

Population Genetics with Statistics

BY YUEJIAN MO

12-21-2017

Abstract

Population genetics is the subfield of genetics, which deals with genetic differences inside and between population, and is primarily found by Sewall Wright, J. B. S Haldane and Ronald Fisher. Fisher also is described as “a genius who almost single-handedly created the foundations for modern statistical science”. So, population genetics has strong relation with statistical. It will be enjoyable to review the work from the development of population genetics and foundation of modern statistical to nowadays. Major fundamental and meaning topics of population genetics and modern statistical are reviewed, especially how to modern statistical methods are being developed for answering population genetics problem. This review attend to review the work of statisticians do for biological problems during 20 century.

On the Origin of Species and The Scope of Biometrika

In 1859, Charles Robert Darwin published *On the Origin of Species*. Darwin's book introduced idea— **survival of the fittest**. Populations evolve over the course of generations through a process of natural selection: a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms. The similarity of some species, showed that modern organism can evolve from organism in the past. For example, different birds in isolative island still are common in anatomy, and different embryo has similarity, even human embryo exist tail during the early development period.

However, Darwin can only explain why modern organism exist. In another words, Darwin can not find a new species, which evolves from old species in human being history. In UK, there were some mathematicians who appreciated Natural Selection. They tried to understand Natural Selection with statistics, meaning finding the form of a numerical statement. Thus, statisticians Francis Galton, Karl Pearson, and Raphael Weldon established journal *Biometrika* to promote the study of biometrics in 1901. In the first issue, they wrote in *The Scope of Biometrika*:

The unit with which such an enquiry must deal, is not an individual out of a race, but a statistically representative sample of a race; and the result must take the form of a numerical statement, showing the relative frequency with which the various kinds of individuals composing the race occur. — *The Scope of Biometrika* (1901)

For K. Pearson, only the probability distribution make sense. For example, Darwin's finches is not the object of research, but the distribution of a finch population. For a kind of finch, if we can measure beak's length of total finch, we will obtain four parameters of distribution function of beak's length, which is represent beak's length. The four parameters in K. Pearson system are:

Definition 1. *The mean*

Definition 2. *The standard deviation*

Definition 3. *Symmetry*

Definition 4. *Kurtosis*

K. Pearson said, assuming survival of the fittest, although we couldn't observe new species's production under environmental pressure in short time, we can find out the change of four parameters of distribution function. The three editors of *Biometrika* published that *Biometrika* would collect data from all over the world, in order to determine these parameters. They expected that parameters of sample vary different environment. In the next 25 years, *Biometrika* published many data from different place. For each group of data, Pearson and his assistants calculated these four parameters and published. For finding out the different, each four parameters will compare with other corresponding data.

At the begin of 20 century, Weldon designed an experiment to confirm survival of the fittest. During the development of china in Southern England, water outside harbour was less silt than inside. Weldon collected hundred of crabs from Plymouth and Dartmouth, and cultured in bottles. Half of crabs lived in water outside harbour, and other crabs lived in water inside harbour. After some time, Weldon measure the shell of living crabs. Reasonably, parameters changed! However, Weldon was dead before finishing this report.

From the articles in *Biometrika*, it is not obvious that they had proved the Natural Selection successfully. After 1925, more and more theory mathematical articles were published. But K. Pearson and other statisticians provided a lot of analysis techniques for biologists. These techniques, which are widely used today for statistical analysis, include the chi-squared test, standard deviation, and correlation and regression coefficients.

Pearson was important in the founding of the school of biometrics, which was a competing theory to describe evolution and population inheritance at the turn of the 20th century. (wikipedia). It is interesting that, the biometric school, unlike the Mendelians, focused not on providing a mechanism for inheritance, but rather on providing a mathematical description for inheritance, but rather on providing a mathematical description for inheritance that was not causal in nature. Pearson criticized biologists who did not focus on the statistical validity of their theories, stating that "before we can accept [any cause of a progressive change] as a factor we must have not only shown its plausibility but if possible have demonstrated its quantitative ability" (The grammar of science).

Nowadays, we have many alternative model organism to confirm Darwin's theory. To confirm it, we can do thousand of generations in short time for the short life cycle of model organism. Using this method, Natural Selection can be used to explain more and more data. So, most scientists think that parameter change in short time is not necessary to prove the evolution process. However, we learn from the revolution of K. Pearson: the object of science research is not itself, but the distribution function in mathematics.

Fincher and Agriculture—The 14 years in Rothamsted Experiment Station

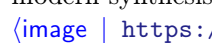
Sir Ronald Aylmer Fisher FRS (17 February 1890 -29 July 1962), was a British statistician and geneticist. In 1919, he began working at the Rothamsted Experimental Station for 14 years, where he analysed its immense data from crop experiments since the 1840s., and developed the analysis of variance (ANOVA).

The usage of chi-square and P value.

Modern population genetics and Statistics

Natural Theory,

modern synthesis

 https://upload.wikimedia.org/wikipedia/commons/thumb/9/97/Modern_Synthesis.svg/800px-Modern_Synthesis.svg.png

References

https://en.wikipedia.org/wiki/Charles_Darwin

(L.) The Scope of Biometrika. (1901). *Biometrika*, 1 (1), 1–2. <https://doi.org/10.1093/biomet/1.1.1>

https://en.wikipedia.org/wiki/Karl_Pearson