Full of subject

E. Contents

- Reminder: connection between probability and machine learning
- Definition of a probability, a random variable
- Explanation of the properties of a probability
- Concept of independence of 2 events
- Difference between marginal and joint probability
- Mean, variance and standard deviation
- Discrete probability vs continuous probability
- Density function
- Probability distributions
- Normal distribution
- Central limit theorem
- Conditional probability and Bayes theorem

Python-based Practice

- Kindly check these two notebooks
  1. probability.ipynb
  2. simulating_coin_flips.ipynb

* Should be run during the session

Home Work

Reading:

- (<u>very important</u>) Two solved examples on computing joint and marginal probabilities: (1) throwing 2 dice and (2) weather forecast in two cities

Practice:

♠ Consider the following joint distribution for the random variables T and N.

	$T=-1$	$T=0$	$T=1$	$T=2$
$N=0$	0.02	0.10	0.03	0.06
$N=1$	0.04	0.02	0.09	0.10
$N=2$	0.09	0.06	0.10	0.11
$N=3$	0.06	0.02	0.09	0.01

- EXERCISE **7.1.17**

  Suppose measurements can only distinguish two values of T, $T > 0$ and $T \leq 0$, and two values of N, $N = 0$ and $N > 0$. Find the joint distribution for these events.

- EXERCISE **7.1.18**

  Suppose measurements can only distinguish two values of T, $T > 0$ and $T \leq 0$, and two values of N, $N \leq 1$ and $N > 1$. Find the joint distribution for these events.

Bernoulli Trials

http://www.math.wichita.edu/history/topics/probability.html#bern-trials

- Boy? Girl? Heads? Tails? Win? Lose? Do any of these sound familiar? When there is the possibility of only two outcomes occuring during any single event, it is called a Bernoulli Trial. Jakob Bernoulli, a profound mathematician of the late 1600s, from a family of mathematicians, spent 20 years of his life studying probability. During this study, he arrived at an equation that calculates probability in a Bernoulli Trial. His proofs are published in his 1713 book *Ars Conjectandi* (Art of Conjecturing).

What constitutes a Bernoulli Trial?

- To be considered a **Bernoulli trial**, an experiment must meet each of three criteria:

- There must be **only 2 possible outcomes**, such as: black or red, sweet or sour. One of these outcomes is called a **success**, and the other a **failure**. Successes and Failures are denoted as S and F, though the terms given do not mean one outcome is more desirable than the other.

- Each outcome has a **fixed probability** of occurring; a success has the probability of ***p***, and a failure has the probability of ***1 - p***.
- Each experiment and result are completely **independent** of all others.

Some examples of Bernoulli Trials

http://en.wikipedia.org/wiki/Bernoulli_trial

- Flipping a coin. In this context, obverse ("heads") denotes success and reverse ("tails") denotes failure. A fair coin has the probability of success 0.5 by definition.
- Rolling a die, where for example we designate a six as "success" and everything else as a "failure".
- In conducting a political opinion poll, choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.
- Call the birth of a baby of one sex "success" and of the other sex "failure." (Take your pick.)

Binomial Distribution

- Describes processes whose trials have only two possible outcomes.

The binomial distribution gives the discrete probability distribution of obtaining exactly n successes out of N Bernoulli trials (where the result of each Bernoulli trial is true with probability p and false with probability 1-p). The binomial distribution is therefore given by

Binomial Distribution

Characteristics of the Binomial Distribution

-The binomial distribution requires that the experiment's trials be independent.

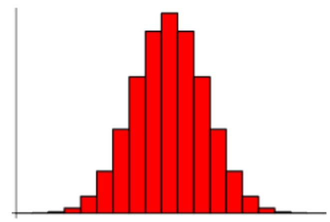
-Each trial has two outcome; a success outcome with probability P a failure out come with probability (1-p)

Binomial Formula

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad \textbf{(4)}$$

where:

n = Random sample size
x = Number of successes (when a success is defined as what we are looking for)
$n - x$ = Number of failures
p = Probability of a success
$q = 1 - p$ = Probability of a failure
$n! = n(n-1)(n-2)(n-3)\ldots(2)(1)$
$0! = 1$ by definition

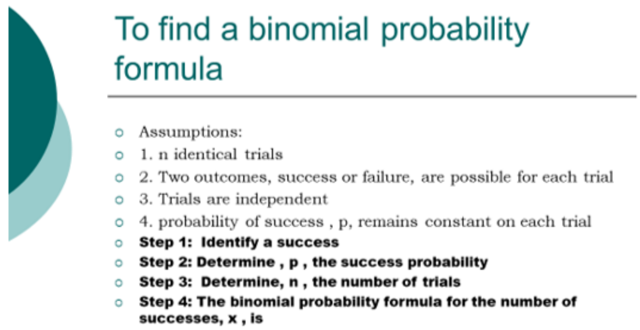


Plot of Binomial probabilities with n = 20 trials, p = 0.5

Python-based Practice

- Kindly check this notebook
  1. binomial-distributions.ipynb

Hint ➔



* Should be run during the session

Poisson Distribution

- Will Explain the same topics covered for binomial distribution
  - 1-intro,properties,function
  - 2- Example
  - 3-expected value &variance(may be with proof)
  - 4-Exercise
  - 5-python (jupyter notebook)
- To be added soon

Uniform Distribution

- Will Explain the same topics covered for binomial distribution
  - 1-intro,properties,function
  - 2- Example
  - 3-expected value &variance(may be with proof)
  - 4-Exercise
  - 5-python (jupyter notebook)
- To be added soon

Exponential Distribution

- Will Explain the same topics covered for binomial distribution
  - 1-intro,properties,function
  - 2- Example
  - 3-expected value &variance(may be with proof)
  - 4-Exercise
  - 5-python (jupyter notebook)
- To be added soon

Central limit theorem

- The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually $n > 30$).

- If the population is normal, then the theorem holds true even for samples smaller than 30. In fact, this also holds true even if the population is binomial, provided that $\min(np, n(1-p)) > 5$, where n is the sample size and p is the probability of success in the population. This means that we can use the normal probability model to quantify uncertainty when making inferences about a population mean based on the sample mean.

Binary Classifier

- The problem of classification predictive modeling can be framed as calculating the conditional probability of a class label given a data sample, for example:

- P(class|data) = (P(data|class) * P(class)) / P(data)
  Where P(class|data) is the probability of class given the provided data.

- Binary classification problem

- Evaluating Binary classifiers
  - Confusion Matrix
  - Precision, Recall, F1 Score
  - PR curve
  - ROC curve and AUC

Maximum Likelihood Estimation

- Explain the importance of maximum likelihood estimation in machine learning through the intuition: regressions don't always work (under/overfitting) so what is the best (more likely) value for a parameter in a given probability distribution?. Define and give example: trying to find the best value of p to describe a sequence HHHHTTHHTH for several coin flips

- To be added soon

Home Work

Reading:

- To be added later

Practice:

- Using Python, please build a naïve Bayes classifier class from scratch with fit and predict functionalities and perform suitable metric for evaluating the results.

- Note: Scikit-learn is not allowed

- *Dataset and problem to be provided*

they contribute to the efficiency of most machine learning algorithms.

C. Student Skills Acquired

- Build, operate and train neural networks through notebook execution.
- Implement Machine Learning algorithms.
- Use the concepts of Machine Learning and Deep Learning in the different frameworks on the market.
- Understand Supervised learning (generative and discriminative methods).
- Understand Unsupervised Learning.
- Understand Reinforcement learning.