

## 1 Introduction

Thomas Bayes: minister, mathematician. Born in London 1701, published “An Essay towards solving a Problem in the Doctrine of Chances” in 1763. It was basically a treatise on card counting.

## 2 Naive Bayes Classification

Basic Bayesian Statistics

- Conditional Probability: The probability of observing some event,  $X$ , given that another event,  $Y$  has already been observed. Denoted:  $P(X | Y)$
- Posterior probability: The probability of an event after all the predictor information has been incorporated. Posterior probabilities reflect the uncertainty of assessing an observation to particular class.

$$P(\theta | X) = \frac{P(X | \theta)P(\theta)}{p(X)}$$

- Prior probability: A probability distribution that expresses one’s belief in an outcome before any evidence is collected.

### 2.1 Bayes Theorem

$$P(\theta | X) = \frac{P(X | \theta)P(\theta)}{P(X)}$$

This is just the posterior probability, derived in Bayes’ original essay.

$P(\theta)$  is the prior distribution

The likelihood  $P(X | \theta)$  is the evidence of  $\theta$  provided by  $X$

$P(X)$  is the probability of  $X$  given all possible  $\theta$ .

## 2.2 Naive Bayes

Exact Bayesian Classification is often impractical, getting the *true* value of the likelihood may be intractable or impossible to acquire. The likelihood is defined as  $P(X_1, X_2, \dots, X_j \mid \theta)$ .

We therefore must approximate  $P(X \mid \theta)$  by calculating the product of the individual conditional probabilities:  $P(X_1 \mid \theta) * P(X_2 \mid \theta) * \dots * P(X_j \mid \theta)$

Classification then becomes assigning the label with the highest likelihood value for a given observation.