AN INTRODUCTION TO BAYESIAN INFERENCE

By Ruth Angus

8 reasons to use BAYESIAN INFERENCE

You can use Bayesian inference if <u>any</u> of these are true

8 reasons to use BAYESIAN INFERENCE

You can use Bayesian inference if any of these are true

1. You want to infer the properties of a physical process by fitting a model to data
2. You want to be explicit about your prior assumptions
3. You take a probabilistic view of the world
4. Your problem is complicated
5. Your problem is hierarchical/structured/layered
6. You would like to compare alternative models
7. You're working at the edge
8. You find frequentist statistics hard

JOINT, MARGINAL and CONDITIONAL PROBABILITIES

JOINT, MARGINAL and CONDITIONAL PROBABILITIES

JOINT, MARGINAL and CONDITIONAL PROBABILITIES

"THE JOINT PROBABILITY OF BAGELS AND CREAM CHEESE"

p(X = 🍩, Y = 🥛🧀)

JOINT, MARGINAL and CONDITIONAL PROBABILITIES

"THE MARGINAL PROBABILITY OF CREAM CHEESE"

p(Y = 🥛🧀)

JOINT, MARGINAL and CONDITIONAL PROBABILITIES

"THE PROBABILITY OF CREAM CHEESE, GIVEN BAGELS"

p(Y = 🥛🧀 | X = 🍩)

OR

"THE PROBABILITY OF CREAM CHEESE, CONDITIONED ON X=🍩"

X

Y		
 	5	1
	3	7

Y \ X	🍩	🍞
🥛🧀	5	1
🥑	3	7

JOINT

1. p(X = 🍞, Y = 🥑) = 7/16
2. p(X = 🍩, Y = 🥛🧀) = 5/16

MARGINAL

3. p(Y = 🥑) = 10/16 = 5/8
4. p(X = 🍩) = 8/16 = 1/2

CONDITIONAL

5. p(X = 🍩| Y = 🥑) = 3/10
6. p(Y = 🥛🧀 | X = 🍩) = 5/8

THE SUM RULE

(MARGINALIZATION)

How did you calculate p(Y = 🥑)?

$$p(Y = \text{🥑}) = p(X = \text{🍩}, Y = \text{🥑}) + p(X = \text{🍞}, Y = \text{🥑})$$

$$p(Y = \text{🥑}) = \sum_{\text{breads}, i=1}^{L} p(X = \text{bread}_i, Y = \text{🥑})$$

(L = number of breads)

$$p(Y=Y_j) = \sum_{i=1}^{L} p(X=X_i, Y=Y_j)$$

THE SUM RULE

(MARGINALIZATION)

How did you calculate p(Y = 🥑)?

$$p(Y = \text{🥑}) = p(X = \text{🍩}, Y = \text{🥑}) + p(X = \text{🍞}, Y = \text{🥑})$$

$$p(Y = \text{🥑}) = \sum_{\text{breads}, i=1}^{L} p(X = \text{bread}_i, Y = \text{🥑})$$

(L = number of breads)

$$p(X=X_i) = \sum_{j=1}^{K} p(X=X_i, Y=Y_j)$$

(K = number of spreads)

THE SUM RULE

(MARGINALIZATION)

$$p(X=x_i) = \sum_{j=1}^{K} p(X=x_i, Y=y_j)$$

THE PRODUCT RULE

Going from a joint to a conditional distribution

How did you calculate p(X = 🍩| Y = 🥑) ?

Y \ X	🍩	🍞
🥛 🧀	5	1
🥑	3	7

THE PRODUCT RULE

Going from a joint to a conditional distribution

How did you calculate $p(X=X_i \mid Y=Y_j)$? $=n_{ij}/r_j$

		X	
		X_1	X_2
Y	Y_1	n_{11}	n_{12}
	Y_2	n_{21}	n_{22}

Q: write down $P(Y=Y_j|X=X_i)$

$=n_{ij}/c_i$

THE PRODUCT RULE

Going from a joint to a conditional distribution

$$p(X=x_i, Y=y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \frac{c_i}{N}$$

$$p(X=x_i, Y=y_j) = p(Y=y_j|X=x_i)\, p(X=x_i)$$

THE PRODUCT RULE

Going from a joint to a conditional distribution

$$p(X=x_i\,,\,Y=y_j) = p(Y=y_j|X=x_i)\,p(X=x_i)$$

LET'S SIMPLIFY THIS NOTATION

$$p(X=x_i, Y=y_j) = p(Y=y_j|X=x_i)\, p(X=x_i)$$

LET'S SIMPLIFY THIS NOTATION

$$p(X,Y) = p(Y|X)\, p(X)$$

NOW WE CAN DERIVE BAYES' RULE

USING THE PRODUCT RULE

$$p(X,Y) = p(Y|X)\, p(X)$$

DERIVE BAYES' RULE

$$p(X|Y) = \frac{p(Y|X)\, p(X)}{p(Y)}$$

BAYES' RULE

$$p(X|Y) = \frac{p(Y|X)\, p(X)}{p(Y)}$$

BAYES' RULE

$$\underbrace{p(\vartheta|D)}_{\text{posterior}} = \frac{\overbrace{p(D|\vartheta)}^{\text{likelihood}}\ \underbrace{p(\vartheta)}_{\text{prior}}}{\underbrace{p(D)}_{\text{evidence}}}$$

evidence

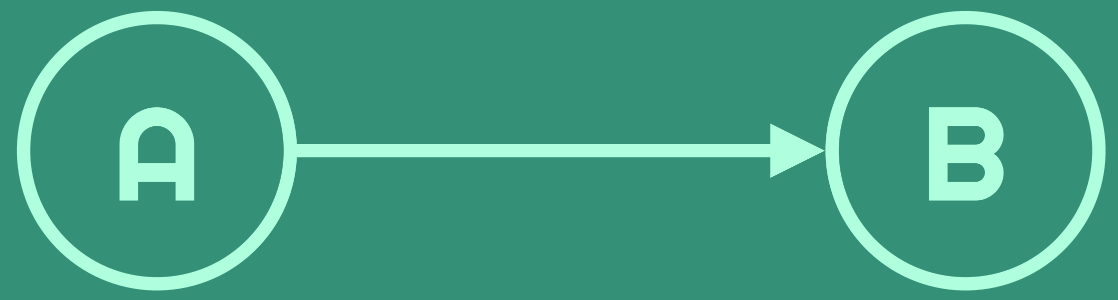
or

fully marginalized likelihood

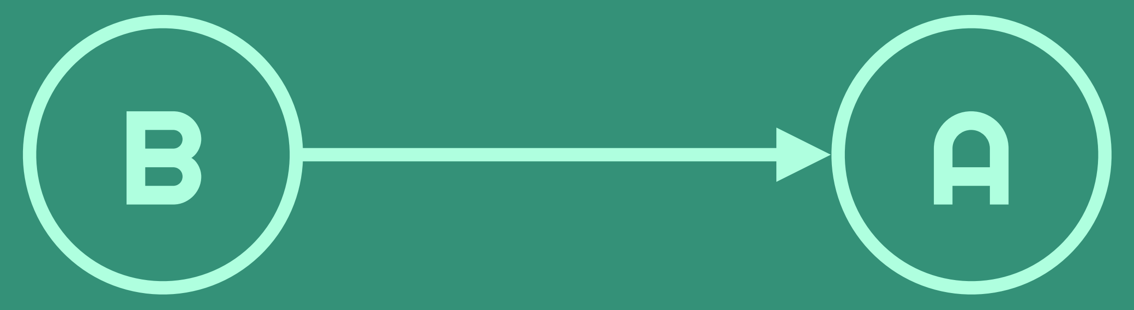
BAYES' RULE

$$\underbrace{p(\vartheta|D)}_{\text{posterior}} = \frac{\overbrace{p(D|\vartheta)}^{\text{likelihood}}\ \underbrace{p(\vartheta)}_{\text{prior}}}{\underbrace{\int p(D|\vartheta)\ p(\vartheta)d\vartheta}_{\substack{\text{evidence} \\ \text{or} \\ \text{fully marginalized likelihood}}}}$$

A BRIEF ASIDE:
PROBABILISTIC GRAPHICAL MODELS
[PGMs]

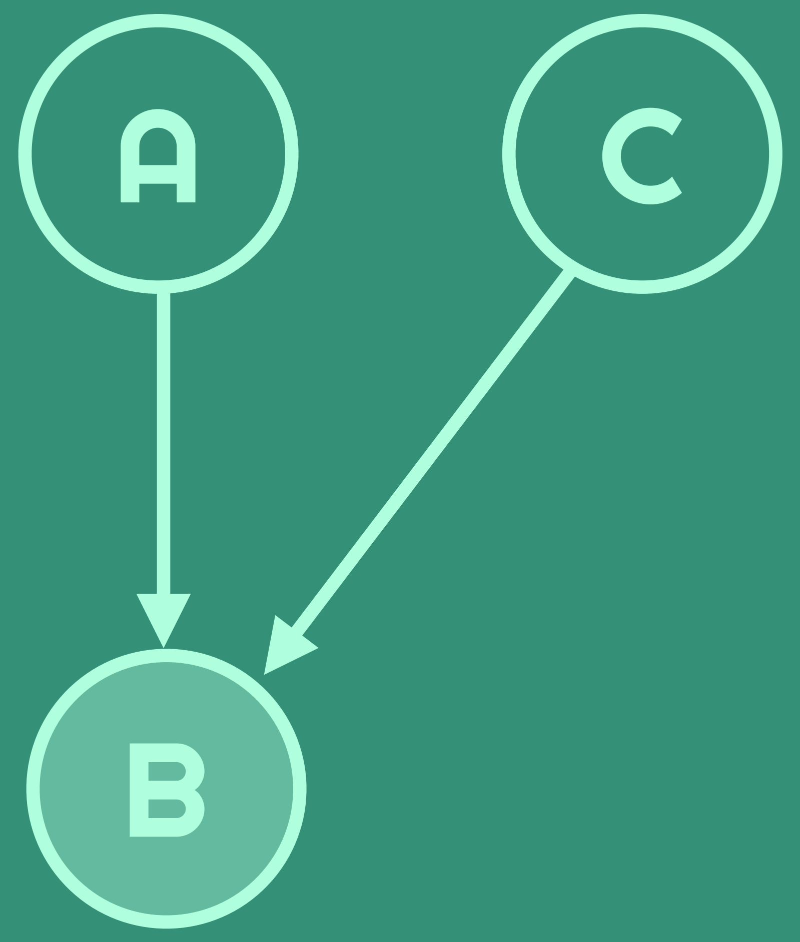


$$p(A,B) = p(A)\,p(B|A)$$



$$p(A,B) = p(B)\, p(A|B)$$

$$p(A,B) = p(A)\,p(B)$$



$$p(A, B, C) = p(A)\, p(C)\, p(B|A,C)$$

A GOOD REFERENCE:

Pattern recognition and machine learning, Bishop

GO TO NOTEBOOK

for an example of Bayesian inference

github.com/ruthangus/Probabilistic-inference-tutorial/