ECE 286: Bayesian Machine Perception

Class 1: Probability Theory

Florian Meyer

Electrical and Computer Engineering Department
University of California San Diego

Event Spaces (i)

- *Possible outcomes* Ω
  - Example: If we consider a dice roll, we have $\Omega = \{1, 2, 3, 4, 5, 6\}$

- *Events* $\mathcal{S}$ to which we want to assign probabilities
  - Each event $\alpha \in \mathcal{S}$ is a subset of Ω
  - Examples: The event $\alpha = \{1\}$ represents the case where the die shows 1

    The event $\alpha = \{2, 4, 6\}$ represents the case of an even outcome

Event Spaces (ii)

- An *event space* $\mathcal{S}$ ***needs to*** satisfy the three basic properties
  - It contains the empty event $\emptyset$, and the trivial event Ω
  - It is closed under union, i.e., if $\alpha, \beta \in \mathcal{S}$, then so is $\alpha \cup \beta$
  - It is closed under complementation, i.e., if $\alpha \in \mathcal{S}$, then so is $\Omega \backslash \alpha$

Axioms of Probability Theory

- A probability distribution P is a mapping from events in $\mathcal{S}$ to real values that satisfies
  - $P(\alpha) \geqslant 0$ for all $\alpha \in \mathcal{S}$
    (Probabilities are not negative)
  - $P(\Omega) = 1$
    (All possible outcomes have the maximal probability of one)
  - If $\alpha, \beta \in \mathcal{S}$ and $\alpha \cap \beta = \emptyset$, then $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$
    (The probability of two disjoint is the sum of their probabilities)

Properties (i)

- Monotonicity: if $\alpha \subseteq \beta$ then $P(\alpha) \leqslant P(\beta)$

  Proof:

- The probability of the empty set: $P(\emptyset) = 0$

  Proof:

Properties (ii)

- The complement rule: $P(\Omega \backslash \alpha) = 1 - P(\alpha)$

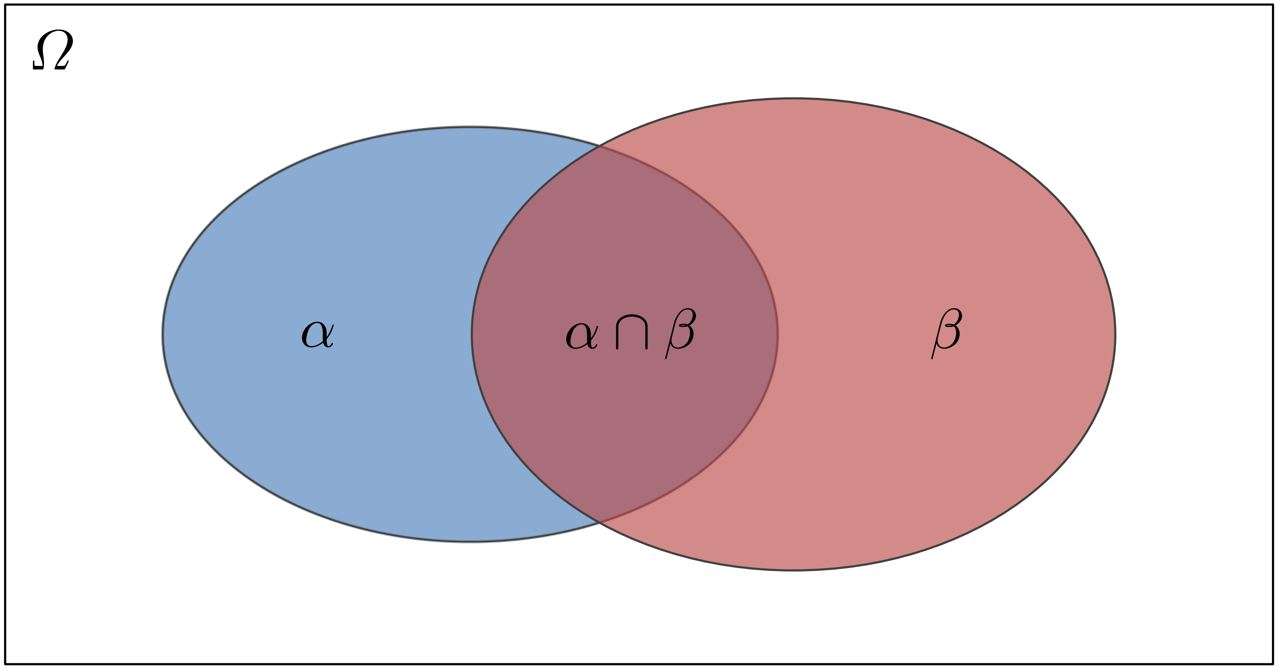
  Proof:

- The sum rule: $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$

  Proof:

A Closer Look at the Sum Rule

$$P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$$



Discrete Random Variables

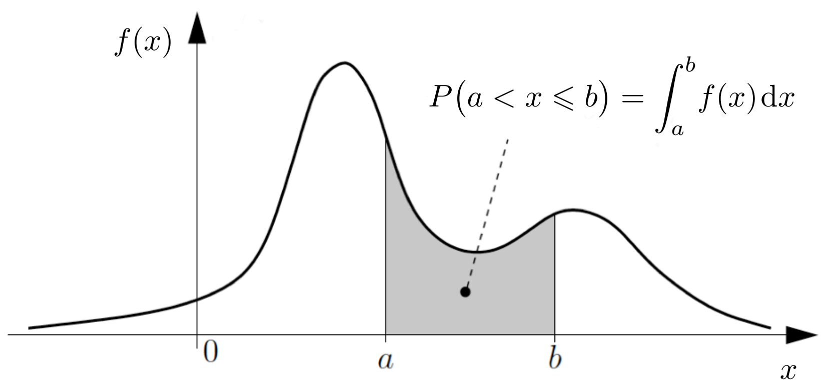
- $\boldsymbol{x}$ denotes a *random variable*
- $\boldsymbol{x}$ can take on a countable number of values in $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_I\}$
- $p_{\boldsymbol{x}}(\boldsymbol{x}_i)$, or $p(\boldsymbol{x}_i)$, is the *probability* that the random variable $\boldsymbol{x}$ takes on value $\boldsymbol{x}_i$
- $p(\cdot)$ is called *probability mass function (pmf)*
- Example: If $\boldsymbol{x}$ is the outcome of a dice roll, we have $\mathcal{X} = \{1, 2, \ldots, 6\}$ and
  $$p(\boldsymbol{x}_i) = 1/6, \forall \boldsymbol{x}_i \in \mathcal{X}$$

Continuous Random Variables

- $\boldsymbol{x}$ takes on values in the continuum
- $f_{\boldsymbol{x}}(\boldsymbol{x})$, or $f(\boldsymbol{x})$, is its *probability density function (pdf)*

$$P(a < x \leqslant b) = \int_a^b f(x)\,\mathrm{d}x$$

- Example:



Joint and Conditional Distributions

- $p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x},\boldsymbol{y})$ or $p(\boldsymbol{x},\boldsymbol{y})$ is the joint pmf of random variables $\boldsymbol{x}$ and $\boldsymbol{y}$
- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are *independent* then

$$p(\boldsymbol{x},\boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y})$$

- $p(\boldsymbol{x}|\boldsymbol{y})$ is the probability of $\boldsymbol{x}$ given (conditioned on) $\boldsymbol{y}$

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x},\boldsymbol{y})/p(\boldsymbol{y}) \qquad p(\boldsymbol{x},\boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})$$

- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are *independent* then

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{x}) \qquad p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y})$$

- Equivalent expressions exist for the pdfs of continuous random variables

Law of Total Probabilities, Marginals

Discrete Case

$$\sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) = 1$$

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\mathcal{Y}} p(\boldsymbol{x},\boldsymbol{y})$$

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\mathcal{Y}} p(\boldsymbol{x}|\boldsymbol{y})\, p(\boldsymbol{y})$$

Continuous Case

$$\int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$$

$$f(\boldsymbol{x}) = \int f(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

$$f(\boldsymbol{x}) = \int f(\boldsymbol{x}|\boldsymbol{y})\, f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

Bayes Rule

- Recall $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})\,p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x})$

- It therefore follows that

likelihood ($\boldsymbol{y}$ is fixed) prior

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{\boxed{p(\boldsymbol{y}|\boldsymbol{x})}\,\boxed{p(\boldsymbol{x})}}{\boxed{p(\boldsymbol{y})}}$$

evidence

Normalization

- For $\boldsymbol{y}$ observed and thus fixed

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

$$= C\, p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})$$

$$\propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})$$

- The constant C ensures that $p(\boldsymbol{x}|\boldsymbol{y})$ sums to one and can be calculated as

$$C = \frac{1}{\sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})}$$

Conditioning

- Law of total probability

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{z}) &= \int p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z})\,\mathrm{d}\boldsymbol{y} \\ &= \int p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{z})\,p(\boldsymbol{y}|\boldsymbol{z})\,\mathrm{d}\boldsymbol{y} \end{aligned}$$

- Bayes rule with background knowledge

$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{z}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})\,p(\boldsymbol{x}|\boldsymbol{z})}{p(\boldsymbol{y}|\boldsymbol{z})}$$

Conditional Independence

- Condition on z, random variables x and y are independent if

$$p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}) = p(\boldsymbol{y}|\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z})$$

- This is equivalent to

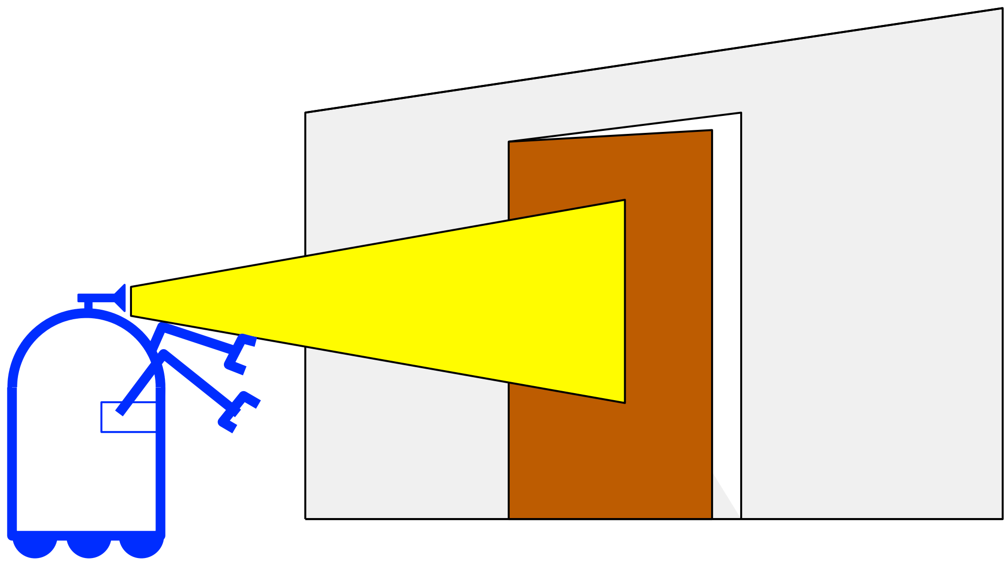
$$p(\boldsymbol{x}|\boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})$$

and

$$p(\boldsymbol{y}|\boldsymbol{z}) = p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x})$$

Simple Example of State Estimation (i)

- Suppose a robot obtains a measurement z of a door
- What is $p(open|z)$?



S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, 2006.

Causal vs. Diagnostic Reasoning

- $p(open|\boldsymbol{z})$ is diagnostic
- $\boxed{p(\boldsymbol{z}|open)}$ is causal
- Often causal knowledge is easier to obtain, e.g., sensor calibration
- Bayesian rule allows up to use causal knowledge:

count frequencies!

$$p(open|\boldsymbol{z}) = \frac{\boxed{p(\boldsymbol{z}|open)}\, p(open)}{p(\boldsymbol{z})}$$

Simple Example of State Estimation (ii)

- Likelihood: $p(\boldsymbol{z}|open) = 0.7, \quad p(\boldsymbol{z}|notopen) = 0.1$
- Prior: $p(open) = p(notopen) = 0.5$

$$
\begin{aligned}
p(open|\boldsymbol{z}) &= \frac{p(\boldsymbol{z}|open)\, p(open)}{p(\boldsymbol{z}|open)\, p(open) + p(\boldsymbol{z}|notopen)\, p(notopen)} \\
&= \frac{0.7 \cdot 0.5}{0.7 \cdot 0.5 + 0.1 \cdot 0.5} \\
&= 0.875
\end{aligned}
$$

- Observation z raises the probability that the door is open