Bayes' Rule

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Fall, 2021

lecture 14

Quiz

Suppose we have the following data matrix:

$$X = \overbrace{\begin{bmatrix} - & \vec{x}_1 & - \\ - & \vec{x}_2 & - \\ & \vdots & \\ - & \vec{x}_N & - \end{bmatrix}}^{d} \left.\vphantom{\begin{bmatrix} a \\ a \\ a \\ a \end{bmatrix}}\right\} N$$

(1) What matrix do we form in order to compute the principal components of this data?

(2) Once we've formed that matrix, how do we get the principal components?

(3) What is the fraction of the total variance captured by the 1st principal component? (Write in terms of the singular values $s_1, \ldots, s_d$)

Suppose we want to use least-squares regression to find weights the $\vec{w}$ that map from X to an output vector Y.

(4) What are the residuals? (ie, write down an expression for the the vector of residuals between the linear prediction and the output vector)

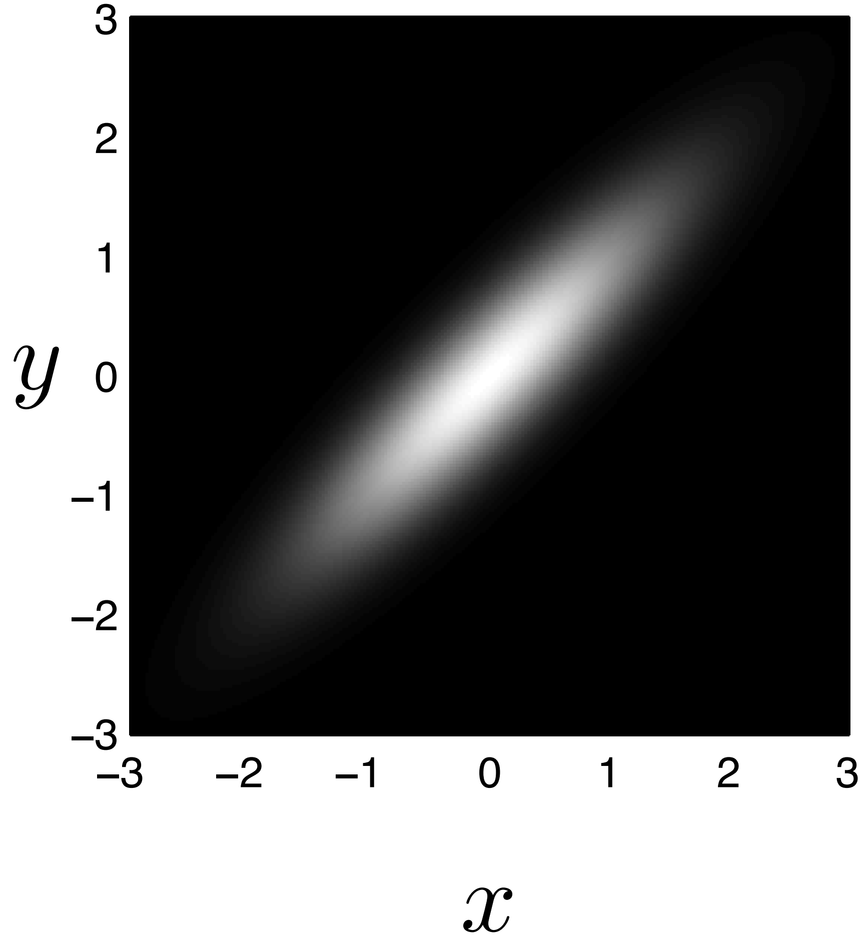
(5) Suppose $[p_1\ p_2\ p_3\ p_4]$ is a discrete probability distribution (PMF).
What **two facts** do we know about the values p_1, p_2, p_3, p_4?

Quick recap

- **Random variable** X takes on different values according to a probability distribution
- discrete: **probability mass function** (pmf)
- continuous: **probability density function** (pdf)
- **marginalization**: summing (“splatting”)
- **conditionalization:** “slicing”
- **expectation:** average of f(x) under P(x)

joint distribution

$$P(x, y)$$

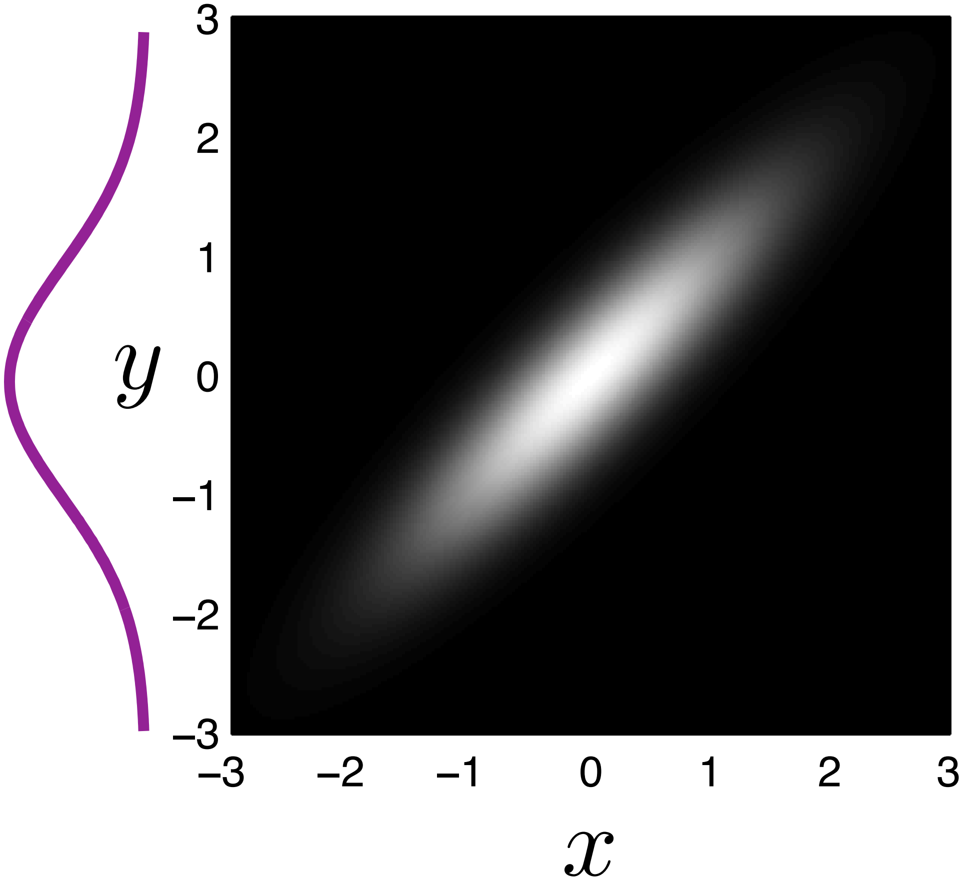


- positive
- sums to 1

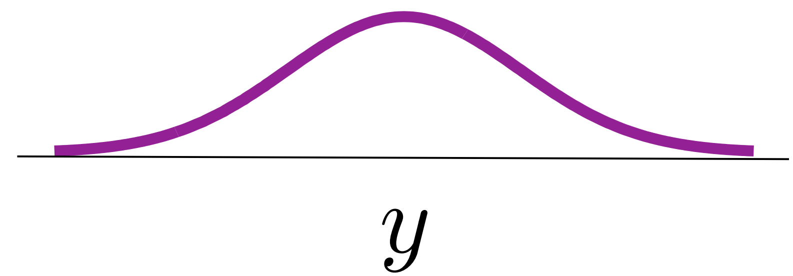
$$\iint P(x, y)\, dx\, dy = 1$$

marginalization ("integration")

$P(x, y)$

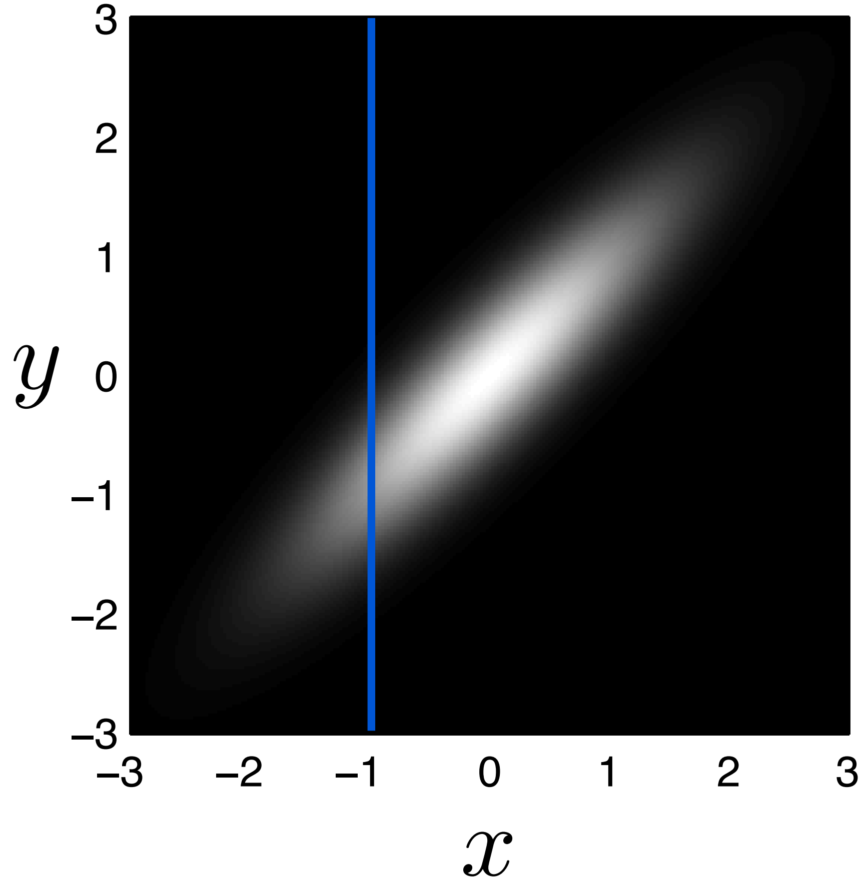


$$P(y) = \int P(x, y) dx$$



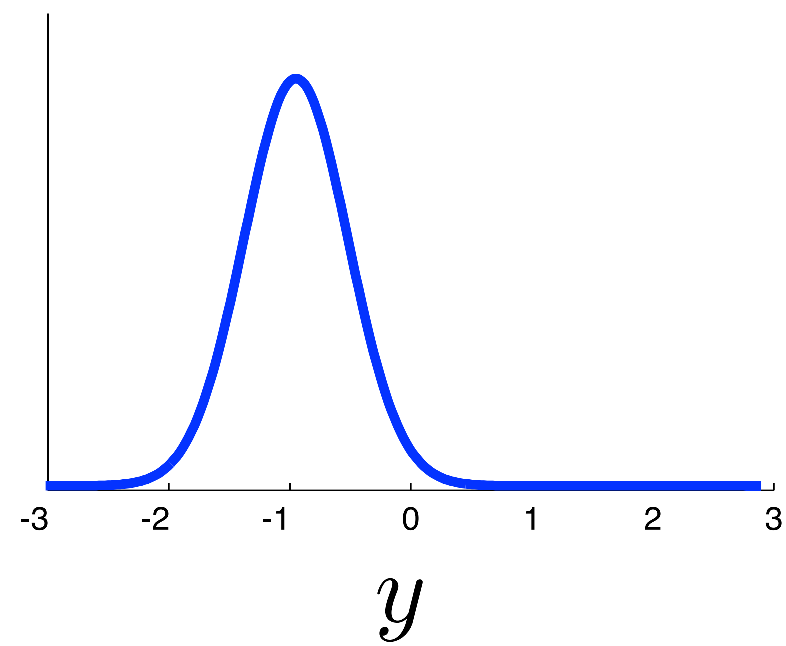
conditionalization ("slicing")

$$P(x, y)$$



$$P(y|x = -1) = \frac{P(y, x = -1)}{P(x = -1)}$$

("joint divided by marginal")



practice problems

P(x,y)

y \ x	1	2	3
1	0.01	0.05	0.04
2	0.08	0.4	0.32
3	0.01	0.05	0.04

1. Is this a joint probability distribution?
2. What is the marginal P(x)?
3. What is the marginal P(y)?
4. What is the conditional P(y | x = 1)?
5. What is the conditional P(x | y = 3)?

practice problems

P(x,y)

y \ x	1	2	3
1	0.01	0.05	0.04
2	0.08	0.4	0.32
3	0.01	0.05	0.04

1. Is this a joint probability distribution?
2. What is the marginal P(x)?
3. What is the marginal P(y)?
4. What is the conditional P(y | x = 1)?
5. What is the conditional P(x | y = 3)?

Expectations ("averages")

Expectation is the weighted average of a function (of a random variable) according to the distribution (of that random variable)

discrete

pmf

$$\mathbb{E}[f(x)] = \sum_i f(x_i)P(x_i)$$

or

continuous

pdf

$$\mathbb{E}[f(x)] = \int f(x)P(x)dx$$

It's really just a dot product! $\mathbb{E}[f(x)] = \vec{P} \cdot \vec{f}$ $\quad \vec{P} = \begin{bmatrix} P(x_1) \\ \vdots \\ P(x_m) \end{bmatrix} \quad \vec{f} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix}$

Several important expectations:

1) Mean: $\mathbb{E}[x]$ - the average value of a random variable

"1st moment" (here we have simply f(x) = x)

if x is discrete, taking on N values: $$\mathbb{E}[x] = \sum_{i=1}^{N} x_i P(x_i)$$

example

$$\begin{array}{cc} x & P(x) \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} \end{array}$$

$$\mathbb{E}[x] = \vec{x} \cdot \vec{P}$$

$$= 1(0.5) + 2(0.3) + 3(0.2) = 1.7$$

Several important expectations:

1) Mean: $\mathbb{E}[x]$ - the average value of a random variable

"1st moment" (here we have simply f(x) = x)

if x is continuous: $$\mathbb{E}[x] = \int x P(x)\, dx$$

- can still think of this as a dot product between two (infinitely tall) vectors of x values and probilities

$$\mathbb{E}[x] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ \end{bmatrix} \bullet \begin{bmatrix} P(x_1) \\ P(x_2) \\ \vdots \\ \\ \end{bmatrix}$$

Several important expectations:

2) $\mathbb{E}[x^2]$ - the average value of squared random variable

"2nd moment" (here f(x) = x^2)

if x is discrete, taking on N values: $$\mathbb{E}[x^2] = \sum_{i=1}^{N} x_i^2 P(x_i)$$

example

$$\begin{array}{ccc} x & x^2 & P(x) \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} & \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} \end{array}$$

$$\mathbb{E}[x^2] = \overrightarrow{x^2} \cdot \overrightarrow{P}$$

$$= 1(0.5) + 4(0.3) + 9(0.2) = 3.5$$

Several important expectations:

3) variance: $\mathbb{E}[(x - \mathbb{E}[x])^2]$

(average squared difference between x and its mean)

mean $\mathbb{E}[x]$

if x is discrete: $\mathrm{var}(x) = \sum_{i=1}^{N} (x_i - \mu)^2 P(x_i)$

if x is continuous: $\mathrm{var}(x) = \int (x - \mu)^2 P(x)\, dx$

practice problems

P(x,y)

y \ x	1	2	3
1	0.01	0.05	0.04
2	0.08	0.4	0.32
3	0.01	0.05	0.04

Q: What are the mean and variance of x?

1) compute P(x)

2) compute $\mathbb{E}[x] = \sum_{x=1}^{3} xP(x)$

3) compute $\mathbb{E}[(x - \mathbb{E}[x])^2)$

Monte Carlo evaluation of an expectation:

1. draw samples from distribution: $x^{(i)} \sim P(x)$ for i = 1 to N

2. average $\mathbb{E}[f(x)] \approx \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)})$

<u>For example, to evaluate the mean:</u>

1) sample values $x^{(i)}$ from P(x)

2) take the average of those samples $\frac{1}{N} \sum x^{(i)}$

A little math: **Bayes' rule**

- very simple formula for manipulating probabilities

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$

conditional probability
"probability of X given that Y occurred"

probability of X

probability of Y

simplified form: $P(X \mid Y) \propto P(Y \mid X)\, P(X)$

A little math: **Bayes' rule**

$$P(X \mid Y) \propto P(Y \mid X)\, P(X)$$

Example: 2 coins

- one coin is fake: "heads" on both sides (H / H)
- one coin is standard: (H / T)

You grab one of the coins at random and flip it. It comes up "heads". What is the probability that you're holding the fake?

$$p(\text{Fake} \mid H) \propto p(H \mid \text{Fake})\, p(\text{Fake})$$

$$(1) \quad (\tfrac{1}{2}) \quad = \tfrac{1}{2} \qquad \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{4}} = \frac{2}{3}$$

$$p(\text{Nrml} \mid H) \propto p(H \mid \text{Nrml})\, p(\text{Nrml})$$

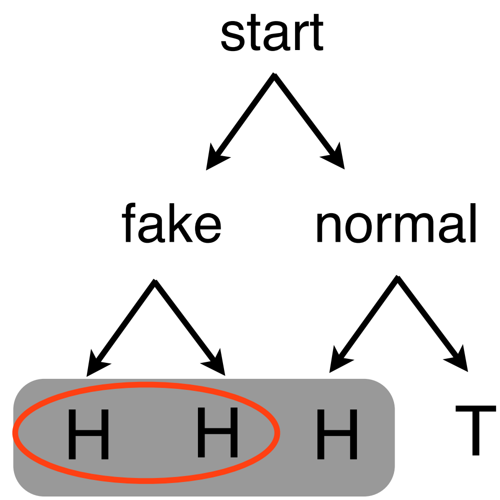
$$(\tfrac{1}{2}) \quad (\tfrac{1}{2}) \quad = \tfrac{1}{4} \qquad \frac{\frac{1}{4}}{\frac{1}{2}+\frac{1}{4}} = \frac{1}{3}$$

probabilities must sum to 1

A little math: **Bayes' rule**

$$P(X \mid Y) \propto P(Y \mid X)\, P(X)$$

Example: 2 coins



$$p(\text{Fake} \mid H) \propto p(H \mid \text{Fake})\, p(\text{Fake})$$

$$(1) \quad (½) \quad = ½ \qquad \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{4}} = \frac{2}{3}$$

$$p(\text{Nrml} \mid H) \propto p(H \mid \text{Nrml})\, p(\text{Nrml})$$

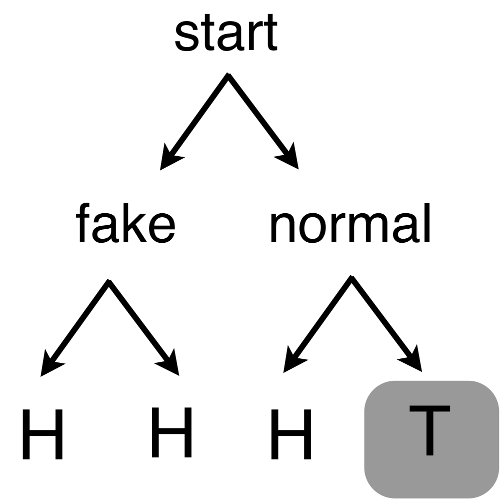
$$(½) \quad (½) \quad = ¼ \qquad \frac{\frac{1}{4}}{\frac{1}{2}+\frac{1}{4}} = \frac{1}{3}$$

probabilities must sum to 1

A little math: **Bayes' rule**

$$P(X \mid Y) \propto P(Y \mid X)\, P(X)$$

Example: 2 coins



Experiment #2: It comes up "tails".
What is the probability that you're holding the fake?

$$p(\text{Fake} \mid T) \propto p(T \mid \text{Fake})\; p(\text{Fake})$$

$$(\,0\,) \qquad (\,½\,) \qquad = 0$$

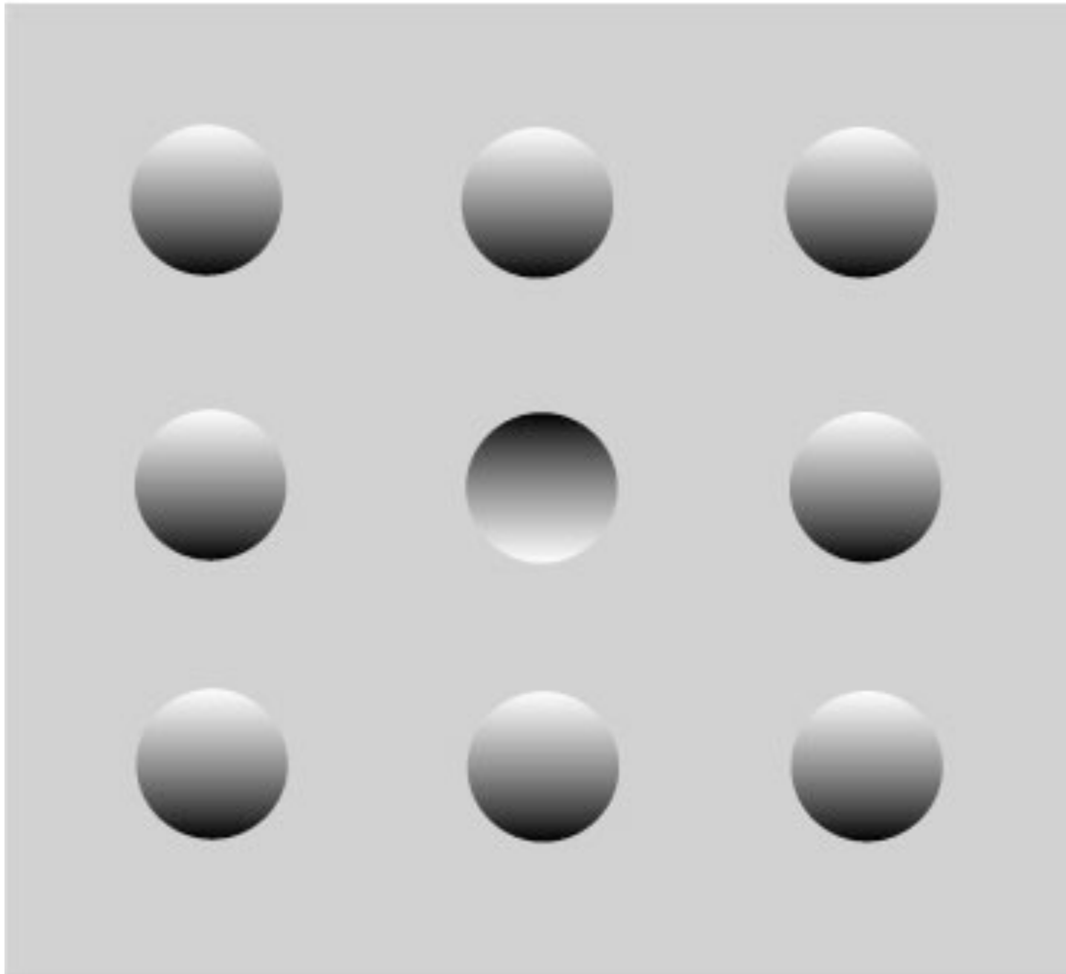
= 0

$$p(\text{Nrml} \mid T) \propto p(T \mid \text{Nrml})\; p(\text{Nrml})$$

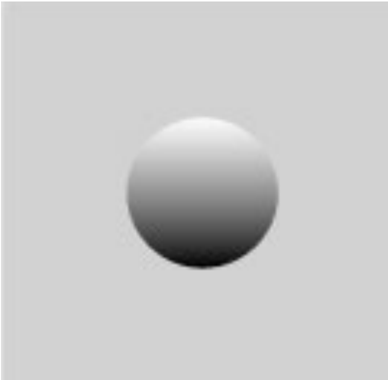
probabilities must
sum to 1

$$(\,½\,) \qquad (\,½\,) \qquad = ¼$$

= 1



Is the middle circle popping “out” or “in”?



$P(\text{image} \mid \text{OUT} \;\&\; \text{light is above}) = 1$
$P(\text{image} \mid \text{IN} \;\&\; \text{Light is below}) = 1$

- Image equally likely to be OUT or IN given sensory data alone

What we want to know: $P(\text{OUT} \mid \text{image})$ vs. $P(\text{IN} \mid \text{image})$

Apply Bayes' rule:

$$P(\text{OUT} \mid \text{image}) \propto P(\text{image} \mid \text{OUT} \;\&\; \text{light above}) \times \overbrace{P(\text{OUT}) \times P(\text{light above})}^{\text{prior}}$$

$$P(\text{IN} \mid \text{image}) \propto P(\text{image} \mid \text{IN} \;\&\; \text{light below}) \times P(\text{IN}) \times P(\text{light below})$$

Which of these is greater?

Bayesian Models for Perception

Bayes' rule: $P(X \mid Y) \propto P(Y \mid X)\,P(X)$

Formula for computing: P(what's in the world | sensory data)

(This is what our brain wants to know!)

X Y

P(world | sense data) ∝ P(sense data | world) P(world)

Posterior

(resulting beliefs about the world)

Likelihood

(given by laws of physics; ambiguous because many world states could give rise to same sense data)

Prior

(given by past experience)

Helmholtz: perception as "optimal inference"

helmholtz 1821-1894

"Perception is our best guess as to what is in the world, given our current sensory evidence and our prior experience."

$$P(\text{world} \mid \text{sense data}) \propto P(\text{sense data} \mid \text{world})\, P(\text{world})$$

Posterior

(resulting beliefs about the world)

Likelihood

(given by laws of physics; ambiguous because many world states could give rise to same sense data)

Prior

(given by past experience)

Helmholtz: perception as "optimal inference"

helmholtz 1821-1894

"Perception is our **best guess** as to what is in the world, given our current **sensory evidence** and our **prior experience**."

$$P(\text{world} \mid \text{sense data}) \propto P(\text{sense data} \mid \text{world})\, P(\text{world})$$

Posterior

(resulting beliefs about the world)

Likelihood

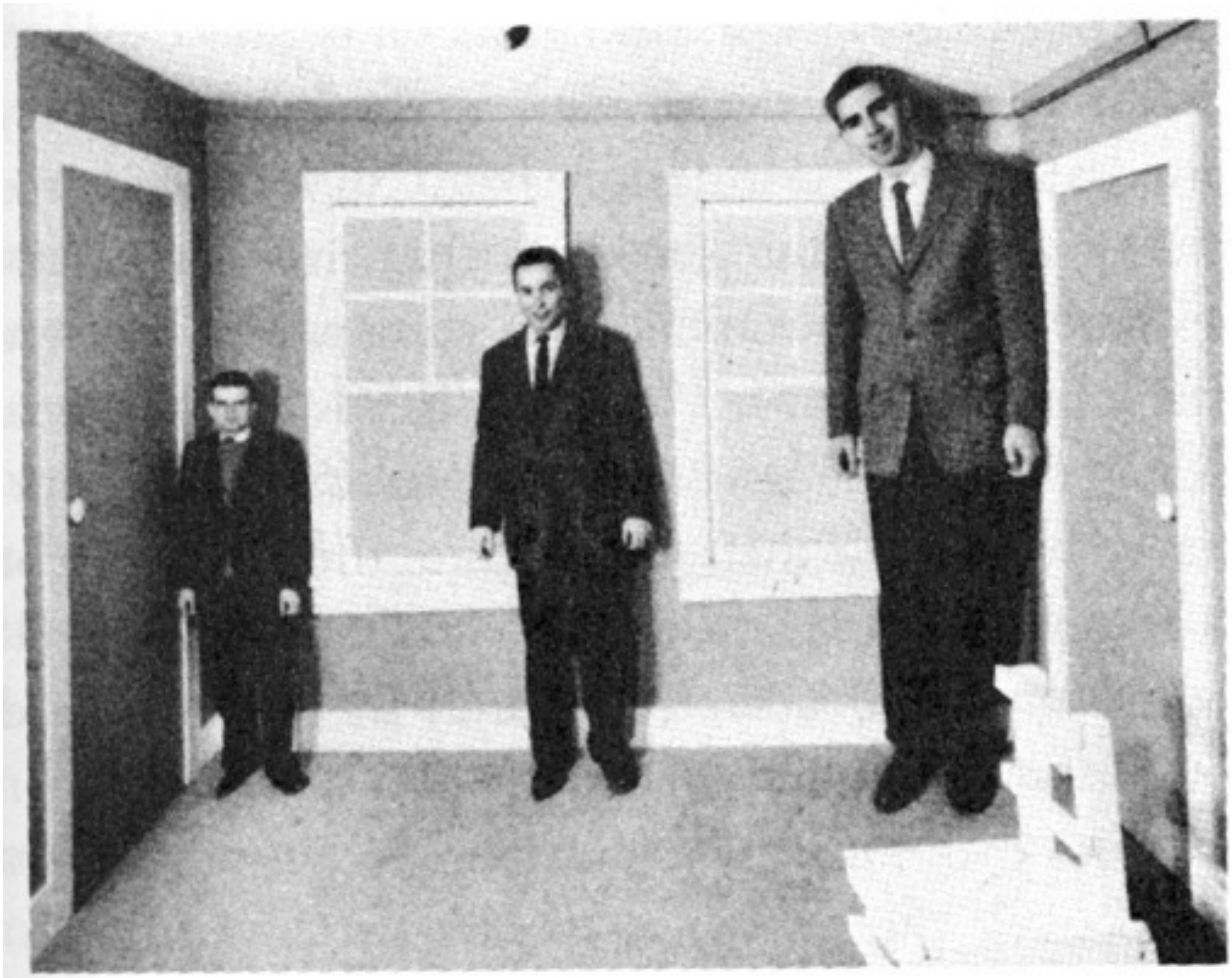
(given by laws of physics; ambiguous because many world states could give rise to same sense data)

Prior

(given by past experience)

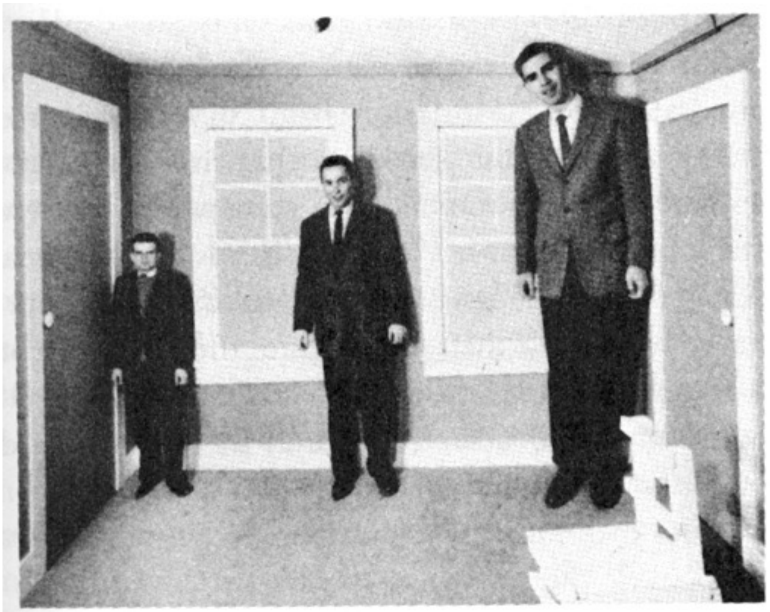
Many different 3D scenes can give rise to the same 2D retinal image

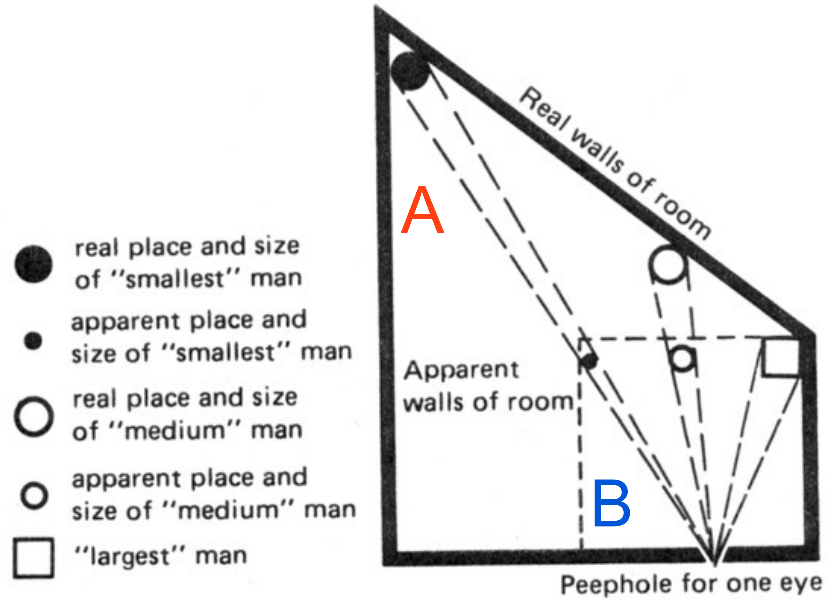
The Ames Room



Many different 3D scenes can give rise to the same 2D retinal image

The Ames Room





How does our brain go about deciding which interpretation?

P(image | A) and P(image | B) are equal!   (both A and B could have generated this image)

Let's use Bayes' rule:

P(A | image) = P(image | A) P(A) / Z
P(B | image) = P(image | B) P(B) / Z

Hollow Face Illusion

http://www.richardgregory.org/experiments/

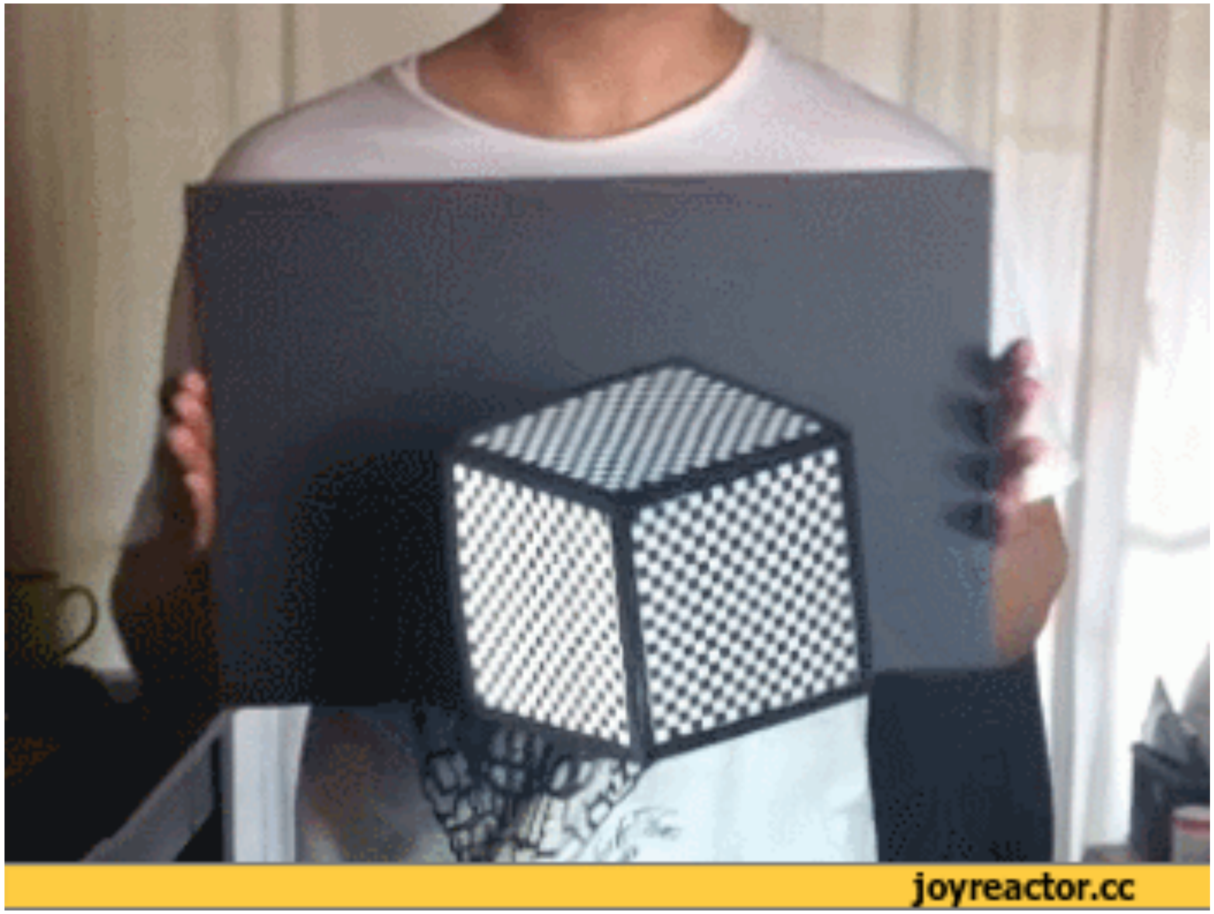
Hollow Face Illusion

Hypothesis #1: face is concave
Hypothesis #2: face is convex

$$P(\text{convex}|\text{video}) \propto P(\text{video}|\text{convex})\, P(\text{convex})$$
$$P(\text{concave}|\text{video}) \propto P(\text{video}|\text{concave})\, P(\text{concave})$$

posterior likelihood prior

$P(\text{convex}) > P(\text{concave}) \Rightarrow$ posterior probability of convex is higher (which determines our percept)



- prior belief that objects are convex is SO strong we can’t over-ride it, even when we know it’s wrong!

(So your brain knows Bayes’ rule even if you don’t!)

Summary

- marginalization (splatting)
- conditionalization (slicing)
- expectation (averaging)

- Monte Carlo evaluation of expectation
- Bayes' rule (prior, likelihood, posterior)
- Bayesian models of perception