Probability Refresher

Keith A. Lewis

Abstract

This note collects salient facts about Probability Theory.

Probability Theory

In order to understand statistics one must first understand *probability theory*.

A *sample space* is a set of *outcomes*. Subsets of a sample space are *events*. A *probability measure* assigns a number between 0 and 1 to events that represents a *degree of belief* that an outcome of a random sample will belong to the event. *Partial information* is modeled by a *partition* of the sample space.

Sample Space

A *sample space* is a set of what can happen in a probability model. An *outcome* is an element of a sample space. An *event* is a subset of a sample space.

A sample space for flipping a coin can be modeled by the set $\{H,T\}$ where the outcome H indicates heads and T indicates tails. Of course any two element set could be used for this.

A sample space for flipping a coin twice can be modeled by the set $\{HH,HT,TH,TT\}$ where each outcome specifies the individual outcomes of the first and second flip. The event 'the first flip was heads' is the subset $\{HH,HT\}$. The partition $\{\{HH,HT\},\{TH,TT\}\}$ represents the partial information of knowing the outcome of the first coin flip. The first event in the partition indicates the first flip was heads. The second event in the partition indicates the first flip was tails.

The first step in any probability model is to specify the possible outcomes. The second step is to assign probabilities to the outcomes.

Measure

A *measure* μ on a set S assigns numbers to subsets of S and satisfies

$$\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F)$$

for any subsets $E, F \subseteq S$ and $\mu(\emptyset) = 0$. Measures do not count twice.

Exercise. *Show if $\nu(E \cup F) = \nu(E) + \nu(F) - \nu(E \cap F)$ for $E, F \subseteq S$ then $\mu = \nu - \nu(\emptyset)$ is measure.*

Solution

By $\mu = \nu - \nu(\emptyset)$ we mean $\mu(E) = \nu(E) - \nu(\emptyset)$ for any subset $E \subseteq S$. Clearly $\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F)$ for any $E, F \subseteq S$. Since $\mu(\emptyset) = \nu(\emptyset) - \nu(\emptyset) = 0$, μ is a measure.

Exercise. *Show if μ is a measure then $\mu(E \cup F) = \mu(E) + \mu(F)$ for any subsets E and F with empty intersection $E \cap F = \emptyset$.*

Solution

Since $\mu(\emptyset) = 0$, $\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F) = \mu(E) + \mu(F) - \mu(\emptyset) = \mu(E) + \mu(F)$.

Exercise. *Show if μ is a measure then $\mu(E) = \mu(E \cap F) + \mu(E \cap F')$ for any subsets E and F where $F' = S \setminus F = \{x \in S : x \notin F\}$ is the* complement *of F in S.*

Solution

Note $(E \cap F) \cup (E \cap F') = E \cap (F \cup F') = E \cap S = E$ and $(E \cap F) \cap (E \cap F') = E \cap (F \cap F') = E \cap \emptyset = \emptyset$ so $\mu(E \cap F) + \mu(E \cap F') = \mu((E \cap F) \cup (E \cap F') = \mu(E)$.

Probability Measure

A *probability measure* P on the sample space Ω is a measure taking values in the interval $[0, 1]$ with $P(\Omega) = 1$. The *probability* $P(E)$ for $E \subseteq \Omega$ represents a *degree of belief* that a random outcome will belong to the event E. This is a somewhat nebulous and controversial notion. How do "random outcomes" occur?

Probability theory originated with games of chance. One way to interpret this is "How much money would you wager on an outcome involving rolling dice or selecting cards from a deck?" Probability theory can also be used to figure out if the dice are weighted or the cards are marked.

Exercise. *Show $P(E \cup F) \leq P(E) + P(F)$ for any events E and F when P is a probability measure.*

Exercise. *Show $P(\cup_i E_i) \leq \sum_i P(E_i)$ for any events (E_i) when P is a probability measure.*

If Ω consists of a finite number of elements $\{\omega_1, \ldots, \omega_n\}$ we can define a probability measure by $P(\{\omega_i\}) = p_i$ where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. Every subset of Ω corresponds to a subset $I \subseteq \{1, \ldots, n\}$. The probability of the event $E = \{\omega_i : i \in I\}$ is $P(E) = \sum_{i \in I} p_i$.

Exercise. *Show this defines a probability measure.*

For the two coin flip model (assuming the coin is fair) we assign probability of 1/4 to each outcome. The probability of the first flip being heads is $P(\{HH, HT\}) = P(\{HH\} \cup \{HT\}) = P(\{HH\} + P(\{HT\}) = 1/4 + 1/4 = 1/2$.

Random Variable

A *random variable* is a symbol that can be used in place of a number when manipulating equations and inequalities with with additional information about the probability of the values it can take on.

Discrete Random Variable

A *discretely distributed random variable* is defined by the values it can take (x_i) together with the probabilities (p_i) that it takes those values $P(X = x_i) = p_i$. The probabilities must satisfy $p_i \geq 0$ and $\sum_i p_i = 1$.

Continuous Random Variable

A *continuously distributed random variable* is defined by a *density function* f where $P(a < X \leq b) = \int_a^b f(x)\,dx$. The density must satisfy $f \geq 0$ and $\int_{\mathbf{R}} f(x)\,dx = 1$.

Cumulative Distribution Function

There are random variables that are neither discrete nor continuous, however all random variables can be defined using a *cumulative distribution function*. The cdf of the random variable X is $F_X(x) = F(x) = P(X \leq x)$. It tells you everything there is to know about X. For example, $P(a < X \leq b) = F(b) - F(a)$.

Exercise. *Show* $P(a \leq X \leq b) = \lim_{x \uparrow a} F(b) - F(x)$.

Hint: $[a, b] = \cap_n (a - 1/n, b]$.

In general $P(X \in A) = \int_A dF(x)$ for sufficiently nice subsets $A \subset \mathbf{R}$ where we are using Riemann–Stieltjes integration.

Exercise: *Show for any cumulative distribution function F that $F(x) \leq F(y)$ if $x < y$, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$, and F is right continuous with left limits.*

Hint: For right continuity use $(-\infty, x] = \cap_n (-\infty, x + 1/n]$.

The cdf $F(x) = \max\{0, \min\{1, x\}\}$ defines the uniformly distributed random variable, U, on the interval $[0, 1]$. For $0 \le a < b \le 1$, $P(a < U \le b) = P(U \in (a, b]) = b - a$ and $P(U < 0) = 0 = P(U > 1)$.

Two random variables, X and Y, have the same *law* if they have the same cdf.

Exercise. *If X has cdf F, then X and $F^{-1}(U)$ have the same law.*

Exercise. *If X has cdf F, then $F(X)$ and U have the same law.*

This shows a uniformly distributed random variable has sufficient randomness to generate any random variable. There are no random, random variables.

The mathematical definition of a random variable is that it is a function $X\colon \Omega \to \mathbf{R}$. Its cumulative distribution function is $F(x) = P(X \le x) = P(\{\omega \in \Omega \mid X(\omega) \le x\})$.

Given a cdf F we can define a random variable having that distribution by $X\colon \mathbf{R} \to \mathbf{R}$ to be the identity function, $X(x) = x$ and let P be the probability measure on $\mathbf{R}$ defined by $P(A) = \int_A dF(x)$.

The mathematical definition is more flexible than defining a random variable by its cumulative distribution function.