Weak Law of Large Numbers and the Central Limit Theorem

Karthik Vaidyanathan

April 29, 2020

Introduction to Measure-Theoretic Probability

Let Ω be a set. We define a $\sigma-$*algebra* $\mathcal{F}$ with the following properties:

1. $\Omega \in \mathcal{F}$.
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
3. $A_1, A_2, ... \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

We can define a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ with the properties that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}[\cup_{i=1}^{\infty}] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ for a collection of disjoint sets $\{A_i\}_{i \in \mathbb{N}}$. We say that $\mathbb{P}$ is a *probability measure*.

So, $(\Omega, \mathcal{F}, \mathbb{P})$ is a *measure space*.

Introduction to Random Variables

A mapping $f : \Omega \to \mathbb{R}$ is *measurable* if for every open set $V \subset \mathbb{R}$, $f^{-1}(V) \in \mathcal{F}$ where $\mathcal{F}$ is the $\sigma -$ algebra of Ω. Let $\mathcal{B}(\mathbb{R})$ be the $\sigma -$ algebra constructed from open sets of $\mathbb{R}$.
We have two spaces:$(\Omega, \mathcal{F}, \mathbb{P}), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We want to create the measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

A random variable $X : \Omega \to \mathbb{R}$ is just a measurable mapping. Thus, for any $A \in \mathcal{B}(\mathbb{R})$, we can define:

$$\mu_X(A) = \mathbb{P}[X \in A] = \mathbb{P}[X^{-1}(A)].$$

Introduction to Lebesgue Integration

Our objective is to now define the quantity $\mathbb{E}[g(X)] = \int_\Omega g(X) d\mathbb{P}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be our measure space. Let μ be the Lebesgue measure on $(\Omega, \mathcal{F})$. Let f be a measurable, non-negative mapping.

For any $A \in \mathcal{F}$:

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$$

Then, $\mu(A) = \int_A d\mu = \int_\Omega \chi_A d\mu$.

Let $s(x) = \sum_{i=1}^n \alpha_i \chi_{A_i}(x)$ where $A_i = \{x : s(x) = \alpha_i\}$. So, s is a *simple function*.
By linearity, $\int_\Omega s d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$.

Introduction to Lebesgue Integration (Ctd.)

Note that by definition, f is non-negative. Thus, there exists a sequence of increasing simple functions $0 \leq s_i \leq s_{i+1}$ that increases to f. Then, $\int_\Omega f d\mu = sup_{0 \leq s \leq f} \int_\Omega s d\mu$.
For a random variable X and a Borel-measurable mapping g, we have that $\int g(X) d\mu$ is defined.

As a consequence of the *Radon-Nikodym Theorem*, we have that the quantity $\frac{d\mu_X}{d\lambda} = f_X$ is defined (where μ_X is the probability measure of X and . This is the *probability density function*.

Therefore, $\mathbb{E}[g(X)] = \int_\Omega g(X) d\mathbb{P} = \int_\Omega g(X) \frac{d\mu_X}{d\lambda} d\lambda$.
We can evaluate the last integral as a Riemann integral.

Statements of the WLLN and the CLT

Let μ represent a probability measure.

WLLN (Weak Law of Large Numbers): Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables with the distribution μ and the characteristic function $\phi = \phi_\mu$ such that $\phi'(0)$ exists. Then, $c = -i\phi'(0) \in \mathbb{R}$ and $\frac{1}{n}\sum_{k=1}^{n} X_k \to c$ in probability.

CLT (Central Limit Theorem): Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables with $0 < Var[X_1] = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] < \infty$. Then, $\frac{\sum_{k=1}^{n}(X_k - \mathbb{E}[X_1])}{\sqrt{n\sigma^2}} \xrightarrow{\mathcal{D}} \chi$, with $\chi \sim N(0,1)$, $\sigma^2 = Var[X_1]$.

But what do these theorems say?

Independence and Sequences of Random Variables

Let $\{X_n\}$ be a sequence of random variables (let X be a random variable). Each r.v. X_n ($n \in \mathbb{N}$) has distribution (probability measure) μ_{X_n}.
We say that random variables X, Y are *independent* if

$$\mathbb{P}[\{X \in A\} \cap \{Y \in B\}] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B] \text{ for all } A, B \in \mathcal{B}(\mathbb{R}).$$

Hence, the term "iid" for a sequence of random variables indicates that the random variables are independent and have common distribution μ_X.

Convergence of Random Variables

Let $\{X_n\}$ and X be as before. Let $\{\mu_n\}, \mu$ be probability measures (former is a sequence). We have several different modes of convergence (let $\epsilon > 0$):

1. Almost Surely (a.s.):
   $X_n \to X \text{ a.s.} \iff \mathbb{P}[\{\omega : X_n(\omega) \not\to X(\omega)\}] = 0.$
2. In Probability:
   $X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{P}[\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}] \to 0$ as $n \to \infty$.
3. Weak: $\mu_n \xrightarrow{w} \mu \iff \int f d\mu_n \to \int f d\mu$ for $f \in C_b(S)$.
4. In Distribution: $X_n \xrightarrow{\mathcal{D}} X \iff \mu_{X_n} \xrightarrow{w} \mu_X$.

Introduction to Characteristic Functions

Given a probability measure μ (on $\mathcal{B}(\mathbb{R})$), let $\varphi_\mu : \mathbb{R} \to \mathbb{C}$ be given as $\varphi_\mu(t) = \int e^{itx}\mu(dx)$. This is the characteristic function of the probability measure μ.

Given a random variable X, the characteristic function is $\varphi_X(t) = \mathbb{E}[e^{itX}]$. Note that for independent variables X, Y, we have $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. Also, given $\alpha \in \mathbb{R}, \varphi_{\alpha X}(t) = \varphi_X(\alpha t)$.

Note that given probability measures μ_1, μ_2 $\varphi_{\mu_1} = \varphi_{\mu_2} \Rightarrow \mu_1 = \mu_2$. This implies that we can characterize random variables with their characteristic functions.

Proving the WLLN

Below is a sketch of the WLLN proof:

1. Show that $\phi'(0) \in \mathbb{C}$ (which implies that $c = -i\phi'(0) \in \mathbb{R}$).
2. Let $S_n = \sum_{k=1}^{n} X_k$. To show that $\frac{1}{n} S_n \xrightarrow{\mathbb{P}} c$, it suffices to show that $\frac{1}{n} S_n \xrightarrow{\mathcal{D}} c$. With characteristic functions, it further suffices to show that $\varphi_{\frac{1}{n} S_n} \to e^{itc} = e^{t\varphi'(0)}$ $\forall t \in \mathbb{R}$.
3. Since $\{X_n\}$ are iid, $\varphi_{\frac{1}{n} S_n}(t) = (\varphi(\frac{t}{n}))^n = (1 + \frac{z_n}{n})^n$ with $z_n = n(\phi(\frac{t}{n}) - 1)$.
4. Show that $z_n \to t\varphi'(0)$ and hence $(1 + \frac{z_n}{n})^n \to e^{t\varphi'(0)}$.

Proving the CLT

Below is a sketch of the CLT proof:

1. Consider the sequence $\{(X_n - \mu)/\sqrt{\sigma^2}\}$ with $\mu = 1,\ \sigma^2 = 1$.
2. Let $S_n = \sum_{k=1}^{n} X_k$ so since the X_n are iid, it follows that $\varphi_{\frac{1}{\sqrt{n}}S_n}(t) = (\varphi(\frac{t}{\sqrt{n}}))^n$. Then it suffices to show that $(\varphi(\frac{t}{\sqrt{n}}))^n \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}t^2}$ $(t \in \mathbb{R})$.
3. Show that $|(\varphi(\frac{t}{\sqrt{n}}))^n - (1 - \frac{t^2}{2n})^n| \leq t^2 r(t/\sqrt{n})$ where $r(t) = \mathbb{E}[X^2 min(t|X|, 1)]$ for $n \geq \frac{2}{t^2}$ $(t \geq 0)$.
4. Show that $\lim_{n\to\infty} r(t/\sqrt{n}) = 0$. This proves the theorem.

Appendix: Radon-Nikodym Theorem

When discussing Lebesgue integration, the function $\frac{d\mu_X}{d\lambda} = f_X$ was mentioned. We briefly discuss the Radon-Nikodym theorem as a way to justify the existence of this function.

Let μ and ν be measures defined on a measure space $(S, \mathcal{S})$. We say that μ is $\sigma-$finite if $\exists$ a collection of disjoint sets $\{X_n\} \subseteq \mathcal{S}$ with $\cup_{n \in \mathbb{N}} X_n = S$ such that $\mu(X_n) < \infty$ for every n.

We say that ν is absolutely continuous with respect to μ (denoted as $\nu \ll \mu$) if for any $A \in \mathcal{S}$, $\mu(A) = 0 \Rightarrow \nu(A) = 0$.

Appendix: Radon-Nikodym Theorem (Ctd.)

The Radon-Nikodym Theorem is as follows: let μ and ν be $\sigma-$finite measures on $(S, \mathcal{S})$ with $\nu \ll \mu$. Then, $\exists f \in \mathcal{L}^0_+$ such that for any $A \in \mathcal{S}$, $\nu(A) = \int_A f d\mu$ (where $\mathcal{L}^0_+$ is the set of continuous, non-negative functions). If $\exists g \in \mathcal{L}^0_+$, then $f = g$ a.e.

For such a function f, we can write $f = \frac{d\nu}{d\mu}$. We call this quantity the $Radon - Nikodym\ derivative$.

Let μ_X and λ be defined as before. Then, both measures are $\sigma-$finite. In addition, $\mu_X \ll \lambda$. Then, the Radon-Nikodym theorem applies and the quantity $f_X = \frac{d\mu_X}{d\lambda}$ is defined.