Strong law of large numbers (First version)

Theorem: Let $X_1, X_2, ..$ be a sequence of independent random variables. Suppose that $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^4) \leq K$ for all i and some uniform constant $K > 0$. Then,

$$\frac{X_1 + X_2 + .. + X_n}{n} \to 0 \quad \text{almost surely.}$$

Proof: We consider the fourth moment.

$$S_n = X_1 + X_2 + .. + X_n$$

$$\mathbb{E}(S_n^4) = \mathbb{E}(X_1 + .. + X_n)^4$$

$$= \mathbb{E}(X_1^4) + .. + \mathbb{E}(X_n^4)$$

$$+ 4\left(\mathbb{E}(X_1 X_2^3) + \mathbb{E}(X_1^3 X_2) + ..\right)$$

$$+ 6\left(\mathbb{E}(X_1^2 X_2^2) + ..\right)$$

$$+ 12\left(\mathbb{E}(X_1^2 X_2 X_3) + ..\right)$$

$$+ 24\left(\mathbb{E}(X_1 X_2 X_3 X_4) + ..\right)$$

We can use independence for mixed terms such as $\mathbb{E}(X_1^2 X_2 X_3) = \mathbb{E}(X_1^2)\mathbb{E}(X_2)\mathbb{E}(X_3) = 0$.

We are left with terms of form

1) $\mathbb{E}(X_i^4)$ but they are bounded by K.

2) $\mathbb{E}(X_i^2 X_j^2) \leq \sqrt{\mathbb{E}(X_i^4)\mathbb{E}(X_j^4)} \leq K$ for all i,j.

by Cauchy-Schwarz.

So, $\mathbb{E}(S_n^4) \leq nK + 6 \cdot \binom{n}{2} K$

$= K(n + 3n(n-1)) = Kn(3n-2) \leq 3Kn^2$

So $\mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) \leq \frac{3K}{n^2}$

$\Rightarrow \mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{4}\right)^4\right) \leq 3K \sum_{n=1}^{\infty} 1/n^2 < \infty.$

So with probability 1, $\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4$ converges

and with probability 1 $S_n/n \to 0$. $\square$

Special case: If $X_1, X_2, ..$ are independent and identically distributed with $\mathbb{E}(X_i^4) < \infty$, then $Y_i = X_i - \mathbb{E}(X_i)$ has expectation 0.

$Y_1, Y_2, ..$ satisfy the conditions above and

$\frac{Y_1 + .. + Y_n}{n} \to 0$, $\frac{(X_1 - m) + .. + (X_n - m)}{n} \to 0$

and $\frac{X_1 + X_2 + .. + X_n}{n} \to m$, where $m = \mathbb{E}(X_i)$.

We can derive the distance to $\mu = \mathbb{E}(X)$ by means of Chebyshev's inequality.

$$P(|X-\mu| \geq c) \leq \frac{Var(X)}{c^2}.$$

It is a special case of Makov's inequality:

$$\mathbb{P}(|X-\mu|^2 \geq c) \leq \frac{1}{c^2} \mathbb{E}(|X-\mu|^2) = \frac{Var(X)}{c^2}.$$

Applying this to $S_n = X_1 + X_2 + .. + X_n$, where $X_1, X_2, ..$ are independent identically distributed (iid) random variables with $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = Var(X_i) < \infty$ gives

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + .. + \mathbb{E}(X_n) = n\mu,$$

$$Var(S_n) = Var(X_1) + .. + Var(X_n) = n\sigma^2 \quad \text{by independence}$$

Further $Var(S_n/n) = Var\left(\frac{S_n - n\mu}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

So $\mathbb{P}(|\frac{1}{n}S_n - \mu| \geq c) \leq \frac{\sigma^2}{n \cdot c} \rightarrow 0$ as $n \rightarrow \infty$

Since c was arbitrary we have

$$\frac{S_n}{n} \rightarrow \mu \quad \text{in probability}.$$

Conditional Expectation

Simple Example: Throw a die. All outcomes $1, 2, \ldots, 6$ are equally likely.

Write X for the r.v. that gives the outcome.

We have $\mathbb{P}(X \leq 3) = 3/6 = 1/2$.

Suppose we additionally know the outcome is even or odd.

The conditional probabilities of the event $\{X \leq 3\}$ are

$$\mathbb{P}(X \leq 3 \mid X \text{ odd}) = \frac{\mathbb{P}(\{X \leq 3\} \cap \{X \text{ odd}\})}{\mathbb{P}(\{X \text{ odd}\})} = \frac{2}{3}$$

$$\mathbb{P}(X \leq 3 \mid X \text{ even}) = \frac{\mathbb{P}(\{X \leq 3\} \cap \{X \text{ even}\})}{\mathbb{P}(\{X \text{ even}\})} = \frac{1}{3}$$

We can also obtain conditional expectations:

$$\mathbb{E}(X \mid X \text{ even}) = \frac{2+4+6}{3} = 4$$

$$\mathbb{E}(X \mid X \text{ odd}) = \frac{1+3+5}{3} = 3.$$

Generally, we define conditional expectations with respect to σ-algebras

Definition (Theorem): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Let X be an integrable random variable.
Then exists a random variable $Y = Y(\omega)$ with the following properties:

(1) Y is $\mathcal{G}$-measurable
(2) Y is integrable
(3) For all $G \in \mathcal{G}$: $\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}$

Moreover Y is almost surely unique.
For any Y, Y' satisfying (1)-(3), $\mathbb{P}(Y = Y') = 1$.
This random variable is called the conditional expectation of X w.r.t. $\mathcal{G}$ and we write

$$Y(\omega) = \mathbb{E}(X \mid \mathcal{G}) \; (= \mathbb{E}(X \mid \mathcal{G})(\omega)).$$

If $\mathcal{G}$ is generated by random variables we write

$$\mathbb{E}(X \mid Z) \text{ instead of } \mathbb{E}(X \mid \sigma(Z))$$

$$\mathbb{E}(X \mid Z_1, Z_2, \dots, Z_n) \text{ — " — } \mathbb{E}(X \mid \sigma(Z_1, \dots, Z_n)).$$

Example: In our die throwing example

$\mathcal{F} = \mathcal{P}(\{1, \dots, 6\})$, $\mathcal{G} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$

$\mathcal{G}$ measurability $\Rightarrow$ Y must be constant on $\{1,3,5\}$ and $\{2,4,6\}$, respectively.

Say $Y(\omega) = \begin{cases} a & \text{for } \omega \in \{1,3,5\} \\ b & \text{for } \omega \in \{2,4,6\} \end{cases}$.

1) $\int_{\emptyset} Y \, d\mathbb{P} = \int_{\emptyset} X \, d\mathbb{P}$ Trivial $0=0$.

2) $\int_{\{1,3,5\}} Y \, d\mathbb{P} = \int_{\{1,3,5\}} X \, d\mathbb{P}$

$\Rightarrow a \, \mathbb{P}(\{1,3,5\}) = 1 \, \mathbb{P}(\{1\}) + 3 \, \mathbb{P}(\{3\}) + 5 \, \mathbb{P}(\{5\})$

$\Rightarrow \frac{1}{2} a = \frac{1+3+5}{6} \Rightarrow a = 3$

(3) $\int_{\{2,4,6\}} Y \, d\mathbb{P} = \int_{\{2,4,6\}} X \, d\mathbb{P}$

$\Rightarrow \frac{1}{2} b = \frac{2+4+6}{6} \Rightarrow b = 4.$

(4) $\int_{\Omega} Y \, d\mathbb{P} = \int_{\Omega} X \, d\mathbb{P}$

Sum of (2) & (3) so also satisfied.

<u>Remark:</u> The "ordinary" expectation is the special case $\mathcal{G} = \{\emptyset, \Omega\}$. Then Y is constant over Ω and

$$\int_{\emptyset} X \, d\mathbb{P} = \int_{\emptyset} Y \, d\mathbb{P} \quad \text{(trivial)}$$

$$\int_{\Omega} X \, d\mathbb{P} = \int_{\Omega} Y \, d\mathbb{P} \Rightarrow Y(\omega) = \int_{\Omega} X \, d\mathbb{P} = \mathbb{E}(X).$$

We may interpret σ algebras as "knowledge" of an event and will investigate sequences $X_1, X_2, \ldots$ of integrable random variables:

Central theme:

Def$^{\underline{n}}$ Let $X_1, X_2, \ldots$ be integrable random variables. The sequence is a <u>martingale</u> if

$$\mathbb{E}(X_{n+1} \mid \sigma(X_1, X_2, \ldots, X_n)) = X_n .$$

Informally, the expectation of the $(n+1)$-th random variable conditioned on "knowing" outcomes $X_1, \ldots, X_n$ is equal to the last observed value.

Properties of conditional expectation:

Th$^{\underline{m}}$ We have

(1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$

(2) If X is $\mathcal{G}$ measurable, then $\mathbb{E}(X \mid \mathcal{G}) = X$ a.s.

(3) Linearity: $\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G})$ a.s.

(4) Positivity: If $X \geq 0$ a.s. then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$ a.s.

Proof:

(1) Since $\Omega \in \mathcal{G}$, we have

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \int_\Omega \mathbb{E}(X \mid \mathcal{G})\, d\mathbb{P} = \int_\Omega X \, d\mathbb{P} = \mathbb{E}(X)$$

(2) X satisfies all conditions in definition / theorem

(3) Note that $a\,\mathbb{E}(X \mid \mathcal{G}) + b\,\mathbb{E}(Y \mid \mathcal{G})$ is $\mathcal{G}$-measurable. The condition then follows by linearity of integration.

(4) Suppose $Y = \mathbb{E}(X \mid \mathcal{G})$ is negative with positive probability: $\mathbb{P}(Y < 0) > 0$.
Then $\exists n$ s.t. $\mathbb{P}(Y \leq -\frac{1}{n}) > 0$. This is hence a $\mathcal{G}$ measurable set and

$$\underbrace{\int_{\{Y \leq -\frac{1}{n}\}} Y \, d\mathbb{P}}_{\leq -\frac{1}{n}\mathbb{P}(\{Y \leq -\frac{1}{n}\}) < 0} = \underbrace{\int_{\{Y \leq -\frac{1}{n}\}} X \, d\mathbb{P}}_{\geq 0}$$

A contradiction. $\square$

Results on convergence carry over:

Th$^{\underline{m}}$ (1) If $X_1, X_2, \ldots$ is a sequence of non-neg. random variables such that $X_n \uparrow X$, then we also have
$\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G})$. (MCT)

(2) If $X_1, X_2, \ldots$ is a sequence of random variables s.t. $|X_n| \leq Y$ for some integrable Y, and $X_n \to X$, then also $\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G})$
(DCT)

(3) If $X_1, X_2, \ldots$ is any sequence of non-negative random variables, then
$\mathbb{E}(\liminf_{n \to \infty} X_n \mid \mathcal{G}) \leq \liminf_{n \to \infty} \mathbb{E}(X_n \mid \mathcal{G})$. (Fatou)

We also get a corresponding analogue of Jensen's inequality:

Th$^{\underline{m}}$ Let $g: I \to \mathbb{R}$ be a convex function on an interval $I \subseteq \mathbb{R}$. Assume $X: \Omega \to I$ and X and $g(X)$ are integrable. Then,

$$\mathbb{E}(g(X) \mid \mathcal{G}) \geq g(\mathbb{E}(X \mid \mathcal{G})) \quad \text{a.s.}$$

Simplification rules:

1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H})$
for sub σ-algebras $\mathcal{G}, \mathcal{H}$ with $\mathcal{H} \subseteq \mathcal{G}$.

2) $\mathbb{E}(Z \cdot X \mid \mathcal{G}) = Z \cdot \mathbb{E}(X \mid \mathcal{G})$
if Z is $\mathcal{G}$-measurable (completely determined by $\mathcal{G}$)

3) $\mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X \mid \mathcal{G})$
if $\mathcal{H}$ is independent of $X, \mathcal{G}$.
Special case: $\mathcal{G} = \{\emptyset, \Omega\}$.
If X is independent of $\mathcal{H}$, then
$\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}(X)$.

These can be proved by verifying conditions of the conditional expectation.