A resource bounded default logic

Gregory R. Wheeler[*]
Centro de Inteligência Artificial (CENTRIA)
Departamento de Informática, Universidade Nova de Lisboa
2829-516 Caparica, Portugal
greg@di.fct.unl.pt

Abstract

This paper presents *statistical default logic*, an expansion of classical (i.e., Reiter) default logic that allows us to model common inference patterns found in standard inferential statistics, including hypothesis testing and the estimation of a populations mean, variance and proportions. The logic replaces classical defaults with ordered pairs consisting of a Reiter default in the first coordinate and a real number within the unit interval in the second coordinate. This real number represents an upper-bound limit on the probability of accepting the consequent of an applied default and that consequent being false. A method for constructing extensions is then defined that preserves this upper bound on the probability of error under a (skeptical) non-monotonic consequence relation.

Introduction

This paper presents a resource bounded default logic. The motivation for the logic is found in a particular approach advanced in (Kyburg & Teng 1999), from the perspective of knowledge representation, of representing the structure of classical statistical inference (Fisher 1956; Lehman 1959) and statistical argumentation (Neyman 1957; Kyburg 1974; Mayo 1996) in terms of classical (Reiter 1980) defaults. In making statistical inferences—a term intended to include hypothesis testing and estimating basic parameters of populations, such as their means, proportions and variances—one accepts a conclusion along with a warning that there is a small, preassigned chance that the conclusion is false. The conditions that ensure the fit between a statistical model and the actual probability of error that one is exposed to by accepting the consequent are defeasible and the behavior of this defeasibility may be captured in terms of defaults, points developed in (Kyburg & Teng 1999).. What is missing from this representational scheme, however, is a means to explicitly record the probability of error and to preserve this measure under (non-monotonic) consequence (Wheeler 2002). This paper presents a default logic that achieves this aim.

[*]My thanks to Thomas Dietterich, Tony Hunter, Henry Kyburg, Choh Man Teng and Gabriel Uzquiano for discussion and comments on earlier drafts. This research was supported in part by the *Institute for Human and Machine Cognition* (IHMC), Pensacola, and grant SES 990-6128 from *The National Science Foundation*.

We will omit discussion of the merits of this approach from a knowledge representational perspective. From the perspective of this paper, statistical reasoning serves as a motivation for the logic. However, it should be noted that the modular construction of the logic—the sharp distinction between the (standard) operation of the fixed point operator, on the one hand, and the mechanism for preserving upper-bound estimates of probabilities on the other—lends itself as a technique, perhaps, to other circumstances calling for a non-monotonic consequence operation that returns a conclusion set whose members must satisfy an additional constraint regarding length of proof.

Statistical Defaults

Although a connection between statistical statements and default logic is developed in (Bacchus *et al.* 1993) and a connection between statistical inference and default logic is suggested in (Tan 1997), the first proposal to represent classical statistical inference in terms of defaults is (Kyburg & Teng 1999). A default is an inference rule of the form

$$\frac{\alpha : \beta_1, ..., \beta_n}{\gamma}, \tag{1}$$

interpreted roughly to mean that given α and the absence of any negated β_i's, conclude γ by default. Kyburg and Teng observe that default rule justifications—the formulae denoted by the β_i's in the default inference form—represent an important structural feature of classical statistical inference, namely the role randomization traditionally plays as a sufficient condition for a sample being representative (Cramér 1951; Moore 1979; Baird 1992).

There is a problem representing statistical inferences directly within Reiter's default logic, however, for representing classical statistical inference as a classical default appears to reward ignorance: given α and no other information whatsoever about how the sample was drawn other than that the sample is consistent with what is known, we may infer that the sample is representative by default. The problem is that there isn't a way for classical defaults to encode the information that would distinguish this feckless statistical inference from one that demands a rigorous attempt to detect a bias in the sample. Put another way, what is missing from Reiter defaults is an explicit means for controlling error.

A statistical default is an inference form that explicitly acknowledges the *upper limit* of its frequency of error.[1] Call a default in the form of

$$\frac{\alpha : \beta_1, ..., \beta_n}{\gamma}\epsilon, \tag{2}$$

an *s-default* and the upper limit on the frequency of error-parameter ϵ an *ϵ-bound* for short, where $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}$ is a Reiter default and $0 \leq \epsilon \leq 1$. The schema (2) is interpreted to say that provided α and no negated β_i's, γ is false no more than ϵ over the long-run application of that rule. (A Reiter default is a special case of a statistical default, namely when $\epsilon = 0$). A statistical default is sound just when the upper limit of the probability of error is *in fact* ϵ. An s-default is a good inference rule if it is sound and ϵ is relatively small, typically less than 0.05.

The parameter ϵ is an essential feature of statistical defaults, just as explicit error control is essential to classical statistical inference in general, and marks an important difference between statistical defaults and Reiter defaults.

The art of constructing a good statistical default, one whose advertised ϵ-bound is an acceptable level and in fact true, depends upon the background knowledge necessary to form the set of s-default justifications that determine the fit between a statistical model and its successful application. Disagreement about the soundness of a statistical default is then similar to disagreement about the soundness of a deductive inference in the sense that a distinction is made between the form of the inference and whether the constituents of the inference are (or not known to be) true. Treating ϵ as an explicit parameter in statistical default forms allows there to be an explicit constituent in the argument form that, much like a premise in a deductive argument, is true or false.

Statistical Default Theories

A statistical default theory is analogous to a default theory[2] in that statistical defaults appear in the object language and an s-default theory induces non-monotonic consequences via a fix-point operator. Given the parameter ϵ, however, how should accepted sentences interact in a statistical default extension? In Reiter's standard formulation, a default extension is deductively closed. But bounds for probability of error for particular statistical inferences do not necessarily carry over unchanged when chained together to form a statistical argument. Accepting a hypothesis H_1 with a confidence 0.95 and accepting another hypothesis H_2 independently with a confidence 0.95 doesn't entail the acceptance of H_1 and H_2 at 0.95. So, one cannot simply close a statistical default extension under deduction and guarantee that the deductive consequences will have the same ϵ-bound as a constituent inference.[3]

Another feature of Reiter's formulation is that an extension is closed under the set of defaults of a default theory.[4] Constructing a candidate default extension by successive applications of defaults results in a set of sentences, all of which are on par: all sentences in an extension are true. But we know that chaining statistical inferences typically *increases* our chance of error. So, again, there is no guarantee that the ϵ-bound of a statistical inference chain is identical to that of any of the constituent inferences. So we cannot close a statistical default extension under the default rules in D either.

The problem is that neither classical logic nor standard default logic restricts the length of permissible sequences of inferences. An inference chain that is deductively valid or one that generates an extension for a given default theory may nevertheless fail to represent an acceptable statistical argument because the error bound of that sequence of inference steps does not fall below the designated error bound for acceptance. So, since a conclusion of a statistical argument is acceptable only if the conclusion is the conclusion of an inference chain bounded in error by a preassigned ϵ, I must define a non-monotonic consequence relation that induces only those non-monotonic consequences that are within a designated error bound.

To allow for bounded inference chains, a *statistical default theory* is defined in terms of a set of s-defaults and a set of *bounded sentences*.

Definition 1. *Bounded sentence*: A sentence ϕ bounded by ϵ is an ordered pair $\langle \phi, \epsilon \rangle$, written $(\phi)_\epsilon$ for short, where ϕ is a sentence in the first-order language $\mathcal{L}$ and $\epsilon \in [0, 1]$. $(\phi)_\epsilon \equiv \phi$, if $\epsilon = 0$.

Definition 2. A *statistical default theory* Δ_s is an ordered pair $\langle W, S \rangle$, where W is a set of bounded sentences, and S a set of statistical defaults.

Statistical inference chains may now be defined with respect to a set of bounded sentences, Π. This is done in two

[1] A trivial corollary of the frequency of error $\hat{\alpha}$ for a statistical inference is the upper limit of the probability of error, denoted by ϵ. So, if $\hat{\alpha} = 0.03$ is understood to mean that the probability of committing a Type I error is 0.03, then $\epsilon = 0.03$ is understood to mean that the probability of committing a Type I error is no more than 0.03.

[2] A default theory is a pair $\langle W, D \rangle$ where W is a set of closed formulae and D is a (countable) set of defaults.

[3] This is a version of Henry Kyburg's lottery paradox (Kyburg 1961; 1997). The paradox is generated by holding three plausible assumptions about rational acceptance: i) If it is very likely that a hypothesis H is true, then it is rational to accept H; ii) If it is rational to accept H and it is rational to accept H', then it is rational to accept $H \wedge H'$; iii) It is irrational to accept any set of propositions that you know are inconsistent. A contradiction is generated by supposing a level $1 - \epsilon$ of rational acceptance that is very close to 1 (e.g., $\epsilon = 0.01$) and then considering a fair n-ticket lottery designed to yield exactly one winner, where $n > \frac{1-\epsilon}{\epsilon}$ (e.g., $n = 1000$). It is rational to accept that each ticket will lose. However, closing this set under conjunction entails that all tickets lose, which contradicts the supposition that one ticket will win.

[4] In standard default logic, an extension E is *closed under a default rule* $\alpha : \beta/\gamma$ if it is not the case that $\alpha \in E$ and $\neg\beta \notin E$ and $\gamma \notin E$; so, if $\alpha : \beta/\gamma$ applies to E then $\gamma \in E$. So an extension E is closed under a set of defaults D if E is closed under every default in D. A sentence γ is *a default conclusion* from E if there is a default $d = \frac{\alpha:\beta_1,...,\beta_n}{\gamma} \in D$ and d applies to E.

steps. Since statistical inference chains may include deductive or non-monotonic inference steps, the first step is to define bounded inference chains composed of purely non-monotonic inference steps (Theorem 1) and, in turn, to define bounded deductive inference chains (Theorem 2). This yields bounded closure conditions for each type of inference chain. From these results I then construct a bounded statistical default extension (Theorem 4).

Before turning to ϵ-bounded s-default closure conditions, it will be useful to collect the sentences appearing in all the pairs of bounded sentences in Π into a set—to crop the tails of the sequences in Π, if you will. Let $Crop(\Pi)$ name that function. So, for example, $Crop(\{(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}\}) = \{\phi_1, ..., \phi_n\}$.

Since rule applicability is relative to a specified bound of error ϵ, I will define a closure condition for statistical defaults on Π bounded by ϵ in the next three definitions. While Π here is treated as an arbitrary set of bounded sentences, when we turn to defining statistical default extensions we will use these operators on bounded-sentences that satisfy additional constraints.

Definition 3. *S-default ϵ-bounded conclusion*: A bounded sentence $(\gamma)_{\epsilon_\gamma}$ is an ϵ-bounded conclusion from a set of bounded sentences, Π, under a statistical default rule $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}\epsilon_s$ if and only if $(\alpha)_{\epsilon_\alpha} \in \Pi$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$, $\epsilon_\alpha + \epsilon_s = \epsilon_\gamma$ and $\epsilon_\gamma \leq \epsilon$.

Definition 4. A set Π of bounded sentences is closed under a particular s-default $s = \frac{\alpha:\beta_1,...,\beta_n}{\gamma}\epsilon_s$ within ϵ if and only if every bounded sentence $(\gamma)_{\epsilon_{\gamma_i}}$ that is an ϵ-bounded default conclusion from Π under s is a member of Π.

Definition 5. *S-default ϵ-bounded closure*: A set Π of bounded sentences is closed under a set S of s-defaults within ϵ if and only if, for every $s \in S$, Π is closed under s within ϵ. Let $Sn_\epsilon(S, \Pi)$, called the *ϵ-bounded s-default closure of* Π *with respect to* S, name an operator on S and Π that produces a set Π' of bounded sentences closed under S within ϵ. (When S is fixed by context I will simply write $Sn_\epsilon(\Pi)$.)

With the closure condition provided by Definition 5, I may now show that if a bounded sentence $(\gamma)_{\epsilon_\gamma}$ appears in the ϵ-bounded statistical default closure of some set of bounded sentences Π, then there is a chain of s-default inferences from Π that has $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded consequence.

Theorem 1 Let S be a set of statistical defaults, Π a set of bounded sentences, $(\gamma)_{\epsilon_\gamma}$ a bounded sentence and $Sn_\epsilon(\Pi)$ be the s-default closure of Π under S within ϵ. Define a statistical default inference chain on Π within ϵ as a sequence of bounded sentences, $\langle(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}\rangle$, such that $(\phi_i)_{\epsilon_{\phi_i}}$ is an ϵ-bounded conclusion from $\Pi \cup \{(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_{i-1})_{\epsilon_{\phi_{i-1}}}\}$, where $1 \leq i \leq n$. If $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$, then there is an s-default inference chain $\langle(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}, (\gamma)_{\epsilon_\gamma}\rangle$ on Π that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion.

Proof. An inductive proof on the length of inference chains is offered. Given an inference chain $\langle(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}\rangle$ on Π, let $\Pi_0 = \Pi, \Pi_1 = \Pi_0 \cup \{(\phi_1)_{\epsilon_{\phi_1}}\}, \Pi_2 = \Pi_1 \cup \{(\phi_2)_{\epsilon_{\phi_2}}\}$, and so on. Suppose $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$. Trivially, if $(\gamma)_{\epsilon_\gamma} \in \Pi$ then $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$. So suppose that $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$ but $(\gamma)_{\epsilon_\gamma} \notin \Pi_0$. By the definition of an ϵ-bounded conclusion from a set of bounded sentences, Π_0, there must be a particular statistical default rule $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}\epsilon_s$ such that $(\alpha)_{\epsilon_\alpha} \in \Pi_0$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi_0)$, $\epsilon_\alpha + \epsilon_s = \epsilon_\gamma$ and $\epsilon_\gamma \leq \epsilon$. Hence, $(\gamma)_{\epsilon_\gamma}$ is an ϵ-bounded s-default conclusion from Π_0. Hence $(\gamma)_{\epsilon_\gamma} \in \Pi_0 \cup \{(\gamma)_{\epsilon_\gamma}\}$. So there is an s-default inference chain, namely $\langle(\gamma)_{\epsilon_\gamma}\rangle$, that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion from Π. I now prove that if $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$ there is an inference chain of length i that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion from Π_i, for $i \geq 0$. By the definition of an ϵ-bounded conclusion from a set of bounded sentences, Π_i, there must be a particular statistical default rule $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}\epsilon_s$ such that $(\alpha)_{\epsilon_\alpha} \in \Pi \cup \{(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_i)_{\epsilon_{\phi_i}}\}$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi \cup (\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_i)_{\epsilon_{\phi_i}})$, $\epsilon_\alpha + \epsilon_s = \epsilon_\gamma$ and $\epsilon_\gamma \leq \epsilon$. Hence, $(\gamma)_{\epsilon_\gamma}$ is an ϵ-bounded s-default conclusion from Π_i, and $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi_i)$. Hence $(\gamma)_{\epsilon_\gamma} \in \Pi_i \cup \{(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_i)_{\epsilon_{\phi_i}}, (\gamma)_{\epsilon_\gamma}\}$. So there is an s-default inference chain, namely $\langle(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_i)_{\epsilon_{\phi_i}}, (\gamma)_{\epsilon_\gamma}\rangle$, that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion from Π. ■

In standard default theories, extensions are closed under logical consequence. I noted at the beginning of this section that this property won't hold for statistical default extensions since the premises of a deductive inference may themselves be s-default conclusions, each individually bounded in error by some ϵ (or other) such that the sum of error bounds of the premises is greater than ϵ.

Definition 6. Given a set of bounded sentences Π and an error parameter ϵ a bounded sentence $(\psi)_{\epsilon_\psi}$ is an *ϵ-bounded consequence* of Π, written $\Pi \Rightarrow_\epsilon (\psi)_{\epsilon_\psi}$ if and only if:

- $(\phi_1)_{\epsilon_1}, ..., (\phi_n)_{\epsilon_n} \in \Pi$,
- $\phi_1 \ldots \phi_n \vdash \psi$, and
- $\epsilon_\psi = \sum_{i=1}^{n} \epsilon_i \leq \epsilon$.

Definition 7. *ϵ-bounded logical closure:* For any set Π of bounded sentences, the operation $Cn_\epsilon(\Pi) = \{(\psi)_{\epsilon_\psi} : \Pi \Rightarrow_\epsilon (\psi)_{\epsilon_\psi}\}$.

I may now show similar results for ϵ-bounded deductive closure, namely that if the bounded sentence $(\gamma)_{\epsilon_\gamma}$ is in the image set of $Cn_\epsilon(\Pi)$, then there is a deductive inference chain defined on Π that has $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded, logical consequence of Π.

Theorem 2 Let Π be a set of bounded sentences, $(\gamma)_{\epsilon_\gamma}$ a bounded sentence and $Cn_\epsilon(\Pi)$ be the ϵ-bound closure of Π. Define a deductive inference chain as a sequence of ϵ-bounded sentences, $\langle(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_n)_{\epsilon_{\psi_n}}\rangle$ such that $(\psi_i)_{\epsilon_{\psi_i}}$ is an ϵ-bounded consequence of $\Pi \cup \{(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_{i-1})_{\epsilon_{\psi_{i-1}}}\}$, where $1 \leq i \leq n$. If $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi)$, then there is a deductive inference chain $\langle(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}, (\gamma)_{\epsilon_\gamma}\rangle$ of deductions on Π that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion.

Proof. An inductive proof on the length of deductive inference chains is offered. Given an inference chain $\langle(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_n)_{\epsilon_{\psi_n}}\rangle$ on Π, let $\Pi_0 = \Pi, \Pi_1 = \Pi_0 \cup \{(\psi_1)_{\epsilon_{\psi_1}}\}, \Pi_2 = \Pi_1 \cup \{(\psi_2)_{\epsilon_{\psi_2}}\}$, and so on. Suppose $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi)$. By the definition of an ϵ-bounded consequence from a set of bounded sentences, Π_0, if $\Pi_0 \Rightarrow_\epsilon (\gamma)_{\epsilon_\gamma}$ then there is a set of bounded sentences $(\phi_1)_{\epsilon_1}, ..., (\phi_n)_{\epsilon_n} \in \Pi_0$ such that $\phi_1 \ldots \phi_n \vdash \gamma$ and $\epsilon_\gamma = \sum_{i=0}^{n} \epsilon_i \leq \epsilon$. Hence, $(\gamma)_{\epsilon_\gamma}$ is an ϵ-bounded consequence from Π_0, and $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi_0)$. Hence $(\gamma)_{\epsilon_\gamma} \in \Pi_0 \cup \{(\gamma)_{\epsilon_\gamma}\}$. So there is a deductive inference chain, namely $\langle(\gamma)_{\epsilon_\gamma}\rangle$, that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion from Π. I now prove that if $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi)$ there is an inference chain of length i that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded consequence from Π_i, for $i \geq 0$. By the definition of an ϵ-bounded consequence from a set of bounded sentences, Π_i, if $\Pi_i \Rightarrow_\epsilon (\gamma)_{\epsilon_\gamma}$ then there is a set of bounded sentences $(\phi_1)_{\epsilon_1}, ..., (\phi_n)_{\epsilon_n} \in \Pi_i$ such that $\phi_1 \ldots \phi_n \vdash \gamma$ and $\epsilon_\gamma = \sum_{i=0}^{n} \epsilon_i \leq \epsilon$. Hence, $(\gamma)_{\epsilon_\gamma}$ is an ϵ-bounded consequence from Π_i, and $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi_i)$. Hence $(\gamma)_{\epsilon_\gamma} \in \Pi_i \cup \{(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_i)_{\epsilon_{\psi_i}}, (\gamma)_{\epsilon_\gamma}\}$. So there is an deductive inference chain, namely $\langle(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_i)_{\epsilon_{\psi_i}}, (\gamma)_{\epsilon_\gamma}\rangle$, that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded consequence from Π_i. ∎

Theorem 3 If Π is a set of 0-bounded sentences, Γ is set of sentences such that $Crop(\Pi) = \Gamma$, then $Cn_\epsilon(\Pi) = Cn(\Gamma)$.

Proof. By the definition of a bounded sentence, for any $(\phi)_\epsilon$, if $\epsilon = 0$ then $(\phi)_\epsilon \equiv \phi$. $\Gamma = Crop(\Pi)$, so there is a one-to-one mapping from every $(\phi_i)_\epsilon \in \Pi$ to every $\phi_i \in \Gamma$ such that $(\phi_i)_\epsilon \equiv \phi_i$. Since for any n membered subset of Π $\sum_{i=0}^{n} \epsilon_i = 0$, then $\Pi \Rightarrow_\epsilon (\psi)_{\epsilon_\psi}$ if and only if $\Gamma \vdash \psi$. ∎

For each of the closure operations Sn_ϵ and Cn_ϵ and any formula ϕ that is in the cropped image set of either operation, the error bound parameter ϵ_ϕ of ϕ is the sum of the error bounds of the constituents participating in the immediate inference step terminating in $(\phi)_{\epsilon_\phi}$. If the bound ϵ_ϕ is less than or equal to the designated bound ϵ, then ϵ_ϕ is included in the extension.

Summing error bounds of constituent inference steps is an imprecise but conservative estimate of the probability of accepting a false conclusion in an n-element inference chain and thereby provides an upper bound on possible error for sound inference chains. Since $\hat{\alpha}$ is by definition the probability of accepting that the outcome of trial (inference) X is true when in fact it is false, let $\hat{\alpha} = 1 - \Pr(X) = \Pr(\bar{X})$. The justification for stating that summing error bounds provides an upper bound on probability of error for a chain of inferences is that we can face no greater risk than the risk of mistakenly accepting each conclusion drawn in a line of reasoning. Summing the error bound parameters provides this upper bound because of a theorem of classical probability, namely that $\Pr(\bar{X} \cup \bar{X}') = \Pr(\bar{X}) + \Pr(\bar{X}') - \Pr(\bar{X}, \bar{X}')$, which holds generally for events $X_1, ..., X_n$. *Boole's inequality*, $\Pr(\bar{X} \cup \bar{X}') \leq \Pr(\bar{X}) + \Pr(\bar{X}')$, follows trivially.

Notice that this procedure of summing ϵ does not yield a probability measure but rather a conservative estimate of $\hat{\alpha}$. We needn't be concerned that the sum of ϵ-bounds is not necessarily a probability, since, as we will soon see, for the sequences of inferences that will be of interest are only those where ϵ takes a value within the unit interval.

When the event in question is accepting the sentence ϕ and the probabilities in question are a measure of the upper bound on the probability of accepting ϕ when ϕ is false, then we may infer that we are exposed to no greater risk of being wrong about ϕ than by accepting the sum of the probabilities of mistakenly accepting each of ϕ's constituents.[5]

What remains is defining a statistical default extension in terms of a set of bounded sentences Π, an error parameter ϵ, and the pair of ϵ-bounded closure conditions Sn_ϵ and Cn_ϵ for statistical defaults and bounded consequence.

A candidate statistical default extension is constructed sequentially, much like a candidate standard default extension. A candidate set on a default theory $\langle W, D\rangle$ is built sequentially by first closing W under consequence, applying all applicable defaults in D to the set of consequences of W, closing that set under consequence, and so on. While a standard default extension candidate is built sequentially by alternatingly closing the set under consequence and the set of defaults until no more defaults can be applied, statistical default extension candidates are built by alternatingly closing a set of bounded sentences under consequence (bounded by a specified threshold parameter ϵ) and the set of statistical defaults (also bounded by ϵ) until no more deductive or default inferences can be made at or below ϵ.

Given a statistical default theory Δ_s I wish to define a statistical default extension Π on Δ_s at ϵ. I offer the following definition.

Definition 8. Where $\Delta_s = \langle W, S\rangle$ at ϵ is a statistical default theory and Π is some set of bounded sentences, let $\mathsf{E}_{\Delta_S}(\Pi)$ be a minimal set satisfying three conditions:

[SD1.] $W \subseteq \mathsf{E}_{\Delta_S}(\Pi)$.
[SD2.] $Cn_\epsilon(\mathsf{E}_{\Delta_S}(\Pi)) = \mathsf{E}_{\Delta_S}(\Pi)$.
[SD3.] $\mathsf{E}_{\Delta_S}(\Pi)$ is closed under S within ϵ, i.e. for all $\frac{(\alpha)_{\epsilon_\alpha} : (\beta_1)_{\epsilon_1}, ..., (\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s \in S$, $(\alpha)_{\epsilon_\alpha} \in \mathsf{E}_{\Delta_S}(\Pi)$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$, $\epsilon_\alpha + \epsilon_s = \epsilon_\gamma$ and $\epsilon_\gamma \leq \epsilon$.

A set of bounded sentences Π is a *statistical extension* for Δ_s at ϵ iff $\mathsf{E}_{\Delta_S}(\Pi) = \Pi$.

Theorem 4 Let Π be a set of bounded sentences, let $(\alpha)_{\epsilon_1}, (\beta)_{\epsilon_2}, (\gamma)_{\epsilon_3}, (\phi)_{\epsilon_4}$, and $(\psi)_{\epsilon_5}$ be ϵ_i-bounded counterparts to sentences $\alpha, \beta, \gamma, \phi$,and ψ in $\mathcal{L}$, and let $\Delta_S = \langle W, S\rangle$ at ϵ be a closed statistical default theory. Define

- *For all* $(\phi_i)_{\epsilon_{\phi_i}} \in W, \epsilon_{\phi_i} = 0$.

[5] But is this measure too conservative? Compare the proposal to sum error bounds to the Bonferroni adjustment (Holm 1979), an adjustment applied to the $\hat{\alpha}$ levels of multiple hypothesis tests performed on the same data set. The adjustment works by dividing the accepted $\hat{\alpha}$ level by the number n of tests performed. The result is that the significance of any one test would need to be $\frac{\hat{\alpha}}{n}$ Unlike the Bonferroni adjustment, the proposal here preserves the significance of each individual test yet retains the property that the sequence of tests has an $\hat{\alpha}$ level no more than the assigned ϵ-bound for acceptance. Hence, the proposal to sum error bounds, while conservative, is not as restrictive as the Bonferroni method.

- $\Pi_0 = W$, and for $i \geq 0$,
- $\Pi_{i+1} = Cn_\epsilon(\Pi_i) \cup \{\gamma | \frac{(\alpha)_{\epsilon_\alpha}:(\beta_1)_{\epsilon_1},...,(\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s) \in S$, where $(\alpha)_{\epsilon_\alpha} \in \Pi_i$ and $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$ and $\epsilon_\alpha + \epsilon_s \leq \epsilon\}$.

Then Π is a statistical extension for Δ_S at ϵ iff $\Pi = \bigcup_{0\leq i\leq\infty}\Pi_i$

Proof. Begin by observing that $\bigcup_{0\leq i\leq\infty}\Pi_i$ enjoys the following properties:

[SD1′.] $W \subseteq \bigcup_{0\leq i\leq\infty}\Pi_i$

[SD2′.] $Cn_\epsilon\left(\bigcup_{0\leq i\leq\infty}\Pi_i\right) = \bigcup_{0\leq i\leq\infty}\Pi_i$

[SD3′.] If $\frac{(\alpha)_{\epsilon_\alpha}:(\beta_1)_{\epsilon_1},...,(\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s \in S$ and $(\alpha)_{\epsilon_\alpha} \in \bigcup_{0\leq i\leq\infty}\Pi_i$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$ and $\epsilon_s + \epsilon_\alpha \leq \epsilon$ then $(\gamma)_{\epsilon_s+\epsilon_\alpha} \in \bigcup_{0\leq i\leq\infty}\Pi_i$.

So, by the minimality of $\mathsf{E}_\Delta(\Pi)$, we have

$$\mathsf{E}_\Delta(\Pi) \subseteq \bigcup_{0\leq i\leq\infty}\Pi_i. \quad (3)$$

($\Rightarrow$) Proof by induction that $\Pi_i \subseteq \Pi$ at ϵ for all $0 \leq i \leq \infty$ from $\bigcup_{0\leq i\leq\infty}\Pi_i \subseteq \Pi$. For short, '$\Pi$' will stand for '$\Pi$ at ϵ'. Clearly $\Pi_0 \subseteq \Pi$, since $\Pi = \mathsf{E}_\Delta(\Pi)$. Assume some $\Pi_i \subseteq \Pi$ and consider $(\gamma)_{\epsilon_\gamma} \in \Pi_{i+1}$. If $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi_i)$, then since $\Pi_i \subseteq \Pi$ we have $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi) = \Pi$. Otherwise there is a default $\frac{(\alpha)_{\epsilon_\alpha}:(\beta_1)_{\epsilon_1},...,(\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s \in S$, where $(\alpha)_{\epsilon_\alpha} \in \Pi_i$ and $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$ and $\epsilon_\alpha + \epsilon_s \leq \epsilon$. Therefore, since $\Pi_i \subseteq \Pi$, $(\alpha)_{\epsilon_\alpha} \in \Pi = \mathsf{E}_\Delta(\Pi)$. Hence, by SD3 $(\gamma)_{\epsilon_\gamma=\epsilon_s+\epsilon_\alpha\leq\epsilon} \in \mathsf{E}_\Delta(\Pi)$. Furthermore, since $\mathsf{E}_\Delta(\Pi) = \Pi$, we have $(\gamma)_{\epsilon_\gamma} \in \Pi$. Therefore, $\bigcup_{0\leq i\leq\infty}\Pi_i \subseteq \Pi$. By equation (3) and the fact that $\Pi = \mathsf{E}_\Delta(\Pi)$ because of the definition of a fixed point, we have $\Pi = \bigcup_{0\leq i\leq\infty}\Pi_i$.

($\Leftarrow$) Proof by induction that $\Pi_i \subseteq \mathsf{E}_\Delta(\Pi)$ for all $i \leq i \leq \infty$ from $\Pi = \bigcup_{0\leq i\leq\infty}\Pi_i \subseteq \mathsf{E}_\Delta(\Pi)$. By appealing to equation (3) we will then have $\Pi = \mathsf{E}_\Delta(\Pi)$ from whence Π is an extension of Δ_s at ϵ. Clearly $\Pi_0 \subseteq \mathsf{E}_\Delta(\Pi)$, so assume $\Pi_i \subseteq \mathsf{E}_\Delta(\Pi)$ and consider $(\gamma)_{\epsilon_\gamma} \in \Pi_i$. If $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi_i)$, then since $\Pi_i \in \mathsf{E}_\Delta(\Pi)$ we have $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\mathsf{E}_\Delta(\Pi)) = \mathsf{E}_\Delta(\Pi)$. Otherwise there is a default $\frac{(\alpha)_{\epsilon_\alpha}:(\beta_1)_{\epsilon_1},...,(\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s \in S$ where $(\alpha)_{\epsilon_\alpha} \in \Pi_i$ and $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$, but $\epsilon_\gamma = \epsilon_\alpha + \epsilon_s\epsilon$ or $(\gamma)_{\epsilon_\gamma} \notin \Pi_i$. Then since $\Pi_i \subseteq \mathsf{E}_\Delta(\Pi)$, $(\alpha)_{\epsilon_\alpha} \in \mathsf{E}_\Delta(\Pi)$. Hence $(\gamma)_{\epsilon_\gamma=\epsilon_\alpha+\epsilon_s\leq\epsilon} \in \mathsf{E}_\Delta(\Pi)$ by SD3. Hence $\Pi_{i+1} \subseteq \mathsf{E}_\Delta(\Pi)$. ■

Statistical Default Consequence: $\mid\!\sim_\epsilon$

I first make some observations about statistical default extensions.

A statistical default extension contains the set W representing the uncontested (i.e., $\epsilon = 0$) world knowledge of the statistical default theory, along with the consequences inferable by logical deduction and s-default rule chaining within the prescribed error bound ϵ. The ϵ-bounded logical closure of the set of uncontested world knowledge W is equivalent to the classical closure of the cropped sentences of W, since, by Theorem 2 and Theorem 3 if Γ is the set such that $\Gamma = Crop(W)$, then $Cn(\Gamma) = Cn_\epsilon(W)$.

A consequence of this definition is that only Π_0 is equivalent to a deductively closed set of propositions and none of the Π_i's for $i > 0$ is closed under either deduction or conjunction.

Note also that for $\epsilon = 0$, the statistical default extensions Π_i of $\Delta_s = \langle W, S\rangle$ at 0 are identical to the standard default extensions of the standard default theory $\Delta = \langle W, D\rangle$, modulo Theorem 2, where D is the set of all default rules in S of confidence 1 ($\epsilon = 0$).[6]

To illustrate how statistical default extensions are constructed, consider the following example.

Example 1 Let $\Delta^1_s = \langle W, S\rangle$ be a statistical default theory, where $W = \emptyset$ and S contains four s-defaults:
$S = \{ \frac{:A}{A}0.01, \frac{:B}{B}0.01, \frac{A:B,C}{C}0.01, \frac{A\wedge B:\neg C}{\neg C}\ 0.01 \}$
For an error bound parameter $\epsilon_1 = 0.02$, there is one statistical default extension Π^1 where $Crop(\Pi^1)$ contains

$$A, B, A \wedge B, C.$$

The bounded sentence A at ϵ_A is included in extension Π^1 by applying the default $\frac{:A}{A}$ and bounded sentence B at ϵ_B is included by applying the default $\frac{:B}{B}$, where each inference has an error bound of 0.01, so $(A)_{0.01}$ and $(B)_{0.01}$. $(A \wedge B)_{\epsilon_{A\wedge B}}$ is included in the extension, since the sum of the error bounds of conjoining A and B is 0.02, that is $(A \wedge B)_{0.02}$. The bounded sentence C at ϵ_C is included by using A, whose error bound is 0.01, to apply the default $\frac{A:B,C}{C}$, whose error bound is also 0.01. Hence $(C)_{0.02}$. The default $\frac{A\wedge B:\neg C}{\neg C}$ cannot be applied because the resulting conclusion $\neg C$ would have an error bound of 0.03, $(\neg C)_{0.03}$ which is above the designated threshold $\epsilon_1 = 0.02$.

For a threshold parameter $\epsilon_2 = 0.03$, there are two statistical default extensions Π^1, which is the same as described above, and Π^2, where $Crop(\Pi^2)$ contains

$$A, B, A \wedge B, \neg C.$$

The default rule that could not be applied before is now applicable with respect to ϵ_2, giving rise to the second (partial) extension Π^2.[7]

Example 2 Let $\Delta^2_s = \langle W_2, S_2\rangle$ be a statistical default theory, where $W_2 = \emptyset$ and $S_2 = \{ \frac{:\neg B,C}{C}0.00, \frac{:C}{C}0.02, \frac{C:B}{B}0.01, \frac{:\neg B}{\neg B}0.03, \frac{:\neg B,A}{A}0.01, \frac{:\neg A}{\neg A}0.01 \}$.

For an error-bound parameter $\epsilon = 0.02$, there is no statistical default extension, since while both $\frac{:\neg B,C}{C}0.00$, $\frac{:C}{C}0.02$ yield C only the bounded sentence $\langle C, 0.00\rangle$ from $\frac{:\neg B,C}{C}0.00$ may

[6] Recall that statistical defaults are standard defaults when $\epsilon = 0$.

[7] The complete cropped extensions: when $\epsilon = 0.02$, Π^1= {A,B,A$\wedge B, C$}; when $\epsilon = 0.03$, $\Pi^1 = \{A, B, A \wedge B, C, A \wedge C, B \wedge C\}$ and Π^2 ={A,B,A$\wedge B, \neg C$}.

be substituted for the antecedent of $\frac{C:B}{B}0.01$ which in turn is applicable in extensions consistent with B. But $\frac{:\neg B, C}{C}0.00$ is applicable only in extensions consistent with $\neg B$.

For an error-bound parameter $\epsilon = 0.03$, there are three extensions. Because this example highlights the role that error bounds play in constructing extensions we display the partial extensions first in uncropped form, then in cropped form.

$\Pi^1 \supset \{\langle C, 0.00\rangle, \langle C, 0.02\rangle, \langle \neg B, 0.03\rangle, \langle A, 0.01\rangle\}$
$\Pi^2 \supset \{\langle C, 0.00\rangle, \langle C, 0.02\rangle, \langle \neg B, 0.03\rangle, \langle \neg A, 0.01\rangle\}$
$\Pi^3 \supset \{\langle C, 0.02\rangle, \langle B, 0.01\rangle, \langle \neg A, 0.01\rangle\}$

And the three corresponding cropped partial extensions at $\epsilon = 0.03$ are:

$Crop(\Pi^1) \supset \{C, \neg B, A\}$
$Crop(\Pi^2) \supset \{C, \neg B, \neg A\}$
$Crop(\Pi^3) \supset \{C, B, \neg A\}$

Like their default logic counterparts, it is not uncommon for statistical default theories to have multiple extensions.

Corollary 1 *The family of statistical default consequence operators are non-monotonic*. If $\Delta_s = \langle W, S\rangle$ is a statistical default theory at ϵ with an extension Π, S' is a non-empty set of defaults and W' a non-empty set of bounded sentences, then $\Delta'_s = \langle S \cup S', W \cup W'\rangle$ at ϵ may have no cropped extension Π' such that $\Pi \subseteq \Pi'$.

Proof. Suppose $\frac{(\alpha)_{\epsilon_\alpha} : (\beta_1)_{\epsilon_1}, \ldots, (\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s \in S$, $(\alpha)_{\epsilon_\alpha} \in W$ and no $\neg\beta_i$'s are in $Crop(\Pi)$. Hence, $\gamma \in Crop(\Pi)$. Now suppose a β_i in W' or as a consequent of an applied default in S'. Then $\gamma \notin \Pi'$, so $\Pi \not\subseteq \Pi'$. ■

A sentence A is a skeptical consequence of a default theory Δ just in case A is in every extension for Δ. I now define an analogous consequence relation for statistical default theories.

Definition 9. *Skeptical Statistical Consequence*: Let $\Delta_s = \langle W, S\rangle$ be a statistical default theory at ϵ and A a sentence. Then A is a skeptical consequence of Δ_s at ϵ—written, $\Delta \mid\!\sim_\epsilon A$—just in case $A \in Crop(\Pi)$ for each extension Π on Δ_s at ϵ.

The consequence relation $\mid\!\sim_\epsilon$ is non-monotonic. Notice, from Corollary 1, that by either augmenting the set of bounded-sentences in the W-component of a default theory or adding new default rules to the S-component a previously induced statistical consequence (at a particular error bound) may then fail to remain supported by the statistical default theory (at that particular error bound).

Corollary 2. *Skeptical statistical default consequence is supra-classical*. If $\Delta_s = \langle W, S\rangle$ is a statistical default theory at ϵ and the set S of s-defaults is empty, then there is one extension of Δ_s which is identical to $Cn(Crop(W))$.

Proof. By the definitions of an s-default extension, an extension $\Pi_0 = W$ and $\Pi_{i+1} = Cn_\epsilon(\Pi_i) \cup \{\gamma | \frac{(\alpha)_{\epsilon_\alpha} : (\beta_1)_{\epsilon_1}, \ldots, (\beta_n)_{\epsilon_n}}{\gamma}\epsilon_s\} \in S$, where $(\alpha)_{\epsilon_\alpha} \in \Pi_i$ and $\neg\beta_1, \ldots, \neg\beta_n \notin Crop(\Pi)$ and $\epsilon_\alpha + \epsilon_s \le \epsilon\}$.. But $S = \emptyset$ by hypothesis, so $Cn(Crop(W))$ for all Π_i. ■

Conclusion

In this paper I've presented a non-monotonic framework for representing arguments composed, at least in part, of classical statistical inferences. Building on Kyburg and Teng's observation that Reiter defaults mirror the general structure of classical statistical inference, an extension of default logic was proposed called *statistical default logic*. It was noted that statistical default logic admits of a skeptical consequence relation that sanctions a single conclusion set bounded in error by ϵ and that this consequence relation is non-monotonic. Thus statistical default logic features a means for preserving the property of a monotonically decreasing measure of the upper limit of the probability of error associated with a sequence of statistical reasoning and yet preserves the non-monotonic behavior of statistical inference.

References

Bacchus, F. A.; Grove, D.; Halpern, J.; and Koller, D. 1993. Statistical foundations for default reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 563–569.

Baird, D. 1992. *Inductive Logic: Probability and Statistics*. Englewood Cliffs, NJ: Prentice Hall.

Besnard, P., and Siegal, P. 1988. The preferential-models approach to nonmonotonic logics. In et. al., P. S., ed., *Non-standard Logics for Automated Reasoning*. London: Academic Press. 137–161.

Brewka, G. 1991. Cumulative default logic: In defense of nonmonotonic inference rules. *Artificial Intelligence* 50:183–205.

Carnap, R. 1962. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 2nd edition.

Cramér, H. 1951. *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Delgrande, J.; Schaub, T.; and Jackson, W. K. 1994. Alternative approaches to default logic. *Artificial Intelligence* 70:167–237.

Fisher, R. 1956. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70.

Kraus, S.; Lehman, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167–207.

Kyburg, Jr., H. E., and Teng, C. M. 1999. Statistical inference as default logic. *International Journal of Pattern Recognition and Artificial Intelligence* 13(2):267–283.

Kyburg, Jr., H. E. 1961. *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press.

Kyburg, Jr., H. E. 1974. *The Logical Foundations of Statistical Inference*. Dordecht: D. Reidel.

Kyburg, Jr., H. E. 1997. The rule of adjunction and reasonable inference. *Journal of Philosophy* 94:109–125.

Lehman, E. L. 1959. *Testing Statistical Hypotheses*. New York: John Wiley.

Levi, I. 1967. Probability kinematics. *The British Journal for the Philosophy of Science* 18:197–209.

Łukaszewicz, W. 1988. Considerations on default logic: an alternative approach. *Computational Intelligence* 4(1):1–16.

Mayo, D. 1996. *Error and the Growth of Knowledge*. Chicago: University of Chicago Press.

Moore, D. 1979. *Statistics*. San Francisco: W. H. Freeman Press.

Neyman, J. 1957. Inductive behavior as a concept of philosophy of science. *Review of the International Statistical Institute* 25:5–22.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13:81–132.

Tan, Y. H. 1997. Default logic and inductive-statistical reasoning. *Synthese* 110:357–379.

Wheeler, G. R. 2002. *Non-monotonicity and Paraconsistency*. Ph.D. Dissertation, Depts. of Philosophy and Computer Science, University of Rochester, New York.