Overview of Probability

Prof. Vojkan Jaksic, McGill University

Notes by Dana Mendelson and Alexandre Tomberg

Summer 2010

Contents

1 Overview of Probability

In the following, we take $(\mathbb{R}, \mathcal{B})$, with $\mathcal{B}$ the Borel σ-field on $\mathbb{R}$. Let μ_n be a sequence of Borel probability measures on $\mathbb{R}$.

Definition 1.1 (Weak Convergence)**.** We say that μ_n converges weakly to a Borel probability measure μ (denoted $\mu_n \xrightarrow{w} \mu$) if for any bounded continuous $f : \mathbb{R} \to \mathbb{R}$

$$\lim_{n\to\infty} \int_{\mathbb{R}} f \, \mathrm{d}\mu_n = \int_R f \, \mathrm{d}\mu.$$

Theorem 1.2 (Lévy-Khinchine)**.** *The following are equivalent*

1. $\mu_n \xrightarrow{w} \mu$
2. *If $F_n(x)$ and $F(x)$ are distribution functions of μ_n, μ, respectively, then at any continuity point x of F, $\lim_{n\to\infty} F_n(x) = F(x)$.*
3. *For all $t \in \mathbb{R}$,*
$$\lim_{n\to\infty} \int_{\mathbb{R}} \mathrm{e}^{itx} \, \mathrm{d}\mu_n(x) = \int_{\mathbb{R}} \mathrm{e}^{itx} \, \mathrm{d}\mu.$$

Definition 1.3 (Notions of convergence)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_n\}_{n\in\mathbb{Z}}$ a sequence of random variables.

1. $X_n \to X$ P **almost surely** (a.s.) if there exists $E \in \mathcal{F}$ with $P(E^c) = 0$ such that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in E$.
2. $X_n \to X$ **in probability** if $\lim_{n\to\infty} P\{|X_n - X| \geq \epsilon\} = 0$ for all $\epsilon > 0$.
3. $X_n \to X$ **in** m^p for $p > 0$ if $\lim \mathbb{E}(|X_n - X|^p) = 0$.
4. $X_n \xrightarrow{w} X$ **in law** if $\mu_{X_n} \xrightarrow{w} \mu_X$.

By Lévy-Khinchine, we also have (4) $\iff$ $\lim_{n\to\infty} \mathbb{E}(\mathrm{e}^{itX_n}) = \mathbb{E}(\mathrm{e}^{itX})$.

There are relations between some of these notions of convergences:

1. Convergence P-a.s. $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in law.
2. Convergence in m^p $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in law.

Further, if $X_n \xrightarrow{w} c$, a constant, convergence in law $\iff$ convergence in probability.

Definition 1.4 (i.i.d. Random Variables)**.** Let $\{X_n\}_{n\in\mathbb{Z}}$ be random variables. They are called identically distributed if $\mu_{X_n} = \mu_{X_m}$ for all n, m. They are independent if for any distinct $n_1, \ldots, n_k$

$$P\left\{X_{n_1}^{-1}(B_{n_1}) \cap \ldots X_{n_k}^{-1}(B_{n_k})\right\} = P\left\{X_{n_1}^{-1}(B_{n_1})\right\} \ldots P\left\{X_{n_k}^{-1}(B_{n_k})\right\}$$
$$\iff \mu_{X_{n_1},\ldots,X_{n_k}} = \mu_{X_{n_1}} \otimes \ldots \otimes \mu_{X_{n_k}}.$$

If the X_n are i.i.d. one can forget the underlying probability space and consider a "specific model":

$$\Omega = \prod_{n\in G} \mathbb{R}, \qquad G = \mathbb{N},\ \text{or}\ \mathbb{Z}$$
$$\mathcal{F} = \bigotimes_{n\in G} \mathcal{B}, \qquad X_n(\omega) = \omega(n)$$
$$P = \bigotimes_{n\in G} \mu.$$

Any other model of i.i.d. X_n with distribution μ is "isomorphic" to this one.

1.1 Law of Large Numbers

Let $(\Omega, \mathcal{F}, P)$ be a measure space and X_n, $n \in \mathbb{N}$ i.i.d. random variables, and let

$$S_n = X_1 + \ldots + X_n.$$

Intuitively, we should have that S_n/n "$\to$" $\mathbb{E}(X_1)$, if this expectation exists.

Theorem 1.5 (Weak Law of Large Numbers v1.00)**.** *Suppose that* $\mathbb{E}(X_1^2) < \infty$*. Then* $S_n/n \to \mathbb{E}(X_1)$ *as* $n \to \infty$*.*

Proof. Consider

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{S_n}{n} - \mathbb{E}[X_1]\right)^2\right] &= \mathbb{E}\left[\left(\frac{\sum_{k=1}^n X_k - \mathbb{E}[X_k]}{n}\right)^2\right] \\
&= \frac{1}{n^2}\sum_{k,j=1}^{n} \mathbb{E}\left[(X_k - \mathbb{E}[X_k])(X_j - \mathbb{E}[X_j])\right] \\
&= \frac{1}{n^2}\sum_{k}^{n} \mathbb{E}[(X_k - \mathbb{E}[X_k])^2] \\
&= \frac{1}{n}\mathbb{E}\left[X_1 - \mathbb{E}[X_1]\right] \underset{n}{\longrightarrow} 0
\end{aligned}$$

□

Theorem 1.6 (Weak Law of Large Numbers v2.00)**.** *Suppose that* $\mathbb{E}|X_1| < \infty$*. Then* $S_n/n \to \mathbb{E}(X_1)$ *as* $n \to \infty$*.*

Proof. We proceed by considering the truncated random variables

$$X_n^c(\omega) = \begin{cases} X_n(\omega) & |X_n(\omega)| \leq c \\ 0 & \text{otherwise} \end{cases}.$$

Let $Y_n^c = X_n - X_n^c$ and note that the X_n^c are again i.i.d. random variables and

$$\mathbb{E}(X_n) = \mathbb{E}(X_n^c) + \mathbb{E}(Y_n^c) =: m_1^c + m_2^c.$$

Then

$$\begin{aligned} \mathbb{E}\left|\frac{S_n}{n} - m\right| &= \mathbb{E}\left|\frac{\sum_{k=1}^n X_k^c - m_1^c}{n} + \frac{\sum_{k=1}^n Y_k^c - m_2^c}{n}\right| \\ &\leq \left(\mathbb{E}\left|\frac{\sum_{k=1}^n X_k^c - m_1^c}{n}\right|^2\right)^{1/2} + \mathbb{E}|Y_1^c - m_2^c|, \end{aligned}$$

using Holder's inequality and the fact that the Y_k^c's are i.i.d. random variables. Now, the first term above goes to zero by Version 1 of the WLLN. Examining the second term above, we find

$$\mathbb{E}|Y_1^c - m_2^c| \leq 2\mathbb{E}|Y_1^c| = 2\int_{\{\omega:|X_1(\omega)|>c\}} |X_1(\omega)|\,\mathrm{d}P$$

hence

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{S_n}{n} - m\right| \leq 2\int_{\{\omega:|X_1(\omega)|>c\}} |X_1(\omega)|\,\mathrm{d}P \underset{c}{\longrightarrow} 0$$

□

Finally, for the ultimate version of the laws of large numbers:

Theorem 1.7 (Kolmogorov Strong Law of Large Numbers)**.** *Let* $\{X_n\}$ *be i.i.d. random variables, then* $\lim_{n\to\infty} S_n/n$ *exists (and is finite) if and only if* $\mathbb{E}(X_1)$ *exists, and in this case,* $\lim_{n\to\infty} S_n/n = \mathbb{E}(X_1)$*.*

Proof. Suppose first that $\lim_{n\to\infty} S_n/n$ exists and is finite. Then

$$\frac{X_n}{n} = \frac{X_1 + \ldots + X_n}{n} - \frac{n-1}{n}\frac{X_1 + \ldots + X_{n-1}}{n-1}$$

We know that $\lim_{n\to\infty} X_n(\omega)/n = 0$ P a.s. Let

$$E_n = \{\omega : \frac{X_n(\omega)}{n} \geq 1\}, \qquad E = \bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} E_j$$

E is the collection of all ω for which ∞ many E_n occur. Since the E_n are independent, by the Borel-Cantelli

$$P(E) = \begin{cases} 0 \\ 1 \end{cases} \qquad \text{depending whether } \sum_{n=1}^{\infty} P(E_n) < \infty \text{ or } = \infty.$$

So since $\lim_{n\to\infty} \frac{|X_n(\omega)|}{n} = 0$ P a.s., $|X_n(\omega)|/n < 1$ for sufficiently large n, $P(E) = 0 \Rightarrow \sum_{n=1}^{\infty} P(E_n) < \infty$, yielding $\sum_{n=1}^{\infty} P\{\omega : |X_1(\omega)| \geq n\} < \infty$. Since

$$\sum_{n=1}^{\infty} P\{|X_1(\omega)| \geq n\} < \infty \iff \int_{\Omega} |X_1| \, \mathrm{d}P < \infty,$$

we are done. □

The opposite direction is more delicate and we will prove it under the assumption that $\mathbb{E}(|X_1|^4) < \infty$.

Proof. Define $S_n = X_1 + \ldots + X_n$ and without loss of generality, let $\mathbb{E}(X_1) = 0$. We consider

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

Since the X_k are i.i.d., terms of the form $\mathbb{E}(X_{k_1}X_{k_2}X_{k_2}X_{k_2}) \neq 0$ if and only if $k_1 = k_2 = k_3 = k_4$ or $k_1 = k_2$ and $k_3 = k_4$ (and permutations of pairwise equality). By this observation

$$\mathbb{E}(S_n^4) = n\mathbb{E}(X_1^4) + 3n(n-1)[\mathbb{E}(X_1)]^2.$$

Thus

$$\begin{aligned} P\left\{\left|\frac{S_n}{n}\right| \geq \frac{1}{j}\right\} &= P\left\{\left|\frac{S_n}{n}\right|^4 \geq \frac{1}{j^4}\right\} \\ &= P\left\{|S_n|^4 \geq \frac{n^4}{j^4}\right\} \\ &\leq \frac{j^4}{n^4}\mathbb{E}(S_n^4) \leq \frac{Cj^4}{n^2} \end{aligned}$$

so $\sum_{n=1}^{\infty} P\left\{\left|\frac{S_n}{n}\right| \geq \frac{1}{j}\right\} < \infty$ and so by the first Borel-Cantelli Lemma,

$$\limsup_{n\to\infty} \left|\frac{S_n}{n}\right| \leq \frac{1}{j}$$

P-a.s. and since j is arbitrary, $\lim_{n\to\infty} S_n/n = 0$ P-a.s. □

1.2 Central limit theorem

Let X_n be i.i.d. random variables with mean m and variance $\mathbb{E}((X-m)^2) = \sigma^2$. A random variables X is called normal (denoted $\mathcal{N}(m, \sigma^2)$) if

$$\mu_X = \frac{1}{\sqrt{2\pi}\sigma} \, \mathrm{e}^{-(x-m)^2/\sigma^2} \, \mathrm{d}x.$$

Theorem 1.8 (Central limit theorem)**.** *Under the hypotheses above,*

$$P\left\{a \leq \frac{S_n}{\sqrt{n}} \leq b\right\} \longrightarrow \frac{1}{\sqrt{2\pi}\sigma} \, \mathrm{e}^{-(x-m)^2/\sigma^2} \, \mathrm{d}x$$

as $n \to \infty$*.*

Proof. By replacing X_n with $\frac{X_n - \mathbb{E}(X_n)}{\sigma}$ we may assume, without loss of generality that $m = 0$ and $\sigma = 1$. Let $\phi_n(t)$ be the characteristic function of $\frac{S_n}{\sqrt{n}}$, then

$$\phi_n(t) = \mathbb{E}(\mathrm{e}^{itS_n/\sqrt{n}}) = [\mathbb{E}(\mathrm{e}^{itX_1/\sqrt{n}})]^n$$

by independence. The Taylor formula then yields

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{itX_1}) &= \int_{\mathbb{R}} \mathrm{e}^{itx} \,\mathrm{d}\mu_{X_1}(x) \\ &= 1 + it \int_{\mathbb{R}} x \,\mathrm{d}\mu_{X_1} - \frac{t^2}{2} \int_{\mathbb{R}} x^2 \,\mathrm{d}\mu_{X_1} - \underbrace{t^2 \int_0^1 (1-s)[\int_0^1 x^2 (\mathrm{e}^{istx} - 1) \,\mathrm{d}\mu_{X_1}] \,\mathrm{d}x}_{r_n(t)}.\end{aligned}$$

Thus

$$\mathbb{E}(\mathrm{e}^{itX_1/\sqrt{n}}) = 1 - \frac{t^2}{n} + r_n(t)\frac{t^2}{n}.$$

So $\lim_{n\to\infty} r_n(t) = 0$ by the Dominated convergence theorem and

$$\phi_n(t) = \left(\frac{(1-t^2)}{n} + \frac{t^2}{n} r_n(t)\right)^n \longrightarrow \mathrm{e}^{-t^2/2} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx} \,\mathrm{e}^{-x^2/2} \,\mathrm{d}x$$

as $n \to \infty$. So by Levy-Khinchine, $S_n/\sqrt{n} \xrightarrow{\text{law}} \mathcal{N}(m, \sigma^2)$. □

We can think of this convergence as follows: $X_1 + \ldots + X_n = nm + \sqrt{n}\epsilon$. The next theorem provides us with a rate of convergence.

Theorem 1.9 (Berry-Esseen Theorem)**.** *Let X_n be i.i.d. random variables with mean* 0 *and variance* 1*. Suppose that $\mathbb{E}(|X_1|^{2+\alpha}) < \infty$ for $\alpha > 0$. Then*

$$\sup_{a,b} \left| P\left\{a \le \frac{S_n}{\sqrt{n}} \le b\right\} - \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/\sigma^2} \,\mathrm{d}x \right| \le \frac{C}{n^\delta}$$

for some $\delta > 0$.

Theorem 1.10 (Law of the iterated logarithm)**.** *Let X_n be i.i.d. mean* 0*, variance* 1 *random variables. Then*

$$P\left\{\omega \,\middle|\, \limsup_{n\to\infty} \frac{X_1 + \ldots + X_n}{\sqrt{n \log(\log(n))}} = \sqrt{2}\right\} = 1$$

$$P\left\{\omega \,\middle|\, \liminf_{n\to\infty} \frac{X_1 + \ldots + X_n}{\sqrt{n \log(\log(n))}} = -\sqrt{2}\right\} = 1$$

1.3 Large Deviations

Let X_n be i.i.d. random variables with mean m. We want to find the probability of a rare even. If $l > m$, what are the asymptotics of

$$P\left\{\frac{X_1 + \ldots + X_n}{n} \ge l\right\}.$$

Exercise 1. Let $X_k = \mathcal{N}(0,1)$, then $\frac{X_1+\ldots+X_n}{n} \sim \mathcal{N}(0, 1/n^2)$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log P\left\{ \frac{X_1 + \ldots + X_n}{n} \geq l \right\} = \frac{-l^2}{2}.$$

Solution. Let $Z \sim \mathcal{N}(0,1)$, we will first show the following estimate for any $x > 0$:

$$\frac{1}{\sqrt{2\pi}} \frac{1}{x + \frac{1}{x}} \mathrm{e}^{-\frac{x^2}{2}} \leq P\{Z \geq x\} \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \mathrm{e}^{-\frac{x^2}{2}} \tag{1}$$

To do so, let $\phi(t) = \frac{1}{2\pi} \mathrm{e}^{-\frac{t^2}{2}}$ be the density function of the standard normal. Then, integrating by parts, $\forall x > 0$, we get

$$\begin{aligned} P\{Z \geq x\} &= \frac{1}{\sqrt{2\pi}} \int_x^\infty \phi(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{t} (t\phi(t))\,\mathrm{d}t \\ &= -\frac{1}{t}\phi(t)\Big|_x^\infty - \int_x^\infty \frac{1}{t^2}\phi(t)\,\mathrm{d}t \leq \frac{\phi(x)}{x}. \end{aligned}$$

On the other hand,

$$P\{Z \leq x\} \geq \frac{\phi(x)}{x} - \frac{1}{x^2} \int_x^\infty \phi(t)\,\mathrm{d}t = \frac{\phi(x)}{x} - \frac{P\{Z \geq x\}}{x^2},$$

which, after a rearrangement, yields the left hand side of (**??**).

Now, since all X_i's are independent normal,

$$P\left\{ \frac{X_1 + \ldots + X_n}{n} \geq l \right\} = P\left\{ \frac{X_1 + \ldots + X_n}{\sqrt{n}} \geq l\sqrt{n} \right\} = P\left\{ X_1 \geq l\sqrt{n} \right\}.$$

Thus by using our estimate (**??**) and taking the logarithm, we have that

$$\log\left(\frac{1}{\sqrt{2\pi}} \frac{1}{l\sqrt{n} + \frac{1}{l\sqrt{n}}} \right) - \frac{l^2}{2} \leq \frac{1}{n} \log P\left\{ \frac{X_1 + \ldots + X_n}{n} \geq l \right\} \leq \log\left(\frac{1}{\sqrt{2\pi}} \frac{1}{l\sqrt{n}} \right) - \frac{l^2}{2}.$$

Finally, taking $n \to \infty$ yields the desired result. □

We define the μ-common distribution: $\psi(\theta) = \log[\mathbb{E}(\mathrm{e}^{\theta X_1})] = \log \int_{\mathbb{R}} \mathrm{e}^{\theta x}\,\mathrm{d}\mu(x)$, for $\theta \in \mathbb{R}$. We have that $\psi(\theta) \in (-\infty, \infty]$ but usually, $\psi(\theta) < \infty$, so we will take this as an assumption. Since $\psi(\theta)$ is finite, we can apply the dominated convergence theorem, yielding

$$\psi'(\theta) = \frac{\displaystyle\int_{\mathbb{R}} x\,\mathrm{e}^{\theta x}\,\mathrm{d}\mu(x)}{\displaystyle\int_{\mathbb{R}} \mathrm{e}^{\theta x}\,\mathrm{d}\mu(x)}, \qquad \psi''(\theta) = \frac{\displaystyle\int_{\mathbb{R}} x^2\,\mathrm{e}^{\theta x}\,\mathrm{d}\mu(x) - \left(\int_{\mathbb{R}} \mathrm{e}^{\theta x}\,\mathrm{d}\mu(x)\right)^2}{\left(\displaystyle\int_{\mathbb{R}} \mathrm{e}^{\theta x}\,\mathrm{d}\mu(x)\right)^2} > 0,$$

so $\psi(\theta)$ is convex except in trivial cases. We now consider the Legendre transform of $\psi(\theta)$,

$$h(\alpha) = \sup_{\theta\in\mathbb{R}} (\theta\alpha - \psi(\theta)).$$

$h(\alpha)$ is convex and positive, i.e. $h(\alpha) \in [0, \infty]$. Further, since $\psi''(\theta) > 0$, $\psi'(\theta)$ is increasing, hence $(\theta m - \psi(\theta))'$ is decreasing and $h(m) = 0$. Let now $\alpha \in [\overline{\theta}, \underline{\theta}]$, where

$$\overline{\theta} = \lim_{\theta \to -\infty} \psi'(\theta), \qquad \underline{\theta} = \lim_{\theta \to \infty} \psi'(\theta)$$

then $(\alpha\theta - \psi(\theta))' = \alpha - \psi'(\theta) = 0$ for a unique $\theta_\alpha = (\psi')^{-1}(\alpha)$. If $\alpha \notin [\overline{\theta}, \underline{\theta}]$, then $h(\alpha) = \infty$.

Theorem 1.11 (Cramer)**.** *Let X_n be i.i.d. random variables, then for any closed $F \in \mathbb{R}$,*

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} \in F\right\} \leq -\inf_{\alpha \in F} h(\alpha)$$

and for any open $O \in \mathbb{R}$,

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} \in O\right\} \geq -\inf_{\alpha \in O} h(\alpha)$$

Exercise 2. Bernoulli case, where $P\{X_n = 0\} = \frac{1}{2} = P\{X_n = 1\}$. First we have

$$\int \mathrm{e}^{\theta x}\, \mathrm{d}\mu = P\{X_n = 0\}\, \mathrm{e}^{0 \cdot \theta} + P\{X_n = 1\}\, \mathrm{e}^{1 \cdot \theta} = \frac{1}{2}(1 + \mathrm{e}^{\theta}),$$

hence, $\psi(\theta) = \log \frac{1}{2}(1 + \mathrm{e}^{\theta})$. Then

$$h(\alpha) = \sup_{\theta \in \mathbb{R}} (\theta\alpha - \log \frac{1}{2}(1 + \mathrm{e}^{\theta})).$$

We compute

$$(\theta\alpha - \log \frac{1}{2}(1 + \mathrm{e}^{\theta}))' = \alpha - \frac{\mathrm{e}^{\theta}}{1 + \mathrm{e}^{\theta}} = 0$$

which has a unique solution for $\alpha \in [0, 1]$ when $\theta = \ln(\alpha) - \ln(1 - \alpha)$, yielding

$$h(\alpha) = \ln(2) - \alpha \ln \alpha - (1 - \alpha)\ln(1 - \alpha), \qquad \alpha \in [0, 1]$$

and $h(\alpha) = \infty$ for $\alpha \notin [0, 1]$.

Let now $\Omega = \{0, 1\}^{\mathbb{N}} = \{\omega_1, \ldots, \omega_n, \ldots\}$ for $\omega_k \in [0, 1]$. Let $P = \bigotimes_{n=1}^{\infty} \nu$ where $\nu(\{1\}) = \nu(\{0\}) = \frac{1}{2}$ and define random variables on the space by setting $X_n(\omega) = \omega_n$. Then

$$\frac{S_n}{n} = \frac{\omega_1 + \ldots + \omega_n}{n}, \qquad \omega \in \Omega, \text{ “microstates”}$$

and $z \in [0, 1]$ is called a macrostate. We wish to look for a partition of phase space such that

$$\frac{\omega_1 + \ldots + \omega_n}{n} \in (z - \epsilon, z + \epsilon).$$

This partition of the space counts the multiplicity of microstates corresponding to a given macrostate. The Boltzman entropy yields that

$$\lim_{n \to \infty} \frac{1}{n} \log P\left\{\omega \,\middle|\, \frac{\omega_1 + \ldots + \omega_n}{n} \in (z - \epsilon, z + \epsilon)\right\} \simeq \ln(2) - z \ln z - (1 - z) ln(1 - z).$$

2 General setting

In the previous section, we went through three levels of theorems about sums of random variables:

1. Laws of large numbers;
2. Central limit theorems, dealing with fluctuations of the convergence above;
3. Large deviations principle (LDP) and entropy, dealing with the rate of appearance of deviations larger than those described by CLT.

All of these were stated for sums of i.i.d. random variables. Now, we will procede to states them in the general setting of dynamical systems.

2.1 Language and notation

The language used in this setting is common to probability theory, dynamical systems and classical statistical mechanics.

- $(\Omega, \mathcal{F})$ measurable space – phase space of the system.
- Observables (random variables) are measurable functions $X : \Omega \to \mathbb{R}$.
- States are associated to probability measures on $(\Omega, \mathcal{F})$. If X is an observable, $P(X) = \int_\Omega X \,\mathrm{d}P$ is the expectation of X in the state P.
- The dynamics of the system are described by a bijection $T : \Omega \to \Omega$, T, T^{-1} both measurable. Given $\omega \in \Omega$,
$$(\omega, T\omega, T^2\omega, T^3\omega, \ldots)$$
is the trajectory of the system. We also alow ourselves to go backwards in time by $T^{-n}\omega$. Thus, T induces a *flow* on observables and states:
$$\begin{aligned} X_n &= X \circ T^n \text{ is the observable } X \text{ at time } n, \text{ and} \\ P_n &= P \circ T^{-n} \text{ i.e. } P(E) = P\left(T^{-n}(E)\right) \end{aligned}$$
is the state P at time n. We have the obvious duality
$$\int_\Omega X_n \,\mathrm{d}P = \int_\Omega X \,\mathrm{d}P_n.$$

Important fact P is assumed to be a stationary state, and thus T-invariant, i.e.

$$P\left(T^{-1}(E)\right) = P(E) \qquad \forall E \in \mathcal{F}.$$

2.2 Conditional expectation

$(\Omega, \mathcal{F}, P)$; $\mathcal{G} \subset \mathcal{F}$ is a σ-subalgebra;

$$X \in L^1(\Omega, \mathcal{F}, P); \qquad \mathbb{E}\left(|X|\right) < \infty.$$

Take $E \in \mathcal{G}$ and define

$$\phi(E) = \int_E X \, \mathrm{d}P.$$

Then, ϕ is a signed measure on $(\Omega, \mathcal{G})$, $\phi << P|_{\mathcal{G}}$, so by LRN theorem, $\exists!$ $\mathcal{G}$-measurable function $\mathbb{E}\left(X|\mathcal{G}\right)$, such that

$$\int_E X \, \mathrm{d}P = \int_E \mathbb{E}\left(X|\mathcal{G}\right) \mathrm{d}P, \quad \forall E \in \mathcal{G}.$$

We call the function $\mathbb{E}\left(X|\mathcal{G}\right)$ – the conditional expectation of X with respect to the σ-subalgebra $\mathcal{G}$. $\mathbb{E}\left(X|\mathcal{G}\right)$ is uniquely specified by requiring that:

- $\mathbb{E}\left(X|\mathcal{G}\right)$ is $\mathcal{G}$-measurable,
- and $\forall E \in \mathcal{G}$,
$$\int_E X \, \mathrm{d}P = \int_E \mathbb{E}\left(X|\mathcal{G}\right) \mathrm{d}P.$$

Definition 2.1. We will write $\mathcal{F} \overset{\circ}{=} \mathcal{G}$ to mean that the σ-algebras $\mathcal{F}$ and $\mathcal{G}$ have the same sets of non zero P-measure.

Properties of Condition Expectation

1. Linear in X.
2. Consistent with the standard definition of expectation:
$$\int_\Omega \mathbb{E}\left(X|\mathcal{G}\right) \mathrm{d}P = \mathbb{E}(X).$$
3. If $\mathcal{G} \overset{\circ}{=} \{\emptyset, \Omega\}$, then $\mathbb{E}\left(X|\mathcal{G}\right) = \mathbb{E}(X)$.
4. If $\mathcal{G} \overset{\circ}{=} \{\emptyset, A, A^c, \Omega\}$ and $X = \chi_E$ for some $E \in \mathcal{F}$, then
$$\mathbb{E}\left(X|\mathcal{G}\right) = \alpha\chi_A + \beta\chi_{A^c}.$$
And $\alpha = \dfrac{P(A \cap E)}{P(A)}$, because
$$P(A \cap E) = \int_A X \, \mathrm{d}P = \int_A \mathbb{E}\left(X|\mathcal{G}\right) \mathrm{d}P = \alpha P(A).$$
Thus,
$$\mathbb{E}\left(X|\mathcal{G}\right) = \frac{P(A \cap E)}{P(A)}\chi_A + \frac{P(A^c \cap E)}{P(A^c)}\chi_{A^c}.$$
5. If $X \in L^2(\Omega, \mathcal{F}, P)$, then $\mathbb{E}\left(X|\mathcal{G}\right)$ is the orthogonal projection of X onto the closed subspace of $\mathcal{G}$-measurable functions in $L^2(\Omega, \mathcal{F}, P)$.

2.3 Birkhoff ergodic theorem

Back to $(\Omega, \mathcal{F}, P, T)$, form
$$\mathcal{A} = \{E \in \mathcal{F} \mid T(E) = E\},$$
the σ-algebra of T-invariant events.

Theorem 2.2 (Bikhoff Ergodic Theorem)**.** *If $X_k = X \circ T^k$, then*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X_k = \mathbb{E}\left(X|\mathcal{A}\right) \ P\text{-a.s.}$$

If $\mathcal{A} \overset{\circ}{=} \{\emptyset, \Omega\}$, T is called *ergodic* and
$$\mathbb{E}\left(X|\mathcal{A}\right) = \mathbb{E}(X) \ P\text{-a.s.}$$
This is a general form of a LLN.

LLN for i.i.d. X, X_n i.i.d. with common distribution μ,
$$\Omega = \prod_{n\in\mathbb{Z}} \mathbb{R}, \quad \mathcal{F} = \bigotimes_{n\in\mathbb{Z}} \mathcal{B}, \quad P = \bigotimes_{n\in\mathbb{Z}} \mu$$
$$X(\omega) = \omega_0, \quad [T(\omega)]_j = \omega_{j+1}(\text{left shift}),$$
Then
$$\text{Kolmogorov LLN} \iff T \text{ is ergodic}$$

2.4 Central limit theorem

WLOG
$$\frac{X_0 + \ldots + X_{n-1} - n\mathbb{E}(X)}{\sqrt{n}} \cong \frac{X_0 + \ldots + X_{n-1}}{\sqrt{n}},$$
and then
$$\mathbb{E}\left(\left[\frac{X_0 + \ldots + X_{n-1}}{\sqrt{n}}\right]^2\right) = \frac{1}{n}\mathbb{E}\left(\left[X_0 + \ldots + X_{n-1}\right]^2\right)$$
$$= \mathbb{E}\left(X_0^2\right) + 2\sum_{k=1}^{n-1} \mathbb{E}\left(X_0 X_k\right).$$
We have thus a necessary condition that
$$\sigma^2 = \mathbb{E}\left(X_0^2\right) + 2\sum_{k=1}^{n-1} \mathbb{E}\left(X_0 X_k\right)$$
converges absolutely.

Let $\delta > 0$ be s.t. $\mathbb{E}\left(|X|^{2+\delta}\right) < \infty$, assuming $\mathbb{E}(X) = 0$. Consider the σ-fields $\sigma(X_{-n}, X_{-n-1}, \ldots) = \mathcal{F}_{-n}$ – the minimal σ-field with respect to which all $\{X_{-k}\}_n^\infty$ are measurable. Let
$$\rho(n) = \sup\Big\{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-n}, B \in \sigma(x)\Big\}.$$

If

$$\sum_n [\rho(n)]^{\frac{\delta}{2(2+\delta)}} < \infty,$$

then

$$\sigma^2 = \mathbb{E}\left(X_0^2\right) + 2\sum_{k=1}^{n-1} \mathbb{E}\left(X_0 X_k\right)$$

converges absolutely and

$$\frac{X_0 + \ldots + X_{n-1}}{\sigma\sqrt{n}} \xrightarrow{\text{in law}} \mathcal{N}(0,1).$$

The proof of this fact uses the martingale CLT.

2.5 Large Deviations

$(\Omega, \mathcal{F}, P)$, T ergodic,

$$S_n = X_0 + \ldots + X_{n-1} \qquad \mathbb{E}(X) = 0.$$

Let us look at $P\left\{\frac{S_n}{n} \in A\right\}$. If $0 \notin A$, $P\left\{\frac{S_n}{n} \in A\right\}$ should go to 0 fast. How fast?

Assumption: $\forall \theta \in \mathbb{R}$, the limit

$$Y(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}(\mathrm{e}^{\theta S_n})$$

exists and is differentiable.

Then $Y(\theta)$ is (strictly) convex, so we can form the legendre transform

$$h(\alpha) = \sup_{\theta \in \mathbb{R}} \left(\alpha\theta - Y(\theta)\right) \quad \longleftarrow \text{ still convex!}$$

So, $h(\alpha) = 0$ if and only if $\alpha = 0$, then $h = \mathbb{E}(X)$.

Conclusion: For any closed $F \subset \mathbb{R}$, and open $O \subset \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} \in F\right\} \leq -\inf_{\alpha \in F} h(\alpha),$$
$$\limsup_{n\to\infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} \in O\right\} \geq -\inf_{\alpha \in O} h(\alpha).$$