Working Paper

Metric Entropy and Nonasymptotic Confidence Bands in Stochastic Programming

Georg Ch. Pflug

WP-96-034
April 1996

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

Metric Entropy and Nonasymptotic Confidence Bands in Stochastic Programming

Georg Ch. Pflug

WP-96-034
April 1996

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

Talagrand has demonstrated in his key paper, how the metric entropy of a class of functions relates to uniform bounds for the law of large numbers. This paper shows how to calculate the metric entropy of classes of functions which appear in stochastic optimization problems. As a consequence of these results, we derive via variational inequalities confidence bands for the solutions, which are valid for any sample size. In particular, the linear recourse problem is considered.

1 Introduction

Consider a stochastic program of the expectation type:

$$\min_{x \in X} \left[c(x) + \int f(x,\omega)\, P(d\omega) \right], \tag{1}$$

where $c(x)$ denote the fixed costs and $f(x,\omega)$ the uncertain costs. The feasible set X is a subset of $\mathbb{R}^d$.

For solving (1), one uses typically the empirical approximation by sampling an i.i.d. sample $s = \{s_i\}_{i=1}^n$ from P and considering

$$\min_{x \in X} \left[c^T x + \int f(x,\omega)\, P_n(s)(d\omega) \right], \tag{2}$$

where $P_n(s)$ is the empirical measure

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

The natural question of measuring the approximation quality of the emprical approximation

$$F_n(x) = \int f(x,\omega)\, P_n(s)(d\omega) = \frac{1}{n} \sum_{i=1}^n f(x, s_i) \tag{3}$$

(and its argmin) to the "true" function

$$F(x) = \int f(x,\omega)\, P(d\omega)$$

(and its argmin) has been addressed by many authors. Almost sure epi-convergence and uniform convergence of F_n to F was proved under various assumptions (see [1], [6]). In [5], Pflug proved the convergence in distribution of

$$\sqrt{n}(F_n(x^* + t/\sqrt{n}) - F(x^* + t/\sqrt{n}))$$

to a Gaussian process in regular situations. Here x^* is the unique minimizer of F. Results of this type lead to confidence results of the following type: Let X_n a minimizer of the empirical program (2). Then

$$\lim_{n \to \infty} P\{\sqrt{n}\|X_n - x^*\| > M\} \le K_1 \exp(-K_2 M^2), \tag{4}$$

where K_1 and K_2 are constants depending on the limiting normal distribution. The practical use of (4) is very limited, since it is valid only for large (and often extremely large) n.

The aim of this paper is to discuss nonasymptotic confidence bounds which are valid for all n and therefore applicable for any sample size. The main keys for deriving such bounds are Talagrand's inequality and variational inequalities.

Talgrand's inequality gives a bound for

$$P\{\sup_{x\in X} \sqrt{n}|F_n(x) - F(x)| > M\} \tag{5}$$

and also for

$$P\{\sup_{x,y\in X} \sqrt{n}|F_n(x) - F(x) - F_n(y) + F(y)| > M\|x-y\|\}. \tag{6}$$

By a variational inequality, a bound for

$$P\{\sqrt{n}\|X_n - x^*\| > M\} \tag{7}$$

can be derived which is true for all n.

The paper is organized as follows: In the section 2, we will show how (7) can be inferred from (5) or (6). Section 3 discusses the key inequality due to Talagrand. In section 4, we apply the result to the linear recourse problem where

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

Finally, a comparison to large deviation results is made in section 5.

2 Variational inequalities and confidence bounds

Let F be the l.s.c. objective function and $\tilde{F}$ some approximant of F. Variational inequalities deal with the question how the approximation error between F and $\tilde{F}$ relates to the approximation error between argmin (F) and argmin $(\tilde{F})$.

Suppose that F fulfills the following growth condition

$$F(x) \geq \inf_y F(y) + c \cdot [\text{dist}(x, \text{ argmin } F)]^\gamma. \tag{8}$$

Lemma 1.

If

$$\sup_x |F(x) - \tilde{F}(x)| \leq \epsilon \tag{9}$$

then for each minimizer $\tilde{x}^*$ of $\tilde{F}$

$$\text{dist}(\tilde{x}^*, \text{ argmin } F) \leq \left[\frac{2\epsilon}{c}\right]^{1/\gamma}. \tag{10}$$

If however for all x, y

$$|F(x) - \tilde{F}(x) - F(y) + \tilde{F}(y)| \leq \epsilon\,\|x-y\|, \tag{11}$$

then for each minimizer $\tilde{x}^*$ of $\tilde{F}$

$$\text{dist}(\tilde{x}^*,\ \text{argmin } F) \le \left[\frac{\epsilon}{c}\right]^{1/(\gamma-1)} \tag{12}$$

(see Shapiro (1994)).

Proof. Let $x^* \in$ argmin F such that $\|\tilde{x}^* - x^*\| = \text{dist}(\tilde{x}^*,\ \text{argmin } F)$. Then, if (9) is fulfilled,

$$\begin{aligned} 0 \ge \tilde{F}(\tilde{x}^*) - \tilde{F}(x^*) &\ge F(\tilde{x}^*) - F(x^*) - |F(x^*) - \tilde{F}(x^*)| - |F(\tilde{x}^*) - \tilde{F}(\tilde{x}^*)| \\ &\ge c\,\|x^* - \tilde{x}^*\|^\gamma - 2\epsilon \end{aligned}$$

whence

$$\|x^* - \tilde{x}^*\| \le \left[\frac{2\epsilon}{c}\right]^{1/\gamma}. \tag{13}$$

If however (11) is true, then

$$\begin{aligned} 0 \ge \tilde{F}(\tilde{x}^*) - \tilde{F}(x^*) &\ge F(\tilde{x}^*) - F(x^*) - |F(x^*) - \tilde{F}(x^*) - F(\tilde{x}^*) + \tilde{F}(\tilde{x}^*)| \\ &\ge c\,\|x^* - \tilde{x}^*\|^\gamma - \epsilon\,\|x^* - \tilde{x}^*\| \end{aligned}$$

whence

$$\|x^* - \tilde{x}^*\| \le \left[\frac{\epsilon}{c}\right]^{1/(\gamma-1)}. \tag{14}$$

□

Variational inequalities build the bridge between confidence bounds for the objective function and confidence bounds for the minimizers: Suppose we may establish that the empirical approximation F_n of the true objective function F satisfies

$$P(\sup_{x\in X} |F_n(x) - F(x)| \le \frac{M}{\sqrt{n}}) \ge 1 - \alpha \tag{15}$$

for all n. If $F(x)$ fulfills the growth condition (8) with $\gamma = 2$ and if it has a unique minimizer x^*, then for each $X_n^* \in$ argmin F_n we have by (10)

$$P(\ \|X_n^* - x^*\| \le \frac{\sqrt{2M}}{\sqrt[4]{n}\sqrt{c}}) \ge 1 - \alpha. \tag{16}$$

for all n. For establishing the sharper bound, recall the definition of L^1-differentiability. The mapping $x \mapsto f(x,\omega)$ is called L^1-differentiable, if there is a vector of functions $\nabla_x f(x,\omega)$ such that

$$\lim_{y\to x} \frac{1}{\|y-x\|} \int |f(y,\omega) - f(x,\omega) - (y-x)^T \nabla_x f(x,\omega)|\ P(d\omega) = 0. \tag{17}$$

If $\nabla_x f(x,\omega)$ is the L^1-derivative of $f(x,\omega)$, then $F(x)$ is differentiable and

$$\nabla_x F(x) = \int \nabla_x f(x,\omega)\ P(d\omega).$$

Set $\nabla_x F_n(x) = \frac{1}{n}\sum_{i=1}^n \nabla_x f(x, s_i)$; compare (3). If X is convex, then

$$\sup_{y \neq x; x, y \in X} \frac{1}{\|y - x\|} |F_n(x) - F_n(y) - F(x) + F(y)| \leq \sup_{x \in X} |\nabla_x F_n(x) - \nabla_x F(x)| \quad (18)$$

Therefore a bound of the form

$$P(\sup_{x \in X} |\nabla_x F_n(x) - \nabla_x F(x)| \leq \frac{M}{\sqrt{n}}) \geq 1 - \alpha \quad (19)$$

implies the sharper bound

$$P(\|X_n^* - x^*\| \leq \frac{M}{c\sqrt{n}}) \geq 1 - \alpha. \quad (20)$$

3 Metric entropy

The notion of metric entropy plays an important role in topology, functional analysis and probability:

Definition 1. A set $A \subset \mathbb{R}^d$ is said to be of covering type (v, V), if for every $\epsilon > 0$ one can find at most $N_\epsilon = \lfloor (V/\epsilon)^v \rfloor$ balls $S_1, S_2, \ldots, S_{N_\epsilon}$, each with diameter ϵ, which cover A, i.e. $A \subseteq \bigcup_{i=1}^{N_\epsilon} S_i$.

Example. The unit cube in $\mathbb{R}^d$ is of covering type $(d, 2\sqrt{d})$.

Definition 2. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A famililiy $\mathcal{F}$ of $L_2(P)$-functions is called of covering type (v, V), if for every $\epsilon > 0$ there are at most $N_\epsilon = \lfloor (V/\epsilon)^v \rfloor$ pairs of functions $(g_1, h_2), \ldots, (g_{N_\epsilon}, h_{N_\epsilon})$ with the properties

(i) $g_i(\omega) \leq h_i(\omega)$ for $1 \leq i \leq N_\epsilon$;

(ii) $\int (h_i(\omega) - g_i(\omega))^2 \, P(d\omega) \leq \epsilon^2$;

(iii) For each $f \in \mathcal{F}$ there is a index $i \in \{1, \ldots, N_\epsilon\}$ such that

$$g_i(\omega) \leq f(\omega) \leq h_i(\omega).$$

Property (iii) may be expressed in the following way:

$$\mathcal{F} \subseteq \cup_{i=1}^{N_\epsilon} [g_i, h_i]$$

where $[g_i, h_i]$ denotes the interval of functions lying between g_i and h_i.

The covering type is essential for uniform confidence bands as was demonstrated by Talagrand (1994):

Theorem 1. Let $|f(w)| \leq C_0$ for all $f \in \mathcal{F}$. Suppose that $\mathcal{F}$ is countable and of covering type (v, V). Then

$$P\left\{\sup_{f\in\mathcal{F}}\left|\int f(\omega)\, P(d\omega) - \int f(\omega)\, P_n(s)(d\omega)\right| \geq \frac{M}{\sqrt{n}}\right\}$$
$$\leq \left(K(V/2C_0)\frac{M}{2C_0\sqrt{v}}\right)^v \exp(-M^2/2C_0). \tag{21}$$

where $K(\cdot)$ is a universal function.

The assumption that $\mathcal{F}$ is countable is not crucial, it only ensures the measurability of the supremum.

In our applications, the class $\mathcal{F}$ is is a parametric family of functions depending smoothly on a parameter $x \in X \subset \mathbb{R}^k$

$$\mathcal{F}_X = \{f(x, \omega) : x \in X \subset \mathbb{R}^n\}.$$

Introduce the following rather weak assumption:

Assumption A1.

(i) $x \mapsto f(x, \omega)$ is lower semicontinuous for every ω;

(ii) $x \mapsto \int f(x, \omega)\; P(d\omega)$ is continuous.

This assumption guarantees that for each closed ball B in $\mathbb{R}^d$ $\omega \mapsto \sup_{x\in B} f(x, \omega)$ is measurable. This can be seen as follows: Let Q^d be the set of rationals in R^d. By Lemma 6 of the appendix, we may represent f as the monotone limit of a sequence $(f^{(k)})$ of continuous functions; $f(x,\omega) =\uparrow \lim_k f^{(k)}(x,\omega)$. Obviously, for each k, the function

$$\omega \mapsto \sup_{x\in B} f^{(k)}(x, \omega) = \sup_{x\in B\cap Q^d} f^{(k)}(x, \omega)$$

is measurable. By Lemma 7 of the Appendix,

$$\sup_{x\in B} f^{(k)}(x, \omega) \uparrow \sup_{x\in B} f(x, \omega),$$

which shows that the latter functions is also measurable.

Lemma 2. Suppose that X is the closure of $X \cap Q^d$. Under assumption A1, the function $s \mapsto \sup_{x\in X} |\int f(x, \omega)\, P_n(s)(d\omega) - \int f(x, \omega)\, P(d\omega)|$ is measurable. Thus the supremum in (21) may be taken over the uncountable set X.

Proof. By continuity,

$$\sup_{x\in X\cap Q^q} \left|\int f^{(k)}(x, \omega)\, P_n(s)(d\omega) - \int f(x, \omega)\, P(d\omega)\right|$$
$$= \sup_{x\in X} \left|\int f^{(k)}(x, \omega)\, P_n(s)(d\omega) - \int f(x, \omega)\, P(d\omega)\right|$$

and it is clear that this function is measurable. Since

$$\begin{aligned}
&\sup_{x\in X} |\int f^{(k)}(x,\omega)\, P_n(s)(d\omega) - \int f(x,\omega)\, P(d\omega)| \\
= \ & \max(\sup_{x\in X}[\int f^{(k)}(x,\omega)\, P_n(s)(d\omega) - \int f(x,\omega)\, P(d\omega)]^+, \\
&\inf_{x\in X} -[\int f^{(k)}(x,\omega)\, P_n(s)(d\omega) - \int f(x,\omega)\, P(d\omega)]^-)
\end{aligned}$$

an application of Lemma 4 of the Appendix implies that

$$\begin{aligned}
&\sup_{x\in X} |\int f(x,\omega)\, P_n(s)(d\omega) - \int f(x,\omega)\, P(d\omega)| \\
= \ & \lim_k \sup_{x\in X} |\int f^{(k)}(x,\omega)\, P_n(s)(d\omega) - \int f^{(k)}(x,\omega)\, P(d\omega)|
\end{aligned}$$

is the limit of measurable functions and hence measurable. □

The aim of this paper is to derive results about covering types of interesting classes of functions

$$\mathcal{F}_X = \{f(x,\omega) : x \in X\}.$$

In particular, we will relate the covering type of $\mathcal{F}_X$ to the covering type of X. In view of the sharper bound (19) we will also consider the class of all L^1-derivatives of functions from $\mathcal{F}_X$

$$\mathcal{F}_X^{\nabla} = \{\nabla_x f(x,\omega) : x \in X\}.$$

Definition 3. For a ball B in $\mathbb{R}^d$, define the diameter of $\{f(x,\cdot); x \in B\}$ as

$$\text{diam}^2\{f(x,\cdot); x \in B\} = \int \left(\sup_{y\in B} f(y,\omega) - \inf_{y\in B} f(y,\omega)\right)^2 P(d\omega).$$

Lemma 3. Suppose that for each ball B in $\mathbb{R}^d$ the following inequality holds

$$\text{diam}^2\{f(x,\cdot); x \in B\} \le C[\text{diam}(B)]^\beta.$$

If X is of covering type (v, V), then $\mathcal{F}_X$ is of covering type $(v/\beta, V^\beta C)$.

Proof. Let $\epsilon = C\,\eta^\beta$. We may cover X by balls $B_1, B_2, \ldots B_{N_\epsilon}$ each of diameter η, where $N_\eta = \lfloor \left(\frac{V}{\eta}\right)^v \rfloor$. The intervals $[\inf_{x\in B_i} f(x,\cdot), sup_{x\in B_i} f(x,\cdot)]$ cover $\mathcal{F}_X$ and have each diameter not more than ϵ. Since $N_\eta = \lfloor \left(\frac{V}{\eta}\right)^v \rfloor = \lfloor \left(\frac{V\,C^{1/\beta}}{\epsilon^{1/\beta}}\right)^v \rfloor = \lfloor \left(\frac{V^\beta}{\epsilon}\right)^{v/\beta} \rfloor$ the Lemma follows. □

Introduce the symbol

$$\partial_S(f(\cdot,\omega)) = \sup_{x\in S} f(x,\omega) - \inf_{x\in S} f(x,\omega)$$

for the variation of $f(\cdot,\omega)$ within S. If $x \mapsto f(x,\omega)$ is Lipschitz continuous with Lipschitz constant $L_S(f(\cdot,\omega)) := \sup_{x,y\in S; x\neq y} \frac{|f(x,\omega)-f(y,\omega)|}{\|x-y\|}$, then trivially

$$\partial_S(f(\cdot,\omega)) \leq \operatorname{diam}(S) L_S(f(\cdot,\omega)). \tag{22}$$

Lemma 4.

(i)

$$\operatorname{diam}^2\{f(x,\cdot) : x \in S\} \leq \operatorname{diam}^2(S) \int L_S^2(f(\cdot,\omega))\, P(d\omega).$$

(ii) For a finite number of random functions $f_1(x,\omega), \ldots, f_K(x,\omega)$

$$\operatorname{diam}^2\{\max_k f_k(x,\cdot) : x \in S\} \leq \operatorname{diam}^2(S) \int \max_k L_S^2(f_k(\cdot,\omega))\, P(d\omega).$$

Proof. (i) follows from (22). For the proof of (ii) notice that

$$\partial_S(\max_k f_k(x)) \leq \max_k \partial_S(f_k). \tag{23}$$

In order to show (23) suppose that $\partial_S(\max_k f_k(x)) = f_i(x^*) - f_j(y^*)$. Then

$$\begin{aligned} \partial_S(\max_k f_k(x)) &= f_i(x^*) - f_j(y^*) \leq f_i(x^*) - f_i(y^*) \\ &\leq \partial_S(f_i) \leq \max_k \partial_S(f_k). \end{aligned}$$

Therefore, the assertion (ii) follows. □

4 An application for linear recourse problems

In this section we consider the linear recourse problem, where the functions $f(x,\omega)$ is of the form

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

We make the following assumption:

Assumption A2.

(i) There exists a measurable function $\tilde{u} : \Omega \to \mathbb{R}^m$ such that

$$\tilde{u}(\omega) \in \left\{W(\omega)^T u \leq q(\omega)\right\} \subseteq \{u : \|u\| \leq C_1\},$$

(ii) The function $b : X \times \Omega \to \mathbb{R}^m$ is differentiable w.r.t x and satisfies $\|b(x,\omega)\| \leq C_0$ a.s. and $\int \sup_{x\in X} \|\nabla b(x,\omega)\|^2 \, P(d\omega) = C_2^2 < \infty$.

Theorem 2.

Let Assumption A2 be fulfilled. If X is of covering type (v, V), then

$$\mathcal{F}_X = \{f(x,\omega) : x \in X\} = \{\min\{q(\omega)^T y | W(\omega)y = b(x,\omega), y \geq 0\} : x \in X\}$$

is of covering type (v, C_1C_2V).

Proof.

By duality, we may write f as the solution of the dual program, i.e. the maximum of a finite number K of functions.

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

(see [6]). Here v_k are the vertices of the dual feasible polyhedron and K is their maximal number. Since $\|b(x,\omega)\| \leq C_0$ and $\|v_k(\omega)\| \leq C_1$, we get

$$|f(x,\omega)| \leq C_0C_1.$$

Moreover, by Lemma 4 (ii),

$$\mathrm{diam}^2\{f(x,\omega)) : x \in S\} \leq \mathrm{diam}^2(S) \int C_1^2 \sup_{x\in X} \|\nabla b(x,\omega)\|^2 \, P(d\omega) \leq \mathrm{diam}^2(S) C_1^2 C_2^2. \tag{24}$$

The assertion follows now from Lemma 3. □

The most important special case is that of a linear $B(x,\omega)$:

$$b(x,\omega) = h(\omega) - T(\omega)x.$$

Since $\|\nabla b(x,\omega)\| = \|T(\omega)\|$, independent of x, the constant C_2 is here simply

$$C_2^2 = \int \|T(\omega)\|^2 \, P(d\omega).$$

5 Entropy of classes of discontinuous functions

Theorem 2 deals with the covering types of classes of Lipschitz continuous functions. However, in view of (19) we are even more interested in classes of derivatives. Since the derivatives of maxima, as the occur in the linear recourse problem are not longer continuous, we will consider now classes of functions having jumps.

To begin with, let $\mathcal{F}_X = \{\mathbb{1}_{\{H(x,\omega)\geq 0\}}, x \in X\}$

Lemma 5. If

(i) $x \mapsto H(x,\omega)$ is Lipschitz continuous for all ω with Lipschitz constant L.

(ii) The random variables $H(x,\omega)$ have densities g_x, which are uniformly bounded by C_1: $g_x(u) \leq C_1$.

and X is of covering type (v, V), then $\mathcal{F}_X$ is of covering type $(2v, \sqrt{LC_1V})$.

Proof. Let B be the ball with center x and radius ϵ. We have

$$\mathbb{1}_{\{\inf_{x\in B} H(x,\omega)>0\}} \leq \mathbb{1}_{\{H(x,\omega)>0\}} \leq \mathbb{1}_{\{\sup_{x\in B} H(x,\omega)>0\}}.$$

Notice that

$$\mathrm{diam}^2\{\mathbb{1}_{\{\sup_{x\in B} H(x,\omega)>0\}} : x \in B\} = \int [\mathbb{1}_{\{\sup_{x\in B} H(x,\omega)>0\}} - \mathbb{1}_{\{\inf_{x\in B} H(x,\omega)>0\}}]^2 \ P(d\omega)$$
$$= P\{\sup_{y\in B} H(y,\omega) > 0 \geq \inf_{y\in B} H(y,\omega)\} \leq P\{|H(x,\omega)| \leq L\cdot\epsilon\}$$
$$\leq \int_{-L\epsilon}^{L\epsilon} g_x(u)\ du \leq 2C_1L\epsilon.$$

An application of Lemma 3 finishes the proof. □

Let us now turn to the covering types of the L^1-derivatives of the functions $f(x,\omega) = \max_k f_k(x,\omega)$. Notice that the L^1-derivative of $\max_k f_k(x,\omega)$ is $\sum_{k=1}^K \nabla_x f_k(x,\omega)\mathbb{1}_{\{f_k(x,\omega)=\max_\ell f_\ell(x,\omega)\}}$. Let us therefore consider the class $\mathcal{F}_X^\nabla = \{\sum_{k=1}^K \nabla_x f_k(x,\omega)\mathbb{1}_{\{f_k(x,\omega)=\max_\ell f_\ell(x,\omega)\}}\}$.

Theorem 3. Suppose that

(i) $x \mapsto \nabla f(x,\omega)$ is Lipschitz continuous with constant L and bounded by C_0,

(ii) The random variables $d_j(x,\omega) = f_j(x,\omega) - \max_{i\neq j} f_i(x,\omega)$ have densities which are bounded by a constant C_1,

(iii) Let $\#\{j : d_j(y,\omega) > 0$ for some y such that $\|x-y\| \leq \epsilon/2 \leq K_1$ if ϵ is sufficiently small.

If X has covering type (v, V), then $\mathcal{F}^\nabla$ has covering type $(2v, C_0\sqrt{C_1KLV})$.

Proof. Let B be the ball with center x and diameter ϵ. Let $A_j(x) = \{\omega : d_j(x,\omega) > 0\}$, $A_j^+ = \{\omega : \sup_{y\in B} d_j(y,\omega) > 0\}$ and $A_j^- = \{\omega : \inf_{y\in B} d_j(y,\omega) > 0\}$. Let $D = \bigcup_j (A_j^+ \setminus A_j^-)$.

Let $h(\omega) = \sum_j \sup_{y\in B} \nabla f_j(y,\omega)\mathbb{1}_{A_j(x)\cap D^c} + C_0\mathbb{1}_D$
and $g(\omega) = \sum_j \inf_{y\in B} \nabla f_j(y,\omega)\mathbb{1}_{A_j(x)\cap D^c} - C_0\mathbb{1}_D$. We have that for all $x \in B$

$$g(\omega) \leq \sum_j \nabla f_j(x,\omega)\mathbb{1}_{A_j(x)} \leq h(\omega).$$

Since $\int [h(\omega) - g(\omega)]^2\ P(d\omega) \leq L^2\epsilon^2 + 2C_0^2C_1KL\epsilon$, we get the desired result. □

6 A comparison to large deviation results

We recall here Sanov's uniform large deviations results: For simplicity, we consider the univariate situation only. Let $\mathcal{P}$ be the class of all probability measures on $\mathbb{R}$. For $Q \in \mathcal{P}$ let $G_Q(u) = Q(-\infty, u]$ its distribution function and $g_Q(u)$ its Lebesgue-density (if existent). Let $T(Q)$ be some functional on $\mathcal{P}$ and

$$K(\Omega_\epsilon, P) = \inf\{\int \log \frac{dQ}{dP} dQ : Q \in \Omega_\epsilon \text{ such that } Q << P\}$$

where

$$\Omega_\epsilon = \{Q \in \mathcal{P} : T(Q) \geq \epsilon\}.$$

Suppose that $Q \mapsto T(Q)$ is uniformly continuous for the distance $\sup_u |G_Q(u) - G_P(u)|$. Sanov's theorem asserts that

$$\lim_{n\to\infty} \frac{-1}{n} \log P\{T_P(P_n) \geq \epsilon\} = K(\Omega_\epsilon, P) \tag{25}$$

for all continuity points ϵ of $K(\Omega_\epsilon, P)$. (Sanov (1957), see Shorack/Wellner (1986), p.792). For an application in our context, suppose that P has Lebesgue-density and that $\mathcal{F}$ is a class of P-integrable functions such that $\sup_{f\in\mathcal{F}} \int |f'(x)|dx < \infty$. Let, for $Q \in \mathcal{P}$

$$T_P(Q) = \sup_{f\in\mathcal{F}} |\int f\, dQ - \int f\, dP| = \sup_{f\in\mathcal{F}} |\int f'(u)G_Q(u)\, du - \int f'(u)G_P(u)\, du|.$$

Then, by Sanov's theorem,

$$\begin{aligned} &\lim_{n\to\infty} \frac{-1}{n} \log P\{\sup_{f\in\mathcal{F}} |\int f(u)\, P_n(du) - \int f(u)\, P(du)| \geq \epsilon\} \\ &= \inf\{\int \log \frac{dQ}{dP} dQ : Q << P \text{ and } \sup_{f\in\mathcal{F}} |\int f\, dP - \int f\, dQ| \geq \epsilon\} \end{aligned} \tag{26}$$

It seems to be difficult to calculate the exact value of the right hand side. However, a bound is easy to find. Suppose that all $f \in \mathcal{F}$ are bounded by C. By the Kullback-Cziszar-Kemperman inequality

$$\|P - Q\|^2 \leq K(Q, P),$$

where $\|P - Q\|$ is the variational distance, (see, for instance Devroye, p. 10), we have

$$\sup_{f\in\mathcal{F}} |\int f\, dP - \int f\, dQ| \leq C\|P - Q\| \leq C\sqrt{K(Q, P)}$$

and therefore

$$\begin{aligned} &\limsup_{n\to\infty} \frac{-1}{n} \log P\{\sup_{f\in\mathcal{F}} |\int f(u)\, P_n(du) - \int f(u)\, P(du)| \geq \epsilon\} \\ &= \inf\{\int \log \frac{dQ}{dP} dQ : Q << P \text{ and } \sup_{f\in\mathcal{F}} |\int f\, dP - \int f\, dQ| \geq \epsilon\} \\ &\geq \inf\{K(Q, P) : C\sqrt{K(Q, P)} \geq \epsilon\} = \frac{\epsilon^2}{C^2}. \end{aligned}$$

Both, Talagrand's inequality (21) and Sanov's limit theorem (25) deal with the probability of deviations from the mean. If we rewrite Sanov's theorem in the form

$$P\{\sup_{f\in\mathcal{F}} |\int f(u)\, P_n(du) - \int f(u)\, P(du)| \geq \epsilon\} = \exp[-n\frac{\epsilon^2}{C^2}(1+o(1))]$$

the relation to Talagrand's inequality becomes apparent: The large deviations result deals with a fixed deviation of ϵ and concerns the tail behavior, whereas Talagrands inequality considers shrinking deviations of size $M/\sqrt{n}$ and focusses on the central behavior. Formally, one may set $\epsilon = M/\sqrt{n}$ to get the same rate in both results. However, notice that the large deviation theorem gives only a rate and not a bound: For every arbitrary large constant $K > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log P\{T_P(P_n) \geq \epsilon\} = \lim_{n\to\infty} \frac{1}{n} \log K \cdot P\{T_P(P_n) \geq \epsilon\}.$$

But of course, the most striking advantage of Talagrand's inequality is that it is uniform in n.

7 Appendix

Lemma 6. A function f is lower semicontinuous if and only if it is the monotone limit of a sequence of continuous functions $f^{(k)}$

$$f(x) =\uparrow \lim_k f^{(k)}(x).$$

Proof. If f is the monotone limit of continuous functions $f^{(k)}$, its epigraph is the intersection of the epigraphs of $f^{(k)}$, which are closed. Therfore the epigraph of f is also closed and this is equivalent to the property that f is l.s.c. Conversely, let $A_i^{(k)}$ be a (non-disjunct) dissection of $\mathbb{R}^d$ into cubes of diameter $1/k$. Let

$$\tilde{f}^{(k)} = \sum_i \inf_{x\in A_i^{(k)}} f(x)\, \mathbb{1}_{x\in A_i^{(k)}}.$$

By the l.s.c. property,

$$\uparrow \lim_k \tilde{f}^{(k)}(x) = f(x). \tag{27}$$

It is easy to modify the functions $\tilde{f}^{(k)}$ such that they become continuous and still (27) holds. □

Lemma 7. Let $f(x)$ be a function, which is the pointwise limit of a monotone sequence of continuous functions $f(x) =\uparrow \lim_k f^{(k)}(x)$. Then, for a compact set X,

$$\lim_k \sup_{x\in X} f^{(k)}(x) = \sup_{x\in X} f(x) \tag{28}$$

$$\lim_k \inf_{x\in X} f^{(k)}(x) = \inf_{x\in X} f(x) \tag{29}$$

Proof. Since $f^{(k)}(x) \uparrow f(x)$, it follows that $\lim_k \sup_{x\in X} f^{(k)}(x) \le \sup_{x\in X} f(x)$ and $\lim_k \inf_{x\in X} f^{(k)}(x) \le \inf_{x\in X} f(x)$. On the other hand, if $f(x^*) \ge \sup_{x\in X} f(x) - \epsilon$, then $\lim_k f^{(k)}(x^*) = f(x^*)$ and therefore $\lim_k \sup_{x\in X} f^{(k)}(x) \ge \lim_k f^{(k)}(x^*) = f(x^*) \ge \sup_{x\in X} f(x) - \epsilon$. Since ϵ is arbitrary, (28) follows. Let now $f^{(k)}(x^{(k)}) = \inf_{x\in X} f^{(k)}(x)$ and $f^* = \sup_k f^{(k)}(x^{(k)})$. Let x^* be a cluster point of the sequence $(x^{(k)})$. W.l.o.g. we may even assume that this is a limit point. Since $f^{(\ell)}(x^{(k)}) \le f^*$ for all $\ell \le k$, we get by continuity of $f^{(\ell)}$ that $f^{(\ell)}(x^*) \le f^*$ and therefore $f(x^*) \le f^*$. This implies that $\inf_{x\in X} f(x) \le f(x^*) \le f^* = \lim_k \inf_{x\in X} f^{(k)}(x)$ and also (29) is shown. □

References

[1] Attouch H, Wets R.J.B. (1990). Laws of large numbers for random lsc functions. Seminaire Anal. Convexe 20, Exp. 13, Université de Montpellier.

[2] Bonnans J.F., Ioffe A.D. (1993). Second-order sufficiency and quadratic growth for non isolated minima. INRIA Report 1853

[3] Devroye L. (1987). A course in density estimation. Birkhuser Verlag, Basel.

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

[5] Pflug G. Ch. (1996). Asymptotic stochastic programs. Math. of OR. to appear

[6] Pflug G. Ch., Ruszczynski A., Schultz R. (1995). On the Glivenko-Cantelli Problem in Stochastic Programming: Linear Recourse. IIASA Working Paper WP-95-003

[7] Shapiro A. (1994). Quantitative stability in stochastic programming. Math. Programming 67, 99 - 108.

[8] Talagrand M. (1994). Sharper bounds for Gaussian and Emprircal Processes. Ann. Probab. 22 (1), 18 - 76

[9] Shorack G. R., Wellner J. A. (1986). Emprirical processes with applications to statistics. J. Wiley and Sons, New York.