

Direct Inference from Imprecise Frequencies¹

Paul D. Thorn

Abstract: It is well known that there are, at least, two sorts of cases where one should not prefer a direct inference based on a narrower reference class, in particular: cases where the narrower reference class is gerrymandered, and cases where one lacks an evidential basis for forming a precise-valued frequency judgment for the narrower reference class. I here propose (1) that the preceding exceptions exhaust the circumstances where one should not prefer direct inference based on a narrower reference class, and (2) that minimal frequency information for a narrower (non-gerrymandered) reference class is sufficient to yield the defeat of a direct inference for a broader reference class. By the application of a method for inferring relatively informative expected frequencies, I argue that the latter claim does not result in an overly incredulous approach to direct inference. The method introduced here permits one to infer a relatively informative expected frequency for a reference class R' , given frequency information for a superset of R' and/or frequency information for a sample drawn from R' .

Keywords: Direct inference, Statistical syllogism, Reference class problem, Imprecise probability.

1 Introduction

Instances of direct inference proceed from two premises. The first premise states that a given object, c , is an element of a given set, R (the reference class). The second (major) premise states something about the relative frequency of elements of R among another set, T (the target class). Typically, the major premise states that the relative frequency of T among R is some value, r (though other sorts of statistical statement, e.g., imprecise-valued frequencies or expected frequencies, might also serve). The conclusion of the direct inference, in the typical case, is then that the probability that c is in T is also r . In order to permit a concise expression of such direct inferences, I use “PROB” to denote the personal probabilities that are rational for a given agent, given the agent’s evidence, and “freq” to denote a function that takes pairs of sets and returns the relative frequency of the first set among the second. Given these conventions, typical instances of direct inference satisfy the following schema:

From $c \in R$ and $\text{freq}(T|R) = r$ infer that $\text{PROB}(c \in T) = r$.

Instances of the preceding schema are, of course, defeasible. For example, it is usually assumed (for good reason) that an instance of the schema is defeated in cases where one is in a position to formulate a direct inference of the following form, where $s \neq r$, and R' is not gerrymandered and narrower than R (i.e., $R' \subset R$) (cf. Venn 1866, Reichenbach 1949, Kyburg 1974, Bacchus 1990, Pollock 1990, Kyburg & Teng 2001):

¹ The final publication is available at Springer: <http://www.springer.com/us/book/9783319537290>

From $c \in R'$ and $\text{freq}(T|R') = s$ infer that $\text{PROB}(c \in T) = s$.

In Section 2, I consider the difficulty of arbitrating between similar pairs of direct inferences, in cases where one is not in a position to make a precise-valued frequency judgment for the narrower reference class, R' . I here maintain that such cases fall into two categories. In the first category, one's frequency 'information' for the narrower reference class is *fully uninformative*, and thereby has no bearing on what conclusion one should adopt concerning the relevant singular proposition ($c \in T$). In these cases, direct inference based on the broader reference class is licensed, provided there are no other defeaters for the inference. On the other hand, if one's frequency information for the narrower reference class is informative (to even the slightest degree), I maintain that direct inference based on the broad reference class is defeated. In Sections 3 and 4, I address the worry that the preceding proposal is overly incredulous, yielding the defeat of too many direct inferences, in the presence of scant frequency information for a narrower reference class.

2 Imprecise-valued Frequency Judgments

Elsewhere (Thorn 2012), I advocated the view that even modest frequency information concerning a (non-gerrymandered) reference class is sufficient to trigger the defeat of a direct inference based on a broader reference class.² In particular, I claimed that direct inference to a conclusion about the probability that $c \in T$ based on frequency information for a reference class R is defeated, if one is warranted in accepting that $\text{freq}(T|R') \in V$, for some $V \subset \{0/|R'|, 1/|R'|, \dots, |R'|/|R'|\}$, where $c \in R'$, R' is non-gerrymandered, and $R' \subset R$.³ The preceding claim conflicts with the most developed competing accounts of direct inference, including those of Kyburg (1974), Bacchus (1990), Pollock (1990), and Kyburg and Teng (2001). Indeed, assuming that direct inference based on gerrymandered reference and target classes are set aside, the preceding accounts all entail the doctrine that direct inference based on a narrower reference class yields the defeat of a direct inference based on a broader class *only if* the conclusions of the two direct inferences are inconsistent. The doctrine maintained by my opponents has a sound motivation (given other auxiliary features of the respective accounts), as it serves to prevent *imprecise-valued* frequency information for a narrow reference class from yielding the defeat of a direct inference based on *precise-valued* frequency information for a broad reference class. For example, consider pairs of direct inferences of the following form:

From $c \in R$ and $\text{freq}(T|R) = 0.5$ infer that $\text{PROB}(c \in T) = 0.5$.

From $c \in R'$ and $\text{freq}(T|R') \in [0.4, 0.6]$ infer that $\text{PROB}(c \in T) \in [0.4, 0.6]$.

² The view was not loudly proclaimed within (Thorn 2012), since the view invites the objection that I will presently consider, to which I had no answer at the time of writing that article.

³ In fact, the view of (Thorn 2012) is committed to an even stronger doctrine about the conditions under which direct inferences are defeated by frequency information for a narrower reference class. For the sake of comprehensibility, details of the stronger doctrine are omitted here, though I believe that the method of inferring expected frequencies introduced in Section 4 is capable of addressing worries concerning the stronger doctrine that parallel the worries discussed in the body of the paper.

Faced with the preceding pair of direct inferences, the doctrine held by my opponents permits the conclusion that $\text{PROB}(c \in T) = 0.5$. The present conclusion is plausible. But as Stone (1987) pointed out, the doctrine (that direct inference based on a narrower reference class yields the defeat of a direct inference based on a broader class *only if* the conclusions of the two direct inferences are inconsistent) permits implausible conclusions in the face of examples of the following form (assuming that the second direct inference incorporates the most precise estimate for $\text{freq}(T|R')$ that is warranted):

From $c \in R$ and $\text{freq}(T|R) = 0.5$ infer that $\text{PROB}(c \in T) = 0.5$.

From $c \in R'$ and $\text{freq}(T|R') \in [0, 0.5]$ infer that $\text{PROB}(c \in T) \in [0, 0.5]$.

The standard accounts of direct inference are overly *credulous* in the face of the preceding example, permitting inference to the conclusion that $\text{PROB}(c \in T) = 0.5$. Unlike the standard accounts, my account yields the defeat of the first direct inference in both of the two preceding examples. It would appear, then, that my account skirts credulity in the second example, while demanding inappropriate incredulity in the first. As it turns out, the *prima facie* incredulousness of my view (in the first example) can be addressed by an auxiliary method for reasoning about the value of the *expected* frequency for a narrower reference class R' , on the basis of frequency information for a broader reference class R . Before illustrating the method, a few words about expected frequencies are in order.

The doctrine that it is statements of expected frequency that are the proper major premises for direct inference may be found in (Bacchus 1990). Within Bacchus's account of direct inference, the doctrine functions to prevent highly uninformative frequency information for a narrower reference class from defeating a direct inference based on an informative frequency statement for a broad reference class. Expected frequencies are apt to perform this function, in virtue of the deductive connections, and lack thereof, between frequencies and expected frequencies. In particular, for all T , R , and r : $\text{PROB}(\text{freq}(T|R) = r) = 1$ implies $E[\text{freq}(T|R)] = r$ (Thorn 2012). The preceding implication explains why it is generally correct to use point-valued frequency statements as major premises for direct inference. On the other hand, in the case where $\text{PROB}(\text{freq}(T|R) \in S) = 1$, we are not generally in a position to infer that $E[\text{freq}(T|R)] \in S$. Rather the most we can deduce, in general, is that $E[\text{freq}(T|R)] \in U$, where U is the smallest interval such that $S \subseteq U$ (Thorn 2012). The lack of an implication, in the latter case, explains why fully uninformative frequency information for a narrow reference class does not defeat direct inferences based on broader reference classes. For example, the judgment that $\text{freq}(T|\{c\}) \in \{0,1\}$ does not result in the defeat of a direct inference for a broader reference class that would yield the conclusion that $\text{PROB}(c \in T) = r \notin \{0,1\}$, since the most one may generally deduce from $\text{PROB}(\text{freq}(T|\{c\}) \in \{0,1\}) = 1$ is that $E[\text{freq}(T|\{c\})] \in [0,1]$. In such cases, the doctrine that statements of expected frequency are the proper major premises of direct inference may be regarded as 'deflating' the inferential role of imprecise frequencies.⁴

In addition to deflating the inferential content of imprecise frequencies, the doctrine that statements of expected frequency are the proper major premises of direct inference provides an avenue to 'inflating' the inferential role of relatively imprecise, though non-vacuous, frequencies, thereby addressing a worry about the incredulousness of the doctrine proposed above (i.e., that

⁴ Further reasons in favor of the doctrine that expected frequencies are the proper major premises of direct inference are presented in (Thorn *forthcoming*).

modest frequency information concerning a narrow reference class is sufficient to trigger the defeat of direct inferences based on broader reference classes). The latter worry is addressed via auxiliary methods of inferring an expected frequency for a narrow reference class R' based on a (relatively) precise-valued frequency judgment for a reference class R ($R' \subset R$). A method for inferring an expected frequency for R' based on frequency information for R was proposed in (Thorn *forthcoming*). In the following section, I provide a simple example that illustrates the kind of conclusions the method permits. In Section 4, I address a major limitation of the method described in Section 3.

3 Imprecise Frequencies Based on Descriptive Statistics

Suppose we are warranted in accepting that $\text{freq}(T|R) = 0.5$ and that $\text{freq}(T|R') \in \{0.4, 0.6\}$ (and we are not warranted in accepting that $\text{freq}(T|R') = 0.4$ or that $\text{freq}(T|R') = 0.6$). To simplify matters, suppose we also know that $|R| = 100$ and $|R'| = 10$. In that case, we can assign a probability to the claim that $\text{freq}(T|R') = 0.4$, by direct inference, and, similarly, to the claim that $\text{freq}(T|R') = 0.6$.

As a basis for assigning a probability to $\text{freq}(T|R') = 0.4$, notice that R' is an element of $\{s : s \subseteq R \wedge |s| = 10 \wedge \text{freq}(T|s) \in \{0.4, 0.6\}\}$. Next notice that we are in a position to deduce the value of the following frequency: $\text{freq}(\{s : \text{freq}(T|s) = 0.4\} \mid \{s : s \subseteq R \wedge |s| = 10 \wedge \text{freq}(T|s) \in \{0.4, 0.6\}\})$. In particular, the value of this frequency is $((\binom{50}{4} \cdot \binom{50}{6}) / ((\binom{50}{4} \cdot \binom{50}{6}) + (\binom{50}{4} \cdot \binom{50}{6}))) = 0.5$. We are thus in a position to formulate a direct inference of the following form:

From $R' \in \{s : s \subseteq R \wedge |s| = 10 \wedge \text{freq}(T|s) \in \{0.4, 0.6\}\}$ and $\text{freq}(\{s : \text{freq}(T|s) = 0.4\} \mid \{s : s \subseteq R \wedge |s| = 10 \wedge \text{freq}(T|s) \in \{0.4, 0.6\}\}) = 0.5$ infer that $\text{PROB}(R' \in \{s : \text{freq}(T|s) = 0.4\}) = 0.5$ (i.e., $\text{PROB}(\text{freq}(T|R') = 0.4) = 0.5$).

A similar direct inference yields the conclusion that $\text{PROB}(\text{freq}(T|R') = 0.6)$ is also 0.5. Taken together the conclusions of the two direct inferences license a conclusion about the value of $E[\text{freq}(T|R')]$. Namely, $E[\text{freq}(T|R')] = \sum_i i \cdot \text{PROB}(\text{freq}(T|R') = i) = 0.4 \cdot \text{PROB}(\text{freq}(T|R') = 0.4) + 0.6 \cdot \text{PROB}(\text{freq}(T|R') = 0.6) = 0.4 \cdot 0.5 + 0.6 \cdot 0.5 = 0.5$.

Recall that (above) I endorsed the doctrine that direct inference to a conclusion about the probability that $c \in T$ based on frequency information for a reference class R is defeated, if one is warranted in accepting that $\text{freq}(T|R') \in V$, for some $V \subset \{0/|R'|, 1/|R'|, \dots, |R'|/|R'|\}$, where $c \in R'$, R' is non-gerrymandered, and $R' \subset R$. The method employed in the preceding example uses frequency information for the relevant R , in order to make a point-valued judgment of the expectation of $\text{freq}(T|R')$. The method thereby goes some distance in addressing the worry that the proposed thesis yields incredulity about the value of $\text{PROB}(c \in T)$ in cases where one's information concerning the possible values of $\text{freq}(T|R')$ is modest. Taken together, the proposed method and the proposed thesis also yield an appropriate degree of incredulity, in the face of the sort of example introduced by Stone (1987): In the example given above, the illustrated method does not entitle one to infer that $\text{PROB}(c \in T) = 0.5$, but only that $\text{PROB}(c \in T)$ is close to 0.5. The exact conclusion one is permitted to draw depends on one's information concerning the size of R and the size of R' . For example, if $|R| = 100$ and $|R'| = 10$, then the described method permits one to infer that $E[\text{freq}(T|R')] \approx 0.4885$ and $\text{PROB}(c \in T) \approx 0.4885$. Similarly, if $|R| = 1,000$ and $|R'| =$

100, then the described method permits one to infer that $E[\text{freq}(T|R')] \approx 0.4652$ and $\text{PROB}(c \in T) \approx 0.4652$.

As explained in (Thorn *forthcoming*), the method employed in the preceding example is also applicable in cases where the values of $|R|$ and/or $|R'|$ are unknown, and where the range of possible values of $\text{freq}(T|R')$ is greater than two, and the range of possible values of $\text{freq}(T|R)$ is greater than one. In presenting the proposed method, I did not consider its application to cases where one's information concerning the possible values of $\text{freq}(T|R')$ was derived by an inductive inference, based on a sample of the elements of R' .⁵ Since cases of the latter sort are common, I now provide a sketch of how I think we should reason about the expectation of $\text{freq}(T|R')$ in such cases.

4 Imprecise Frequencies Based on Sampling

In many cases, we form frequency judgments on the basis of counting, actuarial records, etc. that warrant acceptance of *descriptive* statistical statements of the form $\text{freq}(T|R') \in V$. In other cases, our frequency judgment, for some group, is formed by an inductive inference from an observed sample of members of the group. It is typically proposed that such inductive inferences are underwritten by some form of the Law of Large Numbers, based on the idea that it is reasonable to proceed *as if* the values of the elements of our sample were independent and identically distributed. I favor a less standard view, where induction is underwritten by a combinatorial version of the Law of Large Numbers, along with direct inference, and proceeds without the assumption that the values of the elements of our sample are independent and identically distributed. I will illustrate the ideas that follow according to my preferred view, though similar points could also be expressed within the more standard framework.

According to the view that I prefer, inductive inference proceeds from the combinatorial fact that almost all sufficiently large subsets of a set agree with the set, within a small margin, on the relative frequency of any given characteristic (cf. Williams 1947, Kyburg 1974, Stove 1986, McGrew 2001, Thorn 2014). The following result, reported by McGrew (2001), illustrates the described combinatorial fact:

Theorem. $\forall T, R: \exists \epsilon, \delta, n: n \geq 1/(4\epsilon^2\delta) \Rightarrow$
 $\text{freq}(\{s : \text{freq}(T|s) \approx_\epsilon \text{freq}(T|R)\} | \{s : s \subseteq R \wedge |s| \geq n\}) > 1 - \delta.$ ⁶

By appeal to results such as the preceding, it is possible to underwrite inductive inference via direct inference. Indeed, results such as the preceding, are sufficient to generate the major premises for direct inferences of the following form, where S is our observed sample of the elements of R , and n is sufficiently large, so that ϵ and δ (above) are nearly zero:

From $S \in \{s : s \subseteq R \wedge |s| \geq n\}$ and $\text{freq}(\{s : \text{freq}(T|s) \approx \text{freq}(T|R)\} | \{s : s \subseteq R \wedge |s| \geq n\})$
infer that $\text{PROB}(S \in \{s : \text{freq}(T|s) \approx \text{freq}(T|R)\}) \approx 1$
(i.e., that $\text{PROB}(\text{freq}(T|S) \approx \text{freq}(T|R)) \approx 1$).

⁵ Thanks are due to Christian Wallmann for drawing my attention to this problem (Wallmann *ms*).

⁶ Note that $x \approx_\epsilon y$ if and only if $|x - y| < \epsilon$.

Assuming we have observed that the value of $\text{freq}(T|S)$ is r , we may employ the conclusion of the preceding direct inference to conclude that $\text{PROB}(\text{freq}(T|R) \approx r) \approx 1$, and thus that $E[\text{freq}(T|R')] \approx r$.

As just illustrated, we are sometimes in a position to make an inductive inference about the (approximate) value of $E[\text{freq}(T|R')]$, given frequency information for a sample of the elements of R' . As we saw in the preceding section, it is also possible to reason to a conclusion about the value of $E[\text{freq}(T|R')]$, given the value of $\text{freq}(T|R)$. The problem that I will now address is that of adjudicating the two sorts of inference to the value of $E[\text{freq}(T|R')]$. Intuitively, a conclusion about the value of $E[\text{freq}(T|R')]$ based on a (very) large sample of the elements of R' takes precedence over a competing inference based on the value of $\text{freq}(T|R)$. In such cases, it is, I think, obvious that $\text{PROB}(\text{freq}(T|S) \approx \text{freq}(T|R'))$ should not differ substantially from $\text{PROB}(\text{freq}(T|S) \approx \text{freq}(T|R') \mid \text{freq}(T|S) \approx \text{freq}(T|R))$. After all, although we expect S to agree with both R' and R , regarding T , learning that S disagrees with R should not change our assessment of the probability that it will agree with R' . Rather: evidence that S disagrees with R is evidence that R' is an unrepresentative subset of R , with respect to T .

Beyond such intuitive considerations, it is possible to integrate the two sorts of reasoning concerning the value of $E[\text{freq}(T|R')]$. It is important that the two sorts of reasoning can be integrated, since despite the presumed preference for the inductive inference, in the case where our sample, S , of R' is large, there are also cases where our sample is quite small. In such cases, both the frequency of T among our sample of R' , and the value of $\text{freq}(T|R)$ may be relevant to drawing a conclusion about the value of $E[\text{freq}(T|R')]$. Beyond this, the application of the method of the preceding section yields implausible conclusions about the value of the $E[\text{freq}(T|R')]$ in cases where we have sample-based information concerning the value of $\text{freq}(T|R')$.⁷ For example, suppose we know $|R| = 10,000$, $|R'| = 1,000$, and $\text{freq}(T|R) = 0.8$, and we have drawn a sample of the elements of R' that tells us that it is (virtually) certain that $\text{freq}(T|R') \in [0.45, 0.55]$ (and there is no $S \subset [0.45, 0.55]$, such that it is certain that $\text{freq}(T|R') \in S$). In this case, the method presented in the preceding section yields the conclusion that $E[\text{freq}(T|R')] \approx 0.5497$. This conclusion is implausible, as it takes too little account of our sample-based evidence bearing on the value of $\text{freq}(T|R')$. The example shows that the method presented in the preceding section is quite limited in its proper domain of application.⁸ Beyond this, it is clear that we need an alternative to this method, that is applicable in cases where our judgments about the possible values of $\text{freq}(T|R')$ are based on a sample of the elements of R' , if we are going to address worries about the incredulity of the doctrine advocated in Section 2 (i.e., the doctrine that minimal frequency information regarding R' is sufficient to defeat a direct inference based on R).⁹

In fact, the sort of reasoning described in the preceding section can be integrated with sample-based inductive inference. The trick to seeing how the two sorts of reasoning may be

⁷ Once again, thanks are due to Christian Wallmann for formulating this problem, and presenting me with illustrative examples (Wallmann *ms*).

⁸ I maintain that the conclusion that $E[\text{freq}(T|R')] \approx 0.5497$ would not be implausible, if the information that $\text{freq}(T|R') \in [0.45, 0.55]$ was a descriptive statistic. In that case, our information that $\text{freq}(T|R) = 0.8$ would indicate that $\text{freq}(T|R')$ is very probably 0.55, given the relative proportions of the subsets of R whose frequency of T lies in $[0.45, 0.55]$. Indeed, the vast majority of the subsets of R , whose frequency of elements of T is in $[0.45, 0.55]$, are subsets whose frequency of T is 0.55.

⁹ It may be that variants of the method described here could be used to integrate reasoning of the sort described in the preceding section with other sorts of reasoning (i.e., other than sample-based reasoning) that are capable of providing reasons for assigning probabilities to $\text{freq}(T|R')$ taking various values.

integrated is to notice that inductive inference, based on the likes of Theorem 3 (or on some short run variant of the Law of Large Numbers), generally licenses conclusions about the probability that the value of $\text{freq}(T|R')$ lies within an interval that spans $\text{freq}(T|S)$, for our sample S . For example, given a sample of sufficient size, we may apply Theorem 3 to infer that the probability is at least 0.95 that $\text{freq}(T|R')$ lies within $\text{freq}(T|S) \pm 0.05$. In fact, we will generally be in a position to infer various probabilities regarding the possible values of $\text{freq}(T|R')$ – that the probability is at least 0.98 that $\text{freq}(T|R')$ lies within $\text{freq}(T|S) \pm 0.1$, for example.

Rather than describing a general schema for integrating the two sorts of reasoning, I here present an example that illustrates the proposed method of integrating the two sorts of reasoning. The example consists of an elaboration of the example considered in the preceding section. In the original variant of the example, we were warranted in accepting that $\text{freq}(T|R) = 0.5$ and that $\text{freq}(T|R') \in \{0.4, 0.6\}$, along with the fact that $|R| = 100$ and $|R'| = 10$. For the variant, suppose that instead of knowing that $\text{freq}(T|R') \in \{0.4, 0.6\}$, we observed a two element sample from R' , and found that neither element of the sample is an element of T . The small sample size, in this case, will simplify the needed calculations, while providing an apt illustration of the inferences that we ought to make in the described circumstances. After walking through this ‘toy’ example in detail, I will briefly present some additional examples that show how the method performs in the case where R , R' , and our sample are much larger.

The first step in dealing with the toy example is to calculate the frequencies with which two element subsets of a ten element population (in this case R'), are guaranteed to agree with that population, to various degrees, on the frequency of T . This results in the following conclusions (that are the strongest ones that can be drawn, in the present case): (i) at least $4\bar{6}\%$ (of the two element subsets of a ten element set) differ (from the set with respect to the frequency of T) by no more than 0.2, (ii) at least $5\bar{3}\%$ differ by no more than 0.3, (iii) at least $5\bar{5}\%$ differ by no more than 0.4, (iv) at least $8\bar{6}\%$ differ by no more than 0.5, (v) at least $9\bar{7}\%$ differ by no more than 0.6, and (vi) none differ by more than 0.8. It is also possible that none of the two element subsets of a ten element set differ from the set (at all) on the frequency of T (as is the case when no elements of R' are in T). Now given the preceding, and given that $\text{freq}(T|S) = 0$, for our sample S , it is reasonable to draw the following conclusions (via direct inference): (i') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2\}) \in [0.4\bar{6}, 1]$, (ii') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2, 0.3\}) \in [0.5\bar{3}, 1]$, (iii') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2, 0.3, 0.4\}) \in [0.5\bar{5}, 1]$, (iv') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}) \in [0.8\bar{6}, 1]$, (v') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}) \in [0.9\bar{7}, 1]$, and (vi') $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}) = 1$.

I now propose that we treat the preceding conclusions ((i') through (vi')) as expressing higher order probabilities about the set of values in which $\text{freq}(T|R')$ lies. I describe these (imprecise) probabilities as “higher order probabilities”, since (as we will see in a moment) I will also consider assignments of first order probability to the possible values of $\text{freq}(T|R')$, according to varied assumptions about the set in which $\text{freq}(T|R')$ lies. A difficulty with applying the probabilities expressed by (i') through (vi') is that they are imprecise. However, since the final goal is to form a judgment about the *possible* values of $E[\text{freq}(T|R')]$, we can use the imprecise probabilities in order to reason by cases. In particular, we treat (the set of) upper bounds specified by (i') through (vi') as point-valued probabilities, in order to infer a lower bound on $E[\text{freq}(T|R')]$. Similarly, we treat the lower bounds specified by (i') through (vi') as point-valued probabilities, in order to infer an upper bound on $E[\text{freq}(T|R')]$. Taken as point-valued probabilities, the set of upper bounds specified by (i') through (vi') entail that $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2\}) = 1$. Taken similarly, the lower bounds entail that: (i*) $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2\}) = 0.4\bar{6}$, (ii*)

$\text{PROB}(\text{freq}(T|R') = 0.3) = 0.0\bar{6}$, (iii*) $\text{PROB}(\text{freq}(T|R') = 0.4) = 0.0\bar{2}$, (iv*) $\text{PROB}(\text{freq}(T|R') = 0.5) = 0.3\bar{1}$, (v*) $\text{PROB}(\text{freq}(T|R') = 0.6) = 0.\bar{1}$, and (vi*) $\text{PROB}(\text{freq}(T|R') \in \{0.7, 0.8\}) = 0.0\bar{2}$. Our reasoning according to the two cases proceeds as follows.

For the first case, we assume that $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2\}) = 1$. In this case, it is reasonable to infer the value of $E[\text{freq}(T|R')]$ by application of the method of the preceding section, which requires making three direct inferences in order to draw three conclusions about the probability that $\text{freq}(T|R')$ is 0, 0.1, and 0.2, respectively. Given the conclusions of the aforementioned direct inferences (whose description is omitted here), it follows that the value of $E[\text{freq}(T|R')]$ (according to the first case) is (approximately) 0.1816. For the second case, we assume that (i*) $\text{PROB}(\text{freq}(T|R') \in \{0, 0.1, 0.2\}) = 0.4\bar{6}$, (ii*) $\text{PROB}(\text{freq}(T|R') = 0.3) = 0.0\bar{6}$, etc. For each of (i*) through (vi*), we compute the value of $E[\text{freq}(T|R')]$, by application of the method of the preceding section, on the assumption that the object of the respective probability statement obtains (i.e., on the assumption that $\text{freq}(T|R') \in \{0, 0.1, 0.2\}$, and then on the assumption that $\text{freq}(T|R') = 0.3$, etc.). We then form a weighted average of the resulting values of $E[\text{freq}(T|R')]$, according to the (higher order) probabilities associated with (i*) through (vi*). So, for (i*), we have $E[\text{freq}(T|R')] \approx 0.1816$, with weight $0.4\bar{6}$, and for (ii*), we have $E[\text{freq}(T|R')] = 0.3$, with weight $0.0\bar{6}$, etc. Averaging the respective values of $E[\text{freq}(T|R')]$ according to the described weights yields the result that $E[\text{freq}(T|R')]$ is (approximately) 0.3520 (according to the second case). The pair of conclusions, $E[\text{freq}(T|R')] \approx 0.1816$ and $E[\text{freq}(T|R')] \approx 0.3520$, correspond to the upper and lower bounds specified by (i') through (vi'), so it is correct to use these conclusions as bounds on $E[\text{freq}(T|R')]$, namely: $E[\text{freq}(T|R')] \in [0.1816, 0.3520]$.

The preceding illustrates my proposed approach to integrating the two sorts of reasoning about the value of $E[\text{freq}(T|R')]$. It is important to note that the integrated method ratifies the intuition that a conclusion about the value of $E[\text{freq}(T|R')]$ based on a (very) large sample of the elements of R' takes precedence over a competing inference based on the value of $\text{freq}(T|R)$. Indeed, as the size of our sample of R' increases, the size of the smallest set V such that we are warranted in inferring that $\text{PROB}(\text{freq}(T|R') \in V) \geq r$, for some $r > 0$, will shrink. At the same time, the warranted values of r for $\text{PROB}(\text{freq}(T|R') \in V) \geq r$, for various fixed V , will increase. As a result, the impact of our judgment of the value of $\text{freq}(T|R)$ upon our conclusion about the possible values of $E[\text{freq}(T|R')]$, as licensed by the proposed method will decrease as a function of the size of our sample of R' . For example, suppose we know that 10,000,000 Bavarians voted in the last German federal election, with about 20% casting their vote for the Social Democratic Party (SPD). In addition, suppose we know that 100,000 of these voters were from Nuremberg. Finally, suppose we drew a sample of 1,000 of the 100,000 Nurembergers, and found that 40% of these voters cast their vote for the SPD. In this case, we can apply the proposed method to conclude that the expected frequency of voters from Nuremberg that cast their vote for the SPD is in the interval $[0.3864, 0.3991]$, which closely approximates our sample frequency (which is 0.4). Note that the narrowness of the inferred interval licensed by the method is primarily a function of the size of the sample drawn from the relevant R' , rather than the relative size of the sample in relation to the size of R' . For example, if we had drawn a 1,000 element sample of the set of 1,000,000 voters from Munich (and found that 40% of these voters cast their vote for the SPD), then the method would have licensed the conclusion that the expected frequency of voters from Munich that cast their vote for the SPD is in $[0.3864, 0.3991]$.¹⁰ It should also be observed that the relative size of our sample need not be enormous, in order to exert a significant influence on

¹⁰ Though not identical, the bounds licensed in the two cases differ by less than 0.0001.

the conclusions licensed by the proposed method. For example, in a variant of the above example where we drew a sample of only 100 of the 100,000 Nurembergers, the method would license the conclusion that the expected frequency of voters from Nuremberg that cast their vote for the SPD is in the interval $[0.3569, 0.3941]$.

It should be acknowledged that the computations required in applying the proposed method are not of the sort that could be performed on the back of an envelope, save in cases where all of the relevant sets are quite small (as in the toy example considered above). However, if we are willing to accept modest approximation, the method can be applied in cases where the size of R , R' , and our sample are relatively large (as in the examples of the preceding paragraph), using a typical modern personal computer, with a reasonable run time (i.e., more than a minute, but less than a day).¹¹

5 Conclusion

In the present paper, I articulated an objection to the view that modest frequency information concerning a (non-gerrymandered) reference class, R' , is sufficient to trigger the defeat of a direct inference based on a broader reference class, R . In particular, the view appears to imply an overly incredulous account of direct inference. As a means of addressing this objection, I appealed to a method of inferring the expectation of $\text{freq}(T|R')$ by appeal to the value of $\text{freq}(T|R)$. I then raised an objection to that method, noting that it cannot be used to draw plausible conclusions in cases where our frequency information about R' is based on an inductive inference from a sample of the elements of R' . In order to address this problem, I introduced a new method for inferring the expectation of $\text{freq}(T|R')$. This method integrates the two sorts of inference concerning the value of $\text{freq}(T|R')$, that is, reasoning based on the value of $\text{freq}(T|R)$, and reasoning based on the frequency of T among a sample drawn from R' . While the new method is sensitive to both sorts of information (i.e., about the superset and the sample), information based on a large sample correctly trumps information based on a superset.

Acknowledgments Work on this paper was supported by DFG Grant SCHU1566/9-1 as part of the priority program “New Frameworks of Rationality” (SPP 1516). For comments that motivated the preparation of this paper, I am thankful to participants at EPSA 2015, including Michael Baumgartner, Martin Bentzen, Seamus Bradley, Bert Leuridan, Jan-Willem Romeijn, Gerhard Schurz, and Jon Williamson. I am also grateful for discussions with Christian Wallmann, which motivated the proposal presented in Section 4.

¹¹ The bottleneck in applying the method to large sets derives from the requirement of computing large binomial coefficients. Computing such coefficients, $\binom{n}{k}$, for large values of n and k is problematic. For example, $\binom{10,000,000}{1,000,000}$ is greater than $10^{1,000,000}$. Nevertheless, large binomial coefficients can be computed in linear time, $\mathcal{O}(k)$, assuming the cost of multiplication is not dependent on the size of the factors multiplied. Although the latter assumption is clearly false, it is possible to execute accurate calculations of large binomial coefficients (i.e., accurate to some reasonable number of significant digits), where the cost of multiplication increases very slowly, as a function of k . As a further point of reference, note that, at present, the fastest computers in the world are tens of millions of times faster than a typical personal computer. It should also be observed that the need for repeated calculations of the same binomial coefficient could be eliminated by the use of a lookup table, which is feasible assuming we store approximate (though highly accurate) values of the relevant binomial coefficients.

References

- Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge*. Cambridge, Massachusetts: MIT Press.
- Kyburg, H. (1974). *The Logical Foundations of Statistical Inference*. Dordrecht: Reidel Publishing Company.
- Kyburg, H., & Teng, C. (2001). *Uncertain Inference*. Cambridge: Cambridge University Press.
- McGrew, T. (2001). Direct Inference and the Problem of Induction. *The Monist* 84: 153-174.
- Pollock, J. (1990). *Nomic Probability and the Foundations of Induction*. Oxford: Oxford University Press.
- Reichenbach, H. (1949). *A Theory of Probability*. Berkeley: Berkeley University Press.
- Stone, M. (1987). Kyburg, Levi, and Petersen. *Philosophy of Science* 54 (2): 244-255.
- Stove, D. (1986). *The Rationality of Induction*. Oxford: Clarendon Press.
- Thorn, P. (2012). Two Problems of Direct Inference. *Erkenntnis* 76 (3): 299-318.
- Thorn, P. (2014). Defeasible Conditionalization. *Journal of Philosophical Logic* 43 (2-3): 283-302.
- Thorn, P. (*forthcoming*). On the Preference for More Specific Reference Classes. *Synthese*.
- Venn, J. (1866). *The Logic of Chance*. New York: Chelsea Publishing Company.
- Wallmann, C. (ms). *A Bayesian solution to the conflict of narrowness and precision in direct inference*.
- Williams, D. (1947). *The Ground of Induction*. Cambridge, Massachusetts: Harvard University Press.