Practice "Core Competency" Qualifying Exam (Summer 2018)

Name (Print): ____________________

This exam contains 19 problems. Answer all of them. Point values are in parentheses. You **must show your work** to get credit for your solutions — correct answers without work will not be awarded points. This is a closed book/notes exam. No calculators will be allowed in the exam.

1. (10 points) Consider a routine screening test for a disease. Suppose the frequency of the disease in the population (base rate) is 0.5%. The test is highly accurate with a 5% false positive rate and a 10% false negative rate [a false positive happens when a test result indicates that the disease is present (the result is positive), but it is, in fact, not present. Similarly, a false negative happens when a test result indicates that the disease is not present (the result is negative), but it is, in fact, present].

   You take the test and it comes back positive. What is the probability that you have the disease?

2. (15 points) (5 + 5 + 5) Suppose that $X_1, \ldots, X_n$ are i.i.d Exp$(1/\mu)$, where $\mathbb{E}(X_1) = \mu > 0$.

   (i) Find the mean and variance of $\bar{X}_n = \sum_{i=1}^{n} X_i/n$. Hence, find the asymptotic distribution of $\bar{X}_n$ (properly standardized).

   (ii) Let $T = \log \bar{X}_n$. Find the corresponding asymptotic distribution of T (properly standardized).

   (iii) How can the asymptotic distribution of T be used to construct an approximate $(1-\alpha)$ confidence interval (CI) for μ? Explain your answer and give the desired CI.

3. (10 points) (6 + 4) Let $W_1, W_2, \ldots, W_k$ be unbiased estimators of a parameter θ with $\text{Var}(W_i) = \sigma_i^2$ and $\text{Cov}(W_i, W_j) = 0$ if $i \neq j$.

   (a) Show that among all estimators of the form $\sum_{i=1}^{k} a_i W_i$, where a_i's are constants and $\mathbb{E}_\theta(\sum_i a_i W_i) = \theta$, the estimator $W^* = \frac{\sum_i W_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$ has minimum variance.

   (b) Show that $\text{Var}(W^*) = \frac{1}{\sum_i 1/\sigma_i^2}$.

4. (10 points) Suppose that the radius of a circle is a random variable having the following probability density function:

   $$f(x) = \frac{1}{8}(3x+1), \qquad 0 < x < 2$$

   and 0 otherwise. Determine the probability density function of the area of the circle.

5. (10 points) (5 + 5) Suppose that $Y_1, \ldots, Y_n$ are i.i.d Poisson(λ), $\lambda > 0$ unknown. Assume that n is even, i.e., $n = 2k$ for some integer k. Consider

   $$\hat{\lambda} = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2.$$

   (a) Is $\hat{\lambda}$ an unbiased estimator of λ (show your steps)?

   (b) Is $\hat{\lambda}$ a consistent estimator of λ, as $k \to \infty$ (show your steps)?

6. (20 points) (7 + 7 + 6) Consider observed response variables $Y_1, \ldots, Y_n \in \mathbb{R}$ that depend linearly on covariates $x_1, \ldots, x_n$ as follows:

$$Y_i = \beta x_i + \epsilon_i, \quad \text{for } i = 1, \ldots, n.$$

Here, the ϵ_i's are independent Gaussian noise variables, but we do not assume they have the same variance. Instead, they are distributed as $\epsilon_i \sim N(0, \sigma_i^2)$ for possibly different variances $\sigma_1^2, \ldots, \sigma_n^2$. The unknown parameter of interest is β.

   (a) Suppose that the error variances $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Find the MLE $\hat{\beta}$ for β in this case and derive an explicit formula for $\hat{\beta}$. Show that $\hat{\beta}$ minimizes a certain weighted least-squares criterion.

   (b) Show that the estimate $\hat{\beta}$ in part (a) is unbiased, and derive a formula for the variance of $\hat{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$.

   (c) Compute the Fisher information $I(\beta)$ in this model (still assuming $\sigma_1^2, \ldots, \sigma_n^2$ are known constants). Compare this value with the variance of $\hat{\beta}$ derived in part (b).

7. (10 points) (7 + 3) Suppose that $X \sim \text{Poisson}(\lambda)$ and its parameter $\lambda > 0$ has a prior distribution Gamma(α, β) given by density

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-y\beta} y^{\alpha-1}, \quad \text{for } y \geq 0, \quad \text{(and 0 otherwise)}.$$

   (a) Find the posterior distribution of λ given the observation X, and identify the distribution with its parameters.

   (b) Find the mean of this posterior distribution.

8. (16 points) (4 × 4) Suppose $X_1, X_2 \overset{i.i.d.}{\sim} Ber(p)$ for some unknown parameter $p \in (0, 1)$. Find an unbiased estimator for the following functions of p, if there exists one.

   (a) $g(p) = 2p$.

   (b) $g(p) = p(1 - p)$.

   (c) $g(p) = p^2$.

   (d) $g(p) = p^3$.

9. (10 points) (2 + 8) Suppose X_1, X_2 are i.i.d. random variables from a distribution with 0 and variance 1.

   (a) If $F = N(0, 1)$, show that $\frac{X_1+X_2}{\sqrt{2}} \overset{d}{=} X_1$.

   (b) Now suppose that $\frac{X_1+X_2}{\sqrt{2}} \overset{d}{=} X_1$. Show that $F = N(0, 1)$.

10. (9 points) Suppose that $X_1, \cdots, X_n \overset{i.i.d.}{\sim} N(0, 1)$, and A is a $n \times n$ matrix which is symmetric (i.e., $A^\top = A$) and idempotent (i.e., $A^2 = A$). Find the distribution of $\sum_{i,j=1}^n X_i X_j A(i, j)$. Assume if necessary that $\sum_{i=1}^n A(i, i) = s$.

11. (15 points) (4 + 5 + 6) Suppose that $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Ber(\lambda/n)$.

   (a) What is the distribution of $\sum_{i=1}^n X_i$.

   (b) Compute $\lim_{n\to\infty} \mathbb{P}(\sum_{i=1}^n X_i = k)$, where k is any fixed nonnegative integer, and hence show that $\sum_{i=1}^n X_i$ converges in distribution to a random variable Y.

   (c) Compute $\mathbb{E}[Y(Y-1)]$, where Y is as in part (b).

12. (15 points) (5 + 4 + 6) Suppose that $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$. Let $V_1 := \max(U_1, U_2), V_2 := \min(U_1, U_2)$.

   (a) Find $\mathbb{P}(V_1 \geq x, V_2 \leq y)$, where $x, y \in [0,1]$.

   (b) Hence or otherwise find the joint density for (V_1, V_2).

   (c) Hence or otherwise compute $\mathbb{E}(V_1^2 + V_2^2)$.

13. (15 points) (5×3) Suppose $X_1, X_2, X_3 \overset{i.i.d.}{\sim} N(0,1)$. Let (Y_1, Y_2, Y_3) be defined as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

   (a) Find the joint distribution of (Y_1, Y_2, Y_3).

   (b) Show that $Y_1^2 + Y_2^2 + Y_3^2 = X_1^2 + X_2^2 + X_3^2$.

   (c) Hence or otherwise derive the distribution of $(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2$, where $\bar{X} := \frac{X_1 + X_2 + X_3}{3}$.

14. (20 points) (3+3+6+3+5) Suppose $f : [0, \infty) \mapsto \mathbb{R}$ be a function such that $f(x+y) = f(x)f(y)$.

   (a) Show that $f(x) \geq 0$ for all real $x \geq 0$.

   (b) Show that $f(0) \in \{0, 1\}$.

   (c) Show that for any nonnegative rational number r one has $f(r) = c^r$, where $c \in [0, \infty)$.

   (d) If f is assumed to be continuous, show that $f(x) = c^x$ for all real $x \geq 0$.

   (e) Suppose X is a nonnegative random variable, such that

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$$

   for every $s, t \geq 0$. If X has a continuous distribution function, name the distribution of X.

15. (10 points) (6+4) Suppose X is a random variable taking values in $[0, 1]$.

   (a) Show that $\text{Var}(X) \leq \frac{1}{4}$.

(b) Find a random variable X for which equality holds in part (a).

16. (10 points) Farmers in the Hudson Valley pack apples into bags of approximately 10 pounds, but due to the variation in apples the actual weight varies. We may model the weight of a bag as uniformly distributed in $[9.5, 10.5]$ and independent of other bags. The farmers load 1200 bags onto a truck with maximal admissible load of 13000 pounds. Find a simple approximation to the probability that the truck is overloaded, expressed in terms of the Normal distribution.

17. (15 points) $(4 + 4 + 4 + 3)$ Let X and Y be i.i.d. $\mathcal{N}(0, 1)$ random variables. Consider

$$Z := \text{sign}(Y) \cdot X$$

where $\text{sign}(y) := 1$ if $y > 0$ and $\text{sign}(y) := -1$ if $y \leq 0$.

(i) Find the distribution of Z.

(ii) Compute the covariance of X und Z.

(iii) Determine $P[X + Z = 0]$.

(iv) Are X and Z independent? (Give a precise mathematical argument.)

18. (10 points) Suppose Σ is a non negative definite matrix of size $n \times n$ with real entries. Show that $\text{tr}(\Sigma^2) \geq n \det(\Sigma)^{2/n}$.

19. (10 points) Suppose for every $n \geq 1$ A_n is a real symmetric matrix of size $n \times n$, whose eigenvalues $(\lambda_1, \cdots, \lambda_n)$ satisfies the following properties:

(i) $\max_{i=1}^n |\lambda_i| \stackrel{n\to\infty}{\to} 0$.

(ii) $\sum_{i=1}^n \lambda_i^2 = 1$.

Find the asymptotic distribution of $\sum_{i,j=1}^n A_n(i,j)X_iX_j$, where $\{X_i\}_{i\geq 1}$ is a sequence of i.i.d. $N(0, 1)$.