Basic measure theory definitions

Def. Given a set Ω, a σ-algebra (or σ-field) $\mathcal{F}$ on Ω is a collection of subsets of Ω such that that contains Ω, is closed under complementation and is closed under countable unions (that is, $A^c \in \mathcal{F}$ for any $A \subseteq \mathcal{F}$ and $\bigcup_n A_n \in \mathcal{F}$ for any set $\{A_n\}_{n\in\mathbb{N}}$ of elements in $\mathbb{F}$). It follows that $\mathcal{F}$ is closed under countable intersections, and contains the empty set. Elements of a $\mathcal{F}$ are called measurable sets.

Def. A measurable space (or Borel space) is a tuple $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F}$ a σ-algebra on Ω.

Def. A measure on a measurable space $(\Omega, \mathcal{F})$ is a map $\mu : \mathcal{F} \to [0, \infty]$, where $\mu(\emptyset) = 0$, and $\mu(\bigsqcup_n A_n) = \sum_n \mu(A_n)$ for $\{A_n\}_{n\in\mathbb{N}}$ a countable collection of pairwise disjoint sets in $\mathcal{F}$. This last property is called countable additivity (or σ-additivity). A probability measure is a measure with total measure $\mu(\Omega) = 1$.

Def. A measure space is a triple $(\Omega, \mathcal{F}, \mu)$, where μ is a measure on measurable space $(\Omega, \mathcal{F})$. A probability space is measure space $(\Omega, \mathcal{F}, p)$, where p is a probability measure, in which case Ω is called the sample space, and its elements are called outcomes, and $\mathcal{F}$ is called the event space, and its elements, the measurable sets, are called events.

Def. An $(\mathcal{F}, \mathcal{E})$-measurable function is a map $f : \Omega \to E$, where $(\Omega, \mathcal{F}), (E, \mathcal{E})$ are two measurable spaces, such that for every $B \in \mathcal{E}$, $f^{-1}(B) \coloneqq \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F}$.

Def. An E-valued random variable is an $(\mathcal{F}, \mathcal{E})$-measurable function $X : \Omega \to E$, where $(\Omega, \mathcal{F}, p)$ is a probability space (the "underlying" space), and $(E, \mathcal{E})$ is a measurable space (sometimes called the "sample space" ambiguously with Ω). Often, $(E, \mathcal{E})$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Def. The distribution of (or law of, or measure induced by) random variable X is the push-forward of the probability measure p on $(\Omega, \mathcal{F})$ to probability measure $p_X \coloneqq p \circ X^{-1}$ on $(E, \mathcal{E})$. That is, for all $B \in \mathcal{E}$, the "probability of X taking on a value in B" is

$$\text{“}p(X \in B)\text{”} \coloneqq p_X(B) = p(X^{-1}(B)) = p(\{\omega : X(\omega) \in B\}) = \int_\Omega \mathbf{1}_{X^{-1}(B)}\, \mathrm{d}p = \int_E \mathbf{1}_B\, \mathrm{d}p \circ X^{-1} = \int_B \mathrm{d}p_X$$

The distribution exists, since X is a measurable function, so $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{E}$. Checking that $p_X(E) = 1$, and is countably additive, we see that the distribution is a probability measure on $(E, \mathcal{E})$.

In practice, the probability space induced by the random variable, $(E, \mathcal{E}, p_X)$, is what is used when working with random variable, and the underlying probability space $(\Omega, \mathcal{F}, p)$ is often not even mentioned.

Def. Given measures μ, ν on $(\Omega, \mathcal{F})$ say ν is absolutely continuous with respect to μ (written $\nu \ll \mu$) iff $\mu(A) = 0 \implies \nu(A) = 0, \forall A \in \mathcal{F}$.

Def. If $p_X(C) = 1$ for some countable set in $C \subseteq E$, then X is called discrete.

For real-valued random variable X, if $p_X \ll \lambda$ (that is, $p_X(X \in B) = 0$ for all $B \in \mathcal{E}$ with Lebesgue measure 0) then X is called continuous.

Theorem (Radon-Nikodým). *For any measures μ, ν on $(\Omega, \mathcal{F})$ if $\nu \ll \mu$, there exists a $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable nonnegative function $\rho : \Omega \to \mathbb{R}^{\geq 0}$, such that $\nu(A) = \int_A \rho\, \mathrm{d}\mu$, $\forall A \in \mathcal{F}$, which is uniquely defined μ-almost everywhere. This function is called the Radon-Nikodým derivative of ν with respect to μ, written $\rho = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.*

Def. Given random variable X with values in $(E, \mathcal{E})$, the density of X with respect to a reference measure μ on $(E, \mathcal{E})$ is the Radon-Nikodým derivative $\frac{\mathrm{d}p_X}{\mathrm{d}\mu}$, if such exists. Very often, the following (abuse of ?) notation is used: the symbol used for the probability distribution of a random variable is also overloaded to also mean the density with respect to a dominating measure (usually the Lebesgue measure λ when E is the real numbers). Also, this dominating measure is simply denoted $\mathrm{d}x$, rather than e.g. $\lambda(\mathrm{d}x)$. So what might be written more formally as $\int_S \frac{\mathrm{d}p_X}{\mathrm{d}\lambda}\, \mathrm{d}\lambda$ or using a different symbol ρ_X for the density, $\int_S \rho(x)\lambda(\mathrm{d}x)$ is written instead as instead as something like $\int_S p_X(x)\, \mathrm{d}x$.

Def. For a real-valued random variable X (where the observation space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, or a subset) we can also define the cumulative distribution function $F_X(x) \coloneqq p(X \leq x)$, $\forall x \in \mathbb{R}$.

Any real-valued random variable can be written as a sum of a discrete and a continuous random variable. The density is the *convolution* of the densities of the two random variables.

Def (Lebesgue/abstract integral). Given measure space $(\Omega, \mathcal{F}, \mu)$ and function X, write the integral of a function X with respect to μ as $\int X\, \mathrm{d}\mu$ or $\int X(\omega)\mu(\mathrm{d}\omega)$.

Assuming $X : \Omega \to \mathbb{R}$, this integral is defined as (Lanchier, 2017, section 1.2)

- For a simple function (one with finite range): $\int \sum a_i \mathbf{1}_{A_i}\, \mathrm{d}\mu \coloneqq \sum a_i \mu(A_i)$.
- For a positive function $X : \Omega \to \mathbb{R}^+$: since any positive measurable function can be written as the pointwise limit of a nondecreasing sequence of functions in $\mathcal{S}(\Omega, \mathcal{F}) \coloneqq$ the set of simple measurable functions, $\int X\, \mathrm{d}\mu \coloneqq \sup_{s\in\mathcal{S}(\Omega,\mathcal{F})} \int s\, \mathrm{d}\mu$.
- Let $\int X\, \mathrm{d}\mu \coloneqq \int_{X>0} X\, \mathrm{d}\mu - \int_{X<0} -X\, \mathrm{d}\mu$ for any integrable function (defined as a function where both these terms are finite).

Def. The expectation (or expected value) of a real-valued random variable $X : \Omega \to \mathbb{R}$ (a measurable function from probability space $(\Omega, \mathcal{F}, p)$ to measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined as $\mathbb{E}[X] = \mathbb{E}_p[X] \coloneqq \int_\Omega X\, \mathrm{d}p$.

Given measurable function $\phi : \mathbb{R} \to \mathbb{R}$, then if ϕ is positive or integrable, $\mathbb{E}[\phi(X)] \coloneqq \int_\mathbb{R} \phi\, \mathrm{d}p_X = \int_\Omega \phi \circ X\, \mathrm{d}p$.

By definition, if X is a continuous random variable, $p_X \ll \lambda$, so by RN, it has a density $\rho_X = \frac{\mathrm{d}p_x}{\mathrm{d}\lambda}$, where for all $B \in \mathcal{B}(\mathbb{R})$, $\int_B \mathrm{d}p_X = \int_B \rho_X\, \mathrm{d}\lambda$. Call this the pdf of X.

Kolmogorov Axioms

K1 The probability of event is a nonnegative real: $p(A) \in \mathbb{R}^{\geq 0}$, $\forall A \in \mathcal{F}$

K2 The probability that at least one elementary event in the sample space occurs is 1: $p(\Omega) = 1$

K3 p is σ-additive $\mu(\bigsqcup_n A_n) = \sum_n \mu(A_n)$ for $\{A_n\}_{n\in\mathbb{N}}$ a countable collection of pairwise disjoint sets in $\mathcal{F}$

Laws of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables with identical finite expected value $\mathbb{E}[X_i] = m < \infty$, for all i. The laws of large numbers say that the sample average $\bar{X}_n \coloneqq \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to m.

The Weak LLN (aka Khinchin's law) states that $\bar{X}_n \xrightarrow{\text{Pr}} m$. That is, it converges *in probability*:

$$\forall\,\epsilon > 0,\ \lim_{n\to\infty} p_{\bar{X}_n}\left(\left|\bar{X}_n - m\right| < \epsilon\right) = 1, \quad \text{that is,}\ \lim_{n\to\infty} p\left(\{\omega \in \Omega : \left|\bar{X}_n(\omega) - m\right| < \epsilon\}\right) = 1,$$

The weak law says that for large n, there is high probability that the sample average is arbitrarily close to m. It is possible however that $|\bar{X}_n - m| \geq \epsilon$ for an infinitely often. The strong law says that this almost surely will not occur.

The Strong LLN (aka Kolmogorov's law) states that $\bar{X}_n \xrightarrow{\text{a.s.}} m$. That is, it converges *almost surely*:

$$p_{\bar{X}_n}\left(\lim_{n\to\infty} \bar{X}_n = m\right) = 1 \qquad \text{that is, } p\left(\{\omega \in \Omega : \lim_{n\to\infty} \bar{X}_n(\omega) = m\}\right) = 1$$

Informally, the two laws can be summarized each as a statement about a sequence of estimates $\{Y_n\}_{n\in\mathbb{Z}}$, and some fixed value m:

- **Weak**: In the limit of $n \to \infty$, [the probability that Y_n is close to m] is one.
- **Strong**: [The proposition that in the limit of $n \to \infty$, Y_n equals m], has probability one.

Both the strong and weak LLN hold in the case where the random variables are iid samples from the same underlying distribution, but there are other cases where the weak holds but the strong does not (see Billingsley, 1995, Example 5.4, p.71), or the example below.

Example (weak LLN holds but strong does not). Let $(\Omega, \mathcal{B}, p)$ be the probability space defined on the unit interval $\Omega = [0, 1]$ with Borel sets $\mathcal{B}$ and Lebesgue measure p. Let $\{Y_n\}_{n\in\mathbb{N}}$ be a set of real-valued (in fact, binary) random variables where $Y_n : \Omega \to \{1, 0\}$ is defined as

$$Y_n(\omega) = \begin{cases} 1 & \text{if } \omega \in A_n \\ 0 & \text{else} \end{cases}, \ \text{for } \omega \in \Omega$$

where $A_1 = [0, \frac{1}{2}]$, $A_2 = [\frac{1}{2}, 1]$, $A_3 = [0, \frac{1}{4}]$, $A_4 = [\frac{1}{4}, \frac{1}{2}]$, $A_5 = [\frac{1}{2}, \frac{3}{4}]$, $A_6 = [\frac{3}{4}, 1]$, $A_7 = [0, \frac{1}{8}], \ldots$.

So, $p_{Y_n}(1) = p(\{\omega : Y_n(\omega) = 1\}) = p(A_n)$ (the probability of getting 1 at step n is the length of the interval A_n).

For the sequence Y_n:

- **the weak LLN holds**. That is, $Y_n \xrightarrow{\text{Pr}} 0$.

  As $n \to \infty$, the intervals become ever smaller, and $p_{Y_n}(1) \to 0$ so $p_{Y_n}(0) \to 1$. So, $p(\{\omega : |Y_n(\omega) - 0| < \epsilon\}) \to 1$ and we have that the sequence $Y_n \xrightarrow{\text{Pr}} 0$. [Explicitly: we need to say this converges for all $\epsilon > 0$. Note first that for $\epsilon > 1$ it is trivial: the set $\{\omega : |Y_n(\omega) - 0| < \epsilon\} = \Omega$, which has measure 1 for all n. Then note that for any $0 < \epsilon < 1$ the set $\{\omega : |Y_n(\omega) - 0| < \epsilon\} = \{\omega : Y_n(\omega) = 0\}$, which has measure $1 - p(A_n)$, which approaches 1 as $n \to \infty$.]

- **the strong LLN does not hold**. That is, it is *not* true that $Y_n \xrightarrow{a.s.} 0$

  In fact, for *any* $\omega \in \Omega$, the sequence of A_n's which contain ω continue to occur infinitely often as n grows. So, there is no set $W \subset \Omega$ such that $\lim Y_n(\omega)$ converges at all. So, $p(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\}) = 0 \neq 1$; the strong law fails catastrophically.

Example. Some more similar examples:

- For an example where the strong still law does not hold, but where $\exists\omega \in \Omega$ such that $Y_n(\omega)$ converges, intersect each A_n with the interval $[\frac{1}{3}, \frac{2}{3}]$. Then the $p(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\}) = \frac{1}{3} \neq 1$.
- For an example where the strong law does hold, set $A_n = [0, \frac{1}{n}]$ for all n. Then $p(\{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\}) = p([0, 1] \setminus \{0\}) = 1$.
- For another example where the strong law does hold, but where the set of problematic points $\{\omega : \lim_{n\to\infty} Y_n(\omega) \neq 0\}$ is infinite (even, uncountable) take $A_n = C_n$, the nth set in the recursive definition of the Cantor ternary set. Since the Cantor set has measure zero, the strong law still holds.

References

Billingsley, Patrick (1995). *Probability and Measure*. Third edition. Wiley.

Lanchier, Nicolas (2017). "Basics of Measure and Probability Theory". In: *Stochastic Modeling*. Universitext. Springer International, pp. 3–24.