Mathematical methods in communication July 2nd, 2009

Appendix 1

Lecturer: Haim Permuter *Scribes: Liran Sharoni and Zahi Gradstein*

I. Convergence of random variables

A. Convergence in distribution

Suppose that $F_1, F_2, ...$ is a sequence of cumulative distribution functions corresponding to random variables $X_1, X_2, ...$, and that F is a distribution function corresponding to a random variable X. We say that the sequence X_n converges towards X in distribution, if:

$$\lim_{n\to\infty} F_n(a) = F(a) \tag{1}$$

for every real number a at which F is continuous function continuous. Since $F(a) = P(X \leq a)$, this means that the probability that the value of X is in a given range is very similar to the probability that the value of X_n is in that range, provided n is sufficiently large. Convergence in distribution is often denoted by adding the letter $\mathcal{D}$ over an arrow indicating convergence:

$$X_n \xrightarrow{\mathcal{D}} X \tag{2}$$

Convergence in distribution is the weakest form of convergence. it is implied by all other modes of convergence stated below, and hence, it is the most common and often the most useful form of convergence of random variables.

B. Convergence in probability

To say that the sequence X_n of random variables converges towards X in probability means:

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0 \tag{3}$$

for every $\varepsilon > 0$. Formally, pick any $\varepsilon > 0$ and any $\delta > 0$. Let P_n be the probability that X_n is outside a tolerance ε of X. Then, if X_n converges in probability to X then there exists a value N such that, for all $n \geq N$, P_n is itself less than δ.

Convergence in probability is often denoted by adding the letter P over an arrow indicating convergence:

$$X_n \xrightarrow{\mathbb{P}} X \tag{4}$$

Convergence in probability implies convergence in distribution. Convergence in probability is the notion of convergence used in the weak law of large numbers.

There are more kinds of Convergences such as :

- ”Almost sure convergence”.
- ”Sure convergence”.
- ”Convergence in mean”.

But the only type relevant to the Course is Convergence in probability .

II. Markov's Inequality

Theorem 1 (Markov's Inequality.)

For a random variable X, a function g : $\mathbb{R} \to \mathbb{R}$ and $\mathbb{E}(\mathrm{g}(X) < \infty$ s.t

$$\mathrm{P}(|\mathrm{g}(X)| \geq a) \quad \leq \quad \frac{\mathbb{E}(|\mathrm{g}(X)|)}{a} \ , \ \forall a > 0 \tag{5}$$

Proof:

First of all, let us define an Indicator function

$$\mathbf{1}_A \triangleq \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

then

$$a\mathbf{1}_{(|\mathrm{g}(X|)\geq a)} \quad \overset{(a)}{=} \quad \begin{cases} a & \text{if } |\mathrm{g}(X)| \geq a \\ 0 & \text{if } |\mathrm{g}(X)| < a \end{cases} \tag{6}$$

$$\overset{(b)}{\leq} \quad |\mathrm{g}(X)| \tag{7}$$

Where:

(a) Where the event 'A' is : $|\mathrm{g}(X)| \geq a$

(b)

- When $|\mathrm{g}(X)| \geq a$ the indicator's value is a and then the inequality $|\mathrm{g}(X)| \geq a$ is confirmed.
- When $|\mathrm{g}(X)| < a$ the indicator's value is 0, meaning $|\mathrm{g}(X)| \geq 0$, which is true $\forall \mathrm{g}(X)$.

applying expectation on both sections of the inequality we get:

$$\mathbb{E}[a\mathbf{1}_{(|\mathrm{g}(X)|\geq a)}] \quad = \quad a\mathbb{E}[\mathbf{1}_{(|\mathrm{g}(X)|\geq a)}] \tag{8}$$

$$\overset{(c)}{=} \quad a\mathrm{P}(|\mathrm{g}(X)| \geq a)\cdot 1 + a\mathrm{P}(|\mathrm{g}(X)| < a)\cdot 0 \tag{9}$$

$$= \quad a\mathrm{P}(|\mathrm{g}(X)| \geq a) \tag{10}$$

$$\overset{(d)}{\leq} \quad \mathbb{E}(|\mathrm{g}(X)|) \tag{11}$$

Where:

(c) By definition of expectation.

(d) Expectation of the right side of the inequality.

Hence:

$$\mathrm{P}(|\mathrm{g}(X)| \geq a) \quad \leq \quad \frac{\mathbb{E}(|\mathrm{g}(X)|)}{a} \tag{12}$$

III. Chebyshev's Inequality

Theorem 2 (Chebyshev's Inequality.)

For a random variable X, a function $\mathrm{g} : \mathbb{R} \rightarrow \mathbb{R}$ and $\mathbb{E}(\mathrm{g}(X) < \infty$ s.t

$$\mathrm{P}(|X - \mu| \geq \varepsilon) \quad \leq \quad \frac{Var(X)}{\varepsilon^2},\ \forall \varepsilon > 0 \tag{13}$$

Proof:

Applying a specific $\mathrm{g}(X) = (X - \mu)^2$ and $a = \varepsilon^2$, on Markov's inequality

we get:

$$\mathrm{P}(|X - \mu| \geq \varepsilon) \quad \leq \quad \frac{Var(X)}{\varepsilon^2},\ \forall \varepsilon > 0 \tag{14}$$

IV. Law Of Large Numbers

Theorem 3 (Law Of Large Numbers.) Given $X_1, X_2, ...$an infinite sequence of i.i.d. random variables with finite expected value $\mathbb{E}(X_1) = \mathbb{E}(X_2) = ... = \mu < \infty$, we are interested in the convergence of the sample average

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n). \tag{15}$$

The weak law of large numbers states:

$$\overline{X}_n \xrightarrow{p} \mu \qquad \text{for} \qquad n \rightarrow \infty \tag{16}$$

Proof: (using Chebyshev's inequality)

This proof uses the assumption of finite variance $Var(X_i) = \sigma^2$, $\forall i$.The independence of the random variables implies no correlation between them, and we have that

$$\begin{aligned}
Var(\overline{X}_n) &\stackrel{(a)}{=} Var(\tfrac{1}{n}(X_1 + \cdots + X_n)) && (17)\\
&\stackrel{(b)}{=} \frac{1}{n^2} Var(X_1 + \cdots + X_n) && (18)\\
&\stackrel{(c)}{=} \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n)) && (19)\\
&\stackrel{(d)}{=} \frac{n\sigma^2}{n^2} && (20)\\
&= \frac{\sigma^2}{n}. && (21)
\end{aligned}$$

Where:

(a) By definition : $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

(b) Using Variance's quality :$Var(aX) = a^2 Var(X)$.

(c) $X_1, X_2, ..., X_n$ are i.i.d.

(d) $Var(X_i) = \sigma^2,\ \forall i$.

The common mean μ of the sequence is the mean of the sample average:

$$\mathbb{E}(\overline{X}_n) = \mu. \tag{22}$$

Using Chebyshev's inequality on $\overline{X}_n$ results in

$$P(\left|\overline{X}_n - \mu\right| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}. \tag{23}$$

This may be used to obtain the following:

$$P(\left|\overline{X}_n - \mu\right| < \varepsilon) \quad = \quad 1 - P(\left|\overline{X}_n - \mu\right| \geq \varepsilon) \tag{24}$$

$$\geq \quad 1 - \frac{\sigma^2}{n\varepsilon^2}. \tag{25}$$

As n approaches infinity, the expression approaches 1. And by definition of convergence in probability, we have obtained:

$$\overline{X}_n \xrightarrow{P} \mu \qquad \text{for} \qquad n \to \infty. \tag{26}$$

V. History and Background

A. LLN

- An Italian Gerolamo Cardano and Indian mathematician Brahmagupta stated without proof that the accuracies of empirical statistics tend to improve with the number of trials. This was then formalized as a law of large numbers.The LLN was first proved by Jacob Bernoull.
- The Law Of Large Numbers has two forms, the Weak Law Of Large Numbers , which is implemented in this course and was presented above, and the Strong Law Of Large Numbers which states :

$$\overline{X}_n \xrightarrow{\text{a.s.}} \mu \qquad \text{for} \qquad n \to \infty. \tag{27}$$

Where "a.s" states "almost sure convergence" that is,

$$\mathrm{P}\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1 \tag{28}$$

B. Andrey Markov

Andrey (Andrei) Andreyevich Markov was a Russian mathematician. He is best known for his work on theory of stochastic processes. His research later became known as Markov chains.

C. Pafnuty Chebyshev

Pafnuty Lvovich Chebyshev was a Russian mathematician. Chebyshev is known for his work in the field of probability, statistics and number theory.