Aspects of T. N. Thiele's Contributions to Statistics

Steffen L. Lauritzen
Aalborg University, Department of Mathematical Sciences
Fredrik Bajers Vej 7E
DK-9220 Aalborg, Denmark
steffen@math.auc.dk

1. Introduction

T. N. Thiele (1838–1910) was a Danish astronomer, actuary, applied mathematician, and statistician. He was an original mind with numerous deep and original contributions to statistics that typically were too far ahead of his time to be appreciated by his contempories.

He was professor of Astronomy at the University of Copenhagen from 1875 until 1907. He was actively involved in the process of founding the life insurance company Hafnia in 1972, the Danish Mathematical Society in 1973, and the Danish Society of Actuaries in 1901.

Hald (1981) gives an excellent and thorough description of his main contributions to statistics together with biographical and other information, see also Hald (1998).

Below we give a brief overview of his main contributions and describe and discuss a few of his less well known, but original and important ideas.

2. Overview of main contributions

His most well-known contribution to statistics is the discovery and use of the *half-invariants* of a distribution, later known as the *cumulants,* a term introduced by Fisher and Wishart (1931) as an abbreviation of *cumulative moment function* first used by Fisher (1929).

Already in his first book (Thiele 1889), he defines the half-invariants recursively from the moments, proves their additivity properties, relates them to series expansions in terms of the normal density, and interprets the first two half-invariants as position and spread, and the next standardized cumulants as related to the shape of the distribution, the third and fourth describing skewness and peakedness. The general definition and derivation of half-invariants is given in Thiele (1899), using operator calculus.

Further important contributions in Thiele (1889) include the analysis of skew distributions and their series expansions, the canonical form of the linear normal model via free functions, justification of the method of least squares by an invariance argument, the additive model for two-way layouts, strong emphasis on model criticism via the analysis of residuals and more, see Hald (1981). The elementary text (Thiele 1897) is more accessible, but the most original and therefore difficult material has been left out so it is also less interesting. Unfortunately it is the the elementary text which was subsequently published in an English version (Thiele 1903), whereas the advanced text from 1889 only exists in Danish.

His first paper on the method of least squares (Thiele 1880) is a brilliant *tour de force.* There he derives both Brownian motion with independent and normally distributed increments and variances proportional to time as well as recursive computational procedures for filtering and prediction now known as the Kalman filter and smoother (Kalman and Bucy 1961). He also derives iterative methods for estimating the ratio between system and measurement noise. Clearly this paper was far too advanced for the contemporaries to appreciate. The contents have been described in further detail elsewhere (Lauritzen 1981).

3. Structural equations and causality

Both his book from 1889 and the elementary versions from 1897 and 1903 begin with a chapter that describes the relationship between causality and observation. Thiele, in accordance with the philosophy at the time, is viewing the world as purely deterministic, in the sense that the future state of affairs in principle can be derived from the present circumstances using laws of nature. Science is concerned with discovering and describing and these laws. Random errors are consequences of imperfect description of the present circumstances, whereas systematic errors are consequences of imperfect description of the laws of nature.

It is of some interest that in his first book he formalizes this relation through introducing what we today would call a structural equation. The last few lines on page 2 of Thiele (1889) read as follows, in the author's translation but using Thiele's original notation:

> When the phenomenon under observation can be described by a number o and similarly its acting causes by numbers $v_1, v_2, \dots v_n$ for the essential circumstances and $u_1, u_2, \dots u_\infty$ for its ignored, inessential circumstances, the law of causality in its strong form can be expressed as
>
> $$o = f(v_1, v_2, \dots v_n, u_1, u_2, \dots u_\infty), \tag{1}$$
>
> but because of the actual, reduced observation it is conceived to be
>
> $$o = F(v_1, v_2, \dots v_n). \tag{2}$$
>
> Thus, according to the basic hypothesis of observation, the inessential circumstances are assumed to be constant. The errors that appear as consequence of this assumption are called *accidental errors;* they are unavoidable unless observations exist where all real circumstances can be treated as being essential. As the function F in general is unknown it cannot be avoided that the observations commonly be treated under a different and incorrect assumption
>
> $$o = \Phi(v_1, v_2, \dots v_n) \tag{3}$$
>
> about the shape of the function. Errors which are consequences of this have a less innocent character than accidental errors. They are distinguished as *systematic errors.* These should as far as possible be identified and corrected; when this has been unsuccessfully done, the remaining systematic errors bear witness about the shortcomings of the treatment of this task.

The formal expression in (1) of the causal relationship could be the first explicit structural equation (Wright 1921, Haavelmo 1943). Modern notation would combine the inessential circumstances $u_1, u_2, \dots u_\infty$ into a single but complex error variable ε (Pearl 1998). Interestingly, most modern approaches to causal inference are based on the same deterministic point of view. Thiele does not consider systems of structural equations other than indirectly, through his treatment of the method of least squares.

Note also the emphasis on model criticism in the last sentence. In the remainder of the book this is repeatedly stressed in connection with analysis of model residuals.

4. Likelihood

It has perhaps been less noticed that Thiele in his book from 1889 indeed conceives of a concept that he denotes "Rimelighed" (likelihood) which is a quantitative measure for comparing the correctness of hypothetical values of the unknown probability of success $p/(p+q)$ in $m+l$ trials, of which

m have been observed to be successful. The special term "Rimelighed" is used to distinguish it from probability.

Indeed Thiele produces a diagram of the level curves of the binomial probability as a function both of the theoretical and empirical frequency in the case of $n = m + l = 10$, explaining that for fixed theoretical frequency it displays the binomial probability, whereas one must use the variation along the other axis to explain the 'Bayes' rule' for estimating a proportion. He then writes on page 55 of Thiele (1889) (the author's translation)

> A similar consideration is the basis for the earlier accepted theory of "posterior probabilities". When a single relative frequency is observed and one is uncertain about which of all the infinitely many possible probabilities between 0 and 1 to derive from this frequency, then the thought appears that although all probabilities are possible, they may apparently not all be equally *likely* as hypothetical causes of the observed frequency; the likelihood must for each of them be expressible in terms of a number, and it is not obviously incorrect to measure this likelihood with the probability $\phi(m/(m+l))$ for the observed frequency, calculated under the assumption of $p/(p+q)$ being the unknown value for the probability. That one for these measures of correctness of the hypotheses has used the term "the probability of the correctness of the individual hypotheses" would not have done any harm unless one additionally had permitted calculations with these "probabilities" using all the rules from the direct calculus of probability, despite the fact that the definition of probability does not apply to them.

The early statistical literature does contain various arguments that seem to be related to likelihood, although it mostly is used in connection with the Bayesian argument (Edwards 1974). Thiele seems to be slightly confused when interpreting the diagram, but in the quotation above he is very clearly making the point that likelihood and probability are quite different concepts. On the other hand, Thiele seems unaware of the likelihood only being well-defined up to a multiplicative constant.

However, even though Thiele is well on his way towards getting the full idea of likelihood, he does not use it anywhere else in the book. It seems reasonable to take this fact as an indication that he has not quite understood the universality of the idea. For example, he later spends considerable effort justifying the method of least squares by an argument which is non-Bayesian and eventually produces the canonical form of the linear model and an associated invariance argument (Hald 1981).

Had he understood the universality of the idea of likelihood, he would at least have tried to use it it on the method of least squares, possibly running into the usual difficulties explaining why the degrees of freedom rather than the number of observations should be used in the denominator when estimating the variance.

Epilogue

Thiele was an original mind and as such not always well understood by his contempories. It did not reinforce the spread of his ideas that he primarily wrote in the Danish language although there is a French version of Thiele (1880). Even in Danish his writings were not very clear, but nevertheless there is still much inspiration to be found in Thiele's best writings and the author of this paper is currently translating some of his most important works into English for others to enjoy.

REFERENCES

Edwards, A. W. F. (1974). The history of likelihood. International Statistical Review, 42, 9–15.

Fisher, R. A. (1929). Moments and product moments of sampling distributions. Proceedings of the London Mathematical Society, Series 2, 30, 199–238.

Fisher, R. A. and Wishart, J. (1931). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. Proceedings of the London Mathematical Society, Series 2, 33, 195–208.

Gauss, C. F. (1823). Theoria Combinationes Observationum Erroribus Minimis Obnoxiae. Dieterich, Göttingen.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. Econometrica, 11, 1–12.

Hald, A. (1981). T. N. Thiele's contributions to statistics. International Statistical Review, 49, 1–20.

Hald, A. (1998). A History of Mathematical Statistics from 1750 to 1930. John Wiley and Sons, New York.

Kalman, R. E. and Bucy, R. (1961). New results in linear filtering and prediction. Journal of Basic Engineering, 83 D, 95–108.

Lauritzen, S. L. (1981). Time series analysis in 1880: A discussion of contributions made by T. N. Thiele. International Statistical Review, 49, 319–31.

Pearl, J. (1998). Graphs, causality, and structural equation models. Sociolocical Methods and Research, 27, 226–84.

Thiele, T. N. (1880). Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en 'systematisk' Karakter. Vidensk. Selsk. Skr. 5. Rk., naturvid. og mat. Afd., 12, 381–408. French version: Sur la Compensation de quelques Erreurs quasi-systématiques par la Méthode des moindres Carrés. Reitzel, Copenhague, 1880.

Thiele, T. N. (1889). Forelæsninger over almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode. Reitzel, København. 121 pp.

Thiele, T. N. (1897). Elementær Iagttagelseslære. Gyldendal, København. 127 pp.

Thiele, T. N. (1899). Om Iagttagelseslærens Halvinvarianter. Oversigt over Det Kongelige Videnskabernes Selskabs Forhandlinger, 3, 135–41.

Thiele, T. N. (1903). Theory of Observations. Dayton, London. Reprinted in Annals of Mathematical Statistics 2, 165–308, 1931.

Wright, S. (1921). Correlation and causation. Journal of Agricultural Research, 20, 557–85.

RESUMÉ

On decris des aspects des travails de T. N. Thiele (1838-1910) concernant la théorie de statistique, particulièrement des idées moins connues faisant prévoir la théorie de vraisemblance.