Lecture 2

Cezar Ionescu

Probability theory

- The first question of probability theory is "probability of what?".
  - Richard von Mises (1883-1953): "The term probability is meaningful for us only with regard to a clearly defined collective[1] (or population." (*Mathematical Theory of Probability and Statistics*, page 17).
  - Sir Harold Jeffreys (1891-1989): "there is a valid primitive idea expressing the degree of confidence that we may reasonably have in a proposition" (*Theory of Probability, 3rd Ed.*, page 15).

- The currently accepted mathematical model was developed by Kolmogorov in 1933 and covers both interpretations.
  - Ingredients:
    * Ω, the set of all possible outcomes
    * `Event`, the set of all events $\text{Event} \subseteq \mathcal{P}(\Omega)$[2]
    * $p : \text{Event} \to [0, 1]$, the probability measure

- **Examples**:
  1. Drawing a card from a standard deck. Any of the 52 cards is a possible result, so $\Omega = \{2♣, 2♦, 2♥, 2♠, ...\}$. An event is what we want to assign a probability to, for example "the card drawn is a spade", or "a red card between 7 and 10". Any event can be represented by a subset of Ω; when Ω is finite, we take *all* subsets, so that $\text{Event} = \mathcal{P}(\Omega)$.
  2. Choosing a point of the unit disc. This can be seen as an idealised model for a game of darts. $\Omega = \{(x, y) \mid x, y \in \mathbb{R}, x^2 + y^2 \leq 1\}$. Events are of the form "the point will be chosen from this or that subset of the unit disc", so in principle we would like again to have $\text{Event} = \mathcal{P}(\Omega)$. It is an annoying but inescapable mathematical fact that we cannot, since that would make it impossible to define a good probability measure, so we have to limit the number of subsets that we can assign probabilities to. The mathematical choice is that of a *Borel σ-algebra*, which is "almost as big" as $\mathcal{P}(\Omega)$. We can be even less ambitious and choose `Event` to be the set of all *computable* subsets of Ω (i.e., all those whose characteristic function $X : \Omega \to \{0, 1\}$ can be implemented as a program).

[1] Intuitively, a *collective* is an infinite sequence of results of a repeated experiment.
[2] $\mathcal{P}(X)$ denotes the set of all subsets of X, including $\varnothing$ and X itself.