

Math 341 / 650 Spring 2018

Midterm Examination Two

Johnson

Professor Adam Kapelner

Monday, April 16, 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 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 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. Box in your final answers. Good luck!

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	<code>qbeta(p, α, β)</code>	<code>d-(x, α, β)</code>	<code>p-(x, α, β)</code>	<code>r-(α, β)</code>
betabinomial	<code>qbetabinom(p, n, α, β)</code>	<code>d-(x, n, α, β)</code>	<code>p-(x, n, α, β)</code>	<code>r-(n, α, β)</code>
betanegativebinomial	<code>qbeta_nbinom(p, r, α, β)</code>	<code>d-(x, r, α, β)</code>	<code>p-(x, r, α, β)</code>	<code>r-(r, α, β)</code>
binomial	<code>qbinom(p, n, θ)</code>	<code>d-(x, n, θ)</code>	<code>p-(x, n, θ)</code>	<code>r-(n, θ)</code>
exponential	<code>qexp(p, θ)</code>	<code>d-(x, θ)</code>	<code>p-(x, θ)</code>	<code>r-(θ)</code>
gamma	<code>qgamma(p, α, β)</code>	<code>d-(x, α, β)</code>	<code>p-(x, α, β)</code>	<code>r-(α, β)</code>
geometric	<code>qgeom(p, θ)</code>	<code>d-(x, θ)</code>	<code>p-(x, θ)</code>	<code>r-(θ)</code>
inversegamma	<code>qinvgamma(p, α, β)</code>	<code>d-(x, α, β)</code>	<code>p-(x, α, β)</code>	<code>r-(α, β)</code>
negative-binomial	<code>qnbinom(p, r, θ)</code>	<code>d-(x, r, θ)</code>	<code>p-(x, r, θ)</code>	<code>r-(r, θ)</code>
normal (univariate)	<code>qnorm(p, θ, σ)</code>	<code>d-(x, θ, σ)</code>	<code>p-(x, θ, σ)</code>	<code>r-(θ, σ)</code>
poisson	<code>qpois(p, θ)</code>	<code>d-(x, θ)</code>	<code>p-(x, θ)</code>	<code>r-(θ)</code>
T (standard)	<code>qt(p, ν)</code>	<code>d-(x, ν)</code>	<code>p-(x, ν)</code>	<code>r-(ν)</code>
uniform	<code>qunif(p, a, b)</code>	<code>d-(x, a, b)</code>	<code>p-(x, a, b)</code>	<code>r-(a, b)</code>

Table 1: Functions from R (in alphabetical order) that can be used on this exam. The hyphen in columns 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 1 This question is about car purchasing.



It is a dirty secret in car dealerships that everyone gets a slightly different price based on the negotiations that go on behind closed doors. Here are 20 sale prices of new 2018 Honda Accords at a car dealership in Queens:

31700 32400 31100 33500 30700 32700 31100 31700 33500 32900
32900 32200 32400 33100 31200 33000 33400 34600 31700 33100

Computing the sample average and standard deviation and rounding, we get $\bar{x} = \$33,000$ and $s = \$900$. We'll assume that the final negotiated price is distributed normally because it's most likely the sum of many different negotiation factors.

- (a) [5 pt / 5 pts] Given an uninformative prior for θ and assuming $\sigma = s$, what is the posterior for θ given the data above? Provide numerical values for the parameters.

$$P(\theta) \propto 1 \Rightarrow \theta | x \sim N\left(\bar{x}, \left(\frac{s}{\sqrt{n}}\right)^2\right) = N\left(33,000, \left(\frac{900}{\sqrt{20}}\right)^2\right) = N(33,000, 201^2)$$

(b) [3 pt / 8 pts] Find $\hat{\theta}_{\text{MAE}}$.

$$\hat{\theta}_{\text{MAE}} = \$33,000$$

(c) [5 pt / 13 pts] Compute the shrinkage factor ρ for the point estimate you computed for θ in (b).

$$\rho = 0$$

(d) [8 pt / 21 pts] Given this data and your prior, what is the probability you pay more than \$35,200 for your car? Estimate the numerical probability (do *not* express your final answer using functions from Table 1).

$$X^* | X \sim N\left(\bar{x}, \left(\sqrt{\frac{\sigma^2}{n} + \sigma^2}\right)^2\right) = N(33,000, 122^2)$$

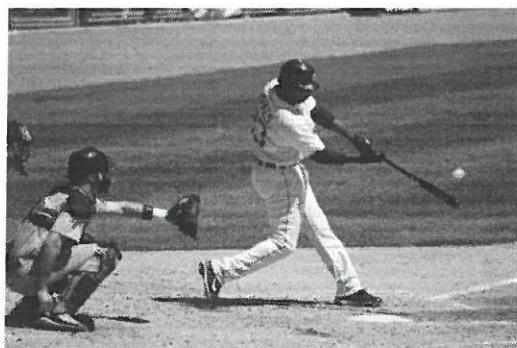
$$\sqrt{\frac{100^2}{20} + 100^2}$$

$$\sqrt{11}$$

$$\sqrt{850500}$$

$$P(X^* > 35,200 | X) = P\left(\frac{X^* - 33,000}{122} > \frac{35,200 - 33,000}{122}\right) = P(Z > 2.38) \approx 1\%$$

Problem 2 This question is about “batting averages” in baseball.



Every hitter's *sample* batting average (BA) is defined as:

$$BA := \frac{\text{sample \# of hits}}{\text{sample \# of at bats}}$$

Let a hitter's *true* batting average be called θ . Each player has a different θ but we focus in this problem on one specific player.

We assume that each at bat (for any player) are conditionally $\overset{iid}{\sim}$ based on the players' true batting average, θ . So if a player has n at bats, then each successful hit in each at bat can be modeled via $X_1 | \theta, X_2 | \theta, \dots, X_n | \theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$ i.e. the standard assumption.

What should we use as a prior? Looking at the entire dataset for 6,061 batters who had 100 or more at bats, I fit the beta distribution PDF to the sample batting averages using the maximum likelihood approach and I'm estimating $\alpha = 42.3$ and $\beta = 127.7$

- (a) [4 pt / 25 pts] Would this prior be uninformative? Yes / no and why.

No. $\alpha_0 = \alpha + \beta = 170$ which represents a lot of pseudodata seen prior to the actual observed data.

- (b) [9 pt / 34 pts] A new draft pick starts his career with 10 at bats of which 9 are hits.

Using only this small sample and the prior, test whether or not he is "better than average" where the "average" is determined by the historical average of players that came before him. You can express your answer using functions from Table 1 if necessary. Note: without a computer you will not be able to compute the p -value.

$$H_0: \theta \leq .25$$

$$H_a: \theta > .25$$

$$p_{\text{nc}} = P(\theta_0 | x) = P(\theta \leq .25 | x) = \boxed{p_{\text{beta}}(.25, 51.3, 128.7)}$$

$$\theta | x \sim \text{Beta}(\alpha + x, \beta + n - x) = \text{Beta}(42.3 + 9, 127.7 + 1) = \text{Beta}(51.3, 128.7)$$

$$\theta_0 = E(\theta) = \frac{42.3}{42.3 + 127.7} = .25$$

- (c) [3 pt / 37 pts] Without computing the Bayesian p -value in (b) are you able to determine if the null is rejected? Yes / no.

- (d) [9 pt / 46 pts] Derive an expression for the Bayes factor for the null and alternative in (b). You can express your answer using functions from Table 1 if necessary or make sure the expression can be computed via numerical integration (i.e. no symbols are allowed).

$$B = \frac{P_{H_1}(x)}{P_{H_0}(x)} = \frac{\int_{H_1} P(x|\theta) P_{H_1}(\theta) d\theta}{\int_{H_0} P(x|\theta) P_{H_0}(\theta) d\theta} = \frac{\int_{.25}^1 \binom{10}{9} \theta^9 (1-\theta)^1 \frac{1}{B(51.3, 128.7)} \theta^{51.3} (1-\theta)^{128.7} d\theta}{\int_0^{.25} \binom{10}{9} \theta^9 (1-\theta)^1 \frac{1}{B(51.3, 128.7)} \theta^{51.3} (1-\theta)^{128.7} d\theta}$$

$$= \frac{\int_{.25}^1 \theta^{50.3} (1-\theta)^{127.7} d\theta}{\int_0^{.25} \theta^{50.3} (1-\theta)^{127.7} d\theta}$$

- (e) [9 pt / 55 pts] Using only this small sample and the prior, derive an expression for the probability that he will have more than 9 hits out of the next 15 at bats. You can express your answer using functions from Table 1 if necessary.

$$X^* | X \sim \text{Beta Binom}(n^*, \alpha', \beta') = \text{Beta Binom}(15, 51.3, 128.7)$$

$$P(X^* > 9 | X) = 1 - P(X^* \leq 9 | X) = 1 - p_{\text{beta binom}}(9, 15, 51.3, 128.7)$$

Problem 3 Consider the exponential model $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(\theta) := \theta e^{-\theta x}$.

- (a) [5 pt / 60 pts] Derive $f_{X_1, \dots, X_n}(X_1, \dots, X_n; \theta)$ i.e. the joint density function for all X_1, \dots, X_n and simplify.

$$p(x|\theta) = f(x|\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

- (b) [3 pt / 63 pts] Is your answer in (a) the same as $\mathcal{L}(\theta; X_1, \dots, X_n)$, i.e. the likelihood function? Yes / No.
- (c) [8 pt / 71 pts] Demonstrate that the conjugate model for the exponential likelihood is gamma. Find the values of the posterior parameters given the prior parameters are the standard α and β .

$$p(\theta|x) \propto p(x|\theta) p(\theta) = \theta^n e^{-\theta \sum x_i} p(\theta)$$

$$\text{If } p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta} \propto \text{Gamma}(\alpha, \beta) \text{ or "Gamma}(\alpha, \beta)\text{"}$$

$$\Rightarrow p(\theta|x) \propto \theta^n e^{-\theta \sum x_i} \theta^{\alpha-1} e^{-\beta\theta} = \theta^{n+\alpha-1} e^{-(\beta + \sum x_i)\theta} \propto \text{Gamma}(n+\alpha, \beta + \sum x_i)$$

- (d) [8 pt / 79 pts] Using your answers from (a) and (b), derive the Fisher information for θ given X_1, \dots, X_n and simplify. To do so, you may need to know the expectation of the exponential distribution. Here it is: if $W \sim \text{Exp}(\theta)$ then $\mathbb{E}[W] = \frac{1}{\theta}$.

$$\begin{aligned} \ell(\theta; x) &= n \ln(\theta) - \theta \sum x_i & I(\theta) &= E[-\ell''(\theta; x)] \\ \ell'(\theta; x) &= \frac{n}{\theta} - \sum x_i & &= E\left[-\frac{n}{\theta^2}\right] \\ \ell''(\theta; x) &= -\frac{n}{\theta^2} & &= E\left[\frac{n}{\theta^2}\right] \\ & & &= \frac{n}{\theta^2} \end{aligned}$$

- (e) [8 pt / 87 pts] Demonstrate that the Jeffrey's prior for the exponential likelihood can be expressed using the gamma PDF function and find the numerical values of its hyperparameters.

$$p_J(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta^2}} \propto \frac{1}{\theta} = \theta^{-1} = \text{Gamma}(0, 0)$$

- (f) [2 pt / 89 pts] Is the Jeffrey's prior for the exponential model *proper*? Yes / no.
 (g) [4 pt / 93 pts] Given the Jeffrey's prior and a sample of $n = 3$ with values 3.6, 1.2, 5.9 find an expression for $\hat{\theta}_{\text{MAP}}$. You can express your answer using functions from Table 1 if necessary otherwise round to two digits. If you did not find Jeffrey's prior, use the Haldane prior instead.

$$\theta \sim \text{Gamma}(0, 0) \Rightarrow \theta | x \sim \text{Gamma}(n, \sum x_i) = \text{Gamma}(3, 10.7)$$

$$\begin{aligned} \sum x_i &= 3.6 + 1.2 + 5.9 = 10.7 \\ n &= 3 \end{aligned}$$

$$\hat{\theta}_{\text{MAP}} = \text{Mode}[\theta | x] = \frac{\alpha' - 1}{\beta'} = \frac{3 - 1}{10.7} = \frac{2}{10.7} = \boxed{0.187}$$

- (h) [4 pt / 97 pts] Given the Jeffrey's prior and a sample of $n = 3$ with values 3.6, 1.2, 5.9 find an expression for $\hat{\theta}_{\text{MAE}}$. You can express your answer using functions from Table 1 if necessary otherwise round to two digits. If you did not find Jeffrey's prior, use the Haldane prior instead.

$$\hat{\theta}_{\text{MAE}} = \text{Med}[\theta | x] = \text{qgamma}(0.5, 3, 10.7)$$

- (i) [3 pt / 100 pts] If you employ the Haldane prior for the exponential likelihood model, is the posterior *always proper*? Yes / no.