

29 December 2005

By: Vlad Tarko, Senior Editor, Sci-Tech News



What Is Ergodicity?

Individual behavior and ensembles

Why are election polls often inaccurate? Why is racism wrong? Why are your assumptions often mistaken? The answers to all these questions and to many others have a lot to do with the non-ergodicity of human ensembles. Many scientists agree that ergodicity is one of the most important concepts in statistics. So, what is it? Ergodicity is usually described in terms of objective properties of an ensemble of objects, and the discussion often gets lost in mathematical subtleties and thus it is often difficult to understand. Nonetheless, I will describe it in bayesian, subjectivist terms; hopefully this will make the concept very accessible. Suppose you are concerned with determining what the most visited parks in a city are. One idea is to take a momentary snapshot: to see how many people are this moment in park A, how many are in park B and so on. Another idea is to look at one individual (or few of them) and to follow him for a certain period of time, e.g. a year. Then, you observe how often the individual is going to park A, how often he is going to park B and so on. Thus, you obtain two different results: one statistical analysis over the entire ensemble of people at a certain moment in time, and one statistical analysis for one person over a certain period of time. The first one may not be representative for a longer period of time, while the second one may not be representative for all the people. The idea is that an ensemble is ergodic if the two types of statistics give the same result. Many ensembles, like the human populations, are not ergodic. The importance of ergodicity becomes manifest when you think about how we all infer various things, how we draw some conclusion about something while having information about something else. For example, one goes once to a restaurant and likes the fish and next time he goes to the same restaurant and orders chicken, confident that the chicken will be good. Why is he confident? Or one observes that a newspaper has printed some inaccurate information at one point in time and infers that the newspaper is going to publish inaccurate information in the future. Why are these inferences ok, while others such as "more crimes are committed by black persons than by white persons, therefore each individual black person is not to be trusted" are not ok? The answer is that the ensemble of articles published in a newspaper is more or less ergodic, while the ensemble of black people is not at all ergodic. If one searches how many mistakes appear in an entire newspaper in one issue, and then searches how many mistakes one news editor does over time, one finds the two results almost identical (not exactly, but nonetheless approximately equal). However, if one takes the number of crimes committed by black people in a certain day divided by the total number of black people, and then follows one random-picked black individual over his life, one would not find that, e.g. each month, this individual commits crimes at the same rate as the crime rate determined over the entire ensemble. Thus, one cannot use ensemble statistics to properly infer what is and what is not probable that a certain individual will do. Or take an even clearer example: In an election each party gets some percentage of votes, party A gets a%, party B gets b% and so on. However, this does not mean that over the course of their lives each individual votes with party A in a% of elections, with B in b% of elections and so on. These were examples of why, in some cases - the non-ergodic cases, one cannot use ensemble statistics to infer something about a particular individual. There is also a complementary problem, faced by the scientists doing opinion polls. They gather data from a very small number of individuals and try to infer the characteristics of the entire ensemble. In order to do this as accurately as possible they don't simply pick the individuals at random; they partition the human ensemble on the basis of some criteria (such as age or income) and afterwards they randomly pick individuals inside each partition being careful that each partition is being

represented. It is worth noting that the so-called margin of error of the opinion polls is not really a margin of error. This margin of error is computed assuming that the human ensemble (or more precisely, the partitions they establish) is (are) ergodic. But in reality they are not. A similar problem is faced by scientists in general when they are trying to infer some general statement from various particular experiments. When is a generalization correct and when it isn't? The answer concerns ergodicity. If the generalization is done towards an ergodic ensemble, than it has a good chance of being correct.