UNIVERSITY OF CAMBRIDGE
Department of Engineering

▶ Introduction

- Probability is the mathematics of uncertain events
- Statistics is the science of collecting and analysing data

▶ Axiomatic approach

- A sample space Ω is the set of all possible outcomes of a random experiment
- An event $\mathcal{A}$ is a subset of Ω
- A probability $\mathbb{P}$ is a measure $\mathbb{P} : \Omega \to \mathbb{R}$. It is subject to the following three axioms:
  - $\mathbb{P}[\mathcal{A}] \geq 0 \quad$ for all $\mathcal{A} \subseteq \Omega$
  - $\mathbb{P}[\Omega] = 1$
  - $\mathbb{P}[\mathcal{A} \cup \mathcal{B}] = \mathbb{P}[\mathcal{A}] + \mathbb{P}[\mathcal{B}] \quad$ for all $\mathcal{A}, \mathcal{B} \subset \Omega$ with $\mathcal{A} \cap \mathcal{B} = \varnothing$
- Consequences:
  - Monotonicity: if $\mathcal{A} \subseteq \mathcal{B}$ then $\mathbb{P}[\mathcal{A}] \leq \mathbb{P}[\mathcal{B}]$
  - Probability of the empty set: $\mathbb{P}[\varnothing] = 0$
  - Complement rule: $\mathbb{P}[\mathcal{A}^{\complement}] = 1 - \mathbb{P}[\mathcal{A}]$
  - Numeric bound: $0 \leq \mathbb{P}[\mathcal{A}] \leq 1$, for all $\mathcal{A} \subseteq \Omega$
  - Addition law: $\mathbb{P}[\mathcal{A} \cup \mathcal{B}] = \mathbb{P}[\mathcal{A}] + \mathbb{P}[\mathcal{B}] - \mathbb{P}[\mathcal{A} \cap \mathcal{B}]$
  - Sum rule: $\mathbb{P}[\mathcal{A} \cap \mathcal{B}] + \mathbb{P}[\mathcal{A} \cap \mathcal{B}^{\complement}] = \mathbb{P}[\mathcal{A}]$

▶ Conditional probability

- The probability of " $\mathcal{A}$ given $\mathcal{B}$ " is defined as $\mathbb{P}[\mathcal{A}|\mathcal{B}] = \dfrac{\mathbb{P}[\mathcal{A} \cap \mathcal{B}]}{\mathbb{P}[\mathcal{B}]}$
- Consequences:
  - Product rule: $\mathbb{P}[\mathcal{A} \cap \mathcal{B}] = \mathbb{P}[\mathcal{A}|\mathcal{B}]\,\mathbb{P}[\mathcal{B}]$
  - Law of Total Probability: $\mathbb{P}[\mathcal{A}] = \mathbb{P}[\mathcal{A}|\mathcal{B}]\,\mathbb{P}[\mathcal{B}] + \mathbb{P}[\mathcal{A}|\mathcal{B}^{\complement}]\,\mathbb{P}[\mathcal{B}^{\complement}]$
  - Bayes' rule: $\mathbb{P}[\mathcal{B}|\mathcal{A}] = \dfrac{\mathbb{P}[\mathcal{A}|\mathcal{B}]\,\mathbb{P}[\mathcal{B}]}{\mathbb{P}[\mathcal{A}]} = \dfrac{\mathbb{P}[\mathcal{A}|\mathcal{B}]\,\mathbb{P}[\mathcal{B}]}{\mathbb{P}[\mathcal{A}|\mathcal{B}]\,\mathbb{P}[\mathcal{B}] + \mathbb{P}[\mathcal{A}|\mathcal{B}^{\complement}]\,\mathbb{P}[\mathcal{B}^{\complement}]}$
- Independence: $\mathcal{A}$ and $\mathcal{B}$ independent iff $\mathbb{P}[\mathcal{A} \cap \mathcal{B}] = \mathbb{P}[\mathcal{A}]\,\mathbb{P}[\mathcal{B}]$ (hence $\mathbb{P}[\mathcal{A}|\mathcal{B}] = \mathbb{P}[\mathcal{A}]$)

▶ Discrete random variables

- A discrete random variable X takes values from a discrete set $\mathbb{X}$ (called the support of X)
- The probability mass function (PMF), $P_{\mathrm{X}} : \mathbb{X} \to [0, 1]$, is defined as $P_{\mathrm{X}}(x) = \mathbb{P}[\mathrm{X} = x]$
- The cumulative distribution function (CDF) is $F_{\mathrm{X}}(x) = \mathbb{P}[\mathrm{X} \leq x]$. It has the following properties:
  - $F_{\mathrm{X}}(a) \leq F_{\mathrm{X}}(b)$ if $a \leq b$
  - $\lim_{x \to -\infty} F_{\mathrm{X}}(x) = 0$ and $\lim_{x \to \infty} F_{\mathrm{X}}(x) = 1$
  - $F_{\mathrm{X}}(b) - F_{\mathrm{X}}(a) = \mathbb{P}[a < \mathrm{X} \leq b]$
- The joint PMF, $P_{\mathrm{XY}} : \mathbb{X} \times \mathbb{Y} \to [0, 1]$, is defined as $P_{\mathrm{XY}}(x, y) = \mathbb{P}[\mathrm{X} = x \cap \mathrm{Y} = y]$, and:
  - The conditional PMF is $P_{\mathrm{X|Y}}(x|y) = \dfrac{P_{\mathrm{XY}}(x, y)}{P_{\mathrm{Y}}(y)}$, and the product rule $P_{\mathrm{XY}}(x, y) = P_{\mathrm{X|Y}}(x|y)\,P_{\mathrm{Y}}(y)$
  - Marginalisation: $P_{\mathrm{X}}(x) = \sum_{y \in \mathbb{Y}} P_{\mathrm{XY}}(x, y)$
  - Bayes' rule: $P_{\mathrm{Y|X}}(y|x) = \dfrac{P_{\mathrm{X|Y}}(x|y) P_{\mathrm{Y}}(y)}{P_{\mathrm{X}}(x)} = \dfrac{P_{\mathrm{X|Y}}(x|y) P_{\mathrm{Y}}(y)}{\sum_{\xi \in \mathbb{Y}} P_{\mathrm{X|Y}}(x|\xi) P_{\mathrm{Y}}(\xi)}$
- Independence: X and Y independent iff $P_{\mathrm{XY}}(x, y) = P_{\mathrm{X}}(x) P_{\mathrm{Y}}(y)$ for all $x, y \in \mathbb{X} \times \mathbb{Y}$ (hence $P_{\mathrm{X|Y}}(x|y) = P_{\mathrm{X}}(x)$)

▶ Expectation and Entropy

- The expectation is defined as $\mathbb{E}[g(\mathrm{X})] = \sum_{x \in \mathbb{X}} g(x) P_{\mathrm{X}}(x)$. In particular, $\mathbb{E}[\mathrm{X}] = \sum_{x \in \mathbb{X}} x\, P_{\mathrm{X}}(x)$
- The expectation is linear: $\mathbb{E}[a\mathrm{X} + b\mathrm{Y}] = a\,\mathbb{E}[\mathrm{X}] + b\,\mathbb{E}[\mathrm{Y}]$ for all $a, b \in \mathbb{R} \times \mathbb{R}$
- For two independent variables X and Y, $\mathbb{E}[\mathrm{X}\,\mathrm{Y}] = \mathbb{E}[\mathrm{X}]\,\mathbb{E}[\mathrm{Y}]$
- The variance is defined as $\mathrm{Var}[\mathrm{X}] = \mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])^2] = \mathbb{E}[\mathrm{X}^2] - \mathbb{E}[\mathrm{X}]^2$
- The entropy is defined as $\mathbb{H}[\mathrm{X}] = \mathbb{E}[-\log_2 P_{\mathrm{X}}(\mathrm{X})] = -\sum_{x \in \mathbb{X}} P_{\mathrm{X}}(x) \log_2 P_{\mathrm{X}}(x)$