[23] Strassen, V. (1964). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, 3, 211–226.

[24] Strassen, V. (1967). *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, Vol. 2. University of California Press, Berkeley, Calif., pp. 315–343.

[25] Whitt, W. (1980). *Math. Operat. Res.*, 5, 67–85.

(BROWNIAN MOTION
CENTRAL LIMIT THEOREM
COMPETING RISKS
EDF STATISTICS
INVARIANCE CONCEPTS IN STATISTICS
LAW OF ITERATED LOGARITHM
LAWS OF LARGE NUMBERS
LIKELIHOOD RATIO
MARKOV PROCESSES
MARTINGALES
OCCUPANCY PROBLEMS
SEQUENTIAL ANALYSIS
STOCHASTIC APPROXIMATION
TIME SERIES
U-STATISTICS)

C. C. HEYDE

INVARIANT PRIOR DISTRIBUTIONS

The controversy surrounding Bayesian inference*, and its acceptability as a scientific methodology of statistical inference, has centered on its requirement that prior information about statistical parameters be explicitly introduced and described in terms of a probability distribution. (*See* INFERENCE, STATISTICAL for further background on the Bayesian approach.) A common objection is that the seeming arbitrariness and subjectivity of the prior distribution is at variance with the desire that statistical inference be entirely "objective."

The *logical Bayesian* view holds that a prior distribution represents partial logical information about unknown parameters, of the same objective status as a statistical model. In particular, it is supposed that, for any model, there is a specific prior distribution representing "complete ignorance." The program of determining such ignorance priors has been presented most cogently by Jeffreys [18]. (*See* JEFFREYS' NONINFORMATIVE PRIOR.) An important strand in this program is the idea of *invariant prior distributions*.

INVARIANT PRIOR PROBABILITY ASSIGNMENTS

Let P_θ be the distribution of a certain observand X over a space $(\mathscr{X}, \mathscr{A})$, given that a parameter Θ, with possible values in $\tilde{\Theta}$, takes the value θ. Define $\mathscr{P} = \{P_\theta : \theta \in \tilde{\Theta}\}$. We shall assume that $\mathscr{P}$ is dominated by a σ-finite measure μ on $\mathscr{A}$ and write $f(x \mid \theta) = dP_\theta(x)/d\mu$. By the *model* $\mathscr{M} = (X, \Theta, \mathscr{P})$ we shall understand the specification of the variables X and Θ, and distributions $\mathscr{P}$, together, implicitly, with the *parametrization* of $\mathscr{P}$, i.e., the mapping associating the appropriate $P_\theta \in \mathscr{P}$ with the value θ of Θ. We shall term the parametrized family $\mathscr{P}$ the *distribution model* associated with $\mathscr{M}$.

The task set is to associate, with each model $\mathscr{M}$, an ignorance prior distribution $\Pi_{\mathscr{M}}$ for its parameter. The possibility is explicitly allowed that ignorance may not be representable by a "proper" probability distribution, but by a general σ-finite measure giving possibly infinite "probability" to the whole parameter space. For example, ignorance about an unrestricted real parameter might perhaps be represented by Lebesgue measure, the "uniform distribution" on $\mathbb{R}^1$. Bayes's formula, stated as $d\Pi(\theta \mid x) \propto f(x \mid \theta) d\Pi(\theta)$, is formally applicable to such "improper" distributions*, and will often yield proper posterior distributions.

In nineteenth-century applications of Bayes's theorem* it was common to take the uniform distribution, in one or several dimensions, as a suitable representation of ignorance about a parameter with values in a Euclidean space. This practice largely followed Laplace* [19, 20]. However, as pointed out by Fisher* [9, Chap. II], this naive procedure leads to inconsistencies if applied to different parametrizations of the same problem. For example, if Θ is an unknown probability, an alternative parameter is $\Phi = \sin^{-1}\sqrt{\Theta}$. But a uniform distribution for Φ implies a nonuniform density for Θ,

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}.$$

Jeffreys attempted to circumvent these difficulties by searching for rules assigning $\Pi_{\mathscr{M}}$ to $\mathscr{M}$ in an invariant way. The main desiderata for such a rule may be set out as follows:

1. **Parameter invariance (PI).** Let $\mathscr{M} = (X, \Theta, \mathscr{P})$ and let $\Phi = \phi(\Theta)$ be a (smooth) invertible function, or *recoding*, of Θ. The model $\mathscr{M}_1 = (X, \Phi, \mathscr{P})$ differs from $\mathscr{M}$ in its parametrization, but describes an equivalent situation. So we should require that $\Pi_{\mathscr{M}_1}(\Phi \in A) = \Pi_{\mathscr{M}}(\Theta \in \phi^{-1}(A))$.
   Assuming Euclidean parameter spaces, and the existence of densities $\pi_{\mathscr{M}}$ and $\pi_{\mathscr{M}_1}$ with respect to Lebesgue measure, this requirement becomes
   $$\pi_{\mathscr{M}_1}(\phi) = \pi_{\mathscr{M}}(\theta).|J(\theta)|^{-1},$$
   where $J(\theta)$ is the Jacobian $\det(\partial\phi(\theta)/\partial\theta)$, and $\phi = \phi(\theta)$.
2. **Data invariance (DI).** Now let $Y = y(X)$ be a recoding of X, and let $\mathscr{M}_2 = (Y, \Theta, \mathscr{Q})$ be the induced model for observand Y and parameter Θ. Again the essential situation is unchanged, and we therefore require that $\Pi_{\mathscr{M}_2}(\Theta \in A) = \Pi_{\mathscr{M}}(\Theta \in A)$.

As noted by Dickey [8], these invariance requirements do not relate specifically to ignorance: identical considerations apply for subjective prior distributions representing genuine knowledge. The key additional assumption is:

3. **Context invariance (CI).** If $\mathscr{M} = (X, \Theta, \mathscr{P})$ and $\mathscr{M}' = (X', \Theta', \mathscr{P})$ are two different models having the same distribution model, we should require that $\Pi_{\mathscr{M}}(\Theta \in A) = \Pi_{\mathscr{M}'}(\Theta' \in A)$. In other words, no features of the structure, meaning, or context of a model, other than its distribution model, should be taken into account. This principle thus formalizes ignorance as the irrelevance of context.

When (CI) is assumed, we may write $\Pi_{\mathscr{P}}$ instead of $\Pi_{\mathscr{M}}$. The criteria (PI), (DI), and (CI) together impose strong restrictions on the assignment of prior distributions.

JEFFREYS'S AND HARTIGAN'S RULES

Jeffreys [18, Sec. 3.10] proposed the rule $\pi_{\mathscr{P}}(\theta) = |I(\theta)|^{1/2}$, where $I(\theta)$ is the *Fisher information matrix** of $\mathscr{P}$, with (i, j) entry $E_\theta[(\partial l/\partial\theta_i)(\partial l/\partial\theta_j)]$, where $l = l(X, \theta) = \log f(X \mid \theta)$. When it exists, this satisfies conditions (PI), (DI), and (CI). In the one-parameter case, Jeffreys's rule is equivalent to assigning a uniform distribution to that parametrization in which the information is constant, in accordance with a suggestion of Perks [24].

Hartigan [13] considered rules directly associating "inverse" distributions for the parameter with arbitrary data values in a specific model. In these terms, requirement (PI), for example, becomes $\Pi_{\mathscr{M}_1}(\Phi \in A \mid X = x) = \Pi_{\mathscr{M}}(\Theta \in \phi^{-1}(A) \mid X = x)$. When $\Pi_{\mathscr{M}}(\cdot \mid X = x)$ is supposed calculated from a fixed prior $\Pi_{\mathscr{M}}$ using Bayes's theorem, this may be rephrased as *relative parameter invariance* (RPI), requiring that $\Pi_{\mathscr{M}_1}(\Phi \in A) \propto \Pi_{\mathscr{M}}(\Theta \in \phi^{-1}(A))$ [or $\pi_{\mathscr{M}_1}(\phi) \propto \pi_{\mathscr{M}}(\theta) \times |J(\theta)|^{-1}$], where the implicit multiplier, which drops out on forming posteriors, may depend arbitrarily on the models and parametrizations. Similarly, we can introduce (RDI) and (RCI).

Hartigan suggested a rule satisfying (RPI), (RDI), and (RCI) which, for the one-parameter case, yields prior density $\pi_{\mathscr{P}}(\theta)$ with: $(d/d\theta)\log\pi_{\mathscr{P}}(\theta) = E_\theta(l_1 l_2)/E_\theta(l_2)$, where $l_i = (\partial^i/\partial\theta^i)\log f(X \mid \theta)$ $(i = 1, 2)$. He called this the *asymptotically locally invariant* (ALI) prior density. The rule may be extended to the multiparameter case, yielding simultaneous differential equations that may, however, be insoluble.

Hartigan also introduced several further invariance criteria. These are all satisfied for the Jeffreys and ALI assignment rules. New relatively invariant prior densities may be constructed by the formula $\pi(\theta) \propto \{\pi^J(\theta)\}^\alpha \{\pi^H(\theta)\}^\beta$, where $\alpha + \beta = 1$, and π^J, π^H are the Jeffreys and ALI densities.

Table 1 gives the Jeffreys and ALI invari-

Table 1 Examples of the Various Invariant Prior Densities

Family of Densities			Type of Invariant Prior Density				
Form	Sample Space	Parameter Space	Relative	Inner	Outer	Jeffreys's	ALI
$(2\pi)^{-1/2}\exp[-\frac{1}{2}(x-\theta)^2]$	$-\infty < x < \infty$	$-\infty < \theta < \infty$	1	1	1	1	1
$\theta^{-1}(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2/\theta^2)$	$-\infty < x < \infty$	$0 < \theta < \infty$	θ^k	θ^{-1}	θ^{-1}	θ^{-1}	θ^{-3}
$\theta_2^{-1}(2\pi)^{-1/2}\exp[-\frac{1}{2}(x-\theta)^2/\theta_2^2]$	$-\infty < x < \infty$	$-\infty < \theta < \infty$, $0 < \theta_2 < \infty$	θ_2^k	θ_2^{-2}	θ_2^{-1}	θ_2^{-2}	θ_2^{-5}
$\lvert\theta\rvert^{1/2}(2\pi)^{-(1/2)n}\exp(-\frac{1}{2}x^T\theta x)$	$x \in \mathbb{R}^n$	θ positive definite $n \times n$ matrix	$\lvert\theta\rvert^k$	$\lvert\theta\rvert^{-1}$	[a]	$\lvert\theta\rvert^{-1}$	1
$(2\pi)^{-(1/2)n}\exp[-\frac{1}{2}(x-K\theta)^T(x-K\theta)]$ $K, n \times k$ matrix of rank k	$x \in \mathbb{R}^n$	$\theta \in \mathbb{R}^k$	1	1	1	1	1
$x^{\theta-1}e^{-x}/\Gamma(\theta)$	$0 < x < \infty$	$0 < \theta < \infty$	[a]	[a]	[a]	$[(d^2/d\theta^2)\log\Gamma(\theta)]^{1/2}$	1
$\theta^x e^{-\theta}/x!$	x nonnegative integer	$0 < \theta < \infty$	[a]	[a]	[a]	$\theta^{-1/2}$	θ^{-1}
$\theta_1^{x_1}\theta_2^{x_2}\ldots\theta_r^{x_r}n!/(x_1!\ldots x_r!)$	x_i integers $x_i \geqslant 0$, $\sum x_i = 1$	$\theta_i > 0, \sum\theta_i = 1$	[a]	[a]	[a]	$(\theta_1\theta_2\ldots\theta_r)^{-1/2}$	$(\theta_1\theta_2\ldots\theta_r)^{-}$
$\theta^x(1-\theta)^{n-x}n!/\{x!(n-x)!\}$	x integer, $0 \leqslant x \leqslant n$	$0 < \theta < 1$	[a]	[a]	[a]	$[\theta(1-\theta)]^{-1/2}$	$[\theta(1-\theta)]^{-1}$
$\theta^r(1-\theta)^{x-r}(x-1)!/\{(r-1)!(x-r)!\}$	x integer, $x \geqslant r$	$0 < \theta < 1$	[a]	[a]	[a]	$(1-\theta)^{-1/2}\theta^{-1}$	$(1-\theta)^{-1}$

[a] Means either that the method is not defined for the family of densities considered, or that it does not determine a prior density for the family.

Source. Hartigan [13], reproduced with permission.

ant priors for some familiar families of distributions. (The columns headed "Relative," "Inner," and "Outer" are explained in the following and in the section "Group Models.")

SELF-CONSISTENCY

Suppose that, for the model $\mathscr{M} = (X, \Theta, \mathscr{P})$, we have a recoding of X, written as $Y = g \circ X$, with the property that, whenever X has a distribution in $\mathscr{P}$, so does Y, and vice versa. We obtain an induced recoding of Θ, $\Phi = \bar{g} \circ \Theta$, such that $X \sim P_\theta$ if and only if $g \circ X \sim P_{\bar{g} \circ \theta}$. Then the model $\mathscr{M}' = (Y, \Phi, \mathscr{P})$ has exactly the same distribution model $\mathscr{P}$ as $\mathscr{M}$. We call g ($\bar{g}$) an *equivariant* recoding of X (Θ), and say that $\mathscr{M}$, or $\mathscr{P}$, is *equivariant* under g and $\bar{g}$. The collection of all equivariant recodings of X (Θ) forms a transformation group $\mathscr{G}$ ($\bar{\mathscr{G}}$).

For $\Phi = \bar{g} \circ \Theta$, $\bar{g} \in \bar{\mathscr{G}}$, it follows from (DI) and (PI) that $\Pi_{\mathscr{M}'}(\Phi \in \bar{g} \circ A) = \Pi_{\mathscr{M}}(\Theta \in A)$. But if criterion (CI) holds, $\Pi_{\mathscr{M}'}(\Phi \in \bar{g} \circ A) = \Pi_{\mathscr{M}}(\Theta \in \bar{g} \circ A)$. So when (PI), (DI), and (CI) all apply, $\Pi_{\mathscr{M}}$ ($= \Pi_{\mathscr{P}}$) must be *invariant* under $\bar{\mathscr{G}}$: that is, $\Pi_{\mathscr{P}}(\bar{g} \circ A) = \Pi_{\mathscr{P}}(A)$ for all $\bar{g} \in \bar{\mathscr{G}}$. Essentially, this argument has been given by Jaynes [17] and Villegas [32].

If $\bar{\mathscr{G}}$ is *transitive* on Θ (so that, for any values θ_1, θ_2 of Θ, there exists $\bar{g} \in \bar{\mathscr{G}}$ with $\theta_2 = \bar{g} \circ \theta_1$), the condition of invariance under $\bar{\mathscr{G}}$ determined $\Pi_{\mathscr{P}}$ uniquely (up to a multiple). This must then agree with Jeffreys's prior, since that certainly satisfies (DI), (PI), and (CI). Frequently, however, $\bar{\mathscr{G}}$ will be small, and there will be numerous invariant distributions. The theory of Brillinger [5] is of relevance to the general characterization of $\bar{\mathscr{G}}$.

If we only insist on the weaker criteria (RPI), (RDI), and (RCI), the self-consistency requirement becomes $\Pi_{\mathscr{P}}(\bar{g} \circ A) = \alpha(\bar{g}).\Pi_{\mathscr{P}}(A)$ ($g \in \bar{\mathscr{G}}$), for some multiplier α, which must be a homomorphism from $\bar{\mathscr{G}}$ into the multiplicative group of positive reals. Such *relatively invariant* priors will include the ALI prior when it exists.

To clarify ideas, suppose that Θ is the unknown weight, in ounces, of a certain potato, and X is the reading, also in ounces, on a balance used to weigh it. Assume that X is normally distributed about Θ, with unknown standard deviation Φ ounces. Now let $X' = bX$, $\Theta' = b\Theta$, $\Phi' = b\Phi$, where $b = 1/35{,}840$, be the same quantities measured in tons rather than ounces. Then the requirement (RCI) demands a proportional formal formula for the prior density of (Θ', Φ') as for that of (Θ, Φ): this is satisfied for the relatively invariant priors $\pi(\theta, \phi) \propto \phi^k$, with invariance if $k = -2$. Clearly, in this context it would *not*, under any reasonable opinion, be irrelevant to the form of the prior distribution whether our measurements were in tons or in ounces. This reflects the fact that we are not entirely ignorant about the weight of potatoes, and demonstrates the strength of the requirement of context invariance.

EXPONENTIAL MODELS

Consider a distribution model with densities constituting a *regular exponential family** of order k:

$$f(x \mid \theta) = \exp\{a(x) + b(\theta) + \phi(\theta)^T t(x)\} \quad (\theta \in \tilde{\Theta}),$$

where

$$\tilde{\Theta} = \left\{\theta: \int \exp\{a(x) + \phi(\theta)^T t(x)\} d\mu(x) < \infty\right\},$$

$\tilde{\Phi} = \phi(\tilde{\Theta})$ is an open convex subset of $\mathbb{R}^k$.

Since the *canonical parameter* $\Phi = \phi(\Theta)$ seems to have a special status, one might assign a suitable prior distribution to Φ, and transfer it to an arbitrary parameter Θ by means of (PI). Specifically, we suppose that a prior distribution over Φ is assigned which depends only on its domain $\tilde{\Phi}$, and on no other feature of the model.

However, if Φ is a canonical parameter, so is any affine transformation of Φ. So the rule above is only self-consistent if the prior assigned is relatively invariant under all affine transformations preserving $\tilde{\Phi}$. The uniform distribution over $\tilde{\Phi}$, which is Hartigan's ALI prior, has this property, and is the only such distribution if $\tilde{\Phi} = \mathbb{R}^k$. For $k = 1$ and $\tilde{\Phi} = (0, \infty)$, $\pi(\phi)$ must have the form ϕ^λ, while for $\tilde{\Phi} = (-1, 1)$, any prior symmetric about 0 is permissible. These results are due to Huzurbazar [16], although his analysis of the case $k > 1$ appears suspect.

An alternative almost identical approach focuses on the *mean-value parameter* $\Psi = \psi(\Theta)$, where $\psi(\theta) = E_\theta(t(X))$, again taking values in a convex subset of $\mathbb{R}^k$, and unique up to an affine transformation.

For the binomial model: $f(x \mid \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, $(x = 0, 1, \ldots, n;\ 0 < \theta < 1)$ we have $\Phi = \log\{\Theta/(1-\Theta)\}$, $\Psi = n\Theta$. The first approach above yields $\pi(\phi) \propto 1$, equivalent to $\pi(\theta) \propto \{\theta(1-\theta)\}^{-1}$ [compare Jeffreys' prior: $\pi(\theta) \propto \{\theta(1-\theta)\}^{-1/2}$]. The second approach justifies any prior for Θ symmetric about $\frac{1}{2}$.

In this case a third approach is to assign, say, the discrete uniform distribution directly to the *canonical statistic* X. This arises as the marginal distribution when $\pi(\theta) = 1$. This is essentially the justification of this prior given by Bayes.

For an exponential family any equivariant recoding must induce an affine transformation of both Φ and Ψ, preserving their domains [1], so that the considerations above do not conflict with those of the preceding section. In particular, the uniform distribution for either Φ or Ψ will be relatively invariant under any equivariant recoding.

Huzurbazar [16] extends these ideas to nonregular families, where the range of the distribution varies with the parameter.

GROUP MODELS

In an attempt to weaken (CI), and the strong implications of self-consistency, we might (following Fraser [10]) suppose to be given, along with a model $\mathscr{M} = (X, \Theta, \mathscr{P})$, a subgroup G of the group $\mathscr{G}$ of equivariant recodings of X. Only for $g \in G$ shall we regard the transformed model $\mathscr{M}' = (g \circ X, \bar{g} \circ \Theta, \mathscr{P})$ as equivalent in context to $\mathscr{M}$, and so require identical, or proportional, prior densities for Θ and $\bar{g} \circ \Theta$. We call $(G, \mathscr{P})$ the *group model* associated with $\mathscr{M}$. An analysis parallel to that of the section "Self-Consistency" now implies only the invariance, or relative invariance, of $\Pi_{\mathscr{M}}$ under the smaller group $\bar{G} = \{\bar{g} : g \in G\}$.

In typical applications, G and $\bar{G}$ are locally compact topological groups. There then exists a *left Haar* measure μ on the Borel subsets of $\bar{G}$, determined up to a multiple, satisfying $\mu(\bar{g}S) \equiv \mu(S)$ $(\bar{g} \in \bar{G})$, and a *right Haar* measure ν, where $\nu(S) = \mu(S^{-1})$, for which $\nu(S\bar{g}) \equiv \nu(S)$. Moreover, $\nu(\bar{g}S) \equiv \Delta(\bar{g})^{-1}\nu(S)$, where Δ is the *modular function* of $\bar{G}$. For further background, see, e.g., Nachbin [21].

We call the group model *transitive* if $\bar{G}$ is transitive on $\tilde{\Theta}$. Then any measure m on $\bar{G}$ induces a measure Π on $\tilde{\Theta}$ by the rule: $\Pi(A) = m(\{\bar{g} \in \bar{G} : \bar{g} \circ \theta_0 \in A\})$, where $\theta_0 \in \tilde{\Theta}$ is a fixed reference point. If m is left Haar, then Π is invariant under $\bar{G}$, while the condition that Π should not depend on the choice of reference point θ_0 is satisfied when m is right Haar. Villegas [33, 34] terms these induced distributions the *inner* and *outer* priors, respectively. The inner prior, being invariant, agrees with Jeffreys's rule, as is apparent in Table 1. The outer prior is relatively invariant with multiplier $\Delta(\bar{g})^{-1}$.

A quantity $Q \equiv q(X, \Theta)$ is termed *invariant* under G if $q(g \circ x, \bar{g} \circ \theta) \equiv q(x, \theta)$ $(g \in G)$. For a transitive model, the distribution of $Q = q(X, \theta)$ under P_θ is the same for all θ; this remains true conditional on the maximal invariant statistic $A = a(X)$ under G, which is ancillary*. Fraser [10], Stein [25], Hora and Buehler [15], and Bondar [3] show that when the outer prior distribution is used, the posterior distribution of a G-invariant quantity Q is identical with its sampling distribution (conditional on A). For example, for the model $X_i \sim \mathscr{N}(M, \Sigma^2)$ independently $(i = 1, 2, \ldots, n)$, with a typical $g \in G$ operating as $g \circ (X_i) = (a + bX_i)$, $\bar{g} \circ (M, \Sigma) = (a + bM, b\Sigma)$ $(b > 0)$, the outer prior, with $\pi(\mu, \sigma) \propto \sigma^{-1}$, implies that, in the posterior distribution, the t-statistic $n^{1/2}(\bar{x} - M)/s$ has Student's distribution on $(n-1)$ degrees of freedom. [The inner prior density $\pi(\mu, \sigma) \propto \sigma^{-2}$ yields the "wrong" degrees of freedom n; however, Villegas [33, 34] attempts an objective justification for inference based on such inner priors: *see* INNER INFERENCE.]

The result above implies the identity of Fraser's fiducial* (structural) distribution for Θ with the posterior based on the outer prior distribution. Moreover, equivariant Bayesian confidence intervals* based on this prior will possess the classical confidence property. It also yields best equivariant procedures in decision theory [36].

Such connections with other modes of inference induced Jeffreys and others to abandon the inner prior in favor of the outer prior in group models, as producing more "objective" results. Note, however, that this cannot be done consistently if the strong condition (CI) is assumed, since there exist distribution families equivariant under the action of two different groups, with different modular functions, for which no prior can be outer under both groups. As an example, let $\mathbf{X}$ have the zero-mean p-variate normal distribution, with dispersion matrix $\mathbf{\Sigma}$. This distribution model is equivariant under the group G of all nonsingular $(p \times p)$ matrices, with $g \circ X = gX$, $g \circ \Sigma = g\Sigma g^T$; and also under its subgroups G_1 (respectively, G_2) of all lower (respectively, upper) triangular matrices with positive diagonal. These yield three different transitive group models, all inducing different outer priors for $\mathbf{\Sigma}$ [23].

SOME DIFFICULTIES

We now present some "paradoxes" relating to the specification of an ignorance prior within a specific model. (Stone and Springer [30] consider the consistency of joint specification of priors in models related to each other.)

Strong Inconsistency

Let $\mathbf{X}_i$ $(i = 1, 2, \ldots, N)$ be a random sample from the general p-variate normal distribution $\mathcal{N}(\mathbf{M}, \mathbf{\Lambda})$. This model is equivariant under $\mathbf{X}_i \to \mathbf{A}\mathbf{X}_i + \mathbf{b}$, $\mathbf{M} \to \mathbf{A}\mathbf{M} + \mathbf{b}$, $\mathbf{\Lambda} \to \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T$. [$\mathbf{A}$ nonsingular $(p \times p)$, $\mathbf{b} \in \mathbb{R}^p$.] A sufficient statistic is $(\overline{\mathbf{X}}, \mathbf{S})$, where

$$N\overline{\mathbf{X}} = \sum_{i=1}^{N} \mathbf{X}_i,$$

$$(N-1)\mathbf{S} = \sum_{i=1}^{N} \mathbf{X}_i\mathbf{X}_i^T - N\overline{\mathbf{X}}\overline{\mathbf{X}}^T.$$

Relatively invariant priors for this group have density of the form $\pi(\boldsymbol{\mu}, \boldsymbol{\lambda}) \propto |\boldsymbol{\lambda}|^{-(1/2)v}$, being inner for $v = p + 2$ and outer for $v = p + 1$.

Consider $Q_1 = N(\overline{\mathbf{X}} - \mathbf{M})^T\mathbf{S}^{-1}(\overline{\mathbf{X}} - \mathbf{M})$ (Hotelling's T^2*) and $Q_2 = N^{1/2}(\overline{X}_1 - M_1)/S_{11}^{1/2}$ (Student's t for a single data component). Then Q_1 is invariant, but Q_2 is not. The sampling distributions for both Q_1 and Q_2 are constant, being respectively $\{p(N-1)/(N-p)\} \cdot F_{p,N-p}$ and t_{N-1}.

It may be shown that the posterior distribution for Q_1 is

$$\{p(N-1)/(N-v+1)\}F_{p,N-v+1},$$

while that for Q_2 is $\{(N-1)/(N-v+1)\}^{1/2}.t_{N-v+1}$ [11]. The choice $v = p + 1$ yields the "correct" distribution for Q_1, as follows from the general results of the section "Group Models"; however, it gives the "wrong" distribution $\{(N-1)/(N-p)\}^{1/2}t_{N-p}$ for Q_2. If we consider the (non-equivariant) interval estimator for μ_1, of the form $(\overline{X}_1 \pm k(S_{11}/N)^{1/2})$, this will have constant sampling coverage probability $\gamma_1 = \Pr(|t_{N-1}| < k)$, and different constant posterior probability

$$\gamma_2 = \Pr\left(|t_{N-p}| < k\{(N-p)/(N-1)\}^{1/2}\right).$$

This conflict between the "objective posterior" and "sampling" interpretations is called "strong inconsistency" by Stone [28].

The choice $v = 2$ eliminates the inconsistency for Q_2, but introduces it for Q_1. No prior can eliminate both inconsistencies simultaneously.

An example with similar behavior is presented by Dempster [7].

Marginalization Paradoxes [6]

Let X_{ij} $(i = 1, 2;\ j = 1, 2, \ldots, N)$ be independent, X_{ij} having the normal distribution $\mathcal{N}(M_i, \Sigma^2)$. This model is equivariant under

$(X_{ij}) \to (a_i + bX_{ij})$, $(M_1, M_2, \boldsymbol{\Sigma}) \to (a_1 + bM_1, a_2 + bM_2, b\boldsymbol{\Sigma})$ $(b > 0)$. Relatively invariant prior densities have the form $\pi(\mu_1, \mu_2, \sigma) \propto \sigma^{\lambda}$, with $\lambda = -3$ giving the inner, and $\lambda = -1$ the outer prior.

Define $\Xi_i = M_i/\boldsymbol{\Sigma}$, $Z_i = \overline{X}_i/S$ $[N\overline{X}_i = \sum_j X_{ij}$, $S^2 = \sum_{i,j}(X_{ij} - \overline{X}_i)^2]$. The marginal posterior distribution for Ξ_1, under a relatively invariant prior, has density at ξ_1 proportional to

$$\int_0^{\infty} \omega^{2n-4-\lambda} \exp\left[-\tfrac{1}{2}\left\{\omega^2 + n(z_1\omega - \xi_1)^2\right\}\right] d\omega, \quad (1)$$

depending only on the value z_1 of Z_1. It seems that it ought, therefore, to be possible to reproduce this marginal posterior if only Z_1 is observed. Now the sampling density of Z_1, which depends only on the value ξ_1 of Ξ_1, is proportional to

$$\int_0^{\infty} \omega^{2n-2} \exp\left[-\tfrac{1}{2}\left\{\omega^2 + n(z_1\omega - \xi_1)^2\right\}\right] d\omega, \quad (2)$$

so that any posterior density for Ξ_1, using Bayes' theorem* with data Z_1 alone, would contain (2) as a factor. However, examining (1), we see that this will not hold, unless $\lambda = -2$—neither inner nor outer.

An almost identical argument applies when (Ξ_1, Ξ_2) are considered jointly, with posterior distribution governed by (Z_1, Z_2); only in this case the choice $\lambda = -3$ is needed to avoid the inconsistency. Thus there can be *no* relatively invariant prior which simultaneously avoids all such paradoxes.

FINITE ADDITIVITY

Many of the problems associated with ignorance priors stem from their impropriety. An alternative approach, still in its infancy, is to insist on propriety, but allow distributions which are only finitely additive. Some relevant theory is given by Heath and Sudderth [14].

We focus on assignments of inverse distributions $\{\Pi_x\}$ for parameter Θ given data $X = x$, which can be regarded as posteriors based on a finitely additive prior. This prior need not be uniquely determined; moreover, the inverse distributions need not, with finite additivity, be constructed using Bayes' theorem. General coherence* properties for proper priors imply that it is now impossible to have an interval estimator $I(X)$ for Θ for which, simultaneously, $P_\theta(\theta \in I(X)) \leqslant \gamma_1$, all θ, and $\Pi_x(\Theta \in I(x)) \geqslant \gamma_2$, all x, where $\gamma_1 < \gamma_2$. That is, strong inconsistency cannot occur. Similarly, the marginalization paradox is avoided, since the "un-Bayesian" look of the marginal posteriors is deceptive: they *are* true posteriors, based on the appropriate, finitely additive marginal prior distribution [31].

If $(G, \mathscr{P})$ is a transitive group model, the appropriate extension of the theory of the section "Group Models" [based on Hartigan's versions of (PI), (DI), and (CI)] yields the *posterior equivariance* requirement:

$$\Pi_{\mathscr{P}}(\Theta \in \bar{g} \circ A \mid X = g \circ x) \equiv \Pi_{\mathscr{P}}(\Theta \in A \mid X = x) \qquad (g \in G).$$

This property does hold for formal posteriors based on improper, countably additive, relatively invariant priors. It is, therefore, of interest to enquire when such formal posteriors are true posteriors for some proper, finitely additive prior, since they cannot then be subject to the difficulties of the section "Some Difficulties." This will be so if and only if the formal prior used is outer, and the group $\overline{G}$ has the technical property known as *amenability* [4, 12], implying that it can support a proper right-invariant finitely additive distribution, which then induces the required proper prior. This approach, therefore, gives some justification to the use of the improper outer prior (although no justification for the inner prior), but only for "well-behaved" groups. In particular, the group of the section "Strong Inconsistency" is nonamenable, as must be the case for Q_2 to provide strong inconsistency for the outer prior having $v = p + 1$. For such groups, the posterior equivariance requirement is simply not satisfiable within the framework of finite additivity.

CONCLUSION

Logical Bayesianism is not currently popular. Certainly, the claim that a unique ignorance prior distribution exists for any problem must remain suspect as long as the relevant theory produces a whole range of choices in some problems, and no prior free from all objections in others. Nevertheless, much current Bayesian practice uses, overtly or covertly, priors supposed to represent "vague prior knowledge" with respect to a given model; often an outer prior is supposed. From a subjectivist viewpoint, such formal priors may be regarded as approximations to diffuse but proper real priors, although this argument requires much care and its general validity depends, at least, on the amenability of the underlying group [26, 27, 29]. Alternatively, the need is felt for a "zero" or "reference" prior to which an informative subjective distribution can be compared [2, 22, 35]. The formalization of ignorance thus remains the central object of a continuing quest by the knights of the Bayesian round table: inspiring them to imaginative feats of daring, while remaining, perhaps, forever unattainable.

References

[1] Barndorff-Nielsen, O., Blaesild, P., Jensen, J. L., and Jørgensen, B. (1982). Exponential transformation models. *Proc. R. Soc. Lond. A*, **379**, 41–65.

[2] Bernardo, J.-M. (1979). Reference posterior distributions for Bayesian inference (with Discussion). *J. R. Statist. Soc. B*, **41**, 113–147.

[3] Bondar, J. V. (1972). Structural distributions without exact transitivity. *Ann. Math. Statist.*, **43**, 326–339.

[4] Bondar, J. V. and Milnes, P. (1981). Amenability: a survey for statistical applications of Hunt–Stein and related conditions on groups. *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **57**, 103–128.

[5] Brillinger, D. R. (1963). Necessary and sufficient conditions for a statistical model to be invariant under a Lie group. *Ann. Math. Statist.*, **34**, 492–500.

[6] Dawid, A. P., Stone, M., and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. R. Statist. Soc. B*, **35**, 189–233.

[7] Dempster, A. P. (1963). On a paradox concerning inference about a covariance matrix. *Ann. Math. Statist.*, **34**, 1414–1418.

[8] Dickey, J. M. (1973). Discussion of Dawid, Stone and Zidek (1973). *J. R. Statist. Soc. B*, **35**, 219–221.

[9] Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver & Boyd, London.

[10] Fraser, D. A. S. (1961). The fiducial method and invariance. *Biometrika*, **48**, 261–280.

[11] Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *J. R. Statist. Soc. B*, **25**, 368–376.

[12] Greenleaf, P. (1969). *Invariant Means on Topological Groups*. Van Nostrand Reinhold, New York.

[13] Hartigan, J. (1964). Invariant prior distributions. *Ann. Math. Statist.*, **35**, 836–845.

[14] Heath, D. and Sudderth, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.*, **6**, 333–345.

[15] Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.*, **37**, 643–656.

[16] Huzurbazar, V. S. (1976). *Sufficient Statistics*. Marcel Dekker, New York.

[17] Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.*, **SSC-4**, 227–241.

[18] Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford (1st ed., 1939).

[19] Laplace, P. S. de (1774). Mémoire sur la probabilité des causes par les évènements. *Mém. Acad. R. Sci. Paris (Savants Étrangers)*, **6**, 621–656.

[20] Laplace, P. S. de (1820). *Théorie Analytique des Probabilités*, 3rd ed. Courcier, Paris.

[21] Nachbin, L. (1965). *The Haar Integral*. Van Nostrand, New York.

[22] Novick, M. R. (1969). Multiparameter Bayesian indifference procedures (with Discussion). *J. R. Statist. Soc. B*, **31**, 29–64.

[23] Nussbaum, M. (1976). Structural distributions in the multivariate linear model. *Math. Operat. Statist.*, **7**, 679–683.

[24] Perks, W. (1947). Some observations on inverse probability including a new indifference rule. *J. Inst. Actuaries*, **73**, 285–334.

[25] Stein, C. (1965). Approximation of improper prior measures by prior probability measures. In *Bernoulli, 1713*; *Bayes, 1763*; *Laplace, 1813*, J. Neyman and L. M. LeCam, eds. Springer-Verlag, Berlin, pp. 217–240.

[26] Stone, M. (1965). Right Haar measure for convergence in probability to quasiposterior distributions. *Ann. Math. Statist.*, **36**, 440–453.

[27] Stone, M. (1970). Necessary and sufficient condition for convergence in probability to invariant posterior distributions. *Ann. Math. Statist.*, **41**, 1349–1353.

[28] Stone, M. (1976). Strong inconsistency from uniform priors (with Discussion). *J. Amer. Statist. Ass.*, **71**, 114–125.

[29] Stone, M. (1980). Review and analysis of some inconsistencies related to improper priors and finite additivity. In *Logic, Methodology and Philosophy of Science VI*, L. J. Cohen, J. Løs, H. Pfeiffer, and K.-P. Podewski, eds. North-Holland, Amsterdam.

[30] Stone, M. and Springer, B. G. F. (1965). A paradox involving quasi prior distributions. *Biometrika*, **52**, 623–627.

[31] Sudderth, W. D. (1980). Finitely additive priors, coherence, and the marginalization paradox. *J. R. Statist. Soc. B*, **42**, 339–341.

[32] Villegas, C. (1971). On Haar priors. In *Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott, eds. Holt, Rinehart and Winston, Toronto, pp. 409–414.

[33] Villegas, C. (1977). Inner statistical inference. *J. Amer. Statist. Ass.*, **72**, 453–458.

[34] Villegas, C. (1981). Inner statistical inference II. *Ann. Statist.*, **9**, 768–776.

[35] Zellner, A. (1977). Maximal data information prior distributions. In *New Developments in the Applications of Bayesian Methods*, A. Aykac and C. Brumat, eds. North-Holland, Amsterdam, pp. 211–232.

[36] Zidek, J. V. (1969). A representation of Bayes invariant procedures in terms of Haar measure. *Ann. Inst. Statist. Math.*, **21**, 291–308.

(ANCILLARITY
BAYESIAN INFERENCE
FIDUCIAL PROBABILITY
HAAR DISTRIBUTIONS
SUFFICIENCY)

A. P. DAWID

INVENTORY THEORY

Inventory theory involves a class of mathematical models devoted to the analysis of systems in which stock is maintained to meet an external or an internal demand. The economic motives for maintaining inventories are discussed below.

Economies of Scale

If the marginal cost of producing or ordering units from an outside supplier is a nonincreasing function of the number of units produced, then it is economical to produce or order in lots and store units for future use. The conventional approach for modeling this phenomenon has been to assume that the cost of ordering y units, say $C(y)$, is of the form

$$C(y) = K\delta(y) + cy,$$

where

$$\delta(y) = \begin{cases} 1 & \text{if } y > 0 \\ 0 & \text{if } y = 0. \end{cases}$$

The fixed cost component, K (often called the setup cost), is incurred if a positive order is placed, independent of the magnitude of the order. The average cost per unit of ordering $y > 0$ units is $[C(y)/y] = (K/y) + c$, which is clearly decreasing in y.

Nonstationarities

If costs of production increase over time, it may be advantageous to retain stocks to avert the higher production costs. Similarly, if the value of the inventory is increasing, stocks are held in order to take advantage of higher future prices. Maintaining inventories may also be advantageous if demands are increasing as a function of time.

Uncertainties

A fundamental motive for carrying inventories is to provide a buffer against uncertainty. The most significant source of uncertainty is future demands. When the demand is random, a consequence of the analysis is that optimal policies retain safety stocks to provide a hedge against the uncertainty. In some systems, uncertainty in supply can be a dominant factor. Another common source of uncertainty is in the time required to replenish stocks.

HISTORICAL BACKGROUND

The simple economic lot size model, which forms the basis for much of the subsequent research in the field, appears to be due to Harris [12], although the term "Wilson lot size formula" is often used because of Wilson's [35] later and more comprehensive analysis. Interest in mathematical inventory models seems to have arisen during World War II, but papers did not start appearing in the open literature until the early 1950s. Dvoretzky et al. [8, 9] discussed the existence of optimal policies and the mathematical structure of inventory models, while Arrow et al. [1] applied renewal theory* to the problem of computing optimal policies. Whitin [34] discussed some of the economic and operational issues of inventory management. He also derived a model which has enjoyed considerable application in practice.

The collection of articles appearing in Arrow et al. [2] provided an important cornerstone to the development of modern inventory theory. The articles treat, in a rigorous and in-depth fashion, the techniques of dynamic programming* and stationary analysis which formed the basis for much of the later research work. The text by Hadley and Whitin [10] provided an excellent reference source for the significant work done up until that time.

Since approximately 1960, well over 1000 papers on inventory control models have appeared in the open literature. These have appeared primarily (although not exclusively) in the operations research journals. The most notable of these are *Management Science*, *Operations Research**, *Naval Research Logistics Quarterly**, and *AIIE Transactions* in the United States, *INFOR* in Canada, *O.R. Quarterly* and *International Journal of Production Research* in Great Britain, *Opsearch* in India, *Cahiers de Recherche Operationelle* in Belgium, and the *Journal of the Japanese OR Society* in Japan.

MODELS WITH KNOWN DEMAND

EOQ Model and Extensions

The classical economic order quantity model may be derived under the following assumptions:

1. Demand is known with certainty and fixed at λ units per unit time.
2. Shortages are not permitted.
3. Costs are levied against:
   a. Ordering at \$$K$ per order
   b. Holding at \$$h$ per unit per unit time

The objective of the analysis is to minimize the average cost per unit time. If batches of size Q are ordered every Q/λ units of time, it follows that the cost rate, $C(Q)$, is given by

$$C(Q) = \frac{K\lambda}{Q} + \frac{hQ}{2}.$$

This follows since one setup is incurred each Q/λ units of time and the average number of units in stock at any point in time is $Q/2$. The minimizing value of Q is

$$Q^* = \sqrt{\frac{2K\lambda}{h}},$$

which is commonly known as the EOQ formula (due to Harris [12]).

There are a number of relatively straightforward extensions of this model which we will only mention. When items are produced internally rather than ordered from an outside supplier, assuming that the rate of production is finite (rather than infinite as we have done above), might be more appropriate. Another common extension is to allow shortages to accrue at a cost per unit backordered per unit time or, if appropriate, at a cost per lost sale. In many circumstances, suppliers might offer quantity discounts for larger orders. The reader interested in these extensions of the simple lot size model should refer to Chap. 2 of Hadley and Whitin [10].

Production Planning Models

When demand is nonstationary, that is, changing over time, the structure of the optimal policy is quite different from the static EOQ model discussed above.

Suppose that known requirements for a