SDS 8102: Advanced Mathematical Statistics

Homework 1

Fall 2026

MBZUAI

Due Date: Sep 11, 2026 (11:59 PM)

Guidelines:

- Start working on the homework early. If you have questions regarding notation, problem statements, or possible errors, please contact us as soon as possible. You *must begin your email* with the subject line `HW [homework number]-CLARIFICATION-SDS 8102 (FALL 2026)`. Without the subject line, the email may get ignored.
- Prepare a LaTeX file for each assignment and only submit your final PDF file in the LMS system. Make sure *your name is clearly written* in the header. Start the solution to each main problem on a *new page*. Any code used in your solutions should be included in an appendix. Submit one combined PDF containing all of your solutions.
- The problems may be adapted from course materials, textbooks, and other standard references. If an external resource plays a substantial role in developing your solution, *please cite it appropriately.*
- The goal of the problems is to develop both intuition and technical skills. You may explore beyond what is explicitly asked. If you introduce additional assumptions, clearly state them. You are also welcome to include short remarks about interesting observations that arise while solving the problems.
- Problems or parts marked [**Bonus**] are optional and are not required for full credit. If you attempt a bonus problem, clearly label the corresponding solution.
- Some problems contain hints. These are intended only as suggestions for getting started; you are free to use a different approach.
- Solutions will be evaluated primarily on mathematical correctness and clarity of reasoning. There is no need to spend excessive effort on minor presentation details.
- **AI use:** AI-generated solutions are strictly discouraged in this course. You may use AI for minor polishing of the presentation. However, you should produce the core logic by yourself.

1 Uniform Convergence of the Empirical CDF on a Compact Support (30 points)

This exercise will help you develop technical skills and intuition that are useful in advanced statistical theory and research.

Let $F : \mathbb{R} \to [0, 1]$ be a continuous cumulative distribution function supported on $[0, 1]$. Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F,$$

and define the empirical cumulative distribution function

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}, \qquad x \in [0, 1].$$

During the lecture, we showed that, for every fixed $x \in [0, 1]$,

$$\widehat{F}_n(x) \xrightarrow{p} F(x).$$

The goal of this problem is to strengthen pointwise convergence to uniform convergence:

$$\sup_{x \in [0,1]} \left| \widehat{F}_n(x) - F(x) \right| \xrightarrow{p} 0.$$

1. **The MGF at a fixed point.** [4 points]

   Fix $x \in [0, 1]$, and define

   $$Y_i(x) = \mathbb{1}\{X_i \leq x\}, \qquad \Delta_i(x) = Y_i(x) - F(x).$$

   (a) Show that $Y_i(x)$ has a Bernoulli distribution with success probability $F(x)$. [1 point]

   (b) Compute the moment-generating function

   $$M_{\Delta_i(x)}(\lambda) = \mathbb{E}\left[e^{\lambda \Delta_i(x)}\right].$$

   [2 points]

   (c) Show that

   $$\widehat{F}_n(x) - F(x) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i(x).$$

   [1 point]

2. **An MGF bound for a centered Bernoulli random variable.** [5 points]

   Let $Y \sim \text{Bernoulli}(p)$. Using its MGF, show that, for every $\lambda \in \mathbb{R}$,

   $$\mathbb{E}\left[e^{\lambda(Y-p)}\right] \leq \exp\left(\frac{\lambda^2}{8}\right).$$

   You may use the fact that if

   $$g(\lambda) = \log \mathbb{E}\left[e^{\lambda(Y-p)}\right],$$

   then

   $$g(0) = g'(0) = 0 \qquad \text{and} \qquad g''(\lambda) \leq \frac{1}{4}.$$

3. **A concentration inequality at a fixed point.** [6 points]

Fix $x \in [0,1]$ and $\varepsilon > 0$.

(a) Using Markov's inequality and the independence of $\Delta_1(x), \dots, \Delta_n(x)$, show that, for every $\lambda > 0$,

$$\Pr\left(\widehat{F}_n(x) - F(x) > \varepsilon\right) \le \exp\left(-\lambda n\varepsilon + \frac{n\lambda^2}{8}\right).$$

[2 points]

Hint: Apply the exponential transformation to $\widehat{F}_n(x) - F(x)$ and then use Markov's inequality.

(b) Optimize the bound over $\lambda > 0$ to prove that

$$\Pr\left(\widehat{F}_n(x) - F(x) > \varepsilon\right) \le e^{-2n\varepsilon^2}.$$

[2 points]

(c) Apply the same argument to the lower tail and conclude that

$$\Pr\left(\left|\widehat{F}_n(x) - F(x)\right| > \varepsilon\right) \le 2e^{-2n\varepsilon^2}.$$

[2 points]

4. **Discretizing the CDF.** [3 points]

Fix an integer $m \ge 1$. Since F is continuous and supported on $[0,1]$, show that there exist points

$$0 = x_0 \le x_1 \le \cdots \le x_m = 1$$

such that

$$F(x_j) = \frac{j}{m}, \qquad j = 0, \dots, m.$$

For $j = 1, \dots, m-1$, one possible choice is

$$x_j = \inf\left\{x \in [0,1] : F(x) \ge \frac{j}{m}\right\},$$

with $x_0 = 0$ and $x_m = 1$.

Explain where the continuity of F is used.

5. **From the grid to the entire interval.** [5 points]

Let $x \in [x_{j-1}, x_j]$. Use the monotonicity of F and $\widehat{F}_n$ to show that

$$\widehat{F}_n(x_{j-1}) \le \widehat{F}_n(x) \le \widehat{F}_n(x_j)$$

and

$$F(x_{j-1}) \le F(x) \le F(x_j).$$

Deduce that

$$\left|\widehat{F}_n(x) - F(x)\right| \le \max\left\{\left|\widehat{F}_n(x_{j-1}) - F(x_{j-1})\right|, \left|\widehat{F}_n(x_j) - F(x_j)\right|\right\} + \frac{1}{m}.$$

Conclude that

$$\sup_{x\in[0,1]}\left|\widehat{F}_n(x) - F(x)\right| \le \max_{0\le j\le m}\left|\widehat{F}_n(x_j) - F(x_j)\right| + \frac{1}{m}.$$

6. **Uniform concentration.** [4 points]

Show that, for every $\varepsilon > 1/m$,

$$\Pr\left(\sup_{x\in[0,1]}\left|\widehat{F}_n(x) - F(x)\right| > \varepsilon\right) \le 2(m+1)\exp\left[-2n\left(\varepsilon - \frac{1}{m}\right)^2\right].$$

Hint: Use the union bound over the grid points $x_0, \ldots, x_m$.

7. **Uniform convergence in probability.** [3 points]

For any fixed $\varepsilon > 0$, choose m sufficiently large so that

$$\frac{1}{m} < \frac{\varepsilon}{2}.$$

Show that

$$\Pr\left(\sup_{x\in[0,1]}\left|\widehat{F}_n(x) - F(x)\right| > \varepsilon\right) \le 2(m+1)e^{-n\varepsilon^2/2} \to 0 \text{ as } n \to \infty.$$

Remark 1: The above leads to *weak consistency* of $\widehat{F}_n$ (We will learn this in later lectures.)

$$\boxed{\sup_{x\in[0,1]}\left|\widehat{F}_n(x) - F(x)\right| \xrightarrow{p} 0.}$$

Remark 2: almost-sure uniform convergence. [No points]

One can also use the bound from the previous remark and the first Borel–Cantelli lemma to prove that

$$\sup_{x\in[0,1]}\left|\widehat{F}_n(x) - F(x)\right| \xrightarrow{\text{a.s.}} 0.$$

This is a special case of the Glivenko–Cantelli theorem.

2 Practice with Sufficiency, Completeness, and UMVUEs (10 points)

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta), \qquad \theta > 0,$$

where $n \geq 2$. Define

$$T = X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

The goal of this problem is to show that T is a complete sufficient statistic and then use Rao–Blackwellization to construct the UMVUE of θ.

1. **Sufficiency and the distribution of T.** [5 points]

   (a) Write down the joint density of $X_1, \ldots, X_n$, and use the Neyman–Fisher factorization theorem to show that T is sufficient for θ. [2 points]

   (b) Find the CDF and density of T. [2 points]

   (c) Show that

   $$\mathbb{E}_\theta[T] = \frac{n}{n+1}\theta.$$

   [1 point]

2. **Completeness of T.** [5 points]

   Let $g : (0, \infty) \to \mathbb{R}$ be a Borel-measurable function such that

   $$\mathbb{E}_\theta[|g(T)|] < \infty \qquad \text{and} \qquad \mathbb{E}_\theta[g(T)] = 0$$

   for every $\theta > 0$.

   (a) Using the density of T, show that

   $$\int_0^\theta g(t) n t^{n-1} \, dt = 0$$

   for every $\theta > 0$. [2 points]

   (b) Define

   $$h(t) = g(t) n t^{n-1}.$$

   Show that $h \in L^1_{\text{loc}}((0, \infty))$, i.e., for any compact set $K \subseteq (0, \infty)$ we have

   $$\int_K |h(t)| \, dt < \infty.$$

   Also, show that

   $$\int_a^b h(t) \, dt = 0$$

   for every $0 < a < b < \infty$. [1 point]

(c) Use the lemma below to show that $g(t) = 0$ for Lebesgue-almost every $t > 0$. Deduce that

$$\Pr_\theta(g(T) = 0) = 1$$

for every $\theta > 0$, and conclude that T is complete. Using this construct a UMVUE of θ. [2 points]

Hint: You may use the following standard result: If $h \in L^1_{\mathrm{loc}}((0,\infty))$ and

$$\int_a^b h(t)\,dt = 0$$

for every $0 < a < b < \infty$, then $h = 0$ Lebesgue-almost everywhere.

3 Admissibility and the James–Stein Estimator (15 points)

Let

$$X \sim N_d(\theta, I_d), \qquad \theta \in \mathbb{R}^d,$$

and consider estimating θ under squared-error loss

$$L(\theta, \delta) = \|\delta - \theta\|^2.$$

The risk of an estimator δ is

$$R(\theta, \delta) = \mathbb{E}_\theta \left[\|\delta(X) - \theta\|^2 \right].$$

The usual estimator is

$$\delta_0(X) = X.$$

1. **Risk, domination, and admissibility.** [2 points]

   (a) Show that

   $$R(\theta, \delta_0) = d$$

   for every $\theta \in \mathbb{R}^d$. [1 point]

   (b) State the definitions of domination and admissibility. Explain what must be shown in order to prove that δ_0 is inadmissible. [1 point]

2. **Stein's identity.** [3 points]

   Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be written as

   $$g(x) = \big(g_1(x), \dots, g_d(x)\big).$$

   Under suitable integrability and differentiability conditions, show, using integration by parts, that

   $$\mathbb{E}_\theta \left[(X - \theta)^\top g(X) \right] = \mathbb{E}_\theta \left[\operatorname{div} g(X) \right],$$

   where

   $$\operatorname{div} g(x) = \sum_{j=1}^{d} \frac{\partial g_j(x)}{\partial x_j}.$$

   Hint: First prove the coordinate-wise identity

   $$\mathbb{E}_\theta \left[(X_j - \theta_j) g_j(X) \right] = \mathbb{E}_\theta \left[\frac{\partial g_j(X)}{\partial x_j} \right].$$

3. **Risk of a perturbed estimator.** [3 points]

   Consider an estimator of the form

   $$\delta_g(X) = X + g(X).$$

Use Stein's identity to show that

$$R(\theta, \delta_g) = d + \mathbb{E}_\theta \left[\|g(X)\|^2 + 2 \operatorname{div} g(X) \right].$$

This expression is known as Stein's unbiased risk identity.

4. **A family of shrinkage estimators.** [4 points]

For $a > 0$, consider

$$\delta_a(X) = \left(1 - \frac{a}{\|X\|^2}\right) X.$$

Thus,

$$g_a(x) = -\frac{a}{\|x\|^2} x.$$

(a) Show that

$$\|g_a(x)\|^2 = \frac{a^2}{\|x\|^2}.$$

[1 point]

(b) Show that

$$\operatorname{div} g_a(x) = -\frac{a(d-2)}{\|x\|^2}.$$

[1 point]

(c) Deduce that

$$R(\theta, \delta_a) = d + \left[a^2 - 2a(d-2)\right] \mathbb{E}_\theta \left[\frac{1}{\|X\|^2}\right].$$

[2 points]

5. **The James–Stein estimator and inadmissibility.** [3 points]

Assume $d \geq 3$.

(a) Determine all values of $a > 0$ for which

$$R(\theta, \delta_a) < R(\theta, \delta_0)$$

for every $\theta \in \mathbb{R}^d$. [1 point]

(b) Find the value of a that minimizes

$$a^2 - 2a(d-2),$$

and hence obtain the James–Stein estimator

$$\boxed{\delta_{\mathrm{JS}}(X) = \left(1 - \frac{d-2}{\|X\|^2}\right) X.}$$

[1 point]

(c) Conclude that X is inadmissible under squared-error loss whenever $d \geq 3$. [1 point]

Conceptual remarks.

- The domination result holds even though every coordinate X_j is individually an admissible estimator of θ_j in the corresponding one-dimensional problem.
- The proof requires $d \geq 3$, in part because

$$\mathbb{E}_\theta \left[\frac{1}{\|X\|^2} \right] < \infty$$

only when $d \geq 3$.

- The argument above does not prove that the James–Stein estimator is itself admissible. In fact, it is dominated by the positive-part James–Stein estimator

$$\delta^+_{\mathrm{JS}}(X) = \left(1 - \frac{d-2}{\|X\|^2}\right)_+ X.$$

4 Two Routes to UMP Tests: Discreteness and Parameter-Dependent Support (15 points)

This problem studies two settings in which UMP tests arise for different reasons. The Poisson model illustrates the role of randomization in discrete models, while the uniform model illustrates hypothesis testing when the support of the distribution depends on the unknown parameter.

1. **A randomized UMP test in the Poisson model.** [7 points]

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda), \qquad \lambda > 0,$$

and consider testing

$$H_0 : \lambda \leq \lambda_0 \qquad \text{versus} \qquad H_1 : \lambda > \lambda_0.$$

Define

$$S = \sum_{i=1}^{n} X_i.$$

(a) Find the distribution of S. Show that the family of distributions of S has a monotone likelihood ratio in S. [2 points]

(b) Let c be an integer satisfying

$$\Pr_{\lambda_0}(S > c) \leq \alpha \leq \Pr_{\lambda_0}(S \geq c).$$

Find $\gamma \in [0,1]$ such that the randomized test

$$\varphi_{\mathrm{P}}(S) = \begin{cases} 1, & S > c, \\ \gamma, & S = c, \\ 0, & S < c, \end{cases}$$

has size exactly α. [2 points]

(c) Show that

$$\sup_{\lambda \le \lambda_0} \mathbb{E}_\lambda[\varphi_{\mathrm{P}}(S)] = \mathbb{E}_{\lambda_0}[\varphi_{\mathrm{P}}(S)] = \alpha.$$

Use the monotone likelihood ratio property to conclude that φ_{P} is a UMP level-α test. [2 points]

(d) Write the power function

$$\beta_{\mathrm{P}}(\lambda) = \mathbb{E}_\lambda[\varphi_{\mathrm{P}}(S)]$$

in terms of Poisson probabilities. [1 point]

2. **A UMP test in the uniform model.** [6 points]

Let

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, \theta), \qquad \theta > 0,$$

and consider testing

$$H_0 : \theta \le \theta_0 \qquad \text{versus} \qquad H_1 : \theta > \theta_0.$$

Define

$$T = Y_{(n)} = \max_{1 \le i \le n} Y_i.$$

(a) Find the distribution of T. Determine c_α such that the test

$$\varphi_{\mathrm{U}}(T) = \mathbb{1}\{T > c_\alpha\}$$

has size α. Show that

$$c_\alpha = \theta_0 (1 - \alpha)^{1/n}.$$

[2 points]

(b) Fix $\theta_1 > \theta_0$. Compare the joint likelihood under $\theta = \theta_1$ with the joint likelihood under $\theta = \theta_0$. Show that the likelihood ratio has the following structure:

$$\frac{f_{\theta_1}(y_1, \dots, y_n)}{f_{\theta_0}(y_1, \dots, y_n)} = \begin{cases} \left(\dfrac{\theta_0}{\theta_1}\right)^n, & T \le \theta_0, \\ +\infty, & \theta_0 < T \le \theta_1, \end{cases}$$

where the second line represents observations that are possible under θ_1 but impossible under θ_0. [2 points]

(c) Use the Neyman–Pearson lemma to show that φ_{U} is most powerful against every fixed $\theta_1 > \theta_0$. Conclude that it is UMP for the composite alternative $H_1 : \theta > \theta_0$.

Derive its power function. [2 points]

3. **Comparing the two UMP constructions.** [2 points]

Briefly answer the following questions.

(a) Why may randomization be necessary to obtain an exact level-α test in the Poisson model but not in the uniform model? [1 point]

(b) The Poisson UMP test follows from a monotone likelihood ratio, whereas the uniform model has parameter-dependent support. Explain how the Neyman–Pearson lemma nevertheless produces a UMP test in the uniform model. [1 point]

5 When Method of Moments Fails: M- and Z-Estimation for Cauchy Data (15 points)

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Cauchy}(\theta, 1), \qquad \theta \in \mathbb{R},$$

with density

$$f_\theta(x) = \frac{1}{\pi\{1 + (x - \theta)^2\}}.$$

The parameter θ is the location of the distribution.

1. **Failure of the ordinary method of moments.** [2 points]

   Show that $\mathbb{E}_\theta[X]$ does not exist. Explain why the usual first-moment equation

   $$\overline{X} = \mathbb{E}_\theta[X]$$

   cannot be used to define a method-of-moments estimator of θ.

2. **A bounded-moment Z-estimator.** [3 points]

   Although X does not have a finite mean, the random variable $\mathbb{1}\{X \le t\}$ is bounded.

   (a) Show that

   $$\mathbb{E}_\theta[\mathbb{1}\{X \le \theta\}] = \frac{1}{2}.$$

   [1 point]

   (b) Consider the estimating equation

   $$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{X_i \le t\} - \frac{1}{2}\right) = 0.$$

   Show that any solution is a sample median. Thus, the sample median is a Z-estimator based on a bounded moment condition. [2 points]

   Hint: For a CDF function F, the the median m is a defined as the quantity such that

   $$\lim_{x \to m-} F(x) \le 1/2 \le F(m).$$

3. **The median as an M-estimator.** [2 points]

   Show that every sample median minimizes

   $$Q_n(t) = \sum_{i=1}^{n} |X_i - t|.$$

   Conclude that the sample median is both an M-estimator and a Z-estimator.

Hint: Examine the left and right derivatives, or the sub-gradient, of $Q_n(t)$. You may use the fact that if t^* is a minimizer then $0 \in \partial Q_n(t)$. Go through the definition [Link] of subgradient and its properties [Link]. A helpful fact:

$$\partial |x| = \begin{cases} \{-1\}, & x < 0, \\ [-1, 1], & x = 0, \\ \{1\}, & x > 0. \end{cases}$$

4. **The maximum likelihood estimator.** [4 points]

   (a) Show that maximizing the likelihood is equivalent to minimizing

   $$L_n(t) = \sum_{i=1}^{n} \log\left\{1 + (X_i - t)^2\right\}.$$

   Hence, the Cauchy MLE is an M-estimator. [1 point]

   (b) Differentiate the objective and show that every interior local minimizer satisfies

   $$\sum_{i=1}^{n} \frac{X_i - t}{1 + (X_i - t)^2} = 0.$$

   Thus, the Cauchy MLE is also a Z-estimator. [1 point]

   (c) Show that the score contribution

   $$\psi(x, t) = \frac{x - t}{1 + (x - t)^2}$$

   is bounded as a function of x. Explain why an extremely large observation has limited effect on the estimating equation. [2 points]

5. **Nonconvexity and multiple roots.** [2 points]

   Compute

   $$\frac{\partial^2}{\partial t^2} \log\left\{1 + (x - t)^2\right\}$$

   and show that it can be negative for some values of $x - t$. Conclude that the Cauchy negative log-likelihood is not necessarily convex and that the likelihood equation may have multiple solutions.

   Explain why solving the score equation is not, by itself, sufficient to identify the global MLE.

6. **Sensitivity to a single outlier.** [2 points]

   Fix $X_1, \ldots, X_{n-1}$, and replace X_n by M, where $M \to \infty$. Compare the behavior of the following quantities:

   $$\overline{X}, \qquad \text{median}(X_1, \ldots, X_{n-1}, M), \qquad \frac{M - t}{1 + (M - t)^2}.$$

Explain what this comparison reveals about the robustness of the sample mean, sample median, and Cauchy likelihood estimator.