2 Exchangeability and its ramifications

A. PHILIP DAWID

2.1 Introduction

Bruno de Finetti's concept of exchangeability [14], and its associated mathematics, form one of the cornerstones of subjectivist Bayesian inference, providing Bayesians with a good reason to take seriously the frequentist's model of independent observations from a common but unknown distribution. There is indeed a touch of magic about de Finetti's famous theorem: assuming nothing more than that our attitudes to a sequence of observations would be unchanged if they were arranged in a different order, it pulls the frequentist model out of a hat.

There are many ways of understanding de Finetti's theorem, and correspondingly many ways of generalizing it. One fruitful conception emphasizes the aspect of invariance under the action of a group: that of all finite rearrangements of the sequence of variables. This leads to generalizations in which we deal with other groups, acting on other structures, generating different kinds of statistical model. This chapter presents a brief survey of some of these generalizations and their applications. It is largely a summary of work that has previously been presented elsewhere [3, 5–7, 9, 10, 12, 13].

2.2 de Finetti's Theorem

At the purely mathematical level, de Finetti's theorem (henceforth dFT) supplies a characterization of those joint distributions P for an infinite sequence $X = (X_1, X_2, \ldots)$ of random variables, all defined on the same space $\mathcal{X}_0$, having the property of *exchangeability*: for any $n = 1, 2, \ldots$ and any permutation π of $(1, 2, \ldots, n)$, the joint distribution of the permuted sequence $(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$ is exactly the same as that of the unpermuted sequence $(X_1, X_2, \ldots, X_n)$. dFT shows that, for exchangeability to hold, it is necessary and sufficient that there exist a distribution (which, to avoid confusion, we shall sometimes term a *law*) ν over the space $\mathcal{Q}$ of distributions Q on $\mathcal{X}_0$, such that, for (suitably measurable) $A \subseteq \mathcal{X} := \mathcal{X}_0^\infty$,

$$P(X \in A) = \int_{\mathcal{Q}} Q^\infty (X \in A)\, d\nu(Q) \tag{2.1}$$

where Q^∞ is the distribution of $X = (X_1, X_2, \ldots)$ when the (X_i) are independent and identically distributed, each with distribution Q. The law ν can be uniquely recovered from P as the limit, as

$n \to \infty$, of the law of the empirical distribution, ν_n, of $(X_1, \ldots, X_n)$—see Goldstein's chapter in this volume [20].

For a binary sample space $\mathcal{X}_0 = \{0, 1\}$, dFT shows that any exchangeable distribution can be obtained from the model of Bernoulli trials with probability parameter q, by mixing with respect to a probability distribution ν for $q \in [0, 1]$. For more general $\mathcal{X}_0$ the mixing is over the set of distributions Q on $\mathcal{X}_0$, which will typically be a large nonparametric class. But—important mathematical niceties aside—the story is basically the same [21].

2.2.1 Intersubjective modelling

If your subjective joint distribution for $(X_1, X_2, \ldots)$ is exchangeable, you would not care if some demon rearranged the sequence in a different order: your opinions for the rearranged sequence would be exactly the same as they were for the original ordering. (In contrast, this would *not* hold if, for example, you thought there were some sort of time trend in the original sequence.)

We can regard this exchangeability property as the natural Bayesian explication of the intuitive concept of 'repeated trials of the same phenomenon under constant conditions', which forms the basis of the frequentist approach to probability and statistics. Indeed, this Bayesian approach, starting only with the very natural judgement of exchangeability as input, logically *derives*, via dFT, the model of independent and identically distributed trials as output. By providing a deeper justification for modelling assumptions that are typically—and thoughtlessly—simply taken as obvious and not in need of deeper analysis, it thus supplies a firmer basis for frequentist statistics than is available from that theory itself.

To expand on this point, consider a bevy of Bayesians who, while holding differing subjective probability distributions for X, all agree on exchangeability. Then they will all agree on the relevance of the independent and identically distributed model—the differences between them being relegated to their varying choices for the prior law for the 'parameter' Q of the model. We can thus justify the frequentist's statistical model of independent and identically distributed variables as an *intersubjective model* [5], conjured into existence by the simple and intuitive exchangeability judgement: that rearranging the variables should have no effect on judgements about them. Since this model comprises the common part of every exchangeable judgement, and since its parameter Q can be recovered from sufficiently extensive observation, it has at least as much claim to 'objectivity' as any other conception of that elusive term.

2.3 Group invariance

Exchangeability of a joint distribution P over $\mathcal{X} = \mathcal{X}_0^\infty$ can be restated as requiring that P be *invariant* under the group of all finite permutations acting on $\mathcal{X}$. dFT shows that the set of such invariant distributions forms a convex simplex, and that the extreme points of this simplex are the independent and identically distributed distributions—so that any member of the set has a unique representation as a convex combination of these extreme points.

Interesting extensions of this characterization arise when we apply it to other transformation groups, acting on sample spaces possibly other than those of infinite sequences. Thus let X be an uncertain quantity taking values in a sample space $\mathcal{X}$, let G be a group of transformations acting on $\mathcal{X}$, and let $\mathcal{P}$ denote the set of all distributions P for X that are invariant under G, so that $P(X \in A) = P(gX \in A)$ for all $g \in G$ and $A \subseteq \mathcal{X}$. Then $\mathcal{P}$ is a convex set. Let $\mathcal{T}$ be the set of extreme points of $\mathcal{P}$, i.e. those distributions in $\mathcal{T}$ that cannot be represented as a non-trivial mixture (convex combination) of distributions in $\mathcal{P}$. The extension of (2.1) to this more general context is then: for any $P \in \mathcal{P}$ there exists a unique law (distribution) ν over $\mathcal{T}$ such that, for $A \subseteq \mathcal{X}$,

$$P(X \in A) = \int_{\mathcal{T}} \theta(X \in A)\, d\nu(\theta). \tag{2.2}$$

Here $\theta \in \mathcal{T}$ is a distribution over $\mathcal{X}$. A friendlier notation renames this to P_θ, reinterpreting θ as a label for P_θ, with $\mathcal{T}$ the set of such labels, so yielding

$$P(X \in A) = \int_{\mathcal{T}} P_\theta(X \in A)\, d\nu(\theta). \tag{2.3}$$

The statistical interpretation of this mathematical property is as follows. Suppose You have made a personal judgement that Your opinions would not be affected if You were to be presented with gX $(g \in G)$, rather than X. Then You must act as though You entertained the *statistical model* $\mathcal{P} = \{P_\theta : \theta \in \mathcal{T}\}$ for X, together with a prior distribution ν over its *parameter-space* $\mathcal{T}$. Again, if we consider the bevy of Bayesians all of whom agree in regarding invariance under G as appropriate, we can justify $\mathcal{P}$ as the intersubjective model engendered solely by this judgement of group invariance.

At this level of generality, the sample space $\mathcal{X}$ need not be a set of infinite sequences, and G need not be a permutation group. The link with frequentist understandings of modelling is then broken—but this is to very positive effect. By no longer insisting on any connection with the idea of 'repeated trials of the same phenomenon under constant conditions' this approach supplies a justification for statistical model building that is totally unavailable from the frequentist perspective.

2.3.1 Sufficiency

An added bonus of constructing statistical models through considerations of invariance is that we can use them to construct sufficient statistics: under suitable conditions a maximal invariant under the action of the group will be sufficient for the associated intersubjective model [17]. Thus for the case of exchangeability, where we observe just the first n variables $(X_1, \ldots, X_n)$, the maximal invariant under the permutation group is their *order statistic*—which is indeed sufficient for the model of all independent and identically distributed distributions for the (X_i). For a binary sample space, this reduces to the counts of 0s and 1s, which are sufficient for the Bernoulli model.

The above approach to model-building through invariance identifies the members of the intersubjective statistical model as the extreme points of the convex set of all invariant distributions. Another approach [18, 25, 26] starts by specifying what are the sufficient statistics, and suitably relating these, both algebraically and probabilistically, across different sample sizes. The set of all distributions consistent with these properties will again be a simplex, and its extreme points can be regarded as the associated statistical model—an *extreme point model*. When, as here, both approaches are possible they lead to the same model [5].

2.4 Other symmetry groups

Staying for the moment with the infinite product sample space $\mathcal{X} = \mathcal{X}_0^\infty$, we can entertain different groups acting on it.

2.4.1 Larger group

For a symmetry group G that contains the group of all finite permutations, the corresponding intersubjective model would still involve independent and identically distributed variables, but would impose additional structure on their common distribution.

One such larger symmetry group is that of all finite orthogonal transformations, where, for any $n = 1, 2, \ldots$, and any orthogonal transformation R of $\mathcal{R}^n$, the distributions of $X^n := (X_1, \ldots, X_n)^{\mathrm{T}}$ and of RX^n are judged to be the same. It was asserted by Freedman [19] and shown by Kingman [24] that a joint distribution has this property if and only if it can be expressed as a mixture, over some law for the parameter $\phi \geq 0$, of the joint distributions $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Norm}(0, \phi)$. That is, the model of independent and identically distributed normal variables with mean 0 arises from a judgement of rotational symmetry. With finitely many observations $(X_1, \ldots, X_n)$, the maximal invariant under this rotation group is $\sum_{i=1}^n X_i^2$, which is thus a sufficient statistic for this model.

Extending this result, Adrian Smith [31] showed that the independent and identically distributed normal model with both parameters unconstrained arises similarly from the assumption of invariance under the subgroup of finite orthogonal transformations that preserve the unit vector. The maximal invariant is now $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$, which is sufficient for this model.

2.5 Smaller group

Alternatively, we can consider 'restricted exchangeability', where we only require invariance under some smaller group of permutations. Often these will respect some additional structure in the sample space.

2.5.1 Partial exchangeability

Consider binary variables laid out in a semi-infinite two-way array: $(X_{ij} : i = 1, \ldots, k;$ $j = 1, 2, \ldots)$. An interpretation could be that we have a number of different coins, labelled by i, with possibly different biases, and can toss each of them over and over, with j labelling the toss. Then it could be appropriate to judge the problem invariant under any permutation of tosses j of the same coin (i.e. for fixed i), but not if we permute across the coins. This is the property of *partial exchangeability* [15]. The associated intersubjective statistical model has, as parameter, a vector $\mathbf{p} \in [0, 1]^k$, and then has all the (X_{ij}) independent, with $\text{Prob}(X_{ij} = 1 | \mathbf{p}) = p_i$. That is, we simply assign different probabilities to the different coins.

As an intermediate position between full and partial exchangeability, we can permit further invariance of the problem under permutations of the label i of the coins. This might be appropriate if the coins were taken randomly from the output of a mint whose quality control is less than perfect, so that different coins might have different biases, but in an unsystematic way. We now allow a potentially infinite number of such coins (so $i = 1, 2, \ldots$), with invariance under all finite permutations of i.

The associated intersubjective statistical model can be expressed hierarchically [6]:

1. The parameter is a distribution Π on $[0, 1]$.
2. Given Π, variables $(P_i : i = 1, 2, \ldots)$ are independent and identically distributed according to Π.
3. Given (Π and) the (P_i), the X_{ij} are independent, with

$$\text{Prob}(X_{ij} = 1) = P_i.$$

The model conditional on the P_i is thus exactly the same as for partial exchangeability; but now, corresponding to the new assumption of exchangeability of the coins, the (P_i), which were previously entirely arbitrary 'fixed effects', are themselves modelled as 'random effects', drawn independently from a common distribution.

2.5.2 Row–column exchangeability

In the above example the labelling j of the tosses has no objective meaning that carries across the coins: there is, for example, no special relationship between the third toss of coin 1 and the third toss of coin 2, and this indeed is why it can make sense to consider permuting the tosses of coin 1 while leaving those of coin 2 in place.

In other contexts there may be such correspondences of j-values across different i. Thus suppose a number of students (labelled by i) answer a number of questions (labelled by j). Then each of i and j has an intrinsic meaning across the levels of the other.

In such a case we might want to consider both students and questions as separately exchangeable. That is, our attitudes to the problem are considered unchanged when we replace the array (X_{ij}) by $X_{\rho(i)\sigma(j)}$, where ρ and σ are arbitrary finite permutations acting respectively on i and j. Note that such permutations preserve the integrity of individual students and questions.

Analysis of this problem is subtle. Aldous [1] showed that, for a doubly infinite array of binary variables, the associated intersubjective statistical model can be represented hierarchically as follows:

1. The parameter is a function $F : [0, 1]^2 \to [0, 1]$.
2. Given F, generate random (P_{ij}) in $[0, 1]$ as

$$P_{ij} = F(\alpha_i, \beta_j)$$

   where the αs and βs are all independent and identically distributed with the uniform distribution over $[0, 1]$.
3. Given (F and) the (P_{ij}), generate the (X_{ij}) independently, with

$$\text{Prob}(X_{ij} = 1) = P_{ij}.$$

We can regard α_i as a measure of the quality of student i, and β_j as a measure of the difficulty of question j. The independent and identically distributed property reflects the assumed separate exchangeability of students and of questions. Note that, for any joint distribution of this form, the sets $(X_{ij} : i \in I, j \in J)$ and $(X_{ij} : i \in I', j \in J')$ are independent of each other whenever there is both no overlap of students $(I \cap I' = \emptyset)$ and no overlap of questions $(J \cap J' = \emptyset)$. Such an array is termed *dissociated* [27]. Moreover, any dissociated row–column exchangeable array can be represented in the above form.

However, this representation, while appealing in many ways, is not ideal, in that different choices for F can lead to identical joint distributions of the (X_{ij}), so that the parameter F is not identifiable (it is in fact identifiable modulo transformations of its arguments by separate functions each preserving the uniform distribution [22, 23]). Also, for a finite data array, the maximal invariant under the permutation group (which by, the general theory, is a sufficient statistic for the model) is hard to describe.

2.5.3 Spherical matrix models

We can combine the above generalizations of exchangeability—orthogonal tranformations rather than permutations, and two-way arrays rather than sequences. Thus suppose we have a doubly infinite array (X_{ij}) of real variables whose distribution can be regarded as *spherical*, i.e. the distribution of any finite submatrix is invariant under the actions of both left- and right-multiplication by an orthogonal matrix. Then the intersubjective model comprises those spherical distributions that are also dissociated. Moreover (assuming finite second moments), any such distribution can be expressed in the form

$$X_{ij} = \lambda_0 U_{ij} + \sum_{m=1}^{\infty} \lambda_m V_{im} W_{mk}$$

where the λs are non-negative real constants with $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\sum_{m=1}^{\infty} \lambda_m^2 < \infty$, and all the Us, Vs and Ws are independent Norm$(0, 1)$ variables [1, 4]. It is rather easier to see that the group-induced sufficient statistic, based on a finite submatrix $(X_{ij} : 1 \leq i \leq I, 1 \leq j \leq J)$, is the unordered set of singular values of the matrix.

2.6 Second-order behaviour

For many purposes it is sufficient to confine attention to the joint first- and second-order moment structure of our variables. Thus if $(X_1, X_2, \ldots)$ is an exchangeable sequence with finite variance, then it is easy to see that there must exist constants μ, γ_0, γ_1 such that, for all $i \neq j$:

$$\mathbb{E}(X_i) = \mu \tag{2.4}$$

$$\text{var}(X_i) = \gamma_0 \tag{2.5}$$

$$\text{Cov}(X_i, X_j) = \gamma_1. \tag{2.6}$$

More generally, we call the sequence (X_i) *second-order exchangeable* if the properties (2.4)–(2.6) hold. This is equivalent to requiring that the mean and dispersion structure of the sequence be invariant under the group of finite permutations.

In this case, define $\phi_1 = \gamma_1$, $\phi_0 = \gamma_0 - \gamma_1$. It is readily checked that $\text{var}(\overline{X}_n) = \phi_0 + \phi_1/n$, whence, since n can be arbitrarily large, $\phi_0 \geq 0$. Also we find $\text{var}(X_n - \overline{X}_n) = (1 - 1/n)\phi_1$, so that $\phi_1 \geq 0$.

Now, for any real μ and $\phi_0, \phi_1 \geq 0$, consider uncorrelated variables $Z, Y_1, Y_2, \ldots$, with zero means, and

$$\text{var}(Z) = \phi_0$$

$$\text{var}(Y_i) = \phi_1$$

and define

$$X_i = \mu + Z + Y_i. \tag{2.7}$$

Then it is easy to see that the (X_i) satisfy (2.4)–(2.6). Moreover, we can recover Z as the mean square limit of $\overline{X}_n - \mu$ as $n \to \infty$, and then Y_i as $X_i - Z - \mu$. Hence an infinite sequence (X_i) is second-order exchangeable if and only if it can be represented as in (2.7), where $Z, Y_1, Y_2, \ldots$ are uncorrelated with zero means and $\text{var}(Y_i)$ is the same for all i. This can be regarded as a second-order (and indeed much simpler) variant of de Finetti's theorem: rather than the (X_i) being independent and identically distributed, after suitable conditioning, now they are uncorrelated with constant variance, after suitable partialling out of Z. See Goldstein [20, Section 7].

When we consider finite rather than infinite second-order exchangeable sequences, it need no longer be true that $\phi_0 \geq 0$, in which case there can be no real variable Z with $\text{var}(Z) = \phi_0$. Even so, for computing variances of linear functions of the (X_i) we can still proceed *as though* we had a representation of the form (2.7).

2.7 Extension to experimental layouts

Again, we can usefully extend the above ideas to other groups, acting on structures other than the sequence. We illustrate this here for the special case in which a number of workers, labelled by w, each operate a number of machines, labelled by m, for a number of different runs, labelled by r. However the theory extends straightforwardly to general distributive block structures [2, 11], which include most of the classical experimental layouts (in particular, all simple orthogonal block structures [28, 29]).

2.7.1 Fully random model

Let X_{mwr} be a measure of the quality of the rth run produced by worker w when operating machine m. In such a case it might sometimes be reasonable for attitudes about all the (X_{mwr}) to be invariant under the following permutations:

1. Permutations of r for a fixed combination of m and w—corresponding to exchangeability of the different runs made by a given worker on a given machine.
2. Permutations of w for fixed m—corresponding to a judgement of exchangeability of the workers with each other.
3. Permutations of m for fixed w—exchangeability of the machines one with another.

If we focus only on the dispersion structure, and assume this is invariant under the group generated by all the above permutations, then, analogous to (2.7), we obtain the following synthetic representation:

$$X_{mwr} = \mu + \alpha_m + \beta_w + \gamma_{mw} + \epsilon_{mwr} \tag{2.8}$$

where μ is a constant, and the other variables appearing on the right-hand side are mutually uncorrelated random variables, all terms involving the same Greek letter having the same variance. As long as the array is infinitely extendible in all directions, these variances will all be non-negative, and the representation (2.8) becomes a genuine equality, with the terms on the right-hand side identifiable as mean square limits of functions of the Xs.

The 'parameters' of the model (2.8) are μ and the respective variances ϕ_α, ϕ_β, ϕ_γ, ϕ_ϵ of the random terms. We can thus consider this model, where these parameters can vary freely, as the intersubjective second-order model corresponding to the imposed symmetries. We see that our simple symmetry assumption delivers, for free, the usual 'random effects' model for this layout.

We can also use the symmetries (specifically, utilizing group representation theory) to deliver for free the appropriate decomposition of data (from a finite balanced layout) into main effects and interactions:

$$X_{mwr} = X_{\ldots} + (X_{m..} - X_{\ldots}) + (X_{.w.} - X_{\ldots}) + (X_{mw.} - X_{m..} - X_{.w.} + X_{\ldots}) + (X_{mwr} - X_{mw.}) \tag{2.9}$$

(where a dot indicates an average over the range of the replaced subscript in the data). Symmetry arguments can also be invoked to specify meaningful null hypotheses. For example, one possible interpretation of the assertion that there are 'no differences between the workers' is that the problem would remain invariant under the still larger group that (in addition to permutations of the machine label m) allowed permutations of *all* the runs on a given machine, irrespective of whether or not these were produced by the same worker. This delivers the model obtained from (2.8) by omitting the terms β_w and γ_{mw}. It thus corresponds to the 'null hypothesis' $\phi_\beta = \phi_\gamma = 0$, which in turn, in the context of the data-decomposition (2.9), is equivalent to the equality of the mean squares for workers, for machine–worker interaction, and for runs—thus suggesting appropriate statistical tests. See [9] for further details.

Another use of symmetry [3] is to make predictive inferences of various kinds—for example, for the production of a machine featuring in our experiment operated on by a new worker.

2.7.2 A mixed model

However, the symmetry approach is not just a different way of deriving and thinking about known results. It can also lead to new perspectives, models and analyses.

Thus suppose, in the above example, we have a finite number M of machines, and, recognizing that these are of various different kinds, are no longer willing to assume invariance under permutations of the index m—while still retaining (second-order) exchangeability between workers w, and between the runs r for any (m, w) combination. The intersubjective model generated by these reduced symmetry assumptions is now [8]

$$X_{mwr} = \mu_m + \alpha_{mw} + \epsilon_{mwr} \tag{2.10}$$

where:

1. $\mu_1, \ldots, \mu_M$ are arbitrary constants.
2. α_{mw} is $(\boldsymbol{\alpha}_w)_m$, where $\boldsymbol{\alpha}_w$ is a random $(M \times 1)$ vector, with mean $\mathbf{0}$ and arbitrary $(M \times M)$ dispersion matrix $\boldsymbol{\Phi}_\alpha$ (the same for all w).
3. ϵ_{mwr} is a real random variable, with variance ϕ_m depending on m alone.
4. The distinct vectors $\boldsymbol{\alpha}_w$ and scalars ϵ_{mwr} are all mutually uncorrelated.

In this model, the parameters are the 'fixed effect' machine means $\mu_1, \ldots, \mu_M$, the $(M \times M)$ dispersion matrix $\boldsymbol{\Phi}_\alpha$ of the 'random across-machine worker effects', $(\boldsymbol{\alpha}_w)$, and the within-machine variances ϕ_m $(m = 1, \ldots, M)$ of the 'random run effects', (ϵ_{mwr}). All these quantities can be consistently estimated from data on the M machines and indefinitely many workers and runs.

This model, fully justified by the symmetry assumptions from which it derives, differs from standard formulations of the 'mixed model'. Once again, the symmetries can be used to guide hypothesis generation, data-analysis and predictive inference. For example, the associated data-decomposition is

$$X_{mwr} = X_{m\cdot\cdot} + (X_{mw\cdot} - X_{m\cdot\cdot}) + (X_{mwr} - X_{mw\cdot})$$

where the component terms are independent, and have a special dispersion structure induced from that of (2.10). The symmetry analysis now leads to new tests of 'no worker effect', and even to a test of exchangeability between the machines (i.e. the model considered in Section 2.7.1).

2.8 Population genetics

An application of the symmetry approach to modelling arises in population genetics. We are interested in a collection of genetic markers (loci on the genome), labelled $m = 1, \ldots, M$. For simplicity, we suppose each of these has two possible variants (alleles), coded 0 and 1. We can sample data from a population of individuals, which itself is divided into many subpopulations (e.g. different ethnic groups), labelled $s = 1, \ldots, S$. We assume random mating within each subpopulation s. For each pair (s, m) there is a *gene pool*, comprising all the genes (with values 0 or 1) at marker m possessed by all the individuals in subpopulation s. We thus have an array (X_{smg}) of binary variables, where X_{smg} is the allele of the gth gene within the gene pool for marker m in subpopulation s.

Reasonable symmetry assumptions to impose on this array are:

1. Exchangeability of the genes within each gene pool (note that, because of random mating, there is no reason to expect the relationship between the two genes comprising the genotype of the same individual to differ in any way from the relationship between two genes belonging to different individuals).
2. Exchangeability of the subpopulations.

However we do not impose exchangeability of the markers.

Any joint distribution satisfying these symmetry assumptions can be represented in the following hierarchical way:

1. Generate, by some random process, a distribution Π over the space $[0, 1]^M$.
2. Given Π, generate vectors $(\mathbf{P}_s)$ in $[0, 1]^M$, independently from distribution Π.
3. Given (Π and) the $(\mathbf{P}_s)$, generate independent binary variables X_{smg} with $P(X_{smg} = 1) = P_{sm}$ (the m-entry of $\mathbf{P}_s$).

Thus $\mathbf{P}_s$ describes the frequency distributions of the different markers, within subpopulation s.

Also, although it does not follow directly from symmetry, it might be reasonable from scientific considerations (especially if the different markers are on different chromosomes) to assume that, for a random vector $\mathbf{P} \sim \Pi$, its components (P_m) are independent (though not necessarily identically distributed). For example, we might take, for Π, a distribution in which $P_m \sim \beta(a_{m0}, a_{m1})$, independently; then Π is determined by the (a_{mj}) $(m = 1, \ldots, M; j = 0, 1)$. However this extra specialization is inessential.

The intersubjective model corresponding to the assumed symmetry properties would take as its parameter the distribution Π (or, in the above specialization, the (a_{mj})), and so be described by levels 2 and 3 of the above hierarchy. The full hierarchical model of a single Bayesian would be obtained by adding in, as level 1, a subjective prior distribution (essentially unconstrained) for Π (or for the (a_{mj})).

Note that in this intersubjective model, because of the random nature of P_{ms} we will have exchangeability, but not independence, between the genes within a common gene pool,

However, the full hierarchical model can also be deconstructed in the following, equally valid, way. We take as our 'parameter' the quantities (P_{sm}), and as our 'model' the final stage 3 of the hierarchy. We flesh out the hierarchy with a joint prior distribution for the (P_{sm}), obtained by combining levels 1 and 2 of the hierarchy. (The first stage is again essentially arbitrary, but the second is now not, so imposing constraints on the form of the joint 'prior' distribution of the (P_{sm}).) In this description, the genes within a given gene pool are now independent.

We thus see that it is not meaningful to ask whether or not the genes in a gene-pool are *really* independent—the answer depends on a somewhat arbitrary choice we have to make as to where to draw the line between model and prior. This realization takes some of the heat out of the apparently discrepant modelling assumptions made in [16] and [30], and their application to criminal identification by means of DNA profiling [12, 13].

References

[1] Aldous, David J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, **11**, 581–598.

[2] Bailey, Rosemary A. (1981). Distributive block structures and their automorphisms. In *Combinatorial Mathematics VIII* (ed. K. L. McAveny), Volume 884 of *Lecture Notes in Mathematics*. Springer Verlag, Berlin.

[3] Dawid, A. Philip (1977). Invariant distributions and analysis of variance models. *Biometrika*, **64**, 291–297.

[4] Dawid, A. Philip (1978). Extendibility of spherical matrix distributions. *Journal of Multivariate Analysis*, **8**, 559–566.

[5] Dawid, A. Philip (1982). Intersubjective statistical models. In *Exchangeability in Probability and Statistics* (ed. G. Koch and F. Spizzichino), pp. 217–232. North-Holland Publishing Company, Amsterdam.

[6] Dawid, A. Philip (1985). Probability, symmetry and frequency. *British Journal for the Philosophy of Science*, **36**, 107–128.

[7] Dawid, A. Philip (1986). A Bayesian view of statistical modelling. In *Bayesian Inference and Decision Techniques* (ed. P. K. Goel and A. Zellner), Chapter 25, pp. 391–404. Elsevier Science Publishers B.V. (North-Holland), Amsterdam.

[8] Dawid, A. Philip (1986). Symmetry analysis of the mixed model. Research Report 53, Department of Statistical Science, University College London.

[9] Dawid, A. Philip (1988). Symmetry models and hypotheses for structured data layouts. *Journal of the Royal Statistical Society, Series B*, **50**, 1–34.

[10] Dawid, A. Philip (1993). Taking prediction seriously. *Bulletin of the International Statistical Institute*, **55**(3), 3–13.

[11] Dawid, A. Philip (1994). Distributive block structures: Mathematical properties. Research Report 133, Department of Statistical Science, University College London.

[12] Dawid, A. Philip (1997). Modelling issues in forensic inference. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pp. 182–186. American Statistical Association (Alexandria, VA).

[13] Dawid, A. Philip and Pueschel, John (1999). Hierarchical models for DNA profiling using heterogeneous databases (with Discussion). In *Bayesian Statistics 6* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), Oxford, pp. 187–212. Oxford University Press.

[14] de Finetti, Bruno (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7**, 1–68. English translation "Foresight: Its logical laws, its subjective sources", in *Studies in Subjective Probability* (1964) (H. E. Kyburg and H. E. Smokler, eds.), pp. 93–158. Wiley, New York.

[15] de Finetti, Bruno (1938). Sur la condition d'équivalence partielle. In *Colloque Consacré à la Théorie des Probabilités*, Volume VI, pp. 5–18. Herman et Cie, Paris.

[16] Evett, Ian W., Foreman, Lindsey A., and Weir, Bruce S. (2000). Letter to the Editor (with responses by A. Stockmarr and B. Devlin). *Biometrics*, **56**, 1274–1275.

[17] Farrell, Roger H. (1962). Representation of invariant measures. *Illinois Journal of Mathematics*, **6**, 447–467.

[18] Freedman, David A. (1962). Invariants under mixing which generalize de Finetti's theorem. *The Annals of Mathematical Statistics*, **33**, 916–923.

[19] Freedman, David A. (1963). Invariants under mixing which generalize de Finetti's theorem: Continuous time parameter. *The Annals of Mathematical Statistics*, **34**, 1194–1216.

[20] Goldstein, Michael (2012). Observables and models: exchangeability and the inductive argument. In *Bayesian Theory and Applications* (P. Damien, P. Dellaportas, N. G. Polson and D. A. Stephens, eds.), pp. 3–18. Oxford University Press, Oxford.

[21] Hewitt, Edwin and Savage, Leonard J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501.

[22] Hoover, Douglas N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, New Jersey.

[23] Hoover, Douglas N. (1982). Row-column exchangeability and a generalized model for exchangeability. In *Exchangeability in Probability and Statistics* (ed. G. Koch and F. Spizzichino), pp. 281–291. North-Holland, Amsterdam.

[24] Kingman, John F. C. (1972). On random sequences with spherical symmetry. *Biometrika*, **59**, 492–493.

[25] Lauritzen, Steffen L. (1988). *Extremal Families and Systems of Sufficient Statistics*. Number 49 in Lecture Notes in Statistics. Springer-Verlag, Heidelberg.

[26] Martin-Löf, Per (1974). Repetitive structures and the relation between canonical and microcanonical distributions in statistics and statistical mechanics. In *Proceedings of Conference on Foundational Questions in Statistical Inference* (ed. O. E. Barndorff-Nielsen, P. Blæsild, and G. Schou), Volume 1, Aarhus, pp. 271–294.

[27] McGinley, William G. and Sibson, Robin (1975). Dissociated random variables. *Mathematical Proceedings of the Cambridge Philosophical Society*, **77**, 185–188.

[28] Nelder, John A. (1965). The analysis of randomized experiments with orthogonal block structure: I. *Proceedings of the Royal Society of London, Series A*, **283**, 147–162.

[29] Nelder, John A. (1965). The analysis of randomized experiments with orthogonal block structure: II. *Proceedings of the Royal Society of London, Series A*, **283**, 163–178.

[30] Roeder, Kathryn, Escobar, Michael, Kadane, Joseph B., and Balazs, Ivan (1998). Measuring heterogeneity in forensic databases using hierarchical Bayes models. *Biometrika*, **85**, 269–287.

[31] Smith, Adrian F. M. (1981). On random sequences with centered spherical symmetry. *Journal of the Royal Statistical Society, Series B*, **43**, 208–209.