

PROBABILITY THEORY

MATH 154

Unit 14: Weak law of large numbers

14.1. A sequence of random variables X_k defines a **stochastic process**. There does not need to be anything stochastic about it. X_k could be the same for example. Or take $Y_k(x) = \cos(2\pi(x + k\sqrt{2}))$ on $([0, 1], \mathcal{A}, dx)$, which shows correlations. The cumulative process $S_n = X_1 + X_2 + \cdots + X_n$ has a chance to converge if we scale it: S_n/n . For example, if X_i is the outcome of a dice, then S_n/n is the average of all n dice. Experience tells us that it is the average of the distribution: $m = (1 + 2 + 3 + 4 + 5 + 6)/6 = 21/6 = 3.5$. Under which conditions can we justify this? We will see later that even in the just mentioned cos process, we still have $\frac{1}{n} \sum_{k=1}^n Y_k \rightarrow 0 = E[Y_k]$.

14.2. The weak law of large numbers holds for pairwise uncorrelated random variables. This is a remarkably weak assumption.

Theorem 1 (Golden Theorem). *Assume $X_i \in \mathcal{L}^2$ are pairwise uncorrelated, have a common mean $E[X_i] = m$ and $M = \sup_n \frac{1}{n} \sum_{i=1}^n \text{Var}[X_i] < \infty$. Then $\frac{S_n}{n} \rightarrow m$ in probability.*

Proof. Since $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}[X, Y]$ and X_n are pairwise uncorrelated, we get $\text{Var}[X_n + X_m] = \text{Var}[X_n] + \text{Var}[X_m]$ and by induction $\text{Var}[S_n] = \sum_{i=1}^n \text{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\text{Var}\left[\frac{S_n}{n}\right] = E\left[\frac{S_n^2}{n^2}\right] - \frac{E[S_n]^2}{n^2} = \frac{\text{Var}[S_n]}{n^2} = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] .$$

The right hand side converges to zero for $n \rightarrow \infty$. With Chebychev's inequality we obtain

$$P\left[\left|\frac{S_n}{n} - m\right| \geq \epsilon\right] \leq \frac{\text{Var}\left[\frac{S_n}{n}\right]}{\epsilon^2} \leq \frac{M}{n\epsilon^2} .$$

□

14.3. As an application in analysis, this leads to a constructive proof of a **theorem of Weierstrass** which states that polynomials are dense in the space $C[0, 1]$ of all continuous functions on the interval $[0, 1]$.

Theorem 2. *For every $f \in C[0, 1]$, the **Bernstein polynomials***

$$B_n(x) = \sum_{k=1}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

converge uniformly to f . If $f(x) \geq 0$, then also $B_n(x) \geq 0$.

Proof. For $x \in [0, 1]$, let X_n be a sequence of independent $\{0, 1\}$ -valued random variables with mean value x . In other words, we take the probability space $(\{0, 1\}^{\mathbb{N}}, \mathcal{A}, \mathbb{P})$ defined by $\mathbb{P}[\omega_n = 1] = x$. Since $\mathbb{P}[S_n = k] = \binom{n}{k} x^k (1-x)^{n-k}$, we can write $B_n(x) = \mathbb{E}[f(\frac{S_n}{n})]$. We estimate with $\|f\| = \max_{0 \leq x \leq 1} |f(x)|$

$$\begin{aligned} |B_n(x) - f(x)| &= |\mathbb{E}[f(\frac{S_n}{n})] - f(x)| \leq \mathbb{E}[|f(\frac{S_n}{n}) - f(x)|] \\ &\leq 2\|f\| \cdot \mathbb{P}[|\frac{S_n}{n} - x| \geq \delta] \\ &\quad + \sup_{|x-y| \leq \delta} |f(x) - f(y)| \cdot \mathbb{P}[|\frac{S_n}{n} - x| < \delta] \\ &\leq 2\|f\| \cdot \mathbb{P}[|\frac{S_n}{n} - x| \geq \delta] + \sup_{|x-y| \leq \delta} |f(x) - f(y)|. \end{aligned}$$

The second term in the last line is called the **continuity module** of f . It converges to zero for $\delta \rightarrow 0$. By the Chebychev inequality and the proof of the weak law of large numbers, the first term can be estimated from above by

$$2\|f\| \frac{\text{Var}[X_i]}{n\delta^2},$$

a bound which goes to zero for $n \rightarrow \infty$ because the variance satisfies $\text{Var}[X_i] = x(1-x) \leq 1/4$. $\square$

14.4. In the weak law of large numbers, we only assumed the random variables to be uncorrelated. Under the stronger condition of independence and the moment assumption $X^4 \in \mathcal{L}^1$, the convergence can be accelerated:

Theorem 3 (Complete golden theorem). *Assume $X_i \in \mathcal{L}^4$ have common expectation $\mathbb{E}[X_i] = m$ and satisfy $M = \sup_n \|X\|_4 < \infty$. If X_i are independent, then $S_n/n \rightarrow m$ in probability. We even have complete convergence: $\sum_{n=1}^{\infty} \mathbb{P}[|\frac{S_n}{n} - m| \geq \epsilon] < \infty$ for all $\epsilon > 0$.*

Proof. We can assume without loss of generality that $m = 0$. Because the X_i are independent, we get

$$\mathbb{E}[S_n^4] = \sum_{i_1, i_2, i_3, i_4=1}^n \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}].$$

Again by independence, a summand $\mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$ is zero if an index $i = i_k$ occurs alone, it is $\mathbb{E}[X_i^4]$ if all indices are the same and $\mathbb{E}[X_i^2] \mathbb{E}[X_j^2]$, if there are two pairwise equal indices. Since by Jensen's inequality $\mathbb{E}[X_i^2]^2 \leq \mathbb{E}[X_i^4] \leq M$ we get

$$\mathbb{E}[S_n^4] \leq nM + n(n-1)M.$$

Use now the Chebyshev-Markov inequality with $h(x) = x^4$ to get

$$\mathbb{P}[|\frac{S_n}{n}| \geq \epsilon] \leq \frac{\mathbb{E}[(S_n/n)^4]}{\epsilon^4} \leq M \frac{n + n^2}{\epsilon^4 n^4} \leq 2M \frac{1}{\epsilon^4 n^2}.$$

$\square$