alpha = 5\%$.

$p_{val} := P(H_0 \mid X) = P(\rho \in [0.48, 0.52] \mid X) \approx 2.5\% < 5\% \Rightarrow$ Reject $H_0$;

estimate from illustration above

This sample is likely not equal in proportion among men & women.




(m) [2 pt / 38 pts] Above is the post-burned and thinned samples from the Gibbs sampler's conditional distribution $\mathbb{P}(I_1 \mid \rho, I_2, \ldots, I_n, \theta_M, \theta_F, \sigma_M^2, \sigma_F^2, X_1, X_2, \ldots, X_n)$. Estimate the value of $\hat{\hat{I}}_{1,\text{MMSE}}$ to two digits.

$0.05$

(n) [1 pt / 39 pts] Estimate the probability that the first subject is a male. 5%

(o) [4 pt / 43 pts] Explain a step-by-step method for drawing $X_*$, a new observation from the random variable model that produced the $X_1, \ldots, X_n$ data observations. Use the notation found in Table 2 if applicable.

① Draw an iid sample from the burned-and-thinned chain, $[\theta_M^S, \theta_F^S, \sigma_M^{2\,S}, \sigma_F^{2\,S}, \rho^S, I_1^S, \ldots, I_n^S]$

② Draw $I_* \sim \text{Bern}(\rho^S)$ using rbinom$(1, \rho^S)$

③ If $I_* = 1$, draw $X_*$ from $\mathcal{N}(\theta_M^S, \sigma_M^{2\,S})$ via rnorm$(\theta_M^S, \sigma_M^{2\,S})$

or if $I_* = 0$, draw $X_*$ from $\mathcal{N}(\theta_F^S, \sigma_F^{2\,S})$ via rnorm$(\theta_F^S, \sigma_F^{2\,S})$

Problem 2 Human birth weight is known to be normally distributed.

We measure $\{8.28, 7.65, 8.88, 7.80, 7.58, 6.96, 7.44, 7.34, 6.89, 6.97\}$, a sample of $n = 10$ birth weights measured in pounds. Its associated sample statistics are: $\bar{x} = 7.58$ and $s^2 = 0.39$. We cannot assume we know the true mean nor the true variance of the random variable that produced this data set. Assume Jeffrey's prior going forward.

(a) [2 pt / 45 pts] Find $\hat{\hat{\theta}}_{MMAE}$ to the nearest two digits.

$$P(\theta|X) = T_{n-1}\left(\bar{x}, \frac{s}{\sqrt{n}}\right) = T_9\left(7.58, \overbrace{\frac{\sqrt{0.39}}{\sqrt{10}}}^{0.197}\right) \Rightarrow \hat{\hat{\theta}}_{MMAE} = \boxed{7.58}$$

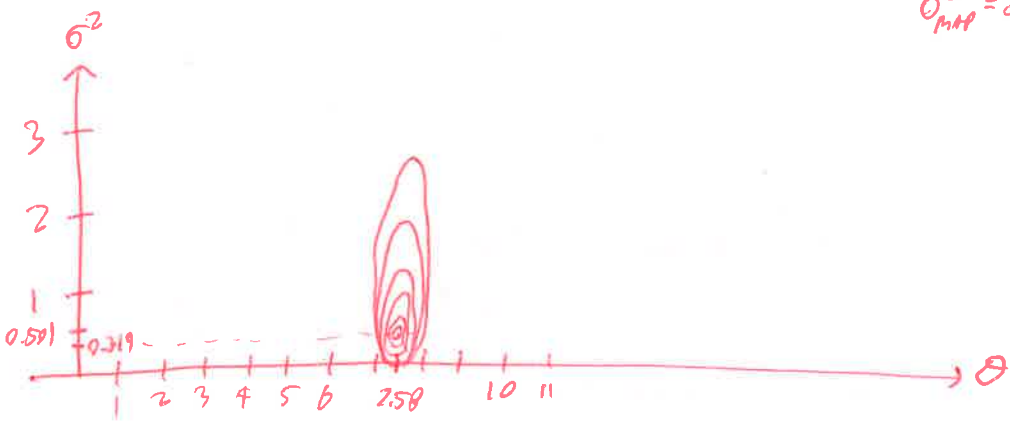
(b) [3 pt / 48 pts] Find $\hat{\hat{\sigma}}^2_{MMSE}$ to the nearest two digits.

$$P(\sigma^2|X) = \text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) = \text{InvGamma}\left(4.5, \overbrace{\frac{9 \cdot 0.39}{2}}^{1.755}\right)$$

$$\hat{\sigma}^2_{MMSE} := E[\sigma^2|X] = \frac{\beta}{\alpha - 1} = \frac{1.755}{4.5-1} = \boxed{0.501}$$

(c) [4 pt / 52 pts] Plot the bivariate density of $\mathbb{P}(\theta, \sigma^2 \mid X)$ as best as you can.

$\hat{\sigma}^2_{MAP} = \frac{\beta}{\alpha+1} = \frac{1.755}{4.5+1} = 0.319$




(d) [4 pt / 56 pts] Compute the Bayesian $p$-val for the theory that this sample's mean is underweight i.e. $H_a : \theta < 7.72$lb. $\Rightarrow H_0: \theta \le 7.72lb$

$p_{val} = P(H_0 \mid x) = P(\theta \le 7.72lb \mid x) = pt.scaled(7.72, 9, 7.58, 0.197)$

(e) [4 pt / 60 pts] Find an expression for the probability the next child in this sample will be underweight.

$P(X_* \mid x) = T_{n-1}\left(\bar{x}, \sqrt{\frac{n+1}{n} s^2}\right) = T_9\left(7.58, \underbrace{\sqrt{\frac{11}{10} 0.39}}_{0.655}\right)$

$P(X_* < 7.72 \mid x) = pt.scaled(7.72, 9, 7.58, 0.655)$

**Problem 3** Below are some pure computation problems based on theory from this class. Solve for them using precise mathematical notation (no approximations with decimals). Simplify if possible.

(a) [4 pt / 64 pts] $\int_0^\infty \underbrace{x^{\overset{-16-1}{-17}} e^{-\pi/x}}_{\text{Kernel of InvGamma}(16,\pi)} dx = \frac{\Gamma(16)}{\pi^{16}}$

(b) [3 pt / 67 pts] $B(4,8)$ $= \frac{\Gamma(4)\Gamma(8)}{\Gamma(4+8)} = \frac{\Gamma(4)\Gamma(8)}{\Gamma(12)} = \frac{3!\,7!}{11!} = \frac{\cancel{2}\cdot\cancel{3}}{11\cdot 10\cdot \cancel{9}\cdot \cancel{8}} = \frac{1}{1320}$

(c) [5 pt / 72 pts] $\displaystyle\sum_{x=0}^{n} \frac{\Gamma(x+\alpha)\,\Gamma(n-x+\beta)}{x!(n-x)!}$ where $n \in \mathbb{N}$ and $\alpha > 0, \beta > 0$

Kernel of BetaBinom$(n,\alpha,\beta) = \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)}$

$= \frac{1}{B(\alpha,\beta)} \frac{n!}{x!(n-x)!} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$

$= \frac{B(\alpha,\beta)\,\Gamma(n+\alpha+\beta)}{n!}$

(d) [5 pt / 77 pts] $\displaystyle\int_{\mathbb{R}} \left((x-\pi)^2+2\right)^{-n/2} dx$ where $n \in \mathbb{N}$

$(n-1)+1$

Kernel of $T_{n-1}(\theta, s) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{\pi(n-1)}\,s}\left(1+\frac{1}{n-1}\left(\frac{x-\theta}{s}\right)^2\right)^{-n/2}$

$\left((x-\pi)^2+2\right)^{-n/2}\cdot\left(\frac{1}{2}\right)^{-n/2} = \left(1+\left(\frac{x-\pi}{\sqrt{2}}\right)^2\right)^{-n/2} = \left(1+\frac{1}{n-1}\left(\frac{x-\pi}{\sqrt{\frac{2}{n-1}}}\right)^2\right)^{-n/2} \Rightarrow \theta = \pi,\ s = \sqrt{\frac{2}{n-1}}$

$\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{\pi(n-1)}\sqrt{\frac{2}{n-1}}} \cdot 2^{-n/2} = \boxed{\frac{\Gamma(\frac{n}{2})\,2^{-\frac{n+1}{2}}}{\Gamma(\frac{n-1}{2})\sqrt{\pi}}}$

**Problem 4** This question is about ratings on youtube. Each video which is voted on is either up-voted or down-voted. A video rating is the total number of thumbs up ratings over the total number of ratings. For example if a movie gets 5080 thumbs up and 960 thumbs down ratings, then it has a 5080/(5080 + 960) = 84.1% approval rating.

But there is a question: how should we order videos by *true* approval rating $\theta \in (0,1)$? For example, here is a table of four videos we wish to order:

| Video Name | # Up votes | # Down votes | $n$ | Approval Rating |
|---|---|---|---|---|
| A | 0 | 1 | 1 | 0.0% |
| B | 4 | 0 | 4 | 100.0% |
| C | 25 | 2 | 27 | 92.6% |
| D | 5080 | 960 | 6040 | 84.1% |

Table 1: Table of videos with their youtube ratings.

(a) [1 pt / 78 pts] Order the movies in Table 1 by name from best to worst using the MLE estimate of its true approval rating. Your answer must be in the format "$A > B > C > D$" where $A$ is the highest-rated and $D$ is the lowest-rated.

$B > C > D > A$

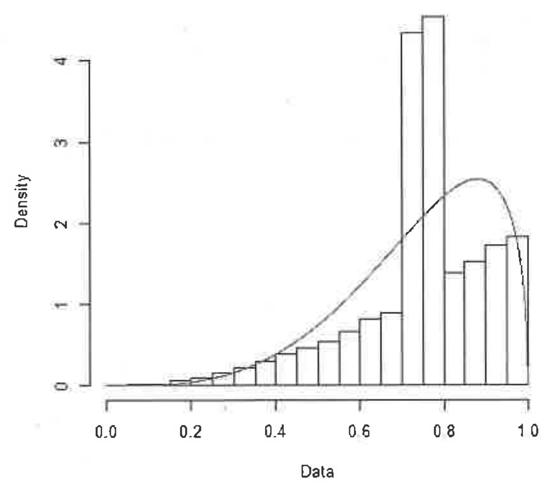
(b) [3 pt / 81 pts] Why is what you did in (a) a poor way to order the four movies?

$\hat{\theta}_{A,MLE}, \hat{\theta}_{B,MLE}$ are unstable due to low sample size

(c) [1 pt / 82 pts] We are now going to use some previous data to create a prior for the true approval rating. What is this kind of procedure is called (two words)?

empirical Bayes

Below is a histogram of the approval ratings of $n_0 = 30,000$ videos of which there are more than 200 votes each. The curve displayed atop the histogram is the best fit beta density. I used R's `fitdistrplus` package which creates a fit via the MLE's of $\alpha$ and $\beta$. I include estimates in output from R below the plot.




```
Parameters:
        estimate Std. Error
shape1 4.283762 0.03567291
shape2 1.442157 0.01073980
```

(d) [3 pt / 85 pts] Besides the fact that the curve does not fit the empirical distribution (given by the histogram), what is wrong with the estimates of $\alpha$ and $\beta$ given above? Hint: think about pseudocounts.

$n_0 = 4.28 + 1.44 = 5.72 << 30{,}000$. This prior is much weaker than expected given the massive amount of prior data.

(e) [3 pt / 88 pts] Given that a movie has $n$ total votes and $x$ of those are thumbs up, what is the posterior distribution of the true approval rating $\theta$ given the data coupled with the prior constructed above in the illustration before question (d)?

$P(\theta|x) = Beta(x + 4.28, n - x + 1.44) \Rightarrow \hat{\theta}_{MMSE} = \frac{x+4.28}{n+5.72}$

(f) [4 pt / 92 pts] Order the movies in Table 1 from best to worst using the Bayesian estimate which minimizes mean squared error. Your answer must be in the format "$A > B > C > D$" where $A$ is the highest-rated and $D$ is the lowest-rated. Compute explicitly. No credit unless work is shown.

$\hat{\theta}_{A,MMSE} = \frac{0+4.28}{1+5.72} = 0.637$, $\hat{\theta}_{B,MMSE} = \frac{4+4.28}{4+5.72} = 0.852$,

$\hat{\theta}_{C,MMSE} = \frac{25+4.28}{27+5.72} = 0.895$, $\hat{\theta}_{D,MMSE} = \frac{5080+4.28}{6040+5.72} = 0.841$

$\Rightarrow C > B > D > A$

Problem 5 Continuing the question from before, there is reason to believe that the average approval rating is trending over time. To test this, we sample the same number $n$ samples every day for $t \in \{1, \dots, T\}$ days and assume that $X_t \overset{ind}{\sim} \text{Binomial}(n, \theta_t)$ where $\theta_t := \beta_0 + \beta_1 t$.

The likelihood is: $\mathbb{P}(X_1, \dots, X_T \mid n, T, \beta_0, \beta_1) = \prod_{t=1}^{T} \binom{n}{x_t} (\beta_0 + \beta_1 t)^{x_t} (1 - \beta_0 - \beta_1 t)^{n - x_t}$.

We'll assume Laplace priors for $\beta_0$ and $\beta_1$ i.e. $\mathbb{P}(\beta_0, \beta_1) \propto 1$ and that $n$ is known.

(a) [3 pt / 95 pts] Find $k(\beta_0 \mid \beta_1, X_1, \dots, X_T, n, T)$.

$\prod_{t=1}^{T} (\beta_0 + \beta_1 t)^{X_t} (1 - \beta_0 - \beta_1 t)^{n - X_t}$

(b) [3 pt / 98 pts] Find $k(\beta_1 \mid \beta_0, X_1, \dots, X_T, n, T)$.

$\prod_{t=1}^{T} (\beta_0 + \beta_1 t)^{X_t} (1 - \beta_0 - \beta_1 t)^{n - X_t}$

(c) [2 pt / 100 pts] If you were to create a Gibbs sampler using both $k(\beta_0 \mid \beta_1, X_1, \dots, X_T, n, T)$ and $k(\beta_1 \mid \beta_0, X_1, \dots, X_T, n, T)$, what is the name of one algorithm that could be used when sampling $\beta_0$?

grid sampling