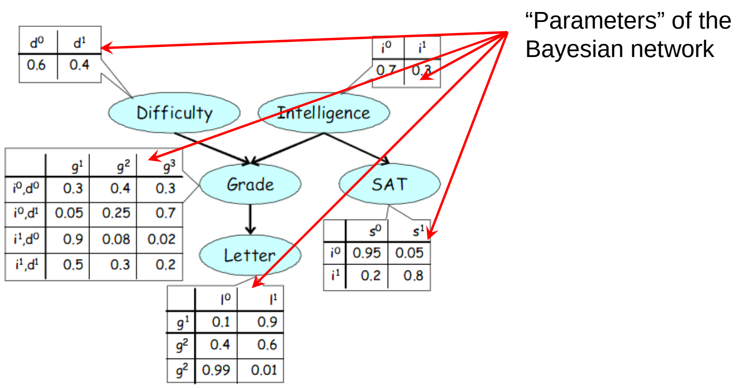
Mathematical Tools: Maximum Likelihood Estimation

Lecture 4

Instructor: Su-In Lee

Parameter estimation

- Assumptions
  - Fixed network structure
  - Fully observed instances of the network variables: D = {d[1],…,d[M]}
  - Maximum likelihood estimation (MLE)



The *Thumbtack* example

- Parameter learning for a single variable.

- Variable
  - X: an outcome of a thumbtack toss
  - Val(X) = {head, tail}

- Data
  - A set of thumbtack tosses: x[1],…x[M]

Maximum likelihood estimation

- Say that P(x=head) = Θ, P(x=tail) = 1-Θ
  - P(<HHTHTH…HHTT>; Θ) =

    Mh heads, Mt tails

- **Definition:** The *likelihood* function
  - L(Θ : D) = P(D; Θ)

- Maximum likelihood estimation (MLE)
  - Given data D=<M_h heads, M_t tails>, find Θ that maximizes the likelihood function L(Θ : D).

MLE for the *Thumbtack* problem

- Given data D=<M_h heads, M_t tails>,
  - MLE solution $\Theta^* = M_h / (M_h+M_t)$.

- Proof:

MLE for general problems

- Learning problem setting
  - A set of random variables X from unknown distribution P*
  - Training data D = M instances of X: {d[1],…,d[M]}

- A *parametric model* P(X; Θ) (a 'legal' distribution)

- Define the *likelihood function*:
  - L(Θ : D) =

- Maximum likelihood estimation
  - Choose parameters Θ* that satisfy:

MLE for Bayesian networks

- Likelihood decomposition:
- The local likelihood function for X_i is:

Bayesian network with table CPDs

Acknowledgement

- This set of slides is based on the following materials:
  - "Probabilistic Graphical Models: Principles and Techniques" by Daphne Koller and Nir Friedman