Lecture 16, Nov 1, 2023

Probability Review

- Example: a man has M pairs of pants and L shirts; over time we observe that out of N observations, $n_{ps}(i, j)$ is the number of times he wore pants i with shirt j, $_p$ is the number of pants he wore pants i, and $n_s(j)$ is the number of times he wore shirt j
  - Let $f_{ps}(i, j) = \dfrac{n_{ps}(i, j)}{N}$ be the likelihood of wearing pants i with shirt j
  - Let $f_p(i) = \dfrac{n_p(i)}{N}$, $f_s(i) = \dfrac{n_s(j)}{N}$
  - Note that all the frequencies/likelihoods are nonnegative, and summing over all i or j gets us 1
  - $n_p(i) = \sum_{j=1}^{L} n_{ps}(i, j), n_s(j) = \sum_{i=1}^{M} n_{ps}(i, j)$
  - Therefore $f_p(i) = \sum_{j=1}^{L} f_{ps}(i, j), f_s(j) = \sum_{i=1}^{M} f_{ps}(i, j)$
    - This is known as *marginalization* or the *sum rule*
- There are two main ways of thinking about probability: *frequentist*, which examines the relative frequency of events occurring in a large number of trials; or *Bayesian*, which thinks about probability in terms of belief and uncertainty
- Let $\mathcal{X}$ be the set of all possible outcomes; let $f_x(\cdot)$ be the *probability density function* (PDF), a real-valued function that is nonnegative and sums to 1
  - $f_x(\cdot)$ and $\mathcal{X}$ define a discrete random variable (DRV) x
  - The PDF defines the notion of probability: $\Pr(x = \bar{x}) = f_x(\bar{x})$
  - We typically use x to denote the variable and a value it can take, e.g. $\Pr(x) = f_x(x)$; we will often also drop the subscript
- We can form a joint PDF $f_{xy}(x, y)$ over two variables
  - $f_{xy}(x, y) \geq 0$ as usual
  - $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f_{xy}(x, y) = 1$
  - *Marginalization* (*sum rule*): $f_x(x) = \sum_{y \in \mathcal{Y}} f_{xy}(x, y)$
    - f_x is fully defined by f_{xy}; consider this as a definition
  - *Conditioning* (*product rule*): $f_{x|y}(x, y) = f(x|y) = \dfrac{f_{xy}(x, y)}{f_y(y)}$
    - This is like taking a slice of the joint probability distribution, and normalizing so it becomes a proper distribution
    - "Probability of y given x"
    - $f(x, y) = f(x|y)f(y) = f(y|x)f(x)$ is known as *Bayes' rule*

Theorem

Total probability theorem:

$$f_x(x) = \sum_{y \in \mathcal{Y}} f_{x|y}(x|y) f_y(y)$$

- Continuous random variables have a continuous range of values and must be integrated instead of summed
  - Note that we will assume $f_x(x)$ is uniformly bounded and piecewise continuous; this excludes things like delta functions and distributions that go to infinity
  - For a CRV it makes no sense to talk about the probability of x being exactly some value
  - Instead we use intervals $\Pr(x \in [a, b]) = \int_a^b f_x(x) \, \mathrm{d}x$

  - All other properties for discrete random variables apply, with integrals instead of sums
- We can mix continuous and discrete random variables in a joint probability distribution