Reconciliation, Coherence, and P-Values

George Casella[1]

Martin T. Wells[2]

Cornell University, Ithaca, NY 14853

1

Summary

For the one-sided testing problem, necessary and sufficient conditions are given for reconciliation of evidence. That is, conditions under which the frequentist p-value is equal to a Bayesian posterior probability are detailed. Moreover, it is also observed that when reconciliation is obtained, it must result in a coherent inference.

1. Introduction

In hypothesis testing, evidence is based on a post-experimental measure of evaluation which, to a frequentist, takes the form of the observed level of significance, that is, the p-value. To a Bayesian, evidence is based on the assessed posterior probability of the null hypothesis (H_0). It is shown in Casella and Berger (1987) that in the location problem it is possible for a frequentist and Bayesian to reconcile their measures of evidence assessment. They show that for a wide class of prior distributions the infimum of the Bayesian posterior probability of H_0 is equal to the p-value; in other cases the infimum is less than the p-value.

For the one-sided testing problem

$$(1.1) \qquad H_0 : \theta \leq \theta_0 \qquad \text{vs.} \qquad H_1 : \theta > \theta_0 \,,$$

Casella and Berger (1987) show that if $X = x$ is observed, where $X \sim f(x-\theta)$, then the frequentist p-value, $p(x)$, is equal to a Bayesian posterior probability using a Lebesgue prior. We have

$$(1.2) \qquad p(x) = P(X \geq x | \theta_0) = \int_x^\infty f(t - \theta_0)dt = \int_{-\infty}^{\theta_0} f(x-\theta)d\theta = P(\theta \leq \theta_0 | x)$$

where we see that $f(x-\theta)$ is also the posterior distribution for θ starting from $\pi(\theta)d\theta = d\theta$.

After hearing a talk about the above result, Morris DeGroot noted that this sufficient condition was also necessary for reconciliation of evidence in the location case. This follows because, if evidence can be reconciled, there exists a prior distribution $\pi(\theta)$ that yields $p(x) = P(\theta \leq \theta_0 | x)$. In the location case this means we must have

$$(1.3) \qquad \int_x^\infty f(x-\theta_0)dx = \frac{\displaystyle\int_{-\infty}^{\theta_0} f(x-\theta)\pi(\theta)d\theta}{\displaystyle\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta)d\theta} = \frac{\displaystyle\int_{-\infty}^{\theta_0} f(x-\theta)\pi(\theta)d\theta}{m_\pi(x)} \,,$$

where $m_\pi(x)$ is the marginal distribution of x. Now (1.3) must hold for every value of θ_0, and if we differentiate both sides of (1.3) with respect to θ_0, we have

$$(1.4) \qquad f(x-\theta_0) = \frac{f(x-\theta_0)\pi(\theta_0)}{m_\pi(x)} \,,$$

so it follows that $\pi(\theta_0) = m_\pi(x)$, which only happens if $\pi(\theta) = $ constant. Thus, in the location case, evidence can be reconciled if and only if $\pi(\theta) = $ constant.

After outlining the above argument, Professor DeGroot then asked if, in general, necessary conditions for reconciliation could be deduced. This paper answers that question.

The main result of the paper is that evidence can be reconciled if and only if $\pi(\theta)$ is the right-invariant Haar measure of the underlying group G. Furthermore, it follows from the results of Heath and Suddereth (1989), that the inferences based on right-invariant priors are coherent. Thus, if evidence can be reconciled, the p-value will give coherent inferences.

In Section 2 we define needed notation, and Section 3 contains the main theorem, the equivalence of the p-value and the posterior probability from the right-invariant prior. Section 4 contains some examples, and Section 5 contains a concluding discussion, and shows the relationship between evidence reconciliation and coherence. An appendix contains some necessary technical lemmas.

2. Preliminaries

The random variable X is assumed to take values in a space $\mathfrak{X}$, having density $f(x|\theta)$ with respect to a σ-finite measure μ. The unknown parameter is $\theta \in \Theta \subset \mathbb{R}^1$. We wish to test

$$(2.1) \qquad\qquad H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \notin \Theta_0 .$$

Specifically, we let $\Theta_0 = (-\infty, \theta_0)$. In addition assume:

(i) There is a group G acting on $\mathfrak{X}$ which induces a group $\bar{G}$ on Θ, with group identities e and $\bar{e}$, respectively.

(ii) G and $\bar{G}$ are isomorphic and are locally compact Hausdorff topological groups.

(iii) $\bar{g}\Theta_0 = \Theta_0$, $\bar{g}\Theta = \Theta$.

(iv) $\displaystyle\int_G f(x|\bar{g}\theta) d\mu^r(g) < \infty$ a.s., where μ is a right invariant measure on G.

(v) The measure μ is relatively invariant under the action of G with multiplier ν; that is, ν is a continuous homomorphism from G to $(0, \infty)$ and $\mu(gA) = \mu(A)\nu(g)$ for all measurable subsets A of $\mathfrak{X}$.

(vi) $f(gx|\bar{g}\theta) = f(x|\theta)\nu(g)$.

For examples of groups that are used in statistics see Berger (1985, Chapter 6). The following definition is a standard one for invariant testing problems.

Definition: The testing problem (2.1), with density $f(x|\theta)$, is *invariant* under G if (iii) and (vi) are satisfied.

Although our discussion is for testing problems with $\theta \in \mathbb{R}$, it is clear that analogous results will hold for vector-valued parameters. In this case the rejection region must be redefined as a region in space rather than an interval. Also, in Section 3 we will show how to proceed in the presence of nuisance parameters.

3. Reconciling Evidence

In this section we give the main theorem, which characterizes the equivalence between p-values and posterior probabilities. Recall that $\Theta_0 = (-\infty, \theta_0)$, so a test of $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$ often have critical region of the form $R_\alpha = \{x : x > c_\alpha\}$, for some constant c_α. (Canonical exponential families and location families will have critical region of the form of R_α.) For such regions, the p-value is given by

$$(3.1) \qquad p(x) = P_{\theta_0}(X \geq x) = \int_x^\infty f(x|\theta_0) dx .$$

A simple transformation gives

$$(3.2) \qquad p(x) = \int_{x \cdot \theta_0^{-1}}^\infty f(y|\bar{e}) dy$$

where "$\cdot$" is the group operation and "-1" is the group inverse.

For a Bayesian approach to the testing problem, we use a prior π. Then a Bayes rule (posterior probability) is

$$(3.3) \qquad \delta_B^\pi(x) = \frac{1}{m_\pi(x)} \int_{\Theta_0} f(x|\theta)\pi(\theta) d\theta ,$$

where m_π is the marginal density of x,

$$(3.4) \qquad m_\pi(x) = \int_\Theta f(x|\theta)\pi(\theta) d\theta .$$

The goal is to show $p(x) = \delta_B^\pi(x)$. Define the ratio $\lambda(\theta) = \pi(\theta)/\pi^r(\theta)$, where π^r is a right invariant Haar measure. In what follows, we will examine the structure of δ_B^π in greater detail. We can write

$$
\begin{aligned}
\delta_B^\pi(x) &= \frac{1}{m_\pi(x)} \int_{\Theta_0} f(x|\theta)\pi(\theta) d\theta \\
&= \frac{1}{m_\pi(x)} \int_{\Theta_0} f(x|\theta)\lambda(\theta)\pi^r(\theta) d\theta \\
&= \frac{K}{m_\pi(x)} \int_{\Theta_0^{-1}} f(x|y^{-1})\lambda(y^{-1})\pi^\ell(y) dy \qquad \text{(by Lemma A.1)} \\
&= \frac{K}{m_\pi(x)} \int_{\Theta_0^{-1}} f(yx|\bar{e})\lambda(y^{-1}) J_Y^\ell(x)\pi^\ell(y) dy \qquad \text{(by Lemma A.3)}
\end{aligned}
$$

(3.5)

$$= \frac{K}{m_\pi(x)} \int_{\Theta_0^{-1}} f(yx|\bar{e})\lambda(y^{-1}) \frac{\pi^\ell(x)}{\pi^\ell(yx)} \pi^\ell(y)dy \qquad \left(\begin{array}{l} \text{definition of } \pi^\ell, \text{ the left-} \\ \text{invariant Haar measure} \end{array} \right)$$

$$= \frac{K\nu(x)\pi^\ell(x)}{m_\pi(x)} \int_{x \cdot \Theta_0^{-1}} f(y|\bar{e})\lambda(y^{-1}x^{-1})dy . \qquad \text{(by Lemma A.2)}$$

$$\left(\text{In Lemma A.2 let } t(yg) \equiv s(yg)\lambda\big((y)^{-1}\big). \text{ Therefore } t(y) \equiv s(y)\lambda^*(y) = s(y)\lambda(y^{-1}g^{-1}). \right)$$

Using a similar argument it follows that

$$(3.6) \qquad m_\pi(x) = K\nu(x)\pi^\ell(x) \int_{\mathfrak{G}} f(y|\bar{e})\lambda(y^{-1}x^{-1})dy \equiv K\nu(x)\pi^\ell(x)C_\lambda(x) .$$

Note that in the special case where $\pi = \pi^r$ (that is, $\lambda \equiv 1$) we have the following reductions:

(i) $$\qquad m_{\pi^r}(x) = K\nu(x)\pi^\ell(x) \qquad (\text{i.e., } C_\lambda(x) \equiv 1 \ \forall x) ;$$

(ii) $$\qquad \delta_B^{\pi^r}(x) = \int_{x \cdot \Theta_0^{-1}} f(y|\bar{e})dy ,$$

and hence $\delta_B^{\pi^r}(x) \equiv p(x)$. Thus, if $p = \delta_B^\pi$ for an arbitrary prior π we need $\delta_B^{\pi^r} = \delta_B^\pi$ to occur. It can also be seen

$$\delta_B^{\pi^r} = \delta_B^\pi \ \Leftrightarrow \ \int_{x \cdot \Theta_0^{-1}} f(y|\bar{e}) = \frac{1}{C_\lambda(x)} \int_{x \cdot \Theta_0^{-1}} f(y|\bar{e})\lambda(y^{-1}x^{-1}) \qquad \forall x$$

$$\Leftrightarrow \qquad 1 = \frac{\lambda(y^{-1}x^{-1})}{C_\lambda(x)} \qquad \forall x$$

$$\Leftrightarrow \quad \pi(y^{-1}x^{-1}) = C_\lambda(x)\pi^r(y^{-1}x^{-1})$$

$$\Leftrightarrow \quad \pi \text{ is proportional to } \pi^r \text{ up to a multiplicative constant .}$$

Therefore π is also a right-invariant Haar measure. We have therefore established the following theorem:

Theorem 3.1. The Bayes estimator $\delta_B^\pi(x) = p(x)$ for all x iff π is proportional to the right-invariant Haar measure on G, hence is a right-invariant Haar measure on G.

In many testing problems one must have procedures that account for nuisance parameters. The t-test (see Example 4.2) is perhaps the most classical example of a test with a nuisance parameter. To study the reconciliation of p-values and posterior probabilities in multiparameter problems we need the concept of the composite model. We shall refer the reader to Barndorff-Nielsen (1983) for a detailed exposition of these models. The main reason for using composite models is that they are quite useful for handling multiparameter problems. Consider a model with an underlying density function $f(x \mid \theta, g)$ relative to an invariance measure on a sample space $\mathfrak{X}$. Let G be a group action on $\mathfrak{X}$. Under some set of conditions, contained in Theorem 5.1 of Barndorff-Nielsen (1983), it is shown under the principles of G-sufficiency (Barnard, 1963), inference for θ is to be drawn from the marginal model. It follows from Theorem 5.1 of Barndorff-Nielsen (1983) that the marginal model's density function is given by

$$f^M(x \mid \theta) = \int_G f(x \mid \theta, g)\beta(dg) \ ,$$

where $\beta(\cdot)$ is the right invariant measure on G. Therefore, under the principle of G-sufficiency inference should be based on f^M. Hence, p-values for the one-sided tests discussed in the beginning of this section will be of the form

$$p(x) = \int_x^\infty f^M(y \mid \theta_0)dy = \int_G^\infty f(y \mid \theta_0, g)\beta(dg)dy \ .$$

If one integrates the posterior $\pi(\theta, g \mid x)$ with respect to $\beta(dg)$ prior on the nuisance parameters, it may be seen that the p-value and the posterior probability of the null hypothesis are equal. Example 4.2 illustrates this relationship. Further results on the agreement between the measures of evidence assessment for multiparameter problems are mentioned in Tsui and Weerahandi (1989).

4. Examples

In this section we look at a few special cases to illustrate the result of Theorem 3.1. As we will see, Theorem 3.1 does more than assert the equivalence of a p-value and posterior probability. It gives us a way to calculate p-values and, moreover, can give us a way to define p-values.

Example 4.1: Scale families. Let $X \sim \left(\frac{1}{\sigma}\right) f\left(\frac{X}{\sigma}\right)$, and suppose we want to test $H_0 : \sigma \le \sigma_0$ vs. $H_1 : \sigma > \sigma_0$ based on observing X=x. The right-invariant Haar measure for the scale groups is $\pi(\sigma) = \frac{1}{\sigma} d\sigma$, and we have

$$\delta_B^\pi(x) = \frac{\displaystyle\int_0^{\sigma_0} \left[\frac{1}{\sigma} f(x/\sigma)\right] \frac{1}{\sigma} d\sigma}{\displaystyle\int_0^{\infty} \left[\frac{1}{\sigma} f(x/\sigma)\right] \frac{1}{\sigma} d\sigma}$$

(4.1)
$$= \int_x^{\infty} \frac{1}{\sigma_0} f(t/\sigma_0) \, dt$$

$$= p(x) ,$$

showing the equivalence of the p-value and the Bayes rule. □

A slightly more complicated case is location-scale, where the definition of the p-value may not be immediately evident. In the next example, however, we know what to expect for the p-value.

Example 4.2: Normal location-scale. For testing $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$, where $X_1, \cdots, X_n$ are iid $n(\theta, \sigma^2)$, both unknown, the right invariant prior is $\frac{1}{\sigma} d\theta d\sigma$. The joint density of the sample, by sufficiency a function only of $\bar{X} = \frac{1}{n} \Sigma X_i$ and $S^2 = \frac{1}{n-1} \Sigma (X_i - \bar{X})^2$, is given by

(4.2)
$$f(\bar{x}, s^2 \mid \theta, \sigma^2) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} e^{-\frac{n}{2} \frac{(\bar{x}-\theta)^2}{\sigma^2}} \frac{(n-1)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}} \frac{(s^2)^{\frac{n-1}{2}-1}}{(s^2)^{\frac{n-1}{2}}} e^{-\frac{(n-1)}{2\sigma^2 s^2}} d\bar{x} ds^2 .$$

Therefore, the posterior probability of H_0 is

$$(4.3) \qquad P(H_0 \,|\, \bar{x}, s^2) = \frac{\displaystyle\int_0^\infty \int_{-\infty}^{\theta_0} \frac{1}{\sigma^{n+1}} \, e^{-\frac{1}{2\sigma^2}\left(n(\bar{x}-\theta)^2 + (n-1)s^2\right)} \, d\theta \, d\sigma}{\displaystyle\int_0^\infty \int_{-\infty}^{\infty} \frac{1}{\sigma^{n+1}} \, e^{-\frac{1}{2\sigma^2}\left(n(\bar{x}-\theta)^2 + (n-1)s^2\right)} \, d\theta \, d\sigma} \; .$$

The integral over σ is straightforward to evaluate, yielding

$$P(H_0 \,|\, \bar{x}, s^2) = \frac{\displaystyle\int_{-\infty}^{\theta_0} \left(n(\bar{x}-\theta)^2 + (n-1)s^2\right)^{-n/2} d\theta}{\displaystyle\int_{-\infty}^{\infty} \left(n(\bar{x}-\theta)^2 + (n-1)s^2\right)^{-n/2} d\theta}$$

$$(4.4)$$

$$= \frac{\displaystyle\int_{\frac{\sqrt{n}(\bar{x}-\theta_0)}{s}}^{\infty} \left(1 + t^2/(n-1)\right)^{-n/2} dt}{\displaystyle\int_{-\infty}^{\infty} \left(1 + t^2/(n-1)\right)^{-n/2} dt}$$

$$= P\left(T_{n\text{-}1} > \frac{\sqrt{n}(\bar{x}-\theta_0)}{s}\right),$$

where T_{n-1} is a Student's t random variable with n–1 degrees of freedom. This last expression is the usual p-value for this situation, showing the reconciliation of evidence. □

There are situations where the p-value does not have a "usual" definition, but by Theorem 3.1, it can be unambiguously defined as the Bayes posterior probability arising from a right-invariant prior. The next example illustrates such a case.

Example 4.3: General location-scale. For a sample $X_1,\cdots,X_n$ from a general location-scale family, the posterior distribution of θ, using the right-invariant Haar measure, is given by

$$(4.5) \qquad \pi(\theta \,|\, X_1,\cdots,X_n) = \frac{\displaystyle\int_0^\infty \prod_{i=1}^{n} f\big((x_i-\theta)y\big) dy}{\displaystyle\int_{-\infty}^{\infty} \int_0^\infty \prod_{i=1}^{n} f\big((x_i-\theta)y\big) dy \, d\theta} \;,$$

and the posterior probability of $H_0 : \theta \leq \theta_0$ is

$$(4.6) \qquad P(H_0 | X_1, \cdots, X_n) = \int_{-\infty}^{\theta_0} \pi(\theta | X_1, \cdots, X_n) d\theta \ ,$$

which can be taken as the definition of the p-value. In fact, this expression for the p-value agrees with the one arising from conditional likelihood-based calculations.

From Fisher (1934), or Barndorff-Nielsen (1983), the conditional distribution of $\hat{\theta}$ and $\hat{\sigma}$, the maximum likelihood estimators of θ and σ, is given by

$$(4.7) \qquad g(\hat{\theta}, \hat{\sigma} | \theta, \sigma, a_1, \cdots, a_n) = \frac{\hat{\sigma}^{n-2} \prod\limits_{i=1}^{n} \frac{1}{\sigma} f\left(\frac{\hat{\sigma}}{\sigma} \left(a_i + \frac{\hat{\theta} - \theta}{\hat{\sigma}} \right) \right)}{\displaystyle\int_{-\infty}^{\infty} \int_{0}^{\infty} \hat{\sigma}^{n-2} \prod\limits_{i=1}^{n} f\left(\frac{\hat{\sigma}}{\sigma} \left(a_i + \frac{\hat{\theta} - \theta}{\hat{\sigma}} \right) \right) d\hat{\sigma} d\hat{\mu}}$$

where $a_i = (x_i - \hat{\theta})/\hat{\sigma}$, $i = a, \cdots, n$, are ancillary statistics. It is straightforward to verify that the p-value for $H_0 : \theta \leq \theta_0$, using (4.7), is equal to (4.6). More precisely,

$$(4.8) \qquad P(H_0 | x_1, \cdots, x_n) = \int_{-\infty}^{\theta_0} \pi(\theta | x_1, \cdots, x_n) d\theta = \int_{\frac{\hat{\theta}_0 - \theta_0}{\hat{\sigma}_0}}^{\infty} \int_{0}^{\infty} g(\hat{\theta}, \hat{\sigma} | \theta, \sigma, a_1, \cdots, a_n) d\hat{\sigma} d\hat{\theta} = \text{p-value} \ ,$$

where $\hat{\theta}_0$ and $\hat{\sigma}_0$ are the observed values of the random variables $\hat{\theta}$ and $\hat{\sigma}$. Thus, the posterior probability against the right Haar measure agrees with the p-value of conditional likelihood inference.

5. Discussion

Reconciliation of Bayesian and frequentist measures of evidence has, in recent years, received much attention. I.J. Good (1987) notes that the p-value and Bayesian posterior probability of the null hypothesis are both here to stay, so the relationship between them needs to be taken seriously. Good notes that these relationships form a large part of the problem of pure rationality, namely, the extent Bayesian and non-Bayesian may be synthesized.

DeGroot (1973) also had an interest in the problem of reconciling evidence. His approach was to construct alternative distributions and find priors for which the p-value and posterior probabilities match. As in this article, DeGroot's work was essentially on the one-sided testing problem.

Berger and Selke (1987) discuss the relationship between the p-value and the posterior probability of a point null hypothesis. They show that in the two-sided problem these measures of evidence can be quite different. Their main conclusion is that the p-value tends to overstate the evidence against the null hypothesis, that is, the p-value tends to be smaller than the Bayesian posterior probability.

The disagreement between the one-sided and the two-sided problem is, now, not at all surprising. Hwang *et al.* (1990) prove that the p-value is an admissible decision procedure in the one-sided testing problem, while it is inadmissible in the two-sided problem. The proofs of these admissibility results follow from the fact that the generalized Bayes rules form a complete class of procedures and that one-sided p-values are generalized Bayes rules while the two-sided p-values are not. In particular, in the case of one-sided hypothesis tests we have shown that the p-value is equal to the Bayes procedure against the right-invariant Haar measure prior. Moreover, in Example 4.2 it is shown that the p-value from the t-test is generalized Bayes, hence it is admissible.

The proof of the complete class results in Hwang *et al.* (1990) show that admissible procedures are limits of Bayes rules and that the limits of Bayes rules are generalized Bayes rules. There is an interesting relationship between the limits of Bayes rules and right Haar measure priors. Take a sequence of priors (w.r.t. π^r) on G,

$$\{\pi_n\} = \left\{\pi^r(G_n)^{-1} \mathbb{I}(G_n)\right\}$$

where $\{G_n\}$ is an increasing sequence of sets whose union is G. Then for each $x \in \mathfrak{X}$ the sequence of posterior distributions $\pi_n(\cdot\,|x)$ converge weakly to $\pi^r(\cdot\,|x)$, the posterior induced from the prior measure π^r. In general, the convergence of $\pi_n(\cdot\,|x)$ to $\pi^r(\cdot\,|x)$ will not be uniform. The idea of convergence of the posterior led Stone (1990) to examine the distance

$$d_n(x) = \sup_{A \subset G} \left| \pi_n(A\,|x) - \pi^r(A\,|x) \right|$$

and consider X to be the r.v. with marginal distribution induced from the Bayesian joint distribution which gives θ the prior π_n and X the distribution P_θ. Stone showed that $d_n(X) \to 0$ in probability if and only if G is amenable and $|G_n|$ is properly chosen. From these results we see that not all limits of proper Bayes rules will converge to the π^r-Bayes rule unless G and G_n are chosen appropriately.

Another result of interest which follows for Theorem 3.1 is that the p-value, in the one-sided problem, is coherent in the sense of Heath and Sudderth (1978, 1989). For the coherence framework the hypothesis-testing problem may be formulated as in Berger (1985, pp. 120-122). It is shown in Heath and Sudderth (1978) that if one uses posterior inference based on the right invariant Haar measure then the resulting inference will be coherent. Specifically, when using the p-value as an inference procedure in the one-sided hypothesis testing problem the resulting inference is coherent. An excellent discussion on the concept of coherence and the related notion of rationality can be found in Berger (1985, pp. 120-122).

Using somewhat different reasoning, Fraser (1961, 1968) has implicitly recommended the use of right invariant Haar priors in structural inference. This follows because Fraser's structural probability distribution is exactly the corresponding invariant posterior distribution. Confidence procedures may also be based on this distribution, as in Hora and Buehler (1966), Bondar (1977), and Chang and Villegas (1986).

Appendix

The following three technical lemmas are needed to establish Theorem 3.1.

Lemma A.1. There exists a constant K such that for any integrable function t,

$$(A.1) \qquad \int_A t(x^{-1})\pi^r(x)dx = K \int_{A^{-1}} t(x)\pi^\ell(x)dx \ ,$$

where π^r and π^ℓ are right and left invariant Haar measures, respectively.

Proof. If $x = y^{-1}$ and $q(y) = \pi^r(y^{-1})J(y)$, where J is the Jacobian of the transformation, then as in Berger (1985, p.411), q is a left invariant Haar measure. Since the right and left Haar densities are unique up to a multiplicative constant, $q(y) = K\pi^\ell(y)$. Therefore

$$(A.2) \qquad \int_A t(x^{-1})\pi^r(x)dx = \int_{A^{-1}} t(y)\pi^r(y^{-1})J(y)dy = K \int_{A^{-1}} t(y)\pi^\ell(y)dy \ . \qquad \square$$

Lemma A.2. For the multiplier ν in (v) of Section 2 and for any integrable function t,

$$(A.3) \qquad \int_A t(yg)\pi^\ell(y)dy = \nu(g) \int_{gA} t(y)\pi^\ell(y)dy \ .$$

Proof. As in Berger (1985, p. 412), $\pi^\ell(yg)J_g^r(y) = \nu(g)\pi^\ell(y)$, where $J_g^r(y)$ is the Jacobian of the transformation $y \rightarrow gy$. Now by applying this identity and the transformation $y = xg$ the results follows. $\qquad \square$

Lemma A.3. The density function can be written $f(x|\bar{g}) = f\left(g^{-1}(x)\big|\bar{e}\right)J_{g^{-1}}^\ell(x)$, where $J_{g^{-1}}^\ell$ is the Jacobian of the transformation $x \rightarrow g^{-1}x$.

Proof. See Berger (1985, p. 410).

For more information on Haar measure see the book by Nachbin (1965).

References

Barnard, G. A. (1963). Some logical aspects of the fiducial argument. *J. Roy. Statist. Soc., Ser. B* **25**, 111-114.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Second Edition. New York: Springer-Verlag.

Casella, G. and Berger, R. L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. Amer. Statist. Assoc.* **82**, 112-139.

Chang, T. and Villegas, C. (1986). On a theorem of Stein relating Bayesian and classical inference in group models. *Canad. J. Statist.* **14**, 289-296.

Heath, D. and Suddereth, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6**, 333-345.

Heath, D. and Suddereth, W. (1989). Coherent inference from improper priors and finitely additive priors. *Ann. Statist.* **17**, 907-919.

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., and Farrell, R. (1990). Estimation of accuracy in testing. Biometrics Unit Technical Report BU-1027-M, Cornell University.

Nachbin (1965). *The Haar Integral.* Van Nostrand.

Stone, M. (1970). Necessary and sufficient conditions for convergence in probability to invariant posterior distributions. *Ann. Math. Statist.* **41**, 1349-1353.

Tsui, K. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses on the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **84**, 602-609.