ERRATUM

Erratum to: Model change and methodological virtues in scientific inference

Erich Kummerfeld · David Danks

Erratum to: Synthese
      DOI 10.1007/s11229-014-0408-3

Appendix 1: Notation

Let X represent a sequence of data, and let X_B^t represent an i.i.d. subsequence of length t of data generated from distribution B.[1] Let $\mathbf{F}$ be a framework (in this case, a set of probability distributions or densities).[2] Let $M_{\mathbf{F}}$ be a method that takes a data sequence X as input and outputs a distribution $B \in \mathbf{F}$; we will typically drop the subscript $\mathbf{F}$ from M as we will be dealing with a single framework at a time. Concretely, $M[X_B^t] = O$ means that M outputs O after observing the sequence X_B^t. Let D be a distance metric over distributions (e.g., the Anderson-Darling test). Let $D_\delta(A, B)$ be shorthand for the following inequality: $D(A, B) < \delta$. Finally, let $[X, Y]$ denote the concatenation of sequence X with sequence Y.

---

[1] We conjecture that the i.i.d. assumption could be eliminated by defining probability distributions over sequences of arbitrary length, though this complication would not add conceptual clarity.

[2] Let any $P( )$ functions be either a probability distribution function or probability density function, as appropriate.

The online version of the original article can be found under doi:10.1007/s11229-014-0408-3.

E. Kummerfeld (✉) · D. Danks
Carnegie Mellon University, Pittsburgh, PA 15213, USA
e-mail: ekummerfeld@gmail.com; ekummerf@andrew.cmu.edu

D. Danks
e-mail: ddanks@andrew.cmu.edu

    🠦 Springer

Definition A distribution A is *absolutely continuous* with respect to another distribution B iff $\forall x\ P_B(x) = 0 \Rightarrow P_A(x) = 0$.[3] That is, if B gives probability 0 to some event x, then A also gives probability 0 to that same event. Let $AC(B)$ be the set of distributions which are absolutely continuous with respect to B except for B itself. Let $AC(B, \delta)$ be as $AC(B)$ except restricted to those distributions which are more than distance $\delta > 0$ away from B.

Definition An estimator M is *consistent* if $\forall B \in \mathbf{F}\forall \delta > 0\ \lim_{n\to\infty} P_B(D_\delta(M[X_B^n], B))=1$. That is, for all distributions in the framework, the probability that M's output is arbitrarily close to the target distribution approaches 1 as the amount of data increases to infinity.

Definition An estimator M can be forced to make *arbitrary errors* if $\forall B_1 \in \mathbf{F}\ \forall \delta > 0\forall B_2 \in AC(B_1, \delta)\cap\mathbf{F}\forall\epsilon > 0\ \forall n_2\exists n_1\ P_{B_1,n_1,B_2,n_2}(D_\delta(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)) \leq \epsilon$. That is, consider any distribution B_2 which is in the framework, is absolutely continuous with respect to B_1, and is more than δ away from B_1 (though there might be no such distribution). Then for any amount of data n_2 from B_2, there is an amount of data n_1 from B_1 such that M's output will still be arbitrarily unlikely to be arbitrarily close to B_2 after seeing the $n_1 + n_2$ data.

Appendix 2: Lemma: Consistency $\Rightarrow$ Arbitrary Errors (within AC)

Proof We prove the contrapositive. If we assume M does not make arbitrary errors and pass the negation through all the quantifiers, then we have:

$$\exists B_1 \in \mathbf{F}\ \exists \delta > 0\exists B_2 \in AC(B_1, \delta)\cap\mathbf{F}\ \exists\epsilon > 0\ \exists n_2\forall n_1\ P_{B_1,n_1,B_2,n_2}(D_\delta(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)) > \epsilon$$

Define $v(B_1, B_2, \delta) = D(B_1, B_2) - \delta$; for convenience, we omit the arguments to v. In general, $v > 0$ since $B_2 \in AC(B_1, \delta)$. Note that $v + \delta = D(B_1, B_2)$. Since D is a distance, the triangle inequality holds for it, so for any other distribution $C \in \mathbf{F}$, $D(B_1, B_2) \leq D(C, B_1)+D(C, B_2)$. It follows that $v+\delta \leq D(C, B_1)+D(C, B_2)$. Consider the case where $C = M[X_{B_1}^{n_1}, X_{B_2}^{n_2}]$. For this inequality to be satisfied, it must be the case that if $D_\delta(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)$ is true (i.e. $D(C, B_2) < \delta$), then $D_v(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_1)$ is false (i.e. $D(C, B_1) > v$). The fully quantified inequality above thus entails a statement about the distance of M's output from B_1:

$$\exists B_1 \in \mathbf{F}\ \exists \delta > 0\exists B_2 \in AC(B_1, \delta)\cap\mathbf{F}\ \exists\epsilon > 0\ \exists n_2\forall n_1\ P_{B_1,n_1,B_2,n_2}(D_v(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_1)) < 1-\epsilon$$

Since this inequality holds for all n_1, we know: $\lim_{n_1\to\infty} P_{B_1,n_1,B_2,n_2}(D_v(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_1)) \leq 1 - \epsilon$.

The probability distribution here depends on the probabilities of sequences. Let $Y = [X_{B_1}^{n_1}, X_{B_2}^{n_2}]$, and let Y_i be the i'th element of Y. Since we have i.i.d. samples in each subsequence, we have: $P_{B_1,n_1,B_2,n_2}([X_{B_1}^{n_1}, X_{B_2}^{n_2}]) = \prod_{i=1}^{n_1} P_{B_1}(Y_i) \prod_{j=1}^{n_2} P_{B_2}(Y_{n_1+j})$.

---

[3] Absolute continuity is typically defined in terms of measures rather than distributions, but we need only the more specific notion.

Note that since $B_2 \in AC(B_1)$, if $\prod_{j=1}^{n_2} P_{B_2}(Y_{n_1+j}) > 0$, then $\prod_{j=1}^{n_2} P_{B_1}(Y_{n_1+j}) > 0$. Therefore, for fixed n_2, as $n_1 \to \infty$, this product converges to $\prod_{i=1}^{n_1} P_{B_1}(Y_i) \prod_{j=1}^{n_2} P_{B_1}(Y_{n_1+j}) = \prod_{i=1}^{n_1+n_2} P_{B_1}(Y_i) = P_{B_1}([X_{B_1}^{n_1+n_2}])$.

Because the two distributions over sequences are the same in the limit, we can conclude $\lim_{n_1 \to \infty} P_{B_1,n_1,B_2,n_2}(D_\nu(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_1)) = \lim_{n_1 \to \infty} P_{B_1}(D_\nu(M[X_{B_1}^{n_1+n_2}], B_1))$. Combining this with the previous inequality yields: $\lim_{n_1 \to \infty} P_{B_1}(D_\nu(M[X_{B_1}^{n_1+n_2}], B_1)) \leq 1 - \epsilon$. Since $\epsilon > 0$, this implies that $\exists B \in \mathbf{F} \, \exists \delta^* > 0 \lim_{n \to \infty} P_B(D_{\delta^*}(M[X_B^n], B)) \neq 1$ (where $\delta^* = \nu$). Hence, M is not consistent. $\qquad \square$

Appendix 3: Construction: Diligence $\Rightarrow \neg$ Arbitrary Errors

We construct the formal definition of diligence from that of "arbitrary errors" (AE) in a way that makes it clear that diligent methods are not subject to arbitrary errors. The negation of AE is:

$$\exists B_1 \in \mathbf{F} \, \exists \delta > 0 \exists B_2 \in AC(B_1, \delta) \cap \mathbf{F} \, \exists \epsilon > 0 \, \exists n_2 \forall n_1 \; P_{B_1,n_1,B_2,n_2}(D_\delta(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)) > \epsilon$$

This condition is, however, insufficiently weak to capture diligence, as we want to avoid such errors for *all* pairs of distributions in the framework, not just for some absolutely continuous pair. We thus strengthen the negation of AE by converting the three leading existential quantifiers into universal quantifiers and extending the domain of the universal quantifier over B_2 to include those distributions which are not absolutely continuous with respect to B_1:

Definition An estimator M is *diligent* if

$$\forall B_1 \in \mathbf{F} \, \forall \delta > 0 \forall B_2 \in \mathbf{F} \backslash B_1 \, \exists \epsilon > 0 \, \exists n_2 \forall n_1 \; P_{B_1,n_1,B_2,n_2}(D_\delta(M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)) > \epsilon.$$

That is, for any pair of distributions in the framework, there is an amount of data n_2 from B_2 such that M's output will be arbitrarily close to B_2 with positive probability after seeing $n_1 + n_2$ data, for *any* amount of data n_1 from B_1.

Definition A framework $\mathbf{F}$ is *nontrivial* iff there exists some $B \in \mathbf{F}$ such that $AC(B) \cap \mathbf{F} \neq \emptyset$.

Clearly, diligence implies the negation of AE for all nontrivial frameworks. We thus have the key theorem for this paper:

Theorem No statistical estimator for a (nontrivial) framework is both consistent and diligent.

Proof Assume M is both consistent and diligent. Its consistency implies that AE holds for it. Its diligence, along with the nontriviality of the framework, implies that $\neg$AE holds for it. Contradiction, and so no M can be both consistent and diligent for a nontrivial framework. $\qquad \square$

Appendix 4: Generalizing Diligence

A natural generalization of diligence yields a novel methodological virtue: Uniform Diligence. Uniform diligence is a strengthening of regular (pointwise) diligence in the same way that uniform consistency is a strengthening of pointwise consistency. Instead of requiring only that, for each B_1, B_2 and δ, there be some n_2, Uniform Diligence requires that there be some n_2 which works for *all* such combinations.

Definition An estimator M is *uniformly diligent* if

$$\exists n_2 \forall B_1 \in \mathbf{F} \, \forall \delta > 0 \, \forall B_2 \in \mathbf{F} \backslash B_1 \, \exists \epsilon > 0 \, \forall n_1 \, P_{B_1, n_1, B_2, n_2} (D_\delta (M[X_{B_1}^{n_1}, X_{B_2}^{n_2}], B_2)) > \epsilon.$$

Obviously, consistency and uniform diligence are also incompatible, as the latter is a strengthening of diligence. The following chart shows three different ways of ordering the quantifiers in the definition of Diligence, producing methodological virtues of varying strength. The weakest, Responsiveness, is not incompatible with consistency. For space and clarity, $\mathbf{B}$ is used in place of $\forall B_1 \in \mathbf{F} \, \forall \delta > 0 \, \forall B_2 \in \mathbf{F} \setminus B_1 \, \exists \epsilon > 0$.

Responsiveness	Diligence	Uniform Diligence
$\mathbf{B} \forall n_1 \exists n_2$	$\mathbf{B} \exists n_2 \forall n_1$	$\exists n_2 \mathbf{B} \forall n_1$