Sufficient statistics (TPE 1.6.32, 1.6.33))

(a) Consider two statistical models (i.e. two classes of distributions) $\mathcal{P}_0, \mathcal{P}_1$ on the same sample space $\mathcal{X}$, such that $\mathcal{P}_0 \subseteq \mathcal{P}_1$. Let $T$ be a sufficient statistics for $\mathcal{P}_1$. Show that it is sufficient for $\mathcal{P}_0$ as well.

(b) Continuing from the previous point, assume that, for any $\mathcal{P} \in \mathcal{P}_1$ there exists $\mathcal{P}' \in \mathcal{P}_0$ such that, for any $N \subseteq \mathcal{X}$ with $\mathcal{P}'(N) = 0$, we have $\mathcal{P}(N) = 0$. Show that, if $T$ is sufficient for $\mathcal{P}_0, \mathcal{P}_1$, and is complete for $\mathcal{P}_0$, then it is complete for $\mathcal{P}_1$.

(c) Let $\mathcal{P}$ be the family of distributions of $n$ i.i.d. random variables $X_1, \ldots, X_n$, with some common density $p(\cdot)$ on $\mathbb{R}$. (In other words $\mathcal{P} = \{p^{\otimes n} : p$ is a density on $\mathbb{R}\}$ is a class of probability distributions on $\mathbb{R}^n$.) Prove that the order statistics $X_{(1)}, \ldots, X_{(n)}$ is complete.

[Hint: you can use the submodel $\mathcal{P}_0$ formed by density of the form $\exp\{\theta_1 \sum_{i=1}^n x_i + \cdots + \theta_n \sum_{i=1}^n x_i^n - \sum_{i=1}^n x_i^2\}$. You can also use the fact that if $\sum_{i=1}^n a_i^k = \sum_{i=1}^n b_i^k$ for all $k \leq n$, then $(a_i)_{i \leq n}$ is a permutation of $(b_i)_{i \leq n}$.

(d) Continuing from the previous point, determine an UMVU estimator of $\mathbb{P}(X_1 \leq x) = \int_{-\infty}^x f(t) dt$ in the class $\mathcal{P}$.

Unbiased estimation from Binomials (TPE 2.1.17)

Let $\mathbb{P}_\theta = \text{Binom}(n, \theta)$, $\theta \in \Theta = [0, 1]$. We consider unbiased estimation of $g(\theta) = \theta^3$.

(a) Show that, for $n \leq 2$, no unbiased estimator exists. What happens for $n = 3$?

(b) Use the orthogonality condition to construct an UMVU estimator for $n > 3$.

Logistic regression

Consider a logistic regression model, where we are given i.i.d. pairs $(Y_i, X_i)$, $i \leq n$, with $X_i \in \mathbb{R}^d$ a feature vector, and $Y_i \in \{0, 1\}$ a label (or response variable). The distribution of the $X_i$ given by $p_{X_i}$, and the distribution of $Y_i$ given $X_i$ is given by

$$
\mathbb{P}_\theta(Y_i = y_i | X_i = x_i) = \frac{e^{y_i(\theta \cdot x_i)}}{1 + e^{(\theta \cdot x_i)}},
$$

(1)
(a) Derive an expression for the Fisher information matrix $I_F(\theta)$.

(b) Assume $p_X \sim N(0, I_d)$. Show that

$$I_F(\theta) = c_0 I_d + c_1 \theta \theta^T,$$

where $c_0 = c_0(\|\theta\|)$ and $c_1 = c_1(\|\theta\|_2)$ are scalars that depend on the norm of $\theta$. Provide expressions for $c_0$, $c_1$ in terms of one-dimensional integrals.

(c) Generalize the previous formula for $p_X \sim N(0, \Sigma)$.

[Hint: In solving this problems, it might be useful to remember Gaussian integration by parts. If $X \sim N(0, \Sigma)$ takes values in $\mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable and such that the expectations below make sense, then

$$E\{X_i f(X)\} = \sum_{j=1}^{d} \Sigma_{ij} E\left\{\frac{\partial f}{\partial X_j}(X)\right\}.$$  

]