Chapter 1

Probability

Probability Space

A *probability space* is a triplet $(\Omega, \mathcal{F}, P)$ in which

- Ω is a set called the *sample space*. The elements of Ω are called *outcomes*.
- $\mathcal{F}$ is a family of subsets of Ω, called *events*, forming a σ-algebra (also called a σ-field), i.e.,
  1. Ω is an event.
  2. If A is an event, then so is the complement, A^c, of A in Ω.
  3. If $A_1, A_2, \ldots$ is a (finite or infinite) sequence of events, then so is $A_1 \cup A_2 \cup \cdots$ (and also $A_1 \cap A_2 \cap \cdots$ by the above condition and De Morgan's Law).
- A *probability law*, P, which is a function that maps events to real numbers. In particular, the event A is mapped to $P(A)$ which is called the *probability* of the event A.

The notion of probability space is typically viewed in the context of a *random experiment* which is a process that produces an uncertain outcome in the sample space. The sample space, the events, and the probability law are then associated with the random experiment. We say that an event occurs if the experiment produces an outcome that belongs to that event.

The probability law satisfies the following three axioms:

Axiom 1 For every event A, $P(A) \geq 0$.

Axiom 2 $P(\Omega) = 1$.

Axiom 3 If $A_1, A_2, \ldots$ is a (finite or infinite) sequence of disjoint events, i.e., for all $k \neq l$ we have $A_k \cap A_l = \emptyset$, the empty set, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

Based on the axioms of probability, we can deduce the following properties:

Property 1 If A is an event and A^c is its complement in Ω, then

$$P(A^c) = 1 - P(A).$$

Property 2 $P(\emptyset) = 0$.

Property 3 If A is an event then $0 \leq P(A) \leq 1$.

Property 4 If A_1 and A_2 are events and $A_1 \subseteq A_2$, then

$P(A_1) \leq P(A_2)$.

This follows from the fact that A_1 and $A_2 - A_1$ are disjoint events[a] and $A_1 \cup (A_2 - A_1) = A_2$. Hence, from Axioms 3 and 1, we have

$$P(A_2) = P(A_1 \cup (A_2 - A_1)) = P(A_1) + P(A_2 - A_1) \geq P(A_1).$$

---

[a]Notice that $A_2 - A_1 = (A_1 \cup A_2^c)^c$ and as A_1 and A_2 are events so is $(A_1 \cup A_2^c)^c$ as the events form a σ-algebra and taking complements and unions of elements in a σ-algebra results in an element in the σ-algebra.

Property 5 If A_1 and A_2 are two events, not necessarily disjoint, then $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

Indeed, let $B = A_1 \cap A_2$, $C_1 = A_1 - B$, and $C_2 = A_2 - B$. Since the events C_1 and B are disjoint and their union is A_1, from Axiom 3 we get $P(A_1) = P(C_1 \cup B) = P(C_1) + P(B)$. Similarly, $P(A_2) = P(C_2 \cup B) = P(C_2) + P(B)$. Furthermore, as B, C_1, and C_2 are disjoint and their union is $A_1 \cup A_2$, Axiom 3 gives $P(A_1 \cup A_2) = P(B \cup C_1 \cup C_2) = P(B) + P(C_1) + P(C_2)$. Eliminating $P(C_1)$ and $P(C_2)$ in the above three equations, we get

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(B) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2). \end{aligned}$$

This property can be generalized to $n \geq 3$ events $A_1, A_2, \ldots, A_n$ as follows:

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_n) = & \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\
& + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \cdots \\
& -(-1)^t \sum_{1 \leq i_1 < i_2 \cdots < i_t \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_t}) + \cdots \\
& -(-1)^n P(A_1 \cap A_2 \cap \cdots \cap A_n).
\end{aligned}
$$

Example 1 The probability that a fair die shows 6 is $1/6$. What is the probability that at least one of two fair dice shows 6?

Let A_1 be the event that the first die shows 6 and A_2 be the event that the second die shows 6. We are interested in the probability of the event $A_1 \cup A_2$. We have $P(A_1) = P(A_2) = 1/6$. Axiom 3 is not applicable since A_1 and A_2 are not disjoint. What is true is that $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. To solve the problem, we need to know $P(A_1 \cap A_2)$. This may be given or deduced by assuming independence, a concept which will be defined later. In this case, $P(A_1 \cap A_2) = \frac{1}{6} \times \frac{1}{6}$ and $P(A_1 \cup A_2) = \frac{1}{6} + \frac{1}{6} - \frac{1}{6} \times \frac{1}{6} = \frac{11}{36}$.

Property 6 Subadditivity of Probability

Let $A_1, A_2, \ldots$ be events. Then,

$$P(A_1 \cup A_2 \cup \cdots) \leq P(A_1) + P(A_2) + \cdots .$$

Indeed, let $D_1 = A_1$ and $D_k = A_k - \cup_{l=1}^{k-1} A_l$ for $k \geq 2$. Then, the events $D_1, D_2, \ldots$ are disjoint and $A_1 \cup A_2 \cup \cdots = D_1 \cup D_2 \cup \cdots$. We conclude that

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots) &= P(D_1 \cup D_2 \cup \cdots) \\ &= P(D_1) + P(D_2) + \cdots \text{(Axiom 3)} \\ &\leq P(A_1) + P(A_2) + \cdots \\ &\qquad \text{(Property 4 as } D_k \subseteq A_k\text{).} \end{aligned}$$

Property 7 Continuity of Probability

Let $A_1, A_2, \ldots$ be events such that $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \lim_{k\to\infty} A_k$. Then, $P(A) = \lim_{k\to\infty} P(A_k)$.

Similarly, let $A_1, A_2, \ldots$ be events such that $A_1 \supseteq A_2 \supseteq \cdots$ and $A = \lim_{k\to\infty} A_k$. Then, $P(A) = \lim_{k\to\infty} P(A_k)$.

We prove the first part. Since $A = A_1 \cup A_2 \cup \cdots$, A is an event. From the proof of Property 6,

$$P(A) = P(A_1 \cup A_2 \cup \cdots) = P(D_1) + P(D_2) + \cdots .$$

As $A_1 \subseteq A_2 \subseteq \cdots$, we have for $k \geq 2$

$A_k = A_{k-1} \cup D_k = A_{k-2} \cup D_{k-1} \cup D_k = \cdots = D_1 \cup D_2 \cup \cdots \cup D_k$.

Since $D_1, D_2, \ldots$ are disjoint,

$$P(A_k) = P(D_1) + P(D_2) + \cdots + P(D_k).$$

The result follows from the above expressions of $P(A)$ and $P(A_k)$ by pushing k to infinity. The proof of the second part is similar.

Remark: Continuity of probability basically says that the orders of the function P and the limit can be interchanged:

$$P(\lim_{k\to\infty} A_k) = \lim_{k\to\infty} P(A_k)$$

if the events are nested, i.e., $A_1 \subseteq A_2 \subseteq \cdots$ or $A_1 \supseteq A_2 \supseteq \cdots$. This should not be taken for granted as not every function admits this.

For example, let Ω be the set of positive integers and $A_k = \{1, 2, \ldots, k\}$ for $k = 1, 2, \ldots$. Then, $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \{1, 2, \ldots\}$. Let F be the function of sets of integers defined by

$$F(S) = \begin{cases} 0 & \text{if } S \text{ is finite} \\ 1 & \text{if } S \text{ is infinite.} \end{cases}$$

Then, $F(A_k) = 0$ for any $k \geq 1$ and, therefore, $\lim_{k\to\infty} F(A_k) = 0$. However, $F(\lim_{k\to\infty} A_k) = F(A) = 1$.

Notice that the function F is not a probability law as it violates the axioms of probability. Indeed,

$$F(\text{set of odd positive integers}) = F(\text{set of even positive integers}) = 1.$$

Since the sets of odd integers and even integers are disjoint and their union is $\{1, 2, \ldots\}$, then by Axiom 3, $P(\{1, 2, \ldots\}) = 1 + 1 = 2$ which violates Axiom 2.

Conditional Probability

Let A and B be events with $P(B) > 0$. The *conditional probability* of A given B is defined as $P(A|B) = \frac{P(A \cap B)}{P(B)}$. It is easy to check that the conditional probability is a probability, i.e, it satisfies the three axioms of probability:

Axiom 1 For every event A, $P(A|B) \geq 0$.

Axiom 2 $P(\Omega|B) = 1$.

Axiom 3 If $A_1, A_2, \ldots$ is a (finite or infinite) sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \cdots |B) = P(A_1|B) + P(A_2|B) + \cdots .$$

Intuitively, $P(A|B)$ is the probability of event A if event B occurs.

Bayes' Rule Let A and B be events such that $P(A) > 0$ and $P(B) > 0$. Then,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Principle of Total Probability Let $B_1, B_2, \ldots$ be disjoint events such that $B_1 \cup B_2 \cup \cdots = \Omega$, i.e., $B_1, B_2, \ldots$ is a partition of Ω. Assume that $P(B_i) > 0$ for all $i \geq 1$. Then, for an event A,

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots .$$

Independence The events A and B are *independent* if $P(A \cap B) = P(A)P(B)$, or, equivalently, $P(A) = P(A|B)$ if $P(B) > 0$. Intuitively, this means that the probability that event A occurs is not affected by whether or not event B occurs.

The events $A_1, A_2, \ldots$ are said to be independent if

$$P(A_{l_1} \cap A_{l_2} \cap \cdots \cap A_{l_m}) = P(A_{l_1})P(A_{l_2}) \cdots P(A_{l_m})$$

for all positive and finite m and distinct $l_1, l_2, \ldots, l_m$.

For example, the events A, B, and C are independent if and only if

$$\begin{aligned} P(A \cap B) &= P(A)\, P(B) \\ P(A \cap C) &= P(A)\, P(C) \\ P(B \cap C) &= P(B)\, P(C) \\ P(A \cap B \cap C) &= P(A)\, P(B)\, P(C). \end{aligned}$$

The first three conditions are not sufficient to establish independence. For example, toss a fair coin three times. Let A be the event that tosses 1 and 2 give the same result, i.e., either HH or TT. Let B be the event that tosses 1 and 3 give the same result and C be the event that tosses 2 and 3 give the same result. Then,

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4},$$

and

$$P(A \cap B \cap C) = \frac{1}{4}.$$

The first three conditions hold but not the fourth. The events A, B, and C are not independent.

Also, the fourth condition does not imply the first three. Consider an unfair die which shows each of the numbers $1, 2, \ldots, 6$ with probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{24}, \frac{1}{24}, \frac{1}{24}$, respectively.

Let $A = B = \{1, 2\}$ and $C = \{1, 3\}$. Then,

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

and

$$P(A \cap B \cap C) = P(\{1\}) = \frac{1}{8}.$$

Hence, $P(A \cap B \cap C) = P(A)P(B)P(C)$ but $P(A \cap B) = P(A) = \frac{1}{2}$, which is not $P(A)P(B)$. The events A, B, and C are not independent.

Let $A_1, A_2, \ldots$ be a sequence of independent events. Let $B_1, B_2, \ldots$ be events obtained by taking complements, intersections, and unions of distinct events from $A_1, A_2, \ldots$. We will show that $B_1, B_2, \ldots$ are independent.

First, suppose that each B_i is either A_i or A_i^c. Suppose that the events $B_1, B_2, \ldots$ are not independent. Then, for some $m \geq 2$ and distinct $l_1, l_2, \ldots, l_m$,

$$P(\cap_{j=1}^{m} B_{l_j}) \neq \prod_{j=1}^{m} P(B_{l_j}).$$

Since $A_1, A_2, \ldots$ are independent, $B_{l_j} = A_{l_j}^c$ for some j. Now choose m and $l_1, l_2, \ldots, l_m$ such that the above inequality holds with the minimum number of l_j's such that $B_{l_j} = A_{l_j}^c$. Let t denote this minimum number. Then, any collection of the events $B_1, B_2, \ldots$ containing fewer than t events B_i for which $B_i = A_i^c$ are independent.

Without loss of generality, assume $B_{l_m} = A^c_{l_m}$. Then, the events $B_{l_1}, B_{l_2}, \ldots, B_{l_{m-1}}$ are independent as among them there are only $t-1$ events $B_{l_j} = A^c_{l_j}$. Hence,

$$P(\cap_{j=1}^{m-1} B_{l_j}) = \prod_{j=1}^{m-1} P(B_{l_j}).$$

Also, the events $B_{l_1}, B_{l_2}, \ldots, B_{l_{m-1}}, A_{l_m}$ are independent as among them there are only $t-1$ events $B_{l_j} = A^c_{l_j}$. Hence,

$$P((\cap_{j=1}^{m-1} B_{l_j}) \cap A_{l_m}) = \prod_{j=1}^{m-1} P(B_{l_j}) \times P(A_{l_m}).$$

We have

$$\begin{aligned} P(\cap_{j=1}^{m-1} B_{l_j}) &= P((\cap_{j=1}^{m-1} B_{l_j}) \cap B_{l_m}) + P((\cap_{j=1}^{m-1} B_{l_j}) \cap B_{l_m}^c) \\ &= P((\cap_{j=1}^{m-1} B_{l_j}) \cap B_{l_m}) + P((\cap_{j=1}^{m-1} B_{l_j}) \cap A_{l_m}) \end{aligned}$$

which implies that

$$\begin{aligned} P(\cap_{j=1}^{m} B_{l_j}) &= P(\cap_{j=1}^{m-1} B_{l_j}) - P((\cap_{j=1}^{m-1} B_{l_j}) \cap A_{l_m}) \\ &= \prod_{j=1}^{m-1} P(B_{l_j}) - \prod_{j=1}^{m-1} P(B_{l_j}) P(A_{l_m}) \\ &= \prod_{j=1}^{m-1} P(B_{l_j})(1 - P(A_{l_m})) = \prod_{j=1}^{m} P(B_{l_j}). \end{aligned}$$

This contradicts the assumption we started with.

We conclude that complementing preserves independence. It is also obvious from the definition of independence that intersections preserve independence, i.e., if $B_1, B_2, \ldots$ are intersections of distinct events from $A_1, A_2, \ldots$, then $B_1, B_2, \ldots$ are independent. It follows that unions, which by De Morgans' Law can be expressed in terms of intersections and complements, also preserve independence.

Limit Superior of Events

Let $A_1, A_2, \ldots$ be a sequence of subsets of Ω. We define the limit superior of this sequence as

$$\limsup_{k\to\infty} A_k = \cap_{l=1}^{\infty} \cup_{k=l}^{\infty} A_k.$$

An outcome ω belongs to the limit superior if and only if it belongs to $\cup_{k=l}^{\infty} A_k$ for every l, which is the case if and only if for every $l \geq 1$, there is some $k \geq l$ such that $\omega \in A_k$. This is precisely the case if and only if ω belongs to an infinite number of A_k's. Hence, the limit superior of a sequence of events is the set of outcomes that belong to an infinite number of the events. From its definition of union and intersection of events, the limit superior is an event which we denote by $A_{\text{i.o.}}$, where "i.o." stands for *infinitely often*.

Borel-Cantelli Lemma

Let $A_1, A_2, \ldots$ be a sequence of events.

a) If $\sum_{k=1}^{\infty} P(A_k) < \infty$, then $P(A_{\text{i.o.}}) = 0$.

b) If $\sum_{k=1}^{\infty} P(A_k) = \infty$ and the events $A_1, A_2, \ldots$ are independent, then $P(A_{\text{i.o.}}) = 1$.

Proof a)

$$
\begin{aligned}
P(A_{\text{i.o.}}) &= P(\cap_{l=1}^{\infty} \cup_{k=l}^{\infty} A_k) \\
&= \lim_{l\to\infty} P(\cup_{k=l}^{\infty} A_k) \\
&\qquad \text{(Property 7 as } \cup_{k=1}^{\infty} A_k \supseteq \cup_{k=2}^{\infty} A_k \supseteq \cdots) \\
&\le \lim_{l\to\infty} \sum_{k=l}^{\infty} P(A_k) \qquad \text{(Property 6)} \\
&= 0 \qquad \text{(by convergence of } \textstyle\sum_{k=1}^{\infty} P(A_k)).
\end{aligned}
$$

b) By De Morgan's Law and Property 7, each applied twice,

$$
\begin{aligned}
P(A^c_{\text{i.o.}}) &= P((\cap_{l=1}^{\infty} \cup_{k=l}^{\infty} A_k)^c) = P(\cup_{l=1}^{\infty}(\cup_{k=l}^{\infty} A_k)^c) \\
&= P(\cup_{l=1}^{\infty} \cap_{k=l}^{\infty} A_k^c) = \lim_{l\to\infty} P(\cap_{k=l}^{\infty} A_k^c) \\
&= \lim_{l\to\infty} \left(\lim_{m\to\infty} P(\cap_{k=l}^{m} A_k^c) \right) \\
&= \lim_{l\to\infty} \left(\lim_{m\to\infty} \prod_{k=l}^{m} P(A_k^c) \right) \qquad \text{(by independence)} \\
&= \lim_{l\to\infty} \left(\lim_{m\to\infty} \prod_{k=l}^{m} (1 - P(A_k)) \right) \\
&\le \lim_{l\to\infty} \left(\lim_{m\to\infty} \prod_{k=l}^{m} e^{-P(A_k)} \right) \qquad \text{(as } 1 - x \le e^{-x} \text{ for } x \ge 0) \\
&= \lim_{l\to\infty} \left(\lim_{m\to\infty} e^{-\sum_{k=l}^{m} P(A_k)} \right) \\
&= 0 \qquad \text{(by divergence of } \textstyle\sum_{k=1}^{\infty} P(A_k)).
\end{aligned}
$$

Example 2 Pick a real number x in the interval $(0, 1)$. Suppose you have an infinite number of friends, $F_1, F_2, \ldots$. You challenge each friend to guess an interval of given length that contains x. In particular, for each $k = 1, 2, \ldots$, you ask F_k to choose an interval of length l_k in the set $[0, 1)$. We ask what is the probability that an infinite number of your friends make the right guess.

Let A_k be the event that x lies in the interval chosen by F_k. Then, $P(A_k) = l_k$.

The Borel-Cantelli Lemma says that if the sum of the lengths of the intervals is finite, then the probability that an infinite number of your friends make the right guess is zero. For example, this is the case if $l_k = \frac{1}{2^k}$.

The Borel-Cantelli Lemma also says that if your friends choose the intervals independently of each other and if the sum of the lengths of the intervals is infinite, then the probability that an infinite number of your friends make the right guess is one. For example, this is the case if $l_k = \frac{1}{k}$.

Now, suppose that F_k chooses the interval $[0, 1/k)$. Then, the events $A_1, A_2, \ldots$ are not independent as $P(A_2) = \frac{1}{2}$, $P(A_3) = \frac{1}{3}$, and $P(A_2 \cap A_3) = \frac{1}{3}$. Although the sum of the lengths of the intervals is infinite, $P(A_{\text{i.o.}}) = 0$ as F_k for all $k > 1/x$ fails in making the right guess. Hence, the probability that an infinite number of your friends make the right guess is zero.

Finally, suppose that $l_k = l$ for all k, where $0 < l \leq 1$, and $F_2, F_3, \ldots$ choose the same interval as F_1. Then, the sum of the lengths of the intervals is infinite and the probability that an infinite number of your friends make the right guess is l.

Why σ-algebra? A natural question to ask is why not define probabilities for all subsets of the sample space. The answer is that this may not be consistent with the axioms of probability. In the following we give an example of the difficulties that may arise if probabilities are defined for all subsets of the sample space.

We consider the sample space $[0, 1)$ and assume that we have a uniform probability distribution over this continuous sample space.

If E is a subset of this sample space and χ is a real number, then we define

$$E_\chi = \{(\eta + \chi) - \lfloor(\eta + \chi)\rfloor : \eta \in E\},$$

i.e., E_χ is the set of fractional parts obtained by adding χ to the elements in E. Clearly, $E_\chi \subseteq [0, 1)$. We say that E_χ is a translation of E by χ.

First, we argue that if χ_1 and χ_2 differ by an integer, then $E_{\chi_1} = E_{\chi_2}$. Indeed, $(\eta + \chi_1) - \lfloor(\eta + \chi_1)\rfloor$ is the fractional part of $\eta + \chi_1$ and $(\eta + \chi_2) - \lfloor(\eta + \chi_2)\rfloor$ is the fractional part of $\eta + \chi_2$. If χ_1 and χ_2 differ by an integer, then their fractional parts are the same.

Next, we argue that there is a one-to-one correspondence between the elements in E and the elements in E_χ. Clearly, every element in E is translated to an element in E_χ. It remains to show that no two elements in E are translated to the same element in E_χ. Indeed, if η_1 and η_2 are elements in E that are translated to the same element, then

$$(\eta_1 + \chi) - \lfloor(\eta_1 + \chi)\rfloor = (\eta_2 + \chi) - \lfloor(\eta_2 + \chi)\rfloor.$$

This gives

$$\eta_1 - \eta_2 = \lfloor(\eta_1 + \chi)\rfloor - \lfloor(\eta_2 + \chi)\rfloor.$$

The right hand side is an integer. Hence, $\eta_1 - \eta_2$ is an integer. But E, being a subset of $[0, 1)$, has no two distinct elements that differ by an integer. We conclude that $\eta_1 = \eta_2$ which establishes the one-to-one correspondence between the elements in E and the elements in E_χ. Since the distribution is uniform over $[0, 1)$, $P(E_\chi) = P(E)$.

This proves that all translations of the same set have the same probability.

Now let V be a subset of $[0, 1)$ such that for every real number α, there is a unique real number ϕ in V that differs from α by a rational number[a]. The set V is called a *Vitali set*. For example, there is exactly one rational number in V and exactly one number of the form $\sqrt{2} - q$, where q is some rational number, in V. Clearly, the difference between any two numbers in V is not a rational number. Indeed, if $\phi_1, \phi_2 \in V$ are distinct and they differ by a rational number q, then $\alpha = \phi_1 + q$ is a real number that differs from both ϕ_1 and ϕ_2 by rational numbers, contradicting the uniqueness statement in the definition of V.

[a] The construction relies on the Axiom of Choice as we are choosing from each set of the form $\{\alpha + q : q \text{ is rational}\}$, where α is a real number, a unique representative $\phi \in [0, 1)$ to include in the set V.

Consider all translations of V by rational numbers and let $V_{q_1}, V_{q_2}, \ldots$ be all such distinct translations of V where $q_1, q_2, \ldots$ are rational numbers[a].

[a] Recall that the rational numbers are countable, i.e., can all be listed as a sequence.

We make two observations:

1. $\cup_{i=1}^{\infty} V_{q_i} = [0,1)$. Indeed, suppose $\alpha \in [0,1)$. Then, $\alpha = \phi + q$ for some real number $\phi \in V$ and rational number q. Since $\lfloor \alpha \rfloor = 0$, $\alpha = (\phi + q) - \lfloor (\phi + q) \rfloor$, which implies that $\alpha \in V_q$.

2. The sets $V_{q_1}, V_{q_2}, \ldots$ are disjoint. Indeed, suppose $\alpha \in V_{q_i} \cap V_{q_j}$. Then $\alpha = (\phi_i + q_i) - \lfloor (\phi_i + q_i) \rfloor$ and $\alpha = (\phi_j + q_j) - \lfloor (\phi_j + q_j) \rfloor$ for some real numbers ϕ_i and ϕ_j in V. Equating these two expressions for α, we get

$$\phi_i - \phi_j = \lfloor (\phi_i + q_i) \rfloor - \lfloor (\phi_j + q_j) \rfloor - q_i + q_j.$$

   The right hand side is the sum and difference of integers and rational numbers. Hence, ϕ_i and ϕ_j differ by a rational number. Since they are in V, they are identical, i.e., $\phi_i = \phi_j$. Hence,

$$q_i - q_j = \lfloor (\phi_i + q_i) \rfloor - \lfloor (\phi_j + q_j) \rfloor$$

   and q_i and q_j differ by an integer. Therefore, $V_{q_i} = V_{q_j}$.

Now, we will show that assigning probabilities $P(V_{q_1}), P(V_{q_2}), \ldots$ to the sets $V_{q_1}, V_{q_2}, \ldots$ is not consistent with the axioms of probability. From the first observation and the second axiom of probability, we have $P(\cup_{i=1}^{\infty} V_{q_i}) = P([0,1)) = 1$. From the second observation and the third axiom of probability, we have $P(\cup_{i=1}^{\infty} V_{q_i}) = \sum_{i=1}^{\infty} P(V_{q_i})$. We conclude that $\sum_{i=1}^{\infty} P(V_{q_i}) = 1$. Since translations of a set have equal probabilities, we obtain a contradiction since either $P(V) = 0$, in which case $\sum_{i=1}^{\infty} P(V_{q_i}) = 0$, or $P(V) > 0$, in which case $\sum_{i=1}^{\infty} P(V_{q_i}) = \infty$. In both cases, the sum is not equal to 1.