Bayesian Analysis (2011) **6**, Number 4, pp. 603–638

Neutral-data comparisons for Bayesian testing

Dan J. Spitzner*

Abstract. A novel approach to evidence assessment in Bayesian hypothesis testing is proposed, in the form of a "neutral-data comparison." The proposed assessment is similar to a Bayes factor, but, rather than comparing posterior to prior odds, it compares the posterior odds of the observed data to that calculated on "neutral" data, which arise as part of the elicitation of prior knowledge. The article develops a general theory of neutral-data comparisons, motivated largely by the Jeffreys-Lindley paradox, and develops methodology for specifying and working with neutral data in the context of Gaussian linear-models analysis. The proposed methodology is shown to exhibit exceptionally strong asymptotic-consistency properties in high dimensions, and, in an application example, to accommodate challenging analysis objectives using basic computational algorithms.

Keywords: Bayesian hypothesis testing; Bayes factors; Bayesian asymptotic-consistency; model choice in high dimensions; analysis of variance.

1 Introduction

The problem of interest is to test a null hypothesis, H_0, about a parameter, $\boldsymbol{\theta}$, against an alternative, H_1, on the basis of data, $\boldsymbol{Y}$. The conventional approach to this problem is to calculate a Bayes factor, whose purpose is to report an assessment of the strength of evidence in $\boldsymbol{Y}$ about H_0. Supposing that H_1 is the negation of H_0, and assuming a suitable prior distribution under each hypothesis, the Bayes factor is

$$BF_0 \quad = \quad \frac{P[H_0|\boldsymbol{Y}]/(1-P[H_0|\boldsymbol{Y}])}{\rho_0/(1-\rho_0)}, \tag{1}$$

writing $\rho_0 = P[H_0]$ to denote the prior null probability. A larger value of BF_0 indicates stronger support for H_0, and may perhaps be interpreted within standard categories such as "positive," "strong," or "very strong" support.

The form of (1) as a ratio highlights that BF_0 is a *comparison* between the posterior and prior odds of H_0. This article proposes an alternative ratio, a "neutral-data comparison," which modifies BF_0's comparison by replacing the baseline with an alternative

*Department of Statistics, University of Virginia, Charlottesville, VA, spitzner@virginia.edu

 DOI:10.1214/11-BA623

quantity that reflects "neutrality" between H_0 and H_1. Specifically, a neutral-data comparison replaces the denominator in (1) with the posterior odds calculated on "neutral data," $\tilde{\boldsymbol{Y}}$, to yield

$$NDC_0 \quad = \quad \frac{P[H_0|\boldsymbol{Y}]/(1-P[H_0|\boldsymbol{Y}])}{P[H_0|\tilde{\boldsymbol{Y}}]/(1-P[H_0|\tilde{\boldsymbol{Y}}])}. \tag{2}$$

Two questions that immediately arise are: "What are neutral data?" and "Why does it make sense to substitute $P[H_0|\tilde{\boldsymbol{Y}}]$ for ρ_0?" A starting point to answering the first question is the following definition: *neutral-data are imaginary data, identified as a component of prior knowledge, that exhibit evidence neither for H_0 nor H_1*. In what follows, this definition is developed into guidelines for specifying $\tilde{\boldsymbol{Y}}$ in practice. The second question is answered using Good's (1950) "device of imaginary results," which is described below and which will serve as a cornerstone for interpreting NDC_0.

One appealing feature of NDC_0 is its potential to bridge longstanding gaps between estimation and testing, particularly with regard to the use of vague priors. In estimation, vague priors are often formulated using reference prior analysis (*cf.* Berger, Bernardo, and Sun, 2009), which typically prescribes impropriety, or by suitably transforming the parameter and then setting a prior scale parameter to some arbitrarily large value. In testing, however, these formulations of the prior do not readily translate for use with Bayes factors, since BF_0 may either depend on an arbitrary normalizing constant (of an improper prior), or, when the scale parameter is large, may suffer the "Jeffreys-Lindley paradox" (*cf.* Lindley, 1957), by which it becomes more sensitive to the diffuseness of the prior than to the data. It is shown below that a neutral-data comparison sidesteps these difficulties, and exhibits sensitivity to prior diffuseness at a level comparable to that observed in estimation. Neutral-data comparisons thus induce a certain degree of consistency between estimation and testing with regard to the types of priors that are suitable for use in practice. Other practical benefits of NDC_0 are possible as well, for it is shown below that neutral-data comparisons admit natural default settings, but readily incorporate prior knowledge about H_0 and H_1; they inspire model-choice procedures with especially good asymptotic-consistency properties, and they admit a straightforward computational method with which to carry out complicated Gaussian linear-models analyses.

1.1 Related literature

Existing methodologies with which neutral-data comparisons most closely align are the calibration technique of Spiegelhalter and Smith (1982), the "intrinsic Bayes factor" of Berger and Pericchi (1996), and the "expected-posterior prior" of Pérez and Berger (2002), each of which uses observed or imaginary "training samples" to calibrate or construct a Bayes factor in a way that roughly resembles the use of neutral data here. There are sharp differences, however, not just in the way that imaginary data are used, but in what they are intended to represent. Whereas neutral data originate in the elicitation of prior knowledge, Spiegelhalter and Smith (1982) and Berger and Pericchi (1996) derive their training samples by objective criteria, to reflect minimal data configurations that are required for carrying out inference. Pérez and Berger (2002) work with objective criteria, too; but, in addition, they consider subjective elicitation of imaginary training samples, in a formulation whereby such data arise "from beliefs as to how a training sample would behave" (p. 495). In contrast, neutral data locate a balance between H_0 and H_1, which is a distinct aspect of prior knowledge. Conceptual differences aside, the statistics constructed by these existing methods do intersect with neutral-data comparisons in certain substantial ways, which are examined in Section 2.1, below. Other related techniques worthy of mention are O'Hagan's (1995) "fractional likelihood" and Aitkin's (1991) "posterior Bayes factors."

A subset of related literature that is currently quite active explores broad criteria for asymptotic consistency of BF_0 as it relates to the choice of a prior. This is discussed in Berger, Ghosh, and Mukhopadhyay (2003), García-Donato and Sun (2007), Liang et al. (2008), Casella et al. (2009), Guo and Speckman (2009), Maruyama and George (2010), and many others. Johnson and Rossell's (2010) asymptotic analysis under "non-local" priors is particularly relevant, in that one of their central focuses is the balance in emphasis between H_0 and H_1, a concept embodied here in the definition of $\tilde{\boldsymbol{Y}}$. A close asymptotic connection between NDC_0 and non-local priors is made in Section 3.2, below. The use of asymptotic consistency to evaluate Bayesian procedures is conceptually justified in Diaconis and Freedman (1986), and is by now widely accepted as guidance for developing methodology. For instance, Berger and Pericchi (1996) state a preference for procedures that match valid Bayes factors asymptotically. A similar point of view motivates the well-known BIC criterion of Schwarz (1978), and it will play an important role here as well.

1.2 Objectives and outline

As indicated in the discussion below (2), the purpose of this article is to motivate NDC_0 both conceptually and as an applied tool, using ideas that are consistent with Bayesian thinking. Its goals are to establish a foundation for theoretical and methodological development, to provide basic guidelines for eliciting neutral data in practice, and to demonstrate interesting capabilities of neutral-data comparisons. The approach to achieving these goals is described in the following outline of discussion.

Section 2 lays out a theory and interpretation of NDC_0 in broad generality. The entry point is a reinterpretation of the Jeffreys-Lindley paradox as a particular type of incoherency that is revealed upon using the device of imaginary results to check elicitation of the parameter ρ_0. Subsequently, NDC_0 is interpreted as (*i*) a reasonably good approximation to an ideal Bayes factor; and (*ii*), an alternative to BF_0 that is better suited to accommodate the above type of incoherency. The conceptual setup is also extended to the scenarios involving multiple hypotheses.

The remainder of the article is concerned with implementation of the theory laid out in Section 2. Its scope is limited to testing linear hypotheses under a Gaussian model, and many results are furthermore asymptotic in nature. Despite its narrower focus, the context is still broad enough to achieve the article's applied goals, and it carries special importance as a widely-used analysis framework. Within this discussion, Section 3 describes how neutral data may be elicited from prior knowledge that is expressed in the traditional form of a probability distribution. Section 4 investigates neutral-data comparisons in high-dimensional analysis, where they are shown to inspire a model-choice procedure that rivals the strongly-performing "sure independence screening" procedure of Fan and Lv (2008). Performance criteria in that section are set at a high bar, in considering sparse signals and dimensionality that increases at an *exponentially* fast rate. Section 5 demonstrates the use of neutral-data comparisons in an example application of Bayesian "analysis of variance." The example highlights that neutral-data comparisons are computationally attractive in their manner of achieving complicated analysis objectives such as partition analysis, whose typical implementation using product-partition models can be delicate. (See, *e.g.*, Crowley, 1997, for basic ideas.) Section 6 offers concluding discussion, and an appendix compiles all technical derivations.

2 Motivation

In this section, the motivation for NDC_0 is laid out in steps, beginning with an exploration of key distinctions between the behavior of NDC_0 and BF_0, and leading to the interpretations of NDC_0 indicated in Section 1.2. It is assumed throughout that a suitable setting for $\tilde{\boldsymbol{Y}}$ exists and is specified (possibly from techniques described later in Section 3) in accordance with the definition of neutral data given below (2). To be clear, $\tilde{\boldsymbol{Y}}$ is held fixed throughout this section, for reasons that are explained in Section 2.4.

2.1 Comparison of BF_0 and NDC_0

Exploration and comparisons of BF_0 and NDC_0 are aided by the alternative formulas

$$BF_0 = \frac{\pi(\boldsymbol{Y}|H_0)}{\pi(\boldsymbol{Y}|H_1)} \quad \text{and} \quad NDC_0 = \frac{\pi(\boldsymbol{Y}|H_0)/\pi(\boldsymbol{Y}|H_1)}{\pi(\tilde{\boldsymbol{Y}}|H_0)/\pi(\tilde{\boldsymbol{Y}}|H_1)}, \tag{3}$$

where $\pi(\boldsymbol{y}|H) = \int_H \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|H)d\boldsymbol{\theta}$ is a marginal density for the data, given the hypothesis H. These are equivalent to the formulas (1) and (2), and highlight that both BF_0 and NDC_0 may be calculated without having specified ρ_0. Independence from ρ_0 is often touted as a benefit to the use of BF_0, and we see now that NDC_0 shares the same property.

The formulas in (3) furthermore show that NDC_0 is insensitive to individual renormalization of the conditional priors, while BF_0 is not. To see this, consider revising $\pi(\boldsymbol{\theta}|H_1)$ to $\pi^*(\boldsymbol{\theta}|H_1) = c\pi(\boldsymbol{\theta}|H_1)$, for some constant c, a revision that might be thought logically inconsequential given that the relative probabilities within H_1 are unaffected; yet, it is clear from (3) that only NDC_0 is left unchanged, while BF_0 is rescaled by the factor $1/c$. For perspective on this behavior, consider the unconditional posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \rho_0\pi(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|H_0) + (1-\rho_0)\pi(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|H_1)$, which clearly becomes modified, as does BF_0, upon renormalizing $\pi(\boldsymbol{\theta}|H_1)$ but not $\pi(\boldsymbol{\theta}|H_0)$. By working with $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$, the source of the behavior becomes clear: disproportionate renormalization changes in the relative probabilities between H_0 and H_1. Accordingly, the sensitivity of $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ to renormalization is not so strange; hence, neither is such sensitivity of BF_0, but the insensitivity of NDC_0 to renormalization is now seen to be a strong and surprising property. What is more is that it is possible for $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ to inherit this insensitivity property from NDC_0, by a route that is described in Section 2.5, below.

The statistics BF_0 and NDC_0 are, in addition, drastically different in their sen-

sitivity to prior scale parameters. To see this, suppose $\boldsymbol{Y}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \boldsymbol{I})$, where $\boldsymbol{\theta}$ is ν_1-dimensional, and that $\pi(\boldsymbol{\theta}|H_1)$ is revised to $\pi^*(\boldsymbol{\theta}|H_1) = \tau^{-\nu_1}\pi(\boldsymbol{\theta}/\tau|H_1)$, for some scale parameter $\tau > 0$. Suppose also that both the observed and neutral posterior densities $\pi(\boldsymbol{\theta}|\boldsymbol{Y}, H_1)$ and $\pi(\boldsymbol{\theta}|\tilde{\boldsymbol{Y}}, H_1)$ peak sharply at unique modes in H_1, respectively denoted $\hat{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_1$, which admits accurate study of BF_0 and NDC_0 using Laplace approximations to $\pi(\boldsymbol{Y}|H_1)$ and $\pi(\tilde{\boldsymbol{Y}}|H_1)$ (*cf.*, Kass and Raftery, 1995, sec. 4.1); the resulting expressions are

$$\begin{aligned} BF_0 &\approx \tau^{\nu_1} \times \frac{|\boldsymbol{I} + \tau^{-2}\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1/\tau)|^{1/2}\pi(\boldsymbol{Y}|H_0)}{(2\pi)^{\nu_1/2}\pi(\boldsymbol{Y}|\hat{\boldsymbol{\theta}}_1)\pi(\hat{\boldsymbol{\theta}}_1/\tau|H_1)} \\ NDC_0 &\approx \frac{|\boldsymbol{I} + \tau^{-2}\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1/\tau)|^{1/2}}{|\boldsymbol{I} + \tau^{-2}\boldsymbol{S}(\tilde{\boldsymbol{\theta}}_1/\tau)|^{1/2}} \frac{\pi(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}}_1)\pi(\tilde{\boldsymbol{\theta}}_1/\tau|H_1)}{\pi(\boldsymbol{Y}|\hat{\boldsymbol{\theta}}_1)\pi(\hat{\boldsymbol{\theta}}_1/\tau|H_1)} \frac{\pi(\boldsymbol{Y}|H_0)}{\pi(\tilde{\boldsymbol{Y}}|H_0)}, \end{aligned} \tag{4}$$

writing $\boldsymbol{S}(\boldsymbol{t})$ to denote the Hessian matrix of $\log\pi(\boldsymbol{t}|H_1)$. As $\tau \to \infty$, typically $\hat{\boldsymbol{\theta}}_1 \to \boldsymbol{Y}$, $\tilde{\boldsymbol{\theta}}_1 \to \tilde{\boldsymbol{Y}}$, and all of $\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1/\tau)$, $\boldsymbol{S}(\tilde{\boldsymbol{\theta}}_1/\tau)$, $\pi(\hat{\boldsymbol{\theta}}_1/\tau|H_1)$, and $\pi(\tilde{\boldsymbol{\theta}}_1/\tau|H_1)$ tend to nonzero constants (this is not true of non-local priors, which are considered later in Section 3.2); hence, the expressions in (4) show that NDC_0 typically stabilizes, while BF_0 increases without bound, at the rate τ^{ν_1}, to eventually assess arbitrarily strong support for H_0, regardless of $\boldsymbol{Y}$. This behavior of BF_0 is an illustration of the Jeffreys-Lindley paradox, about which more is said below. Moreover, NDC_0 is shown in (4) to depend on τ predominantly through $\hat{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_1$, statistics representative of the estimation context (*e.g.*, $\hat{\boldsymbol{\theta}}_1$ itself might be reported as a point estimate of $\boldsymbol{\theta}$, conditional on H_1). By viewing these statistics through their connection to $\hat{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_1$, NDC_0 is seen to be sensitive to prior diffuseness at a level comparable to estimation, while the sensitivity of BF_0 is seen to be much stronger.

2.2 The operational and authentic priors

Although the Jeffreys-Lindley paradox prevents straightforward translation to testing of priors that are suitable for estimation, it has long been accepted as part of Bayesian thinking, and is justified in various ways. In prediction, Smith and Spiegelhalter (1980, p. 216) point out that the Jeffreys-Lindley paradox allows BF_0 to "function as a *fully automatic Occam's Razor*—cutting back to the simpler model when there is nothing lost by so doing." In testing, the Jeffreys-Lindley paradox typically serves as a caution against using priors that place a substantial proportion of mass far from H_0. Jeffreys (1961, p. 251), referring to a point-null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{0}$, argues as much in stating, "the mere fact that it has been suggested that [the parameter] is zero corresponds to

some presumption that it is fairly small." Jeffreys's statement thus indicates that a vague prior, which puts relatively little prior mass in regions where where the parameter is "fairly small," must be interpreted in a careful way. To this end, a vague prior is here interpreted as no more than an "operational prior," a prior that represents what prior knowledge is readily available, and what is to be used as input to data analysis, but which may ultimately paint an incomplete picture. In other words, a vague prior is a partial elicitation of prior knowledge in which prior mass that is far from H_0, and possibly elsewhere, is described only loosely. By this interpretation, there is, at least conceptually, a second prior to consider, an "authentic prior," which is a prior that precisely and completely describes the picture of prior knowledge, but is somehow not fully accessible. Differences between the operational and authentic priors are common in practice, and often arise when the analyst faces insurmountable challenges to fully eliciting prior knowledge, perhaps due to limited resources, or lack of access to an expert, which puts the authentic prior out of reach. In these situations, a vague prior may be the best description of prior knowledge that is available, and NDC_0 is to provide a vehicle with which to exploit it.

It will be useful to consider these ideas in a more formal way, paralleling discussion that appears in Robert (1993). Working again in the setup surrounding (4), consider the impact of increasing τ on the probabilities associated with individual subregions in H_1. In an echo of Jeffreys's statement, above, Robert (1993, p. 603) suggests that a prior should "give sufficient weight to the range of values of [the parameter] which actually caused H_0 to be tested." This "range of values" would, in the present context, be taken to be some particular subset $\tilde{\Theta}_1 \subset H_1$ that lies adjacent to H_0, one that captures the values within a "'reasonable' range" from H_0. (By implication, there is also a subset of values in H_1 that lie *beyond* a reasonable range from H_0, which shall be assumed to have infinite volume.) Robert observes that as $\tau \to \infty$, $P[\tilde{\Theta}_1|H_1] \to 0$; hence, unless $\rho_0 \to 0$ also (at the same rate as $P[\tilde{\Theta}_1|H_1] \to 0$), a severely uneven balance between H_0 *versus* H_1 would eventually arise in the operational prior, at least with respect to the portion of H_1 that is in $\tilde{\Theta}_1$. This means that, as $\tau \to \infty$, eventually $\rho_0/(1-\rho_0)$ may very well *misrepresent* the prior odds, which consequently diminishes BF_0 as a sensible comparison of posterior to prior odds.

Such uneven balance between H_0 *versus* H_1 is not present in the authentic prior. Under that prior, write $\tilde{P}$ to denote "authentic" probabilities, and set $\tilde{\rho}_0 = \tilde{P}[H_0]$. The

statistics of present consideration are

$$\widetilde{BF}_0 = \frac{\tilde{P}[H_0|\boldsymbol{Y}]/(1-\tilde{P}[H_0|\boldsymbol{Y}])}{\tilde{\rho}_0/(1-\tilde{\rho}_0)} \quad \text{and} \quad \widetilde{NDC}_0 = \frac{\tilde{P}[H_0|\boldsymbol{Y}]/(1-\tilde{P}[H_0|\boldsymbol{Y}])}{\tilde{P}[H_0|\tilde{\boldsymbol{Y}}]/(1-\tilde{P}[H_0|\tilde{\boldsymbol{Y}}])}, \quad (5)$$

each of which represents an ideal assessment that the analyst would want to report. Interpreting as above, the conditional probability $\tilde{P}[\tilde{\Theta}_1|H_1]$ falls near one, and $\tilde{\rho}_0/(1-\tilde{\rho}_0)$ correctly represents the prior odds; hence, $\widetilde{BF}_0$ is indeed a sensible comparison of posterior to prior odds. For reasons that are discussed next, so is $\widetilde{NDC}_0$, for it is argued below that the two statistics are identical, $\widetilde{NDC}_0 = \widetilde{BF}_0$.

2.3 The device of imaginary results

The role of the device of imaginary results in these arguments is to allow meaningful examination of the relationship between a prior null probability, ρ_0 or $\tilde{\rho}_0$, and its corresponding posterior probability calculated on neutral data, $P[H_0|\tilde{\boldsymbol{Y}}]$ or $\tilde{P}[H_0|\tilde{\boldsymbol{Y}}]$. As for a definition of this device, Good (1950) formulates it as an exercise to assist the analyst in checking the suitability of the prior: the analyst is to imagine a data set that is interesting in some way, then apply Bayes's rule and check whether the resulting (imaginary) posterior distribution exhibits sensible characteristics in light of the given input. In the present context, the "interesting" data set is $\tilde{\boldsymbol{Y}}$, neutral data, and the characteristic of the prior that is under check is ρ_0 or $\tilde{\rho}_0$. The two are connected by the neutrality of $\tilde{\boldsymbol{Y}}$: since neutral data exhibits evidence for neither hypothesis, one would expect the posterior update made on such data to leave knowledge about H_0 *versus* H_1 unchanged. In other words, one would expect the device of imaginary results to yield $\rho_0 = P[H_0|\tilde{\boldsymbol{Y}}]$ and $\tilde{\rho}_0 = \tilde{P}[H_0|\tilde{\boldsymbol{Y}}]$.

Nevertheless, it is implied from the Jeffreys-Lindley paradox that the presence of an arbitrarily large scale parameter would render one of these criteria, $\rho_0 = P[H_0|\tilde{\boldsymbol{Y}}]$, impossible to satisfy. This is clear from the form of the Bayes factor in (1), for, writing $BF_0(\tilde{\boldsymbol{Y}})$ to denote BF_0 evaluated on $\tilde{\boldsymbol{Y}}$, the criterion $\rho_0 = P[H_0|\tilde{\boldsymbol{Y}}]$ implies $BF_0(\tilde{\boldsymbol{Y}}) = 1$, but this is impossible since $BF_0(\tilde{\boldsymbol{Y}}) \to \infty$ as $\tau \to \infty$, regardless of what is specified as neutral data. On the other hand, to reflect the accurate, complete picture of prior knowledge embodied in the authentic prior, it is expected that $\tilde{\rho}_0 = \tilde{P}[H_0|\tilde{\boldsymbol{Y}}]$ (*i.e.*, $\widetilde{BF}_0(\tilde{\boldsymbol{Y}}) = 1$, using analogous notation) would indeed hold true. An immediate consequence is that $\widetilde{NDC}_0 = \widetilde{BF}_0$, as was claimed in the discussion under (5).

2.4 Interpretation of NDC_0

The pieces are now in place to expound on the two interpretations of NDC_0 from Section 1.2, which are reiterated using present notation as follows.
Interpretation 1: NDC_0 is a reasonably good approximation to $\widetilde{BF}_0$.
Interpretation 2: NDC_0 is an adjustment to BF_0 that corrects for possible misrepresentation of the balance between H_0 *versus* H_1 by ρ_0.

Interpretation 1 formally treats the operational prior as an approximation to the authentic prior. Hence, since $\widetilde{NDC}_0 = \widetilde{BF}_0$, both BF_0 and NDC_0 are approximations to $\widetilde{BF}_0$, and the sensitivity comparison of Section 2.1 suggests a preference for NDC_0. That is, the approximation of NDC_0 is at least "good" relative to BF_0, when the operational prior is vague. Nevertheless, the types of sensitivity examined in Section 2.1 are quite specific, and there is ample room for investigations of accuracy beyond the scope of that discussion. Further support for Interpretation 1 is provided in Section 3, below, in an investigation that ties the accuracy of NDC_0 to the analyst's choice of $\tilde{\boldsymbol{Y}}$.

Interpretation 2 is attractive since it need not involve consideration of any prior other than the operational prior. By this interpretation, a failure of $\rho_0 = P[H_0|\tilde{\boldsymbol{Y}}]$ indicates that only one of ρ_0 or $P[H_0|\tilde{\boldsymbol{Y}}]$ may accurately represent the balance between H_0 *versus* H_1, but the behavior $P[\tilde{\Theta}_1|H_1] \to 0$ as $\tau \to \infty$, observed in Section 2.2, casts doubt on ρ_0 serving in that way. Further manipulation of the quantities involved will moreover lend positive support for $P[H_0|\tilde{\boldsymbol{Y}}]$ as the better representative of the balance between H_0 *versus* H_1. For this, consider solving (1) for ρ_0, and substituting $\tilde{\boldsymbol{Y}}$ for $\boldsymbol{Y}$; the result is

$$\rho_0 = \frac{P[H_0|\tilde{\boldsymbol{Y}}]}{BF_0(\tilde{\boldsymbol{Y}})}\left[\frac{1}{1 - P[H_0|\tilde{\boldsymbol{Y}}]\{1 - 1/BF_0(\tilde{\boldsymbol{Y}})\}}\right], \tag{6}$$

which identifies the factors at play in attempting to achieve the criterion $\rho_0 = P[H_0|\tilde{\boldsymbol{Y}}]$. Suppose, now, that $P[H_0|\tilde{\boldsymbol{Y}}]$ is set to a representative value of the balance between H_0 *versus* H_1, say $P[H_0|\tilde{\boldsymbol{Y}}] = 1/2$. Subsequently, as $\tau \to \infty$, it will follow that $BF_0(\tilde{\boldsymbol{Y}}) \to \infty$, and so formula (6) provides that $\rho_0 \to 0$. According to Section 2.2, this behavior of ρ_0 is precisely what is needed to make sense of the corresponding behavior $P[\tilde{\Theta}_1|H_1] \to 0$. In other words, $P[H_0|\tilde{\boldsymbol{Y}}]$ is the preferred representative of the balance between H_0 *versus* H_1 because sensible patterns emerge upon putting it in that role. As a bonus, these patterns suggest an interpretation of ρ_0 as representing the balance between H_0 *versus* the *portion* of H_1 that is in $\tilde{\Theta}_1$.

A similar pattern clarifies the role of $\tilde{\boldsymbol{Y}}$ in eliciting prior knowledge. In the setup

surrounding (4), consider if $\tilde{\boldsymbol{Y}}$ was not held fixed, but was instead allowed to vary with τ in such a way that fixes $BF_0(\tilde{\boldsymbol{Y}}) = 1$. From (4), this would require $\pi(\tilde{\boldsymbol{Y}}|H_0) = O(1/\tau^{\nu_1})$, hence $\|\tilde{\boldsymbol{Y}}\| \to \infty$; but then surely $\tilde{\boldsymbol{Y}}$ would eventually cease to represent neutrality, for its divergence would lead $\tilde{\boldsymbol{\theta}}_1$, an imaginary conditional estimate of $\boldsymbol{\theta}$, to simultaneously shift far from H_0, to even the unreasonable range of values beyond $\tilde{\Theta}_1$, eventually. Careful elicitation of $\tilde{\boldsymbol{Y}}$ thus requires a certain degree of separation from τ, and is best carried out under this general guideline: avoid deliberately tying neutral data to loose descriptions of prior knowledge. Stated differently, in the presence of differences between the operation and authentic priors, the role of $\tilde{\boldsymbol{Y}}$ is to open an *alternative* channel through which to elicit prior knowledge, to potentially offset unresolved imprecision in descriptions of prior knowledge that are obtained through other channels.

Finally, observe that each of Interpretations 1 and 2 identify a quantity that is to represent the balance between H_0 *versus* H_1: in Interpretation 1, that quantity is $\tilde{\rho}_0$; in Interpretation 2, it is $P[H_0|\tilde{\boldsymbol{Y}}]$. In practice, either quantity would be specified by the analyst, and either would presumably be set to the same value, since they each represent the same concept. To reflect this notion, it will be convenient to write $\tilde{\rho}_0 = P[H_0|\tilde{\boldsymbol{Y}}]$, and to simplify terminology by using the term "corrected prior null probability" as a catchall to refer either to $\tilde{\rho}_0$, $P[H_0|\tilde{\boldsymbol{Y}}]$, or to "the balance between H_0 *versus* H_1." Note that there is heuristic intuition that arises by setting $\tilde{\rho}_0 = P[H_0|\tilde{\boldsymbol{Y}}]$, for such equality suggests that $\tilde{\rho}_0$ might double as a "phantom" prior null probability, a (potentially nonsensical) null probability that remains unchanged upon application of the operational prior's posterior calculation to $\tilde{\boldsymbol{Y}}$. It is worthwhile to note, too, that setting $\tilde{\rho}_0 = 1/2$ reduces NDC_0 to just the posterior odds, whose interpretation is satisfying in certain ways discussed in Lavine and Schervish (1999).

2.5 Derived posterior probabilities

Recall the formulas in (3), which show that neither BF_0 nor NDC_0 need involve ρ_0 in its calculation. In contrast, any comprehensive data analysis would involve consideration of $P[H_0|\boldsymbol{Y}]$, as a component of the full posterior distribution, whose calculation does indeed require a precise setting for ρ_0. As (6) shows, it is possible to derive a value of ρ_0 from $P[H_0|\tilde{\boldsymbol{Y}}]$, which might then be used to calculate $P[H_0|\boldsymbol{Y}]$ by standard formulas. Nevertheless, it is convenient to avoid formula (6) entirely, and instead derive $P[H_0|\boldsymbol{Y}]$ from values of $P[H_0|\tilde{\boldsymbol{Y}}]$ and NDC_0, by solving (2). The result is

$$P[H_0|\boldsymbol{Y}] \;=\; \left\{1 + \left(\rho_0^{-1} - 1\right)/BF_0\right\}^{-1} = \left\{1 + \left(\tilde{\rho}_0^{-1} - 1\right)/NDC_0\right\}^{-1}, \qquad (7)$$

where $\tilde{\rho}_0 = P[H_0|\tilde{\boldsymbol{Y}}]$, which is written to highlight parallels between the pairs (ρ_0, BF_0) and $(\tilde{\rho}_0, NDC_0)$.

In addition, the posterior distribution, when derived in this way, inherits the strong insensitivity of NDC_0 to individual renormalization of $\pi(\boldsymbol{\theta}|H_0)$ and $\pi(\boldsymbol{\theta}|H_1)$, discussed in Section 2.1. To see this, first deduce that $\pi(\boldsymbol{\theta}|\boldsymbol{Y}) = P[H_0|\boldsymbol{Y}]\pi(\boldsymbol{\theta}|\boldsymbol{Y}, H_0) + (1 - P[H_0|\boldsymbol{Y}])\pi(\boldsymbol{\theta}|\boldsymbol{Y}, H_1)$; each of $\pi(\boldsymbol{\theta}|\boldsymbol{Y}, H_0)$ and $\pi(\boldsymbol{\theta}|\boldsymbol{Y}, H_1)$ is automatically insensitive to renormalization, and (7) shows that $P[H_0|\boldsymbol{Y}]$ is insensitive as well, since it is a function of $\tilde{\rho}_0$ and NDC_0 only.

2.6 Multiple hypotheses

The above concepts extend to situations in which there are multiple hypotheses under consideration. Following convention, the hypotheses in this context may be called "hypothesis-models," or just "models," and the context itself may be called "model choice." Suppose there are up to a countable number of models to consider, the j'th of which is denoted as $\mathcal{M}_j$, and that comparisons are made in pairs, $\mathcal{M}_j$ *versus* $\mathcal{M}_k$. The proposed formulation of neutral-data comparisons permits neutral data to on the models being compared; hence, write $\tilde{\boldsymbol{Y}}_{jk}$ to denote neutral data elicited with respect to the comparison $\mathcal{M}_j$ *versus* $\mathcal{M}_k$. The neutral-data comparison for $\mathcal{M}_j$ *versus* $\mathcal{M}_k$ is defined according to

$$\begin{aligned} NDC_{jk} &= \frac{P[\mathcal{M}_j|\boldsymbol{Y}, \mathcal{M}_j \cup \mathcal{M}_k]/P[\mathcal{M}_k|\boldsymbol{Y}, \mathcal{M}_j \cup \mathcal{M}_k]}{\tilde{\rho}_{jk}/\tilde{\rho}_{kj}} \\ &= \frac{\pi(\boldsymbol{Y}|\mathcal{M}_j)/\pi(\boldsymbol{Y}|\mathcal{M}_k)}{\pi(\tilde{\boldsymbol{Y}}_{jk}|\mathcal{M}_j)/\pi(\tilde{\boldsymbol{Y}}_{jk}|\mathcal{M}_j)}, \end{aligned} \tag{8}$$

in which $\tilde{\rho}_{jk} = P[\mathcal{M}_j|\tilde{\boldsymbol{Y}}_{jk}, \mathcal{M}_j \cup \mathcal{M}_k]$ is the corrected conditional prior probability of $\mathcal{M}_j$, given $\mathcal{M}_j \cup \mathcal{M}_k$. The analogous uncorrected prior null probability is $\rho_{jk} = P[\mathcal{M}_j|\mathcal{M}_j \cup \mathcal{M}_k]$.

When extending ideas to model choice, a critical issue is that the $\tilde{\boldsymbol{Y}}_{jk}$ must be constructed in such a way as to preserve the standard rules of probability. In other words, it is required that both sets of probabilities $\tilde{\rho}_{jk}$ and ρ_{jk} are coherent across all model comparisons, and that they are connected by the analogue to (6) given by

$$\rho_{jk}/\rho_{kj} = \frac{\tilde{\rho}_{jk}/\tilde{\rho}_{kj}}{BF_{jk}(\tilde{\boldsymbol{Y}}_{jk})}, \tag{9}$$

where $BF_{jk}(\tilde{\boldsymbol{Y}}_{jk}) = \pi(\tilde{\boldsymbol{Y}}_{jk}|\mathcal{M}_j)/\pi(\tilde{\boldsymbol{Y}}_{jk}|\mathcal{M}_k)$ is the Bayes factor for $\mathcal{M}_j$ *versus* $\mathcal{M}_k$,

calculated on $\tilde{\boldsymbol{Y}}_{jk}$. Alternatively, in terms of $\tilde{\boldsymbol{Y}}_{jk}$ only, the required criteria are

$$\begin{aligned} BF_{jk}(\tilde{\boldsymbol{Y}}_{jk}) &= \frac{1}{BF_{kj}(\tilde{\boldsymbol{Y}}_{kj})} \text{ and} \\ BF_{jl}(\tilde{\boldsymbol{Y}}_{jl}) &= BF_{jk}(\tilde{\boldsymbol{Y}}_{jk})BF_{kl}(\tilde{\boldsymbol{Y}}_{kl}), \text{ whenever } \mathcal{M}_j \Rightarrow \mathcal{M}_k \Rightarrow \mathcal{M}_l. \end{aligned} \tag{10}$$

The analogue to (7) in model choice is

$$\frac{P[\mathcal{M}_j|\boldsymbol{Y}]}{P[\mathcal{M}_k|\boldsymbol{Y}]} = \frac{\rho_{jk}}{\rho_{kj}} BF_{jk} = \frac{\tilde{\rho}_{jk}}{\tilde{\rho}_{kj}} NDC_{jk}, \tag{11}$$

where $BF_{jk} = BF_{jk}(\boldsymbol{Y})$, which allows ratios of posterior probabilities to be calculated from either the pairs (ρ_{jk}, BF_{jk}) or $(\tilde{\rho}_{jk}, NDC_{jk})$. The unconditional posterior probabilities $P[\mathcal{M}_j|\boldsymbol{Y}]$ are implied from these ratios, and may often be calculated efficiently using standard Markov chain Monte Carlo algorithms, as is demonstrated in the application example of Section 5, below.

3 Eliciting neutral data

Having now established the basic concepts of neutral-data comparisons, the remainder of the article is concerned with the development of an associated methodology, and is henceforth carried out in the limited context of testing linear hypotheses in Gaussian settings. This section's main goal is to develop guidelines for specifying neutral data in practice, but its setup is also exploited for the secondary purpose of exploring accuracy in the context of Interpretation 1 from Section 2.4.

3.1 The model and asymptotic setup

The basic hypotheses are now $H_0 : \boldsymbol{\theta} = \mathbf{0}$ *versus* $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$, for a $\nu_{n,1}$-dimensional parameter, $\boldsymbol{\theta}$. The model is

$$\boldsymbol{Y}_n|\boldsymbol{\theta}, \sigma^2 \sim N(\boldsymbol{\theta}, n^{-1}\sigma^2\boldsymbol{I}) \quad \text{and} \quad \boldsymbol{\theta}|H_1, \sigma^2 \sim N(\mathbf{0}, \sigma^2\tau_n^2\boldsymbol{I}), \tag{12}$$

where σ may or may not be known. When σ is unknown, it is supposed there is a statistic $\hat{\sigma}_n^2$, independent of $\boldsymbol{Y}_n$, for which $\nu_{n,2}\hat{\sigma}_n^2/\sigma^2 \sim \chi^2_{\nu_{n,2}}$; an associated prior is specified according to $\lambda/\sigma^2 \sim \chi^2_\kappa$, for parameters $\kappa, \lambda \geq 0$, where $\kappa = 0$ and $\lambda = 0$ are each understood as a limit. (The parameter σ is treated here in the simplest way possible, but an alternative setup, used often in linear-models analysis, would treat σ separately under H_0 and H_1, and similar results would be obtained. See Spiegelhalter

and Smith, 1982, p. 378, for discussion.) The parameter n in (12) specifies the precision of the model, given σ, whose exact (*i.e.*, non-asymptotic) setting may be determined by the conventions of the analysis context, as is done in the application example of Section 5, below. Note also that the prior may depend on n, through τ_n, but most of the properties deduced below will hold when τ_n is fixed at "some large value."

Asymptotic analysis will consider the behavior of $NDC_{n,0}$ as $n \to \infty$, $n\tau_n^2 \to \infty$, and $\nu_{n,2} \to \infty$, where $NDC_{n,0}$ is the neutral-data comparison for H_0 *versus* H_1 under the model (12). Depending on the context, $\nu_{n,1}$ may be held fixed at $\nu_{n,1} = \nu_1$ or may vary with n, the purpose the varying $\nu_{n,1}$ being to identify leading constants, and not to study the impact of dimensionality, which is treated in Section 4. Although $\boldsymbol{\theta}$ may depend on n through its dimensionality, $\nu_{n,1}$, its individual entries are fixed, and so subscripting is omitted. Also fixed are the parameters σ, κ, and λ.

It will be convenient to define standardized versions of the observed and neutral data as $\boldsymbol{Z}_n = n\boldsymbol{Y}_n/\sigma$ and $\tilde{\boldsymbol{Z}}_n = n\tilde{\boldsymbol{Y}}_n/\sigma$. The model provides that $\|\boldsymbol{Z}_n\|^2 \sim \chi^2_{\nu_{n,1}}(n\|\boldsymbol{\theta}\|^2/\sigma^2)$ and $\nu_{n,2}\hat{\sigma}_n^2/\sigma^2 \sim \chi^2_{\nu_{n,2}}$, given $\boldsymbol{\theta}$ and σ^2; subsequently, Chebyshev's inequality implies

$$\begin{aligned} \|\boldsymbol{Z}_n\|^2 &= \nu_{n,1} + n\|\boldsymbol{\theta}\|^2/\sigma^2 + O\left(\sqrt{2\nu_{n,1} + 4n\|\boldsymbol{\theta}\|^2/\sigma^2}\right) \\ \hat{\sigma}_n^2 &= \sigma^2 + O\left(\sqrt{2\sigma^4/\nu_{n,2}}\right). \end{aligned} \tag{13}$$

The mode of convergence here is "in probability," which is too weak for the Bayesian concept of asymptotic consistency, and must be strengthened to the stronger mode of "almost sure" convergence. To this end, almost sure convergence is assumed in (13), and in other expressions like it that will follow, as part of the construction of the model. Refer to Spitzner (2008, sec. 4.1) for detailed discussion of this point.

When σ is known, the relevant neutral-data comparison (2) has

$$-2\log NDC_{n,0} = w_n\left\{\|\boldsymbol{Z}_n\|^2 - \|\tilde{\boldsymbol{Z}}_n\|^2\right\}, \tag{14}$$

where $w_n = 1/\{1 + 1/(n\tau_n^2)\}$. When σ is unknown, set

$$F_n = \frac{n\|\boldsymbol{Y}_n\|^2/\nu_{n,1}}{\hat{\sigma}_n^2 + \lambda/\nu_{n,2}} = \frac{\sigma^2\|\boldsymbol{Z}_n\|^2/\nu_{n,1}}{\hat{\sigma}_n^2 + \lambda/\nu_{n,2}}, \tag{15}$$

and suppose that neutral data take the form of a neutral value, $\tilde{F}_n$, of F_n; the neutral-data comparison has $-2\log NDC_{n,0} = Q_n(F_n) - Q_n(\tilde{F}_n)$, where

$$Q_n(F) = -(\nu_{n,1} + \nu_{n,2} + \kappa)\log\left\{1 - w_n\frac{(\nu_{n,1}/\nu_{n,2})F}{1 + (\nu_{n,1}/\nu_{n,2})F}\right\}. \tag{16}$$

The expansion (13) provides intuition for specifying neutral data: either $\|\tilde{\boldsymbol{Z}}_n\|^2$ is to lie "near" $\nu_{n,1}$, or $\tilde{F}_n$ is to lie "near" one. Accordingly, write $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_{n,1} + r_n\sqrt{\nu_{n,1}}$ or $\tilde{F}_n = 1 + r_n/\sqrt{\nu_{n,1}}$, for some r_n, which parameterizes "how near" a neutral-data statistic is set from its reference value. Treated this way, straightforward analysis precisely delineates conditions for asymptotic-consistency of $NDC_{n,0}$: if $r_n \to \infty$ and $\nu_{n,1} = o(\nu_{n,2})$, then (13) provides that

$$NDC_{n,0} \approx \begin{cases} \exp\{w_n r_n \sqrt{\nu_{n,1}}/2\} & \text{under } H_0 \\ \exp\{-w_n n\|\boldsymbol{\theta}\|^2/(2\sigma^2)\} & \text{under } H_1, \text{ if } r_n = o(n/\sqrt{\nu_{n,1}}); \end{cases} \tag{17}$$

asymptotic-consistency fails under H_0 when $r_n = O(1)$.

3.2 Eliciting neutral data from a probability distribution

Working strictly with asymptotic-consistency criteria, one possible approach to eliciting neutral data would be to decide on a suitable rate at which it is desirable to establish H_0, and then use (17) to set r_n in such a way as to achieve it. This may provide a suitable working method in some contexts, but to expand the possibilities for understanding and developing $NDC_{n,0}$ it is desirable to connect the elicitation of neutral data to traditional expressions of prior knowledge. That connection is made by working formally with Interpretation 1 from Section 2.4, by which $\tilde{\boldsymbol{Y}}_n$ is thought to satisfy $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) = 1$. (The notation here follows Section 2, but with the subscript n added.) The criterion $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) = 1$ links $\tilde{\boldsymbol{Y}}_n$ to the authentic prior, and by that criterion it will be shown that certain isolated characteristics of the prior determine the rate at which $\|\tilde{\boldsymbol{Z}}_n\|^2 \to \infty$. A suitable setting for neutral data is thus possible, using traditional tools, by focusing specifically on those characteristics.

For simplicity, the technique is developed supposing that σ is known; generalizations to unknown σ involve additional technical complication, but yield essentially the same results. Operationally, settings derived in the former case may be translated to the latter by extracting r_n from the parameterization $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_{n,1} + r_n\sqrt{\nu_{n,1}}$ and substituting in $\tilde{F}_n = 1 + r_n/\sqrt{\nu_{n,1}}$.

Write $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)$ to denote the authentic conditional prior under H_1, with σ known. The following theorem provides the desired analysis for cases where the authentic prior is of a standard type, such that $\tilde{\pi}(\mathbf{0}|H_1, \sigma) > 0$, although its statement (*iii*) covers more general cases also. The symbol "$\asymp$" used in statement (*ii*) is asymptotic similarity: $a_n \asymp b_n$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$.

Theorem 1. *Consider testing* $H_0 : \boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ *under the model (12), with* σ *known and* $\nu_{n,1} = \nu_1$ *fixed. Suppose* $\widetilde{BF}_{n,0}$ *is defined from a twice-differentiable prior density* $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)$ *such that* $s_n(\boldsymbol{t}) = n\{\|\boldsymbol{t}\|^2 - 2\boldsymbol{t}^T\boldsymbol{y}\}/(2\sigma^2) - \log\tilde{\pi}(\boldsymbol{t}|H_1, \sigma)$ *is, for any* $\boldsymbol{y}$*, locally convex except possibly at* $\boldsymbol{t} = \mathbf{0}$*. Define* NDC_0 *according to (14), with* $n\tau_n \to \infty$*.*

(i) *Suppose* $\tilde{\pi}(\mathbf{0}|H_1, \sigma) > 0$ *and* $1/\tau_n = O(1)$*. If* $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_1 \log(n/\sigma^2) - 2\log\tilde{\pi}(\mathbf{0}|H_1, \sigma) - \nu_1\log(2\pi)$*, then* $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) \to 1$*, and also* $NDC_{n,0} \approx \widetilde{BF}_{n,0}$ *under* H_0*, but* $NDC_{n,0} \approx C_0\widetilde{BF}_{n,0}$ *under* H_1*, where* $C_0 = \tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)/\tilde{\pi}(\mathbf{0}|H_1, \sigma)$*.*

(ii) *Suppose* $\tilde{\pi}(\mathbf{0}|H_1, \sigma) > 0$*. If* $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_1\log n + O(1)$*, then* $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) \asymp 1$*, and also* $-2\log NDC_{n,0} \approx -2\log\widetilde{BF}_{n,0}$ *(or* $NDC_{n,0} \asymp \widetilde{BF}_{n,0}$ *if* $1/\tau_n = O(1)$*) under both* H_0 *and* H_1*.*

(iii) *If* $(\log n)/\|\tilde{\boldsymbol{Z}}_n\|^2 = O(1)$*, but still* $\|\tilde{\boldsymbol{Z}}_n\|^2/n \to 0$*, then* $-2\log NDC_{n,0} \approx -2\log\widetilde{BF}_{n,0}$*, under* H_1*. If* $(\log n)/\|\tilde{\boldsymbol{Z}}_n\|^2 \to 0$*, then* $\widetilde{BF}_{n,0}/NDC_{n,0} \to 0$*, under* H_1*.*

Statement (*i*) of Theorem 1 provides a very close examination of accuracy between $NDC_{n,0}$ and $\widetilde{BF}_{n,0}$ when neutral data are selected so that $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) = 1$. It shows the approximation to be quite accurate under H_0 and under portions of H_1 that are near H_0, thereby providing strong conceptual support for Interpretation 1 of $NDC_{n,0}$. Statement (*ii*) suggests a specific *default* choice for neutral data, which serves in that capacity by covering the wide range of standard scenarios relevant to that statement. Stated for both $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_{n,1} + r_n\sqrt{\nu_{n,1}}$ and $\tilde{F}_n = 1 + r_n/\sqrt{\nu_{n,1}}$, assuming a common r_n, that setting has

$$\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_{n,1}\log n \quad \text{and} \quad \tilde{F}_n = \log n. \tag{18}$$

Further properties of this setting are explored in Section 3.3, below. Statement (*iii*) identifies limits to the accuracy of $NDC_{n,0}$ when $\|\tilde{\boldsymbol{Z}}_n\|^2 \to \infty$ at rates possibly faster than the proposed default setting in (18).

It is of interest to carry out a similar examination of the following class of non-standard priors.

Definition 1. *A prior density* $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)$ *is a* regular non-local prior density with spherical contours near zero *if it is twice-differentiable and*

$$\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma) \approx c^{-\nu_1/2}\exp\left\{-\frac{1}{2}f\left(\frac{\sigma^2}{\|\boldsymbol{\theta}\|^2}\right)\right\} \text{ as } \|\boldsymbol{\theta}\|^2 \to 0, \tag{19}$$

where c is a normalizing constant, and f is an increasing, twice-differentiable function such that $f_0(x) = f(e^x)$ is convex, and furthermore $\gamma(x) = xf'(x)/f(x) \to \gamma$, as $x \to \infty$, for some $0 \leq \gamma < \infty$.

Each prior of the type in Definition 1 has $\tilde{\pi}(\mathbf{0}|H_1, \sigma) = \mathbf{0}$, and the function $f(x)$ controls the rate at which $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma) \to \mathbf{0}$ as $\boldsymbol{\theta} \to \mathbf{0}$. Johnson and Rossell (2010) study the specific cases where $f(x) \propto \log x$ and $f(x) \propto x^\gamma$ for $\gamma > 0$. They suggest that non-local priors are intuitively attractive for imposing continuity in $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)$ as $\boldsymbol{\theta} \to \mathbf{0}$, and furthermore show that such continuity improves the asymptotic performance of $\widetilde{BF}_{n,0}$ in establishing H_0. Moreover, they show that the degree of improvement is determined by the rate at which $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma) \to 0$. This last pattern is reminiscent of the asymptotic behavior observed in (17), by which the divergence rate of $NDC_{n,0}$ is determined by the rate at which $\|\tilde{\boldsymbol{Z}}_n\|^2 \to \infty$. The following theorem ties these properties together, while complementing the results of Theorem 1.

Theorem 2. *Consider testing $H_0 : \boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ under the model (12), with σ known and $\nu_{n,1} = \nu_1$ fixed. Suppose $\widetilde{BF}_{n,0}$ is defined from a prior density $\tilde{\pi}(\boldsymbol{\theta}|H_1, \sigma)$, of the type in Definition 1, and let $f(x)$ be the function in (19), such that $\gamma(x) = xf'(x)/f(x) \to \gamma \geq 0$. Suppose further that $s_n(\boldsymbol{t}) = n\{\|\boldsymbol{t}\|^2 - 2\boldsymbol{t}^T\boldsymbol{y}\}/(2\sigma^2) + f(\sigma^2/\|\boldsymbol{t}\|^2)/2$ is, for any $\boldsymbol{y}$, locally convex except at $\boldsymbol{t} = \mathbf{0}$.*

(i) If $\tilde{\boldsymbol{Z}}_n$ is such that $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) \to 1$, then $\|\tilde{\boldsymbol{Z}}_n\|^2$ solves

$$\|\tilde{\boldsymbol{Z}}_n\|^2 \approx \frac{(\gamma+1)^2}{2\gamma+1}\left[f\left(\frac{n}{\tilde{a}^2\|\tilde{\boldsymbol{Z}}_n\|^2}\right) + \nu_1 \log n\right], \tag{20}$$

where $\tilde{a} = 1 + \gamma/(\gamma+1)$, hence $1 \leq \tilde{a} < 2$. The solution is unique up to the accuracy of the approximation.

(ii) At each n, the equation

$$a_n^2\|\boldsymbol{Z}_n\|^2\left(1 - \frac{1}{a_n}\right) = \left(\frac{n}{a_n^2\|\boldsymbol{Z}_n\|^2}\right) f'\left(\frac{n}{a_n^2\|\boldsymbol{Z}_n\|^2}\right) \tag{21}$$

is solved uniquely for $a_n > 0$. Subsequently, for such a_n,

$$-2\log\widetilde{BF}_{n,0} \approx -\left[(\gamma+1)f\left(\frac{n}{a_n^2\|\boldsymbol{Z}_n\|^2}\right) + \nu_1 \log n\right], \tag{22}$$

under H_0.

Statement (i) of Theorem 2 is especially useful to translate the behavior of the prior near H_0 into a setting for neutral data. The idea here is that, to specify neutral data,

prior knowledge might be elicited sufficiently well as to fix a suitable function $f(x)$, from which it becomes possible to solve (20) for the desired quantity $\|\tilde{\boldsymbol{Z}}_n\|^2$. Such elicitation may or may not yield sufficient information with which to incorporate $f(x)$ into the operational prior, since all attention is focused on a neighborhood near H_0. Supposing that it does not, our interest is to examine the accuracy of $NDC_{n,0}$ as an approximation to $\widetilde{BF}_{n,0}$ when $f(x)$ is used only to specify $\|\tilde{\boldsymbol{Z}}_n\|^2$. The accuracy of that approximation under H_1 is described in Theorem 1.*iii*; its accuracy under H_0 is described by the following result.

Corollary 1. *In the setting of Theorem 2, define $NDC_{n,0}$ according to (14), with $\|\tilde{\boldsymbol{Z}}_n\|^2$ as in (20), and $n\tau_n \to \infty$.*

(i) *Suppose $f(x) = k\log x$ for $k > 0$. Then $\|\tilde{\boldsymbol{Z}}_n\|^2 = (k+\nu_1)\log n$, $a_n = 1/2 + \{k/\|\boldsymbol{Z}_n\|^2 + 1/4\}^{1/2}$, $-2\log\widetilde{BF}_{n,0} \approx -(k+\nu_1)\log n$, and $-2\log NDC_{n,0} \approx -2\log\widetilde{BF}_{n,0}$, under H_0.*

(ii) *Suppose $f(x) = kx^\gamma$ for $k, \gamma > 0$. Then $\|\tilde{\boldsymbol{Z}}_n\|^2 \approx C_1 n^{\gamma/(\gamma+1)}$ and $-2\log\widetilde{BF}_{n,0} \approx -C_2 n^{\gamma/(\gamma+1)}$, where*

$$C_1 = \left\{ \frac{k(\gamma+1)^2}{\tilde{a}^{2\gamma}(2\gamma+1)} \right\}^{1/(\gamma+1)} \quad \text{and} \quad C_2 = \frac{k(\gamma+1)}{\gamma^{\gamma/(\gamma+1)}}, \tag{23}$$

and $-2\log NDC_{n,0} \approx (C_1/C_2)\{-2\log\widetilde{BF}_{n,0}\}$, under H_0.

Corollary 1 shows that $NDC_{n,0}$ approximates $\widetilde{BF}_{n,0}$ reasonably well (at least on a logarithmic scale), even when the operational and authentic priors are substantially different in the region near H_0. Accuracy does decline in moving from $f(x) \propto \log x$ to $f(x)\log x^\gamma$, and in the latter case it is interesting that accuracy depends on the leading constant k, through C_1 and C_2. Upon further examination, it is seen that $C_1/C_2 \to 1$ as $\gamma \to 0$ and $C_1/C_2 \to 1/(4k)$ as $\gamma \to \infty$, thus suggesting that accuracy is fine-tuned at $k = 1/4$. At that setting, numerical exploration provides that C_1/C_2 has a unique minimum of 0.75 at $\gamma = 0.5$, and other contours of $C_1/C_2 = 0.80$ at $\gamma = 0.15$ and 1.56, $C_1/C_2 = 0.90$ at $\gamma = 0.04$ and 5.16, and $C_1/C_2 = 0.95$ at $\gamma = 0.01$ and 12.13.

3.3 Comparisons

Taking into account both Theorem 1 and Corollary 1, the results of Section 3.2 suggest that Interpretation 1 of $NDC_{n,0}$ is quite reasonable in the sense that its scheme to set $\tilde{\boldsymbol{Y}}$ so that $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n) = 1$ puts $NDC_{n,0}$ well within the vicinity of $\widetilde{BF}_{n,0}$. Nevertheless,

that analysis uncovers disagreements between $NDC_{n,0}$ and $\widetilde{BF}_{n,0}$ also, which, upon careful examination, indicate that the direction of any bias of $NDC_{n,0}$ toward H_0 or H_1 varies by situation: on one hand, for standard prior types such that $\tilde{\pi}(\boldsymbol{\theta}|H_1,\sigma)$ decreases as $\boldsymbol{\theta}$ shifts from zero, the constant $C_0 < 1$ in Theorem 1.*i* indicates that $NDC_{n,0}$ is biased toward H_1; on the other hand, for a non-local prior with $f(x) = (1/4)\log x$, the ratio $C_1/C_2 < 1$ in Corollary 1.*ii* indicates a bias of $NDC_{n,0}$ toward H_0.

In addition to bias comparisons with $\widetilde{BF}_{n,0}$, comparisons of $NDC_{n,0}$ with closely related methods are also possible. For this purpose, consider the repeated sampling case in which $\boldsymbol{Y}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i$ for independent $\boldsymbol{X}_i \sim N(\boldsymbol{\theta},\sigma^2\boldsymbol{I})$. In this context, a minimal training sample, as both Spiegelhalter and Smith (1982) and Berger and Pericchi (1996) would define it, is an individual $\boldsymbol{X}_i$; hence, if the operational prior has $\pi(\boldsymbol{\theta}|H_1,\sigma) \propto 1$, respective Bayes factors prescribed by those authors' techniques are

$$\begin{aligned} BF^{SS}_{n,0} &= n^{\nu_{n,1}/2}\exp\left\{-\|\boldsymbol{Z}_n\|^2/2\right\} \\ BF^{AI}_{n,0} &= n^{\nu_{n,1}/2}\hat{A}_n^{-1}\exp\left\{-\|\boldsymbol{Z}_n\|^2/2\right\}, \end{aligned} \tag{24}$$

superscripting "SS" for "Spiegelhalter and Smith," and "AI" for Berger and Pericchi's "arithmetic intrinsic" Bayes factor, where

$$\hat{A}_n = \frac{1}{n}\sum_{i=1}^{n}\exp\left\{-\|\boldsymbol{X}_i\|^2/(2\sigma^2)\right\} \to 2^{-\nu_{n,1}/2}\exp\left\{-\|\boldsymbol{\theta}\|^2/(4\sigma^2)\right\}.$$

The BIC criterion of Schwarz (1978) also implies an approximation to a Bayes factor, which in this context is identical to $BF^{SS}_{n,0}$. Pérez and Berger (2002) discuss few details of Bayes factors derived by subjective expected-posterior priors, but those derived by objective criteria for the most part copy $BF^{AI}_{n,0}$.

By applying the device of imaginary results, each formula in (24) implies a corresponding setting for neutral data. Write $\tilde{\boldsymbol{\mathcal{X}}}_n = (\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_n)$ to denote a neutral-data version of $\boldsymbol{\mathcal{X}}_n = (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$. It is seen that $BF^{SS}_{n,0}(\tilde{\boldsymbol{\mathcal{X}}}_n) = 1$ when calculated on neutral data such that $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_1\log n$, which matches the default setting (18), and yields $NDC_0 = BF^{SS}_{n,0}$. The neutral data implied analogously by the intrinsic Bayes factor are unusual, since $\|\tilde{\boldsymbol{Z}}_n\|^2$ varies among the solutions to $BF^{AI}_{n,0}(\tilde{\boldsymbol{\mathcal{X}}}_n) = 1$. Another interesting choice sets $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_1\log n - 2\log\hat{A}_n$, which makes $BF^{AI}_{n,0}(\tilde{\boldsymbol{\mathcal{X}}}_n) = 1$ and $NDC_0 = BF^{AI}_{n,0}$, but it, too, is unusual since it formulates neutral data partly as a function of the observed data. Yet, the potential benefit of this latter choice is apparent, for if the authentic prior has $\boldsymbol{\theta}|H_1,\sigma^2 \sim N(\boldsymbol{0},2\sigma^2\boldsymbol{I})$, which is Berger and Pericchi's "intrinsic prior" for this context, then $NDC_{n,0} \approx \widetilde{BF}_{n,0}$ under both H_0 and H_1.

The possibility suggested above of a data-dependent choice for neutral data is intriguing, but tangential to the present focus on a more straightforward neutral-data concept; further discussion is deferred to Section 6. Absent such a choice, the default setting $\|\tilde{\boldsymbol{Z}}_n\|^2 = \nu_1 \log n$ puts $NDC_{n,0}$ subject to a criticism shared with $BF^{SS}_{n,0}$, but not $BF^{AI}_{n,0}$, of being asymptotically biased toward H_1. Yet, as has been shown, $NDC_{n,0}$ is more flexible than $BF^{SS}_{n,0}$ in that its emphasis on H_0 or H_1 can be adjusted by a suitable non-default choice of $\tilde{\boldsymbol{Y}}_n$. This becomes critically important in Section 4, below, where $NDC_{n,0}$ is adjusted to emphasize H_0 much more strongly than does either of $BF^{SS}_{n,0}$ or $BF^{AI}_{n,0}$, in order to achieve asymptotic-consistency in high-dimensions. It is important to keep in mind, too, that criticisms of asymptotic bias relative to a Bayes factor apply only to Interpretation 1 of $NDC_{n,0}$, since Interpretation 2 motivates $NDC_{n,0}$ as a valid Bayesian procedure in its own right, not just an approximation to one.

4 Model choice in high-dimensions

Discussion here and in the next section examines the performance of neutral-data comparisons in non-trivial applications. The present section considers high-dimensional performance in a context where the goal is to select the components of a "true model" from a huge pool of candidate components. This goal is relevant to a variety of scientific disciplines, most notably genetics, where the number of candidate components may be in the hundreds or thousands, or even higher. (See, *e.g.*, Fan and Lv, 2010, for a recent survey.)

To study this context, the Gaussian setup of Section 3.1 is extended to describe individual components, $\boldsymbol{Y}_{n,i}$ and $\hat{\sigma}^2_{n,i}$, collected independently across $i = 1, \ldots, p_n$. These have

$$\boldsymbol{Y}_{n,i} | \boldsymbol{\theta}_i, \sigma_i^2 \sim N(\boldsymbol{\theta}_i, n^{-1}\sigma_i^2 \boldsymbol{I}) \quad \text{and} \quad \nu_{n,2i}\hat{\sigma}^2_{n,i}/\sigma_i^2 | \sigma_i^2 \sim \chi^2_{\nu_{n,2i}}, \tag{25}$$

in which $\boldsymbol{\theta}_i$ is of dimension $\nu_{n,1i}$. The setup is now sufficiently complicated that it will be convenient to write $\boldsymbol{\mathcal{Y}}_n$ to denote "all of the data" (*i.e.*, all $\boldsymbol{Y}_{n,i}$ and $\hat{\sigma}^2_{n,i}$ across $i = 1, \ldots, p_n$). Component-specific hypotheses are

$$H_{0i} : \boldsymbol{\theta}_i = \boldsymbol{0} \quad \textit{versus} \quad H_{1i} : \boldsymbol{\theta}_i \neq \boldsymbol{0}. \tag{26}$$

The prior is specified such that the $\boldsymbol{\theta}_i$ and σ_i^2 are independent across i, and each has

$$\boldsymbol{\theta}_i | H_{1i}, \sigma_i^2 \sim N(\boldsymbol{0}, \sigma_i^2 \tau^2_{n,i} \boldsymbol{I}) \quad \text{and} \quad \lambda_i/\sigma_i^2 \sim \chi^2_{\kappa_i}. \tag{27}$$

Each component is thus an independent copy of the scenario in Section 3.1. The testing problem is defined in the model-choice context of Section 2.6. Relevant hypothesis-models are those of the form

$$\mathcal{M}_{n,j} = \mathcal{M}_n(A_j) = \left\{ \bigcup_{i \in A_j} H_{0i} \right\} \cap \left\{ \bigcup_{i \notin A_j} H_{1i} \right\} \tag{28}$$

across all index-subsets $A_j \subset \{1, \ldots, p_n\}$. Collectively, the parameters $\boldsymbol{\theta}_i$ are consistent with exactly one of these models, $\mathcal{M}_n^* = \mathcal{M}_n(A_n^*)$, where $A_n^* = \{i : H_{0i} \text{ is true}\}$. This is the "true" model. Accordingly, the goal of analysis is to select a promising candidate for $\mathcal{M}_n^*$ and report with it an assessment of evidence in favor of that candidate actually being $\mathcal{M}_n^*$.

This problem is difficult when p_n is large, and it is especially difficult when A_n^* is large also, in which case the true-model configuration is said be "sparse." Relevant performance criteria for asymptotic evaluation take $p_n \to \infty$ as $n \to \infty$, and consider both the rate at which p_n increases and the strength at which the true H_{1i} reveal themselves to be true. One particularly challenging set of such criteria is formulated by Fan and Lv (2008), which describes sparsity in "ultra-high" dimensions: fix constants $a > 0$ and $b > 0$ such that $a < 1 - b$; ultra-high dimensionality allows $\log p_n = O(n^a)$, and sparsity in this context requires only the existence of some $c > 0$ such that $\min_{i \notin A_n^*} \left\{ \|\boldsymbol{\theta}_i\|^2 / \sigma_i^2 \right\} \geq c n^{-b}$. Their strategy for carrying out data analysis in ultra-high dimensions is to first apply "sure independence screening" (SIS) to reduce the context to mere "high-dimensionality," in which p_n increases only at a polynomial rate; once high-dimensionality is achieved, any of a number of well-known procedures can take over the analysis. SIS itself works by first specifying some postulated number, d_n, of true H_{1i}, and then selecting for A_n^* the component indices associated with the smallest $p_n - d_n$ values of $\|\boldsymbol{Y}_{n,i}\|$. Fan and Lv identify (for the case $\nu_{n,1i} = 1$) a certain polynomial rate for d_n that will guarantee convergence to one of the probability that the indices selected for A_n^* by SIS are truly in A_n^*.

The following describes how neutral-data comparisons yield a procedure that performs as well as SIS in ultra-high dimensions. To formulate the procedure, first observe that the neutral-data comparison, $NDC_{n,0i}$, for H_{0i} *versus* H_{1i} has

$$-2 \log NDC_{n,0i} \quad = \quad Q_{n,i}(F_{n,i}) - Q_{n,i}(\tilde{F}_{n,i}), \tag{29}$$

where $F_{n,i}$ and $Q_{n,i}$ are identical to (15) and (16), but defined with the subscript i added to all of $\boldsymbol{Y}_n$, $\hat{\sigma}_n$, $\nu_{n,1}$, $\nu_{n,2}$, λ, κ, and w_n, for which $w_{n,i} = 1/\{1 + 1/(n\tau_{n,i}^2)\}$.

Independence assures that the coherency criteria (10) are satisfied for pairwise comparisons of the models (28), provided that each $\tilde{F}_{n,i}$ is independent of any particular comparison. Here, these quantities are parameterized as $\tilde{F}_{n,i} = 1 + r_n/\sqrt{\nu_{n,1i}}$, using the same r_n across all components, whose preferred settings are identified in Theorem 3, below. The proposed model-choice procedure derives from the formula

$$P[\mathcal{M}_n(A)|\boldsymbol{\mathcal{Y}}_n] \quad = \quad \left\{\prod_{i\in A} P[H_{0i}|\boldsymbol{Y}_{n,i}, \hat{\sigma}_{n,i}]\right\}\left\{\prod_{i\notin A} P[H_{1i}|\boldsymbol{Y}_{n,i}, \hat{\sigma}_{n,i}]\right\}, \qquad (30)$$

where each $P[H_{0i}|\boldsymbol{Y}_{n,i}, \hat{\sigma}_{n,i}] = \left\{1 + \left(\tilde{\rho}_{n,0i}^{-1} - 1\right)/NDC_{0,i}\right\}^{-1}$, as in (7), for component-specific $\tilde{\rho}_{n,0i}$. The procedure itself proceeds by searching among subsets $A \subset \{1, \ldots, p_n\}$ for that which maximizes (30). If it can be shown that $P[\mathcal{M}_n^*|\boldsymbol{\mathcal{Y}}_n] \to 1$, then, since probabilities sum to one, it will follow that A_n^* eventually maximizes (30), and the procedure will select it.

The following result gives conditions under which the desired asymptotic-consistency is achieved.

Theorem 3. *Assume the high-dimension model (25) and prior (27), specified with* $\kappa_i, \lambda_i = 0$. *Suppose the component posterior probabilities in (30) are defined with* $NDC_{n,0i}$ *as in (29),* $\tilde{\rho}_{n,i} = 1/2$, *and* $\tilde{F}_{n,i} = 1 + r_n/\sqrt{\nu_{n,1i}}$. *Suppose further that each* $\nu_{n,1*} < \nu_{n,1i} < \nu_{n,1}^*$, $nc_i < \nu_{n,2i}$, $w_{n*} < w_{n,i} < w_n^*$ *for positive bounds* $\nu_{n,1*}$, $\nu_{n,1}^*$, c_i, w_{n*}, *and* w_n^* *such that* $\nu_{n,1}^* = o(n)$, c_i *is independent of* n, *and* $w_{n*} \to 1$. *Then* $P[\mathcal{M}_n^*|\boldsymbol{\mathcal{Y}}_n] \to 1$ *whenever all of*

$$-\frac{\sqrt{\nu_{n,1}^*}\log(1-w_n^*)}{r_n} = o(1), \frac{\log p_n}{r_n\sqrt{\nu_{n,1*}}} = o(1), \text{ and } \frac{r_n\sqrt{\nu_{n,1*}}}{n\min\limits_{i\notin A_n^*}\left\{\|\boldsymbol{\theta}_i\|^2/\sigma_i^2\right\}} = o(1).$$

Applying Theorem 3 to Fan and Lv's performance criteria (for $\nu_{n,1i} = 1$, say), it is seen that by setting $r_n = n^{(1+a-b)/2}$, and the $w_{n,i}$ so that $-\log(1-w_n^*) = O(r_n)$, the conditions of the theorem are satisfied, and the proposed procedure achieves asymptotic-consistency. Moreover, since (30) is associated with a full posterior distribution, the procedure not only *screens* components to lower dimensions, but provides a coherent testing and estimation methodology that works *in* ultra-high dimensions.

It is interesting to compare the required settings for asymptotic-consistency between the current model-choice problem and that of testing a single component, H_{0i} *versus* H_{1i}, discussed in Section 3. Observe that the current problem requires a faster rate, $r_n = n^{(1+a-b)/2}$, than the default setting (18), where the rate of r_n is $\log n$. Comparing with (17), the model-choice setting represents an adjustment toward more quickly

	Main effects			*2-way interactions*			*3-way int.*
Effect group	Onc.	Material	Form	M × O	F × O	M × F	M × F × O
d.f. (p_g)	12	1	2	12	24	2	24

Table 1: *Effect groupings and degrees of freedom of the simulated tumor data.*

establishing any individual H_{0i}, an adjustment that might reasonably be interpreted to accomodate for multiple testing. (See, *e.g.*, Scott and Berger, 2006, for related discussion.) It is also interesting to interpret $r_n = n^{(1+a-b)/2}$ in the context of Corollary 1.*ii*: each $NDC_{n,0i}$ is seen to approximate a Bayes factor formulated from a non-local prior, as in Definition 1, with $f(x) \propto x^\gamma$, where $\gamma = (1 + a - b)/(1 - a + b) > 0$.

5 Linear-models analysis

This next section showcases the capabilities of neutral-data comparisons in the applied context of "analysis of variance" (ANOVA). Bayesian solutions to ANOVA often focus on estimation (*cf.* Gelman, 2005), but it is shown in what follows how neutral-data comparisons can beneficially supplement such solutions by adding a testing component.

5.1 Simulated tumor data

The example analysis works with a data set described in Hoaglin, Mosteller, and Tukey (1991, secs. 6A, 7A). The data set compiles measurements obtained from thirteen oncologists as they repeatedly judged, by touch only, the cross-sectional areas of six simulated tumors. These six tumors are distinct from each other and represent all possible combinations of two materials (cork or rubber) and three forms (small, oblong, or large). Each was judged twice per oncologist. The data set thus comprises two replications of seventy-eight possible factor combinations, totaling 156 measurements altogether. The layout is a standard (balanced) three-way factorial ANOVA, in which the factors are "oncologists," "materials," and "forms."

A linear-models formulation of ANOVA uses linear transformations to identify independent components of a factorial structure. The components are grouped by "effect" into an "ANOVA table," which also records the number of components per group as the effect's "degrees of freedom." An effect is either a "main effect," of which there is one per factor, or an "interaction effect," of which each may involve two or more

factors. For the simulated tumor data, there are seven effect groups: three main effects, three two-way interactions, and one three-way interaction. These and their associated degrees of freedom are listed in Table 1. A related concept is the "grand mean," or overall average, which is an additional component that represents the "center" of the model; each effect-component is understood as a shift from center. The grand mean has one degree of freedom. See Hoaglin, Mosteller, and Tukey (1991) for details of the construction of these components.

Traditional targets of inference in ANOVA are each effect-group's "finite population standard deviation" (FPSD), using terminology from Gelman (2005). Each FPSD is a rescaled total Euclidean distance from zero of the group's component mean parameters, where rescaling standardizes FPSD across groups with respect to degrees of freedom. When an FPSD is positive, the effect is said to be "present." In addition to an effect's overall FPSD, certain linear relationships among its components may be interesting as well. Pairwise differences between a main effect's individual components are interesting for the possibility of "separating" the levels of a factor into subgroups, in the manner of a partition analysis. Similarly, the individual components of an interaction effect are interesting for providing details of any non-additive relationships in the factorial structure.

5.2 The linear model and hypotheses

The model (25), of Section 4, serves to describe the individual components of the ANOVA scheme. Among p_n total components, the index-values i are partitioned into subsets $B_1, \ldots, B_{G_n}$, corresponding to $G_n - 1$ effect-groups, plus one additional group for the grand mean. The g'th subset has $p^*_g = |B_g|$ degrees of freedom. In ANOVA, the dimension, $\nu_{n,1i}$, of component $\boldsymbol{Y}_{n,i}$, is typically one, but there is little additional complication to working generally, provided that components in the same group have the same dimension, $\nu_{n,1i} = \nu^*_{n,1g}$ across $i \in B_g$. The FPSD for group g is $FPSD_g = \left\{ \sum_{i \in B_g} \|\boldsymbol{\theta}_i\|^2 / (p^*_g \nu^*_{1g}) \right\}^{1/2}$. ANOVA also defines "variability due to error," which refers to a common variance parameter $\sigma^2 = \sigma^2_1 = \cdots = \sigma^2_{p_n}$. Associated with this is an aggregated statistic $\hat{\sigma}^2_n$, independent of the $\boldsymbol{Y}_{n,i}$, such that $\nu_{n,2}\hat{\sigma}^2_n / \sigma^2 \sim \chi^2_{\nu_{n,2}}$, where $\nu_{n,2}$ is the "error degrees of freedom." A convention in ANOVA is to regard the error degrees of freedom as an indicator of precision; hence, the index of precision in (25) is set to $n = \nu_{n,2}$. The simulated tumor data has $\nu_{n,2} = 78$.

The testing framework here is model-choice, with hypothesis-models defined in the

following way by linear constraints. It is supposed that $\mathcal{M}_j$ identifies A_j groups, and on each group $g \in A_j$ is placed u_j^g linearly independent constraints: the model is

$$\mathcal{M}_j : \sum_{i \in B_g} c_{ij}^{gl} \boldsymbol{\theta}_i = \mathbf{0} \text{ for } l = 1, \ldots, u_j^g, \text{ across } g \in A_j, \tag{31}$$

for coefficients c_{ij}^{gl} of the l'th constraint. A group indexed by $g \notin A_j$ has $u_j^g = 0$. The continuous portion of the prior is specified conditionally, given each $\mathcal{M}_j$, as a Gaussian prior on the model's "free parameters:" To describe this portion of the prior, it is assumed, without loss of generality, that the coefficients, c_{ij}^{gl}, have been orthonormalized within groups by the Gram-Schmidt procedure (*cf.* Bellman, 1960), so that

$$\sum_{i \in B_g} c_{ij}^{gl} c_{ij}^{gm} = 1 \text{ if } l = m \text{ or } 0 \text{ if } l \neq m \tag{32}$$

across $l, m = 1, \ldots, u_j^g$. It is furthermore possible to augment the coefficients c_{ij}^{gl} for $l = 1, \ldots, u_j^g$ with additional coefficients c_{ij}^{gl} for $l = u_j^g, \ldots, p_g^*$ such that (32) holds across $l, m = 1, \ldots, p_g^*$. The conditional prior is now

$$\sum_{i \in B_g} c_{ij}^{gl} \boldsymbol{\theta}_i | \mathcal{M}_j, \sigma^2 \sim N(\mathbf{0}, n^{-1} \sigma^2 \tau_g^{*2} \boldsymbol{I}), \tag{33}$$

independently across $l = u_j^g + 1, \ldots, p_g^*$, for each g. The discrete portion of the prior is implied from the discussion below.

5.3 Neutral-data comparisons and posterior probabilities

As part of the analysis setup, the models $\mathcal{M}_j$ are assumed to be highly connected in the sense that it is possible to move from any one model $\mathcal{M}_j$ to any other model $\mathcal{M}_k$ through a series of "simple steps," each made by adding or removing a subset of constraints from a single group. Consequently, once $NDC_{n,jk}$ is formulated for $\mathcal{M}_j$ and $\mathcal{M}_k$ related by a single such simple step, the formulation immediately extends to general comparisons through the formula (8), provided that the coherency criteria (10) hold.

Suppose now that $\mathcal{M}_j$ and $\mathcal{M}_k$ are such that $\mathcal{M}_j$ is just $\mathcal{M}_k$ with the addition of $u_j^h - u_k^h$ constraints to group $h \in A_j$. By the construction in Section 5.2, it may be assumed without loss of generality that $c_{ij}^{hl} = c_{ik}^{hl}$ for $l = u_k^h + 1, \ldots, u_j^h$. The second formula in (8) then yields $-2\log NDC_{n,jk} = Q_{n,jk} - \tilde{Q}_{n,jk}$, where, writing $\nu_{n,1T} = \sum_{i=1}^{p_n} \nu_{n,1i}$,

$$Q_{n,jk} \quad = \quad (\nu_{n,1T} + \nu_{n,2} + \kappa) \log \left\{ 1 + w_{n,h}^* \frac{S_{n,1}}{1 + S_{n,2}} \right\}, \tag{34}$$

for which $w^*_{n,h} = 1/\{1 + 1/(n\tau^{*2}_{n,h})\}$,

$$S_{n,1} = \frac{\nu^*_{n,1h}}{\nu_{n,2}} \sum_{l=u^h_k+1}^{u^h_j} F^{hl}_{n,j}, \quad \text{and} \quad S_{n,2} = \sum_{g=1}^{G_n} \frac{\nu^*_{n,1g}}{\nu_{n,2}} \left(\sum_{l=1}^{u^g_k} F^{gl}_{n,k} + \sum_{l=u^g_k+1}^{p^*_g} \frac{F^{gl}_{n,k}}{1 + n\tau^{*2}_g} \right),$$

with

$$F^{gl}_{n,j} \quad = \quad \frac{n \| \sum_{i \in B_g} c^{gl}_{ij} \boldsymbol{Y}_{n,i} \|^2 / \nu^*_{n,1g}}{\hat{\sigma}^2_n + \lambda/\nu_{n,2}},$$

and $\tilde{Q}_{n,jk}$ defined as in (34), but with each $F^{gl}_{n,j}$ replaced with a neutral version, $\tilde{F}^{gl}_{n,j}$. In this example, neutral-data are specified according to $\tilde{F}^{gl}_{n,j} = \log\nu_{n,2}$, using the analogue to default setting (18), having set $n = \nu_{n,2}$. The coherency criteria (10) are guaranteed by this setting since each $\tilde{F}^{gl}_{n,j}$ is independent of the model against which $\mathcal{M}_j$ might be compared.

Posterior probabilities $P[\mathcal{M}_j|\boldsymbol{\mathcal{Y}}_n]$, are calculated from the $NDC_{n,jk}$ and ratios $\tilde{\rho}_{n,jk}/\tilde{\rho}_{n,kj} = 1$, through (11), by exploiting the connectedness of the models in a Metropolis-Hastings algorithm. The reader should consult Robert and Casella (1999) for full technical details of this type of algorithm. The present version transitions through the models in simple steps, to form a Markov chain whose limiting distribution is identical to the posterior distribution of model probabilities. This yields probabilities associated with the discrete portion of the posterior distribution; but, at each transition, the algorithm additionally draws from the continuous portion of the posterior by direct simulation, conditionally on the current model, so that the final output of the algorithm represents a sample from the full posterior distribution. The results presented below were calculated from one million iterations of this algorithm, after a burn-in of 25,000 iterations.

5.4 Analysis of the simulated tumor data

Two versions of the analysis are considered. Analysis 1 tests each effect only for its presence or absence: effect-group g is either "totally constrained," with $u^g_j = p^*_g$, or "totally free," with $u^g_j = 0$. (The grand-mean group is always totally free.) Analysis 2 incorporates additional hypothesis-models that accommodate the detailed linear relationships discussed at the end of Section 5.1. Specifically, for each two-way interaction group, Analysis 2 incorporates the associated component-specific tests of H_{0i} *versus* H_{1i}, which describe detailed patterns of non-additivity; and, for each main-effect group,

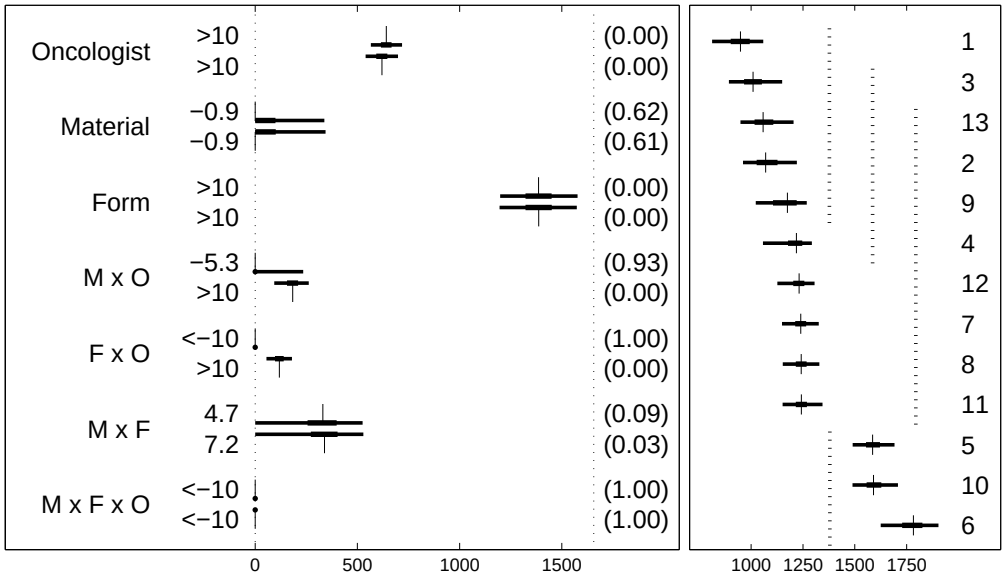

Figure 1: *ANOVA diagram for the simulated tumor data of Hoaglin, Mosteller, and Tukey (1991). The left panel displays posterior quantiles of $FPSD_g$ in pairs corresponding to Analysis 1 (top) and Analysis 2 (bottom). Thin lines connect 5% and 95% quantiles; thick lines connect 25% and 75% quantiles; and vertical bars identify the medians. The evidence for the presence of each effect is listed as $-2\log NDC_{n,jk}$ next to its effect-group label. The far right column lists pairs of posterior probabilities of no effect. The right panel displays posterior quantiles are displayed of the underlying means for oncologists, grouped by vertical dotted lines in such a way that there is no "strong" evidence of a mean difference between any pair in the group.*

it formulates a partition analysis, which is done by incorporating linear constraints on the mean parameters of factor-level pairs. To accommodate the partition analyses, corresponding transitions of the Metropolis-Hastings algorithm are made by either merging two partition cells (hence adding a constraint), or breaking a single partition cell in two (hence removing a constraint). The three-way interaction group is treated identically (presence or absence only) between the two analyses. The quantities $w^*_{n,h}$ in (34) are always set to one, which effectively takes the prior variance parameters to have been set arbitrarily large.

Analysis results are indicated in Figure 1, in a format suggested in Gelman (2005), where the posterior quantiles of FPSD are plotted for each group. These estimates

Partition	$P[\mathcal{M}_j\|\boldsymbol{\mathcal{Y}}_n]$	$-2\log NDC_{n,1j}$
(1,2,3,13)(4,7,8,9,11,12)(5,10)(6)	0.035	0.00
(1)(2,3,9,13)(4,7,8,11,12)(5,10)(6)	0.016	1.53
(1,3)(2,9,13)(4,7,8,11,12)(5,10)(6)	0.014	1.76
(1)(2,3,13)(4,7,8,9,11,12)(5,10)(6)	0.014	1.88
(1,3)(2,9,13)(4,7,8,11,12)(5)(6)(10)	0.013	2.01
(1,2,3,9,13)(4,7,8,11,12)(5)(6)(10)	0.012	2.15
(1,3)(2,13)(4,7,8,9,11,12)(5,10)(6)	0.011	2.36
(1,3,13)(2,4,7,8,9,11,12)(5,10)(6)	0.010	2.58
(1)(2,3,13)(4,7,8,11,12)(5,10)(6)(9)	0.009	2.67
(1)(2,3,13)(4,7,8,9,11,12)(5)(6)(10)	0.008	3.04

Table 2: *Posterior modal partitions of oncologist means with posterior probabilities and comparisons.*

are reported with an associated value of $-2\log NDC_{n,jk}$, in a configuration for testing the "absence" ($\mathcal{M}_j$) *versus* "presence" ($\mathcal{M}_k$) of the effect. The results are displayed in pairs, corresponding to Analyses 1 and 2. The strength of evidence indicated by each neutral-data comparison value is interpreted using standard categories described in Kass and Raftery (1995): values of $-2\log NDC_{n,jk}$ greater than 2 indicate "positive" evidence for $\mathcal{M}_k$; those greater than 6, indicate "strong" evidence; and those greater than 10, "very strong" evidence. The negations of these values indicate the strength of evidence for $\mathcal{M}_j$.

Several interesting observations can be made. First, differences between the two analyses are most apparent in results associated with the "M x O" and "F x O" interactions, where evidence points in different directions: Analysis 1 suggests the absence of the effects and Analysis 2 suggests their presence. This seems to reflect a greater flexibility of the Analysis 2 to explore the detailed structure of an effect. Surprisingly, almost no differences are seen between the analysis results associated with the main effects. Certain results from Analysis 2 are displayed in the right panel of Figure 1, which aim to describe the inferred structure of the "Oncologist" main effect. The results are presented in a manner resembling a classical mean-separation procedure (partly to illustrate its limitations): oncologists are arranged in the order of their estimated mean responses, calculated here as posterior medians, and are then grouped in such a way that there is no "strong" evidence for mean differences in any pair of oncologists in the same group. It is not unusual that such mean-separation groups will overlap, as

happens here, and that creates logical confusion when interpreting the results. An alternative, preferred report of results is shown in Table 2, which lists actual partitions of the oncologists. The first several posterior *modal* partitions–those with highest posterior probability–are displayed in order, and they are accompanied by evidence assessments of the top partition *versus* each partition below it. Though considerable ambiguity yet remains, this alternative report provides a clear, logically consistent interpretation of the analysis output.

6 Conclusions

This article has presented neutral-data comparisons as a new way of assessing evidence in a Bayesian framework, which is sensible and powerful, even when used with a vague prior. The statistic NDC_0 is interpreted as both an approximation to $\widetilde{BF}_0$ and an alternative to BF_0, which is motivated to accommodate a type of incoherency, revealed by the device of imaginary results, that leads to the Jeffreys-Lindley paradox. Guidelines for eliciting neutral data have been formulated, and an applied methodology for neutral-data comparisons has been developed for testing linear hypotheses in Gaussian linear-models analysis. Within this context, neutral-data comparisons are shown to possess strong asymptotic-consistency properties in high-dimensions, and they have been demonstrated to accommodate sophisticated testing problems, such as partition analysis, using straightforward computational algorithms.

It is worthwhile to remark on possible specifications for neutral data, and special issues that arise, in alternative applied contexts than are explored here. In some applications, the choice of neutral data may be obvious; for instance, when testing one-sided hypotheses of a normal mean, $H_0 : \theta \leq 0$ *versus* $H_1 : \theta > 0$, based on data $Y|\theta \sim N(\theta, \sigma^2)$, under a prior that is symmetric about zero, the obvious setting for neutral data is $\tilde{Y} = 0$. In the analogous discrete-data setting, however, the obvious choice may require additional care. To see this, consider a simple treatment-*versus*-control model in which Y_1 and Y_2 are incidence counts from n_1 and n_2 independent trials. Suppose the problem is to test $H_0 : \theta_1 \leq \theta_2$ *versus* $H_1 : \theta_1 > \theta_2$, where θ_1 and θ_2 denote the respective incidence probabilities. A plausible choice for neutral data would set $\tilde{Y}_1 = n_1\hat{\theta}$ and $\tilde{Y}_2 = n_2\hat{\theta}$, where $\hat{\theta} = (Y_1 + Y_2)/(n_1 + n_2)$. Yet, notice this presents another example, in addition to that of Section 3.3, of neutral data being set partly as a function of the observed data. Another conceptual issue arises as well, which is that this setting allows non-integer values, hence neutral data that falls outside of the

the data space. Accommodation of these complexities is not far-fetched, and represents an important task for future investigations: it is conjectured that the data-dependency issue may be understood through some formal concept for "updating" neutral data from what is observed, and that the discrete-data issue may be resolved through a continuous generalization of the model.

Also of interest for future investigation is the development of an exact theory that broadens, refines, and explains the many asymptotic results presented here. One direction of immediate interest would examine the prior structure for ANOVA used in Gelman (2005), where τ_g^* is not fixed, necessarily, but is equipped with a hyperprior whenever the number of free-parameters exceeds two. The usual goal of such hierarchical constructions is to reduce expected quadratic loss, and this begs the question of what impact the addition of a complicated testing component would have on a procedure's decision-theoretic properties. The development of associated computational algorithms is also important, and that task is expected to require special care as scenarios become more complex. For instance, one would want to avoid running an intensive computational algorithm multiple times on observed and neutral data. The ideas discussed here lay a sturdy groundwork for carrying out all of these investigations, and for continued development of neutral-data comparisons as a useful applied tool with a meaningful interpretation.

Appendix

Proof. (Theorem 1) A Laplace approximation to the Bayes factor at $\boldsymbol{y}_n$ (which might be $\boldsymbol{Y}_n$ or $\tilde{\boldsymbol{Y}}_n$) is $\widetilde{BF}_{n,0}(\boldsymbol{y}_n) \approx (2\pi)^{-\nu_1/2}|S_n(\boldsymbol{t}_n^*)|^{1/2}\exp\{s_n(\boldsymbol{t}_n^*)\}$ for $S_n(\boldsymbol{t}) = (n/\sigma^2)\boldsymbol{I} - \boldsymbol{S}_{\tilde{\pi}}(\boldsymbol{t})$, writing $\boldsymbol{S}_{\tilde{\pi}}(\boldsymbol{t})$ to denote the Hessian matrix of $\log\tilde{\pi}(\boldsymbol{t}|H_1,\sigma)$, and $\boldsymbol{t}_n^*$ such that $\nabla s_n(\boldsymbol{t}_n^*) = \boldsymbol{0}$. Note that $\|\nabla s_n(\boldsymbol{t})\| = (n/\sigma^2)\|(\boldsymbol{t}-\boldsymbol{y}_n) - (\sigma^2/n)\nabla\log\tilde{\pi}(\boldsymbol{t}|H_1,\sigma)\|$, and so if $\|\nabla\log\tilde{\pi}(\boldsymbol{y}_n|H_1,\sigma)\| = O(1)$, it must be that $|\boldsymbol{t}_n^* - \boldsymbol{y}_n| \to 0$. Subsequently, $|S_n(\boldsymbol{t}_n^*)| = (n/\sigma^2)^{\nu_1}\{1+o(1)\}$, and $\widetilde{BF}_{n,0}(\boldsymbol{y}_n) \approx \exp\{-(1/2)\|\boldsymbol{z}_n\|^2 + (\nu_1/2)\log(n/\sigma^2) - \log\tilde{\pi}(\boldsymbol{t}_n^*|H_1,\sigma) - (\nu_1/2)\log(2\pi)\}$, writing $\boldsymbol{z}_n = \sqrt{n}\boldsymbol{y}_n/\sigma$. The statements are readily checked upon noting that $\boldsymbol{Y}_n \to \boldsymbol{\theta}$, and, by (13), $\|\boldsymbol{Z}_n\|^2 = n\|\boldsymbol{Y}_n\|^2/\sigma^2 \approx n\|\boldsymbol{\theta}\|^2/\sigma^2$ under H_1. □

Proof. (Theorem 2) Note the following properties of $f(x)$ and $f_0(y) = f(e^y)$, which are implied from $\gamma(x) \to \gamma \geq 0$: (A.) $\frac{d}{dy}\log f_0(y) \to \gamma$ and $\frac{d}{dy}\log f_0'(y) \to \gamma$; (B.) if $y_{1,n}, y_{2,n} \to \infty$ as $n \to \infty$, then $\{\log f_0(y_{2,n}) - \log f_0(y_{1,n})\}/\{y_{2,n} - y_{1,n}\} \to \gamma$; (C.) $xf''(x)/f'(x) \to \gamma - 1$; (D.) $\{\log f(x)\}/\log x \to \gamma$; and (E.) $(\log x)/f(x) = O(1)$. Prop-

erty A is deduced by writing $\gamma(x)$ as $\frac{d}{dy}\log f_0(y) = f_0'(y)/f_0(y)$ evaluated at $y = \log x$, and then applying l'Hôpital's rule to $\frac{d}{dy}\log f_0'(y) = f_0''(y)/f_0'(y)$ (unless $f_0''(y)$ is bounded, in which case the limit $\gamma = 0$ is deduced directly). Property B is a consequence of the mean-value theorem applied to A. Property C follows from A by expanding $\frac{d}{dy}\log f_0'(y) = 1 + e^y f''(e^y)/f'(e^y)$. Property D follows from A by applying l'Hôpital's rule to $\{\log f_0(y)\}/y = \{\log f(x)\}/\log x$, for $y = \log x$. Property E follows from the convexity of $f_0(y)$; this requires $1/\{xf'(x)\} = O(1)$, hence E follows by applying l'Hôpital's rule to $(\log x)/f(x) \approx \{1/x\}/f'(x) = O(1)$, starting at the right or left side according to whether $f'(x) \to 0$.

To prove statement (i), define $\tilde{s}_n(\boldsymbol{t}) = n\{\|\boldsymbol{t}\|^2 - 2\boldsymbol{t}^T\tilde{\boldsymbol{Y}}_n\}/(2\sigma^2) + f(\sigma^2/\|\boldsymbol{t}\|^2)/2$, and consider the system of equations in $\tilde{\boldsymbol{t}}_n$ and $\tilde{\boldsymbol{Z}}_n$ given by

$$\nabla\tilde{s}_n(\tilde{\boldsymbol{t}}_n) = \boldsymbol{0} \quad \text{and} \quad (2\pi/c)^{-\nu_1/2}|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)|^{1/2}\exp\{\tilde{s}_n(\tilde{\boldsymbol{t}}_n)\} = 1,$$

where $\nabla\tilde{s}_n(\boldsymbol{t}) = n\{\boldsymbol{t} - \tilde{\boldsymbol{Y}}_n\}/\sigma^2 - \sigma^2(\boldsymbol{t}/\|\boldsymbol{t}\|^4)f'(\sigma^2/\|\boldsymbol{t}\|^2)$ and $\tilde{S}_n(\boldsymbol{t})$ is the Hessian matrix of $\tilde{s}_n(\boldsymbol{t})$. The second equation is to provide a Laplace approximation to $\widetilde{BF}_{n,0}(\tilde{\boldsymbol{Y}}_n)$, and requires a solution for which $\tilde{\boldsymbol{t}}_n \to 0$. Writing $\tilde{\boldsymbol{t}}_n = \tilde{a}_n\tilde{\boldsymbol{Y}}_n$, the equations become a system in $\tilde{a}_n > 0$ and $\|\tilde{\boldsymbol{Z}}_n\|^2$ that is equivalent to

$$\begin{aligned} \|\tilde{\boldsymbol{Z}}_n\|^2 &= \frac{(\tilde{\gamma}_n+1)^2}{2\tilde{\gamma}_n+1}\left[f\left(\frac{n}{\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2}\right) + \log|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)| - \nu_1\log(2\pi/c)\right] \\ \tilde{\gamma}_n &= \gamma\left(\frac{n}{\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2}\right)\Bigg/\left[1 + \frac{\log|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)| - \nu_1\log(2\pi/c)}{f(n/\{\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2\})}\right], \end{aligned} \tag{35}$$

where $\tilde{a}_n = 1 + \tilde{\gamma}_n/(\tilde{\gamma}_n + 1)$. In translating to (35), an intermediate expression is

$$\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2(1 - 1/\tilde{a}_n) = g_n(\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2), \tag{36}$$

where $g_n(x) = (n/x)f'(n/x)$; since $f(x)$ increases, hence $xf'(x) > 0$, this requires $\tilde{a}_n > 1$ (although $\tilde{a}_n \to 1$ is possible). The requirement $\tilde{\boldsymbol{t}}_n \to 0$ translates to $n/(\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2) \to \infty$. To evaluate $|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)|$, substitute (36) and apply $xf''(x)/f'(x) \to \gamma - 1$ (by Property C, above) to see that $\tilde{S}_n(\tilde{\boldsymbol{t}}_n) \approx \{n/(\sigma^2\tilde{a}_n)\}\{\boldsymbol{I} + 2(\tilde{a}_n - 1)(1 + 2\gamma)\boldsymbol{Z}_n\boldsymbol{Z}_n^T/\|\boldsymbol{Z}_n\|^2\}$. Sylvester's determinant theorem then provides $|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)| = \{n/(\sigma^2\tilde{a}_n)\}^{\nu_1}\{1+2(\tilde{a}_n-1)(1+2\gamma)\}$. Subsequently, Properties D and E imply that $\{\log|\tilde{S}_n(\tilde{\boldsymbol{t}}_n)|\}/f(n/\{\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2\})$ has a limit that is never negative, and is only positive when $\gamma > 0$; therefore, $\tilde{\gamma}_n \to \gamma$. Property B, with $y_{1,n} = \log\{n/(\tilde{a}^2\|\tilde{\boldsymbol{Z}}_n\|^2)\}$ and $y_{2,n} = \log\{n/(\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2)\}$, then provides that $f(n/\{\tilde{a}_n^2\|\tilde{\boldsymbol{Z}}_n\|^2\}) \approx f(n/\{\tilde{a}^2\|\tilde{\boldsymbol{Z}}_n\|^2\})$. The first equation in (35) is thus asymptotically equivalent to (20). Property E, above, and the monotonicity of $f(x)$ imply that $\|\tilde{\boldsymbol{Z}}_n\|^2$ must be unique, to the accuracy of approximation.

To prove statement (*ii*), first note that any positive solution to (21) has $a_n = 1/2 + \{g_n(a_n^2\|\boldsymbol{Z}_n\|^2) \,/\, \|\boldsymbol{Z}_n\|^2 + 1/4\}^{1/2}$. Since $f(x)$ increases, $xf'(x) > 0$, and the convexity of $f_0(x)$ implies that $g_n(x) = f_0'(\log(n/x))$ does not increase in x; hence each a_n must be unique. Notice also that, since $\|\boldsymbol{Z}_n\|^2 \sim \chi^2_{\nu_1}$ under H_0, $n/a_n^2 \to \infty$, $1/a_n = O(1)$, and $a_n \to \infty$ when $\gamma > 0$. Next define $\hat{s}_n(\boldsymbol{t}) = n\{\|\boldsymbol{t}\|^2 - 2\boldsymbol{t}^T\boldsymbol{Y}_n\}/(2\sigma^2) + f(\sigma^2/\|\boldsymbol{t}\|^2)/2$, and note that the $\boldsymbol{t}_n^*$ such that $\nabla\hat{s}_n(\boldsymbol{t}_n^*) = 0$ is $\boldsymbol{t}_n^* = a_n\boldsymbol{Y}_n$. The property $a_n^2/n \to 0$ implies $\|\boldsymbol{t}_n^*\| \to 0$, which admits the Laplace approximation $\widetilde{BF}_0 \approx (2\pi/c)^{-\nu_1/2}|\hat{S}_n(\boldsymbol{t}_n^*)|^{1/2}\exp\{\hat{s}_n(\boldsymbol{t}_n^*)\}$, where $\hat{S}_n(\boldsymbol{t})$ is the Hessian matrix of $\hat{s}_n(\boldsymbol{t})$. We have

$$\hat{s}_n(\boldsymbol{t}_n^*) \;=\; (1/2)a_n^2\|\boldsymbol{Z}_n\|^2(1 - 2/a_n) + (1/2)f(n/\{a_n^2\|\boldsymbol{Z}_n\|^2\}), \tag{37}$$

and $|\hat{S}_n(\boldsymbol{t}_n^*)| \approx \{n/(\sigma^2 a_n)\}^{\nu_1}\{1 + 2(a_n - 1)(1 + 2\gamma)\}$ (*cf.* the approximation to $\tilde{S}_n(\tilde{\boldsymbol{t}}_n)$ in the proof of statement *i*). To complete the proof, substitute $a_n^2\|\boldsymbol{Z}_n\|^2(1 - 1/a_n) = g_n(a_n^2\|\boldsymbol{Z}_n\|^2) = \gamma(n/\{a_n^2\|\boldsymbol{Z}_n\|^2\})f(n/\{a_n^2\|\boldsymbol{Z}_n\|^2\})$ in (37), and $\log|\hat{S}_n(\tilde{\boldsymbol{t}}_n)| \approx \nu_1 \log n$, and recall that $a_n \to \infty$ when $\gamma > 0$. □

Proof. (Corollary 1) Each statement is deduced by direct substitution into formulas (20, 21, 22). □

Lemma 1. *Assume the high-dimension model (25) and prior (27), specified with* $\kappa_i, \lambda_i = 0$. *Suppose the component posterior probabilities in (30) are defined with* $NDC_{n,0i}$ *as in (30), for corrected prior null probabilities* $\tilde{\rho}_{n,i}$. *Suppose further that each* $\nu_{n,1i} = o(n)$ *and there are fixed constants* $c_i > 0$ *for which each* $\nu_{n,2i}/n > c_i$ *for all* n. *Then* $P[\mathcal{M}_n^*|\boldsymbol{\mathcal{Y}}_n] \to 1$ *whenever both*

$$B_{n,0} \;=\; \sum_{i \in A_n^*} \left(\frac{1 - \tilde{\rho}_{n,i}}{\tilde{\rho}_{n,i}}\right)\left(\frac{1}{1 - w_{n,i}}\right)^{\nu_{n,1i}/2} \exp\left\{-\frac{1}{2}Q_{n,i}(\tilde{F}_{n,i})\right\} \tag{38}$$

and

$$B_{n,1}(\boldsymbol{\theta}) \;=\; \sum_{i \notin A_n^*} \left(\frac{\tilde{\rho}_{n,i}}{1 - \tilde{\rho}_{n,i}}\right) \exp\left\{\frac{1}{2}Q_{n,i}(\tilde{F}_{n,i}) - nc_i U_{w_{n,i}}\left(\frac{\|\boldsymbol{\theta}_i\|^2/\sigma_i^2}{2c_i}\right)\right\} \tag{39}$$

converge to zero, where $U_w(\xi)$ *is defined for* $\xi > 0$ *according to*

$$U_w(\xi) \;=\; \xi\log\left\{\frac{t_0(\xi)}{t_w(\xi)}\right\} + \log\left[\frac{1 - t_0(\xi)}{\{1 - t_w(\xi)\}\{1 - wt_w(\xi)\}}\right] \tag{40}$$

with

$$t_w(\xi) \;=\; \left(\frac{1 + w}{2w}\right)\left(\frac{1 + \xi}{2 + \xi}\right)\left\{1 - \sqrt{1 - \frac{4w}{(1 + w)^2}\left(\frac{\xi}{1 + \xi}\right)\left(\frac{2 + \xi}{1 + \xi}\right)}\right\} \tag{41}$$

and $t_0(\xi) = \lim_{w\to 0} t_w(\xi) = \xi/(1+\xi)$. *Furthermore,* $U_w(\xi)$ *is positive, strictly increasing in* ξ, $U_w(\xi) = \xi\log(1+w) + o(\xi)$ *as* $\xi \to 0$, *and* $U_w(\xi) \to -\log(1-w)$ *as* $\xi \to \infty$.

Proof. (LEMMA 1) Substitute $P[H_{0i}|\boldsymbol{Y}_{n,i}, \hat{\sigma}_{n,i}] = NDC_{n,0i}(\tilde{\rho}_{n,i}^{-1}-1)/\{1+NDC_{n,0i}(\tilde{\rho}_{n,i}^{-1}-1)\}$, then apply $\log(1+x) \le x$ to establish that $-\log P[\mathcal{M}_n^*|\boldsymbol{\mathcal{Y}}_n] \le B_n$, where

$$\begin{aligned} B_n &= \sum_{i\in A_n^*}\left(\frac{1-\tilde{\rho}_{n,i}}{\tilde{\rho}_{n,i}}\right)\exp\left[\frac{1}{2}\{Q_{n,i}(F_{n,i}) - Q_{n,i}(\tilde{F}_{n,i})\}\right] \\ &\quad + \sum_{i\notin A_n^*}\left(\frac{\tilde{\rho}_{n,i}}{1-\tilde{\rho}_{n,i}}\right)\exp\left[-\frac{1}{2}\{Q_{n,i}(F_{n,i}) - Q_{n,i}(\tilde{F}_{n,i})\}\right]. \end{aligned}$$

Since the terms of this sum are nonnegative and independent, an extension of the Borel-Cantelli lemmas (*cf.* Billingsley, 1995, prob. 22.3, p. 294) provides that B_n will converge almost surely whenever $E[B_n]$ converges. It will be shown that $E[B_n] \le \{B_{n,0} + B_{n,1}(\boldsymbol{\theta})\}\{1+o(1)\}$.

Having set each $\lambda_i = 0$, $F_{n,i}$ follows a non-central $F_{\nu_{n,1i},\nu_{n,2i}}(\delta_{n,i})$ distribution with non-centrality parameter $\delta_{n,i} = n\|\boldsymbol{\theta}_i\|^2/\sigma_i^2$. Johnson et al. (1994, vol. 2, p. 484) provide that the density of $G_{n,i} = (\nu_{n,1i}/\nu_{n,2i})F_{n,i}$ is

$$\pi_{n,i}(g) = \sum_{k=0}^{\infty}\left\{\frac{(\delta_{n,i}/2)^k e^{-\delta_{n,i}/2}}{k!}\right\}\frac{1}{B(\nu_{n,1i}/2+k, \nu_{n,2i}/2)}\frac{g^{\nu_{n,1i}/2+k-1}}{(1+g)^{(\nu_{n,1i}+\nu_{n,2i})/2+k}},$$

where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ is the beta function. Note that $Q_{n,i}(F_{n,i}) = -(\nu_{n,1i}+\nu_{n,2i})\log\{1 - w_{n,i}G_{n,i}/(1+G_{n,i})\}$, having set $\kappa_i = 0$. If $\delta_{n,i} = 0$, integrate with respect to the change of variable $t = (1-w_{n,i})g$ to see that $E[e^{\frac{1}{2}Q_{n,i}(F_{n,i})}] = (1-w_{n,i})^{-\nu_{n,1i}/2}$. This shows that $B_{n,0}$ is the expected value of the first term in the definition of B_n. If $\delta_{n,i} > 0$, integrate with respect to the change of variable $t = g/(1+g)$ to see that $E\left[e^{-\frac{1}{2}Q_{n,i}(F_{n,i})}\right]$ is

$$\sum_{k=0}^{\infty}\left\{\frac{(\delta_{n,i}/2)^k e^{-\delta_{n,i}/2}}{k!}\right\}{}_2F_1\left(-\frac{\nu_{n,1i}+\nu_{n,2i}}{2}, \frac{\nu_{n,1i}}{2}+k, \frac{\nu_{n,1i}+\nu_{n,2i}}{2}+k, w_{n,i}\right), \quad (42)$$

where ${}_2F_1(-\gamma,\alpha,\alpha+\beta,w) = \int_0^1(1-wt)^{\gamma}t^{\alpha-1}(1-t)^{\beta-1}dt/B(\alpha,\beta)$ is an evaluation of the hypergeometric function. A well known representation of this function has

$${}_2F_1(-\gamma,\alpha,\alpha+\beta,w) = \sum_{k=0}^{\infty}\frac{(-\gamma)_k(\alpha)_k}{(\alpha+\beta)_k}\frac{w^k}{k!}, \quad (43)$$

writing $(x)_0 = 1$ and $(x)_k = \prod_{j=1}^k(x+j-1)$, which, upon substituting $\gamma = (\nu_{n,1i}+\nu_{n,2i})/2$ and $\beta = \nu_{n,2i}/2$, shows that $E\left[e^{-Q_{n,i}(F_{n,i})/2}\right]$ becomes smaller with any increase of $\nu_{n,2i}$. It therefore suffices to check only the case where each $\nu_{n,2i} = nc_i$.

Set $\alpha = \delta_{n,i}/(2n) = \|\boldsymbol{\theta}_{n,i}\|^2/(2\sigma_{n,i}^2)$, $\beta = \nu_{n,2i}/n = c_i$, and $\xi = \alpha/\beta$; then note the Laplace expansion

$$\int_0^1 t^{\alpha n\{1+o(1)\}}\{(1-t)(1-wt)\}^{\beta n\{1+o(1)\}}dt \approx \frac{\exp h_w(t_w(\xi))}{\sqrt{|h_w''(t_w(\xi))|/(2\pi)}}, \tag{44}$$

where $h_w(t) = n[\alpha\log t + \beta\log\{(1-t)(1-wt)\}]$, for which $t_w(\xi)$, defined in (41), is its maximizing argument. Reflecting that the series (42) is weighted by Poisson probabilities, its relevant terms are those for which k is within a slowly increasing multiple of $\sqrt{\delta_{n,i}/2}$ from $\delta_{n,i}/2$. Hence, the expansion above provides that

$$E\left[e^{-\frac{1}{2}Q_{n,i}(F_{n,i})}\right] \approx \sqrt{\frac{|h_0''(t_0(\xi))|}{|h_{w_{n,i}}''(t_{w_{n,i}}(\xi))|}}\exp\{-\beta n U_{w_{n,i}}(\xi)\},$$

for $U_w(\xi)$ defined in (40). A straightforward calculus exercise yields $|h_0''(t_0(\xi))| < |h_w''(t_w(\xi))|$. It follows that $B_{n,1}(\boldsymbol{\theta})$ bounds the expected value of the second term in the definition of B_n. Another calculus exercise will establish the stated properties of $U_w(\xi)$. □

Proof. (THEOREM 3) In the context of Lemma 1, write $D_{n,1}(\boldsymbol{\theta}) = \min_{i\notin A_n^*}\left\{\|\boldsymbol{\theta}_i\|^2/\sigma_i^2\right\}$ and note that the property $U_w(\xi) = \xi\log(1+w) + o(\xi)$ as $\xi \to 0$ implies the existence of a constant $c > 0$ such that

$$2n\min_{i\notin A_n^*}\left\{c_i U_{w_{n,i}}\left(\{\|\boldsymbol{\theta}_i\|^2/\sigma_i^2\}/\{2c_i\}\right)\right\} \geq cnD_{n,1}(\boldsymbol{\theta}),$$

for sufficiently large n. Noting also that each $Q_{n,i}(\tilde{F}_{n,i}) \approx w_{n,i}\nu_{n,1i}\tilde{F}_{n,i}$, the setup of this scenario therefore establishes that $\log B_{n,0}$ is bounded above by

$$-\frac{1}{2}w_{n*}r_n\sqrt{\nu_{n,1*}}\left\{1 - \frac{2\log|A_n^*|}{w_{n*}r_n\sqrt{\nu_{n,1*}}} + \frac{\sqrt{\nu_{n,1*}}\{1 + w_{n*}^{-1}\log(1-w_n^*)\}}{r_n}\right\}\{1+o(1)\}$$

and $\log B_{n,1}(\boldsymbol{\theta})$ is bounded above by

$$-\frac{c}{2}nD_{n,1}(\boldsymbol{\theta})\left\{1 - \frac{2\log(p_n - |A_n^*|)}{cnD_{n,1}(\boldsymbol{\theta})} - \frac{\nu_{n,1}^* + r_n\sqrt{\nu_{n,1}^*}}{cnD_{n,1}(\boldsymbol{\theta})}\right\}\{1+o(1)\},$$

for sufficiently large n, which diverge to $-\infty$ under the stated conditions. □

References

Aitkin, M. (1991). "Posterior Bayes factors." *Journal of the Royal Statistical Society - Series B*, 53: 111–142.

Bellman, R. (1960). *Introduction to matrix analysis*. New York: McGraw-Hill.

Berger, J. O., Bernardo, J. M., and Sun, D. (2009). "The formal definition of reference priors." *Annals of Statistics*, 37: 905–938.

Berger, J. O., Ghosh, J. K., and Mukhopadhyay, N. (2003). "Approximations and consistency of Bayes factors as model dimension grows." *Journal of Statistical Planning and Inference*, 112: 241–258.

Berger, J. O. and Pericchi, L. (1996). "The intrinsic Bayes factor for model selection and prediction." *Journal of the American Statistical Association*, 91: 109–122.

Billingsley, P. (1995). *Probability and measure*. New York: Wiley, 3rd edition.

Casella, G., Girón, F. J., Martinez, M. L., and Moreno, E. (2009). "Consistency of Bayesian procedures for variable selection." *Annals of Statistics*, 37: 1207–1228.

Crowley, E. M. (1997). "Product partition models for normal means." *Journal of the American Statistical Association*, 92: 192–198.

Diaconis, P. and Freedman, D. (1986). "On the consistency of Bayes estimates." *Annals of Statistics*, 14: 1–26.

Fan, J. and Lv, J. (2008). "Sure independence screening for ultrahigh dimensional feature space." *Journal of the Royal Statistical Society - Series B*, 70: 849–911.

— (2010). "A selective overview of variable selection in high dimensional feature space." *Statistica Sinica*, 20: 101–148.

García-Donato, G. and Sun, D. (2007). "Objective priors for model selection in one-way random effects models." *The Canadian Journal of Statistics*, 35: 303–320.

Gelman, A. (2005). "Analysis of variance–why it is more important than ever (with discussion)." *Annals of Statistics*, 33: 1–53.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. London: Griffin.

Guo, R. and Speckman, P. (2009). "Bayes factor consistency in linear models." In *The 2009 International Workshop on Objective Bayes Methodology in Philadelphia, PA, June 5-9, 2009*. http://stat.wharton.upenn.edu/statweb/Conference/OBayes09/AbstractPapers/speckman.pdf.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1991). *Fundamentals of Exploratory Analysis of Variance*. Wiley: New York.

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press, 3rd edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Wiley: New York, 2nd edition.

Johnson, V. E. and Rossell, D. (2010). "On the use of non-local prior densities for default Bayesian hypothesis tests." *Journal of the Royal Statistical Society - Series B*, 72: 143–170.

Kass, R. E. and Raftery, A. E. (1995). "Bayes factors." *Journal of the American Statistical Association*, 90: 773–795.

Lavine, M. and Schervish, M. J. (1999). "Bayes factors: what they are and what they are not." *The American Statistician*, 53: 119–112.

Liang, F., Paulo, R., Molina, G., Clyde, C. A., and Berger, J. O. (2008). "Mixtures of g priors for Bayesian variable selection." *Journal of the American Statistical Association*, 103: 410–423.

Lindley, D. V. (1957). "A statistical paradox." *Biometrika*, 44: 187–192.

Maruyama, Y. and George, E. I. (2010). "gBF: A Fully Bayes Factor with a Generalized g-prior." *arXiv:0801.4410v2*.

O'Hagan, A. (1995). "Fractional Bayes factors for model comparisons." *Journal of the Royal Statistical Society - Series B*, 57: 99–138.

Pérez, J. M. and Berger, J. O. (2002). "Expected-posterior prior distributions for model selection." *Biometrika*, 89: 491–511.

Robert, C. P. (1993). "A note on Jeffreys-Lindley paradox." *Statistica Sinica*, 3: 603–608.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer: New York.

Schwarz, G. (1978). "Estimating the dimension of a model." *Annals of Statistics*, 6: 461–464.

Scott, J. G. and Berger, J. O. (2006). "An exploration of aspects of Bayesian multiple testing." *Journal of Statistical Planning and Inference*, 136: 2144–2162.

Smith, A. F. M. and Spiegelhalter, D. J. (1980). "Bayes factors and choice criteria for linear models." *Journal of the Royal Statistical Society - Series B*, 42: 213–220.

Spiegelhalter, D. J. and Smith, A. F. M. (1982). "Bayes factors for linear and log-linear models with vague prior information." *Journal of the Royal Statistical Society - Series B*, 44: 377–387.

Spitzner, D. J. (2008). "An asymptotic viewpoint on high-dimensional Bayesian testing." *Bayesian Analysis*, 3: 121–160.

Acknowledgments

The author is grateful to Susie Bayarri, Merlise Clyde, Tao Huang, the Associate Editor, and anonymous reviewers, whose comments were invaluable in developing the ideas presented here.