Probability Theory and Statistics

Shuai

Contents

Chapter 1

Introduction

1.1 A note on mathematical abstraction

Maybe the most intimidating things about mathematics is the horrible symbols and terminology used everywhere. I have such experience, so I know it.

However, it is nothing different with learning a new language – you learn it by being used to it, by using it, by becoming natural resident of it. Give yourself some time, you will find it out.

1.2 Clarification on Concept of Probability and Statistics

It takes me some time to understand the difference between probability and statistics. At the first time I learnt them, I learnt through a Chinese book called *Probability Theory and Mathemtical Statistics*. I thought they are kind of the same subjects. However, as time passed and I learnt more math, I found they are not.

Probability Theory deals with the abstraction of the natural sense of probability – I guess it will rain tomorrow probably since the cloud looks like this shape. It is a scientific theory to accommdate another major problem of human being – how to quantify uncertainty.

Statistics deals with data, is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments.[**?**] It focuses on extract useful imformation from data – mean, variance, coorelation all tells something important.

Remark The word statistics, when referring to the scientific discipline, is singular, as in ”Statistics is an art.” This should not be confused with the word statistic, referring to a quantity (such as mean or median) calculated from a set of data, whose plural is statistics (”this statistic seems wrong” or ”these statistics are misleading”).[**?**]

Chapter 2

Probability

2.1 Probability Theory, from intuition to abstraction

This is the text at the very beginning of the book *Probability Theory* written by Loeve. Mathematics could be pure, but every part of science comes from reality. Do not be confused by its abstractness and generality.

Probability theory is concerned with the mathematical analysis of the intuitive notion of "chance" or "randomness", which, like all notions, is born of experience. The quantitative idea of randomness first took form at the gaming tables, and probability theory began, wich Pascal and Fermat(1654), as a theory of games of chance. Since then, the notion of chance has found its way into almost all branches of knowledge. In particular, the discovery that physical "observables", even those which describe the behavior of elementary particles, were to be considered as subject to laws of change made an inverstigation of the notion of chance basic to the whole problem of rational interpretation of nature.[**?**]

A theory becomes mathematical when it sets up a mathematical model of the phenomena with which it is concerned, that is, when, to describe the phenomena, it uses a collection of well-defined symbols and operations on the symbols. As the number of phenomena, together with their known properties, increases, the mathematical model evolves from early crude notions upon which our intuition was built in the direction of higher generality and abstractness.[**?**]

In this manner, the inner consistency of the model of random phenomena became doubtful, and this forced a rebuilding of the whole structure in the second quarter of this century, starting with a formulation in terms of axioms and definitions. Thus there appeared a branch of pure mathematics – probability theory – concerned with the construction and inverstigation per se of the mathematical model of randomness.[**?**]

2.2 Interpretations of Probability

Intuitively, the probability $P(\alpha)$ of an event α quantifies the degree of confidence that α will occur. If $P(\alpha) = 1$, we are certain that one of the outcomes in α occurs. However, this description does not provide an answer to waht the numbers mean. There are two common interpretations for probabilities.[?]

The frequentist interpretation views probabilities as frequencies of events. More precisely, the probability of an event is the fraction of times the event occurs if we repeat the experiment indefinitely. This interpretation gives probabilities a tangible semantics. When we discuss concrete physical systems(for example, dice, coin, flips, and card games) we can envision how these frequencies are defined.[?]

The frequentist interpretation fails, however, when we consider events such as "It will rain tomorrow afternoon." Although the time span of "Tomorrow afternoon" is somewhat ill defined, we expect it to occur exactly once. It is not clear how we define the frequencies of such events. Several attempts have been made to define the probability for such an event by finding a *reference class* of similar for which frequencies are well defined, but they are not satisficatory.[?]

An alternative interpretation views probabilities as subjective degrees of belief. Under this interpretation, the statement $P(\alpha) = 0.3$ represents a subjective statement about one's own degree of belief that the event α will come about.[?] What does it mean by saying subjective degrees of belief?

2.3 Subjective Probability

I still do not understand this term completely. I will write down what I learnt for the time being.

The main idea of subjective probability is when we are dealing with something we are not sure, we should deal with it rationally. if a decision makers subjective probabilities do not cohere he/she may incur sure loss; a competitor can set us a Dutch book to drain up his/her account

The following text is referenced from here[?].

Suppose a person has assigned $P(C) = \frac{2}{5}$ to some event C. Then the *odds* against C would be

$$O(C) = \frac{1 - P(C)}{P(C)} = \frac{1 - \frac{2}{5}}{\frac{2}{5}} = \frac{3}{2}$$

Moreover, if that person is willing to bet, he or she is willing to accept either side of the bet: 1. win 3 units if C occurs and lose 2 if does not occur or 2. win 2 units if C does not occurs and lose 3 if it does. if that is not the case, then that person should review his or her subjective probability of event C.

This is really much like two children dividing a candy bar as equal as possible; One divides it and the other gets to choose which of the two parts seems most desirable; that is the larger. Accordingly, the child dividing the candy bar tries extremely hard to cut it as equal as possible. Clearly, this is exactly what the person selecting the subjective probability does as he or she must be willing to take either side of the bet with the odds established.

Let us now say the reader is willing to accept that the subjective probability $P(C)$ as the fair price for event C, given that you will win one unit in case C occurs and, of course, lose $P(C)$ if it does not occur. Then it turns out, all rules(definitions and theorems) on probability follow for subjective probabilities.

We only prove one theorem.

Theorem 2.3.1 *If C_1 and C_2 are mutually exclusive, then*

$$P(C_1 \cup C_2) = P(C_1) + P(C_2)$$

Proof Suppose a person thinks a fair price for C_1 is $p_1 = P(C_1)$ and that for C_2 is $p_2 = P(C_2)$. However, that person believes the fair price for $C_1 \cup C_2$ is p_3 which differs from $p_1 + p_2$. Say, $p_3 < p_1 + p_2$ and let the difference be $d = (p_1 + p_2) - p_3$. A gambler offers this person the price $p_3 + \frac{d}{4}$ for $C_1 \cup C_2$. That peron takes the offer because it is better than p_3. The gambler sells C_1 at a discount price $p_1 - \frac{d}{4}$ and sells C_2 at a discount price of $p_2 - \frac{d}{4}$ to that person. Being a rational person with those given prices of $p_1, p_2 and p_3$, all three of these deals seem very satisficatory. However, that person received $p_3 + \frac{d}{4}$ and paid $p_1 + p_2 - \frac{d}{2}$. Thus before any bets are paid off, that person has

$$p_3 + \frac{d}{4} - (p_1 + p_2 - \frac{d}{2}) = p_3 - p_1 - p_2 + \frac{3d}{4} = -\frac{d}{4}$$

That is, the person is down $\frac{d}{4}$ before any bets are settled.

- Suppose C_1 happens: the gambler has $C_1 \cup C_2$ and the person has C_1; so they exchange units and the person is still down $\frac{d}{4}$. The same thing occurs if C_2 happens.
- Suppose neither C_1 or C_2 happens, then the gambler and that person receive zero, and the person is still down $\frac{d}{4}$.
- Of course, C_1 and C_2 can not occur together since they are mutually exlusive.

Thus we see that it is bad for that person to assign

$$p_3 = P(C_1 \cup C_2) < p_1 + P_2 = P(C_1) + P(C_2)$$

because the gambler can put that person in a position to lose $(p_1 + p_2 - p_3)/4$ on matter what happens. This is sometimes referred to as a **Dutch book**.

I also found a book called *Subjective Probability: The Real Thing.* Do not read it yet. Seems good.

2.4 Intuitive Background

2.4.1 Events

The primary notion in the understanding of nature is that of *event* – the occurrence or nonocurrence of a phenomenon. The abstract concept of event pertains only to its occurrence or nonocurrence and not to its nature. This is the concept we intend to analyze.[**?**]

Remark About notation used. $A_1 \cap A_2$ is same with $A_1 A_2$. But $A_1 \cup A_2$ can be replaced by $A_1 + A_2$ when A_1 and A_2 are disjoint.

In science, or, more precisely, in the investigation of "laws of nature," events are classified into conditions and outcomes of an experiment. *Conditions* of an experiment are events which are known or are made to occur. *Outcomes* of an experiment are events which *may* occur when the experiment is performed, that is, when its conditions occur. All(finite) combinations of outcomes by means of "not", "and", "or", are outcomes; in the terminology of sets, the outcomes of an experiment form a *field*(or an "algebra" of sets). The condition of an experiment together with its field of outcomes, constitute a *trial*. Any (finite) number of trials can be combined by "conditioning", as following:

The collective outcomes are combinations by means of "not", "and", "or", of the outcomes of the constituent trials. The conditions are conditions of the first constituent trials together with conditions of the second to which are added the observed outcomes of the first, and so on. Thus, given the observed outcomes the preceding trials, every constituent trial is performed under supplementary conditions: it is conditioned by the observed outcomes. When, for every constituent trial, any outcome occurs if, and only if, it occurs without such conditioning, we say that the trials are *completely independent*. If, moreover, the trials are identical, that is, have the same conditions and the same field of outcomes, we speak of *repeated trials* or equivalently,, *identical and completely independent trials*. The possibility of repeated trials is a basic assumption in science, and in games of chance: *evry trial can be performed again and again, the knowledge of past and present outcomes having no influence upon future ones.*[**?**]

2.4.2 Random Events and Trial

Science is essentially concerned with permanencies' in repeated trials. For a long time *deterministic trials* only, where conditions(causes) determine completely the outcomes(effects). Although another type of permanency has been oberseved in games of chance, it is only recently that *Home sapiens* was led to think of a rational interpretation of nature in terms of these permanencies: nature plays the greatest of all games of chance with oberserver.[**?**]

And the investigation in the games of chance leads to the concept of random event: Let the frequency of an outcome A in n repeated trials be the ratio $\frac{n_A}{n}$ of the number n_A of occurrences of A to the total number n of trials. If , in repeating a trial a large number of times, the observed frequencies of any one of its outcomes A cluster about some number, the trial is then said to be random. The outcomes of a random trial are called *random(chance) events.*[**?**]

2.4.3 Random Variables

For a physicist, the outcomes are, in general, values of an obervable. The concept of random variabe is more general than that of random event. In fact, we can assign to every random event A a random variable. Then the observed value tells us whether or not A occurd, and conversely. Furthermore, we can do calculus on them, such as computing its expectation.[**?**]

What's more, a physical obervable may have an infinite number of possible values, and then the foregoing simple definitions do not appy. The evolution of probability theory is due precisely to the consideration of more and more complicated observables.[**?**]

2.5 Axiomization of Intuition

This is the axioms of the finite case.

Let Ω or the *sure event* be a space of points ω; the empty set (set containing no points ω) or the *impossible event* will be denoted by $\emptyset$. Let α be a noempty class of sets in Ω, to be called *random events* or, simply, events, since no other type of events will be considered. Events will be denoted by $A, B, \cdots$ with or without affixes. Let P or *probability* be a numerical function defined on α; the value of P for a event A will be called the *probability* of A and will be denoted by PA. The pair (α, P) is called a *probability field* and the triplet (Ω, α, P) is called a *probability space*.[?]

Then the following two axiom abstracts the intuitive nature of probability:[?]

Axiom I: α is a field: complements A^c, finite intersections $\bigcap\limits_{k=1}^{n} A_k$, and finite unions $\bigcup\limits_{k=1}^{n} A_k$ of events are events.

Axiom II: P on α is normed, nonnegative, and finitely additive:

$$P\Omega = 1, PA \geq 0, P\sum_{k=1}^{n} A_k = \sum_{k=1}^{n} PA_k$$

2.6 Sample Random Variable

Let the probability field (α, P) be fixed. In order to introduce the concept of random variables, it will be cconvenient to begin with very special ones, which permit operations on events to be transformed into ordinary algebraic operations.[?]

Note, this is an example to show concretely what is random variable.

To every event A we assign a function I_A on Ω with values $I_A(\omega)$, such that $I_A(\omega) = 1$ *or* 0 according as ω belongs or does not belong to A; I_A will be called the *indicator* of A(in terms of occurrences, $I_A = 1$ *or* 0, according as A occurs or does not occur). Thus, $I_A^2 = I_A$ and the boundary cases are those of $I_\emptyset = 0$ and $I_\Omega = 1$.[?]

The following properties are immediate[?]:

- if $A \subset B$, then $I_A \leq I_B$, and conversely;
- if $A = B$, then $I_A = I_B$, and conversely;
- $I_{A^c} = 1 - I_A$, $I_{AB} = I_A I_B, I_{A+B} = I_A + I_B$
- $I_{A\cup B} = I_{A+A^cB} = I_A + I_B - I_{AB}$

Linear combinations $X = \sum\limits_{j=1}^{m} x_j I_{A_j}$ of indicators of events A_j of a finite partition of Ω, where the x_j are (finite) numbers, are called *simple random variables*, to be denoted by captials $X, Y, \cdots$, with or without affixes. The set of values PA_j which correspond to the values x_j of X, assumed all distinct, is called the *probability distribution* and the A_j form the *partion* of X.[?]

2.7 Bayesian Probability & Statistics

This part is just nonsense for the time being.

I found the following materials when I try to understand bayesian probability and bayesian statistics.

Broadly speaking, there are two views on Bayesian probability that interpret the 'probability' concept in different ways. For objectivists, probability objectively measures the plausibility of propositions, i.e. the probability of a proposition corresponds to a reasonable belief everyone (even a "robot") sharing the same knowledge should share in accordance with the rules of Bayesian statistics, which can be justified by requirements of rationality and consistency. For subjectivists, probability corresponds to a 'personal belief'. For subjectivists, rationality and coherence constrain the probabilities a subject may have, but allow for substantial variation within those constraints. The objective and subjective variants of Bayesian probability differ mainly in their interpretation and construction of the prior probability.[**?**]

Subjective probability is the foundation of Bayesian methods. [**?**]

Chapter 3

Statistics

3.1 Intuition of Correlation Coefficient

This section is only made for the clarification of the concept of Correlation Coefficient.

The concept I learnt from textbook is like this:

Definition If X and Y are jointly distributed random variables with expectations μ_X and μ_Y, respectively, the Correlation Coefficient of X and Y is:

$$\rho_{X,Y} = coor(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

The intuition behind it is linear coorelation: If the random variables are positively associated—that is, when X is larger than its mean, Y tends to be larger than its mean as well—the covariance will be positive. If the association is negative—that is, when X is larger than its mean, Y tends to be smaller than its mean – the covariance is negative.[**?**]

Actually, this is called *Pearson's product-moment coefficient.*[**?**] There are other ways or intuition to analyze the relationship between variables, such as mutual information.

Bibliography

[1] R.V. Hogg, J.W. McKean, and A.T. Craig. *Introduction to Mathematical Statistics*. Pearson Education, Limited, 2012. http://books.google.com.hk/books?id=xFlVYAAACAAJ.

[2] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009. http://books.google.com.hk/books?id=7dzpHCHzNQ4C.

[3] Loeve. *Probability Theory I*. F.W.Gehring P.r.Halmos and C.c.Moore. Springer, 1978.

[4] J.A. Rice. *Mathematical Statistics and Data Analysis*. Advanced series. Thomson/Brooks/Cole, 2007. http://books.google.com.hk/books?id=EKA-yeX2GVgC.

[5] wikipedia. Bayesian statistics. http://en.wikipedia.org/wiki/Statistics.

[6] wikipedia. Correlation and dependence. http://en.wikipedia.org/wiki/Correlation_and_dependence.

[7] wikipedia. Statistics. http://en.wikipedia.org/wiki/Bayesian_probability.