Introduction to Computational Cognitive Science
AS.050.202 (Fall 2019)

Lab 4
(09/20/19)

Grusha Prasad and Najoung Kim

Plan

- Go over HW1
- GCM notation review + other questions
- Bayesian Inference

Note: Najoung will not be having office hours today. Email her if you wanted to meet her.

GCM notation

$$s_{ij} = e^{-cd_{ij}}$$

$$P(R_i = A \mid i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}$$

GCM notation

How similar is **i** to all members of category A

$$s_{ij} = e^{-cd_{ij}}$$

$$P(R_i = A \mid i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}$$

Exponentiating the negative distance between i and j

How similar is **i** to all previously observed examples

Given that a person has seen **i**, what is the probability that the person assigns **i** to category **A**

GCM notation

How similar is **i** to all members of category A

$$s_{ij} = e^{-cd_{ij}}$$

$$P(R_i = A \mid i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}$$

Exponentiating the negative distance between i and j

How similar is **i** to all previously observed examples

Given that a person has seen **i**, what is the probability that the person assigns **i** to category **A**

How does the GCM take prior knowledge into account?

Bayesian inference

When we apply Bayesian inference for categorization, we are combining information from different sources to help decide what the most appropriate category is.

What are the different sources of information?

Bayesian inference

When we apply Bayesian inference for categorization, we are combining information from different sources to help decide what the most appropriate category is.

What are the different sources of information?

- Information about how probable different categories are
- Information from the input.
- Information about the different generative processes that could have generated the input.

Bayesian inference

When we apply Bayesian inference for categorization, we are combining information from different sources to help decide what the most appropriate category is.

What are the different sources of information?

- Information about how probable different categories are **Prior**
- Information from the input.
- Information about the different generative processes that could have generated the input. **Likelihood**

What is likelihood?

$\mathbf{L}_{\mathbf{Hx}}(\text{Data}) = \mathbf{P}(\text{Observing the data} \mid \text{Hyp}_X \text{ is TRUE})$

“How likely am I to observe the data if I assume that Hypothesis$_X$ is true”

What is likelihood?

Consider the hypothesis (H1):

- There are three categories in the world: A, B and C
- All three categories have 20 members:
  - $A = \{A_1, A_2 \ldots . A_{20}\}$, $B = \{B_1, B_2 \ldots . B_{20}\}$, $C = \{C_1, C_2 \ldots . C_{20}\}$
- The sampling process is random

What is:

- $\mathbf{L_{H1}}(\{A_1\})$
- $\mathbf{L_{H1}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H1}}(\{A_{18}, C_6, D_{10}\})$

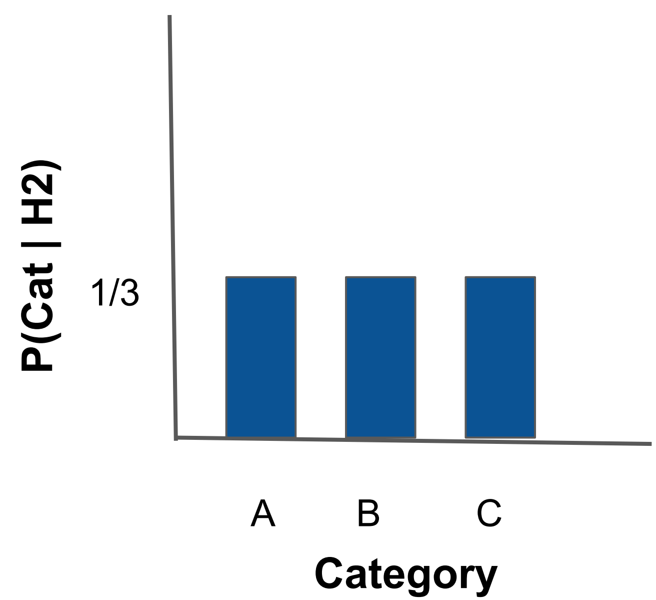
What is likelihood?

Consider the hypothesis (H1):

- There are three categories in the world: A, B and C
- All three categories have 20 members:
  - $A = \{A_1, A_2 \ldots. A_{20}\}$, $B = \{B_1, B_2 \ldots. B_{20}\}$, $C = \{C_1, C_2 \ldots. C_{20}\}$
- The sampling process is random

What is:

- $\mathbf{L_{H1}}(\{A_1\})$
- $\mathbf{L_{H1}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H1}}(\{A_{18}, C_6, D_{10}\})$



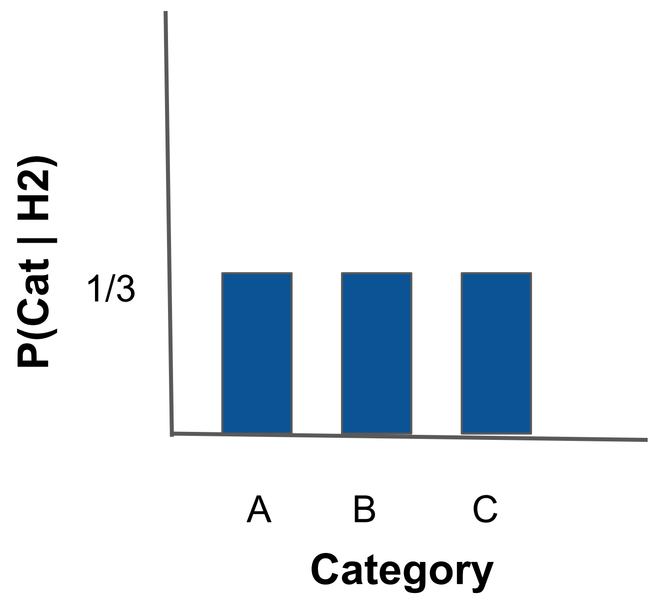
What is likelihood?

Consider the hypothesis (H1):

- There are three categories in the world: A, B and C
- All three categories have 20 members:
  - $A = \{A_1, A_2 \ldots.\ A_{20}\}$, $B = \{B_1, B_2 \ldots.\ B_{20}\}$, $C = \{C_1, C_2 \ldots.\ C_{20}\}$
- The sampling process is random

What is:

- $\mathbf{L_{H1}}(\{A_1\})$ $1/3$
- $\mathbf{L_{H1}}(\{A_2, B_5, B_2, C_{20}\})$ $(1/3)^4$
- $\mathbf{L_{H1}}(\{A_{18}, C_6, D_{10}\})$ 0



What is likelihood?

Consider the hypothesis (H1):

- There are three categories in the world: A, B and C
- All three categories have 20 members:
  - $A = \{A_1, A_2 \ldots . A_{20}\}$, $B = \{B_1, B_2 \ldots . B_{20}\}$, $C = \{C_1, C_2 \ldots . C_{20}\}$
- The sampling process is random

What is:

- $\mathbf{L_{H1}}(\{A_1\})$ 1/3
- $\mathbf{L_{H1}}(\{A_2, B_5, B_2, C_{20}\})$ $(1/3)^4$
- $\mathbf{L_{H1}}(\{A_{18}, C_6, D_{10}\})$ 0

$$P(x \mid H) = \frac{1}{|H|} \text{ if } x \in H \text{, 0 otherwise}$$

$$P(X \mid H) = \prod_{x \in X} P(x \mid H)$$

What is likelihood?

Consider the hypothesis (H2):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 \ldots A_{50}\}$, $B = \{B_1, B_2 \ldots B_{25}\}$, $C = \{C_1, C_2 \ldots C_{25}\}$
- The sampling process is random.

What is:

- $\mathbf{L_{H2}}(\{A_1\})$
- $\mathbf{L_{H2}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H2}}(\{A_{18}, C_6, D_{10}\})$

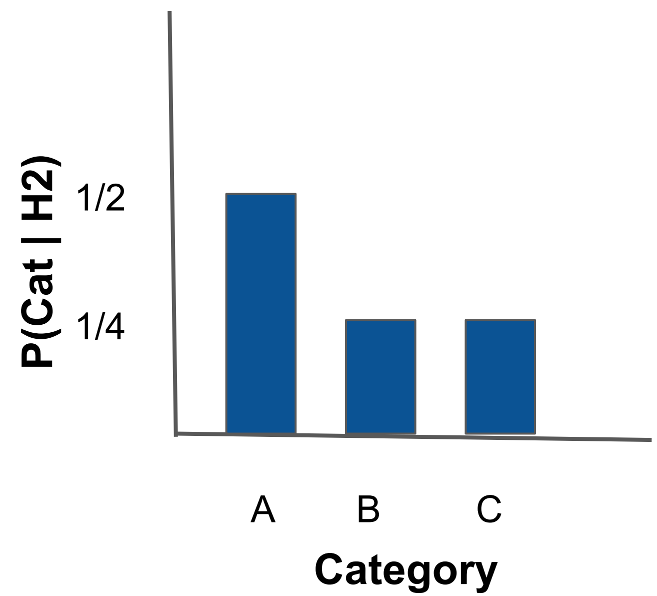
What is likelihood?

Consider the hypothesis (H2):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 \ldots A_{50}\}$, $B = \{B_1, B_2 \ldots B_{25}\}$, $C = \{C_1, C_2 \ldots C_{25}\}$
- The sampling process is random.

What is:

- $\mathbf{L_{H2}}(\{A_1\})$
- $\mathbf{L_{H2}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H2}}(\{A_{18}, C_6, D_{10}\})$



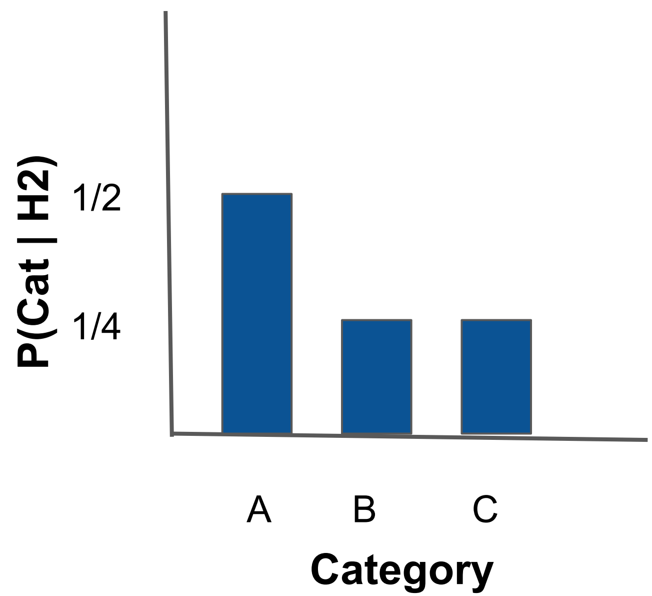
What is likelihood?

Consider the hypothesis (H2):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 \ldots A_{50}\}$, $B = \{B_1, B_2 \ldots B_{25}\}$, $C = \{C_1, C_2 \ldots C_{25}\}$
- The sampling process is random.

What is:

- $\mathbf{L_{H2}}(\{A_1\})$ 1/2
- $\mathbf{L_{H2}}(\{A_2, B_5, B_2, C_{20}\})$ $1/2 * (1/4)^3$
- $\mathbf{L_{H2}}(\{A_{18}, C_6, D_{10}\})$ 0



What is likelihood?

Consider the hypothesis (H3):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 .... A_{20}\}$, $B = \{B_1, B_2 .... B_{20}\}$, $C = \{C_1, C_2 .... C_{20}\}$
- The sampling process is biased, such that A is half as likely to be picked as B or C

What is:

- $\mathbf{L_{H3}}(\{A_1\})$
- $\mathbf{L_{H3}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H3}}(\{A_{18}, C_6, D_{10}\})$

What is likelihood?

Consider the hypothesis (H3):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 \ldots . A_{20}\}$, $B = \{B_1, B_2 \ldots . B_{20}\}$, $C = \{C_1, C_2 \ldots . C_{20}\}$
- The sampling process is biased, such that A is half as likely to be picked as B or C

What is:

- $\mathbf{L_{H3}}(\{A_1\})$
- $\mathbf{L_{H3}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H3}}(\{A_{18}, C_6, D_{10}\})$

$P(B) = P(C) = X$

$P(A) = X/2$

$P(A) + P(B) + P(C) = 1$

$2X + X/2 = 1$

$X = 2/5 = P(B) = P(C)$

$P(A) = 1/5$

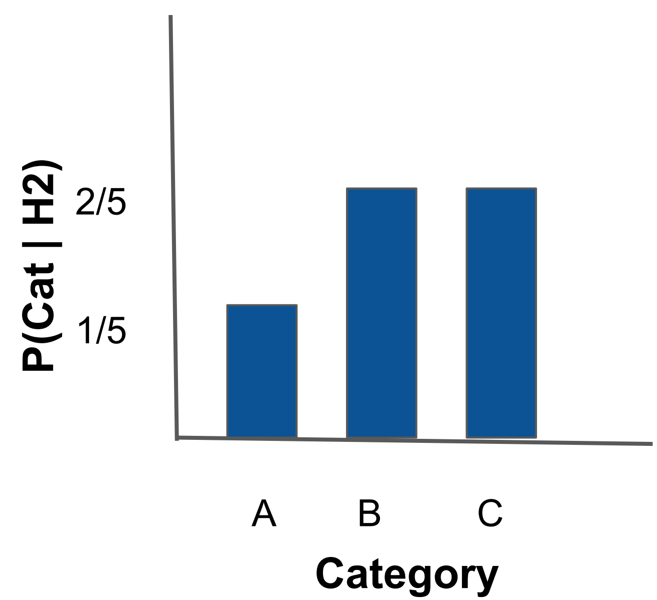
What is likelihood?

Consider the hypothesis (H3):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 \ldots. A_{20}\}$, $B = \{B_1, B_2 \ldots. B_{20}\}$, $C = \{C_1, C_2 \ldots. C_{20}\}$
- The sampling process is biased, such that A is half as likely to be picked as B or C

What is:

- $\mathbf{L_{H3}}(\{A_1\})$
- $\mathbf{L_{H3}}(\{A_2, B_5, B_2, C_{20}\})$
- $\mathbf{L_{H3}}(\{A_{18}, C_6, D_{10}\})$



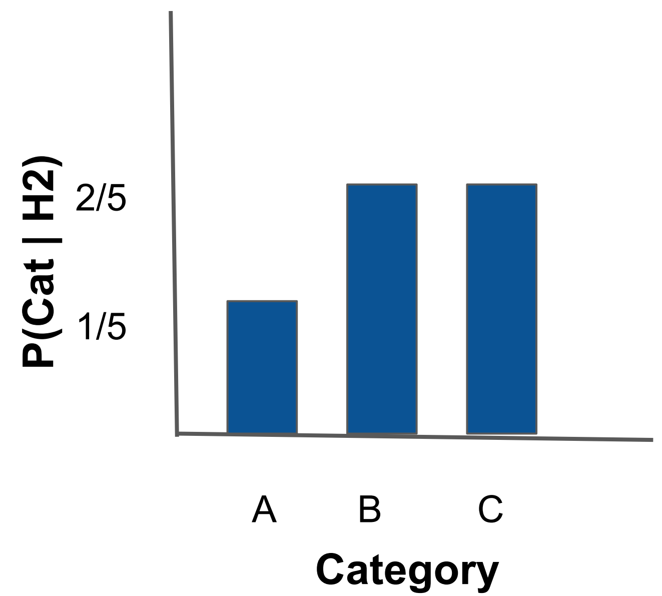
What is likelihood?

Consider the hypothesis (H3):

- There are three categories in the world: A, B and C
- $A = \{A_1, A_2 .... A_{20}\}$, $B = \{B_1, B_2 .... B_{20}\}$, $C = \{C_1, C_2 .... C_{20}\}$
- The sampling process is biased, such that A is half as likely to be picked as B or C

What is:

- $\mathbf{L_{H3}}(\{A_1\})$ 1/5
- $\mathbf{L_{H3}}(\{A_2, B_5, B_2, C_{20}\})$ $1/5 * (2/5)^3$
- $\mathbf{L_{H3}}(\{A_{18}, C_6, D_{10}\})$ 0



What is likelihood?

Which of the three hypotheses maximizes the likelihood of:

- $\{A_1\}$
- $\{A_2, B_5, B_2, C_{20}\}$

What is likelihood?

Which of the three hypotheses maximizes the likelihood of:

- $\{A_1\}$ H_2 (0.5)
- $\{A_2, B_5, B_2, C_{20}\}$ H_3 (0.128)

What is likelihood?

$\mathbf{L}_{\mathbf{Hx}}(\text{Data}) = \mathbf{P}(\text{Observing the data} \mid \text{Hyp}_X \text{ is TRUE})$

"How likely am I to observe the data if I assume that Hypothesis$_X$ is true"

The likelihood of observing the data if hypothesis is true, is affected by our assumptions about:

- The distribution of data in the world under Hypothesis$_X$
- The sampling process under Hypothesis$_X$

What is prior probability?

It is the probability of each of our hypotheses being true before we seeing any data

What are some possible sources of information on which we can base our priors?

What is prior probability?

It is the probability of each of our hypotheses being true before we seeing any data

What are some possible sources of information on which we can base our priors?

- Based on frequency — i.e. how often in the past were each of the hypotheses true?
- Based on world knowledge (e.g., you know that the sampling process is random because you wrote the code for it)
- Based on innate knowledge/ preferences

Bayesian inference

Posterior

Likelihood

Prior

$$P(R_i = A \mid i) = \frac{P(i \mid R_i = A)P(R_i = A)}{P(i \mid R_i = A)P(R_i = A) + P(i \mid R_i = B)P(R_i = B)}$$

Evidence (assuming there are only two categories in the world) — i.e. P(Data)

Bayesian inference

Posterior

Likelihood

Prior

$$P(R_i = A \mid i) = \frac{P(i \mid R_i = A)P(R_i = A)}{P(i \mid R_i = A)P(R_i = A) + P(i \mid R_i = B)P(R_i = B)}$$

Evidence (assuming there are only two categories in the world) — i.e. P(Data)

Why do we usually ignore the evidence term?

Bayesian word learning

https://tommccoy1.github.io/techviz/bayes/birds_intro.html

Questions to think about:

- How is the prior probability calculated?
- How is the likelihood calculated?
- When do certain hypotheses get discarded?
- What happens if the learner assumes that the sampling procedure is unbiased, but you give them a biased sample?

The dress

https://tommccoy1.github.io/techviz/bayes/dress_intro.html

Questions to think about:

- When is the prior probability the same, but the likelihood different?
- When is the likelihood the same, but prior probability different?

The dress

https://tommccoy1.github.io/techviz/bayes/dress_intro.html

Questions to think about:

- When is the prior probability the same, but the likelihood different?

  Same person, different settings of the slider

- When is the likelihood the same, but prior probability different?

  Different person, same settings of the slider

Bayesian belief update

$$Posterior\ probabilty = \frac{Likelihood * Prior}{Evidence}$$

$$Posterior_i = \frac{Likelihood * Posterior_{i-1}}{Evidence} \rightarrow Posterior_{i+1} = \frac{Likelihood * Posterior_i}{Evidence}$$

We update our beliefs based on what we see in the environment