Homework assignment 4 – Mathematical Statistics

Because there is some R involved, students should hand in the answers via an assignment on canvas. Deadline is October 12 at 11 am.

Let $X_1, \ldots, X_n$ be i.i.d. with probability density function given by

$$f(x) = \begin{cases} \frac{4\theta^4}{x^5}, & \text{if } x \geq \theta \\ 0, & \text{if } x < \theta \end{cases} \qquad \text{where } \theta > 0 \text{ is an unknown parameter}$$

a) Determine the **moment estimator of θ**, based on the dataset $X_1, \ldots, X_n$, and determine its expectation (is it unbiased?) and its variance (expressed in θ and n).

b) Show that $\widehat{\theta} = \min(X_1, \ldots, X_n)$ is the **maximum likelihood estimator (MLE) of θ**. (Start by deriving the likelihood function and explicitly mention its domain).

c) Show that the MLE $\widehat{\theta}$ in part b. is an asymptotically unbiased estimator and a consistent estimator of θ.

d) Which approximate distribution does the moment estimator (in part a.) have for large n? Use this approximate distribution to construct an approximate confidence interval for θ, at level 1-α

To detect doping in professional sports, urine samples with ratio testosterone/epitestosterone > 4 are suspicious and are subject to additional testing. Suppose that this ratio for a "clean" professional athlete follows a normal distribution with mean μ and standard deviation σ.

e) Show that the probability that a clean athlete will have a suspicious ratio is

$$1 - \Phi\left(\tfrac{4-\mu}{\sigma}\right)$$

with Φ the cumulative distribution function (c.d.f.) of the standard normal distribution.

f) Given i.i.d. data $X_1, \ldots, X_n$ from clean professional athletes, provide the formula for the 95%-confidence interval of the mean ratio and the formula for the 90%-confidence interval of the standard deviation. For these formulas you do not need to insert the explicit expressions of the quantiles.

g) Read the dataset in the supplementary file athletes.csv into R. Estimate the probability in part (e) by replacing μ and σ by the corresponding MLE and compute the confidence intervals in part f. (The best would be if you print the R script and attach it to your homework.)

Grading:	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**Total**
	1	2	½ + 1½	½ +1	1	1	1+1/2	**10**