FIDUCIAL THEORY AND INVARIANT PREDICTION[1]

By R. B. Hora and R. J. Buehler

University of Oklahoma and University of Minnesota

1. Introduction. We are concerned with the prediction of future observations y and functions $\psi(y)$ given past observations x when the joint distribution of x and y given an unknown parameter ω satisfies certain invariance properties. Our results are analogous to those given in an earlier paper on estimation (Hora and Buehler (1966), hereafter referred to as HB-1), and therefore some details are omitted from proofs. An identity is given involving the expectation of invariant functions $H(x, y, \omega)$ with respect to a distribution which we call the joint fiducial distribution of y and ω given x. The identity is used to define "best" invariant predictors in terms of fiducial expectations, and to establish sufficient conditions for prediction limits obtained from the fiducial distribution of y to be prediction analogues of confidence limits. The relationship to consistency criteria for fiducial distributions is indicated.

For regression models, prediction analogues of confidence limits were discussed by Eisenhart (1939), and later in textbooks such as Mood (1950), Section 13.3. In the interest of simplicity the present paper does not include a regression structure, which would be a fairly straightforward extension.

Weiss (1955) gave a general method of determining "confidence sets" for future observations y using a sufficient statistic $T(x, y)$ for ω. Our construction in Section 7 below is similar to his; his method would apply in certain cases lacking our group structure, but the present method applies in some cases when his sufficient statistic is lacking.

Kitagawa (1951) considered estimators α_1^* and α_2^* of a parameter α, based respectively on a past and a future sample and considered the accuracy of prediction of α_2^*. Later Kitagawa (1957) gave a theory of fiducial prediction quite close in spirit to the present work, but depending heavily on the theory of exponential families of distributions and on sufficient statistics, which are not required in the present treatment. Kitagawa's (1957) statistic h in Definition 2.3 and in equations (4.01) and (4.02) is an ancillary function of two sufficient statistics, and plays the same role as the quantity $t^{-1}y$ in the Appendix below. Kudō (1956) applied Kitagawa's theory in obtaining the fiducial distribution of the maximum of a future sample from a normal population. This case falls within the present scope since $\psi(y) = \max(y_1, \cdots, y_m)$ is an invariantly predictable function, as defined below in Section 5.

The Bayesian analysis of the prediction problem has been discussed for example by Fisher (1956), Chapter 3, Section 2, and by Jeffreys (1961), Chapter 3.

In the present paper attention is restricted to continuous variates. Similar

Received 16 May 1966; revised 28 November 1966.

[1] Research supported by National Science Foundation Grant GP3816.

problems in the prediction of discrete variates have been considered by Roy (1960) and Thatcher (1964).

2. Distributional assumptions. In HB-1 we gave five assumptions regarding the distribution of x given ω. Recall that in our earlier notation, $\mathcal{G} = \{g\}$ is a group which transforms both $\mathfrak{X} = \{x\}$ and $\Omega = \{\omega\}$, a is an ancillary statistic labeling the orbits $\mathcal{G}x$, and t is a conditionally sufficient statistic given a. We now make three additional assumptions regarding the distribution of y given x and ω.

ASSUMPTION 6. $(\mathcal{Y}, B_Y)$ is a measurable space, and for each $g \varepsilon \mathcal{G}$, $gy (y \varepsilon \mathcal{Y})$ is a measurable one-to-one transformation of $\mathcal{Y}$ onto itself.

ASSUMPTION 7. ξ is a measure on $\mathcal{Y}$ such that for each $g \varepsilon \mathcal{G}$, $Y \varepsilon B_Y$, $\xi(Y) = J(g)\xi(gY)$, that is, the Radon-Nikodym derivative $J(g) = \xi(dy)/\xi(g(dy))$ exists and does not depend on y.

ASSUMPTION 8. For each $\omega \varepsilon \Omega$ there is a probability distribution on $\mathfrak{X} \times \mathcal{Y}$ such that the conditional distribution on $\mathcal{Y}$ given x and ω has the form $P^{\omega}(Y \mid x) = \int_Y f_2(\omega^{-1}y \mid \omega^{-1}t, a)J(\omega)\xi(dy)$, all $Y \varepsilon B_Y$.

Loosely speaking, models of this kind will arise when elements g of $\mathcal{G}$ transform both $\mathfrak{X}$ and $\mathcal{Y}$, and when a family of distributions P^g is defined for a given P by $P^g(gX \times gY) = P(X \times Y)$. The assumptions are not actually symmetric with respect to X and Y however, since each orbit $\mathcal{G}y$ (unlike each orbit $\mathcal{G}x$) need not be in one-to-one correspondence with $\mathcal{G}$. For example if $\mathcal{G}$ consists of changes of both location and scale, then $\mathfrak{X}$ must be at least two-dimensional, while $\mathcal{Y}$ maybe one-dimensional. Our earlier location and scale parameter examples extend in obvious ways to include a space $\mathcal{Y} = \{y\}$ of future observations, with gy defined analogously to gx.

3. Bayesian and fiducial distributions. In this section we define fiducial distributions to be equivalent to posterior distributions when the prior measure equals the right Haar measure ν on the group $\mathcal{G}$. The definitions are then justified in several ways.

Assumptions 4 and 8 give a joint density for (t, a, y, ω) of the form

$$f(\omega^{-1}t, \omega^{-1}y \mid a)J(\omega)\lambda(da)\mu(dt)\xi(dy)\nu(d\omega) \tag{1}$$

where

$$f(t, y \mid a) = f_1(t \mid a)f_2(y \mid t, a) \tag{2}$$

and where f_1 is the f of Assumption 4. Note that the total measure of (1) will be infinite when ν is unbounded. We shall require the following three conditional probability elements implied by (1):

$$\text{for } y, \omega \mid t, a: \quad f(\omega^{-1}t, \omega^{-1}y \mid a)J(\omega)\Delta(t)\xi(dy)\nu(d\omega); \tag{3}$$

$$\text{for } \omega \mid t, a: \quad f_1(\omega^{-1}t \mid a)\Delta(t)\nu(d\omega); \tag{4}$$

$$\text{for } y \mid t, a: \quad \Delta(t)\xi(dy) \int_\Omega f(\omega^{-1}t, \omega^{-1}y \mid a)J(\omega)\nu(d\omega). \tag{5}$$

The expression (4) is just the fiducial distribution of ω given x in Fraser's

(1961) sense, and could be derived by use of the conditional pivotal quantity $\omega^{-1}t$.

Two justifications for calling (5) the fiducial distribution of future observations y given past observations x are: (a) consistency with Fisher (1935), (1956), pp. 113–116, in special cases which are straight-forward to verify; and (b) an alternative derivation given in the Appendix, using a pivotal argument which is Fisherian in spirit. Previous generalizations of Fisher's examples have been given by Kitagawa (1957), Sprott (1963), and Ramsey (1963). The present expression is claimed to be more general in that it does not require the existence of sufficient statistics (Kitagawa), nor is it univariate in nature (Sprott), nor is it restricted to location and scale parameter families (Ramsey).

The relationship of (3) to (4) and (5) will be our justification for calling (3) the joint fiducial distribution of y and ω given x. Special cases have been given previously by Ramsey (1963).

4. An expectation identity. The following result generalizes our previous identity, and is used below in Sections 6 and 7.

THEOREM 1. *If Assumptions* 1 *through* 8 *are satisfied and if*

$$H(gx, gy, g\omega) = H(x, y, \omega) \tag{6}$$

then

$$E^{y,t \mid a,\omega}H(x, y, \omega) = E_f^{y,\omega \mid x}H(x, y, \omega) \tag{7}$$

where $E^{y,t \mid a,\omega}$ *denotes conditional expectation with respect to* (y, t) *given* (a, ω) *and where* $E_f^{y,\omega \mid x}$ *denotes expectation with respect to the fiducial distribution* (3) *of* (y, ω) *given* $x = (t, a)$.

PROOF. Both sides of (7) can be shown to equal $\int \int H(t, a, y, e)f(t, y \mid a) \cdot \mu(dt)\xi(dy)$ where $H(t, a, y, \omega) = H(x, y, \omega)$ and e is the identity.

5. Invariantly predictable functions. We now consider the problem of prediction of a function $\psi(y)$ of the future observations y. In order to exploit the assumed invariance properties of the family of distributions we find it necessary to restrict attention to "invariantly predictable functions" defined by

$$\psi(y_1) = \psi(y_2) \quad \text{implies} \quad \psi(gy_1) = \psi(gy_2), \quad \text{all} \quad g \in \mathcal{G}. \tag{8}$$

The transformations $\mathcal{G}' = \{g'\}$ of $\{\psi\}$ defined by

$$g'\psi(y) = \psi(gy) \tag{9}$$

then form a group homomorphic to $\mathcal{G}$.

Some examples of functions ψ satisfying (8) are found in Table 1 below. In addition, any function which is constant on each orbit $\mathcal{G}y$ will satisfy (8) trivially; prediction of such functions is not of interest however, since their distributions are the same for all ω.

6. Best invariant predictors. To treat the prediction problem from a decision theoretic viewpoint (generalized to allow for the future observations y) we wish to construct invariant functions on $\mathfrak{X} \times \mathcal{Y} \times \Omega$ to represent the "loss" incurred when $\hat{\psi}(x)$ is the predicted value, $\psi(y)$ is the observed value, and ω is the true parameter value. A predictor $\hat{\psi}(x)$ of an invariantly predictable function $\psi(y)$ will be called "invariant" if it satisfies

$$\hat{\psi}(gx) = g'\hat{\psi}(x). \tag{10}$$

LEMMA 1. *If $\psi(y)$ and $\hat{\psi}(x)$ satisfy* (8) *and* (10), *if $\Phi(\cdot, \cdot)$ is a function on $\Psi \times \Psi$ (where $\Psi = \{\psi\}$), and if*

$$H(x, y, \omega) = \Phi(\omega'^{-1}\hat{\psi}(x), \omega'^{-1}\psi(y)) \tag{11}$$

where $\omega' \varepsilon \mathcal{G}'$ is the image of $\omega \varepsilon \Omega$ defined in (9), *then $H(x, y, \omega) = H(gx, gy, g\omega)$.*

THEOREM 2. *If* (i) *Assumptions* 1 *through* 8 *are satisfied*, (ii) *$\psi(y)$ is invariantly predictable*, (iii) *$\Phi(\cdot, \cdot)$ is a real-valued function on $\Psi \times \Psi$*, (iv) *there is a unique value $\hat{\psi}_0(x)$ of $\hat{\psi}$ which minimizes*

$$E_f^{\,y,\omega \mid x}\Phi(\omega'^{-1}\hat{\psi}, \omega'^{-1}\psi(y)),$$

then $\hat{\psi}_0(x)$ minimizes $E^{x,y \mid \omega}\Phi(\omega'^{-1}\hat{\psi}(x), \omega'^{-1}\psi(y))$ amongst the class of invariant predictors $\hat{\psi}(x)$.

PROOF. Use Lemma 1 and Theorem 1.

COROLLARY. *When Ψ is a subset of the real line and when for some $\tau(\omega), \varphi(\omega) > 0$, $\omega'^{-1}\psi = \{\varphi(\omega)\}^{\frac{1}{2}}\psi + \tau(\omega)$, then*

$$\hat{\psi}_0(x) = E_f\{\varphi(\omega)\psi(y)\}/E_f\varphi(\omega) \qquad (E_f \equiv E_f^{\,y,\omega \mid x}) \tag{12}$$

is the minimum mean square error invariant predictor of $\psi(y)$, that is, it minimizes $E^{x,y \mid \omega}\{\hat{\psi}(x) - \psi(y)\}^2$ in the class of invariant predictors.

PROOF. The theorem applies with $\Phi(u, v) = (u - v)^2$.

Table 1 gives several examples wherein the Corollary applies. Refer to HB-1 for notation in the first and third columns. In the second column $Q = \sum \sum a_{ij}y_iy_j$ where $\sum_i a_{ij} = 0$(all i) and $\sum_j a_{ij} = 0$ (all j), and Q_1 and Q_2 are similar functions of y_{1i} and y_{2i}, respectively (the definition of the y's should be evident). In the final example, $\varphi = \sigma_1^{-4r}\sigma_2^{-4s}$.

7. Fiducial prediction limits. In this section we show that invariant predictability of $\psi(y)$ essentially ensures that limits obtained from the fiducial distribution of $\psi(y)$ will be prediction analogues of confidence limits.

TABLE 1

Example	$\psi(y)$	$g'\psi(y) = \psi(gy)$	$\phi(\omega'^{-1}\hat{\psi}, \omega'^{-1}\psi)$	"best" invariant predictor
3.1	$y_{\max}$	$\psi(y) + \alpha$	$(\hat{\psi} - \psi)^2$	$E_f\psi$
3.2	$\sum a_iy_i$	$\beta\psi(y) + \alpha\sum a_i$	$(\hat{\psi} - \psi)^2/\sigma^2$	$E_f(\sigma^{-2}\psi)/E_f\sigma^{-2}$
3.2	Q	$\beta^2\psi(y)$	$(\hat{\psi} - \psi)^2/\sigma^4$	$E_f(\sigma^{-4}\psi)/E_f\sigma^{-4}$
3.3	$\sum \sum b_{ij}y_{ij}$	$\beta\psi + \sum \sum \alpha_ib_{ij}$	$(\hat{\psi} - \psi)^2/\sigma^2$	$E_f(\sigma^{-2}\psi)/E_f\sigma^{-2}$
3.4	$Q_1^rQ_2^s$	$\beta_1^{2r}\beta_2^{2s}\psi$	$(\hat{\psi} - \psi)^2\varphi$	$E_f(\varphi\psi)/E_f\varphi$

For any real valued $\psi(y)$, any observed x, and any probability level γ, an upper fiducial prediction limit $\bar{\psi}(x, \gamma)$ is defined by

$$P_f\{\psi(y) \leqq \bar{\psi}(x, \gamma) \mid x\} = \gamma \tag{13}$$

where P_f denotes fiducial probability. The fiducial prediction limits may not possess the frequency interpretation expressed by

$$P\{\psi(y) \leqq \bar{\psi}(x, \gamma) \mid \omega\} = \gamma \quad \text{for all } \omega. \tag{14}$$

THEOREM 3. *If* (i) *Assumptions* 1 *through* 8 *are satisfied,* (ii) $\psi(y)$ *is real-valued and invariantly predictable,* (iii) (13) *has a unique solution for* $\bar{\psi}(x, \gamma)$ *for all* $x \in \mathfrak{X}$, $0 < \gamma < 1$, *and* (iv) $g'\psi$ *increases as* ψ *increases for each* $g' \in \mathcal{G}'$, *then* (14) *is satisfied.*

PROOF. Let $I(x, y)$ be the indicator function which equals 1 or 0 according as $\psi(y)$ is $\leqq$ or $> \bar{\psi}(x, \gamma)$. It can be shown that $I(gx, gy) = I(x, y)$, and the result follows from Theorem 1.

COROLLARY. *Under the conditions of the theorem, Bayesian limits for* $\psi(y)$, *based on prior measure* ν, *have the frequency interpretation* (14).

A result announced by Hall and Novick (1963) is more general than this corollary in that it includes a regression parameter, but more special in being restricted to certain location and scale parameter models.

8. Remarks on some consistency criteria for fiducial distributions. Fisher (1956) states, "The concept of probability involved in the fiducial argument is entirely identical with the classical probability of the early writers, such as Bayes," (p. 51), and later on p. 125, "This fiducial distribution supplies information of exactly the same sort as would a distribution *a priori*." Such claims can be tested by actually using a fiducial density, say $f_f(\omega \mid x)$, as if it were a conventional distribution. Thus for any given density of y given x and ω, say $f(y \mid x, \omega)$ (usually this is just $f(y \mid \omega)$) we may form a joint density of (y, ω) given x as

$$f(y, \omega \mid x) = f_f(\omega \mid x) f(y \mid x, \omega). \tag{15}$$

Let $f(y \mid x)$ denote the integral of (15) over ω, and let $f(\omega \mid x, y) = f(y, \omega \mid x)/f(y \mid x)$.

One consistency criterion proposed by Lindley (1958) (and also studied by Sprott (1960) and Fraser (1962)) tests the equality of $f(\omega \mid x, y)$ and the fiducial distribution of ω given (x, y). A related criterion of Sprott (1960) requires that any posterior or fiducial distribution contain all the symmetries of the likelihood function. It is clear from the relationship to Bayesian analysis that the distributions in Section 3 satisfy the criteria of both Lindley and Sprott.

Let us say that the above density $f(y \mid x)$ is the fiducial distribution of y given x obtained by the "integral method," which was used by Sprott (1963) and seems to be implicit in Fisher (1956), p. 126. Either by direct calculation or by the Bayesian interpretation it is easily seen that $f(y \mid x)$ agrees with our definition (5), and this may be regarded as a further consistency property of the fiducial distributions here defined. A somewhat more stringent criterion proposed by Buehler (1963) requires that prediction limits for $\psi(y)$ obtained by the integral method should have the frequency property (14). Theorem 3 shows that in the present framework, invariant predictability of $\psi(y)$ (with mild regularity conditions) is a sufficient condition, and generalizes the results announced by Buehler (1963).

APPENDIX

Derivation of the fiducial distribution of y given x by the pivotal method. From (1) the distribution of (t, y) given (a, ω) is

$$f(\omega^{-1}t, \omega^{-1}y \mid a)\mu(dt)J(\omega)\xi(dy). \tag{A.1}$$

Defining $v = t^{-1}y$ we obtain the joint distribution of (t, v) given (a, ω) as

$$f(\omega^{-1}t, \omega^{-1}tv \mid a)\mu(dt)J(t^{-1}\omega)\xi(dv). \tag{A.2}$$

To obtain the marginal distribution of v given (a, ω), we integrate the last expression with respect to t, and after changing the integration variable to $z = \omega^{-1}t$ this yields

$$\xi(dv) \int f(z, zv \mid a)J(z^{-1})\mu(dz). \tag{A.3}$$

In this form it is clear that the distribution of v given a does not depend on ω, so that v is a conditional pivotal quantity. Next consider t to be fixed, and transform the variate v to y where $v = t^{-1}y$. This "pivotal argument" yields the fiducial distribution of y given $x = (t, a)$ in the form

$$\xi(dy)J(t) \int f(z, zt^{-1}y \mid a)J(z^{-1})\mu(dz). \tag{A.4}$$

On changing the integration variate from z to ω where $z = \omega^{-1}t$, the last expression will be seen to agree with (5). Thus the "pivotal argument" leads to the same result as the Bayesian analysis.

REFERENCES

BUEHLER, R. J. (1963). A new test of fiducial consistency (Abstract). *Ann. Math. Statist.* **34** 1112.

EISENHART, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Statist.* **10** 162–186.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics* **6** 391–398.

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

FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.

FRASER, D. A. S. (1962). On the consistency of the fiducial method. *J. Roy. Statist. Soc. Ser. B* **24** 425–434.

HALL, W. J. and NOVICK, M. R. (1963). A note on classical and Bayesian prediction intervals for location, scale and regression models (Abstract). *Ann. Math. Statist.* **34** 1619.

HORA, R. B. and BUEHLER, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37** 643–656.

JEFFREYS, H. (1960). *Theory of Probability* (3rd ed.). Oxford Univ. Press.

KITAGAWA, T. (1951). Successive process of statistical inferences (2). *Memoirs of the Faculty of Science, Kyusyu Univ. Ser A* **6** 55–95.

KITAGAWA, T. (1957). Successive process of statistical inference associated with an additive family of sufficient statistics. *Bull. Math. Statist. Fukuoka* **7** 92–112.

KUDŌ, A. (1956). On the confidence interval of the extreme value of a second sample from a normal universe. *Bull Math. Statist.* **6** 51–56.

LINDLEY, D. V. (1958). Fiducial distributions and Bayes theorem. *J. Roy. Statist. Soc. Ser. B* **20** 102–107.

MOOD, A. M. (1950). *Introduction of the Theory of Statistics*. McGraw-Hill, New York.

RAMSEY, F. L. (1963). Some fiducial prediction problems. MS Thesis, Iowa State University, Ames, Iowa.

ROY, A. D. (1960). Some notes on pistimetric inference. *J. Roy. Statist. Soc. Ser. B* **22** 338–347.

SPROTT, D. A. (1960). Necessary restrictions for distributions *a posteriori*. *J. Roy. Statist. Soc. Ser. B* **22** 312–318.

SPROTT, D. A. (1963). A transformation model for the investigation of fiducial distributions. *Bull. Inst. Internat. Statist.* **40** 856–869.

THATCHER, A. R. (1964). Relationships between Bayesian and confidence limits for predictions. *J. Roy. Statist. Soc. Ser. B* **26** 176–192.

WEISS, L. (1955). A note on confidence sets for random variables. *Ann. Math. Statist.* **26** 142–144.