NTHU MATH 2820 Midterm Examination - A April 22, 2025

Note. There are 6 problems in total. **For problems 3 to 6, to ensure consideration for partial scores, write down necessary intermediate steps.** Correct answers with inadequate or no intermediate steps **may result in zero credit**.

Some useful formula.

- Suppose that X follows a binomial(n, p) distribution. The probability mass function (pmf) of X is
$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$
The mean of X is np, and the variance of X is $np(1-p)$.

- Suppose that X follows a hyper-geometric(r, n, m) distribution. The pmf of X is
$$p(x) = \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}}, \text{ for } x = 0, 1, \ldots, \min(r, n), \text{ and } r - x \leq m.$$
The mean of X is $\frac{rn}{n+m}$, and the variance of X is $\frac{rnm(n+m-r)}{(n+m)^2(n+m-1)}$.

- Suppose that X follows a Poisson(λ) distribution. The pmf of X is
$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ for } x = 0, 1, 2, \ldots.$$
The mean of X is λ, and the variance of X is λ.

- Suppose that X follows a uniform(a, b) distribution. The probability density function (pdf) of X is
$$f(x) = \frac{1}{b-a}, \text{ for } a < x < b,$$
and zero otherwise. The mean of X is $\frac{a+b}{2}$, and the variance of X is $\frac{(b-a)^2}{12}$.

- Suppose that X follows an exponential(λ) distribution. The pdf of X is
$$f(x) = \lambda e^{-\lambda x}, \text{ for } 0 < x < \infty,$$
and zero otherwise. The mean of X is $1/\lambda$, and the variance of X is $1/\lambda^2$.

- Suppose that X follws a gamma(α, λ) distribution. The pdf of X is
$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \text{ for } 0 < x < \infty,$$
and zero otherwise. The mean of X is α/λ, and the variance of X is α/λ^2.

---

1. (*14 pts in total. for each question, 2 pts if correct; 0 pt if blank; −1 pt if wrong*) For the following statements, please answer true or false.

   (a) Suppose that $X_1, \dots, X_n$ is an independent and identically distributed (i.i.d.) sample from a Cauchy distribution. Let $\overline{X}_n = (\sum_{i=1}^{n} X_i)/n$. By the central limit theorem, $\overline{X}_n$ converges in distribution to a normal distribution as $n \to \infty$.

   (b) A random variable X with exponential distribution must obey the memoryless property, i.e., $P(X > s + t | X > s) = P(X > t)$, for $s, t > 0$.

   (c) Let $X_1, \dots, X_n$ be an i.i.d. sample from a distribution with cdf F, which has a finite mean and finte variance. When n is large, a histogram of the values in the sample will be approximately normal, even if F is not normal.

   (d) Under regularity conditions, the variance of maximum likelihood estimator (MLE) can always reach the Cramer-Rao lower bound (i.e., the ratio of the MLE's variance to the Cramer-Rao lower bound converges to one) when the sample size tends to infinity.

   (e) The minimum sufficient statistics contain no ancillary information in the data.

   (f) According to the definition of MLE, it is impossible that the estimated value of MLE falls outside the parameter space.

   (g) Because the likelihood has the same form as the joint pdf/pmf, the integration/summation of the likelihood over the parameter space should equal one.

2. (*15pts, 5pts for each*) For each of the random observations X given below,

   - determine the type of the joint distribution (e.g., normal, exponential, gamma, beta, uniform, Poisson, hyper-geometric, Bernoulli, binomial, multinomial, negative binomial, geometric, or others) that best models X, and
   - identify the known and unknown parameter(s) in the chosen distribution.

   (a) A certain electronics company produces a particular type of vacuum tube. Suppose that, on the average, $100p$ tubes of 100 are defective, where p is unknown and $0 < p < 1$. The company packs the tubes in boxes of 20. The company offers a money-back guarantee for a box of tubes if the box contains at least 2 defective tubes. Suppose that the company manufactures and sells n boxes of tubes each month, where n is a fixed and known number. Let Z_i, $i = 1, \dots, k$, denote the number of boxes returned among the n boxes manufactured in the ith month. Set $X = (Z_1, \dots, Z_k)$.

   (b) Let Z equal the time (in minutes) between calls that are made over the public safety radio. On some days, and during a period of one hour on each day, the following observations of Z were made: $Z_1, \dots, Z_k$, where k is a constant determined before the collection of the data. Suppose that the calls arrive randomly in accordance with an approximate Poisson process with an unknown constant occurrence rate λ (which represents the average number of calls per minute). Set $X = (Z_1, \dots, Z_k)$.

   (c) A biologist wants to estimate the size, N, of a population of turtles in a lake. She captures 20 turtles on her first visit to the lake, and marks their backs with paint. In the following k weeks, she returns to the lake once a week and captures 15 different turtles. Let Z_i, $1 \le i \le k$, denote the number of the turtles with paint on their backs in the capture of ith week. It is assumed that each turtle in the lake has the same chance of being captured. Set $X = (Z_1, \dots, Z_k)$.

3. Let $X_1, \ldots, X_n$ be i.i.d. random variables from uniform distribution $U(0, \theta)$ and set

$$Y_n = \max(X_1, \ldots, X_n) \quad \text{and} \quad Z_n = n(\theta - Y_n).$$

Notice that Y_n is the MLE of θ, but it always underestimates θ.

(a) (*8pts*) Prove that Y_n converges in probability to θ by showing that for any small enough $\epsilon > 0$,

$$\lim_{n\to\infty} P(|Y_n - \theta| < \epsilon) = 1.$$

(b) (*8pts*) Prove that Z_n converge in distribution to Z, where Z has the exponential(λ) distribution with parameter $\lambda = 1/\theta$.

[**Hint**. **(i)** The cdf of exponential(λ) is $1 - e^{-\lambda x}$, for $x \geq 0$. **(ii)** $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$.]

4. Let $X_1, \ldots, X_n$ be i.i.d. random variables with the pdf

$$f(x|\theta) = (\theta + 1)x^\theta, \quad 0 \leq x \leq 1,$$

where $-1 < \theta < \infty$.

(a) (*2pts*) Examine whether the pdf is a one-parameter exponential family.

(b) (*5pts*) Find the method of moments estimator of θ.

(c) (*6pts*) Find the MLE of θ.

(d) (*4pts*) Show that the Fisher information contained in $X_1, \ldots, X_n$ is

$$\frac{n}{(\theta + 1)^2}.$$

(e) (*3pts*) Identify the asymptotic sampling distribution of the MLE.

5. Consider the following method of estimating λ for a sample of i.i.d. Poisson random variables, i.e., $X_1, \ldots, X_n$ are i.i.d. from Poisson(λ). Observe that

$$p_0 = P(X_1 = 0) = e^{-\lambda}.$$

Letting Y denote the number of zero from the sample $X_1, \ldots, X_n$, the λ might be estimated by

$$\tilde{\lambda} = -\log\left(\frac{Y}{n}\right).$$

(a) (*2pts*) Show that $Y \sim$ binomial(n, p_0). [**Hint.** Use indicator functions $I_i = I(X_i = 0)$, $i = 1, \ldots, n$.]

(b) (*6pts*) Use the δ-method to obtain approximate expressions for the variance and the bias of the estimator $\tilde{\lambda}$. [**Hint.** δ-method: if Z and W are random variables such that $Z = g(W)$, then $\mathrm{E}(Z) \approx g(\mathrm{E}(W)) + \frac{1}{2}g''(\mathrm{E}(W))\mathrm{Var}(W)$ and $\mathrm{Var}(Z) \approx [g'(\mathrm{E}(W))]^2\mathrm{Var}(W)$.]

(c) (*4pts*) The MLE of λ is $\hat{\lambda} = \overline{X}$. Use the results of (b) to show that the relative efficiency of the estimator $\tilde{\lambda}$ to the MLE $\hat{\lambda}$ is

$$\frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}}.$$

(d) (*3pts*) Show that the relative efficiency in (c) is always smaller than 1. [**Hint.** $1 - e^{-\lambda} - \lambda e^{-\lambda} = P(X_1 \geq 2)$.]

(e) (*5pts*) When the sample size n is large, which of $\tilde{\lambda}$ and $\hat{\lambda}$ is better in terms of mean square error? Explain.

6. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a uniform distribution on the interval $(\theta, 2\theta)$, where $\theta > 0$. Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of $X_1, \ldots, X_n$.

   (a) (*4pts*) Use the factorization theorem to show that $(X_{(1)}, X_{(n)})$ is sufficient. [**Hint.** Use an indicator function of θ to write the joint pdf of $X_1, \ldots, X_n$.]

   (b) (*4pts*) Show that $(X_{(1)}, X_{(n)})$ is minimal sufficient. [**Hint.** Express $X_{(1)}$ and $X_{(n)}$ as functions of the likelihood.]

   (c) (*4pts*) Show that $X_{(n)}/X_{(1)}$ is an ancillary statistic. [**Hint.** Let $Z_i = X_i/\theta$, $i = 1, \ldots, n$. Then, $Z_{(n)}/Z_{(1)} = X_{(n)}/X_{(1)}$.]

   (d) (*3pts*) Is $(X_{(1)}, X_{(n)})$ complete? Explain.