



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



# Schneeweiß: Abraham Wald

Sonderforschungsbereich 386, Paper 439 (2005)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



# Abraham Wald

Hans Schneeweiss

Department of Statistics, University of Munich  
Akademiestr. 1, 80799 München, Germany

## Abstract

This paper grew out of a lecture presented at the 54<sup>th</sup> Session of the International Statistical Institute in Berlin, August 13 - 20, 2003, Schneeweiss (2003). It intends not only to outline the eventful life of Abraham Wald (1902 - 1950) in Austria and in the United States but also to present his extensive scientific work. In particular, the two main subjects, where he earned most of his fame, are outline: Statistical Decision Theory and Sequential Analysis. In addition, emphasis is laid on his contributions to Econometrics and related fields.

Abraham Wald is best known, indeed he is famous, for having founded Statistical Decision Theory and also for having developed the theory of sequential sampling. But he also contributed to many other fields of Statistics often giving decisive impulses or even originating new directions of research.

In Statistics proper one might mention: asymptotic maximum likelihood theory, nonparametric statistics, tolerance intervals, optimal experimental designs, discriminant analysis, statistical quality control, random walks, the problem of incidental parameters, linear models with errors in the variables, and many more.

Econometricians know him for his work on seasonal adjustment, on index number theory, on the identification problem of econometric models, on the problem of estimating such models, and on the famous Wald test as an alternative to the likelihood ratio test.

But he also made major contributions to mathematical economic theory

and to game theory. Interesting enough, he started his academic career as a pure mathematician working in the field of geometry.

This enormous width of interest is certainly due to an eventful life, a life typical for many emigrants from the German sphere of influence in the late thirties of the 20<sup>th</sup> century.

Abraham Wald was born on October 31, 1902 as a citizen of the Austrian - Hungarian Empire in Klausenburg in a German speaking area (Siebenbürgen) then belonging to Hungary. After World War I this region fell to Romania, the city changed its name to Cluj, and Wald became a Romanian citizen.

Born into an orthodox Jewish family, it was contrary to his convictions to attend a public school, where classes were given on Saturdays. He therefore was taught privately, but nevertheless got a degree from a secondary school, which enabled him in 1927 to take up studies at the University of Vienna.

He studied Mathematics, but he did not attend many courses, actually only three. This was possible under the system of complete academic freedom, which then prevailed in German and Austrian universities. Instead he tried his hand to solve open mathematical problems. In the course of these studies he came into contact with Karl Menger and sat in his lectures on geometry and later participated in his Mathematical Colloquium.

This was the beginning of a very productive period in Wald's life, where he published a number of important papers on geometry and topology mostly in "Ergebnisse eines Mathematischen Kolloquiums" (Results of a Mathematical Colloquium).

I cannot go into any details. Suffices it to mention just a few of his discoveries.

He contributed to Hilbert's "Grundlagen der Geometrie" (Foundations of Geometry); he was able to axiomatically characterize a concept of betweenness in metric spaces; he extended Steinitz's theorem on the sums of a series of vectors to spaces of infinite dimension; and - according to Karl Menger (1952) his masterpiece - he gave a new, coordinate free, foundation

of differential manifolds using a novel concept of curvature.

Of greater interest to statisticians might be his characterization of Lebesgue measure as a measure  $\mu$  which assigns the value 1 to every unit cube.

Abraham Wald might have become a great geometer had not fate intervened in his career. After having got his Ph. D. in 1931 he looked for a position at the University of Vienna, but due to the adverse political and economic situation of that time no such position was available for him. Through Karl Menger's intervention, Wald became a private lecturer in mathematics to Karl Schlesinger, a banker with great interest in the mathematical foundations of economics.

It was here that Abraham Wald learned about the concept of a Walrasian equilibrium in a pure exchange economy and also in an economy with production facilities. The economy was described by a set of supply and demand functions for each commodity and for each trader, relating commodities supplied and demanded to a vector of prices for these commodities under the assumption of perfect competition. In equilibrium, demand and supply had to match, leading to a system of equations for the unknown quantities of the commodities traded and their prices. It turns out that the number of equations equals the number of unknowns. But this is certainly not enough to guarantee the existence of a solution. Wald was the first to give sufficient conditions for the existence of a unique solution with nonnegative prices.

This early work (1936) in equilibrium theory was much later taken up by economists in the early fifties culminating in an extended theory of mathematical economics. Nobel Laureate Gerard Debreu acknowledged Wald's work in this field in his Nobel Lecture 1983.

Eventually Wald got a position as a consultant in the Austrian Institute for Business Cycle Research, which then was headed by Oskar Morgenstern. Morgenstern acquainted him with the problem of seasonal adjustment of time series thus starting off Wald's first genuinely statistical contribution.

When the Institute applied a then popular method of Person's to the

series of unemployment data, the result was plainly wrong. The allegedly adjusted series did not only not eliminate seasonal variations, it even turned them to an opposite seasonal movement. Wald was able to show that Person's method only worked correctly if the seasonal pattern was invariant over time. However, with a slowly changing seasonal pattern results such as those observed could easily turn up.

Wald then designed a method that allowed for slow movements of the amplitude in the seasonal component. Suppose a time series  $x_{ij}$ ,  $i = 1, \dots, n, j = 1, \dots, 12$ ,  $i$  denoting the year and  $j$  the month, is decomposable into a smooth, a seasonal, and a random component:

$$x_{ij} = m_{ij} + s_{ij} + r_{ij}.$$

Then for a constant seasonal pattern  $s_{ij} = s_j$ ,  $\bar{s} = 0$ , but for a seasonal pattern with slowly changing amplitude

$$s_{ij} = a_{ij}s_j,$$

where  $a_{ij}$ , the amplitude series, varies slowly around the value 1 and is almost constant over any period of twelve consecutive months.

Wald's method for seasonal adjustment takes this particular seasonal model into account. After eliminating the smooth component by a 12-month moving average, the resulting series  $x_{ij}^*$  is averaged over the years  $i = 1, \dots, n$  for every month  $j$  yielding the mean  $\bar{x}_{.j}^*$  as an estimate of  $\bar{a}_{.j}s_j$ . Due to the slow variation of the amplitude series  $a_{ij}$  the averages  $\bar{a}_{.j}$  are approximately all the same:  $\bar{a}_{.j} \approx \bar{a}$ , where  $\bar{a}$  is the overall average of the series  $a_{ij}$ . Without loss of generality we can take  $\bar{a} = 1$  and so  $\bar{x}_{.j}^*$  is an estimate of  $s_j$ :

$$\bar{x}_{.j}^* \approx s_j.$$

Finally, for any fixed  $i$  and  $j$ , the amplitude  $a_{ij}$  is estimated by a local least squares procedure over twelve months:

$$\sum_{k=j-6}^{j+5} (x_{ik}^* - a_{ij}\bar{x}_{.k}^*)^2 \rightarrow \min_{a_{ij}},$$

which yields the estimate

$$\hat{a}_{ij} = \frac{\sum_k x_{ik}^* \bar{x}_{\cdot k}^*}{\sum_k (\bar{x}_{\cdot k}^*)^2}.$$

The seasonal component then is approximately given by

$$s_{ij} \approx \hat{a}_{ij} \bar{x}_{\cdot j}^*.$$

Subtracting this from the original series  $x_{ij}$  results in a seasonally adjusted time series.

In his book "Berechnung und Ausschaltung von Saisonschwankungen" (1936), Wald explains in depth every single step of this procedure and carefully accounts for the various approximations that appear along the line of calculations. The presentation here is somewhat simplified.

Since the time of Wald's book many other models of seasonal movement have been considered and corresponding adjustment methods have been developed. It befits to pay tribute to the genius loci by mentioning the Berlin procedure of seasonal adjustment. This method starts from a local harmonic decomposition of the seasonal component and assumes a local polynomial for the smooth component and is certainly quite different from Wald's approach. But even this method is based on the same general idea which also underlies Wald's method, namely that the seasonal component just as the smooth component varies slowly over time. Also the techniques involved, though different in detail, are based on the same principles: local smoothing and local least squares.

Another important contribution of Wald's to economic statistics is his work on price index numbers or rather on the index of cost of living, Wald (1937, 1939a). In order to measure the change of prices from period 1 to period 2, statisticians usually compute the Laspeyres or the Paasche price index. Either of them is given by the ratio of the expenditures for a fixed bundle of commodities under the two price vectors  $p_1$  and  $p_2$  prevailing in periods 1 and 2, respectively. If  $q$  is the fixed vector of quantities of this

commodity bundle, then

$$I_{12} = \frac{q^\top p_2}{q^\top p_1}.$$

This index number does not take into account that consumers can and will adjust their consumption to a changing price system according to their preferences. A true cost of living index does not start from a fixed commodity bundle, but from a fixed utility level. It is given by the ratio of the expenditures for two optimal commodity bundles  $q_1, q_2$  under the two price vectors  $p_1, p_2$  such that the utility derived from each of the two commodity bundles is the same:

$$I_{12} = \frac{q_2^\top p_2}{q_1^\top p_1}, \quad u(q_1) = u(q_2).$$

(A commodity bundle  $q$  is optimal under a price system  $p$  and a total expenditure  $e$  if it maximizes utility under the budget constraint  $q^\top p = e$ ). Typically  $q_1$  is the quantity vector observed in period 1. But  $q_2$  is not observed. It is an imputed quantity vector, constructed such that it has the same utility as  $q_1$  while minimizing expenditure under the price system of period 2.

For an economist, this cost of living index comes quite natural, but for a statistician the problem arises of how to compute such an index. It would be easy to compute it if the utility function of the consumer were known. But utilities are something of the mind and are not obviously revealed. To simplify matters, Wald assumes that the utility function is quadratic, at least approximately so in the neighborhood of  $q_1$ :

$$u(q) = q^\top Aq + a^\top q$$

with a symmetric coefficient matrix  $A$  and a coefficient vector  $a$ . It then turns out that, given a price vector  $p$ , the quantity vector  $q$  that maximizes utility under the budget constraint  $q^\top p = e$  is a linear function of total expenditure  $e$ :

$$q = be + c.$$

This is the system of so-called Engel functions, which together form the Engel curve. They can be estimated from a survey of family budgets. Furthermore, it is possible to construct the true cost of living index from the estimated coefficients of the Engel curves of periods 1 and 2. Let  $b_t$  and  $c_t$  be those coefficient vectors for period  $t$  under the price vector  $p_t$ ,  $t = 1, 2$ , then

$$I_{12} = \sqrt{\frac{b_1^\top p_2}{b_2^\top p_1}} + \frac{c_1^\top p_2 - c_2^\top p_1 \sqrt{b_1^\top p_2 / b_2^\top p_1}}{q_1^\top p_1 \left(1 + \sqrt{(b_1^\top p_2)(b_2^\top p_1)}\right)}.$$

If  $c_1 = c_2 = 0$ , i.e., if the Engel curve passes through the origin, this index boils down to Fisher's price index number, the quadratic mean of Laspeyres and Paasche index.

So here is a formula for the true cost of living index, which can actually be used to compute the index. But did it replace the much simpler Laspeyres or Paasche formulas? Certainly not in official statistics. The reason may be twofold. First, many more data need to be collected for the construction of the true index in order to be able to do the necessary econometric estimation of the Engel functions, second, the computations are based on a rather specific and perhaps too restrictive model for the utility function. There is also the idea that one should distinguish between a price index, which just measures price changes and can simply be computed using Laspeyres or Paasche, and a cost of living index, which measures the effect of price changes on utility and rests on additional assumptions about utilities.

A simpler approximation to the cost of living index for more than two periods is the construction of chain index numbers, as e.g. proposed by EUROSTAT with its "harmonized consumer price index".

Nevertheless, Wald's contribution to the theory of price index numbers is still of great interest. It links pure economic theory to empirical concepts.

In those Viennese years Wald came across another statistical problem of a completely different, almost philosophical, kind. Philosophers of the Viennese positivistic school like Hans Reichenbach, but also Karl Popper, had tried to analyze the phenomenon of "randomness". In this context Richard



von Mises introduced the concept of a "Kollektiv" (collective). By this von Mises understood - in the simplest case - an infinite series of zeros and ones which followed each other in a completely irregular way as if being the realization of a series of i.i.d. random variables.

This idea was made precise by the following two postulates: (1) The relative frequency of ones in a beginning section of the series converges with growing size of the section to a fixed number  $p$ , and (2) for any subsection selected from the original series by some selection rule the relative frequency converges to the same number  $p$ . The selection rule should be such that the selection of an element of the series does not depend on the value of the selected element (nor on the values of any element following that one). Otherwise the rule can be quite arbitrary. Examples are selecting every third element or every element succeeding a "1", but not selecting every "1".

The postulated independence of the frequency limit  $p$  should hold with respect to all such selection rules. But, as Wald showed, this requirement is too strong and leads to inconsistencies. A collective in this general sense does not exist. The set of selection rules must be restricted in order for a collective (in a restricted sense) to exist. Wald (1938) showed, among other things, that, for any given *countable* set of selection rules (and for any  $p$ ), collectives - and, in fact, more than countably many collectives - do exist, which obey the two postulates for all selection rules of the set.

Wald argues that the restriction to a countable set of selection rules is so weak a requirement that it is sufficient for all practical purposes. Indeed, if a selection rule is given by a mathematical law and if mathematical laws are formulated within a system of formal logic then there cannot be more than countably many selection rules.

It has been said that Wald's result is only of historical interest as the von Mises approach to probability theory has been superseded by the more effective Kolmogorov axiomatization. But there are still interesting questions open to investigation that surround von Mises' concept of a collective. E.g.,

how can we decide whether a given sequence is a collective with respect to a given set of selection rules. Is the sequence of the decimals of  $\pi$  a collective? Also in studying the randomness of random number generators the concept of a collective lurks behind the corner.

Wald's econometric work came to be known in the U.S., and in 1937 he was invited by Alfred Cowles to become a staff member of the Cowles Commission. Wald hesitated, but soon events in Austria made it impossible for him to stay any longer. In 1938 Austria came under Nazi rule, the "Ergebnisse" ceased to be published, Karl Menger and Oskar Morgenstern left Austria and Abraham Wald was dismissed from the Business Cycle Research Institute. He went to Romania and from there to the U.S.

In a way, he was lucky. Most of his relatives that had stayed at home were eventually murdered.

In America, Wald joined the Cowles Commission at the University of Chicago as research staff member, but in the same year he went to Columbia University at the invitation of Harold Hotelling.

Nevertheless his econometric work continued. Together with Henry B. Mann he published 1943a in *Econometrica* an important paper "on the statistical treatment of linear stochastic difference equations". The authors proved consistency and asymptotic normality of the Quasi-ML estimator, a result which was fundamental to the theory of dynamic simultaneous equation models of econometrics.

In another paper published in the famous No.10 Cowles Commission Monograph (1950) Wald gives a new, somewhat unusual, characterization of the identification of a linear simultaneous equation system. Such a system is given by the matrix equation

$$Ax = u$$

$x$  being an observable stochastic  $p$ -vector and  $u$  an unobservable stochastic  $q$ -vector with  $\mathbb{E}u = 0$ ,  $\mathbb{V}u = \Sigma$  and  $A$  a  $(q \times p)$ -matrix of unknown coefficients. Any linear transformation  $A^* = CA$ ,  $u^* = Cu$  will lead to a similar system

with the same empirical content. Thus  $A$  and  $\Sigma$  are not identifiable from given data  $x_t, t = 1, 2, \dots$ . However economic theory usually provides lots of restrictions on  $A$  (and sometimes also on  $\Sigma$ ) and if these are rich enough,  $A$  (and  $\Sigma$ ) will be identifiable. Wald states necessary and sufficient conditions for the unknown parameters of  $A$  and  $\Sigma$  to be identifiable.

At Columbia University, Wald delved into Statistics proper. He published a series of papers in rather diverse statistical fields some of them coauthored by Jacob Wolfowitz, his student and friend, with whom he shared a long and fruitful period of collaboration.

In these papers, among many other things, Wald together with Wolfowitz invented a method for constructing confidence bands for an unknown continuous distribution function; he studied the moment problem; he and Wolfowitz designed a (nonparametric) runs test of "whether two samples are from the same distribution", a test that should not be missing in any text book on nonparametric methods; with the help of the score function, Wald constructed asymptotically shortest (in the sense of Neyman) confidence intervals for an unknown parameter; together with Mann he suggested a formula for the number of class intervals to be used in a  $\chi^2$ -goodness-of-fit test; also together with Mann he wrote that most useful paper on stochastic order relationships, where they extended Landau's  $o$  and  $O$  notation to a stochastic  $o_p$  and  $O_p$  notation; and he proved several optimality properties of a general parametric test procedure, which has become known as the Wald test (1943b).

Let there be a family of distributions given by the densities  $f(x, \theta)$ ,  $x = (x^1, \dots, x^m)^\top$ ,  $\theta = (\theta^1, \dots, \theta^k)^\top$  and let  $\hat{\theta}$  be the ML estimator from a sample  $x_1, \dots, x_n$  with asymptotic covariance matrix  $\Sigma(\theta)$ . In order to test the null hypothesis

$$g(\theta) = o,$$

where  $g(\theta) = (g^1(\theta), \dots, g^r(\theta))^\top$ ,  $r \leq k$ , construct the test statistic

$$W = n g^\top(\hat{\theta}) \left[ \frac{\partial g(\hat{\theta})}{\partial \theta^\top} \Sigma(\hat{\theta}) \frac{\partial g^\top(\hat{\theta})}{\partial \theta} \right]^{-1} g(\hat{\theta}).$$

Under  $H_0$ ,  $W$  is asymptotically  $\chi^2$ - distributed with  $r$  degrees of freedom. This can be used to construct a critical region of size  $\alpha$ . Among other optimality properties, this test is an asymptotically most stringent test.

Let  $\pi^*(\theta)$  be the upper envelop of the power functions  $\pi(\theta)$  of all tests of size  $\alpha$ . A test of size  $\alpha$  with power function  $\pi_0(\theta)$  is most stringent if the maximal distance between  $\pi_0(\theta)$  and  $\pi^*(\theta)$  is minimal with respect to all other tests of size  $\alpha$ .

But perhaps his most important paper (according to Wolfowitz, 1951) is his *Annals of Mathematical Statistics* (1939b) article "Contributions to the theory of statistical estimation and testing hypothesis", where he designed a common approach to these two main problems of Statistics, estimation and hypothesis testing. That approach was in effect a decision theoretic one, although a fully developed decision theory lay still in the future. But all the main concepts of decision theory were there, like loss and risk functions, Bayes solution, minimax solution, admissibility etc., though not always under these names.

Things changed when America entered the war. A Statistical Research Group (SRG) was founded at Columbia University with the aim of dealing with statistical problems that were of military relevance. In particular, Wald was asked to analyze a sequential sampling procedure for quality control that was suggested by Milton Friedman and Samuel S. Wilks from another SRG at Princeton. Wald succeeded in designing a simple and effective sequential sampling plan that could actually be put to use for quality inspection in the war economy. His work was classified and was not to be published in a journal before the end of the war (there was a paper published by the SRG in 1943). In 1947, the famous book on "Sequential Analysis" appeared, which summarized all the results in sequential sampling up to this time.

The book is easy to read. Wald first develops a sequential likelihood ratio test for a simple hypothesis  $H_0$  against a simple alternative  $H_1$  with given error probabilities,  $\alpha$  and  $\beta$ , of the first and second kind. ( $\alpha$  and  $\beta$  are the probabilities of wrongly rejecting  $H_0$  and  $H_1$ , respectively, when these hypotheses are true). Items are sampled one by one, and each time the probability (density) ratio  $f_1(x_1, \dots, x_n)/f_0(x_1, \dots, x_n) = \lambda_n$  is computed, where  $f_i(x_1, \dots, x_n)$  is the probability (density) of drawing the observed sample  $(x_1, \dots, x_n)$  under  $H_i$ ,  $i = 0, 1$ . Two positive constants  $A$  and  $B$  with  $A > B$  are chosen. If  $\lambda_n$  comes to lie between  $A$  and  $B$ , another item is drawn. If  $\lambda_n \geq A$ , sampling is terminated and  $H_1$  is accepted; if  $\lambda_n \leq B$ ,  $H_0$  is accepted. It can be shown that the sampling process ends with probability 1.  $A$  and  $B$  are chosen so that the two error probabilities  $\alpha$  and  $\beta$  are met. After a very careful discussion, Wald concludes that  $A$  and  $B$  can be determined, to a satisfactory approximation, by setting

$$A = \frac{1 - \beta}{\alpha}, \quad B = \frac{\beta}{1 - \alpha}.$$

In his book, Wald can only prove that this test is near optimal in the sense that under both hypotheses the average sample size is almost minimal. That it is, in fact, optimal was proved later (1948) in a paper together with Wolfowitz by using more sophisticated tools of decision theory.

The simple sequential likelihood ratio test can now be generalized to more complex testing problems, in particular to those of acceptance sampling. Suppose a lot of some mass produced items, ammunition, say, has to be inspected whether it can be accepted or must be rejected. Suppose further that two ratios for defective items,  $p_0$  and  $p_1$ , have been chosen such that a "good" lot, i.e., a lot with  $p \leq p_0$ , should be accepted with high probability at least  $1 - \alpha$  and a "bad" lot, with  $p \geq p_1$ , should be accepted with low probability at most  $\beta$ , then the sequential likelihood ratio test for the hypotheses  $H_0 : p = p_0$  against  $H_1 : p = p_1$  is carried out just as described above.

It is this baffling simplicity of the test procedure which made it a favorite one among practitioners. Still, its use is not so wide spread as one might have thought. The main reason seems to be that the sequential sampling procedure as such is often quite expensive. Only when the inspection costs (as opposed to the costs of sampling) are high, e.g., when inspection leads to the destruction of the inspected item, as in ammunition testing, will the sequential procedure be profitable. Another reason is that the gain in efficiency, as measured by the reduction of average sample size, is not all too impressive if the proportion  $p$  of defective items in the lot is neither very low nor very high; but just in these intermediate cases one would highly need protection against low quality. Finally, nowadays, quality control is built into the production process itself rather than introduced afterwards. On the other hand, for ethical reasons, sequential sampling has gained new importance in clinical trials.

Without any doubt Wald's most important contribution to Statistics is his ingenious idea of founding Statistics on the basis of Decision Theory. This theory, which he developed in his early 1939 article and later on expanded in his book "Statistical Decision Functions" (1950), has become the paradigm of modern Statistics. It unifies and generalizes the theories of estimation and of hypothesis testing. It is so well known that a few indications of its core ingredients should be sufficient.

We start with a family of distributions  $f(x, \theta)$  on a sample space  $X$  characterized by an unknown parameter  $\theta \in \Theta$ . A decision  $d$  from a decision space  $D$  has to be chosen. A nonnegative loss (or weight) function  $W(\theta, d)$  is given that determines the loss due to making decision  $d$  when  $\theta$  is the actual parameter value.

A (randomized) *decision function*  $\delta$  is a mapping from the sample space  $X$  into the space of probability measures on  $D$  ( $D$  being endowed with the structure of a measurable space). For any sample  $x$ ,  $\delta(x)$  is a probability distribution on  $D$ , and, for any subset  $\bar{D} \subset D$ ,  $\delta(x)[\bar{D}]$  is the probability

that the decision  $d$  to be chosen will come from  $\bar{D}$ .

From the loss function the *risk function*  $r(\theta, \delta)$  is derived as the expected loss of adopting a decision function  $\delta$  when  $\theta$  is the actual parameter value

$$r(\theta, \delta) = \mathbb{E}_\theta [\mathbb{E}_{\delta(x)}(W(\theta, d)|x)] = \int \int W(\theta, d)d\delta(x)[d]f(x, \theta)dx.$$

The risk function is the basis for making decisions. In comparing two decision functions  $\delta_1$  and  $\delta_2$ ,  $\delta_1$  is said to be (uniformly) *better* than  $\delta_2$  if

$$r(\theta, \delta_1) \leq r(\theta, \delta_2)$$

for all  $\theta$  with inequality for at least one  $\theta$ . A decision function  $\delta$  is *admissible* if there is no better decision function. A class  $C$  of decision functions is said to be *complete* if for any decision function  $\delta$  not in  $C$  there exists a decision function in  $C$  which is better than  $\delta$ .

Clearly in searching for a "best" decision function one can restrict one's search to a complete class. This explains the importance of complete classes. However, one needs a further criterion in order to choose a "best" decision function among those of a complete class.

Suppose a prior distribution  $\pi$  on  $\Theta$  is given. A *Bayes solution* with respect to  $\pi$  is a decision function  $\delta_B$  that minimizes the expected risk

$$\int r(\theta, \delta)d\pi(\theta).$$

A *minimax solution* is a decision function  $\delta_M$  that minimizes the maximum risk

$$\sup_{\theta} r(\theta, \delta).$$

Finally, a *least favorable prior distribution*  $\pi_0$  maximizes

$$\inf_{\delta} \int r(\theta, \delta)d\pi(\theta).$$

The ultimate goal of the theory is to find minimax solutions. Various criteria for the existence of a minimax solution have been given. In addition,

admissible decision functions and complete classes have been characterized. Under appropriate conditions, the class of Bayes solutions is complete and a minimax solution is a Bayes solution with respect to a least favorable prior distribution. This fact shows the important role of Bayes solutions in the theory, even if a prior distribution in the statistical sense does not exist (because  $\theta$  is not random) or is unknown to the statistician.

The conditions needed are mostly of a topological nature. The topologies are typically defined in an intrinsic way, i.e., derived from the given decision model. Just to give a flavor of what these conditions are about I state one of Wald's results: Let  $W(\theta, d)$  be bounded. The space  $D$  can be endowed with a metric by defining distance in the following way:

$$r(d_1, d_2) = \sup_{\theta} |W(\theta, d_1) - W(\theta, d_2)|.$$

If  $D$  is compact, then there exists a minimax solution  $\delta^0$  and to each prior distribution there exists a corresponding Bayes solution. Furthermore if  $\pi^0$  is a least favorable prior, then  $\delta^0$  is a Bayes solution with respect to  $\pi^0$ .

Wald also dealt with the problem of when one could dispense with randomized decision functions. This is possible, e.g., when  $\Theta$  and  $D$  are finite and the distribution of  $x$  is absolutely continuous. The same is true for the estimation problem, where  $\Theta = D$  is a convex set of  $\mathbb{R}^n$  with a loss function  $W(\theta, d)$  which is convex in  $d$ .

There is a strong link between Statistical Decision Theory and Game Theory. The statistical decision problem is a two-person zero-sum game of the statistician, whose strategies are the decision functions, and Nature, whose strategies are the elements  $\theta \in \Theta$ , the risk function  $r(\theta, \delta)$  being the pay-off function of Nature. Mixed strategies of the statistician are the decision functions, and mixed strategies of Nature are the prior probability distributions.

The book on "Theory of Games and Economic Behavior" by J. von Neumann and O. Morgenstern appeared in 1944 and certainly influenced the final shape of Statistical Decision Theory. But the main ideas of that theory



were already present in Wald's 1939 paper mentioned above, and the specific statistical elements of the theory go far beyond Game Theory.

Statistical Decision Theory has had an enormous impact on modern Statistics. It has become its basis and background. This is true even in those more recent branches of Statistics, where statisticians, due to the complexity of a problem, do no more look for optimal solutions but are satisfied with a procedure that just "works".

It should, however, not be forgotten that there has always been a small school of statisticians that do not adhere to decision theory.

Statistical Decision Theory was certainly not the only field where Wald contributed to the development of Statistics with innovative ideas and novel approaches. Among his many other contribution to Statistics, some of them mentioned above, let me just pick out a problem which was very much discussed for quite a while in Econometrics. It is the problem of estimating a linear relationship

$$\eta = \alpha + \beta\xi$$

when the variables are measured with errors:

$$\begin{aligned}x &= \xi + \delta \\y &= \eta + \epsilon,\end{aligned}$$

$\delta$  and  $\epsilon$  being the measurement errors (or errors in the variables) with expectation 0 and being independent of the error free variables  $\xi$  and  $\eta$ .

Wald's approach looks very simple. Just subdivide the sample  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , into two groups and join the two centers of gravity by a straight line. The subdivision must be independent of the errors and the  $x$ -coordinates of the two centers of gravity must differ by a positive amount in the limit as  $n \rightarrow \infty$ .

Wald (1940) proved that under these conditions  $\alpha$  and  $\beta$  could be estimated consistently. He also gave small-sample confidence region for  $\alpha$  and  $\beta$ , when the errors were normally distributed.

This simple solution to a long-standing problem was often misunderstood, although Wald himself was very clear about the conditions for his result to hold. It was thought that dividing the sample such that the  $x_i$  of the first group were all smaller than those of the second group would provide a subdivision of the required kind. But this subdivision is not independent of the errors, as it depends on the observable variables  $x_i$ , which contain the errors  $\delta_i$ . Consequently the estimator derived from this subdivision is not consistent. On the other hand, taking the first  $\frac{n}{2}$  sample points for the first group and the rest for the second does, in fact, provide an independent subdivision. But now the centers of gravity converge to each other as  $n \rightarrow \infty$  and again the estimator is not consistent.

It is known that when all variables of the model are jointly normally distributed the model is not identifiable and cannot be estimated consistently. So no wonder Wald's method does not work unless some extra information on the latent  $\xi_i$  (or  $\eta_i$ ) is provided. This extra information may come, e.g., as *a priori* knowledge revealing that a certain subsample has values  $\xi_i$  which are all smaller than those of the rest of the sample. In such a case Wald's method produces a consistent estimator.

If the distribution of the  $\xi_i$  is bimodal and the distance of the two modes is large as compared to the range of the distribution of  $\delta$ , then Wald's method with a subdivision according to the magnitude of the  $x_i$  will result in an estimator which is at least approximately consistent.

The method has been improved by considering a subdivision into three groups leaving out the middle group.

It has been said (Wolfowitz, 1952) that Abraham Wald's lectures were clear and lucid. Judging from his publications, this can certainly be confirmed. They are always precise and rigorous and some of them, aimed at a broader audience, are particularly easy to read. But even the more difficult papers are written, without exception, with mathematical rigor, concise, and to the point.

Clearly Abraham Wald was dedicated to his work. After his early years as a pure mathematician he became, what one might call, a full-blooded statistician combining mathematical thinking with practical intuition.

His work was recognized when he became president of the Institute of Mathematical Statistics in 1948 and, in the same year, vice president of the American Statistical Association.

But he also liked to relax in his home and garden. He enjoyed long hikes, and he was very fond of his family. He married Lucille Lang in 1941, and they had two children, Betty and Robert.

When he and his wife died in a plane crash in India in 1950, being on a lecturing tour at the invitation of the Indian government, the statistical community lost one of its most productive and most ingenious members.

## Acknowledgements

I should like to thank Peter Wilrich for fruitful discussion of part of the paper.

## References

### Publications on Abraham Wald

1. Hotelling, H. (1951), Abraham Wald, *American Statistician* **5**, 18-19.
2. Menger, K. (1952), The formative years of Abraham Wald and his work in geometry, *Annals of Mathematical Statistics* **23**, 14-20.
3. Morgenstern, O. (1951), Abraham Wald, 1902-1950, *Econometrica* **19**, 361-367.
4. Tintner, G. (1952), Abraham Wald's contributions to econometrics, *Ann. Math. Statistics* **23**, 21-28.

5. Wolfowitz, J. (1952), Abraham Wald, 1902-1950, *Ann. Math. Statistics* **23**, 1-13.
6. Schneeweiss, H. (2003), Abraham Wald, *Bulletin of the International Statistical Institute 54th Session, Proceedings* **LX 3**, 124-126.

## Publications of Abraham Wald

1. The publications of Abraham Wald, *Ann. Math. Statistics* **23** (1952), 29-33.
2. Über einige Gleichungssysteme der mathematischen Ökonomie, *Zeitschrift für Nationalökonomie* **7** (1936), 637-670.
3. *Berechnung und Ausschaltung von Saisonschwankungen*, (1936) Springer, Wien.
4. Zur Theorie der Preisindexziffern, *Zeitschrift für Nationalökonomie* **8** (1937), 179-219.
5. Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung, *Actualités Scientifiques et Industrielles* **735** (1938), *Colloque Consacré à la Théorie des Probabilités*, Hermann et Cie., 79-99.
6. A new formula for the index of cost of living, *Econometrica* **7** (1939a), 319-331.
7. Contributions to the theory of statistical estimation and testing hypotheses, *Annals of Math. Stat.* **10** (1939b), 299-326.
8. The fitting of straight lines if both variables are subject to error, *Annals of Math. Stat.* **11** (1940), 284-300.
9. On the statistical treatment of linear stochastic difference equations (with H. B. Mann), *Econometrica* **11** (1943a), 173-220.

10. Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Am. Math. Soc.* **54** (1943b), 426-482.
11. *Sequential Analysis*, (1947) John Wiley, New York.
12. *Statistical Decision Functions*, (1950) John Wiley, New York.