May 1964

FIDUCIAL THEORY AND INVARIANT ESTIMATION

R. B. Hora and R. J. Buehler

Technical Report No. 42

University of Minnesota
Minneapolis, Minnesota

1. Introduction. Estimation of location and scale parameters has been studied by Fisher (1934) and Pitman (1939). Fisher gave the fiducial distributions; Pitman considered the problem in greater detail, showing how estimators having various desirable properties could be described in terms of the ("Fisher-Pitman") fiducial distributions. Fraser's (1961a,b) approach to fiducial theory, using transformation groups, is useful in providing a precise mathematical framework which is consistent with Fisher's in the case of location and scale parameters, and apparently in most other cases as well.

The purpose of the present paper is to show how certain of Pitman's results can be generalized by using Fraser's theory. Our final results (Section 5) on "best" invariant estimators can actually be stated without any reference to fiducial distributions. Thus fiducial theory can be regarded as a convenient (but not actually essential) tool for obtaining desirable estimators.

In Section 2, assumptions on the class of distributions are spelled out in detail. These are similar to Fraser's, but slightly more elaborate in order to bring the ancillary statistic and the conditionally sufficient statistic explicitly into the notation. Theorem 2.1 establishes the identity $E_f H = E_a H$ where H is an invariant function of the observations and parameters, E_f denotes expectation with respect to the fiducial distribution and E_a denotes conditional expectation given any value of the ancillary statistic. The theorem is not restricted to location and scale parameter cases.

Section 3 gives four examples of location and scale parameter families whose generality increases from the case of one location parameter to that of two location and two scale parameters. The latter includes the Behrens-Fisher problem as a special case. It is to be emphasized that an assumption of the existence of sufficient statistics is not needed at any point. It may also be noted that the variates need be neither independent nor identically distributed. The group structure is the essential ingredient which makes these other assumptions unnecessary.

In Section 4 a definition is given of an "invariantly estimable function" $\psi(\omega)$ where ω is a point in the parameter space. The concept (but not the term) can be found in Lehmann (1959), p 243. It is shown that when ψ is invariantly estimable there exists a group of transformations of $\{\psi\}$ which is homomorphic with the group of transformations of $\Omega = \{\omega\}$. This result is used to define invariant functions of the observations and parameters in terms of invariant estimators of ψ.

In Section 5 the estimation problem is formulated in terms of decision theory and in a number of cases loss functions are found which lead to minimum mean square error invariant estimators.

2. <u>Generalization of Pitman's expectation identity</u>. In this section we give assumptions which are essentially the same as those of Fraser (1961a,b) and generalize an identity of Pitman (1939).

2.1 <u>Assumptions</u>.

<u>Assumption 1.</u> $(\mathcal{X}, B_X)$ is a measurable space, and $\mathcal{G} = \{g\}$ is a group of one-to-one measurable transformations of $\mathcal{X}$ onto itself.

<u>Assumption 2.</u> $(\mathcal{G}, B_G)$ is a measurable space on which there exists a left Haar measure μ having the invariance property

$$\mu(gG) = \mu(G) \qquad \text{all } g\varepsilon\mathcal{G}, \ G\varepsilon B_G . \tag{2.1}$$

<u>Assumption 3.</u> For each ω in a parameter space Ω, P^{ω} is a probability measure on $(\mathcal{X}, B_X)$ such that for each $g\varepsilon\mathcal{G}$ and each $\omega\varepsilon\Omega$ there exists a unique $\omega_g\varepsilon\Omega$ for which

$$P^{\omega}(X) = P^{\omega_g}(gX) \qquad \text{all } X\varepsilon B_X . \tag{2.2}$$

Let g^* be the one-to-one function on Ω to Ω defined by

(2.3) $$g^*\omega = \omega_g .$$

It is easily seen that $\mathcal{G}^* = \{g^*\}$ is a group of transformations of Ω which is isomorphic to $\mathcal{G}$ when $g_1^* \cdot g_2^*$ is defined by $(g_1^* \cdot g_2^*)\omega = \omega_{(g_1 \cdot g_2)}$.

<u>Assumption 4.</u> $\mathcal{G}^*$ is exactly transitive on Ω; that is, for any $\omega_1, \omega_2 \varepsilon \Omega$ there is a unique $g^* \varepsilon \mathcal{G}^*$ such that $g^*\omega_1 = \omega_2$.

For any $x \varepsilon \mathcal{X}$ the set of points gx generated by all $g \varepsilon \mathcal{G}$ is known as the <u>orbit</u> of x. Let a be an index which takes a fixed value on each orbit and different values on different orbits, and let $\mathcal{A}$ be the set of all labels a. A σ-field B_A of subsets A of $\mathcal{A}$ is defined as the set of all A such that

(2.4) $$\{x \mid a(x) \varepsilon A\} \varepsilon B_X .$$

where a(x) denotes the a-value determined by x.

Any point x can be determined by its a-value (or orbit) together with a second coordinate t describing the position on the orbit.

<u>Assumption 5.</u> The t-values on each orbit take all values in a set $\mathcal{J}$, so that there is a one-to-one correspondence between $\mathcal{X}$ and the direct product space $\mathcal{J} \times \mathcal{A}$ which we denote by

(2.5) $$x \leftrightarrow (t,a) .$$

Moreover it is assumed that

(2.6) $$gx \leftrightarrow (t_g, a)$$

where t_g depends only on t and g and not on a.

It is of course automatic that a-values are invariant under the transformation $\mathcal{G}$. Also note that in view of (2.5), we may say that t is conditionally sufficient for ω, given a.

For any $g \varepsilon \mathcal{G}$ we may define a transformation g' of $\mathcal{J}$ onto itself by

(2.7) $$g't = t_g .$$

Defining $g_1' \cdot g_2'$ as $(g_1 \cdot g_2)'$ gives group $\mathcal{G}' = \{g'\}$ isomorphic to $\mathcal{G}$.

Assumption 6. $\mathcal{G}'$ is exactly transitive on $\mathcal{J}$.

Isomorphisms between $\mathcal{G}$, $\mathcal{G}^*$ and $\mathcal{G}'$ have previously been indicated. By virtue of assumptions 4 and 6 we can now indicate isomorphisms between these spaces and $\mathcal{J}$ and Ω. Let $t_o \varepsilon \mathcal{J}$ and $\omega_o \varepsilon \Omega$ be any fixed reference points. For any $t \varepsilon \mathcal{J}$ we have, by exact transitivity of $\mathcal{G}'$ on $\mathcal{J}$ a unique $g_t' \varepsilon \mathcal{G}'$ such that $t = g_t' t_o$. For any $\omega \varepsilon \Omega$ we have, by exact transivity of $\mathcal{G}^*$ on Ω a unique $g_\omega^* \varepsilon \Omega$ such that $\omega = g_\omega^* \omega_o$. Thus by using the fixed reference points t_o and ω_o, one-to-one correspondences (and incidentally isomorphisms) are established between $\mathcal{J}$ and $\mathcal{G}'$ and between Ω and $\mathcal{G}^*$. The correspondences may be summarized by

(2.8) $$\mathcal{J} \overset{t_o}{\longleftrightarrow} \mathcal{G}' \longleftrightarrow \mathcal{G} \longleftrightarrow \mathcal{G}^* \overset{\omega_o}{\longleftrightarrow} \Omega$$

The notation can now be simplified by defining

(2.9) $$t \equiv g_t' \,\varepsilon\, \mathcal{G}' \overset{t_o}{\longleftrightarrow} t \,\varepsilon\, \mathcal{J}$$

(2.10) $$\omega^* \equiv g_\omega^* \,\varepsilon\, \mathcal{G}^* \overset{\omega_o}{\longleftrightarrow} \omega \,\varepsilon\, \Omega$$

(2.11) $$\omega \,\varepsilon\, \mathcal{G}' \overset{\omega_o}{\longleftrightarrow} \omega \,\varepsilon\, \Omega$$

The correspondences further define measurable spaces $(\mathcal{G}', B_{G'})$, $(\mathcal{G}^*, B_{G^*})$, $(\mathcal{J}, B_T)$ and (Ω, B_Ω) in terms of $(\mathcal{G}, B_G)$.

Assumption 7. Let $B_{T\times A}$ be the minimal σ-field containing cartesian products of all sets in the σ-fields B_T and B_A. We assume the "bimeasurability" conditions

(2.12) $$\{(t,a) \mid t,a) \leftrightarrow x,\ x \varepsilon X\} \ \varepsilon B_{T\times A} \qquad \text{all } X \varepsilon B_X, \text{ and}$$
$$\{x \mid x \leftrightarrow (t,a), (t,a) \varepsilon S, S \varepsilon B_{T\times A}\} \ \varepsilon B_X \qquad \text{all } S \varepsilon B_{T\times A}$$

For any X corresponding to a set $T\times A$ in the product space we may now write

(2.13) $$P^\omega(T \times A) = P^\omega(X)$$

with little danger of confusion.

A marginal probability measure on $(\mathcal{A}, B_A)$ is defined by

(2.14) $$P_2^\omega(A) = P^\omega(\mathcal{J} \times A)$$

From (2.14), (2.2), (2.3), (2.6) and (2.7), $P_2^{\omega}(A) = P^{\omega}(\mathcal{J}\times A) = P^{g^*\omega}(g(\mathcal{J}\times A))$ $= P^{g^*\omega}(g'\mathcal{J}\times A) = P^{g^*\omega}(\mathcal{J}\times A) = P_2^{g^*\omega}(A)$ for all $g^*\varepsilon\mathcal{G}^*$, so that

(2.15) $$P_2^{\omega}(A) = P_2(A) \qquad \text{all } \omega\varepsilon\Omega,$$

that is, the measure is the same for all ω, as indeed was to be expected from the definition of the ancillary.

For any fixed $T\varepsilon B_T$ it is seen that as measures on the space $(\mathcal{A}, B_A)$, $P^{\omega} << P_2$, so that by the Radon-Nikodym theorem there exists a B_A-measurable function $P_1^{\omega}(T|a)$ such that

(2.16) $$P^{\omega}(T\times A) = \int_A P_1^{\omega}(T|a)dP_2(a).$$

When the space $\mathcal{J}$ is Euclidean, the measure $P(\cdot|a)$ actually exists for all a (see for example Lehmann (1959), p 44).

<u>Assumption 8.</u> For any a, ω, $P_1^{\omega} << \mu$ so that there exists a B_T-measurable function $p_1(t|a,\omega)$ such that

(2.17) $$P_1(T|a) = \int_T p_1(t|a,\omega)\, d\mu(t).$$

We may call $p_1(t|a,\omega)$ the conditional density of t given a with respect to the Haar measure μ.

2.2 <u>The expectation identity</u>.

<u>Lemma 2.1</u>. The conditional density $p_1(t|a,\omega)$ can be expressed as $p_1(\omega^{-1}t|a)$ where ω^{-1} denotes the inverse of the element g'_{ω} in $\mathcal{G}'$.

<u>Proof</u>: By (2.16), we have that for all $A\varepsilon B_A$,

$$P^{\omega}(T\times A) = \int_A P_1^{\omega}(T|a)dP_2(a)$$

and similarly for any corresponding g' and g*,

$$P^{g^*\omega}(g'T\times A) = \int_A P_1^{g^*\omega}(g'T|a)dP_2(a)\,.$$

Since it can be easily seen that $P^{\omega}(T\times A) = P^{g^*\omega}(g'T\times A)$, we have that for all $A\varepsilon B_A$,

$$\int_A P_1^{\omega}(T|a)dP_2(a) = \int_A P_1^{g^*\omega}(g'T|a)dP_2(a).$$

Hence, $P_1^{\omega}(T|a) = P_1^{g^*\omega}(g'T|a)$, a.e. P_2. Also by (2.17),

$$P_1^{\omega}(T|a) = \int_{t\varepsilon T} p_1(t|a;\omega)d\mu(t)$$

$$P_1^{g^*\omega}(g'T|a) = \int_{t\varepsilon T} p_1(g't|a;g^*\omega)d\mu(g't)$$

$$= \int_{t\varepsilon T} p_1(g't|a,g^*\omega)\, d\mu(t), \quad \text{by (2.1)}$$

Hence $p_1(t|a,\omega) = p_1(g't|a,g^*\omega)$ a.e. μ. Put $g^* = (\omega^*)^{-1}$ where ω^* is defined by (2.10). Then $g^*\omega$ equals the fixed element ω_0 and $g't$ is expressible as $\omega^{-1}t$ where $\omega \varepsilon \mathcal{G}'$ is defined by (2.11).

Lemma 2.2. If for all x, g, ω,

(2.18) $$H(x,\omega) = H(gx,\ g^*\omega)$$

(where g^* corresponds to g), then $H(x,\omega)$ can be expressed in the form $H'(\omega^{-1}t,a)$.

Proof: From the correspondence $x \leftrightarrow (t,a)$ we may write $H(x,\omega) = H''(t,a,\omega)$. From (2.18), $H''(t,a,\omega) = H''(g't,a,g^*\omega)$. If we again put $g^* = (\omega^*)^{-1}$ so that $g' = \omega^{-1}$, then $H(x,\omega)$ is seen to have the required form.

In order to define the fiducial distribution in the sense of Fraser we must introduce the right Haar measure ν. In view of the isomorphism between spaces Ω, $\mathcal{G}$ and $\mathcal{T}$ we may define ν by

(2.19) $$\nu(T) = \mu(T^{-1}) \qquad T\varepsilon B_T .$$

We shall require also the well known formula

(2.20) $$\mu(Tg') = \Delta(g')\mu(T),$$

where Δ is the "modular function." Following Fraser, we now define the fiducial density of ω given x with respect to measure ν to be

(2.21) $$p_f(\omega|x) = p_1(t|a,\omega)\Delta(t) = p_1(\omega^{-1}t|a)\Delta(t).$$

As is fairly well known, this fiducial distribution is equivalent to a posterior distribution when the prior measure* is given by $\nu(\omega)$. This is

*The examples we consider are typical in that $\nu(\omega)$ is not a probability measure, but an unbounded measure. Wallace (1959) has called such prior measures "quasi-densities." When the group $\mathcal{G}$ is compact, the Haar measure is bounded and the fiducial distribution is a posterior distribution for a true prior density. Examples of this kind arise with rotation groups as indicated in Section 3.3 below.

evident because the likelihood can be factored into the marginal distribution of a times the conditional distribution $p_1(t|a,\omega)$. Since the former does not depend on ω, the posterior distribution is proportional to (2.21), and hence equal to it.

Theorem 2.1. Let E_f denote expectation with respect to the fiducial density (2.21) and E_a denote expectation with respect to the conditional density $p_1(t|a,\omega)$ given the ancillary. If $H(x,\omega)$ satisfies (2.18) and if Assumptions 1 through 8 are satisfied then

$$E_f H(x,\omega) = E_a H(x,\omega). \tag{2.22}$$

Proof: Using Lemmas 2.1, 2.2, and equation (2.1) and twice changing variable of integration gives

$$\begin{aligned} E_a H(x,\omega) &= \int_{t\varepsilon\mathcal{J}} H'(\omega^{-1}t,a)p_1(\omega^{-1}t|a)d\mu(t) \\ &= \int_{t\varepsilon\mathcal{J}} H'(\omega^{-1}t,a)p_1(\omega^{-1}t|a)d\mu(\omega^{-1}t) \\ &= \int_{s\varepsilon\omega^{-1}\mathcal{J}=\mathcal{J}} H'(s,a)p_1(s|a)d\mu(s) \qquad (s=\omega^{-1}t) \\ &= \int_{\omega\varepsilon\Omega} H'(\omega^{-1}t,a)p_1(\omega^{-1}t|a)d\mu(\omega^{-1}t). \end{aligned}$$

But by (2.20) and (2.19) we have

$$d\mu(\omega^{-1}t) = \Delta(t)d\mu(\omega^{-1}) = \Delta(t)d\nu(\omega),$$

so that

$$\begin{aligned} E_a H(x,\omega) &= \int_{\omega\varepsilon\Omega} H'(\omega^{-1}t,a)p_1(\omega^{-1}t|a)\,\Delta(t)d\nu(\omega) \\ &= E_f H(x,\omega). \end{aligned}$$

2.3 A counterexample. We give an example to show that the expectation identity (2.22) can hold for a function H which does not satisfy the invariance condition (2.18). Take $p(x;\theta) = 1/4$ for $\theta < x < \theta+4$ and take

$$H(x,\theta) = \begin{cases} -(x+\theta) & \theta < x \leq \theta+1 \quad \text{or} \quad \theta+3 < x \leq \theta+4 \\ +(x+\theta) & \theta+1 < x \leq \theta+3 \\ 0 \text{ otherwise} & \end{cases}$$

Here there is no ancillary and we may put E for E_a .

$$EH = \frac{1}{4}\left\{\int_{\theta}^{\theta+1} + \int_{\theta+1}^{\theta+3} + \int_{\theta+3}^{\theta+4}\right\} H(x,\theta)dx$$

$$= \frac{1}{4}\left\{(-2\theta - \tfrac{1}{2}) + (4\theta + 4) + (-2\theta - \tfrac{7}{2})\right\} = 0,$$

$$E_f H = \frac{1}{4}\left\{\int_{x-1}^{x} + \int_{x-3}^{x-1} + \int_{x-4}^{x-3}\right\} H(x,\theta)d\theta$$

$$= \frac{1}{4}\left\{(-2x + \tfrac{1}{2}) + (4x - 4) + (-2x + \tfrac{7}{2})\right\} = 0 .$$

3. Examples of invariant specifications. In this section we give examples of specifications which satisfy the assumptions of Section 2. Four location and scale parameter families will be given in some detail (Section 3.1) and some other cases will be mentioned briefly (Section 3.2).

3.1 Details of four location and scale parameter specifications. Location and scale parameters will be denoted by θ and σ respectively. Quantities x, y, θ, α range over $(-\infty,\infty)$ while σ, β range over $(0,\infty)$. Briefly we may describe the four cases thus: Example 3.1: θ; Example 3.2: (θ,σ); Example 3.3: $(\theta_1,\theta_2,\sigma)$; Example 3.4: $(\theta_1,\theta_2,\sigma_1,\sigma_2)$. It may be noted that in none of the examples are the variates assumed to be independent and identically distributed (i.i.d.). Estimation of $\theta_1-\theta_2$ in Example 3.4 is the Behrens-Fisher problem generalized to the non-normal, non-i.i.d. case.

The density for Example 3.4 has the form

$$\sigma_1^{-n}\,\sigma_2^{-m} f((x_1-\theta_1)/\sigma_1,\ldots,(x_n-\theta_1)/\sigma_1,(y_1-\theta_2)/\sigma_2,\ldots,(y_m-\theta_2)/\sigma_2) \tag{3.1}$$

Example 3.3 is obtained by putting $\sigma_1=\sigma_2=\sigma$; Example 3.2 is obtained by deleting the y variates and writing $(\theta_1,\sigma_1) = (\theta,\sigma)$; Example 3.1 is obtained from 3.2 by putting $\sigma=1$. The space $(\mathcal{X},B_{\mathcal{X}})$ is (R_n,B_n) in Examples 3.1 and 3.2 and (R_{n+m},B_{n+m}) in Examples 3.3 and 3.4, where R_k is k-dimensional euclidean space and B_k is the class of Borel sets. Table 3.1 gives the transformations g and g*. To save writing, g is defined only on x_1 and y_1 with the understanding that the definition on other x's and y's is analogous. Table 3.2 gives a suitable (not unique) choice

Table 3.1

Example	Parameters	gx	$g^*\omega$
3.1	θ	$x_1 + \alpha$	$\theta + \alpha$
3.2	θ, σ	$\alpha + \beta x_1$	$\alpha + \beta\theta$, $\beta\sigma$
3.3	θ_1, θ_2, σ	$\alpha_1 + \beta x_1$, $\alpha_2 + \beta y_1$	$\alpha_1 + \beta\theta_1$, $\alpha_2+\beta\theta_2$, $\beta\sigma$
3.4	θ_1, σ_1, θ_2, σ_2	$\alpha_1 + \beta_1 x_1$, $\alpha_2 + \beta_2 y_1$	$\alpha_1 + \beta_1\theta_1$, $\beta_1\sigma_1$, $\alpha_2 + \beta_2\theta_2$, $\beta_2\sigma_2$

Table 3.2

Example	t #	$\Delta(t)$	$d\nu(\omega)$
3.1	t_1	1	$d\theta$
3.2	t_1, t_2	t_2^{-1}	$\sigma^{-1}d\theta d\sigma$
3.3	t_1, t_2, t_3	t_2^{-1}	$\sigma^{-1}d\theta_1 d\theta_2 d\sigma$
3.4	t_1, t_2, t_3, t_4	$t_2^{-1}t_4^{-1}$	$\sigma_1^{-1}\sigma_2^{-1}d\theta_1 d\theta_2 d\sigma_1 d\sigma_2$

$t_1 = x_1$, $t_2 = |x_1 - x_2|$, $t_3 = y_1$, $t_4 = |y_1 - y_2|$.

of the statistic t, the modular function Δ and the right Haar measure element $d\nu$. The left Haar measure element $d\mu$ is deducible from Δ and $d\nu$. Thus in Example 3.4, $d\mu(g) = \beta_1^{-2}\ \beta_2^{-2}d\alpha_1 d\alpha_2 d\beta_1 d\beta_2$.

The ancillary statistic is uniquely determined by the group $\mathcal{G}$ in the sense that the sub-σ-field of B_X associated with B_A (defined by (2.4)) is unique; but it may of course have different representations. The following choices are suitable. Example 3.1: $a = (a_1, \ldots, a_{n-1})$, $a_i = x_{i+1} - x_1$; Example 3.2: $a = (a_o, a_1, \ldots, a_{n-2})$, $a_i = (x_{i+2} - x_1)/(x_2 - x_1)$, $i = 1, \ldots, n-2$, $a_o = 0$ or 1 according as $x_1 - x_2 < 0$ or $\geqq 0$; Example 3.3: $a = (a_o, a_1, \ldots, a_{n-2}, b_1, b_2, \ldots, b_{m-2}, c)$ with the a's as in Example 3.2, $b_j = (y_{j+2} - y_1)/(y_2 - y_1)$, $j = 1, \ldots, m-2$, $c = (y_2 - y_1)/(x_2 - x_1)$; Example 3.4: $a = (a_o, \ldots, a_{n-2}, b_o, \ldots, b_{m-2})$ where $b_o = 0$ or 1 according as

$y_1 - y_2 < 0$ or $\geqq 0$ and the other quantities have the same definitions as in Example 3.3.

3.2 Verification of assumptions. We will consider Example 3.2 only, the other cases being similar. Verification of Assumptions 1 through 3 is straightforward. Exact transitivity of $\mathcal{G}^*$ on Ω (Assumption 4) is easily shown by observing that (θ_1, σ_1) is carried into (θ_2, σ_2) by the unique element $g^*_{\alpha\beta}$ where $\alpha = \theta_2 - \theta_1 \; \sigma_2/\sigma_1$ and $\beta = \sigma_2/\sigma_1$. Assumptions 5 and 6 can be verified using the explicit forms for a and t in the preceding section. In Assumption 7 all sets are Borel sets. To verify $P_1^\omega << \mu$ (Assumption 8) we note that $d\mu = dt_1 dt_2 / t_2^2$ so that $\mu(T) = 0$ implies* $L_2(T) = 0$. Thus for any A, $L_n(T \times A) = 0$, whence from the assumed density analogous to (3.1), $P^\omega(T \times A) = 0$, and from (2.16), $P_1^\omega(T|a) = 0$.

3.3 Other examples of invariant specifications. We briefly mention some other cases which will not be considered in the later sections. We hope to consider the rotational families in a later paper.

Clearly the location and scale parameter discussion could be extended to cases involving more than two σ's and more than two θ's in a straightforward manner.

For any bivariate distribution of variates (x,y) which is not symmetrical about the origin, one may consider the parametric family generated by rotation through an angle α about the origin. It is possible to obtain the fiducial distribution of α given n of bivariate observations. Indeed the fiducial distribution equals the posterior distribution given a uniform prior over the interval $(0, 2\pi)$. The special case (x,y) independent normal with $Ex = R$, $Ey = 0$, Var x = Var y = 1, n = 1, was considered by Fisher (1956), p 135. More general cases have been considered by Hora (1964), particularly with regard to the problem of obtaining "best" estimators of α.

* L_k denotes k-dimensional Lebesgue measure.

4. Invariant functions on $\mathcal{X}$ and on $\mathcal{X}\times\Omega$. In this section we give definitions of invariant estimators of the parameter point ω and of certain functions $\psi(\omega)$. The invariant estimators are used to define invariant functions $H(x,\omega)$ to which the results of Section 2 can be applied.

4.1 Estimation of ω. Let $\hat{\omega}(x)$ be a mapping of $\mathcal{X}$ onto Ω. We will say that $\hat{\omega}$ is an invariant estimator of ω if

(4.1) $$\hat{\omega}(gx) = g^*\hat{\omega}(x).$$

This has been called the "principle of cogredience" by Lehmann (1950), Chapter 1, p 17.

Lemma 4.1. If $\hat{\omega}(x)$ satisfies (4.1) and

(4.2) $$H(x,\omega) = \omega^{*-1}\hat{\omega}(x)$$

where $\omega^*\varepsilon\mathcal{G}^*$ is defined by (2.10), then H satisfies (2.18).

Proof: Let ω_1, ω_2 be arbitrary elements of Ω and let g_1, g_2 and g_1^*, g_2^* be the corresponding elements of $\mathcal{G}$ and $\mathcal{G}^*$. Then

$$\begin{aligned} H(g_2x, g_2^*\omega_1) &= [(\omega_2\cdot\omega_1)^*]^{-1}\hat{\omega}(g_2x) \\ &= (\omega_1^*)^{-1}(\omega_2^*)^{-1}\cdot\omega_2^*\hat{\omega}(x) \\ &= (\omega_1^*)^{-1}\hat{\omega}(x) \\ &= H(x,\omega_1). \end{aligned}$$

4.2 Invariantly estimable functions $\psi(\omega)$. Frequently one does not wish to estimate the parameter point ω but only some function of it, say $\psi(\omega)$ with range Ξ. An equivalence relation "$\sim$" among elements of Ω is defined by

(4.3) $$\omega_1 \sim \omega_2 \quad \text{means} \quad \psi(\omega_1) = \psi(\omega_2)$$

We will say that ψ is an invariantly estimable function (compare Lehmann (1959), p 243) if

(4.4) $$\omega_1 \sim \omega_2 \quad \text{implies } g^*\omega_1 \sim g^*\omega_2 \quad \text{all } g^*\varepsilon\mathcal{G}^*.$$

If ψ is a one-to-one function of Ω onto Ξ, then it satisfies (4.4) trivially. If not, a necessary and sufficient condition for ψ to be invariantly estimable is that $\psi(g^*\omega)$ have the form $\varphi(\psi(\omega))$. To illustrate, in Example 3.2, if

$\psi(\omega) = \psi(\theta,\sigma) = \theta$, then $\psi(g^*\omega) = \alpha + \beta\theta$ which depends on ω only through $\psi(\omega) = \theta$, showing that ψ is invariantly estimable, as is also seen directly from (4.4). Similarly $\psi(\theta,\sigma) = \sigma$ is invariantly estimable, but θ/σ is not. Other examples which are easily verified are given in Table 4.1. Note that in Example 3.4 (generalized Behrens-Fisher case), the difference of means, $\theta_1 - \theta_2$, is not invariantly estimable. Fraser (1961b), Section 12, has noted that in the Behrens-Fisher problem, "a fiducial interval for $\mu_1 - \mu_2$ will not be an invariant interval with respect to transformations for the x's and for the y's. In fact, under separate linear transformations for the x's and for the y's the parameter $\mu_1 - \mu_2$ is not transformed but is 'pulled apart'." The present section is intended to formalize and generalize Fraser's observation.

Table 4.1

Example	Invariantly Estimable	Not Invariantly Estimable
3.1	θ^{2n+1}, $n = 0,1,2,\ldots$	θ^{2n}, $n = 1,2,\ldots$
3.2	θ, σ, $c_1\theta + c_2\sigma$	θ/σ, $\theta^3 + \sigma$
3.3	θ_1, θ_2, σ,	$\theta_1{}^3 + \theta_2{}^3$, $(\theta_1 - \theta_2)^2$
	$c_1\theta_1 + c_2\theta_2 + c_3\sigma$	$\theta_1\theta_2$
3.4	θ_1, θ_2, σ_1, σ_2	$\theta_1 \pm \theta_2$, $\sigma_1 \pm \sigma_2$
	$\sigma_1{}^r\sigma_2{}^s$	$(\theta_1 + \sigma_1)/(\theta_2 + \sigma_2)$

4.3 <u>Invariant functions of $\hat{\psi}$ and ω</u>. For any point $\psi \in \Psi$ let ω denote any point of Ω such that $\psi(\omega) = \psi$. Then for any $g^* \in \mathcal{G}^*$ a transformation $g^\dagger$ of Ψ onto Ψ is defined by

(4.5) $$g^\dagger \psi(\omega) = \psi(g^*\omega),$$

and the definition is unique when (4.4) holds. In general, different elements g_1^* and g_2^* may define the same transformation $g^\dagger$, e.g., in Example 3.2, if $\psi(\theta,\sigma) = \sigma$ then both $g_1^* = (\alpha_1,\beta)$ and $g_2^* = (\alpha_2,\beta)$ give $g^\dagger\sigma = \beta\sigma$. When g^* and $g^\dagger$ are in one-to-one correspondence, then there is an automatic isomorphism. We will now show that in any case a group operation can be defined on $\mathcal{G}^\dagger = \{g^\dagger\}$ such that the mapping of $\mathcal{G}^*$ onto $\mathcal{G}^\dagger$ is a homomorphism. The invariant estimation problem can then be directly related to the invariant decision problem described by Lehmann (1959) p 11, if we identify the decision space D with the parameter space Ψ.

An equivalence relation "$\approx$" on $\mathcal{G}^*$ is defined by

(4.6) $$g_1^* \approx g_2^*$$

when $g_1^*\omega \sim g_2^*\omega$ for all $\omega\varepsilon\Omega$. To lighten the notation in the following two lemmas we will write g in place of g^*.

<u>Lemma 4.2</u>. If $g_1 \approx g_2$ and $g_3 \approx g_4$ then $g_1g_3 \approx g_2g_4$.

<u>Proof</u>: Since $g_3 \approx g_4$, $\psi(g_3\omega) = \psi(g_4\omega)$; and using (4.4), $\psi(g_1g_3\omega) = \psi(g_1g_4\omega)$. Since $g_1 \approx g_2$, $\psi(g_1g_4\omega) = \psi(g_2g_4\omega)$. Thus $g_1g_3 \approx g_2g_4$.

<u>Lemma 4.3</u>. If $g_1 \approx g_2$ then $g_1^{-1} \approx g_2^{-1}$.

<u>Proof</u>: For any given g_1, g_2, ω, define $\omega' = g_2^{-1}\omega$ so that $\omega = g_2\omega'$. Since $g_1 \approx g_2$, $\psi(g_1\omega') = \psi(g_2\omega')$; and using (4.4), $\psi(g_1^{-1}g_1\omega') = \psi(g_1^{-1}g_2\omega')$. But also $\psi(g_1^{-1}g_1\omega') = \psi(g_2^{-1}g_2\omega')$ so that $\psi(g_1^{-1}g_2\omega') = \psi(g_2^{-1}g_2\omega')$, and substituting for ω' gives $\psi(g_1^{-1}\omega) = \psi(g_2^{-1}\omega)$.

Using Lemmas 4.3 and 4.4 we may now give unique definitions

(4.7) $$g_1^\dagger \cdot g_2^\dagger = (g_1 \cdot g_2)^\dagger \quad \text{and} \quad (g^\dagger)^{-1} = (g^{-1})^\dagger ,$$

and the mapping of $\mathcal{G}^*$ (or equivalently $\mathcal{G}$) onto $\mathcal{G}^\dagger$ is a homomorphism.

The natural definition for invariance of an estimator $\hat{\psi}(x)$ of (which indeed corresponds to Lehmann's (1959) requirement for an invariant decision procedure) is

(4.8) $$\hat{\psi}(gx) = g^\dagger\hat{\psi}(x).$$

It may be remarked that a ψ-estimator defined in terms of an invariant ω-estimator is invariant; that is, it is easily shown that if $\hat{\omega}(x)$ satisfies (4.1) and $\hat{\psi}(x) = \psi(\hat{\omega}(x))$ then $\hat{\psi}$ satisfies (4.8).

We now give the analog of Lemma 4.1 which is appropriate for the problem of estimation of $\psi(\omega)$.

<u>Lemma 4.4</u>. Assume $\hat{\psi}(x)$ satisfies (4.8). Let $g_1^* \varepsilon \mathcal{G}^*$ correspond to $\omega_1 \varepsilon \Omega$ by (2.10) and let $g_1^\dagger \varepsilon \mathcal{G}^\dagger$ correspond to the equivalence class containing g_1^* according to (4.5). Define

(4.9) $$H(x, \omega_1) = g_1^{\dagger\,-1}\hat{\psi}(x).$$

Then H satisfies (2.18).

<u>Proof</u>: For any g_2 and corresponding g_2^* let $g_2^\dagger$ correspond to the equivalence class of g_2^*. Then

$$\begin{aligned} H(\breve{g}_2 x, g_2^* \omega_1) &= (g_2^\dagger \cdot g_1^\dagger)^{-1}\hat{\psi}(g_2 x) \\ &= (g_1^\dagger)^{-1}(g_2^\dagger)^{-1} g_2^\dagger\,\hat{\psi}(x) \\ &= H(x, \omega_1)\,. \end{aligned}$$

5. <u>Examples of best invariant estimators</u>.

<u>Theorem 5.1</u>. Suppose that $\psi(\omega)$ is invariantly estimable and that the (non-negative) loss when the true value ψ is estimated by the value $\hat{\psi}$ has the form $L(\hat{\psi}, \psi) = \Phi(g^{\dagger\,-1}\hat{\psi})$. Then the loss is minimized in the class of invariant estimators satisfying (4.8) when $\hat{\psi}(x)$ is that value of $\hat{\psi}$ which minimizes

(5.1) $$E_f \Phi(g^{\dagger\,-1}\psi).$$

<u>Proof</u>: Let $\hat{\psi}(x)$ be an invariant estimator, and let $\hat{\psi}_0(x)$ be the value which minimizes (5.1). It can be shown that $\hat{\psi}_0(x)$ also satisfies (4.8). Let Φ, Φ_0 denote the corresponding Φ-values. Then by definition of $\hat{\psi}_0$ we have $E_f(\Phi - \Phi_0) \geq 0$ for all x. By Lemma 4.4 both $g^{\dagger\,-1}\hat{\psi}$ and $g^{\dagger\,-1}\hat{\psi}_0$ satisfy the invariance condition (2.18) and hence so does $(\Phi - \Phi_0)$. By Theorem 2.1, $E_a(\Phi - \Phi_0) = E_f(\Phi - \Phi_0) \geq 0$ for all a. Taking expectation with respect to the distribution of a gives $E\,\Phi \geq E\,\Phi_0$.

Corollary 5.1. When $\Phi(g^{+-1}\hat{\psi})$ has the form

$$\Phi(g^{+-1}\hat{\psi}) = \varphi(\omega)\,(\hat{\psi}-\psi)^2 \tag{5.2}$$

where $\varphi(\omega) > 0$, then

$$\hat{\psi}_o(x) = E_f(\psi(\omega)\,\varphi(\omega))/E_f(\varphi(\omega)) \tag{5.3}$$

is the miminum mean square error invariant estimator of ψ, that is, it minimizes $E(\hat{\psi}-\psi)^2$.

Proof: Clearly $E_f\{\varphi(\omega)(\hat{\psi}-\psi)^2\}$ is minimized when $\hat{\psi} = \hat{\psi}_o$ given by (5.3). Since $\varphi(\omega)$ is a constant for the operator E_a,

$$\varphi(\omega)E_a(\hat{\psi}_o-\psi)^2 = E_f\{\varphi(\omega)(\hat{\psi}_o-\psi)^2\}$$
$$\leq E_f\{\varphi(\omega)(\hat{\psi}-\psi)^2\} = \varphi(\omega)E_a(\hat{\psi}-\psi)^2.$$

Thus $E_a(\hat{\psi}_o-\psi)^2 \leq E_a(\hat{\psi}-\psi)^2$, and therefore $E(\hat{\psi}_o-\psi)^2 \leq E(\hat{\psi}-\psi)^2$.

Table 5.1 gives six examples of invariantly estimable functions $\psi(\omega)$ and the corresponding expressions for $g^{+}\psi(\omega)$ and $g^{+-1}\hat{\psi}$ implied by the definitions of g given in Table 3.1. If the quantity $\lambda = g^{+-1}\hat{\psi}$ equals zero when $\hat{\psi}=\psi$ then reasonable loss functions are $|\lambda|$, λ^2, λ^4, etc. If $\lambda = 1$ when $\hat{\psi}=\psi$ then one may use $|\lambda-1|$, $(\lambda-1)^2$, $(\lambda-1)^4$, etc. Theorem 5.1 would apply in any of these cases.

Table 5.1

Example	$\psi(\omega)$	$g^{+}\psi(\omega) = \psi(g^{*}\omega)$	$\lambda = g^{+-1}\hat{\psi}$
3.1	θ	$\psi+\alpha$	$\hat{\psi}-\psi$
3.2	θ	$\beta\psi+\alpha$	$(\hat{\psi}-\psi)/\sigma$
3.2	σ	$\beta\psi$	$\hat{\psi}/\psi$
3.2	$\theta+\sigma$	$\beta\psi+\alpha$	$(\hat{\psi}-\psi)/\sigma$
3.3	$c_1\theta_1+c_2\theta_2$	$\beta\psi+c_1\alpha_1+c_2\alpha_2$	$(\hat{\psi}-\psi)/\sigma$
3.4	$\sigma_1^{r}\sigma_2^{s}$	$\beta_1^{r}\beta_2^{s}\psi$	$\hat{\psi}/\psi$

Table 5.2 shows how Φ can be chosen in each case so that Corollary 5.1 can be applied to give the minimum mean square error invariant estimator exhibited in the last column. The first three of the six examples were considered by Pitman (1939). Of course it is not to be inferred that Corollary 5.1 would apply in any example. For instance, in Example 3.1 if we take $\psi(\omega) = \theta^3$ instead of θ, then ψ is still invariantly estimable, $g^{+}\psi = (\psi^{1/3} + \alpha)^3$, $\lambda = g^{+-1}\hat{\psi} = (\hat{\psi}^{1/3} - \psi^{1/3})^3$. Here the loss function $\lambda^{2/3}$ leads to $\hat{\psi}(x) = (E_f\psi^{1/3})^3$. Of course this is simply a translation of the solution obtained in Table 5.2; the point is that it cannot be called a minimum mean square error invariant estimator of $\psi = \theta^3$.

Table 5.2

Example	$\psi(\omega)$	$\Phi(\lambda)$	Minimum mean square error invariant estimator
3.1	θ	λ^2	$E_f\psi$
3.2	θ	λ^2	$E_f(\sigma^{-2}\psi)/E_f(\sigma^{-2})$
3.2	σ	$(\lambda-1)^2$	$E_f(\psi^{-1})/E_f(\psi^{-2})$
3.2	$\theta+\sigma$	λ^2	$E_f(\sigma^{-2}\psi)/E_f(\sigma^{-2})$
3.3	$c_1\theta_1 + c_2\theta_2$	λ^2	$E_f(\sigma^{-2}\psi)/E_f(\sigma^{-2})$
3.4	$\sigma_1{}^r\sigma_2{}^s$	$(\lambda-1)^2$	$E_f(\psi^{-1})/E_f(\psi^{-2})$

6.. Acknowledgement. This work was supported by National Science Foundation grants G14831 and G19126.

REFERENCES

Fisher, R. A. (1934). Two new properties of mathematical likelihood. Proceedings of the Royal Society, Series A, 14 285-307.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh.

Fraser, D. A. S. (1961a). On fiducial inference. Ann. Math. Statist. 32 661-676.

Fraser, D. A. S. (1961b). The fiducial method and invariance. Biometrika 48 261-280.

Hora, R. B. (1964). Fiducial probability theory for distributions with a group structure. Technical Report Number 33, Department of Statistics, University of Minnesota, Minneapolis.

Lehmann, E. L. (1950). Notes on the theory of estimation (mimeographed), University of California, Berkeley.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. Wiley, New York.

Pitman, E. J. G. (1939). The estimation of the location and scale parameters of a continuous population of any given form. Biometrika 30 391-421.

Wallace, David L. (1959). Conditional confidence level properties. Ann. Math. Statist. 30 864-876.