Probabilistic models 1 – The language of thought and Bayesian models

Reading: Piantadosi et al.: The logical primitives of thought

Jakub Dotlačil

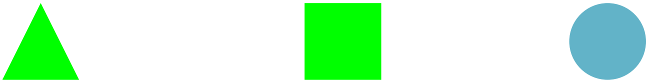
Cognitive Modeling, 2019-2020

This class

- How are concepts/morphemes learned and stored?

  We will focus on rule-based learning and a symbolic system (the Language of Thought)
- How can logic, linguistics and a probability theory be combined to address the question of concept learning?

Motivation



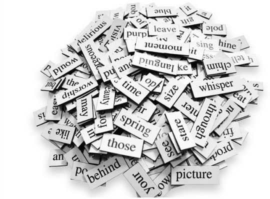
- *wudsy* and pointing to the figure in the middle
- Does *wudsy* expresses the concept 'green,' 'square,' 'green square,' the middle element...?

Motivation

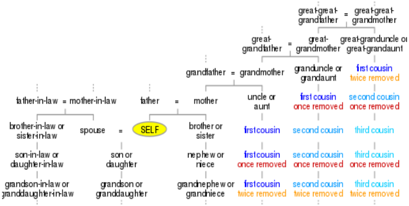
- How are kids to find out what concepts are denoted by words/morphemes?
- Word vs. morpheme:
- tall
- tall-est
- Excluding a few exceptions (cran-berry) we are interested in the meaning of morphemes

Motivation

- Concepts as a heap



- Concepts structured



(Relations)

Motivation

- Concepts as a heap

- Concepts structured

(Numbers)

Motivation

- Concepts as a heap

- Concepts structured

Quantifiers: some, most, all, more than three, at least five…

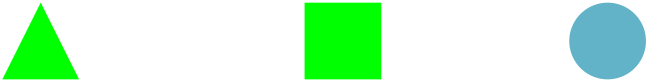
Concept learning/storing

- Engineering: Neural network models, mainly hand in hand with visual cognition (Ronneberger et al., 2015)
- Psychology and linguistics: Prototype theory, conceptual spaces (Berlin and Kay, 1969; Gärdenfors, 2004; Lakoff, 1987)
- Linguistics and applied logic: natural language semantics with formal systems (e.g., first-order logic, higher order logics) (Montague, 1973)

We will look at the combination of applied logic (formal semantics) with cognitive modeling

1. Probability and conditional probability
2. Language of thought and concept learning
3. Concept learning: Experiment
4. Concept learning: Modeling
5. Word acquisition

Probability



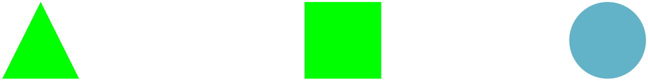
- $P(\blacksquare) = P(\bullet) = P(\blacktriangle) = \frac{1}{3}$
- $P(\blacksquare \cup \bullet) = P(\blacksquare) + P(\bullet) = \frac{2}{3}$ etc.

Probability of space of events S

- $P(S) = 1, P(\emptyset) = 0$
- If events $A_1, A_2, \ldots$ are independent, then

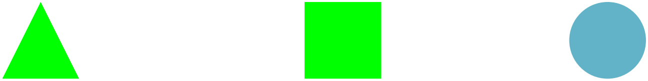
$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$$

Probability



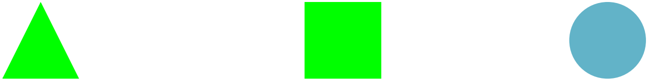
- We hear ‘green …’ What do we conclude?

Probability



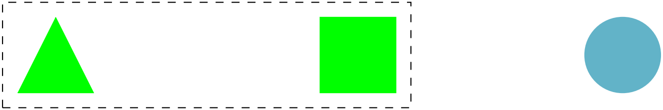
- We hear 'green ...' What do we conclude?
- $P(\blacksquare|green) =?$, $P(\bullet|green) =?$ etc.
- $P(A|B)$ – the probability of A given B
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Probability



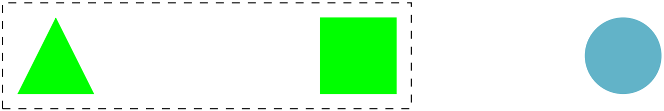
- We hear 'green …' What do we conclude?
- $P(\blacksquare|green) = ?$, $P(\bullet|green) = ?$ etc.
- $P(A|B)$ – the probability of A given B
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(\blacksquare|green) = \frac{P(\blacksquare \cap green)}{P(green)}$

Probability



- We hear 'green …' What do we conclude?
- $P(\blacksquare | green) = \frac{P(\blacksquare \cap green)}{P(green)} = \frac{1}{2}$
- $P(\blacktriangle | green) = \frac{P(\blacktriangle \cap green)}{P(green)} = \frac{1}{2}$

Probability



- We hear 'green …' What do we conclude?
- $P(\blacksquare|green) = \frac{P(\blacksquare \cap green)}{P(green)} = \frac{1}{2}$
- $P(\blacktriangle|green) = \frac{P(\blacktriangle \cap green)}{P(green)} = \frac{1}{2}$

Things to note:

- $P(A|B)$ – the probability of A given B (what information does observing event B provide about event A?)
- Conditional probability is a probability ($P(\cdot|B)$ forms a probability space)
- In fact, we can think of any probability as conditional probability $P(A) = P(A|S)$

Quick exercise on conditional probabilities

Two girls' paradox:

1. Mr. Jones has two children. The older child is a girl (B). What is the probability that both children are girls (A)?
2. Mr. Jones has two children. At least one of them is a girl (C). What is the probability that both children are girls (A)?

Quick exercise on conditional probabilities

Two girls' paradox:

1. Mr. Jones has two children. The older child is a girl (B). What is the probability that both children are girls (A)?
2. Mr. Jones has two children. At least one of them is a girl (C). What is the probability that both children are girls (A)?

Some assumptions – $P(boy) = P(girl) = \frac{1}{2}$, the sex of one kid does not influence the sex of the other kid

Quick exercise on conditional probabilities

Two girls' paradox:

1. Mr. Jones has two children. The older child is a girl (B). What is the probability that both children are girls (A)?
2. Mr. Jones has two children. At least one of them is a girl (C). What is the probability that both children are girls (A)?

Some assumptions – $P(boy) = P(girl) = \frac{1}{2}$, the sex of one kid does not influence the sex of the other kid

- events: $\{bb, bg, gb, gg\}$

Quick exercise on conditional probabilities

Two girls' paradox:

1. Mr. Jones has two children. The older child is a girl (B). What is the probability that both children are girls (A)?
2. Mr. Jones has two children. At least one of them is a girl (C). What is the probability that both children are girls (A)?

Some assumptions – $P(boy) = P(girl) = \frac{1}{2}$, the sex of one kid does not influence the sex of the other kid

- events: $\{bb, bg, gb, gg\}$
- $P(A|B) = \frac{P(\{gg\})}{P(\{gb,gg\})} = \frac{1}{2}$
- $P(A|C) = \frac{P(\{gg\})}{P(\{bg,gb,gg\})} = \frac{1}{3}$

Bayes' rule and the law of total probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Bayes' rule and the law of total probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ (Bayes' rule)

Bayes' rule and the law of total probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ (Bayes' rule)
- $P(B) = \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i)$ (LOTP)

Bayes' rule and the law of total probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ (Bayes' rule)
- $P(B) = \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i)$ (LOTP)
- $P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$ (Bayes' rule+LOTP)

Bayes' rule and coin flipping

- One fair coin ($P(H) = \frac{1}{2}$), one coin which lands Heads with $P(H) = \frac{3}{4}$.
- You pick one coin randomly and land H three times.
- What is the probability that you picked a fair coin?

Bayes' rule and coin flipping

- One fair coin ($P(H) = \frac{1}{2}$), one coin which lands Heads with $P(H) = \frac{3}{4}$.
- You pick one coin randomly and land H three times.
- What is the probability that you picked a fair coin?

$$P(F|A) = \frac{P(A|F)P(F)}{P(A)}$$

Bayes' rule and coin flipping

- One fair coin ($P(H) = \frac{1}{2}$), one coin which lands Heads with $P(H) = \frac{3}{4}$.
- You pick one coin randomly and land H three times.
- What is the probability that you picked a fair coin?

$$\begin{aligned} P(F|A) &= \frac{P(A|F)P(F)}{P(A)} \\ &= \frac{P(A|F)P(F)}{P(A|F)P(F)+P(A|NF)P(NF)} \end{aligned}$$

Bayes' rule and coin flipping

- One fair coin ($P(H) = \frac{1}{2}$), one coin which lands Heads with $P(H) = \frac{3}{4}$.
- You pick one coin randomly and land H three times.
- What is the probability that you picked a fair coin?

$$
\begin{aligned}
P(F|A) &= \frac{P(A|F)P(F)}{P(A)} \\
&= \frac{P(A|F)P(F)}{P(A|F)P(F)+P(A|NF)P(NF)} \\
&= \frac{(1/2)^3 \cdot 1/2}{}
\end{aligned}
$$

Bayes' rule and coin flipping

- One fair coin ($P(H) = \frac{1}{2}$), one coin which lands Heads with $P(H) = \frac{3}{4}$.
- You pick one coin randomly and land H three times.
- What is the probability that you picked a fair coin?

$$
\begin{aligned}
P(F|A) &= \frac{P(A|F)P(F)}{P(A)} \\
&= \frac{P(A|F)P(F)}{P(A|F)P(F)+P(A|NF)P(NF)} \\
&= \frac{(1/2)^3 \cdot 1/2}{(1/2)^3 \cdot 1/2+(3/4)^3 \cdot 1/2} \\
&= 0.23
\end{aligned}
$$

Bayes' rule and the law of total probability

- $P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$ (Bayes' rule+LOTP)
- $P(A|B)$ – the posterior probability
- $P(A)$ – the prior probability
- $P(B|A)$ – the likelihood
- $P(B)$ – the evidence

Probability distributions

Random variable:

- A random variable is a function from S (sample space) to the real numbers R.

Quick exercise:

- You throw a coin twice. Possible outcomes: $S = \{HH, HT, TH, TT\}$
- Random variable X (number of heads):
  $X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0$

When X is discrete – we are interested in **probability mass function** of X

Probability distributions

Random variable:

- A random variable is a function from S (sample space) to the real numbers R.

Quick exercise:

- You throw a coin twice. Possible outcomes: $S = \{HH, HT, TH, TT\}$
- Random variable X (number of heads):

  $X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0$

When X is discrete – we are interested in **probability mass function** of X

$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{1}{2}, P(X = 2) = \frac{1}{4}$

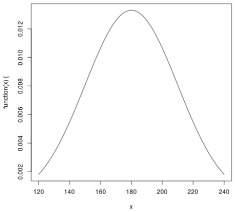
Probability distributions

Random variable:

- A random variable is a function from S (sample space) to the real numbers R.

When X is continuous – we are interested in **probability density function** of X

E.g., say X is the height of a person in USA



$P(X \geq 162) = 0.5, P(X \geq 190) = 0.003$

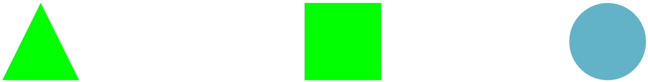
Probability distributions

Some commonly used distributions have a special name:

- Normal distribution (probability density function) $X \sim Norm(\mu, \sigma^2)$
- Bernoulli distribution (probability mass function) $X \sim Bern(p)$

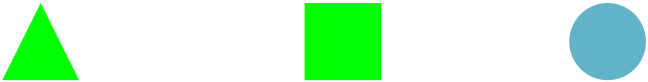
Bernoulli distribution $Bern(p)$: $P(X = 1) = p, P(X = 0) = 1 - p$

Learning concepts as posterior probabilities



- Let's say you hear *wudsy* and see pointing to the figure in the middle
- Does it correspond to 'green,' to 'square,' to 'green square,' to the middle element...?
- h – hypothesis about meaning of *wudsy*
- $P(h)$
- $P(h|I)$ – I - novel information
- We will now explore h. Once we know how to formulate h and the corresponding $P(h)$, $P(h|I)$ follows from Bayes' rule.

Learning concepts as posterior probabilities



- Let's say you hear *wudsy* and see pointing to the figure in the middle
- Does it correspond to 'green', to 'square', to 'green square', to the middle element...?
- h – hypothesis about meaning of *wudsy*
- $P(h)$
- $P(h|I)$ – I - novel information
- We will now explore h. Once we know how to formulate h and the corresponding $P(h)$, $P(h|I)$ follows from Bayes' rule.

$P(\text{‘}cone'|I)$

Language of thought
Fodor, 1975

- A set of computational primitives
- The primitives can be combined into more complex objects by rule-based mechanisms

Language of thought for ‘propositional’ language

CFG:

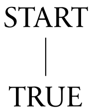
1. START $\rightarrow$ P
2. P $\rightarrow$ P $\wedge$ Q
3. P $\rightarrow$ P $\vee$ Q
4. P $\rightarrow$ TRUE
5. P $\rightarrow$ FALSE
6. P $\rightarrow$ COLOR(o)
7. P $\rightarrow$ SIZE(o)
8. COLOR $\rightarrow$ blue|green|yellow
9. SIZE $\rightarrow$ size1|size2|size3

Language of thought for ‘propositional’ language

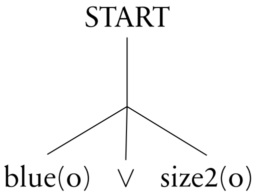
CFG:

1. START $\rightarrow$ P
2. P $\rightarrow$ P $\wedge$ Q
3. P $\rightarrow$ P $\vee$ Q
4. P $\rightarrow$ TRUE
5. P $\rightarrow$ FALSE
6. P $\rightarrow$ COLOR(o)
7. P $\rightarrow$ SIZE(o)
8. COLOR $\rightarrow$ blue|green|yellow
9. SIZE $\rightarrow$ size1|size2|size3

wudsy:



wudsy:

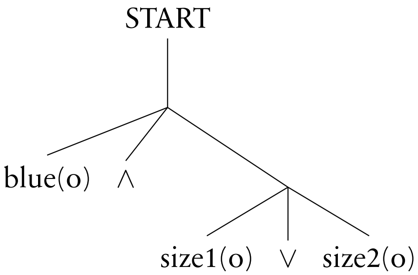


Language of thought for ‘propositional’ language

CFG:

1. START $\rightarrow$ P
2. P $\rightarrow$ P $\wedge$ Q
3. P $\rightarrow$ P $\vee$ Q
4. P $\rightarrow$ TRUE
5. P $\rightarrow$ FALSE
6. P $\rightarrow$ COLOR(o)
7. P $\rightarrow$ SIZE(o)
8. COLOR $\rightarrow$ blue|green|yellow
9. SIZE $\rightarrow$ size1|size2|size3

wudsy:

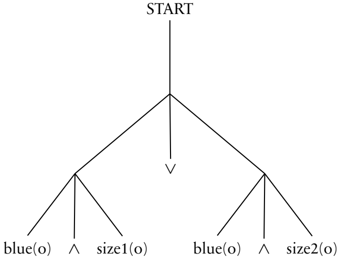


Language of thought for ‘propositional’ language

CFG:

1. START $\rightarrow$ P
2. P $\rightarrow$ P $\wedge$ Q
3. P $\rightarrow$ P $\vee$ Q
4. P $\rightarrow$ TRUE
5. P $\rightarrow$ FALSE
6. P $\rightarrow$ COLOR(o)
7. P $\rightarrow$ SIZE(o)
8. COLOR $\rightarrow$ blue|green|yellow
9. SIZE $\rightarrow$ size1|size2|size3

wudsy:



Probabilistic language of thought

PCFG:

1. START $\rightarrow$ P (1.0)
2. P $\rightarrow$ P $\wedge$ Q ($\frac{1}{6}$)
3. P $\rightarrow$ P $\vee$ Q ($\frac{1}{6}$)
4. P $\rightarrow$ TRUE ($\frac{1}{6}$)
5. P $\rightarrow$ FALSE ($\frac{1}{6}$)
6. P $\rightarrow$ COLOR(o) ($\frac{1}{6}$)
7. P $\rightarrow$ SIZE(o) ($\frac{1}{6}$)
8. COLOR $\rightarrow$ blue|green|yellow ($\frac{1}{3}$)
9. SIZE $\rightarrow$ size1|size2|size3 ($\frac{1}{3}$)

Probabilistic language of thought

wudsy: $P(h) = \frac{1}{6}$

START

TRUE

wudsy: $P(h) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{1944}$

START

blue(o) $\vee$ size2(o)

More complex hypotheses are penalized (they receive lower prior probability)

Learning in the language of thought

- $P(h|l) = \frac{P(l|h) \cdot P(h)}{P(l)}$

PCFG:

1. START $\rightarrow$ P (1.0)
2. P $\rightarrow$ $\neg$ P $(\frac{1}{3})$
3. P $\rightarrow$ TRUE $(\frac{1}{3})$
4. P $\rightarrow$ FALSE $(\frac{1}{3})$

wudsy:

S1	S2	S3	S4
TRUE	FALSE	$\neg$FALSE	$\neg$TRUE

Learning in the language of thought

- $P(h|l) = \frac{P(l|h) \cdot P(h)}{P(l)}$

PCFG:

1. START $\rightarrow$ P (1.0)
2. P $\rightarrow \neg$ P $(\frac{1}{3})$
3. P $\rightarrow$ TRUE $(\frac{1}{3})$
4. P $\rightarrow$ FALSE $(\frac{1}{3})$

wudsy:

S1	S2	S3	S4
TRUE	FALSE	$\neg$FALSE	$\neg$TRUE

Suppose L=TRUE

- $P(S1|L) = \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = \frac{3}{4}$

Learning in the language of thought

- $P(h|l) = \frac{P(l|h) \cdot P(h)}{P(l)}$

PCFG:

1. START $\rightarrow$ P $\quad (1.0)$
2. P $\rightarrow \neg$ P $\quad (\frac{1}{3})$
3. P $\rightarrow$ TRUE $\quad (\frac{1}{3})$
4. P $\rightarrow$ FALSE $\quad (\frac{1}{3})$

wudsy:

S1	S2	S3	S4
TRUE	FALSE	¬FALSE	¬TRUE

Suppose L=TRUE

- $P(S1|L) = \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = \frac{3}{4}$
- $P(S2|L) = \frac{0 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = 0$

Learning in the language of thought

- $P(h|l) = \frac{P(l|h) \cdot P(h)}{P(l)}$

PCFG:

1. START $\rightarrow$ P (1.0)
2. P $\rightarrow \neg$ P $(\frac{1}{3})$
3. P $\rightarrow$ TRUE $(\frac{1}{3})$
4. P $\rightarrow$ FALSE $(\frac{1}{3})$

wudsy:

S1	S2	S3	S4
TRUE	FALSE	¬FALSE	¬TRUE

Suppose L=TRUE

- $P(S1|L) = \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = \frac{3}{4}$
- $P(S2|L) = \frac{0 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = 0$
- $P(S3|L) = \frac{1 \cdot 1/9}{1 \cdot 1/3 + 1 \cdot 1/9} = \frac{1}{4}$
- $P(S4|L) = \frac{0 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/9} = 0$

Likelihood in the model

- $P(l = L|h) = 1$ if h derives L, 0 otherwise
- $f(h) = 1$ if h derives L
- $P(l|h)$ – Bernoulli$(f(h))$
- Alternatives – different parameter for Bernoulli:
  $f(h) = \alpha$ if h derives L
- $P(l|h)$ = Bernoulli$(f(h))$
- Set $\alpha = 0.9$

Likelihood in the model

- $P(l = L|h) = 1$ if h derives L, 0 otherwise
- $f(h) = 1$ if h derives L
- $P(l|h)$ – Bernoulli($f(h)$)
- Alternatives – different parameter for Bernoulli:
  $f(h) = \alpha$ if h derives L
- $P(l|h)$ = Bernoulli($f(h)$)
- Set $\alpha = 0.9$

Suppose L=TRUE

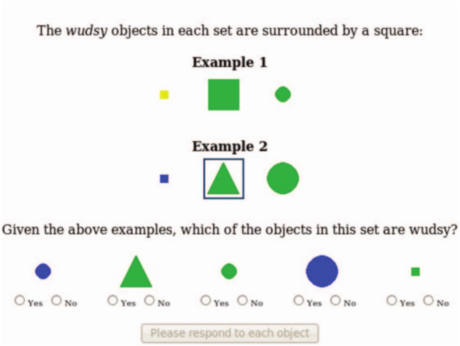
- $P(S1|L) = \frac{0.9 \cdot 1/3}{0.9 \cdot 1/3 + 0.1 \cdot 1/3 + 0.9 \cdot 1/9 + 0.1 \cdot 1/9} = 0.675$
- $P(S2|L) = \frac{0.1 \cdot 1/3}{0.9 \cdot 1/3 + 0.1 \cdot 1/3 + 0.9 \cdot 1/9 + 0.1 \cdot 1/9} = 0.075$
- $P(S3|L) = \frac{0.9 \cdot 1/9}{0.9 \cdot 1/3 + 0.1 \cdot 1/3 + 0.9 \cdot 1/9 + 0.1 \cdot 1/9} = 0.225$
- $P(S4|L) = \frac{0.1 \cdot 1/3}{0.9 \cdot 1/3 + 0.1 \cdot 1/3 + 0.9 \cdot 1/9 + 0.1 \cdot 1/9} = 0.025$

Things to note

- LOT penalizes complex hypotheses
- Learning (via Bayesian update) shifts probabilities to accommodate novel information
- The update by the information can be noisy
- Due to coherency of Bayes' rule, the update can be sequential or in one swoop

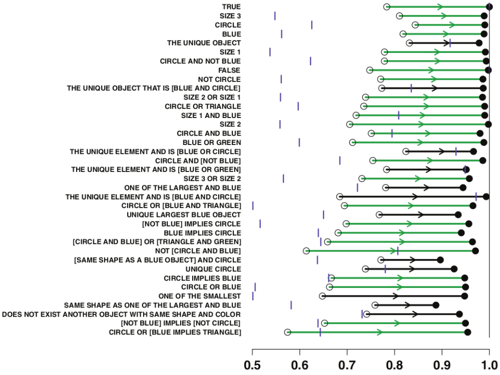
How does the probabilistic model of LOT fit human data?

- Experiment from Piantadosi et al. (2016)



How does the probabilistic model of LOT fit human data?

- Experiment from Piantadosi et al. (2016)



- The model should predict the path of the human learning of individual concepts

Bayesian model structure of cognitive model

$$
\begin{tikzcd}
\mathrm{DirichletDist}(D_{**}) \arrow[dr] & G \arrow[d] & \\
& \mathbf{h} \arrow[d] & \alpha \arrow[dl] \\
& \mathrm{Bernoulli}(f(h, \alpha)) &
\end{tikzcd}
$$

Details about the model: probability distributions

Bernoulli distribution:

- $X \sim Bern(p)$ – response 1 appears with probability p; response 0 appears with probability $1 - p$
- $Bern(f(h, \alpha))$
- $f(h, \alpha)$ – α if h returns x
- $1 - \alpha$ if h returns $y \neq x$

Details about the model: probability distributions

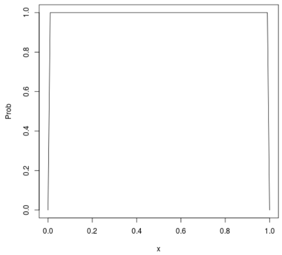
Dirichlet Distribution:

- $p \sim DirichletDist(v)$
- Suppose we only have two rules:
  1. P $\rightarrow$ FALSE
  2. P $\rightarrow$ TRUE

P(Rule2) = p

P(Rule1) = 1-p

p based on DirichletDist(1,1)



Details about the model: probability distributions

Dirichlet Distribution:

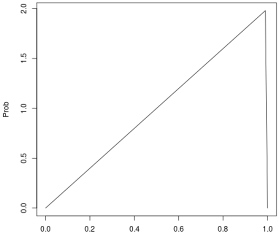
- $p \sim DirichletDist(v)$
- Suppose we only have two rules:
    1. P $\rightarrow$ FALSE
    2. P $\rightarrow$ TRUE

P(Rule2) = p

P(Rule1) = 1-p

p based on DirichletDist(1,1) after Rule2 was used;

DirichletDist(1,2)



Details about the model: probability distributions

Dirichlet Distribution:

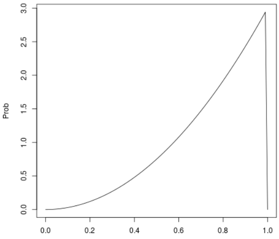
- $p \sim DirichletDist(v)$
- Suppose we only have two rules:
  1. P $\rightarrow$ FALSE
  2. P $\rightarrow$ TRUE

P(Rule2) = p

P(Rule1) = 1-p

p based on DirichletDist(1,1) after Rule2 was used again;

DirichletDist(1,3)



Details about the model: probability distributions

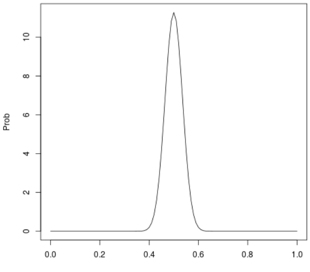
Dirichlet Distribution:

- $p \sim DirichletDist(v)$
- Suppose we only have two rules:
  1. P $\rightarrow$ FALSE
  2. P $\rightarrow$ TRUE

P(Rule2) = p

P(Rule1) = 1-p

p based on DirichletDist(100,100)



Details about the model: probability distributions

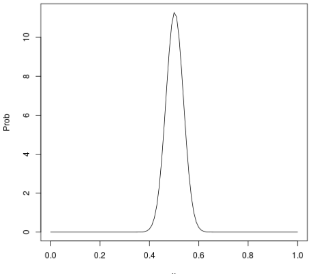
Dirichlet Distribution:

- $p \sim DirichletDist(v)$
- Suppose we only have two rules:
  1. P $\rightarrow$ FALSE
  2. P $\rightarrow$ TRUE

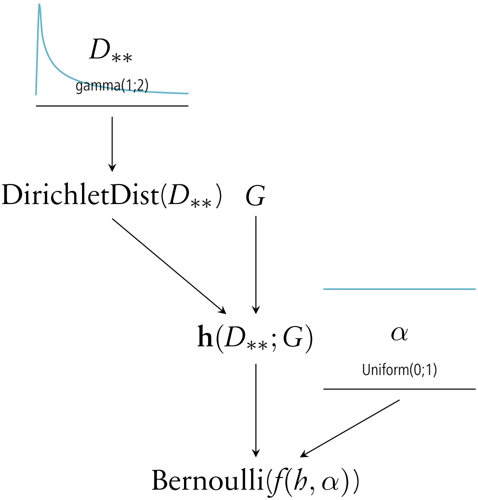
P(Rule2) = p

P(Rule1) = 1-p

p based on DirichletDist(100,100) after Rule2 was used



Bayesian model structure for data analysis



The model is not solved analytically; rather, we sample from posterior and inspect the samples (MCMC techniques) (Kruschke, 2011) (cf. exercise)

Bayesian modeling: details

- The 2nd Bayesian model is trained on part of the data
- The cognitive model is evaluted on held-out data by using the maximum point estimates for parameters D_{**}, α

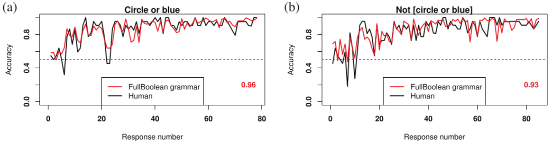
Bayesian modeling: details

- The 2nd Bayesian model is trained on part of the data
- The cognitive model is evaluted on held-out data by using the maximum point estimates for parameters D_{**}, α
- The model fit could be evaluated directly on the same data on which it was trained
- However, model comparison requires different training and testing data sets

Bayesian modeling: results I

Results on selected held-out datasets

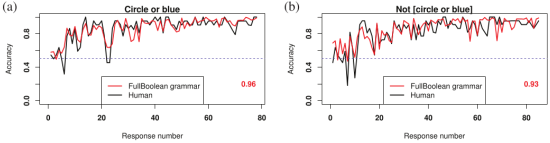
- (a) and (b) as cases of successful fit to human data



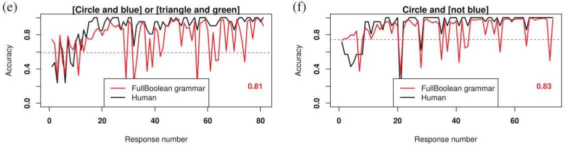
Bayesian modeling: results I

Results on selected held-out datasets

- (a) and (b) as cases of successful fit to human data



- (e) and (f) as cases of successful fit to human data



Bayesian modeling: results II

- Simple Boolean (similar to the one presented on previous slides)
- Full Boolean (Simple Boolean, Biconditional and Simple conditional)
- Horn clauses (all expressions conjunctions of Horn clauses)
- DNF (disjunctive normal form – disjunctions of conjunctions)
- CNF (conjunctive normal form – conjunctions of disjunctions)
- NAND (the only primitive is not-and $(P \uparrow Q = \neg(P \wedge Q))$)
- NOR (the only primitive is not-or $(P \downarrow Q = \neg(P \vee Q))$)

Bayesian modeling: results II

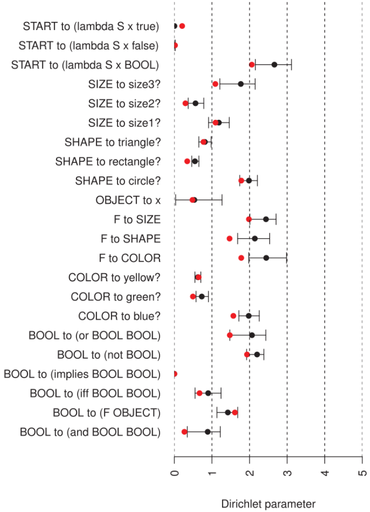
- The fit of the model to the held-out human data

Grammar	H.O.LL	FP	$R^2_{response}$	R^2_{mean}
FullBoolean	−16296.84	27	.88	.60
Biconditional	−16305.13	26	.88	.64
CNF	−16332.39	26	.89	.69
DNF	−16343.87	26	.89	.66
SimpleBoolean	−16426.91	25	.87	.70
Implies	−16441.29	26	.87	.70
HornClause	−16481.90	27	.87	.65
NAND	−16815.60	24	.84	.61
NOR	−16859.75	24	.85	.58

H.O.L.L. - log-likelihood on held-out data, FP - # of free params, R^2 - correlation between model and human responses

- According to H.O.L.L., FullBoolean is the best model

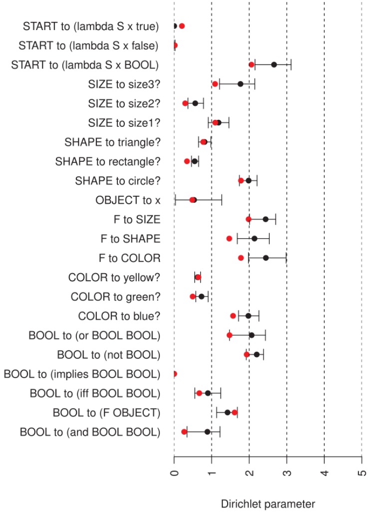
Bayesian modeling: results III



For any left symbol, all expansions form a probability space
e.g., $P(F \rightarrow SIZE|SHAPE|COLOR) \approx \frac{1}{3}$

red dot – max. a posteriori estimate
bars – 95% HPD

Bayesian modeling: results III



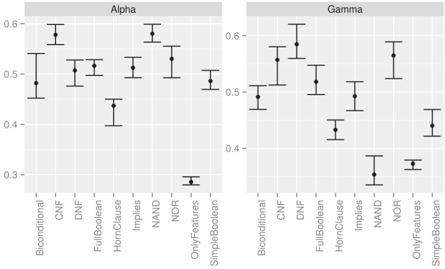
red dot – max. a posteriori estimate
bars – 95% HPD

For any left symbol, all expansions form a probability space
e.g., $P(F \to SIZE|SHAPE|COLOR) \approx \frac{1}{3}$

Numbers close to 1 – reusing rules;
due to prior or data modeled?
why is MAP outside of HDP?
some rules – very low probability – why?

Bayesian modeling: results IV

- Results of other parameters



- Both α and γ are used for responses
- cf. $$\begin{aligned} P(l = T \mid h, \alpha, \gamma) &= \alpha + (1-\alpha)\cdot\gamma \text{ if } h \text{ generates } T \\ &= (1-\alpha)\cdot(1-\gamma) \text{ if } h \text{ generates } F \end{aligned}$$
- A relatively low value for α

Discussion

- The fit to the model is quite impressive
- How strong evidence do we have that FullBoolean is correct?
- Artificialities of learning (kids' rooms are messy)
- Artificialities of concepts
  Blue $\rightarrow$ circle
- *and*, *or*, *nor*, **nand*; *all*, *some*, *no*, **nall*
- The model could almost as easily learn *nor* as *nand*

Learning counting

- A second take: learning counting
- Several grammars based on Language of Thought

One-knower

```
λ S . (if (singleton? S)
          "one"
          undef)
```

Two-knower

```
λ S . (if (singleton? S)
          "one"
          (if (doubleton? S)
              "two"
              undef))
```

Three-knower

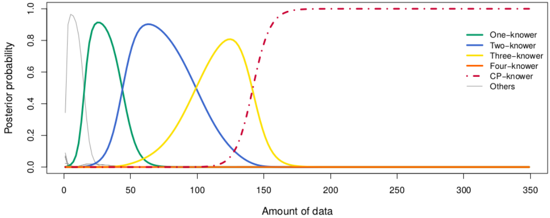
```
λ S . (if (singleton? S)
          "one"
          (if (doubleton? S)
              "two"
              (if (tripleton? S)
              "three"
              undef))
```

CP-knower

```
λ S . (if (singleton? S)
          "one"
          (next (L (set-difference S
                                   (select S)))))
```

Piantadosi: Learning and the language of thought. MIT Thesis.

Learning counting: results



- Recursion acquired after enough evidence accumulated for one, two three, four

Discussion

- Rule-based symbolic models are useful in describing/explaining concept learning
- The models capture adult learning of novel concepts and (less explored) children acquisition

Discussion

- Rule-based symbolic models are useful in describing/explaining concept learning
- The models capture adult learning of novel concepts and (less explored) children acquisition
- Since the compositional/computational system is used beyond word level, the models could form a natural connection between sentence meaning and word meaning

Discussion

- Rule-based symbolic models are useful in describing/explaining concept learning
- The models capture adult learning of novel concepts and (less explored) children acquisition
- Since the compositional/computational system is used beyond word level, the models could form a natural connection between sentence meaning and word meaning
- Interaction between modeling and experimentation

Discussion

- Rule-based symbolic models are useful in describing/explaining concept learning
- The models capture adult learning of novel concepts and (less explored) children acquisition
- Since the compositional/computational system is used beyond word level, the models could form a natural connection between sentence meaning and word meaning
- Interaction between modeling and experimentation
- However, the models are limited:
    - They investigate small sub-domains (numbers, quantifiers,…)

Discussion

- Rule-based symbolic models are useful in describing/explaining concept learning
- The models capture adult learning of novel concepts and (less explored) children acquisition
- Since the compositional/computational system is used beyond word level, the models could form a natural connection between sentence meaning and word meaning
- Interaction between modeling and experimentation
- However, the models are limited:
  - They investigate small sub-domains (numbers, quantifiers,…)
  - They do not study meaning in communication (next lecture)

References I

- Berlin, Brent and Paul Kay (1969). *Basic color terms: Their university and evolution*. California UP.
- Fodor, Jerry A (1975). *The language of thought*. Vol. 5. Harvard university press.
- Gärdenfors, Peter (2004). *Conceptual spaces: The geometry of thought*. MIT press.
- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press/Elsevier.
- Lakoff, George (1987). *Women, fire, and dangerous things*. University of Chicago press.
- Montague, Richard (1973). "The proper treatment of quantification in ordinary English". In: *Approaches to Natural Language*. Ed. by Patrick Suppes Jaakko Hintikka Julius Moravcsik. Dordrecht: Reidel, pp. 221–242.

References II

Ronneberger, Olaf et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.