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 that contains Ω, is closed under complementation, and is closed under countable unions. [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 $\mu : \mathcal{F} \to \mathbb{R} \cup \{\infty\}$, where the following hold:

- nonnegativity: $\mu(A) \in [0, \infty]$ for all $A \in \mathcal{F}$,
- null empty set: $\mu(\emptyset) = 0$,
- countable additivity (or σ-additivity): $\mu(\bigsqcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$, for $\{A_k\}_{k=1}^{\infty}$ a countable collection of pairwise disjoint sets in $\mathcal{F}$.

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.

Def. A 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 a measurable function $X : \Omega \to E$, where $(\Omega, \mathcal{F}, p)$ is a probability space (called sample space), and $(E, \mathcal{E})$ is a measurable space (called observation space).

Def. The law (or the distribution) of 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

$$p_X(B) = p(X^{-1}(B)) = p(\{\omega : X(\omega) \in B\}).$$

[Note: The law 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 law is a probability measure.]

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 $\bar{X}_n \xrightarrow{P} m$. That is, it converges *in probability*:

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

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 $\bar{X}_n \xrightarrow{a.s.} m$. That is, it converges *almost surely*:

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

Both the strong and weak LLN hold in the iid case, but there are other cases where the weak holds but the strong does not (see Billingsley, 1995, Example 5.4, p.71).

References

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