Vanderbilt University Biostatistics Comprehensive Examination

PhD Theory Exam Series 2

May 23–May 26, 2023

Instructions: Please adhere to the following guidelines:

- This exam is scheduled to be administered on Tuesday, May 23 at 9:00am, and will be due on Friday, May 26 at 5:00pm. This deadline is strict: late submissions will not be accepted.
- To turn in your exam, please use your assigned Box folder and e-mail your exam (word-processed or hand-written/scanned) to Dr. Andrew Spieker by the deadline. This level of redundancy is designed to ensure that your exam is received by the deadline. If you would like to e-mail exam drafts along the way, that is perfectly acceptable—do not be concerned about spamming my inbox.
- There are five equally weighted problems of varying length and difficulty. Note that not all sub-questions are weighted equally. You are advised to pace yourself and to not spend too much time on any one problem.
- Answer each question clearly and to the best of your ability. Partial credit will be awarded for partially correct answers.
- Be as specific as possible in your responses.
- You may consult reference material (e.g., course notes, textbooks), though the work you turn in must be your own. This is an *individual effort*. Do not communicate about the exam with anyone. Vanderbilt University's academic honor code applies.
- Please direct clarifying questions by e-mail to Dr. Andrew Spieker and Dr. Bob Johnson.

---

1. [20 pts] Let $\varsigma = \{A_1, A_2, A_3\}$ be a collection of disjoint subsets of $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ defined as:

$$A_1 = \{1, 2, 3\}$$
$$A_2 = \{4, 5, 6\}$$
$$A_3 = \{7, 8\}.$$

Let $\mathcal{F} = \sigma(\varsigma)$ be the σ-algebra with respect to Ω generated by ς.

(a) Write $\mathcal{F}$ as a collection of subsets of Ω. What is its cardinality?

(b) Define the function $X(\omega)$ for $\omega \in \Omega$ as:

$$X(\omega) = \begin{cases} a_k & \text{if } \omega \in A_k \text{ is even} \\ b_k & \text{if } \omega \in A_k \text{ is odd} \\ c & \text{if } \omega = 9 \end{cases}, \qquad \text{for } k = 1, 2, 3.$$

i. What conditions on c, a_k, and b_k (for $k = 1, 2, 3$) are needed—if any—for X to be a random variable defined on the measure space $(\Omega, \mathcal{F})$?

ii. What conditions on c, a_k, and b_k (for $k = 1, 2, 3$) are needed—if any—for X^2 to be a random variable defined on the measure space $(\Omega, \mathcal{F})$?

iii. Is $Y(\omega) = 10$ for $\omega \in \Omega$ a random variable on $(\Omega, \mathcal{F})$?

(c) Define $\mathcal{F}^*$ as the smallest (with respect to cardinality) σ-algebra on Ω such that X is a random variable on $(\Omega, \mathcal{F}^*)$ for all values of c, a_k, and b_k, for $k = 1, 2, 3$. What are the subsets of Ω contained in $\mathcal{F}^*$?

---

2. [20 pts] Let $X_1, X_2, \ldots$ be a sequence of independent Uniform$(0,1)$ random variables and let $N_\lambda \sim$ Poisson(λ) be independent of $X_1, X_2, \ldots$. We are interested in $V_\lambda = \max\{X_1, X_2, \ldots, X_{N_\lambda}\}$ where $V_\lambda = 0$ when $N_\lambda = 0$.

(a) Determine the distribution function (CDF), $F_{V_\lambda}(t)$, and the *mixture* distribution, $f_{V_\lambda}(t)$ of V_λ. Plot these functions for $\lambda = 1.5$.

(b) Determine the moment generating function for V_λ.

(c) Show that $\mathrm{E}[V_\lambda] \longrightarrow 1$ as $\lambda \longrightarrow \infty$.

(d) Show that $\lambda(1 - V_\lambda) \xrightarrow{d} Z$ as $\lambda \longrightarrow \infty$ and determine the distribution of Z.

(e) Determine the generalized inverse distribution function, $F^{-1}_{V_\lambda}(u)$. Plot this function for $\lambda = 1.5$.

(f) Simulate $n = 1000$ values of V_λ for each of $\lambda = 1, 2, 3, 5, 10, 50, 100$ in both of the two ways described below. For each way, plot the empirical distributions overlaying on each of two graphs (one for each method).

i. First simulate N_λ, and then generate $\{X_1, X_2, \ldots, X_{N_\lambda}\}$, and determine V_λ.

ii. Use $F^{-1}_{V_\lambda}(u)$ to simulate.

What appears to be the limiting distribution? Discuss how you could have predicted the result.

---

3. [20 pts] Suppose that $X_1, \ldots, X_n$ are independent and identically distributed Poisson(λ) random variables, with $\lambda > 0$ unknown. Consider the prior distribution,

$$\pi(\lambda) \quad = \quad \frac{1}{6}\alpha^4 \lambda^3 e^{-\alpha\lambda} \mathbf{1}_{(0,\infty)}(\lambda),$$

for some known value $\alpha > 0$.

(a) Derive the posterior mode, $\widetilde{\lambda}_n$, and prove that it is consistent for λ (justify consistency by naming any theorems you invoke).

(b) Prove that for each λ, there is a unique choice of α for which $\widetilde{\lambda}_n$ is unbiased for all n. Then, briefly explain why this fact is of no practical value.

(c) For a given choice of α, does it necessarily follow that there exists some integer n for which $\mathrm{E}[\widetilde{\lambda}_n] = \lambda$ for all λ? Justify your answer.

(d) Explicitly determine the posterior predictive distribution, $p(X^*|X_1, \ldots, X_n)$, where X^* denotes some out-of-sample "future" observation.

(e) What should be the asymptotic posterior predictive distribution? Your argument can be heuristic in this problem.

(f) Formally determine the asymptotic posterior predictive distribution.

For the remainder of the problem, assume that $\lambda = 1$.

(g) A "highest posterior density" 95% credible interval for λ takes the form

$$\mathcal{C} \quad = \quad \{\lambda^* : \pi(\lambda^*|X_1, \ldots, X_n) \geq q\},$$

where q is the largest number such that

$$\int_{\lambda:\pi(\lambda^*|X_1,\ldots,X_n)\geq q} \pi(\lambda^*|X_1, \ldots, X_n) d\lambda^* = 0.95.$$

Conduct a simulation study in which you numerically approximate the coverage associated with a 95% credible interval of this variety for the combination of simulation parameters marked in the table below:

	$n = 10$	$n = 25$	$n = 50$	$n = 100$
$\alpha = 4$				
$\alpha = 8$				
$\alpha = 16$				
$\alpha = 32$				

You may find the R function `hdi()` in the library `HDInterval` useful; you are free to use it. Please include annotated software code as part of your response. Very briefly account for the patterns you see.

(h) Re-run the simulation you conducted in part (g), instead forming the credible intervals in an "equal-tailed" fashion (i.e., based on the 2.5^{th} and 97.5^{th} percentiles of the posterior distribution). Very briefly account for the differences between these results and those of part (g).

---

4. [20 pts] Consider a three-state Markov chain with transition matrix given by:

$$\mathbf{P} = \begin{bmatrix} 0 & \alpha_1 & \beta_1 \\ \alpha_2 & 0 & \gamma_1 \\ \beta_2 & \gamma_2 & 0 \end{bmatrix},$$

specifically letting X_n denote the state (1, 2, or 3) after $n \in \mathbb{Z}^+$ steps from initial state X_0.

(a) Note that $\inf\{n \geq 1 : X_n = 1\}$ counts the number of steps until transitioning to state 1 from another state for the first time. For $i = 1, 2, 3, 4$, derive an explicit expression for $\mathrm{P}(\inf\{n \geq 1 : X_n = 1\} = i | X_0 = 1)$.

(b) Prove the following formula by induction for $i \geq 2$:

$$\mathrm{P}(\inf\{n \geq 1 : X_n = 1\} = i | X_0 = 1) = (\gamma_1\gamma_2)^{\lfloor \frac{i-2}{2} \rfloor} \left((\alpha_1\alpha_2 + \beta_1\beta_2) \frac{(-1)^i + 1}{2} + (\alpha_1\gamma_1\beta_2 + \beta_1\gamma_2\alpha_2) \left(1 - \frac{(-1)^i + 1}{2} \right) \right).$$

Hint: Handle the even and odd cases separately.

(c) Determine an expression for $\mathrm{E}[\inf\{n \geq 1 : X_n = 1\} | X_0 = 1]$ in terms of α_1, α_2, β_1, β_2, γ_1, and γ_2; you may use the following summation formulas without proof:

$$\sum_{j=1}^{\infty} x^{j-1} = \frac{1}{1-x} \text{ and } \sum_{j=1}^{\infty} j x^{j-1} = \frac{1}{(1-x)^2} \text{ for } |x| < 1.$$

Further, specify the values of α_1, α_2, β_1, β_2, γ_1, and γ_2 for which this expectation exists.

(d) Determine all possible combinations of α_1, α_2, β_1, β_2, γ_1, and γ_2 such that the Markov chain is periodic. Compute $\mathrm{E}[\inf\{n \geq 1 : X_n = 1\} | X_0 = 1]$ for each such case using the formula you derived in part (c); can you make a stronger statement about $\inf\{n \geq 1 : X_n = 1\}$ in this case?

(e) Suppose $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, and $\gamma_1 = \gamma_2$. Given this information, determine the exact numeric entries of the Markov matrix. Under this Markov chain, use simulation techniques to approximate the expected number of occurrences of the specific sequence $1 \to 2 \to 3 \to 2 \to 1$ after a total of $n = 24$ steps given each initial state $X_0 = 1$, $X_0 = 2$, and $X_0 = 3$. Include X_0 as part of the chain so that the chain's total length is $n + 1 = 25$. Please include annotated software code as part of your response.

---

5. [20 pts] Let $\mathbf{X}$ denote an $n \times k$ design matrix of covariates, each standardized to have mean zero and variance one (assume $k \leq n$). Further, let $\mathbf{y}$ denote an $n \times 1$ outcome vector. Recall that the ridge-penalized least squares estimator minimizes the following objective function:

$$L_\lambda(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) \quad = \quad ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 + \lambda||\boldsymbol{\beta}||^2,$$

and possesses the following closed-form expression:

$$\widehat{\boldsymbol{\beta}}_\lambda \quad = \quad (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}.$$

You may assume in this problem that the design matrix, $\mathbf{X}$, is fixed by design and that the outcomes have a common variance.

(a) Characterize the set, Λ, of all possible combinations of eigenvalues of $\mathbf{X}^T\mathbf{X}$; justify your answer.

(b) Characterize the set $\Lambda^* \subseteq \Lambda$, of all possible combinations of eigenvalues of $\mathbf{X}^T\mathbf{X}$ such that $\mathbf{X}^T\mathbf{X}$ non-singular; justify your answer.

(c) Characterize all values of $\lambda \in \mathbb{R}$ such that $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ is non-singular; justify your answer.

(d) Justifying your steps, show that for $\lambda > 0$,

$$g(\lambda; \mathbf{X}, \mathbf{y}) := \frac{\partial \widehat{\boldsymbol{\beta}}_\lambda}{\partial \lambda} = -(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\widehat{\boldsymbol{\beta}}_\lambda.$$

(e) Argue that for $\lambda \gg n$, the coefficient path for a single coefficient—namely, $\{(\lambda, \widehat{\beta}_{\lambda;j})\}$, resembles the graph of the hyperbolic function $\widehat{\beta}_{\lambda;j} = k/\lambda$.

For the remainder of the problem, consider the following $n = 5$ independent observations based on three covariates (not yet centered/scaled):

ID	X_1	X_2	X_3	Y
1	-1	1	0	1
2	-1	1	0	2
3	0	0	0	-5
4	0	1	1	5
5	1	0	1	-1

You may use statistical software to aid you in the computationally cumbersome parts of the problems that follow, but you should make clear what you're doing, how you're doing it, and why you're doing it.

(f) Let $\widehat{\boldsymbol{\beta}}_\lambda$ denote a solution (if any) to the penalized normal equations, $\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})\boldsymbol{\beta}$, for a given λ. Argue that $\widehat{\boldsymbol{\beta}}_{\lambda=0}$ exists but is not unique.

(g) Let $\widehat{\mathbf{y}}_\lambda = \mathbf{X}\widehat{\boldsymbol{\beta}}_\lambda$ denote the fitted vector for a given λ. Argue that $\widehat{\mathbf{y}}_{\lambda=0}$ is unique (i.e., the same for any $\widehat{\boldsymbol{\beta}}_{\lambda=0}$ solving the unpenalized normal equations). Justifying your steps, determine $\widehat{\mathbf{y}}_{\lambda=0}$.

(h) It has been shown that in some circumstances, a *negative* ridge penalty is capable of producing a solution that minimizes expected prediction error; however, your answer to part (c) may prompt some concerns. Using R, graph each component of $\widehat{\boldsymbol{\beta}}_\lambda$ as a function of λ over the range `lambda=seq(-pi, 0, 0.2)`. Comment on and account for your findings.

(i) Using R, graph each component of $\widehat{\mathbf{y}}_\lambda$ as a function of λ over the range `lambda=seq(-pi, 0, 0.2)`.

(j) Despite our ability to get around our "problem-case" of $\lambda = 0$, we are apparently *not* able to do so for other problematic values of λ. Discuss how this relates to the formula for effective degrees of freedom given λ:

$$\text{df}(\lambda) \quad = \quad \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda},$$

where σ_i represents the i^{th} singular value of $\mathbf{X}$. *Hint*: Does this formula even make sense for all $\lambda < 0$?
