Working Paper

On the Glivenko-Cantelli Problem in Stochastic Programming: Linear Recourse

Georg Ch. Pflug
Andrzej Ruszczyński
Rüdiger Schultz

WP-95-003
January 1995

IIASA
International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

On the Glivenko-Cantelli Problem in Stochastic Programming: Linear Recourse

Georg Ch. Pflug
Andrzej Ruszczyński
Rüdiger Schultz

WP-95-003
January 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

Integrals of optimal values of random linear programming problems depending on a finite dimensional parameter are approximated by using empirical distributions instead of the original measure. Uniform convergence of the approximations is proved under fairly broad conditions allowing non-convex or discontinuous dependence on the parameter value and random size of the linear programming problem.

Key words: Stochastic Programming, Empirical Measures, Uniform Convergence.

On the Glivenko-Cantelli Problem in Stochastic Programming: Linear Recourse

Georg Ch. Pflug
Andrzej Ruszczyński
Rüdiger Schultz

1 Introduction

Real-world decision problems are usually associated with high uncertainty due to unavailability or inaccuracy of some data, forecasting errors, changing environment, etc. There are many ways to deal with uncertainty; one that proved successful in practice is to describe uncertain quantities by random variables.

Using the probabilistic description of uncertainty within optimization problems leads to *stochastic programming models.* There is a large variety of such models, depending on the nature of information about the random quantitites and on the form of objective and constraints. One of the most popular models, which found numerous applications in operations research practice, is the *two-stage problem.* In its simplest linear form, it can be formulated as follows:

$$\min_{x\in X}\left[c^Tx+\int f(x,\omega)P(d\omega)\right], \tag{1.1}$$

where $X\subset \mathbb{R}^{n_x}$ is the first stage feasible set and $f:\mathbb{R}^{n_x}\times\Omega\mapsto\mathbb{R}$ denotes the *recourse function* dependent on x and on an elementary event in some probability space (Ω,Σ,P). The recourse function is defined as the optimal value of the *second stage problem*

$$f(x,\omega)=\min\left\{q(\omega)^Ty \mid W(\omega)y=b(x,\omega),\ y\geq 0\right\}. \tag{1.2}$$

Here, the vector $y\in\mathbb{R}^{n_y}$ is the second stage decision (which may, in general, depend on x and ω), $q(\omega)$ is a random vector in $\mathbb{R}^{n_y}$, $W(\omega)$ is a random matrix of dimension $m_y\times n_y$ and $b:\mathbb{R}^{n_x}\times\Omega\mapsto\mathbb{R}^{m_y}$ is a measurable function.

There is a vast literature devoted to properties of the two-stage problem (1.1)-(1.2) and to solution methods (see [7, 11] and the references therein). It is usually assumed that W is a deterministic matrix and

$$b(x,\omega)=h(\omega)-T(\omega)x. \tag{1.3}$$

For example, $h(\omega)$ may be interpreted as a random demand/supply and $T(\omega)$ as a certain "technology matrix" associated with the first stage decisions. Then $b(x,\omega)$ is the discrepancy between the technology input/output requirements and the demand/supply observed, and some corrective action y has to be undertaken to account for this discrepancy.

However, in some long-term planning problems in a highly uncertain environment, it is the data referring to the future that are random. For example, in long-term investment planning, where x denotes the investment decisions to be made now, while y represents future actions, the costs q and the technological characteristics W of the future investments are usually uncertain. Moreover, new technologies may appear that may increase our recourse capabilitites. Therefore we focus on the *random recourse* case in a generalized sense, i.e. a situation when besides W and q also the number of columns of W is random.

Next, our model allows much more general relations between the first stage variables and the second stage problem than the linear relation (1.3). In (1.2) we allow, for example, nonlinear and random technologies $T(x, \omega)$; moreover, the supply/demand vector may be dependent on both x and ω. Apart from a broader class of potential applications, such a model appears to be interesting in its own right.

The fundamental question that will be analysed in this paper is the problem of approximation. Namely, given a sample $s = \{s_i\}_{i=1}^{\infty} \in \Omega^{\infty} = \Omega^{\mathbb{N}}$, we consider for $n \in \mathbb{N}$ the *empirical measures*

$$P_n(s) = \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i}, \tag{1.4}$$

where δ_{s_i} denotes point mass at s_i. An empirical measure can be employed to approximate the expected recourse function

$$F(x) = \int f(x, \omega) P(d\omega) \tag{1.5}$$

by the empirical mean

$$F_n(x) = \int f(x, \omega) P_n(s)(d\omega) = \frac{1}{n} \sum_{i=1}^{n} f(x, s_i). \tag{1.6}$$

The main question is the following: *can uniform convergence of F_n to F take place for almost all s (with respect to the product probability P^{∞} on Ω^{∞})?* We shall show that a positive answer to this question can be given for a very broad class of functions $b(x, \omega)$ in (1.2). To this end we shall use some results on the Glivenko-Cantelli problem developed in [9, 25, 26].

Compared with related contributions to the stability of two-stage stochastic programs, the scope of the present paper is novel in two respects: we allow recourse matrices with random entries and random size, and we are able to treat discontinuous and non-convex integrands in the expected recourse function. The tools from probability theory that we use here lead to uniform convergence. The approaches in [5, 10, 18] utilize milder types of convergence (such as epigraphical convergence), and hence they can handle extended-real-valued functions. As in the present paper, the accent in [14] is on convergence of expected recourse functions in the context of empirical measures. The authors obtain consisitency results that cover convex stochastic programs with a fixed recourse matrix W.

Perturbations going beyond empirical measures are studied in [10, 18] for fixed-recourse problems with continuous integrands. Stochastic programs with discontinuous integrands are treated in [1, 21] and in [22], which contains a section on estimation via empirical measures in problems with mixed integer recourse. Further related work concerns various quantitative aspects for stochastic programs involving empirical measures, such as [5, 6, 12, 13, 19, 23, 24]. Because of that, the settings in these papers are more specific than here.

Let us finally mention that the probabilistic analysis of combinatorial optimization problems is another field in mathematical programming, where results developed in the context of the Glivenko-Cantelli problem can be utilized (see, e.g., [8, 15, 16]).

2 The Glivenko-Cantelli problem

Before passing to the main object of our study, we briefly restate the main definitions and results regarding the general Glivenko-Cantelli problem that will be used later. The probability measure P is assumed to be fixed.

Definition 2.1. A class of integrable functions $\varphi_x : \Omega \mapsto \mathbb{R}$, $x \in X$, is called a *P-uniformity class* if

$$\lim_{n\to\infty} \sup_{x\in X} \left| \int \varphi_x(\omega) P(d\omega) - \int \varphi_x(\omega) P_n(s)(d\omega) \right| = 0$$

for P^∞-almost all s.

So, our problem of uniform convergence of (1.6) to (1.5) can be reformulated as the problem of determining whether the family of functions $\omega \mapsto f(x,\omega)$, $x \in X$, is a P-uniformity class.

From now on, having in mind application to stochastic programming, we shall restrict our attention to functions which are measurable with respect to *both* arguments (x,ω). This will allow us to avoid serious technical difficulties associated with non-measurability of sets defined with the use of the existence quantifier.

Following [25], with the simplification mentioned above, we introduce the following definition.

Definition 2.2. Let $\varphi : X \times \Omega \mapsto \mathbb{R}$ be measurable in both arguments. The class of functions $\omega \mapsto \varphi(x,\omega)$, $x \in X$, is called *P-stable* if for each $\alpha < \beta$ and each set $A \in \Sigma$ with $P(A) > 0$ there exists $n > 0$ such that

$$\begin{aligned} &P\{(s_1,\ldots,s_n,t_1,\ldots,t_n) : (\exists x \in X) \\ &\varphi(x,s_i) < \alpha,\ \varphi(x,t_i) > \beta,\ i = 1,\ldots,n\ \} < (P(A))^{2n}. \end{aligned}$$

The main result of [25] reads.

Theorem 2.3. ([25], Theorem 2). *Assume that the function $\varphi(x,\omega) : X \times \Omega \mapsto \mathbb{R}$ is measurable in both arguments. Then the following statements are equivalent:*

(a) the class of functions $\omega \mapsto \varphi(x,\omega)$, $x \in X$, is a P-uniformity class and $\int \varphi(x,\omega)P(d\omega)$, $x \in X$, is bounded;

(b) the class of functions $\omega \mapsto \varphi(x,\omega)$, $x \in X$, is P-stable and there exists v with $\int v(\omega)P(d\omega) < \infty$ such that, for all $x \in X$, $|\varphi(x,\omega)| \leq v(\omega)$ a.s.

Since we shall use this result arguing by contradiction, it is convenient to restate the definition of stability.

Remark 2.4. ([25], Proposition 4). *Let $\varphi : X \times \Omega \mapsto \mathbb{R}$ be measurable in both arguments. The class of functions $\omega \mapsto \varphi(x,\omega)$, $x \in X$, fails to be P-stable if and only if there exist $\alpha < \beta$ and $A \in \Sigma$ with $P(A) > 0$ such that for each $n \in \mathbb{N}$ and almost each $(s_1, \dots, s_n) \in A^n$, for each subset I of $\{1, \dots, n\}$ there is $x \in X$ with*

$$\varphi(x, s_i) < \alpha \quad \text{for} \quad i \in I$$

and

$$\varphi(x, s_i) > \beta \quad \text{for} \quad i \notin I.$$

Stability conditions turn out to be a rather powerful tool for proving various laws of large numbers. As an example, we can consider one of the basic results in the theory of uniform convergence (see, e.g., [20])

Theorem 2.5. *Let $b(x,\omega)$ be jointly measurable on $X \times \Omega$, where X is a compact metric space and $(\Omega, \mathcal{B}, P)$ is a probability space. If $x \mapsto b(x,\omega)$ is continuous for almost all ω and there is an integrable function $g(\omega)$ such that*

$$\sup_{x \in X} |b(x,\omega)| \leq g(\omega) \quad \text{a. s.},$$

then

$$\sup_{x \in X} \left| \int b(x,\omega) P_n(s)(d\omega) - \int b(x,\omega) P(d\omega) \right| \to 0 \quad \text{a. s.}$$

For the direct proof of this result, see [20]. Alternatively, one may use the argument that the family of functions $\omega \mapsto b(x,\omega)$, $x \in X$, is P-stable. In fact, owing to the compactness of X, for each $\epsilon > 0$ there is a finite number of open sets W_i covering X, such that

$$\int \left[\sup_{y \in W_i} b(y,\omega) - \inf_{y \in W_i} b(y,\omega) \right] P(d\omega) < \epsilon$$

for all i. This, however, implies the validity of the Blum-DeHardt conditions for uniformity, which - in turn - entail the stability of the family $\omega \mapsto b(x,\omega)$, $x \in X$ (see [25], p. 839).

Let us use the stability condition to prove some technical lemmas, which will be useful for further considerations.

Lemma 2.6. *Assume that $f : X \times \Omega \mapsto \mathbb{R}$ is measurable in both arguments and the class of functions $\omega \mapsto f(x,\omega)$, $x \in X$, $f(x,\cdot)$, $x \in X$, is P-stable. Then for every measurable function $g : \Omega \mapsto \mathbb{R}$ the class of functions $\omega \mapsto g(\omega)f(x,\omega)$, $x \in X$, is P-stable.*

Proof. Let us use Remark 2.4. Suppose that the set of functions $h(x,\cdot) = g(\cdot)f(x,\cdot)$, $x \in X$, is not P-stable. Then there exist $\alpha < \beta$ and $A \in \Sigma$ with $P(A) > 0$ such that for each n and almost each $(s_1, \ldots, s_n) \in A^n$, for each subset I of $\{1, \ldots, n\}$ there is $x \in X$ with

$$h(x, s_i) < \alpha \quad \text{for} \quad i \in I, \tag{2.1}$$

$$h(x, s_i) > \beta \quad \text{for} \quad i \notin I. \tag{2.2}$$

With no loss of generality we can assume that $\alpha > 0$. Define $q = (1+\beta/\alpha)/2$ and consider the sets

$$B_k^+ = \{\omega \in A : \ q^k < g(\omega) < q^{k+1}\}, \ k = \ldots, -2, -1, 0, 1, 2, \ldots$$
$$B_k^- = \{\omega \in A : \ -q^k < g(\omega) < -q^{k+1}\}, \ k = \ldots, -2, -1, 0, 1, 2, \ldots \ .$$

At least one of them has a positive probability. Let it be B_k^+ for some k (the proof in the case of B_k^- is similar). Since $B_k^+ \subset A$ and $P(B_k^+) > 0$, for almost all $(s_1, \ldots, s_n) \in (B_k^+)^n$ and all possible I, inequalities (2.1) and (2.2) hold. If $i \in I$ then

$$f(x, s_i) < \frac{\alpha}{q^k} = \alpha'.$$

If $i \notin I$ then

$$f(x, s_i) > \frac{\beta}{q^{k+1}} = \beta'.$$

Since $\beta' - \alpha' = (\beta - \alpha)/(2q^{k+1}) > 0$, conditions of Remark 2.4 hold for the family $f(x,\cdot)$, $x \in X$. But then this family cannot be P-stable, a contradiction. □

Lemma 2.7. *Assume that the following conditions are satisfied:*

(i) *the functions $f : X \times \Omega \mapsto \mathbb{R}$ and $g : X \times \Omega \mapsto \mathbb{R}$ are measurable in both arguments;*

(ii) *the families of functions $\omega \mapsto f(x,\omega)$, $x \in X$, and $\omega \mapsto g(x,\omega)$, $x \in X$, are P-uniformity classes;*

(iii) *the expectations $\int f(x,\omega)P(d\omega)$ and $\int g(x,\omega)P(d\omega)$ are bounded for $x \in X$.*

Then the family of functions

$$\omega \mapsto \max\left[f(x,\omega), g(x,\omega)\right], \ x \in X,$$

is a P-uniformity class and there exists $v \in \mathcal{L}^1(\Omega, P)$ such that $|\max\left[f(x,\omega), g(x,\omega)\right]| \le v(\omega)$ a.s..

Proof. At first let us observe that by Theorem 2.3, in particular, there exists $v \in \mathcal{L}^1(\Omega, P)$ such that $\max[|f(x,\omega)|, |g(x,\omega)|] \leq v(\omega)$ a.s., so our second assertion is true. Let us now pass to the P-uniformity assertion. Directly from Definition 2.1 we see that the set of functions

$$\varphi(x,\cdot) = g(x,\cdot) - f(x,\cdot),\ x \in X,$$

is a P-uniformity class. By Theorem 2.3 it is P-stable. Suppose that the family of functions

$$\varphi^+(x,\cdot) = \max[0, \varphi(x,\cdot)],\ x \in X, \tag{2.3}$$

is not P-stable. Then, by Remark 2.4, there exist $\alpha < \beta$ and $A \in \Sigma$ with $P(A) > 0$ such that for each n and almost each $(s_1, \dots, s_n) \in A^n$, for each subset I of $\{1, \dots, n\}$ there is $x \in X$ with

$$\varphi^+(x, s_i) < \alpha \ \text{ for } \ i \in I$$

and

$$\varphi^+(x, s_i) > \beta \ \text{ for } \ i \notin I.$$

Since $\varphi^+(x, s_i) \geq 0$, then $\alpha > 0$, hence $\beta > 0$, too. Thus the above inequalities hold with φ^+ replaced by φ. Then, by virtue of Remark 2.4, the class $\varphi(x,\cdot)$, $x \in X$, cannot be P-stable, a contradiction. Consequently, the family (2.3) is P-stable, and, in view of Theorem 2.3, it is a P-uniformity class. Using the representation

$$\max[f(x,\cdot), g(x,\cdot)] = f(x,\cdot) + \varphi^+(x,\cdot),$$

directly from Definition 2.1 we obtain the desired result. □

Lemma 2.8. *The family of functions*

$$\omega \mapsto \varphi(x,\omega) = \Phi(f(\omega) + g(x)),$$

where $f : \Omega \mapsto \mathbb{R}$ is measurable, $g : X \mapsto \mathbb{R}$ and $\Phi : \mathbb{R} \mapsto \mathbb{R}$ is monotone, is P-stable.

Proof. Let us assume that the assertion is false. Then there exist $\alpha < \beta$ and $A \in \Sigma$ with $P(A) > 0$ such that for each n and almost each $(s_1, \dots, s_n) \in A^n$, for each subset I of $\{1, \dots, n\}$ there is $x \in X$ with

$$\varphi(x, s_i) < \alpha \ \text{ for } \ i \in I, \tag{2.4}$$

$$\varphi(x, s_i) > \beta \ \text{ for } \ i \notin I. \tag{2.5}$$

Replacing I with $\{1, \dots, n\} \setminus I$, we also have, for some $y \in X$,

$$\varphi(y, s_i) > \beta \ \text{ for } \ i \in I, \tag{2.6}$$

$$\varphi(y, s_i) < \alpha \ \text{ for } \ i \notin I. \tag{2.7}$$

With no loss of generality we can assume that Φ is nondecreasing. Define $\Phi^{-1}(u) = \sup\{v : \Phi(v) \leq u\}$. From (2.4) we get

$$f(s_i) + g(x) \leq \Phi^{-1}(\Phi((f(s_i) + g(x))) \leq \Phi^{-1}(\alpha),\ i \in I,$$

while (2.6) implies

$$f(s_i) + g(y) > \Phi^{-1}(\beta), \ i \in I.$$

Thus,

$$g(y) - g(x) > \Phi^{-1}(\beta) - \Phi^{-1}(\alpha) \geq 0.$$

Likewise, from (2.5) and (2.7) we obtain

$$g(x) - g(y) > \Phi^{-1}(\beta) - \Phi^{-1}(\alpha) \geq 0,$$

a contradiction. □

3 Approximating the recourse function

Let us now pass to function (1.5) and its approximation (1.6). We shall make the following assumptions.

(A1) There exist a measurable function $\bar{u} : \Omega \mapsto \mathbb{R}^m$ and $c \in \mathcal{L}^2(\Omega, P)$ such that a.s.

$$\bar{u}(\omega) \in \left\{u : \ W(\omega)^T u \leq q(\omega)\right\} \subseteq \{u : \ \|u\| \leq c(\omega)\}.$$

(A2) The function $b : X \times \Omega \mapsto \mathbb{R}^m$ is measurable in both arguments, there exists $v \in \mathcal{L}^2(\Omega, P)$ such that, for all $x \in X$, $\|b(x,\omega)\| \leq v(\omega)$ a.s., and the family of functions $\omega \mapsto b(x,\omega)$, $x \in X$, is a P-uniformity class.

We are now ready to prove the P-uniformity of empirical approximations (1.6).

Theorem 3.1. *Let $f : X \times \Omega \mapsto \mathbb{R}$ be defined by* (1.2) *and let conditions* (A1) *and* (A2) *hold. Then the family of functions $\omega \mapsto f(x,\omega)$, $x \in X$, is a P-uniformity class and there exists $v \in \mathcal{L}^1(\Omega, P)$ such that, for all $x \in X$, $\|f(x,\omega)\| \leq v(\omega)$ a.s..*

Proof. By (A1) we can use duality in linear programming to get

$$f(x,\omega) = \max\left\{b(x,\omega)^T u \mid W(\omega)^T u \leq q(\omega)\right\}. \tag{3.1}$$

The feasible set of the dual program in (3.1) is a.s. a nonempty bounded polyhedron having finitely many vertices. Then every vertex of the dual feasible set can be expressed as

$$u = B(\omega)^{-1} q_B(\omega), \tag{3.2}$$

where B is a square nonsingular submatrix of $W(\omega)$ of dimension m_y (a basis matrix), and $q_B(\omega)$ is the subvector of $q(\omega)$ that corresponds to the columns in the basis matrix.

Let us denote all possible square submatrices of $W(\omega)$ having dimension m_y by $B_k(\omega)$, $k = 1, \ldots, K = \binom{n_y}{m_y}$. A matrix $B_k(\omega)$ is a *feasible basis matrix* if it is nonsingular and

(3.2) (with $B(\omega) = B_k(\omega)$) yields a feasible point. Now, for each $1 \leq k \leq K$, we define the function

$$v_k(\omega) = \begin{cases} B_k(\omega)^{-T} q_{B_k}(\omega) & \text{if } B_k(\omega) \text{ is a feasible basis matrix,} \\ \bar{u}(\omega) & \text{otherwise.} \end{cases}$$

By (A1), $v_k \in \mathcal{L}^2(\Omega, P)$ for all $k = 1, \ldots, K$. From (3.1) we obtain

$$f(x,\omega) = \max_{k=1,\ldots,K} b(x,\omega)^T v_k(\omega). \tag{3.3}$$

By (A2), for each $j = 1, \ldots, m_y$, the expectation $\int b_j(x,\omega)P(d\omega)$ is bounded for $x \in X$. Hence, by Theorem 2.3 and (A2), the class $b_j(x,\cdot)$ is P-stable, and, by Lemma 2.6, the products $b_j(x,\cdot)v_{kj}(\cdot)$ constitute a P-stable class.

Now, for all $x \in X$,

$$|b_j(x,\omega)v_{kj}(\omega)| \leq v(\omega)v_{kj}(\omega) \text{ a.s.,}$$

and $v \cdot v_{kj} \in \mathcal{L}^1(\Omega, P)$. Therefore, by Theorem 2.3, the products $b_j(x,\cdot)v_{kj}(\cdot)$ form a P-uniformity class. Directly from Definition 2.1, $b(x,\cdot)^T v_k(\cdot)$, $x \in X$, is a P-uniformity class, for every $k = 1, \ldots, K$. Using Lemma 2.7, we conclude that (3.3) is a P-uniformity class and that $\int f(x,\omega)P(d\omega)$ is bounded for $x \in X$. Using Theorem 2.3 again we additionally conclude that an integrable bound on $|f(x,\omega)|$ must exist. □

Roughly speaking, the question whether the optimal value of a linear program is a P-uniformity class has been reduced to the substantially simpler question whether the right hand side is a P-uniformity class. The latter can still be analysed via the stability conditions, as it has been done for the continuous case in Theorem 2.5, but our framework can also handle discontinuous functions.

Example

Assume that in (1.2) the right hand side is defined by the operation of rounding to integers,

$$b_i(x,\omega) = \begin{cases} \lceil b_i(\omega) - T_i(x) \rceil & \text{if } b_i(\omega) - T_i(x) \geq 0 \\ \lfloor b_i(\omega) - T_i(x) \rfloor & \text{if } b_i(\omega) - T_i(x) \leq 0 \end{cases}, \quad i = 1, \ldots, m,$$

where $\lceil a \rceil = \min\{n \in \mathbb{Z} : n \geq a\}$, while $\lfloor a \rfloor = \max\{n \in \mathbb{Z} : n \leq a\}$. If $T(x)$ and $b(\omega)$ are measurable, then, by Lemma 2.8, the family $\omega \mapsto b(x,\omega)$, $x \in X$, is P-stable. Thus, under mild integrability assumptions, $b(x,\omega)$ satisfies condition (A2). Let us point out that the functions $b_i(\cdot,\omega)$ are not even lower semicontinuous here.

4 Problems with random size

Let us now consider the case when $f(x,\omega)$ is the optimal value of the infinite linear programming problem:

$$\min \sum_{i=1}^{\infty} q_i(\omega) y_i$$
$$\sum_{i=1}^{\infty} w_i(\omega) y_i = b(x,\omega) \tag{4.1}$$
$$y_i \geq 0, \ i = 1, 2, \ldots$$

We assume that the infinite sequence $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots)$ with elements $\xi_i(\omega) = (q_i(\omega), w_i(\omega))$, $i = 1, 2 \ldots$, is a random variable in the space Ξ of sequences of $(m_y + 1)$-dimensional vectors; Ξ is equipped with the σ-algebra $\mathcal{A}$ generated by sets of the form $\{\xi : (\xi_1, \ldots, \xi_k) \in B\}$ for all Borel sets $B \in \mathbb{R}^{(m_y+1)k}$ and all k. We shall denote the optimal value of (4.1) by $f(x,\omega) = \varphi(x, \xi(\omega))$.

Next, we define in Ξ the projection operators $\Pi_k, k = 1, 2, \ldots$ by

$$\Pi_k \xi = (\xi_1, \ldots, \xi_k, 0, 0, \ldots).$$

They are, clearly, measurable. For any $\xi \in \Xi$, let

$$J(\xi) = \inf\{k : \ \Pi_k \xi = \xi\}$$

(we take the convention that $\inf \emptyset = \infty$). We make the following assumptions about the distribution of ξ.

(A3) $P\{J(\xi(\omega)) < \infty\} = 1$;

(A4) for all $k \geq j \geq 1$

$$\mathbb{L}(\Pi_j \xi \mid J(\xi) \leq k) = \mathbb{L}(\Pi_j \xi \mid J(\xi) \leq j),$$

where $\mathbb{L}(\cdot, A)$ denotes the conditional probability law under A.

The following two lemmas provide more insight into the nature of our randomly-sized problem.

Lemma 4.1. *If ξ satisfies conditions* (A3) *and* (A4) *then there exists a random variable z with values in Ξ and such that $P\{z_j = 0\} = 0$, $j = 1, 2, \ldots$, and an integer random variable N, independent on z, such that ξ and $\Pi_N z$ have the same distribution.*

Proof. Let ν_j be the conditional distribution of the first j components of ξ, given that $J(\xi) \geq j$. By (A4), ν_j is the distribution of the first j components of ξ under the condition $J(\xi) \geq k$, for every $k \geq j$. Therefore the sequence $\{\nu_j\}$ constitutes a projective family and by Kolmogorov theorem (cf., e.g., [4], Proposition 62.3) there exists a probability measure ν with marginals ν_j.

Let π be the distribution of $J(\xi)$. Consider the pair (z, N) such that $z \in \Xi$ has distribution ν, the integer N has distribution π, and they are mutually independent. Define $\xi' = \Pi_N z$. We shall show that ξ' has the same distribution as ξ. It is sufficient to show that, for each j, $(\xi_1, \ldots, \xi_j)$ and $(\xi_1', \ldots, \xi_j')$ have the same distribution. Since $P\{N = k\} = P\{J(\xi) = k\}$, it suffices to show that

$$\mathbb{L}\{(\xi_1, \ldots, \xi_j) \mid J(\xi) = k\} = \mathbb{L}\{(\xi_1', \ldots, \xi_j') \mid N = k\}.$$

If $k \geq j$, both $(\xi_1, \ldots, \xi_j)$ and $(\xi_1', \ldots, \xi_j')$ have distribution ν_j. If $k < j$, their first k components have distribution ν_k, while the remaining components are zero. □

Lemma 4.2. *Assume* (A1), (A2) *and* (A3). *Then there exists* $v \in \mathcal{L}^1(\Omega, P)$ *such that, for all* $x \in X$, $|f(x,\omega)| \leq v(\omega)$ *a.s..*

Proof. By (A3), with probability 1, $f(x,\omega)$ is defined by the finite dimensional problem

$$f(x,\omega) = \min\{\bar{q}(\omega)^T y \mid \bar{W}(\omega)y = b(x,\omega),\ y \geq 0\},$$

where $\bar{q}(\omega)^T = [q_1(\omega)\ \ldots\ q_{J(\omega)}(\omega)]$ and $\bar{W}(\omega)^T = [w_1(\omega)\ \ldots\ w_{J(\omega)}(\omega)]$. By duality in linear programming,

$$f(x,\omega) = \max\{b(x,\omega)^T u \mid \bar{W}(\omega)^T u \leq \bar{q}(\omega)\}.$$

Our assertion follows from the square integrability of $c(\omega)$ and of the uniform upper bound on $\|b(x,\omega)\|$. □

Let us observe that the above result implies that the expected value $F(x) = \int f(x,\omega)P(d\omega)$ is well-defined and uniformly bounded for $x \in X$.

Lemma 4.3. *The sequence of functions*

$$F^j(x) = E\{\varphi(x, \xi(\omega)) \mid J(\xi(\omega)) \leq j\},\ j = 1, 2, \ldots,$$

is monotonically decreasing.

Proof. Removing columns from a linear program may only increase its optimal value, so, for every j and every $\xi \in \Xi$,

$$\varphi(x, \Pi_j \xi) \geq \varphi(x, \xi).$$

Therefore,

$$F^{j+1}(x) = E\{\varphi(x,\xi) \mid J(\xi) \leq j+1\} \leq E\{\varphi(x, \Pi_j\xi) \mid J(\xi) \leq j+1\}.$$

Next, by (A4),

$$E\{\varphi(x, \Pi_j\xi) \mid J(\xi) \leq j+1\} = E\{\varphi(x,\xi) \mid J(\xi) \leq j\} = F^j(x).$$

Combining the last two relations we obtain the required result. □

5 Approximating the randomly-sized recourse function

Let us now return to our main problem: uniform convergence of empirical approximations (1.6) to the expected recourse function with the recourse problem (4.1).

Theorem 5.1. *Let* $f : X \times \Omega \mapsto \mathbb{R}$ *be defined by* (4.1) *and let conditions* (A1)-(A4) *hold. Then the family of functions* $\omega \mapsto f(x, \omega)$, $x \in X$, *is a P-uniformity class.*

Proof. For the sample $\xi^1, \ldots, \xi^n$ we define

$$I_k = \{1 \leq j \leq n : \; \Pi_k \xi^j = \xi^j\}$$

and denote by n_k the number of elements in I_k. Then we can rewrite (1.6) as

$$F_n(x) = \sum_{k=1}^{\infty} \frac{n_k}{n} \left(\frac{1}{n_k} \sum_{i \in I_k} \varphi(x, \xi^i) \right) = S_n^{1,l}(x) + S_n^{l+1,\infty}(x), \tag{5.1}$$

where

$$S_n^{m,l}(x) = \sum_{k=m}^{l} \frac{n_k}{n} \left(\frac{1}{n_k} \sum_{i \in I_k} \varphi(x, \xi^i) \right). \tag{5.2}$$

Let us consider $S_n^{1,l}$. For every k the collection $\{\xi^i, \; i \in I_k\}$ constitutes a sample of independent observations drawn from the conditional distribution ν_k (under the condition $\Pi_k \xi = \xi$). By the strong law of large numbers, for each $k \leq l$,

$$\lim_{n \to \infty} \frac{n_k}{n} = P\{\Pi_k \xi = \xi\} = p_k, \text{ a. s.,}$$

where $p_k = P\{J(\xi) = k\}$. If $p_k > 0$ then $n_k \to \infty$ a. s. and by Theorem 3.1

$$\frac{1}{n_k} \sum_{i \in I_k} \varphi(x, \xi^i) \to F_k(x), \text{ a.s.,}$$

uniformly for $x \in X$. So, with probability 1, for every $\epsilon > 0$ we can find $N_1(l, \epsilon)$ such that for all $n > N_1(l, \epsilon)$

$$\sup_{x \in X} \left| S_n^{1,l}(x) - \sum_{k=1}^{l} p_k F_k(x) \right| < \epsilon. \tag{5.3}$$

We shall now estimate $S_n^{l,\infty}(x)$. Let us choose $k_0 \leq l$ and consider the random variables

$$\eta^i = \Pi_{k_0} \xi^i, \; i \in \bigcup_{k>l} I_k.$$

Removing columns may only increase the optimal value of (4.1), so $\varphi(x, \xi^i) \leq \varphi(x, \eta^i)$. Thus

$$S_n^{l+1,\infty}(x) = \frac{1}{n} \sum_{k>l} \sum_{i \in I_k} \varphi(x, \xi^i) \leq \frac{1}{n} \sum_{k>l} \sum_{i \in I_k} \varphi(x, \eta^i) = \frac{n_{l+1,\infty}}{n} \frac{1}{n_{l+1,\infty}} \sum_{k>l} \sum_{i \in I_k} \varphi(x, \eta^i), \tag{5.4}$$

where

$$n_{l+1,\infty} = \sum_{k>l} n_k.$$

Again, by the strong law of large numbers,

$$\lim_{n\to\infty} \frac{n_{l+1,\infty}}{n} = \sum_{k>l} p_k \text{ a.s..} \tag{5.5}$$

Next, by (A4) the variables η^i, $i \in \bigcup_{k>l} I_k$, constitute a sample of i.i.d. observations drawn from the conditional distribution ν_{k_0}. Thus, by Theorem 3.1,

$$\lim_{n\to\infty} \frac{1}{n_{l+1,\infty}} \sum_{k>l} \sum_{i\in I_k} \varphi(x,\eta^i) = F_{k_0}(x), \text{ a.s.,} \tag{5.6}$$

uniformly for $x \in X$. Putting together (5.4), (5.5) and (5.6) we can conclude that a.s. we can find $N_2(l,\epsilon)$ such that for all $n > N_2(l,\epsilon)$ and all $x \in X$

$$S_n^{l+1,\infty}(x) \leq \left(\sum_{k>l} p_k\right) |F_{k_0}(x)| + \epsilon. \tag{5.7}$$

On the other hand, by (A1) and the duality in linear programming,

$$\varphi(x,\xi(\omega)) \geq b(x,\omega)^T \bar{u}(\omega).$$

Therefore,

$$\begin{aligned} S_n^{l+1,\infty}(x) &= \frac{1}{n} \sum_{k>l} \sum_{i\in I_k} \varphi(x,\xi^i) \\ &\geq \frac{1}{n} \sum_{k>l} \sum_{i\in I_k} (b^i(x))^T \bar{u}^i \\ &= \frac{n_{l+1,\infty}}{n} \frac{1}{n_{l+1,\infty}} \sum_{k>l} \sum_{i\in I_k} (b^i(x))^T \bar{u}^i \\ &\geq -\frac{n_{l+1,\infty}}{2n} \frac{1}{n_{l+1,\infty}} \sum_{k>l} \sum_{i\in I_k} \left(\|b^i(x)\|^2 + \|\bar{u}^i\|^2\right), \end{aligned} \tag{5.8}$$

where $b^i(x)$ and $\bar{u}^i$ are i.i.d. observations drawn from the distributions of $b(x,\omega)$ and $\bar{u}(\omega)$. By (A2),for all x one has $\|b^i(x)\|^2 \leq (v_i)^2$, where v_i are i.i.d. observations from the upper bound v. Consequently, by the law of large numbers,

$$\frac{1}{n_{l+1,\infty}} \sum_{k>l} \sum_{i\in I_k} \left((v_i)^2 + \|\bar{u}^i\|^2\right) \to E\left\{v^2 + \|\bar{u}\|^2\right\}.$$

Using this relation in (5.8), with a look at (5.5), we conclude that a.s. there is $N_3((l,\epsilon)$ such that for all $n > N_3(l,\epsilon)$ and all x one has

$$S_n^{l+1,\infty}(x) \geq -\frac{1}{2}\left(\sum_{k>l(\epsilon)} p_k\right) E\left\{v^2 + \|\bar{u}\|^2\right\} - \epsilon. \tag{5.9}$$

We can always choose $l(\epsilon)$ so large that for all $x \in X$,

$$\left| \sum_{k>l(\epsilon)} p_k F_k(x) \right| \leq \left(\sum_{k>l(\epsilon)} p_k \right) |F_{k_0}(x)| \leq \epsilon \tag{5.10}$$

and

$$\frac{1}{2} \left(\sum_{k>l(\epsilon)} p_k \right) E\left\{ v^2 + \|\bar{u}\|^2 \right\} \leq \epsilon. \tag{5.11}$$

Then, by (5.1), (5.3), (5.7), (5.9), (5.10) and (5.11), for each $\epsilon > 0$, a.s. there exists $N(\epsilon)$ such that for all $n > N(\epsilon)$,

$$\begin{aligned} \sup_{x\in X} |F^n(x) - F(x)| &\leq \sup_{x\in X} \left| S_n^{1,l(\epsilon)}(x) - \sum_{k=1}^{l(\epsilon)} p_k F_k(x) \right| + \sup_{x\in X} \left| S_n^{l(\epsilon)+1,\infty}(x) \right| + \sup_{x\in X} \left| \sum_{k>l(\epsilon)} p_k F_k(x) \right| \\ &\leq 4\epsilon, \end{aligned}$$

which completes the proof. □

6 Concluding remarks

From the stability theory of general optimization problems it is well-known that uniform convergence of perturbed objective functions can be used as a key ingredient to establish continuity properties of perturbed optimal values and optimal solutions.

Let us assume that F in (1.5) appears in the objective of an optimization problem and that we are interested in asymptotic properties of optimal values and optimal solutions, when F is replaced by the estimates F_n (cf. (1.6)). Assume further that F and F_n ($n \in \mathbb{N}$) are lower semicontinuous and that the optimization problem involving F has a non-empty bounded complete local minimizing set in the sense of [17]. The latter means, roughly speaking, that there is a bounded set of local minimizers which, in some sense, contains all the nearby local minimizers. Both strict local and global minimizers can be treated within this framework (see [17]). Using standard arguments from the stability of optimization problems it is then possible to show that (with probability 1) the optimal values and the optimal solutions are continuous and upper semicontinuous, respectively, as $n \to \infty$ (see, e.g., [22]).

Let us also mention that one possibility to guarantee the boundedness of solution sets is to impose some growth conditions on F. They can also be used to to re-scale the functions, which may allow obtaining uniform convergence on unbounded sets.

Finally, it has to be stressed that in the context of stability of optimization problems with F appearing in the objective, the framework of uniform convergence is not the only one possible; epigraphical convergence (see [2, 3]) requires less from the sequence F_n and may prove to be more flexible. However, the counterpart to the theory of the Glivenko-Cantelli problem has not yet been developed to such an extent as the uniform convergence case.

References

[1] Z. Artstein and R.J.-B. Wets, "Stability of stochastic programs with possibly discontinuous objective functions", *SIAM Journal on Optimization*, to appear.

[2] H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, Boston, 1984.

[3] H. Attouch and R.J.-B. Wets, "Epigraphical processes: laws of large numbers for random lsc functions", technical report, Dept. Mathematics, University of California, Davis 1991.

[4] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, Walter de Gruyter, Berlin, 1974 (English translation *Probability Theory and Elements of Measure Theory*, Academic Press, London, 1981).

[5] J. Dupačová, R.J.-B. Wets, "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems", *The Annals of Statistics* 16(1988) 1517-1549.

[6] Yu.M. Ermoliev and V.I. Norkin, "Normalized convergence in stochastic optimization", *Annals of Operations Research* 30(1991) 187-198.

[7] Yu. Ermoliev and R.J.-B. Wets, *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin, 1988.

[8] S.A. v.d. Geer and L. Stougie, "On rates of convergence and asymptotic normality in the multiknapsack problem", *Mathematical Programming* 51(1991) 349-358.

[9] E. Giné and J. Zinn, "Some limit theorems for empirical processes", *The Annals of Probability* 12(1984) 837-870.

[10] P. Kall, "On approximations and stability in stochastic programming", in: J. Guddat, H. Jongen, B. Kummer and F. Nožička (eds.), *Parametric Optimization and Related Topics*, Akademie-Verlag, Berlin 1987, pp. 387-407.

[11] P. Kall and S.W. Wallace, *Stochastic Programming*, J. Wiley & Sons, Chichester 1994.

[12] Yu.M. Kaniovski, A. King and R.J.-B. Wets, "Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems", Research Report RC 18752, IBM Research Center, Yorktown Heights 1993.

[13] A. King and R.T. Rockafellar, "Asymptotic theory for solutions in statistical estimation and stochastic programming", *Mathematics of Operations Research* 18(1993) 148-162.

[14] A. King and R.J.-B. Wets, "Epi-consistency of convex stochastic programs", *Stochastics and Stochastic Reports* 34(1991) 83-92.

[15] N. Piersma, *Combinatorial Optimization and Empirical Processes*, PhD Dissertation, University of Amsterdam, Tinbergen Institute Research Series 52, 1993.

[16] W.T. Rhee and M. Talagrand, "A concentration inequality for the k-median problem, *Mathematics of Operations Research* 14(1989) 189-202.

[17] S.M. Robinson, "Local epi-continuity and local optimization", *Mathematical Programming* 37(1987) 208-222.

[18] S.M. Robinson and R.J.-B. Wets, "Stability in two-stage stochastic programming", *SIAM Journal on Control and Optimization* 25(1987) 1409-1416.

[19] W. Römisch and R. Schultz, "Lipschitz stability for stochastic programs with complete recourse", *SIAM Journal on Optimization*, to appear.

[20] R.Y Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley and Sons, New York 1992.

[21] R. Schultz, "On structure and stability in stochastic programs with random technology matrix and complete integer recourse", *Mathematical Programming*, to appear.

[22] R. Schultz, "Rates of convergence in stochastic programs with complete integer recourse", *SIAM Journal on Optimization*, submitted.

[23] A. Shapiro, "Asymptotic properties of statistical estimators in stochastic programming", *The Annals of Statistics* 17(1989) 841-858.

[24] A. Shapiro, "Asymptotic behavior of optimal solutions in stochastic programming", *Mathematics of Operations Research* 18(1993) 829-845.

[25] M. Talagrand, "The Glivenko-Cantelli problem", *The Annals of Probability* 15(1987) 837-870.

[26] V.N. Vapnik and A.Y. Červonenkis, "Necessary and sufficient conditions for the uniform convergence of means to their expectations", *Theory of Probability and Applications* 26(1981) 532-553.