International Workshop
"Stochastic Programming for Implementation and Advanced Applications"
(STOPROG-2012)
July 3–6, 2012, Neringa, Lithuania

ISBN 978-609-95241-4-6
L. Sakalauskas, A. Tomasgard, S. W.Wallace (Eds.):
Proceedings. Vilnius, 2012, pp. 94–99

doi:10.5200/stoprog.2012.17

LAW OF SMALL NUMBERS AS CONCENTRATION INEQUALITIES FOR SUMS OF INDEPENDENT RANDOM SETSAND RANDOM SET VALUED MAPPINGS

Vladimir I. Norkin[1], **Roger J-B Wets**[2]

[1]*Department of Operations Research, Institute of Cybernetics, Kiev*
[2]*Department of Mathematics, Univ. of California, Davis*
E-mails: [1]*norkin@i.com.ua (corresponding author);* [2]*rjbwets@ucdavis.edu*

Abstract. In the paper we study concentration of sample averages (Minkowski's sums) of independent bounded random sets and set valued mappings around their expectations. Sets and mappings are considered in a Hilbert space. Concentration is formulated in the form of exponential bounds on probabilities of normalized large deviations. In a sense, concentration phenomenon reflects the law of small numbers, describing non-asymptotic behavior of the sample averages. We sequentially consider concentration inequalities for bounded random variables, functions, vectors, sets and mappings, deriving next inequalities from preceding cases. Thus we derive concentration inequalities with explicit constants for random sets and mappings from the sharpest available (Talagrand type) inequalities for random functions and vectors. The most explicit inequalities are obtained in case of discrete distributions. The obtained results contribute to substantiation of the Monte Carlo method in infinite dimensional spaces.

Keywords: concentration inequality, Talagrand's inequality, random set, random set-valued mapping, Minkowski's averages, large deviation, law of small numbers.

1. Introduction

Concentration inequalities describe tail behavior of the probabilities of large deviations for sums of independent random variables from their mean. They, like limit theorems, determine rate of convergence in the law of large numbers. Moreover, they are valid for any finite sums and thus express, in a sense, a law of small numbers. Classical results of this type are Bernstein's, Chernoff's and Hoeffding's exponential inequalities for bounded random variables. There are extensions of these results to random vectors and random functions (see, e.g. (Talagrand, 1994, 1996), (Pflug, 2003), (Nemirovski, 2004), (Boucheron *et al.*, 2005), (Steinwart and Christmann, 2008), (Shapiro *et al.*, 2009)).

In the present paper we derive such inequalities for random sets (and mappings). First non-asymptotic result of this type was obtained by (Artstein, 1984) for finite dimensional independent random sets. It exponentially bounds the probability that Housdorff distance between sample average set and its mean exceeds a given threshold. The result is essentially finite dimensional since the bound heavily depends on the dimension number, the more the dimension the worse the bounds. Unlike this, contemporary concentration inequalities are dimension free (e.g. are valid for Banach space valued random variables), but exploit different complexity measures for the range of the random object. Asymptotic large deviation result for random sets in a separable Banach space was obtained in (Cerf, 1999). We derive explicit non-asymptotic concentration inequalities for bounded random sets and mappings in a separable infinite dimensional Hilbert space.

In the present paper we sequentially review and update concentration inequalities for bounded random variables, functions, vectors, random sets and set valued mappings, deriving next inequalities from preceding cases. Especially sharp and explicit results are obtained in case of discrete distributions. A closely related asymptotic result, a central limit theorem for random sets with a discrete distribution, was obtained in (Cressie, 1979).

Thus the main contribution of the paper consists in the following. We improve (Artstein's, 1984) concentration results for bounded random sets making the estimates much sharper and equally applicable to finite and infinite dimensional (separable Hilbert space) cases. The estimates still remain non-asymptotic, i.e. valid for any finite sum of independent random sets (unlike (Cerf, 1999)). We derive concentration inequalities from the sharpest available analogues results for random vectors and random functions. Concentration inequalities for random mappings are completely new, although they were obtained only for a discrete distribution. Extensions of these results to general distributions and to unbounded random sets and mappings are still open problems.

2. Concentration inequalities for sums of independent random variables and their extensions

Classical concentration inequalities for random variables are due to Bernstein, Chernoff and Hoeffding (see, e.g. (Boucheron *et al.*, 2005), (Steinwart and Christmann, 2008)).

Next statements extend concentration inequalities to random functions. Further they are used for derivation of concentration inequalities for random vectors and sets.

2.1 Theorem (see, e.g., (Boucheron *et al*.,. 2005)). Let (Y,d) be a separable metric space with metric d, $(\Xi,\Sigma,\mathbf{P})$ be a probability space and $F:=\{f_y:Y\times\Xi\to\mathbb{R}\}$ be a set of functions $f_y:Y\times\Xi\to\mathbb{R}$ continuous in $y\in Y$ for all $\xi\in\Xi$, Σ-measurable in ξ and $\mathbf{E}f_y=0$ for all $y\in Y$. Furthermore, let $b\geq 0$ be a constant such that $\|f_y\|_\infty\leq b$ for all $y\in Y$. For independent random variables $\xi_1,...,\xi_\nu$, we define a random variable $G_\nu:Z\to\mathbb{R}$ by $G_\nu:=\sup_{y\in Y}\left|\sum_{j=1}^{\nu}f_y(\xi_j)\right|$. Then, for all $t>0$ and all $\gamma>0$, we have

$$\mathbf{P}\left\{G_\nu>\mathbf{E}G_\nu+\sqrt{2t\nu}b\right\}\leq e^{-t}.$$

The following theorem presents a refinement of the (Talagrand's, 1994, 1996) inequality.

2.2 Theorem (Talagrand's inequality in i.i.d. case, see (Steinwart and Christmann, 2008)). Let $b\geq 0$, $\sigma\geq 0$, and $n\geq 1$. Moreover, let (Ξ,Σ,μ) be a probability space and F be a countable set of Σ-measurable functions such that $\mathbf{E}_\mu f=0$, $\mathbf{E}_\mu f^2\leq\sigma^2$ and $\|f\|_\infty\leq b$ for all $f\in F$. We write $Z:=\Xi^\nu$ and $\mathbf{P}:=\mu^\nu$, $\mathbf{E}$ denotes expectation over $\mathbf{P}$. Furthermore, for independent random variables $\xi_1,...,\xi_\nu$ we define a random variable $G_n:Z\to\mathbb{R}$ by $G_\nu:=\sup_{f\in F}\left|\sum_{j=1}^{\nu}f(\xi_j)\right|$. Then, for all $t>0$, we have

$$\mathbf{P}\left\{G_\nu\geq\mathbf{E}G_\nu+\sqrt{2t(\nu\sigma^2+2b\mathbf{E}G_\nu)}+2tb/3\right\}\leq e^{-t}.$$

To use Theorems 2.1, 2.2 one needs estimates of the quantity G_ν, which are given in the following lemmas.

2.3 Lemma ($\mathbf{E}G_\nu$ in case of finite Ξ). Suppose the probability space is discrete, Ξ contains N_Ξ elements, and $\sup_{f\in F,\xi\in\Xi}|f(\xi)|\leq M_{F,\Xi}$. Then $\mathbf{E}G_\nu\leq 2M_{F,\Xi}\sqrt{N_\Xi\nu}$.

2.4 Lemma ($\mathbf{E}G_\nu$ in case of finite F, (Boucheron et al., 2005, Theorem 3.3)). Suppose the family F ifs finite and bounded, i.e. it contains $|F|$ number of elements and $\sup_{f\in F,\xi\in\Xi}|f(\xi)|\leq M$. Then $G_\nu\leq M\sqrt{2\ln|F|}\big/\sqrt{\nu}$.

3. Concentration inequalities for sums of random vectors in a separable Hilbert space

In the present section we derive concentration inequalities for sums $S^\nu=\sum_{i=1}^{\nu}\xi_i$ of independent bounded random vectors ξ_i with values in a separable infinite dimension Hilbert space H. In the next section these results will be used to derive concentration inequalities for random sets. Similar results for finite dimensional regular Banach spaces (obtained in a different way) can be found in (Nemirovski, 2004).

Concentration inequalities for sums $S^\nu = \sum_{i=1}^{\nu} \xi_i$ of independent random vectors ξ_i can be obtained from such inequalities (Theorems 2.1, 2.2) for random functions, since for any $s \in H$ holds $\| s \| = \sup_{f\in H:\|f\|\le 1} f\, s$.

3.1 Theorem. Let $\xi_1,\dots,\xi_\nu$ be independent random variables (in a separable Hilbert space) such that $\| \xi_i - \mathbf{E}\xi_i \|_\infty \le b$, $\mathbf{E}\| \xi_i - \mathbf{E}\xi_i \|^2 \le \sigma^2$ for all i, then for any $t > 0$ holds

$$\mathbf{P}\left\{ \frac{1}{\nu} \| S^\nu - \mathbf{E}S^\nu \| > \frac{\sigma}{\sqrt{\nu}} + \frac{\sqrt{2tb}}{\sqrt{\nu}} \right\} \le e^{-t},$$

and (for iid case)

$$\mathbf{P}\left\{ \frac{1}{\nu} \| S^\nu - \mathbf{E}S^\nu \| \ge \frac{\sigma}{\sqrt{\nu}} + \frac{\sqrt{2t\left(\sigma^2 + 2\sigma b/\sqrt{\nu}\right)}}{\sqrt{\nu}} + \frac{2tb}{3\nu} \right\} \le e^{-t}.$$

As an illustration we apply Theorem 3.1 to derive concentration inequalities for frequencies in the law of large numbers. The next theorem establishes rate of convergence (of order $1/\sqrt{\nu}$) and the exponential decay of the probabilities of large deviations for frequencies in the law of large numbers with countable realizations. Later it is used for establishing concentration inequalities for random mappings.

3.2 Theorem. Let $\{\xi_1, \xi_2, \dots\}$ be a sequence of independent random variables taking values from a finite or countable collection $\{\bar\xi_1, \bar\xi_2, \dots\}$ with probabilities $\{p_1, p_2, \dots\} = p$ respectively. Denote $\lambda^\nu = \{\lambda_1^\nu, \dots, \lambda_j^\nu, \dots\}$ a random vector of frequencies such that λ_j^ν is the quotient of $\{\xi\}_{i=1}^{\nu}$ hitting the value $\bar\xi_j$. (If $\{\xi\}_{i=1}^{\nu}$ does not hit the value $\bar\xi_j$ then $\lambda_j^\nu = 0$). Then for any ν and $t > 0$

$$\mathbf{P}\{ \sqrt{\nu} \| \lambda^\nu - p \| > 1 + t \} \le e^{-t^2/4},$$

$$\mathbf{P}\left\{ \sqrt{\nu} \| \lambda^\nu - p \| > 1 + t\sqrt{1 + \frac{2\sqrt{2}}{\sqrt{\nu}} + \frac{2t^2}{3\sqrt{\nu}}} \right\} \le e^{-t^2/2},$$

where $\| \lambda^\nu - p \| = \left(\sum_j (\lambda_j^\nu - p_j)^2 \right)^{1/2}$.

4. Concentration inequalities for bounded random sets

Concentration inequalities from Theorem 3.1 can be extended to random sets. The first result of this type for bounded m-dimensional random sets was obtained by (Artstein, 1984). The estimate of the probability of large deviations of Minkowski's averages from its expectation in (Artstein, 1984) depends on the size of the deviation t through a multiplier $1/t^{m-1}$ standing before the exponential term, so it makes sense only for a finite dimensional case. An asymptotic large deviation result for random sets in a separable Banach space was presented in (Cerf, 1999).

Consider random sets $\{X_i\}$, i.e. Borel-measurable vector functions from some probability space $(\Xi, \Sigma, \mathbf{P})$ to the space of non-empty closed subsets of a separable Hilbert space H with norm $\|\cdot\| = \|\cdot\|_H$. A random vector $\xi : \Xi \to H$ is called a selection of X_i if with probability one holds $\xi \in X_i$. Define the expectation $\mathbf{E}X_i$ of X_i to be

$$\mathbf{E}X_i = \left\{ \mathbf{E}\xi :\ \xi \ \textit{is a selection of} \ X_i \ \textit{and} \ \mathbf{E}\| \xi \|_H < +\infty \right\}.$$

Define the deviation of a set $A \subset H$ from a set $B \subset H$ as $\Delta(A,B) = \sup_{a\in A} \inf_{b\in B} \| a-b \|_H$ and Hausdorff distance between sets A and B as $\mathbf{H}(A,B) = \max\{\Delta(A,B), \Delta(B,A)\}$. In particular for $a \in H$ denote $dist(a,B) = \Delta(a,B)$. Denote coA a convex hull of $A \subset H$, $\| A \|_\infty = \sup_{x\in A} \| x \|$. Define a sample average set

$$S^\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} X_i = \left\{ s = \frac{1}{\nu}\sum_{i=1}^{\nu} \xi_i :\ \xi_i \ is\ a\ selection\ of\ X_i \right\}.$$

In the next theorems we extend and sharpen finite dimensional results by (Artstein, 1984) to a separable infinite dimensional Hilbert space.

4.1 Theorem. Let $X_1,...,X_\nu$ be independent closed valued random sets with $\bar{X}_i = \mathbf{E}coX_i$, $\mathbf{H}(coX_i, \bar{X}_i) \le b < +\infty$ with probability one, $\mathbf{E}\mathbf{H}^2(coX_i, \bar{X}_i) \le \sigma^2$ for all i. Denote

$$D_\nu = \mathbf{H}\left(\frac{1}{\nu}\sum_{i=1}^{\nu} coX_i, \frac{1}{\nu}\sum_{i=1}^{\nu} \bar{X}_i \right).$$

Then for any $t > 0$ we have

$$\mathbf{P}\left\{ D_\nu \ge \mathbf{E}D_\nu + \frac{\sqrt{2t}b}{\sqrt{\nu}} \right\} \le e^{-t};$$

and for iid $\{X_i\}$

$$\mathbf{P}\left\{ D_\nu \ge \mathbf{E}D_\nu + \frac{\sqrt{2t\left(\sigma^2 + 2\sigma b/\sqrt{\nu}\right)}}{\sqrt{\nu}} + \frac{2tb}{3\nu} \right\} \le e^{-t}.$$

If $\{X_i\}$ take on values from a finite collection of size N_X, then the term $\mathbf{E}D_\nu$ in the inequalities can be replaced by $2b\sqrt{N_X}/\sqrt{\nu}$.

In a finite dimensional case, $\{X_i \subset \mathbb{R}^m\}$, the term $\mathbf{E}D_\nu$ in the inequalities can be replaced by $2b\sqrt{m}\left(1+\sqrt{\ln\nu}\right)/\sqrt{\nu}$.

Theorem 4.1 states a concentration result for convex valued random sets $\{coX_i\}$. In a finite dimensional case, $X_i \subset \mathbb{R}^m$, the result can be extended by means of Shapley-Folkman lemma to non-convex valued mappings similar to (Arstein, 1984). In an infinite dimensional case this is not possible. In the next theorem we obtain the desired result for one-sided distance $\Delta\left(\frac{1}{\nu}\sum_{i=1}^{\nu} X_i, \mathbf{E}coX_1\right)$.

4.2 Theorem. Let $X_1,...,X_\nu$ be iid closed valued random sets in a Hilbert space H, $\| X_i \|_\infty \le b < +\infty$ with probability one, $\mathbf{E}\| X_i \|^2 \le \sigma^2$, and $\bar{X} = \mathbf{E}coX_i$. Then for any $t > 0$ and ν we have

$$\mathbf{P}\left\{ \Delta\left(\frac{1}{\nu}\sum_{i=1}^{\nu} X_i, \bar{X} \right) \ge \frac{\sigma}{\sqrt{\nu}} + \frac{\sqrt{2t}b}{\sqrt{\nu}} \right\} \le e^{-t},$$

$$\mathbf{P}\left\{ \Delta\left(\frac{1}{\nu}\sum_{i=1}^{\nu} X_i, \bar{X} \right) \ge \frac{\sigma}{\sqrt{\nu}} + \frac{\sqrt{2t\left(\sigma^2 + 2\sigma b/\sqrt{\nu}\right)}}{\sqrt{\nu}} + \frac{2tb}{3\nu} \right\} \le e^{-t}.$$

5. Concentration inequalities for random sets and mappings in case of a discrete distribution

In this section we establish concentration inequalities for bounded random set valued mappings in the case when the random mapping has a discrete distribution. If the mapping is constant then these inequalities become concentration inequalities for random sets. In a similar setting (Cressie, 1979) obtains a central limit theorem for random sets with finite discrete distribution. Unlikely, our results are non-asymptotic, valid for countable discrete distributions and are extended to random mappings.

Assume that

(i) ξ is a discrete random variable taking at most countable number of values $\{\bar{\xi}_1, \bar{\xi}_2, ...\}$ with probabilities $\{p_1, p_2, ...\} = p$;

(ii) $\{S_1(x), S_2(x), ...\}$ are mappings defined on a set $X \subset \mathbb{R}^n$ with convex values in a Banach space;

(iii) mappings $\{S_j\}$ are uniformly bounded on X, $\| S(X) \| = \sup_j \sup_{x\in X} \| S_j(x) \| < +\infty$.

Define a random mapping $S(\xi, x) := S_j(x)$ if $\xi = \bar{\xi}_j$, and let $\{\xi_i, i = 1, 2, ...\}$ are iid realizations of the random variable ξ, sample average mapping $S^\nu(\xi^\nu, x) = \nu^{-1} \sum_{i=1}^{\nu} S(\xi_i, x)$ and the expectation mapping $\mathbf{E}S(\xi, x) = \sum_j p_j S_j(x) = \left\{ s = \sum_j s_j p_j \,:\, s_j \in S_j(x) \right\}$. Denote

$$S(x) = (S_1(x), S_2(x), ...),\ \| S_j(X) \| = \sup_{x\in X} \| S_j(x) \| < +\infty,\ \ \| S(X) \| = \left(\sum_j \| S_j \|^2 \right)^{1/2} < +\infty .$$

5.1 Theorem. In conditions (i)-(iii) for any $t > 0$ and ν the following estimates hold true for the uniform distance $\mathbf{U}(S^\nu, \mathbf{E}S) = \sup_{x\in X} \mathbf{H}(S^\nu(\xi, x), \mathbf{E}S(\xi, x))$ and graph distance $\mathbf{H}(gphS^\nu, gph\mathbf{E}S)$ between graphs of mappings S^ν and S,

$$\mathbf{P}\left\{ \sqrt{\nu}\ \mathbf{H}(gphS^\nu, gph\mathbf{E}S) > (1+t) \| S(X) \| \right\} \le$$
$$\le \mathbf{P}\left\{ \sqrt{\nu}\ \mathbf{U}(S^\nu, \mathbf{E}S) > (1+t) \| S(X) \| \right\} \ \le\ \exp\left(-t^2/4 \right).$$

These estimates have a non-asymptotic character, do not depend on the distribution p and contain explicit constants. If applied to random sets with discrete distribution, this theorem gives a sharper result than Theorem 4.1.

6. Conclusions

We have obtained a number of new explicit concentration results for vectors (Theorem 3.1), random sets (Theorems 4.1, 4.2) and random set valued mappings (Theorem 5.1) in infinite dimensional spaces. In a sense, concentration phenomenon reflects the law of small numbers, describing non-asymptotic behavior of sample averages. One interesting consequence of Theorem 3.1 is a concentration result for frequencies in the law of large numbers with countable number of realizations (Theorem 3.2). One more interesting consequence (of Theorem 5.1) is that in case of a discrete distribution a uniform law of large numbers for random mappings, i.e. $\lim_{\nu\to\infty} \mathbf{U}(S^\nu, \mathbf{E}S) = 0$, holds with probability one without any semicontinuity assumptions on the involved (convex valued uniformly bounded) mappings $\{S_j(x)\}$ (compare to (Shapiro and Xu, 2007)). All these results contribute to the substantiation of the Monte Carlo method in infinite dimensional spaces. Extensions of these results to general distributions and to unbounded random sets and mappings are still open problems.

Aknowledgements

Vladimir Norkin's work on the paper was supported by a Fulbright scholarship while staying at the Department of Mathematics of the University of California, Davis (October 2010 – September 2011) and by a grant of the Ukrainian State Fund for Fundamental Research Ф40.1/016. Roger Wets' research was supported in part by a grant of the Army Research Office (ARO) W911NS1010246.

References

Artstein, Z. (1984). “Convergence of sums of random sets”, in Ambertzumian, R. and Weil, W. (Ed.), Stochastic Geometry, Geometric Statistics and Stereology, Vol 65 of Teubner-Texte zur Math., Teubner Verlag, Leipzig, pp. 195–248

Boucheron, S.; Bousquet, O. and G. Lugosi (2005). Theory of classification: A survey of some recent advances, *ESAIM: Probability and Statistics* 9: 323–375. http://dx.doi.org/10.1051/ps:2005018

Cerf, R. (1999). “Large deviations for sums of i.i.d. random compact sets”, Proceedings of the American Mathematical Society, Vol 127, pp. 2431–2436. http://dx.doi.org/10.1090/S0002-9939-99-04788-7

Cressie, N. (1979). “A central limit theorem for random sets”, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 49: 37–47. http://dx.doi.org/10.1007/BF00534339

Ledoux, M. (2001*).* The Concentration of Measure Phenomenon, Vol 89 of Mathematical surveys and monographs, American Mathematical Society, Providence.

Nemirovski, A. S. (2004). “Regular Banach spaces and large deviations of random sums”, Available http://www2.isye.gatech.edu/nemirovs/LargeDev2004.pdf

Pflug, G. (2003). “Stochastic optimization and statistical inference”, in Ruszczyński, A. and Shapiro, A. (Ed.), Stochastic Programming, Vol 10 of Handbooks in Operations Research and Management Science, Elsevier, pp. 427–482

Shapiro, A.; Dentcheva, D. and Ruszczyński, A. (2009). Lectures on Stochastic Programming: Modeling and Theory, SIAM, Philadelphia. http://dx.doi.org/10.1137/1.9780898718751

Shapiro, A. and Xu, H. (2007). “Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions”, *Journal of Mathematical Analysis and Applications* 325: 1390–1399. http://dx.doi.org/10.1016/j.jmaa.2006.02.078

Steinwart, I. and Christmann, A. (2008). Support Vector Machines, Springer, New York.

Talagrand, M. (1994). “Sharper bounds for gaussian and empirical processes”, *The Annals of Probability* 22: 28–76. http://dx.doi.org/10.1214/aop/1176988847

Talagrand, M. (1996). “New concentration inequalities in product spaces”, *Inventiones Mathematicae* 126: 505–563. http://dx.doi.org/10.1007/s002220050108