

# AN INTRODUCTION TO BAYESIAN INFERENCE

By Ruth Angus

# 8 reasons to use BAYESIAN INFERENCE

You can use Bayesian inference if any of these are true

1. You have a hypothesis that you want to test

2. You have data that you can use to test your hypothesis

3. You have a way to calculate the probability of your hypothesis being true

4. You have a way to calculate the probability of your hypothesis being false

5. You have a way to calculate the probability of your hypothesis being true given the data

6. You have a way to calculate the probability of your hypothesis being false given the data

7. You have a way to calculate the probability of your hypothesis being true given the data and the probability of your hypothesis being false given the data

8. You have a way to calculate the probability of your hypothesis being true given the data and the probability of your hypothesis being false given the data

# 8 reasons to use BAYESIAN INFERENCE

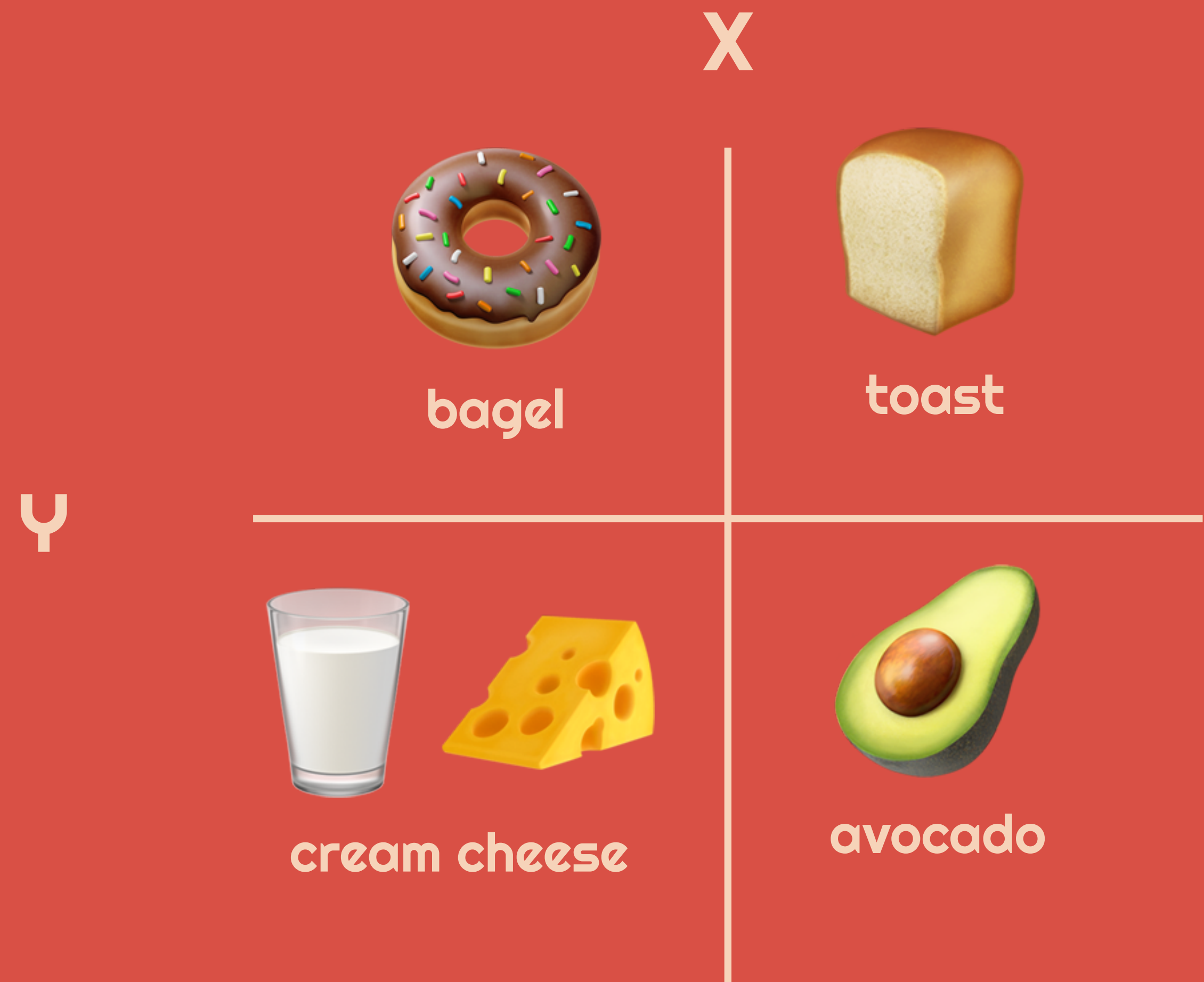
You can use Bayesian inference if any of these are true

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



# JOINT, MARGINAL and CONDITIONAL PROBABILITIES

“THE JOINT PROBABILITY OF BAGELS AND CREAM CHEESE”

$$p(X = \text{🍩}, Y = \text{🥛🧀})$$

# JOINT, MARGINAL and CONDITIONAL PROBABILITIES

“THE MARGINAL PROBABILITY OF CREAM CHEESE”

$$p(Y = \text{🥛🧀})$$

# JOINT, MARGINAL and CONDITIONAL PROBABILITIES

“THE PROBABILITY OF CREAM CHEESE, GIVEN BAGELS”

$$p(Y = \text{🥛🧀} \mid X = \text{🍩})$$

OR

“THE PROBABILITY OF CREAM CHEESE, CONDITIONED ON  $X = \text{🍩}$ ”








X



Y



5	1
3	7

		X	
			
Y	 	5	1
		3	7

## JOINT

$$1. p(X = \text{bread}, Y = \text{avocado}) = 7/16$$

$$2. p(X = \text{donut}, Y = \text{milk} \text{ } \text{cheese}) = 5/16$$

## MARGINAL

$$3. p(Y = \text{avocado}) = 10/16 = 5/8$$

$$4. p(X = \text{donut}) = 8/16 = 1/2$$

## CONDITIONAL

$$5. p(X = \text{donut} \mid Y = \text{avocado}) = 3/10$$

$$6. p(Y = \text{milk} \text{ } \text{cheese} \mid X = \text{donut}) = 5/8$$

# THE SUM RULE

(MARGINALIZATION)

How did you calculate  $p(Y = \text{🥑})$ ?

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

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

(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  $p(Y = \text{🥑})$ ?

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

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

(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  $p(X = \text{🍩} \mid Y = \text{🥑})$  ?

		X	
		🍩	🍞
Y	🥛🧀	5	1
	🥑	3	7

# THE PRODUCT RULE

Going from a joint to a conditional distribution

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

		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

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

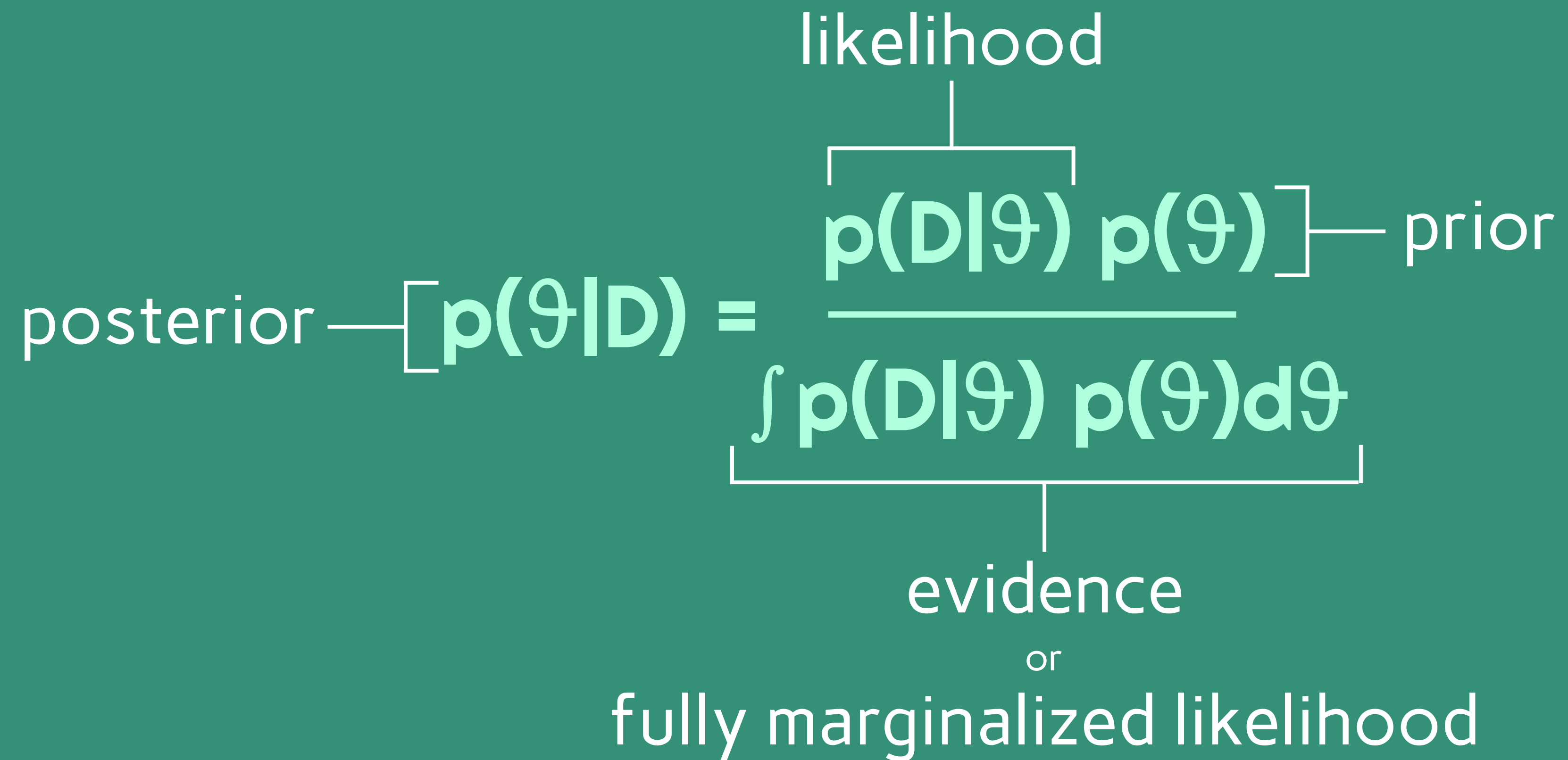
# BAYES' RULE

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

# BAYES' RULE

posterior —  $p(\theta|D) = \frac{\overbrace{p(D|\theta) p(\theta)}^{\text{likelihood}}}{\underbrace{p(D)}_{\text{evidence or fully marginalized likelihood}}} \text{ — prior}$

# BAYES' RULE



The diagram illustrates Bayes' Rule with the following components and annotations:

- posterior** — points to the term  $p(\theta|D)$  in the formula.
- likelihood** — points to the term  $p(D|\theta)$  in the numerator.
- prior** — points to the term  $p(\theta)$  in the numerator.
- evidence or fully marginalized likelihood** — points to the denominator  $\int p(D|\theta) p(\theta) d\theta$ .

$$\text{posterior} \text{ --- } [p(\theta|D) = \frac{p(D|\theta) p(\theta)}{\int p(D|\theta) p(\theta) d\theta}]$$

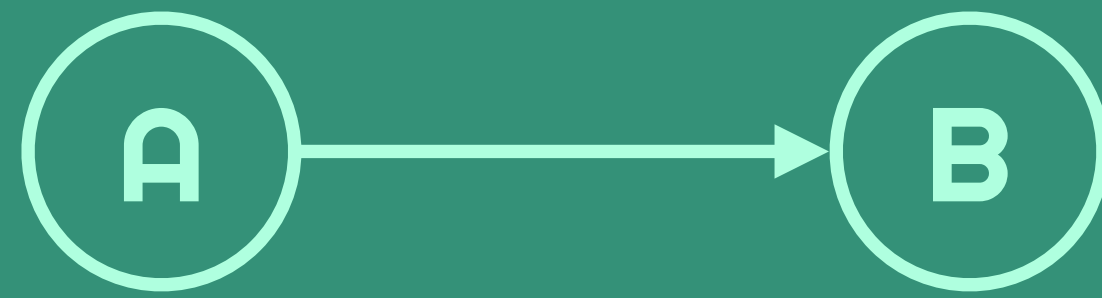
evidence  
or  
fully marginalized likelihood



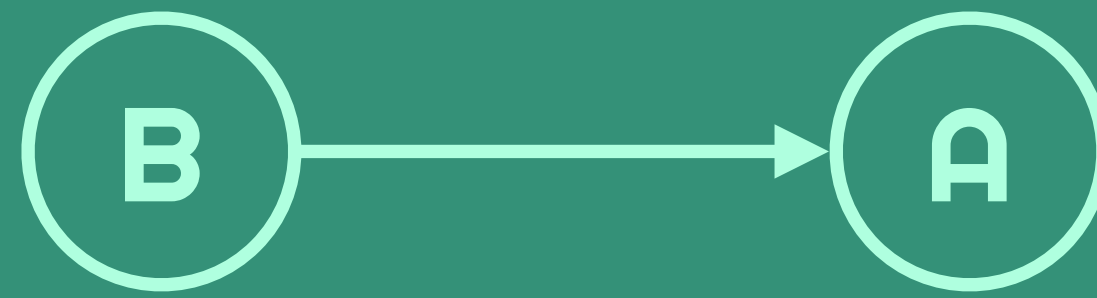
**A BRIEF ASIDE:**

**PROBABILISTIC GRAPHICAL MODELS**

[PGMs]



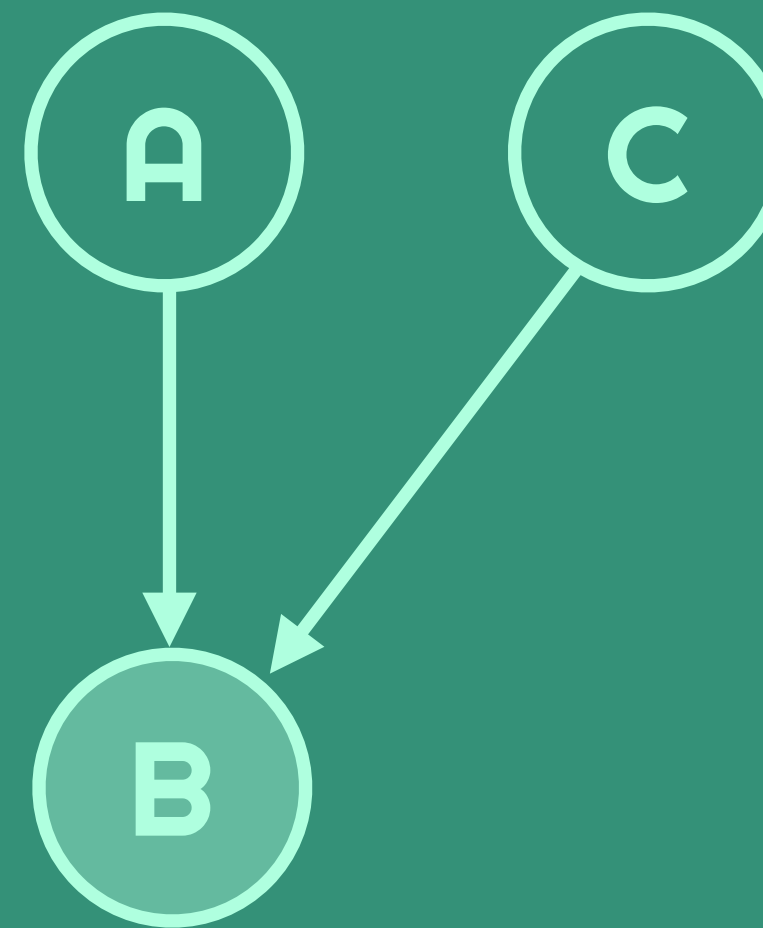
$$p(A,B) = p(A) p(B|A)$$



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



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



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

**A GOOD REFERENCE:**

**Pattern recognition and machine learning, Bishop**

**GO TO NOTEBOOK**

**for an example of Bayesian inference**

[github.com/ruthangus/Probabilistic-inference-tutorial/](https://github.com/ruthangus/Probabilistic-inference-tutorial/)