Problems

The expressions for the probability mass function of the Poisson(λ) distribution, and the density function of the Normal distribution with mean μ and variance σ^2, may be useful:

$$\frac{\lambda^n}{n!}e^{-\lambda} \qquad \text{and} \qquad \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

1. (10 points) For each of the following questions, exactly one answer is correct. Each correct answer gives 1 point, and each incorrect answer results in a 1/2 point reduction. The minimal possible total score for the full problem is 0. *Some of the results may be useful in later problems!*

a) Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $0 < \mathbb{E}(X_1^2) < \infty$. Which of the following statements is a consequence of the strong law of large numbers?

1. $\limsup_{n\to\infty} X_n = +\infty$ almost surely.
2. $\lim_{n\to\infty}(X_1^2 + \cdots + X_n^2)/n = 0$ almost surely.
3. $\sup_{n\geq 1} |X_n| < \infty$ almost surely.

b) Let φ_1 and φ_2 be the characteristic functions of two random variables X_1 and X_2. What does the identity theorem for characteristic functions tell us?

1. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 have the same distribution.
2. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 are equal almost surely.
3. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 are independent.

c) What does it mean for a statistical test to have significance level 0.05?

1. The Null hypothesis will be accepted with probability at least 0.05.
2. If the Null hypothesis is true, then the probability to reject is at most 0.05.
3. If the Null hypothesis is false, then the probability to reject is at least 0.95.

d) The characteristic function φ of a Poisson(λ) random variable is given by

1. $\varphi(u) = \exp(u(e^{\mathrm{i}\lambda} - 1))$.
2. $\varphi(u) = \exp(\lambda(e^{\mathrm{i}u} - 1))$.
3. $\varphi(u) = \exp(\lambda(\mathrm{i}u - 1))$.

e) Suppose a statistical test resulted in a p-value of 0.07. Which of the following statements is correct?

1. The probability that the Null hypothesis is false is 0.07.
2. The Null hypothesis could not be rejected at significance level 0.06.
3. The Null hypothesis could not be rejected at significance level 0.08.

f) Suppose $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$. What is the distribution of $\frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$?

1. $N(0, 1)$.
2. $N(0, 1/n)$.
3. Student t with n degrees of freedom.

g) Suppose $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p) for some $0 < p < 1$. What is the distribution of $X_1 + \cdots + X_n$?

1. Binomial(n, p).
2. Poisson(np).
3. Geometric(n, p).

h) Which of the following formulas is **not** correct in general?

1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A \cap B^c)$.

i) Which of the following statements is correct for any random variables X, Y with unit variance, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$?

1. If $\mathrm{Cov}(X, Y) = 0$ then X and Y are independent.
2. If $X = a + bY$ for some constants $a \in \mathbb{R}$ and $b > 0$, then $\mathrm{Cov}(X, Y) = 1$.
3. If $X = Y^2$, then $\mathrm{Cov}(X, Y) = 1$.

j) What is the value of $\sum_{k=0}^{n-1} 2^k$?

1. $2^n - 1$.
2. $n^2 - 1$.
3. $2^n - n$.

2. (10 points) A team of particle physicists has performed a series of experiments using the Large Hadron Collider (LHC) to study Higgs boson production. The total number of produced Higgs bosons is $N \sim \text{Poisson}(\lambda)$. Unfortunately, due to background effects, other particles are also produced during the experiments. The total number of such particles is $M \sim \text{Poisson}(\gamma)$, with M and N independent. The parameters λ and γ are nonnegative and unknown.

The detector unit can only measure the total number of produced particles, $X = M + N$.

a) Show that $X \sim \text{Poisson}(\lambda + \gamma)$.

Solution:

Method one By the properties of characteristic functions,

$$\varphi_X(u) = \varphi_{M+N}(u) = \varphi_M(u)\varphi_N(u) = \exp(\gamma(e^{\mathrm{i}u}-1))\exp(\lambda(e^{\mathrm{i}u}-1)) = \exp((\lambda+\gamma)(e^{\mathrm{i}u}-1)).$$

The identity theorem for characteristic functions yields the result.

Method two We just have to compute

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{n=0}^{k} \mathbb{P}(N = n)\mathbb{P}(M = n - k) \\
&= \sum_{n=0}^{k} e^{-\lambda}\frac{\lambda^n}{n!}e^{-\gamma}\frac{\gamma^{n-k}}{(n-k)!} \\
&= e^{-\lambda+\gamma}\frac{1}{k!}\sum_{n=0}^{k}\binom{k}{n}\lambda^n\gamma^{n-k} \\
&= e^{-\lambda+\gamma}\frac{1}{k!}(\lambda+\gamma)^k.
\end{aligned}$$

The result is just the probability mass function of a $\text{Poisson}(\lambda + \gamma)$.

The parameter of interest is λ, which controls the number of Higgs bosons produced. To proceed, further information about the background parameter γ is needed. Therefore, the physicists perform a separate experiment from which a random variable $Y \sim N(\gamma, \sigma^2)$ is observed independently of X, where $\sigma^2 > 0$ is known.

b) The joint distribution of (X, Y) given the parameters $\theta = (\lambda, \gamma)$ is of the form

$$\mathbb{P}_\theta(X = n, Y \leq z) = \int_{-\infty}^{z} f(n, y; \theta)dy.$$

Determine $f(n, y; \theta)$.

Solution:

By independence and the form of the Poisson and Normal distribution functions, we have

$$\mathbb{P}(X = n, Y \leq z) = \mathbb{P}(X = n)\mathbb{P}(Y \leq z) = \frac{(\lambda+\gamma)^n}{n!}e^{-(\lambda+\gamma)}\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-\gamma)^2/(2\sigma^2)}dy.$$

Thus

$$f(n, y; \theta) = \frac{(\lambda+\gamma)^n}{n!} e^{-(\lambda+\gamma)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\gamma)^2/(2\sigma^2)}$$

Consider the likelihood function $L(\theta) = f(X, Y; \theta)$.

c) Find the maximum likelihood estimate $(\widehat{\lambda}, \widehat{\gamma})$ of (λ, γ), knowing X and Y, in the case where $0 < Y < X$.

Solution: Change variables to $(\mu, \gamma) = (\lambda + \gamma, \gamma)$. The log-likelihood then becomes

$$g(\mu, \gamma) := \ell(\mu - \gamma, \gamma) = \text{constant } + X \log(\mu) - \mu - \frac{1}{2\sigma^2}(Y - \gamma)^2.$$

The unconstrained maximizer is $(\widehat{\mu}_{\text{unconstr}}, \widehat{\gamma}_{\text{unconstr}}) = (X, Y)$. Hence $(\widehat{\lambda}_{\text{unconstr}}, \widehat{\gamma}_{\text{unconstr}}) = (X - Y, Y)$. Since $X - Y > 0$ and $X > 0$ by assumption, the unconstrained maximizer coincides with the constrained maximizer. That is, $(\widehat{\lambda}, \widehat{\gamma}) = (X - Y, Y)$.

3. (15 points) Consider a series of coin tosses modeled by a sequence $X_1, X_2, \ldots$ of i.i.d. random variables with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$. Here $X_n = 1$ signifies heads and $X_n = -1$ tails. Before each toss, you have the opportunity to bet an amount V_n. If the coin comes up heads ($X_n = 1$), you receive V_n. Otherwise you lose V_n. Therefore the total gain up to and including the nth toss is

$$G_n = \sum_{k=1}^{n} V_k X_k.$$

Note that G_n may be negative, which signifies a loss.

Suppose you use the following strategy: Set $V_1 = 1$. For $k \geq 1$, if $X_k = 1$ you stop, meaning that you set $V_n = 0$ for $n \geq k+1$. If $X_k = -1$ and you have not yet stopped, double your bet, that is, set $V_{k+1} = 2\,V_k$.

Furthermore, let T denote the time you stop, $T = \min\{k \geq 1 : X_k = 1\}$, with $T = \infty$ if $X_k = -1$ for all k.

a) Compute G_n for $1 \leq n < T$.

Solution: Note that $(X_1, \ldots, X_{T-1}, X_T) = (-1, \ldots, -1, 1)$ for $T > 1$, and that $V_k = 2^{k-1}$ for all $k < T$. Thus, for $n < T$,

$$G_n = \sum_{k=1}^{n} 2^{k-1} \times (-1) = -\sum_{k=0}^{n-1} 2^k = 1 - 2^n.$$

b) Compute G_T for $T < \infty$.

Solution: Since $X_T = 1$ and $V_T = 2^T$ we have from the previous problem, if $T > 1$,

$$G_T = G_{T-1} + V_T X_T = 1 - 2^T + 2^T = 1.$$

If $T = 1$, then clearly $G_T = 1 \times 1 = 1$. So in all cases, $G_T = 1$.

c) Compute $\mathbb{P}(T < \infty)$.

Solution: Since $\{T = \infty\} = \{X_n = -1 \text{ for all } n \geq 1\} = \bigcap_{n\geq 1}\{X_n = -1\}$, we have

$$\mathbb{P}(T = \infty) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=1}^{n}\{X_k = -1\}\right) = \lim_{n\to\infty} \prod_{k=1}^{n} \mathbb{P}\left(X_k = -1\right) = \lim_{n\to\infty} 2^{-n} = 0,$$

using the continuity property of probability measures as well as the independence of $\{X_n : n \geq 1\}$. Thus $\mathbb{P}(T < \infty) = 1$.

At this point you may feel that this is too good to be true—and indeed there is a catch:

d) Compute the expected maximal intermediate shortfall. That is, compute $\mathbb{E}\big(\min_{1\leq n\leq T} G_n\big)$.

Solution: From parts (a) and (b) we have $G_n = 1 - 2^n$ for $n < T$, and $G_T = 1$. Thus

$$\min_{1\leq n\leq T} G_n = \begin{cases} 1 - 2^{T-1} & \text{if } T > 1 \\ 1 & \text{if } T = 1. \end{cases}$$

Also, for $n \geq 1$,

$$\mathbb{P}(T = n) = \mathbb{P}(X_1 = \cdots = X_{n-1} = -1 \text{ and } X_n = 1) = \prod_{k=1}^{n} \frac{1}{2} = 2^{-n}.$$

Thus

$$\mathbb{E}\big(\min_{1 \leq n \leq T} G_n\big) = 1 \times \frac{1}{2} + \sum_{n=2}^{\infty} (1 - 2^{n-1}) \times 2^{-n} = \frac{1}{2} + \sum_{n=2}^{\infty} (2^{-n} - 2^{-1}) = -\infty.$$

To make matters worse, in reality you would face a limit on credit. Specifically, assume you are only allowed to use strategies $\{V_n : n \geq 1\}$ such that the gain satisfies $G_n \geq -b$ almost surely for all n, where $b \in (1, \infty)$ is some fixed constant. Modify the above strategy so that you stop as soon as there is a risk of violating the bound. That is, specify V_k as before, except that if you have not stopped at k, and $G_k - 2\,V_k < -b$, then set $V_n = 0$ for all $n \geq k+1$.

Let T' denote the time you stop under this modified strategy.

e) Compute your expected gain $\mathbb{E}(G_{T'})$ under these new rules. In particular, how does this depend on the bound b?

Solution: For $1 \leq n < T$ we have $G_n - 2V_n = 1 - 2^n - 2 \times 2^{n-1} = 1 - 2^{n+1}$. Thus

$$G_n - 2V_n < -b \qquad \Longleftrightarrow \qquad n > \log(1+b) - 1.$$

Let n_0 denote the smallest n for which this happens. Then $T' = T$ on the event $\{T \leq n_0\}$, and $T' = n_0$ on the event $\{T > n_0\}$. Thus,

$$\begin{aligned}
\mathbb{E}(G_{T'}) &= \mathbb{E}(G_T \mathbf{1}_{\{T \leq n_0\}} + G_{n_0} \mathbf{1}_{\{T > n_0\}}) \\
&= 1 \times \mathbb{P}(T \leq n_0) + (1 - 2^{n_0}) \times \mathbb{P}(T > n_0) \\
&= 1 \times (1 - \mathbb{P}(T > n_0)) + (1 - 2^{n_0}) \times \mathbb{P}(T > n_0) \\
&= 1 - 2^{n_0} \mathbb{P}(T > n_0) \\
&= 0,
\end{aligned}$$

using that $\mathbb{P}(T > n_0) = \mathbb{P}(X_1 = \cdots = X_{n_0} = -1) = 2^{-n_0}$. This result is completely independent of the actual value of the bound b!

4. (15 points) Celebrities Saylor Twift and Bustin Jieber are active on Twitter. Saylor has a total of n followers. There is some overlap between Twift and Jieber followers: a fraction $\gamma \in (0,1)$ of Saylor's followers also follow Bustin (that is, $[\gamma n]$ people follow both Saylor and Bustin, where $[x]$ denotes the integer part of a real number x).

Let A denote the event that Bustin tweets "`Saylor Twift got a #beautifulvoice`". If a follower of Saylor reads this tweet, then this follower will re-tweet Saylor's tweets with probability $p_S = 0.3$. Otherwise, the probability is only $q_S = 0.1$. Followers make tweeting decisions independently of each other, and do not re-tweet each other's tweets.

One day, Saylor tweets "`today glimpsed a new world who knew that #statsrulez?`". Let N denote the total number of re-tweets of Saylor's tweet.

a) Assuming that A **did not** happen, find $\mathbb{E}(N)$ and $\mathrm{Var}(N)$.

Solution: If A did not happen, then $N = X_1 + \cdots + X_n$ for i.i.d. random variables $X_i \sim \mathrm{Bernoulli}(q_S)$. Thus $\mathbb{E}(N) = nq_S$, $\mathrm{Var}(N) = nq_S(1-q_S)$.

b) Assuming that A **did** happen, find $\mathbb{E}(N)$ and $\mathrm{Var}(N)$.

Solution: If A did happen, then $N = X_1 + \cdots + X_{[\gamma n]} + Y_1 + \cdots + Y_{n-[\gamma n]}$ for independent random variables $X_i \sim \mathrm{Bernoulli}(p_S)$ and $Y_i \sim \mathrm{Bernoulli}(q_S)$. Thus

$$\begin{aligned}\mathbb{E}(N) &= [\gamma n]p_S + (n - [\gamma n])q_S \\ \mathrm{Var}(N) &= [\gamma n]p_S(1-p_S) + (n-[\gamma n])q_S(1-q_S).\end{aligned}$$

Suppose you work for Twitter and a colleague tells you the value of N. You are currently offline and cannot check specific tweets. You are interested in if A happened or not.

c) Under each of the distributions in a) and b), show that

$$\frac{N - \mathbb{E}(N)}{\sqrt{\mathrm{Var}(N)}}$$

converges weakly to the standard Normal distribution as $n \to \infty$, with p_S, q_S, and γ held fixed.

Hint: It may be helpful to write N as a sum of Bernoulli random variables. Furthermore, you are allowed to use the following result:

Lemma: Let $U_1, U_2, \ldots$ and $V_1, V_2, \ldots$ be two sequences of random variables that both converge weakly to the standard Normal distribution. Suppose U_n and V_n are independent for each n. Let $\varrho_n \in (0,1)$ and assume $\lim_n \varrho_n = \varrho \in (0,1)$. Then $\varrho_n U_n + \sqrt{1-\varrho_n^2}V_n$ converges weakly to the standard Normal distribution.

Solution: (i): If A did not happen, then $N = S_n := X_1 + \cdots + X_n$ for i.i.d. random variables $X_i \sim \mathrm{Bernoulli}(q_S)$. Thus, as $n \to \infty$,

$$\frac{N - \mathbb{E}(N)}{\sqrt{\mathrm{Var}(N)}} = \frac{S_n - nq_S}{\sqrt{n}\sqrt{q_S(1-q_S)}} \to N(0,1)$$

weakly by the Central Limit Theorem.
(ii): If A did happen, then $N = S_n^{(1)} + S_n^{(2)}$, where

$$S_n^{(1)} = X_1 + \cdots + X_{[\gamma n]}$$
$$S_n^{(2)} = Y_1 + \cdots + Y_{n-[\gamma n]}$$

for independent random variables $X_i \sim \text{Bernoulli}(p_S)$ and $Y_i \sim \text{Bernoulli}(q_S)$.
Method one The Central Limit Theorem yields

$$U_n := \frac{S_n^{(1)} - [\gamma n]p_S}{\sqrt{[\gamma n]}\sqrt{p_S(1-p_S)}} \to N(0,1)$$
$$V_n := \frac{S_n^{(2)} - (n-[\gamma n])q_S}{\sqrt{(n-[\gamma n])}\sqrt{q_S(1-q_S)}} \to N(0,1)$$

weakly as $n \to \infty$. Moreover, we have

$$\frac{N - \mathbb{E}(N)}{\sqrt{\text{Var}(N)}} = \varrho_n U_n + \sqrt{1-\varrho_n^2}V_n,$$

where

$$\varrho_n = \left(\frac{\text{Var}(S_n^{(1)})}{\text{Var}(N)}\right)^{1/2} = \left(\frac{[\gamma n]p_S(1-p_S)}{[\gamma n]p_S(1-p_S) + (n-[\gamma n])q_S(1-q_S)}\right)^{1/2}$$
$$\to \left(\frac{\gamma p_S(1-p_S)}{\gamma p_S(1-p_S) + (1-\gamma)q_S(1-q_S)}\right)^{1/2} \in (0,1).$$

Since U_n and V_n are independent for each n, the lemma then gives the desired conclusion.
Method two If we do not one to use the hint we can prove this with Lindeberg's Theorem (4.4 page 63 of the Skript), define

$$Z_{n,i} := \begin{cases} \frac{X_i - \mathbb{E}(X_i)}{\sqrt{\text{Var}(N)}} & i \leq [\gamma n], \\ \frac{Y_{[\gamma n]+i} - \mathbb{E}\left(Y_{[\gamma n]+i}\right)}{\sqrt{\text{Var}(N)}} & [\gamma n] < i \leq n. \end{cases}$$

now we just have to check the three conditions

a) For fixed n the independence comes from the original model.
b) It's clear that $\mathbb{E}\left(Z_{n,i}\right) = 0$, $\mathbb{E}(\mathrm{X}_{\mathrm{n,i}}^2) < \infty$ and that $\sum_{i=1}^n \text{Var}(\mathrm{X}_{\mathrm{n,i}}) = 1$.
c) We just have to see that for all $\epsilon > 0$ and n big enough, $\mathbf{1}_{Z_{n,i} \geq \epsilon} = 0$, thanks to the fact that $\text{Var}(N) \to \infty$ and $0 \leq X, Y \leq 1$.

d) Consider the hypotheses

$$\begin{cases} \text{Null:} & \text{The event } A \text{ happened;} \\ \text{Alternative:} & \text{The event } A \text{ did not happen.} \end{cases}$$

In this problem, you should approximate the corresponding distributions of N by Normals with the same mean and variance. Show that one can find a most powerful test at a given level α of the form

$$\text{reject the Null} \quad \Longleftrightarrow \quad |N + 0.05\,n| > c$$

for some constant c. (You do not have to determine c; it depends on α and the distribution of N under the Null hypothesis.)

Solution: Let μ_0 and σ_0^2 be the mean and variance of N under the Null hypothesis, and let μ_1 and σ_1^2 be the mean and variance of N under the alternative. The f_0 and f_1 denote the corresponding Normal density functions. Under the Normal approximation, the likelihood ratio is

$$\frac{f_1(x)}{f_0(x)} = \frac{(2\pi\sigma_0^2)^{-1/2}\exp(-(x-\mu_0)^2/(2\sigma_0^2))}{(2\pi\sigma_1^2)^{-1/2}\exp(-(x-\mu_1)^2/(2\sigma_1^2))} = \frac{\sigma_1}{\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2} + \frac{(x-\mu_1)^2}{2\sigma_1^2}\right).$$

Thus, for some constants c_i,

$$\begin{aligned}\frac{f_1(x)}{f_0(x)} > c_1 &\Longleftrightarrow -\frac{(x-\mu_0)^2}{\sigma_0^2} + \frac{(x-\mu_1)^2}{\sigma_1^2} > c_2 \\ &\Longleftrightarrow x^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) - 2x\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right) > c_3.\end{aligned}$$

Since a) and b) and the given values $p_S = 0.3$ and $q_S = 0.1$ yield

$$\sigma_0^2 - \sigma_1^2 = [\gamma n](p_S(1-p_S) - q_S(1-q_S)) = 0.12 \times [\gamma n] > 0,$$

we may divide by $\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} > 0$ to get the equivalent statement

$$\Longleftrightarrow x^2 - 2x\left(\frac{\mu_1\sigma_0^2 - \mu_0\sigma_1^2}{\sigma_0^2 - \sigma_1^2}\right) > c_4,$$

and upon completing the square,

$$\begin{aligned}&\Longleftrightarrow \left(x - \frac{\mu_1\sigma_0^2 - \mu_0\sigma_1^2}{\sigma_0^2 - \sigma_1^2}\right)^2 > c_5 \\ &\Longleftrightarrow \left|x - \frac{\mu_1\sigma_0^2 - \mu_0\sigma_1^2}{\sigma_0^2 - \sigma_1^2}\right| > c_6 =: c.\end{aligned}$$

Using again a) and b), we get

$$\mu_1\sigma_0^2 - \mu_0\sigma_1^2 = n\,[\gamma n]\,p_S q_S(q_S - p_S) = -0.006 \times n\,[\gamma n],$$

and thus

$$\frac{\mu_1\sigma_0^2 - \mu_0\sigma_1^2}{\sigma_0^2 - \sigma_1^2} = \frac{-0.006 \times n\,[\gamma n]}{0.12 \times [\gamma n]} = -0.05\,n.$$

The given test is thus most powerful by the Neyman-Pearson lemma, with a suitably chosen c depending on α.

5. **(15 points)** A common measure of the quality of an estimator is the mean squared error (MSE). In this problem, you will see that the MSE may behave in unexpected ways. This was first noticed by Stein in 1956 and developed further by James and Stein in 1961.

Fix $n \geq 2$. Let $X = (X_1, \cdots, X_n)$ be a random vector, where $X_1, \ldots, X_n$ are independent with $X_i \sim N(\theta_i, 1)$ for some unknown parameters $\theta = (\theta_1, \ldots, \theta_n)$.

a) Consider the estimator $\widehat{\theta}_i = X_i$ for $i = 1, \ldots, n$, and set $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$. Compute the mean squared error

$$\mathbb{E}(\|\theta - \widehat{\theta}\|^2)$$

Solution:

$$\mathbb{E}(\|\theta - \widehat{\theta}\|^2) = \mathbb{E}\Big(\sum_{i=1}^{n}(X_i - \theta_i)^2\Big) = \sum_{i=1}^{n}\mathbb{E}\big((X_i - \theta_i)^2\big) = \sum_{i=1}^{n}\mathrm{Var}(X_i) = n.$$

Consider now the alternative estimator $\widehat{\theta}^{\mathrm{JS}} = (\widehat{\theta}_1^{\mathrm{JS}}, \ldots, \widehat{\theta}_n^{\mathrm{JS}})$, where

$$\widehat{\theta}_i^{\mathrm{JS}} = X_i - \frac{n-2}{\|X\|^2}X_i, \qquad i = 1, \ldots, n.$$

b) Let $h : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and such that $h(x) = 0$ whenever $\|x\|$ is sufficiently large (that is, for $h(x) = 0$ whenever $\|x\| > C$ for some constant C). Show that

$$\mathbb{E}\Big((X_i - \theta_i)h(X)\Big) = \mathbb{E}\left(\frac{\partial h}{\partial x_i}(X)\right) \qquad \text{for all } i = 1, \ldots, n. \tag{1}$$

Hint: Use that $\phi'(t) = -t\phi(t)$, where ϕ denotes the $N(0,1)$ density function.

Solution: It is enough to consider $i = 1$. Then,

$$\begin{aligned}
&\mathbb{E}\Big((X_1 - \theta_1)h(X)\Big) \\
&\quad = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi(x_2 - \theta_2)\cdots\phi(x_n - \theta_n) \\
&\qquad \times \int_{-\infty}^{\infty} h(x_1, \ldots, x_n)(x_1 - \theta_1)\phi(x_1 - \theta_1)dx_1 dx_2 \cdots dx_n.
\end{aligned}$$

By the hint and integration by parts (using the assumption that h vanishes for large x to take care of the boundary terms), we obtain

$$\begin{aligned}
\int_{-\infty}^{\infty} h(x_1, \ldots, x_n)(x_1 - \theta_1)\phi(x_1 - \theta_1)dx_1 &= -\int_{-\infty}^{\infty} h(x_1, \ldots, x_n)\phi'(x_1 - \theta_1)dx_1 \\
&= \int_{-\infty}^{\infty}\frac{\partial h}{\partial x_1}(x_1, \ldots, x_n)\phi(x_1 - \theta_1)dx_1.
\end{aligned}$$

Substituting back into the previous expression yields

$$\mathbb{E}\Big((X_1 - \theta_1)h(X)\Big) = \mathbb{E}\left(\frac{\partial h}{\partial x_i}(X)\right).$$

Equation (1) actually holds for more general functions h. In particular, for any $i \in \{1, \dots, n\}$, one can take $h(x) = x_i/|x|^2$ (setting $h(0) = 0$ and $\frac{\partial h}{\partial x_i}(0) = 0$), as long as $n \geq 3$. You may use this fact without proof.

c) For $n \geq 3$, show that $\widehat{\theta}^{\mathrm{JS}}$ has smaller mean squared error than $\widehat{\theta}$. That is, show that

$$\mathbb{E}(\|\theta - \widehat{\theta}^{\mathrm{JS}}\|^2) < \mathbb{E}(\|\theta - \widehat{\theta}\|^2).$$

Solution: Compute:

$$\begin{aligned}
\|\theta - \theta^{\mathrm{JS}}\|^2 &= \sum_{i=1}^n (\theta_i - X_i + \frac{n-2}{\|X\|^2} X_i)^2 \\
&= \sum_{i=1}^n (\theta_i - X_i)^2 + 2(n-2) \sum_{i=1}^n (\theta_i - X_i) \frac{X_i}{\|X\|^2} + \sum_{i=1}^n \frac{(n-2)^2}{\|X\|^4} X_i^2 \\
&= \|\theta - \widehat{\theta}\|^2 + 2(n-2) \sum_{i=1}^n (\theta_i - X_i) \frac{X_i}{\|X\|^2} + \frac{(n-2)^2}{\|X\|^2}.
\end{aligned}$$

Note that $\frac{\partial}{\partial x_i}(x_i/\|x\|^2) = \frac{\|x\|^2 - 2x_i^2}{\|x\|^4}$. Thus by (b) and the subsequent comment,

$$\mathbb{E}\Big((\theta_i - X_i)\frac{X_i}{\|X\|^2}\Big) = -\mathbb{E}\Big(\frac{\|X\|^2 - 2X_i^2}{\|X\|^4}\Big).$$

Taking expectations of the above expression for $\|\theta - \theta^{\mathrm{JS}}\|^2$, we thus obtain

$$\begin{aligned}
\mathbb{E}(\|\theta - \widehat{\theta}^{\mathrm{JS}}\|^2) &= \mathbb{E}(\|\theta - \widehat{\theta}\|^2) - 2(n-2)\mathbb{E}\Big(\frac{n\|X\|^2 - 2\|X\|^2}{\|X\|^4}\Big) + \mathbb{E}\Big(\frac{(n-2)^2}{\|X\|^2}\Big) \\
&= \mathbb{E}(\|\theta - \widehat{\theta}\|^2) - (n-2)^2 \mathbb{E}\Big(\frac{1}{\|X\|^2}\Big).
\end{aligned}$$

If $n \geq 3$, this is strictly less than zero, as required.

This result is counterintuitive, because the X_i are independent. It is tempting to conclude that "the quality of an estimate can be improved by simultaneously estimating independent variables". Food for thought...