3rd Homework

1. (4 points) Construct the Bayesian estimate of the parameter θ of a uniform distribution on the interval [0;θ]. Prove that the sample mean is not a minimax estimator:
a) if the parameter θ has the density $q(t) = \frac{1}{t^2}$ for $t > 1$;
b) if θ is uniformly distributed on [0,1].

2. (6 points) Find the Bayesian estimate for the parameter θ and check whether the estimator $\overline{X}$ or $(\overline{X})^{-1}$ is minimax, if:
a) $X_1, \ldots, X_n$ is a sample from a Bernoulli distribution with parameter θ, and the prior is a Beta distribution $B(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$;
b) a sample from $Exp(\theta)$, and θ has an exponential distribution with parameter β;
c) a sample from a Poisson distribution with parameter θ, where θ has an exponential distribution with parameter a;
d)*(1 point) a sample from a Poisson distribution with parameter θ, where θ takes values 1 and 2 with probabilities 1/3 and 2/3, respectively.

3. (1 points) Find the efficient estimator of the parameter θ among all unbiased estimators, if the sample comes from $Unif[0, \theta]$, using the known information about $X_{(n)}$.

4. (4 points) Let $X_1, \ldots, X_n$ be a sample from a two-parameter shifted exponential distribution with density

$$f_{\beta,\tau}(y) = \begin{cases} \beta^{-1} \exp\left(-\dfrac{y-\tau}{\beta}\right), & y > \tau, \\ 0, & y \le \tau, \end{cases} \qquad \beta > 0, \ \tau \in \mathbb{R}.$$

Find a sufficient statistic , check its completeness and find the efficient estimator among all unbiased estimators :
a) for the parameter τ if β is known;
b) for the parameter β if τ is known;
(*) (2 points) The same questions for the two-dimensional parameter $\boldsymbol{\theta} = (\beta, \tau)$.

5. (2 points) Let $X_1, \ldots, X_n$ be a sample from a uniform distribution on the finite set $\{1, \ldots, \theta\}$, $\theta \in \mathbb{N}$. Prove that the statistic

$$T = \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)}^{n} - (X_{(n)} - 1)^{n}}$$

is the efficient unbiased estimator of the parameter θ among all unbiased estimators.

6. (2 points) Let $X_1, \ldots, X_n$ be a sample from the uniform distribution on the interval $[\theta, 2\theta]$, $\theta > 0$. Prove that the two-dimensional statistic $(X_{(1)}, X_{(n)})$ is not complete.

7.(4 points)
I. (2 points) Let $X_1, \ldots, X_n$ be a sample from $Unif[0, \theta]$. Using the statistic $X_{(n)}$, construct an exact confidence interval of significance level α for the parameter θ.
II. (2 points) Using the statistic $X_{(1)}$, construct an exact confidence interval for θ from a sample of size n from:
a) a uniform distribution on $[\theta, \theta + 1]$;
b)*(1 point) a uniform distribution on $[\theta, 2\theta]$.

II.* (1 point) Out of 400 tested light bulbs, 40 were defective. Find a confidence interval of significance level 0.01 for the probability of defect.

8. (4 points, 2 points per subquestion) Construct confidence intervals for the variance of a normal distribution when the mean is known (assume it is zero) in two ways:
a) Using the statistic $\overline{X^2}$, we obtain

$$\left(\frac{\sum_{k=1}^n X_k^2}{\lambda_{(1+\gamma)/2}}, \frac{\sum_{k=1}^n X_k^2}{\lambda_{(1-\gamma)/2}}\right),$$

where λ_p is the p-quantile of the χ^2 distribution with n degrees of freedom.
b) Using the statistic $\overline{X}^2$, we obtain

$$\left(\frac{n\overline{X}^2}{z^2_{(3+\gamma)/4}}, \frac{n\overline{X}^2}{z^2_{(3-\gamma)/4}}\right),$$

where z_p is the p-quantile of the standard normal distribution.
Explain how each confidence interval is derived and show that it is correct. Describe how the length of each interval changes with increasing sample size.

9. I. (4 points) Let $X_1, \ldots, X_n$ be a sample from $Unif[0, \theta]$. Construct asymptotic confidence intervals for θ:
a) using $X_{(n)}$ (hint: use the statistic $\frac{n(\theta - X_{(n)})}{\theta}$);
b) using asymptotically normal estimators

$$\hat{\theta}_1 = 2\bar{X}, \qquad \hat{\theta}_2 = \sqrt{3\overline{X^2}},$$

construct asymptotic confidence intervals for θ at level α, and show that the second interval is asymptotically shorter than the first.

II.* (1 points) Let $X_1, \ldots, X_n$ be a sample from an exponential distribution with parameter β. Using asymptotically normal estimators

$$\hat{\beta}_1 = \frac{1}{\bar{X}}, \qquad \hat{\beta}_2 = \sqrt{\frac{2}{\overline{X^2}}},$$

construct asymptotic confidence intervals for β at level $1 - \alpha$ and show that the second interval is shorter than the first.

10. (4 points)
a) For a sample from a Poisson distribution with parameter λ, construct the most powerful test of asymptotic size ε to distinguish the hypothesis $\lambda = \lambda_1$ from the alternative $\lambda = \lambda_2$ if $\lambda_1 < \lambda_2$. Compute the limiting power of this test as $n \to \infty$.
b) A player observed a dice game and suspected that six appears in 18% of throws, five in 14%, and the other four faces equally likely (17% each). The player wants to test his hypothesis before participating in n consecutive throws. The only alternative considered is that the die is fair. For $n = 2$, find the most powerful test at level 0.0196.
c)* Denote by $\beta(\varepsilon)$ the power of the most powerful test among all randomized tests of level ε. Show that

$$\beta(\varepsilon) \geq \varepsilon.$$

11. I. (5 points) The table shows the results of a math exam. For each score range, the number of participants in that range is given as a percentage.

Year	0–20	21–40	41–60	61–80	81–100
2018	3.85	31.13	32.96	29.92	2.13
2017	7.92	36.16	25.09	27.49	3.34

The total number of participants each year was $3.9 \cdot 10^5$.
a) (1 point) Since the exam scale has many possible values, we consider the score distribution as continuous. Unfortunately, the data allow reconstructing the empirical distribution function only at a few points. Can the Kolmogorov–Smirnov test convince us to reject the hypothesis, or do the data not contradict it?
b) (2 points) Test the hypothesis of homogeneity at significance level $\alpha = 0.05$.

II. (2 points) Check part b using the χ^2 test.

III.* (1 point) Prove that under the condition $0 \le X_{(1)} \le X_{(n)} \le 1$, the following equality holds:

$$\int_0^1 (F_n(y) - y)^2 \, dy = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^{n} \left(X_{(k)} - \frac{2k-1}{2n} \right)^2 .$$

12. (4 points) Consider the following data of X_i values and their frequencies ν_i:

X_i	1	2	3	4	5	6	> 6
ν_i	4	6	5	4	3	2	1

Test the goodness-of-fit at significance level $\alpha = 0.2$ for the following distributions:
a) Poisson distribution;
b) geometric distribution of the first type.

13. (2 points) Out of 300 applicants, 97 had a grade of "5" at school, and 48 received "5" on the entrance exam in the same subject, with only 18 having "5" both at school and on the exam. Test the hypothesis of independence of receiving a "5" at school and on the exam.

3rd Homework.

李开言 2022111221

LI Kaiyan.

1. (4 points) Construct the Bayesian estimate of the parameter θ of a uniform distribution on the interval $[0;\theta]$.
Prove that the sample mean is not a minimax estimator:
a) if the parameter θ has the density $q(t) = \frac{1}{t^2}$ for $t > 1$;
b) if θ is uniformly distributed on [0,1].

Sol: $R(\bar{X},\theta) = \mathbb{E}_\theta[(\bar{X}-\theta)^2] = \mathbb{E}[\bar{X}^2] - 2\theta\,\mathbb{E}[\bar{X}] + \theta^2$

$\mathbb{E}[\bar{X}^2] = \mathbb{E}[(\frac{1}{n}\sum X_i)^2] = \mathbb{E}[\frac{1}{n^2}\sum X_i^2 + \frac{2}{n^2}\sum_{i\neq j} X_iX_j] = \frac{1}{n^2}\cdot n\cdot\frac{\theta^2}{3} + \frac{2}{n^2}\cdot\frac{n(n-1)}{2}\cdot\frac{\theta^2}{4} = \frac{\theta^2}{n^2}(\frac{n-1}{4}+\frac{1}{3})$

thus $R(\bar{X},\theta) = \frac{\theta^2}{n}(\frac{n-1}{4}+\frac{1}{3}) - 2\theta\cdot\frac{\theta}{2} + \theta^2 = \frac{\theta^2}{n}(\frac{n-1}{4}+\frac{1}{3})$

$\sup_\theta R(\bar{X},\theta) = +\infty$. thus it's not minimax.

a). Likelihood: $L(X,\theta) = \frac{1}{\theta^n}\cdot \mathbb{1}(X_{(n)} \le \theta)$

posterior: $q(\theta|X) \propto L(X,\theta)\,q(\theta) = \frac{1}{\theta^n}\cdot\mathbb{1}(X_{(n)}\le\theta)\cdot\frac{1}{\theta^2} = \frac{1}{\theta^{n+2}}\mathbb{1}\{\theta \ge \max\{X_{(n)},1\}\}$

$$T(X) = \frac{\int \theta\cdot q(\theta|X)\,d\theta}{\int L(X,\theta)\cdot q(\theta)\,d\theta} = \frac{\int_{\max\{1,X_{(n)}\}}^{\infty}\frac{1}{\theta^{n+1}}\,d\theta}{\int_{\max\{1,X_{(n)}\}}\frac{1}{\theta^{n+2}}\,d\theta} = \frac{\frac{(\max\{1,X_{(n)}\})^{-n}}{n}}{\frac{(\max\{1,X_{(n)}\})^{-(n+1)}}{n+1}} = \frac{n+1}{n}\cdot\max\{X_{(n)},1\}$$

b) posterior $q(\theta|X) \propto L(X,\theta)\,q(\theta) = \frac{1}{\theta^n}\cdot\mathbb{1}\{\max\{0,X_{(n)}\}\le\theta\le 1\} = \frac{1}{\theta^n}\mathbb{1}\{X_{(n)}\le\theta\le 1\}$

$$T(X) = \frac{\int_{X_{(n)}}^{1}\frac{1}{\theta^{n-1}}\,d\theta}{\int_{X_{(n)}}^{1}\frac{1}{\theta^n}\,d\theta} = \begin{cases} n>2 & \frac{\frac{\theta^{2-n}}{2-n}\Big|_{X_{(n)}}^{1}}{\frac{\theta^{1-n}}{1-n}\Big|_{X_{(n)}}^{1}} = \frac{\frac{X_{(n)}^{2-n}-1}{n-2}}{\frac{X_{(n)}^{1-n}-1}{n-1}} = \frac{n-1}{n-2}\cdot\frac{X_{(n)}^{2-n}-1}{X_{(n)}^{1-n}-1} \\ n=2 & \frac{-\ln\theta\Big|_{X_{(2)}}^{1}}{-\theta^{-1}\Big|_{X_{(2)}}^{1}} = \frac{-\ln X_{(2)}}{-1+\frac{1}{X_{(2)}}} = X_{(2)}\cdot\frac{\ln(X_{(2)})}{X_{(2)}-1} \\ n=1 & \frac{1-X_{(1)}}{\ln\theta\Big|_{X_{(1)}}^{1}} = \frac{1-X_{(1)}}{-\ln X_{(1)}} \end{cases}$$

2. (6 points) Find the Bayesian estimate for the parameter θ and check whether the estimator $\overline{X}$ or $(\overline{X})^{-1}$ is minimax, if:
a) $X_1, \ldots, X_n$ is a sample from a Bernoulli distribution with parameter θ, and the prior is a Beta distribution $B(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$;
b) a sample from $Exp(\theta)$, and θ has an exponential distribution with parameter β;
c) a sample from a Poisson distribution with parameter θ, where θ has an exponential distribution with parameter a;
d)*(1 point) a sample from a Poisson distribution with parameter θ, where θ takes values 1 and 2 with probabilities 1/3 and 2/3, respectively.

a). likelihood $L(X,\theta) = \prod \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$ $\quad q(\theta) = \frac{\Gamma(\sqrt{n})}{\Gamma(\frac{\sqrt{n}}{2})\cdot\Gamma(\frac{\sqrt{n}}{2})}\theta^{\frac{\sqrt{n}}{2}-1}(1-\theta)^{\frac{\sqrt{n}}{2}-1}$

posterior $q(\theta|X) \propto L(X,\theta)\cdot q(\theta)$

$$T_B(X) = \frac{\int \theta\cdot q(\theta|X)\,d\theta}{\int L(X,\theta)\,q(\theta)\,d\theta} = \frac{\int_0^1 \theta^{\sum X_i+\frac{\sqrt{n}}{2}}(1-\theta)^{n-\sum X_i+\frac{\sqrt{n}}{2}-1}\,d\theta}{\int_0^1 \theta^{\sum X_i+\frac{\sqrt{n}}{2}-1}(1-\theta)^{n-\sum X_i+\frac{\sqrt{n}}{2}-1}\,d\theta} = \frac{B(\sum X_i+\frac{\sqrt{n}}{2}+1,\ n-\sum X_i+\frac{\sqrt{n}}{2})}{B(\sum X_i+\frac{\sqrt{n}}{2},\ n-\sum X_i+\frac{\sqrt{n}}{2})}$$

by the property of Γ, β func. $\frac{B(u+1,v)}{B(u,v)} = \frac{\frac{\Gamma(u+1)\Gamma(v)}{\Gamma(u+v+1)}}{\frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}} = \frac{\Gamma(u+1)}{\Gamma(u)}\cdot\frac{\Gamma(u+v)}{\Gamma(u+v+1)} = u\cdot\frac{1}{u+v}$

that is $T_B(X) = \frac{\sum X_i+\frac{\sqrt{n}}{2}}{\sum X_i+\frac{\sqrt{n}}{2}+n-\sum X_i+\frac{\sqrt{n}}{2}} = \frac{\sum X_i+\frac{\sqrt{n}}{2}}{n+\sqrt{n}}.$

1) consider $\overline{X}$

$R(\overline{X},\theta) = \mathbb{E}_\theta[(\overline{X}-\theta)^2] = Var_\theta(\overline{X}) = \frac{\theta(1-\theta)}{n}$

$\sup_\theta R(\overline{X},\theta) = R(\overline{X},\frac{1}{2}) = \frac{1}{4n}$

2) consider $\overline{X}^{-1}$

we have $P_\theta(\sum X_i = 0) = (1-\theta)^n > 0.$ but $(\overline{X})^{-1}$ undefined

thus $\overline{X}^{-1}$ is not a estimator, let along minimax.

3) check minimax:

denote that $\sum X_i = S$ $\quad \frac{\sqrt{n}}{2} = \alpha.$ $\quad \frac{1}{\sqrt{n}+n} = \beta.$

$R(T_B,\theta) = \mathbb{E}_\theta[(T_B-\theta)^2] = Var_\theta T_B + \mathbb{E}_\theta(T_B-\theta)^2.$

$\mathbb{E}[T_B] = \beta(\alpha+\mathbb{E}_\theta[S]) = \beta(\alpha+n\theta)$ $\quad Var[T_B] = \beta^2 Var_\theta(\alpha+S) = \beta^2 Var_\theta(S) = \beta^2 n\theta(1-\theta)$

thus $R(T_B,\theta) = \beta^2 n\theta(1-\theta) + (\beta(\alpha+n\theta)-\theta)^2 = [(\beta n-1)^2-\beta^2 n]\theta^2 + [\beta^2 n + 2\beta\alpha(\beta n-1)]\theta + \beta^2\alpha^2$

use $\begin{cases}\alpha = \frac{\sqrt{n}}{2}\\ \beta = \frac{1}{\sqrt{n}+n}\end{cases}$ $= [(\frac{-\sqrt{n}}{n+\sqrt{n}})^2 - \frac{n}{(n+\sqrt{n})^2}]\theta^2 + [\frac{n}{(n+\sqrt{n})^2} - \frac{n}{(n+\sqrt{n})^2}]\theta + \frac{n}{4(n+\sqrt{n})^2} = \frac{n}{4(n+\sqrt{n})^2}$

thus Bayes estimator has const. rask. T_B is minimax estimator.

compare: $\frac{1}{4n}$ and $\frac{n}{4(n+\sqrt{n})^2}$:

$\frac{n}{4(n+\sqrt{n})^2} = \frac{n}{4n(\sqrt{n}+1)^2} = \frac{1}{4(\sqrt{n}+1)^2} < \frac{1}{4(\sqrt{n})^2} = \frac{1}{4n}$ thus $\overline{X}$ is not minimax estimator

(b) $f(x_i|\theta) = \theta e^{-\theta x_i}$ $x_i > 0$. $\theta > 0$. $q(\theta) = \beta e^{-\beta\theta}$ $\theta > 0$. $\beta > 0$.

likelihood $f(X|\theta) = \theta^n e^{-\theta \sum x_i}$

$$T_B(x) = \frac{\int_0^\infty \theta^{n+1} e^{-(\beta+\sum x_i)\theta} d\theta}{\int_0^\infty \theta^n e^{-(\beta+\sum x_i)\theta} d\theta} = \frac{\frac{\Gamma(n+2)}{(\beta+\sum x_i)^{n+2}}}{\frac{\Gamma(n+1)}{(\beta+\sum x_i)^{n+1}}} = \frac{n+1}{\beta+\sum x_i}$$

1) consider $T = \bar{X}^{-1}$.

denote $S = \sum X_i$.

$$f(s|\theta) = \frac{\theta^n}{(n-1)!} s^{n-1} e^{-\theta s}$$

$$\mathbb{E}_\theta[\tfrac{1}{S}] = \int_0^\infty \frac{1}{s} f(s|\theta)\, d\theta = \frac{\theta^n}{(n-1)!}\int_0^\infty s^{n-2} e^{-\theta s} ds \overset{t=\theta s}{=\!=} \frac{\theta^n}{(n-1)!}\cdot\frac{1}{\theta^{n-1}}\int_0^\infty t^{n-2} e^{-t} dt = \frac{\theta}{(n-1)!}\cdot(n-2)! = \frac{\theta}{n-1}$$

$(n>1)$

$$\mathbb{E}_\theta[\tfrac{1}{S^2}] = \int_0^\infty \frac{1}{s^2} f(s|\theta)\, d\theta = \frac{\theta^n}{(n-1)!}\cdot\frac{(n-3)!}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)} \quad (n>2)$$

$$R(\bar{X}^{-1},\theta) = n^2\mathbb{E}_\theta[\tfrac{1}{S^2}] - 2n\theta\,\mathbb{E}_\theta[\tfrac{1}{S}] + \theta^2 = \theta^2\frac{n+2}{(n-1)(n-2)} \quad n>2.$$

$\sup_\theta R(\bar{X}^{-1},\theta) = +\infty$ $(\theta \to +\infty)$.

2) consider $T = \bar{X}$

$$R(\bar{X},\theta) = Var_\theta(\bar{X}) + (\mathbb{E}_\theta[\bar{X}] - \theta)^2 = \frac{1}{n\theta^2} + (\tfrac{1}{\theta}-\theta)^2 = \theta^2 - 2 + \frac{n+1}{n\theta^2}$$

$\sup_\theta R(\bar{X},\theta) = +\infty$ $(\theta \to +\infty)$.

thus. neither $\bar{x}$ nor $\bar{x}^{-1}$ is minimax estimator.

(c) $f(x_i|\theta) = \frac{\theta^{x_i}}{(x_i)!} e^{-\theta}$ $x_i = 0, 1, \dots$ $q(\theta) = a\cdot e^{-a\theta}$ $a>0$, $\theta>0$.

$\mathbb{E}_\theta(\bar{X}) = \theta$. $Var_\theta(\bar{X}) = \frac{\theta}{n}$

likelihood: $f(X|\theta) = \frac{\theta^{\sum x_i}}{\prod (x_i)!}\cdot e^{-n\theta}$.

$$T_B(x) = \frac{\frac{a}{\prod(x_i!)}\int_0^\infty \theta^{1+\sum x_i} e^{-\theta(n+a)} d\theta}{\frac{a}{\prod(x_i)!}\int_0^\infty \theta^{\sum x_i} e^{-\theta(n+a)} d\theta} = \frac{\frac{\Gamma(2+\sum x_i)}{(n+a)^{2+\sum x_i}}}{\frac{\Gamma(1+\sum x_i)}{(n+a)^{1+\sum x_i}}} = \frac{1+\sum x_i}{n+a}.$$

$$R(\bar{X},\theta) = \mathbb{E}_\theta(\bar{X}-\theta)^2 = Var(\bar{X}) + (\mathbb{E}_\theta[\bar{X}]-\theta)^2 = \frac{\theta}{n}$$

$\sup_\theta R(\bar{X},\theta) = \infty$. $(\theta \to \infty)$.

for $\bar{x}^{-1}$, $P_\theta(\sum X_i = 0) = e^{-n\theta} > 0$. when $\sum X_i = 0$ $\bar{X}^{-1} = \infty$. thus the risk must be $+\infty$

thus. neither $\bar{x}$ nor $\bar{x}^{-1}$ is minimax estimator.

(d). $f(X_i|\theta) = \frac{\theta^{X_i}}{(X_i)!} e^{-\theta} \quad X_i = 0, 1, \dots \qquad q(\theta) = \begin{cases} 1 & p = \frac{1}{3} \\ 2 & p = \frac{2}{3} \end{cases}$

$\mathbb{E}_\theta(\bar{X}) = \theta$. $Var_\theta(\bar{X}) = \frac{\theta}{n}$

$f(\theta = 1 | X) = e^{-n} \cdot \prod \frac{1}{(X_i)!}$

$f(\theta = 2 | X) = e^{-2n} \cdot 2^{\sum X_i} \cdot \prod \frac{1}{(X_i)!}$

thus. $P(\theta = 1 | X) = \frac{\frac{1}{3} f(\theta=1|X)}{\frac{1}{3} f(\theta=1|X) + \frac{2}{3} f(\theta=2|X)} = \frac{1}{1 + 2^{1+\sum X_i} \cdot e^{-n}}$

$P(\theta = 2 | X) = \frac{\frac{2}{3} f(\theta=2|X)}{\frac{1}{3} f(\theta=1|X) + \frac{2}{3} f(\theta=2|X)} = \frac{2^{1+\sum X_i} \cdot e^{-n}}{1 + 2^{1+\sum X_i} \cdot e^{-n}}$

$T_B(X) = 1 \cdot P(\theta=1|X) + 2 \cdot P(\theta=2|X) = 2 - \frac{1}{1 + 2^{1+\sum X_i} \cdot e^{-n}}$

by c). $R(\bar{X}, \theta) = \frac{\theta}{n} \qquad R(\bar{X}^{-1}, \theta) = +\infty$.

$\sup_\theta R(\bar{X}, \theta) = \frac{2}{n}$.

$\bar{X}^{-1}$ not minimax. $\bar{X}$ might be minimax.

3. (1 points) Find the efficient estimator of the parameter θ among all unbiased estimators, if the sample comes from $Unif[0, \theta]$, using the known information about $X_{(n)}$.

Sol: $f_{X_{(n)}}(x, \theta) = \frac{n x^{n-1}}{\theta^n} \qquad x \in [0, \theta]$

$\mathbb{E}_\theta[X_{(n)}] = \int_0^\theta x \cdot \frac{n x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta$.

if we need $\mathbb{E}_\theta[T(X_{(n)})] = \theta \Rightarrow T_U(X) = \frac{n+1}{n} X_{(n)}$

since $f(x; \theta) = \theta^{-n} \mathbb{1}\{0 \le X_{(n)} \le \theta\} \cdot n! \cdot \mathbb{1}\{0 < X_{(1)} < \dots < X_{(n)}\}$.

by Fisher-Neyman thm. $T(X) = X_{(n)}$ $g(T, \theta) = \theta^{-n} \cdot \mathbb{1}\{0 < t < \theta\}$. $h(x) = n! \cdot \mathbb{1}\{0 < X_{(1)} < \dots < X_{(n)}\}$.

$\Rightarrow X_{(n)}$ is sufficient.

$\mathbb{E}_\theta \varphi(T) = \int_0^\theta \varphi(x) \cdot \frac{n x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \int_0^\theta h(x) \cdot x^{n-1} dx$

if $\mathbb{E}_\theta \varphi(T) = 0 \quad \forall \theta \Rightarrow \int_0^\theta h(x) x^{n-1} dx = 0, \forall \theta$.

$\Rightarrow F(\theta) = \int_0^\theta h(x) x^{n-1} dx \equiv 0 \Rightarrow F'(\theta) = h(\theta) \cdot \theta^{n-1} = 0$ almost everywhere $\forall \theta$.

$\Rightarrow h(\theta) = 0$ almost everywhere $\Rightarrow X_{(n)}$ is complete.

by Lehmann Scheffé. thm. $T_U = \frac{n+1}{n} X_{(n)}$ is efficient.

4. (4 points) Let $X_1, \ldots, X_n$ be a sample from a two-parameter shifted exponential distribution with density

$$f_{\beta,\tau}(y) = \begin{cases} \beta^{-1}\exp\left(-\frac{y-\tau}{\beta}\right), & y > \tau, \\ 0, & y \le \tau, \end{cases} \qquad \beta > 0,\ \tau \in \mathbb{R}.$$

Find a sufficient statistic , check its completeness and find the efficient estimator among all unbiased estimators :

a) for the parameter τ if β is known;
b) for the parameter β if τ is known;
(*) (2 points) The same questions for the two-dimensional parameter $\boldsymbol{\theta} = (\beta, \tau)$.

$$f(x_1, \ldots, x_n; \beta, \tau) = \beta^{-n} \exp\left(-\frac{\sum X_i - n\tau}{\beta}\right) \cdot \mathbb{1}\{\tau < x_{(1)}\}.$$

a). $X_{(1)}$ is sufficient.

$T = X_{(1)}$ $\quad g(t,\tau) = \exp\left(\frac{n\tau}{\beta}\right) \mathbb{1}\{\tau < X_{(1)}\}.$ $\quad h(x) = \beta^{-n}\exp\left(-\frac{\sum X_i}{\beta}\right).$

$$f_{X_{(1)}}(x;\tau) = \frac{n}{\beta} \exp\left(-\frac{n(x-\tau)}{\beta}\right) \mathbb{1}\{x > \tau\}.$$

denote that $X_{(1)} = \tau + Z.$ $\quad Z \sim Exp\left(\frac{n}{\beta}\right).$

$$\mathbb{E}_\tau[X_{(1)}] = \tau + \mathbb{E}[Z] = \tau + \frac{\beta}{n}$$

$$\Rightarrow \hat{\tau} = X_{(1)} - \frac{\beta}{n}$$

similarly. $\mathbb{E}_\tau[h(X_{(1)})] = \int_\tau^\infty h(x) \frac{n}{\beta} \exp\left(-\frac{n(x-\tau)}{\beta}\right) dx = 0 \quad \forall \tau \Rightarrow h(x) = 0$ almost everywhere.

$X_{(1)}$ complete and sufficient by Lehmann Scheffé. thm.

$\Rightarrow \hat{\tau} = X_{(1)} - \frac{\beta}{n}$ is existing unbiased estimator, then it's efficient.

b) denote $Y_i = X_i - \tau.$ (τ is known).

$$\Rightarrow f(Y, \beta) = \beta^{-n} \exp\left(-\frac{\sum Y_i}{\beta}\right) \prod \mathbb{1}\{y_i > 0\}$$

$S = \sum Y_i$ (i.e. $\sum(X_i - \tau)$) is the sufficient statistics.

$$f_S(s;\beta) = \frac{1}{\beta^n} \frac{s^{n-1}}{(n-1)!} e^{-\frac{s}{\beta}} \cdot \mathbb{1}\{s > 0\}$$

$$\mathbb{E}_\beta[S] = \frac{1}{\beta^n} \cdot \frac{1}{(n-1)!} \int_0^{+\infty} s^n \cdot e^{-\frac{s}{\beta}} ds \underset{ds = \beta d\lambda}{\overset{\frac{s}{\beta} = \lambda}{=}} \frac{1}{\beta^n} \cdot \frac{1}{(n-1)!} \int_0^{+\infty} \beta^n \cdot \lambda^n \cdot \beta \cdot e^{-\lambda} d\lambda =$$

$$= \beta \cdot \frac{\Gamma(n+1)}{(n-1)!} = \beta n$$

$\hat{\beta} = \frac{1}{n} S = \frac{1}{n} \sum (X_i - \tau).$ is unbiased estimator.

similarly. $\mathbb{E}_\beta[h(S)] = \int_0^\infty h(s) \cdot \frac{1}{\beta^n} \cdot \frac{s^{n-1}}{(n-1)!} e^{-\frac{s}{\beta}} ds = 0 \quad \forall \beta > 0. \Rightarrow h(s) = 0$

S complete and sufficient by Lehmann Scheffé. thm.

$\hat{\beta} = \frac{1}{n} S = \frac{1}{n} \sum (X_i - \tau).$ is existing unbiased estimator, then it's efficient.

(*) $f(X,\beta,T) = \beta^{-n}\exp\left(-\frac{\sum X_i}{\beta}\right)\exp\left(\frac{nT}{\beta}\right)\mathbb{1}\{T < X_{(1)}\}$

denote $T_1 = X_{(1)}$. $T_2 = \sum X_i$

$g(t_1, t_2, \beta, T) = \beta^{-n}\exp\left(-\frac{t_2}{\beta}\right)\exp\left(\frac{nT}{\beta}\right)\mathbb{1}\{T < t_1\}$ $\qquad h(x) = 1.$

$(X_{(1)}, \sum X_i)$ is sufficient.

similar as (a). $\mathbb{E}[X_{(1)}] = T + \frac{\beta}{n}$. $\qquad \mathbb{E}[\bar{X}] = T + \beta$.

$\mathbb{E}[\bar{X} - X_{(1)}] = \beta(1 - \frac{1}{n})$. $\Rightarrow$ $\hat{\beta} = \frac{n}{n-1}(\bar{X} - X_{(1)})$ unbiased. estimator

set a, b. s.t. $\hat{T} = aX_{(1)} + b\bar{X}$

$\mathbb{E}[\hat{T}] = T \Rightarrow a(T + \frac{\beta}{n}) + b(T+\beta) = T. \Rightarrow \begin{cases} a+b=1 \\ \frac{a}{n} + b = 0 \end{cases} \Rightarrow \begin{cases} a = \frac{n}{n-1} \\ b = -\frac{1}{n-1} \end{cases}$

$\hat{T} = \frac{n}{n-1}X_{(1)} - \frac{1}{n-1}\bar{X}$ unbiased estimator.

check completeness (?).

5. (2 points) Let $X_1, \ldots, X_n$ be a sample from a uniform distribution on the finite set $\{1, \ldots, \theta\}$, $\theta \in \mathbb{N}$. Prove that the statistic

$$T = \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)}^n - (X_{(n)} - 1)^n}$$

is the efficient unbiased estimator of the parameter θ among all unbiased estimators.

Pf: $P(X_{(n)} \le k) = (\frac{k}{\theta})^n$.

$P(X_{(n)} = k) = P(X_{(n)} \le k) - P(X_{(n)} \le k-1) = \frac{k^n}{\theta^n} - \frac{(k-1)^n}{\theta^n} = \frac{k^n - (k-1)^n}{\theta^n}$

$\mathbb{E}_\theta(T) = \sum \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n} \cdot \frac{k^n - (k-1)^n}{\theta^n} = \frac{1}{\theta^n} \sum (k^{n+1} - (k-1)^{k+1}). = \frac{1}{\theta^n}(\theta^{n+1} - 0) = \theta.$

thus T is unbiased

$f(x_1, \ldots, x_n, \theta) = \theta^{-n}\prod_{i=1}^n \mathbb{1}\{1 \le x_i \le \theta\} = \theta^{-n} \cdot \mathbb{1}\{x_{(n)} \le \theta\}.$

thus $X_{(n)}$ is sufficient.

Assume that $\mathbb{E}_\theta(\varphi(X_{(n)})) = \sum \varphi(k) \cdot \frac{k^n - k^{n-1}}{\theta^n} = 0 \Rightarrow \sum_{k=1}^{\theta} \varphi(k)(k^n - k^{n-1}) = 0$

when $\theta = 1 \Rightarrow \varphi(1) = 0$

assume for $\theta = m-1$. $h(1) = h(2) = \cdots h(m-1) = 0$.

when $\theta = m$ $\sum_{k=1}^{m} \varphi(k)(k^n - k^{n-1}) = \varphi(m)(m^n - m^{n-1}) = 0 \Rightarrow \varphi(m) = 0$

thus $\varphi(k) = 0$ $\forall k \in \mathbb{N}$. thus $X_{(n)}$ is complete

by Lehmann - Scheffé thm. T is func. of $X_{(n)}$. and unbiased, then it's efficient.

6. (2 points) Let $X_1, \ldots, X_n$ be a sample from the uniform distribution on the interval $[\theta, 2\theta]$, $\theta > 0$. Prove that the two-dimensional statistic $(X_{(1)}, X_{(n)})$ is not complete.

Pf: denote $Y_i = \frac{X_i}{\theta} - 1$ $\quad Y_i \sim Unif[0,1]$.

$X_{(1)} = \theta(Y_{(1)}+1) \qquad X_{(n)} = \theta(Y_{(n)}+1)$

$f_{Y_{(1)}}(x) = n(1-x)^{n-1}$

$\mathbb{E}[Y_{(1)}] = \int_0^1 xn(1-x)^{n-1}dx \overset{1-x=u}{=\!=\!=} n\left[\int_0^1 u^{n-1}du - \int_0^1 u^n du\right] = n\left[\frac{1}{n} - \frac{1}{n+1}\right] = \frac{1}{n+1}$

$f_{Y_{(n)}}(x) = nx^{n-1}$.

$\mathbb{E}[Y_{(n)}] = n\int_0^1 x^n dx = \frac{n}{n+1}$

set a.b. s.t. $h(X_{(1)}, X_{(n)}) = aX_{(1)} + bX_{(n)} = \theta(aY_{(1)} + bY_{(n)} + a + b)$

let $\mathbb{E}[h(X_{(1)}, X_{(n)})] = 0 \quad \Rightarrow \quad a\left(\frac{n+2}{n+1}\right) + b\left(\frac{2n+1}{n+1}\right) = 0$

let $a = \frac{2n+1}{n+1}$ $\quad b = -\frac{n+2}{n+1}$ $\quad$ we have $\mathbb{E}_\theta(h(X_{(1)}, X_{(n)})) = 0.$ $\forall\theta$.

but. $h(X_{(1)}, X_{(n)}) = \frac{2n+1}{n+1}X_{(1)} - \frac{n+2}{n+1}X_{(n)}$ not always $= 0$.

thus. $(X_{(1)}, X_{(n)})$ not complete.

7.(4 points)
I. (2 points) Let $X_1, \ldots, X_n$ be a sample from $Unif[0, \theta]$. Using the statistic $X_{(n)}$, construct an exact confidence interval of significance level α for the parameter θ.
II. (2 points) Using the statistic $X_{(1)}$, construct an exact confidence interval for θ from a sample of size n from:
a) a uniform distribution on $[\theta, \theta + 1]$;
b)*(1 point) a uniform distribution on $[\theta, 2\theta]$.

II.* (1 point) Out of 400 tested light bulbs, 40 were defective. Find a confidence interval of significance level 0.01 for the probability of defect.

I. $\mathbb{P}(X_{(n)} \le x) = \left(\frac{x}{\theta}\right)^n \quad x \in [0, \theta]$

let $Y = \frac{X_{(n)}}{\theta}$. $\quad \mathbb{P}(Y \le y) = \mathbb{P}(X_{(n)} \le y\theta) = \left(\frac{y\theta}{\theta}\right)^n = y^n \quad y \in [0,1]$ the distribution of Y independent with θ.

note that $\mathbb{P}(\alpha^{1/n} < Y < 1) = \mathbb{P}(Y > \alpha^{1/n}) = 1 - \mathbb{P}(Y \le \alpha^{1/n}) = 1 - (\alpha^{1/n})^n = 1 - \alpha$.

then $\alpha^{1/n} < Y < 1 \Rightarrow \alpha^{1/n} < \frac{X_{(n)}}{\theta} < 1 \Rightarrow \theta \in \left(X_{(n)}, \frac{X_{(n)}}{\alpha^{1/n}}\right)$ which is the exact CI.

II. a) let $Y_i = X_i - \theta$. $\quad Y_i \sim Unif[0,1]$

$\mathbb{P}(Y_{(1)} \ge y) = \mathbb{P}(\forall Y_i \ge y) = (1-y)^n \quad \mathbb{P}(Y_{(1)} < y) = 1 - (1-y)^n$ independent with θ. pivot.

note that $\mathbb{P}(0 < Y_{(1)} < 1 - \alpha^{1/n}) = \mathbb{P}(Y_{(1)} < 1 - \alpha^{1/n}) = 1 - (1 - (1 - \alpha^{1/n}))^n = 1 - \alpha$.

$0 < X_{(1)} - \theta < 1 - \alpha^{1/n} \Rightarrow \theta \in (X_{(1)} - (1 - \alpha^{1/n}), X_{(1)})$ which is the exact CI

b). let $Y_i = \frac{X_i}{\theta} - 1$ $\quad Y_i \sim Unif[0,1]$

similarly as a). $\quad 0 < \frac{X_{(1)}}{\theta} - 1 < 1 - \alpha^{1/n} \Rightarrow \theta < X_{(1)}$ and $\theta > \frac{X_{(1)}}{2 - \alpha^{1/n}} \Rightarrow \theta \in \left(\frac{X_{(1)}}{2 - \alpha^{1/n}}, X_{(1)}\right)$

I^*. define $X_i = \mathbb{1}\{\text{the ith blub is defective}\}$. $X_i \sim Ber(p)$.

we have the sample mean: $\hat{p} = \bar{X} = \frac{40}{400} = 0.1$. $\mathbb{E}X_i = p$. $Var\, X_i = p(1-p)$

by the CLT. $\sqrt{n}\dfrac{\bar{X}-p}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$.

choose z_α. s.t. $\mathbb{P}\left(\left|\dfrac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}}\right| > z_\alpha\right) = \alpha$. $z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2})$.

approximately, $\mathbb{P}\left(\left|\dfrac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}}\right| < z_\alpha\right) = 1-\alpha$. $\Rightarrow p \in \left(\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\alpha = 0.01$ by table $z_\alpha = 2.575$.

thus $p \in \left(0.1 - 2.575\sqrt{\frac{0.09}{400}}, 0.1 + 2.575\sqrt{\frac{0.09}{400}}\right) = (0.061375, 0.138625)$

8. (4 points, 2 points per subquestion) Construct confidence intervals for the variance of a normal distribution when the mean is known (assume it is zero) in two ways:
a) Using the statistic $\overline{X^2}$, we obtain

$$\left(\frac{\sum_{k=1}^n X_k^2}{\lambda_{(1+\gamma)/2}}, \frac{\sum_{k=1}^n X_k^2}{\lambda_{(1-\gamma)/2}}\right),$$

where λ_p is the p-quantile of the χ^2 distribution with n degrees of freedom.
b) Using the statistic $\overline{X}^2$, we obtain

$$\left(\frac{n\overline{X}^2}{z^2_{(3+\gamma)/4}}, \frac{n\overline{X}^2}{z^2_{(3-\gamma)/4}}\right),$$

where z_p is the p-quantile of the standard normal distribution.
Explain how each confidence interval is derived and show that it is correct. Describe how the length of each interval changes with increasing sample size.

Sol: $X_k \sim N(0,\sigma^2)$. σ unknown.

a). $\overline{X^2} = \frac{1}{n}\sum X_k^2$ denote $S = \sum X_k^2$

since $\frac{X_k}{\sigma} \sim N(0,1)$ by the def. of χ^2, $\frac{S}{\sigma^2} = \sum\left(\frac{X_k}{\sigma}\right)^2 \sim \chi^2_n$

induce the notion λ_p, $\mathbb{P}\left(\lambda_{(1-\gamma)/2} \le \frac{S}{\sigma^2} \le \lambda_{(1+\gamma)/2}\right) = \gamma$

$\lambda_{(1-\gamma)/2} \le \frac{S}{\sigma^2} \le \lambda_{(1+\gamma)/2} \Rightarrow \dfrac{S}{\lambda_{(1-\gamma)/2}} \le \sigma^2 \le \dfrac{S}{\lambda_{(1+\gamma)/2}} \Rightarrow \dfrac{\sum X_k^2}{\lambda_{(1-\gamma)/2}} \le \sigma^2 \le \dfrac{\sum X_k^2}{\lambda_{(1+\gamma)/2}}$

CI length $O(\frac{1}{\sqrt{n}})$. When $n\to\infty$. length $\to 0$.

b) $\overline{X}^2 = \frac{1}{n^2}(\sum X_i)^2$.

denote $T = \dfrac{\sqrt{n}\bar{X}}{\sigma}$ since $\bar{X} \sim N(0, \frac{\sigma^2}{n})$. $T \sim N(0,1)$.

$\mathbb{P}\left(z^2_{(3-\gamma)/4} \le T^2 \le z^2_{(3+\gamma)/4}\right) = \mathbb{P}\left(z_{(3-\gamma)/4} \le |T| \le z_{(3+\gamma)/4}\right)$

$= \mathbb{P}\left(-z_{(3+\gamma)/4} \le T \le -z_{(3-\gamma)/4}\right) + \mathbb{P}\left(z_{(3-\gamma)/4} \le T \le z_{(3+\gamma)/4}\right)$

$= [1-\Phi(z_{(3-\gamma)/4}) - (1-\Phi(z_{(3+\gamma)/4}))] + [\Phi(z_{(3+\gamma)/4}) - \Phi(z_{(3-\gamma)/4})]$

$= 2[\Phi(z_{(3+\gamma)/4}) - \Phi(z_{(3-\gamma)/4})] = 2\left[\frac{3+\gamma}{4} - \frac{3-\gamma}{4}\right] = \gamma.$

thus. $\mathbb{P}\left(z^2_{(3-\gamma)/4} \le \frac{n\bar{X}^2}{\sigma^2} \le z^2_{(3+\gamma)/4} \right) = \gamma$

$$\Rightarrow \frac{n\bar{X}^2}{z^2_{(3+\gamma)/4}} \le \sigma^2 \le \frac{n\bar{X}^2}{z^2_{(3-\gamma)/4}}$$

$n\bar{X}^2 \to \sigma^2$ the CI length will tend to const. $\left| \frac{\sigma^2}{z^2_{(3-\gamma)/4}} - \frac{\sigma^2}{z^2_{(3+\gamma)/4}} \right|$.

9. I. (4 points) Let $X_1, \ldots, X_n$ be a sample from $Unif[0, \theta]$. Construct asymptotic confidence intervals for θ:
a) using $X_{(n)}$ (hint: use the statistic $\frac{n(\theta - X_{(n)})}{\theta}$);
b) using asymptotically normal estimators

$$\hat{\theta}_1 = 2\bar{X}, \qquad \hat{\theta}_2 = \sqrt{3\overline{X^2}},$$

construct asymptotic confidence intervals for θ at level α, and show that the second interval is asymptotically shorter than the first.

Sol: $X_i \sim Unif[0, \theta]$

I. a). $\mathbb{P}(X_{(n)} \le x) = \left(\frac{x}{\theta}\right)^n$

consider $Y_n = \frac{n(\theta - X_{(n)})}{\theta}$

$$P(Y_n \le y) = \mathbb{P}\left(\theta - X_{(n)} \le \frac{y\theta}{n}\right) = 1 - \mathbb{P}\left(X_{(n)} \le \theta - \frac{y\theta}{n}\right) = 1 - \left(1 - \frac{y}{n}\right)^n$$

$\left(1 - \frac{y}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-y}$. thus. $Y_n \xrightarrow{d} Exp(1)$. (independent with θ).

we need to find q_1, q_2 s.t. $\mathbb{P}(q_1 \le Y \le q_2) = 1-\alpha$. for Y_n, $q_\alpha = -\ln(1-\alpha)$.

thus we have $\mathbb{P}\left(-\ln(1-\frac{\alpha}{2}) \le Y \le -\ln(\frac{\alpha}{2})\right) = 1-\alpha$.

$$\Rightarrow -\ln(1-\tfrac{\alpha}{2}) \le \frac{n(\theta - X_{(n)})}{\theta} \le -\ln(\tfrac{\alpha}{2}) \Rightarrow \begin{cases} (n + \ln(1-\frac{\alpha}{2}))\theta \ge nX_{(n)} \\ (n + \ln(\frac{\alpha}{2}))\theta \le nX_{(n)} \end{cases} \Rightarrow \frac{nX_{(n)}}{n + \ln\frac{\alpha}{2}} \le \theta \le \frac{nX_{(n)}}{n + \ln(1-\frac{\alpha}{2})}$$

(level $1-\alpha$).

b). $\mathbb{E}X_i = \frac{\theta}{2}$ $\quad \mathbb{E}X_i^2 = \frac{\theta^2}{3}$ $\quad Var\, X_i = \frac{\theta^2}{12}$

1). $\hat{\theta}_1 = 2\bar{X}$.

by CLT. $\sqrt{n}\frac{(\bar{X} - \frac{\theta}{2})}{\sqrt{\frac{\theta^2}{12}}} \xrightarrow{d} N(0,1)$.

the func. $f(x) = 2x$. cont. and $f' \equiv 2 \ne 0$. by Prop. 4.2.

thus. $\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{d} N(0, \frac{\theta^2}{12}\cdot 4) = N(0, \frac{\theta^2}{3})$

asymptotically. $\hat{\theta}_1 \xrightarrow{d} N(\theta, \frac{\theta^2}{3n})$.

construct CI. similar as Problem 8. $\mathbb{P}\left(\left|\frac{\hat{\theta}_1 - \theta}{\theta/\sqrt{3n}}\right| \le z_{1-\alpha/2}\right) = 1-\alpha$.

$$\Rightarrow \theta \in \left(\hat{\theta}_1\left(1-\frac{z_{1-\alpha/2}}{\sqrt{3n}}\right),\ \hat{\theta}_1\left(1+\frac{z_{1-\alpha/2}}{\sqrt{3n}}\right)\right).$$

$$\text{Length}_1 \approx 2\theta_1 \frac{z_{1-\alpha/2}}{\sqrt{3n}}$$

2) $\hat{\theta}_2 = \sqrt{3\overline{X^2}}$ $\quad \mathbb{E}[X_i^2] = \frac{\theta^2}{3}$ $\quad Var\, X_i^2 = \frac{4\theta^4}{45}$

by CLT. $\sqrt{n}\dfrac{(\overline{X^2}-\frac{\theta^2}{3})}{\sqrt{\frac{4\theta^4}{45}}} \xrightarrow{d} N(0,1)$

$g(x) = \sqrt{3x}$ cont. $\quad g'(\frac{\theta^2}{3}) = \sqrt{3}\cdot\dfrac{1}{2\sqrt{\frac{\theta^2}{3}}} = \dfrac{3}{2\theta} \neq 0.$ similarly by Prop. 4.2.

$$\sqrt{n}(\hat{\theta}_2-\theta) \xrightarrow{d} N(0, [g'(\tfrac{\theta^2}{3})]^2\cdot\tfrac{4\theta^4}{45}) = N(0,\tfrac{\theta^2}{5}) \quad \Rightarrow \hat{\theta}_2 \sim N(0,\tfrac{\theta^2}{5n}) \text{ asym.}$$

similarly $\theta \in \left(\hat{\theta}_2\left(1-\frac{z_{1-\alpha/2}}{\sqrt{5n}}\right), \hat{\theta}_2\left(1+\frac{z_{1+\alpha/2}}{\sqrt{5n}}\right)\right).$

$$\text{Length}_2 \approx 2\theta\frac{z_{1-\alpha/2}}{\sqrt{5n}}$$

since $\dfrac{\text{length}_1}{\text{length}_2} = \dfrac{\frac{1}{\sqrt{3}}}{\frac{1}{\sqrt{5}}} = \sqrt{\dfrac{5}{3}} > 1.$ length 2 is shorter

II.* (1 points) Let $X_1,\ldots,X_n$ be a sample from an exponential distribution with parameter β. Using asymptotically normal estimators

$$\hat{\beta}_1 = \frac{1}{\overline{X}}, \qquad \hat{\beta}_2 = \sqrt{\frac{2}{\overline{X^2}}},$$

construct asymptotic confidence intervals for β at level $1-\alpha$ and show that the second interval is shorter than the first.

Sol: $f(x) = \beta e^{-\beta x}$ $\quad x \geq 0.$

$\mathbb{E}X = \frac{1}{\beta}.$ $\quad Var\,X = \frac{1}{\beta^2}$ $\quad \mathbb{E}X^2 = \frac{2}{\beta^2}$ $\quad \mathbb{E}X^4 = \frac{24}{\beta^4}.$ $\quad Var\,X^2 = \frac{20}{\beta^4}$

1) $\hat{\beta}_1 = \frac{1}{\overline{X}}.$

by CLT. $\sqrt{n}(\overline{X}-\frac{1}{\beta}) \xrightarrow{d} N(0,\frac{1}{\beta^2})$

$f(x) = \frac{1}{x}$ $\quad f'(\frac{1}{\beta}) = -\beta^2 \neq 0.$

by Prop 4.2. $\quad \sqrt{n}(\hat{\beta}_1-\beta) \xrightarrow{d} N(0, (\beta^2)^2\cdot\frac{1}{\beta^2}) = N(0,\beta^2).$

$\beta \in \left(\hat{\beta}_1\left(1-\frac{z_{1-\alpha/2}}{\sqrt{n}}\right), \hat{\beta}_1\left(1+\frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\right)$ $\quad \text{Length}_1 \approx 2\beta\frac{z_{1-\alpha/2}}{\sqrt{n}}.$

2). $\hat{\beta}_2 = \sqrt{\frac{2}{\overline{X^2}}}.$

by CLT. $\sqrt{n}(\overline{X^2}-\frac{2}{\beta^2}) \xrightarrow{d} N(0,\frac{20}{\beta^4}).$

$g(x) = \sqrt{\frac{2}{y}}$ $\quad g'(\frac{2}{\beta^2}) = -\dfrac{\sqrt{2}}{2\cdot(\frac{2}{\beta^2})^{3/2}} = -\dfrac{\beta^3}{4} \neq 0.$

by prop.4.2. $\sqrt{n}(\widehat{\beta_2}-\beta) \xrightarrow{d} N(0, \frac{20}{\beta^4}\cdot(-\frac{\beta^3}{4})^2) = N(0, \frac{5\beta^2}{4})$.

$\beta \in (\widehat{\beta_2}(1-\frac{z_{1-\alpha/2}}{\sqrt{\frac{4n}{5}}}), \widehat{\beta_1}(1+\frac{z_{1-\alpha/2}}{\sqrt{\frac{4n}{5}}}))$. $\text{Length}_2 = 2\beta\frac{z_{1-\alpha/2}}{\sqrt{\frac{4}{5}n}}$

$\frac{\text{Length}_1}{\text{length}_2} = \frac{1}{\sqrt{\frac{4}{5}}} = \frac{\sqrt{5}}{2} > 1$. Length 2 shorter.

10. (4 points)
a) For a sample from a Poisson distribution with parameter λ, construct the most powerful test of asymptotic size ε to distinguish the hypothesis $\lambda = \lambda_1$ from the alternative $\lambda = \lambda_2$ if $\lambda_1 < \lambda_2$. Compute the limiting power of this test as $n \to \infty$.
b) A player observed a dice game and suspected that six appears in 18% of throws, five in 14%, and the other four faces equally likely (17% each). The player wants to test his hypothesis before participating in n consecutive throws. The only alternative considered is that the die is fair. For $n = 2$, find the most powerful test at level 0.0196.
c)* Denote by $\beta(\varepsilon)$ the power of the most powerful test among all randomized tests of level ε. Show that

$$\beta(\varepsilon) \geq \varepsilon.$$

Sol: a). $f(x;\lambda) = \prod(e^{-\lambda}\cdot\frac{\lambda^{x_i}}{(x_i)!}) = e^{-n\lambda}\lambda^{\sum x_i}\prod_{i=1}^{n}\frac{1}{x_i!}$

$\Lambda(x) = \frac{f_1(x;\lambda_2)}{f_0(x;\lambda_1)} = \exp(-n(\lambda_2-\lambda_1))\cdot(\frac{\lambda_2}{\lambda_1})^{\sum x_i}$

denote $S_n = \sum X_i$ $\quad \ln\Lambda(S) = -n(\lambda_2-\lambda_1) + S_n\cdot\ln\frac{\lambda_2}{\lambda_1}$.

since $\frac{\lambda_2}{\lambda_1} > 1$. $S_n \nearrow$ $\ln\Lambda(S)\nearrow$. we can use S_n to construct the level.

by N-P lemma. the MP test $\varphi_n(X) = \begin{cases} 1 & S_n > C_n \\ p_n & S_n = C_n \\ 0 & S_n < C_n \end{cases}$,

C_n, p_n s.t. $\alpha_n = P_{\lambda_1}(\varphi_n(X)=1) = P_{\lambda_1}(S_n > C_n) + p_n P_{\lambda_1}(S_n = C_n)$ we need $\lim \alpha_n = \varepsilon$.

i) find the (asymptotic) threshold C_n.

for $S_n \sim Pois(n\lambda)$. $\mathbb{E}[S_n] = n\lambda$. $Var[S_n] = n\lambda$.

for H_0. $\frac{S_n - n\lambda_1}{\sqrt{n\lambda_1}} \xrightarrow{d} N(0,1)$ and let $p_n = 0$.

let $C_n = n\lambda_1 + \sqrt{n\lambda_1}\, z_{1-\varepsilon}$ where $z_{1-\varepsilon} = \Phi^{-1}(1-\varepsilon)$.

then $\alpha_n = \mathbb{P}_{\lambda_1}(S_n \geq C_n) = \mathbb{P}_{\lambda_1}(\frac{S_n - n\lambda_1}{\sqrt{n\lambda_1}} \geq z_{1-\varepsilon}) \xrightarrow{n\to\infty} \varepsilon$.

ii) compute the limiting power

power $\beta_n(\lambda_2) = \mathbb{P}_{\lambda_2}(S_n \geq C_n) = (\frac{S_n - n\lambda_2}{\sqrt{n\lambda_2}} \geq \frac{n\lambda_1 + \sqrt{n\lambda_1}\cdot z_{1-\varepsilon} - n\lambda_2}{\sqrt{n\lambda_2}}) = (\frac{S_n - n\lambda_2}{\sqrt{n\lambda_2}} \geq -\sqrt{n}\frac{\lambda_2-\lambda_1}{\sqrt{\lambda_2}} + z_{1-\varepsilon}\sqrt{\frac{\lambda_1}{\lambda_2}})$

the RHS $\xrightarrow{n\to\infty} -\infty$ $\quad \beta_n(\lambda_2) \to 1$. i.e. $\lim_{n\to\infty}\beta_n(\lambda_2) = 1$.

b).

H_0:

X_i	1	2	3	4	5	6
p_i	0.17	0.17	0.17	0.17	0.14	0.18

H_1:

X_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

2 observation:

$$f_0(i,j) = P_0(Y_1 = i, Y_2 = j). \qquad f_1(i,j) = P_1(Y_1 = i, Y_2 = j) = \frac{1}{36}$$

$$\Lambda(i,j) = \frac{f_1(i,j)}{f_0(i,j)} = \frac{1}{36 \cdot P_0(i)\, P_1(j)}$$

$$\Lambda(5,5) > \Lambda(5,k) \text{ or } \Lambda(k,5) > \Lambda(5,6) \text{ or } \Lambda(6,5) > \Lambda(k,l) > \Lambda(6,k) \text{ or } \Lambda(k,6) > \Lambda(6,6).$$

$k, l \in \{1,2,3,4\}$.

by N-P lemma. $\varphi(X) = \begin{cases} 1 & f_1(x) > k f_0(x) \\ p(k) & f_1(x) = k f_0(x) \\ 0 & f_1(x) < k f_0(x). \end{cases}$

we have $f_0(5,5) = 0.14^2 = 0.0196$ exactly the level we want.

thus. the MP test: $\varphi(y_1, y_2) = \begin{cases} 1 & (y_1, y_2) = (5,5) \\ 0 & 0. \end{cases}$ $\leftarrow$ accept H_1. $\leftarrow$ accept H_0.

now we have the max power $\mathbb{P}_1(\varphi = 1) = \frac{1}{36}$.

11. I. (5 points) The table shows the results of a math exam. For each score range, the number of participants in that range is given as a percentage.

	I_1	I_2	I_3	I_4	I_5
Year	0–20	21–40	41–60	61–80	81–100
2018	3.85	31.13	32.96	29.92	2.13
2017	7.92	36.16	25.09	27.49	3.34

The total number of participants each year was $3.9 \cdot 10^5$.

a) (1 point) Since the exam scale has many possible values, we consider the score distribution as continuous. Unfortunately, the data allow reconstructing the empirical distribution function only at a few points. Can the Kolmogorov–Smirnov test convince us to reject the hypothesis, or do the data not contradict it?

b) (2 points) Test the hypothesis of homogeneity at significance level $\alpha = 0.05$.

Sol: I) a) do some accumulation

H_0: the total number of participant each year is 3.9×10^5. $\qquad H_1$: not H_0

	20	40	60	80	100
2018	3.85	34.98	67.94	97.86	1
2017	7.92	44.08	69.17	96.66	1
|diff|	4.07	9.10	1.23	1.20	0

$D_{obs.} = \max|\text{diff}| = 0.091$

$D_n \geq 0.091$

define the KS-statistics. $T_{N,N} = \sqrt{\frac{N^2}{2N}}\, D_n \geq 441 \times 0.091 \approx 40$

Rayleigh (1/2): $f(x) = 4x \cdot e^{-2x^2}$ $\leftarrow$ 40 is too large for this distribution

thus we can reject the hypothesis that "the total number of participant each year is 3.9×10^5" by KS-test.

b). denote. 2018. $n_1 = N$. theoretical $p = (p_1 \ldots p_5)$ frequency V_{1j}. $j = 1,2,3,4,5$

2017 $n_2 = N$ theorectical $q = (q_1 \ldots q_5)$ frequency V_{2j}. $j = 1,2,3,4,5$.

$H_0: p=q \quad H_1: \text{not } H_0 \quad (\text{homogeneity}).$

$L_{H_0}(p) \propto \prod p_j^{\nu_{1j}+\nu_{2j}}$ MLE: $\tilde{p}_j = \frac{\nu_{1j}+\nu_{2j}}{2N}$

$L_{H_1}(p,q) \propto \prod p_j^{\nu_{1j}} \prod q_j^{\nu_{2j}}$

define $T = \sum_{\substack{i=1,2 \\ j\in[1:5]}} \frac{(\nu_{ij} - N\tilde{p}_j)}{N\tilde{p}_j}$ (if H_0 holds. $T \xrightarrow{d} \chi^2_{s-1}$, $s=5$)

$N\tilde{p}_j = (5.885, 33.645, 29.025, 28.705, 2.735)$ $T_{obs} \approx 3.22$

when, $\alpha = 0.05$. $\chi^2_{4,0.95} \approx 9.49 < 3.22$. we can't reject H_0.

II. (2 points) Check part b using the χ^2 test.

II.

	I_1	I_2	I_3	I_4	I_5	ν_j
2018	3.85	31.13	32.96	29.92	2.13	100
2017	7.92	36.16	25.09	27.49	3.34	100
ν_i	11.77	67.29	58.05	57.41	5.74	200

in χ^2-test. H_0: hypothesis of independent

$\hat{p}_{ij} = \frac{\nu_i \cdot \nu_j}{N^2}$.

$$\chi^2 = \sum\sum \frac{(\nu_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \xrightarrow[n\to\infty]{d} \chi^2_{(s-1)(r-1)}$$

$df = (s-1)(r-1) = (5-1)\times(2-1) = 4.$ $\chi^2_{4,0.95} \approx 9.49$

$\chi^2 \approx 3.22$. (same expression as T in I. b).) also can't reject H_0.

12. (4 points) Consider the following data of X_i values and their frequencies ν_i:

X_i	1	2	3	4	5	6	>6 7
ν_i	4	6	5	4	3	2	1

Test the goodness-of-fit at significance level $\alpha = 0.2$ for the following distributions:
a) Poisson distribution;
b) geometric distribution of the first type.

$\sum X_i\nu_i = 4+12+15+16+15+8+7 = 81$ $\bar{X} = \frac{81}{25} = 3.24$.

a). Likelihood: $L(\lambda) = \prod \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ $\ln L(\lambda) = -n\lambda + (\sum X_i)\ln\lambda - \sum \ln(X_k)$

$\frac{\partial \ln L(\lambda)}{\partial\lambda} = -n + \frac{\sum X_k}{\lambda}$. $\Rightarrow \hat{\lambda} = \bar{X} = 3.24$

compute $E_i = np_i(\hat{\lambda})$

i	1	2	3	4	5	6	>6
E_i	3.17	5.14	5.55	4.50	2.91	1.57	2.16

$\chi^2_{obs} = \sum \frac{(\nu_i - E_i)^2}{E_i} \approx 1.21$ $s=7$ $d=1$. $df = 7-1-1 = 5$.

$\alpha = 0.2$. $\chi^2_{5,0.8} \approx 7.29 > 1.21$. can't reject H_0.

b). $\mathbb{P}(X=k) = p(1-p)^{k-1}$. $\mathbb{E}[X] = \frac{1}{p}$.

$L(p) = p^n(1-p)^{\sum x_k - n}$.

$$\frac{\partial \ln L(p)}{\partial p} = \frac{n}{p} - \frac{\sum x_k - n}{1-p}. \qquad \Rightarrow \quad \hat{p} = \frac{1}{\bar{X}} \approx 0.309.$$

compute E_i. $n=25$.

i	1	2	3	4	5	6	>6
E_i	7.72	5.33	3.69	2.55	1.76	1.22	2.73

$$\chi^2_{obs} = \sum \frac{(\nu_i - E_i)^2}{E_i} \approx 5.63.$$

$\alpha = 0.2$. $\chi^2_{5,0.8} \approx 7.29 > 5.63$. can't reject H_0.

13. (2 points) Out of 300 applicants, 97 had a grade of "5" at school, and 48 received "5" on the entrance exam in the same subject, with only 18 having "5" both at school and on the exam. Test the hypothesis of independence of receiving a "5" at school and on the exam.

Sol: event A: {"5" at school}. B: {"5" on exam}

	B	B^c	sum
A	18	79	97
A^c	30	173	203
sum	48	252	300

H_0: $\exists\, P_1, P_2$, s.t. $P(A,B) = P_1(A)P_2(B)$.

H_1: not independent.

under H_0, $\hat{p}_i = \frac{\nu_i}{n}$ $\quad \hat{p}_j = \frac{\nu_j}{n}$. $\quad \hat{p}_{ij} = \hat{p}_i \hat{p}_j = \frac{\nu_i \nu_j}{n^2}$ $\quad E_{ij} = n\hat{p}_{ij} = \frac{\nu_i \nu_j}{n}$.

we have $E = \begin{pmatrix} 15.52 & 81.48 \\ 32.48 & 170.52 \end{pmatrix}$.

$$\chi^2_{obs} = \sum \frac{(\nu_{ij} - E_{ij})^2}{E_{ij}} \approx \frac{(18-15.52)^2}{15.52} + \frac{(81.48-79)^2}{81.48} + \frac{(32.48-30)^2}{32.48} + \frac{(173-170.52)^2}{170.52} \approx 0.70.$$

$k = (r-1)(s-1) = 1$. $\chi^2_{1,0.95} = 3.84 > 0.70$. in level $\alpha = 0.10$. can't reject H_0.

$\chi^2_{1,0.90} = 2.71 > 0.70$