Probabilistic models of cognition

Charles Kemp

CHD Summer school, 2018

Why build models?

Why build models?

- To help explain some empirical result

Theoretical Intuitions → Experiment → Data Analysis → Model

- To help formulate a theory

Theoretical Intuitions ↕ Model → Experiment → Data Analysis

Modeling approaches

- Formal logic
- Production systems (e.g SOAR, ACT-R)
- Neural networks
- Dynamic systems
- Probabilistic models

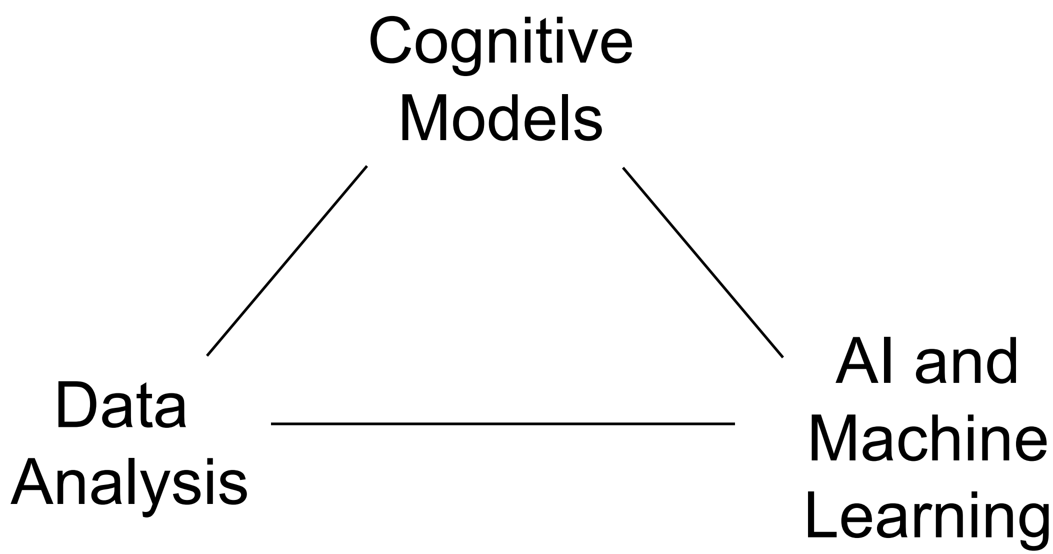
Why build *probabilistic* models?

- For us: probabilities are degrees of belief

- Probability theory captures the right way to update degrees of belief

“The theory of probabilities is at bottom only common sense reduced to calculus; it makes us appreciate with exactitude that which exact minds feel by a sort of instinct without being able oft times to give a reason for it” (Laplace)

Applications of probability theory



Probabilistic machine learning and artificial intelligence

Zoubin Ghahramani[1]

Module 1

INTRODUCTION TO BAYESIAN INFERENCE

Inductive inference

Coughing friend

Inductive inference

John has a cold

Coughing friend

Inductive inference

John has emphysema

Coughing friend

Coughing friend

- d: John is coughing
- h_1: John has a cold
  h_2: John has emphysema
  h_3: John has a stomach upset

Posterior probability

Evidence (Likelihood)

Prior knowledge

$$P(h|d) = \frac{\mathbf{P(d|h)}\ \mathbf{P(h)}}{P(d)}$$

Coughing friend

- d: John is coughing

- h_1: John has a cold
  h_2: John has emphysema
  h_3: John has a stomach upset

Posterior probability

Evidence (Likelihood)

Prior knowledge

$$P(h|d) \propto \mathbf{P(d|h)}\ \mathbf{P(h)}$$

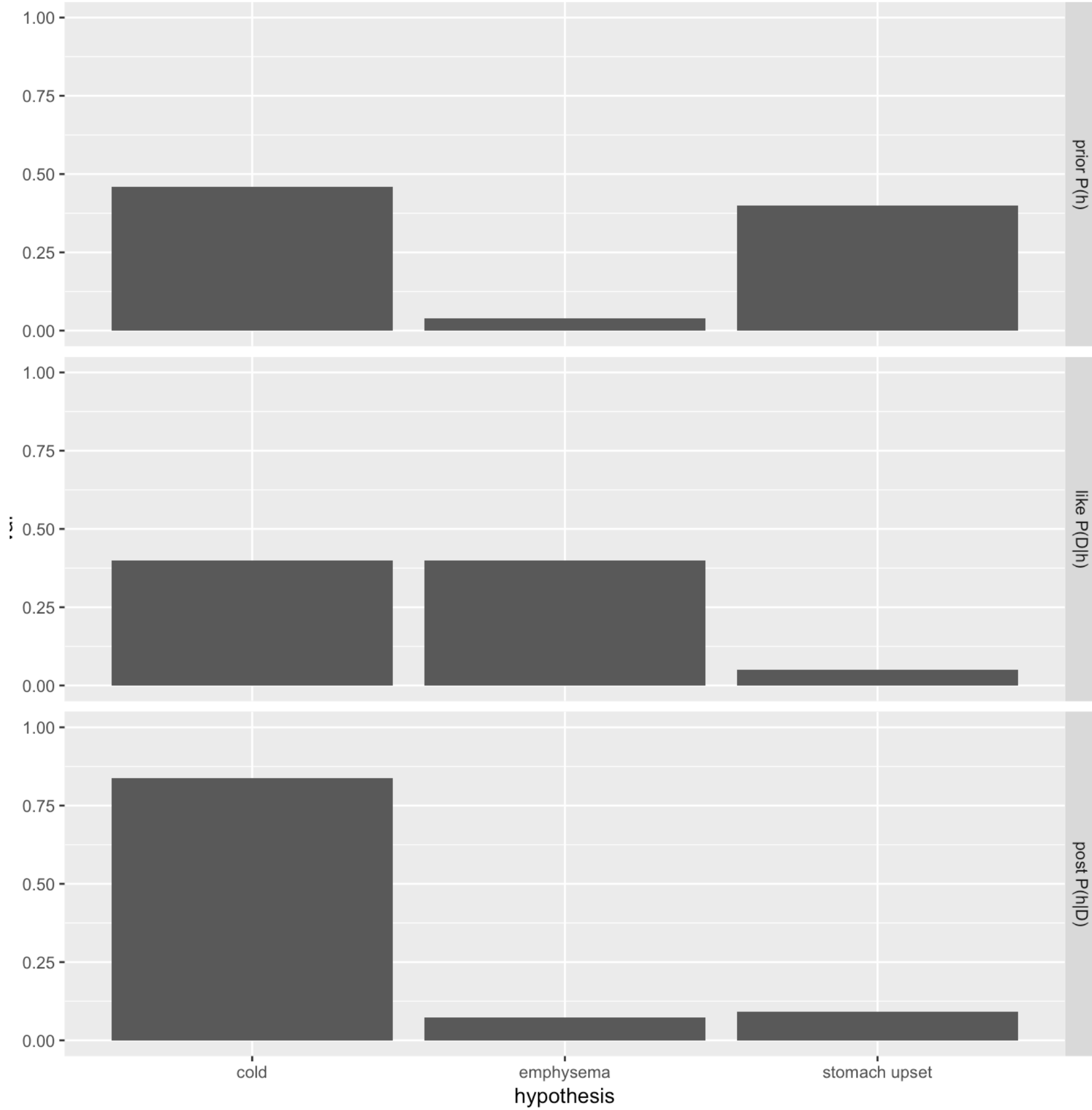
Specifying prior and likelihood

```
h  <- c('cold', 'emphysema', 'stomach upset')
p_h <- c(0.46, 0.04, 0.4)
p_d_given_h <- c(0.4, 0.4, 0.05)
```

prior

likelihood

posterior

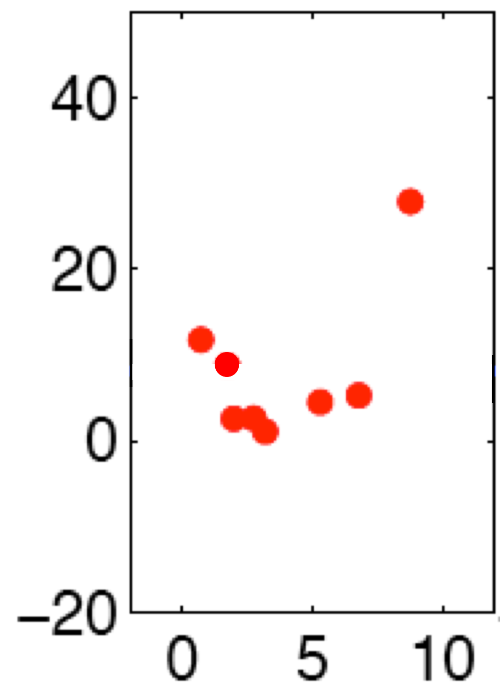


Exercise: Coughing friend

```r
h  <- c('cold', 'emphysema', 'stomach upset')
p_h <- c(0.46, 0.04, 0.4)
p_d_given_h <- c(0.4, 0.4, 0.05)
```

Inductive inference

Coughing friend

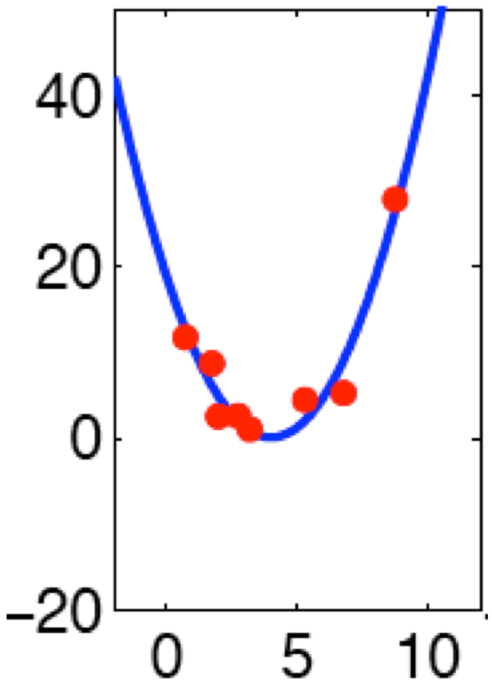


Curve fitting

Inductive inference

John has a cold

Coughing friend

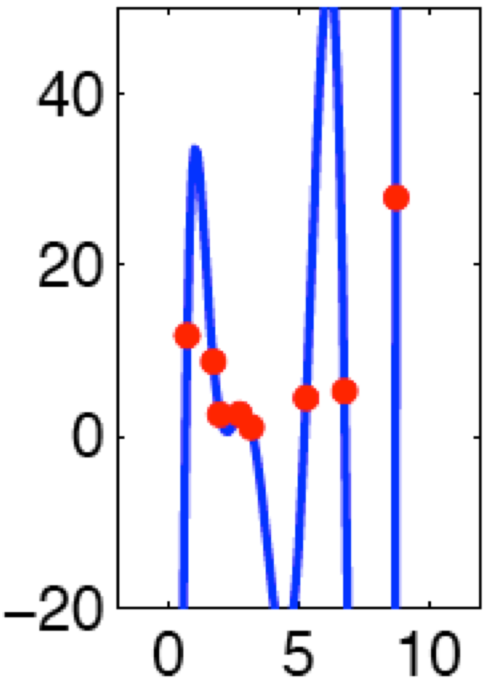


Curve fitting

Inductive inference

John has emphysema

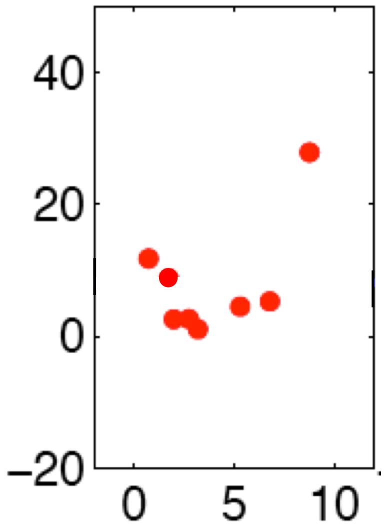
Coughing friend



Curve fitting

Curve fitting

- *D*:



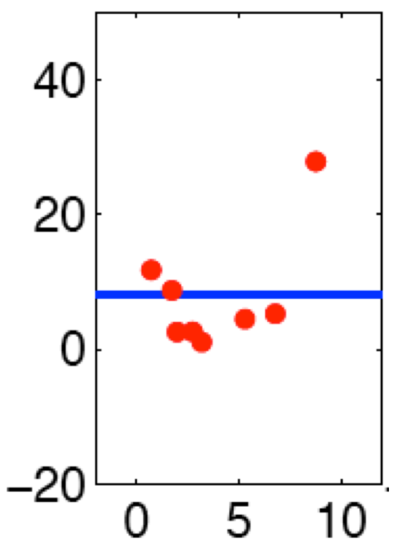
$$P(h|d) \propto \mathbf{P(d|h)}\ \mathbf{P(h)}$$

Posterior probability

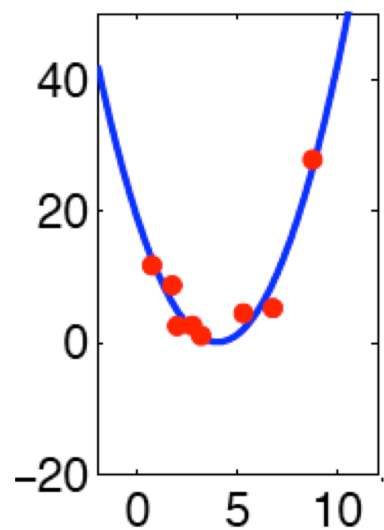
Evidence

Prior knowledge

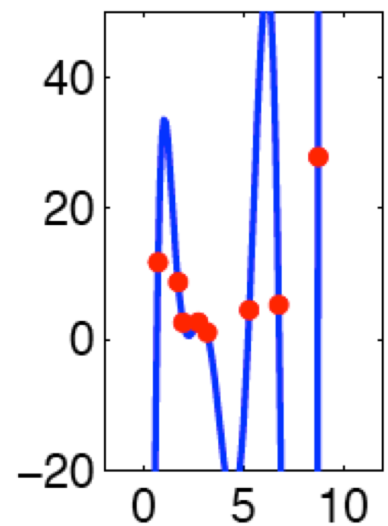
- h_1:

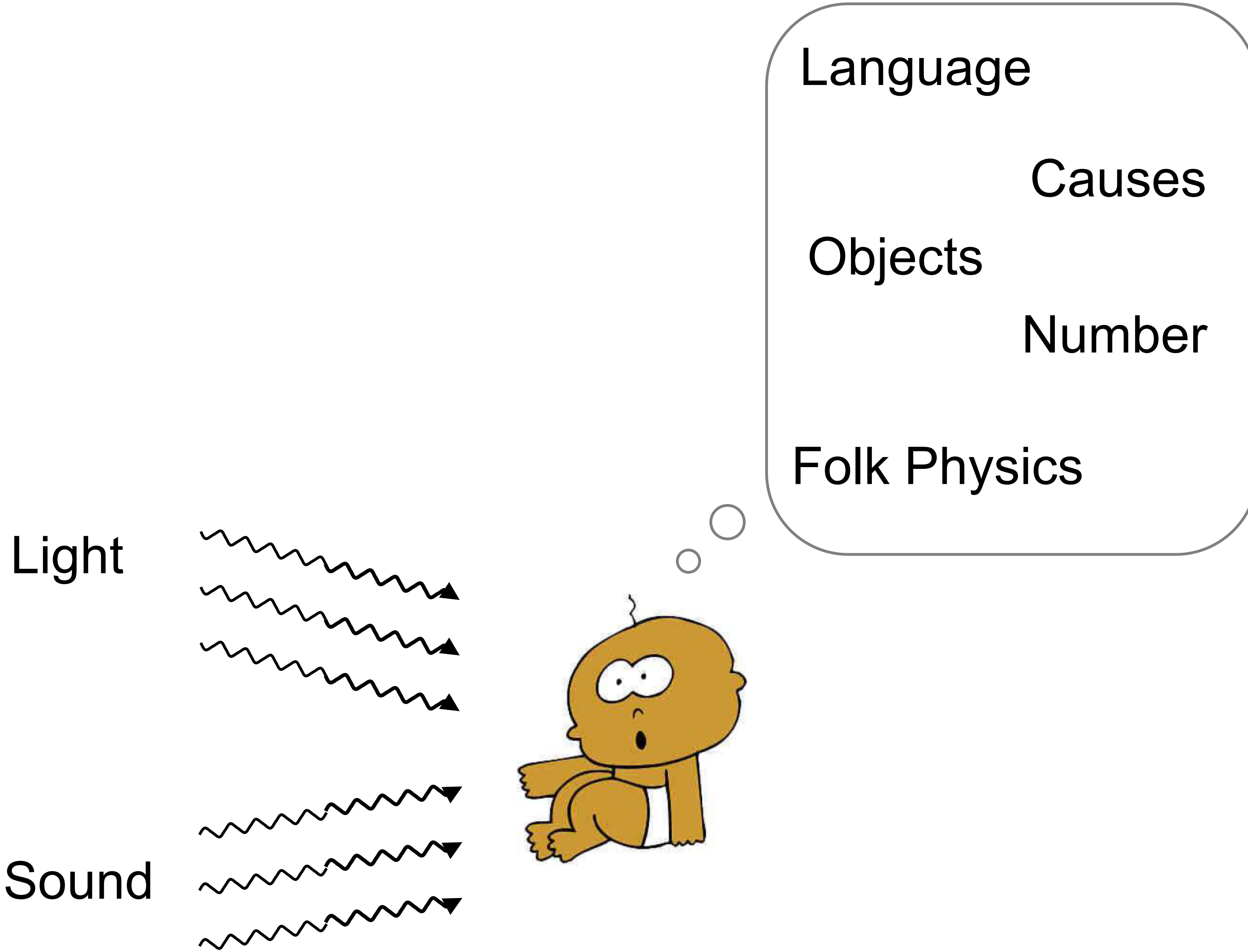


h_2:



h_3:




Language
Causes
Objects
Number
Folk Physics
Light
Sound

Take home messages

- Bayesian models are useful for thinking about inductive problems.
- There are many inductive problems from a wide range of domains including vision, language, motor control, etc.

Module 2

BAYESIAN CONCEPT LEARNING

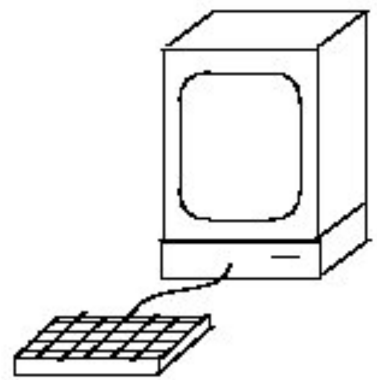
Learning from positive examples

“wombat”

wombat?

wombat?

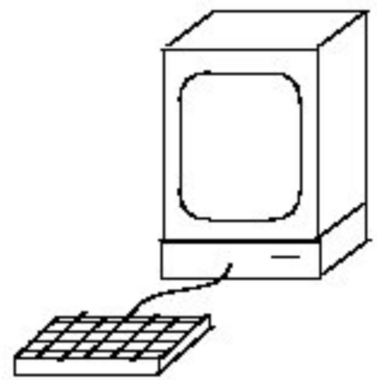
Learning from positive examples



30 70 40 50

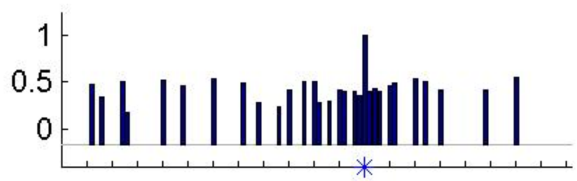
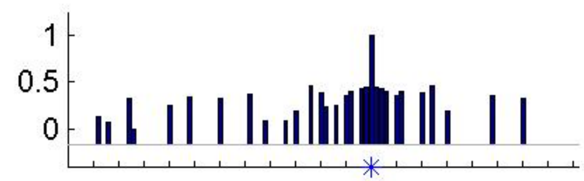
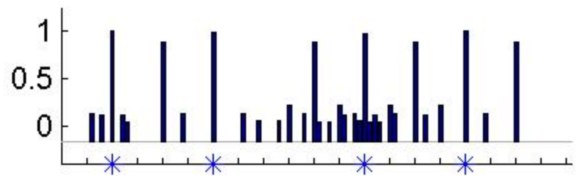
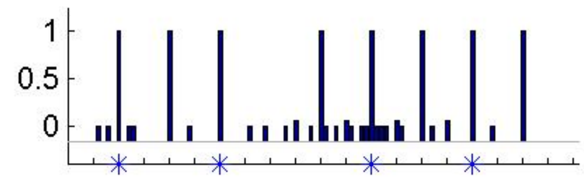
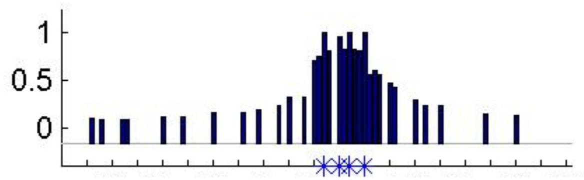
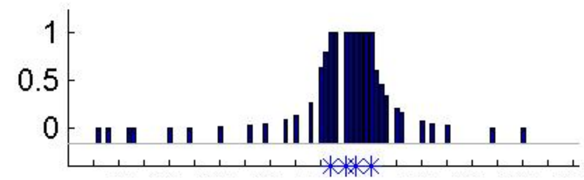
(Tenenbaum)

Learning from positive examples

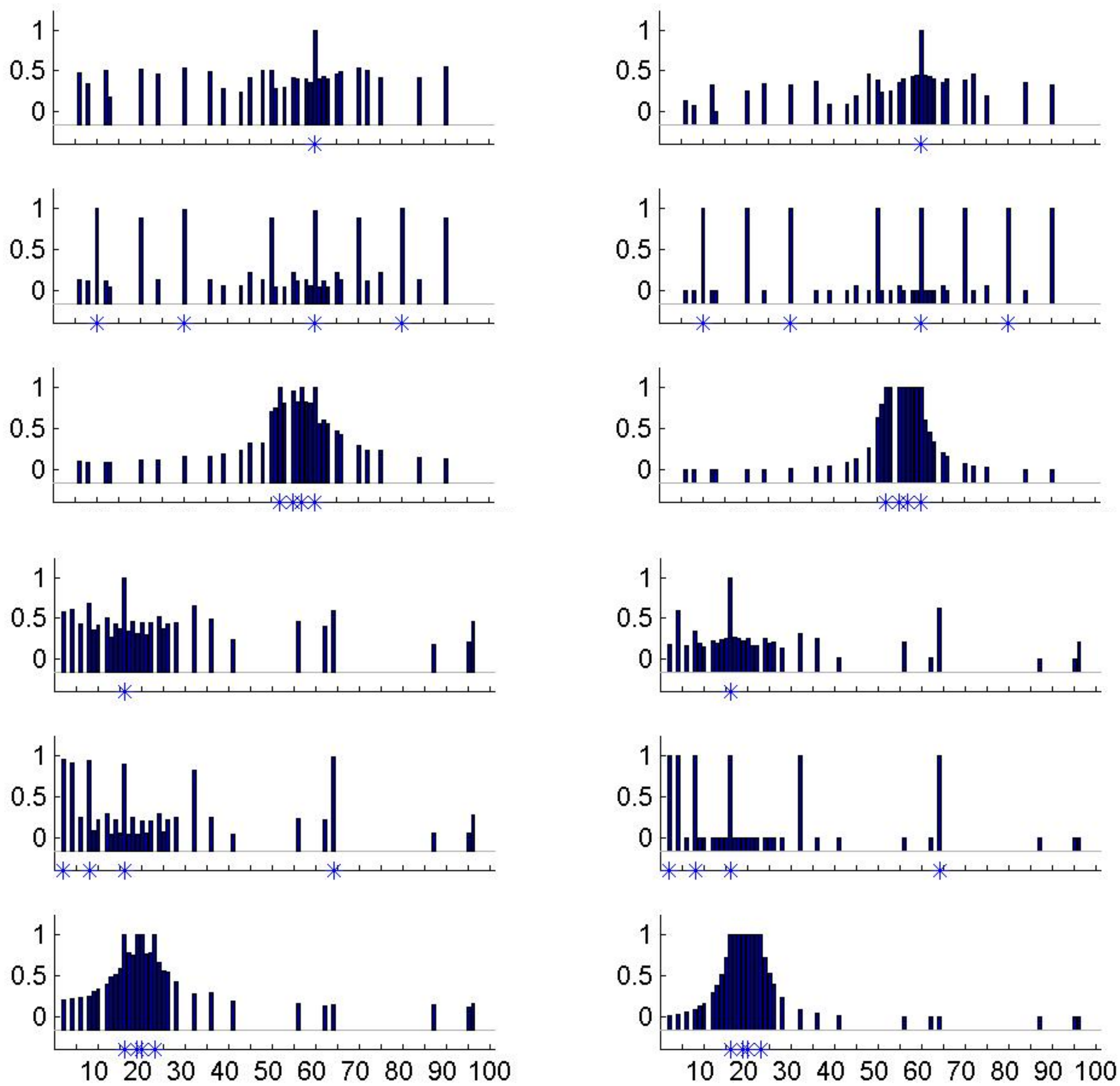


72   68   73

(Tenenbaum)

+ Examples	Human generalization	Bayesian Model
60		
60 80 10 30		
60 52 57 55		

+ Examples	Human generalization	Bayesian Model
60		
60 80 10 30		
60 52 57 55		
16		
16 8 2 64		
16 23 19 20		



Building a Bayesian model

- What are the observed data?
- What are the hypotheses?
- What is the prior?
- What is the likelihood?

Posterior Likelihood Prior

$$P(h|d) \propto \mathbf{P(d|h)}\ \mathbf{P(h)}$$

Bayesian model

- H: Hypothesis space of possible concepts.
  - h_1 = {2, 4, 6, 8, 10, 12, …, 96, 98, 100} ("even numbers")
  - h_2 = {10, 20, 30, 40, …, 90, 100} ("multiples of 10")
  - h_3 = {2, 4, 8, 16, 32, 64} ("powers of 2")
  - h_4 = {50, 51, 52, …, 59, 60} ("numbers between 50 and 60")

- $X = \{x_1, \ldots, x_n\}$: n examples of a concept C.

- Posterior probability of hypothesis h given the data

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

Prior P(h)

- Mathematical properties (24 hypotheses):
  - Odd, even, square, cube, prime numbers
  - Multiples of small integers
  - Powers of small integers
- Raw magnitude (5050 hypotheses):
  - All intervals of integers with endpoints between 1 and 100.

Likelihood P(X|h) (strong sampling)

- Assume that the examples are drawn at random from the hypothesis:

$$P(X|h) = \left[\frac{1}{\text{size}(h)}\right]^n \quad \text{if all } x_i \text{ are in h}$$

$$= 0 \quad \text{otherwise}$$

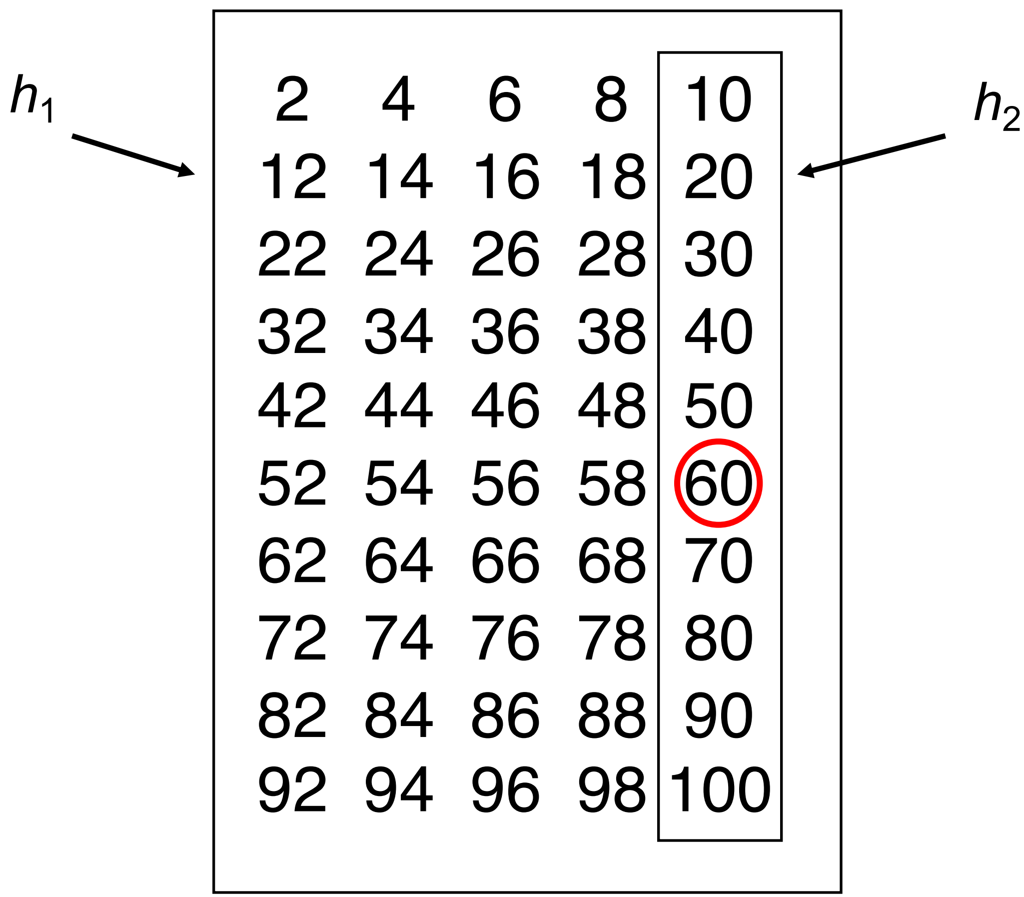
Illustrating the likelihood

h_1

h_2

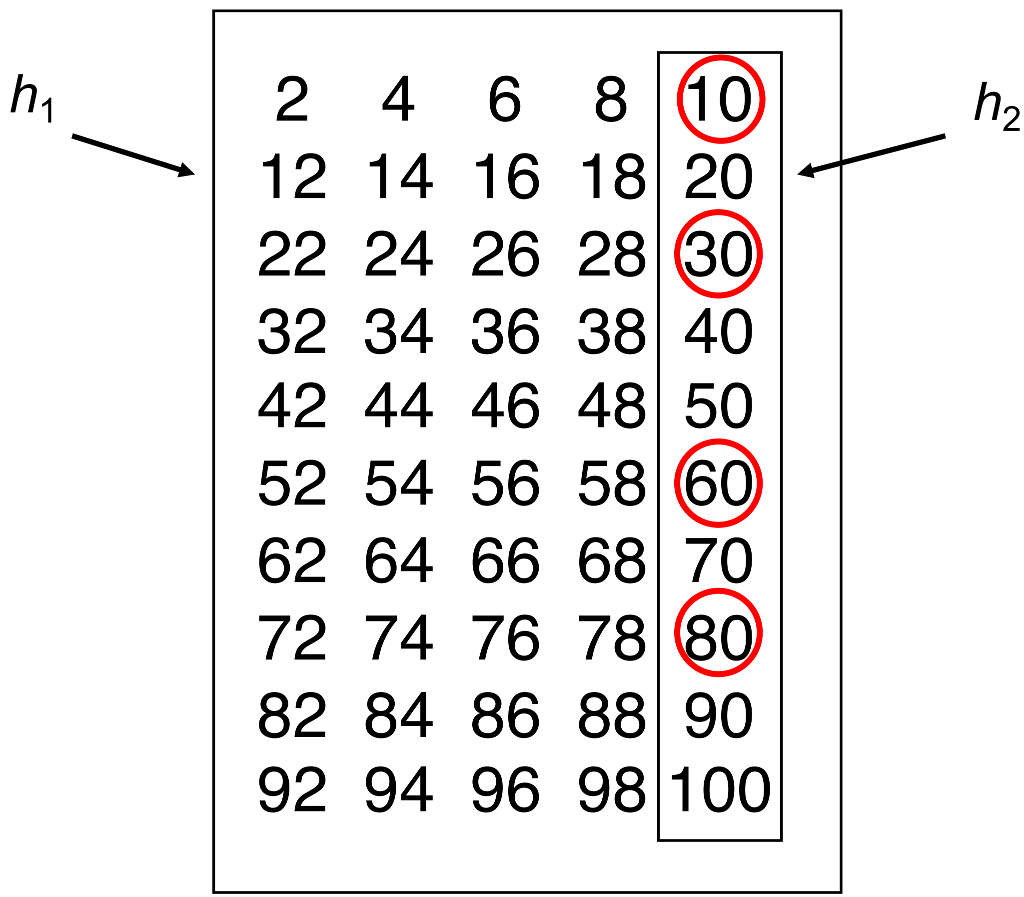
2	4	6	8	10
12	14	16	18	20
22	24	26	28	30
32	34	36	38	40
42	44	46	48	50
52	54	56	58	60
62	64	66	68	70
72	74	76	78	80
82	84	86	88	90
92	94	96	98	100

Illustrating the likelihood



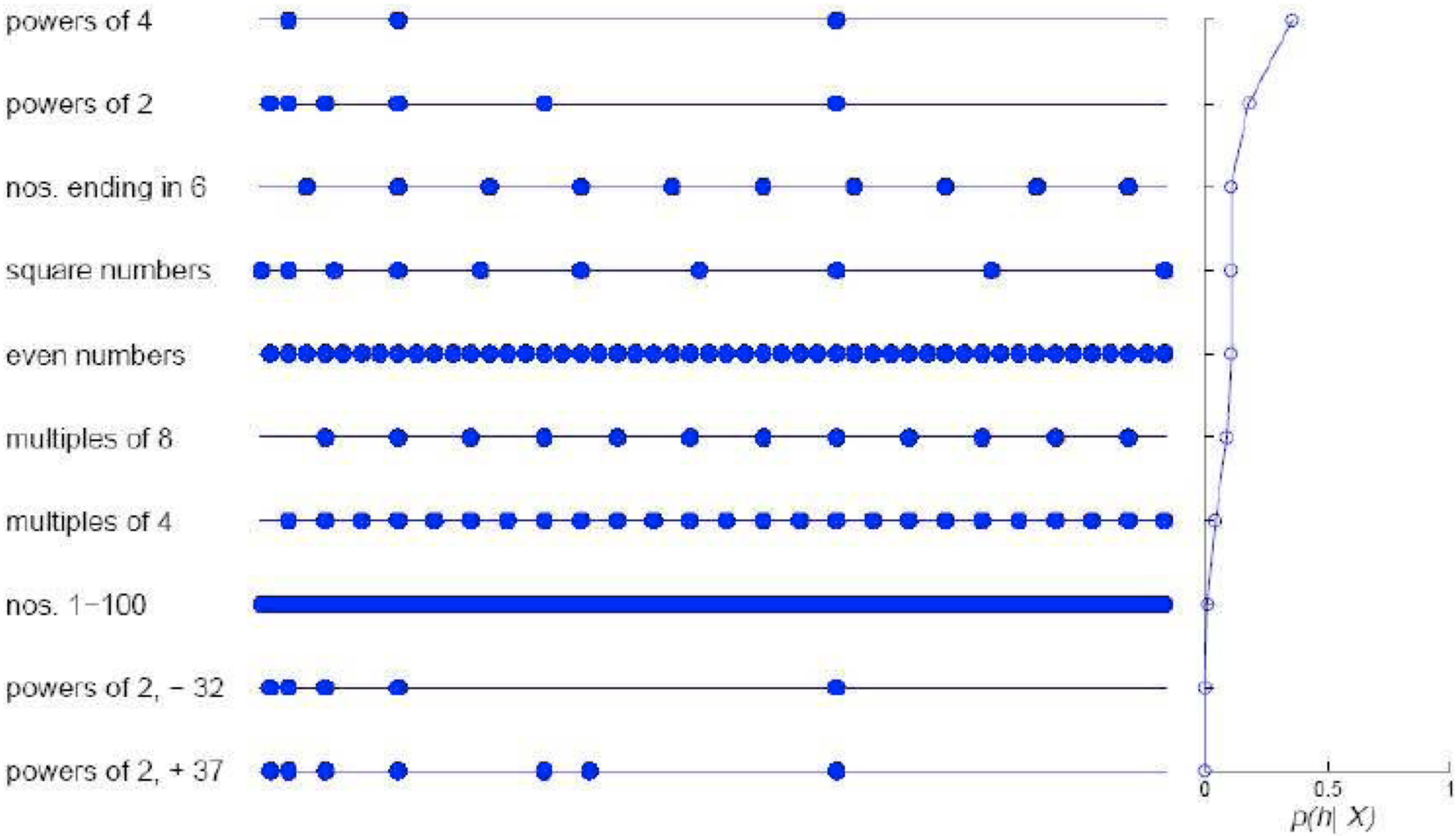
Data slightly more of a coincidence under h_1

Illustrating the likelihood



Data *much* more of a coincidence under h_1

Examples: 16

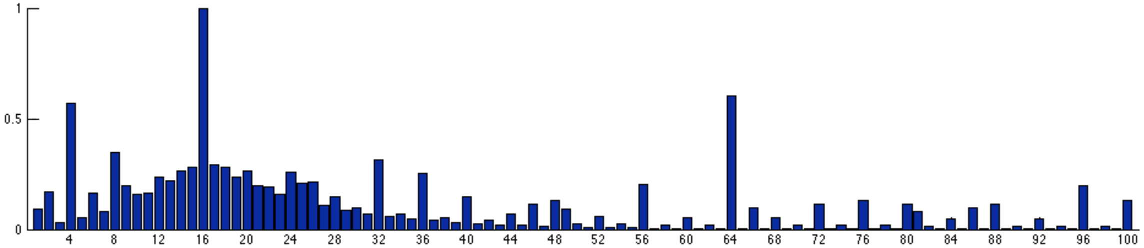


powers of 4
powers of 2
nos. ending in 6
square numbers
even numbers
multiples of 8
multiples of 4
nos. 1−100
powers of 2, − 32
powers of 2, + 37
0
0.5
1
p(h| X)

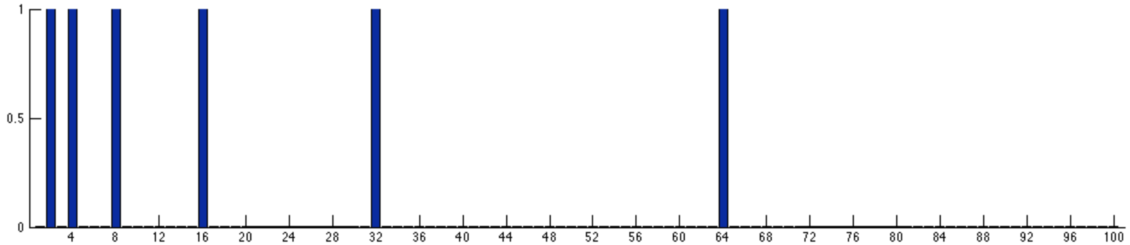
Prediction by hypothesis averaging

$$P(y \in C|X) = \sum_{h \in \mathcal{H}} P(y \in C|h) P(h|X)$$

$$= \sum_{h \in \mathcal{H}_y} P(h|X)$$



Tenenbaum prior, strong sampling



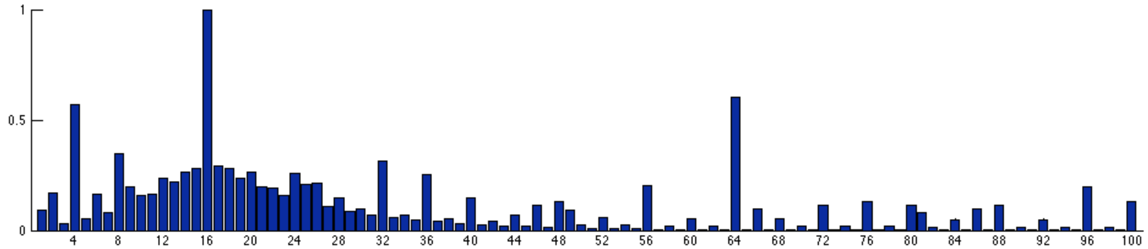
16 8 2 64

Tenenbaum prior, strong sampling

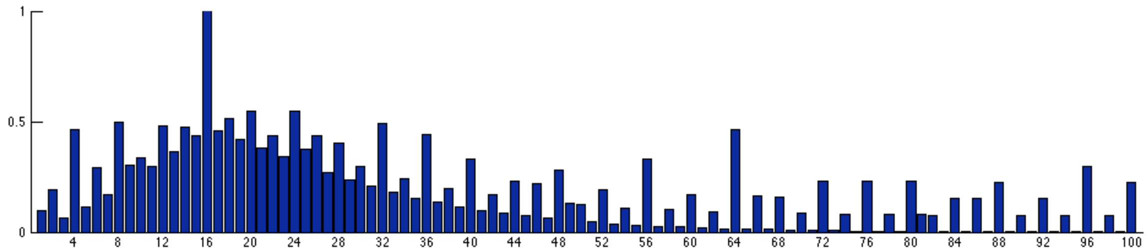
Likelihood P(X|h) (weak sampling)

- Assume that someone has asked whether each x_i is in the concept

$P(X\|h)$	=	1	if all x_i are in h
	=	0	otherwise

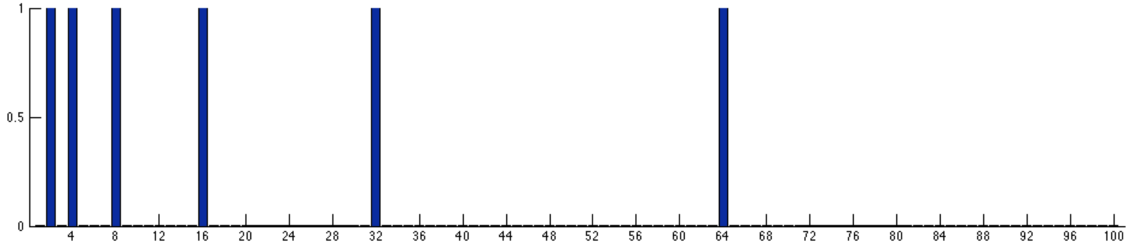


Tenenbaum prior,
strong sampling

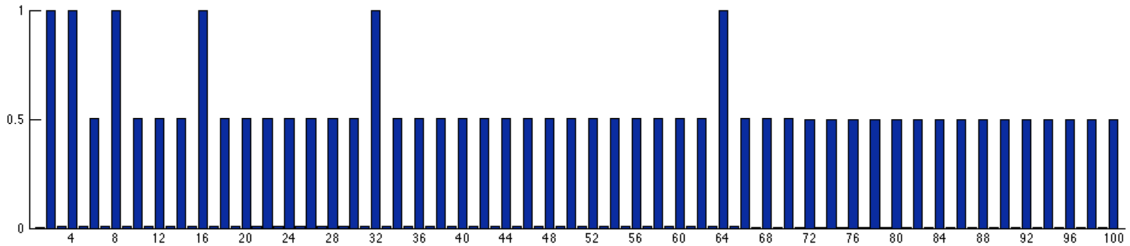


Tenenbaum prior,
weak sampling

16 8 2 64



Tenenbaum prior,
strong sampling

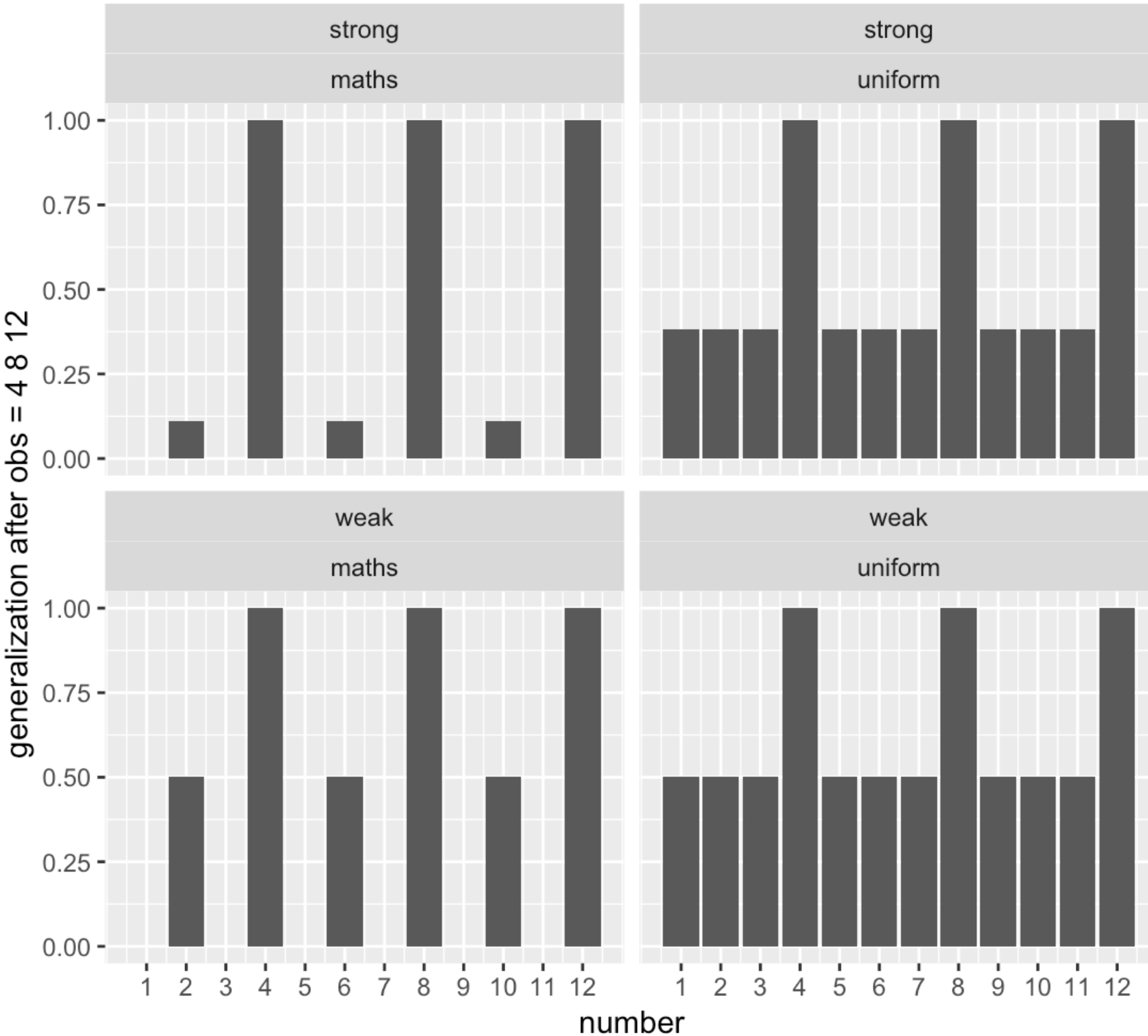


Tenenbaum prior,
weak sampling

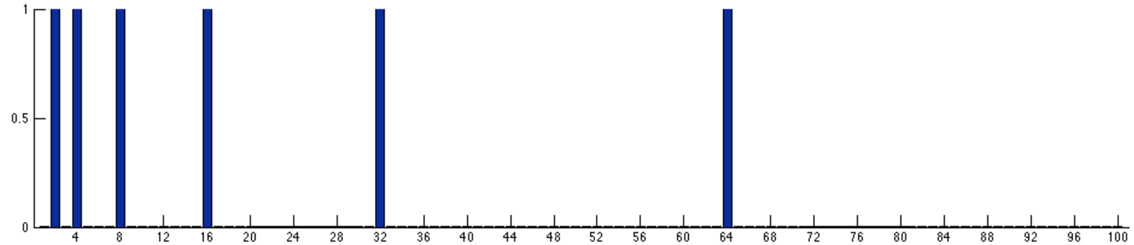
Take home messages

- Bayesian models include a likelihood and a prior.
- The likelihood matters.
- How much does the prior matter?

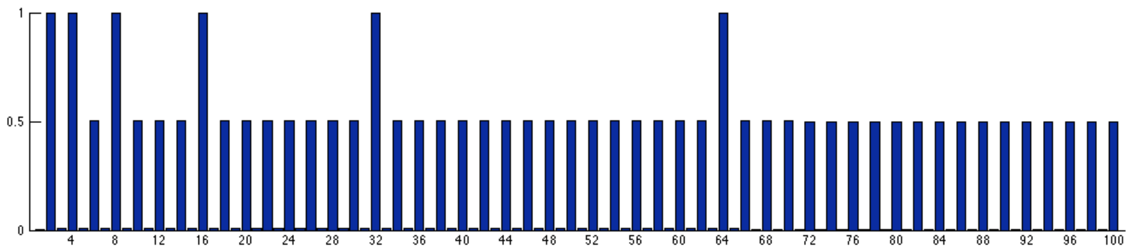
Exercise: number game


strong
maths
strong
uniform
weak
maths
weak
uniform
generalization after obs = 4 8 12
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12
number

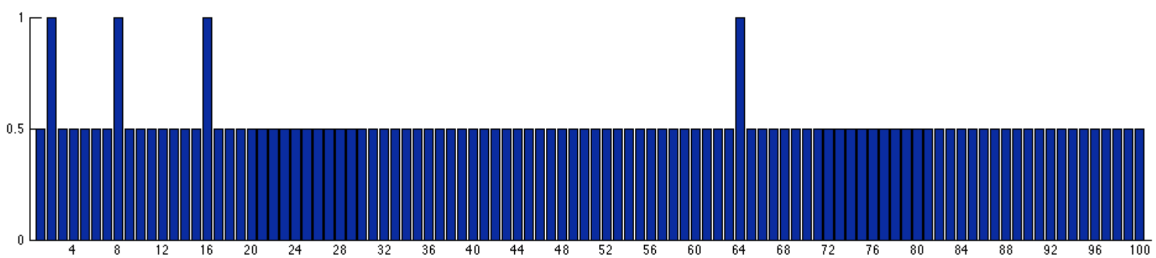
16 8 2 64



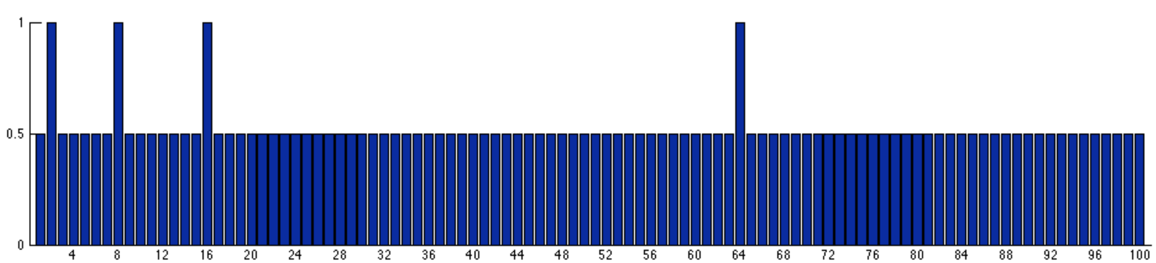
Tenenbaum prior,
strong sampling



Tenenbaum prior,
weak sampling



Uniform prior,
strong sampling



Uniform prior,
weak sampling

Take home messages

- Bayesian models include a likelihood and a prior.
- The likelihood matters.
- The prior matters.
- Think carefully about both when you set up a model!

Module 3

BAYESIAN NETWORKS (DIRECTED GRAPHICAL MODELS)

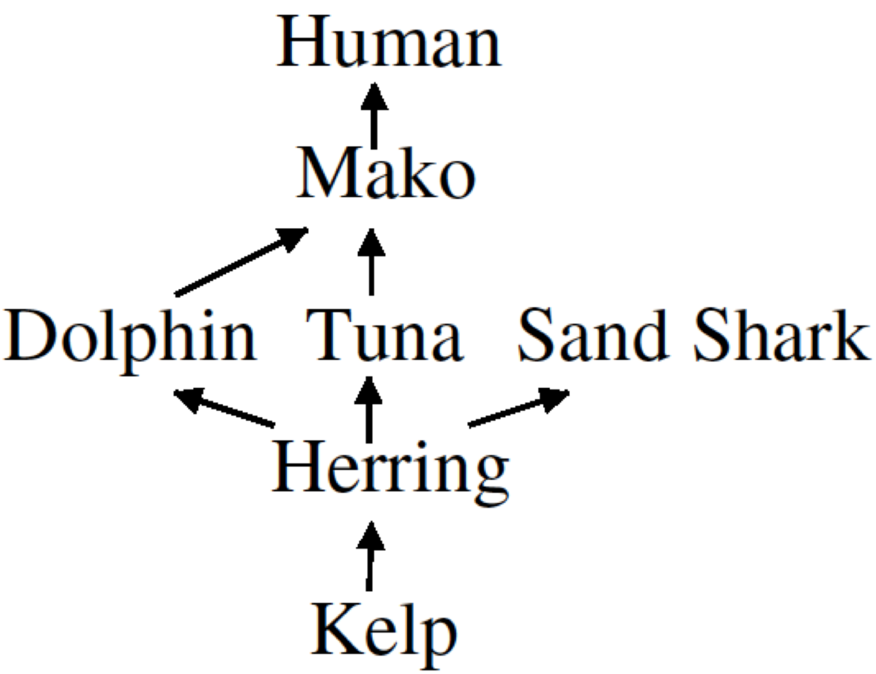
Representing hypothesis spaces and priors

- So far we’ve just enumerated the hypothesis space.

- What if the hypothesis space is huge?

- How do we come up with all the numbers in the prior?

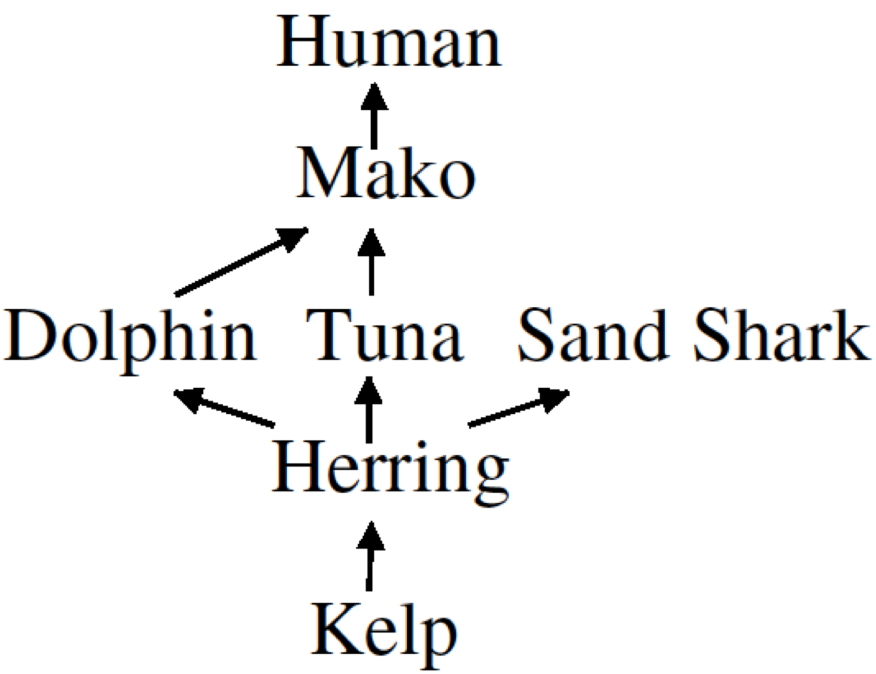
Food web problem

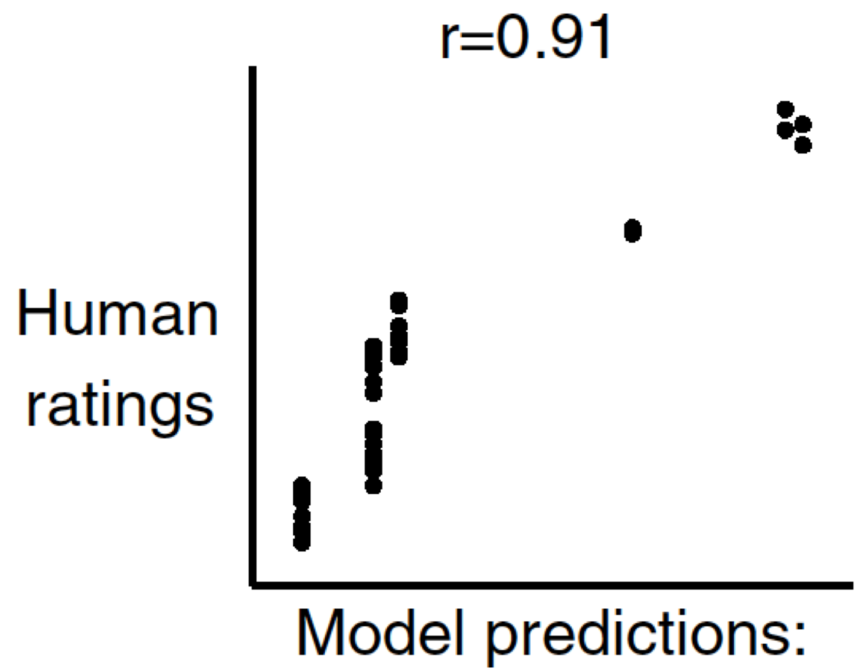


Herring carry a certain disease.

How likely is it that mako have the same disease?

Food web problem





(Shafto et al, 2008)

Building a Bayesian model

- What are the observed data?
- What are the hypotheses?
- What is the prior?
- What is the likelihood?

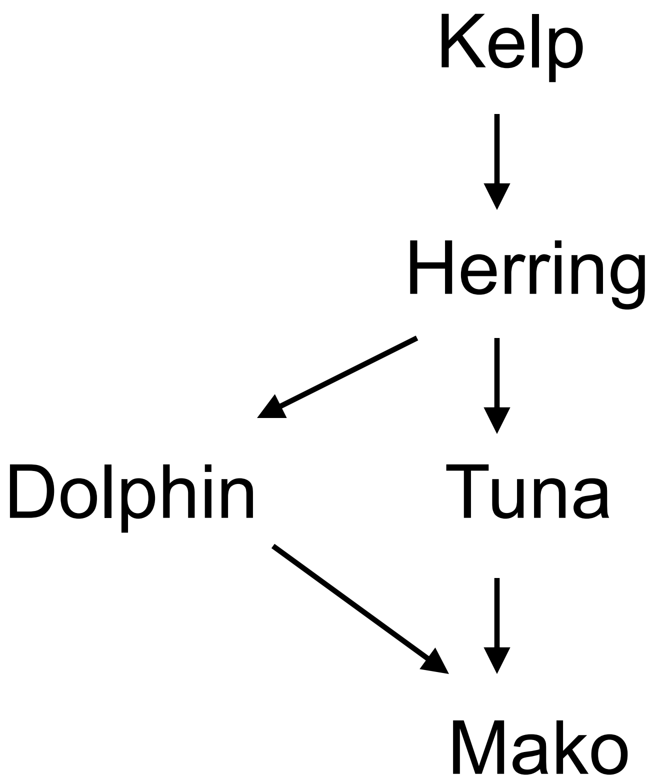
Posterior Likelihood Prior

$$P(h|d) \propto \mathbf{P(d|h)}\ \mathbf{P(h)}$$

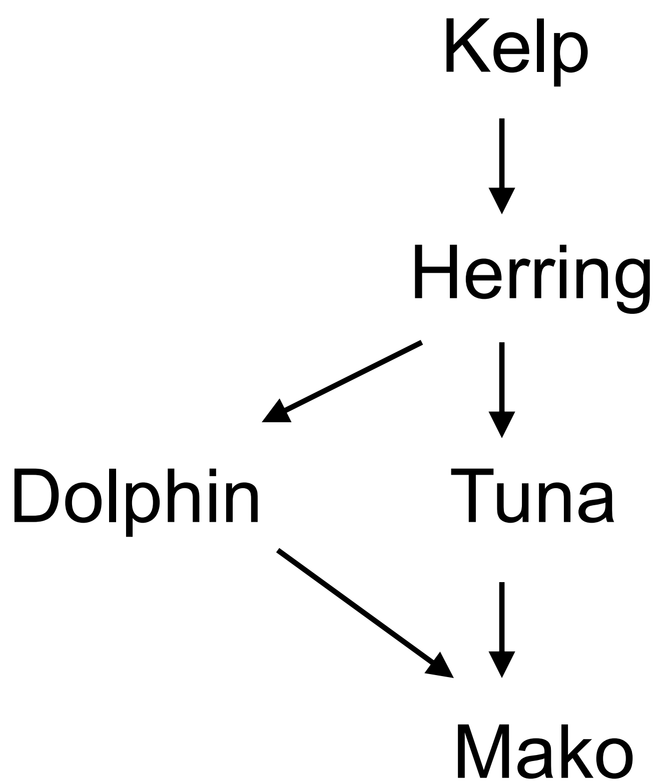
Hypothesis space and prior

	kelp	herring	dolphin	tuna	sandshark	mako	human	prior
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<dbl>
1	1	1	1	1	1	1	1	0.478
2	2	1	1	1	1	1	1	0.0266
3	1	2	1	1	1	1	1	0.00664
4	2	2	1	1	1	1	1	0.00406
5	1	1	2	1	1	1	1	0.0266
6	2	1	2	1	1	1	1	0.00148
7	1	2	2	1	1	1	1	0.00406
8	2	2	2	1	1	1	1	0.00248
9	1	1	1	2	1	1	1	0.0266
10	2	1	1	2	1	1	1	0.00148

Specifying a prior

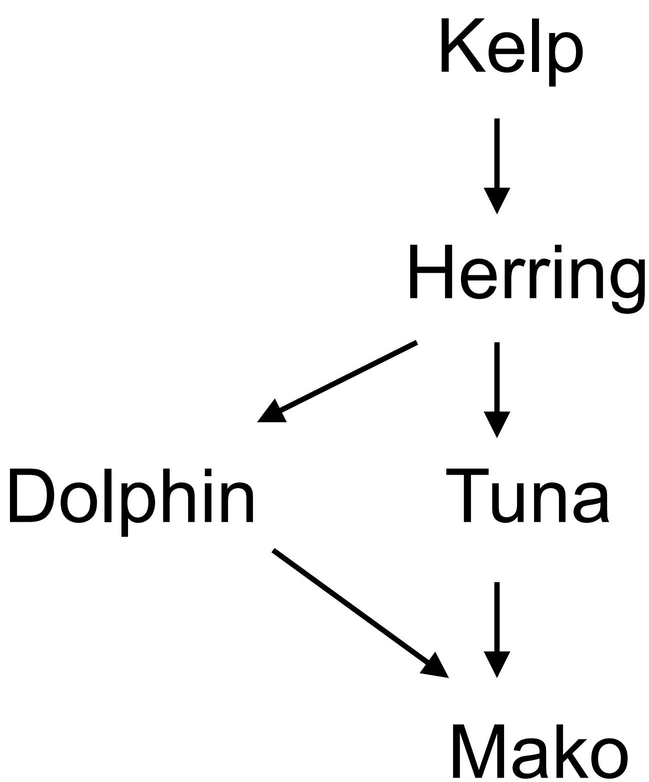


Specifying a prior



$P(K = 1) = 0.1$

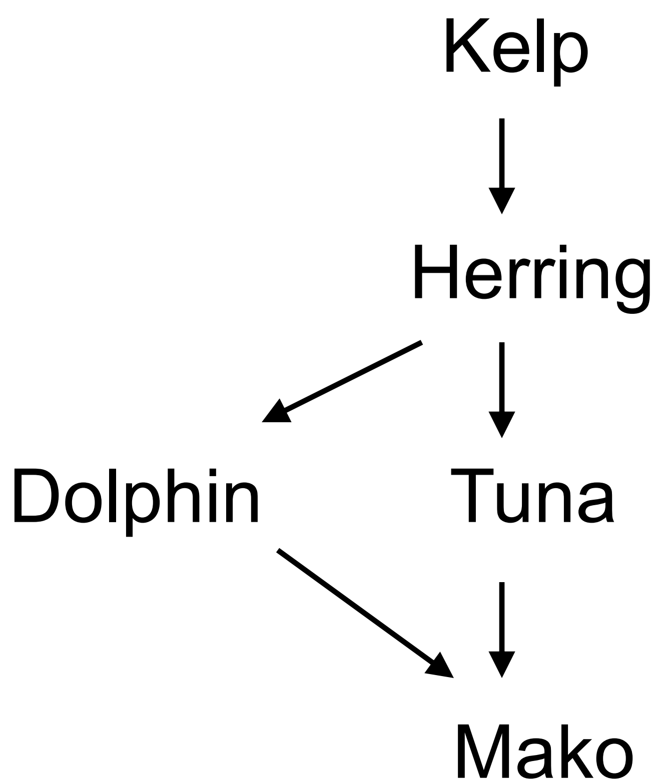
Specifying a prior



$P(K = 1) = 0.1$

K	P(H = 1 |K)
0	0.1
1	0.55

Specifying a prior

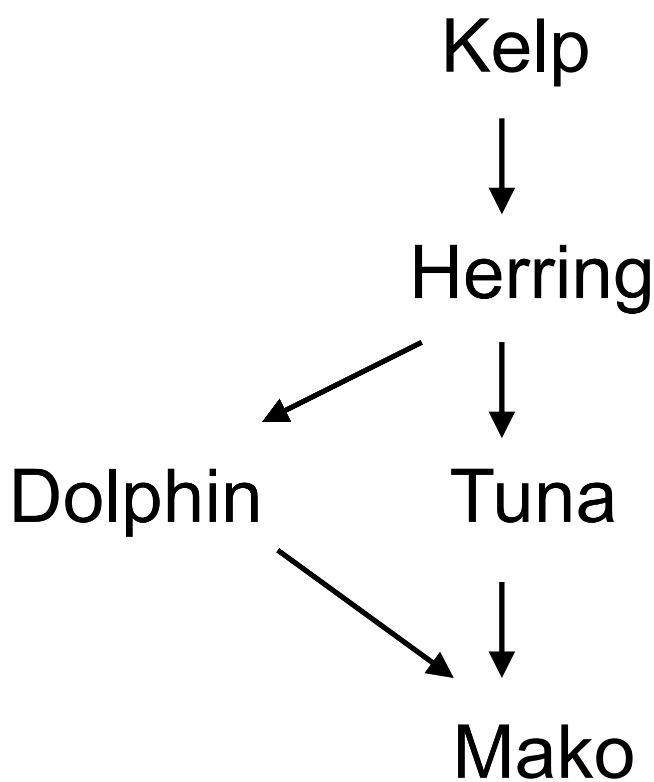


P(K = 1) = 0.1

K	P(H = 1 |K)
0	0.1
1	0.55

D	T	P(M = 1 |D, T)
0	0	0.1
0	1	0.55
1	0	0.55
1	1	0.775

Specifying a prior

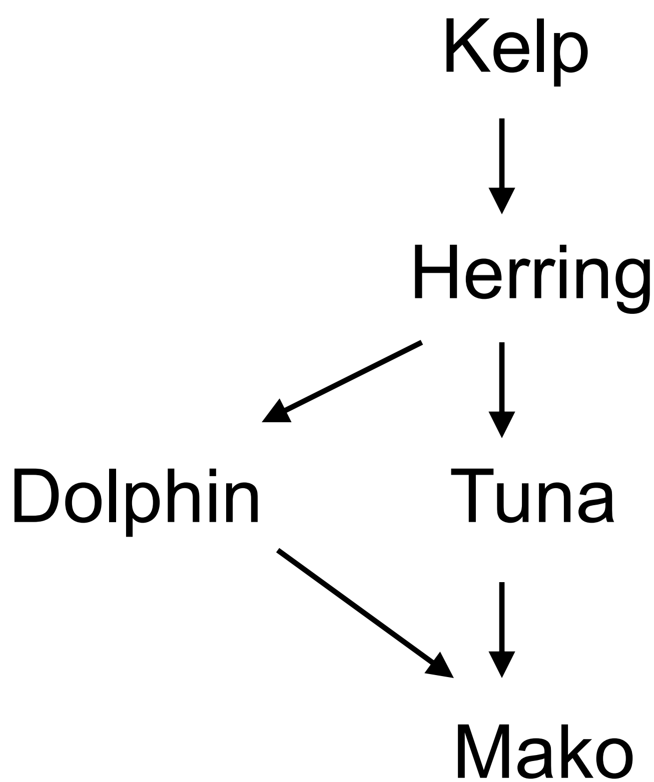


$$P(K, H, D, T, M) = P(K) * P(H|K) * P(D|H) * P(T|H) * P(M|D,T)$$

General formulation

$$P(v_1, \ldots, v_n) = \prod_i P(v_i | \mathrm{pa}(V_i))$$

Specifying a prior



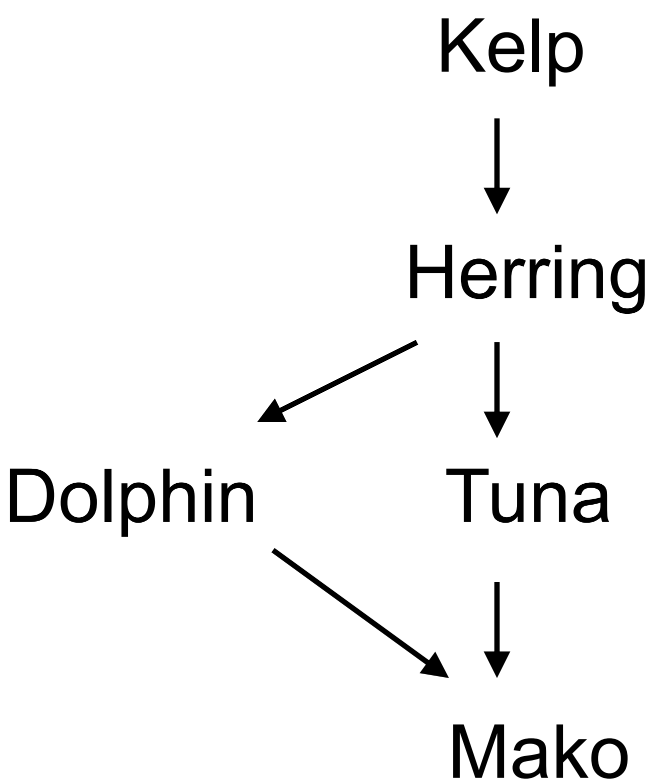
P(K = 1) = 0.1

K	P(H = 1 |K)
0	0.1
1	0.55

D	T	P(M = 1 |D, T)
0	0	0.1
0	1	0.55
1	0	0.55
1	1	0.775

Specifying a prior

b: base rate
t: transmission rate



$P(K = 1) = b$

K	P(H = 1 |K)
0	b
1	$1 - (1-t)(1-b)$

D	T	P(M = 1 |D, T)
0	0	b
0	1	$1 - (1-t)(1-b)$
1	0	$1 - (1-t)(1-b)$
1	1	$1 - (1-t)(1-t)(1-b)$

Exercise: Food web (enumeration)

	kelp	herring	dolphin	tuna	sandshark	mako	human	prior
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<dbl>
1	1	1	1	1	1	1	1	0.478
2	2	1	1	1	1	1	1	0.0266
3	1	2	1	1	1	1	1	0.00664
4	2	2	1	1	1	1	1	0.00406
5	1	1	2	1	1	1	1	0.0266
6	2	1	2	1	1	1	1	0.00148
7	1	2	2	1	1	1	1	0.00406
8	2	2	2	1	1	1	1	0.00248
9	1	1	1	2	1	1	1	0.0266
10	2	1	1	2	1	1	1	0.00148

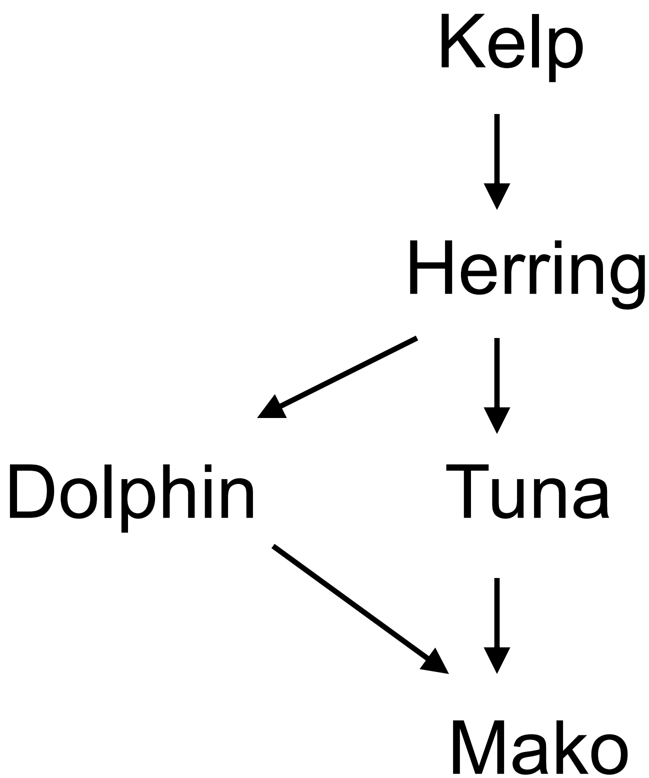
Take home messages

- Bayes nets give us a good way to specify priors

Module 4

INFERENCE BY SAMPLING FROM THE PRIOR

Sampling from the prior



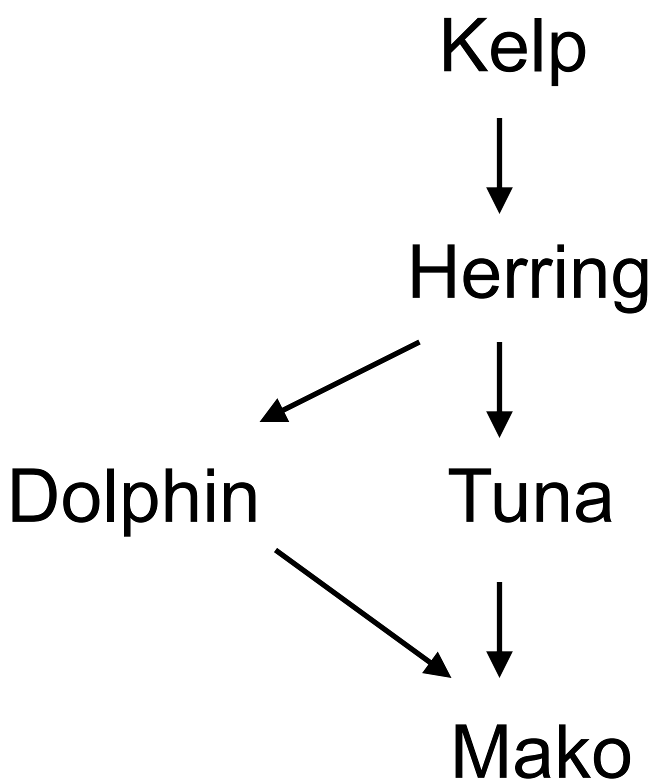
P(K = 1) = 0.1

K	P(H = 1 |K)
0	0.1
1	0.55

D	T	P(M = 1 |D, T)
0	0	0.1
0	1	0.55
1	0	0.55
1	1	0.775

Sample based predictions

```
obs <- list(kelp = 1, mako = 2)
```



1. Collect set of samples from prior
2. Remove all those that are inconsistent with the observations
3. Make predictions based on samples that remain

Sample based predictions

$$P(\text{humans}|obs) = \sum_h P(\text{humans}|h)P(h|obs)$$

$$\propto \sum_h P(\text{humans}|h)P(obs|h)P(h)$$

$$\approx \frac{1}{M}\sum_{i=1}^{M} P(\text{humans}|h^i)P(obs|h^i)$$

Exercise: Food web (sampling from the prior)

Take home messages

- Bayes nets give us a good way to specify priors
- Sampling is often a good way to implement probabilistic inference

Module 5

JAGS

Sample based predictions

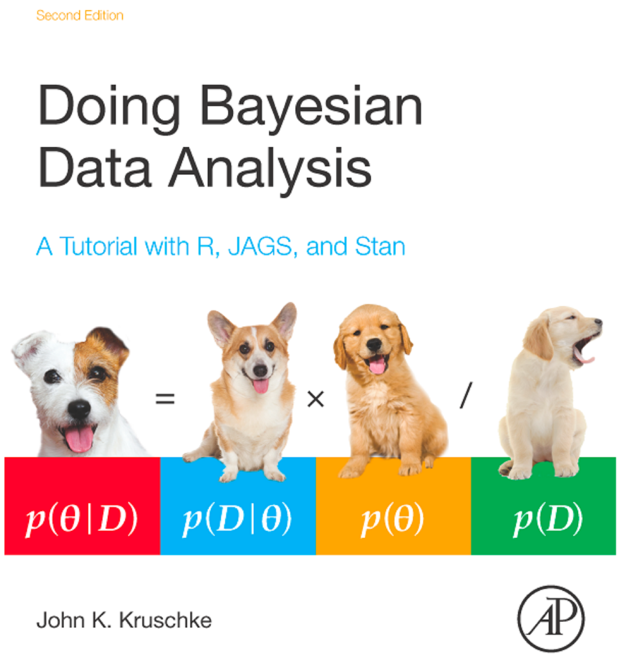
- If we could sample from the posterior $P(h|obs)$

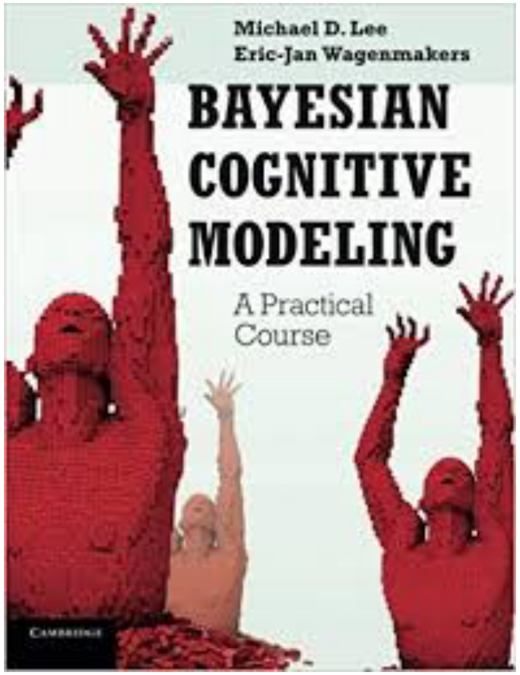
$$P(\text{humans}|obs) = \sum_{h} P(\text{humans}|h)P(h|obs)$$

$$\approx \frac{1}{M}\sum_{i=1}^{M} P(\text{humans}|h^{i})$$

JAGS and STAN

- Widely used for data analysis and cognitive modeling





JAGS : model specification

```
model
    {
       # dcat specifies a discrete categorical distribution
       kelp ~ dcat(p.kelp[1:2])
       herring ~ dcat(p.herring[kelp,1:2])
       dolphin ~ dcat(p.dolphin[herring,1:2])
       tuna ~ dcat(p.tuna[herring,1:2])
       sandshark ~ dcat(p.sandshark[herring,1:2])
       mako ~ dcat(p.mako[dolphin,tuna,1:2])
       human ~ dcat(p.human[mako,1:2])
}
```

JAGS : observations and CPDs

```
# specify observations here
obs <- list(kelp = 1, mako = 2)


 foodwebdata<- c(obs, list(
        p.kelp = zerop,
        p.herring = onep,
        p.dolphin = onep,
        p.tuna = onep,
        p.sandshark = onep,
        p.mako = twop,
        p.human = onep))
```

Running JAGS

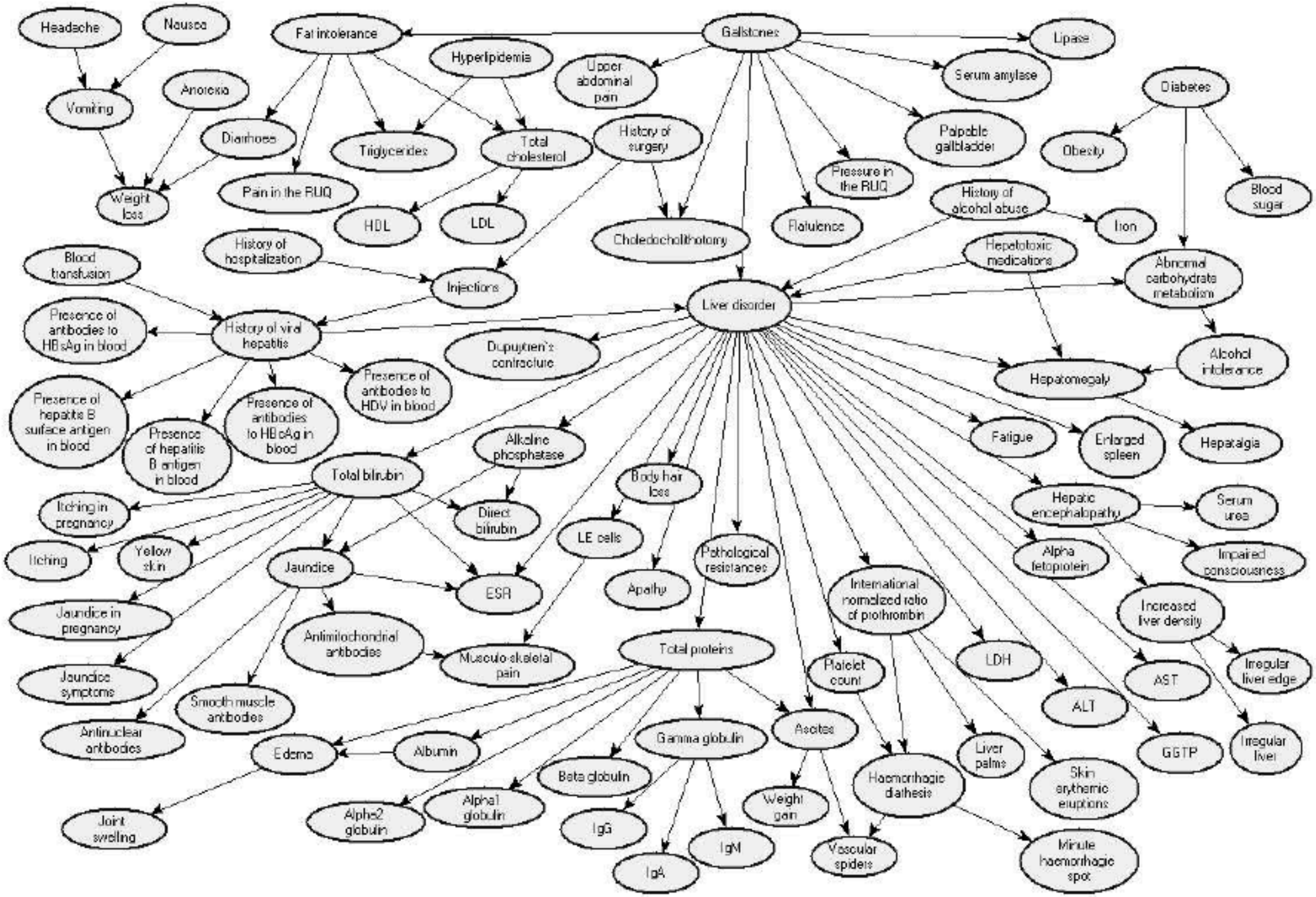
```
# set up the model in JAGS
jags <- jags.model('foodweb.bug', data = foodwebdata,
                   n.chains = 4,
                   n.adapt = 100)

# actually run the model in JAGS (ie sample from the posterior P(h|obs) )
samples <- coda.samples(jags,
             c('kelp', 'herring', 'dolphin', 'tuna', 'sandshark', 'mako', 'human'),
             10000)
```

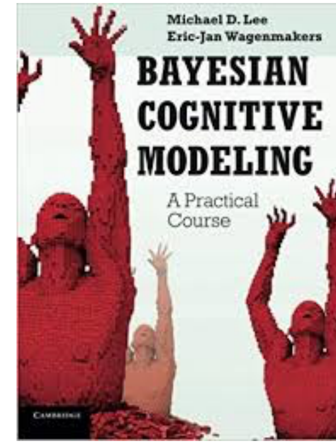
Exercise: Food web (sampling from the posterior using JAGS)

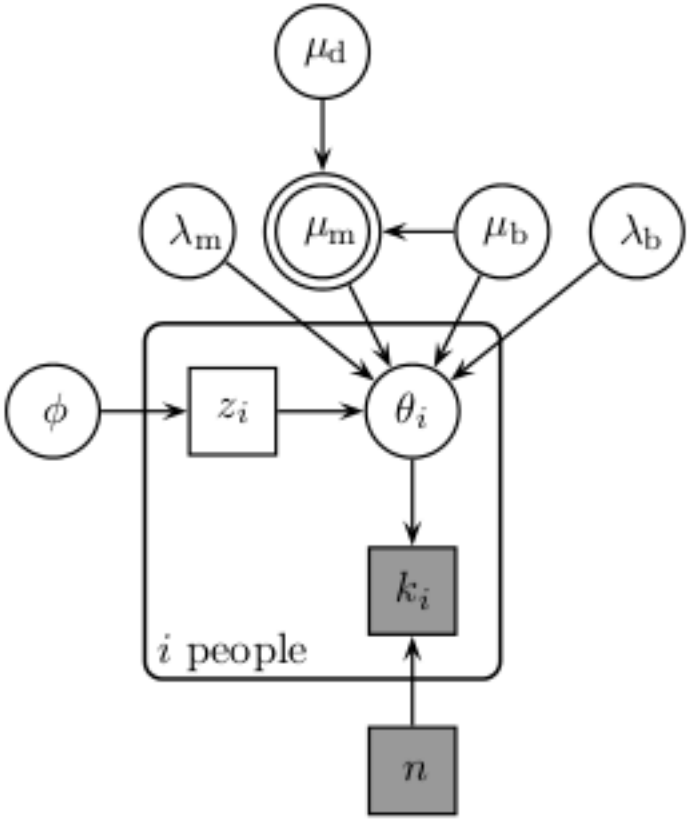
Take home messages

- Bayes nets give us a good way to specify priors
- Sampling is often a good way to implement probabilistic inference
- Tools like JAGS make sampling-based inference relatively simple in many contexts


Headache
Nausea
Fat intolerance
Gallstones
Lipase
Hyperlipidemia
Upper abdominal pain
Serum amylase
Diabetes
Anorexia
Vomiting
Diarrhoea
Triglycerides
Total cholesterol
History of surgery
Palpable gallbladder
Obesity
Weight loss
Pain in the RUQ
Pressure in the RUQ
Blood sugar
HDL
LDL
Choledocholithotomy
Flatulence
History of alcohol abuse
Iron
History of hospitalization
Blood transfusion
Hepatotoxic medications
Abnormal carbohydrate metabolism
Injections
Liver disorder
Presence of antibodies to HBsAg in blood
History of viral hepatitis
Dupuytren's contracture
Alcohol intolerance
Hepatomegaly
Presence of antibodies to HDV in blood
Presence of hepatitis B surface antigen in blood
Presence of hepatitis B antigen in blood
Presence of antibodies to HBcAg in blood
Alkaline phosphatase
Fatigue
Enlarged spleen
Hepatalgia
Total bilirubin
Body hair loss
Hepatic encephalopathy
Serum urea
Itching in pregnancy
Direct bilirubin
LE cells
Itching
Yellow skin
Jaundice
Pathological resistances
Alpha fetoprotein
Impaired consciousness
ESR
Apathy
International normalized ratio of prothrombin
Increased liver density
Jaundice in pregnancy
Antimitochondrial antibodies
Musculo-skeletal pain
Total proteins
Platelet count
LDH
AST
Irregular liver edge
Jaundice symptoms
ALT
Smooth muscle antibodies
Antinuclear antibodies
Gamma globulin
Ascites
Edema
Albumin
Liver palms
GGTP
Irregular liver
Beta globulin
Haemorrhagic diathesis
Skin erythemic eruptions
Alpha1 globulin
Weight gain
Joint swelling
Alpha2 globulin
IgG
IgM
Vascular spiders
Minute haemorrhagic spot
IgA

Bayes nets





$$
\begin{aligned}
\mu_{\mathrm{b}} &\sim \mathrm{Beta}(1,1) \\
\mu_{\mathrm{d}} &\sim \mathrm{Gaussian}(0,0.5)_{\mathcal{I}(0,\infty)} \\
\lambda_{\mathrm{b}} &\sim \mathrm{Uniform}(40,800) \\
\lambda_{\mathrm{m}} &\sim \mathrm{Uniform}(4,100) \\
z_i &\sim \mathrm{Bernoulli}(\phi) \\
\theta_i &\sim \begin{cases} \mathrm{Beta}(\mu_{\mathrm{b}}\lambda_{\mathrm{b}}, (1-\mu_{\mathrm{b}})\,\lambda_{\mathrm{b}}) & \text{if } z_i = 0 \\ \mathrm{Beta}(\mu_{\mathrm{m}}\lambda_{\mathrm{m}}, (1-\mu_{\mathrm{m}})\,\lambda_{\mathrm{m}}) & \text{if } z_i = 1 \end{cases} \\
k_i &\sim \mathrm{Binomial}(\theta_i, n) \\
\mathrm{logit}\mu_{\mathrm{m}} &\leftarrow \mathrm{logit}\mu_{\mathrm{b}} - \mu_{\mathrm{d}} \\
\phi &\sim \mathrm{Beta}(5,5)
\end{aligned}
$$

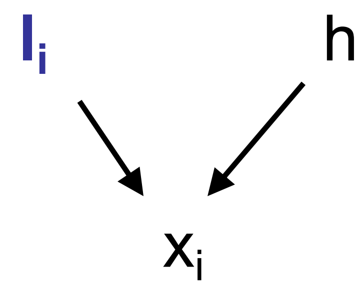
Graphical model for inferring membership of two latent groups, consisting of malingerers and *bona fide* participants.

Bayes nets

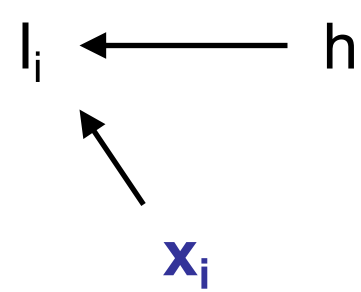
- When you're starting *any* modeling project
  - Try to write down the relevant variables
  - Draw a graph to show how the variables are related to each other

Bayes nets

- Strong sampling

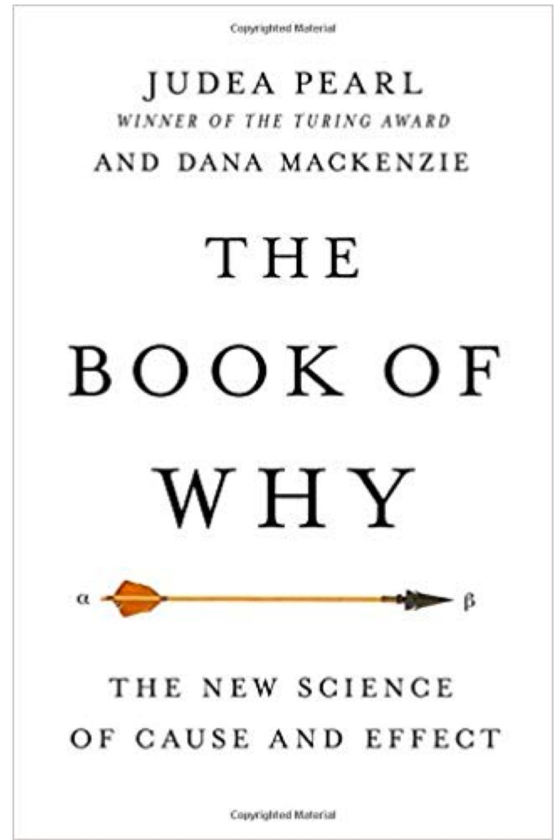


- Weak sampling



Bayes Nets and causal reasoning

- Bayes nets support reasoning about:
  - Interventions
  - Counterfactuals



Take home messages

- Bayes nets are:
  – a useful engineering tool
  – a tool for thinking

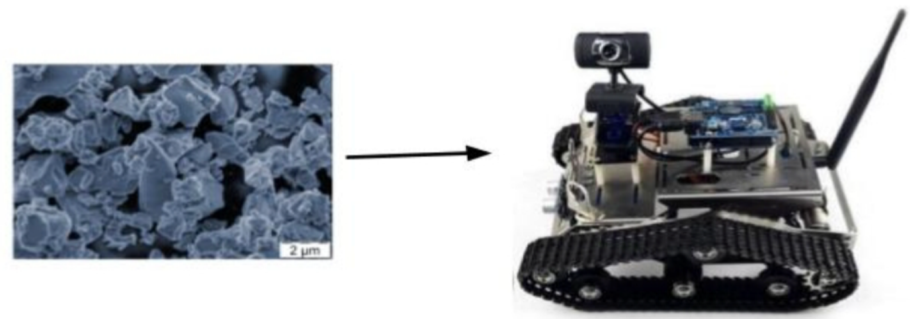
Module 6

SAMPLING FRAMES AND SPHERES OF SODOR

Sampling frames

The robot uses a grapple that can pick up small spheres and test them for the presence of plaxium

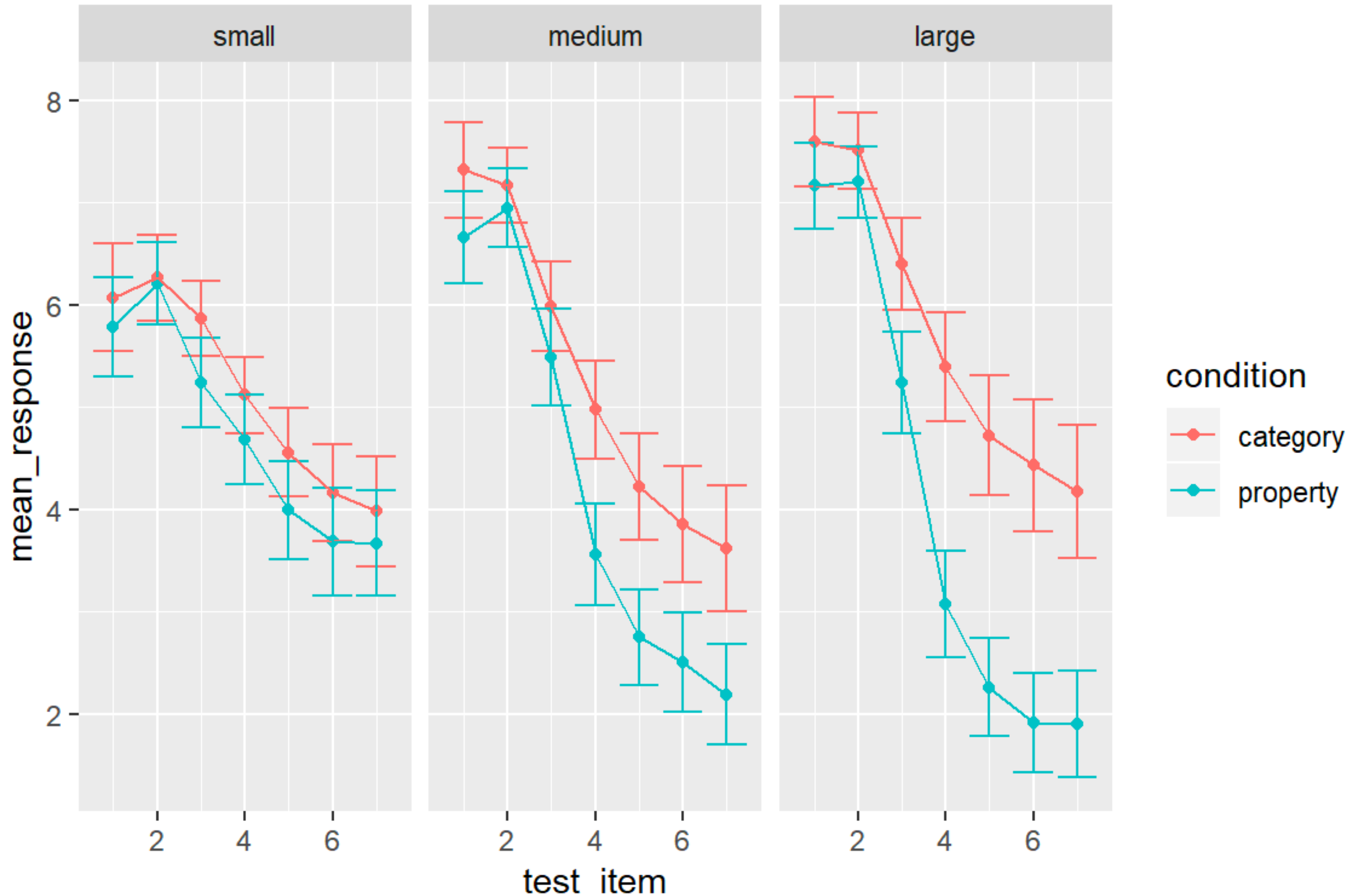
The robot uses a high-resolution scanner to detect plaxium, which triggers the camera to take an image whenever the plaxium coating is found



Category sampling

Property sampling

Human data



small
medium
large
mean_response
test_item
8
6
4
2
2
4
6
condition
category
property

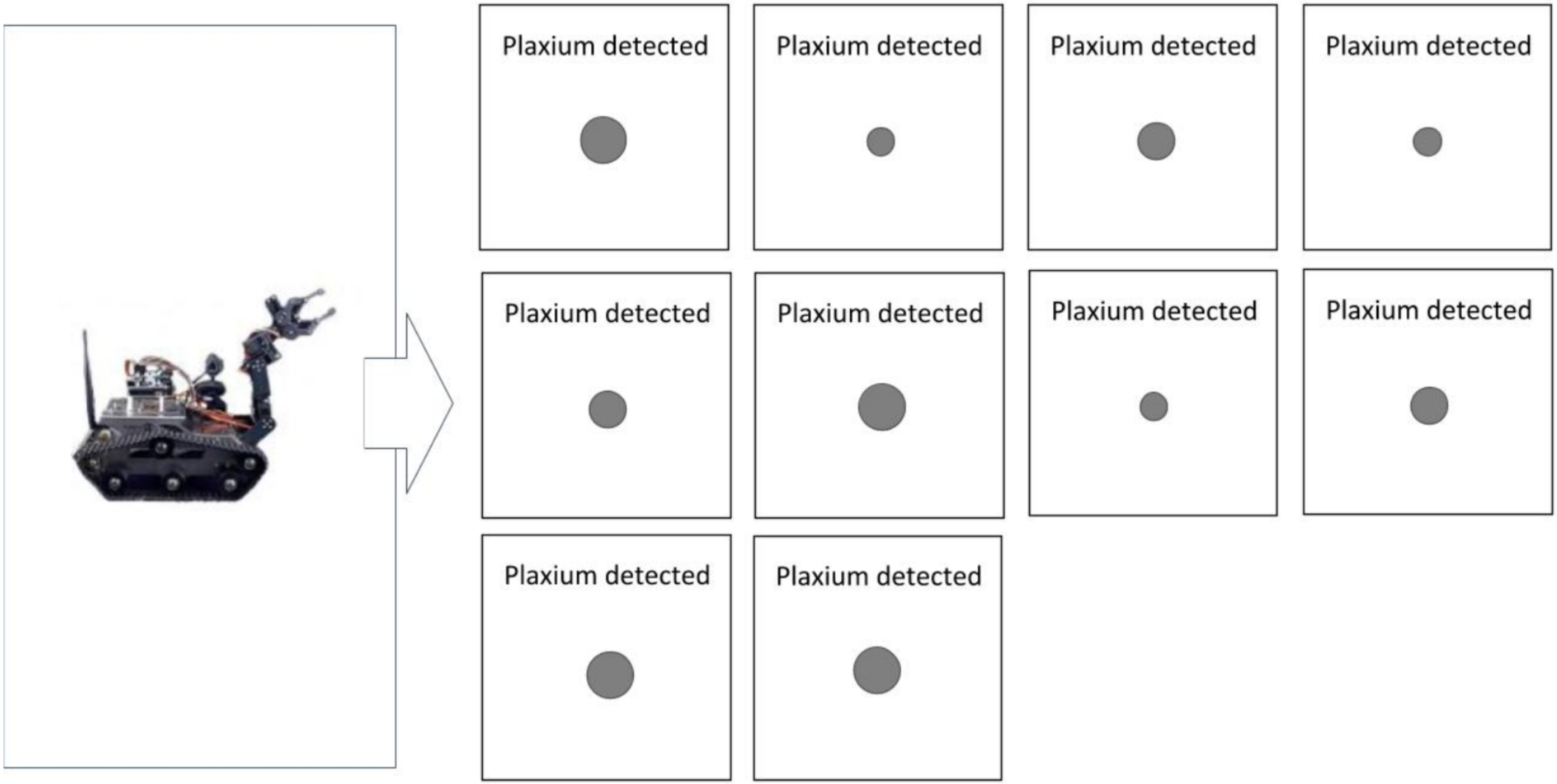
Building a Bayesian model

- What are the observed data?
- What are the hypotheses?
- What is the prior?
- What is the likelihood?

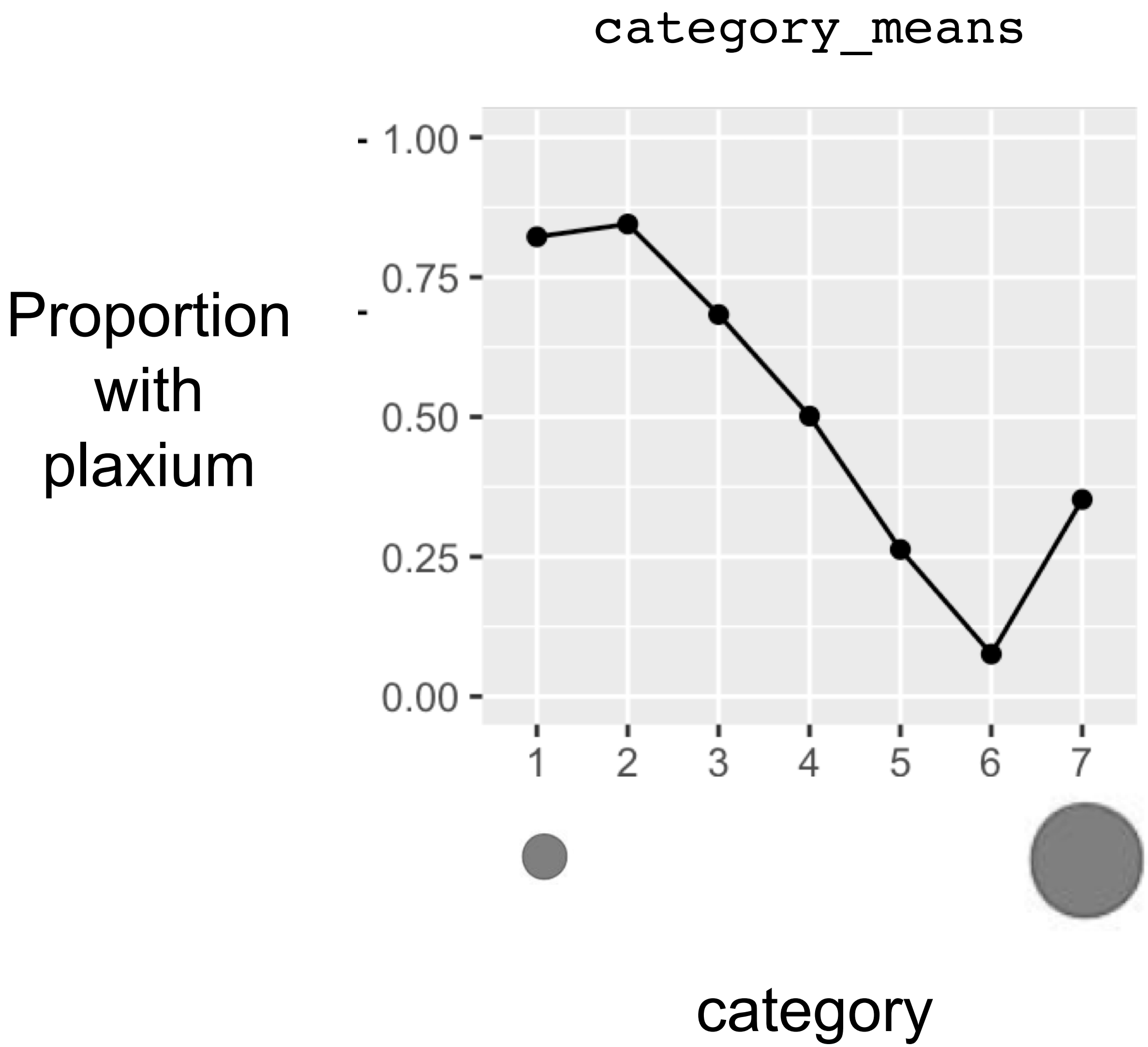
Posterior Likelihood Prior

$$\underbrace{P(h|d)}_{\text{Posterior}} \propto \underbrace{\mathbf{P(d|h)}}_{\text{Likelihood}} \underbrace{\mathbf{P(h)}}_{\text{Prior}}$$

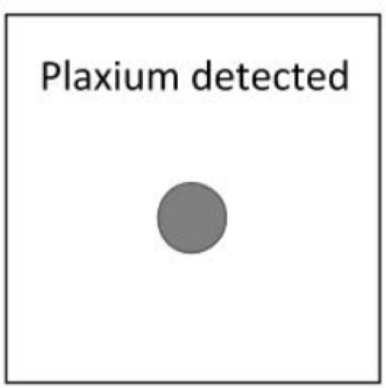
Observations



Hypotheses


category_means
Proportion with plaxium
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
category

Category sampling



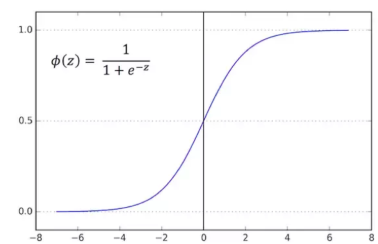

```
category_means
      ↓
plaxium ← **category**
```

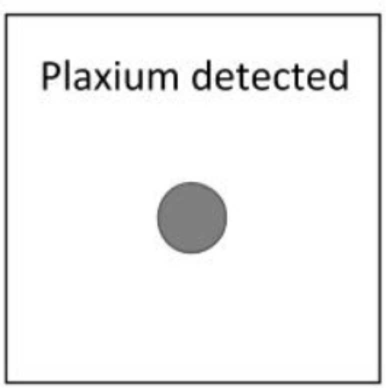
Category sampling

mean_gp

category_means

plaxium ← **category**

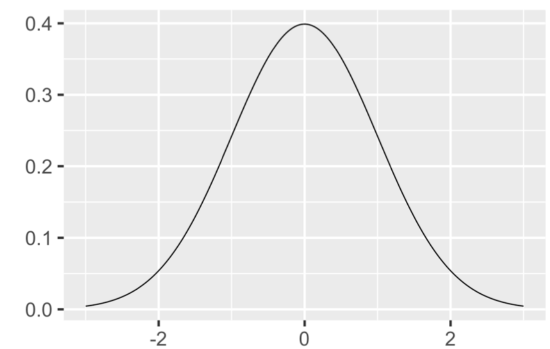




Gaussian distributions

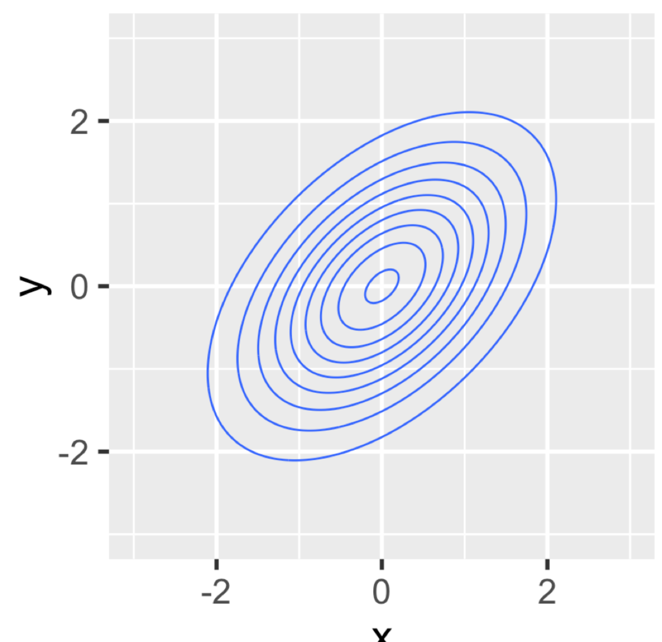
$$m = 0$$

$$\sigma = 1$$

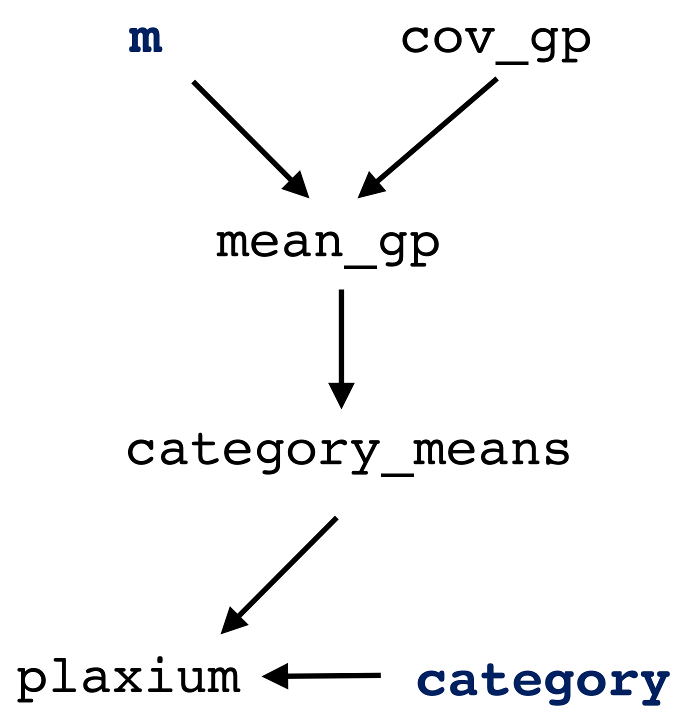


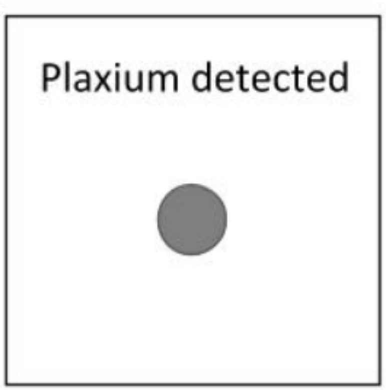
$$m = [0, 0]$$

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

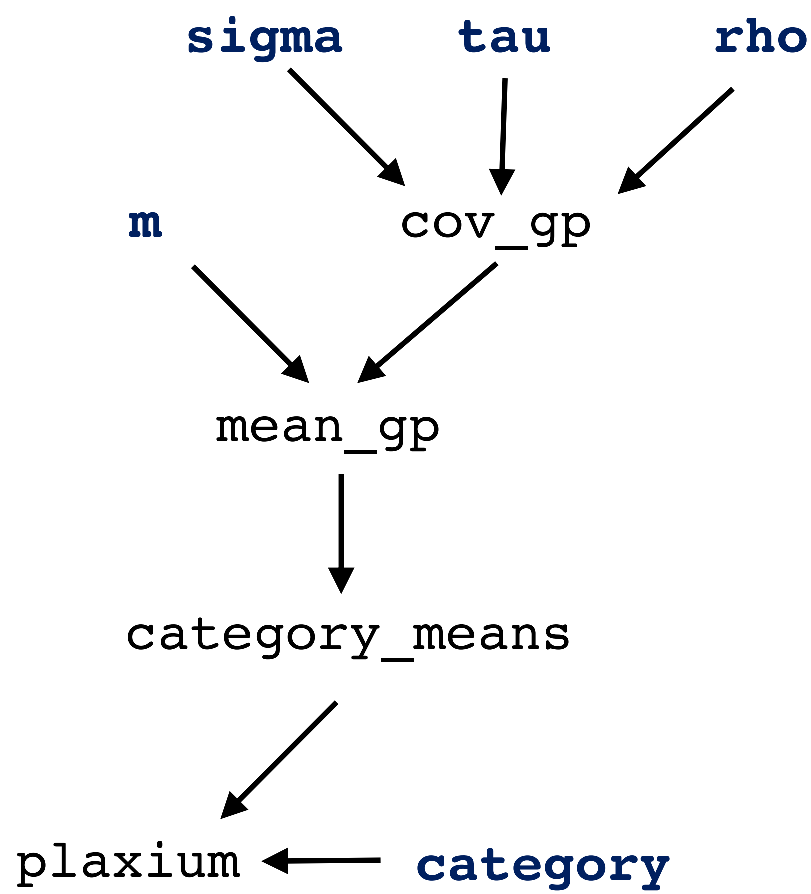


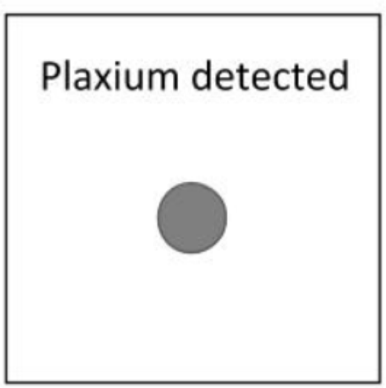
Category sampling



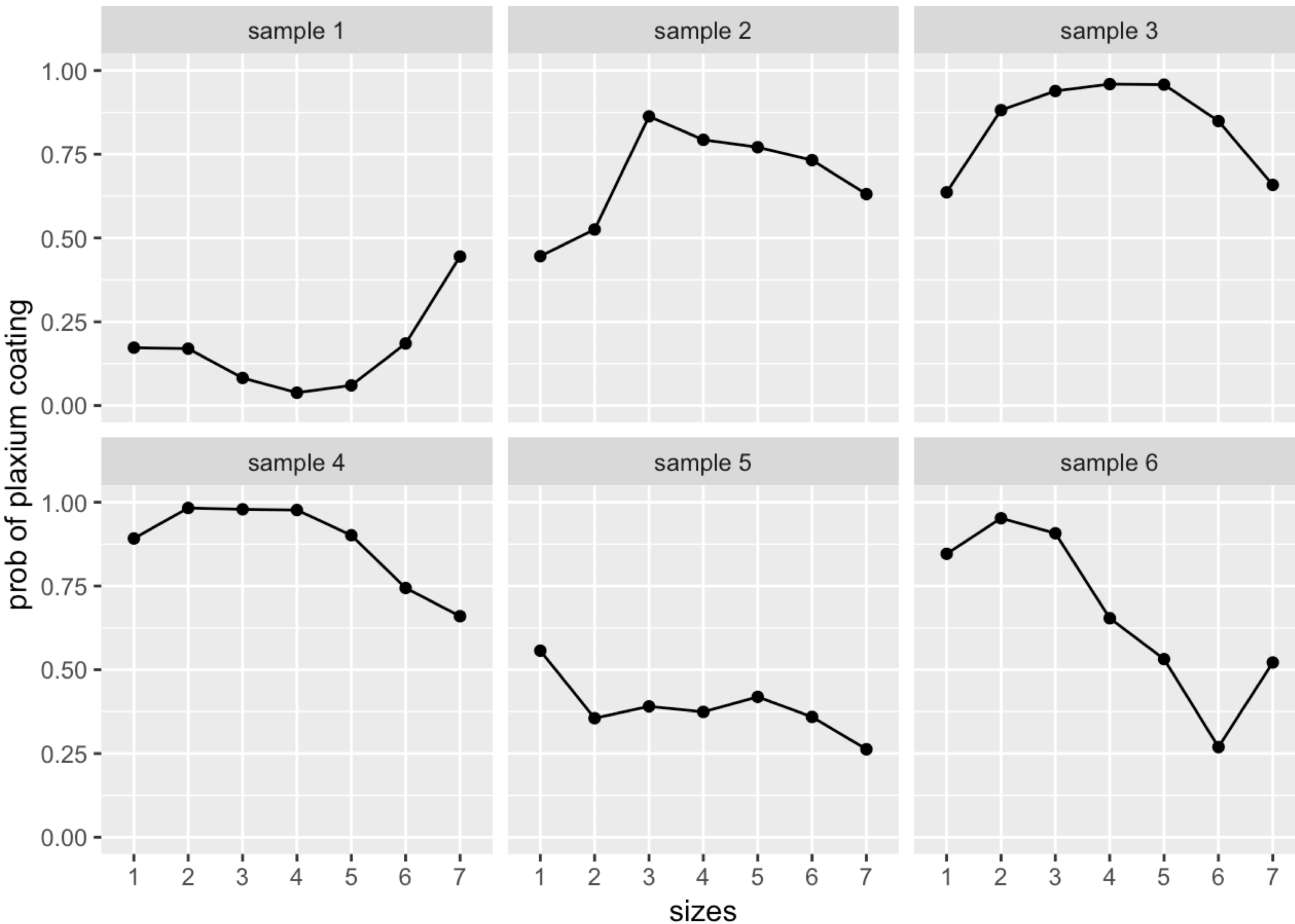


Category sampling

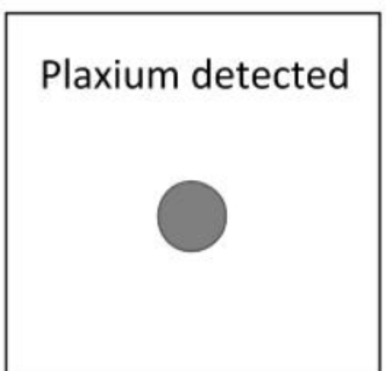




Prior on category means



Property sampling

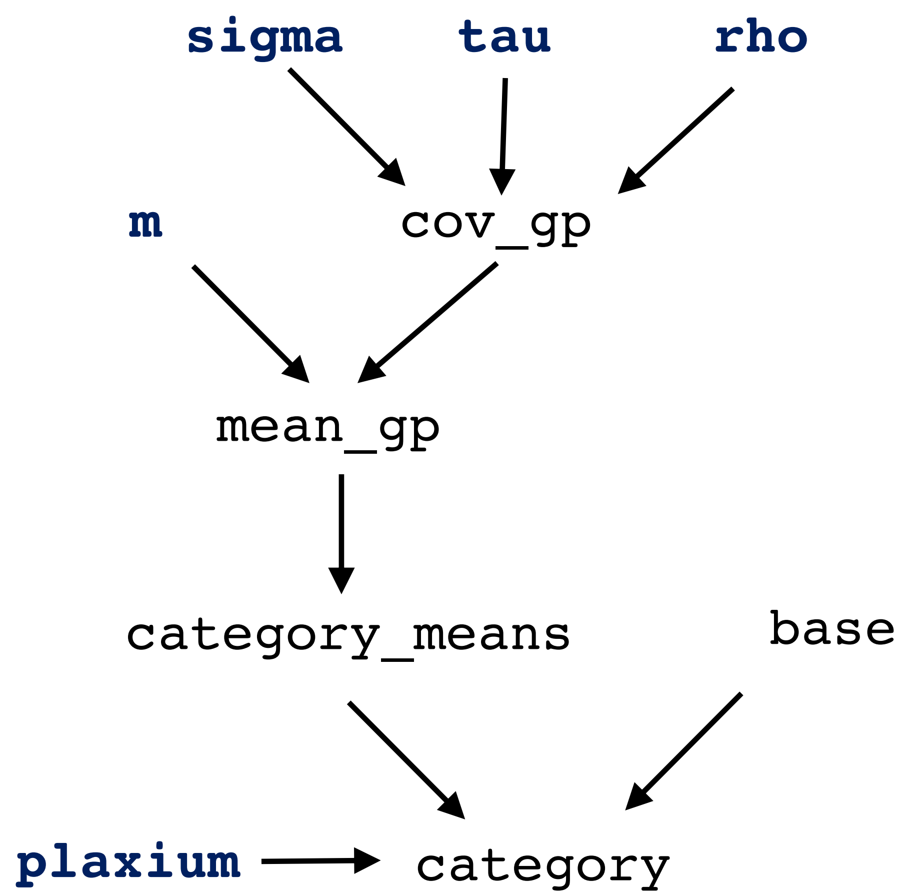


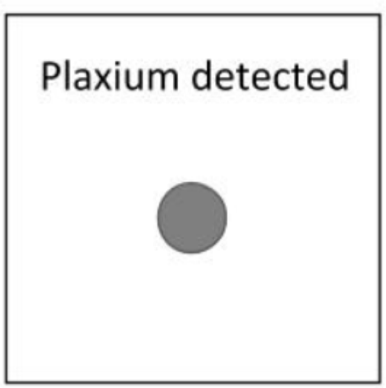
category_means

base

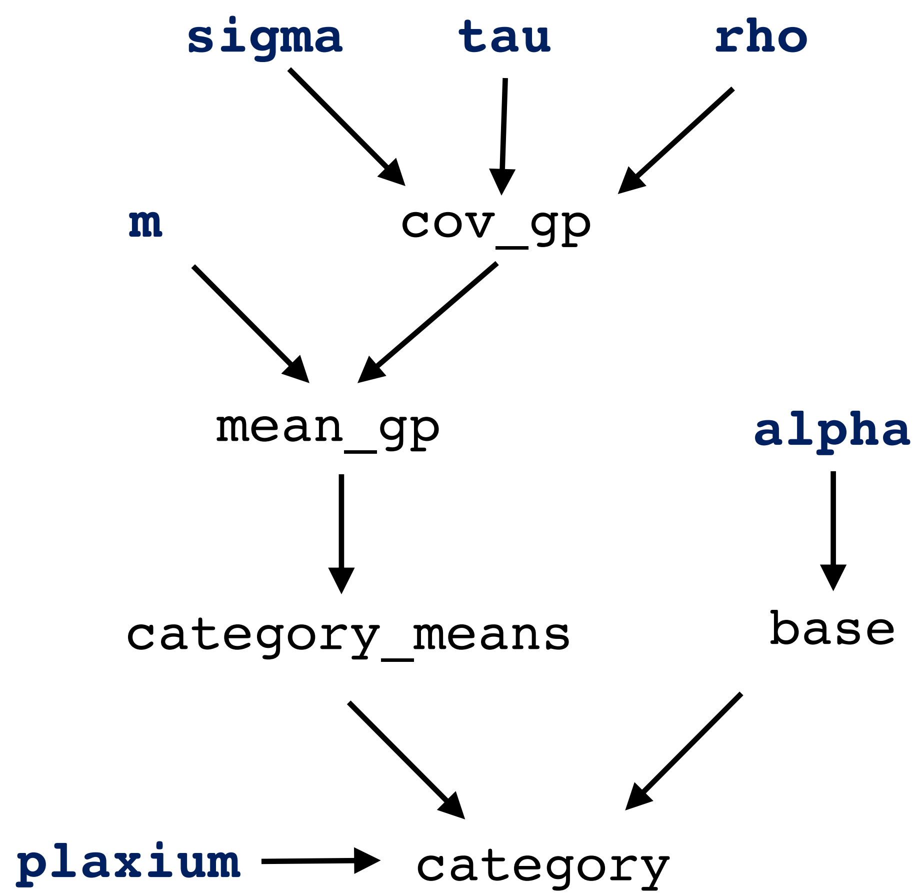
plaxium → category

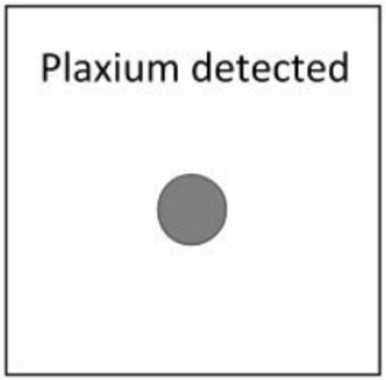
Property sampling





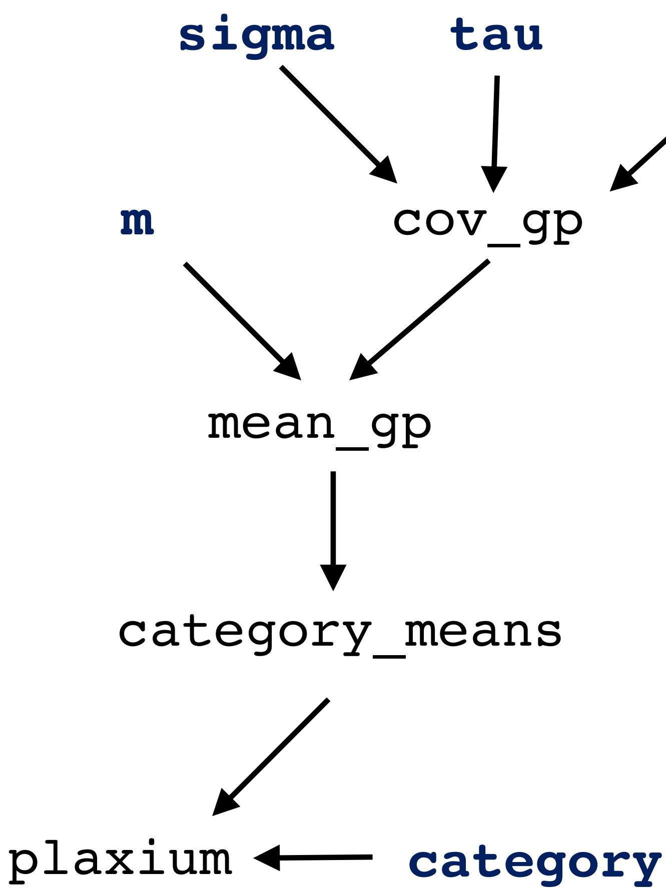
Property sampling





Inference with JAGS

Category sampling



category.bug

```
model {

  # mean and covariance matrix definin
  for(i in 1:ncat) {
    mean_gp[i] <- m
    cov_gp[i,i] <- (sigma^2) + (tau^2)
    for(j in (i+1):ncat) {
      cov_gp[i,j] <- (tau^2) * exp(-rh
      cov_gp[j,i] <- cov_gp[i,j]
    }
  }

  # sample a function from the Gaussia
  cov_gp_inv <- inverse(cov_gp)
  f ~ dmnorm(mean_gp, cov_gp_inv)

  # pass f through logistic function t
  for(i in 1:ncat) {
    category_means[i] <- 1/(1+exp(-f[i
  }
```

Exercise: Sampling frames

Take home messages

- Bayes nets let you build relatively complex models out of simple pieces
- JAGS makes the implementation process relatively painless

TAKING STOCK

Today’s models

- How are the models we’ve discussed useful?
- And what are their limitations?

Bayesian models in general

- Can Bayesian models be useful for thinking about problems you're interested in?

- Where might they be not so useful?