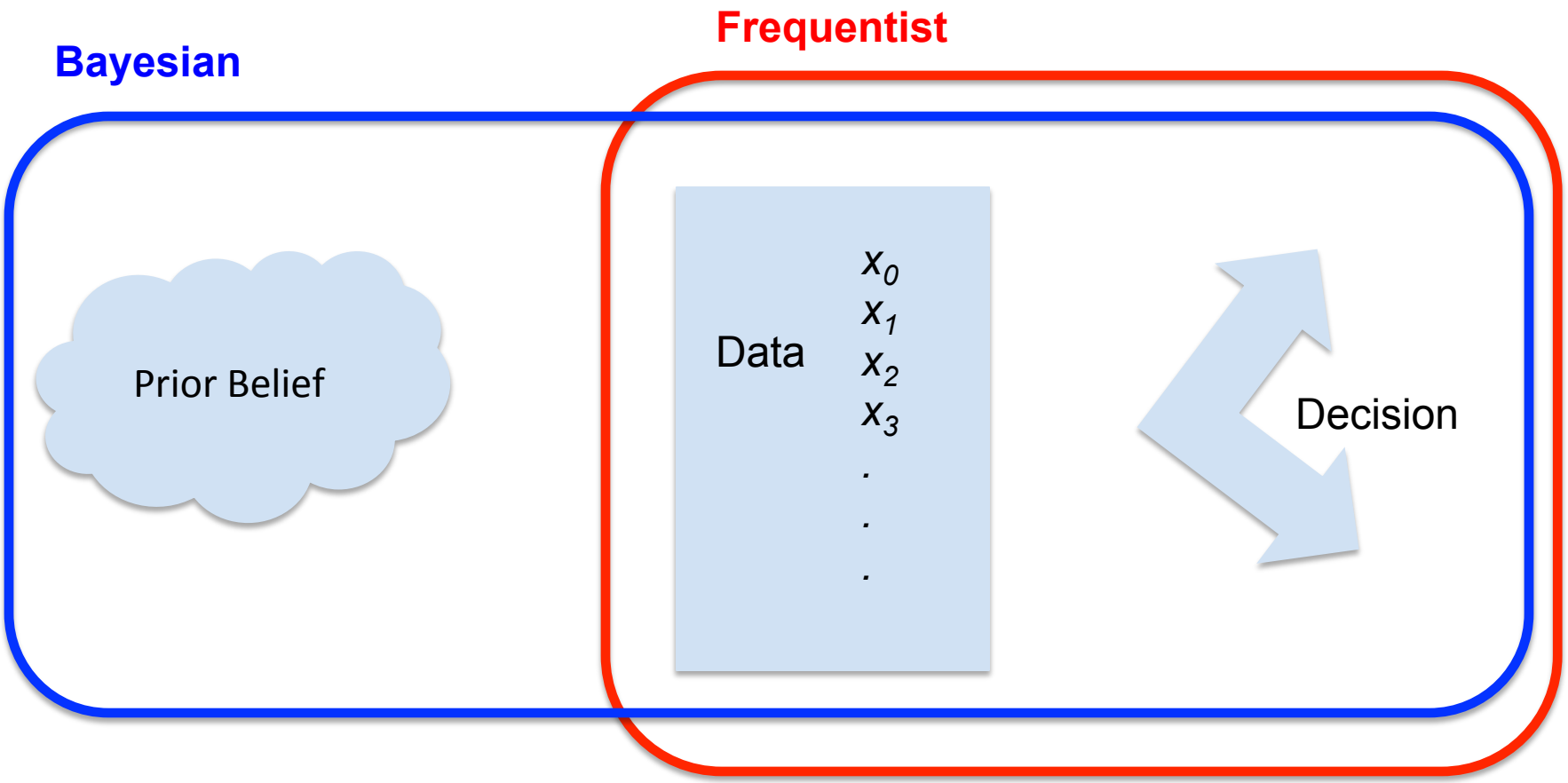
Bayesian
Frequentist
Prior Belief
Data
x_0
x_1
x_2
x_3
.
.
.
Decision

Bayes Theorem

Likelihood
Probability of collecting this data when our hypothesis is true

Prior
The probability of the hypothesis being true before collecting data

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Posterior
The probability of our hypothesis being true given the data collected

Marginal
What is the probability of collecting this data under all possible hypotheses?

	A	~A
B	P(A ∩ B)	
~B		

$$P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A)$$

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

Question

1% of women at age forty who participate in routine screening have breast cancer.

80% of women with breast cancer will get positive mammographies.

9.6% of women without breast cancer will also get positive mammographies.

A woman in this age group had a positive mammography in a routine screening.

What is the probability that she actually has breast cancer?

	Have Cancer (1%)	Do Not Have Cancer (99%)
Positive Test	True Positive: 1% * 80%	False Positive: 99% * 9.6%
Negative Test	False Negative: 1% * 20%	True Negative: 99% * 90.4%

$$P(\text{cancer} \mid \text{positive test}) = \frac{P(\text{positive test} \mid \text{cancer})\, P(\text{cancer})}{P(\text{positive test})}$$

$$P(\text{positive test}) = P(\text{positive test} \mid \text{cancer}) * P(\text{cancer}) + P(\text{positive test} \mid \text{no cancer}) * P(\text{no cancer})$$

$$P(\text{positive test}) = 0.8 * 0.01 + 0.096 * 0.99 = 0.103$$

$$P(\text{cancer}) = 0.01$$

$$P(\text{positive test} \mid \text{cancer}) = 0.8$$

$$P(\text{cancer} \mid \text{positive test}) = \frac{0.8 * 0.01}{0.103} = 0.078$$

Spam Filtering with Naïve Bayes

$$P(spam|words) = \frac{P(spam)P\,(words|spam)}{P(words)}$$

$$P(spam|viagra, rich, ..., friend) = \frac{P(spam)P(viagra, rich, ..., friend|spam)}{P(viagra, rich, ..., friend)}$$

$$P(spam)P(viagra, rich, ..., friend|spam)$$
$$\propto P(spam)P(viagra|spam)P(rich, ..., friend|spam, viagra)$$
$$\propto P(spam)P(viagra|spam)P(rich|spam, viagra)P(..., friend|spam, viagra, rich)$$

...

$$P(viagra|spam)P(rich|spam)...P(friend|spam)$$