Bayesian inference

Marcel Lüthi

Graphics and Vision Research Group
Department of Mathematics and Computer Science
University of Basel

Probabilities: What are they?

Four possible interpretations:

1. Long-term frequencies
   - Relative frequency of an event over time
2. Physical tendencies (propensities)
   - Arguments about a physical situation (causes of relative frequencies)
3. Degree of belief (Bayesian probabilities)
   - Subjective beliefs about events/hypothesis/facts
4. Logic
   - Degree of logical support for a particular hypothesis

Degree of belief: An Example

- Dentist example: *Does the patient have a cavity?*

$P(\text{cavity}) = 0.1$

$P(\text{cavity}|\text{toothache}) = 0.8$

$P(\text{cavity}|\text{toothache, gum problems}) = 0.4$

But the patient either has a cavity or does not

- *There is no 80% cavity!*
- *Having a cavity should not depend on whether the patient has a toothache or gum problems*

These statements do not contradict each other, they summarize ***the dentist's knowledge*** *about the patient*

AIMA: Russell & Norvig, *Artificial Intelligence. A Modern Approach,* 3rd edition,

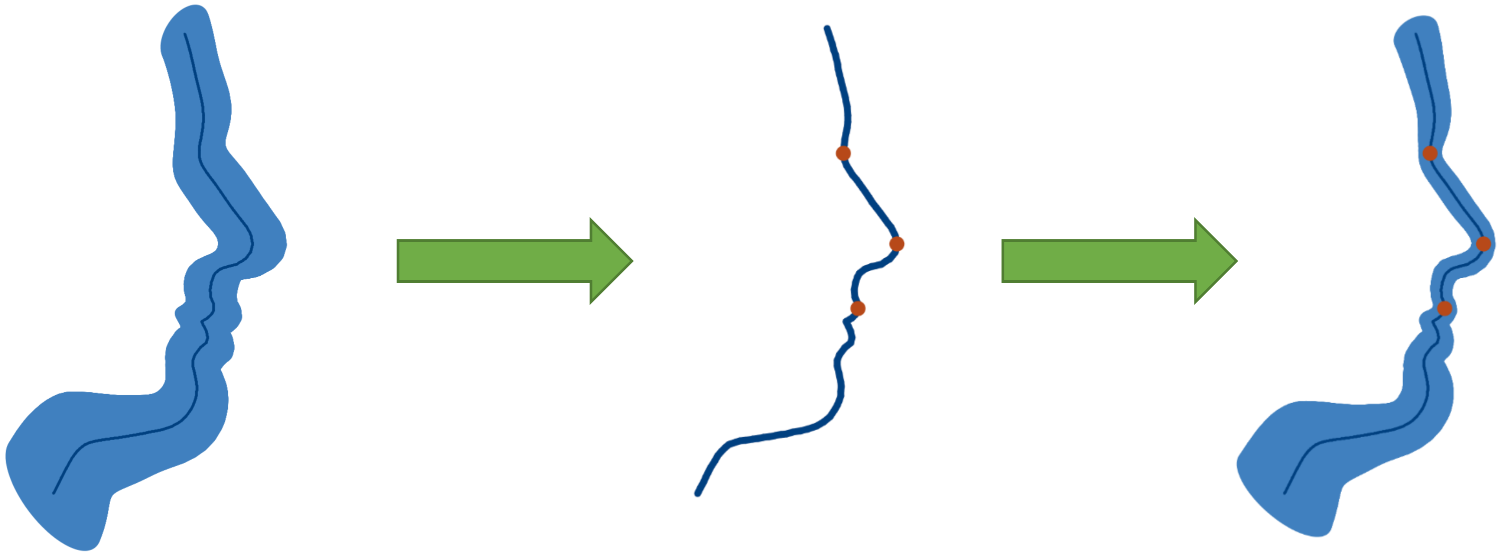
Uncertainty: Bayesian Probability

- Bayesian probabilities rely on a *subjective* perspective:
  - Probabilities express our *current knowledge*.
  - Can *change* when we learn or see more
  - More data -> more *certain* about our result.

Subjectivity: There is no single, real underlying distribution. A probability distribution expresses our knowledge – It is different in different situations and for different observers since they have different knowledge.

  - Subjective != Arbitrary
  - Given belief, conclusions follow by laws of probability calculus

Belief Updates



Model
Face distribution

Prior belief

Observation
Concrete points
Possibly uncertain

More knowledge

Posterior
Face distribution
consistent with observation

Posterior belief

Consistency: Laws of probability calculus!

Two important rules

Probabilistic model: joint distribution of points

$$P(x_1, x_2)$$

Marginal

Distribution of certain points only

$$P(x_1) = \sum_{x_2} P(x_1, x_2)$$

Conditional

Distribution of points conditioned on *known* values of others

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)}$$

Product rule: $P(x_1, x_2) = p(x_1|x_2)p(x_2)$

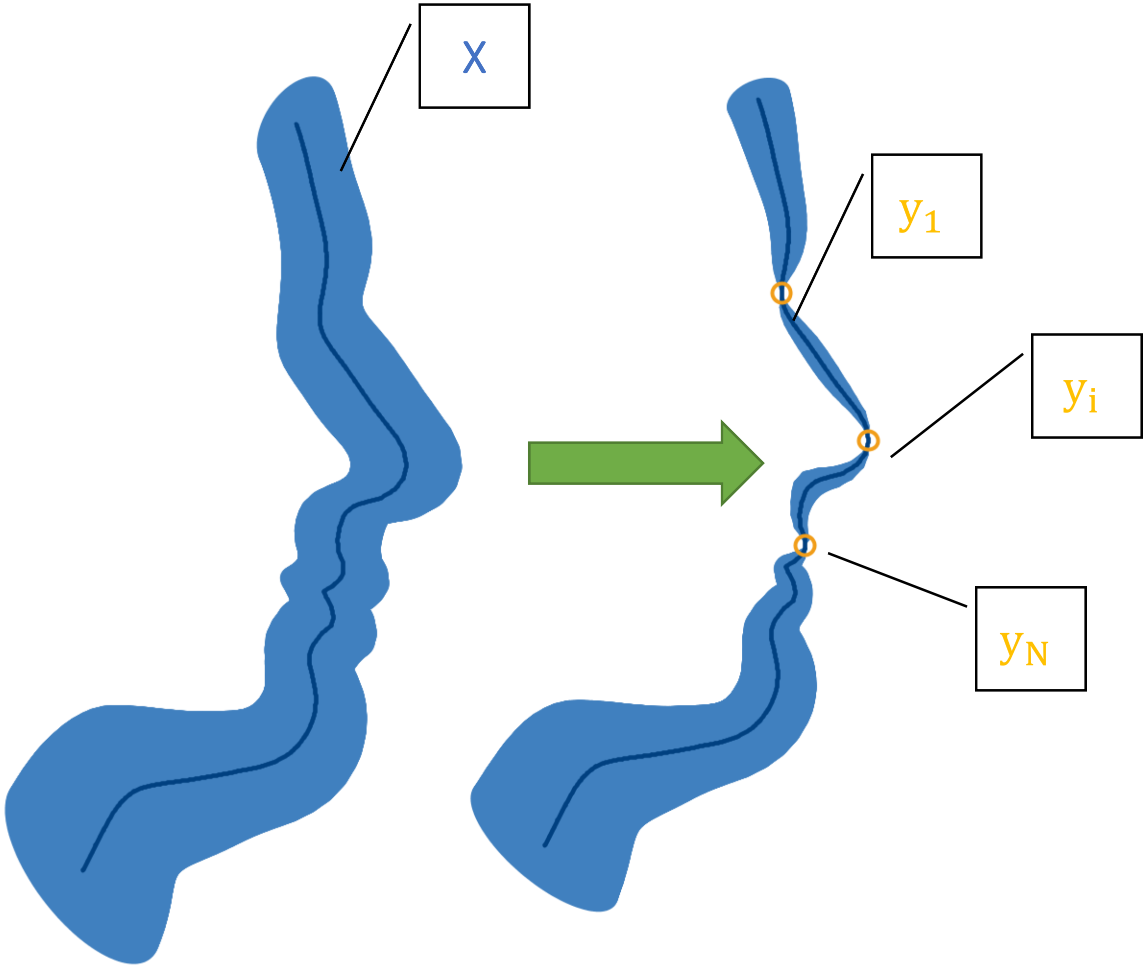
Certain Observation

- Observations are *known* values
- Distribution of X after *observing* $y_1, \dots, y_N$:

$$P(X|y_1 \dots y_N)$$

- Conditional probability

$$P(X|y_1 \dots y_N) = \frac{P(X, y, \dots, y_N)}{P(y_1, \dots, y_N)}$$



Towards Bayesian Inference

- Update belief about X by *observing* $y_1, \dots, y_N$

$$P(X) \rightarrow P(X|y_1, \dots, y_N)$$

- Factorize joint distribution

$$P(X, y_1, \dots, y_N) = P(y_1, \dots, y_N|X)P(X)$$

- Rewrite conditional distribution

$$P(X|y_1, \dots, y_N) = \frac{P(X, y_1, \dots, y_N)}{P(y_1, \dots, y_N)} = \frac{P(y_1, \dots, y_N|X)P(X)}{P(y_1, \dots, y_N)}$$

More generally: distribution of model points X given data Y:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$$

Uncertain Observation

- Observations with *uncertainty*

  Model needs to describe how observations are distributed

  with joint distribution $P(X,Y)$

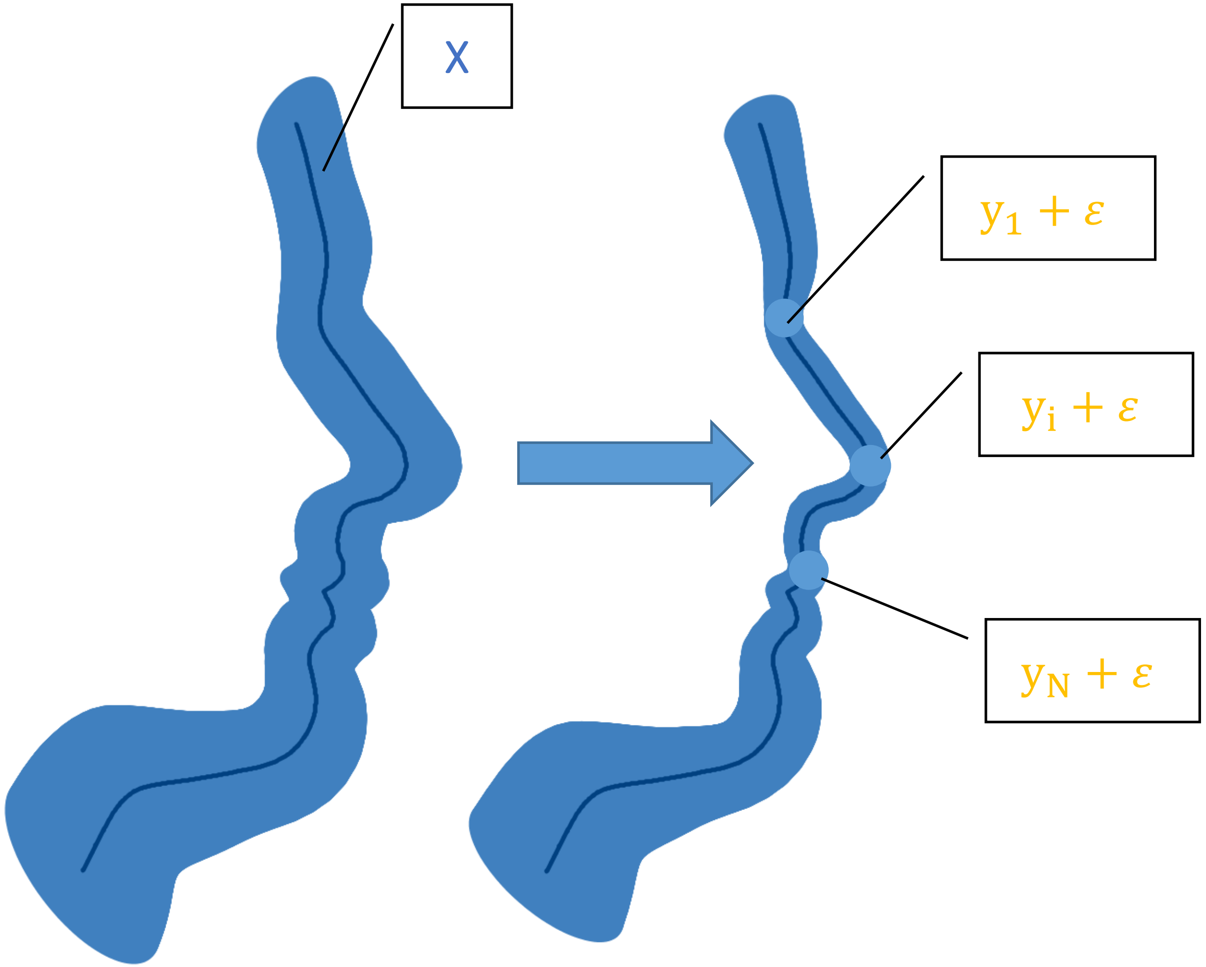
- Still conditional probability

  But joint distribution is more complex

- Joint distribution factorized

$$P(X,Y) = P(Y|X)P(X)$$

  - Likelihood $P(Y|X)$
  - Prior $P(X)$



Likelihood

Joint Likelihood Prior

$$P(X, Y) = P(Y|X)P(X)$$

- *Likelihood x prior:* factorization is more flexible than full joint
  - Prior: distribution of core model *without observation*
  - Likelihood: describes how observations are distributed

Bayesian Inference

- Conditional/Bayes rule: method to update *beliefs*

Likelihood Prior

Posterior

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Marginal Likelihood

- Each observation updates our belief (changes knowledge!)

$$P(X) \to P(X|Y) \to P(X|Y,Z) \to P(X|Y,Z,W) \to \cdots$$

- Bayesian Inference: How beliefs *evolve* with observation
- Recursive: Posterior becomes prior of next inference step

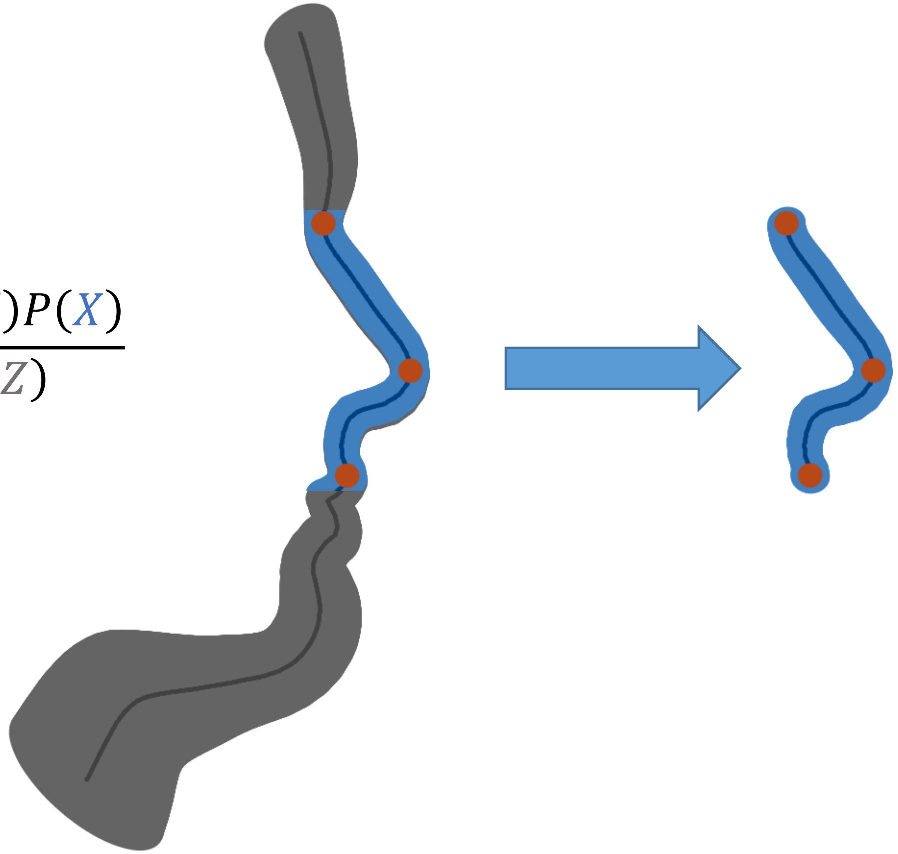
Marginalization

- Models contain irrelevant/hidden variables

  e.g. points on chin when nose is queried

- Marginalize over hidden variables (Z)

$$P(X|Y) = \sum_H P(X,Z|Y) = \sum_H \frac{P(Y,Z|X)P(X)}{P(Y,Z)}$$

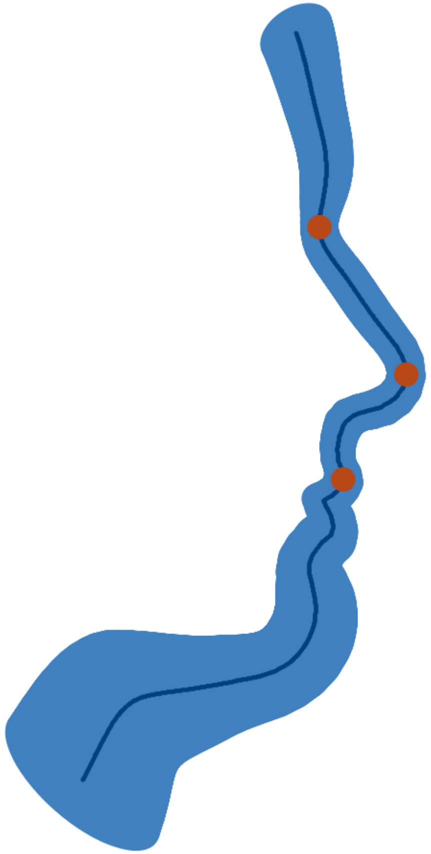


General Bayesian Inference

- Observation of *additional* variables
  - Common case, e.g. image intensities, surrogate measures (size, sex, …)
  - Coupled to core model via likelihood factorization
- General Bayesian inference case:
  - Distribution of data Y
  - Parameters θ

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}$$

$$P(\theta|Y) \propto P(Y|\theta)P(\theta)$$



Summary: Bayesian Inference

- *Belief*: formal expression of an *observer's knowledge*
  - Subjective state of knowledge about the world
- Beliefs are expressed as *probability* distributions
  - Formally not arbitrary: Consistency requires laws of probability
- *Observations* change knowledge and thus beliefs
- Bayesian inference formally updates *prior beliefs* to *posteriors*
  - Conditional Probability
  - Integration of observation via *likelihood x prior* factorization

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$