Math 341 / 650 Spring 2017
Midterm Examination Two

Solutions

Professor Adam Kapelner

Thursday, April 6, 2017

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. $\boxed{\text{Box}}$ in your final answers. Good luck!

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)
normal (multivariate)		dmvnorm($\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)		r-($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$)
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 (nonstandard)	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. 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.

Problem 1 The table below displays an excerpt of historical data of the S&P500 index for American equities going back to 1975, a total of $n = 42$ years.

Year	Beginning Price	Ending Price	Gain or Loss	Percent Gain or Loss
1975	68.56	90.19	21.63	31.55%
1976	90.19	107.46	19.15	19.15%
⋮				
2013	1426.19	1848.36	422.17	29.60%
2014	1848.36	2058.9	210.54	11.39%
2015	2058.9	2043.94	-14.96	-0.73%
2016	2043.94	2238.83	194.89	9.54%

For the yearly percent gains (or losses) we calculate the sample average and sample standard deviation:

$$\begin{aligned} \bar{x} &= 9.90 \\ s &= 15.86 \end{aligned}$$

We also assume that the yearly percent gains (or losses) is modeled via

$$X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\theta,\, \sigma^2\right)$$

(a) [6 pt / 6 pts] Is the model assumption a realistic assumption? To get full credit, you must address two distinct assumptions.

Likely <u>not</u> because:

(a) the returns are not independent as 2015 affects 2016 e.g.

(b) the returns likely have different means and variances based on changing conditions i.e. they're not identically distributed

(c) the returns are likely not Gaussian and likely have bigger tails or possibly not symmetric

(b) [6 pt / 12 pts] Regardless of your answer in (a), assume the normal likelihood model for the duration of the problem. We now assume that σ^2 is known and it is equal to s^2. Provide an uninformative prior for θ below and give some sort of short justification for why you wish to use this prior.

$p(\theta) \propto 1$ because I like the Jeffrey's prior and its advantages

(c) [6 pt / 18 pts] Find the posterior distribution of θ. Provide values for the parameters.

$$\theta | x_1, \ldots, x_n, \sigma^2 \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) = N\left(9.90, \frac{15.86^2}{42}\right) = N(9.90, 2.45^2)$$

(d) [6 pt / 24 pts] Provide an expression for a 99% credible region for θ.

$$CR_{\theta, 99\%} = [\,\text{qnorm}(0.005, 9.90, 2.45),\ \text{qnorm}(.995, 9.90, 2.45)\,]$$

(e) [6 pt / 30 pts] Provide an expression for the probability that the mean return is greater than 11%.

$$P(\theta > 11\% \mid X_1, \ldots, X_n, \sigma^2) = 1 - \text{pnorm}(11, 9.90, 2.45)$$

(f) [6 pt / 36 pts] Provide an expression for the probability that *next year's* return will be greater than 11%.

$$X^* \mid X_1, \ldots, X_n, \sigma^2 \sim N(\theta_p, \sigma_p^2 + \sigma^2) = N\left(\bar{x}, \frac{\sigma^2}{n} + \sigma^2\right) = N\left(9.90, \frac{15.86^2}{42} + 15.86^2\right)$$

$$= N(9.90, 16.05^2)$$

$$P(X^* > 11\% \mid X_1, \ldots, X_n, \sigma^2) = 1 - \text{pnorm}(11, 9.90, 16.05)$$

(g) [6 pt / 42 pts] We now pretend that θ is known and equal to $\bar{x}$. Provide an uninformative prior for σ^2 below and give some sort of short justification for why you wish to use this prior.

$P(\sigma^2) \propto \frac{1}{\sigma^2}$ because I like the Jeffreys prior and its advantages

(h) [6 pt / 48 pts] Find the posterior distribution of σ^2.

$$SSE = (n-1)s^2 = 41 \cdot 15.86^2 = 10313.12$$

$$\sigma^2 \mid X_1, \ldots, X_n, \theta \sim \text{InvGamma}\left(\frac{n}{2}, \frac{SSE}{2}\right) = \text{InvGamma}\left(\underbrace{\frac{42}{2}}_{=21}, \underbrace{\frac{10313.12}{2}}_{=5156.56}\right)$$

(i) [6 pt / 54 pts] Provide an expression for a 99% credible region for σ^2.

$$CR_{\sigma^2, 99\%} = [\text{qinvgamma}(0.005, 21, 5156.56), \text{qinvgamma}(.995, 21, 5156.56)]$$

(j) [6 pt / 60 pts] Test that the standard error return is less than 11% by finding an expression for the p value and indicating the comparison you would do. (Note standard error is the square root of variance).

$H_0: \sigma^2 \geq 11^2$

$H_1: \sigma^2 < 11^2$

$\alpha = 5\%$

$$p_{val} := P(H_0 \mid x_1,\dots,x_n,\theta) = P(\sigma^2 > 11^2 \mid x_1,\dots,x_n,\theta) = 1 - \text{pinvgamma}(11^2, 21, 5156.56)$$

if $p_{val} < \underset{\alpha}{\underset{\shortparallel}{5\%}} \Rightarrow$ Reject H_0

(k) [6 pt / 66 pts] Provide an expression for the probability that next year's return will be less than 11%. If there is no closed form expression or no expression in Table 1, please provide an expression that can be evaluated numerically

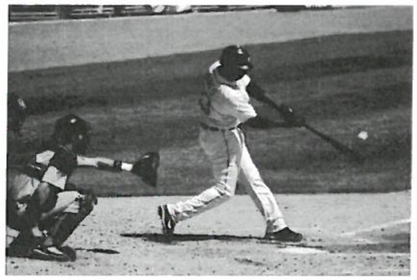
$$P(x^* \mid x_1,\dots,x_n,\theta) = \int_{\Theta} \underbrace{P(x^* \mid \theta, \sigma^2)}_{N(\theta,\sigma^2)} \underbrace{P(\sigma^2 \mid x_1,\dots,x_n,\theta)}_{\text{Invgamma}(21,\,5156.56)} \, d\sigma^2$$

$\theta = 9.90$?

$$= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^* - 9.90)^2} \frac{5156.56^{21}}{\Gamma(21)} e^{-\frac{5156.56}{\sigma^2}} (\sigma^2)^{-22} \, d\sigma^2$$

Problem 2 This question continues our discussion of "batting averages" in baseball. Every hitter's *sample* batting average (BA) is defined as:

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



In class we discussed a means to infer a hitter's true batting average θ by using empirical Bayes with the binomial likelihood model where the prior we considered was $\theta \sim \text{Beta}(\alpha = 78.7, \beta = 224.8)$.

(a) [8 pt / 74 pts] Given that a new batter starts his season with $n = 6$ at bats and 3 hits, find an expression for the probability he has exactly 100 hits during his next 387 at bats.

$X|\theta \sim Bin(n, \theta)$

$\theta \sim Beta(\alpha, \beta)$

$\theta | X \sim Beta(\alpha + x, \beta + n - x)$

$X^* | X \sim BetaBin(n^*, \alpha + x, \beta + n - x) = BetaBin(387, \underbrace{78.7+3}_{81.7}, \underbrace{224.8+3}_{227.8})$

$P(X^* = 100 | X) = dbetabinom(100, 387, 81.7, 227.8)$

Problem 3 This question has some theoretical questions. The parts are independent from each other unless otherwise stated.

(a) [4 pt / 78 pts] If $X_1, \ldots, X_n \mid \theta \overset{iid}{\sim} \text{Geometric}(\theta)$ and $\theta \sim \text{Beta}(\alpha, \beta)$, what is $\mathbb{E}[\theta \mid X_1, \ldots, X_n]$?

$\theta | X \sim Beta(\alpha + n, \beta + \sum x_i)$

$\Rightarrow E[\theta | X] = \frac{\alpha + n}{\alpha + \beta + n + \sum x_i}$

(b) [4 pt / 82 pts] If $X_1, \ldots, X_n \mid \theta \overset{iid}{\sim} \text{Poisson}(\theta)$ and $\theta \sim \text{Gamma}(\alpha,\ \beta)$, what is the kernel of $\mathbb{P}(\theta \mid X_1, \ldots, X_n)$?

$P(\theta | x_1,\ldots,x_n) = \text{Gamma}(\alpha + \sum x_i, \beta + n) = \frac{(\beta+n)^{\alpha + \sum x_i}}{\Gamma(\alpha + \sum x_i)} e^{-(\beta+n)\theta} \theta^{\alpha + \sum x_i - 1} \propto e^{-(\beta+n)\theta} \theta^{\alpha + \sum x_i - 1}$

(c) [2 pt / 84 pts] Under the model in (b), what is the Jeffrey's prior?

$\theta \sim \text{Gamma}(\frac{1}{2}, 0)$

(d) [6 pt / 90 pts] Under the model in (b), find $\mathbb{P}(X^* \mid X_1, \ldots, X_n)$ where X^* is one future realization from the same poisson likelihood model. For the parameters of this distribution, only use notation given in (b).

$X^* | x_1,\ldots,x_n \sim \text{NegBin}\left(\alpha + \sum x_i, \frac{\beta+n}{\beta+n+1}\right)$

(e) [4 pt / 94 pts] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\ \sigma^2)$ with σ^2 known and the prior is $\theta \mid \sigma^2 \sim \mathcal{N}(\mu_0,\ \tau^2)$, what is $\mathbb{P}(\theta \mid X_1, \ldots, X_n, \sigma^2)$?

$\theta | x_1,\ldots,x_n, \sigma^2 \sim N\left(\frac{\frac{\bar{x} n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \right)$

(f) [6 pt / 100 pts] For the model in (e) reparameterize the prior using ρ, the quantity that indicates shrinkage to the prior expectation.

$$\rho = \frac{\sigma^2}{n\tau^2+\sigma^2} \Rightarrow \frac{1}{n\tau^2+\sigma^2} = \frac{\rho}{\sigma^2} \Rightarrow n\tau^2+\sigma^2 = \frac{\sigma^2}{\rho}$$

$$\Rightarrow n\tau^2 = \frac{\sigma^2}{\rho} - \sigma^2 = \sigma^2\left(\frac{1}{\rho}-1\right) = \sigma^2 \frac{1-\rho}{\rho} \Rightarrow \tau^2 = \frac{\sigma^2}{n}\frac{1-\rho}{\rho}$$

$$\Rightarrow \theta \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{n}\frac{1-\rho}{\rho}\right)$$

(g) [6 pt / 106 pts] [Extra Credit] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ with σ^2 known and the prior is a *mixture* of K Gaussians e.g. $\theta_1, \theta_2, \ldots, \theta_K \overset{ind}{\sim} \mathcal{N}(\mu_k, \tau_k^2)$, demonstrate that the posterior is normal and find the posterior parameters. Assume the mixture is uniform i.e., each of the k mixture components has the same probabilitiy i.e. $1/K$.