Confidence Sets and Convergence of Random Functions

Silvia Vogel

Technische Universität Ilmenau
Silvia.Vogel@tu-ilmenau.de

Summary. Universal confidence sets for solutions of optimization problems are sequences of random sets $(C_n)_{n \in N}$ with the property that for each sample size n the set C_n covers the true solution at least with a prescribed probability. Universal confidence sets can be derived making use of uniform concentration-of-measure results for sequences of random functions and knowledge about the limit problem, e.g. a growth condition. We present sufficient conditions for the convergence assumptions and show how estimates for the growth function can be included.

Key words: universal confidence sets, uniform concentration-of-measure, estimates for the growth function
MSC2000: 90C15, 90C31, 62F25, 62F30

1 Introduction

Random optimization problems occur in many frameworks. For instance, in real-life decision problems usually not all parameters, probability distributions etc. are completely known. Replacing the unknown quantities with estimates, one arrives at random surrogate problems. Hence there is the need for assertions that help to evaluate the goodness of the solutions for the surrogate problems. Stability theory of stochastic programming provides qualitative and quantitative results for random optimization problems that can be utilized for that purpose, see [11], [12], [17], [18].

Furthermore, stability theory of stochastic programming can also contribute to asymptotic or non-asymptotic estimation theory. Many statistical estimators are obtained as solutions of random optimization problems while the true value can be regarded as the solution of a suitable deterministic 'limit' optimization problem. Hence assertions about the appropriate convergence of the solution sets of random optimization problems in the almost surely or in probability sense can be employed to derive weak or strong consistency statements. Qualitative stability theory of stochastic programming provides such results under rather general assumptions. There are results e.g. for constrained problems or problems with non-unique solution, so-called non-identifiable problems. Moreover, statistics which are based on dependent samples can be dealt with. Particularly the paper [17] is intended as a toolkit of methods which can be put together to prove consistency for many problems.

If more information about the limit problem and the approximations is available, also 'quantitative' results, e.g. confidence sets, can be derived. In classical statistics confidence sets are usually obtained in the following way: One determines the distribution of a suitable statistic and derives conclusions for the parameter under consideration. If the exact distribution is not available, the asymptotic distribution is used as a surrogate, often without any information about the 'distance' between the asymptotic distribution and and true distribution for the given sample size.

In order to derive the asymptotic distribution, usually differentiability assumptions etc. are imposed, see e.g. [8], [10], [13]. If such properties are not satisfied or hard to prove, again, stability theory of stochastic programming can be employed. Qualitative stability results 'in distribution' provide assertions about convergence and one-sided convergence in distribution of the solution sets. In [9] and [3] results of that kind are employed to derive asymptotic confidence sets. The paper [10] shows how optimization problems with fixed constraint set which converge to a deterministic problem with single-valued solution can be 'blown up' to obtain a suitable random limit problem.

Qualitative stability results for convergence in distribution do, however, not provide reliable confidence sets for small sample sizes n. One way out is to allow for sets which may be a bit larger than the confidence sets that would be obtained for the true, but unknown distribution. So-called outer approximations in probability supplemented with a convergence rate and common tail behavior function have proved to be useful tools to derive suitable supersets [11], [18]. Crucial assumptions for the results in [11], [18] are uniform convergence-of-measure results for the approximating objective and constraint functions. In this paper we will suggest a general approach, which may serve as a bridge between the needed uniform convergence assumptions and available concentration-of-measure inequalities for sequences of suitable random variables, see for instance [7]. Furthermore, the derivation of confidence sets requires some knowledge about the deterministic limit problem, such as growth conditions for the limit objective function. The function which describes the growth condition occurs in the convergence rate. Even if a special form of the objective function can be assumed there are usually certain unknown parameters. We will show how estimates for the growth function can be incorporated into the considerations. A natural consequence are random convergence rates.

The paper is organized as follows. In Section 2 we provide the mathematical model and explain how universal confidence sets can be derived from suitable assertions about convergence in probability with convergence rate and tail behavior function. In order not to overload the paper with technical details, we will confine the considerations to problems with a fixed constraint set. Section 3 contains the convergence result for the solution sets we rely on. Sufficient conditions for the uniform convergence assumptions are dealt with in Section 4. Estimation of the growth function for the limit problem is the topic of Section 5. Section 6 outlines how the results can be combined to an adaptive procedure for the derivation of 'good' confidence sets.

2 Universal Confidence Sets

We investigate solutions of a random optimization problem in a complete separable metric space. Together with the sequence of random optimization problems we consider a deterministic limit problem. As mentioned in the introduction we will only investigate problems with fixed constraints.

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

$(P_0)\ \min\limits_{x\in K} f_0(x)$

is approximated by a sequence of random problems

$(P_n)\ \min\limits_{x\in K} f_n(x, \omega), \qquad n \in N.$

$f_0|E \to R^1$ is a lower semicontinuous function and $f_n|E \times \Omega \to R^1$ are lower semicontinuous random functions which are supposed to be $(\mathcal{B}(E) \otimes \Sigma, \mathcal{B}^1)$-measurable where $\mathcal{B}(E)$ denotes the Borel-σ-field of E. 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 (compare [1], Theorem III.30, Lemma III.37 and 'applications').

By Φ_n we denote the optimal value and by Ψ_n the solution set of the random approximate problems (P_n). Correspondingly, by Φ_0 we denote the optimal value and by Ψ_0 the solution set of the deterministic limit problem (P_0).

Our main concern will be with the solution sets. In the following inequalities $(\beta_{n,\kappa})_{n\in N}$ is a sequence of nonnegative numbers which tends to zero for each $\kappa > 0$ and $\mathcal{H}_n,\ n \in N\}$ is a family of functions with the property that for each $n \in N$ $\lim\limits_{\kappa\to\infty} \mathcal{H}_n(\kappa) = 0$ holds. $U_\alpha X$ denotes an open neighborhood of the set $X \subset E$ with radius α: $U_\alpha X := \{x \in E :\ d(x, X) < \alpha\}$ and $\bar{U}_\alpha X$ means its closure.

$$\forall \kappa > 0\ \forall n \in N :\ P\{\omega :\ \Psi_n(\omega) \setminus U_{\beta_{n,\kappa}} \Psi_0 \neq \emptyset\} \leq \mathcal{H}_n(\kappa) \tag{1}$$

and

$$\forall \kappa > 0\ \forall n \in N:\ P\{\omega:\ \Psi_0 \setminus U_{\beta_{n,\kappa}} \Psi_n(\omega) \neq \emptyset\} \leq \mathcal{H}_n(\kappa). \tag{2}$$

$(\beta_{n,\kappa})_{n\in N}$ is a sequence of nonnegative numbers which tends to zero for each $\kappa > 0$ and $\mathcal{H}_n(\cdot),\ n \in N\}$ is a family of functions with the property that for each $n \in N$ $\lim_{\kappa\to\infty} \mathcal{H}_n(\kappa) = 0$ holds. $U_\alpha X$ denotes an open neighborhood of the set $X \subset E$ with radius α: $U_\alpha X := \{x \in E:\ d(x, X) < \alpha\}$ and $\bar{U}_\alpha X$ means its closure.

In stochastic programming, objective functions and constraint functions which are expectations of random functions are of special interest. If the true, but unknown distribution is replaced with the empirical measure, one often obtains a convergence rate of the form $\hat{\beta}_{n,\kappa} = \frac{\kappa}{\sqrt{n}}$ for these functions. As we shall see later, the rate $\beta_{n,\kappa}$ which occurs in (1) and (2) is an increasing function of $\hat{\beta}_{n,\kappa}$. Hence the neighborhoods grow with increasing κ and become smaller with increasing n, c.f. [18].

Firstly, assume that $\mathcal{H}$ does not depend on n. In order to derive *universal confidence sets to the level* ε, i.e. a sequence of random sets $(C_n)_{n\in N}$ with the property that, for each $n \in N$, the limit solution set Ψ_0 is covered by C_n at least with probability ε, i.e. $\sup_{n\in N} P\{\omega:\ \Psi_0 \setminus C_n(\omega) \neq \emptyset\} \leq \varepsilon$, we can then proceed as follows:

Suppose that a sequence $(\Psi_n)_{n\in N}$ that satisfies inequality (2) is available and choose to $\varepsilon > 0$ a $\kappa = \kappa(\varepsilon) > 0$ such that $\mathcal{H}(\kappa) \leq \varepsilon$. The sets

$$C_n := U_{\beta_{n,\kappa}} \Psi_n,\ n \in N, \tag{3}$$

have the desired property. Of course, one is interested in small confidence sets, hence $(\beta_{n,\kappa})_{n\in N}$ should go to zero as fast as possible and $\mathcal{H}$ should converge to zero as fast as possible if κ tends to infinity.

In Sections 3 and 4 we derive assertions for tail behavior functions which have the form $\mathcal{H}_n(\kappa) = L_n(\kappa)\tilde{\mathcal{H}}(\kappa)$. Then, for given $\varepsilon > 0$ and $n \in N$, we choose $\hat{\kappa}_n(\varepsilon) := \min\{\kappa:\ \mathcal{H}_n(\kappa) \leq \varepsilon\}$. In order to obtain reasonable confidence sets we assume that for each $\varepsilon > 0$ the condition $\lim_{n\to\infty} \beta_{n,\hat{\kappa}_n(\varepsilon)} = 0$ holds. This assumption is often satisfied in applications.

Unfortunately, under reasonable conditions one can only prove inequality (1), which means, roughly spoken, that only a subset of the 'true' solution set Ψ_0 is approximated. However, if Ψ_0 is single-valued and the sets $\Psi_n,\ n \in N$, are uniformly bounded, inequality (1) implies inequality (2), and we can proceed as above. The uniform boundedness condition is satisfied because of the compactness of K. If the solution set to the limit problem is not single-valued one can consider relaxed problems in order to derive inequalities of the form (2), see [18]. When dealing with relaxations, all quantities can depend on κ, especially the optimal values and the solution sets. Here, for sake of simplicity, we will do without relaxations. We will prove needed assertions in the form (1) and assume - for the derivation of confidence sets - that the solution set to the limit problem is single-valued.

Because of the relationship to inner and outer approximations in probability we will call a sequence $(\Psi_{n,\kappa})_{n\in N}$ fulfilling relation (1) an *inner approximation in probability to* Ψ_0 *with convergence rate* $\beta_{n,\kappa}$ *and tail behavior function* $\mathcal{H}_n$ (in short, an *inner* $(\beta_{n,\kappa}, \mathcal{H}_n)$*-approximation*) and a sequence $(\Psi_{n,\kappa})_{n\in N}$ fulfilling (2) an *outer approximation in probability to* Ψ_0 *with convergence rate* $\beta_{n,\kappa}$ *and tail behavior function* $\mathcal{H}$ (in short, an *outer* $(\beta_{n,\kappa}, \mathcal{H})$*-approximation*). Since supersets of outer approximations are again outer approximations, one is especially interested in outer approximations which are also inner approximations. The relationship between the two kinds of approximations and Kuratowski-Painlevè-convergence in probability is discussed in more detail in [18].

As mentioned, in order to obtain reasonable confidence sets one would like to have $\lim_{n\to\infty} \beta_{n,\kappa(\varepsilon)} = 0$ or $\lim_{n\to\infty} \beta_{n,\hat{\kappa}_n(\varepsilon)} = 0$ and the limits should go to zero as fast as possible. These properties are, however, not needed to prove the results in Sections 3, 4, and 5. We only assume throughout the paper that the sequences $(\beta_{n,\kappa})_{n\in N}$ belong to the class B of non-increasing sequences of positive numbers and the functions $\mathcal{H}$ and $\tilde{\mathcal{H}}$ belong to the class H of non-increasing functions which are defined on R^+ and map into R^+.

3 Approximation of the Solution Set

In [18] we proved results about inner and outer approximations of constraint sets and solution sets as well as results about lower and upper approximations of the optimal values, supplemented with a deterministic convergence rate and a tail behavior function. Crucial assumptions are uniform concentration-of-measure conditions for the objective functions and the constraint functions and conditions about the limit problem, which concern the growth of the objective function, some kind of semi-continuity of the objective function and the mutual position of the constraint set and the lower level set of the optimal value.

We will provide a special case of the results in [18], which applies to many practical problems and can demonstrate the role of the convergence conditions and the growth condition without many technical details. We present a result about the inner approximation of the solution set for problems with a fixed constraint set. In this case only the growth condition for the limit objective function has to be taken into account in order to derive the convergence rate. The growth condition will be described with a function μ, which belongs to a set $\Lambda := \{\mu | R^+ \to R^+$: μ is increasing, right-continuous, and satisfies $\mu(0) = 0\}$. For the readers convenience we provide the simple proof. Theorem 1 is, in fact, with respect to the tail behavior function a bit more general than the results in [18], because it allows for functions $\mathcal{H}_n$ that depend on n.

Note that for Theorem 1 the single-valuedness of Ψ_0 is not needed.

Theorem 1 (Inner Approximation of the Solution Set). *Assume that the following assumptions are satisfied:*

(Cf-n) There exist functions $\mathcal{H}_n \in H$ *and for all* $\kappa > 0$ *a sequence* $(\beta_{n,\kappa})_{n\in N} \in B$ *such that*
$\forall n \in N: \; P\{\omega: \sup_{x\in K} |f_n(x,\omega) - f_0(x)| \geq \beta_{n,\kappa}\} \leq \mathcal{H}_n(\kappa).$

(G-f_0) There exists a function $\mu \in \Lambda$ *such that for all* $\delta > 0$
$\forall\, x \in K \setminus U_\varepsilon \Psi_0: \; f_0(x) \geq \Phi_0 + \mu(\delta).$

Then for all $\kappa > 0$ *and* $\tilde{\beta}_{n,\kappa} = \mu^{-1}(2\beta_{n,\kappa})$ *the relation*
$\forall n \in N: P\{\omega: \; U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \subset K \text{ and } \Psi_n(\omega) \setminus U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \neq \emptyset\} \leq \mathcal{H}_n(\kappa)$
holds.

Proof. Let $\kappa > 0$, $n \in N$, and $\omega \in \Omega$ be such that $U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \subset K$ and $\Psi_n(\omega) \setminus U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \neq \emptyset$. Then there is $x_n(\omega) \in \Psi_n(\omega)$ which does not belong to $U_{\tilde{\beta}_{n,\kappa}} \Psi_0$. Furthermore, choose $x_0 \in \Psi_0$. Because of (G-f_0)we have $f_0(x_n(\omega)) - f_0(x_0) \geq \mu(\mu^{-1}(2\beta_{n,\kappa})) \geq 2\beta_{n,\kappa}$. On the other side we have $f_n(x_0, \omega) - f_n(x_n(\omega), \omega) \geq 0$. Consequently, $-f_n(x_n(\omega), \omega) + f_0(x_n(\omega) + f_n(x_0, \omega) - f_0(x_0) \geq 2\beta_{n,\kappa}$ holds. Hence either $-f_n(x_n(\omega), \omega) + f_0(x_n(\omega) \geq \beta_{n,\kappa}$ or $+f_n(x_0, \omega) - f_0(x_0) \geq \beta_{n,\kappa}$, and we can employ (C$f$-n). □

4 Sufficient Conditions for the Uniform Convergence

4.1 Pointwise Approach

A crucial assumption in Theorem 1 and the results in [18] is the (one-sided) uniform convergence in probability with convergence rate and tail behavior function. Some sufficient conditions for these assumptions are presented in [11] and [18]. In the following we will show how further sufficient conditions can be derived. We extend an approach, which was exploited for qualitative stability results in [15] and [6], to the quantitative framework of [18]. The method can be applied to many optimization problems. The constraint set will be covered by finitely many neighborhoods $U^{n,\kappa}\{x_i\}, \; x_i \in E$, which are chosen in such a way that condition (C2) below is satisfied. The index $^{n,\kappa}$ at the neighborhoods is to indicate the dependence on n and κ. It does not mean the radius as in U_α. Because $(\beta_{n,\kappa})_{n\in N}$ should go to zero in order to obtain reasonable confidence sets, the number $L_n(\kappa)$ of neighborhoods needed will increase with increasing n.

For sake of convenience we assume that f_0 is continuous. The upper semicontinuity ensures that condition (C2) can be satisfied. Because of the lower semicontinuity in the proof to Theorem 2 we can assume that the infimum is a minimum.

Theorem 2 (Pointwise Approach). *Assume that there exist a function* $\mathcal{H} \in H$ *and to all* $\kappa > 0$ *sequences* $(\beta^{(a)}_{n,\kappa})_{n\in N}$, $(\beta^{(b)}_{n,\kappa})_{n\in N}$, *sets* $M_{n,\kappa} = \{x^{(1)}_{n,\kappa}, \ldots, x^{(L_n(\kappa))}_{n,\kappa}\}$, *and neighborhoods* $U^{n,\kappa}\{x\}$, $x \in M_{n,\kappa}$, *such that for each* $n \in N$ *the following conditions are satisfied:*

(C1) $K \subset \bigcup_{x_i \in M_{n,\kappa}} U^{n,\kappa}\{x_i\}$,

(C2) $\forall x_i \in M_{n,\kappa}\ \forall x \in U^{n,\kappa}\{x_i\} : \ f_0(x) - f_0(x_i) \le \beta^{(a)}_{n,\kappa}$,

(C3) $\sup_{n\in N} \max_{x_i \in M_{n,\kappa}} P\{\omega : \inf_{x\in U^{n,\kappa}\{x_i\}} f_n(x,\omega) - f_0(x_i) \le -\beta^{(b)}_{n,\kappa}\} \le \mathcal{H}(\kappa)$.

Then
$\sup_{n\in N} P\{\omega : \inf_{x\in K}(f_n(x,\omega) - f_0(x)) \le -\beta^{(a)}_{n,\kappa} - \beta^{(b)}_{n,\kappa}\} \le L_n(\kappa)\mathcal{H}(\kappa)$.

Proof. Assume that for fixed $\kappa > 0$, $n \in N$, and $\omega \in \Omega$ the relation $\inf_{x\in K}(f_n(x,\omega) - f_0(x)) \le -\beta^{(a)}_{n,\kappa} - \beta^{(b)}_{n,\kappa}$ holds. The functions $f_n - f_0$ are lower semicontinuous and $(\mathcal{B}(E) \otimes \Sigma, \mathcal{B}^1)$-measurable. Hence the multifunctions MF_n with $MF_n(\omega) = K \cap \{x \in E : \ f_n(x,\omega) - f_0(x) \le -\beta^{(a)}_{n,\kappa} - \beta^{(b)}_{n,\kappa}\}$ admit measurable selections. Thus we can choose a measurable function x_n with $x_n(\omega) \in K$ such that $f_n(x_n(\omega),\omega) - f_0(x_n(\omega)) \le -\beta^{(a)}_{n,\kappa} - \beta^{(b)}_{n,\kappa}$.
Consequently, there is $x_i \in M_{n,\kappa}$ such that $x_n(\omega) \in U^{n,\kappa}\{x_i\}$. Hence
$f_n(x_n(\omega),\omega) - f_0(x_n(\omega),\omega) \ge f_n(x_n(\omega),\omega) - f_0(x_i) - \beta^{(b)}_{n,\kappa}$
and $f_n(x_n(\omega),\omega) - f_0(x_i) \le -\beta^{(b)}_{n,\kappa}$. It remains to employ (C3). □

Sequences of functions $(f_n)_{n\in N}$ which satisfy the conclusion of the above theorem are called *lower semicontinuous approximations in probability to* f_0 *with convergence rate* $\beta_{n,\kappa} := \beta^{(a)}_{n,\kappa} + \beta^{(b)}_{n,\kappa}$ *and tail behavior function* $\mathcal{H}_n := L_n\mathcal{H}$, in short *lower semicontinuous* $(\beta_{n,\kappa}, \mathcal{H}_n)$*-approximations.* The denotation 'lower semicontinuous' has been chosen because of the relationship to lower semicontinuity of a function of two parameters where n is regarded as parameter. A corresponding result can be proved for an upper semicontinuous approximation, replacing f_n with $-f_n$ and f_0 with $-f_0$. Sequences $(f_n)_{n\in N}$ which are lower and upper $(\beta_{n,\kappa}, \mathcal{H}_n)$-approximations satisfy condition (Cf-n), which describes the uniform convergence in probability with convergence rate and tail behavior function.

4.2 Functions Which Are Expectations

We will illustrate the method investigating functions f_0 which are expectations with respect to a given probability measure. If the probability measure is not completely known it is often estimated by the empirical measure. In this way one arrives at random approximating functions f_n.
Assume that the function f_0 has the form $f_0(x) = \int_{R^m} \varphi(x,z)dP_Z = \mathrm{E}\varphi(x,Z)$ where Z is a random variable with values in R^m and distribution P_Z. φ is

supposed to be measurable with respect to both variables and continuous in each x for almost all z. The discontinuity set $D(x) = \{z : \varphi(\cdot, z)$ is discontinuous in $x\}$ can vary with x.

Furthermore, we assume that to each $x \in E$ there is a neighborhood $U_{\hat{\varepsilon}}\{x\}$ such that $\mathrm{E} \sup_{\tilde{x} \in U_{\hat{\varepsilon}}\{x\}} |\varphi(\tilde{x}, z)|$ exists. $\hat{\varepsilon}$ can depend on x.

P_Z is approximated by the empirical measure P_n based on a sequence of i.i.d. random variables $Z_1, Z_2, \ldots$, and the functions f_n are integrals with respect to P_n: $f_n(x, \omega) = \int_{R^m} \varphi(x, z) dP_n(z, \omega) = \frac{1}{n} \sum_{i=1}^{n} \varphi(x, Z_i(\omega))$.

Then to each $x_i \in K$ and each monotonously decreasing sequence $(\beta^{(a)}_{n,\kappa})_{n \in N}$ we can choose a monotonously decreasing sequence of balls $(U^{n,\kappa}\{x_i\})_{n \in N}$, such that $U^{n,\kappa}\{x_i\} \subset U_{\hat{\varepsilon}}\{x_i\},\ n \in N,\ \kappa > 0$, and the following condition is satisfied:

(C4) $\forall \kappa > 0\ \forall n \in N : \int_{R^m} \sup_{x \in U^{n,\kappa}\{x_i\}} |\varphi(x, z) - \varphi(x_i, z)| dP_Z(z) \leq \beta^{(a)}_{n,\kappa}$.

Since for all $x \in U^{n,\kappa}\{x_i\}$ the relation $f_0(x) \leq \sup_{x \in U^{n,\kappa}\{x_i\}} \int_{R^m} \varphi(x, z) dP_Z(z) \leq \int_{R^m} \sup_{x \in U^{n,\kappa}\{x_i\}} \varphi(x, z) dP_Z(z)$ holds, (C4) implies (C2). Bounds for $L_n(\kappa)$ depend on the shape of f_0 and can be derived under additional assumptions. Below we present an example.

Now we consider (C3) for a fixed $x_0 \in E$. Let $Y^{\varepsilon}_k(\omega) := \inf_{x \in U_{\varepsilon}\{x_0\}} \varphi(x, Z_k(\omega))$. We will employ the following condition.

(CM) To each $\kappa > 0$ there exist a sequence $(\beta_{n,\kappa})_{n \in N} \in B$ and a function $\mathcal{H} \in H$ such that for all $0 < \varepsilon < \tilde{\varepsilon} \leq \hat{\varepsilon}(x_0)$ the inequality
$\sup_{n \in N} P\{\omega : \frac{1}{n} \sum_{k=1}^{n} Y^{\varepsilon}_k(\omega) - \mathrm{E}Y^{\varepsilon}_1 \leq -\beta_{n,\kappa}\} \leq \mathcal{H}(\kappa)$
is fulfilled.

Lemma 1. *Let x_0 be fixed and suppose that (CM) is satisfied. Choose $(U^{n,\kappa}\{x_0\})_{n \in N}$ such that (C4) is fulfilled with $\beta^{(a)}_{n,\kappa} = \beta_{n,\kappa},\ n \in N$, and $U^{n,\kappa}\{x_0\} \subset U_{\tilde{\varepsilon}}\{x_0\}$ holds for all $\kappa > 0$ and $n \in N$.*
Then the relation

$$\sup_{n \in N} P\{\omega : \inf_{x \in U^{n,\kappa}\{x_0\}} f_n(x, \omega) - f_0(x_0) \leq -2\beta_{n,\kappa}\} \leq \mathcal{H}(\kappa) \text{ holds.}$$

Proof. Let $\kappa > 0$ and $n \in N$ be given. Then, due to (C4), we have $-\int_{R^m} \inf_{x \in U^{n,\kappa}\{x_0\}} \varphi(x, z) dP_Z(z) + f_0(x_0) < \beta_{n,\kappa}$. Because of the assumed shape of $U^{n,\kappa}\{x_0\}$ we find $\varepsilon \leq \tilde{\varepsilon}$ such that $U^{n,\kappa}\{x_0\} = U_{\varepsilon}\{x_0\}$.

Now consider $\omega \in \Omega$ such that $\inf_{x \in U^{n,\kappa}\{x_0\}} f_n(x, \omega) - f_0(x_i) \leq -2\beta_{n,\kappa}$ is satisfied. Since $\int_{R^m} \inf_{x \in U^{n,\kappa}\{x_0\}} \varphi(x, z) dP_n(z, \omega) \leq \inf_{x \in U^{n,\kappa}\{x_0\}} f_n(x, \omega)$ holds for all $\omega \in \Omega$, we obtain $\int\limits_{R^m} \inf\limits_{x \in U^{n,\kappa}\{x_0\}} \varphi(x,z)dP_n(z,\omega) - f_0(x_0) \le -2\beta_{n,\kappa}$. Together with $-\int\limits_{R^m} \inf\limits_{x \in U_\varepsilon\{x_0\}} \varphi(x,z)dP_Z(z) + f_0(x_0) < \beta_{n,\kappa}$ this implies $\int\limits_{R^m} \inf\limits_{x \in U_\varepsilon\{x_0\}} \varphi(x,z)dP_n(z,\omega) - \int\limits_{R^m} \inf\limits_{x \in U^{n\kappa)}\{x_0\}} \varphi(x,z)dP_Z(z) \le -\beta_{n,\kappa}$ and eventually $\frac{1}{n}\sum\limits_{k=1}^{n} Y_k^\varepsilon(\omega) - \mathrm{E}Y_1^\varepsilon \le -\beta_{n,\kappa}$. Hence we can employ (CM). □

The assumption (CM) is a concentration-of-measure result which has to be fulfilled for each $\varepsilon > 0$ with the same convergence rate and tail behavior function. Taking into account the special form of the random variables under considerations, we see that the condition is often satisfied if it is fulfilled for a certain $\tilde{\varepsilon}$. Hence the meanwhile large collection of concentration-of-measure results can be exploited, see for instance [7] or [2].

The neighborhoods $U^{n,\kappa}\{x\},\ x \in E$, need not have the same radius. Hence the radius can be fitted to the shape of f_0. We can derive bounds for the number of neighborhoods needed if we have some knowledge about f_0, e.g. a Lipschitz condition. If a Lipschitz condition is satisfied we can even use the same radius for all neighborhoods $U^{n,\kappa}\{x_i\}$, see the following lemma.

Lemma 2. *Let $E = R^p$ and suppose that there is a ball B with radius r which contains K. Additionally, assume that φ fulfils a Lipschitz condition in the following form:*

*(L) $\exists x_0 \in B\ \forall x \in B:\ \varphi(x,z) - \varphi(x_0,z) \le L(z)d(x,x_0)$
with an integrable Lipschitz constant L.*

Then $L_n(\kappa) = \lfloor \frac{r\beta^{(a)}_{n,\kappa}}{\mathrm{E}L} + 1 \rfloor^p$ neighborhoods $U^{n,\kappa}\{x_i\}$ can be chosen such that (C1) and (C2) are fulfilled.

Proof. We consider balls $U^{n,\kappa}(x_i),\ x_i \in R^p$, with radius $r_{n,\kappa}$ which does not depend on x_i. Then we have $\sup\limits_{x \in U^{n,\kappa}\{x_i\}} \varphi(x,z) - \varphi(x_i,z) \le L(z)r_{n,\kappa}$. The condition $\int\limits_{R^m} \sup\limits_{x \in U^{n,\kappa}\{x_i\}} \varphi(x,z)dP_Z(z) - \int\limits_{R^m} \varphi(x_i,z)dP_Z(z) \le \beta^{(a)}_{n,\kappa}$ is fulfilled if $r_{n,\kappa}$ satisfies the condition $r_{n,\kappa} \le \frac{\beta^{(a)}_{n,\kappa}}{\mathrm{E}L}$. Hence $\lfloor \frac{r\beta^{(a)}_{n,\kappa}}{\mathrm{E}L} + 1 \rfloor^p$ balls cover K. □

5 Estimation of the Growth Function

In this section we shall show how adaptive approximations of the function μ can be included. If the growth condition (G-f_0) is satisfied for a given $\mu \in \Lambda$ it is also fulfilled for each smaller $\tilde{\mu} \in \Lambda$. Because μ^{-1} occurs in the convergence rate one is interested in approximating the largest μ. Therefore in the following

we consider functions μ which do not only satisfy condition (G-f_0), but also the following condition (lG-f_0). μ will play the role of a benchmark for the estimate μ_n.
Let $\delta_0 := \sup\{\delta > 0 : K \setminus U_\delta \Psi_0 \neq \emptyset\}$.

(lG-f_0) There exist a function $\mu \in \Lambda$ such that $\forall \delta \in (0, \delta_0]$ the following relations are satisfied:
$\forall x \in K \setminus U_\delta \Psi_0 : \; f_0(x) - \Phi_0 \geq \mu(\delta)$,
$\exists x_\delta \in \mathrm{bdy} U_\delta \Psi_0 : f_0(x) - \Phi_0 \leq \mu(\delta)$.

μ will be estimated by μ_n. For an $\omega \in \Omega$ the approximate objective function $f_n(\cdot, \omega)$ is known. Hence we can determine a growth function $\mu_n(\cdot, \omega) \in \Lambda$ such that $\forall x \in K \setminus U_\delta \Psi_n(\omega) : f_n(x, \omega) - \Phi_n(\omega) \geq \mu_n(\delta, \omega)$.
Again, a smaller function $\tilde{\mu}_n(\cdot, \omega) \in \Lambda$ would do. We only have to make sure that the estimate μ_n does not become too large, because this would result in unjustified small confidence sets if μ is replaced with μ_n. Theorem 3 guarantees that this case can occur only with a small probability if, additionally to (lG-f_0) and (Cf), the following condition (G-f_n) is satisfied.

(G-f_n) There exist continuous, increasing functions $\mu_n,\ \xi_n | R^+ \times \Omega \to R^+$ with trajectories in Λ and $\xi_n(\delta, \omega) \geq \delta$ such that for all $\omega \in \Omega$ and $\forall \delta \in (0, \frac{\delta_0}{2}]$ the following relations are satisfied:
$\forall x \in K \setminus U_\delta \Psi_n(\omega) : f_n(x, \omega) - \Phi_n(\omega) \geq \mu_n(\delta, \omega)$,
$\forall x \in K \setminus U_{\delta + \xi_n(\delta,\omega)} \Psi_n(\omega) \; \forall y \in U_\delta \Psi_n(\omega) : f_n(x, \omega) - f_n(y, \omega) \geq \mu_n(\delta, \omega)$.

Because $f_n(\cdot, \omega)$ is known, for a given $\delta > 0$, we can determine $\mu_n(\delta, \omega)$ and then $\xi_n(\delta, \omega)$. Particularly, if f_n is generated in a way that guarantees certain convexity conditions, μ_n and ξ_n can be determined without too much effort.

We confine the considerations to tail behavior functions which do not depend on n. Therefore we specialize condition (Cf-n) to condition (Cf) below. We abbreviate $\alpha_n(\delta, \omega) := 2\delta + \xi_n(\delta, \omega) + \mathrm{dia}\Psi_n(\omega)$ where $\mathrm{dia}\Psi_n(\omega)$ denotes the diameter of the smallest ball which contains $\Psi_n(\omega)$. Then we can estimate μ in the following way:

Theorem 3 (Approximation of μ). *Suppose that (lG-f_0), (G-f_n), and the following condition are satisfied:*

(Cf) There exist a function $\mathcal{H}$ and to all $\kappa > 0$ a sequence $(\beta_{n,\kappa})_{n \in N} \in B$ such that
$$\sup_{n \in N} P\{\omega : \sup_{x \in K} |f_n(x, \omega) - f_0(x)| \geq \beta_{n,\kappa}\} \leq \mathcal{H}(\kappa).$$

Then for all $\kappa > 0$
$$\sup_{n \in N} P\{\omega : \sup_{0 < \delta < \frac{\delta_0}{2}} (\mu_n(\delta, \omega) - \mu(\alpha_n(\delta, \omega)) > 2\beta_{n,\kappa}\} \leq \mathcal{H}(\kappa).$$

Proof. Let $\kappa > 0$, $n \in N$, and ω be such that $\sup_{0<\delta<\frac{\delta_0}{2}} (\mu_n(\delta, \omega) - \mu(\alpha_n(\delta, \omega)) > 2\beta_{n,\kappa}$. Then there is $\delta \in (0, \frac{\delta_0}{2})$ such that $\mu_n(\delta, \omega) - \mu(\alpha_n(\delta, \omega)) > 2\beta_{n,\kappa}$.

We choose $\hat{x}_n(\omega)$ at the boundary of $U_{\alpha_n(\delta,\omega)}\Psi_0$ such that $\forall x_0 \in \Psi_0: \; f_0(\hat{x}_n(\omega)) - f_0(x_0) \leq \mu(\alpha_n(\delta,\omega))$.

Furthermore, choose $x_n(\omega) \in \Psi_n(\omega)$ with minimal distance to $\hat{x}_n(\omega)$, and $x_0(\omega) \in \Psi_0$ with minimal distance to $\hat{x}_n(\omega)$. Measurability of x_n and x_0 is ensured since Ψ_n is a closed-valued measurable multifunction, Ψ_0 is closed, $\hat{x}_n(\omega)$ can be chosen as a measurable selection, and the distance is a measurable function of its arguments.

We distinguish the cases $d(x_n(\omega), x_0(\omega)) \geq \delta$ and $d(x_n(\omega), x_0(\omega)) < \delta$.

Firstly, assume that $d(x_n(\omega), x_0(\omega)) \geq \delta$. Then, because of $f_n(x_0(\omega),\omega) - f_n(x_n(\omega),\omega) \geq \mu_n(\delta)$ and the choice of $\hat{x}_n(\omega)$, we have the following inequalities:
$2\beta_{n,\kappa} < \mu_n(\delta,\omega) - \mu(\alpha_n(\delta,\omega)) \leq -f_n(x_n(\omega),\omega) + f_n(x_0(\omega),\omega) - f_0(\hat{x}_n(\omega)) + f_0(x_0(\omega))$. Since $f_0(x_0(\omega)) \leq f_0(x_n(\omega))$ and $-f_0(\hat{x}_n(\omega)) \leq -f_0(x_0(\omega))$ we obtain $2\beta_{n,\kappa} < -f_n(x_n(\omega),\omega) + f_0(x_n(\omega)) + f_n(x_0(\omega),\omega) - f_0(x_0(\omega))$. Hence either $-f_n(x_n(\omega),\omega) + f_0(x_n(\omega)) > \beta_{n,\kappa}$ or $f_n(x_0(\omega),\omega) - f_0(x_n(\omega)) > \beta_{n,\kappa}$. Consequently $\sup_{x \in K} |f_n(x,\omega) - f_0(x)| > \beta_{n,\kappa}$ and we employ (Cf).

Secondly, assume that $d(x_n(\omega), x_0(\omega)) < \delta$. Then we have $d(\hat{x}_n(\omega), \Psi_n(\omega)) = d(\hat{x}_n(\omega), x_n(\omega)) \geq d(\hat{x}_n(\omega), x_0(\omega)) - d(x_0(\omega), x_n(\omega)) \geq 2\delta + \xi_n(\delta,\omega) + \mathrm{dia}\Psi_n(\omega) - \delta - \mathrm{dia}\Psi_n(\omega) = \delta + \xi_n(\delta,\omega)$. Consequently $\hat{x}_n(\omega) \in K \setminus U_{\delta+\xi_n(\delta,\omega)}\Psi_n(\omega)$. Otherwise $x_0(\omega) \in U_\delta \Psi_n(\omega)$. Hence we can make use of (G-f_n) and obtain $2\beta_{n,\kappa} < \mu_n(\delta,\omega) - \mu(\alpha_n(\delta,\omega)) \leq f_n(\hat{x}_n(\omega),\omega) - f_n(x_0(\omega),\omega) - f_0(\hat{x}_n(\omega)) + f_0(x_0(\omega))$. Hence either $f_n(\hat{x}_n(\omega),\omega) - f_0(\hat{x}_n(\omega)) > \beta_{n,\kappa}$ or $-f_n(x_0(\omega),\omega) + f_0(x_n(\omega)) > \beta_{n,\kappa}$ and we can proceed as in the first case. □

We are now ready to replace the unknown function μ in the convergence rate by the known function μ_n.

Theorem 4 (Inner Approximation of the Solution Set).
Suppose that (lG-f_0), (G-f_n), and (Cf) are satisfied. Then for all $\kappa > 0$, $n \in N$, *and* $\tilde{\beta}_{n,\kappa}(\omega) = \alpha_n(\mu_n^{-1}(4\beta_{n,\kappa},\omega),\omega)$ *we have*
$\sup_{n \in N} P\{\omega : \tilde{\beta}_{n,\kappa}(\omega) \leq \frac{\delta_0}{2}$ *and* $\Psi_n(\omega) \setminus U_{\tilde{\beta}_{n,\kappa}(\omega)}\Psi_0 \neq \emptyset\} \leq \mathcal{H}(\kappa)$.

Proof: Let $\kappa > 0$, $n \in N$ and $\omega \in \Omega$ be such that $\tilde{\beta}_{n,\kappa}(\omega) \leq \delta_0$ and $\Psi_n(\omega) \setminus U_{\tilde{\beta}_{n,\kappa}(\omega)}\Psi_0 \neq \emptyset$.

Firstly assume that $\sup_{0<\delta<\frac{\delta_0}{2}} (\mu_n(\delta,\omega) - \mu(\alpha_n(\delta,\omega)) > 2\beta_{n,\kappa}$. Then we can employ Theorem 3.

Secondly assume that for all $0 < \delta < \frac{\delta_0}{2}$ the inequality $\mu_n(\delta,\omega) - \mu(\alpha_n(\delta,\omega)) \leq 2\beta_{n,\kappa}$ is fulfilled. Because of $\Psi_n(\omega) \setminus U_{\tilde{\beta}_{n,\kappa}(\omega)}\Psi_0 \neq \emptyset$ there exist $x_n(\omega) \in \Psi_n(\omega)$ with $d(x_n(\omega), \Psi_0) \geq \tilde{\beta}_{n,\kappa}$. Hence, for an $x_0 \in \Psi_0$, $f_0(x_n(\omega)) - f_0(x_0) \geq \mu(\tilde{\beta}_{n,\kappa})$. Furthermore, from $\mu_n(\delta,\omega) - \mu(\alpha_n(\delta,\omega)) \leq 2\beta_{n,\kappa}$ we obtain with the special choice $\delta = \mu_n^{-1}(4\beta_{n,\kappa})$ the inequality $\mu(\tilde{\beta}_{n,\kappa}) = \mu(\alpha_n(\mu_n^{-1}(4\beta_{n,\kappa},\omega),\omega) \geq \mu_n(\mu_n^{-1}(4\beta_{n,\kappa},\omega),\omega)) - 2\beta_{n,\kappa} \leq 2\beta_{n,\kappa}$.

Consequently, $f_0(x_n(\omega)) - f_0(x_0) \geq 2\beta_{n,\kappa}$. Taking into account that $-f_n(x_n(\omega),\omega) + f_n(x_0,\omega) \geq 0$, we see that $-f_n(x_n(\omega),\omega) + f_n(x_0,\omega) + f_0(x_n(\omega)) - f_0(x_0) \geq 2\beta_{n,\kappa}$. Hence either $-f_n(x_n(\omega),\omega) + f_0(x_n(\omega)) \geq \beta_{n,\kappa}$ or $f_n(x_0,\omega) - f_0(x_0) \geq \beta_{n,\kappa}$, and we can employ (C-$f$). □

5.1 Outlook

In order to derive adaptive confidence sets, we can proceed as follows. Let the problem (P_0), a sequence of random approximating problems (P_n), and a prescribed probability level ε_0 be given. For sake of simplicity we assume that the solution set to the problem (P_0) is single-valued and we can deal with inner approximations of the solution sets. Suppose that there exist functions $\mathcal{H}_n \in H$ and to all $\kappa > 0$ (possibly random) sequences $(\beta_{n,\kappa})_{n\in N}$ such that $\forall n \in N P\{\omega: \ \beta_{n,\kappa} \leq \frac{\delta_0}{2}$ and $\Psi_n(\omega) \setminus U_{\beta_{n,\kappa}}\Psi_0 \neq \emptyset\} \leq \mathcal{H}_n(\kappa)$.

Choose a sequence $(\delta_k)_{k\in N}$ of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = 1$. Furthermore, choose n_1, $\hat{\kappa}_{n_1}(\delta_1)$, $\beta_{n_1,\hat{\kappa}_{n_1}(\delta_1)}$, and Ψ_{n_1}. Then we can determine a confidence set $U_{\beta_{n_1,\hat{\kappa}_{n_1}(\delta_1)}}\Psi_{n_1}$ with confidence level δ_1.

The procedure can be repeated with a larger n and the additional restriction $x \in U_{\beta_{n_1,\hat{\kappa}_{n_1}(\delta_1)}}\Psi_{n_1}$. Thus, onthe one hand, in this second stage we can expect a better convergence rate and less effort to determine the solution etc. On the other hand, however, approximations of the constraint sets come into play and we have to take into account additional functions λ and ν which describe the continuity behavior of the objective function and the mutual position of the constraint set and certain level sets for the limit problem, see [18]. The estimation of λ and ν would go beyond the scope of this paper and will be dealt with elsewhere. With suitable estimates for λ, μ, and ν the above procedure can be repeated for the elements of the sequence $(\delta_k)_{k\in N}$ as often as improvements of the solution set are obtained.

Acknowledgement. The author is grateful to an anonymous referee for constructive comments and suggestions.

References

1. G. Castaing, M. Valadier: *Convex Analysis and Measurable Multifunctions.* (Lecture Notes in Mathematics 580, A. Dold, B. Eckmann, eds., Springer, 1977).
2. L. Devroye, G. Lugosi: *Combinatorical Methods in Density Estimation.* (Springer, 2001).
3. O. Gersch: *Convergence in Distribution of Random Closed Sets and Applications in Stability Theory of Stochastic Optimisation.* Dissertation Thesis, Ilmenau University of Technology 2006.
4. W. Hoeffding: *Probability inequalities for sums of bounded random variables.* J. Amer. Statistical Association, 58:13–30, 1963.

5. P. Lachout, E. Liebscher, S. Vogel: *Strong convergence of estimators as ε_n-estimators of optimization problems.* Ann. Inst. Statist. Math., 57:291–313, 2005.
6. P. Lachout, S. Vogel: *On continuous convergence and epi-convergence of random functions. Part II: Sufficient conditions.* Kybernetika, 39:75–98, 2003.
7. G. Lugosi: *Concentration-of-measure inequalities.* (Lecture Notes, Winter School 'Probabilistic Methods in High Dimensional Phenomena', Toulouse 2005).
8. E. L. Lehmann, G. Casella: *Theory of Point Estimation.* 2nd ed. (Springer 1998).
9. G.C. Pflug: *Asymptotic dominance and confidence for solutions of stochastic programs.* Czechoslovak J. Oper. Res., 1:21–30, 1992.
10. G.C. Pflug: *Asymptotic Stochastic Programs.* Math. Oper. Res., 20:769–789, 1995.
11. G.C. Pflug: *Stochastic Optimization and Statistical Inference.* (In: Stochastic Programming, A. Ruszczyński, A. Shapiro, eds., Handbooks in Operations Research and Mangement Science, Vol. 10, Elsevier, 2003) 427–482.
12. W. Römisch: *Stability of Stochastic Programming.* (In: Stochastic Programming, A. Ruszczyński, A. Shapiro, eds., Handbooks in Operations Research and Mangement Science, Vol. 10, Elsevier, 2003), 483 - 554.
13. A. W. van der Vaart, J. A. Wellner: *Weak Convergence and Empirical Processes.* (Springer, 1996).
14. S. Vogel: *A stochastic approach to stability in stochastic programming.* J. Comput. and Appl. Math., Series Appl. Analysis and Stochastics, 56: 65–96, 1994.
15. S. Vogel: *On stability in stochastic programming - Sufficient conditions for continuous convergence and epi-convergence.* (Preprint TU Ilmenau, 1995).
16. S. Vogel: *On semicontinuous approximations of random closed sets with application to random optimization problems.* Ann. Oper. Res., 142: 169–282, 2006.
17. S. Vogel: *Qualitative stability of stochastic programs with applications in asymptotic statistics.* Statistics and Decisions, 23: 219–248, 2005.
18. S. Vogel: *Universal confidence sets for solutions of optimization problems.* to appear in SIAM J. on Optimization.