

# Zhouteng Ye's learning progress on probability and statistics

yeldon

*[2018-05-29 Tue 15:09]*

## Contents

<b>1</b>	<b>History of probability and statistics</b>	<b>1</b>
<b>2</b>	<b>Study of evidence theory</b>	<b>3</b>
2.1	Definition of evidence theory . . . . .	3
2.2	Belief Function . . . . .	3
2.3	The idea of chance . . . . .	4
2.4	Bayesian theory of partial belief . . . . .	4
2.5	Evidence theory and Bayesian theory . . . . .	5
2.6	Support function and weight of evidence . . . . .	5
2.7	Steps of evidence theory approach . . . . .	5
2.8	Recent development of evidence theory . . . . .	6
2.9	Evidence theory in machine learning . . . . .	6
<b>3</b>	<b>Reading lists</b>	<b>6</b>

## 1 History of probability and statistics

**Probability** has been defined as the study of the frequency of appearance of a phenomenon in relation to all possible alternatives.

**Statistics** is the science and art of gathering, analyzing, and making inferences from data.

The history of probability and statistics:

- Randomness was not associated with gaming in early time. Starting from Renaissance, randomness started being considered.

- "Problem of points" gives rise to the science of mathematical probability. Discussion among Paciolo, Pascal, Fermat and Mere not only came up with a convincing, self-consistent solution to the division of the stakes, but also developed concepts that continue to be fundamental in probability to this day.
- In 1657, Huygens wrote the first formal treatise on probability *On reasoning in games of chance* based on the Fermat-Pascal correspondence. In 1713 Jakob Bernoulli pushed the book *The art of conjecturing*. Bernoulli also proved a version of the fundamental law of large numbers.
- John Graunt, "Father of Statistics", was the first person drew statistical inferences from analyses of mass data in 17th century.
- Halley, Newton, Demoivre made pioneering efforts on actuarial mathematics and its relation to life insurance. During which the mathematics of permutation, combination and normal probability curve has been developed.
- After that, probability and statistics entered a new transitional period. mathematicians realized that probability could not be separated from statistics.
- Daniel Bernoulli proposed Petersburg paradox (However, the name took from the resolution of Daniel Bernoulli), calculus was applied to the theory of probability. Euler, Lagrange also advanced the theory of probability by applying differential calculus to it.
- Laplace, "Father of Modern Probability Theory" used the theory of probability to obtain a statistical measure of reliability of numerical results. The results relies on definite causes rather than pure chance.
- Using the probability theory that various types of mathematical distributions has been proposed (Bernoulli, Poisson and normal distribution).
- The beginning of statistical analysis of census data was accomplished in 1829 by Lambert Adolphe Jacques Quetelet.
- Father Gregor Mendel related probability to genetics and hybridization in 1865.

- Galton discovered the law of regression and the correlation coefficient in 1877.
- Beginning in 1894, Karl Pearson applied probability to biology and created the area of study we now call biometric.
- The Russian Andrey Markov developed his chain theory of probabilities
- Norbert Wiener expressing his belief that probability is the link between physics and mathematics, created cybernetics.

## 2 Study of evidence theory

### 2.1 Definition of evidence theory

Evidence theory (Dempster-Shafer theory) is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility and imprecise probability theories.

- Heart of the evidence theory: **Dempster's rule**.
- The theory deals with weights of evidence and with numerical degree of support based on evidence.
- Does not focus on the act of judgment by which such a number is determined.
- Focus on something more amenable to mathematical analysis: the combination on something degrees of belief or support based on on body evidence with those based on an entirely distinct body of evidence.

A **degree of belief** is represented as a **belief function**. Probability values are assigned to sets of possibilities rather than single events.

### 2.2 Belief Function

Suppose  $\Theta$  is a finite set, and let  $2^\Theta$  denote the set of all subsets of  $\Theta$ , Suppose the function  $Bel: 2^\Theta \rightarrow [0, 1]$  satisfies the following conditions:

- (1)  $Bel(\Phi) = 0$
- (2)  $Bel(\Theta) = 1$

- (3) For every positive integer  $n$  and every collection  $A_1, \dots, A_n$  of subsets of  $\Theta$ ,

$$Bel(A_1 \cup \dots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i < j} Bel(A_i \cap A_j) + \dots + (-1)^{n+1} Bel(A_1 \cap \dots \cap A_n)$$

Then  $Bel$  is called a **belief function** over  $\Theta$ .

Only the set functions obey the rules (1) to (3) can be combined by **Dempster's rule of combination**.

### 2.3 The idea of chance

The ideas of **Chance** and **belief** are united under the name **probability**. But they actually have different roles to play. **Chance** arises only when one describes an aleatory (or random) experiment. The outcome varies randomly from one physically independent trial to another.

**Basic rules for chances:** A function  $Ch: 2^X \rightarrow [0, 1]$  is a chance function if and only if it obeys the following rules:

- (1)  $Ch(\Phi) = 0$ .
- (2)  $Ch(X) = 1$ .
- (3) If  $U, V \subset X$  and  $U \cap V = \Phi$ , then  $Ch(U \cup V) = Ch(U) + Ch(V)$ .

The Third rule is \*Rule of additivity for chances.

*Chance can be used to estimate the degrees of belief. But it is merely used as the chance density governing an experiment is usually unknown.*

### 2.4 Bayesian theory of partial belief

the function  $Bel: 2^\Theta \rightarrow [0, 1]$  in Bayesian theory must satisfies the following conditions:

- (1)  $Bel(\Phi) = 0$
- (2)  $Bel(\Theta) = 1$
- (3) If  $A \cap B = \Phi$ , then  $Bel(A \cup B) = Bel(A) + Bel(B)$
- (4) If  $Bel(A) > 0$ , then  $Bel(B|A) = \frac{Bel(B \cap A)}{Bel(A)}$ .

Rule (3) is called **Bayes' rule of additivity**, rule (4) is called **Bayes' rule of conditioning**.

## 2.5 Evidence theory and Bayesian theory

The first two rules from Bayesian theory are identical to the first two rules for belief function. The third rule is different between two theories. All function satisfying Bayesian belief function satisfies the third rule in evidence theory; but not all belief function satisfies Bayes' rule of additivity. *In other words, Bayesian theory is a restrictive spacial case in the theory of evidence; Theory of evidence is a generalization of Bayesian theory.*

## 2.6 Support function and weight of evidence

A belief function  $Bel : 2^\Theta \rightarrow [0, 1]$  is called **Simple support function** if there exists a non-empty subset  $A$  of  $\Theta$  and a number  $s$ ,  $0 \leq s \leq 1$ , such that

- $Bel(B) = 0$ , if  $B$  does not contain  $A$
- $Bel(B) = s$ , if  $B$  contains  $A$ , but  $B \neq \Theta$
- $Bel(B) = 1$ , if  $B = \Theta$

Combination of simple support functions supporting same subset  $A$  leads to a larger class of belief function, called **Separable support function**.

The **Support function** is the function that includes all belief functions that can be obtained by beginning with a separate support function on a certain set of possibilities and then "coarsen" the set of possibility by neglecting to distinguish between certain of its elements. Bayes' belief function should be qualified as **Quasi support function**, which is obtained by the limit of a sequence of support function.

The idea of a **Weight of evidence**  $w$  is closely related to support function. When  $s$  is the degree of support, the relationship between  $s$  can be described as

$$s = 1 - e^{-w}.$$

## 2.7 Steps of evidence theory approach

The ET approach for quantifying the uncertainty in the performance of a system and assessing the safety of the system consists of three steps:

- (a) Model uncertainty. First, models considering each variable separately are constructed. Then a model that considers all variables together is derived.

- Propagate uncertainty through the system. This step results in a model of the uncertainty in the performance of the system.
- Assess the system safety.

## 2.8 Recent development of evidence theory

The classical form of Dempster combination rule in Dempster-Shafer theory is incapable dealing with high conflicting evidence. The modified theory of evidence pursues to handle conflicting evidence. Recent works, for example, Deng (2015) proposed generalized evidence theory (GET) theory with the consideration of two main causes of evidence conflict:

- Sensor reliability caused by disturbances or the condition of equipment.
- Our knowledge is not complete.

Jiang and Zhan (2017) found that generalized combination rules (GCR) proposed by Deng (2015) could also produce unreasonable results when combining generalized basic probability assignment. They modified the GCR is proposed to provide an aspect from geometry to explain the combination rule.

## 2.9 Evidence theory in machine learning

I have not got enough time to read about it, but I do see some paper on this topic, mostly using evidence theory for classifiers.

## 3 Reading lists

- Lightner, J.E., 1991. A brief look at the history of probability and statistics. *The Mathematics Teacher*, 84(8), pp.623-630.
- Ross, S.M., 2014. *Introduction to probability models*. Academic press.
- Shafer, G., 1976. *A mathematical theory of evidence* (Vol. 42). Princeton university press.
- Oberkamp, W.L., 2005, August. Uncertainty quantification using evidence theory. In *Proceedings from the Advanced Simulation & Computing Workshop*.

- Soundappan, P., Nikolaidis, E., Haftka, R.T., Grandhi, R. and Canfield, R., 2004. Comparison of evidence theory and Bayesian theory for uncertain modeling. Reliability engineering & System safety, 85(1-3), pp.295-311.
- Deng, Y., 2015. Generalized evidence theory. Applied Intelligence, 43(3), pp.530-543.
- Jiang, W. and Zhan, J., 2017. A modified combination rule in generalized evidence theory. Applied Intelligence, 46(3), pp.630-640.
- Problem of points on wikipedia
- History of probability on wikipedia
- History of statistics on wikipedia
- Dempster-Shafer theory on wikipedia
- Statistics for Applications on MIT OpenCourWare