(A1, B1) (*14pts, 2pts for each*)

Exam A.
(a) False (b) True (c) False (d) True
(e) False (f) True (g) False

Exam B.
(a) True (b) False (c) False (d) False
(e) False (f) True (g) True

(A2, B2) (*15pts, 5pts for each*)

Exam A.

(a) $Z_1, \ldots, Z_k$ are i.i.d. random variables from binomial(n, q) distribution, where $q = 1 - (1-p)^{20} - 20p(1-p)^{19}$, and p is an unknown parameter.

(b) $Z_1, \ldots, Z_k$ are i.i.d. random variables from exponential(λ) distribution (or gamma$(1, \lambda)$ distribution), where λ is an unknown parameter.

(c) $Z_1, \ldots, Z_k$ are i.i.d. random variables from hyper-geometric$(15, 20, N-20)$ distribution, where N is an unknown parameter.

(**Note**: If "independence" is not stated, your answer only addresses the marginal distributions of $Z_1, \ldots, Z_k$. Since marginal distributions alone are not sufficient to uniquely determine the joint distribution of $Z_1, \ldots, Z_k$, some point will be deducted.)

Exam B.

(a) $Z_1, \ldots, Z_k$ are i.i.d. random variables from exponential(λ) distribution (or gamma$(1, \lambda)$ distribution), where λ is an unknown parameter.

(b) $Z_1, \ldots, Z_k$ are i.i.d. random variables from hyper-geometric$(10, 30, N-30)$ distribution, where N is an unknown parameter.

(c) $Z_1, \ldots, Z_k$ are i.i.d. random variables from binomial(n, q) distribution, where $q = 1 - (1-p)^{10} - 10p(1-p)^9$, and p is an unknown parameter.

(**Note**: If "independence" is not stated, your answer only addresses the marginal distributions of $Z_1, \ldots, Z_k$. Since marginal distributions alone are not sufficient to uniquely determine the joint distribution of $Z_1, \ldots, Z_k$, some point will be deducted.)

(A3, B6) (*16pts*)

(a) (*8pts*) The cdf of Y_n is

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \le y) = P(X_1 \le y, \ldots, X_n \le y) \\ &= \prod_{i=1}^{n} P(X_i \le y) = \left(\frac{y}{\theta}\right)^n \end{aligned}$$

for $0 \le y \le \theta$. So, for small enough ϵ, say $\epsilon \in (0, \theta)$,

$$\begin{aligned} P(|Y_n - \theta| < \epsilon) &= P(\theta - \epsilon < Y_n < \theta + \epsilon) = P(\theta - \epsilon < Y_n < \theta) = 1 - P(Y_n \le \theta - \epsilon) \\ &= 1 - F_{Y_n}(\theta - \epsilon) = 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n \\ &= 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n \longrightarrow 1, \quad \text{as } n \to \infty. \end{aligned}$$

(Note. Compare the result with the consistent property of MLE.)

(b) (*8pts*) Because $0 < Z_n < n\theta$, the cdf of Z_n is

$$\begin{aligned} F_{Z_n}(z) &= P(Z_n \le z) = P(n(\theta - Y_n) \le z) = P\left(\theta - \frac{z}{n} \le Y_n\right) = 1 - P\left(Y_n < \theta - \frac{\epsilon}{n}\right) \\ &= 1 - F_{Y_n}\left(\theta - \frac{z}{n}\right) = 1 - \left(\frac{\theta - z/n}{\theta}\right)^n \\ &= 1 - \left(1 + \frac{(-z/\theta)}{n}\right)^n \end{aligned}$$

for $0 < z < n\theta$. Because

$$F_{Z_n}(z) = 1 - \left(1 + \frac{(-z/\theta)}{n}\right)^n \longrightarrow 1 - e^{-z/\theta}, \quad \text{as } n \to \infty,$$

for any $z \in (0, \infty)$, and $1 - e^{-z/\theta}$ is the cdf of the exponential$(1/\theta)$ distribution, it is proved that Z_n converge in distribution to Z.
(Note. Compare the result with the asymptotic normality property of MLE. Can you see the difference between them?)

(A4, B5) (*20pts*)

(a) (*2pts*) The pdf can be written as

$$\begin{aligned} f(x|\theta) &= \exp\left[\log\left((\theta+1)x^\theta\right)\right] \\ &= \exp\left[\theta\ \log(x) + \log(\theta+1)\right], \end{aligned}$$

for $x \in [0, 1]$, where the support $[0, 1]$ does not dependent on θ. This is a one-parameter exponential with $c(\theta) = \theta$, $T(X) = \log(X)$, and $d(\theta) = \log(\theta + 1)$. Notice that $\sum_{i=1}^n \log(X_i)$ is a sufficient and complete statistic.

(b) (*5pts*) Because

$$\mu_1 = \mathrm{E}_\theta(X_1) = \int_0^1 x\ (\theta+1)x^\theta\ dx = \frac{\theta+1}{\theta+2}\ x^{\theta+2}\Big|_0^1 = \frac{\theta+1}{\theta+2} \quad \Rightarrow \quad \theta = \frac{2\mu_1 - 1}{1 - \mu_1},$$

the moment estimator of θ is

$$\tilde{\theta} = \frac{2\overline{X} - 1}{1 - \overline{X}}.$$

(c) (*6pts*) Because the joint pdf is:

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n \left[(\theta+1)x_i^\theta\right] = (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta,$$

the log-likelihhod function is:

$$l(\theta; x_1, \ldots, x_n) = \log f(x_1, \ldots, x_n|\theta) = n\log(\theta+1) + \theta \sum_{i=1}^n \log(x_i).$$

By setting

$$l'(\theta; x_1, \ldots, x_n) = \frac{n}{\theta+1} + \sum_{i=1}^n \log(x_i) = 0,$$

we can get the solution is

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log(x_i)} - 1.$$

Because

$$l''(\theta; x_1, \ldots, x_n) = -n(\theta+1)^{-2} < 0, \quad \text{for any } \theta,$$

$\hat{\theta}$ is the MLE. Notice that the MLE $\hat{\theta}$ is a function of the sufficient and complete statistic $\sum_{i=1}^{n} \log(X_i)$.

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

$$I_{X_1,\cdots,X_n}(\theta) = \mathrm{E}_\theta[-l''(\theta; X_1, \ldots, X_n)] = \mathrm{E}_\theta\left[\frac{n}{(\theta+1)^2}\right] = \frac{n}{(\theta+1)^2}.$$

Similar calculation can be used to show that the Fisher information contained in *a single observation* X_i is

$$I_{X_1}(\theta) = \frac{1}{(\theta+1)^2}.$$

Notice that $I_{X_1,\cdots,X_n}(\theta) = n\, I_{X_1}(\theta)$. An alternative method to get $I_{X_1}(\theta)$ is as follows. Let $Y = -\log(X_1)$ ($\Rightarrow X_1 = \exp(-Y)$). Then, the pdf of Y is

$$f_Y(y) = f_{X_1}(e^{-y}) \left|\frac{dx_1}{dy}\right| = (\theta+1)e^{-\theta y} \left|-e^{-y}\right| = (\theta+1)e^{-(\theta+1)y},$$

for $0 < y < \infty$. Because Y follows a gamma$(1, \theta+1)$, i.e., exponential$(\theta+1)$, distribution, we have

$$\mathrm{E}_\theta(Y) = \frac{1}{\theta+1} \text{ and } \mathrm{Var}_\theta(Y) = \frac{1}{(\theta+1)^2}.$$

So,

$$\begin{aligned} I_{X_1}(\theta) &= \mathrm{E}_\theta[(l'(\theta; X_1))^2] = \mathrm{E}_\theta\left[\left(\frac{1}{\theta+1} + \log(X_1)\right)^2\right] \\ &= \mathrm{E}_\theta\left[\left(\frac{1}{\theta+1} - Y_1\right)^2\right] = \mathrm{E}_\theta\left[(\mathrm{E}(Y_1) - Y_1)^2\right] = \mathrm{Var}_\theta(Y_1) = \frac{1}{(\theta+1)^2}, \end{aligned}$$

which is identical to the result given above. Actually, from the above calculation, it can also be shown that $\sum_{i=1}^{n} -\log(X_i)$ follows a gamma$(n, \theta+1)$ distribution.

(e) (*3pts*) The asymptotic variance of the MLE $\hat{\theta}$ is

$$\frac{1}{\mathrm{E}_\theta[-l''(\theta; X_1, \ldots, X_n)]} = \frac{1}{n\, I_{X_1}(\theta)} = \frac{(\theta+1)^2}{n},$$

and the asymptotic distribution of the MLE is Normal distribution with mean θ and variance $(\theta+1)^2/n$.

(A5, B4) *(20pts)*

(a) *(2pts)* For $i = 1, \ldots, n$, let $I_i = I(X_i = 0)$, where I is an indicator function. Then $I_1, \ldots, I_n$ are i.i.d. random variables from Bernoulli(p_0), and $Y = \sum_{i=1}^n I_i$. So, $Y \sim$ binomial(n, p_0).

(b) *(6pts)* Let $g(y) = -\log(y/n)$. Then, we have $g'(y) = -y^{-1}$, and $g''(y) = y^{-2}$. Because $Y \sim$ binomial(n, p_0), we have $\mathrm{E}(Y) = np_0$, and $\mathrm{Var}(Y) = np_0(1-p_0)$. By the δ-method, we can get

$$\begin{aligned}
\mathrm{E}(\tilde{\lambda}) = \mathrm{E}[g(Y)] &\approx g(\mathrm{E}(Y)) + \frac{1}{2}g''(\mathrm{E}(Y))\mathrm{Var}(Y) \\
&= g(np_0) + \frac{1}{2}g''(np_0)\,np_0(1-p_0) \\
&= -\log(p_0) + \frac{1}{2}\frac{1}{(np_0)^2}np_0(1-p_0) \\
&= \lambda + \frac{1-p_0}{2np_0} = \lambda + \frac{1-e^{-\lambda}}{2ne^{-\lambda}},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}(\tilde{\lambda}) = \mathrm{Var}(g(Y)) \approx [g'(\mathrm{E}(Y))]^2\mathrm{Var}(Y) \\
&= [g'(np_0)]^2\,np_0(1-p_0) = \left(\frac{-1}{np_0}\right)^2 np_0(1-p_0) = \frac{1-p_0}{np_0} = \frac{1-e^{-\lambda}}{ne^{-\lambda}}.
\end{aligned}$$

The approximate bias is $\mathrm{E}(\tilde{\lambda}) - \lambda \approx \frac{1-p_0}{2np_0} = \frac{1-e^{-\lambda}}{2ne^{-\lambda}}$, which has a rate of convergence $O(n^{-1})$.

(c) *(4pts)* The mean and variance of the MLE $\hat{\lambda}$ are λ (i.e., $\hat{\lambda}$ is unbiased) and λ/n, respectively. By the results of (b), the relative efficiency of $\tilde{\lambda}$ to $\hat{\lambda}$ is

$$\mathrm{eff}(\tilde{\lambda}, \hat{\lambda}) = \frac{\mathrm{Var}(\hat{\lambda})}{\mathrm{Var}(\tilde{\lambda})} = \frac{\lambda/n}{(1-p_0)/(np_0)} \approx \frac{\lambda}{n} \times \frac{ne^{-\lambda}}{1-e^{-\lambda}} = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}.$$

(d) *(3pts)* Because $1 - e^{-\lambda} - \lambda e^{-\lambda} = P(X_1 \geq 2) > 0$, for any $\lambda > 0$, we have

$$1 - e^{-\lambda} > \lambda e^{-\lambda} \;\Rightarrow\; \mathrm{eff}(\tilde{\lambda}, \hat{\lambda}) = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} < 1, \quad \text{for any } \lambda > 0.$$

It shows that $\tilde{\lambda}$ has a large variance than $\hat{\lambda}$, i.e., $\hat{\lambda}$ is a better estimator than $\tilde{\lambda}$ in terms of variance.

(e) *(5pts)* For mean square error (MSE) of an estimator $\hat{\theta}$, notice that

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{bias}(\hat{\theta})]^2.$$

For the MLE $\hat{\lambda}$, its bias is zero so that MSE$(\hat{\lambda})$=Var$(\hat{\lambda})$. For the estimators $\tilde{\lambda}$, by the results of (b), its bias^2 $(= \frac{(1-e^{-\lambda})^2}{4n^2e^{-2\lambda}})$ is $O(n^{-2})$ and its variance $(=\frac{1-e^{-\lambda}}{ne^{-\lambda}})$ is $O(n^{-1})$ when the sample size n is large, so that we can ignore the bias^2 term and only consider Var$(\tilde{\lambda})$. By the result of (d), Var$(\hat{\lambda}) <$ Var$(\tilde{\lambda})$ for any $\lambda > 0$ when n is large. We can therefore conclude that when n is large, the MLE $\hat{\lambda}$ is a better estimator than $\tilde{\lambda}$ in terms of MSE.

(A6, B3) *(15pts)*

(a) *(4pts)* Because $X_{(1)}, \ldots, X_{(n)}$ are the order statistics of $X_1, \ldots, X_n$, we have $\theta < X_{(1)} < \cdots < X_{(n)} < 2\theta$ ($\Rightarrow X_{(n)}/2 < \theta < X_{(1)}$). We can write the joint pdf of $X_1, \ldots, X_n$ as:

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{(\theta,\, 2\theta)}(x_i) = \frac{1}{\theta^n} I_{\left(x_{(n)}/2,\, x_{(1)}\right)}(\theta), \tag{I}$$

where I is the indicator function. Because the joint pdf can be written as a function of $X_{(1)}$, $X_{(n)}$, and θ, by the factorization theorem, $\left(X_{(1)}, X_{(n)}\right)$ is sufficient.

(b) *(4pts)* Notice that by (I), we can express $X_{(1)}$ and $X_{(n)}$ using the likelihood, which treats (I) as a function of θ with $x_{(1)}$ and $x_{(n)}$ being fixed, as follows:

$$x_{(1)} = \sup\{\theta : f(x_1, \ldots, x_n|\theta) > 0\}, \tag{II}$$

and

$$x_{(n)} = 2 \times \inf\{\theta : f(x_1, \ldots, x_n|\theta) > 0\}. \tag{III}$$

For a sufficient statistics $\boldsymbol{T}$, by factorization theorem, there exist functions g and h such that

$$f(x_1, \ldots, x_n|\theta) = g(\boldsymbol{t}, \theta)h(x_1, \ldots, x_n),$$

where $h(x_1, \ldots, x_n) > 0$. Therefore, by (II) and (III), we can express $X_{(1)}$ and $X_{(n)}$ as functions of $\boldsymbol{T}$ as follows:

$$x_{(1)} = \sup\{\theta : g(\boldsymbol{t}, \theta) > 0\},$$

and

$$x_{(n)} = 2 \times \inf\{\theta : g(\boldsymbol{t}, \theta) > 0\}.$$

It shows that the sufficient statistics $\left(X_{(1)}, X_{(n)}\right)$ can be expressed as a function of $\boldsymbol{T}$ for any sufficient statistics $\boldsymbol{T}$, i.e., $\left(X_{(1)}, X_{(n)}\right)$ is minimal sufficient.

(c) *(4pts)* Let $Z_i = X_i/\theta$, $i = 1, \ldots, n$. Then, $Z_1, \ldots, Z_n$ are i.i.d. random variables from the uniform$(1, 2)$ distribution. Because the joint distribution of $Z_1, \ldots, Z_n$ is irrelevant to the parameter θ, the distributions of any transformations of $Z_1, \ldots, Z_n$, including the distribution of $Z_{(n)}/Z_{(1)}$ and the joint distribution of $(Z_{(1)}, Z_{(n)})$, are irrelevant to θ. Because $X_{(n)}/X_{(1)} = (\theta Z_{(n)})/(\theta Z_{(1)}) = Z_{(n)}/Z_{(1)}$, the statistic $X_{(n)}/X_{(1)}$ has a distribution irrelevant to θ, which shows $X_{(n)}/X_{(1)}$ is an ancillary statistic. Notice that $Z_{(1)}$ and $Z_{(n)}$ are random variables but not statistics (they are functions of data and parameter), while $X_{(1)}$ and $X_{(n)}$ are statistics.

(d) *(3pts)* Because we can find a non-constant function (i.e., $X_{(n)}/X_{(1)}$) of $X_{(1)}$ and $X_{(n)}$ such that the distribution of the function is irrelevant to θ, $\left(X_{(1)}, X_{(n)}\right)$ is not complete. [**Note**. A way to prove it by following the definition of completeness is as follows. Let $\mu = \mathrm{E}_\theta(X_{(n)}/X_{(1)})$. Then, μ is a constant over θ because $X_{(n)}/X_{(1)}$ is an ancillary statistic. That is, we can find a non-constant statistic $X_{(n)}/X_{(1)} - \mu$ from $X_{(1)}$ and $X_{(n)}$ such that $\mathrm{E}_\theta(X_{(n)}/X_{(1)} - \mu) = 0$, for any θ.]