Primer on Bayesian inference

ST5 Autonomous robotics

Francis Colas

2022-09-12

Introduction

Autonomous robots

- anatomy
- functions

Introduction

Autonomous robots

- anatomy
- functions

Models

- formalization of expected behavior of
  - sensors, actuators
  - environment (including others in interaction)
- not fully accurate

Introduction

Autonomous robots

- anatomy
- functions

Models

- formalization of expected behavior of
  - sensors, actuators
  - environment (including others in interaction)
- not fully accurate

Aim of this session

- reasoning
- reminders on probabilities
- Bayesian inference
- Bayesian modeling

01

Reasoning

Reasoning

Definition (Merriam-Webster)

- *reasoning*

Reasoning

Definition (Merriam-Webster)

- *reasoning*: the use of *reason*

Reasoning

Definition (Merriam-Webster)

- *reasoning*: the use of *reason*
- *reason*:
  - sanity
  - proper exercise of the mind
  - the power of comprehending, inferring, or thinking

Reasoning

Definition (Merriam-Webster)

- *reasoning*: the use of *reason*
- *reason*:
  - sanity
  - proper exercise of the mind
  - the power of comprehending, inferring, or thinking

Definition (Wikipedia)

- *reasoning*: applying logic to *seek truth* and *draw conclusions* from new or existing information

Reasoning

Definition (Merriam-Webster)

- *reasoning*: the use of *reason*
- *reason*:
  - sanity
  - proper exercise of the mind
  - the power of comprehending, inferring, or thinking

Definition (Wikipedia)

- *reasoning*: applying logic to *seek truth* and *draw conclusions* from new or existing information

Use in robotics

- process sensor information
- know what is going on

Reasoning

Forms of reasoning

- deduction: go from premises to conclusion
- induction: go from cases to generalization
- abduction/retroduction: find out more likely causes of a given effect
- analogical: go from cases to cases by similarity

Reasoning

Forms of reasoning

- deduction: go from premises to conclusion
- induction: go from cases to generalization
- abduction/retroduction: find out more likely causes of a given effect
- analogical: go from cases to cases by similarity
- fallacy: (self-)deception by wrong reasoning

Reasoning

Forms of reasoning

- deduction: go from premises to conclusion
- induction: go from cases to generalization
- abduction/retroduction: find out more likely causes of a given effect
- analogical: go from cases to cases by similarity
- fallacy: (self-)deception by wrong reasoning

Formalization

Reasoning

Forms of reasoning

- deduction: go from premises to conclusion
- induction: go from cases to generalization
- abduction/retroduction: find out more likely causes of a given effect
- analogical: go from cases to cases by similarity
- fallacy: (self-)deception by wrong reasoning

Formalization

- logic: Aristotle, Frege, Hilbert, Gödel, etc.

Reasoning

Forms of reasoning

- deduction: go from premises to conclusion
- induction: go from cases to generalization
- abduction/retroduction: find out more likely causes of a given effect
- analogical: go from cases to cases by similarity
- fallacy: (self-)deception by wrong reasoning

Formalization

- logic: Aristotle, Frege, Hilbert, Gödel, etc.

Issue

- imperfect models ($\sim$ theorems)
- imperfect knowledge ($\sim$ axioms)

$\Rightarrow$ truth-value replaced by plausibility/belief

Cox theorem

Cox theorem (1946)

- plausibility of a proposition as a real number
- common sense reasoning and consistency

$\Rightarrow$ plausibility can be mapped to probability

$\Rightarrow$ reasoning is probability calculus

Cox theorem

Cox theorem (1946)

- plausibility of a proposition as a real number
- common sense reasoning and consistency

$\Rightarrow$ plausibility can be mapped to probability

$\Rightarrow$ reasoning is probability calculus

Bayesian probability theory (E.T. Jaynes)

- formal system of logic under uncertainty
- computing beliefs (state of knowledge)
- using probability computation

Cox theorem

Cox theorem (1946)

- plausibility of a proposition as a real number
- common sense reasoning and consistency

$\Rightarrow$ plausibility can be mapped to probability

$\Rightarrow$ reasoning is probability calculus

Bayesian probability theory (E.T. Jaynes)

- formal system of logic under uncertainty
- computing beliefs (state of knowledge)
- using probability computation

Different from the *frequentist* interpretation of probabilities

02

Reminders on probabilities

Probability values

Probabilities for propositions

- number between 0 and 1

(Random) Variable

- variable we don't know the value of
- values in a given finite domain $\mathcal{D}$: finite set of integers, categories...

Probability distribution

- distribution over the different possible values in $\mathcal{D}$

$$\begin{cases} \mathcal{D}_C \to (0,1) \\ c \mapsto P([C=c]) = \begin{cases} 0.7 & \text{if } c = 8 \\ 0.3 & \text{if } c = 10 \end{cases} \end{cases}$$

informatics mathematics Inria

Probability for continuous variables

Continuous variables

- continuous domain: $(0, 1)$, $\mathbb{R}^+$, $\mathbb{R}$, $\mathbb{R}^n$...

Cumulative probability

- probability for intervals: number between 0 and 1

$$P([D < 5\,\mathsf{min}]) = 0.9$$

- cumulative density function (cdf)

$$\begin{cases} \mathcal{D}_D \to (0, 1) \\ \quad d \mapsto P([D < d]) \end{cases}$$

Probability density function

Probability density function

- derivative of cumulative probability
- can be higher than 1
- e.g. Cauchy distribution:

$$\text{cdf:} \begin{cases} \mathbb{R} \to (0,1) \\ x \mapsto P([X < x]) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2} \end{cases}$$

$$\text{pdf:} \begin{cases} \mathbb{R} \to \mathbb{R}^+ \\ x \mapsto p([X = x]) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} \end{cases}$$

Combinations of variables

Conjunction

- probability of both variables: $P([A = \text{red}], [C = 8]) = 0.36$
- still a variable: $\mathcal{D}_{A,B} = \mathcal{D}_A \times \mathcal{D}_B$
- commutative: $A, B = B, A$

Conditional probability

- assumption on the value of a variable: $P([C = 8] \mid [A = \text{red}]) = 0.9$
- collection of probability distributions

$$P(C \mid A) : \begin{cases} \mathcal{D}_A \times \mathcal{D}_C \to (0,1) \\ \quad\quad a, c \mapsto P([C = c] \mid [A = a]) \end{cases}$$

$$\equiv \begin{cases} \mathcal{D}_A \to [\quad \mathcal{D}_C \to (0,1) \quad] \\ \quad a \mapsto [\quad c \mapsto P([C = c] \mid [A = a]) \quad] \end{cases}$$

informatics mathematics Inria

Probabilities

Notations

- $Pr(A)$
- $P(A)$
- $p(A)$
- $P([A = a])$
- $p(a)$
- $\pi(a)$
- need to be careful:
  - discrete probability/density/cumulative probability
  - variable/value

Continuous/discrete

- unified with measure theory
- still need caution:
  - densities are not probabilities
  - densities can be bigger than 1
  - densities also follow probability operations

03

Bayesian inference

Normalization

Probability distributions

- sum to 1 (law of total probability)

$$\sum_{a \in \mathcal{D}_A} P([A = a]) = 1 \qquad \int_{\mathcal{D}_A} p(a)da = 1 \qquad \sum_{A} p(A) = 1$$

Normalization

Probability distributions

- sum to 1 (law of total probability)

$$\sum_{a \in \mathcal{D}_A} P([A = a]) = 1 \qquad \int_{\mathcal{D}_A} p(a)da = 1 \qquad \sum_{A} p(A) = 1$$

Conditional probability distributions

- don't sum to 1

$$\sum_{B} P(A \mid B) \neq 1 \qquad \sum_{A,B} P(A \mid B) \neq 1 \quad \text{in general}$$

Normalization

Probability distributions

- sum to 1 (law of total probability)

$$\sum_{a \in \mathcal{D}_A} P([A = a]) = 1 \qquad \int_{\mathcal{D}_A} p(a)da = 1 \qquad \sum_{A} p(A) = 1$$

Conditional probability distributions

- don't sum to 1

$$\sum_{B} P(A \mid B) \neq 1 \qquad \sum_{A,B} P(A \mid B) \neq 1 \quad \text{in general}$$

- but of course

$$\forall b \in \mathcal{D}_B, \sum_{a \in \mathcal{D}_A} P([A = a] \mid [B = b]) = 1 \qquad \sum_{A} p(A \mid B) = 1$$

Marginalization rule

Marginalization rule

- “sum rule”
- simple consequence of normalization of distributions

$$\sum_A p(A, B) = p(B)$$

Marginalization rule

Marginalization rule

- "sum rule"
- simple consequence of normalization of distributions

$$\sum_{A} p(A, B) = p(B)$$

Validity

- continuous or discrete variable
- probabilities or distributions
- conditional distributions
- variable conjunctions

$$\sum_{A,B} p(A, B, C \mid D) = p(C \mid D)$$

Bayes rule

Bayes rule

- "product rule"
- several ways to state it
  - $p(B \mid A) = \frac{p(A|B)p(B)}{p(A)}$
  - $p(A, B) = p(A \mid B)p(B) = p(B \mid A)p(A)$
  - $p(A \mid B) = \frac{p(A,B)}{p(B)}$

Bayes rule

Bayes rule

- “product rule”
- several ways to state it
  - $p(B \mid A) = \frac{p(A|B)p(B)}{p(A)}$
  - $p(A, B) = p(A \mid B)p(B) = p(B \mid A)p(A)$
  - $p(A \mid B) = \frac{p(A,B)}{p(B)}$

Validity

- continuous or discrete variables
- probabilities or distributions
- conditional distributions
- variable conjunctions

$$p(A, B, C \mid D) = p(A \mid B, C, D)p(B, C \mid D)$$

Bayesian inference

Bayesian inference

- computation of some (conditional) probability distribution
- based on a factorization of the joint probability distribution
- generic

$$p(A \mid C) = \frac{\sum_B p(A, B, C)}{\sum_{A,B} p(A, B, C)}$$

Implementation challenges

- computation of sums or integrals
- ordering of sums and product

Example: disease test

Variables

- infected or not: i or $\neg i$
- positive test or not: t or $\neg t$

Parameters

- sensibility ($p(t \mid i)$): 96 %
- specificity ($p(\neg t \mid \neg i)$): 99.2 %
- circulation: $p(i) = 0.001$

Inference example (1/2)

Probability of positive test:

$$\begin{aligned}
p(t) &= \sum_i p(t, i) && \text{marginalization} \\
&= \sum_i p(t \mid i) p(i) && \text{Bayes rule} \\
&= p(t \mid \neg i) p(\neg i) + p(t \mid i) p(i) && \text{standard arithmetic} \\
&= (1 - p(\neg t \mid \neg i)) p(\neg i) + p(t \mid i) p(i) && \text{marginalization} \\
&= 0.008 * 0.999 + 0.96 * 0.001 && \text{model} \\
p(t) &= 0.008952
\end{aligned}$$

Inference example (2/2)

Probability of being infected with positive test result:

$$p(i \mid t) = \frac{p(t, i)}{p(t)} \quad \text{Bayes rule}$$
$$= \frac{p(t \mid i)p(i)}{p(t)} \quad \text{Bayes rule}$$
$$= \frac{0.96 * 0.001}{0.008952} \quad \text{model and computation above}$$
$$\approx 0.107$$

Probability of not being infected with negative test result:

$$p(\neg i \mid \neg t) \approx 0.99996$$

04

Bayesian modeling

Modeling

Modeling

- expressing knowledge
- deduce properties

Different paradigms

- equations: classical, differential…
- logic: propositional, first-order…
- probabilities, fuzzy logic…

Bayesian modeling

- uncertain and incomplete knowledge
- plausibility as probability (Cox theorem)
- "subjectivist" use of probabilities

Bayesian modeling

Bayesian modeling

- specify model
- state assumptions

Model

- define variables
- specify joint distribution

Joint distribution

- “exponential” with number of variables
- factorization using independence assumptions

Independence assumptions

Independence

- A independent from B (noted $A \perp B$)
- iff $p(A, B) = p(A)p(B)$

Example

- $\mathcal{D}_A = 1, \ldots, N$ and $\mathcal{D}_B = 1, \ldots, M$
- $p(A, B)$: $N \times M - 1 = 99$ degrees of freedom
- if $A \perp B$, $p(A, B) = p(A)p(B)$: $(N - 1) + (M - 1) = 18$ dof

Conditional independence

- A independent from B conditionally to C ($A \perp B \mid C$)
- iff $p(A, B \mid C) = p(A \mid C)p(B \mid C)$

Independence assumptions

Independence

- A independent from B (noted $A \perp B$)
- iff $p(A, B) = p(A)p(B)$

Example

- $\mathcal{D}_A = 1, \ldots, N$ and $\mathcal{D}_B = 1, \ldots, M$
- $p(A, B)$: $N \times M - 1 = 99$ degrees of freedom
- if $A \perp B$, $p(A, B) = p(A)p(B)$: $(N - 1) + (M - 1) = 18$ dof

Conditional independence

- A independent from B conditionally to C ($A \perp B \mid C$)
- iff $p(A, B \mid C) = p(A \mid C)p(B \mid C)$
- Warning: $A \perp B \mid C \not\Rightarrow A \perp B$

Independence assumptions

Independence

- A independent from B (noted $A \perp B$)
- iff $p(A, B) = p(A)p(B)$

Example

- $\mathcal{D}_A = 1, \ldots, N$ and $\mathcal{D}_B = 1, \ldots, M$
- $p(A, B)$: $N \times M - 1 = 99$ degrees of freedom
- if $A \perp B$, $p(A, B) = p(A)p(B)$: $(N - 1) + (M - 1) = 18$ dof

Conditional independence

- A independent from B conditionally to C ($A \perp B \mid C$)
- iff $p(A, B \mid C) = p(A \mid C)p(B \mid C)$
- Warning: $A \perp B \mid C \not\Longrightarrow A \perp B$
- Warning: $A \perp B \not\Longrightarrow A \perp B \mid C$

Factorization of the joint

Factorization of the joint

- $p(A, B, C)$
- Bayes rule (always true)

$$p(A, B, C) = p(A)p(B \mid A)p(C \mid A, B)$$

Factorization of the joint

Factorization of the joint

- $p(A, B, C)$
- Bayes rule (always true)

$$p(A, B, C) = p(A)p(B \mid A)p(C \mid A, B)$$

- assuming $A \perp B$:

$$p(A, B, C) = p(A)p(B)p(C \mid A, B)$$

Factorization of the joint

Factorization of the joint

- $p(A, B, C)$
- Bayes rule (always true)

$$p(A, B, C) = p(A)p(B \mid A)p(C \mid A, B)$$

- assuming $A \perp B$:

$$p(A, B, C) = p(A)p(B)p(C \mid A, B)$$

- assuming only $C \perp B \mid A$:

$$p(A, B, C) = p(A)p(B \mid A)p(C \mid A)$$

Bayesian networks

Bayesian networks

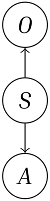
- graphical representation of dependencies
- nodes are variables
- (lack of) edges are [conditional] (in)dependence assumptions

Bayesian networks

Bayesian networks

- graphical representation of dependencies
- nodes are variables
- (lack of) edges are [conditional] (in)dependence assumptions

Example



$$p(O, S, A) = P(S)P(O \mid S)P(A \mid S)$$

Distributions

Discrete distribution

- probability value for each variable value
- parameters: table
- specific case: uniform

Probability distributions on continuous variables

- expression of probability density function
- depends on the domain

Continuous uniform

- only on a bounded interval (a, b)
- no parameter

$$p(x) = \frac{1}{b - a}$$

Examples of Continuous distributions

Gaussian

- only on $\mathbb{R}$
- parameters: mean μ and standard deviation σ

$$p(x) = \mathcal{N}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

multivariate Gaussian

- on real vectors in $\mathbb{R}^d$
- parameters: mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix[a] $\Sigma \in \mathbb{R}^{d \times d}$

$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{d}{2}} \det \Sigma^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\intercal \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

[a] symmetric positive semi-definite

Conditional probability distributions

Conditional probability distributions

- same choice of distributions
- conditional: collection of distributions
- parameters as functions of conditioning variable

Discrete example

$$p(A \mid B): \quad \begin{array}{cccc} & B=0 & B=1 & B=2 \\ A=0 & 0.4 & 0.5 & 0.1 \\ A=1 & 0.6 & 0.5 & 0.9 \end{array}$$

Continuous example

$$p(x \mid y) = \mathcal{N}\left(x; \mu_x(y), \sigma_x(y)\right)$$

05

Conclusion

Probabilities

Probabilities

- represent state of knowledge
- ambiguous notations

Bayesian inference

- Bayes and marginalization rules
- mechanical

Bayesian modeling

- specify actual (conditional) distributions
- specify (conditional) independence assumptions
- modeler's choice and responsibility

Thanks for your attention

Questions?