Borel-Cantelli lemmas and the law of large numbers

Gunnar Englund & Timo Koski
Matematisk statistik
KTH

2008

1 Introduction

Borel-Cantelli lemmas are interesting and useful results especially for proving the law of large numbers in the strong form.
We consider a sequence events $A_1, A_2, A_3, \ldots$ and are intrested in the question of whether infinitely many events occur or if possibly only a finite number of them occur.
We set

$$F_n = \bigcup_{k=n}^{\infty} A_k \text{ and } G_n = \bigcap_{k=n}^{\infty} A_k. \tag{1}$$

If G_n in (1) occurs, this means that all A_k for $k \geq n$ occur. If there is some such n, this means in other words that from this n on all A_k occur for $k \geq n$. With

$$H = \bigcup_{n=1}^{\infty} G_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

we have that if H occurs, then there is an n such that all A_k with $k \geq n$ occur. Sometimes we denote H with $\liminf A_k$.
The fact that F_n occurs implies that there is some A_k for $k \geq n$ which occurs. If F_n in (1) occurs for all n this implies that infinitely many of the A_k:s occur. We form therefore

$$E = \bigcap_{n=1}^{\infty} F_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

If E occurs, then infinitely many of A_k:s occur. Sometimes we write this as $E = \{A_n \text{ i.o.}\}$ where i.o. is to be read as "infinitely often", i.e., infinitely many times. E is sometimes denoted with $\limsup A_k$.
We need a couple of auxiliary results (the lemmas below) of probability calculus (found, e.g., on page 3 of [1]) that are basic in the sense that they are derived directly from the Kolmogorov axioms.

First, we consider a sequence of events $B_k \in \mathcal{F}$ that is *increasing* to $B \in \mathcal{F}$. This means that

$$B_1 \subset B_2 \subset B_3 \subset \ldots \subset B_n \subset B_{n+1} \subset \ldots \subset B$$

and thus

$$\cup_{k=1}^{\infty} B_k = B.$$

We can write this as

$$B_n \uparrow B$$

Lemma 1 *If $B_n \uparrow B$, then $P(B) = \lim_{n\to\infty} P(B_n)$.*

■

Proof: We can write $\cup_{k=1}^{\infty} B_k = \cup_{k=2}^{\infty}(B_k \setminus B_{k-1}) \cup B_1$, since B_k are increasing.

$$P(B) = P\left(\cup_{k=1}^{\infty} B_k\right) = P\left(\cup_{k=2}^{\infty}(B_k \setminus B_{k-1}) \cup B_1\right)$$

But the sets in the decomposition are seen to be disjoint, and hence the axiom of countable additivity yields

$$P\left(\cup_{k=2}^{\infty}(B_k \setminus B_{k-1}) \cup B_1\right) = \sum_{k=2}^{\infty} P\left(B_k \setminus B_{k-1}\right) + P\left(B_1\right)$$

$$= \lim_{n\to\infty} \sum_{k=2}^{n} P\left(B_k \setminus B_{k-1}\right) + P\left(B_1\right)$$

Now we observe that since $B_{k-1} \subset B_k$, we have

$$P\left(B_k \setminus B_{k-1}\right) = P\left(B_k\right) - P\left(B_{k-1}\right).$$

Therefore we get a telescoping series

$$\sum_{k=2}^{n} P\left(B_k \setminus B_{k-1}\right) + P\left(B_1\right) = P\left(B_n\right) - P\left(B_{n-1}\right) + P\left(B_{n-1}\right) - P\left(B_{n-2}\right) + \ldots +$$

$$+P\left(B_2\right) - P\left(B_1\right) + P\left(B_1\right) = P\left(B_n\right).$$

In other words we have shown that

$$P(B) = \lim_{n\to\infty} P\left(B_n\right).$$

■

We take next a sequence of events $B_k \in \mathcal{F}$ that is *decreasing*. This means that

$$B_1 \supset B_2 \supset B_3 \supset \ldots \supset B_k \supset B_{k+1} \supset \ldots \supset \cap_{k=1}^{\infty} B_k$$

and thus we write

$$B_n \downarrow \cap_{k=1}^{\infty} B_k.$$

Lemma 2 *If $B_n \downarrow \cap_{k=1}^{\infty} B_k$, then $P\left(\cap_{k=1}^{\infty} B_k\right) = \lim_{n\to\infty} P\left(B_n\right)$.*

Proof: We use the axiom of probability of complementary events

$$P\left(\cap_{k=1}^{\infty} B_k\right) = 1 - P\left(\left(\cap_{k=1}^{\infty} B_k\right)^c\right). \tag{2}$$

When we apply one of De Morgan's rules we get

$$\left(\cap_{k=1}^{\infty} B_k\right)^c = \cup_{k=1}^{\infty} B_k^c.$$

Now we observe that if $B_k \supset B_{k+1}$, then $B_k^c \subset B_{k+1}^c$, i.e., the complement events of a decreasing sequence of events are an increasing sequence of events. Thus the first lemma implies

$$P\left(\cup_{k=1}^{\infty} B_k^c\right) = \lim_{n\to\infty} P\left(B_n^c\right).$$

Hence we have that

$$\begin{aligned} P\left(\left(\cap_{k=1}^{\infty} B_k\right)^c\right) &= P\left(\cup_{k=1}^{\infty} B_k^c\right) \\ &= \lim_{n\to\infty} P\left(B_n^c\right). \end{aligned}$$

This we shall insert in (2) and get

$$\begin{aligned} P\left(\cap_{k=1}^{\infty} B_k\right) &= 1 - \lim_{n\to\infty} P\left(B_n^c\right) \\ &= 1 - \lim_{n\to\infty}\left(1 - P\left(B_n\right)\right) \\ &= 1 - 1 + \lim_{n\to\infty} P\left(B_n\right) = \lim_{n\to\infty} P\left(B_n\right). \end{aligned}$$

This completes the proof. ■

2 The Borel-Cantelli results

One notes that F_n is a decreasing set of events. This is simply so because

$$F_n = \bigcup_{k=n}^{\infty} A_k = A_n \bigcup \left(\bigcup_{k=n+1}^{\infty} A_k\right) = A_n \bigcup F_{n+1}$$

and thus

$$F_n \supset F_{n+1}.$$

Thus the second lemma above gives

$$P(E) = P\left(\bigcap_{n=1}^{\infty} F_n\right) = \lim_{n\to\infty} P(F_n) = \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right).$$

We have, however, by Boole's inequality that

$$P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k)$$

and this sum $\to 0$ as $n \to \infty$, if the sum $\sum_1^{\infty} P(A_k)$ converges. This implies that we have shown following proposition, also known as the Borel-Cantelli lemma.

Proposition 1 *Borel-Cantelli lemma*
If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then it holds that $P(E) = P(A_n \text{ i.o}) = 0$, i.e., that with probability 1 only finitely many A_n occur.

■

One can observe that no form of independence is required, but the proposition holds in general, i.e., for any sequence of events.
There is a converse to the Borel-Cantelli lemma obtained if we assume that the events $A_1, A_2, \ldots$ are independent .

Proposition 2 *Converse Borel-Cantelli lemma*
If $A_1, A_2, \ldots$ are independent and

$$\sum_{n=1}^{\infty} P(A_n) = \infty,$$

then it holds that $P(E) = P(A_n \text{ i.o}) = 1$, i.e., it holds with probability 1 that infinitely many A_n occur.

■

Proof: We have by independence

$$P(\bigcap_{k=n}^{\infty} A_k^*) = \prod_{k=n}^{\infty} P(A_k^*) = \prod_{k=n}^{\infty} (1 - P(A_k)).$$

Since $1 - x \leq e^{-x}$ we get $1 - P(A_k) \leq e^{-P(A_k)}$ and

$$P(\bigcap_{k=n}^{\infty} A_k^*) \leq \exp(-\sum_{k=n}^{\infty} P(A_k)).$$

If now $\sum_{n=1}^{\infty} P(A_n) = \infty$, then the sum in the exponent diverges and we obtain

$$P(\bigcap_{k=n}^{\infty} A_k^*) = 0.$$

Thus it holds also that

$$P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^*) = 0,$$

which implies by De Morgan's rules that

$$P(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) = 1 - P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^*) = 1 - 0 = 1$$

i.e., that infinitely many A_k:n occur with probability 1. ■

3 Some examples of applications

Exempel 1 Let $X_1, X_2, X_3 \dots$ be independent equidistributed with continuous distribution. We let

$$U_n = \begin{cases} 1 & \text{if } X_n > X_j \text{ for } j = 1, 2, \dots n-1 \\ 0 & \text{annars.} \end{cases}$$

This says simply that $U_n = 1$ if X_n is a "record", i.e., the largest value observed so far. We set $A_n = \{U_n = 1\}$.
We see that $P(U_n = 1) = 1/n$, since the probability that the largest of n values should occur in the round n is $1/n$ for reasons of symmetry. Furthermore, $A_1, A_2, \dots$ are independent. We have

$$P(A_m \cap A_{m+1} \cap \dots A_{m+k}) = P(A_m | A_{m+1} \cap \dots A_{m+k}) P(A_{m+1} \cap \dots A_{m+k})$$

and A_m and $A_{m+1} \cap \dots A_{m+k}$ are, of course, independent since A_m is only concerned with the relative magnitudes of the m first of X-variables.
We get then, as

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

that $P(A_n \text{ i.o.}) = 1$ i.e., infinitely many A_n occur. We get infinitely many records – a result which perhaps (?) is self-evident. Furthermore, we get

$$E(\sum_{n=1}^{\infty} U_n U_{n+1}) = \sum_{n=1}^{\infty} E(U_n U_{n+1}) = \sum_{n=1}^{\infty} P(A_n) P(A_{n+1}) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} < \infty$$

and hence $\sum_1^\infty U_n U_{n+1}$ is finite with probability 1. The conclusion is that there only occurs a finitely number of "double records", i.e., a record two times in a row – this result is by no means trivial.

Exempel 2 Let $X_1, X_2, \dots$ be independent and equidistributed. Then it holds that

$$E(|X_1|) = \int_0^\infty P(|X_1| > x) dx = \sum_{n=0}^{\infty} \int_n^{n+1} P(|X_1| > x) dx$$
$$\leq \sum_{n=0}^{\infty} P(|X_1| > n) \leq 1 + \sum_{n=1}^{\infty} P(|X_n| > n),$$

but also

$$E(|X_1|) = \sum_{n=0}^{\infty} \int_n^{n+1} P(|X_1| > x) dx \geq \sum_{n=0}^{\infty} P(|X_1| > n+1) = \sum_{n=1}^{\infty} P(|X_n| > n).$$

If now $E(|X_1|) < \infty$ we see that $\sum_{n=1}^\infty P(|X_n| > n) < \infty$ which according to Borel-Cantelli lemma implies $P(|X_n| > n \text{ i.o.}) = 0$. On the other hand, if $E(|X_1|) = \infty$ and thereby $\sum_{n=0}^\infty P(|X_n| > n) = \infty$, the converse Borel-Cantelli lemma entails that $P(|X_n| > n \text{ i.o.}) = 1$. If $E(|X_k|)$ are finite (respectively infinite) $|X_n|$ will be larger than n infinitely many times with probability 0 (respectively 1).

4 Proof of the strong form of the law of large numbers

We let $X_1, X_2, \ldots$ be independent equidistributed with $E(X_i) = m$ and $\text{Var}(X_i) = \sigma^2 < \infty$ and define $S_n = X_1 + X_2 + \cdots + X_n$. We are interested in showing the strong form of the law of large numbers (SLLN), i.e., that it holds with probability one that $S_n/n \to m$ as $n \to \infty$. This means that we want to prove that

$$P(\lim_{n\to\infty} \frac{S_n}{n} = m) = 1,$$

i.e., that there exists a set Ω_0 with $P(\Omega_0) = 1$ where for every $\omega \in \Omega_0$ it holds that

$$\lim_{n\to\infty} |\frac{S_n}{n} - m| = 0.$$

We need in other words to prove that for every $\omega \in \Omega_0$ and for every $\varepsilon > 0$ there is $N(\omega, \varepsilon)$ so that if $n \geq N(\omega, \varepsilon)$ holds that $|S_n/n - m| \leq \varepsilon$.
It suffices to prove that $|\frac{S_n}{n} - m| > \varepsilon$ can occur only a finite number of times, i.e.,

$$\lim_{N\to\infty} P(|\frac{S_n}{n} - m| > \varepsilon \quad \text{some } n \geq N) = 0.$$

Note the distinction with regard to the law of large numbers in the weak form, which says that that for all $\varepsilon > 0$

$$P(|\frac{S_n}{n} - m| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

In words: for the law of large numbers in the strong form $|S_n/n - m|$ must be small for all sufficiently large n for all $\omega \in \Omega_0$ where $P(\Omega_0) = 1$.
In tossing a coin we can code heads and tails with 0 and 1, respectively, and we can identify an ω with a number in the intervall $[0, 1]$ drawn at random, where binary expansion gives the sequence of zeros and ones. The law of large numbers says in this case that we will obtain with probability 1 a number such that the proportion of 1:s in sequence converges towards 1/2. There can be "exceptional" -ω - for example the sequence $000\ldots$ is possible, but such exceptional sequences have the probability 0.
After these deliberations of pedagogic nature let us get on with the proof.

Proof of SLLN: Without restriction of generality we can assume that $E(X_i) = m = 0$, since we in any case can consider $X_i - m$. We have $V(S_n) = n\sigma^2$. By Chebyshev's inequality it holds that

$$P(|S_n| > n\varepsilon) \leq \frac{V(S_n)}{(n\varepsilon)^2} = \frac{n\sigma^2}{(n\varepsilon)^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Unfortunately the harmonic series $\sum_1^\infty 1/n$ is divergent so we cannot use Borel-Cantelli lemma directly. But it holds that $\sum_1^\infty 1/n^2 < \infty$ and this means that we can use the lemma for n^2, $n = 1, 2, \ldots$. We have

$$P(|S_{n^2}| > n^2\varepsilon) \leq \frac{\sigma^2}{n^2\varepsilon^2}.$$

In other words it holds by Borel-Cantelli lemma that $P(|\frac{S_{n^2}}{n^2}| > \varepsilon \text{ i.o.}) = 0$ which proves that (with probability 1) $S_n^2/n^2 \to 0$. We have in other words managed to prove that for the subsequence $n^2,\ n = 1, 2, \ldots$ there is convergence with probability 1. It remains to find out what will happen between these n^2. We define therefore

$$D_n = \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|,$$

i.e., the largest of the deviation from S_{n^2} that can occur between n^2 and $(n+1)^2$. We get

$$D_n^2 = \max_{n^2 \leq k < (n+1)^2} (S_k - S_{n^2})^2 \leq \sum_{k=n^2}^{(n+1)^2-1} (S_k - S_{n^2})^2,$$

where we used the rather crude inequality $\max(|x|, |y|) \leq (|x| + |y|)$. This entails

$$E(D_n^2) \leq \sum_{k=n^2}^{(n+1)^2-1} E((S_k - S_{n^2})^2).$$

But $E((S_k - S_{n^2})^2) = (k - n^2)\sigma^2 \leq 2n\sigma^2$ as $n^2 \leq k < (n+1)^2$ and there are $2n$ terms in the sum and this entails

$$E(D_n^2) \leq (2n)(2n)\sigma^2 = 4n^2\sigma^2.$$

With Chebyshev's inequality this gives

$$P(D_n > n^2\varepsilon) \leq \frac{4n^2\sigma^2}{(n^2\varepsilon)^2} = \frac{4\sigma^2}{n^2\varepsilon^2}.$$

In other words, $D_n/n^2 \to 0$ holds with probability 1. Finally this yields for k between n^2 and $(n+1)^2$ that

$$|\frac{S_k}{k}| \leq \frac{|S_{n^2}| + D_n}{k} \leq \frac{|S_{n^2}| + D_n}{n^2} \to 0.$$

This means that we have succeeded in proving that $S_n/n \to 0$ with probability 1. We have done this under additional condition that $\mathrm{Var}(X_i) = \sigma^2$, but with a painstaking effort we can in fact prove that this additional condition is not necessary. ■

Referenser

[1] Z. Brezniak and T. Zastawniak (2007): *Basic Stochastic Processes.* Springer Verlag, London.