Technische Universität Ilmenau
Institut für Mathematik

Preprint No. M **11/12**

Universal confidence sets - estimation and relaxation

Silvia Vogel

2011

Impressum:
Hrsg.: Leiter des Instituts für Mathematik
Weimarer Straße 25
98693 Ilmenau
Tel.: +49 3677 69-3621
Fax: +49 3677 69-3270
http://www.tu-ilmenau.de/math/

ilmedia

Universal Confidence Sets -
Estimation and Relaxation

Silvia Vogel
Technische Universität Ilmenau

Abstract

In [9] we provided universal confidence sets for constraint sets, optimal values and solutions sets of deterministic decision problems. The results assume concentration-of-measure properties for the objective and/or constraint functions and some knowledge about the true problem, such as values of a growth function and a continuity function. If these values are not available, one can try to estimate them from the approximations for the true functions. We show how such estimates can be derived. Furthermore we investigate confidence sets which are obtained via relaxation of certain inequalities. These confidence sets can be derived without any knowledge about the true deterministic problem and yield, with a prescribed high probability, a superset of the true set. We consider such "superset-approximations" for the constraint sets and the solutions sets and discuss the question how their quality may be judged. Furthermore, lower and upper approximations for the optimal value are derived.

Keywords: universal confidence sets, estimates for the growth function, relaxation
MSC2000: 90C15, 90C31, 62F25, 62F30

1 Introduction

In [9] we considered a deterministic decision problem which is approximated by a sequence of random problems. Confidence sets for the constraint set, the optimal value, and the solution set are then derived as suitable neighborhoods of the corresponding sets and values of the approximate problem. The 'radii' of these neighborhoods are determined employing so-called outer (and inner) approximations in probability with convergence rate and tail behavior function. As these notions are quantified (semi-)versions of Kuratowski-Painlevé convergence in probability for sequences of random sets, in this paper we will call them (outer and inner) Kuratowski-Painlevé approximations, in short *KP-approximations*, in order to distinguish them from other approximations considered in the following. For Kuratowski-Painlevé convergence in the deterministic setting see [4], for the "in probability" setting c.f. [7] or [8].

In [9] outer and inner KP-approximations in probability are obtained assuming uniform concentration-of measure properties of the approximating constraint and/or objective functions and some knowledge about the true problem. Uniform concentration-of-measure results can partly be derived from concentration-of-measure assertions for sequences of random variables as provided e.g. in [2] (c.f. [3], [10]). Further results for approximate functions which rely on kernel estimates are given in [6] and forthcoming papers. In the present paper we consider the assumptons concerning knowledge about the true, but unknown model and discuss the question what can be done if this knowledge is not available. One way is to estimate needed parameters from the approximations and to incorporate these estimates into the formulae for the confidence sets. There is, however, also the simple possibility to consider so-called relaxed approximate problems. For instance, in order to derive a confidence set for the constraint set, which is defined via an inequality of the form "≤ 0", we claim instead "$\leq \beta_{n,\kappa}$" for the approximating set and obtain, with a certain probability, a suitable superset of the true set. Here $\beta_{n,\kappa}$ denotes the so-called "convergence rate" of the approximating functions, where n may be interpreted as the size of the underlying sample, and κ is related to the probability level. These approximations usually yield smaller confidence sets than those obtained via incorporation of estimates for growth functions etc.

Often it is desirable to have also inner approximations of the form "$\leq -\beta_{n,\kappa}$", e.g. in order to be able to judge the quality of the outer approximations. Obviously inner approximations of this kind are empty for the solution set. Then, again, inner KP-approximation come into play, and we need estimates for the parameters. Moreover, it may be reasonable in some cases to have a fixed 'radius' which can be attached to the solution set etc. of the approximate problem instead of a set defined by $\leq \beta_{n,\kappa}$.

In the following we will firstly consider the different ways for the derivation of confidence sets. In this framework we shall extend the relaxation approach. Furthermore, lower and upper approximations for the optimal value are derived.

In the second part of the paper we show how parameters for the true problem, needed for the KP-approximations, can be estimated under additional convexity conditions.

The paper is organized as follows. In section 2 we provide the mathematical model, explain how confidence sets may be derived from suitable approximations and discuss examples. In section 3 we consider inner and outer approximations obtained via relaxation. Section 4 deals with estimations of the growth functions and in section 5 the focus is on estimation of the continuity function.

2 Universal Confidence Sets

Let (E, d) be a complete separable metric space and $[\Omega, \Sigma, P]$ a complete probability space. We assume that a deterministic optimization problem

$$(P_0) \quad \min_{x\in\Gamma_0} f_0(x)$$

is approximated by a sequence of random problems

$$(P_n) \quad \min_{x\in\Gamma_n(\omega)} f_n(x,\omega), \quad n \in N.$$

Γ_0 is a nonempty closed subset of E, and the function f_0, which maps into the extended reals $\bar{R}^1 := R^1 \cup \{-\infty\} \cup \{+\infty\}$, is a lower semicontinuous function. For each $n \in N$ and $\kappa > 0$, $\Gamma_n|\Omega \to 2^E$ is a closed-valued measurable multifunction, and $f_n|E \times \Omega \to \bar{R}^1$ is a lower semicontinuous random function, which is supposed to be $(\mathcal{B}(E) \otimes \Sigma, \bar{\mathcal{B}}^1)$-measurable. $\mathcal{B}(E)$ denotes the Borel-σ-field of E and $\bar{\mathcal{B}}^1$ the σ-field which is generated by the Borel sigma field $\mathcal{B}^1$ of R^1 and $\{+\infty\}$, $\{-\infty\}$. Moreover, we assume that all functions are (almost surely) proper functions, i.e. functions with values in $(-\infty, +\infty]$ which are not identically ∞.

Furthermore, we consider constraint sets Γ which are defined by inequality constraints. Let Q_0 be a closed non-empty subset of E and $J = \{1, \ldots, j_M\}$ a finite index set. We consider functions $g_0^j|E \to R^1$, $j \in J$, which are lower semicontinuous in all points $x \in E$, and define $\Gamma_0 := \{x : g_0^j(x) \le 0, j \in J\} \cap Q_0$. Γ_0 is assumed to be non-empty.

The set Q_0 is approximated by a sequence $(Q_n)_{n\in N}$ of closed-valued measurable multifunctions, and the functions g_0^j, $j \in J$, are approximated by sequences $(g_n^j)_{n\in N}$ of functions $g_n^j|E \to R^1$, $j \in J$, which are $(\mathcal{B}(E) \otimes \Sigma, \mathcal{B}^1)$-measurable. Furthermore, we assume that the functions $g_n(\cdot, \omega)$ are lower semicontinuous for all $\omega \in \Omega$.

Eventually, the approximate constraint set Γ_n is defined by
$\Gamma_n(\omega) := \{x \in E : \ g_n^j(x,\omega) \leq 0, \ j \in J\} \cap Q_n(\omega)$. Under our assumptions Γ_n is a closed-valued measurable multifunction.

In contrast to [9] we will not deal with 'relaxed' objective and constraint functions $f_{n,\kappa}$ and $g_{n,\kappa}$, instead we consider approximating functions f_n and g_n which do not depend on κ. Dependence on κ, especially for the resulting constraint sets, could be taken into account, it would, however, require additional technical details and denotations.

The measurability conditions imposed here do not have the weakest form. We use them for sake of simplicity. They are satisfied in many applications and guarantee that all functions of ω needed in the following have the necessary measurability properties. Moreover, the lower semicontinuity assumption of the objective functions f_n can be dropped. Imposing this condition, however, we can omit some technical details in the proofs.

In the following, the optimal values are denoted by Φ.
$\Phi_n(\omega) := \inf\limits_{x \in \Gamma_n(\omega)} f_n(x,\omega)$ is the optimal value for the realization $(P_n(\omega))$ of the approximate problem, while $\Phi_0 := \inf\limits_{x \in \Gamma_0} f_0(x)$ is the optimal value to (P_0). $\Psi_n(\omega)$ and Ψ_0 denote the corresponding solution sets.

We will mostly deal with level sets of functions. They occur in constraint sets which are usually defined by inequality constraints. Furthermore, results about the constraint set can be immediately carried over to the solutions sets using $g := f - \Phi$ and $g \leq 0$, assuming that concentration-of-measure results for the objective functions and the optimal values are available.

In order to simplify notation, we consider one lower semicontinuous function constraint function g_0 and sequences $(g_n)_{n \in N}$ of functions $g_n | E \times \Omega \to \bar{R}^1$ which have the measurability properties explained above. g_0 can be interpreted as $g_0(x) := \sup\limits_{j \in J} g_0^j(x)$ or , later on, as $g_0(x) := f_0(x) - \Phi_0$. If constraints have to be taken into account, we can use
$g_0(x) := \max\{f_0(x) - \Phi_0, \sup\limits_{j \in J} g_0^j(x)\}$. Similarly,
$g_n(x,\omega) := \max\{f_n(x,\omega) - \Phi_n(\omega), \sup\limits_{j \in J} g_n^j(x,\omega)\}$ etc. We will explain the approaches for general sets M_0 and M_n which are defined by inequality constraints with respect to g_0 and g_n:

$$M_0 := \{x \in E : \ g_0(x) \leq 0\},$$
$$M_n(\omega) := \{x \in E : \ g_n(x,\omega) \leq 0\}.$$

Confidence sets for M_0 can be obtained in the following way. Assume that sequences $(M^{sup}_{n,\kappa})_{n\in N},\ \kappa > 0$, are available with the following property:

$$\forall \kappa > 0: \sup_{n\in N} P\{\omega:\ M_0 \setminus M^{sup}_{n,\kappa}(\omega) \neq \emptyset\} \leq \mathcal{H}(\kappa) \tag{1}$$

where $\lim_{\kappa\to\infty} \mathcal{H}(\kappa) = 0$. We can assume that the convergence is monotonous.Then, given a prescribed probability level ε_0, we can choose κ_0 such that $\mathcal{H}(\kappa_0) \leq \varepsilon_0$, and the sequence $(M_{n,\kappa_0})_{n\in N}$ yields for each $n \in N$ a confidence set, i.e. a set which covers the true set M_0 with the prescribed probability $1 - \varepsilon_0$. As this procedure works for each n, G. Ch. Pflug, who considered KP-approximations in [3], introduced the denotation 'universal confidence sets'. Of course in order to derive useful confidence sets, one would like to have approximating sequences that become smaller with each n, i.e. the distance (in a suitable measure) between M_0 and $M^{sup}_{n,\kappa}$ should tend to zero with increasing n for each $\kappa > 0$.

We use the following denotation.

Definition. A sequence $(M^{sup}_{n,\kappa})_{n\in N}$ which satisfies condition (1) is called *superset-approximation in probability with tail behavior function $\mathcal{H}$.*

Examples for superset-approximations in probability are outer Kuratowski-Painlevè approximations and (outer) relaxations as considered in [9]. Here we will classify the results in a unifying scheme of denotations and supplement them by corresponding 'inner' relaxations which can be utilized in order to judge the quality of a confidence set. Furthermore we suggest lower and upper approximations for the optimal values.

Definition. A sequence $(M^{sub}_{n,\kappa})_{n\in N}$ which satisfies the condition

$$\forall \kappa > 0: \sup_{n\in N} P\{\omega:\ M^{sub}_{n,\kappa}(\omega) \setminus M_0 \neq \emptyset\} \leq \mathcal{H}(\kappa) \tag{2}$$

is called *subset-approximation in probability with tail behavior function $\mathcal{H}$.*

Subset-approximations obtained via relaxation will be considered in section 3. Note that inner KP-approximations in general do not satisfy condition (2), because they only *in the limit* tend to be contained in M_0.

3 Relaxation

The results in [9] utilize knowledge about the true, but unknown model, namely certain values of the functions μ and sometimes also λ and ν. There are applications where good estimates for this values are available, e.g. constraint or objective functions with known analytic form and estimated parameters. In general, however, if only a nonparametric estimate for the functions under consideration is available, we have to estimate the values of μ etc. from the estimates for the functions. We will provide such estimates in the following sections.

For given sets M_0 and M_n, defined as in section 2 we introduce the following sets. Let for each $\alpha \in R$,

$$M_0^\alpha := \{x \in E : \; g_0(x) \leq \alpha\}$$

and

$$M_n^\alpha(\omega) := \{x \in E : \; g_n(x, \omega) \leq \alpha\}.$$

Obviously, for $\alpha > 0$ we obtain a superset of the set $M_n(\omega)$ and M_0, respectively, for $\alpha < 0$ a subset. As we will see, M_n^α yields for $\alpha > 0$ also a superset-approximation for M_0 and for $\alpha < 0$ a subset-approximation for M_0. Subset-approximations, however, require the existence of 'inner' points, an assumption which can not be satisfied for the solution set.

Lemma: (i) If there exist a function $\mathcal{H}_1$ and to all $\kappa > 0$ a sequence $(\beta_{n,\kappa}^{(1)})_{n \in N}$ such that
$\sup\limits_{n \in N} P\{\omega : \inf\limits_{x \in UM_0} (g_n(x, \omega) - g_0(x)) \leq -\beta_{n,\kappa}^{(1)}\} \leq \mathcal{H}_1(\kappa)$
for a suitable neighborhood UM_0,
then $(U^{-\beta_{n,\kappa}^{(1)}} M_n)_{n \in N}$ is a subset-approximation for M_0.
(ii) If there exist a function $\mathcal{H}_1$ and to all $\kappa > 0$ a sequence $(\beta_{n,\kappa}^{(1)})_{n \in N}$ such that
$\sup\limits_{n \in N} P\{\omega : \sup\limits_{x \in M_0} (g_n(x, \omega) - g_0(x)) \geq \beta_{n,\kappa}^{(1)}\} \leq \mathcal{H}_1(\kappa)$, then $(U^{\beta_{n,\kappa}^{(1)}} M_n)_{n \in N}$ is a superset-approximation for M_0.

Proof. (i) Let $\omega \in \Omega$, $n \in N$ and $\kappa > 0$ be such that $U^{-\beta_{n,\kappa}^{(1)}} M_n(\omega) \setminus M_0 \neq \emptyset$. Then there is an $x_n(\omega) \in U^{-\beta_{n,\kappa}^{(1)}} M_n(\omega)$ which does not belong to M_0. Hence $g_n(x_n(\omega), \omega) \leq -\beta_{n,\kappa}^{(1)}$ while $g_0(x_n(\omega)) > 0$ and we can employ the assumption.

(ii) Let $\omega \in \Omega$, $n \in N$ and $\kappa > 0$ such that $M_0 \setminus U^{\beta^{(1)}_{n,\kappa}} M_n(\omega) \neq \emptyset$. Then there is an $x_0(\omega) \in M_0$ which does not belong to $U^{\beta^{(1)}_{n,\kappa}} M_n(\omega)$. Hence $g_0(x_0(\omega)) \leq 0$ while $g_n(x_0(\omega)) \geq \beta^{(1)}_{n,\kappa}$. □

It follows from the results in [9] that for an 'outer Kuratowski-Painlevè approximation' $(M_n)_{n\in N}$ the sequence $(U_{\beta^{(1)}_{n,\kappa}} M_n)_{n\in N}$ is a superset-approximation. In the case of convex sets with nonempty interior we can also derive a subset-approximation in the following way: For a convex set M with nonempty interior we define the *α-interior* $U_{-\alpha}M$ by $U_\alpha(U_{-\alpha}M) = M$. Hence for an 'inner' Kuratowski-Painlevè approximation (M_n) of convex sets with nonempty interior the sequence $(U_{-\beta^{(1)}_{n,\kappa}} M_n)_{n\in N}$ is a subset-approximation.

Subset- an superset-approximations are useful tools for the assessment of the goodness of an approximation, because the true set lies in between the two sets with prescribed probability.

It is desirable to have also lower and upper approximations for the optimal value. Based on subset- and superset-approximations we define for nonempty constraint sets
$\Phi^{(u)}_{n,\kappa}(\omega) := \inf\limits_{x\in\Gamma^{sub}_{n,\kappa}(\omega)} f_n(x,\omega)$, $\Phi^{(l)}_{n,\kappa}(\omega) := \inf\limits_{x\in\Gamma^{sup}_{n,\kappa}(\omega)} f_n(x,\omega)$.
Then we have the following results:

Lemma. (i) Assume that a subset-approximation $(\Gamma^{sub}_{n,\kappa})_{n\in N}$ for Γ_0 with tail behavior function $\mathcal{H}_1$ is available and that there exist a function $\mathcal{H}_2$ and to all $\kappa > 0$ a sequence $(\beta^{(2)}_{n,\kappa})_{n\in N}$ such that
$\sup\limits_{n\in N} P\{\omega : \inf\limits_{x\in U\Gamma_0} (f_n(x,\omega) - f_0(x)) \leq -\beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_2(\kappa)$ for a suitable neighborhood $U\Gamma_0$.
Then $\sup\limits_{n\in N} P\{\omega : \Phi^{(u)}_{n,\kappa}(\omega) - \Phi_0 \leq -\beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_1(\kappa) + \mathcal{H}_2(\kappa)$.
(ii) Assume that a superset-approximation $(\Gamma^{sup}_{n,\kappa})_{n\in N}$ for Γ_0 with tail behavior function $\mathcal{H}_1$ is available and that there exist a function $\mathcal{H}_2$ and to all $\kappa > 0$ a sequence $(\beta^{(2)}_{n,\kappa})_{n\in N}$ such that
$\sup\limits_{n\in N} P\{\omega : \sup\limits_{x\in\Gamma_0} (f_n(x,\omega) - f_0(x)) \geq \beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_2(\kappa)$.
Then $\sup\limits_{n\in N} P\{\omega : \Phi^{(l)}_{n,\kappa}(\omega) - \Phi_0 \geq \beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_1(\kappa) + \mathcal{H}_2(\kappa)$.

Proof. (i) Let $\omega \in \Omega$, $n \in N$ and $\kappa > 0$ be such that $\Phi^{(u)}_{n,\kappa}(\omega) - \Phi_0 \leq -\beta^{(2)}_{n,\kappa}$. As the objective functions are supposed to be lower semicontinuous, we find $x_n(\omega) \in \Gamma^{sub}_{n,\kappa}(\omega)$ with $\Phi^{(u)}_{n,\kappa}(\omega) = f_n(x_n(\omega),\omega)$. Furthermore, there is $x_0 \in \Gamma_0$

with $\Phi_0 = f_0(x_0)$. If $x_n(\omega) \notin \Gamma_0$, we can employ the definition of the subset-approximation. If $x_n(\omega) \in \Gamma_0$, we have $f_0(x_n(\omega)) \geq f_0(x_0)$ and proceed as follows: The assumption implies $f_n(x_n(\omega), \omega) - f_0(x_0) \leq -\beta^{(2)}_{n,\kappa}$. Consequently $f_n(x_n(\omega), \omega) - f_0(x_n(\omega)) \leq -\beta^{(2)}_{n,\kappa}$ and we can employ the approximation of f_0. (ii) can be proved in an analogous way. □

Finally we consider the solution sets. A superset-approximation is easy to obtain. Let
$\Psi^{sup-r}_{n,\kappa} := \{x \in \Gamma^{sup}_{n,\kappa} : \ f_n(x, \omega) \leq \Phi^{(u)}_{n,\kappa} + 2\beta^{(2)}_{n,\kappa}\}$.

Theorem. Assume that a superset-appoximation $(\Gamma^{sup}_{n,\kappa})_{n\in N}$ for Γ_0 with tail behavior function $\mathcal{H}_1$ is given. Furthermore, assume that there exist a function $\mathcal{H}_2$ and to all $\kappa > 0$ a sequence $(\beta^{(2)}_{n,\kappa})_{n\in N}$ such that
$\sup\limits_{n\in N} P\{\omega : \inf\limits_{x\in U\Gamma_0} (f_n(x, \omega) - f_0(x)) \leq -\beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_2(\kappa)$ for a suitable neighborhood $U\Gamma_0$. Then
$\sup\limits_{n\in N} P\{\omega : \Psi_0 \setminus \Psi^{sup-r}_{n,\kappa}(\omega) \neq \emptyset\} \leq \mathcal{H}_1(\kappa) + \mathcal{H}_2(\kappa)$.

Proof. Let $\omega \in \Omega$, $n \in N$ and $\kappa > 0$ be such that $\Psi_0 \setminus \Psi^{sup-r}_{n,\kappa}(\omega) \neq \emptyset$. Then there is $x_0(\omega) \in \Psi_0$ that does not belong to $\Psi^{sup-r}_{n,\kappa}(\omega)$. Therefore $f_0(x_0) = \Phi_0$, but $f_n(x_0(\omega), \omega) > \Phi^{(u)}_n + 2\beta^{(2)}_{n,\kappa}$. If $\Phi^{(u)}_{n,\kappa} - \Phi_0 \leq -\beta^{(2)}_{n,\kappa}$ we can employ the foregoing lemma. Otherwise we obtain $\Phi^{(u)}_{n,\kappa} - \Phi_0 > -\beta^{(2)}_{n,\kappa}$, which implies $f_n(x_0(\omega), \omega) - f_0(x_0) > \beta^{(2)}_{n,\kappa}$. □

Thus one can obtain a superset in a simple manner, however, one would like to have also a subset-approximation in order to be able to judge how close the superset comes to the true set. Unfortunately, if we try to define a subset analogously to the above superset-approximation we will usually end up with an empty set. A way out is to derive an inner Kuratowski-Painlevé approximation, which usually does not fulfil the definition for a subset-approximation, but guarantees that at least one element of Ψ_0 lies in a $\beta_{n,\kappa}$-neighborhood of the inner Kuratowski-Painlevé approximation with prescribed high probability. On order to obtain an inner Kuratowski-Painlevé approximation we need estimates for the growth function (at least) of the objective function.

4 Estimation of the Growth Functions

There are many situations where the outer and/or inner growth function μ and $\underline{\mu}$, needed for inner and outer approximations, respectively, are known, because the constraint functions are assumed to be completely known, e.g. in maximum likelihood estimation or if variational inequalities are taken into account. If the constraint functions, however, are obtained via curve estimation, it is desirable to have a method at hand to estimate the growth function μ, or in fact, its inverse μ^{-1}. We will restrict the considerations to the approximation of "outer" growth functions, as needed for inner KP-approximations, cf. [9]. "Inner" growth functions can be obtained in a similar way, regarding $-g_0$ instead of g_0. This case is even easier to deal with, because an empty outer region can not occur under reasonable conditions.

Again, we consider the sets $M_0 = \{x \in R^p : g_0(x) \leq 0\}$ and $M_n(\omega) = \{x \in R^p : g_n(x, \omega) \leq 0\}$ and assume additionally that the functions g_0 and $g_n(\cdot, \omega)$, $n \in N$, are *quasiconvex*. Hence we can make use of the fact that we have convex level sets. These assumptions are not as restrictive as it may seem at the first glance. If, as generally assumed, an increasing growth function μ exists, g_0 can locally be replaced by a strictly quasiconvex function $\tilde{g}_0$ with $\tilde{g}_0 \begin{cases} = 0, \text{if } x \in M_0, \\ \leq g_0(x) \text{ otherwise.} \end{cases}$

If g_n is not strictly quasiconvex, in an appropriate neighborhood of $M_n(\omega)$, a suitable "lower" strictly quasiconvex approximation, if desired of special form, can be used instead of g_n. Note also, that there are methods for estimation under shape constraints, e.g. under convexity assumptions (cf. [1]). Of course, replacing the functions $g_n(\cdot, \omega)$ locally with quasiconvex surrogates will usually increase the radius of the confidence sets.

In order to derive exact confidence sets, we have to make sure, that the confidence sets do not become too small. Obviously, for each function μ or $\underline{\mu}$ also a smaller function will fulfil the corresponding assumptions. Hence we will introduce the greatest possible growth function and the smallest inverse growth function $\mu^{(-1)}$, respectively, as a benchmark. Note that we only need estimates for the values $\mu^{(-1)}(\beta_{n,\kappa})$ with fixed n and κ.

We define the *benchmark* $\hat{\mu}^{-1}(\beta_{n,\kappa})$ by

$$\hat{\mu}^{-1}(\beta_{n,\kappa}) := \inf_{y \in \partial U^{\beta_{n,\kappa}} M_0} d(y, M_0)$$

and the estimate $\mu_n^{-1}(\beta_{n,\kappa},\omega)$ by

$$\mu_n^{-1}(\beta_{n,\kappa},\omega) \ := \begin{cases} \sup\limits_{y\in\partial U^{2\beta_{n,\kappa}}M_n(\omega)} d(y,U^{-\beta_{n,\kappa}}M_n(\omega)), \ \text{if } U^{-\beta_{n,\kappa}}M_n(\omega)\neq\emptyset, \\ \sup\limits_{\substack{x\in\partial U^{2\beta_{n,\kappa}}M_n(\omega)\\ y\in\partial U^{\beta_{n,\kappa}}M_n(\omega)}} d(x,y) \ \text{otherwise.}\end{cases}$$

Then we have the following relationship:

Theorem. Assume that there exist a function $\mathcal{H}$ and for all $\kappa>0$ a sequence $(\beta^{(2)}_{n,\kappa})_{n\in N}$ such that for a suitable neighborhood UM_0 of M_0
$\sup\limits_{n\in N} P\{\omega: \sup\limits_{x\in UM_0}|g_n(x,\omega)-g_0(x)|\geq\beta_{n,\kappa}\}\leq\mathcal{H}(\kappa)$.
Then
$\sup\limits_{n\in N} P\{\omega: \hat{\mu}^{-1}(\beta_{n,\kappa})>\mu_n^{-1}(\beta_{n,\kappa},\omega)\}\leq\mathcal{H}(\kappa)$.

Proof. Let $\omega\in\Omega$, $n\in N$ and $\kappa>0$ be such that $\hat{\mu}^{-1}(\beta_{n,\kappa})>\mu_n^{-1}(\beta_{n,\kappa},\omega)$. Firstly, we assume that $U^{-\beta_{n,\kappa}}M_n(\omega)\neq\emptyset$ and distinguish the following cases:

(i) $U^{\beta_{n,\kappa}}M_0\not\subset U^{2\beta_{n,\kappa}}M_n(\omega)$,

(ii) $U^{-\beta_{n,\kappa}}M_n(\omega)\not\subset M_0$,

(iii) $(U^{-\beta_{n,\kappa}}M_n(\omega)\subset M_0)\ \wedge\ (U^{\beta_{n,\kappa}}M_0\subset U^{2\beta_{n,\kappa}}M_n(\omega))$.

(i) There exists $x_0(\omega)\in U^{\beta_{n,\kappa}}M_0$ that does not belong to $U^{2\beta_{n,\kappa}}M_n(\omega)$. Consequently $g_0(x_0(\omega))\leq\beta_{n,\kappa}$, but $g_n(x_0(\omega),\omega)>2\beta_{n,\kappa}$, which again implies $\sup\limits_{x\in UM_0}|g_n(x,\omega)-g_0(x)|\geq\beta_{n,\kappa}$.

(ii) There exists $x_n(\omega)\in U^{-\beta_{n,\kappa}}M_n(\omega)$ that does not belong to M_0. Hence we have $g_n(x_n(\omega),\omega)\leq-\beta_{n,\kappa}$, but $g_0(x_n(\omega))>0$. This implies $\sup\limits_{x\in UM_0}|g_n(x,\omega)-g_0(x)|\geq\beta_{n,\kappa}$.

(iii) We choose $x_0\in M_0$ and $y_0\in U^{\beta_{n,\kappa}}M_0$ such that $d(x_0,y_0)=\hat{\mu}^{-1}(\beta_{n,\kappa})$ and consider the straight line between x_0 and y_0. Choose $y_n(\omega)\in\partial U^{2\beta_{n,\kappa}}M_n(\omega)$ at the prolongation of the straight line between x_0 and y_0 beyond y_0. Then due to the convexity of the sets under consideration we can proceed as follows: We have
$d(y_n(\omega),M_0)=d(y_n(\omega),x_0)\geq d(y_0,x_0)$. Furthermore, since $U^{-\beta_{n,\kappa}}M_n(\omega)\subset M_0$, the inequality $d(y_n(\omega),z)\geq d(y_n(\omega),x_0)$ holds for all $z\in U^{-\beta_{n,\kappa}}M_n(\omega)$,

hence $d(y_n(\omega), U^{-\beta_{n,\kappa}} M_n(\omega)) \geq d(y_n(\omega), x_0)$. By the definition of $\mu_n^{-1}(\beta_{n,\kappa}, \omega)$ this implies $\mu_n^{-1}(\beta_{n,\kappa}, \omega) \geq d(y_n(\omega), x_0) \geq d(y_0, x_0) = \hat{\mu}^{-1}(\beta_{n,\kappa})$ in contradiction to the assumption.

Now we assume that $U^{-\beta_{n,\kappa}} M_n(\omega) = \emptyset$ and distinguish the following cases:

(i) $U^{\beta_{n,\kappa}} M_0 \not\subset U^{2\beta_{n,\kappa}} M_n(\omega)$,

(ii) $M_0 \not\subset U^{\beta_{n,\kappa}} M_n(\omega)$,

(iii) $(M_0 \subset U^{\beta_{n,\kappa}} M_n(\omega)) \ \wedge \ (U^{\beta_{n,\kappa}} M_0 \subset U^{2\beta_{n,\kappa}} M_n(\omega))$.

In case (i) we can proceed as in the first part of the proof.

(ii) There exists $x_0(\omega) \in M_0$ that does not belong to $U^{\beta_{n,\kappa}} M_n(\omega)$. Consequently $g_0(x_0(\omega)) \leq 0$, but $g_n(x_0(\omega), \omega) > \beta_{n,\kappa}$, which implies $\sup\limits_{x \in U M_0} |g_{n,\kappa}(x, \omega) - g_0(x)| \geq \beta_{n,\kappa}$.
(iii) We choose $x_0 \in M_0$ and $y_0 \in \partial U^{\beta_{n,\kappa}} M_0$ such that $\hat{\mu}^{-1}(\beta_{n,\kappa}) = d(x_0, y_0)$ and consider the straight line between x_0 and y_0. Furthermore, we choose $y_n(\omega) \in \partial U^{2\beta_{n,\kappa}} M_n(\omega)$ on the prolongation of the straight line between x_0 and y_0 beyond y_0. Hence we obtain for a point $z \in U^{\beta_{n,\kappa}} M_n(\omega)$ with maximal distance to $y_n(\omega)$ the following inequality
$d(x_0, y_0) \leq d(x_0, y_n(\omega)) \leq d(z, y_n(\omega)) = \mu_n^{-1}(\beta_{n,\kappa})$ in contradiction to the assumption. $\square$.

5 Estimation of λ

The results concerning the optimal values in [9] contain a so-called continuity function λ. We will show how λ can be approximated in a suitable neighborhood of Γ_n. In order to be able to handle KP-approximations and approximations via relaxation in a unifying way, we deal with suitable neighborhoods $U_r(\beta^{(1)}_{n,\kappa})\Gamma_n,\ r \in \{-1, 1, 2\}$ of Γ_n. We can choose $U_r(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) = U^{r\beta_{n,\kappa}}\Gamma_n(\omega)$ or $U_r(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) = U_{r\beta_{n,\kappa}}\Gamma_n(\omega)$. The neighborhhod $U_1(\beta^{(1)}_{n,\kappa})\Gamma_0$ of Γ_0 is needed for the benchmark.

The following condition will be used:

(Γ-λ) There exist a function $\mathcal{H}_1$ and to all $\kappa > 0$ a sequence $(\beta^{(1)}_{n,\kappa})_{n \in N}$ such that

(Γ-λ-1) $\sup_{n\in N} P\{\omega :\ U_1(\beta^{(1)}_{n,\kappa})\Gamma_0 \setminus U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) \neq \emptyset\} \leq \mathcal{H}_1(\kappa)$ and

(Γ-λ-2) $\sup_{n\in N} P\{\omega :\ U_{-1}(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) \setminus \Gamma_0 \neq \emptyset\} \leq \mathcal{H}_1(\kappa)$.

The smallest possible $\lambda^{-1}(\beta^{(1)}_{n,\kappa})$, denoted by $\hat{\lambda}^{-1}(\beta^{(1)}_{n,\kappa})$, will serve as benchmark:

$$\hat{\lambda}^{-1}(\beta^{(1)}_{n,\kappa}) := -\inf_{y\in U_1(\beta^{(1)}_{n,\kappa})\Gamma_0} f_0(y) + \Phi_0.$$

We propose the following estimate:

$$\lambda_n^{-1}(\beta^{(1)}_{n,\kappa}, \omega) := -\inf_{y\in U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)} f_n(y,\omega) + \inf_{y\in U_{-1}(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)} f_n(y,\omega).$$

Theorem. Let the condition (Γ-λ) and the following condition be satisfied:

(C-f) There exist a function $\mathcal{H}_2$ and to all $\kappa > 0$ a sequence $(\beta^{(2)}_{n,\kappa}(\omega))_{n\in N}$ such that $\sup_{n\in N} P\{\omega : \sup_{x\in U\Gamma_0} |f_n(x,\omega) - f_0(x)| > \beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_2(\kappa)$.

Then
$$\sup_{n\in N} P\{\omega : \lambda_n^{-1}(\beta^{(1)}_{n,\kappa}, \omega) < \hat{\lambda}^{-1}(\beta^{(1)}_{n,\kappa}) - 2\beta^{(2)}_{n,\kappa}\} \leq \mathcal{H}_1(\kappa) + \mathcal{H}_2(\kappa).$$

Proof. Let $\omega \in \Omega$, $n \in N$ and $\kappa > 0$ be such that $\lambda_n^{-1}(\beta^{(1)}_{n,\kappa}, \omega) < \hat{\lambda}^{-1}(\beta^{(1)}_{n,\kappa}) - 2\beta^{(2)}_{n,\kappa}$.

We choose

$x_0 \in \Gamma_0$ such that $f_0(x_0) = \inf_{x\in\Gamma_0} f_0(x)$,

$x_n(\omega) \in U_{-1}(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)$ such that $f_n(x_n(\omega),\omega) = \inf_{x\in U_{-1}(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)} f_n(x,\omega)$,

$y_0 \in U_1(\beta^{(1)}_{n,\kappa})\Gamma_0$ such that $f_0(y_0) = \inf_{y\in U_1(\beta^{(1)}_{n,\kappa})\Gamma_0} f_0(y)$,

$y_n(\omega) \in U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)$ such that $f_n(y_n(\omega),\omega) = \inf_{y\in U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)} f_n(y,\omega)$.

We distinguish the following cases:

(i) If $y_0 \notin U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)$, we obtain $U_1(\beta^{(1)}_{n,\kappa})\Gamma_0 \setminus U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) \neq \emptyset$ and can employ (Γ-λ-1). (ii) If $x_n(\omega) \notin \Gamma_0$, we obtain $U_{-1}(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega) \setminus \Gamma_0 \neq \emptyset$ and can employ (Γ-λ-2).

(iii) Now we assume that $y_0 \in U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)$ and $x_n(\omega) \in \Gamma_0$. Then the assumption implies
$-f_n(y_n(\omega),\omega) + f_n(x_n(\omega),\omega) + f_0(y_0) - f_0(x_0) < -2\beta^{(2)}_{n,\kappa}$.
Furthermore, as $y_0 \in U_2(\beta^{(1)}_{n,\kappa})\Gamma_n(\omega)$, we obtain $f_n(y_0,\omega)) \geq f_n(y_n(\omega),\omega)$, hence $-f_n(y_0,\omega)) \leq -f_n(y_n(\omega),\omega)$. Because of $x_n(\omega) \in \Gamma_0$ we have $f_0(x_n(\omega)) \geq f_0(x_0)$, hence $-f_0(x_n(\omega)) \leq -f_0(x_0)$. Summarizing, we obtain
$-f_n(y_n(\omega),\omega) + f_0(y_n(\omega)) + f_n(x_n(\omega),\omega) - f_0(x_n(\omega)) < -2\beta^{(2)}_{n,\kappa}$.
This implies that either $-f_n(y_n(\omega),\omega) + f_0(y_n(\omega)) < -\beta^{(2)}_{n,\kappa}$ or $f_n(x_n(\omega),\omega) - f_0(x_n(\omega)) < -\beta^{(2)}_{n,\kappa}$. In both cases we can employ (C-f). □

References

[1] M. Birke, H. Dette: *Estimating a convex function in nonparametric regression.* Scand. J. Stat. 34: 384–404, 2007.

[2] G. Lugosi: *Concentration-of-measure inequalities.* (Lecture Notes, Winter School 'Probabilistic Methods in High Dimensional Phenomena', Toulouse, 2005).

[3] G.C. Pflug: *Stochastic Optimization and Statistical Inference.* (In: Stochastic Programming, A. Ruszczyński, A. Shapiro, eds., Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 2003) 427–482.

[4] R.T. Rockafellar, R.J.-B. Wets: *Variational Analysis.* (Springer, 1998).

[5] W. Römisch: *Stability of Stochastic Programming.* (In: Stochastic Programming, A. Ruszczyński, A. Shapiro, eds., Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 2003), 483 - 554.

[6] T. Sinotina, S. Vogel: *Universal confidence sets for the mode of a regression function.* Preprint TU Ilmenau, 2010.

[7] S. Vogel: *A stochastic approach to stability in stochastic programming.* J. Comput. and Appl. Math., Series Appl. Analysis and Stochastics, 56: 65–96, 1994.

[8] S. Vogel: *Qualitative stability of stochastic programs with applications in asymptotic statistics.* Statistics and Decisions, 23: 219–248, 2005.

[9] S. Vogel: *Universal confidence sets for solutions of optimization problems.* SIAM J. on Optimization, 19: 1467–1488, 2008.

[10] S. Vogel: *Confidence Sets and Convergence of Random Functions.* In: In: Festschrift in Celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday, C. Tammer, F. Heyde (Hrsg.), Shaker-Verlag 2008.