International Studies Quarterly

International Studies Quarterly (2014) **58**, 873–875

Information and Uncertainty: Inference in Qualitative Case Studies

Bear F. Braumoeller
The Ohio State University

Drozdova and Gaubatz (2014) represent a welcome addition to the growing literature on quantitative methods designed to complement qualitative case studies. Partly due to its crossover nature, however, the article balances delicately—and ultimately untenably—between within-sample and out-of-sample inference. Moreover, isomorphisms with existing techniques, while validating the methodology, simultaneously raise questions regarding its comparative advantage.

Drozdova and Gaubatz (2014) present an ambitious foray into the realm of information theory, with the goal of formalizing and giving structure to the findings of qualitative case studies. The insight required to recognize a useful application for information analytics and the effort required to realize it are quite laudable. At the same time, the authors conceive of uncertainty in a somewhat unorthodox manner, one that raises the question of the generalizability of the results. Moreover, isomorphisms between these procedures and others already in widespread use, while buttressing the validity of the method, simultaneously raise the question of its newness. While these issues did not warrant rejection of the article, the Editors invited me to submit a revised version of my review as both clarification and commentary. I present my views below.

Uncertain About Uncertainty

Drozdova and Gaubatz (2014) seek to provide a concrete metric of the extent to which the presence of an independent variable, X, alters our uncertainty about the occurrence of a dependent variable, Y. What they mean by "uncertainty" is hinted at in a handful of passages—"a measure of how much knowing about the presence or absence of a given factor reduces uncertainty about the presence or absence of a given outcome" (636–7), "uncertainty about the relative importance of case factors" (637), and "power for discriminating between success and failure" (641).

Simply put, uncertainty is a function of unpredictability, *assuming that we know the probability of success with certainty*. If we know with certainty that $\Pr(Y = 1) = 0.9$, we will be fairly certain about the outcome of Y if it were to happen tomorrow: We could predict a value of 1 and be right 90% of the time. If $\Pr(Y = 1) = 0.5$, on the other hand, we might as well flip a coin: We really do not know anything. If $\Pr(Y = 1) = 1$, Y is perfectly predictable and we are perfectly certain about whether it will occur. The authors call this "outcome uncertainty."

That understanding, however, highlights the importance of my first observation about this paper: *It does not present a statistical technique*. The authors' measure of outcome uncertainty approaches 0 as $\Pr(Y = 1)$ approaches 0 or 1, regardless of the sample size. (Oddly, it never gets there, because taking the log of $\Pr(Y = 0)$ and $\Pr(Y = 1)$ ensures that outcome uncertainty is undefined at those values—a nontrivial problem, given the prevalence of zeros in at least one cell of many small-n 2×2 tables.) In other words, the authors' measure of uncertainty based on $\Pr(Y = 1)$ does not allow for uncertainty *about* $\Pr(Y = 1)$.

This feature of the technique creates a problem: There could be a *lot* of uncertainty regarding the value of $\Pr(Y = 1)$ itself, especially in the small samples for which the method is designed. To take the Krepon and Caldwell example from the article, the baseline probability of ratification ($\Pr(Y = 1)$ in the sample) is 0.57, but the 95% confidence intervals around that estimate range from 0.20 to 0.88. A decision maker looking at this sample, using the authors' techniques, would conclude that ratification is basically a coin flip, when she could easily be looking at the output of a data-generating process that produces ratification with great certainty. 0.57 still provides our best guess, of course, but that guess itself is highly uncertain. By ignoring sampling uncertainty, the authors make their estimates of outcome uncertainty considerably more precise than reality warrants.

Moreover, this understanding of uncertainty undermines the authors' claim that looking at the values of Y conditional on another variable, X, should increase our certainty, even if only by a little. It is possible to find, or invent, situations in which our uncertainty about Y given $X = 1$ equals our unconditional uncertainty about Y—for example, $\Pr(Y = 1) = 0.8$ and $\Pr(Y = 1 \mid X = 1) = 0.2$. Because the uncertainty function is symmetrical around 0.5, our uncertainty about Y is precisely the same if $X = 1$ as it is if we know nothing at all about X. The authors' measure of uncertainty reduction, conditional entropy, *does* give us more information about Y if we have knowledge about X—but only because the measure is based on the reduction of uncertainty associated with *both* possible values of X. That is: In the above example, though we do not increase our certainty about Y if we know that $X = 1$, we *would* be more certain about Y if $X = 0$. The authors' uncertainty measure takes both possibilities into account, weighted by their probability of occurrence.

That fact, though, raises three objections. First, this method aims to serve as a guide to policy, and the next case coming down the block might not have the value of

Bear F. Braumoeller is an Associate Professor in the Department of Political Science at the Ohio State University. His most recent book, *The Great Powers and the International System* (Cambridge University Press), won both the ISA Best Book Award and the J. David Singer Book Award for 2014.

Braumoeller, Bear F. (2014) Information and Uncertainty: Inference in Qualitative Case Studies. *International Studies Quarterly*, doi: 10.1111/isqu.12169

X that a decision maker needs to learn something about $\Pr(Y=1)$. That is, knowing X would not necessarily reduce outcome uncertainty in any given real-world case: It could make it better or worse. Second, there is no guarantee that a decision maker will be interested in uncertainty: In the cases described in the paper—coercive diplomacy, treaty ratification, peacekeeping success—most decision makers will want to increase the probability of success. If they can raise the probability of success from 0.1 to 0.5, they will take that outcome in a heartbeat, despite the fact that 0.5 represents higher outcome uncertainty. Finally, with samples this small, sampling uncertainty will almost certainly swamp any differences in outcome uncertainty. A simple χ^2 test would tell us whether we can reliably distinguish $\Pr(Y=1 \mid X=1)$ from $\Pr(Y=1 \mid X=0)$ in the various applications for which the method is intended. I suspect the answer would be "no" most of the time.

Out-of-Sample Inference

The above may seem like nitpicking, even by academic standards, except for this: It remains unclear, ultimately, whether the cases being examined are the universe of possible/interesting cases or simply a sample. The authors seem to imply both. In their reply to my initial review, they clarified by stating that "The entropy approach does not propose inferences beyond the analyzed data." However, if the authors forswear any out-of-sample predictions or inferences, they undercut the entire rationale for the enterprise—"our approach is…explicitly designed for generating and accumulating policy-relevant knowledge across multiple cases" (Drozdova and Gaubatz, 2014:633). Policy, by its nature, deals with out-of-sample (present and future) cases. If, on the other hand, the goal of measuring changes in outcome uncertainty *is* to aid with out-of-sample inferences, sampling uncertainty cannot be ignored—and with a small number of cases, it will be substantial.

In consequence, the article is irrevocably stuck on the horns of the out-of-sample dilemma: If the findings are not meant to generalize, that fact undercuts the rationale, and if they are, we are stuck with sampling uncertainty.

The Virtue of Simplicity

As a final (and minor) aside, I found the mathematical complexity of a technique designed to provide simple, straightforward metrics perplexing. The authors' aside that

> Shannon's central insight was that the probability of joint events could be translated from a multiplicative to a more intuitive additive measure through the use of logarithms(638)

begs the question, since when are logarithms more intuitive than multiplication? The calculation of conditional probability—

> Conditional probability of outcome Y given the value of factor x_i is calculated by dividing the joint probability that both values co-occur, written $p(x,y)$, by the probability that this factor occurs at all, $p(x_i)$, for all possible combinations(638)

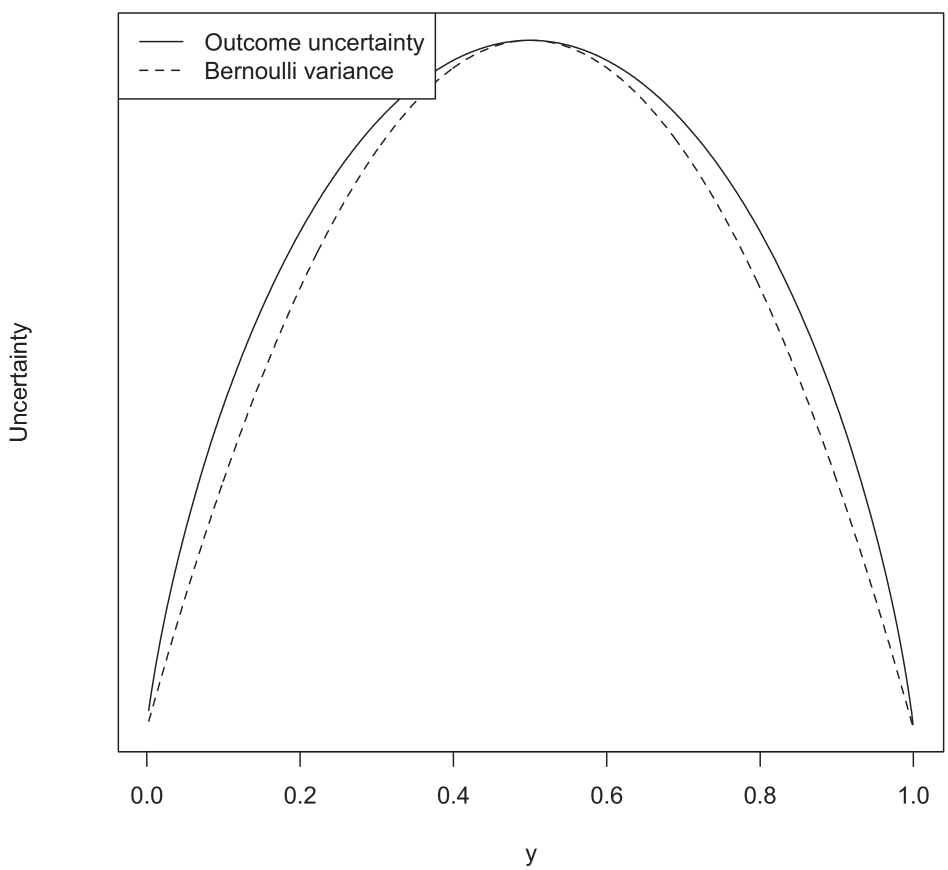

Fig 1. Outcome Uncertainty ($H(Y)$) vs. Bernoulli Variance

seems opaque, until one realizes that the joint probability that both values co-occur could also be written as $\Pr(X|Y)\Pr(Y)$, which gives

$$\Pr(Y|X) = \frac{\Pr(X|Y)\ \Pr(Y)}{\Pr(X)}$$

or, as it is more commonly known, Bayes' theorem.

Finally, the authors' measure of outcome uncertainty performs much the same function as the variance of the Bernoulli distribution, σ^2_y, which is just $\Pr(Y)(1\text{-}\Pr(Y))$. The variance, a measure of how much a variable varies, captures outcome uncertainty directly: The more the variable varies, the more uncertain we are about the outcome of a new trial. The outcome uncertainty metric $H(Y)$ (p. 638) does not map directly to the variance of the Bernoulli distribution, but as Figure 1 demonstrates, it comes awfully close.

It is not clear to me that $H(Y)$ provides a better measure of outcome uncertainty than σ^2_y, though it is not clear that it does not. The authors provide the equations in the Appendix but offer little by way of conceptual or logical explanation. There might be some reason that Shannon's metric captures outcome uncertainty more accurately than the variance of a Bernoulli distribution, but I have no idea why it would. The measure of conditional entropy could be similarly simplified.

This would all merit nothing more than a brief footnote if it were not for the fact that part of the article's value-added resides in the simplicity of the uncertainty reduction metric. That virtue hardly seems compelling if most of the math is isomorphic to, and can readily be approximated by, (much) simpler equations that are better-known to political scientists.

Reference

Drozdova, Katya, and Kurt Taylor Gaubatz. (2014) Reducing Uncertainty: Information Analysis for Comparative Case Studies. *International Studies Quarterly* 58: 633–645.