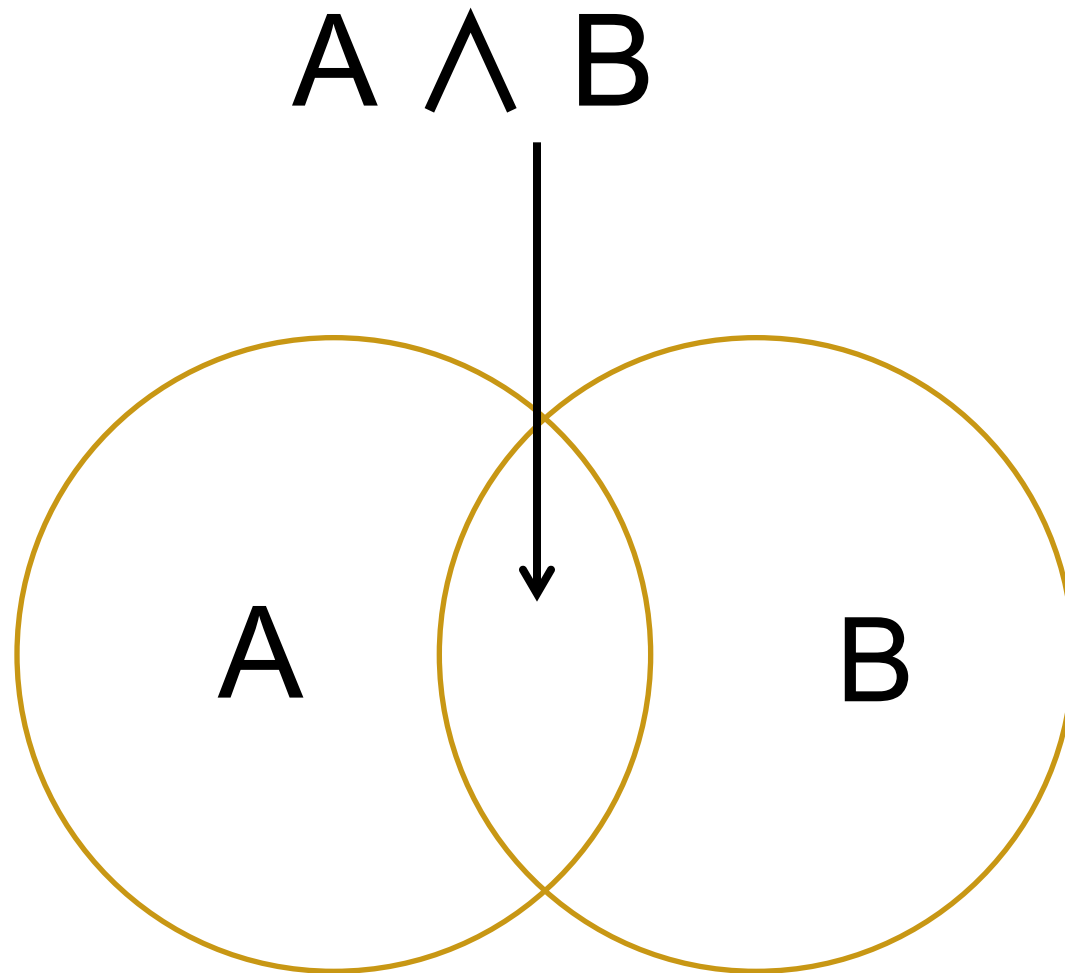# Introduction to Machine Learning

Brown University CSCI 1950-F, Spring 2011
Prof. Erik Sudderth
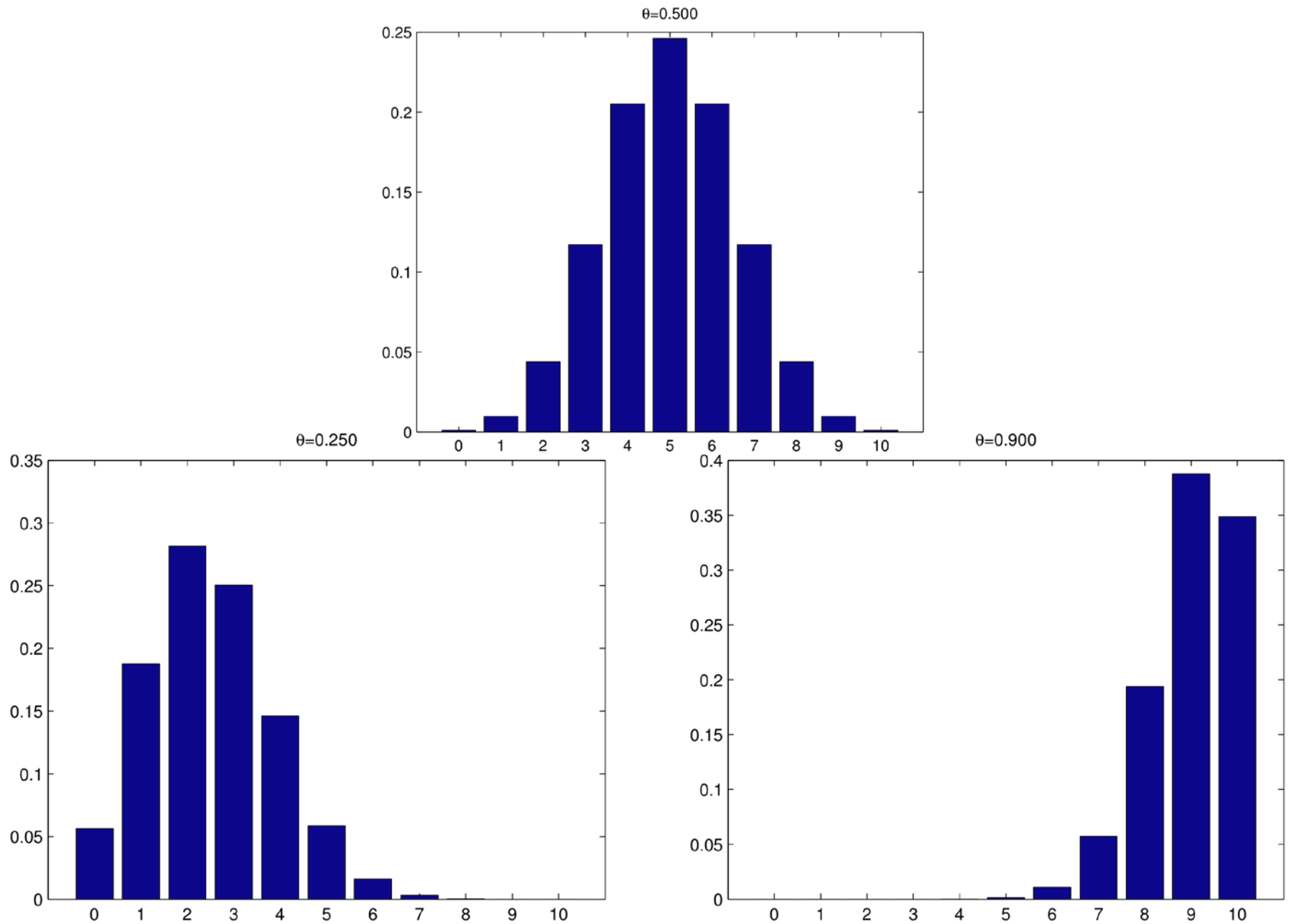
Lecture 3: Probability & Naïve Bayes

Many figures courtesy Kevin Murphy's textbook,
Machine Learning: A Probabilistic Perspective
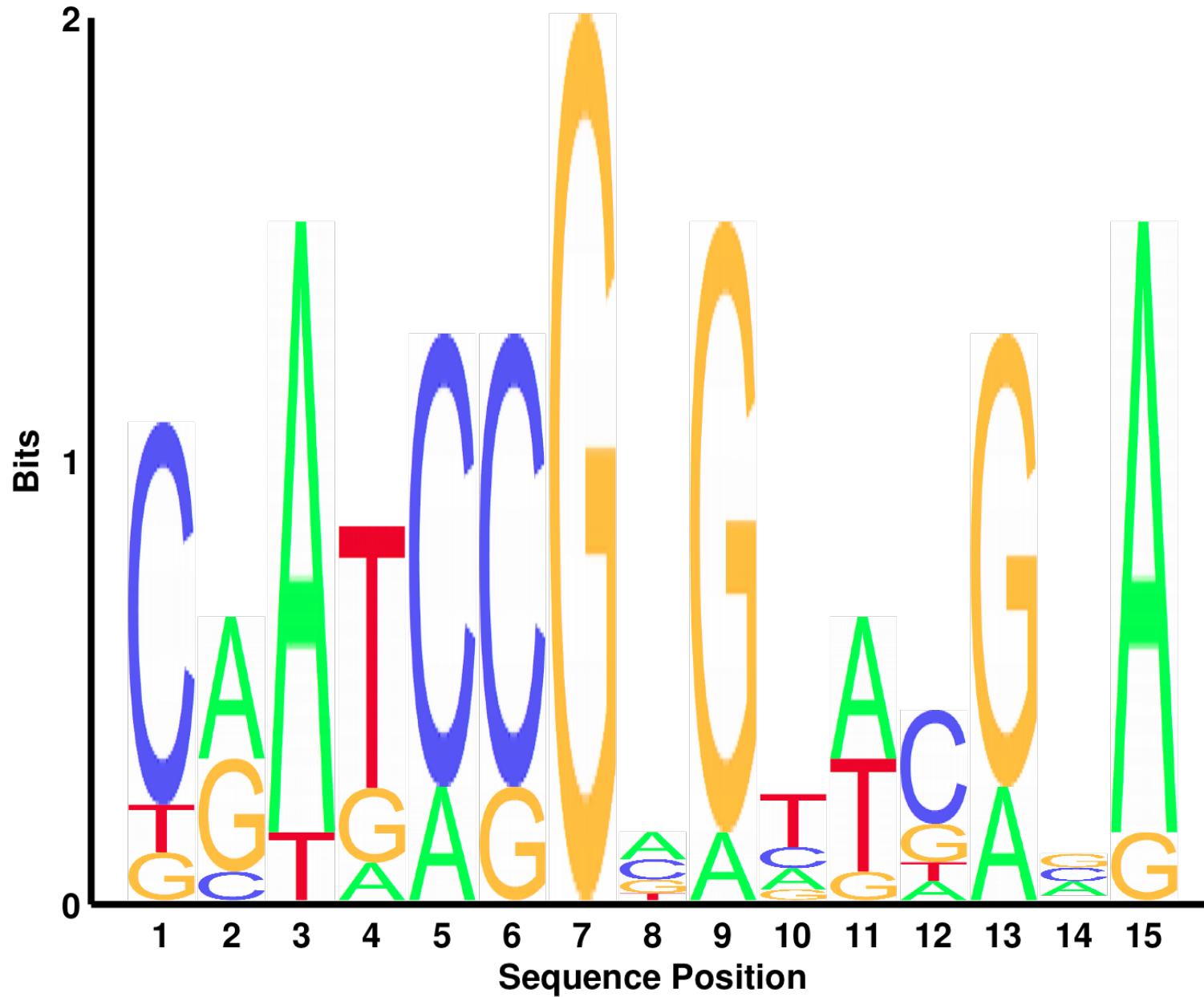
Probability of Two Events

$A \wedge B$
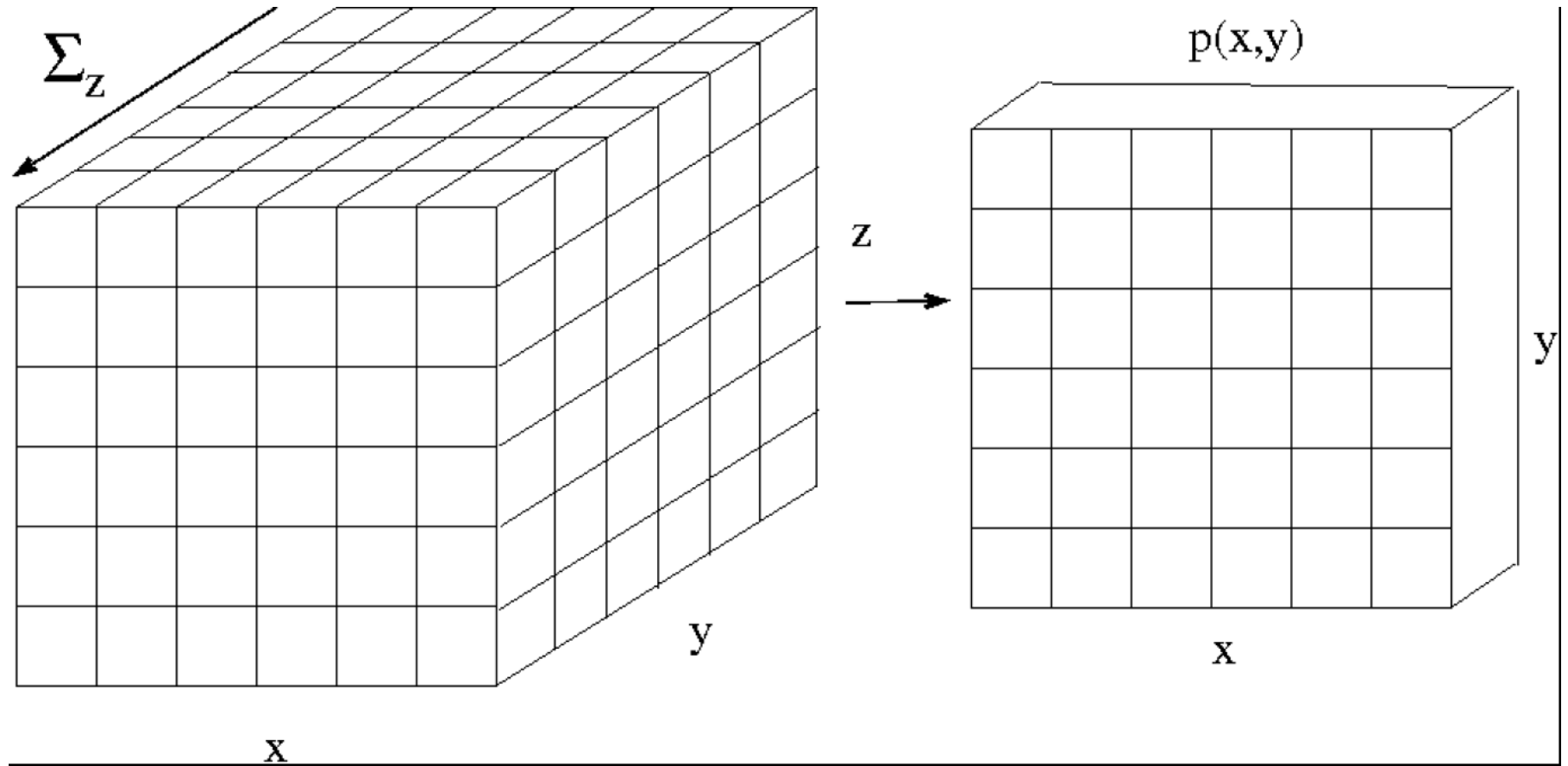
Binomial Distributions

Aligned DNA Sequences

```
c g a t a c g g g g t c g a a
c a a t c c g a g a t c g c a
c a a t c c g t g t t g g g a
c a a t c g g c a t g c g g g
c g a g c c g c g t a c g a a
c a t a c g g a g c a c g a a
t a a t c c g g g c a t g t a
c g a g c c g a g t a c a g a
c c a t c c g c g t a a g c a
g g a t a c g a g a t g a c a
```
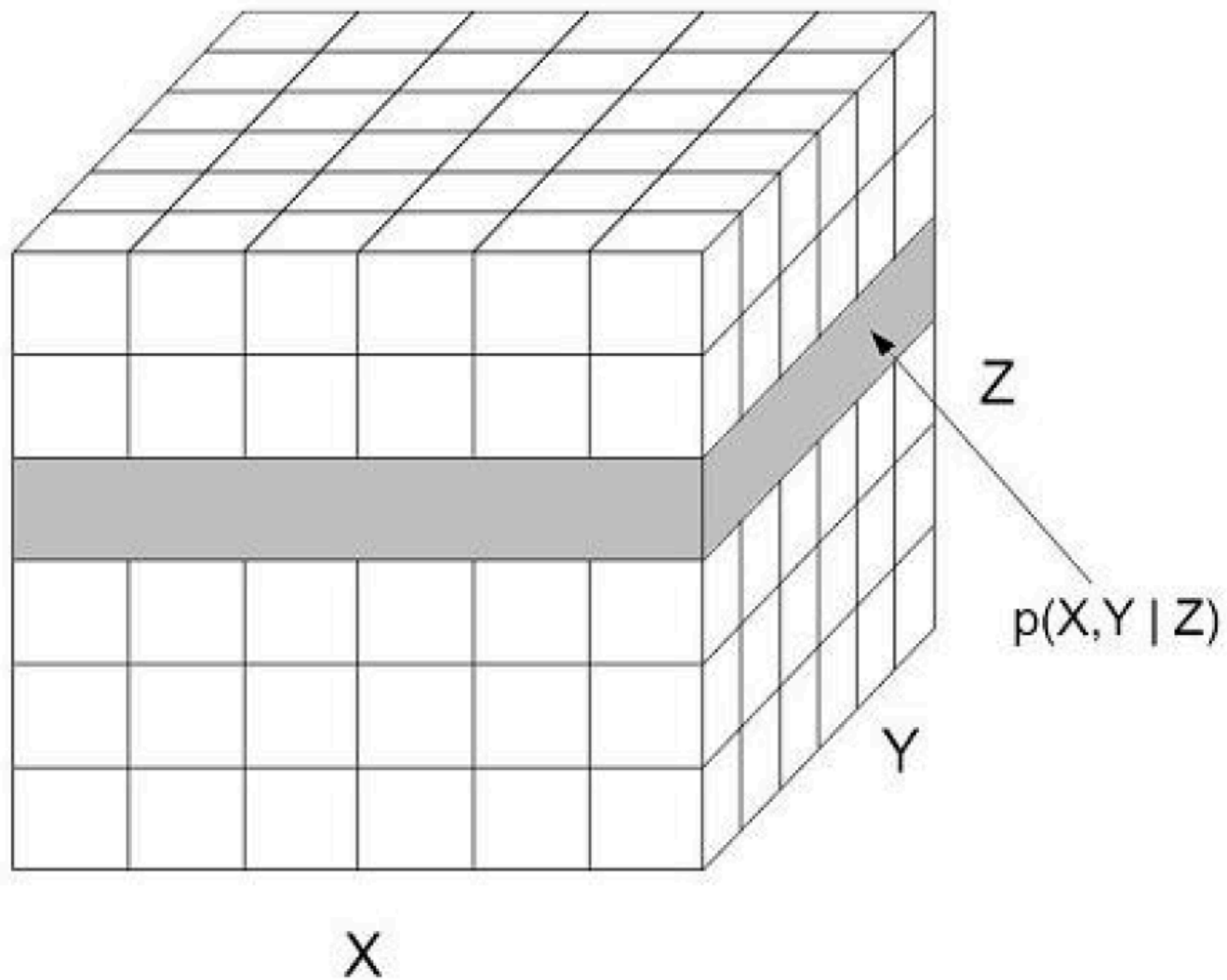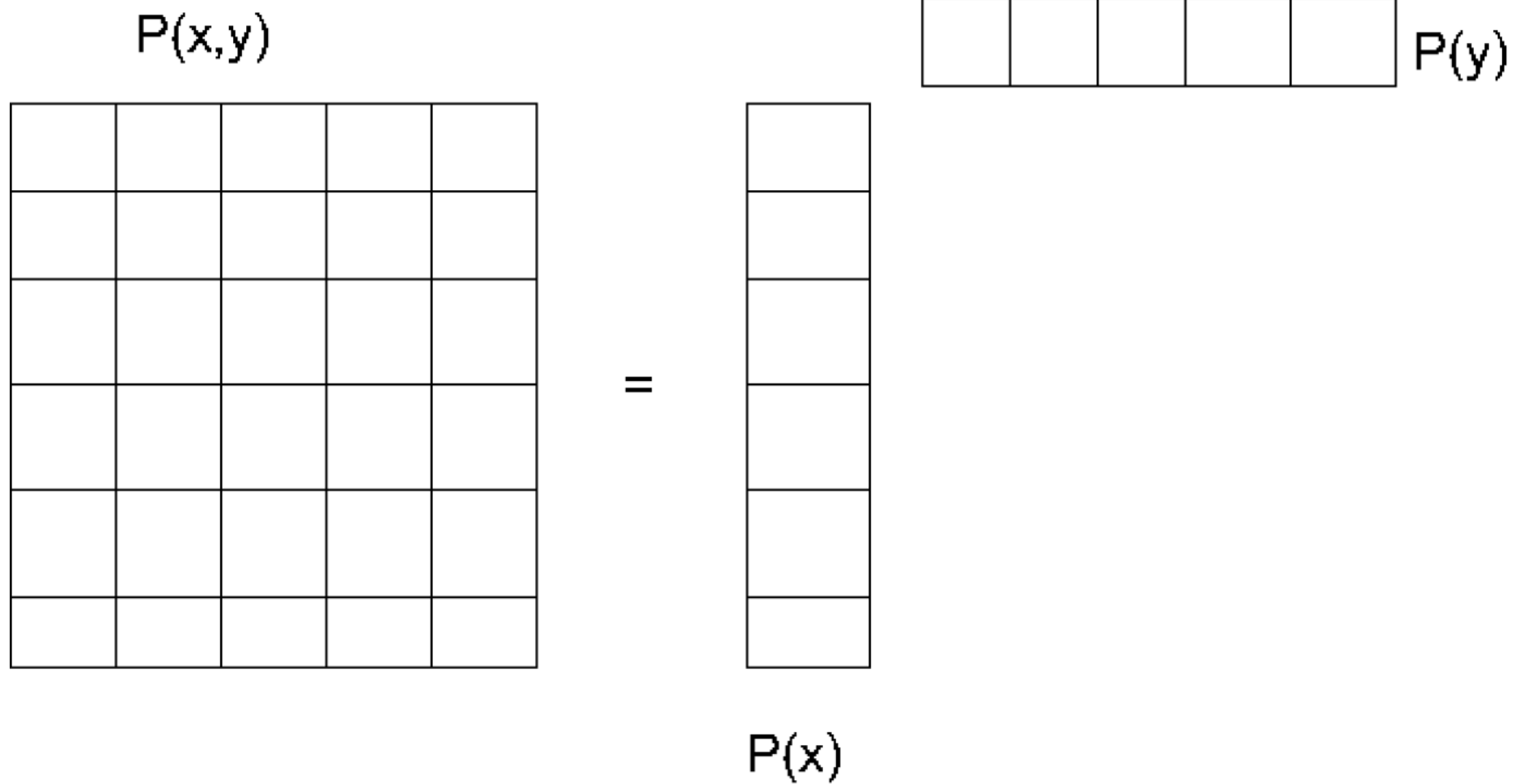
Multinomial Model of DNA

Marginalization

Conditioning

$p(X, Y \mid Z)$

Independence

P(x,y)

=

P(x)

P(y)

Naïve Bayes Spam Classification