TAMS42 (Probability and Statistics) Vinjett 2
—optimization techniques appearing in statistics

Background: It is a basic question in Calculus to investigate the maximal value of a real-valued smooth function $L(\theta)$ defined on some domain $D \subseteq \mathbb{R}$ (here I chose to use θ as the variable for the purpose of notational consistency in statistics). It should be noted that the maximal value of $L(\theta)$ heavily depends on the domain D. A very simply example is:

$$\max_{\{\theta:\theta\in D=[1,2]\}} -\theta^2 = \left[-\theta^2\right]_{\theta=1} = -1, \qquad \max_{\{\theta:\theta\in D=[-1,1]\}} -\theta^2 = \left[-\theta^2\right]_{\theta=0} = 0.$$

Discussion 1: For a function $L(\theta)$ with domain D, what are the usual steps to find the maximal $L(\theta_0)$?

In statistics, for the topic Maximal-Likelihood method, one usually considers the so-called likelihood function $L(\theta)$ in the form

$$L(\theta) = f_1(\theta) \cdot f_2(\theta) \cdot \ldots \cdot f_n(\theta),$$

where each $f_k(\theta)$ is a function of θ defined on a domain $\theta \in D_k, k = 1, 2 \ldots, n$. Therefore the domain of $L(\theta)$ is D which is given as

$$\theta \in D = \bigcap_{k=1}^{n} D_k.$$

To find the maximal value of $L(\theta)$ for $\theta \in D$ is sometimes a little troublesome, which is (I believe) of course related to optimizations

Discussion 2: When a function $L(\theta)$ is given as a product of several functions (see above), what possible difficulties may appear when maximizing $L(\theta)$? Are there good ways to avoid such difficulties? (such as using transforms $e^{L(\theta)}$ or $\ln(L(\theta))$ or ?).

Two specific tasks:

(1). Suppose that a likelihood function $L(\theta)$ is defined as follows

$$L(\theta) = f_1(\theta) \cdot f_2(\theta) \cdot \ldots \cdot f_n(\theta),$$

where for each $1 \leq k \leq n$, the function $f_k(\theta)$ is given as

$$f_k(\theta) = (\theta + 1) \cdot x_k^{\theta}, \text{ for some } 0 < x_k < 1 \text{ and } \theta > -1.$$

Find the point $\theta_0 > -1$ such that $L(\theta)$ reaches its maximal value at $\theta = \theta_0$.

(2). Suppose that a likelihood function $L(\lambda, \theta)$ (NOTE! two variables) is defined as follows

$$L(\lambda, \theta) = f_1(\lambda, \theta) \cdot f_2(\lambda, \theta) \cdot \ldots \cdot f_n(\lambda, \theta),$$

where for each $1 \leq k \leq n$, the function $f_k(\lambda, \theta)$ is given as

$$f_k(\lambda, \theta) = \lambda \cdot e^{-\lambda(x_k - \theta)}, \text{ for some } x_k \geq \theta \text{ and } \lambda > 0, \theta \in \mathbb{R}.$$

One more assumption is that there are at least two constants from $\{x_1, x_2, \ldots, x_n\}$ which are not the same. Find the point (λ_0, θ_0) such that $L(\lambda, \theta)$ reaches its maximal value at $(\lambda, \theta) = (\lambda_0, \theta_0)$.