PROBABILITY ASYMPTOTICS: NOTES ON NOTATION

SVANTE JANSON

Abstract. We define and compare several different versions of the O and o notations for random variables. The main purpose is to give proper definitions in order to avoid ambiguities and mistakes.

1. Introduction

There are many situations where one studies asymptotics of random variables or events, and it is therefore important to have good definitions and notations for random asymptotic properties. Probabilists use often the standard concepts *convergence almost surely* ($\xrightarrow{\text{a.s.}}$), *convergence in probability* ($\xrightarrow{\text{p}}$) and *convergence in distribution* ($\xrightarrow{\text{d}}$); see any textbook in probability theory for definitions. (Two of my favorite references, at different levels, are Gut [2] and Kallenberg [4].)

Other notations, often used in, for example, discrete probability such as probabilistic combinatorics, are probabilistic versions of the O and o notation. These notations are very useful; however, several versions exist with somewhat different definitions (some equivalent and some not), so some care is needed when using them. In particular, I have for many years avoided the notations "$O(\cdot)$ w.h.p." and "$o(\cdot)$ w.h.p." on the grounds that these combine two different asymptotic notations in an ambiguous and potentially dangerous way. (In which order do the quantifiers really come in a formal definition?) I now have changed opinion, and I regard these as valid and useful notations, provided proper definitions are given. One of the purposes of these notes is to state such definitions explicitly (according to my interpretations of the notions; I hope that others interpret them in the same way). Moreover, various relations and equivalences between different notions are given.

The results below are all elementary and more or less well-known. I do not think that any results are new, and they are in any case at the level of exercises in probability theory rather than advanced theorems. Nevertheless, I hope that this collection of various definitions and relations may be useful to myself and to others that use these concepts. (See also the similar discussion in [3, Section 1.2] of many of these, and some further, notions.)

Date: 24 April 2009; typos corrected 30 July 2009.

These notes were written at Institut Mittag-Leffler, Djursholm, Sweden, during the programme "Discrete Probability" 2009. I thank several participants for helpful comments and suggestions.

We suppose throughout that X_n are random variables and a_n positive numbers, $n = 1, 2, \dots$; unless we say otherwise, the X_n do not have to be defined on the same probability space. (In other words, only their distributions matter.) All unspecified limits are as $n \to \infty$.

All properties below relating X_n and a_n depend only on X_n/a_n; we could thus normalize and assume that $a_n = 1$, but for convenience in applications, we will state the results in the more general form with arbitrary positive a_n.

2. O and o

We begin with the standard definitions for non-random sequences. Assume that b_n is some sequence of numbers.

(D1) $b_n = O(a_n)$ if there exist constants C and n_0 such that $|b_n| \le Ca_n$ for $n \ge n_0$. Equivalently,

$$b_n = O(a_n) \iff \limsup_{n\to\infty} \frac{|b_n|}{a_n} < \infty. \tag{1}$$

(D2) $b_n = o(a_n)$ if $b_n/a_n \to 0$. Equivalently, $b_n = o(a_n)$ if for every $\varepsilon > 0$ there exists n_ε such that $|b_n| \le \varepsilon a_n$ for $n \ge n_\varepsilon$.

Remark 1. When considering sequences as here, the qualifier "$n \ge n_0$" is not really necessary in the definition of $O(\cdot)$, and it is often omitted, which is equivalent to replacing lim sup by sup in (1). The only effect of using an n_0 is to allow us to have a_n or b_n undefined or infinite, or $a_n = 0$, for some small n; for example, we may write $O(\log n)$ without making an explicit exception for $n = 1$. Indeed, if everything is well defined and $a_n > 0$, as we assume in these notes, and $|b_n| \le Ca_n$ for $n \ge n_0$, then $\sup_n |b_n/a_n| \le \max(C, \max_{i\le n_0} |b_n/a_n|) < \infty$.

On the other hand, when considering functions of a continuous variable, the two versions of $O(\cdot)$ are different and should be distinguished. (Both versions are used in the literature.) For example, there is a difference between the conditions $f(x) = O(x)$ on $(0, 1)$ (meaning $\sup_{0<x<1} |f(x)/x| < \infty$, i.e., a uniform estimate on $(0, 1)$), and $f(x) = O(x)$ as $x \to 0$ (meaning $\limsup_{x\to0} |f(x)/x| < \infty$, i.e., an asymptotic estimate for small x); the former but not the latter entails that f is bounded also close to 1. (As shown here, when necessary, the two versions of O can be distinguished by adding qualifiers such as "$n \to \infty$" or "$x \to 0$" for the asymptotic version and "$n \ge 1$" or "$x \in (0, 1)$" for the uniform version. Often, however, such qualifiers are omitted when the meaning is clear from the context.)

3. Convergence in probability

The standard definition of convergence in probability is as follows.

(D3) $X_n \xrightarrow{\text{p}} 0$ if for every $\varepsilon > 0$, $\mathbb{P}(|X_n| > \varepsilon) \to 0$. Equivalently,

$$X_n \xrightarrow{\text{p}} 0 \iff \sup_{\varepsilon>0} \limsup_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0.$$

Remark 2. More generally, one defines $X_n \overset{\mathrm{p}}{\longrightarrow} a$ for a constant a similarly, or by $X_n \overset{\mathrm{p}}{\longrightarrow} a \iff X_n - a \overset{\mathrm{p}}{\longrightarrow} 0$. If the random variables X_n are defined on the same probability space, one further defines $X_n \overset{\mathrm{p}}{\longrightarrow} X$ for a random variable X (defined on that probability space) by $X_n \overset{\mathrm{p}}{\longrightarrow} X$ if $X_n - X \overset{\mathrm{p}}{\longrightarrow} 0$.

It is well-known that convergence in probability to a constant is equivalent to convergence in distribution to the same constant. (See e.g. [1; 2; 4] for definition and equivalent charaterizations of convergence in distribution.) In particular,

$$X_n \overset{\mathrm{p}}{\longrightarrow} 0 \iff X_n \overset{\mathrm{d}}{\longrightarrow} 0. \tag{2}$$

4. With high probability

For events, we are in particular interested in typical events, i.e., events that occur with probability tending to 1 as $n \to \infty$. Thus, we consider an event $\mathcal{E}_n$ for each n, and say that:

(D4) $\mathcal{E}_n$ holds *with high probability* (w.h.p.) if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$.

This too is a common and useful notation.

Remark 3. A common name in probabilistic combinatorics for this property has been "almost surely" or "a.s.", but that conflicts with the well established use of this phrase (and abbreviation) in probability theory where it means probability *equal* to 1. In my opinion, "almost surely" (a.s.) should be reserved for its probabilistic meaning, since giving it a different meaning might lead to confusion. (In these notes, a.s. is used in the standard sense.) Another alternative name for (D4) is "asymptotically almost surely" or "a.a.s.". This name is commonly used, for example in [3], and the choice between the synonymous "w.h.p." (often written whp) and "a.a.s." is a matter of taste. (At present, I prefer w.h.p., so I use it here.)

Definition (D3) of convergence in probability can be stated using w.h.p. as:

$$X_n \overset{\mathrm{p}}{\longrightarrow} 0 \iff \text{for every } \varepsilon > 0,\ |X_n| \le \varepsilon \text{ w.h.p.} \tag{3}$$

5. O_{p} and o_{p}

A probabilistic version of O that is frequently used is the following:

(D5) $X_n = O_{\mathrm{p}}(a_n)$ if for every $\varepsilon > 0$ there exists constants C_ε and n_ε such that $\mathbb{P}(|X_n| \le C_\varepsilon a_n) > 1 - \varepsilon$ for every $n \ge n_\varepsilon$.

In other words, X_n/a_n is bounded, up to an exceptional event of arbitrarily small (but fixed) positive probability. This is also known as X_n/a_n being *bounded in probability.*

The definition (D5) can be rewritten in equivalent forms, for example as follows.

Lemma 1. *The following are equivalent:*

(i) $X_n = O_{\mathrm{p}}(a_n)$.

(ii) *For every* $\varepsilon > 0$ *there exists* C_ε *such that* $\mathbb{P}(|X_n| \le C_\varepsilon a_n) > 1 - \varepsilon$ *for every* n.
(iii) *For every* $\varepsilon > 0$ *there exists* C_ε *such that* $\limsup_{n\to\infty} \mathbb{P}(|X_n| > C_\varepsilon a_n) < \varepsilon$.
(iv) *For every* $\varepsilon > 0$ *there exists* C_ε *such that* $\sup_n \mathbb{P}(|X_n| > C_\varepsilon a_n) < \varepsilon$.
(v) $\lim_{C\to\infty} \limsup_{n\to\infty} \mathbb{P}(|X_n| > Ca_n) = 0$.
(vi) $\lim_{C\to\infty} \sup_n \mathbb{P}(|X_n| > Ca_n) = 0$.

Proof. (i) $\implies$ (ii) follows by increasing C_ε in (D5) such that $\mathbb{P}(|X_n| \le C_n a_n) > 1 - \varepsilon$ for $n = 1, \dots, n_0$ too.

(ii) $\implies$ (i) is trivial.

(i) $\iff$ (iii) $\iff$ (v) and (ii) $\iff$ (iv) $\iff$ (vi) are easy and left to the reader. □

Remark 4. Another term equivalent to "bounded in probability" is *tight*; thus, $X_n = O_p(a_n)$ if and only if the family $\{X_n/a_n\}$ is tight. By Prohorov's theorem [1; 4], tightness is equivalent to relative compactness of the set of distributions. Hence, $X_n = O_p(a_n)$ if and only if every subsequence of X_n/a_n has a subsequence that converges in distribution; however, different convergent subsequences may have different limits. In particular, if X_n/a_n converges in distribution, then $X_n = O_p(a_n)$.

The corresponding o_p notation can be defined as follows.

(D6) $X_n = o_p(a_n)$ if for every $\varepsilon > 0$ there exists n_ε such that $\mathbb{P}(|X_n| \le \varepsilon a_n) > 1 - \varepsilon$ for every $n \ge n_\varepsilon$.

The definition (D6) too has several equivalent forms, for example as follows.

Lemma 2. *The following are equivalent:*

(i) $X_n = o_p(a_n)$.
(ii) *For every* $\varepsilon > 0$, $\mathbb{P}(|X_n| > \varepsilon a_n) \to 0$.
(iii) $\sup_{\varepsilon>0} \limsup_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon a_n) = 0$.
(iv) *For every* $\varepsilon > 0$, $|X_n| \le \varepsilon a_n$ *w.h.p.*
(v) $X_n/a_n \xrightarrow{\text{p}} 0$.

Proof. (i) $\iff$ (ii) follows by standard arguments which we omit.

(ii) $\iff$ (iv) is immediate by the definition (D4) of w.h.p.

(ii) $\iff$ (v) is immediate by the definition (D3) of $\xrightarrow{\text{p}}$.

(Further, (iv) $\iff$ (v) follows by (3).) □

6. Using arbitrary functions $\omega(n)$

Some papers use properties that are stated using an arbitrary function (or sequence) $\omega(n) \to \infty$. (Or, equivalently, stated in terms of an arbitrary sequence $\delta_n := 1/\omega(n) \to 0$; see for example [4, Lemma 4.9], which is essentially the same as (i) $\iff$ (iii) in the following lemma.) They are equivalent to O_p or o_p by the following lemmas. (I find the O_p and o_p notation more transparent and prefer it to using $\omega(n)$.)

Lemma 3. *The following are equivalent:*

(i) $X_n = O_{\mathrm{p}}(a_n)$.

(ii) *For every function* $\omega(n) \to \infty$, $|X_n| \le \omega(n) a_n$ *w.h.p.*

(iii) *For every function* $\omega(n) \to \infty$, $|X_n|/(\omega(n)a_n) \xrightarrow{\mathrm{p}} 0$.

Proof. (i) $\Longrightarrow$ (ii). For every $\varepsilon > 0$, choose C_ε as in Lemma 1(iii). Then $\omega(n) > C_\varepsilon$ for large n, and thus

$$\limsup_{n\to\infty} \mathbb{P}(|X_n| > \omega(n)a_n) \le \limsup_{n\to\infty} \mathbb{P}(|X_n| > C_\varepsilon a_n) < \varepsilon.$$

Hence, $\limsup_{n\to\infty} \mathbb{P}(|X_n| > \omega(n)a_n) = 0$, which is (ii).

(ii) $\Longrightarrow$ (i). If $X_n = O_{\mathrm{p}}(a_n)$ does not hold, then, by the definition (D5), there exists $\varepsilon > 0$ such that for every C there exist arbitrarily large n with $\mathbb{P}(|X_n| > Ca_n) \ge \varepsilon$. We may thus inductively define an increasing sequence n_k, $k = 1, 2, \dots$, such that $\mathbb{P}(|X_{n_k}| > ka_{n_k}) \ge \varepsilon$. Define $\omega(n)$ by $\omega(n_k) = k$ and $\omega(n) = n$ for $n \notin \{n_k\}$. Then $\omega(n) \to \infty$ and $\mathbb{P}(|X_n| > \omega(n)a_n) \not\to 0$, so (ii) does not hold.

(ii) $\Longrightarrow$ (iii). If $\varepsilon > 0$, then $\varepsilon\omega(n) \to \infty$ too, and thus by (ii) $|X_n| \le \varepsilon\omega(n)a_n$ w.h.p. Thus $X_n/(\omega(n)a_n) \xrightarrow{\mathrm{p}} 0$ by (3).

(iii) $\Longrightarrow$ (ii). Take $\varepsilon = 1$ in (3). □

Remark 5. Lemma 3 generalizes the corresponding result for a non-random sequence $\{b_n\}$: $\{b_n\}$ is bounded $\iff$ $|b_n| \le \omega(n)$ for every $\omega(n) \to \infty$ $\iff$ $|b_n|/\omega(n) \to 0$ for every $\omega(n) \to \infty$.

Lemma 4. *The following are equivalent:*

(i) $X_n = o_{\mathrm{p}}(a_n)$.

(ii) *For some function* $\omega(n) \to \infty$, $|X_n| \le a_n/\omega(n)$ *w.h.p.*

(iii) *For some function* $\omega(n) \to \infty$, $\omega(n)|X_n|/a_n \xrightarrow{\mathrm{p}} 0$.

Proof. (i) $\Longrightarrow$ (ii). By the definition (D6), for every k there exists n_k such that if $n \ge n_k$, then $\mathbb{P}(|X_n| > k^{-1}a_n) < k^{-1}$. We may further assume that $n_k > n_{k-1}$, with $n_0 = 1$. Define $\omega(n) = k$ for $n_k \le n < n_{k+1}$. Then $\omega(n) \to \infty$ and $\mathbb{P}(|X_n| > \omega(n)^{-1}a_n) < \omega(n)^{-1}$. Since $\omega(n)^{-1} \to 0$, this yields (ii).

(ii) $\Longrightarrow$ (i). Let $\varepsilon > 0$. Since $\omega(n)^{-1} \le \varepsilon$ for large n, (ii) implies that $|X_n| \le \varepsilon a_n$ w.h.p. Thus $X_n = o_{\mathrm{p}}(a_n)$ by Lemma 2.

(ii) $\Longrightarrow$ (iii). If $\omega(n)$ is as in (ii), then $\omega(n)^{1/2}|X_n|/a_n \le \omega(n)^{-1/2}$ w.h.p.; since $\omega(n)^{-1/2} \to 0$, this implies $\omega(n)^{1/2}X_n/a_n \xrightarrow{\mathrm{p}} 0$, so (iii) holds with the function $\omega(n)^{1/2} \to \infty$.

(iii) $\Longrightarrow$ (ii). Take $\varepsilon = 1$ in (3). □

7. O_{L^p} and o_{L^p}

The following notations are less common but sometimes very useful. Recall that for $0 < p < \infty$ the L^p norm of a random variable X is $\|X\|_{L^p} := \big(\mathbb{E}\,|X|^p\big)^{1/p}$. Let $p > 0$ be a fixed number. (In applications, usually $p = 1$ or $p = 2$.)

(D7) $X_n = O_{L^p}(a_n)$ if $\|X_n\|_{L^p} = O(a_n)$.
(D8) $X_n = o_{L^p}(a_n)$ if $\|X_n\|_{L^p} = o(a_n)$.

In other words, $X_n = O_{L^p}(a_n) \iff \mathbb{E}\,|X_n|^p = O(a_n^p)$ and $X_n = o_{L^p}(a_n) \iff \mathbb{E}\,|X_n|^p = o(a_n^p)$; in particular, $X_n = O_{L^1}(a_n) \iff \mathbb{E}\,|X_n| = O(a_n)$ and $X_n = o_{L^1}(a_n) \iff \mathbb{E}\,|X_n| = o(a_n)$.

$X_n = o_{L^1}(a_n)$ thus says that $\mathbb{E}\,|X_n/a_n| \to 0$, which often is expressed as $X_n/a_n \to 0$ *in mean.* More generally, $X_n = o_{L^p}(a_n)$ is the same as $\mathbb{E}\,|X_n/a_n|^p \to 0$, which is called $X_n/a_n \to 0$ *in p-mean* (or in L^p). (For $p = 2$, a common name is $X_n/a_n \to 0$ *in square mean.*)

We may also take $p = \infty$. Since L^∞ is the space of bounded random variables and $\|X\|_{L^\infty}$ is the essential supremum of $|X|$, i.e., $\|X\|_{L^\infty} := \inf\{C : |X| \le C \text{ a.s.}\}$, the definitions (D7)–(D8) can for $p = \infty$ be rewritten as:

(D9) $X_n = O_{L^\infty}(a_n)$ if there exists a constant C such that $|X_n| \le Ca_n$ a.s.
(D10) $X_n = o_{L^\infty}(a_n)$ if there exists a sequence $\delta_n \to 0$ such that $|X_n| \le \delta_n a_n$ a.s.

Remark 6. In applications in discrete probability, typically each X_n is a discrete random variable taking only a finite number of possible values, each with positive probability. In such cases (and more generally if the number of values is countable, each with positive probability), $|X_n| \le Ca_n$ a.s. $\iff$ $|X_n| \le Ca_n$ surely (i.e., for each realization), and $|X_n| \le \delta_n a_n$ a.s. $\iff$ $|X_n| \le \delta_n a_n$ surely.

The notions O_{L^p} and o_{L^p} are useful for example when considering sums of a growing (or infinite) number of terms, since (for $p \ge 1$) such estimates can be added by Minkowski's inequality. For example, if $X_n = \sum_{i=1}^n Y_{ni}$, and $Y_{ni} = O_{L^p}(a_n)$ (uniformly in i) for some $p \ge 1$, then $X_n = O_{L^p}(na_n)$, and similarly for o_{L^p}. Note that the corresponding statement for O_{p} and o_{p} are false. (Example: Let Y_{ni} be independent with $\mathbb{P}(Y_{ni} = n^2) = 1 - \mathbb{P}(Y_{ni} = 0) = 1/n$ and let $a_n = 1$.)

By Lyapunov's (or Hölder's) inequality, $X_n = O_{L^p}(a_n) \implies X_n = O_{L^q}(a_n)$ and $X_n = o_{L^p}(a_n) \implies X_n = o_{L^q}(a_n)$ when $0 < q \le p \le \infty$. Thus the estimates become stronger as p increases. They are, for all p, stronger than O_{p} and o_{p}.

Lemma 5. *Let* $0 < p \le \infty$. *Then* $X_n = O_{L^p}(a_n) \implies X_n = O_{\mathrm{p}}(a_n)$ *and* $X_n = o_{L^p}(a_n) \implies X_n = o_{\mathrm{p}}(a_n)$.

Proof. Immediate from Markov's inequality. □

The converse fails for every $p > 0$. (Example for any $p > 0$: Take X_n with $\mathbb{P}(X_n = e^n) = 1 - \mathbb{P}(X_n = 0) = 1/n$ and let $a_n = 1$.)

Remark 7. For $p < \infty$, $X_n = o_{L^p}(a_n)$ is equivalent to $X_n = o_{\mathrm{p}}(a_n)$ together with the condition that $\{|X_n/a_n|^p\}$ are uniformly integrable, see e.g. [2] or [4].

Another advantage of O_{L^p} and o_{L^p} is that they are strong enough to imply moment estimates:

Lemma 6. *If k is a positive integer with $k \le p$, then $X_n = O_{L^p}(a_n) \implies \mathbb{E}\, X_n^k = O(a_n^k)$ and $X_n = o_{L^p}(a_n) \implies \mathbb{E}\, X_n^k = o(a_n^k)$.* □

In particular, $X_n = O_{L^1}(a_n) \implies \mathbb{E}\, X_n = O(a_n)$ and $X_n = o_{L^1}(a_n) \implies \mathbb{E}\, X_n = o(a_n)$; further, $X_n = O_{L^2}(a_n) \implies \operatorname{Var} X_n = O(a_n^2)$ and $X_n = o_{L^2}(a_n) \implies \operatorname{Var} X_n = o(a_n^2)$.

8. O w.h.p. and o w.h.p.

Since the basic meaning of O is "bounded by some fixed but unknown constant", my interpretation of "$O(a_n)$ w.h.p." is the following:

(D11) $X_n = O(a_n)$ w.h.p. if there exists a constant C such that $|X_n| \le Ca_n$ w.h.p.

Comparing Definitions (D5) and (D11), we see that the latter is a stronger notion:

$$X_n = O(a_n) \text{ w.h.p.} \implies X_n = O_{\mathrm{p}}(a_n), \tag{4}$$

but the converse does not hold. (In fact, (D11) is the same as (D5) with the restriction that C_ε must be chosen independent of ε.) For example, if $X_n/a_n \xrightarrow{\mathrm{d}} Y$ for some random variable Y, then always $X_n = O_{\mathrm{p}}(a_n)$, see Remark 4, but it is easily seen that $X_n = O(a_n)$ w.h.p. if and only if Y is bounded, i.e., $|Y| \le C$ (a.s.) for some constant $C < \infty$. (In particular, if $X_n = X$ does not depend on n, then always $X_n = O_{\mathrm{p}}(1)$, but $X_n = O(1)$ w.h.p. only if X is bounded.) This also shows that $X_n = O_{L^p}(a_n)$ in general does not imply $X_n = O(a_n)$ w.h.p.

Remark 8. More generally, $X_n = O(a_n)$ w.h.p. if and only if every subsequence of X_n/a_n has a subsequence that converges in distribution to a bounded random variable, with some uniform bound for all subsequence limits.

Remark 9. The property $X_n = O(a_n)$ w.h.p. was denoted $X_n = O_C(a_n)$ in [3]. (A notation that perhaps was not very successful.)

Similarly, the basic meaning of o is "bounded by some fixed but unknown sequence $\delta_n \to 0$"; thus my interpretation of "$o(a_n)$ w.h.p." is the following:

(D12) $X_n = o(a_n)$ w.h.p. if there exists a sequence $\delta_n \to 0$ such that $|X_n| \le \delta_n a_n$ w.h.p.

This condition is the same as Lemma 4(ii) (with $\delta_n = \omega(n)^{-1}$), and thus Lemma 4 implies the following equivalence:

Lemma 7. *$X_n = o(a_n)$ w.h.p. $\iff X_n = o_{\mathrm{p}}(a_n)$.* □

It is obvious from the definitions (D11) and (D12) that $o(a_n)$ w.h.p. implies $O(a_n)$ w.h.p., and we thus have the chain of implications (where the last two are not reversible):

$$o_{\mathrm{p}}(a_n) \iff o(a_n) \text{ w.h.p.} \implies O(a_n) \text{ w.h.p.} \implies O_{\mathrm{p}}(a_n). \tag{5}$$

Warning. I do not think that definition (D11) is the only interpretation of "$O(a_n)$ w.h.p." that is used, so extreme care is needed when using or seeing this notation to avoid confusion and mistakes. (For example, I've heard the interpretation that "$O(a_n)$ w.h.p." should be equivalent to "$O_p(a_n)$".) The risks with "o_p w.h.p." seem smaller; at least, I do not know any other reasonable (non-equivalent) interpretation of it.

9. O and o a.s.

In this section we assume that the random variables X_n are defined together on the same probability space Ω. In other words, the variables X_n are coupled. (In combinatorial situations this is usually *not* the case, since typically each X_n is defined separately on some model of "size" n; however, it happens, for example in a model that grows in size by some random process.) This assumption makes it possible to talk about convergence and other properties a.s., i.e., pointwise (= pathwise) for all points in the probability space Ω except for a subset with probability 0. This means that we consider the sequence $X_n(\omega)$ of real numbers separately for each point ω in the probability space. Hence, we apply definitions (D1) and (D2) for non-random sequences and obtain the following definitions.

(D13) $X_n = O(a_n)$ a.s. if for almost every $\omega \in \Omega$, there exists a number $C(\omega)$ such that $|X_n(\omega)| \le C(\omega)a_n$. In other words, $X_n = O(a_n)$ a.s. if there exists a *random variable* C such that $|X_n| \le Ca_n$ a.s. Equivalently,

$$X_n = O(a_n) \text{ a.s.} \iff \limsup_{n\to\infty} \frac{|X_n|}{a_n} < \infty \text{ a.s.} \tag{6}$$

(D14) $X_n = o(a_n)$ a.s. if for almost every $\omega \in \Omega$, $|X_n(\omega)|/a_n \to 0$. In other words, $X_n = o(a_n)$ a.s. if $X_n/a_n \xrightarrow{\text{a.s.}} 0$.

It is well-known that convergence almost surely implies convergence in probability (but not conversely). Consequently, by (D14) and Lemmas 2 and 7,

$$X_n = o(a_n) \text{ a.s.} \implies X_n = o_p(a_n) \iff X_n = o(a_n) \text{ w.h.p.} \tag{7}$$

The situation for O is more complicated. We first observe the implication

$$X_n = O(a_n) \text{ a.s.} \implies X_n = O_p(a_n). \tag{8}$$

(The converse does not hold, see Example 9 below.) Indeed, if c is any constant, and C is a random variable with $|X_n| \le Ca_n$ as in (D13), then $\mathbb{P}(|X_n| > ca_n) \le \mathbb{P}(C > c)$, and thus Lemma 1(vi) holds because $\mathbb{P}(C > c) \to 0$ as $c \to \infty$. Hence, Lemma 1 yields (8).

However, the following two examples show that neither of $X_n = O(1)$ a.s. and $X_n = O(1)$ w.h.p. implies the other.

Example 8. Let $X_n = X$ be independent of n and let $a_n = 1$. Then $X_n = O(1)$ a.s. for every random variable X (take $C = X$ in (D13)), but

$X_n = O(1)$ w.h.p. only if X is a bounded random variable (i.e., $|X| \le c$ a.s. for some constant c).

Example 9. Let X_n be independent random variables with $\mathbb{P}(X_n = n) = 1/n$ and $\mathbb{P}(X_n = 0) = 1 - 1/n$, and take $a_n = 1$. By the Borel–Cantelli lemma, $X_n = n$ infinitely often a.s., and thus $\limsup_{n\to\infty} X_n = \infty$ a.s.; consequently X_n is not $O(1)$ a.s. On the other hand, $X_n \xrightarrow{\text{p}} 0$, so $X_n = o_{\text{p}}(1)$ and $X_n = O(1)$ w.h.p. by (5).

Warning. In particular, there is no analogue of (7) for O a.s. and O w.h.p. Since "a.s." usually is a strong notion compared to others (for example for convergence), there is an obvious risk of confusion and mistakes here, and it is important to be extra careful when using "$O(a_n)$ a.s." and "$O(a_n)$ w.h.p.".

10. A final warning

Sometimes one sees expressions of the type $X_n = O(a_n)$ or $X_n = o(a_n)$, for some random variables X_n, without further qualifications or explanations. In analogy with Section 8, I think that the natural interpretations of these are the following:

(D15) $X_n = O(a_n)$ if there exists a constant C such that $|X_n| \le Ca_n$ (surely, or a.s.).

(D16) $X_n = o(a_n)$ if there exists a sequence $\delta_n \to 0$ such that $|X_n| \le \delta_n a_n$ (surely, or a.s.).

These notations are thus uniform estimates, and stronger than $X_n = O(a_n)$ w.h.p. and $X_n = o(a_n)$ w.h.p., since no exceptional events of small probabilities are allowed.

Remark 10. As remarked in Remark 6, in typical applications "surely" and "a.s." are equivalent. When they are not, it is presumably best to follow standard probability theory practise and ignore events of probability 0, so the interpretation "a.s" in (D15)–(D16) seems best. In this case, (D15)–(D16) are the same as (D9)–(D10), so $X_n = O(a_n) \iff X_n = O_{L^\infty}(a_n)$ and $X_n = o(a_n) \iff X_n = o_{L^\infty}(a_n)$.

Warning. However, I guess that most times one of these notations is used, (D15) or (D16) is *not* the intended meaning; either there are typos, or the author really means something else, presumably one of the other notions discussed above.

Remark 11. In the special situation that all X_n are defined on a common probability space as in Section 9, another reasonable interpretation of $X_n = O(a_n)$ and $X_n = o(a_n)$ is $X_n = O(a_n)$ a.s. and $X_n = o(a_n)$ a.s., see (D13)–(D14). This is equivalent to allowing random C or δ_n in (D15)–(D16), and is a weaker property. (This emphasizes the need for careful definitions to avoid ambiguities.)

The notations (D15) and (D16) thus risk being ambiguous. If (D15) or (D16) really is intended, it may be better to use the unambiguous notation O_{L^∞} or o_{L^∞}, see Section 7 and Remark 10.

References

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968; 2nd ed. 1999.

[2] A. Gut, *Probability: A Graduate Course.* Springer, New York, 2005. Corrected 2nd printing 2007.

[3] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[4] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se

URL: http://www.math.uu.se/∼svante/