Bayesian Models of Cognition

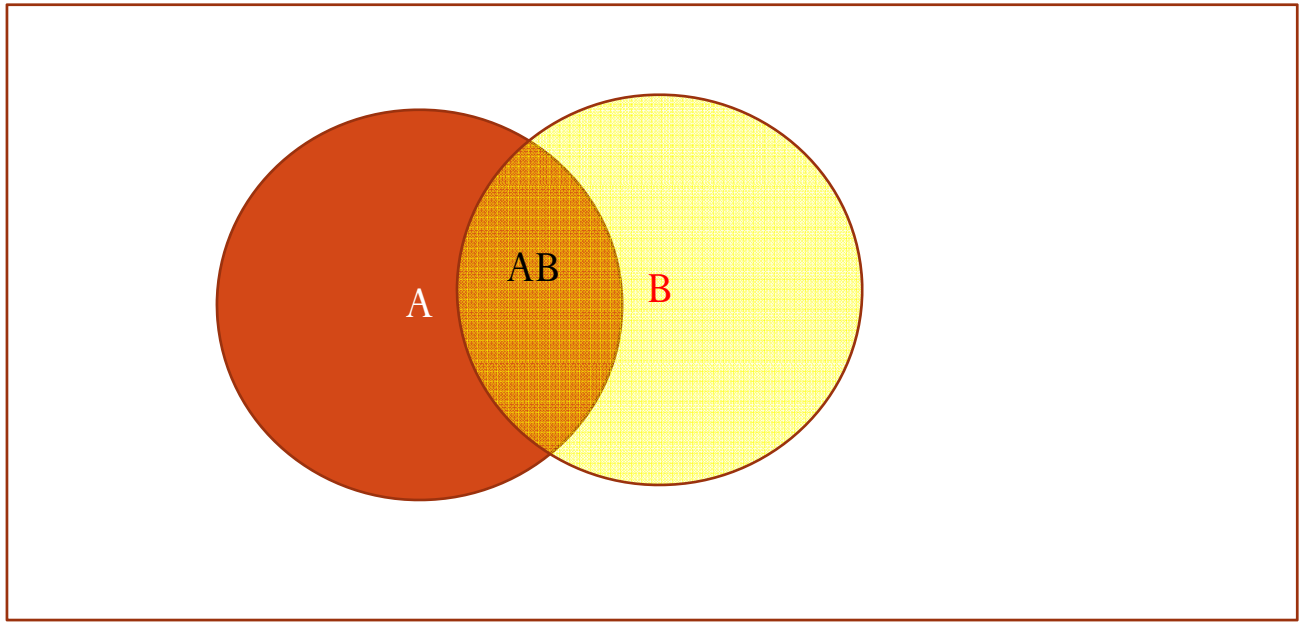
----The Cambridge Handbook of Computational Psychology (Chap.3)

Jake

Clustering Intelligence Club

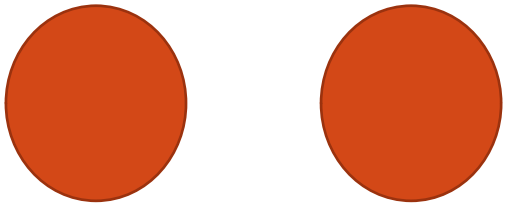
Bayesian Theorem

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)}$$



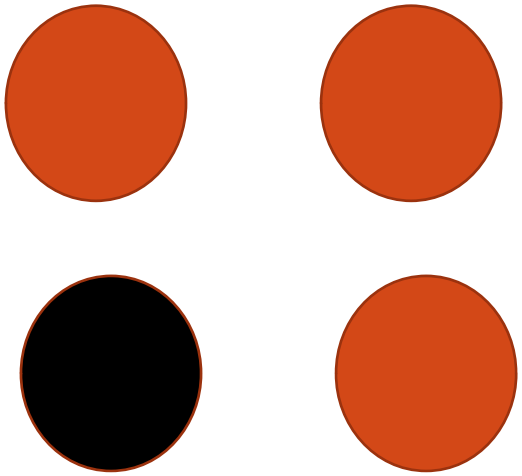
Question 1

- There are 2 children
- What is the probability of one child being a boy?



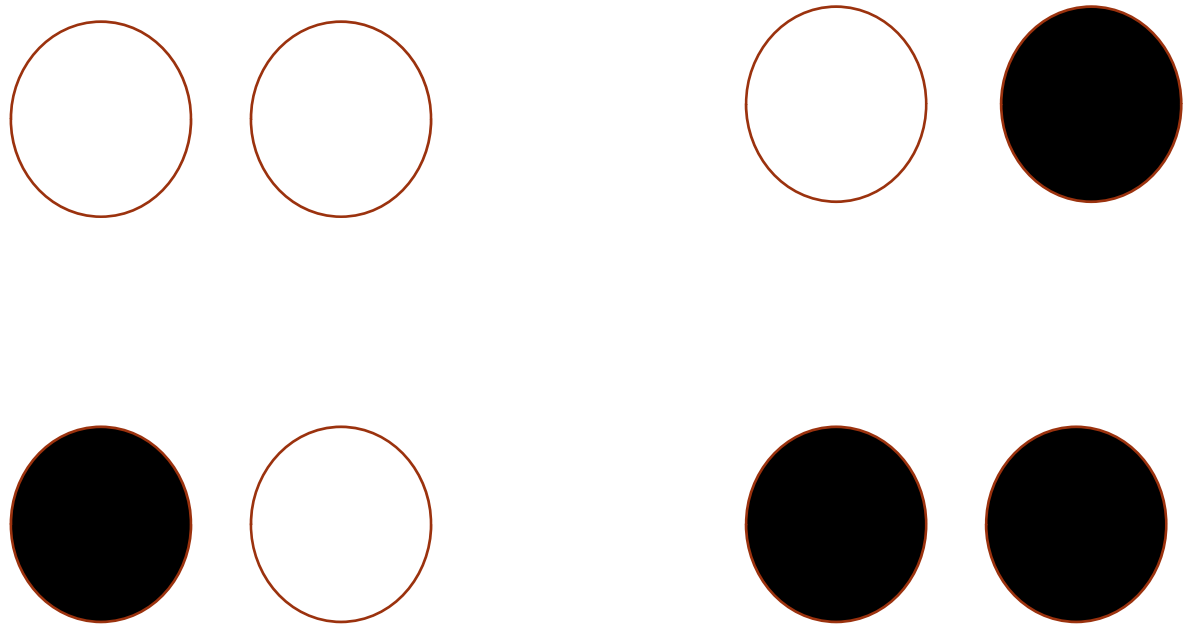
Question 2

- There are 2 children
- One of them is a boy
- What is the probability of the other child being a boy?



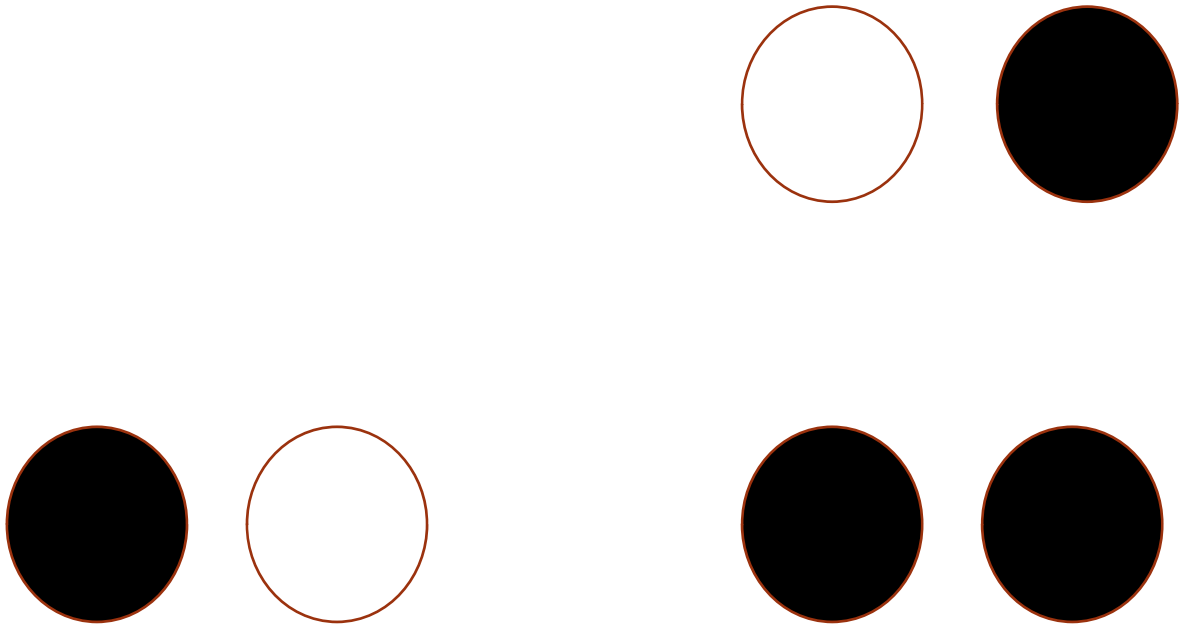
A Question

- Event Space



A Question

- After telling the information



On Bayesian Theorem

- B=One of the child is a boy
- A=The other is a boy
- $P(A|B)=P(AB)/P(B)=(1/4)/(3/4)=1/3$

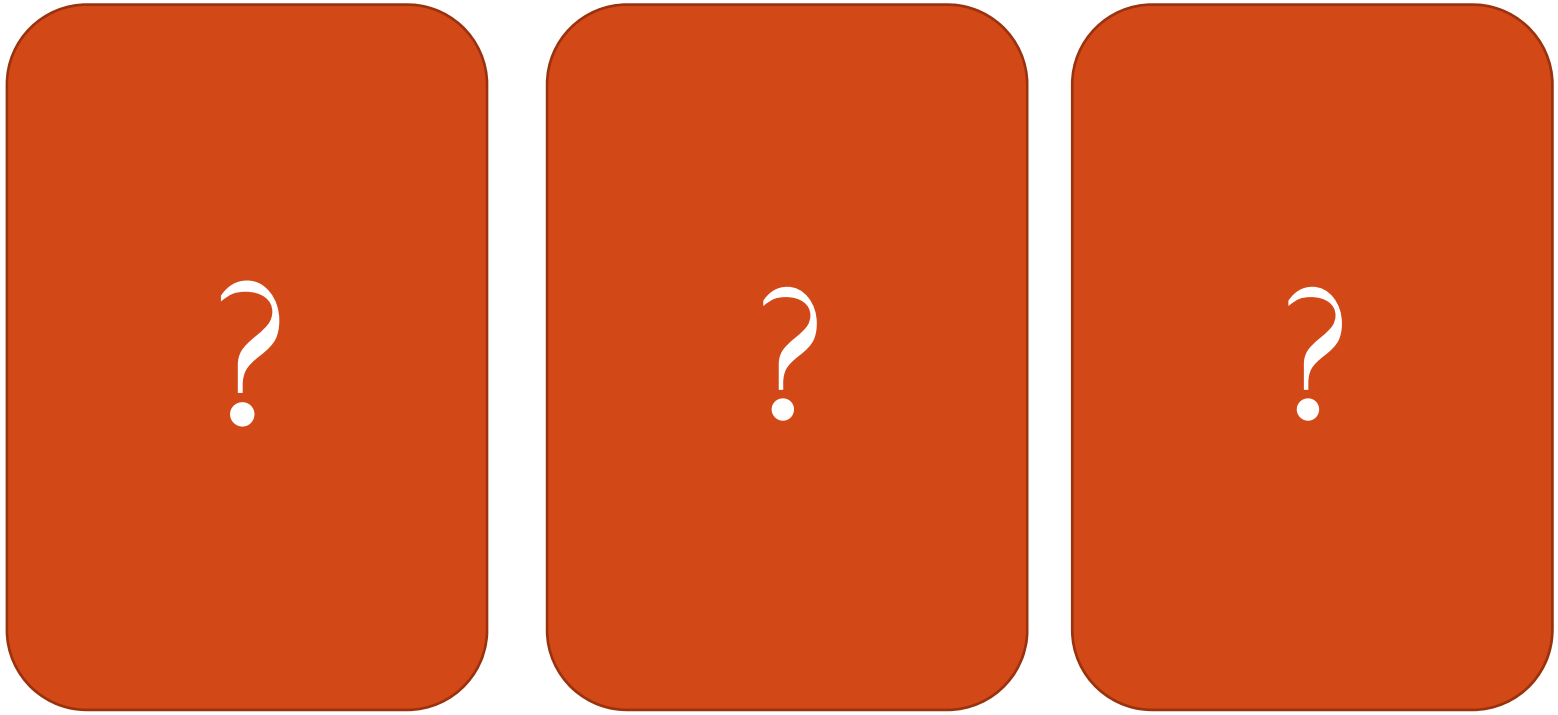
The hard philosophical point

- Why does probability change?
- Probability is not an objective measurement
- Probability is subjective
- So Bayesian is called subjective probability school

Another hard point

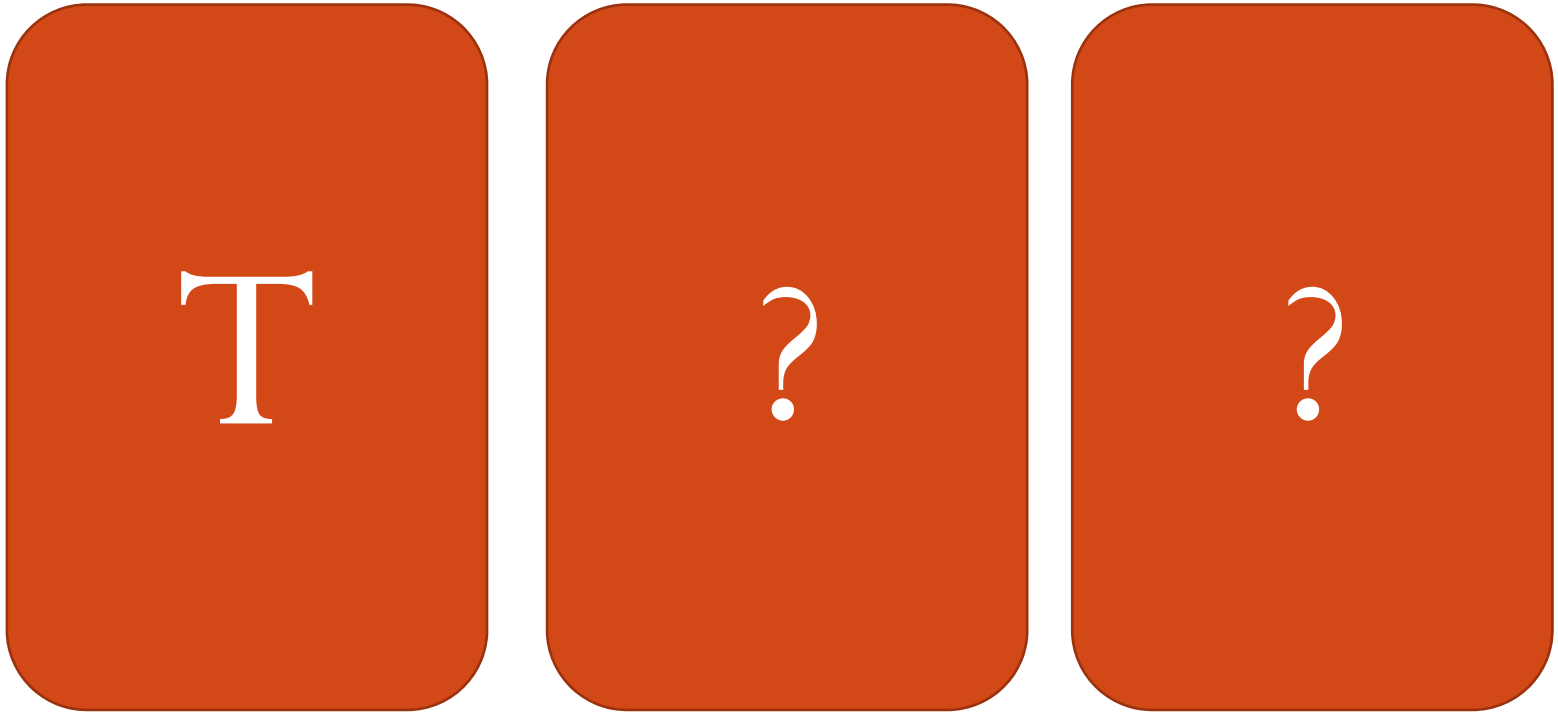
- X is a random variable in {0,1} denotes the first child
- Y is a random variable in {0,1} denotes the second child
- How can we write down Bayesian theorem?
- P(X=1|Y=1)=P(X=1&Y=1)/P(Y=1)=(1/4)/(1/2)=1/2
- P(Y=1|X=1)=P(X=1&Y=1)/P(X=1)=1/2
- But, P(X=1|Y=1)+ P(Y=1|X=1)=1
- P(X=1/\Y=1|X=1VY=1)=P(X=1/\Y=1/\(X=1VY=1))/P(X=1VY=1)=(1/4)/(3/4)=1/3

A Game---Monty Hall Problem



1 Beauty & 2 Tiggers

A Game---Monty Hall Problem



Still stick to your choice or another one?

A Game---Monty Hall Problem

- The answer is you should change your choice
- Because, the other one is beauty with more probability!
- The probability of another is not 1/3, but 2/3!

A Wrong Analysis

- A: A is a beauty
- B: B is a tiger
- $P(A|B)=P(AB)/P(B)=(1/3)/(2/3)=1/2$

- The key point is B will not be selected if B is just the beauty
- The host makes selection non-randomly

A Correct Bayesian Analysis

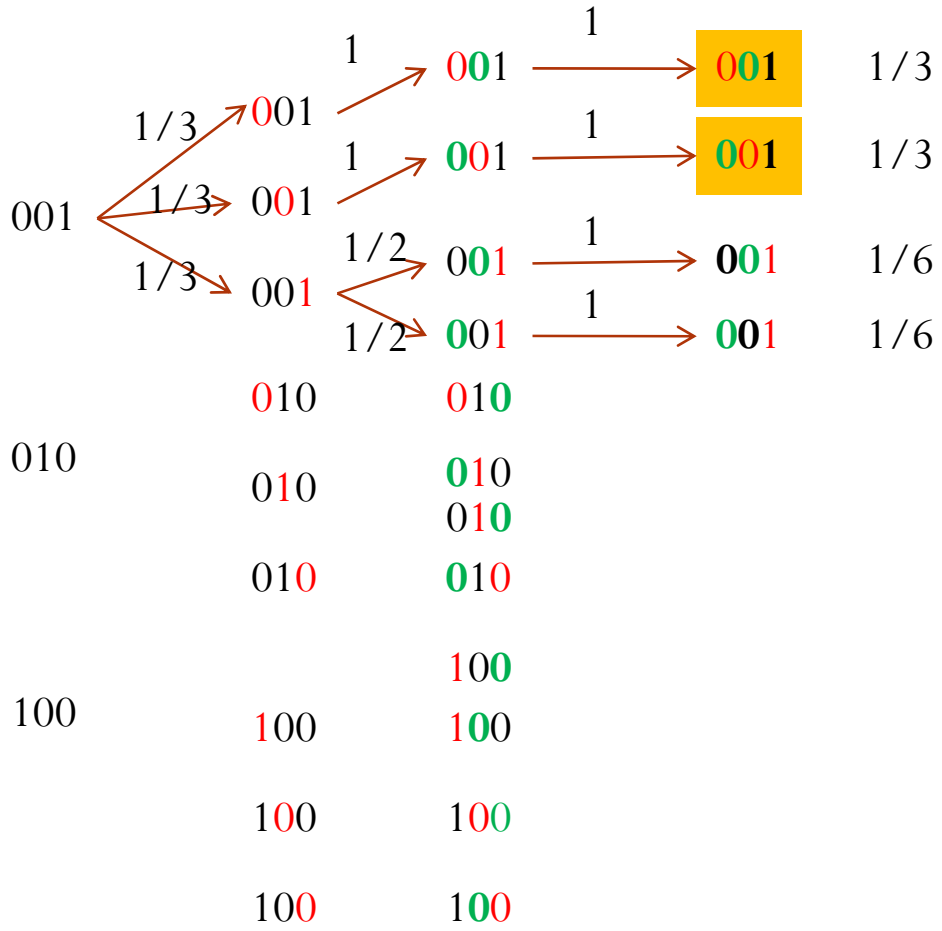
- A: A is a beauty
- B: The host opens door B after your selection
- C: C is a beauty

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|B)P(B) + P(B|C)P(C)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{3}$$

$$P(C|B) = \frac{P(CB)}{P(B)} = \frac{P(B|C)P(C)}{P(B|A)P(A) + P(B|B)P(B) + P(B|C)P(C)} = \frac{2}{3}$$

Event Space---Decision Tree



Modeling Human Inference

$$P(h \mid d) = \frac{P(d \mid h)P(h)}{P(d)} = \frac{P(d \mid h)P(h)}{\sum_{h'} P(d \mid h')P(h')}$$

An Example

- Drawing a coin, get the sequence:
  - HHHHHHHHHH
  - HHTHTHTTHT
- Comparing Hypotheses
- Parameter Estimation
- Model Selection

Comparing Hypotheses

- Hypothesis 1:
  - Producing head with probability 0.5
- Hypothesis 2:
  - Producing head with probability 0.9
- Which hypothesis do you prefer when you observe:
  - HHHHHHHHHH
  - HHTHTHTTHT

Comparing Hypotheses

- Bayesian Formula:

$$P(h_1 \mid d) = \frac{P(d \mid h_1)P(h_1)}{P(d)} = \frac{P(d \mid h_1)P(h_1)}{P(d)}$$

- h1: θ1=0.5
- h2: θ2=0.9
- d: observation

Comparing Hypotheses

$$P(h_1) = 1/2, P(h_2) = 1/2$$

$$P(d \mid h_i) = P((\#(H) = N_H) \mid h_i) = \theta_i^{N_H} (1-\theta_i)^{N_T}$$

$$P(h_i \mid d) = \frac{P(d \mid h_i) P(h_i)}{P(d)}$$

$$\frac{P(h_1 \mid d)}{P(h_2 \mid d)} = \frac{P(d \mid h_1) P(h_1)}{P(d \mid h_2) P(h_2)} = \frac{\theta_1^{N_H} (1-\theta_1)^{N_T} 0.5}{\theta_2^{N_H} (1-\theta_2)^{N_T} 0.5} = \frac{0.5^{N_H} 0.5^{N_T} 0.5}{0.9^{N_H} 0.1^{N_T} 0.5}$$

- It is 0.00280075 when HHHHHHHHHH is observed
- It is 165.382 when HHTHTHTTHT is observed

Comparing Hypotheses

Posterior odds Likelihood ratio Prior odds

$$\frac{P(h_1 \mid d)}{P(h_2 \mid d)} = \frac{P(d \mid h_1)}{P(d \mid h_2)} \cdot \frac{P(h_1)}{P(h_2)}$$

$$\log \frac{P(h_1 \mid d)}{P(h_2 \mid d)} = \log \frac{P(d \mid h_1)}{P(d \mid h_2)} + \log \frac{P(h_1)}{P(h_2)}$$

Parameter Estimation

- I don't know the parameter θ in [0,1]
- Observe the sequences
  - HHHHHHHHHH
  - HHTHTHTTHT
- What is the most probable value of θ?
- It is the infinite hypotheses comparison version.

Parameter Estimation

$$p(\theta)=1, \int_0^1 p(\theta)d\theta=1$$

$$P(d\mid\theta)=\theta^{N_H}(1-\theta)^{N_T}$$

$$P(d)=\int_0^1 P(d\mid\theta)d\theta=\int_0^1 \theta^{N_H}(1-\theta)^{N_T}\,d\theta$$

$$p(\theta\mid d)=\frac{P(d\mid\theta)p(\theta)}{P(d)}=\frac{\theta^{N_H}(1-\theta)^{N_T}}{\int_0^1 \theta^{N_H}(1-\theta)^{N_T}\,d\theta}=\frac{(N_H+N_T+1)!}{N_H!N_T!}\theta^{N_H}(1-\theta)^{N_T}=\beta(N_H+1,N_T+1)$$

Estimation

$$\hat{\theta}=\int_0^1 \theta p(\theta\mid d)d\theta$$ posterior mean estimation

$$\hat{\theta}=\arg\max_{\theta}(p(\theta\mid d))$$ Maximum a posterior estimation

Model Selection

- A special hypothesis comparison
- Hypothesis (model) 1:
  - θ is 0.5
- Hypothesis(model) 2:
  - Θ is in [0,1]
- Which hypothesis should I prefer when I watch the sequence?

Model Selection

- A special hypothesis comparison
- Hypothesis (model) 1:
  - θ is 0.5
- Hypothesis(model) 2:
  - Θ is in [0,1]
- Which hypothesis should I prefer when I watch the sequence?

Model Selection

- Compute the posterior odds

Posterior odds Likelihood ratio Prior odds

$$\frac{P(h_1 \mid d)}{P(h_2 \mid d)} = \frac{P(d \mid h_1)}{P(d \mid h_2)} \cdot \frac{P(h_1)}{P(h_2)} = \frac{0.5^{N_H} 0.5^{N_T} \times 0.5}{P(d \mid h_2) \times 0.5}$$

$$P(d \mid h_2) = \int_0^1 P(d \wedge \theta \mid h_2) d\theta = \int_0^1 P(d \mid \theta, h_2)\, p(\theta \mid h_2) d\theta = \int_0^1 \theta^{N_H} (1-\theta)^{N_T}\, d\theta$$

$$= \frac{N_H! N_T!}{(N_H + N_T + 1)!}$$

$$\log \frac{P(h_2 \mid d)}{P(h_1 \mid d)} = \sum_{i=1}^{N_H} \log i + \sum_{i=1}^{N_T} \log i - (N_H + N_T) \log(0.5) - \sum_{i=1}^{N_H + N_T + 1} \log i$$

Model Selection

- Compute the posterior odds

Posterior odds Likelihood ratio Prior odds

$$\frac{P(h_1 \mid d)}{P(h_2 \mid d)} = \frac{P(d \mid h_1)}{P(d \mid h_2)} \cdot \frac{P(h_1)}{P(h_2)} = \frac{0.5^{N_H} 0.5^{N_T} \times 0.5}{P(d \mid h_2) \times 0.5}$$

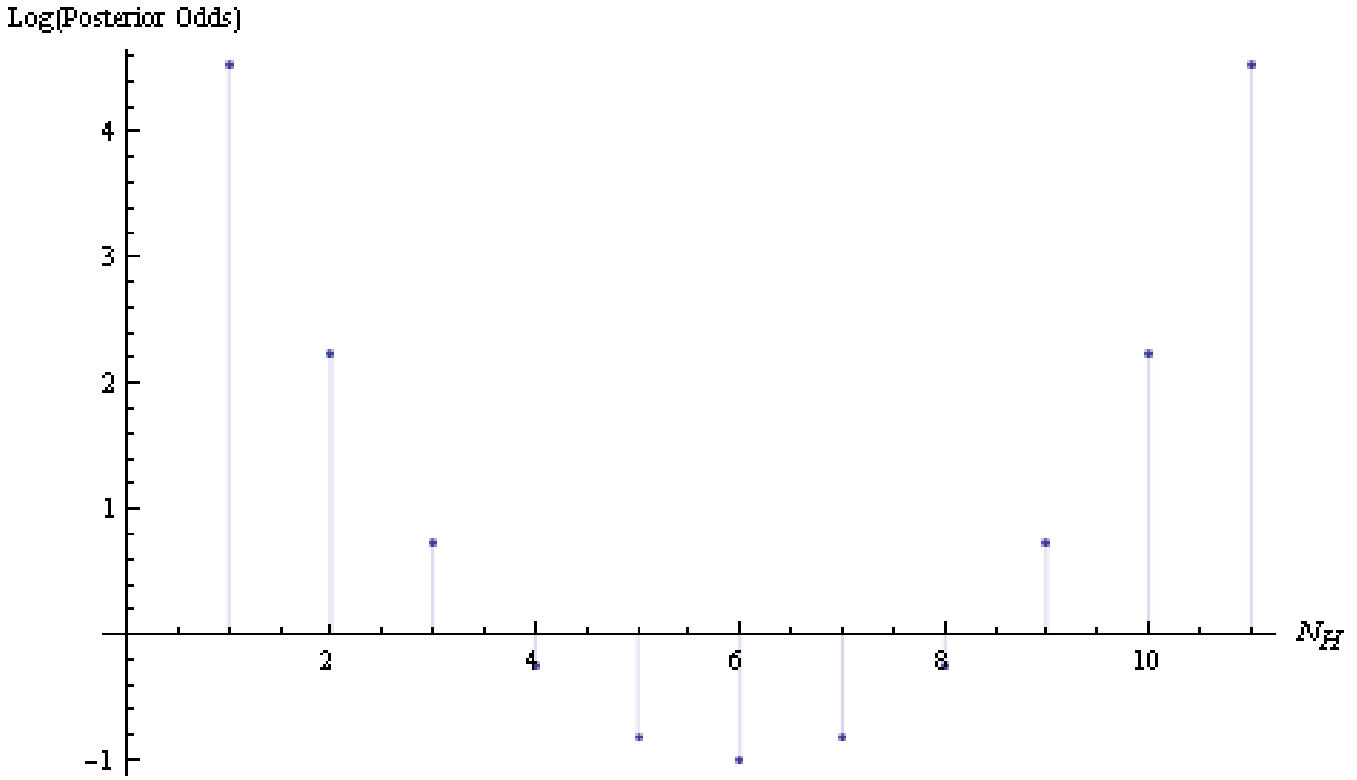
$$P(d \mid h_2) = \int_0^1 P(d \wedge \theta \mid h_2) d\theta = \int_0^1 P(d \mid \theta, h_2)\, p(\theta \mid h_2) d\theta = \int_0^1 \theta^{N_H} (1-\theta)^{N_T}\, d\theta$$

$$= \frac{N_H! N_T!}{(N_H + N_T + 1)!}$$

$$\log \frac{P(h_2 \mid d)}{P(h_1 \mid d)} = \sum_{i=1}^{N_H} \log i + \sum_{i=1}^{N_T} \log i - (N_H + N_T) \log(0.5) - \sum_{i=1}^{N_H + N_T + 1} \log i$$

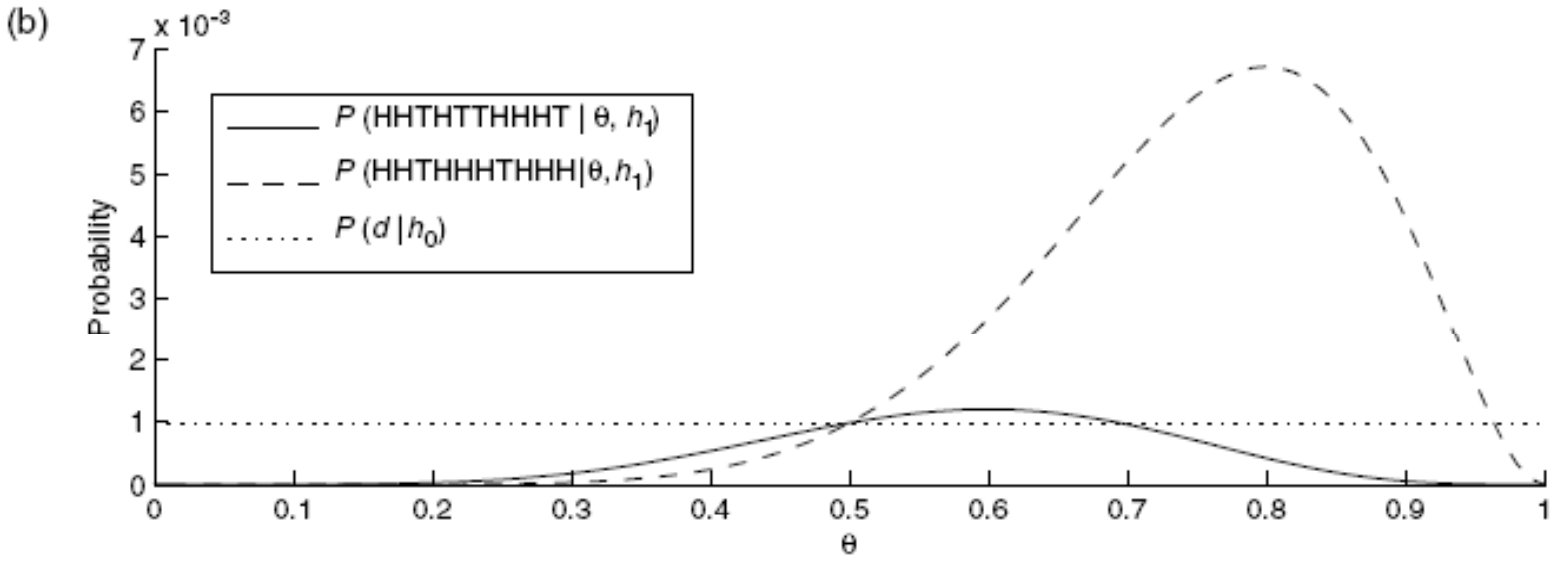
Model Selection

- $N_H+N_T=10$



Bayesian Occam's Razor

- It seems that more complex model has advantage, for example:

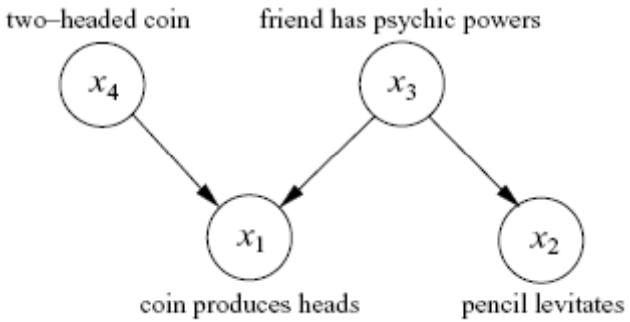


Bayesian Occam's Razor

- However, the computation of P(d | h) takes average for all h.
- So, complex models have no more advantages
- This is Bayesian Occam's Razor

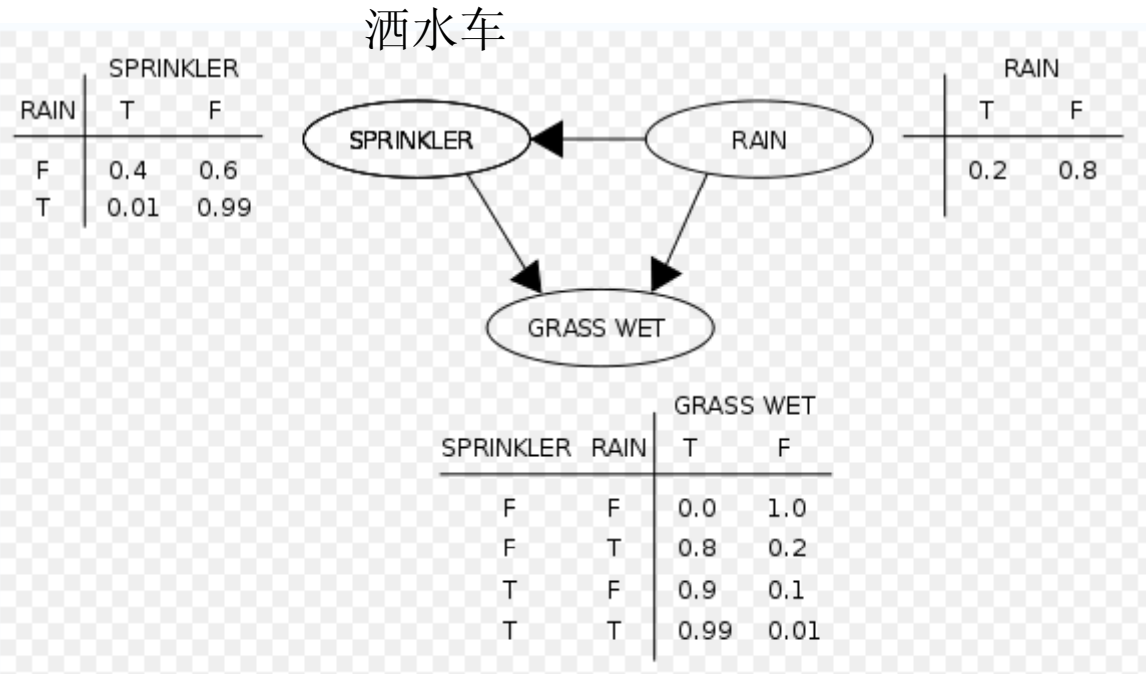
$$P(d \mid h_2) = \int_0^1 P(d \wedge \theta \mid h_2) d\theta = \int_0^1 P(d \mid \theta, h_2)\, p(\theta \mid h_2) d\theta$$

Bayesian Network



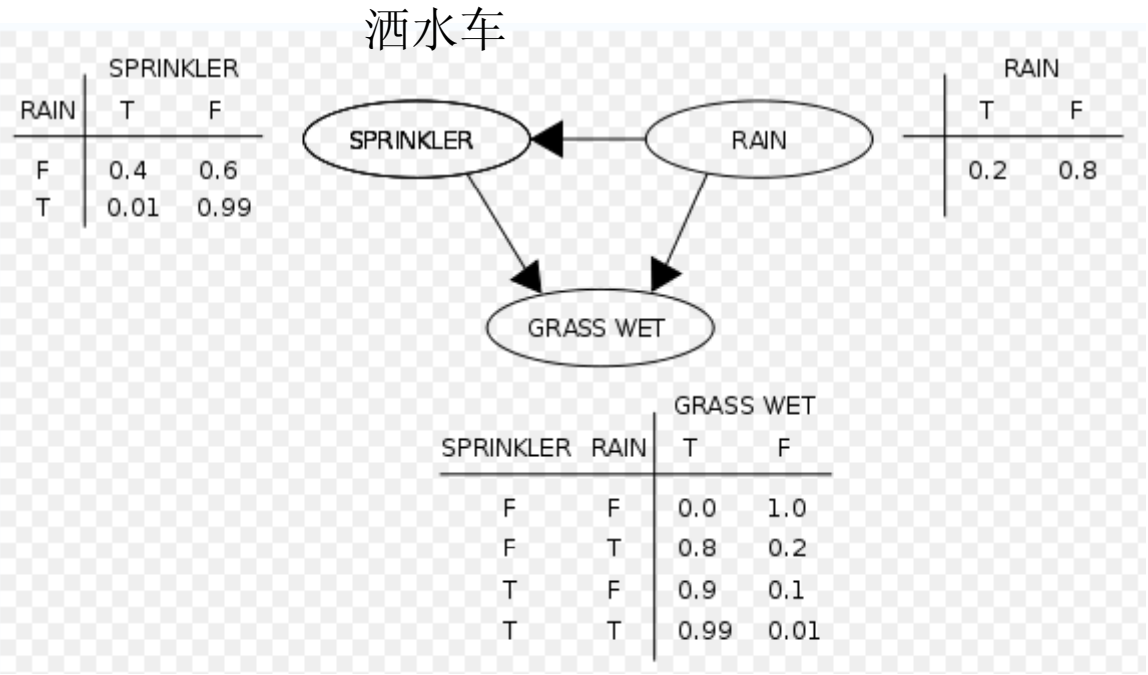
- P(X1,X2,X3,X4)=P(X1|X3,X4)P(X2|X3)P(X3)P(X4)
- General Rule

Bayesian Network



- P(SRG)=P(S|R)P(G|S,R)P(R)
- P is distribution function but not probability

Bayesian Network

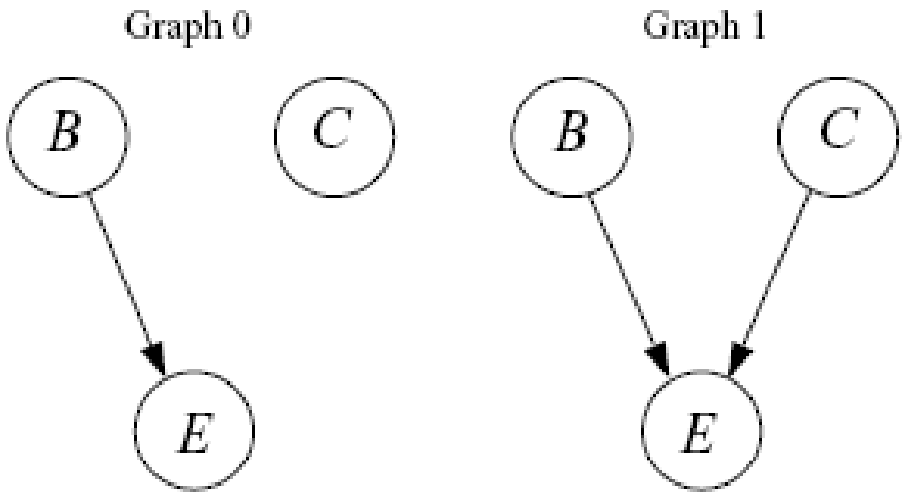


RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99

- Exercise: What is the probability of R=T|G=T
- P(R=T&G=T)/P(G=T)=

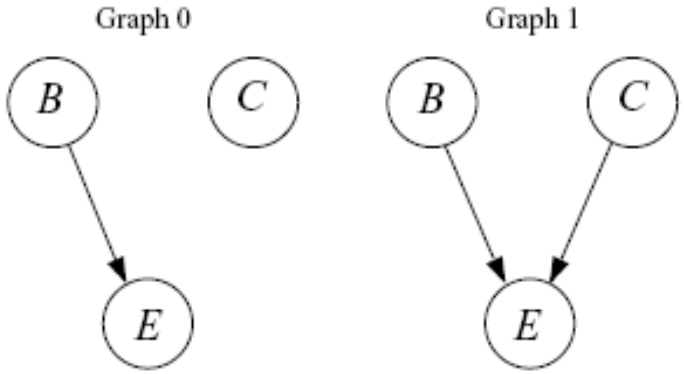
Causal Graph

- Inference the network's structure
- Is the given edge existed?



Causal Graph

- By giving a contingency table



	Effect present (e^+)	*Effect absent* (e^-)
Cause present (c^+)	$N(e^+, c^+)$	$N(e^-, c^+)$
Cause absent (c^-)	$N(e^+, c^-)$	$N(e^-, c^-)$

Causal Graph

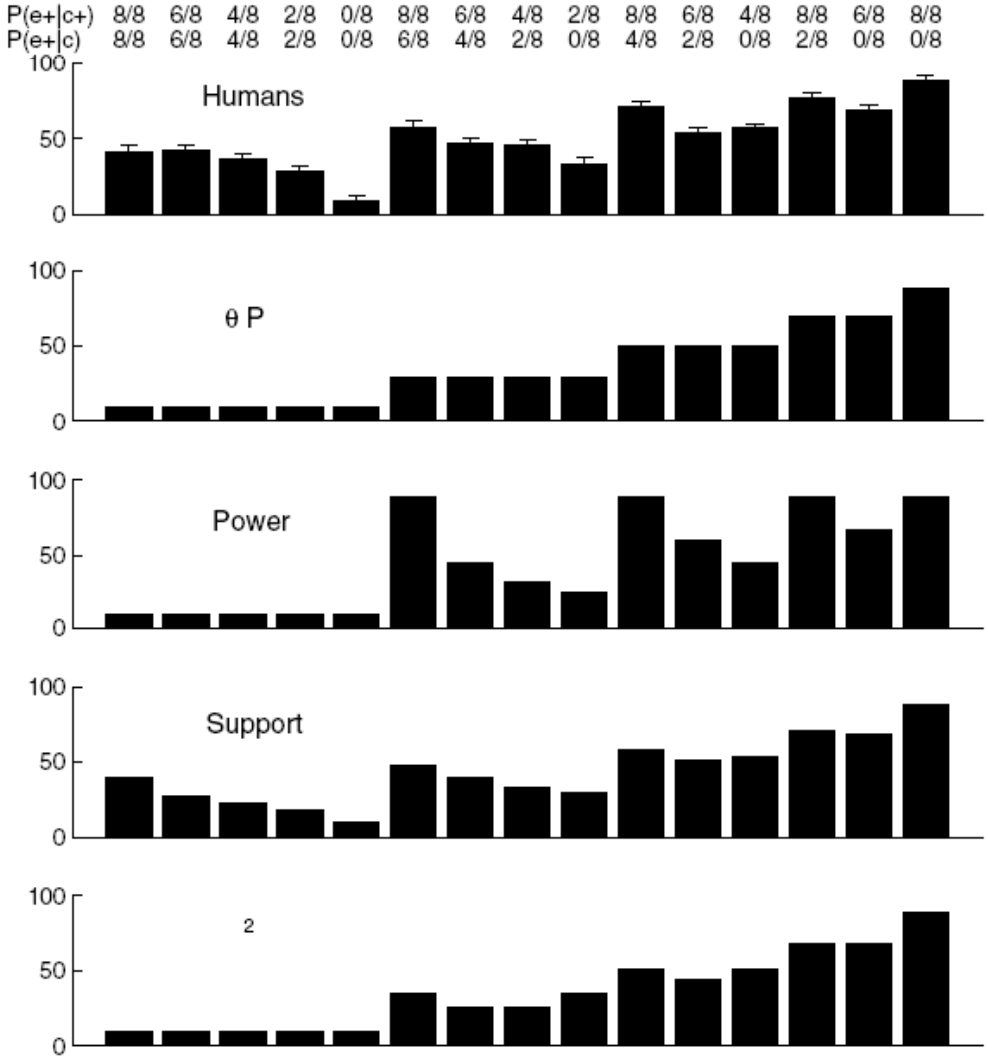
- Different indicators

	Effect present (e^+)	*Effect absent* (e^-)
Cause present (c^+)	$N(e^+, c^+)$	$N(e^-, c^+)$
Cause absent (c^-)	$N(e^+, c^-)$	$N(e^-, c^-)$

$$\Delta P = \frac{N(e^+, c^+)}{N(e^+, c^+) + N(e^-, c^+)} - \frac{N(e^+, c^-)}{N(e^+, c^-) + N(e^-, c^-)} = P(e^+ \mid c^+) - P(e^+ \mid c^-),$$

$$\text{power} = \frac{\Delta P}{1 - P(e^+ \mid c^-)},$$

Causal Graph



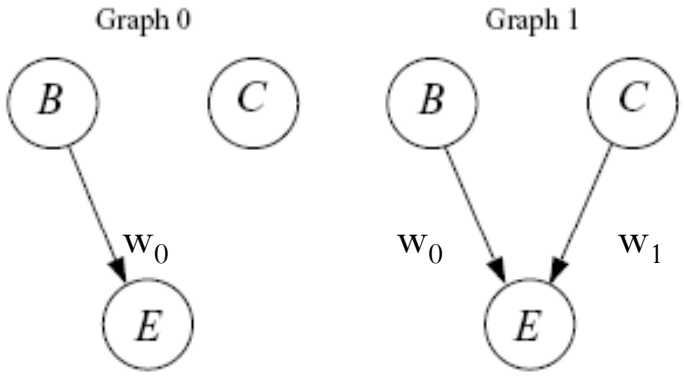
Graph selection – Model Selection

- Support is computed as (w_i as parameter):

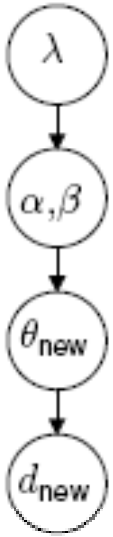
$$\text{support} = \log \frac{P(d \mid \text{Graph } 1)}{P(d \mid \text{Graph } 0)}$$

$$P(d \mid \text{Graph } 1) = \int_0^1 \int_0^1 P_1(d \mid w_0, w_1, \text{Graph } 1) \times P(w_0, w_1 \mid \text{Graph } 1)\, dw_0\, dw_1 \quad (3.24)$$

$$P(d \mid \text{Graph } 0) = \int_0^1 P_0(d \mid w_0, \text{Graph } 0) \times P(w_0 \mid \text{Graph } 0)\, dw_0. \quad (3.25)$$



Hierarchical Bayesian Model

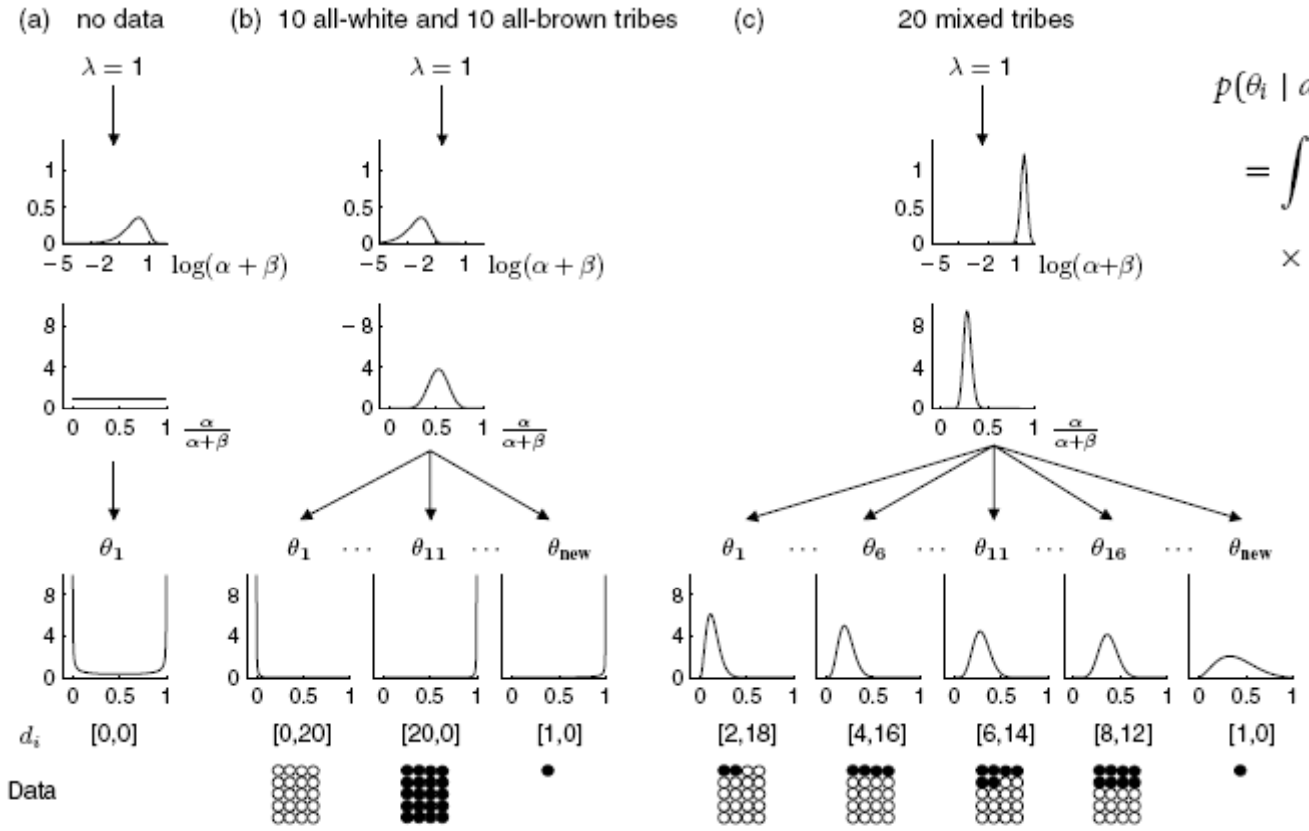


- In Bayesian Model, you should know the parameter distribution as the prior knowledge
  - i.e., θ distributes on [0,1] evenly.
- However, this may be unknown (or changeable)
- You just know θ follows a Beta distribution Beta(alpha,beta)
- But the parameters alpha and beta should be estimated
- This is a second order problem of Bayesian inference.
- You know the prior distribution of alpha and beta follows an exponential distribution with fixed parameter lambda

Hierarchical Bayesian Model

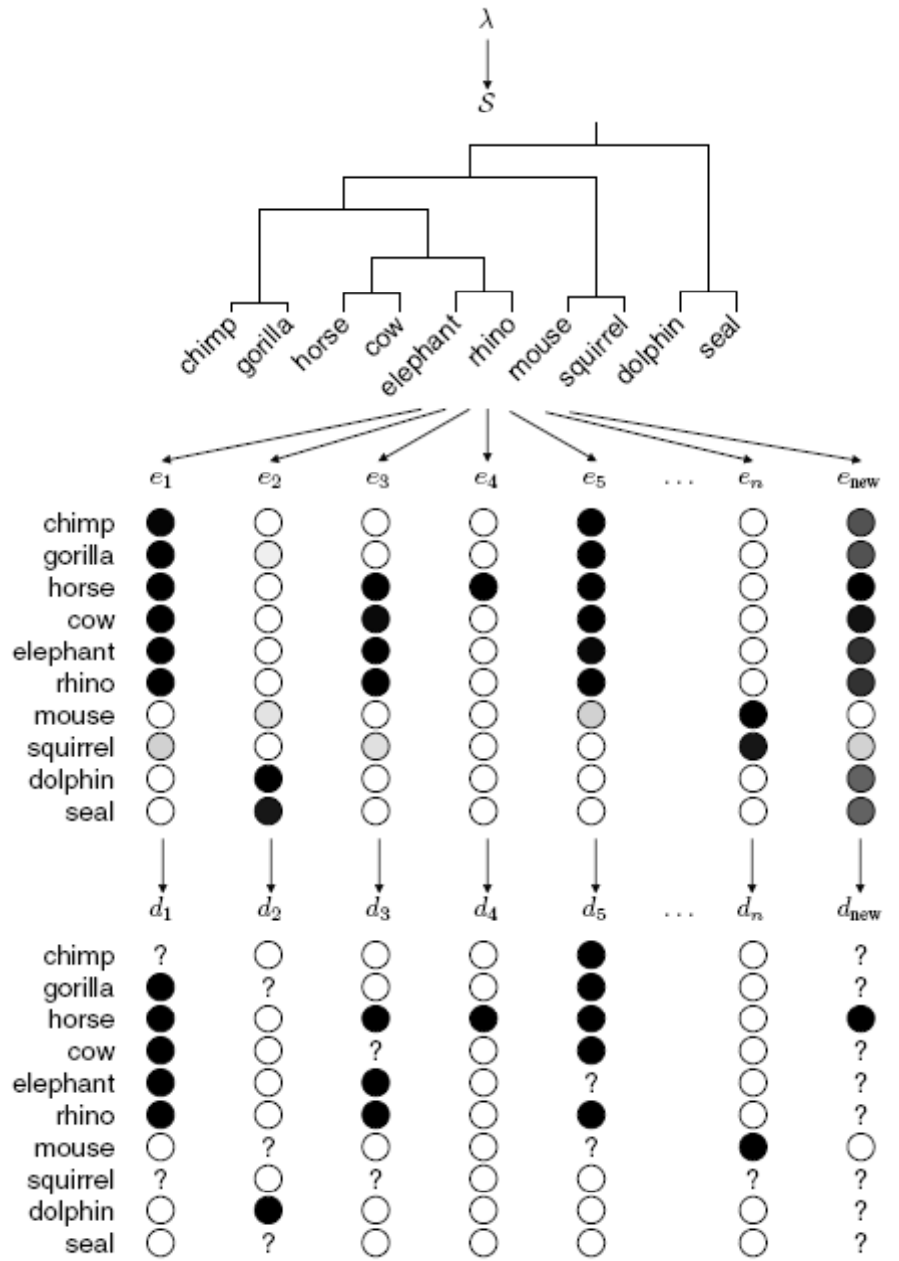
- Update information about alpha and beta
- Suppose you have 10 different coins, you can perform 20 experiments on each coin
- The alpha and beta can be updated accordingly

Hierarchical Bayesian Model



$$p(\theta_i \mid d_1, d_2, \ldots, d_n) = \int p(\theta_i \mid \alpha, \beta, d_i) \times p(\alpha, \beta \mid d_1, d_2, \ldots, d_n) d\alpha d\beta$$

Figure 3.8. Inferences about the distribution of features within tribes. (a) Prior distributions on θ, $\log(\alpha+\beta)$ and $\frac{\alpha}{\alpha+\beta}$. (b) Posterior distributions after observing 10 all-white tribes and 10 all-brown tribes. (c) Posterior distributions after observing 20 tribes. Black circles indicate obese indiviuals and the rate of obesity varies among tribes. Reproduced with permission from Kemp, Perfors, and Tenenbaum (2007).

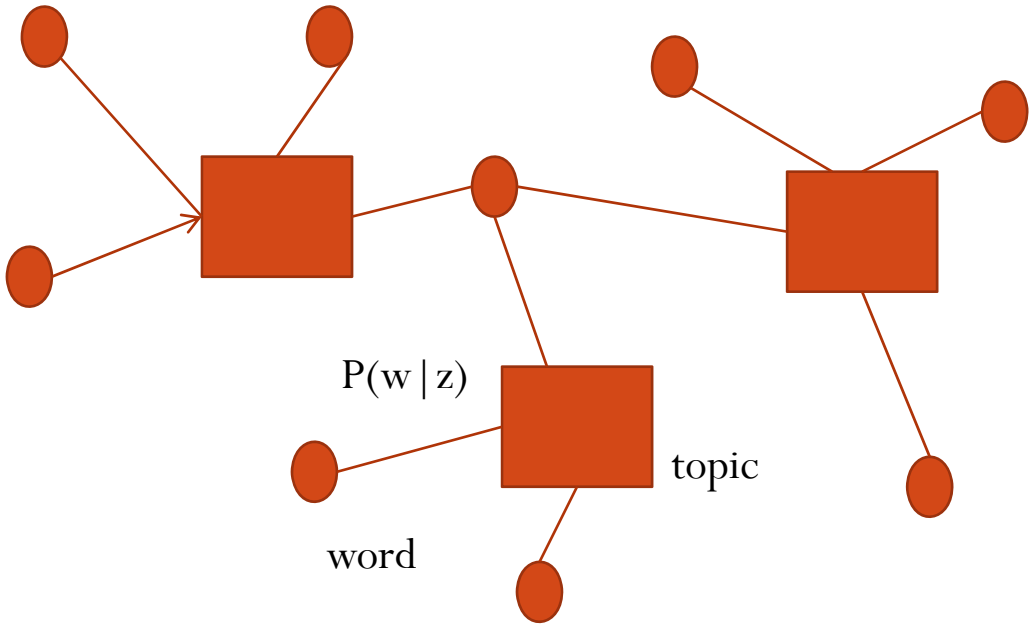


Hierarchical Bayesian Model

Monte Carlo Simulation

- Probability theory
  - Mathematical Theory
- Statistics
  - Inference about probability according data
- Monte Carlo Simulation
  - Generating Data according to probability model (distribution function)

The topic-word model

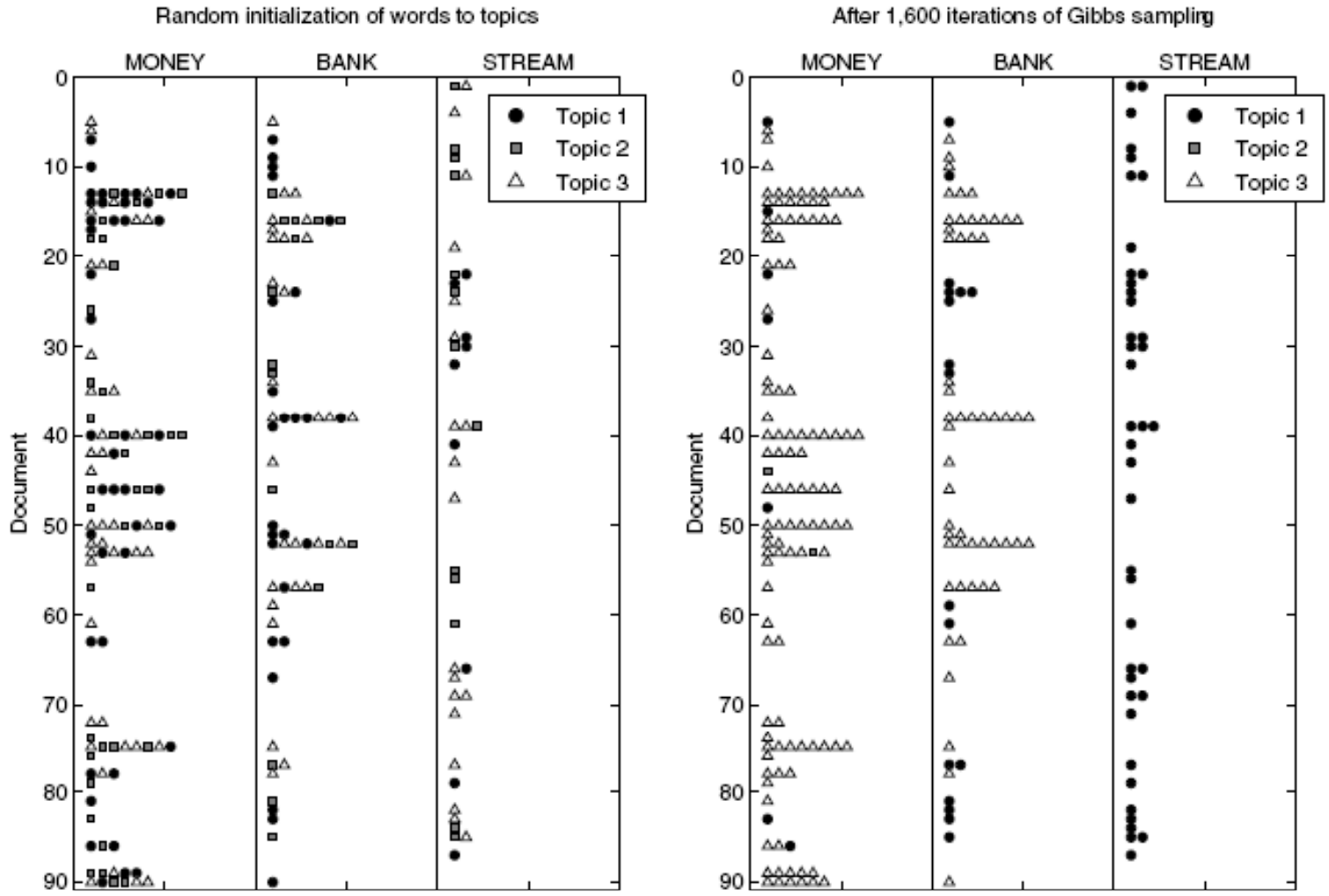


Articles as training data

Inference:

$$P(z \mid w_1) = \frac{P(w_1 \mid z)P(z)}{\sum_{z'=1}^{T} P(w_1 \mid z')P(z')}$$

$$P(w_2 \mid w_1) = \sum_{z=1}^{T} P(w_2 \mid z)P(z \mid w_1)$$


Random initialization of words to topics
After 1,600 iterations of Gibbs sampling
MONEY
BANK
STREAM
Document
0
10
20
30
40
50
60
70
80
90
Topic 1
Topic 2
Topic 3