Math 341 / 650 Spring 2022
Midterm Examination ~~One~~ Two

Professor Adam Kapelner

Wednesday, April 27

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 **one** page (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 Measuring the speed of light using a laser is known to be unbiased i.e. the mean is $\theta = 299792458$ m/s.

$\Sigma(x_i-\theta)^2 =$
6.761×10^{13}

However, there is variance σ^2 which we seek to understand. We do 3 experiments and find measurements of the speed of light to be $x_1 = 303884490$, $x_2 = 296562208$ and $x_3 = 306150981$. Assume these measurements are $\overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$.

(a) [4 pt / 4 pts] Find $\hat{\sigma}^2_{MLE}$.

$$= \frac{\Sigma(x_i-\theta)^2}{n} = 2.254 \times 10^{13}$$

(b) [4 pt / 8 pts] What is the Jeffrey's prior for $\sigma^2 \mid \theta$? Inv Gamma $(0, 0)$.

(c) [4 pt / 12 pts] Using Jeffrey's prior for $\sigma^2 \mid \theta$, find the posterior distribution $\mathbb{P}(\sigma^2 \mid X, \theta)$.

$$\text{InvGamma}\left(\frac{n}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2}\right) = \text{InvGamma}\left(1.5,\ 3.381 \times 10^{13}\right)$$

(d) [6 pt / 18 pts] Write the PDF of $\mathbb{P}(\sigma^2 \mid X, \theta)$ explicitly.

$$P(\sigma^2 \mid X, \theta) = \frac{(3.381 \times 10^{13})^{1.5}}{\Gamma(1.5)} (\sigma^2)^{-2.5} e^{-\frac{3.381 \times 10^{13}}{\sigma^2}}$$

(e) [6 pt / 24 pts] Compute $\hat{\sigma}^2_{MAP}$. $= \frac{\beta}{\alpha+1}$

$$= \frac{3.381 \times 10^{13}}{1.5+1} = 1.352 \times 10^{13}$$

(f) [6 pt / 30 pts] Express the $CR_{\sigma^2, 90\%}$ using Table 1.

$$\left[\text{qinvgamma}(.05, 1.5, 3.381 \times 10^{13}),\ \text{qinvgamma}(.95, 1.5, 3.381 \times 10^{13})\right]$$

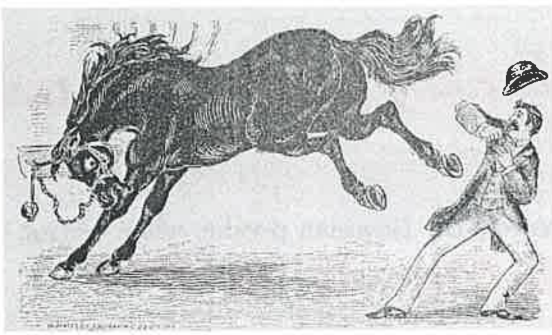
(g) [6 pt / 36 pts] Express the probability the next experiment will have a measurement different from the truth by more than 1,000 m/s using Table 1.

$$P(X_* \mid X, \theta) = T_{n+n_0}\left(\theta, \frac{n\hat{\sigma}^2_{MLE} + n_0\sigma_0^2}{n+n_0}\right) = T_n\left(\theta, \hat{\sigma}^2_{MLE}\right) = T_3(299792458, 2.254 \times 10^{13})$$

$$P(|X_* - 299792458| > 1000 \mid X, \theta) = P(X_* > 299793458) + P(X_* < 299791458)$$

$$= \text{pt.scaled}\Big(299793458, 3, 299792458, \underbrace{\sqrt{2.254 \times 10^{13}}}_{4747282}\Big) - \text{pt.scaled}\Big(299791458, 3, 299792458, \underbrace{\sqrt{2.254 \times 10^{13}}}_{4747282}\Big)$$

Problem 2 Ladislaus Josephovich Bortkiewicz was a Polish economist and statistician who published a book in 1898 about the Poisson distribution. A famous dataset in this book is the "Prussian horse-kicking data" famous.



The data gave the number of soldiers killed by being kicked by a horse each year in each of 16 different cavalry corps over a 20-year period. Bortkiewicz showed that those numbers followed a Poisson distribution and hence we can assume the following data for the general corps is $\overset{iid}{\sim}$ Poisson (θ).

$\sum x_i = 16$

Year (18–)	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
# kicks	0	2	2	1	0	0	1	1	0	3	0	2	1	0	0	1	0	1	0	1

(a) [4 pt / 40 pts] In this example, what is the interpretation of the parameter θ? Answer using English sentence(s).

expected/mean # of soldiers in the general corps killed per year

(b) [4 pt / 44 pts] What is Haldane's prior for θ? Gamma $(0, 0)$

(c) [2 pt / 46 pts] Is Haldane's prior for θ proper? Yes / no.

(d) [4 pt / 50 pts] Using Haldane's prior for θ, find the posterior distribution $\mathbb{P}(\theta \mid X)$.

$$\text{Gamma}\left(\Sigma x_i, n\right) = \text{Gamma}(16, 20)$$

(e) [2 pt / 52 pts] Is $\mathbb{P}(\theta \mid X)$ proper? Yes / no.

(f) [4 pt / 56 pts] Compute $\hat{\hat{\theta}}_{MMSE}$.

$$\hat{\hat{\theta}}_{MMSE} = \frac{\alpha + \Sigma x_i}{\beta + n} = \frac{\Sigma x_i}{n} = \frac{16}{20} = 0.8$$

(g) [5 pt / 61 pts] Express the $CR_{\theta,95\%}$ below using Table 1.

$$\left[\text{qgamma}(0.025, 16, 20),\ \text{qgamma}(0.975, 16, 20)\right]$$

(h) [7 pt / 68 pts] Express the Bayesian p-value when testing the hypothesis that $\theta > 1$ using Table 1.

$H_a: \theta > 1 \Rightarrow H_0: \theta \leq 1$

$$p_{val} = P(H_0 \mid x) = P(\theta \leq 1 \mid x) = \text{pgamma}(1, 16, 20)$$

(i) [7 pt / 75 pts] Express the probability the number of kicks in 1895 is three or more using Table 1.

$$P(X_* \geq 3 \mid X) = 1 - P(X_* \leq 2 \mid X) = 1 - \text{pnbinom}\left(2, 16, \tfrac{20}{21}\right)$$

0.952

$$P(X_* \mid X) = \text{ExtNegBin}(r, p) = \text{ExtNegBin}\left(16, \tfrac{20}{21}\right)$$

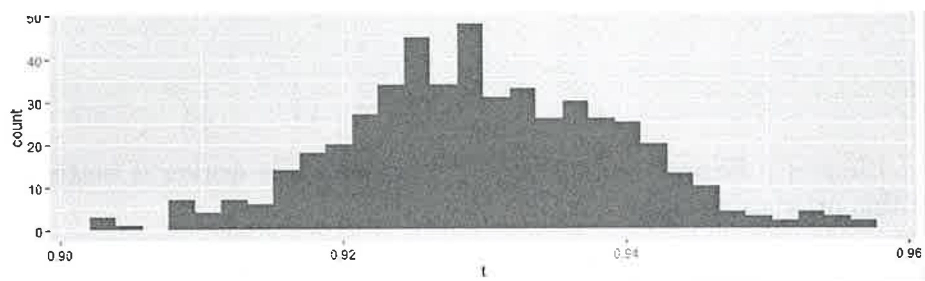
where $r = \Sigma x_i + \alpha = 16$,

$$p = \frac{\beta + n}{\beta + n + 1} = \frac{20}{21}$$

Problem 3 Every worker at the chocolate factory must wrap the chocolate desserts.

Each dessert takes a variable amount of time to wrap. We can assume since wrapping has many factors summed, an $\overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ is a good model where σ^2 is known to be 0.01 seconds-squared. However, every worker has a different θ.

A new worker is hired and we naturally want to see how fast this worker wraps on average. For the first three wrappings, the worker does it in 0.23 seconds, 0.21 seconds and 0.24 seconds for an average of $\bar{x} = 0.2267$ seconds. This is remarkably fast relative to other workers so it is unrealistic to assume this is a good estimate of this new worker's true θ. Here are other worker's lifetime averages:



Fitting a normal distribution to this data and using maximum likelihood, we find the average worker lifetime wrapping time average is 0.930 seconds with a variance of 0.0000927 seconds-squared. We would like to use the empirical Bayes estimation to draw inference for the new worker's θ.

(a) [5 pt / 80 pts] If we use the normal fit of other workers' lifetime averages to create a prior, this prior will be equivalent to how many pseudoobservations n_0?

$$p(\theta|\sigma^2) = N\left(\mu_0, \underbrace{\frac{\sigma^2}{n_0}}_{\tau^2}\right) = N\left(0.930, \frac{0.01}{n_0}\right) = N(0.930, 0.0000927)$$

$$0.0000927 = \frac{0.01}{n_0} \Rightarrow n_0 = \frac{0.01}{0.0000927} = 107.9$$

(b) [2 pt / 82 pts] Is this prior uninformative? Yes / no.

(c) [5 pt / 87 pts] Compute $\hat{\hat{\theta}}_{MMAE}$ under this prior.

$$p(\theta \mid X, \sigma^2) = N\left(\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right) = N\left(\frac{n\bar{x}+n_0\mu_0}{n+n_0}, \frac{\sigma^2}{n+n_0}\right)$$

$$\frac{n\bar{x}+n_0\mu_0}{n+n_0} = \frac{3 \cdot 0.2267 + 107.9 \cdot 0.930}{3+107.9} = 0.911$$

(d) [5 pt / 92 pts] Calculate the shrinkage ρ for this $\hat{\hat{\theta}}_{MMAE}$ estimate under this prior.

$$\rho = \frac{n_0}{n+n_0} = \frac{107.9}{3+107.9} = 0.973$$

(e) [8 pt / 100 pts] Express the probability that this new worker is better than average under this prior using Table 1.

$\cancel{P(X_u \geq 0.93 \mid X) = 1 - P(X_u \leq 0.93 \mid X) = 1 - pnorm(0.93, 0.911}$ [illegible]

$$P(X_u \leq 0.93 \mid X) = pnorm\left(0.93, 0.911, \underbrace{\sqrt{0.01009}}_{0.1004}\right)$$

$$p(X_u \mid X, \sigma^2) = N\left(\frac{n\bar{x}+n_0\mu_0}{n+n_0}, \frac{\sigma^2}{n+n_0}+\sigma^2\right)$$

$$= N\left(0.911, \underbrace{\frac{0.01}{3+107.9}+0.01}\right) \quad 0.01009$$

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)
T (scaled)	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 exam with their arguments. 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.