Fiducial prediction

by

Peter McCullagh
University of Chicago
October 2004

Summary

A real-valued fiducial process is a process in the probabilistic sense, but the probability distributions are not defined on Borel sets. The adjective 'fiducial' indicates that a group acts on the observation spaces and the probability distributions are defined on the smaller σ-algebra of group-invariant Borel sets. Since a different group action determines a different σ-algebra, the group is an integral part of a fiducial process. A fiducial process obeys the rules of probability theory, but care is required in the calculation of conditional distributions, such as are required for purposes of prediction and Bayesian inference. For a simple fiducial model, the one-step prediction is an ordered pair $(\hat{g}, P)$ consisting of a group element $\hat{g}$ together with the conditional distribution P of $(\hat{g}, Y_{n+1})$ on the invariant σ-algebra on $\mathcal{G}\times\mathcal{R}$. More conventionally, the fiducial prediction is an ordered pair consisting of a reference point or estimator $\hat{g}$ together with the conditional distribution of the pivotal statistic $\hat{g}^{-1}Y_{n+1}$ on $(\mathcal{R}, \mathcal{B})$ given the σ-algebra generated by the observation.

A fiducial model with parameter space Θ is a set of processes in which the group also acts on the parameter space, and all probability distributions are defined on the group-invariant σ-algebra. Several examples of fiducial models are described, mainly with the aim of understanding the nature and limitation of fiducial predictions. A Bayesian analysis is described for fiducial models, and it is shown that if the group acts transitively on the parameter space, the prior has no effect on the conclusion. The Bayesian conclusion does not emerge immediately in the form of a probability distribution on the parameter space. Instead it occurs in the form of an ordered pair $(\hat{g}, P)$ consisting of a reference point $\hat{g} \in \mathcal{G}$ together with the conditional distribution P of $(\hat{g}, \theta)$ on $\mathcal{G}$-invariant sets in $\mathcal{G}\times\Theta$. The Bayesian conclusion may be re-expressed in the form of a conditional confidence density on the parameter space, in agreement with Fisher's (1934) analysis for location and location-scale models. This derivation is entirely probabilistic, without recourse to improper distributions.

Some key words: Bayesian fiducial model; Behrens-Fisher problem; Best linear predictor; Conditional distribution; Confidence distribution; Intrinsic process; Fiducial process; Group-invariant σ-algebra; Kriging; Partition process; Smoothing spline;

1. Introduction

Let $\mathcal{G}$ be a group acting on the observation space $\mathcal{R}^n$. By convention, the term 'distribution on $\mathcal{R}^n$' means a probability distribution defined on Borel subsets $\mathcal{B}_n$. In this paper, a fiducial distribution is a probability distribution defined on the $\mathcal{G}$-invariant Borel subsets $\mathcal{K}_n \subset \mathcal{B}_n$ of $\mathcal{R}^n$. By extension, a fiducial process is defined by a family of mutually consistent finite-dimensional distributions P_n on the observation spaces $(\mathcal{R}^n, \mathcal{K}_n)$. The adjective fiducial is simply a reminder that the group $\mathcal{G}$ acts on each observation space, and that each probability distribution is defined on $\mathcal{G}$-invariant Borel sets.

Beyond the examples used by way of illustration, it is not entirely clear what Fisher meant by the term fiducial distribution, or how a fiducial distribution differs from a confidence density. The view is that 'confidence interval' is simply an American translation of the English 'fiducial interval', was declared by Barnard (1954) to be 'not wholly accurate,' though the nature of the inaccuracy was not fully explained. In his discussion of the papers by Fieller, Creasy and David (1954), Hammersley began with a carefully worded declaration that the term fiducial probability is *familiar in the literature, and there was a time when I thought I knew what it meant.* Brillinger's (1962) remark, that the fiducial method was put forward by Fisher as a general principle but has been illustrated chiefly by means of special examples, was later echoed by Savage (1976, §4.6). It was unclear in Fisher's construction whether fiducial probability was meant to obey the rules of probability in Kolmogorov's sense. Indeed, the interpretation implicit in certain counterexamples Tukey (1957), Lindley (1958), Savage (1976) suggest that it does not. My own instinct is that the fiducial argument, though not mathematically convincing, is sufficiently plausible that it must contain the threads of a logically sound theory.

This paper is not an attempt to explain Fisher or later work by Barnard (1963, 1995), Dempster (1963) or Fraser (1961, 1968) because we begin with a fundamentally different probabilistic foundation. The present definition makes it clear that fiducial probability and fiducial processes do obey the rules of probability. In fact the fiducial distribution is derived directly by the application of conditional probability and Bayes's theorem, so the rules are clear and nothing fundamentally new or mysterious is involved. Nevertheless, the manipulation of fiducial processes and the calculation of conditional distributions are not entirely straightforward.

The simplest sort of fiducial model is an exchangeable real-valued process in which the distribution P_n on the measurable space $(\mathcal{R}^n, \mathcal{K}_n)$ is the marginal distribution of P_{n+1} on $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1})$ under coordinate deletion. This statement implies, of course, that coordinate deletion is $\mathcal{K}$-measurable, and that each $\mathcal{K}_n$ is closed under coordinate permutation, conditions not to be taken for granted. Frequently, $\mathcal{K}_1 = \{\emptyset, \mathcal{R}\}$ is trivial, so each one-dimensional projection Y_r has a distribution P_1 on $\mathcal{K}_1$ that is necessarily trivial and uninformative. So far as prediction is concerned, the observation is a measurable map $Y^{(n)}\colon (\mathcal{R}^{n+1}, \mathcal{K}_{n+1}) \to (\mathcal{R}^n, \mathcal{K}_n)$, and the predictive distribution is the conditional distribution on $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1})$ given the σ-algebra generated by the random variable $Y^{(n)}$. Since this is a probability distribution on $\mathcal{G}$-invariant sets, it is a fiducial predictive distribution. For a process of this sort the σ-algebra generated by Y_{n+1} is trivially independent of the σ-algebra generated by the observation. In other words, the fiducial predictive distribution is not the conditional distribution of Y_{n+1} given the observation because this distribution is trivial and uninformative.

Because of the restriction to $\mathcal{G}$-invariant sets, it is necessary to distinguish between the

value of the observation as a point in $y \in \mathcal{R}^n$ and the associated event $\mathcal{G}y \in \mathcal{K}_n$, the orbit of the point y by the group. The conditional distribution given the observation is the conditional distribution given the σ-algebra generated by the observed event. However, the observation determines both the point and the event, and it is sometimes possible to express the observation in the form $y \cong (\hat{g}, \mathcal{G}y)$ with $\hat{g} \in \mathcal{G}$ as a reference point; see Eaton and Sudderth (1999). The predictive distribution may then be expressed in a more convenient form, consisting of the reference point $\hat{g}$ together with the conditional distribution of the Borel pivotal statistic $\hat{g}^{-1}Y_{n+1}$. Expressions of this sort are convenient, and are frequently used in the construction of confidence intervals and predictive distributions. But ultimately they cannot obscure the fact that a fiducial process is defined on $\mathcal{G}$-invariant sets, and fiducial inferences likewise are defined on $\mathcal{G}$-invariant sets. As a consequence, there are non-trivial measurability constraints governing the ways in which a fiducial distribution may be manipulated and transformed.

Apart from the fact that all processes are defined on $\mathcal{G}$-invariant sets, the theory of fiducial prediction and fiducial inference is merely the theory of conditional probability, which we illustrate by a number of examples.

2. Ten illustrations of fiducial prediction

As the following examples demonstrate, a real-valued fiducial process on $\mathcal{U}$ is a process in the orthodox sense consisting of a family of distributions P_S on $\mathcal{R}^S$, one such distribution for each finite subset $S \subset \mathcal{U}$. The adjective fiducial serves only as a reminder that the distributions are not defined on Borel sets. The family of distributions is mutually consistent under sub-sampling of units, so that for each $S \subset S'$, P_S is the marginal distribution of $P_{S'}$ under deletion of units not in S. As always, prediction is based on the conditional distribution given the σ-algebra generated by the observed segment of the process. In that sense, fiducial prediction is an integral part of probability theory for processes.

Example 1. Gaussian prediction Let (X, Y) be a bivariate random variable, a function taking values in $\mathcal{R}^2$. The joint distribution is such that $Y - X \sim N(0, 1)$ independent of X. The value $X = 7.5$ is observed. What can reasonably be deduced about the likely value of Y? It is natural to answer the question in the form of a predictive distribution or conditional distribution given X, and the obvious predictive answer is $Y \sim N(7.5, 1)$. From this, we conclude if necessary, that the predictive distribution for Y^2 is $\chi^2_1(7.5^2)$ with non-centrality parameter 7.5^2. In particular $E(Y^2 \,|\, X = 7.5) = 7.5^2 + 1$.

The argument just given is a familiar one. It is a correct argument only if the phrase 'taking values in $\mathcal{R}^2$' is understood in the conventional sense that the probability distribution is defined on the Borel subsets of $\mathcal{R}^2$. To see what may happen if this convention is suspended, let $\mathbf{1}_n \subset \mathcal{R}^n$ be the one-dimensional subspace consisting of vectors whose components are equal, and let $\mathcal{K}_n$ be the class of Borel sets $A \subset \mathcal{R}^n$ such that $A + \mathbf{1}_n = A$. Thus $\mathcal{K}_1 = \{\emptyset, \mathcal{R}\}$, and $\mathcal{K}_2$ is the σ-algebra generated by the contrast $Y - X$ as a Borel-measurable map into $(\mathcal{R}, \mathcal{B})$. The coordinate projection $(X, Y) \mapsto X$ is $\mathcal{K}$-measurable $(\mathcal{R}^2, \mathcal{K}_2) \to (\mathcal{R}, \mathcal{K}_1)$, trivially so. The contrast $(X, Y) \mapsto Y - X$ is Borel measurable $(\mathcal{R}^2, \mathcal{K}_2) \to (\mathcal{R}, \mathcal{B})$, in fact an isomorphism $\mathcal{K}_2 \cong \mathcal{B}$. Since the sub-algebra of $\mathcal{K}_2$ generated by the coordinate projections is trivial, every distribution defined on $(\mathcal{R}^2, \mathcal{K}_2)$ has the property that $X, Y, Y - X$ are mutually independent, i.e. the sub-algebras generated by

these random variables are independent. Accordingly, the statement of independence in the preceding paragraph is gratuitous. The conditional distribution on $(\mathcal{R}^2, \mathcal{K}_2)$ given that $X = 7.5$ is the same as the unconditional distribution, so in that sense the observation is uninformative. The pedantically correct statement is that $X = 7.5$ and the conditional distribution on $(\mathcal{R}^2, \mathcal{K}_2)$ is such that $Y - X \sim N(0, 1)$.

Given that $X = 7.5$ has been observed, the fiducial prediction that $Y \sim N(7.5, 1)$ seems direct and unlikely to mislead. The adjective fiducial, meaning a point or other object taken as a basis for comparison, is apposite because it serves as a reminder of the relative nature of the fiducial statement, and that the random variable on which it is based is the pivotal contrast $Y - X$. The fiducial distribution is most definitely not the conditional distribution of the random variable Y given the random variable X because these $\mathcal{K}_1$-projections are trivially independent. The meaning of the fiducial statement is that the conditional distribution of the Borel contrast or pivot $Y - X$ given $X = 7.5$ is $N(0, 1)$. From this, we may deduce the conditional distribution of any Borel-measurable function such as $|Y - X|$. But $Y^2 - X^2$ is not a Borel-measurable function on $(\mathcal{R}^2, \mathcal{K}_2)$, so the final conclusion in the first paragraph of this section does not follow from the information provided.

Example 2. Cauchy prediction The second illustration is less familiar, but the logic of the argument is the same. Let (Y_1, Y_2, Y_3, Y_4) be a random variable in $\mathcal{R}^4$ such that the cross-ratio

$$X = \frac{(Y_4 - Y_3)(Y_2 - Y_1)}{(Y_4 - Y_2)(Y_3 - Y_1)}$$

is independent of (Y_1, Y_2, Y_3) with density

$$f(x) = \frac{x \log|x| + (1 - x) \log|1 - x|}{\pi^2\, x(x - 1)}. \tag{2.1}$$

The value (y_1, y_2, y_3) is observed. How should we predict Y_4?

A simple calculation shows that

$$Y_4 = \frac{y_2(y_1 - y_3)X + y_3(y_2 - y_1)}{(y_1 - y_3)X + y_2 - y_1}, \tag{2.2}$$

a fractional linear transformation of X with coefficients determined by the observed values. The fiducial distribution for Y_4 is obtained by transforming the density (2.1) in the usual way. In particular, if the observed values are $(-1, 0, 1)$, the fiducial density for Y_4 is

$$\frac{(y - 1) \log|y - 1| + (y + 1) \log|y + 1| - 2y \log|2y|}{\pi^2 y(1 - y^2)}, \tag{2.3}$$

which has a spike at each of the observed values. In this example $\mathcal{G}$ is the group of real fractional linear transformations.

Expression (2.1) is in fact the density of the cross-ratio of four Cauchy random variables, conditionally independent given the parameter. In the conventional formulation, however, the cross-ratio is not independent of (Y_1, Y_2, Y_3), and the fiducial distribution is not the conditional distribution of Y_4 given the values observed. Nonetheless, expression (2.3) is the fiducial density estimator, the predictive distribution for the next value based on a sample of size three from an exchangeable sequence.

Example 3. Stationary random walk Let $X = (X_1, X_2, \ldots)$ be a real-valued process on the integers such that the differences $Z_r = X_r - X_{r-1}$ are independent and identically distributed $N(0,1)$. The term 'real-valued process' means that each component X_r is a real-valued random variable, but it does not imply that the finite-dimensional distributions are defined on Borel sets. On the contrary, $\mathcal{K}_n$ is the class of Borel sets such that $A+\mathbf{1}_n = A$, i.e. translation-invariant in $\mathcal{R}^n$. The finite-dimensional distributions are defined on $(\mathcal{R}^n, \mathcal{K}_n)$ by the statement that successive differences are independent $N(0,1)$. Suppose the value $(X_1, \ldots, X_n)$ is observed, and we wish to predict the next pair. Since the current value x_n is known, and the next values are $x_n + Z_{n+1}$ and $x_n + Z_{n+1} + Z_{n+2}$, common sense tells us that the fiducial distribution must be Gaussian with mean (x_n, x_n) and covariance matrix $\mathrm{cov}(X_{n+r}, X_{n+s} \,|\, \mathrm{data}) = \min(r, s)$.

We note that coordinate deletion $\mathcal{R}^{n+1} \to \mathcal{R}^n$ is $\mathcal{K}$-measurable, so the measurable spaces $\{(\mathcal{R}^n, \mathcal{K}_n)\}$ form a projective system on which a real-valued process may be defined. The bivariate random variable (X_{n+1}, X_{n+2}) has a distribution on $(\mathcal{R}^2, \mathcal{K}_2)$ defined by the statement that the difference is $N(0,1)$. Furthermore (X_{n+1}, X_{n+2}) is independent of $(X_1, \ldots, X_n)$, so the conditional and unconditional distributions are the same. By contrast, the fiducial distribution on the Borel subsets of $\mathcal{R}^2$, is based on the bivariate Borel contrast $(X_{n+1} - X_1, X_{n+2} - X_1)$, whose conditional distribution given the data is bivariate Gaussian with mean $(x_n - x_1, x_n - x_1)$ and covariance as described above.

Example 4. Smoothing spline Let $\mathcal{U}$ be the set or population of statistical units, and let $x{:}\mathcal{U} \to \mathcal{R}$ be a given real-valued function. It may be helpful to think of $x(u) = x_u$ as the spatial position of the unit, so the alternative term 'covariate' may be misleading if narrowly interpeted. Let Y be the 'intrinsic' zero-mean stationary Gaussian process whose generalized covariance function is

$$K_{ij} = \delta_{ij} + |x_i - x_j|^3$$

for $i, j \in \mathcal{U}$. Given the values of the process on $n \geq 2$ units $u_1, \ldots, u_n$ at points $x(u_1), \ldots, x(u_n)$, we aim to predict the value for a new unit u such that $x(u) = t$.

Let $\mathcal{X} \subset \mathcal{R}^{\mathcal{U}}$ be the two-dimensional subspace spanned by the constant vector and x, and, for each sample, let $\mathcal{X}_n \subset \mathcal{R}^n$ be the restriction of $\mathcal{X}$ to the sampled units. Let $\mathcal{K}_n \subset \mathcal{R}^n$ be the class of Borel sets such that $A + \mathcal{X}_n = A$. The measurable spaces $(\mathcal{R}^n, \mathcal{K}_n)$ form the projective system on which the finite-dimensional distributions are defined. The adjective 'intrinsic' means that the process has a non-zero kernel, in this case the two-dimensional kernel $\mathcal{X}$. In effect, the process is defined on the Borel quotient space $\mathcal{R}^{\mathcal{U}}/\mathcal{X}$. The generalized covariance function $K{:}\mathcal{U} \times \mathcal{U} \to \mathcal{R}$ is a symmetric matrix such that for each contrast $c \in \mathcal{X}_n^0$ taking the value zero on $\mathcal{X}_n$, the quadratic form $c^{\mathrm{T}}Kc$ is positive for $c \neq 0$. Each contrast $c \in \mathcal{X}_n^0$ determines a Borel variable $c^{\mathrm{T}}Y$ whose distribution is zero-mean Gaussian with variance $c^{\mathrm{T}}Kc$.

Suppose that $n = 3$, that the observed x-values are $1, 2, 3$ and that we wish to predict for a new unit such that $x(u) = 4$. Let $Z = CY$ where C is the matrix

$$C = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}.$$

We find that $\mathrm{var}(Z_1) = \mathrm{var}(Z_2) = 14$ with covariance -2. Accordingly, the conditional distribution of Z_2 given $Z_1 = z_1$ is Gaussian with mean $-z_1/7$ and variance 13.7. In other

words, the linear combination

$$\begin{aligned} Z_2 + Z_1/7 &= Y_4 - 2Y_3 + Y_2 + (Y_3 - 2Y_2 + Y_1)/7 \\ &= Y_4 - (13Y_3 - 5Y_2 - Y_1)/7 \end{aligned}$$

is Gaussian with zero mean and variance 13.7, independent of Z_1, i.e. independent of the observation (Y_1, Y_2, Y_3). The combination $\tilde{Y}_4 = (13Y_3 - 5Y_2 - Y_1)/7$ is the best linear predictor of Y_4, but it is not a Borel-measurable statistic, nor is it the conditional mean of Y_4 given the data. The fiducial distribution for Y_4 is simply $N(\tilde{y}_4, 13.7)$, but the probabilistic meaning is a relative one, namely that $Y_4 - \tilde{Y}_4$ is a Borel-measurable Gaussian random variable independent of (Y_1, Y_2, Y_3).

Once again, the fiducial distribution is not the conditional distribution of the random variable $Y(u)$ given the observed data. For the particular covariance function shown above, the mean of the fiducial distribution is a cubic spline in $x(u)$ (Wahba, 1990).

Example 5. Interpolation and spatial Kriging Many models used in the study of stationary spatial processes have a non-zero kernel, a vector subspace $\mathcal{X}$ consisting of certain functions $\mathcal{R}^2 \to \mathcal{R}$. For exaple, the thin-plate generalized covariance function for a planar Gaussian process

$$K(x, x') = |x - x'|^2 \log |x - x'|$$

has a kernel of dimension three consisting of the polynomials of degree ≤ 1, the constant and two linear functions. Denote by $\mathcal{X}_n$ the restricton of $\mathcal{X}$ to the finite set of sampled sites. The finite-dimensional distributions of the process are defined on $(\mathcal{R}^n, \mathcal{K}_n)$, where $\mathcal{K}_n$ is the class of Borel sets $A \subset \mathcal{R}^n$ such that $A + \mathcal{X}_n = A$. Given values $Y_1, \ldots, Y_n$ observed at spatial locations $x_1, \ldots, x_n$, the fiducial distribution for the value of Y at any other position x^* may be computed in the standard manner using contrasts in $\mathcal{X}^0$. The mean of the fiducial distribution is a certain linear combination $a_1 Y_1 + \cdots + a_n Y_n$ having the property that $Y(x^*) - a_1 Y_1 - \cdots - a_n Y_n$ is a contrast in $\mathcal{X}^0$ that is independent of $(Y_1, \ldots, Y_n)$. For the covariance function shown above, this linear combination is called the thin-plate interpolating spline (Wahba, 1990). More generally, for a Gaussian process having a non-zero kernel, prediction using the fiducial mean is called universal Kriging (Stein, 1999; Cressie, 1993).

Example 6. An exchangeable Gaussian process Let $N_n(\theta \mathbf{1}_n, I_n)$ be the Gaussian distribution on $\mathcal{R}^n$, let $\mathcal{K}_n$ be the σ-algebra of Borel subsets of $\mathcal{R}^n$ such that $A + \mathbf{1}_n = A$, and let $P_n(A) = N_n(\theta \mathbf{1}_n, I_n)(A)$ be the restriction to $\mathcal{K}_n$. By construction P_n does not depend on the parameter θ. The distributions P_n on $(\mathcal{R}^n, \mathcal{K}_n)$ are mutually compatible under coordinate permutation and coordinate deletion, so there exists a real-valued exchangeable process $Y_1, Y_2, \ldots$ having finite-dimensional distributions P_n. Each contrast $\sum c_j Y_j$ such that $\sum c_j = 0$ is a Borel-measurable function, a random variable with distribution $N(0, \sum c_j^2)$. A sum or average such as $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$ is $\mathcal{K}_1$-measurable, and thus trivial as a random variable. The contrast

$$D = \bar{Y}_n - \bar{Y}_m = (Y_1 + \cdots + Y_n)/n - (Y_{n+1} + \cdots + Y_{n+m})/m$$

is Borel-measurable $(\mathcal{R}^{m+n}, \mathcal{K}_{m+n}) \to (\mathcal{R}, \mathcal{B})$ with distribution $N(0, 1/m + 1/n)$, independent of the projections $(Y_1, \ldots, Y_n)$ and $(Y_{n+1}, \ldots, Y_{n+m})$.

After observing the value $(y_1, \ldots, y_n)$ as a point in $(\mathcal{R}^n, \mathcal{K}_n)$, the fiducial prediction for the next value is based on the conditional distribution on $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1})$ given the observed value. But the contrast $D = Y_{n+1} - \bar{Y}_n$ is independent of the observed value, so the fiducial distribution for Y_{n+1} is $N(\bar{y}_n, (n+1)/n)$. Likewise, the conditional distribution of the contrast $\bar{Y}_\infty - \bar{Y}_n$ given the observed value is $N(0, 1/n)$, so the fiducial distribution for the infinite average is $N(\bar{y}_n, 1/n)$. These statements are relatively uncontroversial provided that they are understood in the sense of contrasts: the conditional distribution of the Borel increment $\bar{Y}_\infty - \bar{Y}_n$ given $(Y_1, \ldots, Y_n)$ is $N(0, 1/n)$. Equivalently, since $\theta = \bar{Y}_\infty$, the increment $\theta - \bar{Y}_n$ is distributed as $N(0, 1/n)$ given $(Y_1, \ldots, Y_n)$.

Example 7. Student's process Let $X_1, X_2, \ldots$ be a real-valued process such that, for each $n \geq 2$, the configuration vector with components $(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)/(s_n\sqrt{n-1})$ is uniformly distributed on the unit sphere in the $n-1$-dimensional space of residuals. In this example $\mathcal{G}$ is the location-scale group acting component-wise on $\mathcal{R}^n$, and $\mathcal{K}_n$ is the σ-algebra of $\mathcal{G}$-invariant Borel subsets of $\mathcal{R}^n$. The configuraton vector in $\mathcal{R}^n$ is a maximal invariant, and thus an isomorphism of $\mathcal{K}_n$ with the Borel subsets of the unit $n-1$-dimensional sphere.

It follows by the usual argument that the pivotal ratio

$$\frac{X_{n+1} - \bar{X}_n}{s_n\sqrt{1 + 1/n}}$$

is distributed as Student's t on $n-1$ degrees of freedom, independently of the observation $(X_1, \ldots, X_n)$, here regarded as a measurable function $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1}) \to (\mathcal{R}^n, \mathcal{K}_n)$. The fiducial predictive distribution is such that, given the observation $(X_1, \ldots, X_n)$, the transition to the next value is determined stochastically by

$$X_{n+1} = \bar{X}_n + s_n\sqrt{1 + 1/n}\,\epsilon_n \tag{2.4}$$

where $\epsilon_n \sim t_{n-1}$. A similar calculation for the average of m subsequent values shows that the ratio $(\bar{X}_m - \bar{X}_n)/(s_n\sqrt{1/m + 1/n})$ is distributed as t_{n-1} independently of $(X_1, \ldots, X_n)$. The limiting average $\bar{X}_\infty$ is thus finite, and its fiducial or transition distribution given $(X_1, \ldots, X_n)$ is such that $\bar{X}_\infty = \bar{X}_n + s_n\epsilon_n/\sqrt{n}$.

For $n \geq 2$, expression (2.4) tells us how to generate subsequent values of the process in a consistent manner, as if this were a process defined on Borel sets. It does not tell us how to get the process started, so in that sense the prescription is incomplete. The joint distributions are uniquely determined on the projective system $(\mathcal{R}^n, \mathcal{K}_n)$, and on this system the process is infinitely exchangeable. There does not exist an infinitely exchangeable Borel extension.

Example 8. Linear regression by fiducial prediction Let $\mathcal{U}$ be the set or population of units, let $x{:}\mathcal{U} \to \mathcal{R}$ be a given function, and let $\mathcal{X} = \text{span}\{\mathbf{1}, x\}$ in $\mathcal{R}^{\mathcal{U}}$. For each finite subset $S = \{u_1, \ldots, u_n\} \subset \mathcal{U}$ of units, let $\mathcal{X}_S$ be the restriction of $\mathcal{X}$ to S. In concrete terms, $\mathcal{X}_S \subset \mathcal{R}^S$ is the span of the columns of the linear regression model matrix for the sampled units. We write $\mathcal{X}_S \subset \mathcal{R}^S$ rather than $\mathcal{X}_n \subset \mathcal{R}^n$ because the subspace depends on the the particular subset of units sampled. Let $\mathcal{K}_S$ be the class of Borel subsets $A \subset \mathcal{R}^S$ such that $A + \mathcal{X}_S = A$.

The fiducial linear regression model is a process in the orthodox sense, consisting of a probability distribution P_S defined on $(\mathcal{R}^S, \mathcal{K}_S)$, one such distribution for each finite subset

of units. For each $S \subset S'$, the coordinate projection map $T\colon \mathcal{R}^{S'} \to \mathcal{R}^S$ that deletes those components not in S' is $\mathcal{K}$-measurable, and $P_S(A) = P_{S'}(T^{-1}A)$, so P_S is the marginal distribution of $P_{S'}$. The particular process considered here is defined by

$$P_S(A) = N_n(0, I_n)(A) \tag{2.5}$$

for a sample S containing n units and $A \in \mathcal{K}_S$. In other words, the probability assigned by P_S to $A \in \mathcal{K}_S$ is the same as the probability assigned by the standard Gaussian distribution.

Let $\mathcal{X}_S^0$ be the vector space of contrasts, i.e. linear functionals taking the value zero on $\mathcal{X}_S$. For each contrast c, the linear combination $c^{\mathrm{T}}Y$ is zero-mean Gaussian with variance $c^{\mathrm{T}}c$. Let $\bar{Y}_n, \bar{x}_n$ be the sample means, and $\hat{\beta}$ the least-squares regression coefficient, all computed on the sampled units S. Now consider a new unit such that $x(u^*) = x^*$, and let $S' = S \cup \{u^*\}$ be the extended sample. The linear combination

$$Y(u^*) - \bar{Y}_n - \hat{\beta}(x^* - \bar{x}_n)$$

is a contrast in $\mathcal{X}_{S'}^0$ distributed normally with zero mean and variance $1+1/n+(x^*-\bar{x}_n)^2/s_x^2$, and independent of the observed value, i.e. independent of the coordinate projection $(\mathcal{R}^{S'}, \mathcal{K}_{S'}) \to (\mathcal{R}^S, \mathcal{K}_S)$. The fiducial prediction for $Y(u^*)$ is thus Gaussian with mean equal to the best linear predictor $\bar{Y}_n + \hat{\beta}(x^* - \bar{x}_n)$ and variance shown above.

This example may be modified in an obvious way by restricting each P_S to sets that are also scale-invariant, i.e. by requiring $A + \mathcal{X}_S = A$ and $\lambda A = A$ for each real $\lambda > 0$. The translation-scale group consisting of scalar multiples plus elements of $\mathcal{X} \subset \mathcal{R}^{\mathcal{U}}$ acts on the spaces $\mathcal{R}^S$ in the obvious way. The group acts component-wise in the sense that coordinate deletion commutes with group actions, so the spaces $(\mathcal{R}^S, \mathcal{K}_S)$ determine a projective system. Definition (2.5) is now understood in the sense of scale-invariant sets only. A modification of the preceding argument along the lines of example 8 leads to the conclusion that the standardized linear combination

$$\frac{Y(u^*) - \bar{Y}_n - \hat{\beta}(x^* - \bar{x}_n)}{s_n\sqrt{1 + 1/n + (x^* - \bar{x}_n)^2/s_x^2}}$$

is a Borel variable distributed as t_{n-2} and independent of the observation. The same argument shows that the average of m units all with the same covariate value is such that

$$\frac{\bar{Y}_m(u^*) - \bar{Y}_n - \hat{\beta}(x^* - \bar{x}_n)}{s_n\sqrt{1/m + 1/n + (x^* - \bar{x}_n)^2/s_x^2}}$$

is also distributed as Student's t_{n-2} independent of the observation.

In the case of the translation-scale group, we could define P_S to be the restriction of a different Borel model such as

$$P_S(A) = N_n\big(\alpha\mathbf{1}_n + \beta x,\, \sigma^2 I_n + \tau_0^2 J_n + \tau_1^2(x\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n x^{\mathrm{T}}) + \tau_2^2 x x^{\mathrm{T}}\big)(A)$$

or a proper mixture of these. However, all of these Borel models have the same restriction, i.e. they assign the same probabilities to sets in $\mathcal{K}_S$. In other words, sets $A \in \mathcal{K}_S$ have the same probability for each parameter value $(\alpha, \beta, \sigma, \tau)$, so the fiducial conclusions are unaffected. In that sense, Fisher's goal of inferential statements without committing to a prior distribution on the parameter space is achieved.

Example 9. A partition process In the peceding examples a group $\mathcal{G}$ of transformations $\mathcal{R} \to \mathcal{R}$ also acts component-wise $\mathcal{R}^n \to \mathcal{R}^n$, and $\mathcal{K}_n$ is infinite for sufficiently large n. In this section, $\mathcal{G}$ is the group of invertible Borel-measurable transformations $\mathcal{R} \to \mathcal{R}$, so each orbit in $\mathcal{R}^n$ is a partition of the index set $\{1, \ldots, n\}$. Thus, for $n = 2$ the two orbits are

$$12 = \{(x_1, x_2) \mid x_1 = x_2\}$$
$$1|2 = \{(x_1, x_2) \mid x_1 \neq x_2\}$$

For $n = 3$ there are five orbits, one in which all three components are equal, three in which exactly two components are equal, and one in which all three components are distinct. In general, each orbit in $\mathcal{R}^n$ is determined by an equivalence relation on $\{1, \ldots, n\}$, a symmetric binary matrix such that $E_{ij} = 1$ if $x_i = x_j$ and zero otherwise. The set $\mathcal{E}_n$ of group orbits in $\mathcal{R}^n$ is the finite set of partitions of $\{1, \ldots, n\}$. Each $\mathcal{G}$-invariant Borel set in $\mathcal{R}^n$ is a union of orbits, so $\mathcal{K}_n$ is in 1–1 correspondence with the power set of $\mathcal{E}_n$. Accordingly, a process on $\{(\mathcal{R}^n, \mathcal{K}_n)\}$ is determined by a family of probability distributions p_n on $\mathcal{E}_n$ such that p_n is the marginal distribution of p_{n+1} under deletion of the last element.

Consider the Poisson-Dirichlet model (Kingman, 1993, ch. 9) with index λ for which the distribution on orbits is given by the Ewens formula

$$p_n(E) = \frac{\Gamma(\lambda)\ \lambda^{|E|}}{\Gamma(n+\lambda)} \prod_{b \in E} (|b| - 1)!$$

(Ewens, 1972). The product runs over the blocks of E, $|b|$ denotes block size, and $|E|$ is the number of elements (blocks) in E. It may be verified that p_n is the marginal distribution of p_{n+1}, and that each p_n is invariant under permutation of elements, so the process is exchangeable.

Suppose now that $n = 3$ and that the observation is $(3.1, 2.4, 3.1)$, regarded as a point in $(\mathcal{R}^3, \mathcal{K}_3)$. How do we predict the next value? The observation corresponds to the partition $13|2$, and the conditional distribution on $\mathcal{E}_4$ given that $E_3 = 13|2$ is as follows:

$$p_4(134|2 \mid E_3) = 2/(3+\lambda); \quad p_4(13|24 \mid E_3) = 1/(3+\lambda); \quad p_4(13|2|4 \mid E_3) = \lambda/(3+\lambda),$$

with zero mass on the remaining points. The fiducial distribution for X_4 is thus 3.1 with probability $2/(3+\lambda)$, 2.4 with probability $1/(3+\lambda)$, and an indeterminate other value with probability $\lambda/(3+\lambda)$. For general n, the fiducial distribution is a weighted mixture of the empirical distribution with weight $n/(n+\lambda)$ and an indeterminate other value with weight $\lambda/(n+\lambda)$. This may seem to be an odd sort of conclusion, but it is nothing more than a re-statement of the conditional distribution on $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1})$ given the observed value in $(\mathcal{R}^n, \mathcal{K}_n)$.

Example 10. Ordered partitions The preceding example may be modified by choosing a smaller group, thereby increasing the number of orbits. Let $\mathcal{G}$ be the group of continuous strictly increasing transformations $\mathcal{R} \to \mathcal{R}$. An orbit in $\mathcal{R}^n$ is an ordered partition of $\{1, \ldots, n\}$. The orbit labelled $13|24$ is the set of points in $\mathcal{R}^4$ for which $x_1 = x_3 < x_2 = x_4$, which is distinct from the orbit $24|13$. If $\mathcal{O}_n$ is the set of orbits in $\mathcal{R}^n$, we find that $\mathcal{O}_2$ contains three orbits, $\mathcal{O}_3$ has 13 orbits, $\mathcal{O}_4$ has 75, and so on. Under the Poisson-Dirichlet model, the distribution on orbits is invariant under permutation of elements, and thus

invariant under permutation of blocks. The probability of the ordered permutation $E \in \mathcal{O}_n$ is given by

$$q_n(E) = \frac{\Gamma(\lambda)\ \lambda^{|E|}}{\Gamma(n+\lambda)\,|E|!} \prod_{b \in E} (|b|-1)!.$$

The observation $x = (3.1, 2.4, 3.1)$ in the preceding section corresponds to the ordered partition $2|13$. Given this orbit, the conditional distribution on $\mathcal{O}_4$ is concentrated on the five ordered partitions

$$24|13, \quad 2|134, \quad 4|2|13, \quad 2|4|13, \quad 2|13|4$$

with weights proportional to $(1, 2, \lambda/3, \lambda/3, \lambda/3)$ respectively. For general n, the fiducial distribution for X_{n+1} is a weighted mixture of the empirical distribution with weight $n/(n+\lambda)$, the remainder being distributed with equal mass on the $n+1$ open intervals determined by the ordered response values.

In the preceding two examples, we observe that the fiducial distribution given the observed value is a probability distribution. Formally, it is a probability distribution on $(\mathcal{R}^{n+1}, \mathcal{K}_{n+1})$ with positive mass on a rather small class of subsets.

Example 11. Multivariate exchangeable model We consider here an exchangeable model, similar to examples 6 and 7, in which each component Y_i takes values in $\mathcal{R}^q$. In the conventional parametric version $Y_i \sim N_q(\mu, \Sigma)$ with independent components. Two fiducial versions of this model are considered, one in which $\mathcal{G} = GA(\mathcal{R}^q)$ is the general affine group with composition $[a, b][c, d] = [a+bc, bd]$, where $a, c \in \mathcal{R}^q$ and b, d are invertible matrices. In the second fiducial model $\mathcal{G}$ is the sub-group of affine lower triangular transformations with positive diagonal values. In each case the group acts transitively on the parameter space, so all parameter points determine the same probabilities for sets in $\mathcal{K}$. The probabilities for the Gaussian fiducial model are defined by

$$P_n(A) = N_{nq}(0, I_{nq})(A)$$

for $\mathcal{G}$-invariant Borel sets A.

Assume that $n > q$ and that we aim to predict Y_{n+1} having observed $(Y_1, \ldots, Y_n)$ as a point in $\mathcal{R}^{nq}$ and as an event in $\mathcal{K}_n$. To understand the orbits of the general affine group, consider the transformation $\hat{g}\colon \mathcal{R}^{nq} \to \mathcal{G}$ with $\hat{g} = [g_1, g_2]$

$$\hat{g}_1 = y_1, \quad \hat{g}_2 = [y_2 - y_1, \ldots, y_{q+1} - y_1],$$

in which $\hat{g}_2$ is non-singular with probability one. It may be verified that $\hat{g}(hy) = h\hat{g}(y)$ for each $h \in \mathcal{G}$, so $\hat{g}$ is equivariant. From this, we see that

$$\hat{g}^{-1}(y_1, \ldots, y_{n+1}) = (a_1, \ldots, a_{n+1}) = (0, e_1, \ldots, e_q, a_{q+2}, \ldots, a_{n+1})$$

is a maximal invariant in which $[e_1, \ldots, e_q]$ is the identity matrix, and $a_{q+2}, \ldots, a_{n+1}$ are points in $\mathcal{R}^q$. The σ-algebra generated by the observation is the σ-algebra generated by $(a_{q+2}, \ldots, a_n)$, so the exercise for prediction is to compute the conditional distribution of $a_{n+1} = \hat{g}^{-1} Y_{n+1}$ given $(a_{q+2}, \ldots, a_n)$. In principle, this is a straightforward exercise because

the fiducial model determines the joint distribution of $(a_{q+2}, \ldots, a_{n+1})$ as a multivariate Cauchy distribution.

The lower triangular version of the preceding model may be tackled in a similar way, and the conclusions are different because the group is different. Let M be the matrix $[y_2 - y_1, \ldots, y_{q+1} - y_1]$, and let $MM^{\mathrm{T}} = LL^{\mathrm{T}}$ be the Cholesky decomposition in which L is a lower triangular matrix with positive diagonal entries. Now take $\hat{g} = [y_1, L]$, so that $\hat{g}$ is equivariant, and $\hat{g}^{-1}y$ is a maximal invariant as before. The exercise for prediction is to compute the conditional distribution of $\hat{g}^{-1}Y_{n+1}$ given $\hat{g}^{-1}(Y_1, \ldots, Y_n)$.

3. Transformation of fiducial variables

Formally, a fiducial distribution is a pair consisting of a group $\mathcal{G}$ and a probability distribution $(\mathcal{S}, \mathcal{K}, P)$. The group acts on the space $\mathcal{S}$, $\mathcal{K}$ is a σ-algebra of $\mathcal{G}$-invariant subsets, and P is a probability distribution on $\mathcal{K}$. In the construction of a real-valued fiducial process where $\mathcal{S} = \mathcal{R}^n$, it was observed that an arbitrary transformation $\mathcal{R}^{n+1} \to \mathcal{R}^n$, even an arbitrary linear transformation, is not necessarily $\mathcal{K}$-measurable. The reason that coordinate deletion is $\mathcal{K}$-measurable is that this operation commutes with the group actions on $\mathcal{R}^n$ and $\mathcal{R}^{n+1}$.

In the Bayesian analysis that follows, $\mathcal{G}$ is the parameter space, and the conclusion involves a fiducial distribution on $\mathcal{S} = \mathcal{G} \times \mathcal{G}$. It is of some interest to know what sorts of transformations are permitted for such distributions. Since a fiducial distribution is simply a probability distribution on a slightly unfamiliar σ-algebra, the answer to this question is solely a matter of measurability of transformations. We consider here the general case in which $\mathcal{S}, \mathcal{S}'$ are two sets on which the group acts. If P is a distribution on $(\mathcal{S}, \mathcal{K})$ and T is a measurable transformation $(\mathcal{S}, \mathcal{K}) \to (\mathcal{S}', \mathcal{K}')$, then $TP = PT^{-1}$ is the marginal distribution on $(\mathcal{S}', \mathcal{K}')$.

Up to this point, we have not distinguished between the group element $g \in \mathcal{G}$ and the action $g\colon \mathcal{R}^n \to \mathcal{R}^n$. Implicitly, the action was taken to be faithful, meaning that distinct group elements give rise to distinct transformations. In general, the action need not be faithful, so several distinct group elements may generate the same action. To make this aspect clear in the notation, we write $\rho(g)\colon \mathcal{S} \to \mathcal{S}$ and $\rho'(g)\colon \mathcal{S}' \to \mathcal{S}'$ for the two actions corresponding to the same group element g. By definition, $\rho(1) = 1$ and $\rho(gh) = \rho(g)\rho(h)$ preserving identity and composition, so $\rho(g^{-1}) = \rho(g)^{-1}$ is invertible. If $\rho(g) = 1$ for every g, the action is said to be trivial, and $\mathcal{K}$ is the σ-algebra of Borel sets. Otherwise $\mathcal{K}, \mathcal{K}'$ are the σ-algebras of $\mathcal{G}$-invariant Borel sets.

A Borel-measurable transformation $T\colon \mathcal{S} \to \mathcal{S}'$ is $\mathcal{K}$-measurable if and only if, for each $g \in \mathcal{G}$ and $x \in \mathcal{S}$, the points Tx and $T\rho(g)x$ are on the same orbit in $\mathcal{S}'$. Equivalently, for each $g \in \mathcal{G}$ there exists a $g' \in \mathcal{G}$ such that $T\rho(g) = \rho'(g')T$. To see why the condition is sufficient, let $A \in \mathcal{K}'$. Then

$$\rho(g^{-1})T^{-1}A = (T\rho(g))^{-1}A = (\rho'(g')T)^{-1}A = T^{-1}\rho'(g'^{-1})A = T^{-1}A$$

showing that the inverse image $T^{-1}A$ is $\mathcal{G}$-invariant and thus in $\mathcal{K}$.

To see why the condition is necessary, suppose that there exists a point $x \in \mathcal{S}$ and a group element g such that $y = Tx$ and $y' = T\rho(g)x$ are on different orbits. The orbits $\mathcal{O}(y)$ and $\mathcal{O}(y')$ are non-overlapping sets in $\mathcal{K}'$, so the inverse images $T^{-1}(\mathcal{O}y)$ and $T^{-1}(\mathcal{O}y')$

are non-overlapping subsets of $\mathcal{S}$. However, $x \in T^{-1}(\mathcal{O}y)$ while $\rho(g)x \in T^{-1}(\mathcal{O}y')$ implies $\rho(g)x \notin T^{-1}(\mathcal{O}y)$, so $T^{-1}(\mathcal{O}y)$ is not $\mathcal{G}$-invariant and thus not in $\mathcal{K}$. In other words, T is not $\mathcal{K}$-measurable.

If the measurability condition is satisfied by $g' = g$, i.e. if $T\rho(g) = \rho'(g)T$, we say that T commutes with the group actions. Commutativity is sufficient but not necessary for measurability. Even if both group actions are faithful, (i.e. $\rho(g) = \rho(h)$ implies $g = h$ and the same for ρ'), the two conditions are not equivalent. For an example, let $\mathcal{G}$ be the general affine group acting on itself and on the real line in the usual way, $(g_1, g_2)(a, b) = (g_1 + g_2 a, g_2 b)$ and $(g_1, g_2)y = g_1 + g_2 y$. The map $T\colon (a, b) \mapsto a + 3$ from $\mathcal{G}$ to $\mathcal{R}$ does not commute with the group actions, but the condition $T\rho(g) = \rho'(g')T$ is satisfied by $g' = (g_1 + 3 - 3g_2, g_2)$. Now let $\mathcal{S} = \mathcal{G} \times \mathcal{G}$ and $\mathcal{S}' = \mathcal{R}^2$ on which the group acts componentwise, let $\mathcal{K}, \mathcal{K}'$ be the $\mathcal{G}$-invariant Borel sets, and let $T^2\colon \mathcal{S} \to \mathcal{S}'$ also be defined componentwise. Since $T\rho(g) = \rho'(g')T$, the extended function T^2 satisfies the measurability condition even though it does not commute with the group actions.

Consider the case where $\rho(g) = 1$ is the trivial group action on $\mathcal{S} = \mathcal{R}^n$, and ρ' is a non-trivial action on the same space. Then $\mathcal{K}$ is the class of Borel sets and $K' \subset \mathcal{K}$ is a proper sub-algebra. The identity map $\mathcal{S} \to \mathcal{S}$ does not commute with the group actions, but the measurability condition is satisfied by $g' = 1$ for each g. The identity map acts on probability distributions by restriction to $\mathcal{G}$-invariant sets.

For a second example, suppose $\mathcal{S} = \mathcal{G} \times \mathcal{G}$ on which the group acts componentwise by composition on the left, and let $\mathcal{S}' = \mathcal{G}$ with the trivial group action, so that $\mathcal{K}'$ is the class of Borel sets. Then the map $(g_1, g_2) \mapsto g_1^{-1} g_2$ is $\mathcal{K}$-measurable whereas $(g_1, g_2) \mapsto g_1$ is not.

For a third example, let $\mathcal{G} = \mathcal{C}$ be the non-zero complex numbers with multiplication, and let $T\colon z \mapsto |z|$ be a mapping from the group into the positive real numbers. The group also acts on $\mathcal{R}^+$ by multiplication by $\rho'(g) = |g|$, and we find that $(Tg)(z) = |gz| = \rho'(g)Tz$. Thus T commutes with the group actions on $\mathcal{C}$ and $\mathcal{R}^+$. Let $\mathcal{S} = \mathcal{G} \times \mathcal{G}$, $\mathcal{S}' = \mathcal{R}^+ \times \mathcal{R}^+$ on which the group acts component-wise, and let $\mathcal{K}, \mathcal{K}'$ be the associated $\mathcal{G}$-invariant σ-algebras. Then $T^2\colon (z_1, z_2) \mapsto (|z_1|, |z_2|)$ also commutes with the group actions, so T^2 is a measurable transformation $(\mathcal{S}, \mathcal{K}) \to (\mathcal{S}', \mathcal{K}')$. What this means is that a fiducial distribution $(\mathcal{S}, \mathcal{K}, P)$ for a complex-valued parameter or random variable θ induces a fiducial distribution $(\mathcal{S}', \mathcal{K}', PT^{-1})$ for the sub-parameter $|\theta|$, and the same is true for the complementary function $\theta/|\theta|$.

By way of contrast, consider the map $T\colon z \mapsto \Re(z)$ as a map $\mathcal{C} \to \mathcal{R}$, and suppose that $\rho'(g) = |g|$ as before. In this case T does not commute with the group actions, and the weaker condition also fails. The map $(\theta, \theta') \mapsto (\Re(\theta), \Re(\theta'))$ is not $\mathcal{K}$-measurable. In other words, a fiducial distribution for a complex-valued parameter does not determine a fiducial distribution for either the real or imaginary parts.

The posterior distribution in a Bayesian fiducial model is a pair consisting of a reference point $\hat{\theta} \in \mathcal{G}$ together with a fiducial distribution on $\mathcal{G}$-invariant Borel sets in $\mathcal{G} \times \mathcal{G}$. The additional structure provided by the reference point means that any transformation of posterior fiducial distributions must satisfy further properties in addition to measurability. The additional condition is simply that T must commute with the group actions. This restriction helps to explain why fiducial distributions do not exist for arbitrary sub-parameters, but only for natural sub-parameters. See Savage (1976, p. 467).

4. Parametric Bayesian inference for fiducial models

4.1 *General remarks*

Every fiducial model is associated with a group $\mathcal{G}$, and if $\mathcal{G}$ acts transitively the parameter space the restriction to $\mathcal{G}$-invariant Borel sets has the effect of eliminating the dependence on the parameter. In that sense a fiducial model has a parameter space consisting of a single point only, so Bayesian inference in the conventional sense is necessarily trivial. Nonetheless, it is sometimes convenient to re-introduce the group as a parameter space in the conventional sense of a set used to index events in the tail σ-algebra of the process. To do so, we merely interpret the conventional model as a joint distribution on $\mathcal{G}$-invariant Borel sets in $\mathcal{R}^n \times \Theta$ with $\Theta \equiv \mathcal{G}$. In conventional terms, the product

$$f_n(y;\theta)\,\pi(\theta)$$

is the joint density at the point $(y,\theta) \in \mathcal{R}^n \times \mathcal{G}$ with respect to some suitable product measure $dy\,d\theta$. The mapping $(y,\theta) \mapsto \theta^{-1}y$ from $\mathcal{R}^n \times \mathcal{G}$ into $\mathcal{R}^n$ is a maximal invariant. If $\mathcal{K}_n^{\mathcal{G}}$ denotes the set of $\mathcal{G}$-invariant Borel sets, the same mapping is also measurable $(\mathcal{R}^n \times \mathcal{G}, \mathcal{K}_n^{\mathcal{G}}) \to (\mathcal{R}^n, \mathcal{B}_n)$. In fact $\mathcal{K}_n^{\mathcal{G}} \cong \mathcal{B}_n$ are isomorphic. In other words, each conventional joint distribution on the Borel subsets of $\mathcal{R}^n$ determines a $\mathcal{G}$-invariant distribution on $\mathcal{R}^n \times \mathcal{G}$, and vice-versa.

By restricting the joint distribution to $\mathcal{G}$-invariant subsets, fiducial projection eliminates the effect of the prior distribution without eliminating the parameter space. The observation corresponds to the projection $(y,\theta) \mapsto y$ as a measurable map $(\mathcal{R}^n \times \mathcal{G}, \mathcal{K}_n^{\mathcal{G}}) \to (\mathcal{R}^n, \mathcal{K}_n)$. Under certain conditions, the posterior or conditional distribution given the observation is a pair consisting of a reference point $\hat{g} \in \mathcal{G}$ together with the conditional distribution of the pair $(\hat{\theta},\theta)$ defined on $\mathcal{G}$-invariant subsets of $\mathcal{G} \times \mathcal{G}$. equivalently, the conclusion may be expressed as a point estimate $\hat{g}$ together with the conditional predictive distribution of $\hat{\theta}^{-1}\theta$ defined onBorel subsets of $\mathcal{G}$. That is to say, the traditional fiducial solution is obtained by direct application of Bayes's theorem to the fiducial model defined on $\mathcal{G}$-invariant sets.

4.2 *Bayesian fiducial distribution*

Let $\mathcal{G}$ be a topological group and let $(\mathcal{G}^2, \mathcal{K}_2)$ be the measurable space consisting of $\mathcal{G}$-invariant Borel sets. That is to say, an element $g \in \mathcal{G}$ acts on the set $\mathcal{G} \times \mathcal{G}$ by composition on the left

$$g\colon (a,b) \mapsto (ga, gb)$$

and the sets in $\mathcal{K}_2$ are such that $gA = A$ for each $g \in \mathcal{G}$. Equivalently, the map given by $h\colon (a,b) \mapsto a^{-1}b$ is a maximal invariant, and $h\colon (\mathcal{G}^2, \mathcal{K}_2) \to (\mathcal{G}, \mathcal{B})$ is an isomorphism of $\mathcal{K}_2$ with the Borel subsets of $\mathcal{G}$. Consequently, each probability distribution on $(\mathcal{G}, \mathcal{B})$ determines a probability distribution on $(\mathcal{G}^2, \mathcal{K}_2)$ and vice-versa.

Under suitable conditions, the Bayesian analysis leads to a conclusion in the form of a fiducial distribution unaffected by the choice of prior distribution. A fiducial distribution is an ordered pair consisting of an element $\hat{g} \in \mathcal{G}$ together with a probability distribution $(\mathcal{G}^2, \mathcal{K}_2, Q)$ or the equivalent probability distribution Qh^{-1} on $(\mathcal{G}, \mathcal{B})$. In conventional terms, $\hat{g}$ is a point estimate, θ is the parameter, and the ordered pair $(\hat{g},\theta)$ is a random variable taking values in $(\mathcal{G}^2, \mathcal{K}_2)$. The conditional distribution of $(\hat{g},\theta)$ given the data is the fiducial distribution on which probabilistic conclusions are based. The conclusion is

usually expressed in the ordered-pair form consisting of the point estimate $\hat{g}$ together with the conditional error distribution Qh^{-1} of the pivotal statistic $\hat{g}^{-1}\theta$.

4.3 *Examples*

Three examples are studied here, all having independednt and identically distributed components when expressed in the conventional form as distributions on Borel sets. The three groups are translation, location-scale and fractional linear, and in each case the group serves as the parameter space in the Bayesian fiducial model.

Example 1. Translation model Consider the conventional Borel model in which $\Theta = \mathcal{R}$, and for each $\theta \in \Theta$, the sequence $Y_1, Y_2, \ldots$ is independent and identically distributed with density $f(\cdot - \theta)$. Let π be the prior density, so the joint density at (y, θ) in $\mathcal{R}^n \times \Theta$ is

$$f_n(y, \theta)\, dy\, d\theta = \prod_{i=1}^{n} f(y_i - \theta)\, \pi(\theta)\, dy\, d\theta.$$

All subsequent calculations refer to the distribution P_n defined by restriction of F_n to the $\mathcal{G}$-invariant Borel subsets of $\mathcal{R}^n \times \Theta$. The maximal invariant is the space of linear contrasts spanned by differences such as $y_i - y_j$ and $y_i - \theta$, whereas the observation as an event in $(\mathcal{R}^n, \mathcal{K}_n)$ is determined by the differences $y_i - y_j$ alone. The major part of the Bayesian exercise consists in finding the conditional distribution given the observation, which is equivalent to finding the distribution of $y_1 - \theta$ or $\bar{y}_n - \theta$ given the differences $\{y_i - y_j\}$. The solution, first derived by Fisher (1934), is that the conditional density of $\theta - y_1$ at t is proportional to

$$h(t) \propto f(-t) f(y_2 - y_1 - t) \cdots \times f(y_n - y_1 - t).$$

The Bayesian fiducial solution consists of the arbitrary reference point $y_1 \in \mathcal{G}$ together with the conditional distribution on $(\mathcal{G}, \mathcal{B})$ with density h. The conclusion for θ is usually expressed in the translated form $h(y_1 + t)$, which is simply the normalized likelihood function.

Example 2. Affine and location-scale model Let f be a given density function on the real line, and let $\mathcal{G}$ be general affine group. Consider the conventional model with independent and identically distributed components such that $\theta^{-1} Y_1$ has density f. Given a prior distribution π, the joint density on $\mathcal{R}^n \times \mathcal{G}$ is

$$f_n(y, \theta) = L(\theta; y)\pi(\theta) = |\theta_2|^{-n} \prod_{r=1}^{n} f((y_r - \theta_1)/\theta_2)\, \pi(\theta),$$

where $L(\theta; y)$ is the likelihood function. On the assumption that response values are distinct with probability one we make a transformation to new variables as follows:

$$t_1 = (\theta_1 - y_1)/(y_2 - y_1), \quad t_2 = \theta_2/(y_2 - y_1), \quad a_r = (y_r - y_1)/(y_2 - y_1), \quad (r = 1, \ldots, n),$$

leaving the group component θ unchanged. This transformation is such that $t = (t_1, t_2)$ is the group element that transforms the reference point $\hat{\theta} = (y_1, y_2 - y_1)$ to the parameter

(θ_1, θ_2) by right composition, i.e. $\hat{\theta}t = \theta$. Furthermore, $(a_1, a_2) = (0, 1)$, $(t_1, t_2, a_3, \ldots, a_n)$ constitutes a maximal invariant, and the σ-algebra generated by a is the σ-algebra generated by the observation as a random variable taking values in $(\mathcal{R}^n, \mathcal{K}_n)$. The joint density of the transformed variables is

$$\prod_{r=1}^{n} f(t^{-1}a_r)\,|t_2|^{-n-1}\,da\,dt_1\,dt_2 \times \pi(\theta)\,d\theta_1\,d\theta_2, \qquad (**)$$

where $da = da_3 \cdots da_n$. Restriction to $\mathcal{G}$-invariant sets is equivalent to working with the maximal invariant statistic (t, a), which is distributed independently of the parameter, and thus independent of the prior. The conditional distribution given the data is determined by the equivariant reference point $\hat{\theta} = (y_1, y_2 - y_1)$ together with the conditional distribution of the group increment $t = \hat{\theta}^{-1}\theta$ given a. The conditional density of the 'error' t is proportional to

$$\prod_{r=1}^{n} f(t^{-1}a_r)\,|t_2|^{-n-1}\,dt_1\,dt_2.$$

Instead of expressing the conclusion in the form of a point estimate $\hat{\theta}$ together with an error distribution on $\mathcal{G}$, it is conventional to shift the distribution so that the origin or group identity is transferred to $\hat{\theta}$. The distribution itself is otherwise unaffected. The density at $\theta \in \mathcal{G}$ of the shifted distribution is proportional to

$$\prod_{r=1}^{n} f(\theta^{-1}y)\,|\theta_2|^{-n-1}\,d\theta_1\,d\theta_2 = L(\theta)\,d\theta_1\,d\theta_2/|\theta_2|$$

which is the usual form of the fiducial prediction for θ. While this expression is a simple and convenient summary, the unfortunate consequence is that the fiducial distribution appears to be a distribution on the Borel subsets of the parameter space. The fact of the matter is that the fiducial distribution is the joint distribution of $(\hat{\theta}, \theta)$ given the observation, and the joint distribution is defined on the $\mathcal{G}$-invariant Borel subsets of $\Theta \times \Theta$. Consequently, all transformations are subject to the measurability conditions described in the preceding section. In effect, only transformations that commute with the group actions are measurable.

This derivation is an orthodox but unusual application of conditional probability. The primary Fisherian goal of a conclusion unaffected by the choice of prior distribution is clearly satisfied, but the cost is that the conclusion does not emerge immediately in the form of a probability distribution on the parameter space.

For the location-scale sub-group with $\theta_2 > 0$, the preceding argument needs to be modified slightly. The simplest device is to take $\hat{\theta} = (y_1, |y_2 - y_1|)$ as the equivariant reference point, in which case $a_2 = \pm 1$ with probability one half each. The remainder of the argument is unaffected in the sense that $(\hat{\theta}^{-1}\theta, a)$ is a maximal invariant, and the σ-algebra generated by a coincides with the σ-algebra generated by the observation. The conditional density of the error $t = \hat{\theta}^{-1}\theta$ is the expression shown above, but restricted to the sub-group.

Example 3. Fractional linear models Let f be a given density function on the real line, and let $\mathcal{G}$ be the real fractional linear group $y \mapsto (ay+b)/(cy+d)$ with $ad-bc \neq 0$. In this context, the real line is the extended real line, compactified by the inclusion of a single point at infinity. The group element is a real 2×2 matrix with components $\theta = (a,b,c,d)$ and inverse $\theta^{-1} = (d,-b,-c,a)$. Proportional matrices are not distinguished, so $\mathcal{G}$ is the quotient group modulo the normal subgroup of scalar multiples. The Jacobian of the transformation $y \mapsto \theta y$ is $J_\theta(y) = |ad-bc|/(cy+d)^2$, and $J_{\theta^{-1}}(x) = |ad-bc|/(a-cx)^2$.

Consider the conventional model with independent and identically distributed components such that $\theta^{-1}Y_1$ has density f. In other words, $\mathcal{G}$ is the parameter space, and for each $\theta \in \mathcal{G}$, the marginal density of each component is

$$f(y;\theta)\,dy = f(\theta^{-1}y)J_{\theta^{-1}}(y)\,dy.$$

The likelihod function is a product of n such factors. If, for each pair of points θ, θ', equality of the distributions $f(\cdot;\theta)$ and $f(\cdot;\theta')$ implies $\theta = \theta'$, the model is said to be identifiable. Identifiability depends on f: in general, the models described here are not identifiable.

For any three distinct points (y_1, y_2, y_3) there exists a unique group element $\hat{\theta}$ that sends $(0,1,-1)$ to (y_1, y_2, y_3). The coefficients are

$$\hat{\theta} = \begin{pmatrix} y_1(y_2-y_3) & y_2(y_1-y_3)+y_3(y_1-y_2) \\ y_2-y_3 & 2y_1-y_2-y_3 \end{pmatrix}$$

Since it is unique and $g\hat{\theta}$ sends $(0,1,-1)$ to (gy_1, gy_2, gy_3), the induced group element is equivariant, meaning that $\hat{\theta}(gy) = g\hat{\theta}(y)$ for each $g \in \mathcal{G}$. Consequently, for any point $y \in \mathcal{R}^n$ whose first three components are distinct there exists an invertible transformation sending (y,θ) to $(\hat{\theta}^{-1}y, \hat{\theta}^{-1}\theta, \theta)$. As in the preceding example, the first two components determine the maximal invariant, and the σ-algebra generated by the first component is the σ-algebra generated by the observation.

The transformation from (a,t,θ) to (y,θ) such that $y = (\theta t^{-1})a = \hat{\theta}a$ is invertible with Jacobian

$$\frac{\partial(y,\theta)}{\partial(a,t,\theta)} = 2|t|^{-2}\prod_{r=1}^{n} J_{\theta t^{-1}}(a_r)$$

in which the first three components of a are $(0,1,-1)$. The group element t is taken in the form of a 2×2 matrix whose $(2,2)$-component is one, $|t| = t_1 - t_2 t_3$ and $dt = dt_1\,dt_2\,dt_3$. The density function of the transformed variables is

$$|t|^{-2}\prod_{r=1}^{n} f(t^{-1}a_r)J_{t^{-1}}(a_r)\,da\,dt \times \pi(\theta)\,d\theta.$$

As before, the joint distribution restricted to $\mathcal{G}$-invariant sets is equivalent to the marginal distribution of the maximal invariant (a,t), and the prior distribution has no effect on this calculation. The conditional distribution given the observation is determined by the conditional distribution of t given a, and the conditional density is proportional to

$$|t|^{-2}\prod_{r=1}^{n} f(t^{-1}a_r)J_{t^{-1}}(a_r)\,dt.$$

Since the observation also determines the reference point $\hat{\theta}$, one inferential conclusion takes the traditional form of an estimator together with the error distribution shown above. The error distribution relative to the reference point is the conditional distribution of the pivotal invariant $t = \hat{\theta}^{-1}\theta$, or the transition element from the reference point to the parameter. Alternatively, the density of the shifted distribution at θ is proportional to

$$|\theta|^{-2} L(\theta; y),$$

which is the fiducial distribution on the parameter space. An alternative non-probabilistic interpretation of this conclusion is available in the form of a conventional parametric model and a specific improper prior with density proportional to $|\theta|^{-2}$ on the group.

The parameter space serves in a fiduciary capacity, summarizing the information required for purposes of prediction. The fiducial predictive density for Y_{n+1} given the data $y^{(n)}$ is

$$\frac{\int_{\mathcal{G}} f(y;\theta)|\theta|^{-2} L(\theta; y^{(n)})\, d\theta}{\int |\theta|^{-2} L(\theta; y^{(n)})\, d\theta}$$

In this respect, the fiducial distribution behaves as if it were a posterior distribution obtained in the conventional manner. In general, however, a fiducial distribution may not be transformed in an arbitrary manner, so the fiducial conclusion is not equivalent to the conventional Bayesian conclusion in the form of a posterior distribution on the group. One cannot usually construct a fiducial predictive distribution for arbitrary functions such as Y_{n+1}^2, for example.

Example 4. A numerical example Let $n = 3$, and let the response values be $(-4, 1, 3)$, so $\bar{y}_n = 0$ and $s_n^2 = 13$. The task is to predict the next value, i.e. to compute the predictive distribution or fiducial density for Y_4 given the observation. In particular, we want to calculate the probabilities for the four intervals $(-\infty, -4), (-4, 1), (1, 3), (3, \infty)$. The answer depends, of course, on the model used. Four fiducial models are considered, consisting of two choices for f, normal and Cauchy, and two choices for $\mathcal{G}$, affine and fractional linear.

For the gaussian/affine model, the predictive distribution is Student's t on two degrees of freedom, centered at the sample mean with scale parameter $s_n\sqrt{1 + 1/n} = \sqrt{52/3}$. For the Cauchy/affine model, the fiducial density for θ is proportional to

$$L(\theta; y^{(3)})/|\theta_2| = |\theta_2|^{-1} \prod f(y_r; \theta)$$

and the predictive density at y for the next value is proportional to

$$\int_{\mathcal{G}} f(y;\theta)\, L(\theta; y^{(3)})\, d\theta_1\, d\theta_2/|\theta_2|.$$

These densities are plotted in the left panel of Fig. 1. The probabilities for the four intervals determined by the data points are

Gaussian	$0.219, 0.365, 0.143, 0.273$
Cauchy	$0.201, 0.299, 0.201, 0.299$

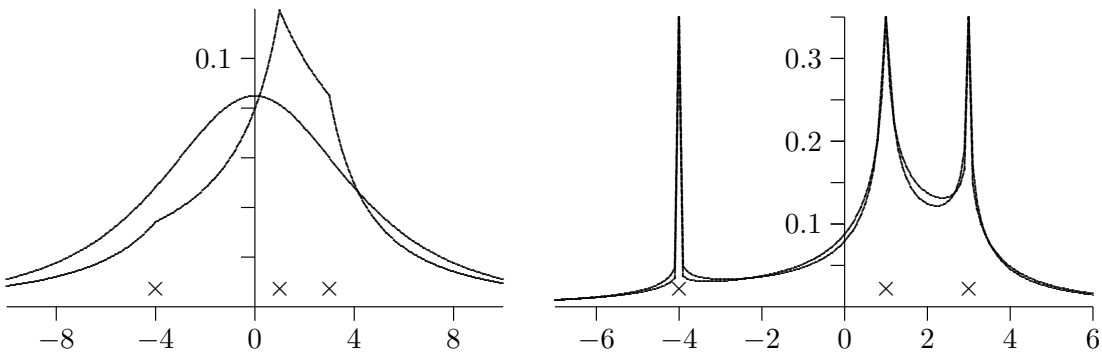

Fig. 1: Predictive densities for Y_4 given $y^{(3)} = (-4, 1, 3)$ in the Gaussian and Cauchy fiducial models: The left panel is for the loction scale group, the right panel for the fractional linear group.

The predictive probabilities for the Cauchy location-scale or affine models for $n = 3$ are such that the first and third intervals have equal probabilities, and by symmetry the same is true for the second and fourth intervals. This pattern persists regardless of the observed values, resulting in a singularity if any two values coincide. In the symmetric case, if $y_2 = (y_1 + y_3)/2$, each interval has probability 1/4 under the Cauchy model, whereas the Gaussian predictive probabilities are 0.261 for the internal intervals.

For the Cauchy model with the fractional linear group, the group element

$$\hat{\theta}^{-1} = \begin{pmatrix} y_2 - y_1 & -y_3(y_2 - y_1) \\ y_3 - y_1 & -y_2(y_3 - y_1) \end{pmatrix}$$

sends (y_1, y_2, y_3) to $(1, \infty, 0)$. The cross-ratio

$$X = \hat{\theta}^{-1} Y_4 = (Y_4 - Y_3)(Y_2 - Y_1)/\big((Y_4 - Y_2)(Y_3 - Y_1)\big)$$

is a maximal invariant whose distribution on $(\mathcal{R}, \mathcal{B})$ is given by (2.1). Since X is distributed independently of the observation, and $\hat{\theta}$ is determined by the observation, the predictive distribution for Y_4 is obtained by transforming the distribution (2.1) by the group element $\hat{\theta}$ as in (2.2). For $n = 3$, the predictive density has spikes at the observed values, as illustrated in Fig. 1b. The extreme irregular behaviour of these densities is a small-sample phenomenon.

For any exchangeable sequence, invariance under permutation implies that the cross-ratio has a distribution that is invariant under the finite group of transformations

$$x, \quad 1/x, \quad 1 - x, \quad 1/(1 - x) \quad (x - 1)/x, \quad x/(x - 1).$$

Consequently the distribution is symmetric about $\frac{1}{2}$, and each of the intervals

$$(-\infty, -1],\ (-1, 0],\ (0, \tfrac{1}{2}], (\tfrac{1}{2}, 1],\ (1, 2],\ (2, \infty]$$

has probability 1/6. The distribution on $(0, \frac{1}{2}]$ determines the distribution on the remaining intervals. In the predictive distribution, each cyclic interval determined by the observed values has probability 1/3.

It has been observed empirically that, for continuously distributed independent components, the density of the cross-ratio is invariably decreasing on $(0, \frac{1}{2})$ with an integrable singularity at the origin. Regardless of the parent distribution, the density at $0 < x < 1$ is approximately $f(x)^\gamma$ where $\gamma \geq 1$ and f is the density (2.1). This is certainly not an exact expression, but it is more than adequate as an approximation. For the normal distribution $\gamma \simeq 3/2$, and $\gamma \leq 3/2$ for the entire t-family provided that the index exceeds 0.5. The predictive densities for the Gaussian and Cauchy fractional linear models are virtually indistinguishable.

The family of distributions generated by fractional linear transformation of the Cauchy distribution is the same as the family generated by the location-scale sub-group, so the correspondence between group element and distribution is not 1–1. To each distribution there correspond infinitely many group elements, so the model indexed by the group is not identifiable. In other words, there are at least two distinct fiducial models corresponding to the conventional two-parameter Cauchy model, and these models give rise to different predictions. The predictive density shown in Fig. 2 could also be obtained by using the improper prior $d\mu\, d\sigma/\sigma^2$ in the conventional two-parameter Cauchy model.

5. Predictive distribution and confidence distribution

In the examples described in the preceding section, the observation determines a reference point $\hat{\theta} \in \mathcal{G}$ together with the conditional distribution of either $(\hat{\theta}, \theta)$ or $\hat{\theta}^{-1}\theta$. Denote by P the conditional distribution of the pair $(\hat{\theta}, \theta)$ on $(\mathcal{G}^2, \mathcal{K}_2)$, and by Q the conditional distribution of the pivotal statistic $\hat{\theta}^{-1}\theta$ on $(\mathcal{G}, \mathcal{B})$. These distributions are equivalent in the sense that the pivotal map $(\hat{\theta}, \theta) \mapsto \hat{\theta}^{-1}\theta$ is an isomorphism of σ-algebras, so P determines Q and vice-versa. With a minor change of notation, these same remarks apply to prediction.

Let $(\hat{\theta}, Q)$ be the reference point and conditional distribution of the pivotal statistic $\hat{\theta}^{-1}\theta$ on Borel subsets of $\mathcal{G}$. The confidence distribution or predictive distribution for the parameter θ is the shifted distribution $Q\hat{\theta}^{-1}$ defined on the Borel subsets of the group. This is the traditional and most natural way to express the conclusions either for predicting the next value or for inference concerning parameter values determining events in the tail σ-algebra. In fact, most of the conclusions in the preceding section were expressed in this predictive form.

The group element g sends the pair $(\hat{\theta}, Q)$ to $(g\hat{\theta}, Q)$, shifting the reference point but leaving the distribution of the pivotal statistic invariant. The confidence distribution is transformed from $Q\hat{\theta}^{-1}$ to $Q\hat{\theta}^{-1}g^{-1}$, leading to a commutative diagram

$$\begin{array}{ccc} (\hat{\theta}, Q) & \longmapsto & Q\hat{\theta}^{-1} \\ g\downarrow & & g\downarrow \\ (g\hat{\theta}, Q) & \longmapsto & Q\hat{\theta}^{-1}g^{-1} \end{array}$$

in which the horizontal arrow is the map from fiducial distribution to confidence distribution. Each group element g transforms the confidence distribution in the standard manner by composition with the inverse image.

It is of interest to ask whether a confidence distribution may be transformed by transformations other than those in $\mathcal{G}$. In the general set-up, $\mathcal{S}, \mathcal{S}'$ are two spaces on which the group acts, Q is the distribution of the pivotal statistic $\hat{\theta}^{-1}Y$, $F = Q\hat{\theta}^{-1}$ is the

confidence distribution on $(\mathcal{S}, \mathcal{B})$, and $T: \mathcal{S} \to \mathcal{S}'$ is a map. The group actions are denoted by $\rho(g), \rho'(g)$. For each $g \in \mathcal{G}$, the action $\rho(g)$ sends the confidence distribution F to $F\rho(g)^{-1}$, so logical consistency demands that the action on $\mathcal{S}'$ should be compatible with that in $\mathcal{S}$. In other words, the marginal distributions FT^{-1} and $F\rho(g)^{-1}T^{-1}$ must be related by the action of the group in $\mathcal{S}'$:

$$F\rho(g)^{-1}T^{-1} = FT^{-1}\rho'(g)^{-1}.$$

or $F(T\rho(g))^{-1} = F(\rho'(g)T)^{-1}$ on Borel sets. In other words, the only transformations to be considered for confidence distributions or predictive distributions are those that commute with the group actions.

The consequences of this conclusion are far-reaching. In particular, for a spatial model whose kernel includes the constant functions, a predictive distribution for Y_{n+1} does not determine a predictive distribution for $\log Y_{n+1}$. Likewise a predictive distribution for (Y_{n+1}, Y_{n+2}) determines a predictive distribution for all linear combinations but it does not determine a predictive distribution for the ratio.

In the Behrens-Fisher two-sample problem, the parameter space consists of two pairs $(\mu, \sigma), (\mu', \sigma')$, so the natural group for a fiducial model is $GA(\mathcal{R}) \times GA(\mathcal{R})$. The Bayesian calculations described in section 4 give rise to a reference point plus a fiducial distribution on the group, and the confidence transformation produces a distribution for the parameter, defined on Borel subsets of the group. The sub-parameter $(\mu, \sigma, \mu', \sigma') \mapsto (\mu, \mu')$ is natural in the sense that the transformation commutes with the obvious group actions, so one obtains a joint confidence distribution for the two means. But the transformation $(\mu, \sigma, \mu', \sigma') \mapsto \mu - \mu'$ does not satisfy the measurability condition in section 3, so there is no confidence distribution for the mean difference. However, there is a joint confidence distribution for the variances, and a confidence distribution for the variance ratio. The situation is entirely different if the group is restricted to $\sigma = \sigma'$, and if the parameter space is similarly restricted the transformation to $\mu - \mu'$ commutes with the group actions. Fisher's approval of Behrens's solution to the two-sample problem with unequal variances, shows that the version of fiducial inference described here does not cover all that Fisher had in mind.

References

Barnard, G.A. (1963) Some logical aspects pf the fiducial argument. *J. R. Statist. Soc.* B **25**, 111–114.

Barnard, G.A. (1995) Pivotal models and the fiducial argument (with discussion). *Int. Statist. Rev.* **63**, 309–323.

Brillinger, D.R. (1962) Examples bearing on the definition of fiducial probability with a bibliography. *Ann. Math. Statist.* **33**, 1349–1355.

Cressie, N.A.C. (1993) *Statistics for Spatial Data.*

Dempster, A.P. (1963) On direct probabilities. *J. R. Statist. Soc.* B **25**, 100–110.

Eaton, M.L. and Sudderth, W.D. (1999) Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli* **5**, 833–854.

Ewens, W. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

Fisher, R.A. (1934) Two new properties of mathematical likelihood. *Proc. Roy. Soc. London* **A144**, 285–307.

Fisher, R.A. (1935) The logic of uncertain inference. *J. R. Statist. Soc.***98**, 39–54.

Fraser, D.A.S. (1968) *The Structure of Inference.* J. Wiley & Sons, New York.

Kingman, J.F.C. (1993) *Poisson Processes.* Oxford Scientific Publications.

Lindley, D.V. (1958) Fiducial distributions and Bayes's theorem. *J. R. Statist. Soc.* B **20**, 102–107.

Savage, L.J. (1976) On re-reading R.A. Fisher. *Ann. Statist.* **4**, 441–500.

Stein, M.L. (1999) *Interpolation of Spatial Data.* Springer-Verlag, New York.

Tukey, J.W. (1957) Some examples with fiducial relevance. *Ann. Math. Statist.* **28**, 687–695.

Wahba, G. (1990) *Spline Models for Observational Data.* SIAM monograph No. 59.

Zabell, S. (1992) R.A. Fisher and the fiducial argument. *Statistical Science* **7**, 369–387.