Intro to LLN

A. Veretennikov

Warm-up: Bienaimé–Chebyshev and variance

Bienaimé–Chebyshev's inequality: for any r.v. $\xi \ge 0$, and any constant $c > 0$ (notation is used, $a \wedge b \equiv \min(a, b)$),

$$\boxed{P(\xi \ge c) \le \frac{E\xi}{c} \wedge \frac{E\xi^2}{c^2} \wedge \dots}$$

Proof: (for squares and higher moments similarly)

$$P(\xi \ge c) = E1(\xi \ge c) \le E1(\xi \ge c)\frac{\xi}{c} \le E\frac{\xi}{c}.$$

Bienaimé–Chebyshev with variance: for *any* r.v. ξ with a finite variance $\sigma^2 = E(\xi - E\xi)^2$, and any constant $c > 0$,

$$\boxed{P(|\xi - E\xi| \ge c) \le \frac{\mathrm{var}(\xi)}{c^2}.}$$

Variance and higher moments

Variance of a sum (reminder): for independent r.v.'s

$$\mathrm{var}(n^{-1}\sum_{k=1}^{n}\xi_k) = n^{-2}\sum_{k=1}^{n}\mathrm{var}(\xi_k).$$

Higher moments of a sum: for any sequence of IIDRV's $(\xi_k,\ k \geq 1)$ with a finite moment $E|\xi_k|^{2m}$, with some even integer number $2m > 0$, there exists $C > 0$ such that

$$E\left|\sum_{k=1}^{n}(\xi_k - E\xi_k)\right|^{2m} \leq Cn^m.$$

Proof

Warm-up: LLN for IIDRV's with a finite variance

LLN–1: for IIDRV's with $\sigma^2 := \mathrm{var}(\xi_1) < \infty$, for any $\epsilon > 0$,

$$P\left(\left|n^{-1}\sum_{k=1}^{n}\xi_k - E\xi_1\right| > \epsilon\right) \to 0, \qquad n \to \infty.$$

Proof: due to $E(\xi - E\xi)^2 =: \mathrm{var}(\xi)$, follows from the line,

$$P\left(\left|n^{-1}\sum_{k=1}^{n}\xi_k - E\xi_1\right| > \epsilon\right) \le \frac{\mathrm{var}(\sum_{k=1}^{n}(\xi_k - E\xi_1))}{n^2\epsilon^2} = \frac{n\sigma^2}{n^2\epsilon^2}.$$

Def: we say that $\eta_n \xrightarrow{P} \eta_0$ (called *convergence in probability*), iff $\forall\epsilon > 0$, $P(|\eta_n - \eta_0| \ge \epsilon) \to 0$, $n \to \infty$.

So, *LLN-1 is a theorem about convergence in probability.*

Khintchine's LLN (under the 1st moment)

LLN–2 [Khintchine]: for IIDRV's with only $E\xi_1 < \infty$,

$$n^{-1}\sum_{k=1}^{n}\xi_k - E\xi_1 \stackrel{P}{\to} 0, \qquad n \to \infty.$$

NB: by $E\xi_1 < \infty$ we always understand $E|\xi_1| < \infty$.

NB: Since for any $c > 0$, in continuous case,

$$\int_{-\infty}^{\infty}|x|p(x)\,dx = \lim_{n\to\infty}\int_{-cn}^{cn}|x|p(x)\,dx,$$

if the left hand side is finite, then *it is natural to accept without proof that in general,*

$$\boxed{\lim_{n\to\infty} E|\xi|1(|\xi| \le cn) = E|\xi|} \sim \boxed{\lim_{n\to\infty} E|\xi|1(|\xi| > cn) = 0.}$$

Khintchine's LLN: Proof

Proof **[from AYV Lecture notes 1997]**: I. We may assume $E\xi_1 = 0$. By considering smooth approximations f to $1(|x| \ge \epsilon)$, it easily follows that it suffices to show convergence for $f \in C_b^2$,

$$Ef(n^{-1}\sum_{k=1}^{n}\xi_k) - f(0) \to 0, \; n \to \infty. \tag{1}$$

Hence, assume $|f'(x+z) - f'(x)| \le g(|z|)$, with $0 \le g$ bounded, monotone increasing on $[0, +\infty)$ and $\lim_{z\to 0} g(|z|) = 0$.
II. Let $S_m := n^{-1}\sum_{k=1}^{m}\xi_k$, $1 \le m \le n$, and let us use "telescopic sums" (with $S_0 = 0$),

$$Ef(n^{-1}\sum_{k=1}^{n}\xi_k) - f(0) = \sum_{j=1}^{n} E(f(S_j) - f(S_{j-1})).$$

Proof ctd

III. We have, for every j, as $E(f(S_{j-1})\xi_j \mid S_{j-1}) = 0$,

$$\begin{aligned}
\big|E(f(S_j) - f(S_{j-1})) &= E(f(S_{j-1} + \xi_j/n) - f(S_{j-1}))\big| \\
&= \left|E\left(E(\int_0^1 f'(S_{j-1} + \alpha\xi_j/n)\,d\alpha\,\xi_j/n \mid S_{j-1})\right)\right|
\end{aligned}$$

$$\le E\left(E(\int_0^1 \left|f'(S_{j-1} + \alpha\xi_j/n) - f'(S_{j-1})\right| d\alpha\,|\xi_j|/n \mid S_{j-1})\right)$$

$$\begin{aligned}
&\le E\left(E(g(|\xi_j|/n)|\xi_j|/n \mid S_{j-1})\right) = n^{-1}E|\xi_1|g(|\xi_1|/n) \\
&= n^{-1}E|\xi_1|g(|\xi_1|/n)(1(|\xi_1|/n < \delta) + 1(|\xi_1|/n \ge \delta)).
\end{aligned}$$

Given $\epsilon > 0$, choose here $\delta > 0$ such that $g(\delta) \le \epsilon$. Then, $n^{-1}E|\xi_1|g(|\xi_1|/n)1(|\xi_1|/n < \delta) \le \epsilon\, E|\xi_1|/n$, while $n^{-1}E|\xi_1|1(|\xi_1|/n \ge \delta) \to 0$, by Lemma. This proves (1).

Characteristic functions

Weak LLN can be also regarded as a theorem about *weak convergence*, which was already implicitly used above in the first proof of LLN-2. This is only because LLN is convergence in probability *to some constant*, and not to a "nontrivial" r.v.

Def.: weak convergence $\xi_n \Longrightarrow \xi_0$ means that for every bounded continuous function f,

$$Ef(\xi_n) \to Ef(\xi_0), \quad n \to \infty.$$

Def.: characteristic function of a r.v. ξ is called the function

$$\varphi_\xi(\lambda) := E\exp(i\lambda\xi) \equiv E\sin(\lambda\xi) + iE\cos(\lambda\xi), \quad \lambda \in R.$$

Weak convergence

Theorem (on equivalence): $\xi_n \Longrightarrow \xi_0$ iff for every $\lambda \in R$,

$$\varphi_{\xi_n}(\lambda) \to \varphi_{\xi_0}(\lambda), \quad n \to \infty.$$

Sketch of the proof: approximate any continuous function (on any bounded closed interval) by trig polynomials.

Proof of LLN-2 via characteristic functions

Khintchine's Proof of LLN-2

The original proof is based on truncation method and Bienaimé-Chebyshev with variance.

Exponential convergence rate

The problem: well, $P(|n^{-1}\sum_{k=1}^n \xi_k - E\xi_1| \geq \epsilon) \to 0$, $n \to \infty$. But can we say what is the *rate of convergence?*

Exponential case: assume $\exists\ c > 0$ such that

$$E\exp(c|\xi_1|) < \infty. \tag{2}$$

Exponential rate theorem: *under (2), for any $\epsilon > 0$ there exists $\lambda > 0$ such that*

$$P(|n^{-1}\sum_{k=1}^n \xi_k - E\xi_1| \geq \epsilon) \leq C\exp(-\lambda n). \tag{3}$$

Moreover, if $\xi_1 \not\equiv$ const, then at least for $\epsilon > 0$ small enough, the l.h.s. in (3) admits a similar lower bound, $C'\exp(-\lambda' n)$.

Exponential convergence: proof

McMillan's Theorem

Polynomial convergence rate

Polynomial case: assume $\exists\ m = 1, 2, \ldots,$ such that

$$E|\xi_1|^{2m} < \infty. \tag{4}$$

Polynomial rate theorem: *under (4), for any* $\epsilon > 0$ *there exists* $C > 0$ *such that*

$$P(|n^{-1}\sum_{k=1}^{n}\xi_k - E\xi_1| \geq \epsilon) \leq Cn^{-m}. \tag{5}$$

Intermediate *sub*-exponential rates exist under conditions like $E\exp(c|\xi_1|^p) < \infty$, $0 < p < 1$. Under (2), there is no faster than exponential rate for small ϵ (because of the lower bounds in (3)); however, for large ϵ the probability in the l.h.s. may even decay to identical zero, say, if ξ_1 is *bounded*.

Polynomial convergence: proof

Kolmogorov's LLN

LLN–3 [Kolmogorov]: for IIDRV's with only $E\xi_1 < \infty$,

$$P\left(n^{-1}\sum_{k=1}^{n}\xi_k \to E\xi_1, \qquad n \to \infty\right) = 1.$$

The claim is apparently different: here not probabilities of some events tend to zero (for a proper comparison we should consider probabilities of complementary events, which tend to one), but some limit exists a.s.
Proof: is based on *the Borel–Kantelli Lemma* about convergence almost surely (i.e. with probability one). This type of convergence in Probability Theory is called *strong*, unlike *weaker* "convergence in probability".
Counterexample [who?]: if $E\xi_1 = \infty$, then
$P(\exists\, n' \to \infty,\ (n')^{-1}\sum_{k=1}^{n'}\xi_k \to \infty) = 1$.

Doob–Kolmogorov's inequalities

Proof of LLN–3

Etemadi’s Proof of LLN–3