Some Remarks on Consistency and Strong Inconsistency of Bayesian Inference

Kociecki, Andrzej

National Bank of Poland

8 February 2011

SOME REMARKS ON CONSISTENCY AND STRONG INCONSISTENCY OF BAYESIAN INFERENCE

Andrzej Kocięcki
National Bank of Poland
e-mail: andrzej.kociecki@nbp.pl

First version: February 2011
This version: February 2011

Abstract: The paper provides new sufficient conditions for consistent and coherent Bayesian inference when a model is invariant under some group of transformations. Building on our theoretical results we reexamine an example from Stone (1976) giving some new insights. The priors for multivariate normal models and Structural Vector AutoRegression models that entail consistent and coherent Bayesian inference are also discussed.

I. INTRODUCTION

This paper contributes to the line of research whose main goal is to determine conditions under which the Bayesian inference is not strongly inconsistent in Stone's (1976) sense, coherent in de Finetti's sense (Dutch book) or coherent in the sense of Heath and Sudderth (1978) (precise definition to be given later). Informally, when strong inconsistency or incoherence arises, an adoption of the frequency approach and Bayesian inference for the same model may result in very different and sometimes paradoxical conclusions. Although a Bayesian inference that corresponds to a posterior with a proper prior, which is a countable additive measure, is always coherent and not strongly inconsistent, an inference with improper prior need not be coherent and may be strongly inconsistent. Since improper priors are widely used, the problem is very serious. The literature on strong inconsistency and/or incoherence contains many interesting but also striking results. One of them concerns the use of the Jeffreys' prior when we have a sample of n random variables $y_1,\dots,y_n$, where $y_i \in \mathbb{R}^m$ ($m \geq 2$) and $y_i \sim i.i.d.\ N(0,\Sigma)$. For example, Eaton and Sudderth (1995), Eaton and Freedman (2004) showed that the Jeffreys' prior in such a case (i.e. proportional to $|\Sigma|^{-\frac{1}{2}(m+1)}$) leads to incoherent posteriors in the sense of Heath and Sudderth (1978), incoherent in de Finetti's sense (Dutch book) and strongly inconsistent in the Stone's (1976) sense. Moreover, Eaton and Sudderth (1993, 1998), Eaton and Freedman (2004) showed that predictive distributions in this model under the Jeffreys' prior are incoherent in de Finetti's sense, in the sense of Heath and Sudderth (1978) and strongly inconsistent in the Stone's (1976) sense. Those conclusions are far–reaching and imply that incoherence may arise in basic models even with standard priors. It also suggests that the priors in Simultaneous Equations Models (SEM) or Vector AutoRegression (VAR) models that are proportional to $|\Sigma|^{-\frac{1}{2}(m+1)}$ are incoherent and strongly inconsistent. Unfortunately, the matters are even worse. Except Bernardo's reference prior (see e.g. Yang and Berger (1994)), all non–informative priors for SEM or VAR models proposed in the literature have the form $|\Sigma|^{\alpha}$, for some $\alpha \in \mathbb{R}$ (see e.g. Drèze (1976), Zellner (1977), and entries 3–7 as listed in table 1 in Keyes and Levy (1996)). Perhaps surprisingly, Eaton and Sudderth (1998), Eaton (2008) showed that any prior in the form $|\Sigma|^{\alpha}$ leads to incoherent and strongly inconsistent predictive distributions in multivariate normal model i.e. a coherent predictive distribution is built on the prior that can not be cast in the form $|\Sigma|^{\alpha}$, for any $\alpha \in \mathbb{R}$. We think the above claims make the concepts of strong inconsistency and incoherence very interesting research topic.

In all considerations about strong inconsistency and incoherence special role is reserved for the prior induced by the right Haar invariant measure on a group that acts in a model (see section IV for formal statement)[1]. For reference, let us call such a prior the right Haar prior. It turns out that any predictive distribution that is not based on the right Haar prior is incoherent and strongly inconsistent, see e.g. Eaton and Sudderth (1998,1999), Eaton and Freedman (2004), Eaton (2008). Parallel results of Eaton and Sudderth (2002,2004) and Eaton and Freedman (2004) indicate that any Bayesian posterior that is not derived under the right Haar prior must be strongly inconsistent and incoherent. However the question whether the Bayesian posterior or predictive distribution derived under the right Haar prior is not strongly

[1] One may say 'again'. It is remarkable in how many problems the right Haar invariant measure turns out to be really the "right" choice. For an incomplete survey see e.g. Helland (2004) and references therein.

inconsistent (i.e. consistent) or coherent, has no definite answer. Some further conditions are needed.

One criterion for coherence of Bayesian posterior under the improper prior was given by Heath and Sudderth (1978). The posterior is coherent if the improper prior may be replaced with some finitely additive (proper) prior which gives the same posterior as that under the improper prior, see Heath and Sudderth (1978), or the posterior is approximable by proper prior, see Heath and Sudderth (1989). Neither criterion is easy to apply in practice. Some other coherency conditions are available in Lane and Sudderth (1983) but are suitable only when a parameter space or a sample space is compact. On the other hand Heath and Sudderth (1978) showed that when a parameter space is equal to a sample space (very unrealistic), a model belongs to the translation family and a group acting in the model is amenable[2], a posterior derived under the right Haar prior is coherent. However, the results of Eaton and Sudderth (1998,1999,2002,2004) clearly suggest that the above assertion depends crucially on the amenability of a group, which appears as the essential sufficient condition for the coherent inference.

Since the group of nonsingular matrices with matrix multiplication as a group composition (so–called the general linear group) is not amenable, we are in an uncomfortable position. It is so because such a group is the most natural group acting in the multivariate normal model with zero mean. However according to the available theory represented by Eaton and Sudderth (1998,1999,2002,2004), the lack of amenability implies that the right Haar prior in this case is strongly inconsistent. Essentially this is the reason why the Jeffreys' prior in the multivariate normal model is condemned by those authors to strong inconsistency and incoherence. Our contribution is, to some extent, a partial re–validation of the Jeffreys' prior. We provide an alternative sufficient condition, which does not refer to amenability, but guarantees that the right Haar prior leads to consistent and coherent Bayesian inference. By our criterion, the coherence of Bayesian inference may be preserved even though the group is not amenable. Since non–amenable groups are basic in all theoretical considerations (e.g. general linear group, affine group), our contribution has obvious merits.

[2] Amenability of a group is a rather technical notion. There are many equivalent definitions of the amenability see e.g. Bondar (1977), Heath and Sudderth (1978), Eaton and Sudderth (1999), Lehmann and Casella (1998), p. 422. More complete discussion is available in Bondar and Milnes (1981). We note that amenability of a group is precisely the condition for Hunt–Stein theorem, see e.g. Lehmann (1986), pp. 519–522.

II. Notation and Assumptions

All results in our paper are restricted for invariant models with respect to some group of transformations. See e.g. Lehmann (1986), chapter 6, for the theory and our assumption 1 for the mathematical definition. By G we will denote this underlying group acting in a model. Basic material on groups, group actions and other related notions (sufficient for our problem) may be found in Eaton (1989). By e we denote the identity element in a group G. We will extensively use the concepts of Haar measures and integrals. Traditional reference is Nachbin (1965), but Eaton (1989) and Wijsman (1990) are also useful. We will use the following unified notation, μ_G: the left Haar invariant measure on a group G; ν_G: the right Haar invariant measure on a group G; $\eta_{\mathcal{S}}$: a probability measure on a space $\mathcal{S}$; $\lambda_{\mathcal{S}}$: a $\sigma-$finite measure on a space $\mathcal{S}$. In particular the Lebesgue measure on a space $\mathcal{S}$ will be denoted as (ds), where $s \in \mathcal{S}$. We shall denote σ–algebra of Borel subsets of a space $\mathcal{S}$ as $\mathcal{B}_{\mathcal{S}}$.

We will not differentiate between groups and its domain spaces. Thus $GL_m = \{g \in \mathbb{R}^{m\times m} \mid \det(g) \neq 0\}$ signifies both the general linear group with matrix multiplication as a group composition, and (seen as a space) the space of $m\times m$ nonsingular matrices. Analogous remark relates to LT_m^+: the group (space) of $m\times m$ lower triangular matrices with positive elements on the diagonal; and $O_m = \{g \in \mathbb{R}^{m\times m} \mid g'g = gg' = \mathrm{I}_m\}$: the group (space) of orthogonal matrices ($\mathrm{I}_m : (m\times m)$ is the identity matrix). Obviously, a group composition in LT_m^+ and O_m is the usual matrix multiplication. Lastly, a space of $m\times m$ positive definite symmetric matrices will be denoted as PD_m.

Let Y be a random variable (with realization y) taking on values in $\mathcal{Y}$ (a sample space). Let $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ be a family of probability measures on $\mathcal{Y}$ indexed by the parameter $\theta \in \Theta$ i.e. a model. In the sequel, probability measure P_θ will be also denoted as $P(\cdot \mid \theta)$. We assume that P_θ has a density $p(y \mid \theta)$ with respect to some dominating measure $\lambda_{\mathcal{Y}}$, which is relatively left invariant. That is $\lambda_{\mathcal{Y}}(gB) = \chi(g)\cdot\lambda_{\mathcal{Y}}(B)$, for all $B \in \mathcal{B}_{\mathcal{Y}}$, $g \in G$, where the notation gB is explained in assumption 1 and $\chi : G \to \mathbb{R}^+$ is the multiplier, see e.g. Wijsman (1990), pp. 127–130. Moreover, let $\{\Pi_y \mid y \in \mathcal{Y}\}$ be a family of posterior distributions on Θ determined by a model $\mathcal{P}$ and a fixed $\sigma-$finite prior measure π. That is $\Pi_y(B) = [m(y)]^{-1}\cdot\int_{\theta\in B} p(y \mid \theta)\pi(d\theta)$, for any $B \in \mathcal{B}_\Theta$, where $m(y) = \int_\Theta p(y \mid \theta)\pi(d\theta)$. Having measurable spaces $(\Theta, \mathcal{B}_\Theta)$ and $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$, we assume that the map $y \mapsto \Pi_y(B)$

is $\mathcal{B}_{\mathcal{Y}}$ −measurable for each $B \in \mathcal{B}_{\Theta}$. We say that Bayesian inference is proper if $m(y) < \infty$, a.e. $[\lambda_{\mathcal{Y}}]$. We shall call $\{\Pi_y \mid y \in \mathcal{Y}\}$ derived under specific π, a Bayesian inference. Depending on situation, Π_y will be also denoted as $\Pi(\cdot \mid y)$.

We use the symbol "$\circ$" to denote the abstract group operation on some sets. On the other hand the group composition will not be symbolically distinguished from the usual matrix operation e.g. $g \circ \theta$ but gh, for $g, h \in G$ (group) and $\theta \in \Theta$ (set). With abuse of notation, but following Eaton (1989) and a large body of the literature, both an action of a group on $\mathcal{Y}$ and its induced action on Θ will be denoted as $g \circ y$ and $g \circ \theta$, respectively. With this in mind, the assumption that a model $\mathcal{P}$ is G −invariant reads

Assumption 1 (model G –invariance): $P_{g\circ\theta}(Y \in gB) = P_{\theta}(Y \in B)$ for all $g \in G$, where $B \in \mathcal{B}_{\mathcal{Y}}$ and $gB = \{g \circ y \mid y \in B\}$.

We say that G acts transitively on Θ if for each $\theta_1, \theta_2 \in \Theta$ there is a $g \in G$ such that $\theta_2 = g \circ \theta_1$. In other words, transitivity means that given $\theta_0 \in \Theta$, every $\theta \in \Theta$ can be represented as $\theta = g \circ \theta_0$, for some $g \in G$.

Assumption 2 (transitivity): G acts transitively on Θ.

Assumption 2 is restrictive. It holds only in very simple models (like the multivariate normal model with unknown mean and covariance), but more complicated models including VAR or SEM violate this assumption. However transitivity is assumed (literally) in all works cited in the references in the similar context. Otherwise, it appears to be difficult to obtain sharp theoretical results.

Define the stabilizer $\mathrm{Stab}_s = \{g \in G \mid g \circ s = s\} \subseteq G$, for each $s \in \mathcal{S}$. In fact, the stabilizer is a subgroup of G. When $\mathrm{Stab}_s = \{e\}$, $\forall\ s \in \mathcal{S}$, we say that a group G acts freely on $\mathcal{S}$.

Assumption 3 ($G - \Theta$ freeness): $\mathrm{Stab}_{\theta} = \{e\}$; $\forall\ \theta \in \Theta$.

Assumption 4 ($G - \mathcal{Y}$ freeness): $\mathrm{Stab}_{y} = \{e\}$; $\forall\ y \in \mathcal{Y}$.

Assumption 3 may be violated even in basic models. For example in the multivariate normal model with mean 0 and unknown covariance that was explicitly mentioned in section I, when the underlying group is GL_m. However there are models for which assumption 3 is fulfilled, see section VI. Sometimes, whether the assumption holds or not may be a matter of the clever reparameterization of the original model. For instance, if instead of the parameterization of the normal model from section I we use an alternative model $y_i = \theta \varepsilon_i$, for $i = 1, \ldots, n$, where $y_i \in \mathbb{R}^m$ ($m \geq 2$), $\theta \in LT_m^+$ and $\varepsilon_i \sim i.i.d.\ N(0, \mathrm{I}_m)$, then the $G - \Theta$ freeness assumption is satisfied (provided that $G = LT_m^+$).

Assumption 4 holds in standard cases before making a sufficiency reduction of the sample space (with possibility that we exclude some "singular" points of zero measure from the sample space). See e.g. Bondar (1976), Chang and Eaves (1990) for some clarifying discussion. We note that assumption 4 appears under the label "FP" in Eaton and Sudderth (1999) and "unitary" in Fraser (1968), p. 49.

In general there is a possibility to weaken assumptions 3 and 4 by requiring that both Stab_y, for each $y \in \mathcal{Y}$, and Stab_θ, for each $\theta \in \Theta$, are compact subgroups of G, see e.g. Eaton and Sudderth (2002). However in such a case our lucid framework would be lost. In fact assumptions 3 and 4 are critical for our proof method of the main theorem.

Let $\mathcal{W}$ be a space of some cross section on $\mathcal{Y}$. That is $\mathcal{W}$ is a subset of $\mathcal{Y}$ that intersects each orbit of $\mathcal{Y}$ in exactly one point[3]. In other words, a cross section is in one–to–one correspondence with the orbit space, see e.g. Wijsman (1986,1990). It follows that $\mathcal{W}$ is also a space of maximal invariant, see e.g. Eaton (1989), p. 28, for definition of the maximal invariant. From assumption 4 it follows that the sample space is subject to the following factorization

$$\mathcal{Y} = G \times \mathcal{W} \tag{1}$$

Hence there is a bijection $y \leftrightarrow (g, w)$, $g \in G$, $w \in \mathcal{W}$, and a group G acts on $G \times \mathcal{W}$ according to the rule $\overline{g} \circ (g, w) := (\overline{g}g, w)$, for every $\overline{g} \in G$ (see e.g. Wijsman (1986,1990)).

[3] The orbit of $\mathcal{Y}$ is defined as $Gy := \{g \circ y \mid g \in G\} \subseteq \mathcal{Y}$, for each $y \in \mathcal{Y}$.

Since we shall work with measures we must ensure that there is a one–to–one correspondence between Borel subsets in $\mathcal{Y}$ and those in $G \times \mathcal{W}$ (i.e. bimeasurability)

Regularity condition (RC):

a) Let $\mathcal{Y}$ and Θ be complete separable metric locally compact topological spaces (i.e. Polish and locally compact). Moreover, let G be Polish locally compact topological group acting continuously on both $\mathcal{Y}$ and Θ.

b) For each $y \in \mathcal{Y}$ define the set $Gy := \{g \circ y \mid g \in G\}$. Let the bijective map $\gamma : \{y\} \times G \to Gy$, $\gamma_y(g) = g \circ y$ be a homeomorphism i.e. γ_y and γ_y^{-1} are continuous.

By RC, it is meaningful to state the next assumption

Assumption 5: A model $\mathcal{P}$ is dominated by a product measure $\mu_G \otimes \eta_{\mathcal{W}}$ (remember that we employ our unified notation for measures).

The density with respect to $\mu_G \otimes \eta_{\mathcal{W}}$ corresponding to P_θ will be denoted as $p(g,w \mid \theta)$.

Assumption 5 is very useful in our proof technique of the main theorem, yet it holds under our previous assumptions. In fact it can be proved. Here is an informal proof in the continuous case. Assume that a model is $G-$invariant. Usually there exists a density $p^*(y \mid \theta)$ with respect to Lebesgue measure i.e. $P_\theta(dy) = p^*(y \mid \theta)(dy)$. Evidently, Lebesgue measure is relatively left invariant with multiplier $\chi : G \to \mathbb{R}^+$ i.e. $d(g \circ y) = \chi(g)(dy)$, for every $g \in G$ (just compute the Jacobian). By our RC and using Theorem 2 in Bondar (1976), if $p^*(y \mid \theta)$ is a density with respect to (dy), then $\chi(g) \cdot p^*(g,w \mid \theta)$ is a density with respect to $\mu_G \otimes \lambda_{\mathcal{W}}$, where $\lambda_{\mathcal{W}}$ is some $\sigma-$finite (in general not a probability) measure on $\mathcal{W}$. But setting $\eta_{\mathcal{W}}(dw) = p(w \mid \theta)\lambda_{\mathcal{W}}(dw)$ where $p(w \mid \theta) = \int \chi(g) \cdot p^*(g,w \mid \theta)\mu_G(dg)$ we can easily verify that $p(g,w \mid \theta) = \chi(g) \cdot p^*(g,w \mid \theta) / p(w \mid \theta)$ is a density with respect to $\mu_G \otimes \eta_{\mathcal{W}}$ as stipulated by assumption 5. In fact, $p(g,w \mid \theta)$ is a version of the conditional density of g given w (and θ), so as

$$\int p(g,w \mid \theta)\mu_G(dg) = 1 \tag{2}$$

Under assumption 5, the $G-$invariance of a model implies that densities are subject to the following identity, see e.g. Zidek (1969) or Dawid et al. (1973)

$$p(g,w \mid \theta) = p(\overline{g}g, w \mid \overline{g} \circ \theta)\,;\ \text{ a.e. } [\mu_G \otimes \eta_{\mathcal{W}}], \qquad \text{for each } \overline{g} \in G \tag{3}$$

III. Strong Inconsistency

Let a model $\mathcal{P}$ be fixed. Assume that Bayesian delivers the posterior Π_y using some prior π (which may be such that $\pi(\Theta) = \infty$ provided that $m(y) < \infty$ a.e. $[\lambda_{\mathcal{Y}}]$). Formal definition of the strong inconsistency is as follows

Definition 1: *A Bayesian inference i.e.* $\{\Pi_y \mid y \in \mathcal{Y}\}$ *derived under specific* π, *is said to be strongly inconsistent* (*with a model* $\mathcal{P}$) *iff there is a bounded,* $(\mathcal{B}_{\mathcal{Y}} \times \mathcal{B}_{\Theta})$–*measurable function* $\phi : \mathcal{Y} \times \Theta \to \mathbb{R}$ *such that*:

$$\inf_{\theta} \int \phi(y,\theta) P(dy \mid \theta) > \sup_{y} \int \phi(y,\theta) \Pi(d\theta \mid y)$$

Strong inconsistency is a very undesirable (finite sample) property first noticed by Stone (1976). His ideas were formalized in the form of definition 1 by Lane and Sudderth (1983). For interpretation, intuition and discussion see Heath and Sudderth (1978), Sudderth (1994), Eaton and Freedman (2004), Eaton and Sudderth (2004), Eaton (2008). Moreover, Eaton and Freedman (2004) proved that strong inconsistency (in our countably additive setup) is equivalent to de Finetti's incoherence or existence of the Dutch book (that can be made against Π_y). Thus throughout the paper we will use the terms "consistency" and "coherence" interchangeably. Overall, by the compelling arguments used in the above cited works, an avoidance of Bayesian inference that is strongly inconsistent should be recommended.

We say that the Bayesian inference is consistent with a model $\mathcal{P}$ (in short, consistent), or equivalently that the Bayesian inference is coherent, iff for every bounded, $(\mathcal{B}_{\mathcal{Y}} \times \mathcal{B}_{\Theta})$–measurable function $\phi : \mathcal{Y} \times \Theta \to \mathbb{R}$ we have

$$\inf_{\theta} \int \phi(y,\theta) P(dy \mid \theta) \leq \sup_{y} \int \phi(y,\theta) \Pi(d\theta \mid y) \tag{4}$$

If $\pi(\Theta) < \infty$ (and π is countably additive)[4], the corresponding Bayesian inference is coherent, see e.g. Heath and Sudderth (1978). However the converse does not hold. There are coherent inferences which are based on improper prior: not every coherent inference may be derived from a proper (countably additive) prior see e.g. Lane and Sudderth (1983). The problem is to find a class of improper priors subject to some constraints on the inferential framework, which result in coherent Bayesian inference. This is the aim of our paper. To proceed further we only need the last assumption

Assumption 6: $\phi : \mathcal{Y} \times \Theta \to \mathbb{R}$ is $G-$invariant i.e. $\phi(g \circ y, g \circ \theta) = \phi(y, \theta)$, for every $g \in G$.

Assumption 6 is crucial to obtain the main theorem. We note that it is not universally acceptable but was also adopted in Eaton and Sudderth (1999,2002).

IV. A Prior Induced by the Right Haar Invariant Measure

Being in our model setup (defined with our assumptions), we know that the only class of improper priors that may entail coherent inference is the prior induced by the right Haar invariant measure, see Eaton and Sudderth (2002,2004) and Eaton and Freedman (2004). To prepare the ground for the next section we should clarify the notion of this prior. If a group G acts transitively and freely on a parameter space then there is a bijection between a group G and a parameter space Θ. Intuitively, the parameter space is just a group (seen as a space). Since by our RC a) a group G possesses a unique (up to a constant) right Haar invariant measure it is natural to find the induced measure on Θ. To this end, let us define a continuous bijective function $f : G \to \Theta$ defined as $f(g) = g \circ \theta_0$, where $\theta_0 \in \Theta$ may be chosen arbitrarily. If G is second countable (e.g. a subspace of $\mathbb{R}^n$, which will usually be the case in applications) then by our RC a) it follows that $f : G \to \Theta$ is a homeomorphism (see e.g. lemma 2.3.17 in Wijsman (1990)). In such a case we are in a position to define the induced measure (i.e. a prior) on Θ. If ν_G denotes the right Haar measure on G, the induced prior measure π on Θ is defined as $\pi(B) := \nu_G(f^{-1}(B))$, for all $B \in \mathcal{B}_\Theta$. Then

[4] The requirement of the countable additivity should be emphasized. Indeed, not every finitely additive (proper) prior has a posterior for a given model, see e.g. Heath and Sudderth (1989).

$$\int_{g\in f^{-1}(B)} h(g\circ\theta_0)\nu_G(dg) = \int_{\theta\in B} h(\theta)\pi(d\theta) \tag{5}$$

where h is any integrable function, see e.g. Lehmann (1986), p. 43. We note that the induced prior measure π is the same for any $\theta_0 \in \Theta$ used in f mapping (see e.g. Lehmann and Casella (1998), pp. 249–250). In fact when a group G acts transitively and freely on Θ, then the only measure that is independent of the reference point θ_0 is the right invariant Haar measure, see Villegas (1981). This constitutes the additional self–evident virtue of the right Haar invariant measure. Thus we may indifferently set $\theta_0 = e$ (this remark will be used in section VI). Putting $B = \Theta$ in (5) we have

$$\int_G h(g\circ\theta_0)\nu_G(dg) = \int_\Theta h(\theta)\pi(d\theta) \tag{6}$$

Then we say that the prior π is induced by the right Haar invariant measure. The formula (6) will prove essential in analytical integral manipulations.

V. Main Result

It is hard to talk about coherence in the case when the Bayesian inference is improper. One may say that propriety is a prerequisite of any Bayesian inference (not only a coherent one). The following lemma provides very useful property of the prior induced by the right Haar invariant measure. Namely, Bayesian inference derived under such a prior is proper

Lemma (Bondar (1977)): *Under assumptions 1, 2, 3, 4, 5 and* RC:
$\int p(y \mid g\circ\theta_0)\nu_G(dg) < \infty \;\; a.e. \; [\lambda_y]$

We are in a position to state the main result in our paper

Theorem: *Let a model $\mathcal{P}$ be given. Under assumptions 1, 2, 3, 4, 5, 6 and* RC, *the Bayesian inference derived under the prior induced by the right Haar invariant measure on G is consistent with a model $\mathcal{P}$ (coherent).*

Proof: see appendix.

VI. STONE'S EXAMPLE

It is inevitable in the context of our theorem to discuss the Stone's (1976) example B. The reason is that this example, when superficially taken, may be considered as a counter–example to our theorem. In fact this example convinced many researchers that amenability of a group is needed to reach the coherence of Bayesian inference. We argue that this is not true. Instead what is crucial for coherence (consistency) of Bayesian inference (except the transitivity assumption) is the $G-\Theta$ freeness assumption.

Stone (1976) in his example B considers the following model

$$y_i = \theta u_i, \qquad i = 1,2. \tag{7}$$

where $y_i = (y_{1i}, y_{2i})'$ is vector of observations and $u_i \sim N(0, \mathrm{I}_2)$. Moreover, $\theta \in \mathbb{R}^{2\times 2}$ is assumed to be nonsingular. Let us write $y = [y_1, y_2] \in \mathcal{Y}$. Then the data sampling density with respect to Lebesgue measure reads

$$p(y \mid \theta) = k \cdot |\,\theta\,|^{-2}\ \mathrm{etr}\{-\tfrac{1}{2} yy'(\theta\theta')^{-1}\} \tag{8}$$

where k is a normalizing constant and $\mathrm{etr} := \exp\{tr\{\cdot\}\}$. As a natural group acting in the model (7) we take $G = GL_2$. Evidently, (7) is GL_2-invariant with the action of GL_2 on the sample space defined as $y \mapsto g \circ y := gy$ and on the parameter space $\theta \mapsto g \circ \theta := g\theta$, $g \in GL_2$. The prior induced by the right (= left) Haar invariant measure on GL_2 is $\pi(d\theta) = |\,\theta\,|^{-2}\,(d\theta)$. To see this, put $\theta_0 = e \equiv \mathrm{I}_2$ in (6) (I_2 is the identity element in the group GL_2) and note that $g \circ e := g\,\mathrm{I}_2 = g$ and $\nu_{GL_2}(dg) = |\,g\,|^{-2}\,(dg)$ (see e.g. Eaton (1989), p. 9). In this case the posterior distribution of θ is

$$\Pi(d\theta \mid y) \propto |\,\theta\,|^{-4}\ \mathrm{etr}\{-\tfrac{1}{2} yy'(\theta\theta')^{-1}\}(d\theta) \tag{9}$$

Denote the posterior density with respect to Lebesgue measure as

$$p(\theta \mid y) \propto |\,\theta\,|^{-4}\ \mathrm{etr}\{-\tfrac{1}{2} yy'(\theta\theta')^{-1}\} \tag{10}$$

Now Stone uses a one–to–one decomposition $y = TC$, where $T \in LT_2^+$ and $C \in O_m$. Inserting $y = TC$ into (10) we get

$$p(\theta \mid T, C) = p(\theta \mid T) \propto |\theta|^{-4} \operatorname{etr}\{-\tfrac{1}{2}TT'(\theta\theta')^{-1}\} \tag{11}$$

Then Stone pretends that the model is

$$y \equiv T = \begin{bmatrix} y_{11} & 0 \\ y_{21} & y_{22} \end{bmatrix} = \theta[u_1, u_2] \tag{12}$$

and argues that the prior for θ induced by the right Haar invariant measure on GL_2 leads to strong inconsistency of Bayesian inference. There are two closely related flaws in this reasoning. The first one was suggested in a comment on Stone's article i.e. Pratt (1976). That is the model (7) is unidentified i.e. $p(y \mid \theta) = p(y \mid \theta h)$, for all $h \in O_m$. In a rejoinder, Stone (1976) comments on Pratt's remark: "*Pratt asks about the significance of the nonidentifiability of θ. I am sure it is what makes the example work*". In our opinion, this is only partially true. We believe that it is nonidentifiability together with unrecognizing the effect of changing the pattern of the data matrix y (to the triangular one) that makes the example work. Let us clarify this claim. Leaving the nonidentifiability of θ aside, when $y \in LT_2^+$ it is no longer true that the natural group operating in the example is still GL_2. When the sample space is LT_2^+, the only group that preserves the pattern of the data is LT_2^+ (or its subgroup). Hence we must assume that the model is LT_2^+-invariant. Then nonidentifiability of θ enters the scene. Essentially, being in our invariance framework, there is only way to deal with it. We should assume that $\theta \in LT_2^+$, which makes the model identified. Then we arrive exactly at our setup i.e. $G = LT_2^+$ acts freely on the sample space and transitively and freely on the parameter space. In such a case the prior induced by the right Haar invariant measure on LT_2^+ (see below for an explicit formulation) makes the Bayesian inference consistent. This is ensured by our theorem. But it is so not because the group LT_2^+ is amenable (which is), but because its action on the parameter space is free.

When $\theta \in LT_2^+$ we could alternatively parameterize the density $p(y \mid \theta)$ in terms of positive definite covariance $\Sigma = \theta\theta'$, which is isomorphic to $\theta \in LT_2^+$. Thus

$$p(y \mid \Sigma) = k \cdot |\Sigma|^{-1} \operatorname{etr}\{-\tfrac{1}{2}yy'\Sigma^{-1}\} \tag{13}$$

It may be useful to find the implied prior for Σ from that for θ (which in turn is induced by the right Haar invariant measure). Since it is of some interest we derive the prior in the general case where $y:(m\times n)=[y_1,\ldots,y_n]\in\mathcal{Y}$, hence $\theta\in LT_m^+$. In order to satisfy the $G-\mathcal{Y}$ freeness assumption we must have $n\geq m$ and $\operatorname{rank}(y)=m$. Note that $\nu_{LT_m^+}(dg)=\prod_{i=1}^m g_{ii}^{-(m-i+1)}\cdot(dg)$, where g_{ii} are diagonal elements of $g\in LT_m^+$ (see e.g. Eaton (1989), p. 17). The Jacobian from $\theta\in LT_m^+$ to $\Sigma\in PD_m$ is given by $J(\theta\to\Sigma)=2^{-m}\cdot\prod_{i=1}^m\theta_{ii}^{-(m-i+1)}=2^{-m}\cdot\prod_{i=1}^m|\,\Sigma^{[1\mapsto i]}\,|^{-\frac{1}{2}}$, where $\Sigma^{[1\mapsto i]}:=(\sigma_{jk})$; $j,k=1,\ldots,i$ ($\Sigma^{[1\mapsto i]}$ is a leading principal submatrix of Σ consisting the first i rows and columns of Σ). Then using results from section IV and by the usual integral transformation technique

$$\int p(y\mid g\circ\theta_0)\nu_{LT_m^+}(dg)=\int p(y\mid g\theta_0)\nu_{LT_m^+}(dg) \tag{14}$$
$$=\int p(y\mid g)\nu_{LT_m^+}(dg)\quad\text{[definition of the right Haar integral, see e.g. Eaton (1989), p. 7]}$$
$$\equiv\int p(y\mid\theta)\nu_{LT_m^+}(d\theta)=\int_{LT_m^+}p(y\mid\theta)\cdot\prod_{i=1}^m\theta_{ii}^{-(m-i+1)}(d\theta)=2^{-m}\int_{PD_m}p(y\mid\Sigma)\cdot\prod_{i=1}^m|\,\Sigma^{[1\mapsto i]}\,|^{-1}(d\Sigma)$$

It follows that the prior for the covariance $\Sigma\in PD_m$ (induced by the right Haar invariant measure on LT_m^+) is

$$\pi(d\Sigma)\propto\prod_{i=1}^m|\,\Sigma^{[1\mapsto i]}\,|^{-1}(d\Sigma) \tag{15}$$

When used in Stone's example B and assuming $\theta\in LT_2^+$, (15) can not lead to strong inconsistency. In particular, (15) is exactly the prior recommended by Eaton and Sudderth (2010), proposition 4.1, for an $m-$variate normal model with mean 0 and covariance Σ (written in a more elegant form). The use of (15) makes Bayesian inference consistent with a multivariate normal model, yet as emphasized by Eaton and Sudderth (1998) and evident from (15), it can not be cast in the form $\pi(d\Sigma)\propto|\,\Sigma\,|^{\alpha}\,(d\Sigma)$, for any $\alpha\in\mathbb{R}$ ($m>1$). Needless to say, staying within theoretical framework of Eaton and Sudderth (1998,1999,2002), the Jeffreys' prior i.e. $\pi(d\Sigma)\propto|\,\Sigma\,|^{-\frac{1}{2}(m+1)}\,(d\Sigma)$, is prohibited being strongly inconsistent (incoherent).

However as noted by Eaton and Sudderth (1999,2010) themselves, there is one conundrum connected with the above reasoning (which was also noticed much earlier by Stone (1965)). When one parameterizes the model (7) with $\theta\in LT_2^+$, it becomes only LT_2^+–invariant. In particular since permutation matrix does not belong to LT_2^+, the prior (15) is not invariant under the permutation of variables in the model. But this is an absolute minimum we require from the prior that possesses any attribute of invariance. Hence the prior (15) is consistent only with a model that is LT_2^+–invariant. The reason why the available theory is helpless to resolve satisfactorily this puzzle is the fact that GL_m is not amenable (for $m > 1$) and amenability of a group is among the most important conditions to get the coherent inference. Since the amenability plays no role in our framework one may hope for some resolution. Indeed, this is the case.

We will demonstrate that standard Jeffreys' prior for Σ may be consistent and that nonidentifiability of a model has nothing to do with the consistency. Without loss of generality, consider a variant of the model (7) in the form

$$\beta y_i = u_i, \qquad i = 1,2. \tag{16}$$

where $\beta \in \mathbb{R}^{2\times 2}$ is nonsingular. Clearly, if juxtaposed with (7), we identify $\beta = \theta^{-1}$. Incidentally, (16) is the model considered by Villegas (1971) and more importantly a simplistic version of the Structural VAR (SVAR) model (with no lags). Since (16) accommodates the SVAR specification, from the economic standpoint, (16) is more natural parameterization than (7) is. The model (16) is still GL_2–invariant with the same action on the sample space as in (7) but the (left) action of GL_2 on the parameter space is defined as $\beta \mapsto g\circ\beta := \beta g^{-1}$, where $g \in GL_2$. Of course (16) remains unidentified. Now the data sampling density with respect to Lebesgue measure reads

$$p(y \mid \beta) = k \cdot |\, \beta \,|^2 \ \mathrm{etr}\{-\tfrac{1}{2} yy'\beta'\beta\} \tag{17}$$

As before, let us consider the general case of (16) where $y : (m \times n) = [y_1, \dots, y_n] \in \mathcal{Y}$, hence $\beta \in GL_m$. Moreover, to have a free action of GL_m on the sample space we assume $n \geq m$ and $\mathrm{rank}(y) = m$. Since $\mu_{GL_m}(dg) = \nu_{GL_m}(dg) = |\, g \,|^{-m}\ (dg)$ (see e.g. Eaton (1989), p. 9), denoting $W = \beta'\beta$ we get

$$\int p(y \mid g\circ\beta_0)\nu_{GL_m}(dg) = \int p(y \mid g\circ e)\nu_{GL_m}(dg) \qquad \text{[see section IV]} \tag{18}$$

$$= \int p(y \mid g^{-1})\nu_{GL_m}(dg) \qquad [\, g\circ\beta := \beta g^{-1},\ e \equiv \mathrm{I}_m \,]$$

$$= \int p(y \mid g)\mu_{GL_m}(dg) \qquad \text{[see e.g. Nachbin (1965), p. 80]}$$
$$\equiv \int p(y \mid \beta)\mu_{GL_m}(d\beta)$$
$$= \int_{GL_m} p(y \mid \beta) \mid \beta'\beta \mid^{-\frac{1}{2}m} (d\beta) = \pi^{\frac{1}{2}m^2}[\Gamma_m(\tfrac{m}{2})]^{-1} \cdot \int_{PD_m} p(y \mid W) \mid W \mid^{-\frac{1}{2}(m+1)} (dW)$$

where the last equality follows by noting that $p(y \mid \beta)$ is only a function of $\beta'\beta$ and using Hsu lemma (see e.g. Anderson (2003), p. 539), where $\Gamma_m(\cdot)$ denotes the multivariate gamma function i.e. $\Gamma_m(a) = \pi^{\frac{1}{4}m(m-1)}\prod_{i=1}^{m}\Gamma(a - \frac{i-1}{2})$. Further, since the covariance matrix is equal to W^{-1} i.e. $\Sigma = W^{-1}$, taking into account the Jacobian $J(W \to \Sigma) = \mid \Sigma \mid^{-(m+1)}$, we obtain

$$\int p(y \mid \beta)\mu_{GL_m}(d\beta) = \pi^{\frac{1}{2}m^2}[\Gamma_m(\tfrac{m}{2})]^{-1} \cdot \int_{PD_m} p(y \mid \Sigma) \mid \Sigma \mid^{-\frac{1}{2}(m+1)} (d\Sigma) \tag{19}$$

The formulas (18) and (19) may be summarized as follows. The right Haar invariant measure on GL_m in extension of (16) to $m-$variate case, induces the left Haar invariant prior on the parameter space i.e. $\pi(d\beta) \propto \mid \beta \mid^{-m} (d\beta)$. The latter implies the Jeffreys' prior $\pi(d\Sigma) \propto \mid \Sigma \mid^{-\frac{1}{2}(m+1)} (d\Sigma)$ in an $m-$variate normal model with mean 0 and covariance Σ. More importantly, since the conditions of our theorem hold, both the prior $\pi(d\beta) \propto \mid \beta \mid^{-m} (d\beta)$ and the implied Jeffreys' prior $\pi(d\Sigma) \propto \mid \Sigma \mid^{-\frac{1}{2}(m+1)} (d\Sigma)$, lead to coherent Bayesian inference[5]. Needless to say, in contrast to (15) both $\pi(d\beta) \propto \mid \beta \mid^{-m} (d\beta)$ and the implied Jeffreys' prior are invariant with respect to permutation of the variables in a model. To the extent that the formal arguments for use of the Jeffreys' prior in a multivariate model with zero mean and unknown covariance "justify" its adoption in general linear regressions, SEM and VAR models, our coherence arguments "justify" the use of the prior $\pi(dA, dA_1, \ldots, dA_p) \propto \mid A \mid^{-m} (dA)(dA_1)\ldots(dA_p)$ in the general SVAR model

$$Ay_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t\,; \qquad t = 1,\ldots,T \tag{20}$$

where $y_t \in \mathbb{R}^m$, $A : (m \times m)$ nonsingular, $A_i : (m \times m)$ and $\varepsilon_t \mid y_{t-1}, y_{t-2}, \ldots \sim N(0, \mathrm{I}_m)$.

[5] The assertion applies to a generalization of (16), when $y : (m \times n) = [y_1, \ldots, y_n] \in \mathcal{Y}$, $\beta \in GL_m$. The corresponding priors for (16) proper are given by setting $m = 2$.

VII. CONCLUSION

We have established conditions which guarantee that Bayesian inference is consistent and coherent. In comparison with analogous conditions available in the literature we have replaced the technical requirement of the group amenability with more pleasant $G-\Theta$ freeness assumption. In consequence our conditions allow us to cope with cases where the underlying group operating in a model is not amenable (e.g. a group of general nonsingular matrices). In this sense, our framework may be more useful in practice.

In the course of our analysis we reexamined the Stone's (1976) example pointing to some flaws in his arguments. In the context of this example we encountered the prior for multivariate normal model recently proposed by Eaton and Sudderth (2010). The latter was given in more intuitive terms than in the original source. Lastly, we demonstrated that the Jeffreys' prior in the multivariate normal model may be coherent. This last (seemingly innocent) conclusion is in fact quite new in the literature. We also proposed the prior for the Structural VAR models that result in the coherent inference.

We think that attractiveness of our approach lies also in its simplicity. Compare the analogous framework in Eaton and Sudderth (1999,2002). Of course this happens at the cost of our restrictive $G-\Theta$ freeness assumption.

APPENDIX:

We denote with a bar above those elements in integrals that are fixed with respect to the integration process. This will facilitate to keep track of the algebraic manipulations used below. We need to prove (4). To this end we have

$$\int \phi(\overline{y},\theta)\Pi(d\theta \mid \overline{y}) = \frac{\int \phi((\overline{g},\overline{w}), g\circ\theta_0)p(\overline{g},\overline{w} \mid g\circ\theta_0)\nu_G(dg)}{\int p(\overline{g},\overline{w} \mid g\circ\theta_0)\nu_G(dg)} \qquad \text{[by (6)]}$$

Note that the denominator exists by our lemma so that Π_y is proper which means that $\int \phi(\overline{y},\theta)\Pi(d\theta \mid \overline{y}) < \infty$. Continuing

$$\int \phi(\overline{y},\theta)\Pi(d\theta \mid \overline{y}) = \frac{\int \phi((\overline{g},\overline{w}), g\circ\theta_0)p(\overline{g},\overline{w} \mid g\circ\theta_0)\nu_G(dg)}{\int p(\overline{g},\overline{w} \mid g\circ\theta_0)\nu_G(dg)}$$

$$= \frac{\int \phi((g^{-1}\overline{g},\overline{w}),\theta_0)p(g^{-1}\overline{g},\overline{w} \mid \theta_0)\nu_G(dg)}{\int p(g^{-1}\overline{g},\overline{w} \mid \theta_0)\nu_G(dg)} \qquad \text{[by (3) and assumption 6]}$$

$$= \frac{\int \phi((g\overline{g},\overline{w}),\theta_0)p(g\overline{g},\overline{w} \mid \theta_0)\mu_G(dg)}{\int p(g\overline{g},\overline{w} \mid \theta_0)\mu_G(dg)} \qquad \text{[see e.g. Nachbin (1965), p. 80]}$$

$$= \frac{\Delta(\overline{g}^{-1})\cdot\int \phi((g,\overline{w}),\theta_0)p(g,\overline{w} \mid \theta_0)\mu_G(dg)}{\Delta(\overline{g}^{-1})\cdot\int p(g,\overline{w} \mid \theta_0)\mu_G(dg)} \qquad \text{[see e.g. Eaton (1989), p. 7]}$$

$$= \frac{\int \phi((g,\overline{w}),\theta_0)p(g,\overline{w} \mid \theta_0)\mu_G(dg)}{\int p(g,\overline{w} \mid \theta_0)\mu_G(dg)}$$

In the next to last line, $\Delta : G \to \mathbb{R}^+$ denotes the right hand modulus of G, see e.g. Eaton (1989), p. 7.

By expression (2)

$$\int \phi(\overline{y},\theta)\Pi(d\theta \mid \overline{y}) = \int \phi((g,\overline{w}),\theta_0)p(g,\overline{w} \mid \theta_0)\mu_G(dg) \equiv f(\overline{w})$$

From the latter we conclude $\int \phi(\overline{y},\theta)\Pi(d\theta \mid \overline{y})$ does not depend on g, hence

$$\sup_y \int \phi(y,\theta)\Pi(d\theta \mid y) = \sup_{g,w} \int \phi((g,w),\theta)\Pi(d\theta \mid g,w) = \sup_w \int \phi((g,w),\theta_0)p(g,w \mid \theta_0)\mu_G(dg)$$

Since

$$\int \phi(y,\overline{\theta})P(dy \mid \overline{\theta}) = \int \phi((g,w),\overline{g}\circ\theta_0)p(g,w \mid \overline{g}\circ\theta_0)\mu_G(dg)\eta_{\mathcal{W}}(dw) =$$

$$= \int \phi((\overline{g}^{-1}g,w),\theta_0)p(\overline{g}^{-1}g,w \mid \theta_0)\mu_G(dg)\eta_{\mathcal{W}}(dw) \qquad \text{[by (3) and assumption 6]}$$

$$= \int \phi((g,w),\theta_0)p(g,w \mid \theta_0)\mu_G(dg)\eta_{\mathcal{W}}(dw) \qquad \text{[see e.g. Eaton (1989), p. 6]}$$

$$= \int \phi(y,\theta_0)P(dy \mid \theta_0)$$

It follows that $\int \phi(y,\theta)P(dy \mid \theta)$ does not depend on θ (θ_0 was arbitrary). Thus

$$\inf_{\theta}\int\phi(y,\theta)P(dy\mid\theta)=\int\phi(y,\theta)P(dy\mid\theta)=\int\phi((g,w),\theta_0)p(g,w\mid\theta_0)\mu_G(dg)\eta_{\mathcal{W}}(dw)\leq$$

$$\leq\{\sup_{w}\int\phi((g,w),\theta_0)p(g,w\mid\theta_0)\mu_G(dg)\}\cdot\int\eta_{\mathcal{W}}(dw)=\sup_{w}\int\phi((g,w),\theta_0)p(g,w\mid\theta_0)\mu_G(dg)$$

$$=\sup_{y}\int\phi(y,\theta)\Pi(d\theta\mid y)\qquad\qquad Q.E.D.$$

REFERENCES:

Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, *Third edition*, John Wiley & Sons, New Jersey.

Bondar, J.V. (1976), "Borel Cross–Sections and Maximal Invariants", *The Annals of Statistics*, 4, pp. 866–877.

Bondar, J.V. (1977), "A Conditional Confidence Principle", *The Annals of Statistics*, 5, pp. 881–891.

Bondar, J.V. and P. Milnes (1981), "Amenability: A Survey for Statistical Applications of Hunt–Stein and Related Conditions on Groups", *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* , 57, pp. 103–128.

Chang, T. and D. Eaves (1990), "Reference Priors for the Orbit in a Group Model", *The Annals of Statistics*, 18, pp. 1595–1614.

Dawid, A.P., M. Stone and J.V. Zidek (1973), "Marginalization Paradoxes in Bayesian and Structural Inference", *Journal of the Royal Statistical Society*, *series B*, 35, pp. 189–213.

Drèze, J.H. (1976), "Bayesian Limited Information Analysis of the Simultaneous Equations Model", *Econometrica*, 44, pp. 1045–1075.

Eaton, M.L. (1989), *Group Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics, vol. 1., Institute of Mathematical Statistics, Hayward, California.

Eaton, M.L. (2008), "Dutch Book in Simple Multivariate Normal Prediction: Another Look", in: D. Nolan and T. Speed, eds., *Probability and Statistics*: *Essays in Honor of David A. Freedman,* Institute of Mathematical Statistics, Beachwood, Ohio.

Eaton, M.L. and D.A. Freedman (2004), "Dutch Book Against Some 'Objective' Priors", *Bernoulli*, 10, pp. 861–872.

Eaton, M.L. and W.D. Sudderth (1993), "Prediction in a Multivariate Normal Setting: Coherence and Incoherence, *Sankhyā* , *series A*, 55, pp. 481–493.

Eaton, M.L. and W.D. Sudderth (1995), "The Formal Posterior of a Standard Flat Prior in MANOVA is Incoherent", *Journal of the Italian Statistical Society*, 2, pp. 251–270.

Eaton, M.L. and W.D. Sudderth (1998), "A New Predictive Distribution for Normal Multivariate Linear Models", *Sankhyā* , *series A*, 60, pp. 363–382.

Eaton, M.L. and W.D. Sudderth (1999), "Consistency and Strong Inconsistency of Group–Invariant Predictive Inferences", *Bernoulli*, 5, pp. 833–854.

Eaton, M.L. and W.D. Sudderth (2002), "Group Invariant Inference and Right Haar Measure", *Journal of Statistical Planning and Inference*, 103, pp. 87–99.

Eaton, M.L. and W.D. Sudderth (2004), "Properties of Right Haar Predictive Inference, *Sankhyā* , 66, pp. 487–512.

Eaton, M.L. and W.D. Sudderth (2010), "Invariance of Posterior Distributions Under Reparametrization", *Sankhyā* , *series A*, 72, pp. 101–118.

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

Heath, D. and W.D. Sudderth (1978), "On Finitely Additive Priors, Coherence and Extended Admissibility, *The Annals of Statistics*, 6, pp. 333–345.

Heath, D. and W.D. Sudderth (1989), "Coherent Inference from Improper and from Finitely Additive Priors", *The Annals of Statistics*, 17, pp. 907–919.

Helland, I.S. (2004), "Statistical Inference under Symmetry", *International Statistical Review*, 72, pp. 409–422.

Keyes, T.K and M.S. Levy (1996), "Goodness of Prediction Fit for Multivariate Linear Models", *Journal of the American Statistical Association*, 91, pp. 191–197.

Lane, D.A and W.D. Sudderth (1983), "Coherent and Continuous Inference", *The Annals of Statistics*, 11, pp. 114–120.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, *Second edition*, Springer–Verlag, New York.

Lehmann, E.L. and G. Casella (1998), *Theory of Point Estimation*, *Second edition*, Springer–Verlag, New York.

Nachbin, L. (1965), *The Haar Integral*, D. Van Nostrand Company, Princeton, New Jersey.

Pratt, J.W. (1976), "Comment" on "Strong Inconsistency from Uniform Priors" by M. Stone, *Journal of the American Statistical Association*, 71, pp. 119–120.

Stone, M. (1965), "Right Haar Measure for Convergence in Probability to Quasi Posterior Distributions", *Annals of Mathematical Statistics*, 36, pp. 440–453.

Stone, M. (1976), "Strong Inconsistency from Uniform Priors" with discussion, *Journal of the American Statistical Association*, 71, pp. 114–125.

Sudderth, W.D. (1994), "Coherent Inference and Prediction in Statistics", in: D. Prawitz, B. Skyrms and D. Westerståhl, eds., *Logic*, *Methodology and Philosophy of Science IX*, Elsevier Science, Amsterdam.

Villegas, C. (1971), "On Haar Priors", in: V.P. Godambe and D.A. Sprott, eds., *Foundations of Statistical Inference*, Holt, Rinehart and Winston, Toronto.

Villegas, C. (1981), "Inner Statistical Inference II", *The Annals of Statistics*, 9, pp. 768–776.

Wijsman, R. A. (1986), "Global Cross Sections as a Tool for Factorization of Measures and Distribution of Maximal Invariants", *Sankhyā* , *series A*, 48, pp. 1–42.

Wijsman, R. A. (1990), *Invariant Measures on Groups and Their Use in Statistics*, Institute of Mathematical Statistics Lecture Notes–Monograph Series, vol. 14., Hayward, California.

Yang, R. and J.O. Berger (1994), "Estimation of a Covariance Matrix Using the Reference Prior", *The Annals of Statistics*, 22, pp. 1195–1211.

Zellner, A. (1977), "Maximal Data Information Prior Distributions", in: A. Aykac and C. Brumat, eds., *New Developments in the Applications of Bayesian Methods*, North–Holland.

Zidek, J.V. (1969), "A Representation of Bayes Invariant Procedures in Terms of Haar Measure", *Annals of the Institute of Statistical Mathematics*, 21, pp. 291–308.