Probability I [gmd9-A]

Elementary probabilities:

Probabilities can be defined (i) axiomatically, (ii) from relative frequencies of events, or (iii) by assignment to elementary events based on one of two criteria [e.g. Papoulis 1991, Sec. 1-2]:

- symmetry (somewhat circularly),
- principle of insufficient reason (somewhat dubiously).

Ambiguities may arise if the range of events is

- discrete and infinite,
- continuous (such as in Bertrand's paradox).

In general, probability densities change under a transformation of variables. Particular assignments are often hard to justify.

In classical equilibrium statistical mechanics, uniform probability densities are assigned to canonical coordinates under specific circumstances. This choice is most readily justifiable in the case of action-angle coordinates.

Canonical transformations leave probability densities invariant. However, the existence of action-angle coordinates is limited to integrable systems. Only a tiny minority of many-body systems are integrable.

Elements of set theory:

A *set* S is a collection of *elements* ϵ_i: $S = \{\epsilon_1, \epsilon_2, \ldots\}$.

All elements of *subset* A are also elements of S: $A \subset S$.

The *empty set* $\emptyset$ contains no elements.

If S contains n elements then the number of subsets is 2^n.

Hierarchy of subsets: $\emptyset \subset A \subset A \subset S$.

Transitivity of subset relation: If $C \subset B$ and $B \subset A$ then $C \subset A$.

Mutuality condition for subsets: $A = B$ iff[1] $A \subset B$ and $B \subset A$.

Union of subsents: $C = A + B$ or $C = A \cup B$.
All elements of S that are contained in A or in B or in both.

Intersection of subsets: $C = AB$ or $C = A \cap B$.
All elements of S that are contained in A and in B.

Union and intersection are

– *commutative*: $A + B = B + A, \quad AB = BA$;
– *associative*: $(A + B) + C = A + (B + C), \quad (AB)C = A(BC)$;
– *transitive*: $A(B + C) = AB + AC$.

Some consequences:
$A + A = A, \quad A + \emptyset = A, \quad A + S = S; \quad AA = A, \quad A\emptyset = \emptyset, \quad AS = A.$

Mutually exclusive subsets have no common elements: $AB = \emptyset$.

Partition $\mathcal{P} = [A_1, A_2, \ldots]$ of S into mutually exclusive subsets:
$S = A_1 + A_2 + \ldots$ with $A_iA_j = \emptyset$ for $i \neq j$.

The *complement* $\bar{A}$ of subset A has all elements of S that are not in A.

Some consequences:
$A + \bar{A} = S, \quad A\bar{A} = \emptyset, \quad \bar{\bar{A}} = A, \quad \bar{S} = \emptyset, \quad \bar{\emptyset} = S.$

Duality principle:
A set identity is preserved if all sets are replaced by their complements.

DeMorgan's law: $\overline{A + B} = \bar{A}\bar{B}, \quad \overline{AB} = \bar{A} + \bar{B}$.

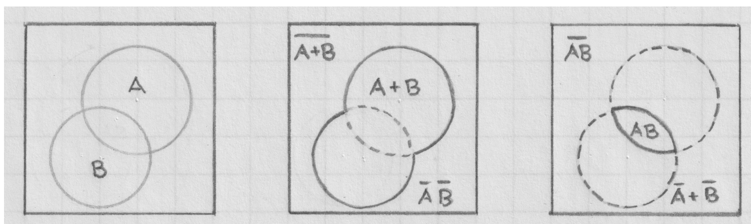
[1]if and only if

Illustration of deMorgan's law in Venn diagrams:

Consider a set S of points inside a square with two subsets A and B that are not mutually exclusive as shown on the left.

The diagram in the center shows the complementary sets $A + B$ and $A \bar{+} B$. The latter is identical to the set $\bar{A}\bar{B}$.

On the right, the complementary sets AB and $\bar{A}B$ are highlighted. The latter is identical to the set $\bar{A} + \bar{B}$.



Infinite sets can be *enumerable* (e.g. the set of integers) or *nonenumerable* (e.g. the set of real numbers).

Consider a set S of points on a line, a plane, or in 3D space with a well-defined distance function between points. This allows the definition of a *neighborhood* of a point.

- A point is *isolated* if there is a neigborhood around it which does not contain any other point from the set S. Otherwise it is an *accumulation* point.
- In a *dense set* all points are accumulation points.
- For an *interior* point of the set S there exists a neighborhood around it which contains points exclusively from S.
- Interior points are necessarily accumulation points, but not all accumulation points are interior points.
- An *open set* contains only interior points.
  Open line segment: $x \in (a, b)$ if $a < x < b$.
- A *closed set* contains all its accumulation points.
  Closed line segment: $x \in [a, b]$ if $a \leq x \leq b$.
- A *region* is an open set such that any two points are connected by a line of interior points.

The nunenumerable set $\mathbb{R}$ of real numbers consists of two interpenetrating dense sets:

– the (enumerable) set $\mathbb{Q}$ of rational numbers,

– the complementary set $\bar{\mathbb{Q}}$ irrational numbers.

Elements of probability theory:

Transcription from set theory:

- set $\rightarrow$ sample space
- subset $\rightarrow$ event
- element $\rightarrow$ elementary event

Sample space S: set of all possible outcomes in an experiment.

Event $A \subset S$: possible outcome of experiment.

Probability axioms [Kolmogorov 1933]:

- $P(A) \geq 0$ for all $A \subset S$,
- $P(S) = 1$,
- $P(A + B) = P(A) + P(B)$ if $AB = \emptyset$.

Some immediate consequences [nex94]

- $P(\emptyset) = 0$,
- $P(\bar{A}) = 1 - P(A)$,
- $P(A + B) = P(A) + P(B) - P(AB)$.

Under idealized circumstances the sample space is divisible into elementary, mutually exclusive, events to which equal probabilities can be assigned for reasons of symmetry:

$$S = A_1 + \cdots + A_N \text{ with } A_i A_j = \emptyset \text{ for } i \neq j \text{ and } P(A_i) = \frac{1}{N}.$$

Joint and conditional probabilities:

Joint probability: $P(AB)$ (event A and event B)

Conditional probability: $P(A|B)$ (event A if event B)

Relations: $P(AB) = P(A|B)P(B) = P(B|A)P(A)$

Simple consequences:

- If $A \subset B$ then $P(A|B) = P(A)/P(B)$
- If $B \subset A$ then $P(A|B) = 1$

Conditional probabilities satisfy probability axioms [nex90].

Bayes' theorem: $P(A|B) = P(B|A)\dfrac{P(A)}{P(B)}$

Statistical independence:

Case of two events A, B.

Criterion: $P(AB) = P(A)P(B)$

Simple consequences:

- $P(A|B) = P(A), \quad P(B|A) = P(B)$
- $P(A\bar{B}) = P(A)P(\bar{B}), \quad P(\bar{A}B) = P(\bar{A})P(B), \quad P(\bar{A}\bar{B}) = P(\bar{A})P(\bar{B})$

Case of three events A, B, C.

Criteria: pairwise statistical independence is not sufficient!

- $P(AB) = P(A)P(B)$
- $P(AC) = P(A)P(C)$
- $P(BC) = P(B)P(C)$
- $P(ABC) = P(A)P(B)P(C)$

Statistical uncertainty and information:

An experiment has n possible outcomes that occur with probabilities $P_1, P_2, \ldots, P_n$.

Properties that must be satisfied by any quantitative measure of *uncertainty*:

1. The uncertainty is a function of the probabilities of all possible outcomes: $\Sigma = \Sigma(P_1, P_2, \ldots, P_n)$.
2. The uncertainty is symmetric under all permutations of the P_i.
3. The maximum uncertainty occurs if all P_i are equal.
4. The uncertainty is zero if one of the outcomes has probability $P_i = 1$.
5. The combined outcome of two independent experiments has an uncertainty equal to the sum of the uncertainties of the outcomes of each experiment.

$$\Rightarrow \; \Sigma(P_1, P_2, \ldots, P_n) = -\sum_{i=1}^{n} P_i \ln P_i = -\langle \ln P \rangle.$$

Information comes in messages: $A_1, A_2, \ldots$. A message carries information only if it contains some news, i.e. something not completely expected.

$P(A)$: probability that message A is sent.
$I(A)$: information gain if message is indeed received.

The less likely the message, the greater the information gain if the message is received:

If $P(A) < P(B)$ then $I(A) > I(B)$, if $P(A) = 1$ then $I(A) = 0$.

If two independent messages are received, then the information gain is the sum of the information gains pertaining to each individual message:

$$P(AB) = P(A)P(B) \;\Rightarrow\; I(AB) = I(A) + I(B).$$

The information content of a message is equal to the change in (statistical) uncertainty at the receiver:

$$P_1, P_2, \ldots, P_n \xrightarrow{A} \bar{P}_1, \bar{P}_2, \ldots, \bar{P}_n \;\Rightarrow\; I(A) = \Sigma(P_1, P_2, \ldots, P_n) - \Sigma(\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_n)$$

Information as used here refers only to the scarcity of events. Any aspects of usefulness and meaningfulness are disregarded.