Homework 10 — Sampling Mean/Variance and Lebesgue–Stieltjes Integration

1) Sampling mean and variance — concepts and main distributional features

Setup. Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Key properties (hold under i.i.d., finite variance):

- **Unbiasedness:** $\mathbb{E}[\bar{X}] = \mu$, $\quad \mathbb{E}[S^2] = \sigma^2$.
- **Precision:** $\operatorname{Var}(\bar{X}) = \sigma^2/n$.
- **Consistency:** $\bar{X} \xrightarrow{p} \mu$ and $S^2 \xrightarrow{p} \sigma^2$ (LLN).
- **Asymptotic normality:** $\sqrt{n}(\bar{X} - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$ (CLT).

Normal-data refinements (when $X_i \sim \mathcal{N}(\mu, \sigma^2)$):

- $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
- $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.
- **Independence:** $\bar{X} \perp S^2$.
- **t–statistic:** $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- (Optional detail) $\operatorname{Var}(S^2) = \dfrac{2\sigma^4}{n-1}$.

2) Lebesgue–Stieltjes (L–S) integration: idea and links to probability/measure theory

Let F:R→R be nondecreasing, right-continuous, of bounded variation.
For measurable f, the **Lebesgue–Stieltjes integral** is

$$\int_a^b f(x)\,dF(x).$$

.

Measure-theoretic view. F induces a measure μ_F on Borel sets via $\mu_F([a,b))=F(b)-F(a)$. Then $\int f\,dF$ is the Lebesgue integral $\int f\,d\mu_F$.

Probability connection. If F is the CDF of X, then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,dF(x).$$

This **unifies** discrete, continuous, and mixed laws. If F has density p on a set and atoms at a_i with masses $\Delta F(a_i)=P(X=a_i)$, then

$$\int f\,dF = \int f(x)\,p(x)\,dx + \sum_i f(a_i)\,\Delta F(a_i).$$

3) Practice — numerical comparison: Lebesgue vs. (density-only) Riemann

Goal. Compute a mean/variance numerically as an L–S integral and contrast it with a "Riemann (density-only)" computation that ignores atoms.

Test distribution (mixed):
With probability 0.7: U(0,1); with probability 0.3: a point mass at 2.

Exact benchmarks (via L–S):

$$\mathbb{E}[X] = 0.7 \cdot \tfrac{1}{2} + 0.3 \cdot 2 = 0.95, \qquad \mathbb{E}[X^2] = 0.7 \cdot \tfrac{1}{3} + 0.3 \cdot 4 = 1.433333\ldots$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1.433333\ldots - 0.9025 = 0.530833\ldots$$

What the density-only Riemann approach would yield (incorrectly, by ignoring the atom):

$$\mathbb{E}_{\mathrm{dens}}[X] = \int_0^1 x \cdot 0.7\,dx = 0.35, \qquad \mathrm{Var}_{\mathrm{dens}}(X) = 0.7 \cdot \tfrac{1}{3} - 0.35^2 = 0.110833\ldots$$

Numerical plan for the code (to be attached):

1. Use trapezoidal/Simpson rule to approximate $\int_0^1 x\,(0.7)\,dx$ and $\int_0^1 x^2\,(0.7)\,dx.$

2. **L–S estimate:** add the atomic contributions 0.3. f(2)with f(x)=x and f(x)=x².
3. **Riemann (density-only) estimate:** omit the atom to show the undercount.

4. Optionally, simulate large samples from the mixture to confirm the L–S values.

What this demonstrates. The L–S integral naturally captures discrete mass $(\sum f(a_i)\Delta F(a_i))$, while a density-only Riemann calculation misses it—hence the large discrepancy above.