1. Let $f(x;\theta), \theta \in \Theta \subset \mathbb{R}^d$ be a family of density functions, and consider the sequence of moments

$$m_{\theta,k} = E_\theta[X^k] \quad \text{for } k = 1, 2, \ldots$$

when these exist, and where E_θ indicates expectation with respect to $f(x;\theta)$. Let $X_1, \ldots, X_n$ be an i.i.d. sample from $f(x;\theta)$ and

$$\widehat{m_k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

the sample k^{th} moment. It is desired to estimate some function ψ of θ, which may equal θ itself. We say $\widehat{\psi}$ is a moment estimator of ψ of order k when $m_{\theta,1}, \ldots, m_{\theta,k}$ exist and for some function g,

$$\widehat{\psi} = g(\widehat{m_1}, \ldots, \widehat{m_k}) \quad \text{when} \quad \psi = g(m_{\theta,1}, \ldots, m_{\theta,k}). \tag{1}$$

a. Find a moment estimator of order 2 for the variance of the distribution $f(x;\theta)$, assuming it exists.

b. Let $f_1(x), \ldots, f_d(x)$ be known density functions and

$$f(x;\theta) = \sum_{i=1}^{d} \theta_i f_i(x)$$

where $\theta \in \Theta$ with

$$\Theta = \left\{ \theta \in \mathbb{R}^d; \sum_{i=1}^{d} \theta_i = 1, \theta_i \geq 0 \right\}.$$

Find a moment estimator for θ, and write what moment assumptions are needed. It is not necessary to write the function g in (1) explicitly, but explain how to calculate the moment estimator for θ.

c. Prove that if $\widehat{\psi}$ is a moment estimator of ψ of order k of the form (1) where g is continuous then $\widehat{\psi}$ is consistent for ψ.

d. Assuming sufficient smoothness on the function g, identify the asymptotic distribution of $\widehat{\psi}$, after properly centering and scaling to assure the limit is non trivial. (No need to prove the distributional convergence.)

2. (a) Let $a_1, \ldots, a_n$ be distinct real numbers and let X have the discrete uniform distribution on the set $\{a_1, \ldots, a_n\}$, i.e.,

$$P(X = a_i) = 1/n \quad \text{for all} \quad i = 1, \ldots, n.$$

Find

i. $E(X)$

ii. $\text{Var}(X)$

iii. a median $\text{med}(X)$ of X, i.e., a number minimizing $E|X - \text{med}(X)|$, which is not necessarily unique.

(b) A scientist has a data set $x_1, \ldots, x_{10}$ of distinct real numbers such that $\sum_{i=1}^{10}(x_i - \overline{x})^2 = 110$, where $\overline{x} = (x_1 + \ldots + x_{10})/10$ is the sample mean.

i. The scientist claims that one of the data points exceeds the sample mean by 11 or more. Could this be true? Why or why not?

ii. You further learn that all the data points are positive and $\overline{x} = 1.4$. The scientist claims that two of the data points exceed the sample mean by 7 or more. Could this be true? Why or why not?

iii. Finally the scientist claims that 5 is a sample median m of the data set. Could this be true? Why or why not? *Hint:* Use Jensen's inequality to bound $|\overline{x} - m|$.