Myles Lavan

Epistemic Uncertainty, Subjective Probability, and Ancient History

Periods from which few data survive pose a major challenge for history in the quantitative mode. Many important historical quantities can be estimated only on the basis of sparse and disparate information. For example, we do not have any census data for the population of the Roman Empire as a whole. Yet we are hardly in a position of complete ignorance. Scattered information, both quantitative and qualitative, allows us to reason in terms of likelihood. The methodological question is how to report our inevitably uncertain and subjective conclusions. Ancient historians tend to frame their debate in terms of point estimates, disputing whether 54 million or 45 million is a better estimate of the population of the Roman Empire in 14 C.E. Estimates of this type are hard to interpret since they convey no information about the margin of error, which is often large and sometimes asymmetrical. Historians may resort to ranges as a concession to uncertainty, but their ranges tend to be arbitrary rather than grounded in any measure of confidence or credibility. They signal the existence of uncertainty without offering any real guidance regarding its magnitude.

The problem of uncertainty becomes particularly acute when historians combine estimates for multiple quantities. Highly uncertain quantities are often estimated on the basis of other quantities about which we have better (if still limited) knowledge. For example, Roman GDP has been modeled as a function of total population and average per capita consumption. Since these numbers

are themselves uncertain, the proliferating uncertainties pose a major challenge to the credibility of any point estimate.[1]

This research note discusses an alternative framework that allows for a more rigorous accounting of uncertainty. It formalizes the probabilism that is already implicit in most historical reasoning by using probability to measure degree of belief. A brief survey of work on the population of the Roman Empire illustrates the methodological problem, paving the way for a discussion of the "subjective" interpretation of probability as degree of belief and its use as a tool of historical analysis, particularly when combining uncertainties. Conceptualizing uncertainty as probability is a useful discipline in itself, but its greatest value lies in the scope for aggregation.[2]

The probabilistic approach discussed herein is familiar to scholars and practitioners in future-oriented fields because it informs much current work in forecasting, risk assessment, and decision analysis. Its unfamiliarity to many historians, even those engaged in quantitative analysis, is probably due to a mistaken belief that uncertainty about the past is qualitatively different from that faced by other disciplines. Ancient historians are not, however, alone in the need to base quantification on subjective assessments of what is likely rather than on hard data; the problem of reliance on subjective assessment is shared by many other fields. Consider an observation about risk analysis from a textbook: "Probabilistic risk analysis treats events with a low intrinsic rate of occurrence, and large amounts of data are seldom available. Since its inception, *expert opinion in the form of subjective probabilities* has been a dominant source of data for failure probabilities" (my emphasis). Ancient historians can learn from the techniques that other disciplines have developed to manage epistemic uncertainty, that is, uncertainty that arises from the limits of our knowledge.[3]

1   For the state of the art in the estimation of Roman GDP, see Walter Scheidel and Steven J. Friesen, "The Size of the Economy and the Distribution of Income in the Roman Empire," *Journal of Roman Studies*, XCIX (2009), 61–91.

2   For an earlier article that applied this approach to a long-standing problem in ancient history, see Lavan, "The Spread of Roman Citizenship, 14–212 C.E.: Quantification in the Face of High Uncertainty," *Past & Present*, 230 (2016), 3–46. Daniel Jew is applying it to the problem of Athenian population in *The Probable Past: Agriculture and Carrying Capacity in Ancient Greece* (New York, forthcoming). This research note expands on the theoretical premises, particularly the underpinning conceptions of epistemic uncertainty and subjective probability.

3   For the quotation, see Tim Bedford and Roger Cooke, *Probabilistic Risk Analysis: Foundations and Methods* (New York, 2001), 191. Historians in a few sub-fields have already adopted

THE POPULATION OF THE ROMAN EMPIRE   Population is relevant to a wide range of questions in the social and economic history of the Roman Empire. Unfortunately, the population data that the Roman state collected through regular censuses are almost entirely lost to us. The most significant exception is a series of census figures for Roman citizens that extends to the year 48 C.E. Since citizens were still concentrated in Italy at that point, the citizen population should be a reasonable proxy for the population of Italy. Yet a major ambiguity about whether these numbers comprise all persons or just adult males leaves even the question of Italy's population hotly contested. Population data that are even more problematical exist for a few other sub-regions—such as northwest Iberia and Egypt—but Roman historians are otherwise dependent on estimates of carrying capacity and on crude judgments about regional variation in population density and about the trajectory of population levels relative to the late medieval and early modern periods.[4]

In 1886, Beloch, a pioneer in the modern study of the Roman population, reckoned a total of 54 million inhabitants in the Empire in 14 C.E. He subsequently revised this number to 70 million (assuming higher populations for Gaul, the Balkans, and North Africa).

---

a probabilistic approach by combining uncertainties through Monte Carlo simulation (discussed further below). For a pioneering application, see Donald Schaefer and Thomas Weiss, "The Use of Simulation Techniques in Historical Analysis: Railroads versus Canals," *Journal of Economic History,* XXXI (1971), 854–884, a precedent followed by several subsequent articles in modern economic history and historical demography. For its use in the CAMSIM microsimulation of kin sets, see James E. Smith and Jim Oeppen, "Estimating Numbers of Kin in Historical England Using Demographic Microsimulation," in David S. Reher and Roger S. Schofield (eds.), *Old and New Methods in Historical Demography* (New York, 1993), 413–425. Archaeologists have increasingly used the method to manage uncertainty about chronologies (see n. 10), and ancient historians have used it in an ad hoc way to examine miscellaneous problems in ancient history. See, for example, Ellen Janssen et al., "Fuel for Debating Ancient Economies: Calculating Wood Consumption at Urban Scale in Roman Imperial Times," *Journal of Archaeological Science: Reports,* XI (2017), 592–599, which uses Monte Carlo simulation to estimate fuel consumed by pottery production and baths in Roman Sagalassos but reverts to traditional interval analysis to estimate total fuel consumption and its impact on local woodland—precisely the type of problem in which epistemic uncertainties could fruitfully be construed as subjective probabilities. None of these contributions ground the method in the Bayesian conception of uncertainty and probability. Instead, the uncertainty tends to be interpreted as aleatory (that is, related to variability or random processes).

4   For the population of Italy, see Scheidel, "Roman Population Size: The Logic of the Debate," in Luuk De Ligt and Simon J. Northwood (eds.), *People, Land, and Politics: Demographic Developments and the Transformation of Roman Italy 300 BC–AD 14* (Leiden, 2008), 17–70.

He also suggested, offhandedly, that the population grew to a peak of around 100 million by the end of the second century. The next important intervention considered even his first estimate too high. In 1978, McEvedy and Jones constructed an extraordinarily ambitious model of the evolution of the global population on a country-by-country basis from 400 B.C.E. to the present. They significantly reduced Beloch's 1886 estimates for several sub-regions, particularly Anatolia and the Levant, to arrive at population levels that seemed plausible in the light of the long-term history of those sub-regions as they understood it. They put the peak population of the Empire c. 200 C.E. at just 46 million—10 percent lower than Beloch's estimate for 14 C.E. and less than half his later estimate for peak population. Their work has provided the starting point for most subsequent research in the field of Roman history.[5]

Reverting closer to Beloch's figures for Anatolia and the Levant and positing slightly higher long-term growth over the first two centuries C.E., Frier proposed a peak population of 61 million in 164 C.E. (a more credible date for the peak, given the "Antonine plague" in the 160s). More recently, Scheidel suggested a peak of 59 to 72 million in 165 C.E., assuming slightly higher populations in the northwestern provinces and allowing for some uncertainty.[6]

We need not delve more deeply into the evidence to observe that the debate has been conducted in a way that obscures the question of uncertainty. Most interventions have taken the form of point estimates without any serious discussion about the margin of error. Scheidel's range at least signals the problem of uncertainty, but it is far from clear how it should be interpreted. Does he mean to rule out the possibility of a population less than 59 million or higher than 72 million? The evidence cannot categorically disprove either McEvedy and Jones' minimal estimate of 46 million or Beloch's maximal one of 100 million, though both now seem much less likely than a figure in the 60s. The proposition of this

5   Julius Beloch, *Die Bevölkerung der griechisch-römischen Welt* (Leipzig, 1886), 507; *idem*, "Die Bevölkerung im Altertum," *Zeitschrift für Sozialwissenschaft,* II (1899), 618, 620; Colin McEvedy and Richard M. Jones, *Atlas of World Population History* (London, 1978), 22.
6   Bruce W. Frier, "Demography," in Alan K. Bowman, Peter Garnsey, and Dominic Rathbone (eds.), *The Cambridge Ancient History. XI. The High Empire, AD 70–192* (New York, 2000), 812–814; Scheidel, "Demography," in *idem*, Ian Morris, and Richard P. Saller (eds.), *The Cambridge Economic History of the Greco-Roman World* (New York, 2007), 45–49.

research note is that formal probabilities offer a better way of representing and managing the uncertainty.

KNOWLEDGE, UNCERTAINTY, AND PROBABILITY   This proposition may seem troubling at first because it contravenes our intuitions about the nature of uncertainty and probability. Uncertainty appears to take two fundamentally different forms, which can be illustrated by, say, predicting the outcome of a coin toss and estimating the distance between Cambridge and St. Andrews. The uncertainty in the first case is (or rather appears to be) the result of a random process and cannot be resolved until the coin is flipped. The uncertainty in the latter case is merely the result of the limits of an individual's knowledge and could be resolved through measurement. The first type of uncertainty is often termed *aleatory* (or objective) and the second type *epistemic* (or subjective). Intuitively, probability seems a natural way of representing aleatory uncertainty (the chance of heads is 50 percent), but it may seem an abuse to apply it to epistemic uncertainty (distance appears to leave no room for probability).

The association of probability with objective randomness is, however, far less secure than it appears. The meaning of probability is a profound and unresolved philosophical question. The two most important positions are the *frequentist* and the *subjective* interpretations. Anyone with some understanding of statistics will recognize the frequentist view, which long dominated introductory textbooks. Frequentists see probability as an attribute of repeated events. The probability of an event is the frequency of its occurrence in a long sequence of similar trials. Hence, the probability of heads in a coin toss is 50 percent because the frequency of heads would approach 50 percent in a suitably long series of tosses. On the frequentist view, it would be nonsensical to speak of the probability that some historical quantity had some value, because it either had it or did not.[7]

The subjective or "Bayesian" interpretation holds that probability represents an observer's degree of belief, given the available information. As such, it is a function not just of the world but also

---

7   For a brief overview of interpretations of uncertainty, see M. Granger Morgan and Max Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (New York, 1990), 48–50.

of a particular state of knowledge. Since knowledge varies from observer to observer, probability is always subjective or personal ("my probability," not "the probability")—hence De Finetti's famous dictum "probability does not exist," meaning that there are no objective probabilities. This seemingly radical view rests on the insight that all uncertainty has an irreducible element of subjectivity. The apparently obvious distinction between aleatory and epistemic uncertainty dissolves under closer inspection. Phenomena that appear random are often the result of processes that are in fact deterministic, though chaotic, in the technical sense that the outcome is highly sensitive to small changes in the initial conditions. Uncertainty about the outcome of a coin toss, for example, is actually epistemic uncertainty about the initial conditions and how they determine the behavior of the coin. Most randomness is thus a result of an observer's lack of knowledge, not inherent in the world itself. Furthermore, an assessment that the probability of heads on any coin toss is 50 percent depends on an unstated, and possibly erroneous, assumption that the coin in question is unbiased. The fundamental insight of the subjectivists is that probabilities conventionally thought to be objective (a property of the world) are always based on assumptions about the generating mechanism (like a coin toss) and hence subjective. Their growing influence is evident in the spread of Bayesian methods throughout a wide variety of fields.[8]

Scholars in fields concerned with forecasting, risk analysis, and decision analysis recognize that predicting the future always involves epistemic as well as aleatory uncertainty. Many of them have embraced the use of probability as a measure of epistemic uncertainty. Various alternative mathematical frameworks have been suggested (including intervals, "imprecise probabilities," "possibilities," and "belief functions") to address doubts about the applicability of probability to all forms of uncertainty. Nevertheless, probability remains the dominant conceptual tool for representing epistemic uncertainty

---

8   Bruno De Finetti, *Theory of Probability: A Critical Introductory Treatment* (New York, 1974). For an introduction to the subjective interpretation, see David Spiegelhalter, "Quantifying Uncertainty," in Layla Skinns, Michael Scott, and Tony Cox (eds.), *Risk (Darwin College Lectures)* (New York, 2011), 17–33. Dennis V. Lindley, *Understanding Uncertainty* (Hoboken, 2006), provides a fuller but still accessible overview. Morgan and Henrion, *Uncertainty*, 57–60, 62–64; Bedford and Cooke, *Probabilistic Risk Analysis*, 33–34; Terje Aven, *Foundations of Risk Analysis* (Chichester, 2012), 47–49, illustrate the personalist interpretation of randomness.

in a wide range of fields engaged in describing the present and predicting the future. Its strongest proponents view it as the only rational framework to deal with uncertainty: "If you want to handle uncertainty, then you must use probability to do it, there is no choice." "It is very firmly our opinion that the uniquely suitable representation of uncertainty, whether aleatory or epistemic, is probability."[9]

Past, Present, and Future   On this subjective interpretation, probability is precisely the right form in which to represent uncertainty about the past. Subjectivists see no qualitative difference between uncertainty about the past, present, and future, which are all equally uncertain from the perspective of an observer in the present. Some archaeologists and historians have already embraced probability for the formal representation of epistemic uncertainty about dating. Absolute chronology construction (such as carbon dating) and phylogeny (both genetic and linguistic) are sub-fields that routinely use Bayesian methods of inference based on subjective probability. Yet these technical fields are remote from the experience of most archaeologists and historians. A few archaeologists have also applied a probabilistic approach to the more quotidian exercise of dating artifacts based on established chronologies of types. For example, the date of deposition of a sherd of African Red Slipware of Hayes form 1—a pottery type associated with the period from 50 to 80 C.E.—can be represented as a probability distribution over that period (using a uniform, normal, or any other distribution depending on assumptions about the processes

9   For the quotations, see Lindley, *Understanding Uncertainty*, 239; Anthony O'Hagan and Jeremy E. Oakley, "Probability Is Perfect, but We Can't Elicit It Perfectly," *Reliability Engineering & System Safety*, LXXXV (2004), 247. Some theorists accept the distinction between the two types of uncertainty and account for them separately, but they use probability in both cases. See, for example, Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach* (Chicago, 1993), 24–25; Stan Kaplan, "Formalisms for Handling Phenomenological Uncertainties: The Concepts of Probability, Frequency, Variability, and Probability of Frequency," *Nuclear Technology*, CII (1993), 137–142 (introducing the "probability of frequency" framework in which aleatory uncertainty takes the form of frequentist probabilities and epistemic uncertainty subjective probabilities); David Vose, *Risk Analysis: A Quantitative Guide* (Chichester, 2008), 47–49. Others who fully embrace the subjectivist perspective that all uncertainty contains an epistemic element and should therefore be represented by subjective probabilities include Lindley, *Understanding Uncertainty*, and Aven, *Foundations of Risk Analysis*. For alternatives to subjective probability, see Franz Huber, "Formal Representations of Belief," in Edward N. Zalt (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016), available at https://plato.stanford.edu/archives/spr2016/entries/formal–belief/.

of production and deposition). Whereas these archaeologists use probability to represent the epistemic uncertainties about the dates of hundreds or thousands of individual sherds or other artifacts, this research note generalizes this approach to epistemic uncertainty about quantities other than dates.[10]

The uncertainties that historians face are clearly epistemic. They can be represented as probabilities, but only if the probabilities are understood as subjective. A probability is meaningful only in relation to a particular state of knowledge. It represents a historian's degree of belief based on a body of evidence, not an objective randomness. Hence, this research note speaks of *assigning* not *estimating* probabilities and employs the term *beliefs* to foreground the subjectivity that is inherent in the encounter with uncertainty. According to this view, beliefs comprise a set of evidence-based probabilistic judgments about historical uncertainties. "Your" beliefs about the past may differ from "mine" if "you" have access to more (or less) information than "I" have or if "you" interpret it

10   For the equivalence of past, present, and future, see Lindley, *Understanding Uncertainty*, 2–7, which offers twenty examples that deliberately conflate the three time frames, as well as the similar remarks of Buck et al., *Bayesian Approach*, 54: "The view adopted in this book is that assessments of probability are subjective and made in the light of experience: there is no difference in kind between the bookmaker's estimate of odds, the architectural historian's view of a date for a medieval building, the doctor's diagnosis, the archaeologist's opinion about the provenance of a pot, or the uncertainty in a scientist's estimate of the distance of the sun from the earth." For absolute chronologies, see Buck and Meson, "On Being a Good Bayesian" and other articles in the special issue "Prehistoric Bayesian Chronologies," *World Archaeology*, XLVII (2015), 575–700. For detailed discussions about the use of probability to date artifacts, see David L. Carlson, "Computer Analysis of Dated Ceramics: Estimating Dates and Occupational Ranges," *Southeastern Archaeology*, II (1983), 8–20; John M. Roberts et al., "A Method for Chronological Apportioning of Ceramic Assemblages," *Journal of Archaeological Science,* XXXIX (2012), 1513–1520; Enrico R. Crema, "Modelling Temporal Uncertainty in Archaeological Analysis," *Journal of Archaeological Method and Theory*, XIX (2012), 440–461; Rinse Willet, "Experiments with Diachronic Data Distribution Methods Applied to Eastern Sigillata A in the Eastern Mediterranean," *Herom*, III (2014), 39–69; Mike J. Baxter and H. E. M. Cool, "Reinventing the Wheel? Modelling Temporal Uncertainty with Applications to Brooch Distributions in Roman Britain," *Journal of Archaeological Science*, LXVI (2016), 120–127. Practical applications include Elizabeth Fentress and P. Perkins, "Counting African Red Slip Ware," in Attilio Mastino (ed.), *L'Africa Romana: Atti del V Convegno di studio Sassari, 11–13 dicembre 1987* (Sassari, 1988), 205–14; Martin Millett, "Pottery: Population or Supply Patterns? The Ager Tarraconensis Approach," in Graeme W. W. Barker and John Lloyd (eds.), *Roman Landscapes: Archaeological Survey in the Mediterranean Region* (London, 1991), 18–26; Andrew Wilson, "Approaches to Quantifying Roman Trade," in Bowman and Wilson (eds.), *Quantifying the Roman Economy: Methods and Problems* (New York, 2009), 213–49.

differently. But describing beliefs in terms of probabilities clarifies our differences and facilitates dialectic.[11]

FROM BELIEFS TO PROBABILITY DISTRIBUTIONS   The uncertainty about historical quantities such as the population of the Roman Empire can usefully be expressed as a probability distribution. This distribution represents a historian's degree of belief in different possible values, given the available evidence.[12]

What do I believe about the peak population of the Roman Empire? Without going into further detail, I am persuaded by Frier and Scheidel that a total population in the 60 million range is more likely than anything lower or higher. Yet identifying a most-likely value is only part of the process of estimation. We also need to ask how wide a range is possible. Unfortunately, past scholarship tended to focus almost exclusively on identifying a most-likely value rather than establishing upper and lower limits. For the purposes of this discussion, 40 million will serve as a minimum, since there is little scope to lower McEvedy and Jones' already minimal figure of 46 million. Establishing a ceiling is more difficult. Most modern scholars find Beloch's suggestion of a population of around 100 million too high, since the combined population of the former territories of the Empire do not appear to have exceeded that level until 1800. Beloch's figure can stand as a maximum, with the proviso that the matter deserves further consideration.[13]

11   For this sense of *belief*, see Lindley, *Understanding Uncertainty*, 12–13. The use of subjective probability should be distinguished from formal Bayesian inference, which involves not just subjective probability but also the use of Bayes' theorem to update a priori probabilities given data: Cailtlin E. Buck et al., *Bayesian Approach to Interpreting Archaeological Data* (Chichester, 1996) explores the potential of Bayesian approaches in archaeology; Buck and Bo Meson, "On Being a Good Bayesian," *World Archaeology*, XLVII (2015), 567–584, is a recent review of progress to date. For many problems, because we lack meaningful data with which to update our subjective beliefs, we can make only the first step, quantifying informed but subjective knowledge as probabilities (analogous to the formulation of "priors" in the Bayesian framework).

12   This research note focuses on uncertain *quantities*, as opposed to another type of uncertainty, which concerns not the value of a quantity but the truth of a proposition. These uncertainties are termed *events* in probability theory. Much uncertainty in history concerns events in this technical sense. Although the discussion herein is limited to quantities, the framework can also accommodate uncertainty about events, by assigning a probability to the proposition that the event is true. On the concept of event, see Lindley, *Understanding Uncertainty*, 12.

13   The estimate of the 1800 population of the Roman Empire's former territories derives from data in McEvedy and Jones, *Atlas of World Population History*, 43, 57, 63, 65, 87, 89, 93, 97, 105, 107, 113, 115, 135, 139, 143, 221, 225, 227.

Fig. 1   Selecting a Distribution
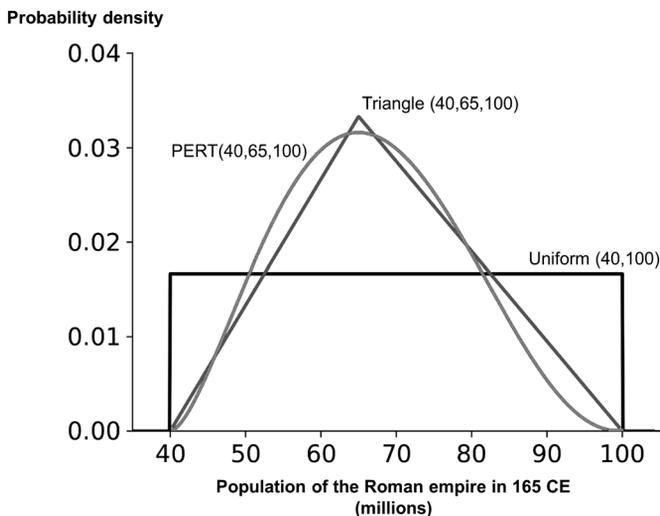
The Probability Density Function   Uncertainty about a continuous quantity (one that can take any value within a range) is represented by *probability density function* (PDF). To facilitate computation, the PDF implicit in our beliefs is approximated by some known mathematical distribution. The simplest available distribution is the *uniform* distribution, which assigns an equal probability to all possible values between a minimum and maximum. Figure 1 illustrates a uniform PDF for the peak population of the Empire, implying that all values between 40 million and 100 million are equally likely. The obvious objection to this representation of the uncertainty about the Roman population is that it fails to take any account of my belief that a value c. 65 million is much more likely than one around 40 or 100 million. The uniform distribution can be useful in expressing an absence of information or introducing a conservative element into an analysis, but it is clearly a crude representation of the state of knowledge in this case.[14]

14   With continuous variables, only intervals can be assigned a discrete probability. The probability of an interval is represented by the area under the probability-density function within that interval. The total area under a PDF always sums to 100 percent. Strictly speaking, Roman population is a discrete rather than a continuous variable (since there are no fractional persons), but the number of possibilities (in the tens of millions) is so large that it can be treated as continuous for convenience. This brief and discursive discussion of probability is intended only to pique the interest of historians. For a formal introduction to probability and probability distributions aimed at archaeologists, see Buck et al., *Bayesian Approach*, 47–65.

A simpler but much better alternative is the *triangle* distribution, which introduces a third parameter, the most-likely value (the point of highest probability). Figure 1 illustrates a triangle distribution with 65 million as the most-likely value. Though still crude, it is a much better representation of the way my degree of belief falls away from the most-likely value toward the minimum and maximum possible values.[15]

The graphic representation also reveals an aspect of the problem obscured by a focus on the most-likely value. Since the possible range extends further above 65 million than it does below, the uncertainty is asymmetrical, thereby making the most-likely value a biased estimator of the actual value. In this case, the distortion is relatively minor, but asymmetry is common in epistemic uncertainty and can be more pronounced (the probabilistic approach has a solution to this problem, as presented below).

The triangle distribution remains imperfect in at least two respects. It exaggerates how quickly my degree of belief declines in the immediate area of the most-likely value (a value of 65 million is not much more likely than a value of 60 million), and it assigns too high a probability to extreme values, especially in the 90 to 100 million range. A curve with attenuated tails would be better.

PERT Numerous distributions could serve that purpose, including the beta, gamma, Weibull, and Burr distributions. Fitting these distributions to particular beliefs can be computationally complex, however, since the parameters that define them are abstract quantities without a real-world interpretation. The most intuitive to manipulate is the PERT (Program Evaluation Review Technique) distribution (a special case of the beta distribution), in which the three parameters have the same interpretation as those for the triangle distribution—minimum, most-likely, and maximum values. Figure 1 illustrates the use of a PERT distribution to represent the uncertainty about the Roman population. It is a marginally better approximation because of its rounded peak and more attenuated tails. However, the computationally simpler triangle distribution may be adequate in many cases. Practitioners in many other fields have found triangle distributions to be a convenient

15 For the use of triangle distributions to represent epistemic uncertainty, see Vose, *Risk Analysis*, 403; Morgan and Henrion, *Uncertainty*, 96.

way to represent subjective probabilities in situations of epistemic uncertainty. In any case, a historian is free to choose from a wide range of distributions to find the best representation of the uncertainty.[16]

This exercise of assigning probabilities is not as outlandish as it may appear. The procedure of encoding beliefs as probability distributions has become well established in other disciplines. Even in fields with much better data, estimation often entails an irreducible element of subjective judgment. Forecasting and risk models regularly include quantities for which values are assignable only via the subjective judgment of experts. The problem is sufficiently widespread to have generated a whole literature devoted to the "elicitation" of expert opinion in the form of probability distributions.[17]

The procedure merely makes explicit a probabilism that is already implicit in historical argument. When historians from Beloch to Frier proposed point estimates for the population of the Empire, they were presumably reporting the value that they judged to be the most likely—the peak of their probability distribution for the quantity. Their rejection of other estimates as less likely implies their assignment of lower probabilities to those values, whereas their rejection of other suggestions as implausible or impossible means that their probability distributions were at, or near, zero (otherwise their rhetoric is misleading). Making the probabilities explicit would help to clarify the positions and focus attention on the degree of uncertainty.

The purpose of the exercise is not to estimate some objective probabilities but rather to use probabilities to represent uncertainty. Kaplan, who helped to develop the probabilistic approach to risk analysis, made the point succinctly: "People often think that putting

16 The more-familiar normal distribution, though often used to represent uncertainty arising from variation and measurement error, is unsuitable in this context since it is strictly symmetrical—the uncertainty in this case being clearly asymmetrical—and extends infinitely in both directions. For further discussion of distributions, see Vose, *Risk Analysis*, 401–410; Paul H. Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, 100 (June 2005), 688–689.

17 For the widespread need to rely on subjective judgments in future-oriented fields, see, for example, Anthony O'Hagan et al., *Uncertain Judgements: Eliciting Experts' Probabilities* (Chichester, 2006), 97–120; Vose, *Risk Analysis*, 393–422; for a brief overview of the elicitation of expert opinion, Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions"; for a fuller synthesis of research in the field, O'Hagan et al., *Uncertain Judgements*.

forth an uncertainty curve is somehow difficult, compared with giving a single number or 'point estimate.' It becomes much easier if we remind ourselves that probability curves 'do not exist,' as De Finetti said. ... They are only a language in which we express our state of knowledge or state of certainty. With this understanding, it is easy to put forth a curve (fat, if necessary) to express our uncertainty. What is far more difficult is to put forth a single number that people are going to believe and use for design and regulatory decisions, as if it were gospel truth."[18] The same argument holds for historical estimates.

COGNITIVE BIASES   Historians should be aware of two kinds of cognitive bias that can affect any attempt to estimate uncertain quantities, not just the formal probabilistic approach developed herein. The most significant bias is *overconfidence*. People estimating uncertain quantities usually produce ranges that are too narrow because they assign probabilities too close to zero for values that are very low or very high. The overconfidence becomes more pronounced as estimation becomes more difficult (the "hard-easy effect"). Overconfidence is a major issue for historians appraising uncertain quantities when their information is scant. Scholars who have worked on the population of the Roman Empire are likely to have been too quick to dismiss relatively low or high values as implausible. The range of plausible values is probably wider than they have suggested.[19]

The second important bias, the *anchoring* effect, arises from one of the "heuristics" identified by Kahnemann and Tversky in their work on the cognitive shortcuts that people take when making decisions based on limited information. Estimates of uncertain quantities are often distorted by preconceived values as a result of a heuristic that first evaluates proposed estimates as high or low and then corrects them—Kahnemann and Tversky's "judgment by anchoring and adjustment." As a rule, such corrections tend to

---

18   Kaplan, "Formalisms for Handling Phenomenological Uncertainties," 141.
19   The study of judgment under uncertainty began with Amos Tversky and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, CLXXXV (1974), 1124–1131. For syntheses of research in the field, focusing on the implications for the formulation of expert knowledge in probabilistic form, see O'Hagan et al., *Uncertain Judgements*, 33–60; Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," 682–684, 685; for more about overconfidence, Ward Edwards et al., *Advances in Decision Analysis: From Foundations to Applications* (New York, 2007), 143–144; Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," 685.

be too small. The implication is that historians are likely to anchor to previously published estimates, regardless of their quality, when forming their opinions. Every estimate of the Roman population's size has probably been anchored to its predecessor. The single largest swing in the modern history of the debate was proposed by Beloch himself, when he updated his own estimate for 14 C.E. from 54 to 70 million. He was presumably more acutely aware of the uncertainties in his own estimate than later readers would have been. Subsequent revisions tended to be more modest. McEvedy and Jones reduced Beloch's estimate by around 14 million in their estimate for 1 C.E. Frier nudged their estimate for 1 C.E. upward by just 5 million, and Scheidel adjusted Frier's estimate for 164 C.E. by just 4 million. Given the existence of the anchoring effect, these corrections may well have been too small.[20]

Considerable research has been devoted to our capacity to suppress these biases. The most fruitful approach involves a training process in which individuals repeatedly estimate a quantity before being confronted with the actual value. It has proved effective in calibrating the probability judgments of forecasters such as meteorologists, but it is of little use to historians, who rarely have the opportunity to compare their estimates to actual values. The best that historians can do is to be aware of the biases affecting their judgment and try to compensate for them. A formally probabilistic approach provides the best framework for doing so.

COMBINING UNCERTAINTIES   Conceptualizing beliefs about uncertain quantities as probability distributions is a useful intellectual discipline. Ancient historians are accustomed to confining their disputes to point estimates, asserting a particular value or range to be the "most likely" without stating their confidence that the actual value was close to their proposed value and without considering how much less likely the rejected rival estimates were. The rigor of thinking in terms of probability distributions forces historians to confront these difficult but important questions. Its real value, however, lies in the aggregation of uncertainties. One

---

20   For the experiment, see Tversky and Kahneman, "Judgment under Uncertainty." O'Hagan, et al., *Uncertain Judgements*, 47–49; Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," 682.

key advantage of probabilities is that they are easy to combine mathematically.

Monte Carlo Simulation   Many future-oriented fields combine uncertainties using Monte Carlo simulation. This technique involves three steps: (1) the construction of a mathematical model to represent the quantity of interest as a function of several better-understood quantities, in which all uncertain quantities are represented as random variables—the quantity of interest being the output variable and the better-understood quantities the input variables; (2) the conversion of observer's beliefs about the input variables to probability distributions, including the assessment of epistemic interdependence between input variables (discussed below); and (3) the generation of a series of scenarios through the random selection of input values from the probability distributions for the input variables. The output that the model produces in each scenario can be regarded as a random sample from the probability distribution for the output variable. As the number of scenarios increases, the distribution of the output values in the sample will converge on the underlying probability distribution of the output variable.[21]

The Monte Carlo method, which emerged in the context of nuclear engineering during and after World War II, has gained wide application in forecasting, risk assessment, and decision analysis. Because it is most often used in contexts that appear to involve aleatory uncertainty, its relevance to historical problems, in which the uncertainty is clearly epistemic, has been underestimated. Yet it offers historians a useful tool to aggregate epistemic uncertainties. For instance, it offers a potential solution to a hitherto intractable problem in Roman history, estimating the proportion of the population that had Roman citizenship before Caracalla's universal grant in 212/13 C.E. —crucial to the assessment of the significance of that decision. Even though the mechanisms by which Roman citizenship was disseminated were relatively well understood, quantification seemed impossible because of the many uncertainties involved. Any estimate of the proportion of persons who were citizens in 212 would require estimates of the number of new citizens created by 200 years of communal and personal grants, service in the army, office holding in provincial cities, and the manumission of

---

21   For an excellent practical guide to Monte Carlo simulation, see Vose, *Risk Analysis*.

slaves (the principal routes to citizenship), as well as of the total population of the empire, all of these variables being themselves uncertain. A traditional point estimate based on most-likely values for each of the input quantities could never hope to command credibility because of the proliferating uncertainties. A probabilistic approach using Monte Carlo simulation makes it possible to account for all the component uncertainties and to assess the aggregate uncertainty about the prevalence of citizenship.[22]

INTERPRETING THE RESULTS   The output of a Monte Carlo simulation is most intuitively grasped through a histogram. Figure 2 shows the result of a simulation of the spread of Roman citizenship. In each random scenario, the model calculates the prevalence of citizens in 212 C.E., expressed as a percentage of the free population. The histogram shows how often different prevalences occurred in a sample of 50,000 scenarios. The shape of the distribution approximates the probability density function for the prevalence of citizens that is implied by the beliefs encoded in the model. The distribution peaks in the interval 22 to 23 percent. The most-likely value is thus around 22.5 percent. But the mean value, 24 percent, is actually the best point estimator of the quantity (termed the *expected value* or *expectation* of the uncertain quantity in probability theory), because it is the probability-weighted average of all possible outcomes. Unlike the most-likely value, it takes account of any asymmetry in the uncertainty (the fact that the distribution is slightly skewed to the right in Figure 2). This is the best solution to the problem of asymmetry noted earlier, which can cause problems for reasoning stemming from most-likely values alone.[23]

In this case, the improvement in point estimation is modest (though it may be more pronounced in situations of greater asymmetry). The overall shape of the distribution is more important. A most-likely value would have been relatively easy to establish using a traditional point-estimate approach. It would also be obvious that a value near the most-likely value is more plausible than any higher or lower value, but a historian would not otherwise

22   Lavan, "Spread of Roman Citizenship."
23   For the expected value of an uncertain quantity, see Lindley, *Understanding Uncertainty*, 137–139.

Fig. 2 Monte Carlo Simulation of the Prevalence of Citizenship

Frequency of outcomes (%)

Citizens / all free provincials in 212 CE
(%)
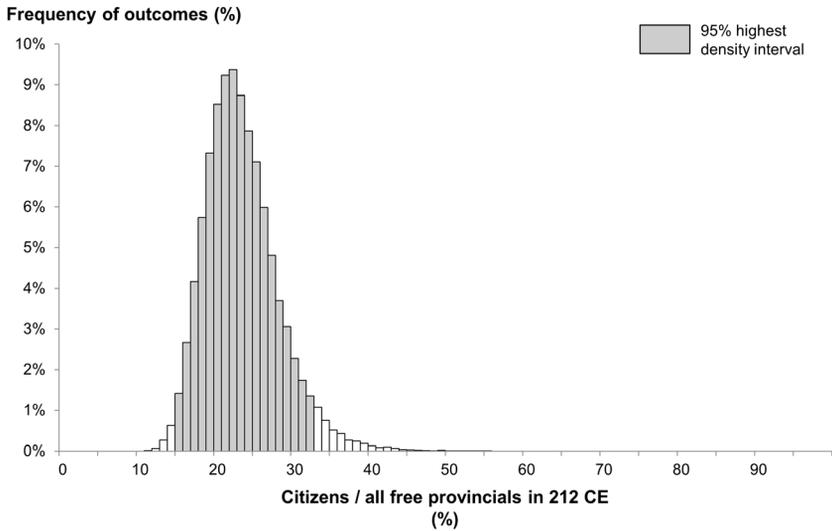
have grounds to establish *how much* less likely outlying values are. The benefit of the Monte Carlo simulation is that it quantifies the decline in plausibility.

Estimation always entails a trade-off between confidence and precision. Ceteris paribus, the wider the range, the more confident we can be that it includes the actual value. But a wider range also contains less information about the quantity. Most strategies for managing uncertainty depend on discounting some possibilities as highly unlikely. In the case of Figure 2, which is a probability density function representing degree of belief, this strategy means discounting the tails of the distribution and reporting a specified *credible interval* (a Bayesian *credible interval*—to be distinguished from a frequentist *confidence interval*, which does not admit a probabilistic interpretation—is a contiguous interval that contains a specified proportion of the total probability mass). Which interval to report is a matter of convention. Many disciplines operate with a 95 percent threshold for estimation (though the threshold is often applied in a frequentist rather than Bayesian framework). Some fields with better data hold themselves to a higher standard, operating with a 99 percent, or even higher, threshold. In a field as data-poor as ancient history, a lower threshold of 80 percent may well be appropriate. Most ancient historians seem to operate with even lower

thresholds in the ranges that they report, but the issue is never discussed. In this case, the 95 percent credible interval for the prevalence of citizenship is 15 to 33 percent. In other words, that range is sufficient to enclose 95 percent of the probability mass. This estimate incorporates the uncertainty about the population of the Empire and the other relevant variables. The resulting range, though broad, represents an important advance in our understanding of an important quantity that had hitherto resisted quantification entirely.[24]

The Monte Carlo simulation demonstrates that beliefs about the input variables and the laws of probability together constrain beliefs about the quantity of interest. The underlying logic is that of *coherence*—the axiom that a set of probabilistic judgments has to be internally consistent to be valid. Our beliefs about the processes that disseminated citizenship (as encoded in the identification of variables and the mathematical model that links them to the prevalence of citizenship) and about the historical values of those variables (as encoded in the input probability distributions) impel us to assign a much higher probability to some possible values of the quantity than to others. In other words, we learned that we already knew enough about the mechanisms of enfranchisement and the demography of the Empire to be confident that the proportion of the population who had citizenship in 212 was between 15 and 33 percent.[25]

The aggregation of probability distributions through Monte Carlo simulation is a better method of manipulating uncertain quantities than traditional approaches that collapse uncertainty by treating all variables as point estimates. In some cases, as in this example, Monte Carlo simulation will produce credible intervals that are usefully narrow, revealing that historians knew more than they realized about the quantity of interest. In other cases, however, even an 80 percent credible interval may be too broad to be informative. But that finding would also be significant, demonstrating the vulnerability of any existing point estimates for the quantity.

EPISTEMIC INTERDEPENDENCE   It is essential to consider whether there is any interdependence between the uncertainties that are

24   For Bayesian credible intervals, see O'Hagan et al., *Uncertain Judgements*, 234–235.
25   For the principle of coherence, see Lindley, *Understanding Uncertainty*, 36–37, 236–237; for the application discussed herein, Aven, *Foundations of Risk Analysis*, 97–98.

being combined. The Monte Carlo approach can cope with interdependence, but only if it is taken into proper account. The interdependence in question is specifically *epistemic*, a matter of the interdependence of historical problems. Two quantities are epistemically interdependent if acquiring new information about one quantity would change the historian's beliefs about the second quantity.[26]

Returning to the problem of the Roman population, the uncertainties about the population of Italy and the population of Iberia are clearly not independent. All estimates for Iberia are informed by an assumption that Iberia was less densely populated than Italy was. If we were to discover that the actual peak population of Italy was toward the top of our range of possible values, we would have to adjust our probability distribution for the population of Iberia accordingly. Modeling the two variables as independent would fail to account for this interdependence and produce a meaningless result. But many other uncertainties can be regarded as independent. Take the peak population of the Empire and the proportion of slaves who were freed. In this case, discovering the exact value of the population would in no way reduce my uncertainty about the freed slaves. The two can thus be treated as independent.

Various strategies are available for managing epistemic interdependence once it has been identified, but unacknowledged epistemic interdependence is the potential Achilles heel of any probabilistic model. The most dangerous pitfall is ignoring strong epistemic interdependence that makes extreme outcomes more likely. The risk of such an error is increased by historians' desire to produce a narrower and hence more informative estimate. Epistemic interdependence needs to be accounted for carefully.[27]

---

26   For epistemic (also termed probabilistic and subjective) interdependence, see Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," 686; O'Hagan et al., *Uncertain Judgements*, 107–108, 243; Lindley, *Understanding Uncertainty*, 52–53.

27   Interdependence either reduces the variance of the resulting probability distribution (that is, produces a narrower credible interval), if the effect of an extreme value for one variable is partly offset by a correspondingly extreme value for the other variable, or increases it, if the effects of extreme values compound each other (as would be the case with the populations of Italy and Spain with regard to the total population of the Empire). For different ways to incorporate interdependence, see Garthwaite et al., "Statistical Methods for Eliciting Probability Distributions," 687; Vose, *Risk Analysis*, 356–364.

SENSITIVITY ANALYSIS AND ITERATION   Monte Carlo simulation is useful not just for arriving at an estimate but also for clarifying the structure of a problem. Various types of sensitivity analysis can be performed to measure how much the individual input variables contribute to the overall uncertainty about the output variable. The point is to identify the most important components of uncertainty. Sensitivity analysis of this sort is particularly important for historians because the Monte Carlo approach works best as an iterative process. The information relevant to problems in ancient history tends to be dispersed and difficult to interpret. Even discounting the discovery of new information, an individual's assessment of the current state of knowledge can only be provisional because it has to rely on the work of others and is likely to omit at least some relevant information—for example, comparable data from other regions or periods. Hence, the assignment of probability distributions to the input variables should be an iterative process. It should start with a rough set of probability distributions (erring on the side of exaggerating the uncertainty) to identify the variables that contribute most to the uncertainty about the quantity of interest. The probability distributions for those variables can then be refined through a deeper review of the evidence, and so on, until the probability distribution for the quantity of interest begins to stabilize. Several iterations may be necessary before reaching a stable estimate (stable in the sense that further analysis of the available evidence is unlikely to change it significantly).[28]

This research note demonstrates the value of subjective probability as a tool of historical analysis. It provides a framework that can accommodate the significant epistemic uncertainty involved in estimates of historical quantities, especially (but not only) regarding periods for which we have limited data. Thinking in terms of probability distributions is always a good discipline because it draws attention to complexities that traditional approaches miss by focusing exclusively on a most-likely value. It becomes even more useful when multiple uncertain quantities are combined in a single analysis, a common occurrence in ancient history. Though it may appear a radical departure from current practice, it builds upon a probabilism that is already latent in historical reasoning.

---

28 For a good overview of the options for sensitivity analysis, see Vose, *Risk Analysis*, 80–88.

Most of the estimates that circulate in ancient history are implicit expressions of their proponents' probability distributions for the quantities in question, insofar as they represent the value judged to be the most likely, given the available evidence. But the traditional best-estimate approach leaves their beliefs about the likelihood of other possible values unclear or unexamined.

These probabilities have to be understood as subjective, in the technical sense. They are not estimates of objective probabilities that exist in the world (indeed, Bayesians admit no objective probabilities) but representations of the epistemic uncertainty about quantities that have a fixed value, albeit an unknown one. As such, the probabilities are both conditional and personal. They are conditional because they depend on a state of knowledge. New information will change the probabilities. They are personal because they rely on an individual scholar's assessment of the evidence. One scholar cannot dictate another's probabilities. Yet their probabilities should converge if they agree on the model and on the probability distributions for the input variables. If they disagree, expressing their beliefs in terms of probability distributions will clarify the area of disagreement and focus further research and debate.

The avowedly subjective character of the framework may trouble some historians. But it merely makes explicit an inherent feature of historical analysis. Historians can present their evidence and their arguments, but they can never coerce their colleagues' assent; other scholars may reach different conclusions from the same evidence. The most that they can hope is that their reasoning will prove persuasive. This condition is no different from my expectation that other scholars will recognize in my probability distributions a careful and honest representation of the state of knowledge. The subjectivity inherent in historical analysis is too often regarded as an intellectual defect to be obfuscated through a misleading rhetoric of objective authority. One of the great merits of this framework is that it acknowledges the irreducible subjectivity in all empirical disciplines and shows that it is not an obstacle to quantitative analysis.