(A1, B1) (*12pts, 2pts for each, 0pt if blank, −1 if wrong*)

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

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

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

Exam A.

(a) We can model $X_1, \ldots, X_{1000}$ as i.i.d. random variables from the exponential(λ) distribution, where λ is the average emission rate and $\lambda \approx \frac{10130}{12070} = 0.83927$.

(b) (X_1, X_2, X_3, X_4)

$$\sim \text{multinomial}(n, p_1, p_2, p_3, p_4),$$

where $n = 1000$ and p_1, p_2, p_3, p_4 are unknown parameters satisfying $p_1+p_2+p_3+p_4 = 1$.

(c) Because the death rates are not constant over the years, Poisson distribution with a constant rate might be inappropriate. It is better to model $X_1, \ldots, X_{10}$ as i.i.d. random variables from the generalized negative binomial (m, k) distribution given in LN, Ch8, p.65, rather than Poisson distribution. Both m and k are unknown parameters.

(**Note**: For (a) and (c), if "independence" is not stated, your answer only addresses the marginal distributions of $X_1, \ldots, X_k$. Since marginal distributions alone are not sufficient to uniquely determine the joint distribution of $X_1, \ldots, X_k$, some point will be deducted.)

Exam B.

(a) (X_1, X_2, X_3, X_4)

$$\sim \text{multinomial}(n, p_1, p_2, p_3, p_4),$$

where $n = 800$ and p_1, p_2, p_3, p_4 are unknown parameters satisfying $p_1+p_2+p_3+p_4 = 1$.

(b) Because the death rates are not constant over the years, Poisson distribution with a constant rate might be inappropriate. It is better to model $X_1, \ldots, X_{10}$ as i.i.d. random variables from the generalized negative binomial (m, k) distribution given in LN, Ch8, p.65, rather than Poisson distribution. Both m and k are unknown parameters.

(c) We can model $X_1, \ldots, X_{1000}$ as i.i.d. random variables from the exponential(λ) distribution, where λ is the average emission rate and $\lambda \approx \frac{8140}{10863} = 0.74933$.

(**Note**: For (b) and (c), if "independence" is not stated, your answer only addresses the marginal distributions of $X_1, \ldots, X_k$. Since marginal distributions alone are not sufficient to uniquely determine the joint distribution of $X_1, \ldots, X_k$, some point will be deducted.)

(A3, B5) *(18pts)*

(a) *(4pts)* The likelihood is

$$L(\theta) = \prod_{i=1}^{n} I_{[\theta,\theta+1]}(X_i) = I_{[Z_n-1,Y_n]}(\theta),$$

where $I_{[a,b]}(u) = 1$ if $a \le u \le b$, and 0, otherwise. The last equality holds because we have $L(\theta) = 1$ if and only if all sample points satisfy $\theta < X_i < \theta + 1$ for all i, which is equivalent to $\theta < Y_n$ and $Z_n < \theta + 1$, or equivalently, $Z_n - 1 < \theta < Y_n$. Thus the maximum of the likelihood on $-\infty < \theta < \infty$ is 1, and the likelihood is maximized for every θ in the interval $[Z_n - 1, Y_n]$. Therefore, every value in $[Z_n - 1, Y_n]$ is an MLE, and hence the MLE is not unique.

(b) *(6pts)* Because $Z_n - 1 \le \theta$, for any $\varepsilon > 0$, we have

$$P(|(Z_n - 1) - \theta| < \varepsilon) \;=\; P(\theta - (Z_n - 1) < \varepsilon) = P(Z_n > \theta + 1 - \varepsilon).$$

For $0 < \varepsilon < 1$,

$$\begin{aligned} P(Z_n \le \theta + 1 - \varepsilon) &= P(X_1 \le \theta + 1 - \varepsilon, \ldots, X_n \le \theta + 1 - \varepsilon) \\ &= \prod_{i=1}^{n} P(X_i \le \theta + 1 - \varepsilon) = \prod_{i=1}^{n} [(\theta + 1 - \varepsilon) - \theta] = (1 - \varepsilon)^n. \end{aligned}$$

Hence,

$$\lim_{n\to\infty} P(|(Z_n - 1) - \theta| < \varepsilon) = \lim_{n\to\infty} [1 - (1 - \varepsilon)^n] = 1.$$

Therefore, $Z_n - 1 \xrightarrow{P} \theta$.
(FYI, we can similarly prove that $Y_n \xrightarrow{P} \theta$. Actually, any MLE $\hat{\theta}$ in $[Z_n - 1, Y_n]$ is a consistent estimator.)

(c) *(8pts)* Let $W_n = n(Y_n - \theta)$. For $0 \le x \le n$, the cdf of W_n is

$$\begin{aligned} P(W_n \le x) &= 1 - P(W_n > x) = 1 - P\big(n(Y_n - \theta) > x\big) \\ &= 1 - P\left(Y_n > \theta + \frac{x}{n}\right) = 1 - P\left(X_1 > \theta + \frac{x}{n}, \ldots, X_n > \theta + \frac{x}{n}\right) \\ &= 1 - \prod_{i=1}^{n} P\left(X_i > \theta + \frac{x}{n}\right) = 1 - \prod_{i=1}^{n} \left[(\theta + 1) - \left(\theta + \frac{x}{n}\right)\right] \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \end{aligned}$$

Because $\left(1 - \frac{x}{n}\right)^n \to e^{-x}$ when $n \to \infty$, we have

$$\lim_{n\to\infty} P(W_n \le x) = 1 - e^{-x}.$$

for $x \ge 0$. This is exactly the cdf of exponential(1). Hence, $n(Y_n - \theta) \xrightarrow{d}$ exponential(1).
(FYI, we can similarly prove that $n(\theta - (Z_n - 1)) \xrightarrow{d}$ exponential(1).)

(A4, B3) (*20pts*)

(a) (*4pts*) The joint pdf is:

$$\prod_{i=1}^{n} f(y_i|\theta) = \frac{1}{\theta^n} r^n \left(\prod_{i=1}^{n} y_i\right)^{r-1} e^{-\frac{\sum_{i=1}^{n} y_i^r}{\theta}} = g(t,\theta)h(y_1,\ldots,y_n),$$

where

$$t = \sum_{i=1}^{n} y_i^r, \quad g(t,\theta) = \frac{1}{\theta^n} e^{-\frac{t}{\theta}}, \quad \text{and} \quad h(y_1,\ldots,y_n) = r^n \left(\prod_{i=1}^{n} y_i\right)^{r-1}.$$

By the factorization theorem, the statistic $T = \sum_{i=1}^{n} Y_i^r$ is sufficient.

(b) (*4pts*) The log-likelihood function is:

$$l(\theta) = -n\log(\theta) - \frac{\sum_{i=1}^{n} y_i^r}{\theta} + \log(h(y_1,\ldots,y_n)).$$

By solving the normal equation:

$$l'(\theta) = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} y_i^r}{\theta^2} = 0,$$

we obtain the solution: $\hat{\theta} = \frac{\sum_{i=1}^{n} y_i^r}{n} = T/n$. Because

$$l''(\theta)|_{\theta=\hat{\theta}} = \left[\frac{n}{\theta^2} - \frac{2\sum_{i=1}^{n} y_i^r}{\theta^3}\right]\Bigg|_{\theta=\hat{\theta}=T/n} = \frac{n^3}{T^2} - \frac{2n^3}{T^2} = -\frac{n^3}{T^2} < 0,$$

this solution is MLE.

(c) (*4pts*) Let $X = Y^r$, then $Y = X^{\frac{1}{r}}$ and $X > 0$. Because $|J| = |\frac{dy}{dx}| = \frac{1}{r}x^{\frac{1}{r}-1}$, for $x > 0$, the pdf of X is

$$f(x|\theta) = \left(\frac{1}{\theta}\right) r x^{\frac{r-1}{r}} e^{-\frac{x}{\theta}} \times \frac{1}{r} x^{\frac{1}{r}-1} = \frac{1}{\theta} e^{-\frac{x}{\theta}},$$

which is the pdf of exponential$(1/\theta)$.

(d) (*3pts*) Because $Y_1^r, Y_2^r, \ldots, Y_n^r$ are i.i.d. from exponential$(1/\theta)$ by (c), we have

$$E(\hat{\theta}) = E(T)/n = n \times E(Y_1^r)/n = n \times \theta/n = \theta,$$

and

$$Var(\hat{\theta}) = Var(T)/n^2 = n \times Var(Y_1^r)/n^2 = n \times \theta^2/n^2 = \theta^2/n.$$

Hence, the bias of $\hat{\theta}$ is $E(\hat{\theta}) - \theta = 0$, and the standard error of $\hat{\theta}$ is

$$\sqrt{Var(\hat{\theta})} = \sqrt{\theta^2/n} = \frac{\theta}{\sqrt{n}}.$$

(e) (*5pts*) Because

$$\frac{d^2}{d\theta^2}\log(f(y|\theta)) = \frac{1}{\theta^2} - \frac{2}{\theta^3}y^r,$$

and $\mathrm{E}(Y^r) = \mathrm{E}(X) = \theta$ (from (c)), we can get the Fisher information of one observation:

$$I(\theta) = -\mathrm{E}\left(\frac{1}{\theta^2} - \frac{2}{\theta^3}Y^r\right) = -\frac{1}{\theta^2} + \frac{2}{\theta^3}\mathrm{E}(Y^r) = -\frac{1}{\theta^2} + \frac{2}{\theta^3}\theta = \frac{1}{\theta^2}.$$

The Fisher information contained in $Y_1, \ldots, Y_n$ is $nI(\theta) = n/\theta^2$, and the asymptotic variance of the MLE is

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

(A5, B6) (*17pts*)

(a) (*6pts*) Let $g(t) = (1/t) - 1$. We have $g'(t) = -t^{-2}$. By δ-method,

$$\begin{aligned} Var(\hat{\theta}_1) &\approx Var(\overline{X}_n) \cdot \{g'[E(\overline{X}_n)]\}^2 = Var(\overline{X}_n) \cdot \frac{1}{[E(\overline{X}_n)]^4} \qquad \text{(I)} \\ &= \frac{\theta^2}{n(2\theta+1)(\theta+1)^2} \cdot (\theta+1)^4 \\ &= \frac{\theta^2(\theta+1)^2}{n(2\theta+1)}. \end{aligned}$$

(b) (*4pts*) From the Hint, we have

$$Var(\hat{\theta}_2) = \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} -\log(X_i)\right) = \frac{1}{n^2} \cdot n \cdot Var(-\log(X_1)) = \frac{\theta^2}{n}.$$

(Note: Two points are subtracted if you use Fisher information to get the answer (i.e., $\frac{1}{nI_{X_1}(\theta)}$) because it is the formula for asymptotic variance, rather than exact variance.)

(c) (*4pts*) The asymptotic efficiency is

$$\lim_{n\to\infty} \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} = \lim_{n\to\infty} \frac{\frac{\theta^2}{n}}{\frac{\theta^2(\theta+1)^2}{n(2\theta+1)}} = \frac{2\theta+1}{(\theta+1)^2} = \frac{2\theta+1}{\theta^2+2\theta+1} < 1.$$

(d) (*3pts*) When the sample size, n, is large enough, the relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is smaller than one, i.e., $\hat{\theta}_2$ has a smaller variance than $\hat{\theta}_1$. The estimator $\hat{\theta}_2$ is better (note that both estimators are consistent).

(A6, B4) (*18pts*)

(a) (*4pts*) The joint pdf of $X_1, \ldots, X_r$ is:

$$\begin{aligned} f(x_1, \ldots, x_r|p) &= \prod_{i=1}^{r} \left(\begin{array}{c} n \\ x_i \end{array} \right) p^{x_i}(1-p)^{n-x_i} \\ &= \left[\prod_{i=1}^{r} \left(\begin{array}{c} n \\ x_i \end{array} \right) \right] \left[p^{\sum_{i=1}^r x_i} \right] \left[(1-p)^{nr-\sum_{i=1}^r x_i} \right] \\ &= \left[\left(\frac{p}{1-p} \right)^{\sum_{i=1}^r x_i} \right] [(1-p)^{nr}] \left[\prod_{i=1}^{r} \left(\begin{array}{c} n \\ x_i \end{array} \right) \right] \\ &= \left\{ \exp \left[\left(\sum_{i=1}^{r} x_i \right) \cdot \log \left(\frac{p}{1-p} \right) \right] \right\} [(1-p)^{nr}] \left[\prod_{i=1}^{r} \left(\begin{array}{c} n \\ x_i \end{array} \right) \right]. \end{aligned}$$

This is an *exponential family* with $C(p) = \log\left(\frac{p}{1-p}\right)$ and $T(X_1, \ldots, X_r) = \sum_{i=1}^r X_i$. This implies T is sufficient and complete.

(b) (*4pts*)

$$E(U) = 1 \cdot P(U=1) = P(X_1 = 0) = \left(\begin{array}{c} n \\ 0 \end{array} \right) p^0(1-p)^n = (1-p)^n = \theta.$$

(c) (*6pts*) For $t \leq n(r-1)$,

$$\begin{aligned} P(U=1|T=t) &= \frac{P(X_1=0, T=\sum_{i=1}^r X_i = t)}{P(T=t)} = \frac{P(X_1=0, \sum_{i=2}^r X_i = t)}{P(T=t)} \\ &= \frac{P(X_1=0)P(\sum_{i=2}^r X_i = t)}{P(T=t)} \\ &= \frac{\left[\left(\begin{array}{c} n \\ 0 \end{array} \right) p^0(1-p)^n \right] \left[\left(\begin{array}{c} n(r-1) \\ t \end{array} \right) p^t(1-p)^{n(r-1)-t} \right]}{\left(\begin{array}{c} nr \\ t \end{array} \right) p^t(1-p)^{nr-t}} \\ &= \frac{\left(\begin{array}{c} n \\ 0 \end{array} \right) \left(\begin{array}{c} n(r-1) \\ t \end{array} \right)}{\left(\begin{array}{c} nr \\ t \end{array} \right)} = \frac{\left(\begin{array}{c} n(r-1) \\ t \end{array} \right)}{\left(\begin{array}{c} nr \\ t \end{array} \right)} \end{aligned}$$

(note that from the Hint, $\sum_{i=2}^r X_i \sim B(n(r-1), p)$ and $T = \sum_{i=1}^r X_i \sim B(nr, p)$). When $t > n(r-1)$, it is impossible that X_1 can be zero. Therefore $P(U=1|T=t) = 0$.

(d) (*4pts*) Because T is sufficient and complete for θ (note that $\theta = (1-p)^n$ is a one-to-one transformation of p) and U is an unbiased estimator of θ, by Rao-Blackwell and Lehmann-Scheffe theorems, $E(U|T)$ is a UMVUE. The UMVUE is:

$$E(U|T) = 1 \cdot P(U=1|T) = \begin{cases} \dfrac{\left(\begin{array}{c} n(r-1) \\ T \end{array} \right)}{\left(\begin{array}{c} nr \\ T \end{array} \right)}, & \text{if } T \leq n(r-1) \\ 0, & \text{if } T > n(r-1) \end{cases}.$$