Mathematics relies on axiomatic framework to develop theory.

Axiom means (as per dictionary) self evident truth.

In mathematics, however, we need to answer "evident to whom?" No simple answer. But evident to someone sensible who reveals it to others and developes theory with these as a starting point.

Hence, in mathematics, axioms are to be treated as assumptions that hold without any requirement of proof. The idea is to put up logical, reasonable set of minimal starting requirements.

Axioms for probability :

When we think about probability theory, we typical think about 3 things:

1. Possible outcomes,
2. Events,
3. Probability values.

Axiomatic framework will define requirements for these 3 elements of probability theory.

When we talk about probability theory, we always have an experiment in mind whether exclusievely specified or not.

- Set of all possible outcomes:

$$\Omega = \{\omega : \text{set of all possible outcomes}\}$$

Nothing outside of Ω is considered to be valid.

- Events

Event, simply, is a collection of possible outcomes. So, potentially any set A s.t. $A \subseteq \Omega$ can be an event. We say that event A has occured, if experiment outcome $\omega \in A$.

All possible subsets may not be events of interest.

Let $\mathcal{F}$ denotes collection of events of interest.

$$\mathcal{F} = \{A : A \subseteq \Omega\}.$$

$\mathcal{F}$ is a set of sets. We need to put some restrictions of $\mathcal{F}$.

1. $\Omega \in \mathcal{F}$ (Certain event is of interest)
2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
   If A is of interest, then A^c is also of interest.
   In other words, if we are interested in occurance of A, then we are interested in non-occurance of A as well.

3. If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

If we are interested in $A_1, A_2, \ldots$ then we are interested in either of them happening.

Lemma 1: $\phi \in \mathcal{F}$.

Lemma 2: $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

Lemma 3: Let $A_1, \dots, A_N \in \mathcal{F}$, then $\bigcup_{n=1}^{N} A_n \in \mathcal{F}$.

finite union

Probability measure:

Probability measure P on measurable space $(\Omega, \mathcal{F})$ is:

1. $P : \mathcal{F} \to [0,1]$, i.e. $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{F}$, and for any $A \notin \mathcal{F}$ prob. is not defined.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, \dots$ are disjoint sets in $\mathcal{F}$, then
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Lemma: P is monotone, i.e. $P(A) \leq P(B)$ whenever $A \subseteq B$.

Proof: $B = A \cup (B \setminus A)$. Note that

$A \cap (B \setminus A) = \phi$.

$\Rightarrow P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$

$\Rightarrow P(B) \geq P(A)$ as $P(B \setminus A) \geq 0$.

Corollary: $P(B \setminus A) = P(B) - P(A)$ whenever $A \subseteq B$.

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

Proof: $A\cup B = \underbrace{A \cup (B \setminus (A\cap B))}_{\text{Mutually disjoint}}$

Moreover, $(A\cap B) \subseteq B$. Thus,

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

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

Continuity of Prob. measure:

Continuity of a fn:

$f: \mathcal{X} \to \mathcal{X}$ is continuous at $x \in \mathcal{X}$ if for every sequence $\{x_n\}$ s.t. $x_n \to x$,

$$\lim_{n\uparrow\infty} f(x_n) = f(x) = f(\lim_{n\uparrow\infty} x_n).$$

How do we define continuity of P?

for any $\{A_n\} \in \mathcal{F}$, s.t. $A_n \to A$ $(A \in \mathcal{F})$

Always happen but not clear why.

then $\lim_{n\uparrow\infty} P(A_n) = P(A) = P(\lim_{n\uparrow\infty} A_n)$.

How should we define limit for sequence of sets?

In general it is not obvious, but lets consider special cases:

Let $\{A_n\}$ be a monotone increasing seq, i.e. $A_n \subseteq A_{n+1}$, then visually you can see that

$$\lim_{n\uparrow\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

Continuity from below:

If $\{A_n\} \in \mathcal{F}$, monotone increasing, and let $A_n \uparrow A$, then $P(A_n) \uparrow P(A)$.

$B_1 = A_1$, and $B_k = A_k \setminus A_{k-1}$

$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k$ and B_k's are disjoint.

$$P(A) = \sum_{k=1}^{\infty} P(B_k) = \lim_{n\uparrow\infty} \sum_{k=1}^{n} P(B_k)$$

$$= \lim_{n\uparrow\infty} P\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\uparrow\infty} P(A_n).$$

Continuity from above:

If $\{A_n\} \in \mathcal{F}$ monotone decreasing, and let $A_n \downarrow A$, then $\lim P(A_n) = P(A)$.

Note that $A = \bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c$.

$$P(A) = P\left(\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \quad \left[A_n^c \text{ is monotone increasing}\right]$$

$$= 1 - \lim_{n\uparrow\infty} P(A_n^c)$$ by continuity from below.

$$= \lim_{n\uparrow\infty} \left[1 - P(A_n^c)\right] = \lim_{n\uparrow\infty} P(A_n).$$

Finite sub-additivity:

$A_1, A_2, \ldots, A_N \in \mathcal{F}$, then $P(\bigcup_{n=1}^{N} A_n) \leq \sum_{n=1}^{N} P(A_n)$.

Define, $B_1 = A_1$, $B_2 = A_2 \setminus A_1, \ldots,$ $B_k = A_k \cap A_{k-1}^c \cdots \cap A_1^c$.

Note that $B_k \subseteq A_k$ and $\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k$.

Moreover, B_k's are disjoint. Thus,

$$P\left(\bigcup_{k=1}^{n} A_k\right) = P\left(\bigcup_{k=1}^{n} B_k\right)$$
$$= \sum_{k=1}^{n} P(B_k) \quad - (\sigma\text{-additivity of } P)$$
$$\leq \sum_{k=1}^{n} P(A_k) \quad - (\text{monotonicity of } P).$$

Conditional Probability:

Consider a prob. space $(\Omega, \mathcal{F}, P)$. Let D be any set $D \in \mathcal{F}$ s.t. $P(D) > 0$.

Probability of A conditioned on D is

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)}.$$

Define, $P_{|D}(A) = P(A \mid D)$.

- Show that $P_{|D}(\cdot)$ is a probability measure on $(\Omega, \mathcal{F})$.

① Clearly, $P_{|D}$ is a function from $\mathcal{F}$ to $[0,1]$

② $P_{|D}(\Omega) = 1$

③ If $A_1, A_2, \ldots$ mutually disjoint, then $P_{|D}(\bigcup A_n) = \sum P_{|D}(A_n)$

Law of total probability:

Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be partition of Ω s.t. $P(A_k) > 0 \ \forall\ k = 1, \ldots, n$. Then,

$$P(B) = \sum_{k=1}^{n} P(B \mid A_k) P(A_k) \quad \forall\ B \in \mathcal{F}.$$

Example:

		Components	defective	Prob.
Box	1	2000	5%	0.05
	2	500	40%	0.4
	3	1000	10%	0.1
	4	1000	10%	0.1

choose box at random and pick an item at random. What is the prob. of finding a defective item?

Baye's Theorem:

$$P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)}.$$

Let the item be defective, what is the prob that it was taken from box 2.

Independent events:

Events A_1 and A_2 are independent if

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

Independance corresponds to prob. measure. Events may be independant with one measure and may not be so according to some other measure.

When should we say events $A_1, A_2, \dots, A_n$ form collection of independent event?

Events $A_1, A_2, \dots, A_n$ are said to be independent if for every sub-collection $A_{\sigma(1)}, \dots, A_{\sigma(k)}$,

$$P\left(\bigcap_{j=1}^{k} A_{\sigma(j)}\right) = \prod_{j=1}^{k} A_{\sigma(j)}, \quad \text{for every } k=1,\dots,n.$$

Total $2^n - n - 1$ conditions to verify!!

Independence of collection of events:

Let $\mathcal{A}_1$ and $\mathcal{A}_2 \subseteq \mathcal{F}$ be two collection of events.

$\mathcal{A}_1$ and $\mathcal{A}_2$ are said to be independent if

$\forall\, A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$, $P(A \cap B) = P(A) \cdot P(B)$.

It is important to note that even if $\mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2$, $\sigma(\mathcal{A}_1)$ need not be independent of $\sigma(\mathcal{A}_2)$. To see this consider the following example:

Experiment: Two dice R and G are thrown, and outcome is observed.

$$\Omega = \{(x_1, x_2) : x_1 \in \{1, 2, \dots, 6\}, x_2 \in \{x_1, \dots, x_6\}\}$$

$\mathcal{F} = \wp(\Omega)$ (power set of Ω)

$P(\{x_1, x_2\}) = \frac{1}{36}$ $\forall\, x_1$ and x_2. (This completely defines the prob. measure).

Consider 3-events:

$A = \{(x_1, x_2) : x_1 = 1,3,5\}$ $\quad P(A) = \frac{1}{2}$

$B = \{(x_1, x_2) : x_2 = 1,3,5\}$ $\quad P(B) = \frac{1}{2}$

$C = \{(x_1, x_2) : x_1 + x_2 = 1,3,5,7,9,11\}$. $\quad P(C) = \frac{1}{2}$

Now, define two collections:

$\mathcal{A}_1 = \{A\}$ and $\mathcal{A}_2 = \{B, C\}$.

Note that $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent collections:

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

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

But $\sigma(\mathcal{A}_1)$ is not independent of $\sigma(\mathcal{A}_2)$

Note that: $B \cap C \in \sigma(\mathcal{A}_2)$. Now,

$$P(A \cap (B \cap C)) = 0 \neq \underbrace{P(A)}_{1/2} \cdot \underbrace{P(B \cap C)}_{1/4}$$

mutually exclusive

Conclusion: Even when collections $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent, $\sigma(\mathcal{A}_1)$ need not be independent of $\sigma(\mathcal{A}_2)$.

One important exception does exist:

Thm: If $\mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2$, and $\mathcal{A}_1$ and $\mathcal{A}_2$ are π-systems, then $\sigma(\mathcal{A}_1) \perp\!\!\!\perp \sigma(\mathcal{A}_2)$.

What is a π-system?

$\mathcal{A}_1$ is called π-system if $\forall$ A and $B \in \mathcal{A}_1$, then $A \cap B \in \mathcal{A}_1$. (Closed under pairwise intersections).

Limits of sequence sets:

Let $A_1, A_2, \ldots$ $(\{A_n\}_{n\geq 1})$ be a sequence of sets.

How should we determine:

1. Whether a limit exists?
2. How should we identify the limiting set?

Convergence typically refers to element coming close as index n becomes large. We can say that sets are becoming similar if they have more and more same elements and less and less different elements. It can be argued that elements in the limiting set must occur in infinitely many sets.

Thus, an element ω may belong to the limiting set if

$$\underbrace{\sum_{n=1}^{\infty} 1_n(\omega)}_{\text{counts \# of times } \omega \text{ has been a member of a set.}} = \infty, \quad \text{where } 1_n(\omega) = 1 \text{ if } \omega \in A_n. \quad -(1)$$

Remark: If $\sum_{n=1}^{\infty} 1_n(\omega) < \infty$ for some ω, then

the ω belongs to only finitely many sets in the sequence and eventually left out with no chance of occuring again. Clearly, such ω can not be a part of the limiting set.

We classify ω satisfying (1) in two types.

$$\overline{A} = \{\omega : \forall n \ \exists k \geq n \text{ s.t. } \omega \in A_k\}$$

$$\underline{A} = \{\omega : \exists n \text{ s.t. } \forall k \geq n \text{ s.t. } \omega \in A_k\}$$

Following observations can be made:

① $\underline{A} \subseteq \overline{A}$,

② $\overline{A} = \{\omega : \limsup 1_{A_n}(\omega) = 1\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$,

③ $\underline{A} = \{\omega : \liminf 1_{A_n}(\omega) = 1\} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

$\overline{A}$ is called $\limsup A_n$, and
$\underline{A}$ is called $\liminf A_n$.

We say that $\lim A_n$ exists and equals A, if $\overline{A} = \underline{A} = A$.

Two most important results:

Borel-Cantelli Lemmas:

Lemma 1 (BC Lemma 1):
In prob space $\{\Omega, \mathcal{F}, P\}$, let $\{A_n\}_{n\geq 1}$ be a seq of valid events $(A_n \in \mathcal{F}, \forall n)$ s.t.
$\sum_{n=1}^{\infty} P(A_n) < \infty$. Then, $P(\limsup A_n) = 0$.

Proof:

$$P(\limsup A_n) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)$$
$$\leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \quad \forall n.$$
$$\leq \sum_{k=n}^{\infty} P(A_k) \quad \text{(Union bound)}$$

Now, $\sum_{k=n}^{\infty} P(A_k) \to 0$ as $n \to \infty$, as $\sum_{k=1}^{\infty} P(A_k) < \infty$.

— x — x — x —

Lemma 2 (BC Lemma 2):

If $\{A_n\}_{n \geq 1}$ is a collection of independent events, then:

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \text{ implies } P(\limsup A_n) = 1.$$

<u>Proof</u>: Note that $(\limsup A_n)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$.

Define, $B_n = \bigcap_{k=n}^{\infty} A_k^c$. Now,

$$B_n \subseteq \bigcap_{k=n}^{N} A_k^c, \quad \forall N \geq n.$$

$$P(B_n) \leq P\left(\bigcap_{k=n}^{N} A_k^c\right) \quad \text{(Monotonicity of Prob. measure)}$$

$$= \prod_{k=n}^{N} P(A_k^c) \quad (A \perp\!\!\!\perp B \Rightarrow A^c \perp\!\!\!\perp B^c).$$

$$= \prod_{k=n}^{N} (1 - P(A_k))$$

$$\leq \prod_{k=n}^{N} \exp\{-P(A_k)\} \quad \because 1 - x \leq e^{-x}.$$

$$= \underbrace{\exp\Big\{-\underbrace{\sum_{k=n}^{N} P(A_k)}_{\to \infty}\Big\}}_{\to 0 \;\cdot\; \forall n}, \quad \forall N.$$

$$P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right)$$

$$\leq \sum_{n=1}^{\infty} \underbrace{P(B_n)}_{=0 \;\forall n} = 0.$$