Jingyi Zhu

550.420 Introduction to Probability

Spring 2017

1 Week 1 (Discussion Sections on 02/2)

1.1 [She02, Sect. 2.2] Sample Space and Event

Definition 1 (Sample Space and Event)**.** Sample space Ω is the set containing *all possible* sample points *of interest*. Event $E \subseteq \Omega$ is a set containing *some* sample points. In particular, $\Omega, \varnothing$ are events, and $\Omega \cap \varnothing = \varnothing$, $\Omega \cup \varnothing = \Omega$.

Example 1 (Choosing Marbles)**.** The experimenter gets to pick two balls out of a box that contains one blue ball and one green ball. The sampling procedure is with replacement.

- If the experimenter is interested in the event $E_1 = \{\text{exactly one blue marble is drawn}\}$, then it'd be more convenient to construct the sample space to be $\Omega = \{$ BG , GG, BB $\}$. Whether the blue marble is drawn in the first try or not does NOT matter for E_1.
- On the contrary, if $E_2 = \{\text{a blue marble drawn in the first try}\}$ is of the experimenter's interest, then it'd be wiser to construct the sample space to be $\Omega = \{$ GB, BG , GG, BB $\}$, as the order does matter now.

Remark 1. The construction of the sample space is always of the experimenter's disposal. Build Ω carefully, and use the *combination* and *permutation* in [She02, Chap. 1] cautiously.

Example 2 (Throwing a Dart)**.** $\Omega = \big\{(x, y) : x^2 + y^2 \leq 1\big\}$ is established on the basis of ignoring "missed" darts. Those "missed" darts does not shed light on the accuracy estimation of our interest. That being said, Ω can be built upon the events of our interests ONLY.

Definition 2 (Experiment)**.** Intuitively, an experiment is a testing/observation procedure that can be infinitely repeated and has a well-defined set of possible outcomes (sample space). For simplicity, we assume that the experimenter can have a vision of *all possible* outcomes. That being said, the sample space is known, while the possibility/likelihood of each event still remains unknown. So the observation process becomes necessary, given that *certain* (other than Ω and $\varnothing$) individual outcomes (called *sample point*s) cannot be predicted with certainty.

In frequentist's point of view, the probability (to appear soon) of an event is the *long-run relative frequency* of the occurance of that event in a large number of experiments. The ultimate goal of experiment is to *assess the likelihood of events*.

Proposition 1 (Laws of Set Operations)**.**

Relating the laws of set union and set intersection to the laws of scalar addition and scalar multiplication makes it convenient for memorizing's sake. For scalar operations, we have (1) commutativity: $a_1 + a_2 = a_2 + a_1$, $a_1a_2 = a_2a_1$; (2) associativity: $(a_1 + a_2) + a_3 = a_1 + (a_2 + a_3)$, $(a_1a_2)a_3 = a_1(a_2a_3)$; (3) distributive laws: $(a_1 + a_2)a_3 = a_1a_3 + a_2a_3$ for any $a_1, a_2, a_3 \in \mathbb{R}$.

(i) Commutative Laws. $E_1 \cup E_2 = E_2 \cup E_1$, and $E_1 \cap E_2 = E_2 \cap E_1$.

(ii) Associative Laws. $(E_1 \cup E_2) \cup E_3 = E_1 \cup (E_2 \cup E_3)$, and $(E_1 \cap E_2) \cap E_3 = E_1 \cap (E_2 \cap E_3)$.

(iii) Distributive Laws. $(E_1 \cup E_2) \cap E_3 = (E_1 \cap E_3) \cup (E_2 \cap E_3)$, $(E_1 \cap E_2) \cup E_3 = (E_1 \cup E_3) \cap (E_2 \cup E_3)$.

(iv) DeMorgan's Law. $(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}$ and $(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$

Relatable 1. $(E_1 \cup E_2 \cup E_3)^{\mathsf{c}} = (E_1 \cup E_2)^{\mathsf{c}} \cap E_3^{\mathsf{c}} = E_1^{\mathsf{c}} \cap E_2^{\mathsf{c}} \cap E_3^{\mathsf{c}}$.

By *induction*, DeMorgan's Law can be applied to a *countable* number of events.

The analogy between set operation and scalar operation above can be interpreted further:

- $\mathscr{P}(E_1 \cup E_2) = \mathscr{P}(E_1) + \mathscr{P}(E_2)$ if E_1 and E_2 are *disjoint* (mutually exclusive, $E_1 \cap E_2 = \varnothing$).
- $\mathscr{P}(E_1 \cap E_2) = \mathscr{P}(E_1) \cdot \mathscr{P}(E_2)$ if E_1 and E_2 are *independent*.
  Generally, $\mathscr{P}(E_1 \cap E_2) = \mathscr{P}(E_1) \cdot (E_2 | E_1)$ when they are not necessarily independent—conditional probability will be covered soon.

1.2 [She02, Sects. 2.3–2.4] Axioms of Probability

Definition 3 (Probability Law)**.** Probability $\mathscr{P}(E)$ is tells us how likely the event E is to occur. $\mathscr{P}$ is a *set function* that maps (*measurable*) subsets of Ω to $[0, 1]$, which satisfies following axioms:

(i) Normalization. $\mathscr{P}(\Omega) = 1$.

(ii) Nonnegativity. $\forall E \subseteq \Omega, 0 \leq \mathscr{P}(E) \leq 1$.

(iii) Countable Additivity. If a *countable* number of events $E_1, E_2, \ldots$ are disjoint, then $\mathscr{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathscr{P}(E_i)$.

Remark 2. $\mathscr{P}(\{\omega\})$ is short for $\mathscr{P}(\omega)$. The probability law always has a *set* as its input. [She02, Sect. 2.6] views probability as a *continuous* set function. Intuitively, probability is a measure of occurrence likelihood (belief).

Further useful properties:

- Complementation. $\mathscr{P}(E) + \mathscr{P}(E^{\mathsf{c}}) = 1$. In particular, $\mathscr{P}(\varnothing) = 0$.
- Monotonicity. $E_1 \subseteq E_2$ implies $\mathscr{P}(E_1) \leq \mathscr{P}(E_2)$.

  Proof. $E_2 = E_1 \cup (E_1^{\mathsf{c}} \cap E_2)$ implies $\mathscr{P}(E_2) \geq \mathscr{P}(E_1)$, due to the nonnegativity of $\mathscr{P}(E_1^{\mathsf{c}} \cap E_2)$. □

- Inclusion-Exclusion Theorem. $\mathscr{P}(A_1 \cup A_2) = \mathscr{P}(A_1) + \mathscr{P}(A_2) - \mathscr{P}(A_1 \cap A_2)$.

  Proof. $\mathscr{P}(E_1 \cup E_2) = \mathscr{P}(E_1 \cup (E_1^{\mathsf{c}} \cap E_2)) = \mathscr{P}(E_1) + \mathscr{P}(E_1^{\mathsf{c}} E_2) = \mathscr{P}(E_1) + (\mathscr{P}(E_2) - \mathscr{P}(E_1 E_2))$.
  Remark 3. "Splicing" strategy enables us to make use of countable additivity.

  □

- $\mathscr{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathscr{P}(E_i) - \sum_{i_1 < i_2} \mathscr{P}(E_{i_1} E_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdot < i_r} \mathscr{P}(E_{i_1} E_{i_2} \cdots E_{i_r}) + \cdots + (-1)^{n+1} \mathscr{P}(E_1 E_2 \cdots E_n)$

Example 3 (Continuous sample space)**.** $([0, 1], \mathcal{B}, \lambda)$.

Example 4 (Discrete sample space (may be infinite))**.** Toss a balanced coin. Build up Ω in your favor.

2 Week 2 (02/06)

2.1 [She02, Sect. 3.2] Conditional Probabilities

Definition 4 (Conditional Probability)**.**

References

[She02] R. Sheldon. *A First Course in Probability*. Pearson Education India, 2002.