Introduction to  
Bartlett (1937) Properties of Sufficiency and Statistical Tests  

D.A.S. Fraser  
York University  

1. Overview  
Fisher’s immense contribution to statistics in the 1920s and 1930s arose in sharp contrast to the general statistics of the time and the developing Neyman-Pearson theory; for an overview, see the published version of the first R.A. Fisher lecture [Bartlett (1965)]. With very strong credentials in the background statistics of the time, Bartlett in this paper and its predecessor [Bartlett (1936)], examines analytically and seriously many of the Fisher concepts against the background theory. From the perspective of the 1965 review paper and even that of the present time, this paper stands as a major initiative in the development of statistics.  

2. Background and Contents  
This study of sufficiency and statistical tests appeared at a time when statistics had a strong basis in the biological sciences with a developing concern for experimentation, particularly in agricultural science. The contemporary statistical forces were in conflicting directions: Fisher (1922, 1925, 1930, 1934, 1935; also 1950) proposing the concepts of sufficiency, information, ancillarity, likelihood, and fiducial; Neyman and Pearson (1933, 1936a, 1936b) developed the formal accept-reject theory of hypothesis testing. Rather than on estimation or testing, this paper focuses on the “structure of small-sample” procedures, on the distributions for inference concerning one parameter in the presence of “irrelevant unknown parameters” and “variation” (error with
a known distribution). It thus provides the seeds for statistical inference as a distinct area of study, largely delineated following the publication of the books by Fisher (1956) and Savage (1954).

The paper uses concepts and theory from both the Fisher and Neyman–Pearson schools and proposes statistical theory for inference in a manner that might now be called unified.

In Sec. 5 exact tests are discussed; these relate to the similar tests of the Neyman–Pearson school and, in part, address the larger issue of the distribution for inference concerning an interest parameter in the presence of "irrelevant unknown" (nuisance) parameters.

As a part of this, Bartlett "state(s) as a general principle that all exact tests of composite hypotheses are equivalent to tests of simple hypotheses for conditional samples." This focuses on what would now be expressed by the factorization

\[ f(y; \psi, \lambda) = f(y_2|y_1; \psi) f(y_1; \psi, \lambda), \]

with \( \psi \) as the interest and \( \lambda \) as the nuisance parameter. In this, \( y_1 \) is sufficient for \( \lambda \), given any particular value for \( \psi \), and is perhaps slightly more general than the prescription: "for the variation in the (response variable) to be independent of the nuisance parameter \( \lambda \), a sufficient ... statistic(,) must exist for \( \lambda \)." The discussion uses the term "variation" in the context of inference, thus anticipating the current separation of variation from effect in general statistical inference, distinct from the special case in the analysis of variance as proposed by Fisher.

As a first example, Bartlett discusses briefly the conditional analysis of independence in the \( 2 \times 2 \) contingency table which has a long and continuing history of proponents and detractors. For some recent views, see Yates (1984).

As a second example, Bartlett considers in Sec. 6 a likelihood ratio test of the homogeneity of sample variances. The starting point, however, is the conditional (also marginal) model given the sufficient statistic for regression parameters; this avoids the usual degrees-of-freedom problem commonly illustrated by the many-means problem [Neyman and Scott (1948)]. As part of approximating the (conditional) likelihood ratio chi-square statistic, he derives corrections of a type now generally called Bartlett corrections; see, e.g., McCullagh (1987).

In the preamble to the example, Bartlett proposes that a procedure "be based directly on the conditional likelihood (for the interest parameter)." This notion of conditional likelihood has only recently been pursued generally, e.g., Cox and Reid (1987); Fraser and Reid (1989), although closely related marginal likelihoods have had longer attention: Fraser (1967, 1968); Kalbfleish and Sprott (1970); Fraser and Reid (1989).

In Sec. 7, Bartlett discusses exact tests of fit, examining initially the normal model. For this, he notes that the conditional distribution of the response \( (y_1, \ldots, y_n) \) given the sufficient statistic \( (\bar{y}, s^2) \) is free of the parameters \( (\mu, \sigma^2) \)
and is thus available for model testing. This procedure anticipates much contemporary theory for model testing and for conditional inference.

At the end of this section, Bartlett briefly mentions a corresponding procedure for the location Cauchy distribution and notes the goodness of fit would be based on the marginal distribution of the configuration statistic; see Fisher (1934). The procedure would now be expressed by the factorization

$$f(y; \psi, \lambda) = f(y_2; \psi)f(y_1 | y_2; \psi, \lambda),$$

in which a marginal variable $y_2$ is used for inference concerning $\psi$. It should be noted that this exact test procedure would contradict Bartlett's general principle cited before Eq. (1). In fact, exact tests can come from marginal (2) as well as conditional (1) models concerning an interest parameter, see, e.g., Fraser and Reid (1989).

Sections 9 and 10 consider many discrete and contingency table examples of conditional inference for an interest parameter in the presence of a nuisance parameter; this would be the current language although Bartlett mainly used the term "exact tests."

Bartlett's paper initiates many of the methods of conditional and marginal inference. The conditional methods apply widely to component parameters in exponential family models, and to generalizations using the exponential as a pattern. The marginal methods apply widely to component parameters in transformation parameter models, and to generalizations using the transformation model as a pattern. For an overview, see Fraser and Reid (1990). The paper also initiates the study of distributional corrections for the likelihood ratio statistic, such as the Bartlett and mean-variance corrections.

3. Personal Background

Maurice S. Bartlett was born in London, June 18, 1910. After studying at Cambridge, he took an appointment at University College, London. In 1934, he left to work at a research station of Imperial Chemical Industries, but returned to the academic world in 1938 as a lecturer in mathematics at Cambridge. In 1947, he took the chair in mathematical statistics at the University of Manchester. In 1960, became professor of statistics at University College, London, and from 1967 to his retirement in 1975, was professor of biostatistics at the University of Oxford. For a detailed interview, see Olkin (1989).

Bartlett's strengths range from mathematics and the foundations through to the realities of statistics in application. His departure from University College to go to Imperial Chemical Industries in 1934 was triggered by the need to teach statistics with an understanding of its applications. He thus had the background and concerns for this early and penetrating investigation toward unity in statistical inference.
References


