PH.D. QUALIFYING EXAMINATION - THEORETICAL STATISTICS

Time: 1:00 pm-5:00 pm, August 27, 2021

General Instructions

- There are two equally-weighted parts in this exam: Part A (STA584) and Part B (STA684). Each part consists of 5 problems with subproblems and has a possible total of 50 points. You are to solve all problems. Your score on each part will be individually converted into a percentage.

- You must show all your work and justifications correctly and completely to receive full credits. Partial credits may be given for partially correct solutions.

- You are assumed to know the properties, e.g., expected values of the frequently used distributions introduced in the textbook. You may use the properties without proof unless solicited.

- If a theorem is employed, you must clearly state the theorem, identify its assumption(s) and conclusion(s), and explain why it is applicable.

- Please print your name on the line below. Begin each problem on a new sheet with the problem number clearly noted. Write on only one side of each sheet. Do not write your name on any sheets.

- When finished, please arrange all pages according to the problem numbers and then number the pages accordingly. Slide your work and this exam booklet into the envelope given to you. Hand in the envelope.

NAME:____

Part A [The parts of each problem are not necessarily given equal weight.]

Problem #01 [9 points]

To earn full credits, you must round your answer to each of the following three problems to four decimal places.

1.a) Suppose that 10% of general population is infected with the coronavirus. Testing for this infection is available, but the test result is known to be 90% accurate when patient is actually infected, and 95% accurate if patient is actually not infected. Veronica was tested once, and the result was positive. Given the fact that she tested positive once, find the probability that she is infected.

1.b) A random variable X has a Poisson distribution with mean 4. Find $E(X|X > 2)$.

1.c) A soft-drink filling machine is regulated so that it discharges an average of 11 ounces per cup. The amount of drink discharged is normally distributed with a standard deviation of 1 ounce. Assume that the amount of drink filled by the machine is not serially correlated. If 12-ounce cups are used for the next 10 drinks, what is the probability that less than 2 cups will overflow?

Problem #02 [11 points]

Let X and Y be discrete random variables with a joint probability distribution given by

$$f(x,y) = \begin{cases} \dfrac{x+2y}{18} & \text{for } (x,y) = (0,1),\ (0,2),\ (1,2),\ (1,3) \\ 0 & \text{otherwise.} \end{cases}$$

2.a) Find $P(X + Y \leq 2.1)$.

2.b) Find $Var(Y|X = 1)$.

2.c) Find the joint distribution of $U = X + Y$and $V = X - Y$.

2.d) Find the marginal distribution and the cumulative distribution function (CDF) of U.

2.e) Find the moment generating function of U.

Problem #03 [11 points]

The joint probability density function (PDF) of X and Y is given by

$$f(x,y) = \begin{cases} kx & \text{for } 0 < y < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

3.a) Verify that $k = 3$. (No partial credits will be given for this sub-problem.)

3.b) Find $P(X < 2Y)$.

3.c) Find $E\left(\frac{Y}{X}\right)$.

3.d) Find the conditional variance of Y given $X = x$, where $0 < x < 1$.

3.e) Are X and Y independent? Verify.

Problem #04 [11 points]

Let X and Y be independent random variables with a continuous uniform distribution on the interval (0, 1).

4.a) Find the CDF and PDF of the random variable $W = -2\ln(X)$. What are the name and the parameter value(s) of the probability distribution of W?

4.b) Define $U = X + Y$. Find the CDF of U.

4.c) Define $V = X - Y$. Find the joint PDF of U and V. Specify and draw a two dimension graph of the support of the joint PDF.

4.d) Find the correlation between X and V.

Problem #05 [8 points]

A random variable X has a pdf given by, for $\alpha > 0$ and $\beta > 0$,

$$f(x) = \begin{cases} kx^{\alpha-1}e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

5.a) Find k in terms of the distribution parameters.

5.b) **Derive** the moment generating function of X.

5.c) Use your result in part (5.b) to find the mean and variance of X.

5.d) Use your result in part (5.b) to find the moment generating function of an exponential random variable with a mean of λ.

5.e) Use your result in part (5.c) to find the mean and variance of a chi-square random variable with n degrees of freedom.

Part B [The parts of each problem are not necessarily given equal weight.]

Problem #06 [7 points]

Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution $f(x;\theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$.

6.a) Show that the likelihood ratio test of $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$ is based upon the statistic $Y = \Sigma_{i=1}^n X_i$. Show that the rejection region is of the form $Y \leq k_1$ or $Y \geq k_2$. Sketch the graph of the rejection region as a function of y.

6.b) Obtain the null distribution of $Y = \Sigma_{i=1}^n X_i$. For $\theta_0 = 0.5$ and $n = 8$, find the significance level of the test that rejects H_0 if $Y \leq 1$ or $Y \geq 7$.

Problem #07 [10 points]

7.a) Suppose X_1 and X_2 constitute a random sample of size 2 from the population with probability density function $f(x;\theta) = \theta x^{\theta-1}$, for $0 < x < 1$ and zero otherwise. Suppose the critical region $x_1 x_2 \leq 0.1$ is used to test $H_0: \theta = 2$ against $H_1: \theta = 1$.

i. Sketch the rejection region and find the power of the test in term of the parameter θ.

ii. What is the significance level of the test? What is the power of the test at $\theta = 1$?

iii. Comment on the power in (ii) relative to the significance level in (ii).

iv. Suppose the power in (ii) is too small, suggest two ways to increase the power. Which method, if any, is better and why?

7.b) State the Neyman-Pearson theorem.

7.c) Suppose a random variable X has the gamma probability density function $f(x;\theta) = \theta^{-2} x e^{-x/\theta}$ for $x > 0$. Consider the simple null hypothesis $H_0: \theta = \theta_0$ against the alternative hypothesis $H_1: \theta < \theta_0$. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from the distribution. Use the Neyman-Pearson theorem to find the most powerful critical region of size α.

Problem #08 [10 points]

8.a) Let X be a random variable of the discrete type with probability mass function $p(x)$, which is positive on the nonnegative integers and is equal to zero elsewhere. Show that $E(X) = \Sigma_{x=0}^{\infty}[1-F(x)]$, where $F(x)$ is the cumulative distribution function of X.

8.b) Define each of the following concepts in respect of a sequence $\{X_n\}$, $(n \geq 1)$ of random variables: **i.** Convergence in probability. **ii.** Convergence in distribution.

8.c) Let X_n have a beta distribution with parameters $\alpha = \beta = n$. The probability density function for X_n is given as $\Gamma(2n)[\Gamma(n)\Gamma(n)]^{-1}x^{n-1}(1-x)^{n-1}$ for $0 < x < 1$. Prove that X_n converges in probability to a constant k. Find the value of k. Furthermore, show that $1 - X_n / 2$ converges in probability to a constant c. Find the value of c.

8.d) If X_n converges in probability to a random variable X and Y_n converges in probability to a random variable Y, prove that $X_n + Y_n$ converges in probability to $X + Y$.

Problem #09 [11 points]

9.a) Define each of the following and give an example to illustrate your definition.

i. Sufficient statistic **ii.** Ancillary statistic

9.b) Suppose a random sample of size n is taken from a generalized negative binomial distribution (GNBD) with probability function $f(x;\theta) = \frac{m}{m+4x}\binom{m+4x}{x}\theta^x(1-\theta)^{m+3x}$, $x = 0,1,2,3,\ldots$,

where $0 < \theta < 0.25$ and $m > 0$ are parameters. The population mean μ and variance σ^2 for the distribution are $\mu = m\theta(1-4\theta)^{-1}$ and $\sigma^2 = m\theta(1-\theta)(1-4\theta)^{-3}$ respectively.

i. If the parameter m is known, find the moment estimator of θ.

ii. If the parameter m is known, find the maximum likelihood estimator of θ.

iii. Find the moment estimators of θ and m. [Hint: Parameter θ has two roots.]

iv. Determine the values of θ for which the population variance σ^2 is greater than the population mean μ. Based on your answer, is σ^2 always greater than μ for the GNBD? Explain. Using this result or otherwise, determine the correct estimator of θ from the two roots in (iii).

Problem #10 [12 points]

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \theta),\ 0 < \theta < \infty,$ where *m* is unknown. Let $Y = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ and let $L[\theta, \delta(y)] = [\theta - \delta(y)]^2$ be the squared-error loss function. We consider decision functions of the form $\delta(y) = by,$ where *b* does not depend upon *y*.

10.a) Show that $R(\theta, \delta) = (\theta^2 / n^2)[(n^2 - 1)b^2 - 2n(n-1)b + n^2].$

10.b) Using your result in (a), show that $b = \frac{n}{n+1}$ yields a minimum risk decision function of the form $\delta(y)$.

10.c) Is the decision function $\delta(Y) = \frac{nY}{n+1}$ unbiased estimator of θ?

10.d) With decision function $\delta(y) = \frac{ny}{n+1}$ and $0 < \theta < \infty,$ determine $\max_{\theta} R(\theta, \delta)$ if it exists.

10.e) Is the decision function in (c) asymptotically unbiased for θ?

STANDARD NORMAL DISTRIBUTION TABLE

Entries represent $\Pr(Z \leq z)$. The value of z to the first decimal is given in the left column. The second decimal is given in the top row.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000