Some Handy Probability Facts

Mike Titelbaum and Alan Hájek

- Kolmogorov's **probability axioms** (sentential form):
  1. **Non-negativity**: For any sentence p, $P(p) \geq 0$.
  2. **Normality**: For any tautology T, $P(\mathsf{T}) = 1$.
  3. **Finite Additivity**: For any mutually exclusive p, q, $P(p \vee q) = P(p) + P(q)$.
- **Conditional probability**: For any p, q, if $P(p) > 0$ then

$$P(q \mid p) = \frac{P(p \,\&\, q)}{P(p)}$$

Consider all of the following to be universally quantified over $p, q, r, \ldots$, except that when a conditional probability expression appears we assume the sentence after the "$\mid$" has a positive unconditional probability.

- $P(\sim p) = 1 - P(p)$
- $P(p \vee q) = P(p) + P(q) - P(p \,\&\, q)$
- If $p \vdash q$ then $P(q) \geq P(p)$, $P(q \mid p) = 1$, and $P(p \supset q) = 1$.
- If $p \dashv\vdash q$ then $P(p) = P(q)$.
- $P(p \,\&\, q) = P(p) \cdot P(q \mid p)$
- For any p (such that $P(p) > 0$), $P(\cdot \mid p)$ is itself a probability function.
- $P(p \supset q) \geq P(q \mid p)$
- $P(p \supset q) = 1$ iff $P(q \mid p) = 1$.
- For any mutually exclusive p, q, if $P(r \mid p) = x$ and $P(r \mid q) = x$ then $P(r \mid p \vee q) = x$.
- **Substitution rules**: If $P(p \equiv q) = 1$, then p can be replaced by q anywhere in a true probability fact to yield another true probability fact. Corollary: If $p \dashv\vdash q$, then p can be replaced by q anywhere in a true probability fact to yield another true probability fact.
- **Law of Total Probability**: $P(p) = P(p \mid q) \cdot P(q) + P(p \mid \sim q) \cdot P(\sim q)$
- **Bayes' Theorem**:

$$P(p \mid q) = \frac{P(q \mid p) \cdot P(p)}{P(q)} = \frac{P(q \mid p) \cdot P(p)}{P(q \mid p) \cdot P(p) + P(q \mid \sim p) \cdot P(\sim p)}$$

  More generally, suppose there is a set of hypotheses $\{h_1, h_2, \ldots, h_n\}$ that is mutually exclusive and exhaustive (this is called a **partition**) and some evidence e. Then for each h_i,

$$P(h_i \mid e) = \frac{P(e \mid h_i) \cdot P(h_i)}{\sum_{j=1}^{n} P(e \mid h_j) \cdot P(h_j)}$$

  The $P(h_i)$ terms are called **priors**, the $P(h_i \mid e)$ terms are called **posteriors**, and the $P(e \mid h_i)$ terms are called **likelihoods**.
- The following are all equivalent:

$$\begin{aligned} P(p \mid q) &> P(p) \\ P(p \mid q) &> P(p \mid \sim q) \\ P(p) &> P(p \mid \sim q) \\ P(q \mid p) &> P(q \mid \sim p) \end{aligned}$$

  When these conditions hold, p and q are **positively correlated**. Replacing the inequalities with equals-signs yields equivalent conditions under which p and q are **probabilistically independent**.

- **Regularity** holds that all non-contradictory sentences have positive probability. Note that Regularity is an additional condition one might impose on probability functions; it doesn't follow from the axioms. Given the probability axioms, Regularity is equivalent to the converse of Normality (it says that anything with a probability of 1 is a tautology).

- **Updating by Conditionalization**: Taking the probability function P_0 and updating it by conditionalization on e generates a probability function P_1 such that

$$P_1(\cdot) = P_0(\cdot \mid e)$$

- **Jeffrey Conditionalization**: When the probability function P_0 is updated by Jeffrey Conditionalization on the elements of a partition $\{e_1, e_2, \ldots, e_n\}$, the resulting probability function P_1 satisfies

$$P_1(\cdot) = \sum_{i=1}^{n} P_0(\cdot \mid e_i) \cdot P_1(e_i)$$

- **Arithmetic Approach to Probabilities**: Given any sentential language, the state descriptions in that language will have non-negative probabilities that sum to 1. Specifying the probabilities of all state descriptions suffices to specify an entire probability distribution over sentences in the language. The probability of a sentence is the sum of the probabilities of all the state descriptions on which that sentence is true.

- **Expected Utilities**: On classical **(evidential) decision theory**, expected utility is calculated by

$$EU(p) = \sum_{i=1}^{n} u(p \,\&\, S_i) \cdot P(S_i \mid p)$$

  Here u is a utility function, $\{S_1, S_2, \ldots, S_n\}$ is a partition of possible states of the world, and p typically represents a proposition (though sometimes p is taken to represent an act).

  On **causal decision theory** the probability $P(S_i \mid p)$ is replaced by the probability of something related to causal influence. On the Gibbard-Harper theory, for instance, it is replaced by $P(p \mathrel{\Box\!\!\rightarrow} S_i)$, where $\Box\!\!\rightarrow$ is the counterfactual conditional.