Problem 1. Let $\widehat{\theta}_n$ be an MLE based on n observations such that $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}(\theta))$ as $n \to \infty$ whatever the true value of $\theta \in \Theta$. Here $I(\theta)$ denotes the Fisher information in the underlying model. Now consider an arbitrary value $\eta \in \Theta$ and define

$$\widetilde{\theta}_n = \begin{cases} \eta, & \text{if } \ n^{1/4}\,|\widehat{\theta}_n - \eta| \le 1, \\ \widehat{\theta}_n, & \text{if } \ n^{1/4}\,|\widehat{\theta}_n - \eta| > 1. \end{cases}$$

Show that

$$\sqrt{n}\,(\widetilde{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \begin{cases} \mathcal{N}(0, I^{-1}(\theta)), & \text{if } \ \theta \ne \eta, \\ \mathcal{N}(0,0) \overset{\text{a.s.}}{=} 0, & \text{if } \ \theta = \eta, \end{cases}$$

i.e. $\widetilde{\theta}_n$ is 'super-efficient'.

Problem 2. For a statistical model $\{f(x|\theta);\, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$ open, let $\lambda(x|\theta) = \log f(x|\theta)$ and let $I(\theta)$ denote the Fisher information matrix in this model (in one observation). For a random sample $X_1, \ldots, X_n$ from $f(x|\theta_0)$ let $\widehat{\theta}_n$ denote the MLE. For a twice differentiable function $g : \Theta \to \mathbb{R}$ let $\ddot{g}$ denote its Hessian matrix. Suppose that

(i) $\widehat{\theta}_n$ is a consistent estimator of θ_0;

(ii) $I(\theta_0) = -\mathrm{E}_{\theta_0}\big[\ddot{\lambda}(X_1|\theta_0)\big]$ and $I(\theta_0)$ is positive definite;

(iii) $\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\lambda(x|\theta)$ is continuous at θ_0 for all x and $j,k = 1,\ldots,d$;

(iv) There exist functions $K_{j,k}(x)$, $j,k = 1,\ldots,d$ with $\mathrm{E}_{\theta_0}|K_{j,k}(X_1)| < \infty$ and

$$\big|U_{j,k}(x|\theta)\big| := \Big|\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\lambda(x|\theta)\Big| \le K_{j,k}(x) \qquad \text{for all } \ x \text{ and } \theta;$$

(v) For every $\epsilon > 0$ and all $j,k = 1,\ldots,d$ we have with $\mathcal{U}_{\theta_0}(\epsilon) = \{\theta : \|\theta - \theta_0\| \le \epsilon\}$ that

$$P_{\theta_0}\Big[\sup_{\theta \in \mathcal{U}_{\theta_0}(\epsilon)} \Big|\frac{1}{n}\sum_{i=1}^n \big[U_{j,k}(X_i|\theta) - \mathrm{E}_{\theta_0}U_{j,k}(X_i|\theta)\big]\Big| > \epsilon\Big] \to 0 \qquad \text{as } \ n \to \infty.$$

Show that then

a) $\quad -\frac{1}{n}\ddot{\lambda}_n(\mathbf{X}|\widehat{\theta}_n) \to I(\theta_0) \qquad$ in probability, as $n \to \infty$,

b) $\quad -n\,\big(\ddot{\lambda}_n(\mathbf{X}|\widehat{\theta}_n)\big)^{-1} \to I^{-1}(\theta_0) \qquad$ in probability, as $n \to \infty$, where $\mathbf{X} = (X_1, \ldots, X)$ and $\lambda_n(\mathbf{X}|\theta)$ denotes the log-likelihood function. As for convergence in probability of matrices we use the Frobenius norm $\|\cdot\|_F$ which for a matrix $A = (a_{ij})$ is defined as $\|A\|_F^2 = \mathrm{tr}(A'A) = \sum_{i,j} a_{ij}^2$, i.e. a sequence of $(d \times d)$-matrices A_n with random entries converges in probability to a (non-random) $(d \times d)$-matrix A iff for every $\epsilon > 0$ we have that $P(\|A_n - A\|_F > \epsilon) \to 0$ as $n \to \infty$.

Problem 3. **(a)** *(Multivariate mean value theorem)* Let $f : \mathbb{R}^d \to \mathbb{R}^k$ be differentiable derivative $\dot{f}(x)$, and assume that $\dot{f}(x)$ is continuous in the neighborhood $\mathcal{U}_{x_0}(r) = \{x \in \mathbb{R}^d : \|x - x_0\| < r\}$. Then the multivariate mean value theorem states that for $x \in \mathcal{U}_{x_0}(r)$ we have

$$f(x) = f(x_0) + \int_0^1 \dot{f}\big(x_0 + t\,(x - x_0)\big)\,dt\;(x - x_0).$$

Verify this equality.

(b) *(Multivariate second order Taylor expansion)* Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice differentiable with Hessian matrix $\ddot{f}(x)$ continuous in $\mathcal{U}_{x_0}(r) = \{x \in \mathbb{R}^d : \|x - x_0\| < r\}$ for some $r > 0$. Then, for $x \in \mathcal{U}_{x_0}(r)$ we have

$$f(x) = f(x_0) + \dot{f}(x_0)(x - x_0) + (x - x_0)' \int_0^1 \int_0^1 s\,\ddot{f}\big(x_0 + st\,(x - x_0)\big)\,ds\,dt\;(x - x_0).$$

(c) Show that for $d = 1$ the formula in (b) reduces to

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} (x - t) f''(t)\,dt.$$

Remark. A generalization of the formula given in (c): If f is $(k+1)$-times differentiable with $f^{(k+1)}(x)$ continuous in an open interval around x_0, then

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \tfrac{1}{k!} f^{(k)}(x_0)(x - x_0)^k + \tfrac{1}{k!} \int_{x_0}^{x} (x - t)^k f^{(k+1)}(t)\,dt.$$

Problem 4. Let $X_1, X_2, \ldots$ be iid from a mixture of Gamma distributions with pdf

$$f(x|\theta) = \Big[(1 - \theta)\,e^{-x} + \theta\,x\,e^{-x}\Big]\;\mathbf{1}\{x > 0\},$$

where $0 < \theta < 1$.

a) Find the method of moment estimator $\tilde{\theta}_n$ of θ and discuss the estimator (i.e. comment on whether it is a good estimator, and justify your assessments).

b) Derive the asymptotic distribution of $\tilde{\theta}_n$.

c) Find an asymptotically efficient estimate by improving $\tilde{\theta}_n$.

Problem 5. Consider a statistical model $\{f(x|\theta);\, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$ open. Let $\widetilde{\theta}_n \in \mathbb{R}^d$ be an estimator of θ with $\widetilde{\theta}_n - \theta_0 = O_P(n^{-1/2})$. On top of the assumptions of the theorem on the asymptotic normality of the MLE (see below) assume that the function $\theta \to I(\theta),\ \theta \in \Theta$, is continuous in a neighborhood of θ_0. Show that the method of scoring estimator

$$\widetilde{\theta}_n^{(1)} = \widetilde{\theta}_n + \frac{1}{n} I(\widetilde{\theta}_n)^{-1} \dot{\lambda}_n(\mathbf{X}|\widetilde{\theta}_n)$$

is asymptotically equivalent to the MLE $\widehat{\theta}_n$, in the sense that $\widetilde{\theta}_n - \widehat{\theta}_n = o_P(n^{-1/2})$ (implying that the method of scoring estimator $\widetilde{\theta}_n$ has the same asymptotic distribution as the MLE, i.e. it is asymptotically efficient). Here $\lambda_n(\mathbf{X}|\theta)$ is as in problem 2.

Using the notation from problem 2, the assumptions underlying the theorem on the asymptotic normality of the MLE (as discussed in STA231B) are as follows:

(i) The parameter space $\Theta \subset \mathbb{R}^d$ is open,

(ii) Second partial derivatives of $f(x|\theta)$ with respect to θ exist and are continuous for all x, and may be passed under the integral sign in $\int f(x|\theta)\, dx$,

(iii) There exists a function $K(x)$ such that $\mathrm{E}_{\theta_0} K(X_1) < \infty$ and

$$\sup_{\theta \in \mathcal{U}_{\theta_0}(r)} \left| \frac{\partial^2}{\partial\theta_j \partial\theta_k} \lambda(x|\theta) \right| \le K(x)$$

(iv) $I(\theta_0) = -\mathrm{E}_{\theta_0}\big[\ddot{\lambda}(X_1|\theta_0)\big]$ and $I(\theta_0)$ is positive definite,

(v) $\{x :\ f(x|\theta) \neq f(x|\theta_0)\}$ has Lebesgue measure $0 \;\;\Rightarrow\;\; \theta = \theta_0$