spl13.tex
**Lecture 14. 7.11.2010**

**Lemma**. (i) If $X \geq 0$ has mean $\mu$ and distribution function $F$,

$$\sum_1^\infty P(|X| \geq n) \leq E|X| \leq 1 + \sum_1^\infty P(|X| \geq n).$$

(ii) $EX = \int_0^\infty (1 - F(x))dx$.

*Proof.* (i) For $i \geq 0$, let $A_i := \{i \leq X < i+1\}$. Then

$$\sum iP(A_i) \leq EX = \int XdP = \sum_i \int_{A_i} dP < \sum (i+1)P(A_i) = 1 + \sum_i iP(A_i).$$

But

$$\sum_i iP(A_i) = \sum_i \sum_{j=1}^i 1P(A_i) = \sum_j \sum_{i \geq j} P(A_i) = \sum_j P(X \geq j).$$

(ii) As the mean exists, $x(1 - F(x)) = \int_x^\infty xdF(u) \leq \int_x^\infty udF(u) \to 0$ (tail of a convergent integral), so $x(1 - F(x)) \to 0$. So

$$EX = \int XdP = \int_0^\infty xdF(x) \qquad \text{(by the transformation formula)}$$

$$= -\int_0^\infty xd(1-F(x)) = -[x(1-F(x))]_0^\infty + \int_0^\infty (1-F(x)dx = \int_0^\infty (1-F(x))dx,$$

integrating by parts. //

**10. The Strong Law of Large Numbers (SLLN).**

**Theorem (Strong Law of Large Numbers, Kolmogorov, 1933)**. For $X_n$ iid, $(X_1 + \ldots + X_n)/n$ converges to a constant $\mu$ a.s. as $n \to \infty$ iff $E|X| < \infty$, and then $\mu = EX$.

*Proof.* First take the case $X_n$ non-negative. If $E|X|(= EX) < \infty$, write $\mu$ for $EX$. Truncate $X_n$ at the value $n$ to obtain $Y_n$:

$$Y_n := X_n \quad (X_n < n), \quad 0 \quad (X_n \geq n).$$

By the Lemma,

$$\sum P(X_n \neq Y_n) = \sum P(X_n \geq n) = \sum P(X_1 \geq n) \leq EX_1 < \infty.$$

So by the first Borel-Cantelli lemma, a.s. only finitely many of the events $X_n \neq Y_n$) occur. So

$$\frac{1}{n}\sum_1^n (X_k - Y_k) \to 0 \qquad a.s.,$$

so it suffices to show that, writing $S_n := \sum_1^n Y_k$,

$$S_n/n = \frac{1}{n}\sum_1^n Y_k \to \mu \qquad a.s. \qquad (*)$$

Choose $q > 1$, and write $n_k$ for the integer part of $q^k$. Since $\sum 1/n_k^2$ is essentially a convergent geometric progression, it is at most a multiple of its first term:

$$\sum_m^\infty 1/n_k^2 \leq C/n_m^2$$

for some constant $C$. Also $n_{k+1}/n_k \to q$ as $k \to \infty$. For $q > 1$, $\epsilon > 0$,

$$\sum P(|S_{n_k} - E(S_{n_k})| > \epsilon) \leq \frac{1}{\epsilon^2}\sum\nolimits_k var(S_{n_k})/n_k^2, \qquad (**)$$

by Chebychev's inequality (L13). Variances add over independent summands, so $var S_n = \sum_1^n var Y_i \leq \sum_1^n E[Y_i^2]$. Substitute this into $(**)$ and change the order of summation on the right from $1 \leq i \leq n_k$ to first $k$ with $n_k \geq i$ and then over $i$. The inner sum gives at most $C/n_k^2 \leq C/i^2$. So

$$\sum P(|S_{n_k} - E(S_{n_k})|/n_k > \epsilon) \leq \frac{C}{\epsilon^2}\sum \frac{1}{i^2}E[Y_i^2].$$

Let $A_{ij} := (j - 1 \leq X_i < j)$; $P(A_{ij}) = P(A_{1j})$, as the $X_i$ are identically distributed. Note that (arguing as in the proof of the Integral Test for convergent series)

$$\sum_{i=j}^\infty 1/i^2 - 1/j^2 \leq \int_j^\infty dx/x^2 = 1/j \leq \sum_{i=j}^\infty 1/i^2 : \quad \sum_{i=j}^\infty 1/i^2 \leq 1/j + 1/j^2 \leq 2/j.$$

Now

$$\begin{aligned}
\sum \frac{1}{i^2} E[Y_i^2] &= \sum_1^\infty \frac{1}{i^2} \sum_{j=1}^{i} E[Y_i^2 I(A_{ij})] \\
&\leq \sum_1^\infty \frac{1}{i^2} \sum_{j=1}^{i} j^2 P(A_{ij}) \\
&= \sum_{j=1}^{\infty} j^2 P(A_{ij}). \sum_{i \geq j} 1/i^2 \\
&\leq \sum_{j=1}^{\infty} j^2 P(A_{ij}).2/j \leq 2[1 + EX] < \infty,
\end{aligned}$$

by the integral-test argument above and the Lemma, (i), as $EX < \infty$, given. Combining,

$$\sum P(|S_{n_k} - E(S_{n_k}|/n_k > \epsilon) < \infty,$$

and so

$$[S_{n_k} - E(S_{n_k}]/n_k \to 0 \qquad a.s. \qquad (k \to \infty),$$

by the first Borel-Cantelli lemma. Also

$$EY_n = E[X_n I_{\{X_n < n\}}] = E[X_1 I_{\{X_1 < n\}}] \to EX_1 = \mu \qquad (n \to \infty),$$

by monotone convergence. Averaging preserves convergence, so

$$\frac{1}{n_k} ES_{n_k} = \frac{1}{n_k} \sum_{i=1}^{n_k} E[Y_i] \to \mu \qquad (n \to \infty)$$

also. Combining,

$$S_{n_k}/n_k \to \mu \qquad (k \to \infty) \qquad a.s.$$

This proves the result along the 'nearly geometric' subsequence $n_k$. It remains to fill in the gaps. Since the $Y_n$ are non-negative, $S_n$ is non-decreasing. So for $n_k \leq m \leq n_{k+1}$,

$$\frac{S_{n_k}}{n_{k+1}} \leq \frac{S_m}{m} \leq \frac{S_{n_{k+1}}}{n_k}.$$

Let $k \to \infty$: since $S_{n_k}/n_k \to \mu$ and $n_{k+1}/n_k \to q$,

$$\mu/q \leq \liminf S_m/m \leq \limsup S_m/m \leq \mu q.$$

Letting $q \downarrow 1$ gives

$$S_m/m \to \mu,$$

which is $(*)$, completing the proof one way in the non-negative case. The general case follows by splitting into positive and negative parts, as usual.

Conversely, if $\Sigma_1^n X_k/n \to \mu$ a.s., then also $\Sigma_1^{n-1} X_k/n = [(n-1)/n].\Sigma_1^{n-1} X_k/(n-1) \to \mu$ also. Subtracting, $X_n/n \to 0$ a.s. Since the events $(|X_n| \geq n)$ are independent, the second Borel-Cantelli lemma gives

$$\sum P(|X_1| \geq n) = \sum P(|X_n| \geq n) < \infty.$$

This gives $E|X| < \infty$ by the Lemma. The conclusion of the first part now applies, and this completes the proof. //

*Note.* 1. Kolmogorov's SLLN of 1933 completes the story begun with Bernoulli's theorem in 1713. It gives precise form to the intuitive idea of the 'Law of Averages' – e.g., thinking about a probability as a long-run frequency. What this essentially says is that (thinking of a random variable as its mean plus a random error) independent errors tend to cancel. Any form of the LLN is really a result about *cancellation*.

2. Independence is not needed here. Strongly dependent errors need not cancel, but weakly dependent errors do (weak dependence can be made precise in many ways!). *Pairwise independence* suffices (N. Etemadi, 1981).

3. There are many proofs of SLLN. The one we give is adapted from [GS], Section 7.5. Others use Kolmogorov's inequality (a *maximal inequality*), or Kolmogorov's three-series theorem (for random series). The SLLN follows from the Martingale Convergence Theorem (below), or more simply from the Reversed Martingale Convergence Theorem ([S], 18.8). Another generalization of SLLN is the Pointwise Ergodic Theorem (due to Birkhoff and Khinchin, which originates in Statistical Mechanics). But the Ergodic Theorem is different: one can have a.s. convergence without the mean being finite.

4. With more moments finite, stronger results can be given (e.g., the Marcinkiewicz-Zygmund SLLN of 1937, with $p$ moments finite, $1 \leq p < 2$.

5. The SLLN generalizes in full to infinite-dimensional situations (Banach spaces).

6. The SLLN (in which we divide by $n$) and the CLT (in which we divide by $\sqrt{n}$ form two of the three main limit theorems of Probability Theory. The third is the Law of the Iterated Logarithm (LIL – Khinchin, 1924), which is intermediate: here we divide by $\sqrt{n \log \log n}$.