Statistics Ph.D. Qualifying Exam: Part II

August 18, 2017

Student Name: ______________________________

1. Answer 8 out of 12 problems. Mark the problems you selected in the following table.

Problem	1	2	3	4	5	6	7	8	9	10	11	12
Selected												
Scores												

2. Write your answer right after each problem selected, attach more pages if necessary. **Do not** write your answers on the back.

3. Assemble your work in right order and in the original problem order. (Including the ones that you do not select)

4. You can use the N(0,1) distribution table as attached.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population with CDF given by

$$F_x(x) = \left(\frac{x}{\theta} - k\right) \quad k\theta < x < (k+1)\theta$$

where k is a known non-negative number and where $\theta > 0$ is an unknown parameter. Let $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ and $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ and let $U = X_{(n)} - X_{(1)}$.

   (a) Find the density function $f_U(u)$ of U.

   (b) Find a function $g(U)$ such that $E(g(U)) = \theta$.

2. Suppose that U_i has a Bernoulli distribution with parameter π for $i = 1, 2, 3$ and W has a Bernoulli distribution with parameter θ. Let

$$X = WU_1 + (1 - W)U_2 \qquad Y = WU_1 + (1 - W)U_3.$$

Further assume that U_1, U_2, U_3 and W are mutually independent.

(a) Suppose that $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a random sample from the joint distribution of X and Y. Then consider the following 2 estimators of π

$$\hat{\pi}_1 = \bar{X} \qquad \hat{\pi}_2 = \frac{1}{2}(\bar{X} + \bar{Y}).$$

Which of these two estimators should be preferred and why?

(b) If $\pi = 1/2$ find an unbiased estimator of θ.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of service times with mean μ and variance σ^2. We are interested in using $\dfrac{1}{\bar{X}}$ to estimate the rate of service, $\dfrac{1}{\mu}$. Use the central limit theorem and the delta method to construct a $100(1-\alpha)\%$ confidence interval for $\dfrac{1}{\mu}$.

4. Consider the linear (fixed effect) model $y_i = \beta_1 + \beta_2 x_i + \epsilon_i, i = 1, \cdots, n$, where $\epsilon_i \sim i.i.d. N(0, \sigma^2)$. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the least square estimates and form the statistic

$$V_n = \frac{\sum_{i=1}^n (\hat{\beta}_1 + \hat{\beta}_2 x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2} \tag{1}$$

(a) Find the distribution of V_n under $H_0 : \beta_2 = 0$.

(b) Find the distribution of V_n under the alternative $H_1 : \beta_2 \neq 0$.

(c) Explain how to obtain the power of the level α test of $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$ using V_n.

5. In a certain laboratory experiment, the time Y (in milliseconds) for a certain blood clotting agent to show an observable effect is assumed to have the negative exponential distribution $f_Y(y) = \alpha^{-1}e^{-y/\alpha}$, $y > 0$, $\alpha > 0$. Let $Y_1, Y_2, \cdots, Y_n$ constitute a random sample of size n from $f_Y(y)$, and let $y_1, y_2, \cdots, y_n$ be the corresponding observed values of $Y_1, Y_2, \cdots, Y_n$. It is of interst to make statistical inferences about the unknown parameter $\theta = Var(Y) = \alpha^2$ using the available data $\mathbf{y} = (y_1, y_2, \cdots, y_n)$.

   (a) Develop an explicit expression for the MLE $\hat{\theta}$ of θ. If the observed value of $S = \sum_{i=1}^{n} Y_i$ is the value $s = 40$ when $n = 50$, compute an appropriate large-sample 95% CI for the parameter θ.

   (b) Find the minimum variance unbiased estimator(MVUE) $\hat{\theta}^*$ of θ and then find the variance of $\hat{\theta}^*$.

   (c) Does the $\hat{\theta}^*$ in Part (b) achieve the CRLB for the variance of any unbiased estimator of θ?

   (d) For any finite value of n, develop the explicit expressions for MSE($\hat{\theta}$,θ) and MSE($\hat{\theta}^*$,θ) in terms of θ. Using the MSE criterion, which estimator do you prefer for finite n and which estimator do you prefer asymptotically (as $n \to \infty$).

6. Let

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is an n-dimensional random vector, $\mathbf{X}$ is an $n \times k$ matrix, $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$, and $\mathbf{V}$ is positive-definite matrix. Suppose that we erroneously assume $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ and find $\widehat{\boldsymbol{\beta}}$ by ordinary (unweighted) least squares.

   (a) Is $\widehat{\boldsymbol{\beta}}$ unbiased for $\boldsymbol{\beta}$? Justify your answer.

   (b) What is $\mathrm{Var}(\widehat{\boldsymbol{\beta}})$?

   (c) If $\widehat{\boldsymbol{\beta}}_V$ is the generalized least squares estimator of $\boldsymbol{\beta}$ (taking $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$), what is $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_V)$?

   (d) Show that for any $\mathbf{c}$

$$\mathrm{Var}(\mathbf{c}'\widehat{\boldsymbol{\beta}}_V) \leq \mathrm{Var}(\mathbf{c}'\widehat{\boldsymbol{\beta}}).$$

7. Let $X_1, X_2, \ldots$ be a sequence of independent Exponential(λ) random variables. Let $N \sim$ Geometric(p) be a geometric random variable that is independent of the X's. Let $X_{(1)}, \ldots, X_{(N)}$ be the order statistics based on a sample of size N.

   (a) Prove that

$$P\{X_{(1)} > a\} = \frac{pe^{-\lambda a}}{1-(1-p)e^{-\lambda a}}.$$

   (b) Hence, otherwise, find $E(X_{(1)})$.

8. Let $\{X_1, \ldots, X_n\}$ be a random sample from the population with density $f(x, \theta)$, where $f(x, \theta)$ is given by: $f(x, \theta) = e^{-(x-\theta)}, x > \theta; = 0$ if otherwise. Denote by $\{Y_i = i(X_{(n-i+1)} - X_{(n-i)})$ if $i = 1, \ldots, n-1, Y_n = n(X_{(1)} - \theta)$, where $X_{(r)}$ is the $r - th$ order statistic, $r = 1, \ldots, n$.

   (a) Show that $\{Y_i, i = 1, \ldots, n\}$ are independently and identically distributed random variables with density $g(y) = e^{-y}, y > 0; = 0, y \leq 0$.

   (b) Use the result above to derive the UMVUE (Uniformly Minimum Varianced Unbiased estimator) of θ.

   (c) Suppose the density of the prior distribution is given by: $h(\theta) = e^{-(a-\theta)}, a > \theta; = 0, a \leq \theta$, where a is a known constant. Derive the Bayese estimator of θ and compare it with the UMVUE of θ.

9. Consider independent random samples $X_{i1}, \ldots, X_{in}$ $(i = 1, 2)$ from the uniform populations:

$$f(x_i) = \frac{1}{\theta_i}, \quad 0 < x_i < \theta_i, \quad \theta_i > 0, \quad i = 1, 2.$$

Let $Y_i = \max(X_{i1}, \ldots, X_{in})$, $i = 1, 2$, and $Y = \max(Y_1, Y_2)$.

(a) Show that the likelihood ratio statistic to test the hypothesis $H_0 : \theta_1 = \theta_2$ is given by

$$Z = \left(Y_1 Y_2 / Y^2\right)^n.$$

(b) Obtain the exact distribution of $-2 \log Z$ under H_0.

10. Let T_n, $n \geq 1$ be a sequence of random variables such that $n^{1/2}(T_n - \theta)$ converges in distribution to $N\big(0, \sigma^2(\theta)\big)$.

(a) If g is differentiable, derive the asymptotic distribution of $g(T_n)$.

(b) Suppose that $X_1, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$. Find g such that the limiting distribution of $g(S_n^2)$ does not depend on σ^2, where $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$ and $\bar{X}_n = \sum_{i=1}^n X_i/n$.

11. Suppose that $X_1, X_2, ...X_n$ form a random sample from a $B(1, \theta)$ distribution for which the θ is unknown $(0 < \theta < 1)$. Let $T = \sum_{i=1}^{n} X_i$; and

$$\begin{aligned} S &= 1 \qquad \text{if } X_1 = 0 \text{ and } X_2 = 0, \\ S &= 0 \qquad \text{otherwise.} \end{aligned}$$

(a) Show that $E(S) = \theta^2$ and find $\text{Var}(S)$.

(b) Compute the Cramer-Rao lower bound of unbiased estimators of θ^2.

(c) Find the UMVUE of θ^2.

(d) Find the asymptotic distribution of the UMVUE found in (c).

12. Suppose that $X_1, X_2, ...X_n$ form a random sample from a uniform distribution with the following p.d.f.

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_n = X_{(n)}$ be the M.L.E. of θ.

Suppose that we would like to test $H_0 : 1 \leq \theta \leq 4$ vs. $H_1 : \theta < 1$ or $\theta > 4$. The critical region is chosen to be $C = \{Y_n < 3 \text{ or } Y_n > 4\}$.

(a) Determine the power function of the test.

(b) Determine the size of the test.

(c) Determine the size of the test. If the sample size is 5, find a test of level of significance $\alpha = 0.10$. Is your test unbiased? Why ?

Table of P(Z < z), Z ~N(0,1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999