The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 204–223

THE WEIGHTED LIKELIHOOD RATIO, LINEAR HYPOTHESES ON NORMAL LOCATION PARAMETERS[1]

BY JAMES M. DICKEY

State University of New York at Buffalo

Raiffa and Schlaifer's theory of conjugate prior distributions is here applied to Jeffrey's theory of tests for a sharp hypothesis, for simple normal sampling, for model I analysis of variance, and for univariate and multivariate Behrens–Fisher problems. Leonard J. Savage's Bayesianization of Jeffrey's theory is given with new generalizations. A new conjugate prior family for normal sampling which allows prior independence of unknown mean and variance is given.

1. Introduction and summary. The subjective, or Bayesian, approach to an hypothesis-testing problem focuses on the ratio of posterior odds to prior odds for the hypothesis if its prior probability is positive. This ratio of odds equals a ratio of likelihood functions which are averaged according to prior distributions under the hypothesis and under its alternative. The theory is here stated for a general loss structure, with the notion of a weighted likelihood ratio replaced by a "weighted utility-likelihood ratio" (Section 2).

Jeffreys' (1948) version of a test, against a vague alternative, of a sharp hypothesis, defined by fixed values of parameters and having positive prior probability, is now known (Dickey and Lientz (1968)) to lead to a weighted likelihood ratio equal, without approximation, to Savage's (1963) ratio of posterior to prior densities; that is, *if* the conditional prior distribution of the nuisance parameters given the sharp hypothesis coincides with the conditional distribution calculated from the distribution under the vague alternative (densities are uniquely defined as elementary derivatives). An analogue of this representation applies to the weighted utility-likelihood ratio (Section 3).

Raiffa and Schlaifer's (1961) conjugate families of prior distributions apply readily to the unknown parameters conditional on a vague alternative to a sharp hypothesis (inducing thereby, we assume, the prior distribution conditional on the sharp hypothesis). One is now left, in practice, with the unavoidable task of assessing realistic bounds on the parameters of the conjugate family and then examining the easily-calculated weighted likelihood ratio for robust inference throughout the bounded region. Any other decision procedure which delivers "more" for less, by ignoring prior opinion within the two hypotheses, is an irresponsible escape. This indictment is directed at the usual tail-area and likelihood-ratio tests for composite hypotheses.

Received November 10, 1969; revised July 1970.

[1] The Bayesian replacement for the usual F test was presented in a talk by the author at the 32nd Annual Meeting of the Institute of Mathematical Statistics, August, 1969. This research was supported by NIGMS-NIH Grant GM-16557.

When it applies, Savage's (1959) notion of "stable estimation" or "precise measurement" requires assessment of the prior density at only one point for calculation of an approximate weighted likelihood ratio. Although a very simple and interesting statistic, the approximate ratio should not form the basis of a formal decision procedure without an evaluation of the quality of the approximation before each "decision".

The conjugate theory and the stable-estimation theory are here presented for a point hypothesis on a univariate normal location parameter (Section 4.1). A new conjugate family is presented which allows prior independence for unknown mean and variance and which yields a weighted likelihood ratio involving Behrens–Fisher densities. A solution is given to the Behrens–Fisher problem (Section 4.2).

I provide a Bayesian replacement for the very popular and elegant, but misleading (Edwards, Lindman, and Savage (1963)), F test for a linear hypothesis within a linear normal model, the Model-I analysis of variance (Section 5.1). A solution is also given to the *multivariate* Behrens–Fisher problem (Section 5.2).

2. Odds and decision theory. Assume a statistical model in which the observed data vector $\mathbf{D} \in R^n$ occurs according to the probability mass or density function (elementary derivative) $\varphi(\mathbf{D} \mid \boldsymbol{\theta})$, depending continuously on an unknown parameter vector $\boldsymbol{\theta} \in R^r$. Assume an individual's opinion about $\boldsymbol{\theta}$ before and after his observation of $\mathbf{D}$ is described by his personal prior and posterior probability distributions $P(S)$ and $P(S \mid \mathbf{D})$. We thus take $\mathbf{D}$ and $\boldsymbol{\theta}$ to have a well-behaved joint personal distribution.

Now, suppose one suspects the unknown parameter $\boldsymbol{\theta}$ of belonging to a given Borel set $H \subset R^r$; or in terms of prior probability, $0 < P(H) < 1$. We consider only cases in which, for the posterior probability too, $0 < P(H \mid \mathbf{D}) < 1$.

Let $\bar{H}$ denote the alternative: $H \cap \bar{H} = \emptyset$ and $P(H) + P(\bar{H}) = 1$. Let O denote odds corresponding to probabilities P,

$$O(H) = P(H)/P(\bar{H}) = P(H)/[1 - P(H)], \tag{2.1}$$

having the immediate properties, $O(\bar{H}) = 1/O(H)$, $P(H) = O(H)/[1 + O(H)]$.

The posterior odds

$$O(H \mid \mathbf{D}) = P(H \mid \mathbf{D})/P(\bar{H} \mid \mathbf{D}), \tag{2.2}$$

are then given, by Bayes' theorem,

$$O(H \mid \mathbf{D}) = L_D(H) \cdot O(H), \tag{2.3}$$

where the "likelihood ratio for H",

$$L_D(H) = \Phi(\mathbf{D} \mid H)/\Phi(\mathbf{D} \mid \bar{H}), \tag{2.4}$$

and

$$\Phi(\mathbf{D} \mid H) = \int \varphi(\mathbf{D} \mid \boldsymbol{\theta})\, dP(\boldsymbol{\theta} \mid H), \tag{2.5}$$

$$\Phi(\mathbf{D} \mid \bar{H}) = \int \varphi(\mathbf{D} \mid \boldsymbol{\theta})\, dP(\boldsymbol{\theta} \mid \bar{H}). \tag{2.6}$$

In case H and $\bar{H}$ are point hypotheses, $L_D(H)$ is the usual likelihood-ratio-test statistic.

Recall briefly the familiar decision-theoretic formulation, e.g., Blackwell and Girshick (1954), in which a reward or utility $U(d, \boldsymbol{\theta})$ depends on the "true" parameter value $\boldsymbol{\theta}$ and on the data $\mathbf{D}$ through the "act" or "decision" $d(\mathbf{D})$. A "Bayes procedure" $d(\cdot)$ calls for an act d which maximizes the posterior expectation over $\boldsymbol{\theta}$ given $\mathbf{D}$ of $U(d, \boldsymbol{\theta})$, hence, yielding the maximum, among procedures $d(\cdot)$, of the prior expectation over $\mathbf{D}$ and $\boldsymbol{\theta}$ of $U(d(\mathbf{D}), \boldsymbol{\theta})$.

In practice, we are left with problems of specifying the statistical model φ, the utility function U and its domains, and the prior distribution of $\boldsymbol{\theta}$. Several realistically possible such specifications should enter into an examination of the robustness of the optimal act d to the specifications. Ideally, data could be reported by a display of the optimal act (or of the weighted likelihood ratio) as a function of prior weightings and of model specifications (Hildreth (1963), Dickey (1970)).

In a two-decision problem with $d = d_H$ or $d_{\bar{H}}$ (maximizing $U(d, \boldsymbol{\theta})$ for every $\boldsymbol{\theta} \in H$ or for every $\boldsymbol{\theta} \in \bar{H}$, respectively) we can assume, without loss of generality, that "wrong" decisions pay nothing,

$$(2.7) \qquad \begin{aligned} U(d_H, \boldsymbol{\theta}) &= 0 \quad \text{all} \quad \boldsymbol{\theta} \in \bar{H}, \qquad \text{and} \\ U(d_{\bar{H}}, \boldsymbol{\theta}) &= 0 \quad \text{all} \quad \boldsymbol{\theta} \in H. \end{aligned}$$

As suggested by Lindley (1961) (for example), this can be achieved mathematically by subtracting from $U(d, \boldsymbol{\theta})$ the function of $\boldsymbol{\theta}$: $U(d_H, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \bar{H}$, and $U(d_{\bar{H}}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in H$.

Then, to choose the maximum of the two posterior expected utilities, we have the *posterior-expected-utility-ratio* criterion,

$$(2.8) \qquad E[U(d_H, \boldsymbol{\theta}) \mid \mathbf{D}]/E[U(d_{\bar{H}}, \boldsymbol{\theta}) \mid \mathbf{D}] \gtrless 1,$$

with either decision permitted for the ratio equal to 1. The denominator in (2.8) is assumed to be positive. (Otherwise, the inequalities would be reversed.)

Since

$$(2.9) \qquad E[U(d, \boldsymbol{\theta}) \mid \mathbf{D}] = E[U \mid H, \mathbf{D}] \cdot P(H \mid \mathbf{D}) + E[U \mid \bar{H}, \mathbf{D}] \cdot P(\bar{H} \mid \mathbf{D}),$$

then by (2.7), we have

$$(2.10) \qquad E[U(d_H, \boldsymbol{\theta}) \mid \mathbf{D}]/E[U(d_{\bar{H}}, \boldsymbol{\theta}) \mid \mathbf{D}] = R_D(H) \cdot L_D(H) \cdot O(H) \gtrless 1,$$

where the *posterior weighted utility ratio*,

$$(2.11) \qquad R_D(H) = E[U(d_H, \boldsymbol{\theta}) \mid H, \mathbf{D}]/E[U(d_{\bar{H}}, \boldsymbol{\theta}) \mid \bar{H}, \mathbf{D}].$$

In case $U(d, \boldsymbol{\theta})$ is constant in $\boldsymbol{\theta}$ within H and $\bar{H}$, then the factor $R_D(H) = R(H) = U(d_H, \boldsymbol{\theta} \in H)/U(d_{\bar{H}}, \boldsymbol{\theta} \in \bar{H})$ does not depend on the data, nor on the prior weightings, and (2.10) is a weighted-likelihood-ratio criterion with threshold $1/[R(H) \cdot O(H)]$. Then $L_D(H)$ is an adequate summary of the data $\mathbf{D}$.

On the other hand, (2.10) can always be considered a weighted-likelihood-ratio criterion by the following device, which justifies the notation,

$$L_{U^*}(H \mid \mathbf{D}) = R_D(H). \tag{2.12}$$

Define the perfect decision,

$$\begin{aligned} d^*(\boldsymbol{\theta}) &= d_H \quad \text{if} \quad \boldsymbol{\theta} \in H \\ &= d_{\bar{H}} \quad \text{if} \quad \boldsymbol{\theta} \in \bar{H}. \end{aligned} \tag{2.13}$$

Then by (2.4), (2.11), $R_D(H)$ has the interpretation (2.12) where $\mathbf{U}^*$ symbolizes the "data" from a hypothetical independent experiment yielding the likelihood function, $\varphi^*(\mathbf{U}^* \mid \theta) \propto U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta})$.

The combined "data" $\mathbf{U}^*$, $\mathbf{D}$ has the product *weighted utility-likelihood ratio*

$$L_{U^*,D}(H) = L_{U^*}(H \mid \mathbf{D}) \cdot L_D(H), \tag{2.14}$$

$$L_{U^*,D}(H) = \Phi^*(\mathbf{D}, \mathbf{U}^* \mid H)/\Phi^*(\mathbf{D}, \mathbf{U}^* \mid \bar{H}), \tag{2.4*}$$

where for some irrelevant $k = k(\mathbf{U}^*, \mathbf{D})$,

$$\Phi^*(\mathbf{D}, \mathbf{U}^* \mid H) = k \int U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta}) \varphi(\mathbf{D} \mid \boldsymbol{\theta})\, dP(\boldsymbol{\theta} \mid H), \tag{2.5*}$$

$$\Phi^*(\mathbf{D}, \mathbf{U}^* \mid \bar{H}) = k \int U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta}) \varphi(\mathbf{D} \mid \boldsymbol{\theta})\, dP(\boldsymbol{\theta} \mid \bar{H}). \tag{2.6*}$$

We have now the analogue of (2.3),

$$E[U(d_H, \boldsymbol{\theta}) \mid \mathbf{D}]/E[U(d_{\bar{H}}, \boldsymbol{\theta}) \mid \mathbf{D}] = L_{U^*,D}(H) \cdot O(H) \gtrless 1. \tag{2.3*}$$

Hence, $L_{U^*,D}(H)$ is an adequate summary of the data, the decision-criterion threshold being $P(\bar{H})/P(H)$.

3. Sharp hypothesis. Consider the case of a sharp hypothesis, namely, H an analytic surface segment in real Cartesian space R^r. In the normal-theory cases considered here, the surface is a linear manifold. Positive probability for this singular set is proposed as a good approximation to many prior opinions.

Let $\boldsymbol{\xi}(\boldsymbol{\theta}) = (\boldsymbol{\eta}', \boldsymbol{\zeta}')' \in R^r$, $\boldsymbol{\eta}(\boldsymbol{\theta}) \in R^q$, and $\boldsymbol{\zeta}(\boldsymbol{\theta}) \in R^{r-q}$ be smooth functions for which the inverse $\theta(\boldsymbol{\xi}) = \theta(\boldsymbol{\eta}, \boldsymbol{\zeta})$ exists and has non-zero Jacobian. We define, for a fixed value $\boldsymbol{\eta}_H$,

$$H: \boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\eta}_H, \qquad \bar{H}: \boldsymbol{\eta}(\boldsymbol{\theta}) \neq \boldsymbol{\eta}_H. \tag{3.1}$$

By an abuse of notation, denote again by P the induced prior measure for $\boldsymbol{\xi}$, and again by H the $\boldsymbol{\xi}$-image of H. Assume as the support of the $\boldsymbol{\xi}$-measure P an analytic surface segment $\Xi \subset R^r$ (possibly R^r itself) of dimension $\dim(\Xi)$, and assume $\dim(H \cap \Xi) < \dim(\Xi) \leq r$. We shall use integration with respect to Lebesgue measure μ on the surface Ξ with differential element denoted $d\boldsymbol{\xi} = d\boldsymbol{\zeta}\, d\boldsymbol{\eta}$; and then on the measure-zero set $H \cap \Xi$, integration with respect to the Lebesgue measure μ_1 with element $d\boldsymbol{\zeta}$. So μ is the product measure of μ_1 and the Lebesgue measure μ_2 with element $d\boldsymbol{\eta}$.

The prior probability measure P is a mixture over H and $\bar{H}$ of the following assumed form. For a set $S \subset R^r$ with Borel-measurable intersection $S \cap \Xi$ (and hence $S \cap H \cap \Xi$),

$$P(S) = P(\bar{H}) \int \int_{S \cap \Xi} f(\eta, \zeta)\, d\zeta\, d\eta + P(H) \int_{S \cap H \cap \Xi} g(\zeta)\, d\zeta. \tag{3.2}$$

The density function f is assumed uniquely defined throughout Ξ as the elementary derivative,

$$f(\xi) = \lim_{\rho \to 0} P[S_\rho(\xi) \mid \bar{H}]/\mu[S_\rho(\xi) \cap \Xi], \tag{3.3}$$

where $S_\rho(\xi)$ denotes the ball of radius ρ centered at ξ. Hence, even though for $\xi \in \bar{H}$, $f(\xi)$ has an interpretation as a conditional density of ξ given $\bar{H}$, if $\xi \in H$, $f(\xi)$ is well defined and not necessarily zero.

Similarly, g is assumed given by

$$g(\zeta) = \lim_{\rho \to 0} P[S_\rho(\eta_H, \zeta) \mid H]/\mu_1[S_\rho(\eta_H, \zeta) \cap H \cap \Xi]; \tag{3.4}$$

and for $\xi \in H$, $g(\zeta)$ has an interpretation as a conditional density of ζ given H.

Note that, with respect to a special dominating measure, the product of μ_1 by μ_2, but with μ_2 modified to have an additional unit mass at $\eta = \eta_H$, P has the density

$$P'(\xi) = P(\bar{H}) f(\xi)[1 - \delta(\eta - \eta_H)] + P(H) g(\zeta)\, \delta(\eta - \eta_H), \tag{3.5}$$

where, as usual, $\delta(\mathbf{0}) = 1$ and $\delta = 0$ for other arguments.

For a sharp hypothesis, equation (2.5) and equation (2.6) take the forms

$$\Phi(\mathbf{D} \mid H) = \int \varphi(\mathbf{D} \mid \eta_H, \zeta) g(\zeta)\, d\zeta, \tag{3.6}$$

$$\Phi(\mathbf{D} \mid \bar{H}) = \int \int \varphi(\mathbf{D} \mid \eta, \zeta) f(\eta, \zeta)\, d\zeta\, d\eta. \tag{3.7}$$

Note the abuse of notation: $\varphi(\mathbf{D} \mid \eta, \zeta)$ for $\varphi(\mathbf{D} \mid \theta(\eta, \zeta))$. The integral (3.6) and integral (3.7), of which $L_D(H)$ is the ratio, are in many cases readily calculated, the integral over η in (3.7) needing no restriction to $\eta \neq \eta_H$.

The proportions Φ^* of the generalized likelihood ratio $L_{U^*,D}(H)$ (2.4*) take forms identical to (3.6) and (3.7) except for the further factor $U(d^*(\theta), \theta)$ in the integrands.

Define for all η, $P'(\eta \mid \bar{H}) = \int f(\eta, \zeta)\, d\zeta$, and define for all η, ζ, $P'(\eta, \zeta \mid \bar{H}, \mathbf{D}) = \varphi(\mathbf{D} \mid \eta, \zeta) \cdot f(\eta, \zeta)/\Phi(\mathbf{D} \mid \bar{H})$, motivating the quite natural definition for all η, $P'(\eta \mid \bar{H}, \mathbf{D}) = \int \varphi(\mathbf{D} \mid \eta, \zeta) \cdot f(\eta, \zeta)\, d\zeta / \Phi(\mathbf{D} \mid \bar{H})$.

THEOREM (Savage's Density Ratio). *If*

$$g(\zeta) = f(\eta_H, \zeta)/\int f(\eta_H, \tilde{\zeta})\, d\tilde{\zeta}, \tag{3.8}$$

then

$$L_D(H) = P'(\eta_H \mid \bar{H}, \mathbf{D})/P'(\eta_H \mid \bar{H}). \tag{3.9}$$

PROOF. (Dickey and Lientz, (1968)). Use (3.8) for g in the numerator $\Phi(\mathbf{D} \mid H)$ (3.6) of L (2.4).

Equation (3.9) was given in special approximate forms by Jeffreys (1948), Lindley (1961), and Savage (1959), (1961). It was discovered, but not published, as a general exact formula by Savage (1963) (see also Patil (1964a) and Dickey (1968)).

Equation (3.9) yields $L_D(H)$ more readily than do (2.5) and (2.6) directly, in case $P'(\boldsymbol{\theta} \mid \bar{H})$ (or $f(\boldsymbol{\eta}, \boldsymbol{\zeta})$) is taken to be one of a "conjugate" family (Raiffa and Schlaifer (1961)) of prior distributions to φ. In this case, the parameters of $P'(\boldsymbol{\theta} \mid \bar{H})$ merely change in a simple way to the parameters of $P'(\boldsymbol{\theta} \mid \bar{H}, \mathbf{D})$, and the numerator of (3.9) is as simple as the denominator.

Savage's approximate form of (3.9) involves an application of "stable estimation" or "precise measurement" to yield an approximate numerator $P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D})$ derived from a uniform (possibly nonintegrable) prior density $f(\boldsymbol{\xi}) \equiv c$. Such a prior density can be a useful approximation to a genuine prior density which is locally constant in a region emphasized by a likelihood function and is not outlandishly greater elsewhere. For then, the posterior density is approximately equal to the normalized likelihood function,

$$P'(\boldsymbol{\xi} \mid \bar{H}, \mathbf{D}) \doteq \psi_D(\boldsymbol{\xi}) = \varphi(\mathbf{D} \mid \boldsymbol{\xi}) / \int \varphi(\mathbf{D} \mid \tilde{\boldsymbol{\xi}})\, d\tilde{\boldsymbol{\xi}}.$$

(Note that prior uniformity in $\boldsymbol{\xi}$ may not mean uniformity in $\boldsymbol{\theta}$, for nonlinear $\boldsymbol{\xi}(\boldsymbol{\theta})$.) Savage's density ratio (3.9) then takes the simple limiting form,

$$L_D(H) \doteq \int \psi_D(\boldsymbol{\eta}_H, \boldsymbol{\zeta})\, d\boldsymbol{\zeta} / P'(\boldsymbol{\eta}_H \mid \bar{H}),$$

so long as we assume that the vicinity of approximate local constancy of $f(\boldsymbol{\xi})$ includes $\boldsymbol{\eta}_H$ for the conditionally likely values of $\boldsymbol{\zeta}$.

A better approximation by Savage (see also Lindley (1961)), follows immediately from the definition of the weighted likelihood ratio (2.4) without the assumption to include $\boldsymbol{\eta}_H$ in the vicinity of prior uniformity,

$$L_D(H) \doteq \int \psi_D(\boldsymbol{\eta}_H, \boldsymbol{\zeta})\, d\boldsymbol{\zeta} / [P'(\boldsymbol{\xi} \mid \bar{H})_{\hat{\boldsymbol{\xi}}} / P'(\boldsymbol{\zeta} \mid H)_{\hat{\boldsymbol{\xi}}_H}],$$

where $\hat{\boldsymbol{\xi}}$ is the maximum likelihood value of $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}_H$ is the maximum likelihood value of $\boldsymbol{\zeta}$ under the constraint $\boldsymbol{\eta} = \boldsymbol{\eta}_H$. Given the assumption (3.8) that $P'(\boldsymbol{\zeta} \mid H) = P'(\boldsymbol{\zeta} \mid \boldsymbol{\eta}_H, \bar{H})$, and given that the nuisance parameter $\boldsymbol{\zeta}$ can be defined such that simultaneously $\hat{\boldsymbol{\xi}}_H = \hat{\boldsymbol{\xi}}$, the unconstrained maximum likelihood value, and $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are approximately prior independent under $\bar{H}$, $P'(\boldsymbol{\xi} \mid \bar{H}) \doteq P'(\boldsymbol{\eta} \mid \bar{H}) P'(\boldsymbol{\zeta} \mid \bar{H})$, then we have

$$L_D(H) \doteq \int \psi_D(\boldsymbol{\eta}_H, \boldsymbol{\zeta})\, d\boldsymbol{\zeta} / P'(\boldsymbol{\eta} \mid \bar{H})_{\hat{\boldsymbol{\eta}}}.$$

In these approximations, the numerator is a function of only the data, and the denominator is a single subjective number to be assessed. Stable estimation is unlikely to yield a realistic approximate posterior distribution for a large number of unknown parameters (Edwards, Lindman, and Savage (1963), page 233).

Dickey and Lientz (1969) show that $L_D(H)$ is invariant with respect to choice of the defining transformations $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ for H, so long as the distribution of the new parameters is the induced one. Cornfield (1966) applies $L_D(H)$ in its original form (2.4) to a comparison problem, violating in his equation (7.8) our assumption (3.8).

If $U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta}) \cdot \varphi(\mathbf{D} \mid \boldsymbol{\theta})$, for $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\eta}, \zeta)$, is continuous in $\boldsymbol{\eta}$ at $\boldsymbol{\eta}_H$ (for the analogue of (3.3)), then $L_{U^*,D}(H)$ reduces to the form analogous to (3.9)

$$(3.9^*) \qquad L_{U^*,D}(H) = P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D}, \mathbf{U}^*)/P'(\boldsymbol{\eta}_H \mid \bar{H}),$$

where the loosely symbolized $P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D}, \mathbf{U}^*)$ is defined similarly to $P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D})$, except for the further factors in the integrands, $U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta})$, $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\eta}, \zeta)$ or $\boldsymbol{\theta}(\boldsymbol{\eta}_H, \zeta)$.

By (2.14),

$$(3.9^{**}) \qquad R_D(H) = L_{U^*}(H \mid \mathbf{D}) = P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D}\,\mathbf{U}^*)/P'(\boldsymbol{\eta}_H \mid \bar{H}, \mathbf{D}).$$

Equation (3.9*) and equation (3.9**) are more valuable as conceptual aids than in practice, where handy prior distributions "conjugate" to $U(d^*(\boldsymbol{\theta}),\boldsymbol{\theta}) \cdot \varphi(\mathbf{D} \mid \boldsymbol{\theta})$ are desired. Yet, a natural, constant, removable, discontinuity of $U = U(d^*(\boldsymbol{\theta}), \boldsymbol{\theta})$ at $\boldsymbol{\eta}_H$, say $U = U_0 \neq 0$ at $\boldsymbol{\eta}_H$ with limit value U_1, can be removed, for use of (3.9*) and (3.9**), by letting

$$(3.10) \qquad \tilde{U}(d^*(\boldsymbol{\theta}),\boldsymbol{\theta}) = (U_1/U_0) \cdot U(d^*(\boldsymbol{\theta}),\boldsymbol{\theta}) \quad \text{if } \boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\eta}_H;$$
$$= U(d^*(\boldsymbol{\theta}),\boldsymbol{\theta}) \quad \text{if } \boldsymbol{\eta}(\boldsymbol{\theta}) \neq \boldsymbol{\eta}_H.$$

Then

$$(3.11) \qquad L_{U^*,D}(H) = (U_0/U_1) \cdot L_{\tilde{U}^*,D}(H), \qquad \text{and}$$

$$(3.12) \qquad L_{U^*}(H \mid \mathbf{D}) = (U_0/U_1) \cdot L_{\tilde{U}^*}(H \mid \mathbf{D}).$$

4. Univariate normal theory, linear hypothesis. In this section we apply the theory of tests just outlined to inference about location parameters of normal populations. Conjugate families of prior distributions are assigned to the unknown population parameters under the alternative to a linear hypothesis. In the theory of conjugate families, devised by Raiffa and Schlaifer (1961), the parameters of the prior distribution combine in the same way with the sufficient statistics from the sample, to form the parameters of the posterior distribution, that the sufficient statistics from two independent samples would combine to form the sufficient statistics of the pooled sample. Lindley (1965) devised the notation used here, whereby prior and posterior parameters are symbolized exactly as the sufficient statistics, but subscripted by 0 (prior) and 1 (posterior). This notation is a limited alternative to the much discussed notation of Raiffa and Schlaifer (1961).

4.1. *A special value for a normal mean.*

Model:

$$\mathbf{D} = (x_1, \cdots, x_n)',$$

independent identically distributed x_i's given μ and σ^2,

$$\text{each} \quad x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2), \tag{4.1}$$

normal with unknown mean μ and with known or unknown variance σ^2.

Hypothesis:

$$H: \mu = \mu_H, \qquad \bar{H}: \mu \neq \mu_H. \tag{4.2}$$

4.1.1. *Variance σ^2 known.*

Prior: $0 < P(H) < 1$; and

$$\mu \mid \sigma^2, \bar{H} \sim N(\bar{x}_0, \sigma^2/n_0), \qquad n_0 > 0. \tag{4.3}$$

Sufficient statistics: n; $\bar{x} = n^{-1}\sum x_i$.

Posterior:

$$\mu \mid \sigma^2, \bar{H}, \mathbf{D} \sim N(\bar{x}_1, \sigma^2/n_1), \qquad \text{where} \tag{4.4}$$

$$n_1 = n_0 + n, \qquad \text{and} \quad \bar{x}_1 = n_1^{-1}(n_0 \bar{x}_0 + n\bar{x}). \tag{4.5}$$

Weighted likelihood ratio:

$$L_D(H) = (n_1/n_0)^{\frac{1}{2}} \exp(\tfrac{1}{2}\sigma^{-2}Q), \qquad \text{where} \tag{4.6}$$

$$Q = n_0(\bar{x}_0 - \mu_H)^2 - n_1(\bar{x}_1 - \mu_H)^2 = n_0 n_1^{-1} n(\bar{x}_0 - \bar{x})^2 - n(\bar{x} - \mu_H)^2. \tag{4.7}$$

If $\bar{x}_0 = \mu_H$, then $Q = -n_1^{-1}n^2(\bar{x} - \mu_H)^2$.

Note that as $n_0 \to 0$ (prior "ignorance" for the alternative $\bar{H}$), $L_D(H) \to \infty$, a point discussed by Cornfield (1966). Edwards, Lindman, and Savage (1963) applied "stable estimation" in the following way, avoiding this pathology. If the prior density under $\bar{H}$ is *approximately* constant in the region emphasized by the likelihood function, then the corresponding posterior distribution is approximately given by (4.4) with $n_0 = 0$, equation (4.4) with the subscripts 1 absent,

$$n^{\frac{1}{2}}\sigma^{-1}(\mu - \bar{x}) \mid \sigma^2, \bar{H}, \mathbf{D} \mathrel{\dot\sim} n^{\frac{1}{2}}\sigma^{-1}(\mu - \bar{x}) \mid \sigma^2, \mu \sim N(0, 1), \tag{4.8}$$

and thus,

$$L_D(H) \doteq [\text{usual (normal) density of } \bar{x} \mid \sigma^2, \mu_H]/P'(\mu \mid \bar{H})_{\mu = \bar{x}}. \tag{4.9}$$

4.1.2. *Variance σ^2 unknown.* The nuisance parameters ζ in Section 3 appear here as σ^2.

Prior: $0 < P(H) < 1$; and

$$\mu, \sigma^{-2} \mid \bar{H} \sim \text{Normal-gamma } (\bar{x}_0, s_0^{\,2}, n_0, \nu_0), \tag{4.10}$$

with parameters $n_0 > 0$, $\nu_0 > 0$, $s_0^{\,2} > 0$ (see also Raiffa and Schlaifer (1961), Section 3.2.5 and Section 11.5); namely,

$$\mu \mid \sigma^2, \bar{H} \sim N(\bar{x}_0, \sigma^2/n_0), \quad \text{as in (4.3)}, \tag{4.11}$$

$$\nu_0 s_0^{\,2} \sigma^{-2} \mid \bar{H} \sim \chi^2_{\nu_0}, \quad \text{chi-squared}, \quad \nu_0 \text{ d.f.}, \tag{4.12}$$

then

$$[n_0(\mu-\bar{x}_0)^2+\nu_0 s_0^2]\sigma^{-2}\,|\,\mu,\bar{H}(\text{or } H) \sim \chi^2_{\nu_0+\Delta(n_0)} \tag{4.13}$$

(where, for later convenience, we define $\Delta(0)=0$ and $\Delta=1$ for nonzero arguments), and

$$(\mu-\bar{x}_0)n_0^{\frac{1}{2}}s_0^{-1}\,|\,\bar{H} \sim t_{\nu_0},\quad \text{Student } t,\quad \nu_0 \text{ d.f.} \tag{4.14}$$

In accordance with (3.8), assume that $\sigma^2\,|\,H$ is distributed identically to the limiting distribution of $\sigma^2\,|\,\mu,\bar{H}$, namely as given by (4.13) with μ set equal to μ_H.

Sufficient statistics: n; $\bar{x}$; $\nu = n-1$, and $\nu_s^2 = SSW = \sum(x_i-\bar{x})^2 = \sum x_i^2 - n\bar{x}^2$.

Posterior:

$$\mu,\sigma^{-2}\,|\,\bar{H},\mathbf{D} \sim \text{Normal-gamma }(n_1,\nu_1,\bar{x}_1,s_1^2), \quad \text{where} \tag{4.15}$$

$$n_1 = n_0+n, \qquad \nu_1 = \nu_0+\nu+\Delta(n_0) \tag{4.16}$$

$$\bar{x}_1 = n_1^{-1}(n_0\bar{x}_0+n\bar{x})(=n_1^{-1}\textstyle\sum_{i=-n_0+1}^{n}\bar{x}_i), \quad \text{and} \tag{4.17}$$

$$\begin{aligned}\nu_1 s_1^2 = SSW_1 &= (\nu_0 s_0^2+n_0\bar{x}_0^2+\nu s^2+n\bar{x}^2)-n_1\bar{x}_1^2\\ &= \nu_0 s_0^2+\nu s^2+n_0 n_1^{-1}n(\bar{x}_0-\bar{x})^2\\ &(=\textstyle\sum_{i=-n_0+1}^{n}(x_i-\bar{x}_1)^2 = \sum_{i=-n_0+1}^{n}x_i^2-n_1\bar{x}_1^2).\end{aligned} \tag{4.18}$$

(The notation x_i, with a nonpositive subscript i, refers to a hypothetical prior sample of size, n_0, sample mean, $\bar{x}_0 = n_0^{-1}\sum_{i=-n_0+1}^{0}x_i$, sample sum of squares about $\bar{x}_0$, $\nu_0 s_0^2 = SSW_0 = \sum_{i=-n_0+1}^{0}(x_i-\bar{x}_0)^2$. This negative-subscript notation is insightful even if strictly nonsensical for n_0 not an integer or $\nu_0 \neq n_0-1$.)

Weighted likelihood ratio: From (3.9) and (4.14), prior ($k=0$) and posterior ($k=1$), $L_D(H) = P_1'(\mu_H)/P_0'(\mu_H)$, where with h_ν being the density of t_ν,

$$\begin{aligned}P_k'(\mu_H) &= n_k^{\frac{1}{2}}s_k^{-1}h_{\nu_k}[n_k^{\frac{1}{2}}s_k^{-1}(\bar{x}_k-\mu_H)]\\ &= \Gamma[\tfrac{1}{2}(\nu_k+1)][\Gamma(\tfrac{1}{2}\nu_k)\pi^{\frac{1}{2}}]^{-1}[n_k/(\nu_k s_k^2)]^{\frac{1}{2}}\\ &\qquad\cdot[1+n_k(\nu_k s_k^2)^{-1}(\bar{x}_k-\mu_H)^2]^{-\frac{1}{2}(\nu_k+1)}.\end{aligned} \tag{4.19}$$

Edwards, Lindman, and Savage (1963) obtained essentially the following stable-estimation approximate form. With the joint prior density of μ, $\log\sigma\,|\,\bar{H}$ approximately constant ($P'(\mu,\sigma^2\,|\,\bar{H}) \mathrel{\dot{\propto}} \sigma^{-2}$) in the region emphasized by the likelihood function, $\nu_0 = n_0 = 0$ in the Normal-gamma theory,

$$n^{\frac{1}{2}}s^{-1}(\mu-\bar{x})\,|\,\bar{H},\mathbf{D} \mathrel{\dot{\sim}} n^{\frac{1}{2}}s^{-1}(\mu-\bar{x})\,|\,\sigma^2,\mu \sim t_\nu \tag{4.20}$$

$$\nu s^2\sigma^{-2}\,|\,\bar{H},\mathbf{D} \mathrel{\dot{\sim}} \nu s^2\sigma^{-2}\,|\,\sigma^2,\mu \sim \chi_\nu^2, \tag{4.21}$$

and

$$L_D(H) \doteq s^{-1}\,[\text{usual } (t) \text{ density of } s^{-1}\bar{x}\,|\,\sigma^2,\mu_H]/P'(\mu\,|\,\bar{H})_{\mu=\bar{x}}, \tag{4.22}$$

in which the numerator is (4.19) with k absent.

4.1.3. *Prior independence of μ and σ^2.* Because of the usually ridiculous prior dependence between μ and σ^2 exhibited by equation (4.11), we here extend Raiffa and Schlaifer's Normal-gamma conjugate family to permit prior independence. Stone (1964) first used such a family of prior distributions.

Prior: $0 < P(H) < 1$; and let $\mu, \sigma^{-2} \mid \bar{H}$ have a density given by the proportionality,

$$(4.23)\quad P'(\mu, \sigma^{-2} \mid \bar{H}) \propto \exp[-\tfrac{1}{2}m_0(\mu - \bar{y}_0)^2] \cdot (\sigma^{-2})^{\frac{1}{2}v_0 - 1} \exp[-\tfrac{1}{2}v_0 s_0^2 \sigma^{-2}] \cdot (\sigma^{-2})^{\frac{1}{2}\Delta(n_0)} \exp[-\tfrac{1}{2}n_0 \sigma^{-2}(\mu - \bar{x}_0)^2],$$

where $\Delta = 1$ for nonzero arguments and $\Delta(0) = 0$. This is a product of a $N(\bar{y}_0, 1/m_0)$ density in μ and a Normal-gamma $(\bar{x}_0, s_0^2, n_0, v_0)$ density in μ, σ^{-2}, but no longer restricted by $n_0 > 0$. We permit prior independence of μ and σ^2 by allowing $n_0 = 0$, for which the former (marginal) distribution of σ^2 (4.12) applies.

The former conditional distribution of σ^2 (4.13) holds. The requirement (3.8) is met by accepting again the first sentence following (4.13).

If $m_0 = 0$ the joint distribution is the former Normal-gamma one.

The former conditional distribution of μ (4.11) has now the analogue,

$$(4.24)\quad \mu \mid \sigma^2, \bar{H} \sim N[(n_0\sigma^{-2} + m_0)^{-1}(n_0\sigma^{-2}\bar{x}_0 + m_0\bar{y}_0), (n_0\sigma^{-2} + m_0)^{-1}].$$

Of course, for $n_0 = 0$ (independence), μ has a marginal distribution given by (4.24) $N(y_0, 1/m_0)$. But if $n_0 > 0$, μ has a marginal density proportional to the product of a Student-t density and the normal-density factor,

$$(4.25)\quad P'(\mu \mid \bar{H}) = C_0 \cdot (n_0^{\frac{1}{2}} s_0^{-1}) h_{v_0}[n_0^{\frac{1}{2}} s_0^{-1}(\mu - \bar{x}_0)] \cdot (2\pi/m_0)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}m_0(\mu - \bar{y}_0)^2],$$

where the normalizing constant C_0 is the reciprocal of the integral of the other factors,

$$(4.26)\quad C_0^{-1} = (s_0^2/n_0 + 1/m_0)^{-\frac{1}{2}} h_{v_0,\infty,\omega_0}[(s_0^2/n_0 + 1/m_0)^{-\frac{1}{2}} \cdot (\bar{x}_0 - \bar{y}_0)],$$

where $h_{v,v^*,\omega}(d)$ denotes the density of the Behrens–Fisher random variable $d_{v,v^*,\omega}$, $0 \leqq \omega \leqq \pi/2$,

$$(4.27)\quad d_{v,v^*,\omega} \sim t_v \cos\omega - t_{v^*} \sin\omega,$$

t_v and t_{v^*} independent. For us, $v^* = \infty$, $t_\infty \sim N(0, 1)$, and

$$(4.28)\quad \tan\omega_0 = (1/m_0)^{\frac{1}{2}}(s_0^2/n_0)^{-\frac{1}{2}}.$$

C_0^{-1} is the density at $(\bar{x}_0 - \bar{y}_0)$ of $(s_0^2/n_0 + 1/m_0)^{\frac{1}{2}} \cdot d_{v,\infty,\omega_0}$ where

$$(4.29)\quad (u^2 + u^{*2})^{\frac{1}{2}} d_{v,v^*,\tan^{-1}(u^*/u)} \sim ut_v - u^* t_{v^*}.$$

The normalizing constant for the joint density (4.23) follows readily from (4.13) and (4.25). Patil (1964b) (1965) and Dickey (1967a) (1968) discuss the calculation of Behrens–Fisher densities.

M. M. Desu has suggested in conversation the choice $\bar{y}_0 = \bar{x}_0$ when $n_0 \neq 0$, for which the prior density (4.23) has the simple property, $E(\mu \mid \sigma^2, \bar{H}) = \bar{x}_0 = E(\mu \mid \bar{H})$.

Posterior (given the data (4.1) with the sufficient statistics following (4.14)): as in (4.23), but with parameters $\bar{x}_1$, s_1^2, n_1, ν_1 determined according to (4.16)–(4.18), and $\bar{y}_1 = \bar{y}_0$, $m_1 = m_0$. Note that prior independence is lost by the transition to the posterior distribution. Note that the six-parameter family (4.23) is closed under sampling also from a normal distribution with variance proportional to σ^2 and another (known) mean, as well as from one with another (known) variance and mean μ. Raiffa and Schlaifer's (1961), Section 11.5, Normal-gamma parametrization permits a unified statement of prior-to-posterior parameter changes for the three kinds of sampling.

Weighted likelihood ratio:

$$L_D(H) = L_D^{(4.19)}(H) \cdot (C_1/C_0), \tag{4.30}$$

where $L_D^{(4.19)}(H)$ was given by (4.19), and C_1 by (4.26) but based on the posterior parameters. In the case of prior independence ($n_0 = 0$),

$$L_D(H) = C_1 \cdot \{n^{\frac{1}{2}} s_1^{-1} h_{\nu_1}[n^{\frac{1}{2}} s_1^{-1}(\bar{x} - \mu_H)]\}. \tag{4.31}$$

4.2. *The Behrens–Fisher Problem.* In this section, the family of Normal-gamma prior distributions is independently applied to two (possibly) different normal populations. It is hoped that future work will exploit the more realistic prior distributions of Section 4.1.3.

$$\text{Model}: \mathbf{D} = (x_1, \cdots, x_n)', \qquad \mathbf{D}^* = (x_1^*, \cdots, x_{n^*}^*)',$$

two independent samples,

$$\text{each} \quad x_i \sim N(\mu, \sigma^2), \qquad \text{each} \quad x_i^* \sim N(\mu^*, \sigma^{*2}), \tag{4.32}$$

unknown μ, μ^*, σ^2, σ^{*2}.

Hypothesis:

$$H: \eta \equiv \mu - \mu^* = 0, \qquad \bar{H}: \eta \neq 0. \tag{4.33}$$

Prior: $0 < P(H) < 1$; and independently as in (4.10),

$$\mu, \sigma^{-2} \mid \bar{H} \sim \text{Normal-gamma} \quad (\bar{x}_0, s_0^2, n_0, \nu_0), \tag{4.34}$$

$$\mu^*, \sigma^{*-2} \mid \bar{H} \sim \text{Normal-gamma} \quad (\bar{x}_0^*, s_0^{*2}, n_0^*, \nu_0^*). \tag{4.35}$$

In particular, independently,

$$n_0^{\frac{1}{2}} s_0^{-1}(\mu - \bar{x}_0) \mid \bar{H} \sim t_{\nu_0}, \qquad n_0^{*\frac{1}{2}} s_0^{*-1}(\mu^* - \bar{x}_0^*) \mid \bar{H} \sim t_{\nu_0^*}. \tag{4.36}$$

Define for $k = 0$, 1, or absent,

$$\tilde{s}_k^2 = s_k^2/n_k + s_k^{*2}/n_k^*. \tag{4.37}$$

Then

$$\eta - (\bar{x}_0 - \bar{x}_0^*) \mid \bar{H} \sim \tilde{s}_0 \cdot d_{\nu_0, \nu_0^*, \omega_0}, \tag{4.38}$$

where $d_{\nu_0,\nu_0^*,\omega_0}$ is the Behrens–Fisher random variable defined by (4.27), and

$$\tan\omega_0 = (s_0^{*2}/n_0^*)^{\frac{1}{2}}(s_0^2/n_0)^{-\frac{1}{2}}. \tag{4.39}$$

Again, assume (3.8) holds.

Sufficient statistics: $\bar{x}$, s^2, n, ν; $\bar{x}^*$, s^{*2}, n^*, ν^*.

Posterior: independently, as in (4.15)–(4.18), $(\mu, \sigma^{-2} \mid \bar{H}, \mathbf{D}, \mathbf{D}^*)$ and $(\mu^*, \sigma^{*-2} \mid \bar{H}, \mathbf{D}, \mathbf{D}^*)$ are distributed Normal-gamma $(\bar{x}_1, s_1^2, n_1, \nu_1)$ and Normal-gamma $(\bar{x}_1^*, s_1^{*2}, n_1^*, \nu_1^*)$. Hence

$$\eta-(\bar{x}_1-\bar{x}_1^*) \mid \bar{H}, \mathbf{D}, \mathbf{D}^* \sim \tilde{s}_1 \cdot d_{\nu_1,\nu_1^*,\omega_1}, \tag{4.40}$$

$$\tan\omega_1 = (s_1^{*2}/n_1^*)^{\frac{1}{2}}(s_1^2/n_1)^{-\frac{1}{2}}. \tag{4.41}$$

Weighted likelihood ratio: $L_D(H) = P_1'(\eta = 0)/P_0'(\eta = 0)$, where

$$P_k'(\eta = 0) = \tilde{s}_k^{-1} h_{\nu_k,\nu_k^*,\omega_k}[\tilde{s}_k^{-1}(\bar{x}_k-\bar{x}_k^*)]. \tag{4.42}$$

Of special interest is $n_0 = n_0^*$, $\nu_0 = \nu_0^*$, $\bar{x}_0 = \bar{x}_0^*$, $s_0^2 = s_0^{*2}$, for which $\omega_0 = \pi/4$, $\tilde{s}_0^2 = 2s_0^2/n_0$, and the denominator of $L_D(H)$ becomes

$$(2s_0^2/n_0)^{-\frac{1}{2}} h_{\nu_0,\nu_0,\pi/4}(0) = (2s_0^2/n_0)^{-\frac{1}{2}} 2[\Gamma(\nu_0+\tfrac{1}{2})/\Gamma(\nu_0+1)] \cdot \{\Gamma[\tfrac{1}{2}(\nu_0+1)]/\Gamma(\tfrac{1}{2}\nu_0)\}^2(\pi\nu_0)^{-\frac{1}{2}}. \tag{4.43}$$

Working as a Ph.D. student under Leonard J. Savage, Patil (1964a) obtained essentially the following stable-estimation approximate form of $L_D(H)$. With the appropriate joint prior density proportional to $\sigma^{-2}\sigma^{*-2}$ for $(\mu, \sigma^2, \mu^*, \sigma^{*2} \mid \bar{H})$ in the region emphasized by the likelihood function (including $\mu-\mu^* = 0$),

$$\eta-(\bar{x}-\bar{x}^*) \mid \bar{H}, \mathbf{D}, \mathbf{D}^* \mathrel{\dot\sim} \tilde{s} d_{\nu,\nu^*,\omega}, \tag{4.44}$$

$$\tan\omega = (s^{*2}/n^*)^{\frac{1}{2}}(s^2/n)^{-\frac{1}{2}}, \tag{4.45}$$

and with $\eta = \mu-\mu^*$,

$$L_D(H) \doteq \tilde{s}^{-1} h_{\nu,\nu^*,\omega}[\tilde{s}^{-1}(\bar{x}-\bar{x}^*)]/P'(\eta \mid \bar{H})_{\eta=\bar{x}-\bar{x}^*}. \tag{4.46}$$

5. Multivariate normal extensions.

5.1. *Linear normal model.* (See also Raiffa and Schlaifer (1961), Chapters 12 and 13.)

Model: $\mathbf{D} = \mathbf{y} = (y_1, \cdots, y_n)'$,

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathrm{N}^{(n)}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{W}^{-1}), \quad n\text{-variate normal}, \tag{5.1}$$

known full-rank "design" matrix $(n \times r)\mathbf{X}$, unknown regression coefficient vector $\boldsymbol{\beta} \in R^r$, known variance structure matrix $(r \times r)\mathbf{W}^{-1} > 0$ (positive-definite symmetric), and the either known or unknown variance scale parameter σ^2.

Define the "products matrix" $\mathbf{N}(r \times r)$ and "generalized least squares" estimate $\mathbf{b}$ of $\boldsymbol{\beta}$ by

$$\mathbf{N} = \mathbf{X}'\mathbf{W}\mathbf{X}, \qquad \mathbf{N}\mathbf{b} = \mathbf{X}'\mathbf{W}\mathbf{y}. \tag{5.2}$$

Define the "residual sum of squares" or "Sum of Squares Within" by

$$(5.3)\qquad \nu s^2 = SSW = \mathbf{W}((\mathbf{y}-\mathbf{X}\mathbf{b})) = \mathbf{W}((\mathbf{y}))-\mathbf{N}((\mathbf{b})), \qquad \nu = n-r,$$

in which double parentheses denote a quadratic form, $\mathbf{M}((\mathbf{x})) = \mathbf{x}'\mathbf{M}\mathbf{x} = \sum\sum m_{ij} x_i x_j$. Recall that $\mathbf{X}\mathbf{b}$ is the projection of $\mathbf{y}$ on the subspace of R^n spanned by the columns of $\mathbf{X}$ with respect to the $\mathbf{W}$-inner-product $(\mathbf{y}, \mathbf{z}) = \mathbf{y}'\mathbf{W}\mathbf{z}$; and SSW is the squared $\mathbf{W}$-distance between $\mathbf{y}$ and $\mathbf{X}\mathbf{b}$, $(\mathbf{y}-\mathbf{X}\mathbf{b}, \mathbf{y}-\mathbf{X}\mathbf{b})$.

Hypothesis:

$$(5.4)\qquad H:\boldsymbol{\eta} = \boldsymbol{\eta}_H, \qquad \bar{H}:\boldsymbol{\eta} \neq \boldsymbol{\eta}_H,$$

where

$$(5.5)\qquad \boldsymbol{\eta} = \mathbf{C}_H\boldsymbol{\beta}, \qquad (q \times r)\mathbf{C}_H \quad \text{full rank.}$$

Denote by $B_H{}^N$ the $\mathbf{N}$-inner-product projection operator of R^r onto the linear manifold H. Then

$$(5.6)\qquad B_H{}^N\boldsymbol{\beta} = \mathbf{B}_K{}^N\boldsymbol{\beta} + B_H{}^N\mathbf{0},$$

where $(r \times r)\mathbf{B}_K{}^N$ is the *matrix* operator for $\mathbf{N}$-projection onto the parallel sub*space* $K = \{\boldsymbol{\beta}: \mathbf{C}_H\boldsymbol{\beta} = \mathbf{0}\}$, namely,

$$(5.7)\qquad \mathbf{B}_K{}^N = \mathbf{I}_r - \mathbf{N}^{-1}\mathbf{C}_H{}'(\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H{}')^{-1}\mathbf{C}_H = \mathbf{N}^{-1}\mathbf{C}'_{H\perp}(\mathbf{C}_{H\perp}\mathbf{N}^{-1}C'_{H\perp})^{-1}C_{H\perp}$$
$$= \mathbf{A}_K(\mathbf{A}_K{}'\mathbf{N}\mathbf{A}_K)^{-1}\mathbf{A}_K{}'\mathbf{N},$$

where, in meaningful notation, $C_{H\perp}((r-q) \times r)$ is any full-rank solution to $\mathbf{C}_{H\perp}\mathbf{N}^{-1}\mathbf{C}_H{}' = \mathbf{0}$, namely $\mathbf{C}_{H\perp} = \mathbf{A}_K{}'\mathbf{N}$ where $\mathbf{A}_K(r \times (r-q))$ is any full-rank matrix having column vectors spanning K. Also,

$$(5.8)\qquad B_H{}^N\mathbf{0} = \mathbf{N}^{-1}\mathbf{C}_H{}'[\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H{}']^{-1}\boldsymbol{\eta}_H,$$

the unique point in H whose projection on K is $\mathbf{0}$, and so

$$(5.8\text{a})\qquad B_H{}^N\mathbf{b} = \mathbf{b} + \mathbf{N}^{-1}\mathbf{C}_H{}'(\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H{}')^{-1}(\boldsymbol{\eta}_H - \mathbf{C}_H\mathbf{b}).$$

Equation (5.8a) and the member of (5.7) in terms of $\mathbf{C}_H$ follow from DeGroot (1970), p. 254, or Rao (1965), p. 49.

The usual "Sum of Squares Between" is then

$$(5.9)\qquad SSB = \mathbf{N}((\mathbf{b} - B_H{}^N\mathbf{b})) = \mathbf{W}((\mathbf{X}\mathbf{b} - \mathbf{X}B_H{}^N\mathbf{b})) = SST - SSW$$
$$= (\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H{}')^{-1}((\mathbf{C}_H\mathbf{b} - \boldsymbol{\eta}_H))$$
$$(= \mathbf{N}((\mathbf{b} - B_H{}^N\mathbf{0})) - \mathbf{N}((B_K{}^N\mathbf{b}))),$$

where

$$(5.10)\qquad SST = W((\mathbf{y} - \mathbf{X}B_H{}^N\mathbf{b})).$$

The "Sum of Squares Total" has the above decomposition by the $\mathbf{W}$-orthogonality of $\mathbf{y} - \mathbf{X}\mathbf{b}$ and $\mathbf{X}(\mathbf{b} - B_H{}^N\mathbf{b})$.

The traditional $\chi_q{}^2$ statistic for H is $\chi^2(\mathbf{D}) = SSB/\sigma^2$, and the traditional $F_{q,\nu}$ statistic for H is $F(\mathbf{D}) = (SSB/q)/(SSW/\nu)$.

5.1.1. *Variance scale σ^2 known.*

Prior: $0 < P(H) < 1$; and

$$\boldsymbol{\beta} \,|\, \sigma^2, \bar{H} \sim \mathrm{N}^{(r)}(\mathbf{b}_0, \sigma^2 \mathbf{N}_0^{-1}), \qquad (r \times r)\mathbf{N}_0 > 0, \tag{5.11}$$

with $\boldsymbol{\beta} \,|\, \sigma^2, H$ distributed as the limiting distribution of $\boldsymbol{\beta} \,|\, \sigma^2, \bar{H}$, from (5.11) given $\mathbf{C}_H \boldsymbol{\beta} = \boldsymbol{\eta}_H$.

Sufficient statistics: $\mathbf{N}$, $\mathbf{b}$ (5.2).

Posterior:

$$\boldsymbol{\beta} \,|\, \sigma^2, \bar{H}, \mathbf{D} \sim \mathrm{N}^{(r)}(\mathbf{b}_1, \sigma^2 \mathbf{N}_1^{-1}), \qquad \text{where} \tag{5.12}$$

$$\mathbf{N}_1 = \mathbf{N}_0 + \mathbf{N}, \qquad \mathbf{b}_1 = \mathbf{N}_1^{-1}(\mathbf{N}_0 \mathbf{b}_0 + \mathbf{N}\mathbf{b}). \tag{5.13}$$

Choose some $\mathbf{C}_H{}^*$ for which

$$\mathbf{C} = (\mathbf{C}_H', \mathbf{C}_H{}^{*\prime})' \tag{5.14}$$

is nonsingular. Assign the nuisance parameters $\boldsymbol{\zeta}$ of Section 3,

$$\boldsymbol{\zeta} = \mathbf{C}_H{}^* \boldsymbol{\beta}, \qquad \boldsymbol{\xi} = (\boldsymbol{\eta}', \boldsymbol{\zeta}')' = \mathbf{C}\boldsymbol{\beta}, \tag{5.15}$$

and for $k = 0$, 1, or absent,

$$\tilde{N}_k = (\mathbf{C}^{-1})' \mathbf{N}_k \mathbf{C}^{-1}, \qquad \tilde{\mathbf{b}}_k = \mathbf{C}\mathbf{b}_k. \tag{5.16}$$

We have the natural definition and straightforward identity, in the usual notation for partitioned matrices,

$$\begin{aligned} SSB_k &= \mathbf{N}_k((\mathbf{b}_k - B_H{}^{N_k}\mathbf{b}_k)) = (\tilde{\mathbf{N}}_k)_{11.2}(((\tilde{\mathbf{b}}_k)_1 - \boldsymbol{\eta}_H)) \\ &= (\mathbf{C}_H \mathbf{N}_k{}^{-1} \mathbf{C}_H')^{-1}((\mathbf{C}_H \mathbf{b}_k - \boldsymbol{\eta}_H)). \end{aligned} \tag{5.17}$$

The prior and posterior distributions,

$$\boldsymbol{\xi} \,|\, \sigma^2, \bar{H}(\text{and } \mathbf{D} \text{ iff } k \neq 0) \sim \mathrm{N}^{(r)}(\tilde{\mathbf{b}}_k, \sigma^2 \tilde{\mathbf{N}}_k{}^{-1}), \tag{5.18}$$

and hence,

$$\boldsymbol{\eta} \,|\, \sigma^2, \bar{H}(\text{and } \mathbf{D} \text{ iff } k \neq 0) \sim \mathrm{N}^{(r)}((\tilde{\mathbf{b}}_k)_1, \sigma^2 (\tilde{\mathbf{N}}_k{}^{-1})_{11}). \tag{5.19}$$

According to our assumption (3.8)

$$\boldsymbol{\zeta} \,|\, \sigma^2, H(\text{and } \mathbf{D} \text{ iff k} \neq 0) \sim \mathrm{N}^{(r-q)}(B_H{}^{\tilde{N}_k}\tilde{\mathbf{b}}_k, \sigma^2(\tilde{\mathbf{N}}_k{}^{-1})_{22\cdot 1}), \qquad \text{where} \tag{5.20}$$

$$B_H{}^{\tilde{N}_k}\tilde{\mathbf{b}}_k = (\tilde{\mathbf{b}}_k)_2 + (\tilde{\mathbf{N}}_k{}^{-1})_{21}(\tilde{\mathbf{N}}_k)_{11\cdot 2}[\boldsymbol{\eta}_H - (\tilde{\mathbf{b}}_k)_1]. \tag{5.21}$$

Weighted likelihood ratio: $L_D(H) = P_1'(\boldsymbol{\eta}_H)/P_0'(\boldsymbol{\eta}_H)$, where, with the standard q-variate normal density denoted $\varphi^{(q)}$, and with the chi-squared density on ν degrees-of-freedom denoted g_ν (density of $\chi_\nu{}^2 = ||\mathrm{N}^{(q)}(\mathbf{0}, \mathbf{I})||^2$, sum of squared coordinates),

$$\begin{aligned} P_k'(\boldsymbol{\eta}_H) &= |\sigma^{-2}(\tilde{\mathbf{N}}_k)_{11\cdot 2}|^{\frac{1}{2}} \varphi^{(q)}\{[\sigma^{-2}(\tilde{\mathbf{N}}_k)_{11\cdot 2}]^{\frac{1}{2}}[(\tilde{\mathbf{b}}_k)_1 - \boldsymbol{\eta}_H]\} \\ &= [\tfrac{1}{2}q\pi^{\frac{1}{2}q}/\Gamma(\tfrac{1}{2}q+1)]^{-1}\sigma^{-2}|(\tilde{\mathbf{N}}_k)_{11\cdot 2}|^{\frac{1}{2}}(SSB_k)^{1-\frac{1}{2}q} g_q(SSB_k/\sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}q}|(\tilde{\mathbf{N}}_k)_{11\cdot 2}|^{\frac{1}{2}} \exp[-\tfrac{1}{2}\sigma^{-2}SSB_k]. \end{aligned} \tag{5.22}$$

(The unique symmetric square root of a positive definite matrix $\mathbf{M}$ is here indicated by $\mathbf{M}^{\frac{1}{2}}$ although it is not necessary to actually calculate $\mathbf{M}^{\frac{1}{2}}$ in practice.) The quantity SSB_k in (5.22), as given by (5.17) is the usual analysis of variance quantity, but with the products matrix $\mathbf{N}$ and the least-squares estimate $\mathbf{b}$ replaced by the prior or posterior information-structure matrix $\mathbf{N}_k$ and the prior or posterior mean $\mathbf{b}_k$. By using the identity

$$(5.23)\qquad |(\tilde{\mathbf{N}}_k)_{11\cdot 2}| = 1/|(\tilde{\mathbf{N}}_k^{-1})_{11}| = 1/|\mathbf{C}_H\mathbf{N}_k^{-1}\mathbf{C}_H'|,$$

we have the result,

$$(5.24)\qquad L_D(H) = (|\mathbf{C}_H\mathbf{N}_0^{-1}\mathbf{C}_H'|/|\mathbf{C}_H\mathbf{N}_1^{-1}\mathbf{C}_H'|)^{\frac{1}{2}}\exp(\tfrac{1}{2}\sigma^{-2}Q),$$

where

$$(5.25)\quad Q = SSB_0 - SSB_1 = [(\tilde{\mathbf{N}}_0)_{11\cdot 2}(\tilde{\mathbf{N}}_1^{-1})_{11}(\tilde{\mathbf{N}})_{11\cdot 2}](((\tilde{\mathbf{b}})_1 - (\tilde{\mathbf{b}}_0)_1)) - SSB.$$

When stable estimation applies under $\bar{H}$, namely approximate constancy of the prior density of $\boldsymbol{\beta}\,|\sigma^2, \bar{H}$ relative to the likelihood function, then the above posterior distributions ($k = 1$) apply approximately with k absent, equation (5.12) becoming, with $\mathbf{N}_0 = \mathbf{0}$,

$$(5.26)\qquad \sigma^{-1}\mathbf{N}^{\frac{1}{2}}(\boldsymbol{\beta}-\mathbf{b})\,|\,\sigma^2, \bar{H}, \mathbf{D} \mathrel{\dot\sim} \sigma^{-1}\mathbf{N}^{\frac{1}{2}}(\boldsymbol{\beta}-\mathbf{b})\,|\,\sigma^2, \boldsymbol{\beta} \sim \mathbf{N}^{(r)}(\mathbf{0}, \mathbf{I}).$$

Hence, assuming for $\mathbf{C}_H{}^* = \mathbf{C}_{H^\perp}$, $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are approximately prior independent under $\bar{H}$,

$$(5.27)\quad L_D(H) \doteq [\text{usual (multiv. normal) density of } \mathbf{C}_H\mathbf{b}\,|\,\sigma^2, \boldsymbol{\beta}, H]/P'(\boldsymbol{\eta}\,|\,\bar{H})_{\boldsymbol{\eta}=\mathbf{C}_H\mathbf{b}}$$
$$= [\text{usual } (\chi_q{}^2) \text{ density of } SSB/\sigma^2\,|\,\sigma^2, \boldsymbol{\beta}, H]/$$
$$P'[\mathbf{N}((\boldsymbol{\beta} - B_H{}^N\boldsymbol{\beta}))/\sigma^2\,|\,\bar{H}]_{SSB/\sigma^2}$$

where the first numerator is equal to (5.22) without k, and the two denominators are related by

$$(5.28)\quad P'[\mathbf{N}((\boldsymbol{\beta} - B_H{}^N\boldsymbol{\beta}))/\sigma^2\,|\,\bar{H}]_{SSB/\sigma^2}$$
$$= \sigma^2[\tfrac{1}{2}q\pi^{\frac{1}{2}q}/\Gamma(\tfrac{1}{2}q+1)]\,|\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H'|^{\frac{1}{2}}\,SSB^{\frac{1}{2}q-1}\cdot P'(\boldsymbol{\eta}\,|\,\bar{H})_{\boldsymbol{\eta}=\mathbf{C}_H\mathbf{b}}$$

formula (5.28) to be found, in essence, in Lindley (1965), pages 95–98.

Note that $\mathbf{N}((\boldsymbol{\beta} - B_H{}^N\boldsymbol{\beta})) = (\mathbf{C}_H\mathbf{N}^{-1}\mathbf{C}_H')^{-1}((\mathbf{C}_H\boldsymbol{\beta} - \boldsymbol{\eta}_H))$.

Edwards, Lindman, and Savage's (1963) equation [22], page 233, is thus corrected and generalized.

5.1.2. *Variance scale σ^2 unknown.* The nuisance parameters $\boldsymbol{\zeta}$ in Section 3 appear here as $\boldsymbol{\zeta}$, σ^2.

Prior: $0 < P(H) < 1$; and

$$(5.29)\qquad \boldsymbol{\beta}, \sigma^{-2}\,|\,\bar{H} \sim (r\text{-variate Normal})\text{-gamma}\quad (\mathbf{b}_0, s_0{}^2, \mathbf{N}_0, \nu_0),$$

with parameters $N_0 > 0$, $\nu_0 > 0$, $s_0^2 > 0$ (see also Raiffa and Schlaifer (1961), Section 3.2.5 and Section 12.4); namely,

$$\beta \mid \sigma^2, \bar{H} \sim N^{(r)}(b_0, \sigma^2 N_0^{-1}), \qquad \text{as in (5.11)}, \tag{5.30}$$

$$\nu_0 s_0^2 \sigma^{-2} \mid \bar{H} \sim \chi^2_{\nu_0}, \tag{5.31}$$

then

$$[N_0((\beta - b_0)) + \nu_0 s_0^2]\sigma^{-2} \mid \beta, \bar{H}(\text{or } H) \sim \chi^2_{\nu_0 + \text{rank}(N_0)} \tag{5.32}$$

(until later, N_0 is of full rank, r), and

$$s_0^{-1} N_0^{\frac{1}{2}}(\beta - b_0) \mid \bar{H} \sim t^{(r)}_{\nu_0}. \tag{5.33}$$

For a treatment of the r-variate t distribution, $t_\nu^{(r)}$, see Raiffa and Schlaifer (1961); $t_\nu^{(r)}$ is distributed like $N^{(r)}(0, I_r)/(\chi_\nu^2/\nu)^{\frac{1}{2}}$, with density in R^r,

$$h_\nu^{(r)}(t) = \Gamma[\tfrac{1}{2}(\nu + r)][\Gamma(\tfrac{1}{2}\nu)\pi^{\frac{1}{2}r}]^{-1}\nu^{-\frac{1}{2}r}[1 + \nu^{-1}||t||^2]^{-\frac{1}{2}(\nu + r)}. \tag{5.34}$$

In accordance with (3.8), assume that β, $\sigma^2 \mid H$ is distributed as the limiting distribution of β, $\sigma^2 \mid \bar{H}$, from (5.29) given $C_H \beta = \eta_H$.

Sufficient statistics: b, s^2, N, ν (5.2)–(5.3).

Posterior:

$$\beta, \sigma^{-2} \mid \bar{H} \sim (r\text{-variate Normal})\text{-gamma} \quad (b_1, s_1^2, N_1, \nu_1), \qquad \text{where} \tag{5.35}$$

$$N_1 = N_0 + N, \qquad \nu_1 = \nu_0 + \nu + \text{rank}\,(N_0), \tag{5.36}$$

$$b_1 = N_1^{-1}(N_0 b_0 + Nb), \tag{5.37}$$

$$\begin{aligned} \nu_1 s_1^2 = SSW_1 &= [\nu_0 s_0^2 + N_0((b_0)) + \nu s^2 + N((b))] - N_1((b_1)) \\ &= \nu_0 s_0^2 + \nu s^2 + [N_0 N_1^{-1} N]((b_0 - b)). \end{aligned} \tag{5.38}$$

In the notation of (5.14)–(5.21), for $k = 0$, 1, or absent,

$$s_k^{-1}[(\tilde{N}_k)_{11\cdot 2}]^{\frac{1}{2}}[\eta - (\tilde{b}_k)_1] \mid \bar{H}(\text{and } D \text{ iff } k = 0) \sim t^{(q)}_{\nu_k}, \qquad \text{and} \tag{5.39}$$

$$\begin{aligned} \zeta, \sigma^{-2} \mid H(\text{and } D \text{ iff } k \neq 0) &\sim ((r-q)\text{-variate Normal})\text{-gamma} \\ &(B_H^{\tilde{N}_k}\tilde{b}_k, \;\; SST_k/(\nu_k + q), \;\; (\tilde{N}_k)_{22}, \;\; \nu_k + q), \qquad \text{where} \end{aligned} \tag{5.40}$$

$$SSW_k = \nu_k s_k^2, \qquad SST_k = SSB_k + SSW_k. \tag{5.41}$$

Weighted likelihood ratio: $L_D(H) = P_1'(\eta_H)/P_0'(\eta_H)$, where,

$$\begin{aligned} P_k'(\eta_H) &= |s_k^{-2}(\tilde{N}_k)_{11\cdot 2}|^{\frac{1}{2}} h^{(q)}_{\nu_k}\{[s_k^{-2}(\tilde{N}_k)_{11\cdot 2}]^{\frac{1}{2}}[(\tilde{b}_k)_1 - \eta_H]\} \\ &= [\tfrac{1}{2}q\pi^{\frac{1}{2}q}/\Gamma(\tfrac{1}{2}q + 1)]^{-1} |(\tilde{N}_k)_{11\cdot 2}|^{\frac{1}{2}}(SSW_k/\nu_k)^{-1}(SSB_k)^{1 - \frac{1}{2}q} \\ &\quad \cdot q^{-1} f_{q,\nu_k}[(SSB_k/q)/(SSW_k/\nu_k)] \\ &= \Gamma[\tfrac{1}{2}(\nu_k + q)][\Gamma(\tfrac{1}{2}\nu_k)\pi^{\frac{1}{2}q}]^{-1} |C_H N_k^{-1} C_H'|^{-\frac{1}{2}}(SSW_k)^{-\frac{1}{2}q} \\ &\quad \cdot [1 + SSB_k/SSW_k]^{-\frac{1}{2}(\nu_k + q)}, \end{aligned} \tag{5.42}$$

where $f_{q,\nu}$ denotes the F density on q and ν degrees of freedom (density of $F_{q,\nu} = \|\mathbf{t}_\nu^{(q)}\|^2/q$).

When stable estimation applies under $\bar{H}$, approximate constancy of the prior density of $\boldsymbol{\beta}$, $\log\sigma \mid \bar{H}$ relative to the likelihood function ($P'(\boldsymbol{\beta}, \sigma^2 \mid \bar{H}) \mathrel{\dot{\propto}} \sigma^{-2}$), then the above posterior distributions ($k = 1$) apply approximately with k absent, the posterior versions of (5.33) and (5.31) becoming, with $\mathbf{N}_0 = \mathbf{0}$ and $\nu_0 = 0$ in (5.36),

$$(5.43) \qquad s^{-1}\mathbf{N}^{\frac{1}{2}}(\boldsymbol{\beta}-\mathbf{b}) \mid \bar{H}, \mathbf{D} \mathrel{\dot\sim} s^{-1}\mathbf{N}^{\frac{1}{2}}(\boldsymbol{\beta}-\mathbf{b}) \mid \sigma^2, \boldsymbol{\beta} \sim \mathbf{t}_\nu^{(r)},$$

$$(5.44) \qquad \nu s^2\sigma^{-2} \mid \bar{H}, \mathbf{D} \mathrel{\dot\sim} \nu s^2 \sigma^{-2} \mid \sigma^2, \boldsymbol{\beta} \sim \chi_\nu^2.$$

Hence, again assuming for $\mathbf{C}_H{}^* = \mathbf{C}_{H^\perp}$, $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are approximately prior independent under $\bar{H}$,

$$(5.45) \quad L_D(H) \doteq s^{-1}[\text{usual (multiv. } t\text{) density of } s^{-1}\mathbf{C}_H\mathbf{b} \mid \sigma^2, \boldsymbol{\beta}, H]/P'(\boldsymbol{\eta} \mid \bar{H})_{\boldsymbol{\eta}=\mathbf{C}_H\mathbf{b}}$$
$$= [\text{usual } (F_{q,\nu}) \text{ density of } (SSB/q)/(SSW/\nu) \mid \sigma^2, \boldsymbol{\beta}, H]/$$
$$P'[(\mathbf{N}((\boldsymbol{\beta}-B_H{}^N\boldsymbol{\beta}))/q)/(SSW/\nu) \mid \bar{H}]_{(SSB/q)/(SSW/\nu)},$$

where the first numerator is equal to (5.42) without k, and the denominators are related by (5.28) with σ^2 replaced by $q(SSW/\nu)$.

The author has vague memories of verbal statements by Leonard J. Savage, at least as early as 1965, that ordinates of F densities are more important than tail areas.

5.1.3. *Prior independence of $\boldsymbol{\beta}$ and σ^2.* A direct multivariate generalization of equation (4.23) extends Raiffa and Schlaifer's (1961) (r-variate normal)-gamma conjugate family (5.29) to yield the following generalizations of equations (4.24)–(4.31):

$$(5.46) \quad \boldsymbol{\beta} \mid \sigma^2, \bar{H} \sim \mathbf{N}^{(r)}[(\sigma^{-2}\mathbf{N}_0+\mathbf{M}_0)^{-1}(\sigma^{-2}\mathbf{N}_0\mathbf{b}_0+\mathbf{M}_0\mathbf{c}_0), (\sigma^{-2}\mathbf{N}_0+\mathbf{M}_0)^{-1}],$$

$$(5.47) \qquad P'(\boldsymbol{\beta} \mid \bar{H}) = C_0 \cdot (s_0{}^{-r}|\mathbf{N}_0|^{\frac{1}{2}}) h_{\nu_0}^{(r)}[s_0{}^{-1}\mathbf{N}_0{}^{\frac{1}{2}}(\boldsymbol{\beta}-\mathbf{b}_0)]$$
$$\cdot (2^{-r}\pi^{-r}|\mathbf{M}_0|)^{\frac{1}{2}} \exp[-\tfrac{1}{2}\mathbf{M}_0((\boldsymbol{\beta}-\mathbf{c}_0))],$$

where the normalizing constant here C_0 satisfies

$$(5.48) \qquad C_0{}^{-1} = h^{(r)}_{\nu_0,\infty;s_0N_0{}^{-1/2},M_0{}^{-1/2}}(\mathbf{b}_0-\mathbf{c}_0),$$

where $h^{(r)}_{\nu,\nu^*;U,U^*}(\mathbf{d})$ is the density of the r-variate Behrens–Fisher random vector,

$$(5.49) \qquad \mathbf{d}^{(r)}_{\nu,\nu^*;U,U^*} \sim \mathbf{U}\mathbf{t}_\nu{}^{(r)} - \mathbf{U}^*\mathbf{t}^{(r)}_{\nu^*}, \quad \text{independent } \mathbf{t}^{(r)}\text{'s}.$$

(Note that in one dimension, $d^{(1)}_{\nu,\nu^*;u,u^*} \sim (u^2+u^{*2})^{\frac{1}{2}} d_{\nu,\nu^*,\tan^{-1}(u^*/u)}$, where $d_{\nu,\nu^*,\omega}$ is the usual Behrens–Fisher random variable.) Dickey (1968) expressed $h^{(r)}_{\nu,\nu^*;U,U^*}$ in terms of a one-dimensional integral.

Using Lemma 3.1 of Dickey (1967c) on partitioned quadratic forms, integrate $\boldsymbol{\zeta}$ out of the density of $\boldsymbol{\beta} \mid \sigma^2$, $\bar{H}$, equation (5.46) ($k = 0$), and its posterior version ($k = 1$), to obtain the prior and posterior densities under $\bar{H}$, $P_k'(\boldsymbol{\eta} \mid \sigma^2)$ ($k = 0, 1$), multivariate normal with mean $\mathbf{C}_H(\sigma^{-2}\mathbf{N}_k+\mathbf{M}_0)^{-1}(\sigma^{-2}\mathbf{N}_k\mathbf{b}_k+\mathbf{M}_0\mathbf{c}_0)$ and variances $\mathbf{C}_H(\sigma^{-2}\mathbf{N}_k+\mathbf{M}_0)^{-1}\mathbf{C}_H'$. The marginal densities of σ^{-2} under $\bar{H}$ satisfy

$$(5.50)\qquad P_k'(\sigma^{-2}) \propto |\sigma^{-2}\mathbf{N}_k+\mathbf{M}_0|^{-\frac{1}{2}}(\sigma^{-2})^{\frac{1}{2}[\nu_k+\mathrm{rank}(N_k)]-1} \cdot \exp\{-\tfrac{1}{2}\sigma^{-2}[\nu_k s_k^2+(\mathbf{N}_k(\sigma^{-2}\mathbf{N}_k+\mathbf{M}_0)^{-1}\mathbf{M}_0((\mathbf{b}_k-\mathbf{c}_0))]\}.$$

Then $L_D(H) = P_1'(\boldsymbol{\eta}_H)/P_0'(\boldsymbol{\eta}_H)$, where

$$(5.51)\qquad P_k'(\boldsymbol{\eta}) = \textstyle\int_0^\infty P_k'(\boldsymbol{\eta}\,|\,\sigma^2)P_k'(\sigma^{-2})\,d\sigma^{-2}.$$

The one-dimensional integral of equation (5.51) is easily calculated numerically after simultaneously diagonalizing $\mathbf{N}_k$ and $\mathbf{M}_0$, as in Dickey (1968). Counting the normalizing constants of $P_k'(\sigma^{-2})$, four numerical integrations are needed for $L_D(H)$. If $\mathbf{N}_0 = \mathbf{0}$ (prior independence of $\boldsymbol{\beta}$ and σ^2), the denominator $P_0(\boldsymbol{\eta}_H)$ is just the multivariate normal density with mean $\mathbf{C}_H\mathbf{c}_0$ and variance $\mathbf{C}_H\mathbf{M}_0^{-1}\mathbf{C}_H'$.

5.2. *The multivariate Behrens–Fisher problem.*
Model:

$$(5.52)\qquad \mathbf{D}^{(a)} = \mathbf{y}^{(a)} = (y_1^{(a)},\cdots,y_{n^{(a)}}^{(a)}),\qquad a = 1,\cdots,A,$$

independently each

$$(5.53)\qquad \mathbf{y}^{(a)}\,|\,\boldsymbol{\beta}^{(a)},\sigma^{(a)2} \sim \mathbf{N}^{(n^{(a)})}(\mathbf{X}^{(a)}\boldsymbol{\beta}^{(a)},\sigma^{(a)2}\mathbf{W}^{(a)-1}),$$

as in (5.1). Every $\sigma^{(a)2}$ and $\boldsymbol{\beta}^{(a)}\in R^{r^{(a)}}$ are unknown.
Hypothesis:

$$(5.54)\qquad H:\boldsymbol{\eta} = \boldsymbol{\eta}_H,\qquad \bar{H}:\boldsymbol{\eta} \neq \boldsymbol{\eta}_H,\qquad \text{where}$$

$$(5.55)\qquad \boldsymbol{\eta} = \textstyle\sum_{a=1}^{A}\mathbf{C}_H^{(a)}\boldsymbol{\beta}^{(a)},\in R^q.$$

Prior: $0 < P(H) < 1$; and independently with (a) given $\bar{H}$, as in (5.29), each

$$(5.56)\qquad \boldsymbol{\beta}^{(a)},\sigma^{(a)-2}\,|\,\bar{H} \sim (r^{(a)}\text{-variate Normal)-gamma}\quad(\mathbf{b}_0^{(a)},s_0^{(a)2},\mathbf{N}_0^{(a)},\nu_0^{(a)}).$$

Assume, again, that (3.8) holds. After obvious modification to the notation (5.14)–(5.16), by (5.39),

$$(5.57)\qquad \boldsymbol{\eta}\,|\,\bar{H} \sim \textstyle\sum_a s_0^{(a)}[(\tilde{\mathbf{N}}_0^{(a)})_{11\cdot 2}]^{-\frac{1}{2}}\mathbf{t}^{(q)}_{\nu_0^{(a)}}+\sum_a(\tilde{\mathbf{b}}_0^{(a)})_1,$$

in which the first sum is a Behrens–Fisher random vector.
Sufficient statistics: $\mathbf{b}^{(a)}$, $s^{(a)2}$, $\mathbf{N}^{(a)}$, $\nu^{(a)}$, for each (a), as in (5.2), (5.3).
Posterior: the analogues of (5.56) and (5.57) are, with $\mathbf{D} = (\mathbf{D}^{(1)},\cdots,\mathbf{D}^{(A)})$,

$$(5.58)\qquad \boldsymbol{\beta}^{(a)},\sigma^{(a)-2}\,|\,\bar{H},\mathbf{D}^{(a)} \sim (r^{(a)}\text{-variate Normal)-gamma}\quad(\mathbf{b}_1^{(a)},s_1^{(a)2},\mathbf{N}_1^{(a)},\nu_1^{(a)}),$$

(posterior parameters for each (a) satisfying (5.36)–(5.38))

$$(5.59)\qquad \boldsymbol{\eta}\,|\,\bar{H},\mathbf{D} \sim \textstyle\sum_a s_1^{(a)}[(\tilde{\mathbf{N}}_1^{(a)})_{11\cdot 2}]^{-\frac{1}{2}}\mathbf{t}^{(q)}_{\nu_1^{(a)}}+\sum_a(\tilde{\mathbf{b}}_1^{(a)})_1.$$

Likelihood ratio:

$$(5.60)\quad L_D(H) = h^{(q)}_{(\nu_1{}^{(a)})_{\text{all } a};\,([(\tilde{N}_1{}^{(a)})_{11\cdot 2}]^{-1/2})_{\text{all } a}}\left[\textstyle\sum_a (\tilde{\mathbf{b}}_1{}^{(a)})_1 - \boldsymbol{\eta}_H\right] \Big/ h^{(q)}_{(\nu_0{}^{(a)})_{\text{all } a};\,([(\tilde{N}_0{}^{(a)})_{11\cdot 2}]^{-1/2})_{\text{all } a}}\left[\textstyle\sum_a (\tilde{\mathbf{b}}_0{}^{(a)})_1 - \boldsymbol{\eta}_H\right],$$

where $h^{(q)}_{(\nu^{(a)})_{\text{all } a};\,(U^{(a)})_{\text{all } a}}(\mathbf{d})$ denotes the density of the Behrens–Fisher random vector, as in (5.49) (the signs of the summands are irrelevant to the distribution),

$$(5.61)\qquad d^{(q)}_{(\nu^{(a)})_{\text{all } a};\,(U^{(a)})_{\text{all } a}} = \textstyle\sum_a \mathbf{U}^{(a)}\mathbf{t}^{(q)}_{\nu^{(a)}}.$$

Dickey (1968), Theorem 2, has expressed this q-variate Behrens–Fisher density as an $(A-1)$-dimensional integral. The stable-estimation version of (5.60) is

$$(5.62)\qquad L_D(H) \doteq h^{(q)}_{(\nu^{(a)})_{\text{all } a};\,([(\tilde{N}^{(a)})_{11\cdot 2}]^{-1/2})_{\text{all } a}}\left[\textstyle\sum_a \mathbf{C}_H{}^{(a)}\,\mathbf{b}^{(a)} - \boldsymbol{\eta}_H\right] / P'(\boldsymbol{\eta}_H \mid \bar{H})_{\hat{\boldsymbol{\eta}}},$$

where $\hat{\boldsymbol{\eta}} = \sum_a \mathbf{C}_H{}^{(a)}\,\mathbf{b}^{(a)}$.

Acknowledgment. I thank the Associate Editor Thomas S. Ferguson for his continual unselfish devotion to research.

REFERENCES

ANDERSON, T. W. (1958). *Introduction to Multivariate Analysis.* Wiley, New York.

BARNARD, G. A. (1964). Unpublished series of lectures on statistical inference, National Institutes of Health.

BARTLETT, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika* **44** 533–534.

BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

CORNFIELD, JEROME (1966). A Bayesian test of some classical hypotheses—with applications to sequential clinical trials. *J. Amer. Statist. Assoc.* **61** 577–594.

DE FINETTI, BRUNO (1937). La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincare* **7** 1–68 (English translation in *Studies in Subjective Probability*, (Kyburg and Smokler, eds.) Wiley, 1964).

DEGROOT, MORRIS H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

DICKEY, JAMES M. (1967a). Expansions of *t* densities and related complete integrals. *Ann. Math. Statist.* **38** 503–510.

DICKEY, JAMES M. (1967b). A Bayesian hypothesis-decision procedure. *Ann. Inst. Statist. Math.* **19** 367–369.

DICKEY, JAMES M. (1967c). Matricvariate generalizations of the multivariate *t* distribution and the inverted multivariate *t* distribution. *Ann. Math. Statist.* **38** 511–518.

DICKEY, JAMES M. (1968). Three multidimensional-integral identities with Bayesian applications. *Ann. Math. Statist.* **39** 1615–1627.

DICKEY, JAMES M. (1970). Scientific reporting and personal probabilities, Student's hypothesis. To be published.

DICKEY, JAMES M. and LIENTZ, B. P. (1968). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Statist.* **41** 214–226.

EDWARDS, WARD, HAROLD LINDMAN, and LEONARD J. SAVAGE (1963). Bayesian statistical inference for psychological research. *Psychological Review* **70** 193–242 (reprinted in *Readings in Mathematical Psychology*, ed. by R. Duncan Luce, R. Bush, and Eugene Galanter, Wiley, 1965 **2** 519–568).

GOOD, I. J. (1965). *The Estimation of Probabilities.* M.I.T. Press, Cambridge, Mass.

GOOD, I. J. (1966). How to estimate probabilities. *J. Inst. Math. Appl.* **2** 364–383.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Hafner, New York.

HILDRETH, CLIFFORD (1963). Bayesian statisticians and remote clients. *Econometrica* **38** 422–438.

JEFFREYS, HAROLD (1940). Note on the Behrens–Fisher formula. *Ann. Eugen.* **10** 48–51.

JEFFREYS, HAROLD (1948). *Theory of Probability* (3rd ed.). Clarendon Press, Oxford, 1961.

LINDLEY, D. V. (1957). A statistical paradox. *Biometrika* **44** 187–192. (See also Bartlett (1957).)

LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 453–468. Univ. of California Press.

LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint: Part-Two Inference*. Cambridge Univ. Press.

LINDLEY, D. V. (1968). The choice of variables in multiple regression. *J. Roy. Statist. Soc. Ser. B* **30** 31–66.

LUENBERGER, D. G. (1969). *Optimization By Vector Space Methods*. Wiley, New York.

PATIL, V. H. (1964a). The Behrens–Fisher problem and its Bayesian solution. *J. Indian Statist. Assoc.* **2** 21–31.

PATIL, V. H. (1964b). Difficulties involved in computing Behrens–Fisher densities, cumulative probabilities, and percentage points from first principles. *J. Indian. Statist. Assoc.* **2** and **3** 110–118.

PATIL, V. H. (1965). Approximation to the Behrens–Fisher distributions. *Biometrika* **52** 267–271.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Harvard Univ. Business School, Boston.

RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

SAVAGE, LEONARD J. (1959). *The Foundations of Statistical Inference*, Joint Statistics Seminar at the University of London. Methuen, London, 1962.

SAVAGE, LEONARD J. (1961). The subjective basis of statistical practice. Unpublished Univ. of Michigan Notes.

SAVAGE, LEONARD J. (1963). Private longhand notes from University of Michigan course on the foundations of statistics.

STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–285.

STONE, M. (1964). Comments on a posterior distribution of Geisser and Cornfield. *J. Roy. Statist. Soc. B* **26** 274–276.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.