Math 617: Day 48

Prop: If $\{X_1, \ldots, X_n\}$ are independent r.v.'s & each $X_i \in L^1$, then $\prod_{i=1}^n X_i \in L^1$ & $E(\prod X_i) = \prod E(X_i)$.

Cor: If $\{X_1, \ldots, X_n\}$ are independent & in L^2, then $\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n)$.

In particular, if the X_i are independent & *identically distributed* (i.e., $P_{X_i} = P_{X_j}$ $\forall\, i,j$) w/ common variance $\sigma^2(X_i) = \sigma^2$, then

$$\sigma^2(X_1 + \cdots + X_n) = n\,\sigma^2$$

If, additionally, the X_i have mean 0, then $\sigma^2 = \sigma^2(X_i) = E(X_i^2) - E(X_i)^2 = E(X_i^2)$, so $\sigma^2\left(\frac{X_i}{\sqrt{n}}\right) = E\left(\left(\frac{X_i}{\sqrt{n}}\right)^2\right) = \frac{1}{n} E(X^2) = \frac{\sigma^2}{n}$

so $\sigma^2\left(\frac{1}{\sqrt{n}} \sum X_i\right) = \sum \sigma^2\left(\frac{X_i}{\sqrt{n}}\right) = \sum \frac{\sigma^2}{n} = \sigma^2$.

As we will see in the Central Limit Thm, $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ actually converges in probability (i.e., in measure) to the Gaussian w/ mean 0 & variance σ^2.

Markov's Inequality: If X is real-valued r.v., then $P(|X| \geq \lambda) \leq \frac{1}{\lambda} E(|X|)$.

Pf: Clearly, for any λ, $\lambda \chi_{\{|X| \geq \lambda\}} \leq |X|$. Taking expectations gives $\lambda P(|X| \geq \lambda) \leq E(|X|)$, so the result follows. ▨

This leads to the very simple (but still useful) *large deviation inequality*:

Chebyshev's Inequality: If X is a scalar-valued r.v., then $P(|X - E(X)| \geq \lambda) \leq \frac{\sigma^2(X)}{\lambda^2}$.

Pf: We can alternatively write $\sigma^2(X) = E(|X - E(X)|^2)$. Then $P(|X - E(X)| \geq \lambda) = P(|X - E(X)|^2 \geq \lambda^2) \leq \frac{\sigma^2(X)}{\lambda^2}$ by Markov's inequality. ▨

Weak Law of Large Numbers: Let $(X_i)_{i=1}^\infty$ be a seq. of independent L^2 scalar (i.e., real- or complex-valued) r.v.'s w/ means (μ_i) & variances (σ_i^2). If $\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \to 0$ as $n \to \infty$, then $\frac{1}{n} \sum (X_i - \mu_i) \to 0$ in probability.

In other words, $E\left(\frac{1}{n} \sum X_i\right) \to \frac{1}{n} \sum E(X_i)$ ... the expectation of the mean goes to the mean of the expectations.

Proof: $\frac{1}{n} \sum (X_i - \mu_i)$ clearly has mean 0 & variance $\frac{1}{n^2} \sum \sigma_i^2$ (by the Corollary). If $\varepsilon > 0$, Chebyshev's Inequality says

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - \mu_i)\right| > \varepsilon\right) \leq \frac{\frac{1}{n^2} \sum \sigma_i^2}{\varepsilon^2} = \frac{\sum \sigma_i^2}{n^2 \varepsilon^2} \to 0 \text{ as } n \to \infty$$

But this is exactly what it means to converge in measure. ▨

To prove the Strong Law of Large Numbers (which says the same thing, but w/ convergence in probability replaced by convergence almost everywhere), we will need a couple of lemmas:

Borel-Cantelli Lemma: Suppose $E_1, E_2, \dots$ are a seq. of events.

① If $\sum P(E_i) < \infty$, then $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n\right) = 0$. In other words, the probability that infinitely many events occur simultaneously is 0.

② If the E_i are independent (i.e., the X_{E_i} are independent random variables) & $\sum P(E_i) = \infty$, then $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n\right) = 1$ (i.e., almost every outcome corresponds to infinitely many events).

Pf: ① $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n\right) \le P\left(\bigcup_{n=k}^{\infty} E_n\right) \le \sum_{n=k}^{\infty} P(E_n) \to 0$ as $k \to \infty$ since $\sum_{n=1}^{\infty} P(E_n)$ converges.

② WTS $0 = P\left(\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n\right)^c\right) = P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} E_n^c\right)$, so it suffices to show $P\left(\bigcap_{n=k}^{\infty} E_n^c\right) = 0 \ \forall k$.

But the E_n are independent, which implies the E_n^c are independent$^{(*)}$, so

$$P\left(\bigcap_{n=k}^{N} E_n^c\right) = P\left(\prod_{n=k}^{N} X_{E_n^c} = 1\right) = \prod_{n=k}^{N} P(X_{E_n^c} = 1) = \prod_{n=k}^{N} P(E_n^c) = \prod_{n=k}^{N} (1 - P(E_n)) \le \prod_{n=k}^{N} \left(e^{-P(E_n)}\right) = e^{-\sum_{n=k}^{N} P(E_n)} \to 0. \ \blacksquare$$

(*) can be proved either by induction (as hinted in the book), or using the principle of inclusion/exclusion:

$$P(E_{n_1}^c \cap \dots \cap E_{n_k}^c) = 1 - P(E_{n_1} \cup \dots \cup E_{n_k}) = 1 - \left[\sum_{i=1}^{k} P(E_{n_i}) - \sum_{i,j} P(E_{n_i}) P(E_{n_j}) + \sum_{i,j,l} P(E_{n_i}) P(E_{n_j}) P(E_{n_l}) - \dots\right]$$

$$= \prod_{i=1}^{k} (1 - P(E_{n_i})) = \prod_{i=1}^{k} (P(E_{n_i}^c)).$$