Axioms of Probability

Without worrying here about the historical origins, philosophical interpretations, or any other aspects of meaning, which we treat in their appropriate places, let us now only describe the resulting formalism, which is due to Kolmogorov (1933):

The ingredients are, first,

the formal mathematical prerequisites

(1) formal language $\mathcal{L}$, & its associated
(2) logic, or rules of valid argument construction (e.g. 1st-order/predicate logic, prob. supplemented by 2nd or higher order logic)
(3) set theory (a theory in predicate/1st-order logic, p. 25 Huber)

formal mathematical prerequisites

(4) truth conditions / semantics / model theory: propositions in L carry meaning & truth value only under an interpretation. There is logical truth & logical implication (Huber p. 11) before any of that, however (enough for proving theorems).

This is the main point for prob. proper.

(5) event space / sample space, a set S. We require not merely S, but a specific class of subsets of S, a σ-algebra $\mathcal{A} \subseteq \mathcal{P}(S)$, satisfying the axioms

(i) $\emptyset \in \mathcal{A}$

(ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

(iii) $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{A} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

The sample space is the generic form of the situation encountered in practice:

In trying to understand & quantify statistical patterns in a given population (a collection under observation) or in trying to quantify & bound uncertainty (objective & physical, or subjective & epistemic) concerning the outcome of a 'random' process, we conduct experiments & make observations of pop's. In probability, it is the experiment which symbolizes the typical setup: a well-designed experiment has a well-defined set of possible outcomes (whose precise determination would give exact/precise knowledge, an ideal only approximable by the uncertainties of imprecise/inexact knowledge).

Probability concerns these possible outcomes S of an experiment.

An element A of the σ-algebra $\mathcal{A}$ is called an event, & symbolizes the sub-sets of possible outcomes, even if a set A is never itself a single outcome (e.g. rolling an even on a single die).

* Ash calls an event some-thing answering a Yes/No quest. about the out-comes of an experiment. (p. 166)

To get a measure of A, in the sense of likelihood, we therefore require:

(6) axiomatic definition of a probability function (or measure):

$$\boxed{P : \mathcal{A} \to \mathbb{R}}$$

satisfying

(i) $0 \leq P(E) \leq 1$ for every $E \in \mathcal{A}$
(i.e. $P : \mathcal{A} \to [0,1]$)

(ii) $P(S) = 1$

(iii) For any sequence of mutually disjoint events $(E_n)_{n \in \mathbb{N}}$ in $\mathcal{A}$,

$$P\left(\bigsqcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

Immediate consequences of the axioms:

(1) $\boxed{P(\emptyset) = 0}$

pf: Define a sequence E_n of disjoint events by

$$E_1 \stackrel{def}{=} S$$
$$E_n \stackrel{def}{=} \emptyset \quad \text{for all } n \geq 2$$

& apply axiom (iii):

$$\begin{aligned} P(S) &= P(S \sqcup \emptyset \sqcup \emptyset \ldots) \\ &= P\left(\bigsqcup_{n=1}^{\infty} E_n\right) \\ &\stackrel{(iii)}{=} \sum_{n=1}^{\infty} P(E_n) \\ &= P(S) + \sum_{n=2}^{\infty} P(\emptyset) \end{aligned}$$

or $1 = 1 + \sum_{n=2}^{\infty} P(\emptyset)$.

$-1 \quad -1$

$$\Longrightarrow 0 = \sum_{n=2}^{\infty} \underbrace{P(\emptyset)}_{\geq 0 \text{ by axiom (i)}}$$

$\Longrightarrow P(\emptyset) = 0$.

QED

(2) $\boxed{P(E \sqcup F) = P(E) + P(F)}$ for any two disjoint events $E, F \in \mathcal{A}$.

pf: Define E_n in $\mathcal{A}$ by

$$E_1 = E$$
$$E_2 = F$$
$$E_n = \emptyset \text{ for } n \geq 3$$

and apply axiom (iii) & (1) above:

$$\begin{aligned} P(E \sqcup F) &= P(E \sqcup F \sqcup \emptyset \sqcup \emptyset \cdots) \\ &= P\left(\coprod_{n=1}^{\infty} E_n\right) \\ &= \sum_{n=1}^{\infty} P(E_n) \\ &= P(E) + P(F) + \underbrace{\sum_{n=3}^{\infty} P(\emptyset)}_{= 0 \text{ by } (1)} \\ &= P(E) + P(F) \end{aligned}$$

QED

(3) Let $E^c \stackrel{\text{def}}{=} S - E \stackrel{\text{def}}{=} \{s \in S \mid s \notin E\}$ be the complement of an event $E \in \mathcal{A}$. (so also $E^c \in \mathcal{A}$ by (ii) of defn of $\mathcal{A}$)

Then,

$$\boxed{P(E^c) = 1 - P(E)}$$

<u>pf:</u> $S = E \sqcup E^c$ so by (2) above

$$1 \stackrel{(ii)}{=} P(S) = P(E \sqcup E^c)$$

$$\stackrel{(2)}{=} P(E) + P(E^c) \quad \underline{\text{QED}}$$

(4) $$\boxed{E \subseteq F \implies P(E) \leq P(F)}$$ (for any events $E, F \in \mathcal{A}$)

("P is increasing, & monotonic")

<u>pf:</u> $F = F \cap S = F \cap (E \sqcup E^c)$

$$= \underbrace{(F \cap E)}_{= E} \sqcup \underbrace{(E^c \cap F)}_{\neq \emptyset \text{ nec.}}$$

$$= E \sqcup (E^c \cap F)$$

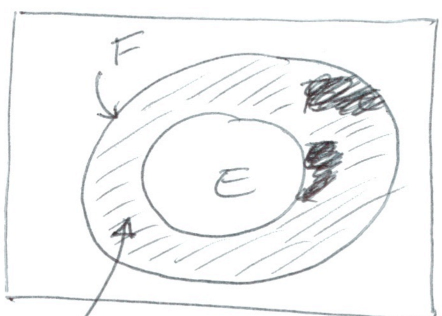
$$\implies P(F) \stackrel{(2)}{=} P(E) + \underbrace{P(E^c \cap F)}_{\geq 0 \text{ by (i)}} \geq P(E) \quad \underline{\text{QED}}$$

Remark: Note another consequence:

$$\boxed{E \subseteq F} \implies F - E = F \cap E^c$$

in the proof of (4) $\left\{ \begin{aligned} &\implies P(F) = P(E) + P(E^c \cap F) \\ &\qquad\qquad = P(E) + P(F-E) \end{aligned} \right.$

$$\implies \boxed{P(F-E) = P(F) - P(E)}$$

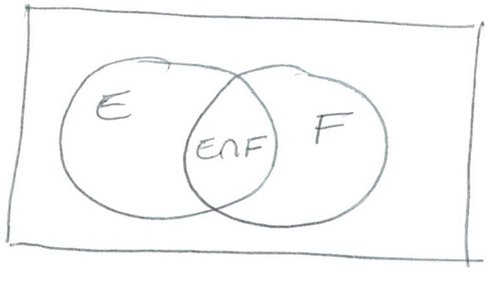


$F \cap E^c = F - E$

$\boxtimes$

(5) For arbitrary events $E, F \in \mathcal{A}$,

$$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F)}$$



inclusion-exclusion identity

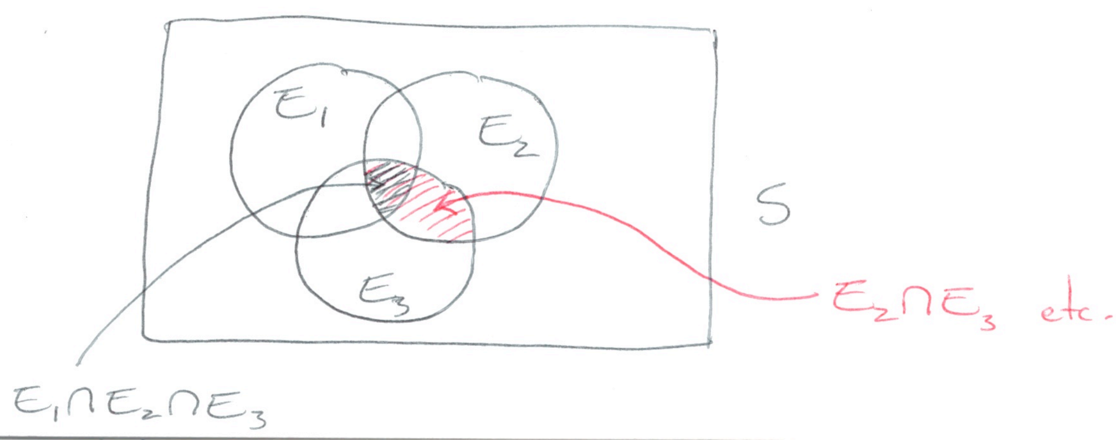
pf:
$$E \cup F = E \sqcup (F \cap E^c)$$
$$= E \sqcup (F - (E \cap F))$$

$$\Rightarrow P(E \cup F) \overset{(2)}{=} P(E) + P(F - (E \cap F))$$
$$\overset{(4)\text{ Remark}}{=} P(E) + P(F) - P(E \cap F)$$

QED

(6) For arbitrary events E_1, E_2, E_3,

$$\boxed{\begin{aligned} P(E_1 \cup E_2 \cup E_3) = {} & P(E_1) + P(E_2) + P(E_3) \\ & - P(E_1 \cap E_2) - P(E_1 \cap E_3) \\ & - P(E_2 \cap E_3) \\ & + P(E_1 \cap E_2 \cap E_3) \end{aligned}}$$



pf: $P(E_1 \cup (E_2 \cup E_3))$

$$\overset{(5)}{=} P(E_1) + P(E_2 \cup E_3) - P(\underbrace{E_1 \cap (E_2 \cup E_3}_{= (E_1 \cap E_2) \cup (E_1 \cap E_3)}$$

$$= P(E_1) + P(E_2 \cup E_3) - P((E_1 \cap E_2) \cup (E_1 \cap E$$

apply (6) to these

$$= P(E_1) + \Big(P(E_2) + P(E_3) - P(E_2 \cap E_3)\Big)$$

$$- \Big(P(E_1 \cap E_2) + P(E_1 \cap E_3) - P(\underbrace{(E_1 \cap E_2) \cap (E_1 \cap E_3)}_{= E_1 \cap E_2 \cap E_3})\Big)$$

$$= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

QED

general inclusion – exclusion identity

(7)

$$\boxed{\begin{aligned} &p(E_1 \cup E_2 \cup \cdots \cup E_n) \\ &= \sum_{\substack{i_1 < \cdots < i_k \\ (k=1,\ldots,n)}} (-1)^{k+1} p(E_{i_1} \cap \cdots \cap E_{i_k}) \\ &= \sum_{i=1}^{n} p(E_i) - \sum_{i_1 < i_2} p(E_{i_1} \cap E_{i_2}) \\ &\quad + \sum_{i_1 < i_2 < i_3} p(E_{i_1} \cap E_{i_2} \cap E_{i_3}) \\ &\quad + \cdots + (-1)^{n+1} p(E_1 \cap \cdots \cap E_n) \end{aligned}}$$

pf: by induction: $n=1$ is trivial, & $n=2$ was the previous point (5), which can serve as the base step. Thus, for the inductive step suppose true for some $n \geq 2$ & consider the $(n+1)$st case:

$$p((E_1 \cup \cdots \cup E_n) \cup E_{n+1})$$

$$= p(E_1 \cup \cdots \cup E_n) + p(E_{n+1}) - p((E_1 \cup \cdots \cup E_n) \cap E_{n+1})$$

$$= p(E_1 \cup \cdots \cup E_n) + p(E_{n+1}) - p((E_1 \cap E_{n+1}) \cup (E_2 \cap E_{n+1}) \cup \cdots \cup (E_n \cap E_{n+1}))$$

by induction hyp. on $p(E_1 \cup \cdots \cup E_n)$

$$= \sum_{i=1}^{n} p(E_i) + p(E_{n+1}) - \sum_{1 \le i_1 < i_2 \le n} p(E_{i_1} \cap E_{i_2}) + \cdots + (-1)^{n+1} p(E_1 \cap \cdots \cap E_n)$$

by induction hyp. on $p((E_1 \cap E_{n+1}) \cup (E_2 \cap E_{n+1}) \cup \cdots \cup (E_n \cap E_{n+1}))$

$$- \Big[\sum_{i=1}^{n} p(E_i \cap E_{n+1}) - \sum_{1 \le i_1 < i_2 \le n} p(E_{i_1} \cap E_{i_2} \cap E_{n+1}) + \cdots + (-1)^{n+1} \underbrace{p(E_1 \cap \cdots \cap E_n \cap E_{n+1})}_{= p((E_1 \cap E_{n+1}) \cap (E_2 \cap E_{n+1}) \cap \cdots \cap (E_n \cap E_{n+1}))} \Big]$$

$= (-1)^{(n+1)+1}$

$$= \sum_{1 \le i_1 < \cdots < i_k \le n+1} (-1)^{n+2} p(E_{i_1} \cap \cdots \cap E_{i_k}) \qquad \boxtimes$$