The Laws of Large Numbers

UNIFORM INTEGRABILITY AND CONVERGENCE CONCEPTS

I. Uniform Integrability (UI)
   A. $|X_n| < Y \in L^r$, $r > 0$, implies $\int_{[|X_n|>t]} |X_n|^r \, d\mathsf{P} \le \int_{[Y>t]} Y^r \, d\mathsf{P} \to 0$ as $t \to \infty$, uniformly in n. This is the *definition* of UI.
   B. X_n UI iff $\forall\epsilon\exists\delta \ni \int_\Lambda |X_n| \, d\mathsf{P} < \epsilon$ for $\mathsf{P}[\Lambda] \le \delta$ (take $\delta = \epsilon/2t$)
   C. $\mathsf{E}|X_n|^p \le c^p < \infty$ implies $|X_n|^r$ UI for each $r < p$... $\delta = (\epsilon/c)^q$, $\frac{1}{p} + \frac{1}{q} = 1$, $q = \frac{p}{p-1}$.
      1. Remark: *not* for $r = p$ (counterexample: $X_n = n1_{(0,1/n]}$)
   D. Main result (Thm 4.5.4, p97): If $X_n \to X$ *in pr.* then $|X_n|^r$ UI iff $X_n \to X$ in L^r iff $\mathsf{E}|X_n|^r \to \mathsf{E}|X|^r$.

II. Vague Convergence
   A. $X_n \to X$ *in pr.* iff $\forall n_k \exists n_{k_i} \ni X_{n_{k_i}} \to X$ *a.e.* (by contradiction)
   B. $X_n \to X$ *a.s.* and $\phi(x)$ continuous implies $\phi(X_n) \to \phi(X)$ *a.s.*
   C. $X_n \to X$ *in pr.* and $\phi(x)$ continuous implies $\phi(X_n) \to \phi(X)$ *in pr.* (use A)
   D. Definition: $X_n \Rightarrow X$ if $\mathsf{E}\phi(X_n) \to \mathsf{E}\phi(X)$ $\forall \phi \in C_b(\mathbb{R})$
      1. Prop: $X_n \to X$ *in pr.* implies $X_n \Rightarrow X$ (use II.C)
      2. Prop: $X_n \Rightarrow X$ implies $F_n(r) \to F(r)$ wherever $F(r) = F(r-)$.
         a. Remark: Even if $X_n \Rightarrow X$, $F_n(r)$ may not converge where $F(r)$ jumps;
         b. Remark: Even if $X_n \Rightarrow X$, $f_n(r) = F_n'(r)$ may not converge to $f(r) = F'(r)$; in fact, either may fail to exist.

III. Implications among these notions: *a.e.*, *in pr.*, L^r, L^p, L^∞, *vg.* ($0<r<p<\infty$):
   A. *a.e.* $\Longrightarrow$ *in pr.* (by Easy Borel-Cantelli)
      1. *in pr.* $\Longrightarrow$ *a.e.* along subsequences
      2. *in pr.* $\not\Longrightarrow$ *a.e.* (counterexample: $X_n(\omega) = 1_{(i/2^j,(i+1)/2^j]}(\omega)$, $n = i + 2^j$)
   B. $L^p \Longrightarrow$ *in pr.* (by Chebychev's inequality)
      1. *in pr.* $\Longrightarrow L^p$ under Uniform Integrability
      2. *in pr.* $\not\Longrightarrow L^p$ (counterexample: $X_n = n^{1/p}1_{(0,1/n]}$)
   C. $L^p \Longrightarrow L^r$ (by Jensen's inequality)
      1. $L^r \not\Longrightarrow L^p$ (counterexample: $X_n = n^{1/p}1_{(0,1/n]}$)
   D. $L^\infty \Longrightarrow L^p$ (simple estimate)
      1. $L^p \not\Longrightarrow L^\infty$ (counterexample: $X_n = n^{1/2p}1_{(0,1/n]}$)
   E. $L^\infty \Longrightarrow$ *a.e.* (uniform cgce implies pointwise cgce)
   F. *in pr.* $\Longrightarrow$ *vg.* (II.D.1 above)
      1. *vg.* $\not\Longrightarrow$ *in pr.* (counterexample: X_n, X on different spaces)
      2. *vg.* $\Longrightarrow$ *a.s.* ($\exists(\Omega, \mathcal{F}, \mathsf{P})$, $X_n, X \ni X_n \to X$ *a.e.*...)

INFINITE COIN-TOSS AND THE LAWS OF LARGE NUMBERS

The traditional interpretation of the *probability* of an event E is its *asymptotic frequency*: the limit as $n \to \infty$ of the fraction of n repeated, similar, and independent trials in which E occurs. Similarly the "expectation" of a random variable X is taken to be its *asymptotic average*, the limit as $n \to \infty$ of the average of n repeated, similar, and independent replications of X. As statisticians trying to make inference about the underlying probability distribution $f(x|\theta)$ governing observed random variables X_i, this suggests that we should be interested in the probability distribution for large n of quantities like the average of the RV's, $\frac{1}{n}\sum_{i=1}^n X_i$.

Three of the most celebrated theorems of probability theory concern this sum. For independent random variables X_i, all with the same probability distribution satisfying $\mathsf{E}|X_i|^3 < \infty$, set $\mu = \mathsf{E}X_i$, $\sigma^2 = \mathsf{E}|X_i - \mu|^2$, and $S_n = \sum_{i=1}^n X_i$. The three main results are:

Laws of Large Numbers:

$$\frac{S_n - n\mu}{\sigma n} \longrightarrow 0 \qquad (\textit{in pr.} \text{ and } \textit{a.s.})$$

Central Limit Theorem:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow N(0,1) \qquad (\textit{vg.})$$

Law of the Iterated Logarithm:

$$\limsup \pm \frac{S_n - n\mu}{\sigma\sqrt{2n\log\log n}} = 1.0 \qquad (\textit{a.s.})$$

Together these three give a clear picture of how quickly and in what sense $\frac{1}{n}S_n$ tends to μ. We begin with the Law of Large Numbers (LLN), in its "weak" form (asserting convergence *in pr.*) and in its "strong" form (convergence *a.s.*). There are several versions of both theorems. The simplest requires the X_i to be IID and L^2; stronger results allow us to weaken (but not eliminate) the independence requirement, permit non-identical distributions, and consider what happens if the RV's are only L^1 (or worse!) instead of L^2.

The text covers these things well; to complement it I am going to: (1) Prove the simplest version, and with it the Borel-Cantelli theorems; and (2) Show what happens with Cauchy random variables, which don't satisfy the requirements (the LLN fails).

IV. Weak version, non-iid, L^2: $\mu_i = \mathsf{E}X_i$, $\sigma_{ij} = \mathsf{E}[X_i - \mu_i][X_j - \mu_j]$
   A. $Y_n = (S_n - \Sigma\mu_i)/n$ satisfies $\mathsf{E}Y_n = 0$, $\mathsf{E}Y_n^2 = \frac{1}{n^2}\Sigma_i\sigma_{ii} + \frac{2}{n^2}\Sigma_{i<j}\sigma_{ij}$;
      1. If $\sigma_{ii} \le M$ and $\sigma_{ij} \le 0$, Chebychev $\Longrightarrow Y_n \to 0$, *in pr.*
      2. (pairwise) IID L^2 is OK
V. Strong version, non-iid, L^2: $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 \le M$, $\mathsf{E}X_iX_j \le 0$.
   A. $\mathsf{P}[|S_n| > n\epsilon] < \frac{Mn}{n^2\epsilon^2} = \frac{M}{n\epsilon^2}$
      1. $\mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M}{n^2\epsilon^2}$, $\Sigma_n\mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M\pi^2}{6\epsilon^2}$
      2. Borel-Cantelli: $\mathsf{P}[|S_{n^2}| > n^2\epsilon \textit{ i.o.}] = 0$, $\frac{1}{n^2}S_{n^2} \to 0$ *a.s.*
      3. $D_n = \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$, $\mathsf{E}D_n^2 \le 2n\mathsf{E}|S_{(n+1)^2} - S_{n^2}| \le 4n^2M$
      4. Chebychev: $\mathsf{P}[D_n > n^2\epsilon] < \frac{4n^2M}{n^4\epsilon^2}$, $D_n \to 0$ *a.s.*
   B. $|S_k/k| \le \frac{|S_{n^2}| + D_n}{n^2} \to 0$ *a.s.*, QED
      1. Bernoulli RV's, normal number theorem, Monte Carlo.
VI. Weak version, pairwise-iid, L^1
   A. Equivalent sequences: $\sum_n \mathsf{P}[X_n \ne Y_n] < \infty$
      1. $\sum_n [X_n - Y_n] < \infty$ *a.s.*
      2. $\sum_{i=1}^n [X_i]$, $a_n\sum_{i=1}^n [X_i]$ converge iff $\sum_{i=1}^n [Y_i]$, $a_n\sum_{i=1}^n [Y_i]$ do
      3. $Y_n = X_n 1_{[|X_n| \le n]}$
VII. Counterexamples: Cauchy,
   A. $X_i \sim \frac{dx}{\pi[1+x^2]} \Longrightarrow \mathsf{P}[|S_n|/n \le \epsilon] \equiv \frac{2}{\pi}\tan^{-1}(\epsilon) \not\to 1$, WLLN fails.
   B. $\mathsf{P}[X_i = n] = \frac{\pm c}{n^2}$, $n \ge 1$; $X_i \notin L^1$, and $S_n/n \not\to 0$ *in pr.* or *a.s.*
   C. $\mathsf{P}[X_i = n] = \frac{\pm c}{n^2 \log n}$, $n \ge 3$; $X_i \notin L^1$, but $S_n/n \to 0$ *in pr.* and not *a.s.*
   D. Medians: for ANY RV's $X_n \to X_\infty$ *in pr.*, then $m_n \to m_\infty$ if m_∞ is unique.

Let X_i be iid standard Cauchy RV's, with

$$\mathsf{P}[X_1 \le t] = \int_{-\infty}^{t} \frac{dx}{\pi[1+x^2]} = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$$

and characteristic function

$$\mathsf{E}\, e^{i\lambda X_1} = \int_{-\infty}^{\infty} e^{i\lambda x} \frac{dx}{\pi[1+x^2]} = e^{-|\lambda|},$$

so S_n/n has characteristic function

$$\mathsf{E}\, e^{i\lambda S_n/n} = \mathsf{E}\, e^{i\frac{\lambda}{n}[X_1+\cdots+X_n]} = \left(\mathsf{E}\, e^{i\frac{\lambda}{n}X_1}\right)^n = (e^{-|\frac{\lambda}{n}|})^n = e^{-|\lambda|}$$

and S_n/n also has the standard Cauchy distribution with $\mathsf{P}[S_n/n \le t] = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$; in particular, S_n/n does not converge almost surely, in probability, or even in distribution.