ON THE EXTREMA OF INTEGRABLE FUNCTIONS

SLAWOMIR DOROSIEWICZ

ABSTRACT. This paper contains the definition of the extremum of integrable functions (e.g., the mode of density function). It seems to be a generalization of well-known standard definition and can be applied in estimation theory to extend the maximum likelihood method.

2000 Mathematics Subject Classification. 90C48.

1. Introduction. The notion of global (local) extremum is defined and well known for every function $f : X \to \mathbb{R}$ (local extremum requires topology on X). In some situations, for example, in mathematical statistics (especially in the estimation theory) it is necessary to extend this definition to include the case of extremum of integrable function. This paper provides the concept of such definition which can be applied to generalize the maximum likelihood method.

2. Extremum of integrable functions. Let (X, M, μ) be a measurable space with bounded, nonnegative, and nonatomic measure $\mu : M \to \mathbb{R}$. Without loss of generality, we assume that μ is a probabilistic measure (i.e., $\mu(X) = 1$). We assume that X is also a topological space with some topology $\mathcal{T}$ such that $(X, \mathcal{T})$ is a Hausdorff space. Let $\mathcal{L}(X, M, \mu)$ denote the set of μ-integrable functions $f : X \to \mathbb{R}$. We identify the functions which differ on the set with measure zero. This is equivalent to division of $\mathcal{L}(X, M, \mu)$ by the relation $\simeq$ defined by

$$f \simeq g \Longleftrightarrow \mu(\{x \in X : f(x) \neq g(x)\}) = 0. \tag{2.1}$$

Let $L^1(X, M, \mu) = \mathcal{L}(X, M, \mu) / \simeq$. For simplicity, the elements of $L^1(X, M, \mu)$ will also be called the integrable functions. Now we define the notion of extremum for this class of objects.

For a given function $f \in L^1(X, M, \mu)$ and a number $\epsilon > 0$, let $Z^+(f, \epsilon)$ (resp., $Z^-(f, \epsilon)$) denote the following optimization problem:

$$\int_S f \, d\mu \longrightarrow \max \ (\text{resp., min}), \tag{2.2}$$

subject to

$$S \in M; \quad \mu(S) = \epsilon. \tag{2.3}$$

DEFINITION 2.1. The function $f \in L^1(X, M, \mu)$ has maximum (resp., minimum) in the point $x^0 \in X$ if for each neighborhood U of x^0, there exists a sequence (ϵ_n) decreasing to zero and a sequence (S_n^*), such that S_n^* is an optimal solution of $Z^+(f, \epsilon_n)$

(resp., $Z^-(f, \epsilon_n)$), and for any $n \in \mathbb{N}$

$$\mu^*(U \cap S_n^*) > 0, \tag{2.4}$$

where μ^* denotes the external measure of sets (the measure μ^* is defined by the formula

$$\mu^*(A) = \inf \sum_n \mu(A_n), \tag{2.5}$$

where the infimum concerns for all finite and infinite sequences $A_1, A_2, \ldots$, such that $A \subset \bigcup_n A_n$ (cf. [2])).

First note that the definition is well defined in $L^1(X, M, \mu)$, that is, the equivalent functions $f \simeq g$ have the equal sets of extrema.

THEOREM 2.2. *If the integrable functions f, g are equal (μ-a.e.), then they have the same sets of maxima (also minima).*

THEOREM 2.3. *For any function $f \in L^1(X, M, \mu)$ the set of maxima (resp., minima) is closed.*

PROOF. Let $(x_\sigma)_{\sigma \in \Sigma}$ be a generalized sequence of maxima of $f \in L^1(X, M, \mu)$ convergent to a point $\bar{x} \in X$. For any neighborhood U of $\bar{x}$, there exists $\sigma_U \in \Sigma$ such that if $\sigma \geq \sigma_U$, then $t_\sigma \in U$. We consider any element $\sigma \geq \sigma_U$. By assumption, the function f has maximum at x_σ. There exists the sequence $(\epsilon)_{n \in \mathbb{N}}$ decreasing to zero, such that $\mu^*(U \cap \Omega_{\epsilon_n}^*) > 0$.

Because U is arbitrary then the axioms of Definition 2.1 are satisfied. Then f has maximum at $\bar{x}$. For the case of minimum the proof is similar. □

The optimal solutions of $Z^+(f, \epsilon)$ (resp., $Z^-(f, \epsilon)$) consists of the points where the values of f are sufficiently large (small). More precisely, the general optimality conditions for the problems $Z^\pm(f, \epsilon)$ (cf. [3]) imply the following lemma.

LEMMA 2.4. *The necessary and sufficient condition for Ω^* to be the optimal solution of $Z^+(f, \epsilon)$ is the existence of a number $\lambda^* \in \mathbb{R}$ satisfying the following conditions:*

$$x \in \Omega^* \Rightarrow f(x) \geq \lambda^*, \qquad x \notin \Omega^* \Rightarrow f(x) \leq \lambda^*, \qquad \mu(\Omega^*) = \epsilon. \tag{2.6}$$

In the case of the problem $Z^-(f, \epsilon)$ the inequalities should have the reverse sense.

THEOREM 2.5. *If M includes the family $\beta(X)$ of Borel subsets of $X, f : X \to \mathbb{R}$ is continuous and has at the point $x_0 \in X$ maximum (resp., minimum) with respect to Definition 2.1, then $f(x_0) = \sup f(X)$ (resp., $f(x_0) = \inf f(X)$).*

PROOF. Assume that $f(x_0) < \sup f(X)$. The set $V = f^{-1}(]f(x_0), \sup f(X)[)$ is nonempty and open, and thus $\lambda(V) > 0$. Fix the neighborhood U of x_0. Consider open and nonempty set $V_0 \subset V$ such that $V_0 \cap U = \varnothing$. The optimal solution Ω_ϵ^* of the problem $Z^+(f, \epsilon)$, where $\epsilon = \lambda(V_0)/2$, satisfies the condition $\lambda(\Omega_\epsilon^* \cap U) = 0$. If $\lambda(\Omega_\epsilon^* \cap U) > 0$, then taking the sets K, K_0 such that $K \subset U, K_0 \subset V_0, \lambda(K) = \lambda(K_0) > 0$, one can construct the set

$$\tilde{\Omega}_\epsilon^* = \Omega_\epsilon^* - (\Omega_\epsilon^* \cap U - K) \cup K_0, \tag{2.7}$$

which satisfies the conditions of $Z^+(f, \epsilon)$ and

$$\int_{\bar{\Omega}_\epsilon^*} f \, d\lambda > \int_{\Omega_\epsilon^*} f \, d\lambda, \tag{2.8}$$

which is in contradiction to the definition of Ω_ϵ^*. This completes the proof. $\qquad\square$

If the function f is continuous, the set of its global extrema includes the set of its extrema with respect to Definition 2.1. The inverse theorem does not hold (see Example 2.6 below).

Additionally, note that by the Weierstrass theorem, Lemma 2.4 imply that if X is compact, f continuous, then the set of f's extrema (with respect to Definition 2.1), is nonempty.

EXAMPLE 2.6. Define

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad \text{with } f(x) = \begin{cases} 0 & \text{if } x \le 0 \text{ or } x \ge 2, \\ x & \text{if } x \in \left]0, \frac{1}{2}\right[, \\ 1 - x & \text{if } x \in \left[\frac{1}{2}, 1\right[, \\ 1 & \text{if } x \in [1, 2]. \end{cases} \tag{2.9}$$

The function f has "ordinary" global maximum at each point of $\{1/2\} \cup [0,1]$, but the maximum with respect of Definition 2.1 is only reached at the points of $[0,1]$. The optimal solution Ω_ϵ^* of the problem (2.2) and (2.3) can be obtained from the following conditions:

$$x \in \Omega_\epsilon^* \Longrightarrow f(x) \ge \lambda_\epsilon^*, \qquad x \notin \Omega_\epsilon^* \Longrightarrow f(x) \le \lambda_\epsilon^*, \tag{2.10}$$

where λ_ϵ^* is chosen in order that the measure Ω_ϵ^* equals ϵ ($\epsilon > 0$). This implies that

(1) If $\epsilon \le 1$, then Ω_ϵ^* is the subset of $[1,2]$ with measure ϵ (more accurately $\mu(\Omega_\epsilon^*) = \epsilon$ and $\mu([1,2] - \Omega_\epsilon^*) = 0$).

(2) If $1 < \epsilon < 2$, then $\Omega_\epsilon^* = [1,2] \cup [1 - \epsilon/2, \epsilon/2]$ (with accuracy to a set of measure zero).

(3) If $\epsilon \ge 2$, then Ω_ϵ^* can be chosen as any set which contains $[0,2]$ and with the measure equal to ϵ (with accuracy to a set of measure zero, that is, $\mu([0,2] - \Omega_\epsilon^*) = 0$).

For sufficiently small ϵ ($\epsilon < 1$) there exist the neighborhoods of the point $1/2$ disjoint with Ω_ϵ^*. The function f does not have a maximum at the point $1/2$ in the sense of Definition 2.1.

EXAMPLE 2.7. We construct a function not equal to a constant on $[0,1]$ and integrable with respect to the Lebesgue measure λ, which has maximum at every point of this interval. Consider the functions ($\mathbb{Z}$ denotes the set of integer numbers)

$$f : \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}, \qquad f(x) = \sup_{k \in \mathbb{Z}} \phi(t - k), \tag{2.11}$$

where

$$\phi : \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}, \qquad \phi(x) = \begin{cases} \dfrac{1}{\sqrt{|x|}} & \text{if } x \ne 0, \\ +\infty & \text{if } x = 0. \end{cases} \tag{2.12}$$

The function f is finite in $]0,1[$ and integrable on $[0,1]$. Let $(\eta_k)_{k\in\mathbb{N}}$ be a sequence of positive numbers such that $\sum_{k\in\mathbb{N}}\eta_k$ converges. Consider the sequences of the functions $(f_k)_{k\in\mathbb{N}}$ and $(g_k)_{k\in\mathbb{N}}$ defined by

$$f_k : \mathbb{R} \longrightarrow \mathbb{R}\cup\{+\infty\}, \qquad f_k(x) = \eta_k f\left(2^{k-1}\left(\frac{x-1}{2^{k-1}}\right)\right),$$

$$g_k : \mathbb{R} \longrightarrow \mathbb{R}\cup\{+\infty\}, \qquad g_k = \sup(f,f_1,\ldots,f_k). \tag{2.13}$$

It is easy to calculate that for given $M > 0$ the value of Lebesgue measure of set

$$(f_k|_{[0,1]})^{-1}([M,+\infty[) \tag{2.14}$$

does not excess

$$\min\left(1, 2\frac{\eta_k^2}{M^2}\right). \tag{2.15}$$

Indeed, we have $\lambda(f_k|_{[0,1]})^{-1}([M,+\infty[) \leq 1$. Because f_k has the period $1/2^{k-1}$, then directly from definition of f_k we obtain

$$\lambda(f_k|_{[0,1]})^{-1}([M,+\infty[) \leq 2^{k-1}\lambda\left(\left[\frac{1}{2^{k-1}} - \frac{\eta_k^2}{M^2 2^{k-1}}, \frac{1}{2^{k-1}} + \frac{\eta_k^2}{M^2 2^{k-1}}\right]\right) = \frac{2\eta_k^2}{M^2}. \tag{2.16}$$

Denoting $A_n(M) = (g_n|_{[0,1]})^{-1}([M,+\infty[)$, we obtain

$$\lambda(A_n(M)) \leq \min\left(1, \frac{2}{M^2} + 4\sum_{k=1}^{n}\frac{\eta_k^2}{M^2}\right). \tag{2.17}$$

For every M and $n \in \mathbb{N}$ we have $A_n(M) \subset A_{n+1}(M)$, then

$$\lambda\left(\bigcup_{n\in\mathbb{N}} A_n(M)\right) = \lim_{n\to\infty}\lambda(A_n(M)). \tag{2.18}$$

It follows that

$$\lambda\left(\bigcup_{n\in\mathbb{N}} A_n(M)\right) \leq \min\left(1, \frac{2}{M^2}\left(1 + 2\sum_{n\in\mathbb{N}}\eta_n^2\right)\right), \tag{2.19}$$

and particularly

$$\lim_{M\to+\infty}\lambda\left(\bigcup_{n\in\mathbb{N}} A_n(M)\right) = 0. \tag{2.20}$$

We define $g(x) = \lim_{n\to+\infty}g_n(x)$ for $x \in [0,1]$. The limit $\lim_{n\to+\infty}g_n(x)$ exists and does not excess M if

$$x \in \bigcup_{m\in\mathbb{N}}\bigcap_{n>m}([0,1] - A_n(M)). \tag{2.21}$$

For each $n \in \mathbb{N}$ we have $A_n(M) \subset A_{n+1}(M)$, this implies that if $x \in [0,1] - \bigcup_{n\in\mathbb{N}} A_n(M)$, then $g(x) \leq M$. The values of g are thus finite on

$$\bigcup_{M>0}\left([0,1] - \bigcup_{n\in\mathbb{N}} A_n(M)\right) = \bigcup_{m\in\mathbb{N}}\bigcap_{n\in\mathbb{N}} A_n'\left(\frac{1}{m}\right). \tag{2.22}$$

Because

$$\lambda\left(\bigcup_{m\in\mathbb{N}}\bigcap_{n\in\mathbb{N}}A'_n\left(\frac{1}{m}\right)\right) = \lim_{m\to\infty}\lambda\left(\bigcap_{n\in\mathbb{N}}A'_n\left(\frac{1}{m}\right)\right),$$

$$\lambda\left(\bigcap_{n\in\mathbb{N}}A'_n\left(\frac{1}{m}\right)\right) = \lim_{n\to\infty}\lambda\left(A'_n\left(\frac{1}{m}\right)\right)$$

$$= \lim_{n\to\infty}\lambda\left([0,1]-\bigcup_{n\in\mathbb{N}}A'_n\left(\frac{1}{m}\right)\right)$$

$$= 1-\min\left(1,2m^2\left(1+\sum_{k\in\mathbb{N}}\eta_k^2\right)\right),$$

(2.23)

then

$$\lambda\left(\bigcup_{m\in\mathbb{N}}\bigcap_{n\in\mathbb{N}}A'_n\left(\frac{1}{m}\right)\right) = 1,$$

(2.24)

thus g is finite λ-a.e. on $[0,1]$.

Dividing, for fixed $k\in\mathbb{N}$, the interval $[0,1]$ on the family of components

$$\left[0,\frac{1}{2^k}\right],\left[\frac{1}{2^k},\frac{3}{2^k}\right],\dots,\left[1-\frac{1}{2^k},1\right],$$

(2.25)

it is easy to see that the following inequalities hold:

$$\int_0^1 f_k(x)\,dx \le C\eta_k,$$

$$\int_0^1 g_n(x)\,dx \le \int_0^1\left(f(x)+\sum_{i=1}^n f_i(x)\right)dx \le C\left(1+\sum_{i=1}^n \eta_i\right),$$

(2.26)

where $C \le \int_0^1 f(x)\,dx$. The convergence of the sum $\sum_i \eta_i$ implies that

$$\sup_{n\in\mathbb{N}}\int_0^1 g_n(x)\,dx < \infty,$$

(2.27)

and consequently, the integrability of g on $[0,1]$. Additionally, for every $x\in[0,1]$ and its neighborhood $U \ni x$, the set

$$U\cap\bigcup_{n\in\mathbb{N}}g_n^{-1}(]M,+\infty[)$$

(2.28)

has nonempty interior. This follows that g has the maximum in the sense of Definition 2.1 at every point of $[0,1]$.

Note the difference of this result from the notion of "ordinary" maximum: the function g has a maximum with respect to Definition 2.1 at every point of $[0,1]$, but g is not a constant, that is, $\lambda(\{x\in[0,1]:g(x)=c\}) \ne 1$ for any constant c.

Determining the extrema in the sense of Definition 2.1 can be applied in mathematical statistics to generalize maximum likelihood method (ML). In this method the

estimators are computed as the extrema of a special function, called likelihood function. In the case when this function is not sufficiently smooth (i.e., differentiable or at least continuous) the above definition of extrema can be useful to construct the generalization of ML method. We start this part of the paper by the presentation of the basic idea of this method.

Let $(\Omega, \mathcal{M}, \mathcal{P})$ be a statistical space, that is, $\Omega \neq \varnothing$—a given set, $\mathcal{M}$-σ-algebra of subsets of Ω, and a family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability distributions defined on $\mathcal{M}$. Suppose that

(1) $\Theta \subset \mathbb{R}^w$ for any $w \in \mathbb{N}$;
(2) the space $(\Omega, \mathcal{M}, \mathcal{P})$ is dominated, that is, there exists a σ-finite measure $\nu : \mathcal{M} \to \mathbb{R}$ such that every distribution from $\mathcal{P}$ is absolutely continuous with respect to ν.

The map

$$L : \Theta \times \Omega \longrightarrow \mathbb{R}, \qquad L(\theta, \omega) = \frac{dP_\theta}{d\nu}(\omega); \tag{2.29}$$

where $dP_\theta/d\nu$ denotes the Radon-Nikodym derivative of measure P_θ with respect to ν, is called likelihood function. "Classical" ML method requires determining the set of maxima of the function $L(\cdot, \omega)$ for every $\omega \in \Omega$. If the map, $(\arg_{\theta \in \Theta} \max L(\theta, \omega)$ denotes the set of elements $\theta \in \Theta$ such that the map $L(\cdot, \omega) : \Theta \to \mathbb{R}$ reaches the maximum).

$$\omega \longmapsto \arg_{\theta \in \Theta} \max L(\theta, \omega) \tag{2.30}$$

is measurable, then it determines the ML estimator of the parameter θ (equivalently probability distribution P_θ). If $L(\cdot, \omega)$ is integrable, but it does not have another properties (such as continuity, differentiability, etc.) it may be difficult to say about its extrema. Definition 2.1 may be useful and allows to generalize the ML method on mentioned class of likelihood functions. We precise the idea of this generalization.

Suppose that for every $\omega \in \Omega$ the function $L(\cdot, \omega)$ is integrable on Θ with respect to the Lebesgue measure. Determining the maxima of $L(\cdot, \omega)$ with respect to Definition 2.1 requires solving the problem $Z^+(L(\cdot, \omega), \epsilon)$ which has the following form:

$$\int_S L(\theta, \omega) \, d\lambda(\theta) \longrightarrow \max, \tag{2.31}$$

subject to

$$S \in \beta(\mathbb{R}^w), \quad \int_S d\lambda = \epsilon. \tag{2.32}$$

For $\omega \in \Omega$, we denote the set of maxima of $L(\cdot, \omega)$ (with respect to Definition 2.1) by $\mathcal{L}(\omega)$.

If the map (multivalued in general)

$$\Omega \ni \omega \longmapsto \mathcal{L}(\omega) \in \Theta \tag{2.33}$$

has at least one measurable selection (the μ-measurable function $\hat{\theta} : \Omega \to \mathbb{R}^w$, which for $\omega \in \Omega$ satisfies $\hat{\theta}(\omega) \in \mathcal{L}(\omega)$), then any one of them may be considered as a (generalized) ML estimator of the parameter θ.

It is well known that ML estimators have, under necessary assumptions, some properties as asymptotical consistence and normality (cf. [1, 4]). Theorem 2.5 allows to have the hope, that the estimators obtained from generalized ML method have often little stronger properties.

References

[1] J. Bartoszewicz, *Wyk/lady ze statystyki matematycznej* [*Lectures in Mathematical Statistics*], 2nd ed., Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1996 (Polish). CMP 1 624 321.

[2] P. Billingsley, *Prawdopodobieństwo i miara* [*Probability and Measure*], Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1987 (Polish). MR 89b:60001.

[3] S. Dorosiewicz, *Modelowanie sieci transportowych. Wybrane zagadnienia* [*Modelling of Transportation Networks. Selected Topics*], Monografie i Opracowania, vol. 437, Warsaw School of Economics, Warsaw, 1997 (Polish).

[4] R. Zieliński, *Siedem wyk/lad/ow wprowadzaj/acych do statystyki matematycznej* [*Intrduction to Mathematical Statistics*], Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1990 (Polish).

SLAWOMIR DOROSIEWICZ: WARSAW SCHOOL OF ECONOMICS, WARSAW, POLAND
E-mail address: doro@sgh.waw.pl