Self-consistent confidence sets and tests of composite hypotheses applicable to restricted parameters

March 8, 2017

David R. Bickel

Ottawa Institute of Systems Biology

Department of Biochemistry, Microbiology, and Immunology

Department of Mathematics and Statistics

University of Ottawa; 451 Smyth Road; Ottawa, Ontario, K1H 8M5

Alexandre G. Patriota

Department of Mathematics and Statistics

University of São Paulo; Rua do Matão, 1010; São Paulo, Brazil, 05508-090

Abstract

Frequentist methods, without the coherence guarantees of fully Bayesian methods, are known to yield self-contradictory inferences in certain settings. The framework introduced in this paper provides a simple adjustment to p values and confidence sets to ensure the mutual consistency of all inferences without sacrificing frequentist validity. Based on a definition of the compatibility of a composite hypothesis with the observed data given any parameter restriction and on the requirement of self-consistency, the adjustment leads to the possibility and necessity measures of possibility theory rather than to the posterior probability distributions of Bayesian and fiducial inference.

1 Introduction

A common criticism of frequentist statistical methods is that they lead to contradictory conclusions in settings where Bayesian methods cannot. Following Kaplan (1996), a method of hypothesis testing or set estimation will be called *deductively cogent* if it cannot make mutually contradictory rejections of hypotheses. Minimal requirements for a deductively cogent method of hypothesis testing are the following:

1. It is *restriction-respecting* in the sense that it cannot reject every hypothesis that is consistent with the restriction imposed and in that it rejects all hypotheses that are inconsistent with the restriction.

2. It is *coherent* in the sense that a hypothesis can only be rejected if every hypothesis implying it is also rejected (Gabriel, 1969).

Standard confidence procedures often fail to meet the first requirement in the presence of parameter restrictions, which are often encountered in physics. For example, if the parameter restriction is a bound on the parameter of interest, then inferences should proceed conditional on that bound. However, confidence intervals can be partially or entirely outside the bound (Mandelkern, 2002a; Fraser, 2011); cf. Zhang and Woodroofe (2003); Marchand and Strawderman (2004); Wang (2007); Marchand and Strawderman (2013). Taking the intersection of the parameter restriction set and the confidence set leads in the former case to truncating the confidence set at the bound, and in the latter case to an empty confidence set. Since parameter values outside a confidence set are considered rejected, an empty confidence set is equivalent to rejecting the entire set of possible parameter values, contradicting the condition that the parameter value lies in that set.

Empty confidence sets also occur for an epidemiological model, a branching process, and Brownian motion (Ball et al., 2002). While an empty confidence set is often interpreted as an indication of model inadequacy, procedures leading to them also lead to very small confidence sets, misleadingly indicating accurate knowledge of the parameter value (Ball et al., 2002). As a result, such confidence sets do not give the estimates of uncertainty that are needed in practice (Mandelkern, 2002a; Wang, 2006).

For an example of violating coherence, one-sided p values are interpreted as attained confidence levels of composite hypotheses, including those concerning the value of an unbounded parameter. Since such attained confidence levels can be smaller for a region than for a subset of that region (Efron and Tibshirani, 1998; Polansky, 2007, pp. 224-227), they do not correspond to coherent hypothesis tests. The fact that frequentist approaches can violate coherence has led many to develop methods complying with the strong likelihood principle, whether using prior distributions (e.g., Schervish (1996); Lavine and Schervish (1999)) or not (e.g., Royall (1997); Bickel (2012); Zhang and Zhang (2013)).

To render existing frequentist methods deductively cogent, this paper instead presents an alternative framework of hypothesis testing and confidence sets. The framework is based on the concept of the compatibility between a hypothesis and the observed data rather than on any likelihood principle.

That data-compatibility measure is specified and illustrated in Section 2 using the most important concepts found in the more theoretical parts of the paper. Additional examples are provided in Section 3, some of which feature bounded parameter problems. The foundational motivation is stated in terms of the axioms of Section 4. Section 5 derives properties of the data compatibility of a hypothesis, including the fact that the data compatibility of a point null hypothesis is the p value divided by the highest p value corresponding to the point null

hypotheses in the parameter space or in the parameter restriction, if any. As a result, the corresponding set estimate is a conservative confidence set. Section 6 introduces the concept of the acceptability of a hypothesis in order to indicate when to accept the hypothesis, when to reject it, and when to take neither of those actions. The restriction-respecting and coherence aspects of that procedure are also proven in the latter section. Finally, Section 7 remarks on the place of the proposed framework in possibility theory and ranking theory.

2 Methodology of data-hypothesis compatibility

2.1 Hypothesis testing

Let θ denote the parameter of interest restricted to a subset $\mathcal{R}$ of the parameter space Θ, x the observed sample of data, $H_0 : \theta = \theta_0$ the hypothesis that the value of θ is θ_0, and $H_0 : \theta \in \mathcal{H}_0$ the hypothesis that the value of θ is in some $\mathcal{H}_0 \subseteq \Theta$. The observed p value corresponding to $H_0 : \theta = \theta_0$ is $p(\theta_0; x)$. Here, $x \mapsto p(\theta_0; x)$ is a function such that the probability law of $p(\theta_0; X)$ weakly converges to $\mathrm{U}(0, 1)$ as the sample size increases given that X is distributed in agreement with $H_0 : \theta = \theta_0$, i.e., $P_{\theta_0, \gamma}(p(\theta_0; X) \leq \alpha) \to \alpha$ as the sample size tends to infinity for all $\alpha \in [0, 1]$ and $\gamma \in \Gamma$, where γ is the nuisance parameter, Γ is the nuisance parameter space and $P_{\theta_0, \gamma}$ is the probability measure of the data X. For an extensive discussion on p values, we refer the reader to Cox (1977).

The compatibility of $H_0 : \theta = \theta_0$ with x given that $\theta \in \mathcal{R}$ is the *c value*

$$c(\theta_0; x | \mathcal{R}) = \begin{cases} 0 & \text{if } \theta_0 \notin \mathcal{R} \\ \frac{p(\theta_0; x)}{\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x)} & \text{if } \theta_0 \in \mathcal{R}. \end{cases} \tag{1}$$

More generally, the compatibility of $H_0 : \theta \in \mathcal{H}_0$ with x given $\theta \in \mathcal{R}$ is the C *value*

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \sup_{\theta_0 \in \mathcal{H}_0} c\left(\theta_0; x | \mathcal{R}\right).$$

It is easy to verify that the compatibility of a hypothesis with the data is 0 whenever they are logically inconsistent, close to 0 whenever all observed p values corresponding to the hypothesis are low, and 1, the highest possible value, for at least one hypothesis that is logically consistent with the parameter restriction.

The absence of a parameter restriction is represented by $\mathcal{R} = \Theta$. Since the degenerate restriction that $\theta \in \Theta$ is necessarily true according to the model, the marginal compatibilities $C\left(\mathcal{H}_0; x | \Theta\right)$ and $c\left(\theta_0; x | \Theta\right)$ are marginal degrees to which their hypotheses are compatible with x. They are abbreviated by $C\left(\mathcal{H}_0; x\right)$ and $c\left(\theta_0; x\right)$, respectively.

The first example compares a simple null hypothesis to a simple alternative hypothesis (cf. Berger, 2003; Wang, 2004) to demonstrate the use of the proposed framework as simply as possible.

Example 1. Comparison of two simple hypotheses, $X \sim \mathrm{N}\left(0, 1\right)$ and $X \sim \mathrm{N}\left(1, 1\right)$, on the basis of a single observation x. In this example, $\mathcal{R} = \{0, 1\}$, Θ is any set of real numbers such that $\mathcal{R} \subseteq \Theta$, $P_{\theta_0} = \mathrm{N}\left(\theta_0, 1\right)$ for $\theta_0 \in \{0, 1\}$, and the two null hypotheses may be restated as $\theta = 0$ and $\theta = 1$. Thus, the usual two-sided p-value function $p\left(\bullet; x\right)$ is given by

$$p\left(\theta_0; x\right) = 2\left(\Phi\left(x - \theta_0\right) \wedge \left(1 - \Phi\left(x - \theta_0\right)\right)\right),$$

where $\wedge$ is the minimum and Φ the standard normal distribution function. Figure 1 displays the following "significance values" of the hypothesis that $\theta = 1$:

6

1. The two-sided p value $p(1; x) = 2(\Phi(x-1) \wedge (1 - \Phi(x-1)))$ appears in solid gray. This does not depend on the hypothesis that $\theta = 0$.

2. The corresponding compatibility of the hypothesis that $\theta = 1$ with x conditional on $\theta \in \{0, 1\}$ appears in solid black. According to equation (1), that compatibility is

$$c(1; x|\{0,1\}) = \begin{cases} \frac{p(1;x)}{p(0;x)} & \text{if } p(1;x) < p(0;x) \\ 1 & \text{if } p(1;x) \geq p(0;x) \end{cases},$$

where $p(0; x) = 2(\Phi(x) \wedge (1 - \Phi(x)))$ is the p value of the hypothesis that $\theta = 0$.

3. The posterior probability that $\theta = 1$ on the basis of 50% prior probability of each of the null hypotheses conditional on $\theta \in \{0, 1\}$ appears in dashed black.

From Figure 1, it can be seen that, given any significance level $\alpha \in [0, 1]$, the p value would erroneously lead to the rejection of the better-supported null hypothesis for sufficiently large $x > 1$ but that the other two quantities take the other null hypothesis into account. Even when observing a value as high as $x = 3$, the c value reasonably indicates no evidence against the null hypothesis that $\theta = 1$ given the information that $\theta \in \{0, 1\}$, information the p value ignores.

Further, for all $x > 1/2$, there is not any $\alpha \in [0, 1]$ such that the compatibility conditional on $\theta \in \{0, 1\}$ is less than α, with the result that it is impossible to reject the better-supported null hypothesis, regardless of how high the significance level is. The posterior probability does not share that feature: being strictly less than 1, it is less than sufficiently high values of α.

In agreement with $c(1; x|\{0,1\})$, Chuaqui (1991, p. 97) recommended the ratio of p val-
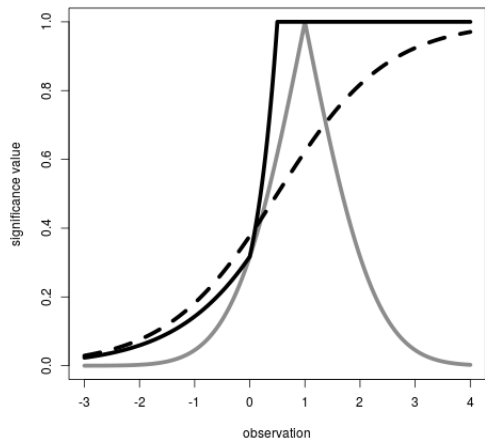
Figure 1: The p value $p(1; x)$ in solid gray, the data compatibility $c(1; x|\{0, 1\})$ in solid black, and the posterior probability that $\theta = 1$ in dashed black as functions of x, the value of the normal observation.

ues for comparing two hypotheses on the basis of the same observation. ▲

2.2 Interval estimation and other set estimation

As there is ambiguity in how formal notation in an English sentence can be understood, a few clarifying remarks may be helpful. The phrase "The hypothesis that $\theta \in \mathcal{H}_0$ is compatible" herein abbreviates "The hypothesis that θ is a member of $\mathcal{H}_0$ is compatible" rather than "The hypothesis that θ, which is a member of $\mathcal{H}_0$, is compatible." More generally, a hypothesis about a parameter value, not the parameter value itself, may be compatible with the data, rejected, accepted, etc.

For the purpose of representing hypotheses, 2^Θ will denote the set of all subsets of Θ. For any $\mathcal{H}_0 \in 2^\Theta$, the hypothesis that $\theta \in \mathcal{H}_0$ is *simple* if $\mathcal{H}_0$ has one member and *composite* if it has multiple members.

8

What it means for a hypothesis to be "compatible" with data is defined in analogy with confidence intervals. For any restriction of θ to a set $\mathcal{R} \in 2^{\Theta} \setminus \{\emptyset\}$, the set

$$\text{CS}\left(\alpha; x | \mathcal{R}\right) = \{\theta_0 \in \mathcal{R} : p\left(\theta_0; x\right) \geq \alpha\} \tag{2}$$

is known as a $(1 - \alpha)(100\%)$-*confidence set* for any $\theta_0 \in \mathcal{R}$ since

$$\lim_{n \to \infty} P_{\theta_0, \gamma}\left(\theta_0 \in \text{CS}\left(\alpha; X | \mathcal{R}\right)\right) = 1 - \alpha$$

for all $\alpha \in\ ]0, 1]$ and $\gamma \in \Gamma$ results from equation (4). It is called *exact* if its coverage is equal to $1 - \alpha$ for all n sufficiently large, which requires X to be continuous (§4.1).

Definition 1. For any $\mathcal{H}_0, \mathcal{R} \in 2^{\Theta} \setminus \{\emptyset\}$, $x \in \mathcal{X}$, and $\alpha \in\ ]0, 1]$, the hypothesis that $\theta \in \mathcal{H}_0$ is α-*compatible* with the observation that $X = x$, conditional on the restriction that $\theta \in \mathcal{R}$, if there is a $\theta_0 \in \mathcal{H}_0$ such that $c\left(\theta_0; x | \mathcal{R}\right) \geq \alpha$, where $c\left(\theta_0; x | \mathcal{R}\right)$ is the c value of the hypothesis that $\theta = \theta_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$. The α-*compatibility set* given $X = x$ and $\theta \in \mathcal{R}$ is

$$\mathcal{H}\left(\alpha; x | \mathcal{R}\right) = \{\theta_0 \in \Theta : c\left(\theta_0; x | \mathcal{R}\right) \geq \alpha\} \tag{3}$$

for all $\mathcal{R} \in 2^{\Theta} \setminus \{\emptyset\}$, $x \in \mathcal{X}$, and $\alpha \in\ ]0, 1]$.

The definition formally explicates the imprecise idea of whether a hypothesis is compatible with the data given any restrictions. As will be seen in Section 5.2, $c\left(\theta_0; x | \Theta\right) = p\left(\theta_0; x\right)$ often holds when there are no restrictions on θ.

9

3 Additional examples

Like Example 1, the following examples illustrate the c value and support the claim that it is more suitable than the p value as a measure of the compatibility between a hypothesis and data. The first example is an idealized version of restricted parameter problem encountered, for example, in physics (§1).

Example 2. Bounded parameter. Fraser (2011) considered a $N(\theta, 1)$ observable variable $X \sim P_\theta = N(\theta, 1)$ with observed value x and the parameter restriction $\theta \geq 0$, and the left-tailed version of the two-tailed p value

$$p(\theta_0; x) = 2\left(\Phi(x - \theta_0) \wedge (1 - \Phi(x - \theta_0))\right)$$

for every $\theta_0 \geq 0$. Thus, if $x \geq 0$, then $p(\theta_0; x) = 1$ holds for a value of $\theta_0 \geq 0$, namely, $\theta_0 = x$. In that case, Corollary 2 applies (see Section 5.2), and $c(\theta_0; x|[0, \infty[) = p(\theta_0; x)$ for all $\theta_0 \geq 0$. On the other hand, if $x < 0$, then Corollary 1 (see Section 5.2) instead gives $c(\theta_0; x|[0, \infty[) = p(\theta_0; x)/\sup_{\theta_1 \geq 0} p(\theta_1; x)$ for all $\theta_0 \geq 0$. This relationship between the compatibility and the p value is seen in Figure 2 for the observation $x = -1$. The exact $(1 - \alpha)(100\%)$-confidence interval is

$$\mathrm{CI}(\alpha; x|[0, \infty[) = \{\theta_0 \geq 0 : p(\theta_0; x) \geq \alpha\} = \left[0 \vee \left(x + \Phi^{-1}(\alpha/2)\right), 0 \vee \left(x + \Phi^{-1}(1 - \alpha/2)\right)\right],$$

with Φ^{-1} denoting the quantile function. By contrast, equation (3) and Theorem 1 (see

Section 5.2) give the α-compatibility interval

$$
\begin{aligned}
\mathcal{H}\left(\alpha; x \mid [0, \infty[\right) &= \left\{\theta_0 \in \Theta : p\left(\theta_0; x\right) \geq \alpha \sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right)\right\} \\
&= \left\{\theta_0 \in \Theta : p\left(\theta_0; x\right) \geq \alpha \sup_{\theta_1 \in [0, \infty[} p\left(\theta_1; x\right)\right\} \\
&= \left[0 \vee \left(x + \Phi^{-1}\left(\frac{\alpha p^+\left(x\right)}{2}\right)\right), 0 \vee \left(x + \Phi^{-1}\left(1 - \frac{\alpha p^+\left(x\right)}{2}\right)\right)\right],
\end{aligned}
$$

where $p^+\left(x\right) = \sup_{\theta_1 \geq 0} p\left(\theta_1; x\right)$. As required by Theorem 2, $\mathcal{H}\left(\alpha; x \mid [0, \infty[\right) = \mathrm{CI}\left(\alpha p^+\left(x\right); x \mid [0, \infty[\right)$. For the observation $x = -1$, the confidence intervals are compared to their compatibility counterparts in Figure 3.

If the variance were unknown, the solution would depend on whether the mean is still of interest or whether the mean-variance pair is the new parameter of interest. In the former case, the variance would be a nuisance parameter, and the t test could be used to obtain the p values on which the compatibility values and intervals are based. They would approach the above results asymptotically. In the latter case, maximization over the mean and variance rather than only over the mean in equation (1) would lead to very different compatibility values and intervals. Both cases are discussed in Example 6. ▲

The next three examples involve discrete observations to illustrate cases in which the p value is not exactly $\mathrm{U}\left(0, 1\right)$ under the null hypothesis.

Example 3. Mandelkern (2002b) and Fraser et al. (2004) discussed a restricted parameter problem for Poisson distributions. In physics, background signal and the event of interest are typically modeled under an additive structure: the count of background signal plus the count of the event signal (see van Dyk, 2014, for an application in the Large Hadron Collider). The observable count is modeled as a sum of two Poisson processes: $X = B + E$, with the count
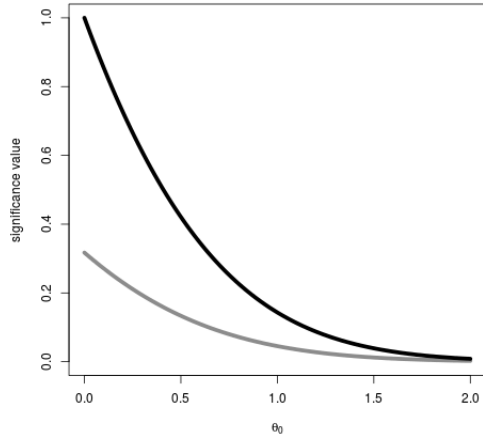
11

Figure 2: The p value $p(\theta_0; -1)$ in gray and the data compatibility $c(\theta_0; -1|\,[0, \infty[)$ in black as functions of θ_0, the parameter value.
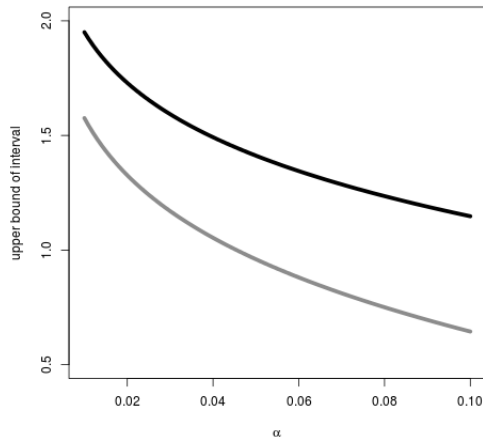
Figure 3: The upper bounds of the α-confidence interval $\mathrm{CI}(\alpha; -1|\,[0, \infty[)$ in gray and of the α-compatibility interval $\mathcal{H}(\alpha; -1|\,[0, \infty[)$ in black as functions of α, the threshold applied to the curves of Figure 2.

of background signal $B \sim \text{Poisson}(b)$ being independent of the count of the event signal $E \sim \text{Poisson}(\mu)$, where $b > 0$ is known and $\mu \geq 0$. Then, $X \sim \text{Poisson}(\theta) \equiv P_\theta$, where $\theta \geq b$. Let $X_1, \ldots, X_n$ be an independent and identical distributed random sample of X. The interest is in testing the null hypothesis $H_0 : \theta = \theta_0$ under the restriction $\theta \geq b$. The mid-p value is

$$p\left(\theta_0; \bar{x}\right) = P_{\theta_0}\left(\left|\bar{X} - \theta_0\right| > x_0\right) + \frac{1}{2} P_{\theta_0}\left(\left|\bar{X} - \theta_0\right| = x_0\right),$$

for $\theta_0 \geq b$, where $x_0 = |\bar{x} - \theta_0|$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean. Under H_0, $n\bar{X} \sim \text{Poisson}(n\theta_0)$, then the mid-$p$ value is easily calculated by

$$p\left(\theta_0; \bar{x}\right) = 1 - P_{\theta_0}\left(n\theta_0 - nx_0 \leq n\bar{X} \leq n\theta_0 + nx_0\right) + \frac{1}{2} P_{\theta_0}\left(n\bar{X} \in \{n\theta_0 + nx_0, n\theta_0 - nx_0\}\right)$$

By equation (1), $c\left(\theta_0; \bar{x}\right| [b, \infty[) = p\left(\theta_0; \bar{x}\right) / \sup_{\theta \geq b} p\left(\theta; \bar{x}\right)$ for all $\theta_0 \geq b$. The relationship between the compatibility and the p value is seen in Figure 4 for $n = 1$, $b = 10$ and the observed sample mean $\bar{x} = 9$. The approximate $(1 - \alpha)(100\%)$-confidence interval and the α-compatibility interval are computed from the equations (2) and (3). ▲

Example 4. Consider a binomial random variable $X \sim P_\theta = \text{Bin}(n, \theta)$, where $\theta \in]0, 1[$, with observed value x. The mid-p value for testing $H_0 : \theta = \theta_0$ is

$$p\left(\theta_0; x\right) = P_{\theta_0}\left(\left|X - n\theta_0\right| > x_0\right) + \frac{1}{2} P_{\theta_0}\left(\left|X - n\theta_0\right| = x_0\right),$$

where $x_0 = |x - n\theta_0|$ which can be written as

Figure 4: The p value $p(\theta_0; 9)$ in gray and the data compatibility $c(\theta_0; 9|[10, \infty[)$ in black as functions of θ_0, the parameter value.
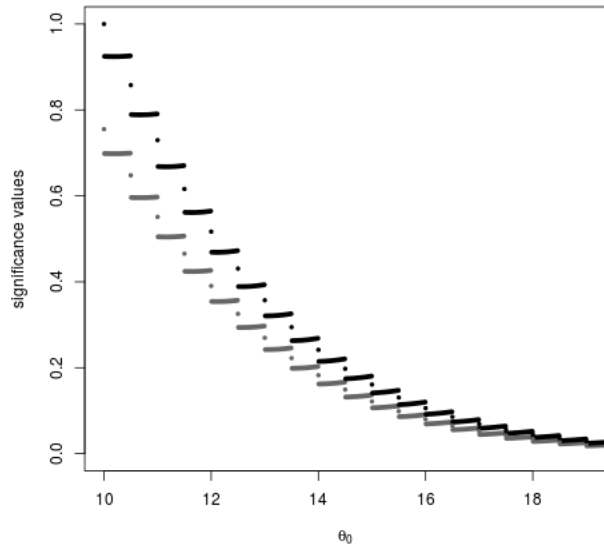
$$p(\theta_0; x) = 1 - P_{\theta_0}(n\theta_0 - x_0 \leq X \leq n\theta_0 + x_0) + \frac{1}{2}P_{\theta_0}(X \in \{n\theta_0 - x_0, n\theta_0 + x_0\}),$$

By equation (1), $c(\theta_0; x|]0, 1[) = p(\theta_0; x) / \sup_{\theta \in]0,1[} p(\theta; x)$ for all $\theta_0 \in]0, 1[$. The relationship between the compatibility and the p value is seen in Figure 5 for $n = 1$ and the observed value $x = 0$. The approximate $(1 - \alpha)(100\%)$-confidence interval and the α-compatibility interval are computed from the equations (2) and (3). ▲

Example 5. Consider a negative binomial random variable $X \sim P_\theta = \text{NBin}(n, \theta)$, where $P_\theta(X = x) = \binom{x+n-1}{n-1}\theta^n(1 - \theta)^x$, with $\theta \in]0, 1[$ and $x \in \{0, 1, 2, \ldots\}$. The mid-$p$ value for testing $H_0 : \theta = \theta_0$ is
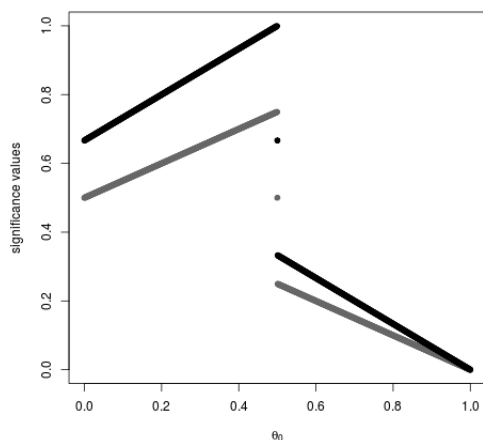
14

Figure 5: The p value $p(\theta_0; 0)$ in gray and the data compatibility $c(\theta_0; 0 | \,]0, 1[)$ in black as functions of θ_0, the parameter value.

$$p(\theta_0; x) = P_{\theta_0}(|X - g(\theta_0)| > x_0) + \frac{1}{2} P_{\theta_0}(|X - g(\theta_0)| = x_0),$$

where $x_0 = |x - g(\theta_0)|$ and $g(\theta_0) = n\theta_0^{-1}(1 - \theta_0)$ is the expectation of X, under H_0. Under H_0, $X \sim \text{Nbin}(n, \theta_0)$, then the mid-$p$ value can be computed by

$$p(\theta_0; x) = 1 - P_{\theta_0}(g(\theta_0) - x_0 \leq X \leq g(\theta_0) + x_0) + \frac{1}{2} P_{\theta_0}(X \in \{g(\theta_0) - x_0, g(\theta_0) + x_0\}).$$

By equation (1), $c(\theta_0; x | \,]0, 1[) = p(\theta_0; x) / \sup_{\theta \in]0,1[} p(\theta; x)$ for all $\theta_0 \in \,]0, 1[$. The relationship between the compatibility and the p value is seen in Figure 6 for $n = 1$ and the observed value $x = 1$. The approximate $(1 - \alpha)(100\%)$-confidence interval and the α-compatibility interval are computed from the equations (2) and (3). ▲

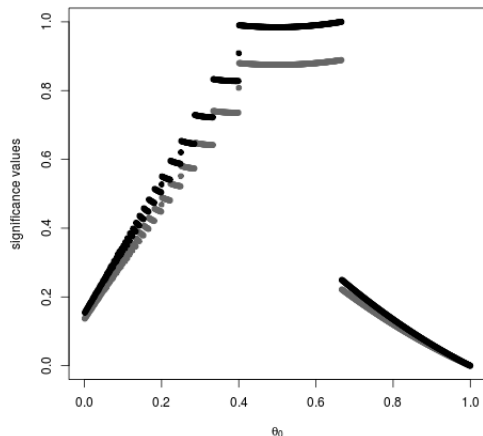Although Examples 3, 4 and 5 all employ the mid-p value method to compute the c value,

15

Figure 6: The p value $p(\theta_0; 1)$ in gray and the data compatibility $c(\theta_0; 1|\,]0, 1[)$ in black as functions of θ_0, the parameter value

they illustrate that the c values are affected qualitatively by the model specification. In the Poisson case, when $n = 1$ and $\bar{x} = 9$, the c value has many points of discontinuity (Figure 4). In the binomial scenario, when $n = 1$ and $x = 0$, there is only one point of discontinuity, which is at $\theta = 0.5$ (Figure 5); in the negative binomial case, when $n = 1$ and $x = 1$, there is one point of discontinuity greater than 0.5 and many smaller than 0.5 (Figure 6).

Example 6. Let $X = (X_1, \ldots, X_n)$ be a random sample from a normal distribution with unknown mean μ and unknown variance $\sigma^2 > 0$. We consider the two cases discussed in Example 2 with no restriction on the parameter space, namely, a) μ and σ^2 are parameters of interest, i.e., $\theta = (\mu, \sigma^2)$ and b) μ is the parameter of interest and σ^2 is the nuisance parameter, i.e., $\theta = \mu$ and $\gamma = \sigma^2$. For case a), the p value for testing simple hypothesis $H_{00} : (\mu, \sigma^2) = (\mu_0, \sigma_0^2)$ is given by

$$p_1\left((\mu_0, \sigma_0^2); x\right) = 2\left(\Phi\left(\sqrt{n}\frac{\bar{x} - \mu_0}{\sigma_0}\right) \wedge \left(1 - \Phi\left(\sqrt{n}\frac{\bar{x} - \mu_0}{\sigma_0}\right)\right)\right).$$

16

According to equation (1), the c value under no restriction is precisely the above p value, namely, $c_1 \left((\mu_0, \sigma_0^2); x \right) = p_1 \left((\mu_0, \sigma_0^2); x \right)$. The hypothesis $H_{00'} : \theta \in \mathcal{H}_{00'} (\mu_0)$, where $\mathcal{H}_{00'} (\mu_0) = \{(\mu_0, \sigma^2) : \sigma^2 > 0\}$, is the hypothesis that $\mu = \mu_0$. The associated C value is

$$C_1 \left(\mathcal{H}_{00'} (\mu_0); x \right) = \sup_{\sigma_0^2 > 0} c_1 \left((\mu_0, \sigma_0^2); x \right) = \lim_{\sigma_0^2 \to \infty} p_1 \left((\mu_0, \sigma_0^2); x \right) = 2 \left(\frac{1}{2} \wedge \left(1 - \frac{1}{2} \right) \right) = 1$$

for all $-\infty < \mu_0 < \infty$. That is, based on C_1, it is not possible to reject the hypothesis that $\theta_0 \in \mathcal{H}_{00'} (\mu_0)$ for any fixed significance value $\alpha \in]0, 1[$. Despite this fact, C_1 is still useful to test hypotheses that actually concern both μ and σ^2, for instance $H_{00''} : \mu \geq 0, \sigma^2 \leq 1$.

For case b), the p value for testing simple hypothesis $H_{01} : \mu = \mu_0$ is given by

$$p_2 (\mu_0; x) = 2 \left(F_{T_{n-1}} \left(\sqrt{n} \frac{\bar{x} - \mu_0}{s_x} \right) \wedge \left(1 - F_{T_{n-1}} \left(\sqrt{n} \frac{\bar{x} - \mu_0}{s_x} \right) \right) \right),$$

where F_{T_k} is the cumulative distribution of a Student-t random variable with k degrees-of-freedom. According to equation (1), the c value under no restriction is $c_2 (\mu_0; x) = p_2 (\mu_0; x)$. The hypothesis $H_{01'} : \theta \in \mathcal{H}_{01'} (\mu_0)$, where $\mathcal{H}_{01'} (\mu_0) = \{\mu_0\}$, is the hypothesis that $\mu = \mu_0$. The associated C value is

$$C_2 \left(\mathcal{H}_{01'} (\mu_0); x \right) = c_2 (\mu_0; x) = p_2 (\mu_0; x) .$$

Figure 7 shows the curves $C_1 \left(\mathcal{H}_{00'} (\mu_0); x \right)$ in black and $C_2 \left(\mathcal{H}_{01'} (\mu_0); x \right)$ in gray as functions of μ_0, for $n = 2$, $x_1 = 1$ and $x_2 = 2$. ▲

17
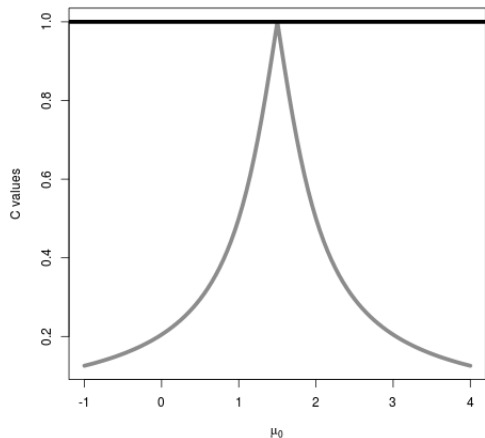
Figure 7: The C values $C_1\left(\mathcal{H}_{00'}(\mu_0); x\right)$ in black and $C_2\left(\mathcal{H}_{01'}(\mu_0); x\right)$ in gray as functions of μ_0 when $n = 2$, $x_1 = 1$ and $x_2 = 2$.

4 Axioms of data-hypothesis compatibility

4.1 Preliminary notation

For convenience, we review some notation introduced in Section 2. The unknown values θ and γ of the parameter of interest and of the nuisance parameter are members of the sets Θ and Γ, respectively. The observed tuple x is a member of some set $\mathcal{X}$ of possible observations.

A function $p\left(\bullet; \bullet\right) : \Theta \times \mathcal{X} \to [0, 1]$ is a *p-value function* if

$$\lim_{n \to \infty} P_{\theta_0, \gamma}\left(p\left(\theta_0; X\right) < \alpha\right) = \alpha \tag{4}$$

for all $\theta_0 \in \Theta$, $\gamma \in \Gamma$, and $0 \leq \alpha \leq 1$. Each $p\left(\theta_0; x\right)$ is the p value for testing the hypothesis that $\theta = \theta_0$ given the observation that $X = x$. While usual p-value functions are isomorphic to confidence distributions (Bickel and Padilla, 2014; cf. Schweder and Hjort, 2002; Xie and Singh, 2013; Nadarajah et al., 2015), the concept of the observed confidence level (Polansky,

18

2007), a belief-type probability according to a confidence distribution, plays no role in the current paper, in which probability is always of the frequency type (see Hacking, 2001).

4.2 Degrees of data-hypothesis compatibility

4.2.1 Axioms of compatibility

The next definition applies the α-compatible concept to composite hypotheses as well as simple hypotheses. Just as a p value can be defined in terms of whether the null hypothesis is rejected at a fixed significance level α, the degree of compatibility with data is defined in terms of whether the null hypothesis is α-compatible with the data at a fixed value of α.

Definition 2. The functions $C\left(\bullet; \bullet | \bullet\right) : 2^{\Theta} \times \mathcal{X} \times 2^{\Theta} \setminus \{\emptyset\} \rightarrow [0, 1]$ and $C\left(\bullet; \bullet\right) = C\left(\bullet; \bullet | \Theta\right) :$ $2^{\Theta} \times \mathcal{X} \rightarrow [0, 1]$ are *compatibility set functions*, and $C\left(\mathcal{H}_0; x | \mathcal{R}\right)$ is the *compatibility* of the hypothesis that $\theta \in \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$ if these conditions hold for all $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^{\Theta}$, and $\mathcal{R} \in 2^{\Theta} \setminus \{\emptyset\}$:

- *Axiom of minimal compatibility.* If $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$, then $C\left(\mathcal{H}_0; x | \mathcal{R}\right) = 0$.

- *Axiom of maximal compatibility.* $C\left(\Theta; x | \Theta\right) = 1$.

- *Axiom of conditional compatibility.* If $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$, then

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \frac{C\left(\mathcal{H}_0 \cap \mathcal{R}; x\right)}{C\left(\mathcal{R}; x\right)}. \tag{5}$$

- *Axiom of compatible hypotheses.* With "$\mathcal{H}_0 \overset{\alpha | \mathcal{R}}{\sim} x$" denoting the hypothesis that "$\theta \in \mathcal{H}_0$ is α-compatible with the observation that $X = x$, conditional on the restriction that

19

$\theta \in \mathcal{R}$,"

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \sup\left\{\alpha \in \left]0, 1\right] : \mathcal{H}_0 \overset{\alpha | \mathcal{R}}{\sim} x\right\}. \tag{6}$$

- *Axiom of evidential compatibility.* For any $\theta_0, \theta_1 \in \mathcal{R}$,

$$\frac{C\left(\{\theta_0\}; x | \mathcal{R}\right)}{C\left(\{\theta_1\}; x | \mathcal{R}\right)} = \frac{p\left(\theta_0; x\right)}{p\left(\theta_1; x\right)}. \tag{7}$$

The functions $c\left(\bullet; \bullet | \bullet\right) : \Theta \times \mathcal{X} \times 2^{\Theta} \backslash \{\emptyset\} \rightarrow [0, 1]$ and $c\left(\bullet; \bullet\right) = c\left(\bullet; \bullet | \Theta\right) : \Theta \times \mathcal{X} \rightarrow [0, 1]$ are *compatibility point functions* if $c\left(\theta_0; x | \mathcal{R}\right) = C\left(\{\theta_0\}; x | \mathcal{R}\right)$ for all $\theta_0 \in \Theta$, $x \in \mathcal{X}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$.

The compatibility $C\left(\mathcal{H}_0; x | \mathcal{R}\right)$ is the degree to which the hypothesis that $\theta \in \mathcal{H}_0$ is compatible with x under the restriction that $\theta \in \mathcal{R}$. This definition gives Definition 1 an axiomatic foundation by connecting the compatibility functions to the p-value function.

4.2.2 Explanations of the axioms

Each axiom has its own motivation. The first two are simply what Jeffreys (1948) calls conventions since the 0 and 1 could be replaced by any positive numbers as long as the second exceeds the first. The rationale for the axiom of conditional compatibility will become clear in light of possibility theory (§7).

The basis of the axiom of compatible hypotheses on Definition 1 specifies what is meant by data-hypothesis compatibility. It makes compatibility similar to a p value in that it is designed to reject hypotheses of sufficiently low values. Equation (6) says the degree of compatibility of a hypothesis with data, conditional on the parameter restriction, is the highest level of α such that the hypothesis remains α-compatible with the data, conditional

20

on the restriction.

The axiom of evidential compatibility might be justified by p-value functions of the form

$$\mathbb{R} \ni \theta' \mapsto p\left(\theta'; x\right) = P_{\theta', \gamma}\left(\tau\left(X\right) \geq \tau\left(x\right)\right), \tag{8}$$

where τ is a function transforming a sample to a real statistic that does not depend on θ' or γ such that the distribution of $\tau\left(X\right)$ does not depend on γ. This occurs most commonly in practice when there is no nuisance parameter γ and when $\tau\left(X\right)$ is a point estimator of θ, implying that $\tau\left(x\right)$ is the observed point estimate. Because $p\left(\bullet; x\right)$ is a function on $\Theta = \mathbb{R}$ according to equation (8), it can be used to compare the hypothesis that $\theta = \theta_0$ to the hypothesis that $\theta = \theta_1$ for any $\theta_0, \theta_1 \in \mathbb{R}$. Comparing the two point hypotheses suggests a likelihood-ratio approach to measuring evidence (Royall, 1997). The relevant likelihood ratio involves f_{θ_0}, the probability mass function on $\{0, 1\}$ that satisfies $f_{\theta_0}\left(0\right) = P_{\theta_0, \gamma}\left(\tau\left(X\right) < \tau\left(x\right)\right)$ and $f_{\theta_0}\left(1\right) = P_{\theta_0, \gamma}\left(\tau\left(X\right) \geq \tau\left(x\right)\right)$. Thus,

$$p\left(\theta_0; x\right) = P_{\theta_0, \gamma}\left(1_{[\tau(x), \infty[}\left(\tau\left(X\right)\right) = 1\right) = f_{\theta_0}\left(1\right),$$

and the analogous probability mass function f_{θ_1} satisfies $p\left(\theta_1; x\right) = f_{\theta_1}\left(1\right)$. As a likelihood ratio based on reduced data, $f_{\theta_0}\left(1\right) / f_{\theta_1}\left(1\right)$ is the strength of the statistical evidence in the observation that $1_{[\tau(x), \infty[}\left(\tau\left(X\right)\right) = 1$ in favor the hypothesis that $\theta = \theta_0$ as opposed to the hypothesis that $\theta = \theta_1$ (Royall, 1997). Requiring the data-compatibility of a hypothesis to be proportional to its strength of the statistical evidence results in

$$\frac{C\left(\{\theta_0\}; x | \mathcal{R}\right)}{C\left(\{\theta_1\}; x | \mathcal{R}\right)} = \frac{f_{\theta_0}\left(1\right)}{f_{\theta_1}\left(1\right)}, \tag{9}$$

which, with $p\left(\theta_0; x\right) = f_{\theta_0}\left(1\right)$ and $p\left(\theta_1; x\right) = f_{\theta_1}\left(1\right)$, yields equation (7). Equating $f_{\theta_0}\left(1\right) / f_{\theta_1}\left(1\right)$ with the strength of statistical evidence is in turn justified by noting that $f_{\theta_0}\left(1\right) / f_{\theta_1}\left(1\right)$ is the Bayes factor in

$$\frac{\text{Prob}\left(\theta = \theta_0 | 1_{[\tau(x),\infty[}\left(\tau\left(X\right)\right) = 1, \theta \in \{\theta_0, \theta_1\}\right)}{\text{Prob}\left(\theta = \theta_1 | 1_{[\tau(x),\infty[}\left(\tau\left(X\right)\right) = 1, \theta \in \{\theta_0, \theta_1\}\right)} = \left(\frac{f_{\theta_0}\left(1\right)}{f_{\theta_1}\left(1\right)}\right)\left(\frac{\text{Prob}\left(\theta = \theta_0 | \theta \in \{\theta_0, \theta_1\}\right)}{\text{Prob}\left(\theta = \theta_1 | \theta \in \{\theta_0, \theta_1\}\right)}\right),$$

the equation relating the posterior odds to the prior odds. This follows the general principle that a measure of support for a hypothesis should agree with Bayes's theorem when a suitable prior is available even though the measure is also applicable without a prior (Edwards, 1992; Bickel, 2013a,b).

This rationale is not entirely convincing, for its equation (9) could only be derived in the special case of equation (8). Further, why should the likelihood ratio be based on the reduction of X to $1_{[\tau(x),\infty[}\left(\tau\left(X\right)\right)$ rather than on X directly, as is more usual when measuring the strength of evidence (Royall, 1997)? That such a data reduction is needed to consider a ratio of p values as a likelihood ratio may shed light on the cryptic comment that the p value is "not very defensible save as an approximation" (Fisher, 1973, p. 71; cf. 74-75).

In view of those shortcomings, the axiom of evidential compatibility may be relaxed by replacing equation (7) with the requirement that $C\left(\{\theta_0\}; x | \mathcal{R}\right)$ be a function of $p\left(\theta_0; x\right)$ that is continuous and strictly increasing but not necessarily linear. This defines a class of alternative measures of data-hypothesis compatibility. In our opinion, the main appeal of the axiom of evidential compatibility is its practical value in uniquely identifying a simple default. Nonetheless, the justification based on equation (8) may have some theoretical value in making a connection to likelihood methods of measuring the strength of evidence.

5   Properties of data-hypothesis compatibility

5.1   Relations between concepts

This lemma connects the concepts of a compatible hypothesis and a compatibility set.

Lemma 1. *For any* $\mathcal{H}_0 \in 2^\Theta$, $\mathcal{R} \in 2^\Theta \backslash \{\emptyset\}$, $x \in \mathcal{X}$, *and* $\alpha \in]0,1]$, *the hypothesis that* $\theta \in \mathcal{H}_0$ *is* α-*compatible with the observation that* $X = x$, *conditional on the restriction that* $\theta \in \mathcal{R}$, *if and only if* $\mathcal{H}_0 \cap \mathcal{H}(\alpha; x | \mathcal{R}) \neq \emptyset$, *where* $\mathcal{H}(\alpha; x | \mathcal{R})$ *is the* α-*compatibility set given* $X = x$ *and* $\theta \in \mathcal{R}$.

Proof. By definition, the hypothesis is α-compatible if and only if $\emptyset \neq \{\theta_0 \in \mathcal{H}_0 : c(\theta_0; x | \mathcal{R}) \geq \alpha\} = \mathcal{H}_0 \cap \{\theta_0 \in \Theta : c(\theta_0; x | \mathcal{R}) \geq \alpha\}$. $\qquad\square$

The compatibility of a hypothesis is now seen to be proportional to the p value.

Lemma 2. *For any* $\theta_0 \in \Theta$ *and* $x \in \mathcal{X}$, *the marginal compatibility of the hypothesis that* $\theta = \theta_0$ *with the observation that* $X = x$ *is*

$$c(\theta_0; x) = \kappa p(\theta_0; x) \tag{10}$$

for some $\kappa \in]0,1]$.

Proof. The axiom of evidential compatibility (7) and $c(\theta_0; x | \mathcal{R}) = C(\{\theta_0\}; x | \mathcal{R})$ give equation (10). $\qquad\square$

5.2   Deriving data-hypothesis compatibility

The compatibility is easily derived from the p-value function using the simple equations of the next two results.

Theorem 1. *The compatibility of the hypothesis that $\theta \in \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$ is*

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \begin{cases} 0 & \text{if } \mathcal{H}_0 \cap \mathcal{R} = \emptyset \\ \frac{\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p(\theta_0; x)}{\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x)} & \text{if } \mathcal{H}_0 \cap \mathcal{R} \neq \emptyset \end{cases}$$

for all $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^{\Theta}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$.

Proof. In the case that $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$, the axiom of minimal compatibility gives $C\left(\mathcal{H}_0; x | \mathcal{R}\right) = 0$. In the $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$ case, Definition 1 and equation (6) yield

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \sup \left\{\alpha \in \,]0, 1] : \theta_0 \in \mathcal{H}_0, c\left(\theta_0; x | \mathcal{R}\right) \geq \alpha\right\} = \sup_{\theta_0 \in \mathcal{H}_0} c\left(\theta_0; x | \mathcal{R}\right). \quad (11)$$

Thus, the axiom of conditional compatibility (5) gives

$$C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x\right) / C\left(\mathcal{R}; x\right) = \frac{\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x\right)}{\sup_{\theta_1 \in \mathcal{R}} c\left(\theta_1; x\right)} \quad (12)$$

Since $C\left(\mathcal{R}; x\right) = C\left(\mathcal{R}; x | \Theta\right)$, equation (12) entails that $C\left(\mathcal{R}; x\right) = \sup_{\theta_1 \in \mathcal{R}} c\left(\theta_1; x\right) / C\left(\Theta; x\right)$. By the axiom of maximal compatibility, $C\left(\mathcal{R}; x\right) = \sup_{\theta_1 \in \mathcal{R}} c\left(\theta_1; x\right)$. Thus, with Lemma 2, equation (12) reduces to $C\left(\mathcal{H}_0; x | \mathcal{R}\right) = \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p\left(\theta_0; x\right) / \sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right)$. $\qquad \square$

Corollary 1. *For any $\theta_0 \in \Theta$, $x \in \mathcal{X}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$, the compatibility of the hypothesis that $\theta = \theta_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$ is given by equation (1).*

Proof. By Definition 2, $c\left(\theta_0; x | \mathcal{R}\right) = C\left(\{\theta_0\}; x | \mathcal{R}\right)$ for all $\theta_0 \in \Theta$. The desired result follows from Theorem 1. $\qquad \square$

24

In the usual setting of testing the simple hypothesis that $\theta = \theta_0$, the parameter is relatively unrestricted, and the compatibility is the p value. That is formally stated as the following direct result of Theorem 1 and Corollary 1.

Corollary 2. *For any $x \in \mathcal{X}$, $\theta_0 \in \mathcal{R}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$ such that $\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x) = 1$, the compatibility of the hypothesis that $\theta = \theta_0$ with the observation that $X = x$, conditional on the restriction that $\theta \in \mathcal{R}$, is $c(\theta_0; x|\mathcal{R}) = p(\theta_0; x)$. Under the same conditions, the compatibility of the hypothesis that $\theta \in \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$ is*

$$C(\mathcal{H}_0; x|\mathcal{R}) = \sup_{\theta_0 \in \mathcal{H}_0} p(\theta_0; x) \tag{13}$$

for all $x \in \mathcal{X}$ and $\mathcal{H}_0 \in 2^{\Theta}$ such that $\mathcal{H}_0 \subseteq \mathcal{R}$.

Corollary 2 justifies the practice of maximizing a p value over all the parameter values of a composite null hypothesis (e.g., Wendell and Schmee, 1996; Silvapulle and Sen, 2011, p. 33; Patriota, 2013).

The next corollary highlights ways conditional compatibility is similar to and different from conditional probability.

Corollary 3. *Given some $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^{\Theta}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$, the compatibility $C(\mathcal{H}_0; x|\mathcal{R})$ of the hypothesis that $\theta \in \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$ satisfies $C(\mathcal{H}_0; x|\mathcal{R}) = 1$ if and only if $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$ and*

$$\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x) = \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p(\theta_0; x). \tag{14}$$

Proof. In the $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$ case, Theorem 1 gives $C(\mathcal{H}_0; x|\mathcal{R}) = 0 \neq 1$. On the other hand,

in the case that $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$, Theorem 1 implies that equation (14) holds if and only if $C(\mathcal{H}_0; x | \mathcal{R}) = 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.3   Conservative error rate control and coverage

The following theorem demonstrates that compatibility controls the Type I error rate and that α-compatibility sets are $(1 - \alpha)(100\%)$-confidence sets that are valid in that their coverage rates are conservative if not exact.

Theorem 2. *For every $x \in \mathcal{X}$, $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$, and $\theta_0 \in \mathcal{R}$, let $p(\theta_0; x)$ denote the* p *value testing $\theta = \theta_0$ as the null hypothesis, and let $c(\theta_0; x | \mathcal{R})$ denote the compatibility of the hypothesis that $\theta = \theta_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$, let $\mathrm{CS}(\alpha; x | \mathcal{R})$ denote the exact confidence set given by equation (2), and let $\mathcal{H}(\alpha; x | \mathcal{R})$ denote the α-compatibility set given $X = x$ and $\theta \in \mathcal{R}$ for any $\alpha \in]0, 1]$. For any $\gamma \in \Gamma$, it follows that $c(\theta_0; x | \mathcal{R}) \geq p(\theta_0; x)$, $\mathrm{CS}(\alpha; x | \mathcal{R}) \subseteq \mathcal{H}(\alpha; x | \mathcal{R})$, and*

$$\mathcal{H}(\alpha; x | \mathcal{R}) = \mathrm{CS}\left(\alpha \sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x); x | \mathcal{R}\right) \tag{15}$$

$$\lim_{n \to \infty} P_{\theta_0, \gamma}(c(\theta_0; X | \mathcal{R}) < \alpha) \leq \alpha \tag{16}$$

$$\lim_{n \to \infty} P_{\theta_0, \gamma}(\theta_0 \in \mathcal{H}(\alpha; X | \mathcal{R})) \geq 1 - \alpha, \tag{17}$$

with the formulas (16) and (17) holding with exact equality if $\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x) = 1$.

Proof. Since $\theta_0 \in \mathcal{R}$, Corollary 1 entails that $c(\theta_0; x | \mathcal{R}) \geq p(\theta_0; x)$ for all $x \in \mathcal{X}$, from which $P_{\theta_0, \gamma}(c(\theta_0; X | \mathcal{R}) \geq p(\theta_0; X)) = 1$ follows, providing

$$P_{\theta_0, \gamma}(c(\theta_0; X | \mathcal{R}) < \alpha) \leq P_{\theta_0, \gamma}(p(\theta_0; X) < \alpha) = \alpha + o(1), \tag{18}$$

26

where o(1) converges to zero as $n \to \infty$, and equation (4) yields formula (16). Applying inequality (18) to equation (3),

$$\text{CS}\left(\alpha; x | \mathcal{R}\right) = \{\theta_0 \in \Theta : p\left(\theta_0; x\right) \geq \alpha\} \subseteq \mathcal{H}\left(\alpha; x | \mathcal{R}\right) \tag{19}$$

for every $x \in \mathcal{X}$. Hence, by equation (4),

$$\lim_{n \to \infty} P_{\theta_0, \gamma}\left(p\left(\theta_0; X\right) \geq \alpha\right) = 1 - \alpha \leq \lim_{n \to \infty} P_{\theta_0, \gamma}\left(\theta_0 \in \mathcal{H}\left(\alpha; X | \mathcal{R}\right)\right),$$

proving formula (17). Corollary 1 and equation (19) imply that $\text{CS}\left(\alpha \sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right); x | \mathcal{R}\right) = \{\theta_0 \in \Theta : c\left(\theta_0; x | \mathcal{R}\right) \geq \alpha\}$ and thus that equation (15) holds. Finally, if $\sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right) = 1$ for all $x \in \mathcal{X}$, then Lemma 1 requires that $c\left(\theta_0; x\right) = p\left(\theta_0; x\right)$ for all $\theta_0 \in \mathcal{R}$ and that $\text{CS}\left(\alpha; x | \mathcal{R}\right) = \mathcal{H}\left(\alpha; x | \mathcal{R}\right)$. In that case, $\lim_{n \to \infty} P_{\theta_0, \gamma}\left(c\left(\theta_0; X | \mathcal{R}\right) < \alpha\right) = \alpha$ and

$$\lim_{n \to \infty} P_{\theta_0, \gamma}\left(\theta_0 \in \mathcal{H}\left(\alpha; X | \mathcal{R}\right)\right) = 1 - \alpha$$

follow from equation (4). □

6 Hypothesis acceptance, rejection, or neither

6.1 Warrant for accepting a hypothesis

While the compatibility of a hypothesis with data does not warrant accepting the hypothesis, a lack of compatibility justifies rejecting it and accepting its negation under the statistical model. That idea leads to the following measure of the degree of warrant for accepting a hypothesis.

Definition 3. A function $W(\bullet; \bullet | \bullet) : 2^{\Theta} \times \mathcal{X} \times 2^{\Theta} \backslash \{\emptyset\} \rightarrow [0, 1]$, called the *warrant set function*, is defined as follows. For all $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^{\Theta}$, and $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$,

$$W(\mathcal{H}_0; x | \mathcal{R}) = 1 - C(\mathcal{R} \backslash \mathcal{H}_0; x | \mathcal{R})$$

is the *warrant* of the hypothesis that $\theta \in \mathcal{H}_0$ given the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$, where $C(\mathcal{R} \backslash \mathcal{H}_0; x | \mathcal{R})$ is the compatibility of the hypothesis that $\theta \in \mathcal{R}$ but $\theta \notin \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$.

From equation (5),

$$W(\mathcal{H}_0; x | \mathcal{R}) = 1 - C(\mathcal{R} \backslash \mathcal{H}_0; x | \mathcal{R}) = 1 - \frac{C((\mathcal{R} \backslash \mathcal{H}_0) \cap \mathcal{R}; x)}{C(\mathcal{R}; x)} = 1 - \frac{C(\mathcal{R} \backslash \mathcal{H}_0; x)}{C(\mathcal{R}; x)}. \quad (20)$$

For example, if $\mathcal{R} = \Theta$, then $W(\mathcal{H}_0; x | \Theta) = 1 - C(\mathcal{H}_0'; x)$, where $\mathcal{H}_0'$ is the complement of $\mathcal{H}_0$, that is, $\mathcal{H}_0' = \Theta \backslash \mathcal{H}_0$. However, it does not follow that $W(\mathcal{H}_0; x | \Theta) = C(\mathcal{H}_0; x)$, as it would if $C(\bullet; x)$ were a probability measure. That is because $C(\bullet; x)$ is a possibility measure (§7), a special case of an upper probability function, which is not an additive measure.

The warrant for a hypothesis corresponding to a set estimate $\mathcal{H}(\alpha; x | \mathcal{R})$ is important as a lower bound on the coverage rate of the set estimator $\mathcal{H}(\alpha; X | \mathcal{R})$, as formally stated in the next theorem.

Theorem 3. *Let $x \in \mathcal{X}$, $\mathcal{R} \in 2^{\Theta} \backslash \{\emptyset\}$, and $\theta_0 \in \mathcal{R}$, and let W denote a warrant function corresponding to $\mathcal{H}(\alpha; x | \mathcal{R})$, the α-compatibility set given $X = x$ for every $\alpha \in]0, 1]$. Assume*

$c\left(\theta_{0};x|\mathcal{R}\right)$ *is continuous as a function of* θ_{0}. *For any* $\alpha \in \,]0,1]$ *and* $\gamma \in \Gamma$,

$$\lim_{n\to\infty} P_{\theta_{0},\gamma}\left(\theta_{0} \in \mathcal{H}\left(\alpha;X|\mathcal{R}\right)\right) \geq W\left(\mathcal{H}\left(\alpha;x|\mathcal{R}\right);x|\mathcal{R}\right), \tag{21}$$

which holds with exact equality if $\sup_{\theta_{1}\in\mathcal{R}} p\left(\theta_{1};x\right) = 1$, *where* $p\left(\theta_{0};x\right)$ *is the* p *value testing* $\theta = \theta_{0}$ *as the null hypothesis for all* $\theta_{0} \in \mathcal{R}$.

Proof. According to the definitions of warrant and the α-compatibility set,

$$
\begin{aligned}
W\left(\mathcal{H}\left(\alpha;x|\mathcal{R}\right);x|\mathcal{R}\right) &= 1 - C\left(\mathcal{R}\backslash\mathcal{H}\left(\alpha;x|\mathcal{R}\right);x|\mathcal{R}\right) \\
&= 1 - C\left(\left\{\theta_{0} \in \mathcal{R} : c\left(\theta_{0};x|\mathcal{R}\right) < \alpha\right\};x|\mathcal{R}\right),
\end{aligned}
$$

Thus, since that C is the relevant compatibility set function,

$$W\left(\mathcal{H}\left(\alpha;x|\mathcal{R}\right);x|\mathcal{R}\right) = 1 - \sup\left\{c\left(\theta_{0};x|\mathcal{R}\right) < \alpha : \theta_{0} \in \mathcal{R}\right\} = 1 - \alpha \tag{22}$$

by the continuity assumption. Formula (21) then results from Theorem 2. The same theorem says $\sup_{\theta_{1}\in\mathcal{R}} p\left(\theta_{1};x\right) = 1$ implies that $\lim_{n\to\infty} P_{\theta_{0},\gamma}\left(\theta_{0} \in \mathcal{H}\left(\alpha;X|\mathcal{R}\right)\right) = 1 - \alpha$, leading to $\lim_{n\to\infty} P_{\theta_{0},\gamma}\left(\theta_{0} \in \mathcal{H}\left(\alpha;X|\mathcal{R}\right)\right) = W\left(\mathcal{H}\left(\alpha;x|\mathcal{R}\right);x|\mathcal{R}\right)$ via equation (22). $\square$

Equation (22) interprets the nominal confidence level $1 - \alpha$ as the degree of warrant for the hypothesis that the observed confidence set $\mathcal{H}\left(\alpha;x|\mathcal{R}\right)$ contains the target value of the parameter.

6.2 Acceptability of a hypothesis

The information in the data-compatibility and warrant of a hypothesis will be combined into a single measure of acceptability in this section. Hypotheses of sufficiently high acceptability are accepted, those with sufficiently negative acceptability are rejected, and the remaining hypotheses are neither accepted nor rejected. What circumstances require an agent to believe a rejected hypothesis to be false or to believe an accepted hypothesis to be true is a complex question (Cohen, 1992) that cannot be entertained here.

For any $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^\Theta$, and $\mathcal{R} \in 2^\Theta \setminus \{\emptyset\}$, recall that $C(\mathcal{H}_0; x | \mathcal{R})$ denotes the compatibility of the hypothesis that $\theta \in \mathcal{H}_0$ with the observation that $X = x$ conditional on the restriction that $\theta \in \mathcal{R}$.

Definition 4. The *acceptability* of the hypothesis that $\theta \in \mathcal{H}_0$ given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$ is the extended real number $A(\mathcal{H}_0; x | \mathcal{R}) \in \{-\infty, \infty\} \cup \mathbb{R}$ such that, for all $\alpha \in]0, 1]$,

$$(\theta_1 \in \mathcal{H}(\alpha; x | \mathcal{R}) \implies \theta_1 \in \mathcal{H}_0) \quad \Longleftrightarrow \quad A(\mathcal{H}_0; x | \mathcal{R}) > \log 1/\alpha \tag{23}$$

$$(\theta_2 \in \mathcal{H}(\alpha; x | \mathcal{R}) \implies \theta_2 \in \mathcal{H}_0') \quad \Longleftrightarrow \quad A(\mathcal{H}_0; x | \mathcal{R}) < -\log 1/\alpha \tag{24}$$

$$\exists \theta_1, \theta_2 \in \mathcal{H}(\alpha; x | \mathcal{R}); \, \theta_1 \in \mathcal{H}_0; \, \theta_2 \in \mathcal{H}_0' \quad \Longleftrightarrow \quad |A(\mathcal{H}_0; x | \mathcal{R})| \leq \log 1/\alpha, \tag{25}$$

where $\mathcal{H}(\alpha; x | \mathcal{R})$ is the α-compatibility set given $X = x$ and $\theta \in \mathcal{R}$. Here, the base of log might be 2 for best interpretability but can be any number greater than 1. At level α, the hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$, is *accepted* if and only if $A(\mathcal{H}_0; x | \mathcal{R}) > \log 1/\alpha$ and is *rejected* if and only if $A(\mathcal{H}_0; x | \mathcal{R}) < -\log 1/\alpha$. In the absence of a restriction ($\mathcal{R} = \Theta$), the acceptability $A(\mathcal{H}_0; x | \Theta)$ is abbreviated as $A(\mathcal{H}_0; x)$.

30

In that way, the acceptability of a general hypothesis over its alternative hypothesis is defined in terms of which values of the parameter of interest are compatible with the observed data and with the given restrictions according to Section 2.2. Formula (23) says a hypothesis is accepted at level α if it is consistent with *all* of the α-compatible parameter values. Likewise, formula (24) says a hypothesis is rejected at level α if it is not consistent with *any* of the α-compatible parameter values. Finally, formula (25) means there is insufficient evidence to accept or reject the hypothesis at level α if it is consistent with some but not all of the α-compatible parameter values.

The last case means there is no arbitrary requirement that every hypothesis be either rejected or accepted. At the same time, the rejection of a null hypothesis for lack of compatibility with other information necessarily implies acceptance of an alternative hypothesis, as this lemma makes clear.

Lemma 3. *These propositions are equivalent for any $\mathcal{H}_0 \in 2^\Theta$, $\mathcal{R} \in 2^\Theta \setminus \{\emptyset\}$, $x \in \mathcal{X}$, and $\alpha \in \,]0,1]$:*

1. *$A\left(\mathcal{H}_0; x | \mathcal{R}\right) < -\log {}^1\!/\!\alpha.$*

2. *The hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$, is rejected at level α.*

3. *The same hypothesis is not α-compatible with the observation that $X = x$, conditional on the restriction that $\theta \in \mathcal{R}$.*

4. *$A\left(\mathcal{H}_0'; x | \mathcal{R}\right) > \log {}^1\!/\!\alpha.$*

5. *The hypothesis that $\theta \in \mathcal{H}_0'$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$, is accepted at level α.*

31

Proof. Propositions 1 and 2 are equivalent by Definition 4: the hypothesis that $\theta \in \mathcal{H}_0$ is rejected if and only if $A(\mathcal{H}_0; x|\mathcal{R}) < -\log{}^1\!/\alpha$. Similarly, Propositions 4 and 5 are equivalent. According to formula (24), Proposition 1 is equivalent to

$$\theta_2 \in \mathcal{H}(\alpha; x|\mathcal{R}) \implies \theta_2 \in \mathcal{H}'_0, \tag{26}$$

which, by formula (23), holds if and only if $A(\mathcal{H}'_0; x|\mathcal{R}) > \log{}^1\!/\alpha$, the definition of accepting the hypothesis that $\theta \in \mathcal{H}'_0$. That establishes the equivalence of Propositions 1 and 4. Lemma 1 entails that Proposition 3 is equivalent to $\mathcal{H}_0 \cap \mathcal{H}(\alpha; x|\mathcal{R}) = \emptyset$, and that equivalence makes the same assertion as formula (26). Therefore, Propositions 2 and 3 are equivalent. $\square$

Thus, whereas the fact that a hypothesis is data-compatible is merely necessary for its acceptance, the fact that its denial is incompatible is sufficient. Calculating the acceptability is facilitated by the next theorem.

Theorem 4. *For any $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^\Theta$, and $\mathcal{R} \in 2^\Theta \backslash \{\emptyset\}$, the acceptability of the hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$, is*

$$A(\mathcal{H}_0; x|\mathcal{R}) = \log \frac{C(\mathcal{H}_0; x|\mathcal{R})}{C(\mathcal{H}'_0; x|\mathcal{R})}; \tag{27}$$

$A(\mathcal{H}_0; x|\mathcal{R}) = -\infty$ *if* $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$, $A(\mathcal{H}_0; x|\mathcal{R}) = \infty$ *if* $\mathcal{H}'_0 \cap \mathcal{R} = \emptyset$, *or*

$$A(\mathcal{H}_0; x|\mathcal{R}) = \log \frac{\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p(\theta_0; x)}{\sup_{\theta_0 \in \mathcal{H}'_0 \cap \mathcal{R}} p(\theta_0; x)} \tag{28}$$

if $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$ *and* $\mathcal{H}'_0 \cap \mathcal{R} \neq \emptyset$.

Proof. For any $\mathcal{H}_0 \in 2^\Theta$, let

$$\widetilde{A}(\mathcal{H}_0) = \log \frac{C(\mathcal{H}_0; x|\mathcal{R})}{C(\mathcal{H}_0'; x|\mathcal{R})}, \tag{29}$$

and let $A(\mathcal{H}_0; x|\mathcal{R})$ denote the acceptability of the hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$. Assume, contrary to the claim, that $A(\mathcal{H}_0; x|\mathcal{R}) \neq \widetilde{A}(\mathcal{H}_0)$. In the case that relation (14) holds, $\widetilde{A}(\mathcal{H}_0) = \log \frac{1}{C(\mathcal{H}_0'; x|\mathcal{R})}$ by Corollary 3. From equation (6) and Lemma 3,

$$
\begin{aligned}
\widetilde{A}(\mathcal{H}_0) &= \log\left(1/\sup\left\{\alpha \in \,]0,1] : \mathcal{H}_0' \overset{\alpha|\mathcal{R}}{\sim} x\right\}\right) \\
&= \log\left(1/\sup\left(\,]0,1] \setminus \left\{\alpha \in \,]0,1] : \neg\mathcal{H}_0' \overset{\alpha|\mathcal{R}}{\sim} x\right\}\right)\right) \\
&= \log\left(1/\sup\left\{\alpha \in \,]0,1] : A(\mathcal{H}_0'; x|\mathcal{R}) \geq -\log \frac{1}{\alpha}\right\}\right) \\
&= \inf\left\{\log \frac{1}{\alpha} \geq 0 : \log \frac{1}{\alpha} \geq -A(\mathcal{H}_0'; x|\mathcal{R})\right\} \\
&= -A(\mathcal{H}_0'; x|\mathcal{R}) = A(\mathcal{H}_0; x|\mathcal{R}),
\end{aligned}
$$

the last equality following from the equivalence of Propositions 1 and 4 of Lemma 3. In the case that relation (14) does not hold, $\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x) > \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p(\theta_0; x)$, yielding

$$\sup_{\theta_1 \in \mathcal{R}} p(\theta_1; x) = \sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} p(\theta_0; x).$$

Thus, Corollary 3 now gives $C(\mathcal{H}_0'; x|\mathcal{R}) = 1$ and $\widetilde{A}(\mathcal{H}_0) = \log C(\mathcal{H}_0; x|\mathcal{R})$ by implication.

From equation (6) and Lemma 3,

$$
\begin{aligned}
\widetilde{A}\left(\mathcal{H}_0\right) &= \log\left(\sup\left\{\alpha \in \,]0,1] : \mathcal{H}_0 \overset{\alpha|\mathcal{R}}{\sim} x\right\}\right) \\
&= \log\left(\sup\left\{\alpha \in \,]0,1] : A\left(\mathcal{H}_0; x|\mathcal{R}\right) \geq -\log\,{}^1\!/\!\alpha\right\}\right) \\
&= \sup\left\{\log\alpha \leq 0 : A\left(\mathcal{H}_0; x|\mathcal{R}\right) \geq \log\alpha\right\} \\
&= \sup\left\{\log\alpha \leq 0 : \log\alpha \leq A\left(\mathcal{H}_0; x|\mathcal{R}\right)\right\} \\
&= A\left(\mathcal{H}_0; x|\mathcal{R}\right).
\end{aligned}
$$

Therefore, $\widetilde{A}\left(\mathcal{H}_0\right) = A\left(\mathcal{H}_0; x|\mathcal{R}\right)$ in both possible cases, contradicting the assumption and establishing equation (27). The rest of the claims follow from Theorem 1. $\qquad\square$

Breaking that into the three major cases sheds light on the interpretation of acceptability.

Corollary 4. *For any $x \in \mathcal{X}$, $\mathcal{H}_0 \in 2^\Theta$, and $\mathcal{R} \in 2^\Theta \backslash \{\emptyset\}$ such that $\mathcal{H}_0 \cap \mathcal{R} \neq \emptyset$ and $\mathcal{H}_0' \cap \mathcal{R} \neq \emptyset$, the acceptability of the hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$, is*

$$
A\left(\mathcal{H}_0; x|\mathcal{R}\right) = \begin{cases}
-\log\sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right) & \text{if } \mathcal{H}_0 \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset, \mathcal{H}_0' \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) = \emptyset \\
\log\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right) & \text{if } \mathcal{H}_0 \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) = \emptyset, \mathcal{H}_0' \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset \\
0 & \text{if } \mathcal{H}_0 \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset, \mathcal{H}_0' \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset;
\end{cases}
$$

$$
\widehat{\mathcal{H}}\left(x|\mathcal{R}\right) = \left\{\theta_1 \in \mathcal{R} : \forall\theta_0 \in \mathcal{R}, p\left(\theta_0; x\right) \leq p\left(\theta_1; x\right)\right\}. \tag{30}
$$

Proof. Corollary 1 implies that $\widehat{\mathcal{H}}\left(x|\mathcal{R}\right) = \left\{\theta_0 \in \mathcal{R} : c\left(\theta_0; x|\mathcal{R}\right) = 1\right\}$. Thus, by equation

(28),

$$
\begin{aligned}
A\left(\mathcal{H}_0; x|\mathcal{R}\right) &= \log \frac{\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right)}{\sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right)} \\
&= \begin{cases} \log\left(1/\sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right)\right) & \text{if } \mathcal{H}_0 \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset \\ \log\left(\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right)/1\right) & \text{if } \mathcal{H}_0' \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset. \end{cases}
\end{aligned}
$$

If both $\mathcal{H}_0 \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset$ and $\mathcal{H}_0' \cap \widehat{\mathcal{H}}\left(x|\mathcal{R}\right) \neq \emptyset$, then $-\log \sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right) = \log \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right)$, which is only possible if $\sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right) = \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c\left(\theta_0; x|\mathcal{R}\right) = 1$. $\qquad\square$

Remark 1. As $A\left(\{\theta_0\}; x|\mathcal{R}\right) = \log p\left(\theta_0; x\right) - \log \sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right)$, the hypothesis that $\theta = \theta_0$ cannot be accepted when $\sup_{\theta_1 \in \mathcal{R}} p\left(\theta_1; x\right) = 1$, since, under this condition, $A\left(\{\theta_0\}; x|\mathcal{R}\right) = \log p\left(\theta_0; x\right) \leq 0$. Thus, in the typical case of testing a simple hypothesis (Corollary 2), its acceptability cannot be positive. That agrees with the idea commonly held by frequentists that evidence might be against a simple hypothesis but can never support it.

As stated in Section 1, every deductively cogent statistical procedure is both restriction-respecting and coherent. Those properties will be proven of the acceptability method (Definition 4) in the next two subsections.

6.3 Acceptability is restriction-respecting

Recall that a restriction-respecting statistical method does not permit the rejection of all hypotheses that are consistent with the restriction but requires the rejection of all hypotheses

that are inconsistent with the restriction (§1). Conditional acceptability is now seen to be restriction-respecting.

Theorem 5. *For any $\alpha \in\,]0,1]$, conditional on the restriction that $\theta \in \mathcal{R}$ for some $\mathcal{R} \in 2^{\Theta} \setminus \{\emptyset\}$, the procedure in Definition 4 rejects the hypothesis that $\theta \in \mathcal{H}_0$ for every $\mathcal{H}_0 \in 2^{\Theta}$ such that $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$ and does not reject every hypothesis that $\theta \in \mathcal{H}_1$ for all $\mathcal{H}_1 \in 2^{\Theta}$ such that $\mathcal{H}_1 \cap \mathcal{R} \neq \emptyset$.*

Proof. Theorem 4 says $A(\mathcal{H}_0; x|\mathcal{R}) = -\infty$ for every $\mathcal{H}_0 \in 2^{\Theta}$ such that $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$. Thus, $A(\mathcal{H}_0; x|\mathcal{R}) < -\log 1/\alpha$, which means $\theta \in \mathcal{H}_0$ is rejected, for all $\alpha \in\,]0,1]$. To prove the other claim, it sufficient to show that for at least one $\mathcal{H}_1 \in 2^{\Theta}$ such that $\mathcal{H}_1 \cap \mathcal{R} \neq \varnothing$ that $\theta \in \mathcal{H}_1$ cannot be rejected. Let $\widehat{\mathcal{H}}(x|\mathcal{R})$ be defined according to equation (30), and denote its complement by $\widehat{\mathcal{H}}'(x|\mathcal{R}) = \Theta \setminus \widehat{\mathcal{H}}(x|\mathcal{R})$. If $\widehat{\mathcal{H}}(x|\mathcal{R}) = \mathcal{R}$, then $\widehat{\mathcal{H}}'(x|\mathcal{R}) \cap \mathcal{R} = \emptyset$ and, according to Theorem 4, $A\left(\widehat{\mathcal{H}}(x|\mathcal{R}); x|\mathcal{R}\right) = \infty$. On the other hand, if $\widehat{\mathcal{H}}(x|\mathcal{R}) \neq \mathcal{R}$, then $\widehat{\mathcal{H}}'(x|\mathcal{R}) \cap \mathcal{R} \neq \emptyset$, and Theorem 4, with equation (30), yields

$$A\left(\widehat{\mathcal{H}}(x|\mathcal{R}); x|\mathcal{R}\right) = \log \frac{\sup_{\theta_0 \in \widehat{\mathcal{H}}(x|\mathcal{R})} p(\theta_0; x)}{\sup_{\theta_0 \in \widehat{\mathcal{H}}'(x|\mathcal{R}) \cap \mathcal{R}} p(\theta_0; x)} \geq 0$$

since for each $\theta_0 \in \widehat{\mathcal{H}}(x|\mathcal{R})$ and $\theta_1 \in \widehat{\mathcal{H}}'(x|\mathcal{R}) \cap \mathcal{R}$, $p(\theta_0; x) \geq p(\theta_1; x)$. Thus, since $A\left(\widehat{\mathcal{H}}(x|\mathcal{R}); x|\mathcal{R}\right) \geq 0$ in both cases, there is no $\alpha \in\,]0,1]$ such that $A\left(\widehat{\mathcal{H}}(x|\mathcal{R}); x|\mathcal{R}\right) < -\log 1/\alpha$, which means $\theta \in \widehat{\mathcal{H}}(x|\mathcal{R})$ cannot be rejected. $\square$

6.4 Acceptability is coherent

In the context of multiple comparisons, Gabriel (1969) called a statistical procedure "coherent" if, for every hypothesis that it rejects, it also rejects all of the hypotheses that imply the truth of the rejected hypothesis (§1). Thus, for every $\mathcal{H}_0 \in 2^{\Theta}$, any *rejection-coherent*

procedure rejects the hypothesis that $\theta \in \mathcal{H}_1$ for every $\mathcal{H}_1 \in 2^\Theta$ such that $\mathcal{H}_1 \subseteq \mathcal{H}_0$ if it rejects the hypothesis that $\theta \in \mathcal{H}_0$. Likewise, for every $\mathcal{H}_1 \in 2^\Theta$, any *acceptance-coherent* procedure accepts the hypothesis that $\theta \in \mathcal{H}_0$ for every $\mathcal{H}_0 \in 2^\Theta$ such that $\mathcal{H}_1 \subseteq \mathcal{H}_0$ if it accepts the hypothesis that $\theta \in \mathcal{H}_1$.

The concepts are applied to compatibility and acceptability in the next two results.

Lemma 4. *Conditional on the restriction that $\theta \in \mathcal{R}$ for some $\mathcal{R} \in 2^\Theta \backslash \{\emptyset\}$, the compatibility of the hypothesis that $\theta \in \mathcal{H}_1$ with the observation that $X = x$ is at most the compatibility of any hypothesis that it implies with the same observation, that is,*

$$C\left(\mathcal{H}_1; x|\mathcal{R}\right) \le C\left(\mathcal{H}_0; x|\mathcal{R}\right)$$

for every $\mathcal{H}_0, \mathcal{H}_1 \in 2^\Theta$ such that $\mathcal{H}_1 \subseteq \mathcal{H}_0$.

Proof. According to Theorem 1, either $C\left(\mathcal{H}_1; x|\mathcal{R}\right) = 0$, in which case $C\left(\mathcal{H}_1; x|\mathcal{R}\right) \le C\left(\mathcal{H}_0; x|\mathcal{R}\right)$, or $C\left(\mathcal{H}_1; x|\mathcal{R}\right) > 0$, in which case $\mathcal{H}_1 \cap \mathcal{R} \ne \emptyset$. Thus, since $\mathcal{H}_1 \subseteq \mathcal{H}_0$, it follows from $\mathcal{H}_1 \cap \mathcal{R} \ne \emptyset$ that $\mathcal{H}_0 \cap \mathcal{R} \ne \emptyset$ and, by Theorem 1, that

$$\frac{C\left(\mathcal{H}_0; x|\mathcal{R}\right)}{C\left(\mathcal{H}_1; x|\mathcal{R}\right)} = \frac{\sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} p\left(\theta_0; x\right)}{\sup_{\theta_0 \in \mathcal{H}_1 \cap \mathcal{R}} p\left(\theta_0; x\right)}.$$

That ratio satisfies $C\left(\mathcal{H}_0; x|\mathcal{R}\right) / C\left(\mathcal{H}_1; x|\mathcal{R}\right) \ge 1$ given that $\mathcal{H}_1 \cap \mathcal{R} \subseteq \mathcal{H}_0 \cap \mathcal{R}$. $\qquad\square$

Theorem 6. *Conditional on the restriction that $\theta \in \mathcal{R}$ for some $\mathcal{R} \in 2^\Theta \backslash \{\emptyset\}$, the procedure in Definition 4 is both rejection-coherent and acceptance-coherent for any $\alpha \in]0, 1]$, and the acceptability of the hypothesis that $\theta \in \mathcal{H}_1$ is at most the acceptability of any hypothesis that it implies, that is,*

$$A\left(\mathcal{H}_1; x|\mathcal{R}\right) \le A\left(\mathcal{H}_0; x|\mathcal{R}\right) \tag{31}$$

37

for every $\mathcal{H}_0, \mathcal{H}_1 \in 2^\Theta$ *such that* $\mathcal{H}_1 \subseteq \mathcal{H}_0$.

Proof. The following statements hold for any $\alpha \in \,]0, 1]$. According to Definition 4, the hypothesis that $\theta \in \mathcal{H}_0$, given the observation that $X = x$ and the restriction that $\theta \in \mathcal{R}$ is rejected at level α if and only if $A(\mathcal{H}_0; x|\mathcal{R}) < -\log 1/\alpha$. That requires that $A(\mathcal{H}_0; x|\mathcal{R}) < 0$, which only obtains when either $\mathcal{H}_0 \cap \mathcal{R} = \emptyset$, in which case $A(\mathcal{H}_0; x|\mathcal{R}) = -\infty$ by Theorem 4, or

$$A(\mathcal{H}_0; x|\mathcal{R}) = \log \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c(\theta_0; x|\mathcal{R}) = -\log 1/C(\mathcal{H}_0; x|\mathcal{R}) \tag{32}$$

by Corollary 4 and Theorem 1. If, on the other hand $A(\mathcal{H}_0; x|\mathcal{R}) > 0$, as required for acceptance $(A(\mathcal{H}_0; x|\mathcal{R}) > \log 1/\alpha)$ then either $\mathcal{H}_0' \cap \mathcal{R} = \emptyset$, in which case $A(\mathcal{H}_0; x|\mathcal{R}) = \infty$ by Theorem 4, or

$$A(\mathcal{H}_0; x|\mathcal{R}) = -\log \sup_{\theta_0 \in \mathcal{H}_0' \cap \mathcal{R}} c(\theta_0; x|\mathcal{R}) = \log 1/C(\mathcal{H}_0'; x|\mathcal{R}) \tag{33}$$

by Corollary 4 and Theorem 1. Whether equation (32) or equation (33) applies, equation (31) follows from Lemma 4. Both rejection coherence and acceptance coherence are immediate consequences of equation (31).

$\square$

7 Connections with possibility theory

In agreement with the classical idea of inference to the best explanation (Peirce, 1998, p. 234), the acceptability $A(\mathcal{H}_0; x|\mathcal{R})$ may be understood as the degree to which the data would evoke surprise were the hypothesis that $\theta \in \mathcal{H}_0$ is known to be false. While that should not be confused with Shackle's degree of potential surprise in the revealed truth of a hypothesis

38

(Shackle, 1961), the concepts share many properties at the mathematical level.

Those relationships may be succinctly expressed in terms of possibility theory and ranking theory, the successors of the the theory of potential surprise:

1. Possibility theory. A function $\text{Poss} : 2^{\Theta} \to [0,1]$ is a *possibility measure* on 2^{Θ} if $\text{Poss}(\emptyset) = 0$, $\text{Poss}(\Theta) = 1$, and $\text{Poss}\left(\bigcup_{j \in \mathcal{J}} \mathcal{H}_{0j}\right) = \sup_{j \in \mathcal{J}} \text{Poss}(\mathcal{H}_{0j})$ for any index set $\mathcal{J}$ such that $\bigcup_{j \in \mathcal{J}} \mathcal{H}_{0j} \in 2^{\Theta}$ and $\mathcal{H}_{0j} \in 2^{\Theta}$ for all $j \in \mathcal{J}$ (Wang, 2008, §4.6). Further, a function $\pi : \Theta \to [0,1]$ such that $\text{Poss}(\mathcal{H}) = \sup_{\theta \in \Theta} \pi(\theta)$ is called a *possibility profile*, and a function $\text{Nec} : 2^{\Theta} \to [0,1]$ is a *necessity measure* on 2^{Θ} if $\text{Nec}(\mathcal{H}) = 1 - \text{Poss}(\mathcal{H}')$ for all $\mathcal{H} \in 2^{\Theta}$ (Wang, 2008, §4.6). Thus, $C(\mathcal{H}_0; x|\mathcal{R}) = \sup_{\theta_0 \in \mathcal{H}_0 \cap \mathcal{R}} c(\theta_0; x|\mathcal{R})$ as a function of $\mathcal{H}_0$ is a possibility measure corresponding to the possibility profile $c(\bullet; x|\mathcal{R})$. Similarly, in view of Definition 3, $W(\mathcal{H}_0; x|\mathcal{R})$ as a function of $\mathcal{H}_0$ is a necessity measure.

2. Ranking theory. If Poss is a possibility measure, then $-\log \text{Poss}(\mathcal{H}_0)$ as a function of $\mathcal{H}_0$ is a *negative ranking function* (Spohn, 2012, §11.8). It follows that

$$\text{Rank}(\mathcal{H}_0) = \log \frac{\text{Poss}(\mathcal{H}_0)}{\text{Poss}(\mathcal{H}_0')}$$

as a function of $\mathcal{H}_0$ is a *two-sided ranking function* (Spohn, 2012, §5.2). Both $-\log C(\mathcal{H}_0; x|\mathcal{R})$ and the potential surprise of $\mathcal{H}_0$ (Shackle, 1961) as functions of $\mathcal{H}_0$ are negative ranking functions. While $-\log C(\mathcal{H}_0; x|\mathcal{R})$ does not measure the potential surprise of learning that $\theta \in \mathcal{H}_0$, it might be seen as the level of surprise of observing that $X = x$ were it known that $\theta \in \mathcal{H}_0$, in accordance with the comments on surprise in Section 7. Since

39

$C\left(\bullet; x|\mathcal{R}\right)$ is a possibility measure and since

$$A\left(\mathcal{H}_0; x|\mathcal{R}\right) = \log \frac{C\left(\mathcal{H}_0; x|\mathcal{R}\right)}{C\left(\mathcal{H}_0'; x|\mathcal{R}\right)}$$

by equation (27), $A\left(\bullet; x|\mathcal{R}\right)$ qualifies mathematically as a conditional two-sided ranking function. However, the interpretation encoded in Definition 4 differs from that of Spohn (2012), who developed ranking theory to model degrees of belief.

The definition of conditional possibility used in the axiom of conditional compatibility (5) is not the only notion of conditional possibility, but it has desirable properties when possibility has quantitative information beyond mere ordering (e.g., Dubois and Prade, 1998; De Baets et al., 1999; Lapointe and Bobée, 2000; Marchioni, 2006). In that case, it is meaningful to say that a hypothesis of possibility value 0.9 is in some sense nine times as possible as a hypothesis of possibility value 0.1. By contrast, when possibility only indicates ordering, the two possibility values compared to each other indicates nothing more than that the hypothesis of possibility value 0.9 is more possible than the hypothesis of possibility value 0.1. Thus, the axiom of conditional compatibility enables us to say a hypothesis that has a data-compatibility value of 0.9 is nine times as compatible with the data observed as is a hypothesis that has a data-compatibility value of 0.1. That enables the use of data-hypothesis compatibility thresholds for hypothesis testing and interval estimation. That lack of quantitative information would render compatibility useless in hypothesis testing and set estimation. Equation (5) also ensures that conditional compatibility is a conditional idempotent probability, a powerful tool in the theory of large deviations (Puhalskii, 1997, 2001).

As precursors to this transformation of the compatibility function of a parameter into a

possibility measure, the p-value function of a parameter and the likelihood function had been transformed into possibility measures. When $\sup_{\theta_1 \in \Theta} p(\theta_1; x) = 1$, possibility theory provides useful interpretations of p values and confidence levels. First, Corollary 2 interprets the p value as the level of compatibility of the null hypothesis with the data or how possible the null hypothesis is in light of the data. Second, Theorem 3 interprets the confidence level as the degree of warrant for the hypothesis or how necessary its logical truth is given the data and the model. These interpretations in terms of possibility and necessity measures are related to previous work. Under broad conditions, the confidence-based methods of Mauris et al. (2001, §2.2), Dubois et al. (2004), Masson and Denœux (2006), and Ghasemi Hamed et al. (2012) likewise lead to interpreting p values as possibility values. Dubois et al. (1997) and Giang and Shenoy (2005) instead used the likelihood function in place of the p-value function $p(\bullet; x)$ for the special case in which Θ is countable. Patriota (2013) defines the s value, a large-sample possibility measure that uses both likelihood and confidence concepts. For confidence regions based on the likelihood ratio statistic, the proposed c value is equivalent to the s value under no restrictions over the parameter space.

Acknowledgments

References

Ball, F., Britton, T., O'Neill, P., 2002. Empty confidence sets for epidemics, branching processes and brownian motion. Biometrika 89 (1), 211–224.

Berger, J., 2003. Could Fisher, Jeffreys and Neyman have agreed on testing? Statistical Science 18 (1), 1–12.

Bickel, D. R., 2012. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. Statistica Sinica 22, 1147–1198.

Bickel, D. R., 2013a. Minimax-optimal strength of statistical evidence for a composite alternative hypothesis. International Statistical Review 81, 188–206.

Bickel, D. R., 2013b. Pseudo-likelihood, explanatory power, and Bayes's theorem [comment on "A likelihood paradigm for clinical trials"]. Journal of Statistical Theory and Practice 7, 178–182.

Bickel, D. R., Padilla, M., 2014. A prior-free framework of coherent inference and its derivation of simple shrinkage estimators. Journal of Statistical Planning and Inference 145, 204–221.

Chuaqui, R., 1991. Truth, Possibility and Probability: New Logical Foundations of Probability and Statistical Inference. North-Holland Mathematics Studies. Elsevier Science, Oxford.

Cohen, L., 1992. An Essay on Belief and Acceptance. Clarendon Press, Oxford.

Cox, D. R., 1977. The role of significance tests. Scandinavian Journal of Statistics 4, 49–70.

De Baets, B., Tsiporkova, E., Mesiar, R., 1999. Conditioning in possibility theory with strict order norms. Fuzzy Sets and Systems 106, 221–229.

Dubois, D., Foulloy, L., Mauris, G., Prade, H., 2004. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing 10 (4), 273–297.

Dubois, D., Moral, S., Prade, H., Jan. 1997. A semantics for possibility theory based on likelihoods. Journal of Mathematical Analysis and Applications 205 (2), 359–380.

Dubois, D., Prade, H., 1998. Possibility theory: Qualitative and quantitative aspects. In: Smets, P. (Ed.), Quantified Representation of Uncertainty and Imprecision. Vol. 1 of Handbook of Defeasible Reasoning and Uncertainty Management Systems. Springer Netherlands, pp. 169–226.

Edwards, A. W. F., 1992. Likelihood. Johns Hopkins Press, Baltimore.

Efron, B., Tibshirani, R., 1998. The problem of regions. Annals of Statistics 26, 1687–1718.

Fisher, R. A., 1973. Statistical Methods and Scientific Inference. Hafner Press, New York.

Fraser, D., Reid, N., Wong, A., 2004. Inference for bounded parameters. Physical Review D 69, 033002.

Fraser, D. A. S., 2011. Is Bayes posterior just quick and dirty confidence? Statistical Science 26, 299–316.

Gabriel, K. R., 1969. Simultaneous test procedures - some theory of multiple comparisons. Ann. Math. Statist. 40, 224–250.

Ghasemi Hamed, M., Serrurier, M., Durand, N., 2012. Representing uncertainty by possibility distributions encoding confidence bands, tolerance and prediction intervals. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7520, 233–246.

Giang, P. H., Shenoy, P. P., 2005. Decision making on the sole basis of statistical likelihood. Artificial Intelligence 165, 137–163.

Hacking, I., 2001. An introduction to probability and inductive logic. Cambridge University Press, Cambridge.

Jeffreys, H., 1948. Theory of Probability. Oxford University Press, London.

Kaplan, M., 1996. Decision Theory As Philosophy. Cambridge University Press, Cambridge.

Lapointe, S., Bobée, B., 2000. Revision of possibility distributions: A Bayesian inference pattern. Fuzzy Sets and Systems 116 (2), 119 – 140.

Lavine, M., Schervish, M. J., 1999. Bayes factors: What they are and what they are not. American Statistician 53, 119–122.

Mandelkern, M., 2002a. Setting confidence intervals for bounded parameters. Statistical Science 17, 149–172.

Mandelkern, M., 2002b. Setting confidence intervals for bounded parameters. Statistical Science 17, 149–159.

Marchand, É., Strawderman, W., 2013. On bayesian credible sets, restricted parameter spaces and frequentist coverage. Electronic Journal of Statistics 7 (1), 1419–1431.

Marchand, É., Strawderman, W. E., 2004. Estimation in restricted parameter spaces: A review. Lecture Notes-Monograph Series 45, 21–44.

Marchioni, E., 2006. Possibilistic conditioning framed in fuzzy logics. International Journal of Approximate Reasoning 43 (2), 133–165.

Masson, M.-H., Denœux, T., 2006. Inferring a possibility distribution from empirical data. Fuzzy Sets and Systems 157 (3), 319–340.

Mauris, G., Lasserre, V., Foulloy, L., Apr. 2001. A fuzzy approach for the expression of uncertainty in measurement. Measurement 29 (3), 165–177.

Nadarajah, S., Bityukov, S., Krasnikov, N., 2015. Confidence distributions: A review. Statistical Methodology 22, 23–46.

Patriota, A. G., 2013. A classical measure of evidence for general null hypotheses. Fuzzy Sets and Systems 233, 74 – 88.

Peirce, C., 1998. The Essential Peirce: Selected Philosophical Writings (1893-1913). The Essential Peirce: Selected Philosophical Writings. Indiana University Press.

Polansky, A. M., 2007. Observed Confidence Levels: Theory and Application. Chapman and Hall, New York.

Puhalskii, A., 1997. Large deviations of semimartingales: a maxingale problem approach i. limits as solutions to a maxingale problem. Stochastics: An International Journal of Probability and Stochastic Processes 61 (3-4), 141–243.

Puhalskii, A., 2001. Large Deviations and Idempotent Probability. Monographs and Surveys in Pure and Applied Mathematics. CRC Press, New York.

Royall, R., 1997. Statistical Evidence: A Likelihood Paradigm. CRC Press, New York.

Schervish, M. J., 1996. P values: What they are and what they are not. American Statistician 50, 203–206.

Schweder, T., Hjort, N. L., 2002. Confidence and likelihood. Scandinavian Journal of Statistics 29, 309–332.

Shackle, G., 1961. Decision, Order and Time in Human Affairs. Cambridge University Press, Cambridge.

Silvapulle, M., Sen, P., 2011. Constrained Statistical Inference: Order, Inequality, and Shape Constraints. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.

Spohn, W., 2012. The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press.

van Dyk, D. A., 2014. The role of statistics in the discovery of a higgs boson. Annual Review of Statistics and Its Application 1, 41–59.

Wang, H., 2004. Improved estimation of accuracy in simple hypothesis versus simple alternative testing. Journal of Multivariate Analysis 90 (2), 269 – 281.

Wang, H., 2006. Modified p-value of two-sided test for normal distribution with restricted parameter space. Communications in Statistics-Theory and Methods 35 (8), 1361–1374.

Wang, H., 2007. Modified p-values for one-sided testing in restricted parameter spaces. Statistics and Probability Letters 77, 625–631.

Wang, Q., 2008. Probability distribution and entropy as a measure of uncertainty. Journal of Physics A: Mathematical and Theoretical 41, 065004.

Wendell, J. P., Schmee, J., 1996. Exact inference for proportions from a stratified finite population. Journal of the American Statistical Association 91 (434), pp. 825–830.

Xie, M.-G., Singh, K., 2013. Confidence distribution, the frequentist distribution estimator of a parameter: A review. International Statistical Review 81 (1), 3–39.

Zhang, T., Woodroofe, M., 2003. Credible and confidence sets for restricted parameter spaces. Journal of Statistical Planning and Inference 115, 479–490.

Zhang, Z., Zhang, B., 2013. A likelihood paradigm for clinical trials. Journal of Statistical Theory and Practice 7, 157–177.