Probability and Statistical Inference

Tianqi Zhang
Emory University

Apr 17th 2025

3 Discrete Probability Measures

3.1 Discrete Probability Measures

We start with discrete, countable latent space Ω.

Definition 3.1.1 (Discrete Probability Measure). A discrete probability measure on sample space Ω, finite or countable, is a sequence of $\{p_\omega\}_{\omega\in\Omega}$ of non-negative real numbers such that:

1. $p_\omega \geq 0$, $\forall\omega \in \Omega$,
2. $\sum_{\omega\in\Omega} p_\omega = 1$

A general definition that works not only for finite Ω but also for countable Ω since it allows in both cases to compute for any random event $A \subseteq \Omega$.

$$P(A) = \sum_{\omega\in A} P_\omega$$

such that it satisfies the **Komolgrov axioms:**

Theorem 3.1.2 Kolmogorov Axioms for Discrete Probability

Let Ω be a finite or countable set. A function $P : \mathscr{P}(\Omega) \to [0, 1]$ is a probability measure if:

1. **Non-negativity:** $P(A) \geq 0$ for all $A \subseteq \Omega$,
2. **Normalization:** $P(\Omega) = 1$,
3. **Countable additivity:** For any sequence of disjoint events $A_1, A_2, \cdots \subseteq \Omega$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Let's consider some common examples of discrete probability measures.

Example 3.1.3 Common Measures

- **Counting Measure:** $\mu(A) = |A|$. Define $\Omega = \mathbb{N}$ and let

$$P(A) = \text{number of elements in } A.$$

  This is called the **counting measure**. It's not a probability measure since the total mass is infinite, but it satisfies many of the same properties and helps build intuition.

- **Dirac Measure at $p \in \Omega$:**

$$\delta_p(A) = \begin{cases} 1, & \text{if } p \in A, \\ 0, & \text{otherwise.} \end{cases}$$

  This is called the **Dirac measure** at p. It concentrates all probability mass at a single point.

- **Uniform Distribution on a Finite Set** Let $\Omega = \{1, 2, \ldots, n\}$ and define

$$P(\omega) = \frac{1}{n} \quad \text{for all } \omega \in \Omega.$$

  This assigns equal weight to each element — a classic finite uniform distribution.

3.2 Basic Properties of Discrete Probability

Let P be a discrete probability measure on Ω and let $A, B \subseteq \Omega$.

Proposition 3.2.1 Basic Rules:

1. **Empty and full space:** $P(\varnothing) = 0, \quad P(\Omega) = 1$
2. **Complement rule:** $P(A^c) = 1 - P(A)$
3. **Difference rule:** If $B \subseteq A$, then

$$P(A \setminus B) = P(A) - P(B)$$

Corollary 3.2.2 Monotonicity:
If $A \subseteq B$, then $P(A) \leq P(B)$

Corollary 3.2.3 Partition Formula:
Suppose Ω is partitioned into disjoint subsets $\{H_i\}$ such that $\bigcup_i H_i = \Omega$. Then for any $A \subseteq \Omega$,

$$P(A) = \sum_i P(A \cap H_i).$$

Corollary 3.2.4 Sylvester Formula (Inclusion-Exclusion):
For any collection of subsets $\{A_i\}$ of Ω

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right)$$

Or a more intuitive version, for any $A, B \subseteq \Omega$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

3.3 Further Properties of Discrete Probability

Probability satisfies a number of important inequalities, logical identities, and limit properties. These results often appear in theoretical proofs and practical applications.

3.3.1 Inequalities and Bounds

Proposition 3.3.1 Boole's Inequality:
For any (finite or countable) collection of events $A_i \subseteq \Omega$, not necessarily disjoint:

$$P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i).$$

Remark 3.1 *This inequality reflects that when events overlap, simply summing their individual probabilities can overestimate the probability of their union. In general measure theory, Boole's inequality expresses the principle of **countable subadditivity**.*

Proposition 3.3.2 Bonferroni's Inequality:
Let $A_1, A_2, \ldots, A_n \subseteq \Omega$. Then:

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

Remark 3.2 *Bonferroni's inequality provides a lower bound on the probability that *all* events occur. It's commonly used in multiple hypothesis testing.*

3.3.2 Set-Theoretic Identities

Proposition 3.3.3 De Morgan's Law:
For any collection $\{A_i\}$,

$$\left(\bigcap_{i} A_i\right)^C = \bigcup_{i} A_i^C.$$

Taking probabilities on both sides:

$$P\left(\bigcap_{i} A_i\right) = 1 - P\left(\bigcup_{i} A_i^C\right).$$

where $A^C \equiv \Omega \setminus A$ denotes the complement of A.