*Year* 2012 *Paper* 95

# A prior-free framework of coherent inference and its derivation of simple shrinkage estimators

David R. Bickel*

*Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, dbickel@uottawa.ca

# A prior-free framework of coherent inference and its derivation of simple shrinkage estimators

David R. Bickel

**Abstract**

The reasoning behind uses of confidence intervals and p-values in scientific practice may be made coherent by modeling the inferring statistician or scientist as an idealized intelligent agent. With other things equal, such an agent regards a hypothesis coinciding with a confidence interval of a higher confidence level as more certain than a hypothesis coinciding with a confidence interval of a lower confidence level. The agent uses different methods of confidence intervals conditional on what information is available. The coherence requirement means all levels of certainty of hypotheses about the parameter agree with the same distribution of certainty over parameter space. The result is a unique and coherent fiducial distribution that encodes the post-data certainty levels of the agent.

While many coherent fiducial distributions coincide with confidence distributions or Bayesian posterior distributions, there is a general class of coherent fiducial distributions that equates the two-sided p-value with the probability that the null hypothesis is true. The use of that class leads to point estimators and interval estimators that can be derived neither from the dominant frequentist theory nor from Bayesian theories that rule out data-dependent priors. These simple estimators shrink toward the parameter value of the null hypothesis without relying on asymptotics or on prior distributions.

# A prior-free framework of coherent inference and its derivation of simple shrinkage estimators

June 28, 2012

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

**Abstract**

The reasoning behind uses of confidence intervals and p-values in scientific practice may be made coherent by modeling the inferring statistician or scientist as an idealized intelligent agent. With other things equal, such an agent regards a hypothesis coinciding with a confidence interval of a higher confidence level as more certain than a hypothesis coinciding with a confidence interval of a lower confidence level. The agent uses different methods of confidence intervals conditional on what information is available. The coherence requirement means all levels of certainty of hypotheses about the parameter agree with the same distribution of certainty over parameter space. The result is a unique and coherent fiducial distribution that encodes the post-data certainty levels of the agent.

While many coherent fiducial distributions coincide with confidence distributions or Bayesian posterior distributions, there is a general class of coherent fiducial distributions that equates the two-sided p-value with the probability that the null hypothesis is true. The use of that class leads to point estimators and interval estimators that can be derived neither from the dominant frequentist theory nor from Bayesian theories that rule out data-dependent priors. These simple estimators shrink toward the parameter value of the null hypothesis without relying on asymptotics or on prior distributions.

**Keywords:** confidence distribution; confidence curve; confidence measure; confidence posterior distribution; fiducial inference; large-scale simultaneous inference; multiple hypothesis testing; multiple comparison procedure; observed confidence level

# 1 Introduction

In the years following the oracle that some form of fiducial inference may play a pivotal role in 21st-century statistics (Efron, 1998), there has been an ongoing resurgence of interest in fiducial distributions that generate confidence intervals (e.g., Schweder and Hjort, 2002; Singh et al., 2005; Polansky, 2007; Singh et al., 2007; Tian et al., 2011; Bityukov et al., 2011; Kim and Lindsay, 2011; Bickel, 2011b, 2012b) and in fiducial distributions more generally (e.g., Hannig et al., 2006; Hannig, 2009; Xiong and Mu, 2009; Gibson et al., 2011; Wang et al., 2012; Zhao et al., 2012). Fiducial inference initially promised an objective alternative to Bayesianism as a form of inductive reasoning (Fisher, 1973) but has historically suffered from problems of understanding the meaning of fiducial probability and from the ability to derive conflicting fiducial probabilities from the same family of sampling distributions (see Wilkinson, 1977). This paper addresses both difficulties by interpreting fiducial probability in terms of the theories of coherent decision making that also undergird Bayesian inference.

The main thesis is that many of the usual applications of confidence intervals in science lead to reasonable inferences that can be improved by enforcing self-consistency in the technical sense of probabilistic coherence, which does not in itself require Bayesian posterior distributions (Hacking, 1967; Goldstein, 1997; Bickel, 2012a). Using a confidence interval procedure is reasonable when the confidence level of the interval estimate computed using the observed data is at least approximately monotonic with the degree of certainty or level of belief that the statistician has in saying the true value of the parameter lies in the interval (Cox, 1958). In other words, higher confidence levels correspond to higher subjective levels of certainty of the statistician adopting the confidence procedure; otherwise, a different procedure should be adopted in the absence of other considerations. If consistent with one another, the certainty levels of that statistician can be encoded as a probability distribution on parameter space known as a confidence distribution. If the same statistician would reasonably and self-consistently use another confidence procedure for another parameter in the data analysis, the levels of certainty of the first parameter can still be represented as a probability distribution, this time a fiducial distribution that need not be a confidence distribution. The situation described here is abstracted by replacing the actual statistician with an artificially intelligent agent that either approximates the certainty levels of the actual decision-makers or that serves to derive the hypothetical consequences of adopting its fiducial distributions for statistical inference.

The metaphor of a decision-making agent that has a unique fiducial distribution for any data set leads to coherent hypothesis tests, point estimates, and interval estimates without the requirement of eliciting the actual levels of belief of any human agent. Since the coherent agent is fully determined by choices of confidence interval and hypothesis testing procedures, the subjectivity involved is no greater than that already present in frequentist inference. While some likelihoodists have criticized frequentism for even that subjectivity (Royall, 1997, §3.7), the subjectivity involved in selecting the rejection region for significance testing coheres with post-positivistic philosophies of science that frankly acknowledge that scientific inference is not a matter of following an algorithm (Polanyi, 1962, §3.1).

Section 2 provides preliminary concepts and propositions, demonstrating that interpreting confidence levels as certainty levels or hypothetical levels of belief leads either to non-coherent estimates and hypothesis testing or to inference on the basis of a confidence distribution of the parameter as if it were a Bayesian posterior distribution. Iterating that reasoning along the lines of Fisher's fiducial argument for multiple parameters leads to merging confidence distributions into a parameter distribution that is coherent in the sense that it is a probability measure. This is fiducial inference in the sense that it is a modern development of fiducial reasoning but without the often impractical requirements involving aspects of conditional inference and without violating the rules of ordinary probability theory, that of the Kolmogorov axioms. The framework proposed in Section 2 also differs from Fisher's in its incorporation of nested confidence sets of vector parameters. Thus, the proposed framework for inference is presented as a realization of the core ideas behind the original fiducial argument, Neyman-Pearson confidence intervals, and theories of coherent decision-making that prescribe minimizing expected loss with respect to a posterior distribution (e.g., von Neumann and Morgenstern, 1953; Savage, 1954). (Following the usage in Dempster (2008), Eaton and Sudderth (2010), and Bickel (2012a), the term "posterior" herein means data-dependent and thus includes but is not limited to a Bayesian posterior relative to some prior.)

Section 3 demonstrates that the resulting framework of fiducial inference can lead to shrinkage in point and interval estimates toward a null hypothesis value in a way that is not possible in the pure frequentist and pure Bayesian approaches. For example, Figure 1 displays the shrunken parameter estimate as an alternative to the usual frequentist estimate computed after testing the null hypothesis. Given the two-sided p-value PV, the maximum-likelihood estimate $\hat{\theta}$ is simply shrunk to $(1 - \mathrm{PV})\,\hat{\theta}$. That value would only be available from Bayes's theorem if the prior depended on the sample size such that the posterior probability of the null hypothesis were equal to PV.

Lastly, remarks elaborating on technical points appear in Section 4, a brief discussion on equating p-values with fiducial probabilities in Section 5, and longer proofs in Appendix A.

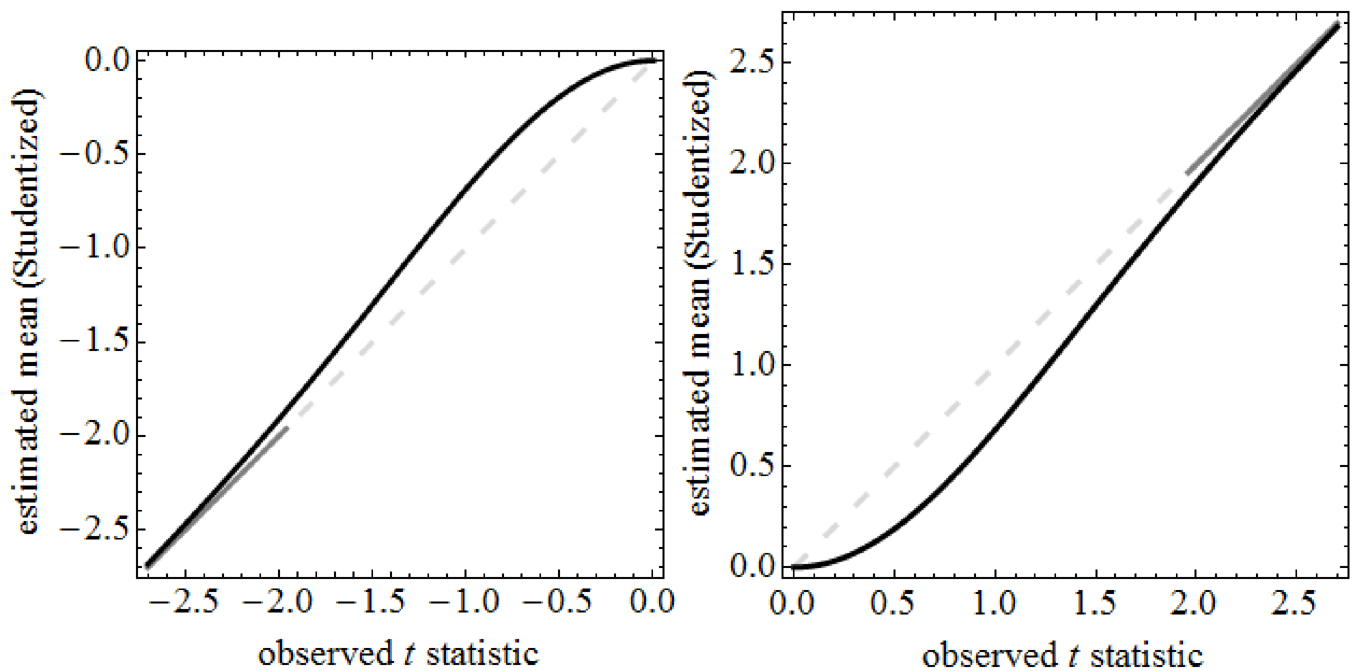


Figure 1: Estimates of the normal mean relative to its standard error as a function of the observed number of standard errors from 0, the null hypothesis value. The black curve is the posterior mean with respect to the fiducial distribution, and the gray line is the MLE, plotted as a solid line wherever the null hypothesis is rejected at the 5% significance level and as a dashed line elsewhere. See Example 8 for details.

# 2 Fiducial distributions

The concept of a fiducial distribution will be introduced in order to ground coherent decision making in the procedure of confidence intervals or more general confidence sets. In this way, the coherence condition will be supplemented with a confidence-based condition in order to prescribe point estimates, interval estimates, hypothesis tests, and other actions that minimize expected loss. The various types of fiducial distributions are formulated as frequentist posteriors: the basic fiducial distribution is defined in Section 2.1, and other fiducial distributions are defined in Section 2.2.

## 2.1 Basic fiducial distributions

### 2.1.1 Fiducial probability as coherent confidence

The basic parameter $\theta$ and nonbasic parameter $\gamma$ are in the parameter sets denoted by $\Theta$ and $\Gamma$, respectively. The distinction between the basic and nonbasic parameters will become clear shortly. For now, it is enough to note that which parameter is basic cannot be a function of which parameter happens to be of interest provided that the background (pre-data) knowledge of the hypothetical agent is fixed (Remark 1).

The observed $n$-tuple $x$ is a member of $\mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^n$. Let $\mathcal{B}(\mathcal{X})$ denote the $\sigma$-field of Borel subsets of $\mathcal{X}$. The family of distributions of the random variable $X$ of outcome $x$ is $\{P_{\theta,\gamma} : \theta \in \Theta, \gamma \in \Gamma\}$, where each $P_{\theta,\gamma}$ is defined on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. The set of all closed interval subsets of $[0, 1]$ will be denoted by $\mathfrak{I}$. Herein, $\subset$ (as opposed to $\subseteq$) means "is a proper subset of." A function $\widehat{\Theta} = \widehat{\Theta}_{\bullet}(\bullet)$ on $\mathcal{X} \times \mathfrak{I}$ is a *procedure of nested confidence sets* for $\theta$ if there is a function $p = p_{\bullet}(\bullet) : \mathcal{X} \times \Theta \to [0, 1]$, such that, for any

$\mathcal{I} \in \mathfrak{I}$,

$$\widehat{\Theta}_x(\mathcal{I}) = \{\theta \in \Theta : p_x(\theta) \in \mathcal{I}\} \tag{1}$$

for all $x \in \mathcal{X}$ and such that the corresponding *nested confidence set estimator* $\widehat{\Theta}_\bullet(\mathcal{I})$ on $\mathcal{X}$ satisfies

$$P_{\theta,\gamma}\left(\theta \in \widehat{\Theta}_X(\mathcal{I})\right) = |\mathcal{I}| \tag{2}$$

for all $\theta \in \Theta$ and $\gamma \in \Gamma$, where $|\bullet|$ is the Lebesgue measure. (Since $\mathcal{I}$ is an interval in this case, $|\mathcal{I}|$ is the width of $\mathcal{I}$.) As a result, $|\mathcal{I}|$ is called the *confidence level* of $\widehat{\Theta}_\bullet(\mathcal{I})$, and $p$ is called the *confidence curve* of $\widehat{\Theta}$ (Birnbaum, 1961; Blaker, 2000).

**Lemma 1.** *If $\widehat{\Theta}$ is the procedure of nested confidence sets for $\theta$ that is defined by some confidence curve $p$, then $p_X(\theta)$ is uniformly distributed between 0 and 1 ($p_X(\theta) \sim \mathrm{U}(0,1)$) for all $\theta \in \Theta$.*

*Proof.* By the definitions of a confidence level and a procedure of nested confidence sets, formulas (1)-(2) yield

$$P_{\theta,\gamma}(p_X(\theta) \in \mathcal{I}) = |\mathcal{I}| \tag{3}$$

for all $\mathcal{I} \in \mathfrak{I}$, $\theta \in \Theta$, and $\gamma \in \Gamma$. Thus, using $\mathcal{I} = [\alpha, 1]$ for any $\alpha \in [0,1]$,

$$P_{\theta,\gamma}(p_X(\theta) \geq \alpha) = P_{\theta,\gamma}(p_X(\theta) \in [\alpha, 1]) = 1 - \alpha.$$

□

Lemma 1 implies that $p_x(\theta_0)$ is the observed p-value for testing the null hypothesis that $\theta = \theta_0$ against alternative hypothesis that $p_X(\theta_0)$ is stochastically less than $\mathrm{U}(0,1)$. For that reason, $p_x(\bullet)$ has usually been called a *p-value function* or a *significance function* in the case of a scalar basic parameter ($\Theta \subseteq \mathbb{R}^1$) (e.g., Fraser, 1991).

Confidence sets used in practice are typically interpreted such that the confidence levels have the same order as the levels of certainty a statistician or scientist would place on the hypotheses that the parameter value is within the confidence sets. To state this formally for a procedure $\widehat{\Theta}$ of nested set estimators, let $\mathcal{H}_x\left(\widehat{\Theta}\right)$ denote the set of nested confidence sets corresponding to $x \in \mathcal{X}$:

$$\mathcal{H}_x\left(\widehat{\Theta}\right) = \left\{\widehat{\Theta}_x\left(\mathcal{I}\right) : \mathcal{I} \in \mathfrak{I}\right\}. \tag{4}$$

Given any two observations $x_1, x_2 \in \mathcal{X}$ and any two parameter subsets $\Theta_1 \in \mathcal{H}_{x_1}\left(\widehat{\Theta}\right)$ and $\Theta_2 \in \mathcal{H}_{x_2}\left(\widehat{\Theta}\right)$, the hypothesis that $\theta \in \Theta_1$ is considered no more certain than ($\preceq$) the hypothesis that $\theta \in \Theta_2$ if and only if the highest confidence level corresponding to the former hypothesis is less than or equal to that corresponding to the latter:

$$\Theta_1 \preceq \Theta_2 \iff \sup\left|\widehat{\Theta}_{x_1}^{-1}\left(\Theta_1\right)\right| \leq \sup\left|\widehat{\Theta}_{x_2}^{-1}\left(\Theta_2\right)\right|. \tag{5}$$

The parameter $\theta$ that defines those hypotheses is called the *basic parameter*.

**Definition 1.** Let $\sigma_x$ denote any $\sigma$-field such that $\mathcal{H}_x\left(\widehat{\Theta}\right) \subset \sigma_x$. For any $x \in \mathcal{X}$, a probability measure $C_x$ on $(\Theta, \sigma_x)$ is a *certainty distribution* that is compatible with a confidence procedure $\widehat{\Theta}$ and with its confidence curve $p$ if

$$C_{x_1}\left(\Theta_1\right) \leq C_{x_2}\left(\Theta_2\right) \iff \Theta_1 \preceq \Theta_2 \tag{6}$$

for all $x_1, x_2 \in \mathcal{X}$; $\Theta_1 \in \mathcal{H}_{x_1}\left(\widehat{\Theta}\right)$; $\Theta_2 \in \mathcal{H}_{x_2}\left(\widehat{\Theta}\right)$.

Other fiducial distributions will be defined in Section 2.2.

**Example 1.** Consider the spherically normal model:

$$X \sim \mathrm{N}\left(\xi\theta, \gamma^2 \mathbf{I}\right),$$

where $X$ is a random column vector of $n$ observable responses, $\theta \in \mathbb{R}^d$ is a column vector of $d < n$ unknown means, $\xi$ is a $n \times d$ design matrix, $\gamma$ is the unknown standard deviation, and $\mathbf{I}$ is the $d \times d$ identity matrix. Thus, $x$ is the fixed column vector of $n$ observed responses. Let $\widehat{\theta}(x)$ and $\widehat{\gamma}(x)$ denote the maximum likelihood estimates of $\theta$ and $\gamma$, respectively, and let $C_x$ be the multivariate $t$ distribution of location $d$-vector $\widehat{\theta}$, scale matrix $\widehat{\gamma}^2(x)\,\xi^{\mathrm{T}}\xi$, and $n-d$ degrees of freedom. If $\vartheta$ is the random variable with distribution $C_x$, i.e., $\vartheta \sim C_x$, then

$$\frac{\left(\vartheta - \widehat{\theta}(x)\right)^{\mathrm{T}} \xi^{\mathrm{T}}\xi \left(\vartheta - \widehat{\theta}(x)\right)}{\widehat{\gamma}^2(x)\, d} \tag{7}$$

is $F_{d,n-d}$, the random variable distributed as the $F$-distribution with $\langle d, n-d \rangle$ degrees of freedom (Box and Tiao, 1992, §2.7.2). Let $c_x$ denote the probability density function equal to the Radon-Nikodym derivative of $C_x$ with respect to the Lebesgue measure. If $\widehat{\Theta}_x$ is defined by the density contours such that $C_x\left(\vartheta \in \widehat{\Theta}_x(\mathcal{I})\right) = |\mathcal{I}|$ and

$$\theta_1 \notin \widehat{\Theta}_x(\mathcal{I}), \theta_2 \in \widehat{\Theta}_x(\mathcal{I}) \implies c_x(\theta_1) < c_x(\theta_2)$$

for all $\mathcal{I} \in \mathfrak{I}$, then $\widehat{\Theta}_X(\mathcal{I})$ is a $100\,|\mathcal{I}|\,\%$ confidence region in the sense that it satisfies formula (2) (Box and Tiao, 1992, §2.9.0). According to formula (6), $C_x$ is a certainty distribution that is compatible with $\widehat{\Theta}$. ▲

The procedure of nested set estimators also provides a general concept of a confidence distribution.

**Definition 2.** For any $x \in \mathcal{X}$, a probability measure $K_x$ on $(\Theta, \sigma_x)$ is a *confidence distribution* that is compatible with $\widehat{\Theta}$ if, for every $\Theta_1 \in \mathcal{H}_x\left(\widehat{\Theta}\right)$,

$$K_x\left(\Theta_1\right) \in \mathcal{K}_x\left(\Theta_1\right), \tag{8}$$

where $\mathcal{K}_x\left(\Theta_1\right) = \left\{|\mathcal{I}| : P_{\theta,\gamma}\left(\theta \in \widehat{\Theta}_X\left(\mathcal{I}\right)\right) = |\mathcal{I}|, \mathcal{I} \in \mathfrak{I}, \widehat{\Theta}_x\left(\mathcal{I}\right) = \Theta_1\right\}$.

The definition specified by formula (8) extends the usual confidence distribution of a scalar parameter defined on the basis of strictly nested confidence intervals (Cox, 1958) to confidence distributions of higher-dimensional basic parameters defined on the basis of confidence sets that could have $\widehat{\Theta}_x\left(\mathcal{I}_1\right) = \widehat{\Theta}_x\left(\mathcal{I}_2\right)$ for some $\mathcal{I}_1 \neq \mathcal{I}_2$. In the former case, $\sigma_x$ is the Borel field over $\Theta$. Polansky (2007), Singh et al. (2007), and Bickel (2011b, 2012a) present alternative definitions of confidence distributions of vector basic parameters. The definition used here is a slight generalization of the "confidence posterior" found in Bickel (2012b, §2.3).

The simplest type of fiducial distribution is a special case of a confidence distribution.

**Definition 3.** For any $x \in \mathcal{X}$, a probability measure $\Pi_x$ on $(\Theta, \sigma_x)$ is a *basic fiducial distribution* that is compatible with $\widehat{\Theta}$ if, for every $\Theta_1 \in \mathcal{H}_x\left(\widehat{\Theta}\right)$,

$$\Pi_x\left(\Theta_1\right) \geq P_{\theta,\gamma}\left(\theta \in \widehat{\Theta}_X\left(\mathcal{I}\right)\right) \tag{9}$$

for all $\mathcal{I} \in \mathfrak{I}$ such that $\widehat{\Theta}_x\left(\mathcal{I}\right) = \Theta_1$.

Formulas (8) and (9) are related by $\Pi_x\left(\Theta_1\right) = \sup \mathcal{K}_x\left(\Theta_1\right)$. In Example 1, $C_x$ is a basic fiducial distribution as well as a certainty distribution. The inequality of formula (9) essentially follows van Berkum et al. (1996); see also Bickel (2012b,d) and references.

Definition 3 sheds light on the relationship between the concepts of a certainty distribution and a basic fiducial distribution. Every basic fiducial distribution is necessarily a certainty distribution, as is clear from fact that formulas (5) and (9) imply formula (6).

The converse is not necessarily true, but satisfaction of a condition usually met in practice is sufficient for a certainty distribution $C_x$ to be a basic fiducial distribution. The confidence procedure $\widehat{\Theta}$ is said to be *potentially invertible* if, for any $\epsilon > 0$, there are an $x \in \mathcal{X}$ and a $\mathfrak{I}_x \subseteq \mathfrak{I}$ that satisfy

$$\left| \cup_{\mathcal{I} \in \mathfrak{I}_x} \mathcal{I} \right| \geq 1 - \epsilon \tag{10}$$

such that the function $\widetilde{\Theta}_x(\bullet) : \mathfrak{I}_x \to \Theta_x$ is invertible (bijective) for some $\Theta_x \subseteq \Theta$, where $\widetilde{\Theta}_x(\mathcal{I}) = \widehat{\Theta}_x(\mathcal{I})$ for all $\mathcal{I} \in \mathfrak{I}_x$. The condition is trivially met when, as in Example 1, $\widehat{\Theta}_x$ is bijective for all $x \in \mathcal{X}$, since in that case $\left| \cup_{\mathcal{I} \in \mathfrak{I}_x} \mathcal{I} \right| = 1$ in formula (10) with $\mathfrak{I}_x = \mathfrak{I}$ and $\widetilde{\Theta}_x = \widehat{\Theta}_x$ for all $x \in \mathcal{X}$ and $\mathcal{I} \in \mathfrak{I}$. The next example illustrates this non-trivial but commonly applicable result:

**Theorem 1.** *With* $\Theta$ *as any interval such that* $\sup \Theta = \infty$*, let* $\widehat{\Theta}$ *be a procedure of nested confidence intervals for* $\theta \in \mathbb{R}$ *that is defined by some confidence curve* $p$*. If* $p_x(\bullet)$ *is a strictly increasing and continuous function such that* $\lim_{\theta \to \infty} p_x(\theta) = 1$ *for all* $x \in \mathcal{X}$*, then* $\widehat{\Theta}$ *is potentially invertible.*

*Proof.* By Lemma 1, $p_X(\theta) \sim \mathrm{U}(0, 1)$ for all $\theta \in \Theta$. It follows that, for any $\epsilon > 0$, there is an $x \in \mathcal{X}$ such that $\lim_{\theta \to \inf \Theta} p_x(\theta) < \epsilon$. For any such $\epsilon$ and $x$, let $\mathfrak{I}_x$ denote the set of all closed interval subsets of $[\epsilon, 1]$. Since $\left| \cup_{\mathcal{I} \in \mathfrak{I}_x} \mathcal{I} \right| = 1 - \epsilon$, inequality (10) clearly holds. The invertibility of $\widetilde{\Theta}_x(\bullet)$ is a consequence of the stated assumption that $p_x(\bullet)$ is strictly increasing and continuous. $\square$

**Example 2.** The observable vector $X$ consists of $n$ independent random variables of distribution $\mathrm{N}(\mu, \sigma^2)$ with $\mu$ and $\sigma$ unknown. Let $\upsilon(\bullet) : \mathcal{X} \times \mathbb{R} \to [0, \infty[$ and $\tau(\bullet) : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ denote the functions such that

$$\upsilon(x; \mu) = \tau^2(x; \mu);$$

$$\tau(x; \mu) = \frac{\hat{\mu}(x) - \mu}{\hat{\sigma}(x) / \sqrt{n}},$$

which is the observed Student $t$ statistic with $\mu$ as the null hypothesis value, for any $x \in \mathcal{X}$. If the basic parameter is $\theta = \mu^2$, then the nonbasic parameter is the pair $\gamma = (\mu/|\mu|, \sigma)$, and $\tau(X; \mu)$ is a pivotal quantity with the Student $t$ distribution of $n-1$ degrees of freedom, implying that $\upsilon(X; \mu) = F_{1,n-1}$. (The ratio $\mu/|\mu|$ is the sign of $\mu$.) A confidence procedure $\widehat{\Theta}$ can then be constructed by defining the confidence curve $p$ according to the upper-tailed p-value

$$p_x(\theta) = P_{\theta,\gamma}(\upsilon(X; 0) \geq \upsilon(x; 0)) \tag{11}$$

for all $x \in \mathcal{X}$, $\theta \geq 0$, and $\gamma \in \{-1, 0, 1\} \times ]0, \infty[$. Because $\Theta = [0, \infty[$ and because formula (11) implies that $p_x(\bullet)$ is strictly increasing and continuous for all $x \in \mathcal{X}$, the conditions of Theorem 1 are met even though $\widehat{\Theta}_X(\bullet)$ is almost surely not invertible. ▲

**Theorem 2.** *Let $\widehat{\Theta}$ be a procedure of nested confidence sets for $\theta$ that is defined by some confidence curve $p$. If $\widehat{\Theta}$ is potentially invertible, then every certainty distribution compatible with $\widehat{\Theta}$ is also a basic fiducial distribution that is compatible with $\widehat{\Theta}$.*

### 2.1.2 P-values as hypothesis probabilities

The next result provides sufficient conditions for equating the certainty level of a simple (point) null hypothesis with a p-value.

**Corollary 1.** *Let $\widehat{\Theta}$ be a procedure of nested confidence intervals for $\theta$ that is defined by some confidence curve $p$. Under the conditions of Theorem 2, every certainty distribution $C_x$ compatible with $\widehat{\Theta}$ is also a basic fiducial distribution that is compatible with $\widehat{\Theta}$ and, if $\Theta = [\theta_0, \infty[$ for some $\theta_0 \in \mathbb{R}$, then, for all $x \in \mathcal{X}$,*

$$C_x(\vartheta = \theta_0) = p_x(\theta), \tag{12}$$

*where $\vartheta \sim C_x$, i.e., $\vartheta$ is the random variable of distribution $C_x$.*

**Example 3.** Example 2, continued. Since the conditions of Corollary 1 are satisfied, the certainty level of the hypothesis that the parameter value equals zero is equal to the p-value of the test with $\theta = 0$ as the null hypothesis:

$$C_x(\vartheta = 0) = p_x(0) = P_{0,\gamma}(\upsilon(X; 0) \geq \upsilon(x; 0)) \tag{13}$$

for all $x \in \mathcal{X}$ and $\gamma \in \{-1, 0, 1\} \times ]0, \infty[$. That is simply the usual two-sided p-value from the single-sample $t$-test, as equation (11) makes clear. ▲

In conclusion, since Bayesian posterior probabilities are typically not equal to two-sided p-values, Corollary 1 prevents certainty theory from being regarded as a special case of Bayesian theory (§5).

Some operating characteristics of testing hypotheses under the equality of the p-value and the certainty level appear in the remainder of this subsection. They do not in themselves warrant the use of the fiducial distribution but rather report some of its repeated-sampling properties. Here, $\delta_{i,j}$ is Kronecker's delta: $\delta_{\theta_0,\theta_0} = 1$ and $\delta_{\theta,\theta_0} = 0$ for $\theta \neq \theta_0$.

**Theorem 3.** *Consider the null hypothesis that $\theta = \theta_0$ for some $\theta_0 \in \Theta$. For a Type I error* *cost* $\ell_I > 0$ *and a Type II error cost* $\ell_{II} > 0$*, the loss function* $L : \Theta \times \{0,1\} \rightarrow \{0, \ell_I, \ell_{II}\}$ *is defined by* $L(\theta, 0) = (1 - \delta_{\theta,\theta_0}) \ell_{II}$ *and* $L(\theta, 1) = \delta_{\theta,\theta_0} \ell_I$*. If the action* $a(x) \in \{0,1\}$ *is chosen to minimize expected loss with respect to a certainty distribution* $C_x$ *that is compatible with a confidence curve* $p$ *and that meets the criteria of Corollary 1 for all* $x \in \mathcal{X}$*, then*

$$a(x) = \begin{cases} 1 & \text{if } p_x(\theta_0) < \alpha(\ell) \\ 0 & \text{if } p_x(\theta_0) > \alpha(\ell), \end{cases} \tag{14}$$

where $\ell = \ell_{\mathrm{I}}/\ell_{\mathrm{II}}$, and $\alpha(\ell) = (1 + \ell)^{-1}$ is the Type I error rate of $a(X)$.

*Proof.* It is known that some algebra leads to

$$a(x) = \arg\min_{b=0,1} \int L(\theta, b)\, d\Pi_x(\theta) = \begin{cases} 1 & \text{if } C_x(\vartheta = \theta_0) < \alpha(\ell) \\ 0 & \text{if } C_x(\vartheta = \theta_0) > \alpha(\ell) \end{cases}$$

for any parameter distribution $\Pi_x$. Corollary 1 implies that $C_x(\vartheta = \theta_0) = p_x(\theta_0)$, which leads to equation (14) by substitution. Because $p_X(\theta_0) \sim \mathrm{U}(0,1)$ by Lemma 1 under $\theta = \theta_0$, the Type I error rate is

$$P_{\theta_0,\gamma}(a(X) = 1) = P_{\theta_0,\gamma}(p_X(\theta) < \alpha(\ell)) = \alpha(\ell). \tag{15}$$

□

For instance, equation (14) implies that the practice of rejecting the null hypothesis at the 0.05 significance level would be appropriate if $\ell = 19$, i.e., if the cost of a Type I error were 19 times as much as the cost of a Type II error.

**Corollary 2.** If $\theta = \theta_0$ in addition to the assumptions of Theorem *3*, then the loss $L(\theta, a(X))$ averaged over the sample space is

$$\int L(\theta_0, a(x))\, dP_{\theta_0,\gamma}(x) = \frac{\ell_I}{1+\ell}. \quad \left[\ell = \frac{\ell_I}{\ell_{II}}\right]$$

*Proof.* By equation (14),

$$\begin{aligned}\int L(\theta_0, a(x))\, dP_{\theta_0,\gamma}(x) &= P_{\theta_0,\gamma}(a(X)=0)\, L(\theta,0) + P_{\theta_0,\gamma}(a(X)=1)\, L(\theta,1) \\ &= P_{\theta_0,\gamma}(a(X)=0)\,(1-\delta_{\theta_0,\theta_0})\,\ell_{\mathrm{II}} + P_{\theta_0,\gamma}(a(X)=1)\,\delta_{\theta_0,\theta_0}\ell_{\mathrm{I}} \\ &= 0 + P_{\theta_0,\gamma}(a(X)=1)\,\ell_{\mathrm{I}}.\end{aligned}$$

According to equation (15), the first factor of the right-hand-side is $\alpha(\ell)$. □

Since the loss function of Theorem 3 may be less applicable when a p-value is reported as a measure of evidence rather than compared to a fixed significance level, quadratic loss of the p-value as a point estimator of a hypothesis truth value is often considered (Bickel, 2012a). In this context, Hwang et al. (1992) and Morgenthaler and Staudte (2005) find that the p-value is not necessarily admissible under the frequentist decision theory of Wald (1961). However, the next theorem indicates that the p-value is often optimal according to theories of minimizing expected loss with respect to the agent's parameter distribution (e.g., Savage, 1954). Its repeated-sampling performance under the null hypothesis is quantified in the corollary.

**Theorem 4.** *Consider the null hypothesis that $\theta = \theta_0$ for some $\theta_0 \in \Theta$. The quadratic loss function $L : \Theta \times [0,1] \to [0,1]$ is defined by $L\left(\theta, \widehat{\delta}\right) = \left(\widehat{\delta} - \delta_{\theta,\theta_0}\right)^2$. If the action $\widehat{\delta}(x) \in [0,1]$ is chosen to minimize expected loss with respect to a certainty distribution $C_x$*

*that is compatible with a confidence curve $p$ and that meets the criteria of Corollary 1 for all $x \in \mathcal{X}$, then $\widehat{\delta}(x) = p_x(\theta_0)$.*

*Proof.* A standard result (e.g., Lad, 1996) is that $\widehat{\delta}(x) = C_x(\vartheta = \theta_0)$ minimizes expected quadratic loss for any parameter distribution $C_x$. Corollary 1 implies that $C_x(\vartheta = \theta_0) = p_x(\theta_0)$. □

**Corollary 3.** If $\theta = \theta_0$ in addition to the assumptions of Theorem *4*, then the loss averaged over the sample space is

$$\int L\left(\theta, \widehat{\delta}(x)\right) dP_{\theta,\gamma}(x) = \frac{1}{3}.$$

*Proof.* Theorem 4 gives $\widehat{\delta}(x) = p_x(\theta_0)$, with the result that

$$\begin{aligned}\int L\left(\theta_0, \widehat{\delta}(x)\right) dP_{\theta_0,\gamma}(x) &= \int \left(p_x(\theta_0) - \delta_{\theta_0,\theta_0}\right)^2 dP_{\theta_0,\gamma}(x) \\ &= \int \left(p_x^2(\theta_0) - 2p_x^2(\theta_0) + 1^2\right) dp_x(\theta_0) \\ &= E\left(U^2 - 2U + 1\right) = E\left(U^2\right) - 2E(U) + 1,\end{aligned}$$

where, by equation Corollary 1, $U \sim \mathrm{U}(0,1)$. Finally, $E(U^2) = 1/3$ and $E(U) = 1/2$. □

## 2.2 Other fiducial distributions

As above, the distribution of $X$ depends on the value of some full parameter. Let $\Phi$ denote a set of *target parameter* values, where each target parameter value is a function of the full parameter value. Hypothesis tests, effect-size estimates, and other actions may depend on the value of the target parameter. In other words, any potential parameter of interest is a function of the target parameter. The possible dependence of the distribution of $X$ on another parameter, called the *nontarget parameter*, is suppressed for notational economy.

Suppose there are measurable spaces $\left(\Phi^{(1)},\Sigma^{(1)}\right)$, $\left(\Phi^{(2)},\Sigma^{(2)}\right)$ and functions $\bullet':\Phi\rightarrow\Phi^{(1)}$, $\bullet'':\Phi\rightarrow\Phi^{(2)}$ such that the function $\boldsymbol{\phi}:\Phi\rightarrow\Phi^{(1)}\times\Phi^{(2)}$ is bijective (invertible), where $\boldsymbol{\phi}(\phi)=(\phi',\phi'')$ for all $\phi\in\Phi$. Thus, $\phi=\boldsymbol{\phi}^{-1}((\phi',\phi''))$ for any $\phi\in\Phi$, with the interpretation that $\phi'$ and $\phi''$ are subparameters of $\phi$ that together contain all the information in $\phi$. Accordingly, $\bullet'$ and $\bullet''$ are called *subparameter functions*.

The general definition of a fiducial distribution is self-referential with the recursion stopping at one or more basic fiducial distributions (Definition 1).

**Definition 4.** Consider a fiducial distribution $\Pi_x^{(1)}$ on $\left(\Phi^{(1)},\Sigma^{(1)}\right)$ and a probability distribution $\Pi_x^{(2)}(\bullet|\phi')$ on $\left(\Phi^{(2)},\Sigma^{(2)}\right)$ for every $\phi\in\Phi$. Let $\pi_x^{(1)}:\Phi^{(1)}\rightarrow[0,\infty[$ and $\pi_x^{(2)}(\bullet|\phi'):\Phi^{(2)}\rightarrow[0,\infty[$ denote the probability density functions defined in terms of Radon-Nikodym differentiation of $\Pi_x^{(1)}$ and $\Pi_x^{(2)}(\bullet|\phi')$ with respect to the same dominating measure. A probability distribution $\Pi_x$ on a measurable space $(\Phi,\Sigma)$ is called the *joint fiducial distribution* that extends $\Pi_x^{(1)}$ and $\Pi_x^{(2)}$ if it corresponds to a probability density function $\pi_x:\Phi^{(1)}\times\Phi^{(2)}\rightarrow[0,\infty[$ such that

$$\pi_x(\phi',\phi'')=\pi_x^{(1)}(\phi')\,\pi_x^{(2)}(\phi''|\phi') \tag{16}$$

for all $\phi\in\Phi$ and if $\mathcal{P}_x^{(2)}=\left\{\Pi_x^{(2)}(\bullet|\phi'):\phi\in\Phi\right\}$ satisfies these conditions:

1. For all $\phi\in\Phi$ such that $\phi''$ is a function of $\phi'$, the probability distribution $\Pi_x^{(2)}(\bullet|\phi')$ is $\Delta_{\phi''}$, the Dirac measure with support at $\phi''$ (probability distribution concentrated at $\phi''$).

2. Let $\Phi^\star$ denote the set of all $\phi\in\Phi$ such that $\phi''$ is not a function of $\phi'$. At least one of the following statements holds:

   (a) Consider the function $\phi^{(1)}:\Phi^{(2)}\rightarrow\Phi^{(1)}$ that satisfies $\phi^{(1)}(\phi'')=\phi'$ for each

$\phi \in \Phi^\star$. For all $\phi \in \Phi^\star$, the *conditional fiducial distribution given* $\phi'$ is defined by

$$\Pi_x^{(2)}(\bullet|\phi') = \Pi_x^{(2)}\left(\bullet|\phi^{(1)}\left(\varphi^{(2)}\right) = \phi'\right), \tag{17}$$

which is the conditional probability distribution of $\varphi^{(2)}$ given $\phi^{(1)}\left(\varphi^{(2)}\right) = \phi'$, where $\varphi^{(2)}$ is the random variable distributed as some fiducial distribution $\Pi_x^{(2)}$.

(b) For all $\phi \in \Phi^\star$, $\Pi_x^{(2)}(\bullet|\phi')$ is a fiducial distribution. In this case, the probability distribution

$$\Pi_x^{(2)} = \Pi_x^{(2)}(\bullet) = \int \Pi_x^{(2)}(\bullet|\phi')\, d\Pi_x^{(1)}(\phi') \tag{18}$$

is called the *marginal fiducial distribution* with respect to $\Pi_x^{(1)}$ and $\mathcal{P}_x^{(2)}$.

Any parameter distribution is a *fiducial distribution* if it is a probability distribution that can be deduced from a basic fiducial distribution or a joint fiducial distribution. ▲

According to the definition and Kolmogorov probability theory, any distribution of a parameter is a fiducial distribution if it is a basic fiducial distribution, a conditional fiducial distribution, a marginal fiducial distribution, or a joint fiducial distribution. While basic fiducial distributions are necessarily confidence distributions, other fiducial distributions are often not confidence distributions.

The joint fiducial distributions of Examples 4 and 5 are well-known posterior distributions derived by Fisher via his fiducial argument and by Jeffreys via improper priors (Jeffreys, 1998, §7.1). Specific instances of fiducial distributions that have no Bayesian counterpart are introduced for the first time in Section 3.

**Example 4.** As in Example 2, the observable vector $X$ consists of $n$ independent random variables of distribution $P_{\theta,\sigma} = \mathrm{N}(\theta, \sigma^2)$ with $\theta$ and $\sigma$ unknown. With the parameterization

$\phi = (\theta, \sigma^2)$, define the subparameter functions such that $\phi' = \sigma^2$ and $\phi'' = \theta$ for all $\phi \in \mathbb{R} \times ]0, \infty[$. Let $\upsilon(\bullet) : \mathcal{X} \times ]0, \infty[ \to ]0, \infty[$ and $\tau(\bullet) : \mathcal{X} \times \mathbb{R} \times ]0, \infty[ \to \mathbb{R}$ denote the pivot functions such that

$$\upsilon(x; \sigma) = \frac{(n-1)\hat{\sigma}^2(x)}{\sigma^2}; \; \tau(x; \theta, \sigma) = \frac{\hat{\theta}(x) - \theta}{\sigma/\sqrt{n}}$$

for all $x \in \mathcal{X}$, $\theta \in \mathbb{R}$, and $\sigma \in ]0, \infty[$, where $\hat{\theta}(x)$ and $\hat{\sigma}^2(x)$ are the usual estimates of the mean and variance. Since $\upsilon(X; \sigma)$ has a $\chi^2$ distribution with $n-1$ degrees of freedom for all $\sigma \in ]0, \infty[$, there is a random variable $\varsigma^2$ of basic fiducial distribution $\Pi_x^{(1)}$ such that

$$\Pi_x^{(1)}(\varsigma \le \sigma) = P_{\theta,\sigma}(\upsilon(X; 1) \ge \upsilon(x; 1))$$

for all $x \in \mathcal{X}$, $\theta \in \mathbb{R}$, and $\sigma \in ]0, \infty[$. Likewise, since $\tau(X; \theta, \sigma)$ has a standard normal distribution for all $\theta \in \mathbb{R}$ and $\sigma \in ]0, \infty[$, there is a random variable $\vartheta(\sigma^2)$ of basic fiducial distribution $\Pi_x^{(2)}(\bullet|\sigma^2)$ such that

$$\Pi_x^{(2)}\left(\vartheta\left(\sigma^2\right) \le \theta|\sigma^2\right) = P_{\theta,\sigma}(\tau(X; 0, \sigma) \ge \tau(x; 0, \sigma))$$

for all $x \in \mathcal{X}$, $\theta \in \mathbb{R}$, and $\sigma \in ]0, \infty[$. The distribution $\Pi_x$ of the resulting random variable $\varphi = (\vartheta(\varsigma^2), \varsigma^2)$ is the joint fiducial distribution according to equation (16). With $\theta$ as the parameter of interest, Yates (1939) eliminated the nuisance parameter $\sigma$ by integration with respect to $\Pi_x^{(1)}$, finding that the posterior mean $\varphi'' = \bar{\vartheta} = \int \vartheta(s^2)\, d\Pi_x^{(1)}(s^2)$ is distributed such that $\left(\bar{\vartheta} - \hat{\theta}(x)\right)\sqrt{n}/\hat{s}(x)$ has the Student $t$ distribution with $n-1$ degrees of freedom. That marginal fiducial distribution is also the confidence distribution for $\theta$ that corresponds to $\tau(X; \theta, \sigma(X))$ as the pivotal quantity (Wilkinson, 1977). ▲

In that example, the fiducial distribution of the parameter of interest is a confidence distribution. That is not always the case, as the next example makes clear.

**Example 5.** For samples of sizes $n_1$ and $n_2$ from two different normal populations of unknown means $(\theta_1, \theta_2)$ and variances $(\sigma_1^2, \sigma_2^2)$, the $n_i$-tuple $X_i = (X_{i,1}, \dots, X_{i,n_1})$ has independently distributed components $X_{i,j} \sim \mathrm{N}(\mu_i, \sigma_i^2)$ for $i = 1, 2$ and $j = 1, \dots, n_i$. The parameter of interest is the difference in means, $\theta = \theta_1 - \theta_2$. As seen in Example 4, marginal inferences may be made about $\theta_i$ on the basis of the random parameter $\bar{\vartheta}_i$, distributed according to the marginal fiducial distribution such that $\left(\bar{\vartheta}_i - \hat{\theta}_i(x)\right)\sqrt{n_i}/\hat{s}_i(x)$ has the Student $t$ distribution with $n - 1$ degrees of freedom for $i = 1, 2$, where $\hat{\theta}_i(x)$ and $\hat{s}_i(x)$ are the observed estimates of the mean and variance for the $i$th sample. Let $\phi = (\phi', \phi'') = (\theta_1, \theta_2)$, let $C'$ denote the fiducial distribution of $\bar{\vartheta}_1$, and let $\Pi_x^{(2)}(\bullet|\phi')$ denote the fiducial distribution of $\bar{\vartheta}_2$ for all $\theta_1 \in \mathbb{R}$. Since marginalization according to equation (16) implies that $\bar{\vartheta}_1$ and $\bar{\vartheta}_2$ are independent, the marginal fiducial distribution of $\vartheta = \bar{\vartheta}_1 - \bar{\vartheta}_2$ is the Behrens-Fisher fiducial distribution of the difference in means (Fisher, 1935). As has often been pointed out, that fiducial distribution does not lead to exact confidence intervals; thus, the fiducial distribution of $\vartheta$ is not a confidence distribution. ▲

Because a sampling model and data set do not lead to a unique fiducial distribution, it is useful at this point to formalize the concept of a hypothetical intelligent agent that ultimately bases its decisions on confidence intervals. Let $\mathcal{P}$ denote the set of all fiducial distributions on $(\Phi, \Sigma)$ that can be constructed with $\mathcal{X}$ as the data space. A *fiducial agent* (FA) is a function $\Pi : \mathcal{X} \to \mathcal{P}$ such that its basic fiducial distributions are derived from the same procedures of nested confidence sets and such that its joint fiducial distributions are related to its other fiducial distributions by equation (16) for the same measurable spaces and subparameter functions ($\bullet'$ and $\bullet''$). Thus, the fiducial distribution of any FA $\Pi$ and observation $x \in \mathcal{X}$ is uniquely specified by $\Pi(x)$, which is denoted by $\Pi_x$ above. Likewise, a Bayesian posterior is uniquely specified by the prior and family of sampling distributions that represent the beliefs of a Bayesian agent. This correspondence between fiducial or Bayesian agents and fiducial or Bayesian posteriors adds formal support to the claim of Fraser (2008) and Hannig (2009) that Bayesian inference faces essentially the same uniqueness problem as fiducial inference. Unique fiducial distributions can alternatively be derived under certain conditions by considering the confidence procedure as part of the model of the physical system (Remark 3).

**Example 6.** Welch (1947) proposed a system of approximate confidence intervals as a solution to the Behrens-Fisher problem (Example 5). The corresponding approximate confidence distribution, as an approximate basic fiducial distribution, represents the posterior beliefs of a different agent than the agent whose posterior beliefs are represented by the fiducial distribution derived in Example 5. The latter agent is a better idealization of statisticians who would order the certainty level of hypotheses according to the confidence levels from the basic fiducial distributions of Example 5 when making inferences about the mean as well as when making inferences about the standard deviation. Such ordering is not coherent with ordering levels of certainty according to the confidence levels of Welch (1947). This has far-reaching implications for statistical practice (Remark 1). ▲

In some cases, the statistician may have difficulty in committing to a single FA. When multiple FAs are equally suitable as representations of the posterior beliefs of a scientist, organization, or other real agent, the most representative FAs may be coherently combined into a single posterior distribution via simple arithmetic averaging (see, e.g., Paris, 1994) or the game-theoretic method of Bickel (2012d). The combined posterior distribution will not necessarily be a fiducial distribution.

# 3 Inference in the presence of a plausible null hypothesis

## 3.1 Certainty based on a plausible null hypothesis

In many applications involving testing the null hypothesis that that $\phi = \phi_0$ for some $\phi_0 \in \Phi$, the parameter value $\phi_0$ is regarded as *a priori* more plausible than any other parameter value, at least for the sake of argument or reporting. That information can be encoded in joint fiducial distributions by using the Dirac measure in place of a basic fiducial distribution, as Definition 4 allows.

A simple and widely applicable way to do that begins by defining $\phi'$ as a distance from the most plausible parameter value. In this setting, $\Phi^{(1)} \subseteq [0, \infty[$, and the subparameter functions $\bullet'$ and $\bullet''$ and a *magnitude* transformation mag satisfy $\phi'' = \phi$ and

$$\phi' = \mathrm{mag}\,(\phi) = D\,(\phi, \phi_0)$$

for all $\phi \in \Phi$, where $D$ is a distance measure. Let $\Pi_x^{(2)}$ denote the basic fiducial distribution that is compatible with a procedure of nested confidence sets for $\phi$. By assumption, $\Pi_x^{(2)}$ meets the conditions of Corollary 1. Since $\mathrm{mag}\,(\phi_0'') = 0$ implies that $\phi_0 = \phi$, Definition 4 requires that $\Pi_x^{(2)}\,(\bullet|0) = \Delta_{\phi_0}$. Let $\tau\,(X; \phi, \gamma)$ be a pivotal quantity that defines the random variable $\vartheta^{(2)}$ of a basic fiducial distribution $\Pi_x^{(2)}$ compatible with a confidence curve $p^{(2)}$ by

$$\Pi_x^{(2)}\left(\vartheta^{(2)} \leq \phi_0\right) = P_{\phi,\gamma}\left(\tau\,(X; \phi_0, \gamma) \geq \tau\,(x; \phi_0, \gamma)\right)$$

for all $x \in \mathcal{X}$, $\phi \in \Phi \backslash \{\phi_0\}$, and $\gamma \in \Gamma$, where $\Gamma$ is the set of possible values of the nontarget parameter, which in this subsection is required to be nonbasic (§2.1). To define $\Pi_x^{(1)}$, the parameter $\phi$ will be broken into its magnitude component $\phi'$ and direction component, a member of

$$\mathcal{D}_{\phi_0} = \{\phi/\phi' : \phi \in \Phi \backslash \{\phi_0\}\}.$$

Let $\upsilon(x;\phi',\delta,\gamma) = |\tau(x;\phi'\delta,\gamma)|$ and $P_{\phi',\delta,\gamma} = P_{\phi'\delta,\gamma}$ for all $x \in \mathcal{X}$, $\phi \in \Phi\backslash\{\phi_0\}$, $\delta \in \mathcal{D}_{\phi_0}$, and $\gamma \in \Gamma$. Suppose there is a random variable $\vartheta^{(1)}$ of basic fiducial distribution $\Pi^{(1)}_{x,\phi_0}$ that is compatible with a procedure of nested confidence sets for $\theta^{(1)}$ that is defined by some confidence curve $p^{(1)}_{\bullet,\phi_0}$. This fiducial distribution must satisfy

$$\Pi^{(1)}_{x,\phi_0}\left(\vartheta^{(1)} \le \phi'\right) = p^{(1)}_{x,\phi_0}(\phi') = P_{\phi',\delta,\gamma}\left(\upsilon(X;\phi',\delta,\gamma) \ge \upsilon(x;\phi',\delta,\gamma)\right)$$

for all $x \in \mathcal{X}$, $\phi \in \Phi\backslash\{\phi_0\}$, $\delta \in \mathcal{D}_{\phi_0}$, and $\gamma \in \Gamma$. Since the fiducial distribution $\Pi^{(2)}_x$ generates conditional fiducial distributions according to equation (17) for $\phi \in \Phi\backslash\{\phi_0\}$, there is a joint fiducial distribution that extends $\Pi^{(1)}_{x,\phi_0}$ and $\Pi^{(2)}_x$. That distribution is denoted by $C_{x,\phi_0}$ and is called a *fiducial distribution in the presence of the plausible null hypothesis that* $\phi = \phi_0$.

Letting $\varphi$ denote the random variable of distribution $C_{x,\phi_0}$, that fiducial distribution is succinctly expressed as a mixture of $\Delta_{\phi_0}$ and $\Pi^{(2)}_x\left(\bullet|\vartheta^{(2)} \ne \phi_0\right)$:

$$\begin{aligned} C_{x,\phi_0}(\bullet) &= C_{x,\phi_0}(\varphi = \phi_0)\, C_{x,\phi_0}(\bullet|\varphi = \phi_0) + C_{x,\phi_0}(\varphi \ne \phi_0)\, C_{x,\phi_0}(\bullet|\varphi > \phi_0) \\ &= \begin{array}{c} \Pi^{(1)}_{x,\phi_0}\left(\vartheta^{(1)} = \phi^+(\phi_0)\right)\Delta_{\phi_0}(\bullet) + \\ \Pi^{(1)}_{x,\phi_0}\left(\vartheta^{(1)} > \phi^+(\phi_0)\right)\Pi^{(2)}_x\left(\bullet|\vartheta^{(2)} \ne \phi_0\right) \end{array} \\ &= \Pi^{(2)}_x\left(\bullet|\vartheta^{(2)} > \phi_0\right) + \left(\Delta_{\phi_0}(\bullet) - \Pi^{(2)}_x\left(\bullet|\vartheta^{(2)} \ne \phi_0\right)\right)\Pi^{(1)}_{x,\phi_0}\left(\vartheta^{(1)} = \phi_0\right). \end{aligned} \tag{19}$$

The certainty level of the plausible null hypothesis is equal to a p-value since $\Pi^{(1)}_{x,\phi_0}$ meets the conditions of Corollary 1 and since $\Phi^{(1)} \subseteq [0, \infty[$. Specifically,

$$C_{x,\phi_0} (\varphi = \phi_0) = \Pi^{(1)}_{x,\phi_0} (\varphi = \phi_0) = p^{(1)}_{x,\phi_0} (\phi'_0) . \tag{20}$$

Simplification in the form of

$$\Pi^{(2)}_x \left( \bullet | \vartheta^{(2)} \neq \phi_0 \right) = \Pi^{(2)}_x (\bullet) \tag{21}$$

(equality up to measure 0) is possible in the case that $\Pi^{(2)}_x \left( \vartheta^{(2)} \neq \phi_0 \right) = 1$, as when $\vartheta^{(2)} \sim \Pi^{(2)}_x$ is continuous. The next example illustrates this.

**Example 7.** In the notation of this subsection, equation (13) of Example 3 says $\Pi^{(1)}_{x,0} \left( \vartheta^{(1)} = 0 \right) = p'_{x,0} (0)$, where $p'_{x,0} (0)$ is the usual two-sided p-value from the single-sample $t$-test of the null hypothesis that the mean is equal to 0, i.e., that $\phi = \phi_0$. Thus, equation (20) equates that p-value with the posterior level of certainty in that hypothesis: $C_{x,0} (\varphi = 0) = p'_{x,0} (0)$. By contrast, the basic fiducial distribution $\Pi^{(2)}_x$ assigns 0 certainty to the hypothesis $\left( \Pi^{(2)}_x \left( \vartheta^{(2)} = 0 \right) = 0 \right)$ since it admits a continuous density function: the continuous random variable $\vartheta^{(2)} \sim \Pi^{(2)}_x$ is proportional to a noncentral $t$ variate according to expression (7), in which $d = 1$ here. Using the same example but with a known variance and in the multivariate setting ($d \geq 2$) of Example 1, Stein (1959) pointed out the discrepancy between $\Pi^{(1)}_{x,0} \left( \vartheta^{(1)} \in \bullet \right)$ and $\Pi^{(2)}_x \left( \left( \varphi^{(2)} \right)^{\mathrm{T}} \varphi^{(2)} \in \bullet \right)$ and favored the former for inference about $\mathrm{mag}(\theta) = \theta^{\mathrm{T}} \theta$ since $\Pi^{(1)}_{x,0}$ corresponds to a confidence procedure for $\mathrm{mag}(\theta)$; Remark 4 briefly surveys the literature on this discrepancy. In the context of the prior plausibility of the null hypothesis value $\phi_0 = 0$ (Bickel, 2012b,d), equation (19) indicates that there can be no conflict between the two distributions: $\Pi^{(1)}_{x,0}$ only pertains to the magnitude of $\theta$, and $\Pi^{(2)}_x$

only pertains to its direction. By contrast, in the context of no prior plausibility of one value of $\phi$ above any other, $\Pi_x^{(2)}$ rather than $C_{x,0}$ would be appropriate for the minimization of expected utility. More formally, $C_{x,0}$ and $\Pi_x^{(2)}$ correspond to the idealized knowledge bases of different agents, one of which may better represent actual knowledge. ▲

## 3.2 Effect-size estimates shrunk toward the null hypothesis

In this subsection, it is assumed that $\Pi_x^{(2)}\left(\vartheta^{(2)} \neq \phi_0\right) = 1$, entailing that equation (21) holds. The conditions of Corollary 1 are also taken for granted with the result that every basic fiducial distribution considered is a confidence distribution.

### 3.2.1 Point estimation

An estimator of a parameter is considered *consistent* if it converges in $P_{\phi,\gamma}$-probability to the true value of the parameter. Similarly, for a scalar parameter ($\phi \in \mathbb{R}$), the significance function is called *asymptotically powerful* (cf. Bickel, 2012a) if it converges in probability to 0 or 1 under the alternative hypothesis ($\phi \neq \phi_0$):

$$q_X(\phi_0) \overset{P_{\phi,\gamma}}{\to} \begin{cases} 1 & \text{if } \phi < \phi_0 \\ 0 & \text{if } \phi > \phi_0. \end{cases} \tag{22}$$

Let $\bar{\bullet}_x$ denote the posterior mean of a parameter with respect to its fiducial distribution for any observation $x \in \mathcal{X}$; again, "posterior" abbreviates "data-dependent" and is not necessarily a Bayesian posterior for a data-independent prior. The posterior means of the $\vartheta^{(2)}$ and $\varphi$

defined in Section 3.1 are their expectation values with respect to their fiducial distributions:

$$\bar{\vartheta}_x^{(2)} = \int \underline{\phi} d\Pi_x^{(2)} \left(\underline{\phi}\right); \tag{23}$$

$$\bar{\varphi}_x = \int \underline{\phi} d\Pi_x \left(\underline{\phi}\right) = p_{x,\phi_0}^{(1)} \left(\phi_0'\right) \phi_0 + \left(1 - p_{x,\phi_0}^{(1)} \left(\phi_0'\right)\right) \bar{\vartheta}_x^{(2)}, \tag{24}$$

where $\underline{\phi} \in \Phi$ is the dummy variable of integration. Setting $\phi_0 = 0$ yields the shrunken parameter estimate advertised in Section 1: $\bar{\varphi}_x = \left(1 - p_{x,0}^{(1)} (0)\right) \bar{\vartheta}_x^{(2)}$, as will be exploited in Example 8.

**Theorem 5.** *Let $\Phi = \mathbb{R}^{d_1}$ for some $d_1 \in \{1, 2, \dots\}$. If $\Pi_x^{(2)} \left(\vartheta^{(2)} \neq \phi_0\right) = 1$, if the alternative-conditional posterior mean $\bar{\vartheta}_X^{(2)}$ is a consistent estimator of $\phi$, and if the p-value $p_{X,\phi_0}^{(1)} (\phi_0)$ is asymptotically powerful, then the alternative-marginal posterior mean $\bar{\varphi}_X$ is a consistent estimator of $\phi$.*

*Proof.* First, the result is easily obtained in the case of a true null hypothesis ($\phi = \phi_0$). Since $\bar{\vartheta}_X^{(2)} \overset{P_{\phi_0,\gamma}}{\to} \phi = \phi_0$, equation (24) immediately yields $\bar{\varphi}_X \overset{P_{\phi_0,\gamma}}{\to} \phi_0$. Next, consider the case of a true alternative hypothesis ($\phi \neq \phi_0$). According to equation (24), for any $\epsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} P_{\phi,\gamma} \left( \left| \bar{\varphi}_X - \bar{\vartheta}_X^{(2)} \right| < \epsilon \right) &= \lim_{n\to\infty} P_{\phi,\gamma} \left( \left| p_{X,\phi_0}^{(1)} \left(\phi_0'\right) \left(\phi_0 - \bar{\vartheta}_X^{(2)}\right) \right| < \epsilon \right) \\ &= \lim_{n\to\infty} P_{\phi,\gamma} \left( \left| \bar{\vartheta}_X^{(2)} - \phi_0 \right| < \frac{\epsilon}{\left| p_{X,\phi_0}^{(1)} \left(\phi_0'\right) \right|} \right), \end{aligned}$$

which is 1 according to equation (23) since $\bar{\vartheta}_X^{(2)} \overset{P_{\phi,\gamma}}{\to} \phi$ by the definition of consistency and since $\left| p_{X,\phi_0}^{(1)} \left(\phi_0'\right) \right|$ by equation (22). Thus, $\bar{\varphi}_X \overset{P_{\phi,\gamma}}{\to} \bar{\vartheta}_X^{(2)}$. Together, $\bar{\vartheta}_X^{(2)} \overset{P_{\phi,\gamma}}{\to} \phi$ and $\bar{\varphi}_X \overset{P_{\phi,\gamma}}{\to} \bar{\vartheta}_X^{(2)}$ are sufficient for $\bar{\varphi}_X \overset{P_{\phi,\gamma}}{\to} \phi$. □

The result is widely applicable. Indeed, for the special case of a scalar basic parameter ($\phi \in \mathbb{R}$), Singh et al. (2007) found that $\bar{\vartheta}_X^{(2)}$ is a consistent estimator of $\phi$ under broad conditions.

**Example 8.** Example 7, continued. For $n \to \infty$, Figure 1 compares the posterior mean based on the fiducial distribution to the maximum-likelihood estimate (MLE), which is the sample mean in this case. The plot illustrates how the fiducial distribution provides a smooth alternative to estimation after testing with respect to a fixed significance threshold. Thus, that practice (Fisher, 1925; Montazeri et al., 2010) may be interpreted as a dirty approximation to coherent frequentist inference. However, in this case, no approximation is warranted on computational grounds since the posterior mean is simply $\bar{\varphi}_x = \left(1 - p_{x,0}^{(1)}\right) \bar{\vartheta}_x^{(2)}$ according to equation (24), where $p_{x,0}^{(1)}(0)$ is the two-sided p-value and $\bar{\vartheta}_x^{(2)}$ is the sample mean.

Smooth shrinkage can also be achieved through methods of frequentist model averaging (FMA) aimed at estimating a parameter (Claeskens and Hjort, 2008). With respect to point estimation, the certainty-distribution approach and FMA have many of the same advantages over estimation after testing and estimation after model selection, their respective threshold-dependent counterparts. However, existing FMA methods require asymptotic approximations that fiducial distributions do not, indicating that the latter may be more reliable for small samples. Nonetheless, fiducial distributions can depend nonetheless on asymptotic confidence intervals when exact confidence intervals are not available. Another advantage of basing point estimation on a joint fiducial distribution is coherence with interval estimates.

### 3.2.2 Interval estimation

Many contexts call for reporting certainty regions, regions that contain the parameter at some specified level of certainty. When the target parameter is a scalar ($\phi \in \mathbb{R}$), the regions are intervals. In that case, it is convenient to define the *certainty curve* as the function $p_{\bullet,\phi_0}(\bullet) : \mathcal{X} \times \Phi \to [0,1]$ such that

$$p_{x,\phi_0}(\phi) = C_{x,\phi_0}(\varphi \leq \phi)$$

for all $x \in \mathcal{X}$ and $\phi \in \Phi$, where $\varphi \sim C_{x,\phi_0}$. Unlike $p^{(1)}_{\bullet,\phi_0}$ and $p^{(2)}$, this $p_{\bullet,\phi_0}$ is not a confidence curve since $C_{x,\phi_0}$ is not a confidence distribution. By equation (19),

$$p_{x,\phi_0}(\phi) = p^{(1)}_{x,\phi_0}(\phi'_0)\, 1_{[\phi_0 \infty)}(\phi) + \left(1 - p^{(1)}_{x,\phi_0}(\phi'_0)\right) p^{(2)}_x(\phi).$$

Inverting $p_{x,\phi_0}$ yields, for any $\beta \in [0,1]$,

$$p^{-1}_{x,\phi_0}(\beta) = \begin{cases} \left(p^{(2)}_x\right)^{-1}\left(\frac{\beta}{1-p^{(1)}_{x,\phi_0}(\phi'_0)}\right) & \text{if } \beta < \left(1 - p^{(1)}_{x,\phi_0}(\phi'_0)\right) p^{(2)}_x(\phi_0) \\ \phi_0 & \text{if } \begin{array}{c} \left(1 - p^{(1)}_{x,\phi_0}(\phi'_0)\right) p^{(2)}_x(\phi_0) \leq \beta \leq \\ \left(1 - p^{(1)}_{x,\phi_0}(\phi'_0)\right) p^{(2)}_x(\phi_0) + p^{(1)}_{x,\phi_0}(\phi'_0) \end{array} \\ \left(p^{(2)}_x\right)^{-1}\left(\frac{\beta - p^{(1)}_{x,\phi_0}(\phi'_0)}{1-p^{(1)}_{x,\phi_0}(\phi'_0)}\right) & \text{if } \beta > \left(1 - p^{(1)}_{x,\phi_0}(\phi'_0)\right) p^{(2)}_x(\phi_0) + p^{(1)}_{x,\phi_0}(\phi'_0) \end{cases} \quad (25)$$

The interval

$$\widehat{\Phi}_{x,\phi_0}(\beta_1,\beta_2) = \left[p^{-1}_{x,\phi_0}(\beta_1), p^{-1}_{x,\phi_0}(\beta_2)\right]$$

is the $(\beta_2 - \beta_1)\,100\%$ *certainty interval centered at* $(\beta_1 + \beta_2)/2$ *in the presence of the plausible null hypothesis that* $\phi = \phi_0$, where $0 \leq \beta_1 \leq \beta_2 \leq 1$.

It is clear from equation (25) that, for any $\phi \in \Phi$, those certainty intervals are almost always shorter than the confidence intervals based on $\Pi_x^{(2)}$. When $\phi$ is close to $\phi_0$, that improvement tends to be substantial. Thus, nested confidence intervals successfully generate interval estimates that smoothly shrink toward the plausible hypothesis value rather than retaining the frequentist coverage property that is appropriate when such a value is unknown.

Bickel (2012c) derived the equivalent of equation (25) with an estimated or approximate Bayesian posterior probability of the null hypothesis in place of the p-value $p_{x,\phi_0}^{(1)}(\phi_0')$. A key difference from the present approach is the interpretation of the interval estimates. Whereas the marginal confidence intervals of Bickel (2012c) may be interpreted as an approximation to the physical distribution of the parameter, that interpretation cannot apply to the above certainty intervals since $p_{x,\phi_0}^{(1)}(\phi_0')$ is not a Bayesian posterior probability for any data-independent prior (§5).

# 4 Remarks

*Remark* 1. Example 6 brings into bold relief the fundamental difference between the proposed use of confidence distributions and frequentism as it is usually practiced: there is no FA that would switch from the one-sample $t$-test to the Welch $t$-test merely due to a change in the parameter of interest. In the fiducial distribution approach, inferences for a given agent cohere with each other regardless of choices of the parameter of interest, whereas many frequentists would instead follow Cox (2006) in changing the system of confidence intervals according to the parameter of interest even in the absence of changes in background information. The objective Bayesian practice of using reference priors that depend on which parameter is of interest (e.g., Berger, 2009) also sacrifices coherence in favor of reducing inference to automatic rules (Bickel, 2012e).

*Remark* 2. The concise term "confidence measure" (Bickel, 2009) for what is here called a confidence distribution is less subject to misunderstanding than other terms in the literature. Many authors call the exact confidence measure a "confidence distribution" (e.g., Efron, 1993; Schweder and Hjort, 2002). By contrast, more recent papers (e.g., Singh et al., 2005, 2007) use "confidence distribution" for the cumulative distribution function (CDF) of an exact confidence measure and use "asymptotic confidence distribution" for the CDF of any confidence measure. To avoid the confusion generated by those different definitions of "confidence distribution," the term "confidence posterior distribution" (Bickel, 2011b, 2012a) has been suggested as a term that emphasizes its use in minimizing posterior expected loss. Polansky (2007, p. 24) coined "observed confidence levels" for probabilities associated with confidence measures.

*Remark* 3. Since Section 2.2 defines the fiducial distribution in terms of the procedure of confidence intervals that contains the relevant information is relevant to the knowledge base of an intelligent agent, it does not extend the statistical model of the physical system. That model remains the family of distributions, which is insufficient to specify a fiducial distribution. However, the basic fiducial distribution has much in common with extended models, including the structural models of Fraser (1968) and the pivotal models of Barnard (1980) and Barnard (1995) (with Barnard (1996)). While a structural model is defined by adding a transformation group to the family of distributions, and a pivotal model is defined by adding a pivot to the family, the two are isomorphic under general conditions (Fraser, 1996). See also McCullagh (2002) and Helland (2004, 2009) for closely related extensions of the physical model. These considerations may play a role in discriminating between agents and their corresponding fiducial distributions (Remark 4). In contrast with both fiducial distributions and extended physical models, Fisher did not intend the fiducial argument to depend on any assumptions in addition to the family of distributions (Dawid and Stone, 1982, comment by Fraser) except for the assumption that any physical prior distribution (§5) is unknown (Fisher, 1973).

*Remark* 4. Previous work related to selecting a fiducial distribution according to the available background information is expressed here in the notation of Example 7. Wilkinson (1977) found the nonzero probabilities of the null hypothesis provided by $\Pi_{x,0}^{(1)}$ appropriate when the null hypothesis has plausibility apart from $x_i$. By contrast, he found the 0 probability of the null hypothesis provided by $\Pi_x^{(2)}$ appropriate in the absence of any pre-data information about $\phi$. Wilkinson (1977, pp. 126-127) reasoned that the null hypothesis would not be of sufficient interest for statistical inference were it implausible, which is consistent with the agent-based theory of the present paper. One way to determine which agent best represents prior information is to require invariance to certain parameter and data transformations. Helland (2004) proposed choosing between $\Pi_{x,0}^{(1)}$ and $\Pi_x^{(2)}$ on the basis of transformation properties; see Remark 3. Similarly, from a subjective Bayesian viewpoint, whether the uniform prior is appropriate depends on an agent's beliefs (Berger, 1985, §4.7.9).

# 5 Discussion

While the general theory of Section 2 is built on elements of frequentist and Bayesian reasoning, it leads to distinctive results that can be derived from neither frequentist theory nor Bayesian theory alone. Specifically, ordering levels of belief according to confidence levels of nested confidence intervals in a framework of maximum expected utility leads to fiducial distributions that are not necessarily confidence distributions or Bayesian posterior distributions.

A striking implication of this fiducial approach is the interpretation of the p-value as the level of agent belief in the null hypothesis (§§2.1.2, 3.1). Given the conditions of Lemma 1 and Corollary 1, the certainty level of a simple null hypothesis is distributed as a p-value under the null hypothesis: $C_X(\vartheta = \theta_0) \sim \mathrm{U}(0, 1)$. That sharply conflicts with the behavior of the Bayesian posterior probability of the null hypothesis, which converges to 1 under the null hypothesis under widely applicable conditions. Thus, while many fiducial distributions are equal to certain objective Bayesian posterior distributions (Jeffreys, 1998, §7.4), the joint fiducial distributions emphasized in Section 3 have no strict Bayesian counterpart.

The discrepancy between the p-value and Bayesian posterior probabilities of the null hypothesis (Berger and Sellke, 1987) has been explained in terms of treating the simple (sharp) null hypothesis as an approximation of a composite null hypothesis centered at the parameter value of the null hypothesis (Gómez-Villegas and Sanz, 1998). From the point of view of effect-size estimation, the low probability of a simple null hypothesis is irrelevant if the estimated effect size is too small to be of any practical significance. For that reason, the impact of the proposed approach on point and interval estimation (§3.2) is more relevant to applications than the probability of the null hypothesis in itself.

The tension between the fiducial probability and a Bayesian posterior probability of the null hypothesis is also alleviated by recalling that the former is only appropriate inasmuch as the physical distribution of the parameter is unknown. As epistemological distributions, fiducial distributions must yield to Bayesian posteriors to the extent that physical priors are known (Bickel, 2011a, 2012b,d). For example, if a physical prior is fully known, then the Bayesian posterior completely replaces the fiducial distribution (Fisher, 1973; Wilkinson, 1977).

## Acknowledgments

This research was partially supported by the Canada Foundation for Innovation, by the Ministry of Research and Innovation of Ontario, and by the Faculty of Medicine of the University of Ottawa.

## Appendix A: Additional proofs

**Proof of Theorem 2** The ordering specified by formulas (5) and (6) implies the existence of a function $\omega : [0,1] \to [0,1]$ such that $C_x(\Theta_1) = \omega\left(\sup\left|\widehat{\Theta}_x^{-1}(\Theta_1)\right|\right)$ for all $x \in \mathcal{X}$ and $\Theta_1 \in \mathcal{H}_x\left(\widehat{\Theta}\right)$. By formulas (1) and (4),

$$C_x(\Theta_1) = C_x\left(p_x(\vartheta) \in \overline{\mathcal{I}}(\Theta_1)\right) = \omega\left(\left|\overline{\mathcal{I}}(\Theta_1)\right|\right), \tag{26}$$

for all $\Theta_1 \in \mathcal{H}_x\left(\widehat{\Theta}\right)$, where $\overline{\mathcal{I}}(\Theta_1)$ is the widest interval in $\widehat{\Theta}_x^{-1}(\Theta_1)$, and $\vartheta$ is the random variable of distribution $C_x$. Therefore, since the confidence procedure is potentially invertible, there is an $x \in \mathcal{X}$ and a $\mathfrak{I}_x \subseteq \mathfrak{I}$ such that, given any $\epsilon > 0$ and $\delta \in ]0,1[$,

$$\omega\left(\left|[\alpha, \alpha+\delta]\right|\right) \leq C_x\left(\alpha \leq p_x\left(\vartheta\right) \leq \alpha+\delta\right) \leq \omega\left(\left|[\alpha, \alpha+\delta]\right| + \epsilon\right)$$

for all $\alpha \in [0, 1-\delta]$. Since $\epsilon$ is arbitrarily small, the function $\omega$ must satisfy

$$\begin{aligned} C_x\left(\alpha \leq p_x\left(\vartheta\right) \leq \alpha+\delta\right) &= \omega\left(\left|[\alpha, \alpha+\delta]\right|\right) \\ &= \omega\left(\delta\right) \end{aligned}$$

for all $\alpha \in [0, 1-\delta]$. Since $\omega\left(\delta\right)$ (the right-hand side) does not depend on $\alpha$ and since $C_x\left(0 \leq p_x\left(\vartheta\right) \leq 1\right) = 1$, $p_x\left(\vartheta\right)$ is uniformly distributed between 0 and 1 for all $x \in \mathcal{X}$, and $\omega\left(\delta\right) = \delta$ for all $\delta \in [0,1]$. Consequently, by formula (26),

$$C_x\left(\Theta_1\right) = \left|\overline{\mathcal{I}}\left(\Theta_1\right)\right| = \sup\left|\widehat{\Theta}_x^{-1}\left(\Theta_1\right)\right| \geq \left|\widehat{\Theta}_x^{-1}\left(\Theta_1\right)\right|$$

for all $\Theta_1 \in \mathcal{H}_x\left(\widehat{\Theta}\right)$. In conclusion, $C_x\left(\Theta_1\right) \geq \left|\mathcal{I}\right|$ for all $\mathcal{I} \in \mathfrak{I}$ such that $\widehat{\Theta}_x\left(\mathcal{I}\right) = \Theta_1$. Substitutions involving formulas (3) and (9) complete the proof.

**Proof of Corollary 1** The first claim follows immediately from Theorems 1 and 2. Since every certainty distribution $C_x$ is also a basic fiducial distribution, the definition of the latter yields

$$C_x\left(\vartheta = \theta_0\right) = C_x\left([\theta_0, \theta_0]\right) \geq P_{\theta,\gamma}\left(p_X\left(\theta\right) \in [0, \alpha]\right)$$

for all $\alpha \in [0,1]$ such that $\left\{\theta \in \Theta : p_x\left(\theta\right) \in [0, \alpha]\right\} = [\theta_0, \theta_0]$. Thus,

$$C_x\left(\vartheta = \theta_0\right) = P_{\theta,\gamma}\left(p_X\left(\theta\right) \leq p_x\left(\theta\right)\right).$$

Lemma 1 then gives formula (12).

## References

Barnard, G., 1980. Pivotal inference and the Bayesian controversy. Trabajos de Estadistica Y de Investigacion Operativa 31, 295–318.

Barnard, G. A., 1995. Pivotal models and the fiducial argument. International Statistical Review / Revue Internationale de Statistique 63, 309–323.

Barnard, G. A., 1996. Corrigenda: Pivotal models and the fiducial argument. International Statistical Review / Revue Internationale de Statistique 64.

Berger, J.O., B. J. S. D., 2009. The formal definition of reference priors. Annals of Statistics 37 (2), 905–938.

Berger, J. O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J. O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of p values and evidence. Journal of the American Statistical Association 82, 112–122.

Bickel, D. R., 2009. A frequentist framework of inductive reasoning. Technical Report, Ottawa Institute of Systems Biology, arXiv:math.ST/0602377.

Bickel, D. R., 2011a. Blending Bayesian and frequentist methods according to the precision of prior information with an application to hypothesis testing. Technical Report, Ottawa Institute of Systems Biology, arXiv:1107.2353.

Bickel, D. R., 2011b. Estimating the null distribution to adjust observed confidence levels for genome-scale screening. Biometrics 67, 363–370.

Bickel, D. R., 2012a. Coherent frequentism: A decision theory based on confidence sets. Communications in Statistics - Theory and Methods 41, 1478–1496.

Bickel, D. R., 2012b. Controlling the degree of caution in statistical inference with the Bayesian and frequentist approaches as opposite extremes. Electron. J. Statist. 6, 686–709.

Bickel, D. R., 2012c. Empirical Bayes interval estimates that are conditionally equal to unadjusted confidence intervals or to default prior credibility intervals. Statistical Applications in Genetics and Molecular Biology 11 (3), art. 7.

Bickel, D. R., 2012d. Game-theoretic probability combination with applications to resolving conflicts between statistical methods. International Journal of Approximate Reasoning 53, 880–891.

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

Birnbaum, A., 1961. Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. Journal of the American Statistical Association 56, pp. 246–249.

Bityukov, S., Krasnikov, N., Nadarajah, S., Smirnova, V., 2011. Confidence distributions in statistical inference. AIP Conference Proceedings 1305, 346–353.

Blaker, H., 2000. Confidence curves and improved exact confidence intervals for discrete distributions. Canadian Journal Of Statistics 28 (4), 783–798.

Box, G. E. P., Tiao, G. C., 1992. Bayesian Inference in Statistical Analysis. Wiley-Interscience.

Claeskens, G., Hjort, N. L., 2008. Model Selection and Model Averaging. Cambridge University Press, Cambridge.

Cox, D. R., 1958. Some problems connected with statistical inference. The Annals of Mathematical Statistics 29, 357–372.

Cox, D. R., 2006. Principles of Statistical Inference. Cambridge University Press, Cambridge.

Dawid, A. P., Stone, M., 1982. The functional-model basis of fiducial inference (with discussion). The Annals of Statistics 10, 1054–1074.

Dempster, A. P., 2008. The dempster-shafer calculus for statisticians. International Journal of Approximate Reasoning 48, 365–377.

Eaton, M., Sudderth, W. D., 2010. Invariance of posterior distributions under reparametrization. Sankhya A 72 (1996), 101–118.

Efron, B., 1993. Bayes and likelihood calculations from confidence intervals. Biometrika 80, 3–26.

Efron, B., 1998. R. A. Fisher in the 21st century, invited paper presented at the 1996 R. A. Fisher lecture. Statistical Science 13, 95–114.

Fisher, R., 1935. The fiducial argument in statistical inference. Annals of Human Genetics 6, 391–398.

Fisher, R. A., 1925. Statistical Methods for Research Workers. Oliver and Boyd, London.

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

Fraser, D. A. S., 1968. The Structure of Inference. John Wiley, New York.

Fraser, D. A. S., 1991. Statistical inference: likelihood to significance. Journal of the American Statistical Association 86, 258–265.

Fraser, D. A. S., 1996. Some remarks on pivotal models and the fiducial argument in relation to structural models. International Statistical Review / Revue Internationale de Statistique 64, 231–236.

Fraser, D. A. S., 2008. Fiducial inference. In: Durlauf, S. N., Blume, L. E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke.

Gibson, G. J., Streftaris, G., Zachary, S., 2011. Generalised data augmentation and posterior inferences. Journal of Statistical Planning and Inference 141, 156–171.

Goldstein, M., 1997. Prior inferences for posterior judgements. Structures and Norms in Science, 55–71.

Gómez-Villegas, M. a., Sanz, L., Jun. 1998. Reconciling Bayesian and frequentist evidence in the point null testing problem. Test 7 (1), 207–216.

Hacking, I., 1967. Slightly more realistic personal probability. Decision, Probability, and Utility.

Hannig, J., 2009. On generalized fiducial inference. Statistica Sinica 19, 491–544.

Hannig, J., Iyer, H., Patterson, P., 2006. Fiducial generalized confidence intervals. Journal of the American Statistical Association 101, 254–269.

Helland, I. S., 2004. Statistical inference under symmetry. International Statistical Review 72, 409–422.

Helland, I. S., 2009. Steps Towards a Unified Basis for Scientific Models and Methods. World Scientific Publishing Company, Singapore.

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., Farrell, R. H., 1992. Estimation of accuracy in testing. The Annals of Statistics 20, 490–509.

Jeffreys, H., 1998. Theory of Probability. Oxford University Press, London.

Kim, D., Lindsay, B. G., 2011. Using confidence distribution sampling to visualize confidence sets. Statistica Sinica 21 (2), 923–948.

Lad, F., 1996. Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction. Wiley-Interscience.

McCullagh, P., 2002. What is a statistical model? (with discussion). The Annals of Statistics 30, 1225–1310.

Montazeri, Z., Yanofsky, C. M., Bickel, D. R., 2010. Shrinkage estimation of effect sizes as an alternative to hypothesis testing followed by estimation in high-dimensional biology: Applications to differential gene expression. Statistical Applications in Genetics and Molecular Biology 9, 23.

Morgenthaler, S., Staudte, R. G., 2005. Conditionally optimal weights of evidence. Acta Mathematicae Applicatae Sinica, English Series 21 (2), 247–256.

Paris, J. B., 1994. The Uncertain Reasoner's Companion: A Mathematical Perspective. Cambridge University Press, New York.

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

Polanyi, M., 1962. Personal Knowledge: Towards a Post-critical Philosophy. University of Chicago Press, Chicago.

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

Savage, L. J., 1954. The Foundations of Statistics. John Wiley and Sons, New York.

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

Singh, K., Xie, M., Strawderman, W. E., 2005. Combining information from independent sources through confidence distributions. Annals of Statistics 33, 159–183.

Singh, K., Xie, M., Strawderman, W. E., 2007. Confidence distribution (CD) – distribution estimator of a parameter. IMS Lecture Notes Monograph Series 2007 54, 132–150.

Stein, C., 1959. An example of wide discrepancy between fiducial and confidence intervals. The Annals of Mathematical Statistics 30, 877–880.

Tian, L., Wang, R., Cai, T., Wei, L.-J., 2011. The highest confidence density region and its usage for joint inferences about constrained parameters. Biometrics 67 (2), 604–10.

van Berkum, E., Linssen, H., Overdijk, D., 1996. Inference rules and inferential distributions. Journal of Statistical Planning and Inference 49, 305–317.

von Neumann, J., Morgenstern, O., 1953. Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Wald, A., 1961. Statistical Decision Functions. John Wiley and Sons, New York.

Wang, C., Hannig, J., Iyer, H. K., 2012. Fiducial prediction intervals. Journal of Statistical Planning and Inference 142 (7), 1980–1990.

Welch, B. L., 1947. The generalization of student's problem when several different population variances are involved. Biometrika 34, 28–35.

Wilkinson, G. N., 1977. On resolving the controversy in statistical inference (with discussion). Journal of the Royal Statistical Society B 39, 119–171.

Xiong, S., Mu, W., 2009. On construction of asymptotically correct confidence intervals. Journal of Statistical Planning and Inference 139 (4), 1394–1404.

Yates, B. F., 1939. An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. Proc. Camb. Phil. Soc. 5, 579–591.

Zhao, S., Xu, X., Ding, X., 2012. Fiducial inference under nonparametric situations. Journal of Statistical Planning and Inference.