CPSC 550: Machine Learning II	2008/9 Term 2

Lecture 7 — Feb 12, 2009

Lecturer: Nando de Freitas | *Scribe: Matt Hoffman*

Notation

- If S is a set, $\overline{S}$ denotes its complement.
- $X_n \xrightarrow{a.s.} X$ means the sequence X_n converges *almost surely* to X.

7.1 Refresher (from last lecture)

Given a sequence of sets $\{X_n\}$, we'll define the set $\{X_n \text{ i.o.}\}$ as the set of values which occur *infinitely often*, i.e.

$$\{X_n \text{ i.o.}\} = \{w \colon w \in X_m \text{ for an infinite number of indices } m\}.$$

More precisely we can write this using the *limit superior*,

$$\{X_n \text{ i.o.}\} = \limsup_{n\to\infty} X_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} X_m = \lim_{n\to\infty} \bigcup_{m=n}^{\infty} X_m \text{ where}$$

$$\bigcup_{m=n}^{\infty} X_m = \sup\{X_m \colon n \le m < \infty\}.$$

7.2 Borel-Cantelli and the Strong Law of Large Numbers

With the definitions from the last lecture, we can state the following lemma.

Lemma 7.1 (Borel-Cantelli). *Given a sequence of events A_n, if the sum over the probabilities of these events is finite, then the probability that infinitely many of these events occur is 0. In other words,*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P(A_n \textit{ i.o.}) = 0.$$

Proof: The probability of events occurring infinitely often is given by

$$\begin{aligned} P(A_n \text{ i.o.}) = P(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m) &= \lim_{n\to\infty} P(\bigcup_{m=n}^{\infty} A_m) \\ &\le \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) \quad \textit{(via the union bound)} \\ &= \sum_{m=1}^{\infty} P(A_m) - \lim_{n\to\infty} \sum_{m=1}^{n} P(A_m) = 0, \end{aligned}$$

where both of these sums are finite by assumption. □

Example. Let $X_n \in \{0,1\}$ be a sequence of independent random variables such that the sum over probabilities $\sum_{n=1}^{\infty} P(X_n = 1) < \infty$ is finite. By the Borel-Cantelli lemma $P(X_n = 1 \text{ i.o.}) = 0$, and moreover

$$\overline{\{X_n = 1 \text{ i.o.}\}} = \{w\colon \lim_{n\to\infty} X_n(w) = 0\},$$

therefore $X_n \xrightarrow{a.s.} 0$, i.e. it converges to 0 almost surely.

We can now show that the Borel-Cantelli lemma with respect to the errors of some estimate can be used to show almost sure convergence.

Theorem 7.2. *A sequence $\{Z_n\}$ of random variables converges almost surely to Z iff for all $\epsilon > 0$ the probability of an error greater than ϵ occurring infinitely often is zero, i.e.*

$$Z_n \xrightarrow{a.s.} Z \iff \forall \epsilon > 0,\ P(|Z_n - Z| \ge \epsilon \textit{ i.o.}) = 0.$$

Proof: We will first assume that $Z_n \xrightarrow{a.s.} Z$ and prove that the probability of an infinite number of errors is zero. For any ϵ, we can construct the set

$$\{|Z_n - Z| \ge \epsilon \text{ i.o.}\} \subset \overline{\{w\colon \lim_{n\to\infty} Z_n(w) = Z(w)\}},$$

where the right-most set contains all the ways in which Z_n doesn't converge to Z. We can then write the probability of this set as

$$P(|Z_n - Z| \ge \epsilon \text{ i.o.}) \le 1 - P(\lim_{n\to\infty} Z_n(w) = Z(w)) = 0,$$

where this equality holds as a result of our assumption.

We will now show that the implication holds in the other direction, namely by assuming that $P(|Z_N - Z| \ge \epsilon \text{ i.o.}) = 0$ for all ϵ and showing that almost sure convergence holds. Let N_k for $k \ge 1$ be the last n such that $|Z_n - Z| \ge \frac{1}{k}$, and in particular we'll define $N_k = \infty$ if no such n exists. Since our assumption holds for all ϵ, we can let $\epsilon = \frac{1}{k}$ and for any k write

$$P(|Z_n - Z| \geq \tfrac{1}{k} \text{ i.o.}) = 0 \implies P(N_k = \infty) = 0.$$

Hence, $P(\lim_{n\to\infty} Z_n(w) = Z(w)) = 1$, and $Z_n \xrightarrow{a.s.} Z$. □

We can now get around to our main reason for introducing the Borel-Cantelli lemma, namely the strong law of large numbers. We will use Borel's version of the SLLN.

Theorem 7.3 (Strong law of large numbers). *If $\{X_n\}$ is a sequence of i.i.d. random variables with finite expected value $\mathbb{E}[X_1] < \infty$, then in the limit the average of these variables converges almost surely to $\mathbb{E}[X_1]$, i.e.*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1].$$

Proof: Let $m = \mathbb{E}[X_1]$ and $S_n = \sum_{i=1}^{n} X_i$. We can write the probability of deviations as

$$\begin{aligned} P\big(|\tfrac{1}{n}S_n - m| \geq \epsilon\big) &= P\big((\tfrac{1}{n}S_n - m)^4 \geq \epsilon^4\big) \\ &\leq \epsilon^{-4}\,\mathbb{E}\big[(\tfrac{1}{n}S_n - m)^4\big] \qquad \text{(by Markov's inequality)} \\ &\leq (n\epsilon)^{-4}\,\mathbb{E}\big[(\textstyle\sum_{i=1}^{n} X_i - m)^4\big]. \end{aligned}$$

As a result, the following implication holds,

$$\begin{aligned} \mathbb{E}\big[(\tfrac{1}{n}S_n - m)^4\big] \leq Kn^2 &\implies P\big(|\tfrac{1}{n}S_n - m| \geq \epsilon\big) \leq \frac{K}{n^2\epsilon^4} \\ &\implies \sum_n P\big(|\tfrac{1}{n}S_n - m| \geq \epsilon\big) \leq \sum_n \frac{K}{n^2\epsilon^4} < \infty. \end{aligned}$$

If this final statement holds, then via Borel-Cantelli it must hold that our sample mean converges almost surely to the true expected value, $\frac{1}{n}S_n \xrightarrow{a.s.} m$. It remains to show that this hypothesis holds, i.e. that for $\mathbb{E}\big[(\sum_{i=1}^{n} Y_i)^4\big] \leq Kn^2$, where we have defined $Y_i = X_i - m$. Because m is the expectation of X_i for all i, we can easily see that

$$\mathbb{E}[Y_i] = 0, \quad \mathbb{E}[Y_1 Y_2 \cdots] = 0, \quad \mathbb{E}[Y_1 Y_2^2 \cdots] = 0, \quad \ldots$$

i.e. the expectation is 0 for every term containing a "singleton". As a result the only terms remaining are those of the form $\mathbb{E}[Y_i^4]$ and $\mathbb{E}[Y_i^2 Y_j^2]$, for which there are n terms of the first kind and $3n(n-1)$ of the second (via a trivial combinatorial argument). Hence the expectation is given by

$$\mathbb{E}\big[(\sum_{i=1}^{n} Y_i)^4\big] = n\mathbb{E}[Y_i^4] + 3n(n-1)\mathbb{E}[Y_1^2 Y_2^2] \leq Kn^2 \text{ for some } K.$$

This only relies on the expectations being finite, so our proof must hold. □