TEACHING INDEPENDENCE AND EXCHANGEABILITY

Lisbeth K. Cordani
Instituto Mauá de Tecnologia, Brasil
Sergio Wechsler
Universidade de São Paulo, Brasil
lisbeth@maua.br

Most part of statistical literature, mainly that written in elementary level, is based on the classical (frequentist) approach. The Bayesian school, even if originated in the 18th century, has only recently seen a strong development of its tools. This development, however, has not been seen in a basic level. The disciplines, as well as the teachers, reflect the classical dominance, which reinforces the current paradigm. Although they have different starting points, both approaches, classical and Bayesian, have tools to analyze data, and we should offer the choice to the student. This article deals with two important concepts, one very useful from the classical point of view, which is the concept of independence, and other related to the Bayesian thought, the concept of exchangeability. Definitions and simple examples are presented to relate both approaches, from an elementary point of view.

INTRODUCTION

The inferential statistical approach, in many school levels, is usually taught through the frequentist or classical point of view, from the theories developed by sir Ronald Fisher and Neyman and Pearson in the first half of the 20th century. The Bayesian School, even if very old in its origin (Bayes' Theorem, 18th century), is now seeing a great development in terms of mathematical statistics but the same has not happened in the didactic and pedagogical environment.

Classical statistics development was contemporary the philosophical approach of positivism and it may be thought that the former was influenced by positivistic ideas. Specifically, classical statistics rests on the belief of objectivity toward scientific results. The Bayesian school, on the other hand, had been criticized for its subjectivity and, due also to computational barriers, it was put aside. Nowadays, given the advance of computational resources, this last point is not evoked anymore.

There is also the dominance of the classical approach in the scholar environment, not only in the available textbooks for inertial reasons, but also in research in different fields of knowledge. As the teachers have nothing different to present to their students, they almost always offer the classical version.

Statistics deals with uncertainty and partial information and one of its scope should be to develop some degree of criticism in the students. However, most of the time, there is a kind of pragmatic usage among their users, in order to have a special rule which guarantees the "establishment of the truth," searching definitive answers, typical of the positivist behaviour. Both the classical and the Bayesian approaches offer tools for data analysis and students should be capable to decide which of them suits better to solve an inferential problem. But, as the current paradigm is the classical school, the Bayesian theory concepts have too little attention.

The disciplines in the field of the Statistics emphasise the concept of *independence*, whether among events or random variables, but rarely refer to the concept of exchangeability, which should be offered in the same circumstances. This is probably linked to the fact that both schools are put apart, the classical school using the notion of *independence* intensively, whereas the Bayesian one uses *exchangeability*. In classical statistics, the samples are often supposed to be formed by *independent* and identically distributed random variables (iid), while in Bayesian statistics they can only be considered as such if conditioned to the parameter value, which is based on the notion of exchangeability.

The word *independence*, in its colloquial sense may mean free from whatever subordination, impartiality, self-sufficiency. However, this does not help to understand the probabilistic concept of independence, which is not of common knowledge. Every teacher of basic probability has faced the students' difficulty to understand the probabilistic meaning of the term *independence*, since this word is usually confused with the notion of exclusion (that is, there is some confusion between *independence* and incompatibility).

Apart from this, the concept of independence itself keeps little connection with the student's life, which means it is an abstract concept, with certain empiric meaning, whose application and interpretation are not easy. On the other hand, the notion of "permutability" (also a mathematical one) has a quicker empiric interpretation.

The term *exchangeability* means what is exchangeable: it is said of two factors that may be changed without affecting the results. Here the "factors" may be instances, as throws of a coin, for example. Such throws are exchangeable if the order in which they are done is irrelevant for the probabilities of possible outcomes – the probabilities being the "results."

A judgment of *exchangeability* of instances is a kind of confession from the observer that he cannot distinguish among them, since he believes they are homogeneous. The examples that follow show that – apart from the fact that *independence* is mathematically stronger than *exchangeability* – this last notion is more concrete and, therefore, easier for students to grasp.

However, the notion of *exchangeability*, introduced by De Finetti in 1930, and retrieved by Savage in 1950, is rarely included in basic courses, avoiding a precise registration of students' difficulties. However, this word itself allows an almost automatic understanding of its meaning in terms of probability. This does not occur with the primitive notion of *independence*.

In this article, we intend to introduce the concept of *exchangeability* to beginners in a basic level, showing them its simplicity and comparing it to the notion of *independence*.

DEFINITIONS

Independence

Two events A and B from a sample space S are said to be independent *if* $P(A \text{ and } B) = P(A).P(B)$.

If $P(B)>0$, the events A and B are said to be independent if $P(A|B)=P(A)$, which is equivalent to say the occurrence of B does not affect the probability of A.

In general, a set of events $A_1, A_2, \ldots, A_n$ $(n \geq 2)$ are said to be independent if $P(Ai_1\, Ai_2 \ldots Ai_m) = P(Ai_1), P(Ai_2), \ldots, P(Ai_m)$, whatever be $1 \leq i_1 < i_2 < \ldots < i_m \leq n$, for any m, with $m = 2,3,\ldots,n$. In a sequence of events $A_1, A_2, \ldots\ldots$, they are said to be mutually independent if, whatever $n \geq 2$, the events $A_1, A_2, \ldots, A_n$ are independent. The terms stochastically independent, statistically independent and collectively independent may be used interchangeably.

Exchangeability

Two events A and B are said to be exchangeable if $P(A^cB) = P(AB^c)$, which means there is indifference with respect to the order, because both intersections describe the occurrence of exactly one of the two events, either at the 1st or at the 2nd. instance.

A finite sequence of events $A_1, A_2, \ldots, A_n$ is said to be exchangeable when the probability of occurrence of exactly k of them, whatever order, is always the same, $(\forall k \leq n)$. It can still be said that a sequence of events $A_1, A_2, \ldots$ is exchangeable whenever $A_1, A_2, \ldots, A_n$ are exchangeable, for every $n \geq 1$.

Intuitively, it means that in n throws of the same coin, each particular sequence of Heads and Tails with m Heads and n-m Tails has the same probability, $\forall m \leq n$. The order is irrelevant.

Both definitions were shown for a sequence of events and may be enlarged for a sequence of random variables (see Bernardo and Smith, 1994, for example). It is not so easy to understand the concept of independence. The introduction of the notion of exchangeability – mathematically weaker – could help. For example it can be shown that withdrawals *with* replacement from an urn with unknown composition are just exchangeable and not independent.

EXAMPLES

Example 1

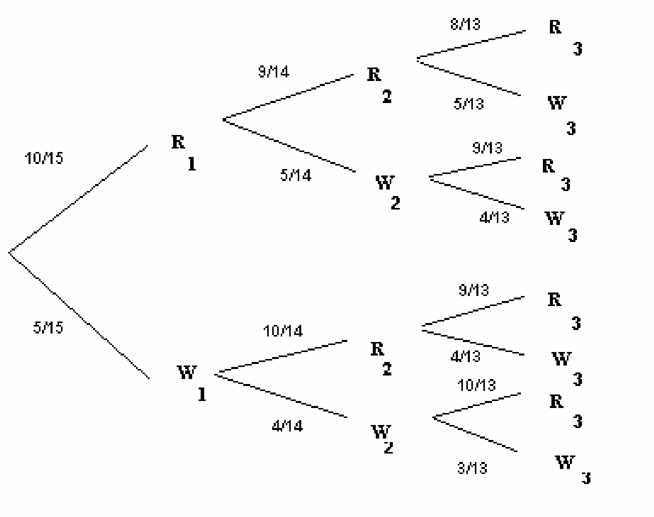
Consider random sampling without replacement of marbles from an urn having known composition, 10 red and 5 white marbles. The marbles are selected one by one and let us define

the event R_i as occurring when the i^{th} sampled marble is red. Analogously, W can be used for the white marble related event. Thus,
R_1 = red marble at the 1st selection
R_2 = red marble at the 2nd selection
R_3 = red marble at the 3rd selection
....
R_k = red marble at the k^{th} selection

The calculation of related probabilities is sometimes hard for beginning students to grasp, especially when there is too much combinatorial reasoning involved. To obtain the probabilities

$$P(R_1),\ P(R_2),\ P(R_3),\ P(R_1R_2),\ P(R_2 R_3\ ),\ P(R_1R_3),\ P(R_2\ |R_1),$$

it may be possible to use a *tree diagram* for better motivation. A tree diagram in such small sized problems is usually easy to construct.



The probabilities values may be easily obtained from the tree diagram constructed for the marbles selected without replacement from the urn:

$P(R_1)$ = 10/15
$P(R_2)$=(10/15)(9/14)+(5/15)(10/14)=10/15
$P(R_3)$=.....=10/15
$P(R_1 R_2)$ = (10/15) (9/14) = 3/7
$P(R_2R_3)$ = ... = 3/7
$P(R_1R_3)$=3/7
$P(R_2|\ R_1)$=9/14.

As $P(R_2|R_1)\ =\ 9/14\ \neq\ 10/15\ =\ P(R_2)$, R_2 and R_1 are not independent, that is, to inform that R_1 has occurred modifies the probability of R_2.

But the probabilities $P(R_i)$ still equal 10/15, for i=1, 2, 3. Furthermore, all intersection probabilities, taken pairwise, of the events R_j, equal 3/7. This happens as the said events are *exchangeable*. This is a well-known example of dependence, with exchangeability: the selections are indistinguishable but nevertheless dependent.

Example 2

Let us consider the same urn of the previous example, that is, an urn having 10 red marbles and 5 white marbles. The marbles are now randomly sampled with replacement from the urn. Letting again the event R_i occur when the i^{th} sampled marble is red, the probabilities

$$P(R_1),\ P(R_2),\ P(R_3),\ P(R_1R_2),\ P(R_2 R_3\ ),\ P(R_1R_3),\ P(R_2\ |R_1),$$

may be easily obtained from the same tree diagram with the appropriate modified values at every branch:

$$P(R_1) = P(R_2) = P(R_3) = (10/15) = 2/3,$$

and also

$$P(R_1R_2) = P(R_2R_3) = P(R_1R_3) = 4/9.$$

And

$$P(R_2|R_1) = (10/15) = 2/3.$$

In this example we obtain the same probabilities $P(R_1)$, $P(R_2)$, and $P(R_3)$ from the previous Example 1. Conditional probabilities, however, are now equal to the unconditional probabilities, e.g., $P(R_2|R_1) = P(R_2)$. This is then a situation where the events are not only *exchangeable*, but also *independent.*

It should be stressed that in both examples above the urn has *known* composition. If the composition is *unknown,* the events are just exchangeable – not independent – even when draws are made *with* replacement.

DISCUSSION

The property that defines exchangeability of n events is that the occurrence of any intersection of k such events has the same probability, i.e., this probability does not depend on their position, but exclusively on k (and a sequence is exchangeable if the property holds for every n). It is a property of symmetry with respect to their order (or their labels). In the examples presented above, the events are exchangeable, even for urns with unknown composition.

According to Barnett (1982), the concept of exchangeability plays the same role in the subjectivist theory as random sampling in Von Mises frequentist theory. In other words, it captures the notion of a sequence of “similar” events. Exchangeability turns the classical i.i.d assumptions into a more realistic structure. Furthermore, Bayesian statistical models use exchangeability instead of independence – the former is weaker. For example, exchangeable withdrawals are used as an assumption in Bayesian modeling the same way as independent withdrawals are used in classical statistics.

Barnett also points out that inference under exchangeable events is usually robust relative to different *a priori* probability measures that may be expressed by different persons. Barnett has in mind the Law of Large Numbers, which is valid for sequences of exchangeable events (sequences of relative frequencies from exchangeable events converge with probability 1 to an unknown, random, limit).

In particular, this teaching approach eases the understanding of the concept of independence, stressing its “geometric” or “mathematical” foundation: As in Geometry, where, Pythagoras Theorem, say, becomes the focus, with the numerical values of a, b and c being irrelevant, in Probability Calculus the numerical values of distribution parameters are irrelevant and attention is given to theorems like, e.g., Chebyshev’s Inequality.

In Statistical Inference, however, not knowing the numerical values of parameters makes an assumption of independence unrealistic. For example, if the probability p of Heads of a coin is (as is the case in Statistical Inference) unknown, predictions about future results are very influenced (dependent on) by knowledge and learning of past results, contradicting the very definition of independence!

In other words, in the Calculus of Probability, the assumption of independence (of coin tosses, say) is innocuous relative to building connections to the real world, whereas in Statistical Inference such an assumption is absurd and, therefore, poorly understood by students. The correct way of teaching is to show that coin tosses are independent, given numerical values (of p). Without knowledge of the numerical value of p, what is to be taught is the exchangeability of the tosses, which in turn implies the aforementioned conditional independence. This is the essence of the celebrated Bruno De Finetti’s Representation Theorem from 1937 (see Kotz and Johnson, 1993, for example).

After long experience, even without a formal research, the authors suggest trying this approach to avoid initial difficulties (which includes semantic confusion, like independence and disjoint events). It is easier for students to understand the ampler concept of *exchangeability* *(indistinguishability of instances)* than the concept of *independence*. As a consequence, it becomes advantageous to teach *exchangeability* before *independence*, the latter being presented as a special case of the former.

ACKNOWLEDGEMENTS

The authors thank the referees for their careful reading and review, which led to an improved final version. Note: An earlier version of this article has been accepted for publication in Spanish Language in Alternativas (Argentina) although the publishing date has not been fixed yet.

REFERENCES

Barnett, V. (1982). *Comparative Statistical Inference* (2nd. edition). Chichester: Wiley.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.

Kotz, S. and Johnson, N. L. (Eds.). (1993). *Breakthroughs in Statistics*. *Volume I*. New York: Springer-Verlag New York, Inc.