Annual Review of Anthropology

Bayesian Statistics in Archaeology

Erik Otárola-Castillo and Melissa G. Torquato

Department of Anthropology, Purdue University, West Lafayette, Indiana 47907, USA;
email: eoc@purdue.edu, mtorquat@purdue.edu

Annu. Rev. Anthropol. 2018. 47:435–53

First published as a Review in Advance on August 3, 2018

The *Annual Review of Anthropology* is online at anthro.annualreviews.org

https://doi.org/10.1146/annurev-anthro-102317-045834

Keywords

Bayesian inference, R-statistics, frequentist, hypothesis testing, archaeostatistics

Abstract

Null hypothesis significance testing (NHST) is the most common statistical framework used by scientists, including archaeologists. Owing to increasing dissatisfaction, however, Bayesian inference has become an alternative to these methods. In this article, we review the application of Bayesian statistics to archaeology. We begin with a simple example to demonstrate the differences in applying NHST and Bayesian inference to an archaeological problem. Next, we formally define NHST and Bayesian inference, provide a brief historical overview of their development, and discuss the advantages and limitations of each method. A review of Bayesian inference and archaeology follows, highlighting the applications of Bayesian methods to chronological, bioarchaeological, zooarchaeological, ceramic, lithic, and spatial analyses. We close by considering the future applications of Bayesian statistics to archaeological research.

INTRODUCTION

Science requires empirical observation to corroborate or reject ideas and theories about how the world works. Fundamental to this process is statistics, the discipline that evaluates how well observations support theoretical expectations. Accordingly, observing the archaeological record and using statistical methods is one way that archaeologists assess the accuracy with which archaeological theory explains past people's behavior. Statistical methods were introduced in archaeological research through texts by Myers (1950), Spaulding (1953), Clarke (1968), and Binford (1964). Although archaeologists have continued to use statistical methods in modern archaeological research, their use has declined since the 1970s (Cowgill 2015). Nevertheless, today, numerous textbooks are dedicated to teaching introductory statistical concepts and techniques to archaeology students (Carlson 2017, Fletcher & Lock 2005).

Currently, the statistical methods most commonly used by archaeologists and other scientists are grouped into the statistical framework known as null hypothesis significance testing (NHST). Pioneers in statistical inference (e.g., Fisher 1925; Neyman & Pearson 1933, p. 294) popularized this statistical framework in the early twentieth century, making NHST methods widely available to scientists of the time. Today's NHST methods are associated with frequently used statistics such as confidence intervals and probability values (p-values). NHST employs these statistics to make probabilistic statements about one's data in relation to a hypothesis. Although this approach is very useful, the fundamental concepts in NHST can seem arbitrary and often confusing. For example, the relationship between p-values and null hypotheses is often unclear and consequently misunderstood.

As an alternative, archaeologists have begun to incorporate Bayesian methods into their tool kit. Bayesian inference offers an alternative and, in some respects, an improved statistical framework over NHST. Although Bayes' theorem, the fundamental core of Bayesian inference, was developed in the eighteenth century, it was only in the 1990s that the scientific use of Bayesian statistics increased significantly (Robert & Casella 2011). Since then, Bayesian statistics has become fundamental to the scientific endeavor in general and increasingly common in anthropological and archaeological science (Gelman et al. 2014, McElreath 2016) (**Figure 1**). One reason for the appeal of Bayesian methods is the relative simplicity of the interpretation compared to NHST. Whereas NHST makes statements about the probability of data with respect to hypotheses about a population, Bayesian inference makes statements about the probability of hypotheses in light of data. For example, Bayesian inference allows researchers to use their data to assign probabilities to their hypotheses, rather than using a p-value to reject or fail to reject a null hypothesis. Instead of confidence intervals, Bayesian credible intervals assign a probability range to estimated parameters.

Here, we provide a simple explanation of Bayesian statistics in comparison to NHST and review how archaeologists have been applying Bayesian statistics to solve problems. Although the

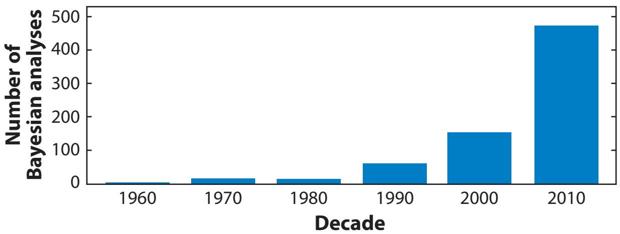

Figure 1

Bar plot showing the number of Bayesian analyses in archaeology per decade between the 1960s and 2010s.

use of Bayesian methods by archaeologists has grown considerably, in particular with respect to radiocarbon dating and chronology, this review widens its lens to include Bayesian applications related to problems in lithic, ceramic, zooarchaeological, bioarchaeological, and spatial analyses. We have compiled a database of ~700 published archaeological manuscripts that rely on Bayesian statistics to make inferences about their respective topics (see **Figure 1**). We focus here on key manuscripts within the archaeological subdisciplines mentioned.

We begin our review with an example that contrasts how NHST and Bayesian inference might function to solve an archaeological research problem, illustrating why the latter approach seems less ambiguous and thus is becoming more popular. Following this example, we more formally define both the NHST and Bayesian frameworks, providing a brief history of both. We then review the applications of Bayesian statistics in the archaeological literature and make predictions about the future of Bayesian and NHST applications in archaeology.

A SIMPLE EXAMPLE COMPARING NHST TO BAYESIAN INFERENCE

We are interested in the effect of projectile-propelling technology on the change in size (length) of stone points excavated from an archaeological site made up of two components from different time periods: early period and late period. For comparison, we have simulated the length measurements of a large sample of ethnographic stone projectile points propelled using three different technologies: (*a*) bow and arrow (arrow tips), (*b*) atlatl (dart tips), and (*c*) hand throwing (spear tips). The archaeological sample is composed of 10 specimens from the early period with a mean length of 6.1 cm and a standard deviation (SD) of 2 cm. The late period sample consists of 9 specimens with a mean length of 13 cm and an SD of 3.2 cm. These are the test statistics of each sample. We refer to these values as "the data."

We know that these size measurements are maximum lengths and are normally distributed. We also know the parameter values (means and SDs) for the distribution of lengths related to each type of technology (arrow tips: mean = 6.9 cm, SD = 2 cm; dart tips: mean = 11 cm, SD = 2 cm; spear tips: mean = 14 cm, SD = 2 cm). We can use these measurements as the hypothesized parameter values for each technological population. In the example, these parameter values are termed "the hypotheses." We then use these values to compare the population parameters to our archaeological samples in order to infer which technology might have propelled the projectile points at each component of the site.

To infer the most likely propelling technology underlying each period, we have the option of using NHST or a Bayesian framework. Below, we describe how we would proceed following each approach.

Null Hypothesis Significance Testing

We can easily use the common z-score or z-test method to evaluate the probability of obtaining such data given that the hypothesized population parameters are true. We know how to compute a z-score and find its associated p-value to conduct this test. If the calculation generates an extreme or unexpected z-score and associated probability (usually a p-value <0.05), then the archaeological data set has a low probability of having been sampled from the hypothesized population. This result provides enough evidence to reject the hypothesis. If, however, the calculation yields an expected z-score value and probability, there is not enough evidence to reject the given hypothesis.

Early period. In this example, we begin by evaluating whether the early period data could have been sampled from a population with hypothesized parameters representative of the arrow tip population. A z-score of -1.26 is computed along with an associated two-tailed p-value of 0.21. Because this is not an extreme value (it is not less than 0.05), we cannot reject the hypothesis that the early period data might have resulted from a population of projectile points propelled by bow and arrow. Conducting the z-score test again for dart tips and spear tips yields very low p-values (9.366013e-15 and 8.359965e-36). We reject these respective hypotheses because these p-values are lower than 0.05 and therefore considered extreme.

Results of the NHST approach to the early period data allow us to make two statements. First, the sample has a very low probability of resulting from a population of dart tips or spear tips. Second, the sample does not have a low probability of resulting from a population of arrow tips. Whereas the first statement seems clear, the second can be difficult to understand, which raises the question, "Do these findings mean that the sample has a high probability of belonging to a population of arrow tips?" The answer is, no, it does not because the NHST framework does not evaluate the probability of a hypothesis; rather, it evaluates the probability of the data in the context of a hypothesis about a population.

Late period. Following a similar logic, we conduct a z-test to evaluate whether the late period data could have been sampled from a population with hypothesized parameters belonging to the three stone-tip populations. Beginning with the arrow tips, we compute a z-score of 5.72 with a two-tailed p-value much lower than 0.05. This result allows the archaeologist to reject the hypothesis that the late period stone points belong to a population characteristic of arrows. The archaeologist also computes z-scores for darts ($z = 1.87$) and spear tips ($z = -0.94$) with associated p-values of 0.06 and 0.35, respectively. Following the NHST logic, the archaeologist can state that the late period sample has a very low probability of resulting from a population of arrow tips. In addition, the sample does not have a low probability of resulting from a population of dart tips or from a population of spear tips. In conclusion, the archaeologist may state that the stone tip size data indicate the occurrence of a shift in weaponry technologies between the early and late periods. However, the results remain unclear. The archaeologist cannot state whether the late period stone points belong to the dart- or spear-tip populations because p-values do not evaluate the probability of the hypotheses. This NHST statistical scenario is very common across the sciences, giving Bayesian approaches some key advantages over the NHST approach.

Hypotheses and Bayesian Inference

In many cases, scientists might be interested in assessing the probability of their hypotheses given the data on hand, especially when multiple hypotheses are involved; yet, the NHST approach does not facilitate this. Bayes' theorem addresses this concern by taking the NHST likelihood statement along with the a priori knowledge of the phenomenon (known as the prior probability) to compute the posterior probability of a hypothesis. Although incorporating prior information is a useful feature of Bayesian inference, many scientists have concerns about adding this information, as it can include the subjective knowledge of experts (Gelman et al. 2014). To avoid subjectivity, or if additional knowledge is not available, many analysts incorporate so-called uninformative, flat, or vague priors, which assign equal probabilities to every state of the estimated parameters. Bayesian inference allows scientists to speak about the probabilities of their hypotheses and more clearly compare multiple hypotheses based on their probability of being correct.

Working under a Bayesian framework, we would need to make our hypotheses explicit in order to compute their probabilities in light of the data on hand. The hypotheses can be outlined as follows:

Hypothesis 1: the observed sample functioned as arrow tips,

Hypothesis 2: the observed sample functioned as dart tips, and

Hypothesis 3: the observed sample functioned as spear tips.

To compute the probability of each hypothesis, we calculated posterior probabilities using the rstan package (Carpenter et al. 2017, McElreath 2016, Stan Dev. Team 2018) in the R computing environment (R Core Team 2018). Stan is one of the leading software packages to conduct Bayesian analysis. We use rstan because it enables us to run Stan through R; our code is provided in the **Supplemental Material**. We evaluate the probabilities of each hypothesis first for the early period and then for the late period.

Supplemental Material >

Early period. For the early period, given the data on hand, the archaeologist can state the following:

1. "There is a 0.97 probability that the early period's average stone tip length falls within the range of arrow tips, 6.9 ± 2 cm."
2. "There is a 0.004 probability that the early period's average stone tip length falls within the range of dart tips, 11 ± 2 cm."
3. "There is a 0.00003 probability that the early period's average stone tip length falls within the range of spear tips, 14 ± 2 cm."

In addition, using Bayesian credible intervals, we can also state, "There is a 0.90 probability that the true average projectile point measurement from the early period is within 5.19–7.78—squarely within the range of arrow tips." This is a subtle but paramount difference in verbiage from the NHST framework. With the Bayesian framework, we can now speak directly about the probabilities of the hypotheses on hand.

Late period. For the late period, given the data on hand, the archaeologist can state the following:

1. "There is a 0.001 probability that the late period's average stone tip length falls within the range of arrow tips, 6.9 ± 2 cm."
2. "There is a 0.16 probability that the late period's average stone tip length falls within the range of dart tips, 11 ± 2 cm."
3. "There is a 0.89 probability that the late period's average stone tip length falls within the range of spear tips, 14 ± 2 cm."

We can also state, "There is a 0.90 probability that the true average projectile point measurement from the late period is within 12.08–16.2 and is well within the range of spear tips." In addition, the probability values of all three hypotheses clearly favor Hypothesis 3 nearly five times more than they favor Hypothesis 2. Unlike with NHST, the Bayesian framework allows us to make clear conclusions and provide our readers with unambiguous statements about the probabilities of the hypotheses.

NHST: GENERAL DESCRIPTION AND BRIEF HISTORY

The general NHST process estimates the value of some parameter, θ, of unknown value that summarizes a measurable characteristic of a population, A. We measure an n-sized sample, D, from A to infer the value of θ. The measurements of D result in variables $d_1, d_2, d_3 \ldots d_n$ ("the data"). We calculate the statistic $\hat{\theta}$ from the sample and hypothesize the value of the population parameter θ. This becomes H_0, the null hypothesis. We would like to know whether D as summarized by $\hat{\theta}$

is an expected outcome of a population under H_0 or an extreme value from a different population (H_A, the alternative hypothesis). To estimate a range of expected outcomes, we use the probability distribution function (PDF), $f(D)$, to compute the probability, P, of values at least as extreme as $\hat{\theta}$ under H_0. This procedure effectively asks, what is the probability of the data, summarized by $\hat{\theta}$, with respect to the null hypothesis? We can symbolize this conditional statement by

$$P(\hat{\theta}|H_0).$$

If $P(\hat{\theta}|H_0)$ is less than some predetermined cutoff probability value, or *p*-value, usually 0.05, there is a very small probability that the data on hand were sampled from a population under H_0. In other words, there is a significant difference between the statistic estimated from the data, $\hat{\theta}$, and the hypothesized population parameter, θ. The null hypothesis, H_0, is thus rejected in favor of H_A. However, if $P(\hat{\theta}|H_0)$ is greater than 0.05, then H_0 is not rejected.

NHST is the most widely used statistical framework in science and archaeology. The origins of NHST statistics are rooted in Karl Pearson's (1922) development of the chi-square goodness of fit test, which enabled scholars to compare competing models and arbitrate between them using the available data. This process of comparing and rejecting models was an important precursor to hypothesis testing in which the *p*-value is calculated and used to reject or fail to reject hypotheses (Salsburg 2001).

In 1925, Ronald A. Fisher's influential work *Statistical Methods for Research Workers* first described *p*-values for determining significant findings. The text's step-by-step layout made it popular among scientists. In Fisher's significance testing, experimental data were compared to a null hypothesis, using the calculated *p*-value as the basis for rejecting the null hypothesis. He described *p*-values as "an arbitrary, but convenient, level of significance for the practical investigator" (Fisher 1929, p. 191). Fisher (1929) assigned a *p*-value of 0.05 as the threshold for significance. The purpose of significance testing was to determine whether an experiment caused an effect or whether it supported the null hypothesis. According to Fisher, experiments that resulted in nonsignificant results should be discarded by researchers (Salsburg 2001). Unfortunately, Fisher never fully explained how *p*-values, as an "arbitrary" value, should be interpreted. Despite the lack of explanation, scholars readily adopted $p\text{-value} = 0.05$ as the significance criterion in their research, establishing one of the key limitations of NHST today.

Later work by Jerzy Neyman and Egon Pearson revolutionized significance testing by introducing the concept of an alternative hypothesis (Salsburg 2001). Although in Fisher's significance test a significant result caused the null hypothesis to be rejected, nonsignificant test results were unclear. Fisher's test compared experimental data to a normally distributed null hypothesis assumed to be true; yet, Neyman and Pearson questioned how researchers could be sure that data were normally distributed if the results of the significance tests were nonsignificant. This critique expanded to a broader question of how researchers could know if they were applying a significance test correctly. The uncertainty surrounding significance testing led Neyman and Pearson to develop hypothesis testing.

Neyman & Pearson (1933) suggested that data could be compared to a hypothesis only if another outcome was possible. Thus, hypothesis testing compares two competing hypotheses: a null hypothesis and an alternative hypothesis. The null hypothesis is tested against a set of alternative hypotheses. Each competing hypothesis is a distribution that the data may fit. Like Fisher's significance test, Neyman & Pearson's hypothesis test uses a *p*-value to test the null hypothesis, but Neyman & Pearson also introduced the concept of the power of a test or the probability of detecting that the alternative hypothesis is true (Salsburg 2001). Comparing two hypotheses allows researchers to choose the hypothesis with the highest probability. In effect, if the null hypothesis does not have a low probability (i.e., it is nonsignificant), it can be selected over the alternative hypothesis. Neyman & Pearson's hypothesis test was expanded in the textbook *Testing Statistical Hypotheses* (Lehmann 1959), which continues to be used today. Additional work by Neyman (1934, Neyman 1937) led to the development of confidence intervals, which are used to determine the accuracy of an estimate (Salsburg 2001).

Although NHST is the most widely used statistical evaluation, numerous behavioral and social science disciplines have questioned its usefulness in recent years (Benjamin et al. 2018, Gelman 2017, Gill 1999; but also see McShane & Gal 2017). Many of the questions raised focus on the lack of clear inference associated with *p*-values and confidence intervals. First, the interpretation of a nonsignificant result is often unclear. Moreover, NHST can never accept a null hypothesis. As Fisher (1935, p. 474) explained, "Tests of significance . . . are capable of rejecting or invalidating hypotheses, in so far as they are contradicted by the data; but that they are never capable of establishing them as certainly true." Scholars often misunderstand this point, attempting to use NHST to verify their null hypothesis. However, failing to reject a null hypothesis is not synonymous with holding the hypothesis as true. Rather, there is simply not enough evidence to state that it is not correct.

BAYESIAN INFERENCE: GENERAL DESCRIPTION AND BRIEF HISTORY

We stated above that NHST makes statements about the probability of data given a hypothesis, described by $P(\hat{\theta}|H_0)$. Bayesian inference makes statements about the probability of hypotheses about a population (H), given a set of data (D). This can be symbolized by $P(H\,|\,D)$. In the context of data and hypotheses, Bayes' theorem functions as follows:

$$P(H\,|\,D) = \frac{P(D|H)P(H)}{P(D)},$$

where $P(H|D)$ is the posterior probability, or the probability of the hypothesis given the data, $P(D|H)$ is the probability of the data given the hypothesis, or the likelihood of the observed data, $P(H)$ is the prior probability of the hypothesis, and $P(D)$ is the probability of the data set in hand out of all possible values of the data.

In this manner, Bayesian statistics offers an alternative statistical framework for evaluating hypotheses. Historically, the development of Bayesian statistics has been attributed to Thomas Bayes' discovery of the inner symmetry of conditional probabilities. Although other scholars, including Bernoulli and de Moivre, discussed this inner symmetry prior to Bayes, they are not credited with its discovery (Salsburg 2001). A conditional probability is the probability of an outcome given that another outcome has already occurred (Diez et al. 2017), e.g., the probability that a stone tool is a projectile point given that it is made out of obsidian. Bayes discovered that the inner symmetry of conditional probabilities allowed scholars to calculate the probability of the before based on the conditional after, in this case the probability that a stone tool is made of obsidian given that it is a projectile point.

Bayes himself never published his findings; however, in a posthumously published paper, Bayes (1763) described the theorems for conditional probability, as well as the selection of a prior probability distribution. Compared to NHST, the use of prior probabilities, or simply priors, is unique to Bayesian statistics. Priors allow for the explicit inclusion of previous knowledge or beliefs about the topic being studied (Buck et al. 1996, Cowgill 1993). The inclusion of prior information allows previous knowledge to interact with new data—something that is especially important in archaeological settings. As Buck and colleagues (1996) discuss, archaeologists interpret the discovery of new artifacts in conjunction with the artifacts that have already been discovered. This is a simple example of using prior information to inform current inferences.

Bayesian inference can be used to compute the probability of a hypothesis or a hypothesized parameter. First, a researcher must establish the prior information related to the hypothesis. This prior information is then transformed into a prior probability, typically in the form of a probability density function, to be used in Bayesian inference. Once the prior probability has been determined, new data are generated through observation or experiment to test the hypothesis. The likelihood of each datum is then computed, providing a measure of support for the data given the hypothesis. The prior then weights the likelihood. The result of the Bayesian analysis is the posterior probability, i.e., the probability of the hypothesis given the observed data (Buck et al. 1996, Salsburg 2001). It is important to note that Bayesian analyses can be iterative, meaning that the posterior probability of one analysis testing a hypothesis can become the prior probability of an additional analysis testing the truth of the same hypothesis in light of new data.

The primary advantage of Bayesian statistics over NHST statistics is the intelligibility of the inferences drawn from a Bayesian analysis. Furthermore, the use of prior probabilities offers practical advantages over NHST testing by formally including prior, expert information (Cowgill 2001). The inclusion of previously gained knowledge results in a more complete understanding of the relevance of the hypothesis. The use of prior probabilities also allows Bayesian inferences to be updated, producing a cyclical effect, as new knowledge becomes prior knowledge for even newer studies.

BAYESIAN INFERENCE IN ARCHAEOLOGY

Despite being developed in the eighteenth century, the scientific use of Bayesian statistics only increased significantly in the 1990s (Robert & Casella 2011). Archaeologists in particular remained generally unaware of Bayesian statistics (Cowgill 1993), with a few important exceptions (Buck et al. 1991, 1992; Cowgill 1989; Kadane & Hastorf 1987; Litton & Leese 1990). The earliest uses of Bayesian statistics by archaeologists focused on the calibration of radiocarbon dates and modeling chronology. Early Bayesian chronological models pervaded archaeological analyses, including lithic, ceramic, spatial, zooarchaeological, and bioarchaeological studies. It was not until the early twenty-first century that Bayesian modeling increased in its use to answer nonchronological questions.

Calibrating Radiocarbon Dates and Chronological Modeling

The most widespread use of Bayesian statistics in archaeology has been in radiocarbon dating archaeological sites. In this context, Bayesian inference has been called an ongoing "revolution" in radiocarbon dating (Bayliss 2009). Bayesian methods combine observed radiocarbon dates with calendar calibration curves and relative dating information such as stratigraphy to estimate the true probability distribution of the dates in question. The results are radiocarbon chronologies with refined precision. Buck and colleagues (1991, 1992) were early champions of Bayesian statistics as a method for calibrating radiocarbon dates. Later development and advancement of calibration software, specifically OxCal (Bronk Ramsey 1994, 2008) and BCal (Buck et al. 1999), made it easier for archaeologists to utilize Bayesian methods. The OxCal and BCal programs revolutionized the use of Bayesian statistics in archaeology by providing a simple way for archaeologists to calibrate radiocarbon dates and develop more complex chronological models.

Although the methodology and software for Bayesian chronological modeling was developed in the early 1990s, most of the studies utilizing these methods have been published within the last decade. The majority of Bayesian chronological modeling has been used to determine site chronologies, seriations, and typologies (Bayliss 2015). Early scholars used Bayesian chronological modeling to date sites in Europe, including Bavaria (Aitchison et al. 1991) and Stonehenge (Bayliss et al. 1997). The use of Bayesian chronologies in the archaeology of the United Kingdom has been prolific, owing to the development of Bayesian calibration methods by British scholars (e.g., Gearey et al. 2009, Griffiths 2014, Hamilton et al. 2015, Rhodes et al. 2003). Additional Bayesian chronologies have focused on other parts of Europe, including Scandinavia (Kammonen et al. 2012, Riede & Edinborough 2012, Wohlfarth et al. 2006) and southern Europe (Alberti 2013, Capuzzo & Barceló 2015, Lull et al. 2013).

In the Near East, Bayesian chronologies have been applied to analyses of funerary contexts in Egypt (Savage 1998), work enhancing Egyptian dynastic chronologies (Bronk Ramsey et al. 2010, Dee et al. 2013, Quiles et al. 2013), and analyses of the Levant (Finkelstein & Piasetzky 2015, Higham et al. 2005, Lee et al. 2013) and Anatolia (Bayliss et al. 2015). Within the last decade, Bayesian methods have spread to the rest of Asia and Oceania with studies of Siberia (Bronk Ramsey et al. 2014, Weber 2012), Southeast Asia (Higham & Higham 2009), China, and the Philippines (Acabado 2009, Long & Taylor 2015), and analyses of radiocarbon dates in Papua New Guinea (Petrie & Torrence 2008), Fiji (Burley & Edinborough 2014, Nunn & Petchey 2013), and Hawai'i (Athens et al. 2014).

North American applications of Bayesian chronologies have covered a range of cultures from the Pleistocene (Buck & Bard 2007, Kopperl et al. 2015) to the Holocene (Goring et al. 2012, Pluckhahn et al. 2015), including Mesoamerican cultures in Mexico and Central America (Beramendi-Orosco et al. 2009, Culleton et al. 2012). In addition, examples of Bayesian applications to chronology in South America include studies conducted in Ecuador (Zeidler et al. 1998) and Peru (Cadwallader et al. 2015, Unkel et al. 2007).

Few applications of Bayesian chronologies exist in the African continent. Examples include chronologies in South Africa (Bonneau et al. 2011, Millard 2006), followed by more recent applications in Zimbabwe (Chirikure et al. 2013) and Libya (Cherkinsky & Di Lernia 2013).

Bioarchaeology and Paleopathology

In bioarchaeological research, the earliest application of Bayesian statistics was to study chronology, as scholars analyzed radiocarbon data to date cemetery usage (Buck et al. 1992, Hey et al. 1999). Bayesian statistics was also used to improve paleodemographic methods for estimating age at death and stature. Konigsberg & Frankenberg (1992, 1994) developed the Rostock protocol, formalizing the use of prior knowledge in age-at-death estimations. This step was revolutionary. To infer age, paleodemographers measure indicators such as stature, dental development, and epiphyseal fusion. Conceptually, this inferential step requires researchers to assume that growth and development cause age. The Bayesian approach inverts this thinking to correspond with biological processes where "age causes growth and development" (Konigsberg & Frankenberg 1994, p. 96).

Since then, many studies have demonstrated the usefulness of Bayesian statistics in age-at-death estimates (Hoppa & Vaupel 2008, Konigsberg & Frankenberg 2002). Other Bayesian age-at-death estimates have been used by numerous scholars studying archaeological populations (e.g., Coqueugniot et al. 2010, Mays 2012, Nagaoka et al. 2008, Séguy et al. 2013). Similarly, a method using prior knowledge, Bayesian statistics, and maximum likelihood was proposed to estimate stature (Konigsberg et al. 1998). Furthermore, Bayesian methods have proven popular in reconstructing juvenile paleodemography, particularly the estimation of age using skeletal (Gowland & Chamberlain 2002, Nagaoka et al. 2012, Rissech et al. 2013, Tocheri et al. 2005) and dental remains (Heuzé & Cardoso 2008, Millard & Gowland 2002).

More recent applications of Bayesian statistics in bioarchaeology have involved the examination of ancient DNA to study evolutionary and dietary trends (Fu et al. 2013b). Specifically, Bayesian evolutionary analysis by sampling trees (BEAST) has been used to analyze mitochondrial DNA from archaeological populations. This approach constructs phylogenetic trees (Shapiro & Hofreiter 2012) and has been used to study pigmentation variation in modern Europeans (Wilde et al. 2014) and the extinction of Neanderthals during the Last Glacial Maximum (Dodge 2011). Other Bayesian methods have been used to create phylogenetic trees to analyze the relationship between modern and ancient humans in China (Fu et al. 2013a). Bayesian methods have been used to study gene flow and population migration throughout human evolution in Europe (Brotherton et al. 2013), the Americas (Ray et al. 2009), Australia (reviewed in van Holst Pellekaan 2013), and Asia (Mellars et al. 2013).

Additionally, some scholars have used Bayesian modeling to reconstruct ancient diets using the food reconstructing using isotopic transferred signals (FRUITS) approach. This Bayesian mixing model method has been used to study diet transitions in the Neolithic (Fernandes et al. 2015) and the consumption of marine resources as a dietary adaptation (Fernandes 2016, Tsutaya et al. 2014). Other Bayesian methods have been used to study maize reliance in the Americas (Coltrain & Janetski 2013) and weaning ages in archaeological samples.

Zooarchaeology

Bayesian statistics is becoming critical in zooarchaeological research. For example, analyses of bone surface modifications (BSM), particularly cut marks, are paramount trace evidence when identifying early instances of stone tool–aided butchery in human evolution. Traditional identification of the causal agent behind BSM is usually based on expert judgment resulting from morphological observations at the micro and macro scale. This approach, however, has not been successful at discriminating between ambiguous marks. Consequently, new protocols that include Bayesian inference are beginning to develop (e.g., Harris et al. 2017, Otárola-Castillo et al. 2018). These studies allowed researchers to correctly match bone marks to causal agents, using probabilities derived from the weight of several types of quantitative and qualitative evidence. Similar protocols are also being developed to probabilistically differentiate between sheep and goat bones during the identification process (Wolfhagen & Price 2017).

More generally, Bayesian inference in zooarchaeology has been focused on several research questions related to (*a*) the exploitation of resources over time and (*b*) the genetic history of animal species. To study resource exploitation, scholars have used Bayesian methods and radiocarbon calibration to calculate estimates of lapsed time intervals in order to determine how long groups practiced specific behaviors, e.g., hunting and scavenging of mastodons by Paleoindians (Fisher 1987), seal hunting in Siberia (Nomokonova et al. 2015), whale hunting in Canada (Béland et al. 2018), and mollusk foraging in southern England (Mannino & Thomas 2001). Similar studies have examined the effects of human activities on species, including the Grevy's Zebra of East Africa (Faith et al. 2013).

Scholars have used Bayesian inference to construct genetic phylogenies and examine genetic mutations in animals in order to better understand animal domestication and the exploitation of animal resources, e.g., the domestication of pigs in Europe (Larson et al. 2007) and the Levant (Meiri et al. 2013), swamp buffalo in China (Yue et al. 2013), and chickens on Norfolk Island (Kraitsek et al. 2013). Other analyses of genetic diversity have traced the spread of domesticated species. Sacks and colleagues (2013) conducted a study of chromosomal mutations, revealing that the domestication of dogs likely occurred in Southeast Asia before expanding westward during the Neolithic. Similar studies have provided information related to the domestication of cattle (Horsburgh et al. 2013, Speller et al. 2013) and river buffalo (Nagarajan et al. 2015).

The analysis of genetic variation in steppe bison, horses, and elk has helped to illuminate trends in the exploitation of animal resources. In reconstructing the genetic history of bison in Beringia, Shapiro and colleagues (2004) found that the observed decline in bison genetic diversity correlated with changes in the environment rather than increases in human populations. Warmuth and colleagues (2013) utilized Bayesian cluster analyses to examine the genetic variation of nonbreed horses in order to track trends in domestication and trade throughout Eurasia. Meiri and colleagues (2014) used a variety of Bayesian methods to study the migration of elk from Siberia to the Bering Isthmus, which paralleled the migration of people of the Americas.

Ceramics

As in other subdisciplines, the earliest application of Bayesian statistics to the study of ceramics was for the analysis of phase chronology as pioneered by Naylor & Smith (1988). This study applied a Bayesian framework to refine the chronology of Danebury Fort, an Iron Age structure in Hampshire, England. Notably, the researchers incorporated qualitative expert knowledge of ceramic design into the Bayesian model to increase chronological precision. This approach demonstrated the flexibility of Bayesian inference to include the prior knowledge of expert archaeologists. Other studies soon followed; for example, the use of Bayesian inference allowed Buck (1993) to formally combine the knowledge of ceramic experts with geochemical analyses in order to identify the source material. Similarly, Papageorgiou & Liritzis (2007) demonstrated the usefulness of Bayesian methods in provenance studies based on the geochemical data from the Aegean Islands.

Bayesian models combined with pottery and ceramic artifacts have informed chronological debates regarding the Iron Age of the Levant (Levy et al. 2006, Smith & Levy 2008). Finkelstein & Piasetzky (2010) used Bayesian modeling to produce a ceramic chronology for the Iron Age Levant based solely on radiocarbon dates. Their model identified six ceramic phases and six transitions, informing a reinterpretation of history in the Levant. Similar Bayesian models have been used to assess the chronology and duration of Famabalasto Black Incised pottery in Pre-Hispanic Argentina (Greco & Palamarczuk 2014), the seriation of ceramics in Fennoscandia (Pesonen et al. 2012), and the dating of celadon porcelain in China (Xie et al. 2009).

Additionally, Bayesian methodologies have been used in the reconstruction of ceramic artifacts. For example, Cooper and colleagues (Cooper et al. 2001, Willis & Cooper 2004) championed the utilization of Bayesian modeling to reconstruct broken pots from archaeological contexts. Their model included three-dimensional geometric measurements of sherds to virtually reconstruct broken artifacts. However, since 2010, there has been very little use of Bayesian models in the three-dimensional reconstruction of ceramics (Rasheed & Nordin 2015).

Lithics

Bayesian statistics can be useful when studying lithics because the use of prior knowledge allows scholars to account for small or nonexistent samples when studying numerous sites. For example, Borradaile (2003) considered prior knowledge to study the remagnetization of a small sample of chalk monuments in England and Israel. Additionally, Arakawa and colleagues (2013) used Bayesian methods to examine changes in subsistence behavior and population size in the American Southwest by comparing the ratio of projectile points to ceramic sherds. However, not all sites in their sample contained projectile points. Restricting the study to the sample of sites with projectile points would have underpowered the analysis. Bayesian methods allowed Arakawa and colleagues to include all sites by estimating model parameters from prior knowledge for sites lacking associated lithic artifacts. This approach permitted a more nuanced inference of the effects of social processes on prehistoric subsistence strategies.

Scholars have often used Bayesian methods to research temporal trends in technology, including the construction of hearths in Australia (Holdaway et al. 2004), the evolution of bow and arrow technology in Scandinavia (Edinborough 2005a,b), and the use of source materials in Hawai'i (Rieth et al. 2013). Bayesian methods have also been used to date chronologically ambiguous collections. For example, Fernández-López de Pablo & Barton (2015) used Bayesian methods to date collections of surface artifacts. They determined the probability of occupation by considering the prior knowledge of previously dated lithic collections compared to the artifacts that were found.

Spatial Archaeology and Geographic Information Systems

Bayesian inference in spatial analyses are relatively underutilized. The primary use of Bayesian statistics in a spatial context has been to combine prior knowledge with geographical data (Kirkinen 1999). For example, Ortman et al. (2007) utilized Bayesian inference to formally quantify prior archaeological knowledge from the Colorado cultural resource database when reconstructing Pueblo settlement patterns. By synthesizing a large amount of relevant data, the researchers were able to estimate settlement data from unexcavated sites. This method is a vast improvement over previous ad hoc methods of combining data because it transparently quantifies the bases of their interpretation. Spatial Bayesian studies generally focus on predictive modeling (Canning 2003, Ford et al. 2009, Hill et al. 2006). More recent research has combined the use of Bayesian methods with data mining techniques to identify useful knowledge in geographic information system databases (Arias 2013). If the positive trend of Bayesian inference continues, the future of spatial analysis will soon include more studies under this framework.

THE FUTURE OF BAYESIAN STATISTICS IN ARCHAEOLOGY

Statistics allow archaeologists to evaluate their theories by comparing them to the empirical world. Just as the applications of NHST in archaeology have increased over time, so should the applications of Bayesian inference. Bayesian methods provide simpler interpretations of the empirical support for one's hypotheses than does NHST, a way to incorporate prior knowledge into one's inference, and a mechanism for updating one's conclusions in light of new evidence. These are highly appealing factors that will undoubtedly promote the application of Bayesian methods to archaeological science. There is a trade-off, however. Bayesian methods are a mathematically and computationally intensive set of tools, more so than NHST. To achieve healthy growth in the application of these methods to archaeology we need (*a*) to ensure an equally healthy growth in our knowledge of mathematics, computation, and their appropriate use; (*b*) to continue expanding the diversity of our collaborations to include archaeological data scientists and archaeological statisticians grounded in both NHST and Bayesian inference, who can guide use; and (*c*) to develop archaeo-statistics as a field. This latter point begins by encouraging students to become not only aware of but also fluent in the mathematics underlying NHST and Bayesian inference. In addition, adaptations to current curricula might be necessary to educate expert archaeo-statisticians who can focus on statistical solutions for the rest of us to use with confidence. A favorable indicator that these changes are taking place within archaeology is the current institutional support. The Society for American Archaeology (SAA), for example, has facilitated the development of interest groups concerned with methodological and data infrastructure development (e.g., the Digital Data and Open Science Interest Groups). In particular, the nascent Quantitative Methods and Statistical Computing in Archaeology (QUANTARCH) Interest Group was formed to develop a community for statistical and quantitative-oriented archaeologists. These developments show a positive outlook for archaeology to continue developing and adopting appropriate statistical methods to empirically evaluate theory with data. This step is critical to ensure the continued success of archaeology as a scientific endeavor to make inferences about the past.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ACKNOWLEDGMENTS

We dedicate this article to the memory of Dr. George L. Cowgill, a pioneering advocate of Bayesian inference in archaeology. We thank the graduate and undergraduate members of the Laboratory for Computational Anthropology and Anthroinformatics (LCA) and the Department of Anthropology at Purdue University. Contributions by M.G.T. were financially supported in part by the National Science Foundation Graduate Research Fellowship Program (Grant No. 106469). E.O.C. acknowledges financial support from the Margo Katherine Wilke Undergraduate Research fund and the College of Liberal Arts' ASPIRE fund at Purdue University. The authors thank the production editor and anonymous reviewers for their time and helpful comments. This manuscript has certainly been improved through the peer-review process. R-code to conduct NHST and Bayesian analysis is available in the **Supplemental Material** and upon request.

Supplemental Material >

LITERATURE CITED

Acabado S. 2009. A Bayesian approach to dating agricultural terraces: a case from the Philippines. *Antiquity* 83:801–14

Aitchison T, Ottaway B, Al-Ruzaiza AS. 1991. Summarizing a group of ^{14}C dates on the historical time scale: with a worked example from the Late Neolithic of Bavaria. *Antiquity* 65:108–16

Alberti G. 2013. A Bayesian ^{14}C chronology of Early and Middle Bronze Age in Sicily. Towards an independent absolute dating. *J. Archaeol. Sci.* 40:2502–14

Arakawa F, Nicholson C, Rasic J. 2013. The consequences of social processes: aggregate populations, projectile point accumulation, and subsistence patterns in the American Southwest. *Am. Antiq.* 78:147–65

Arias V. 2013. *Application of GIS and spatial data modeling to archaeology: a case study in the American Southwest*. PhD Diss., Dep. Anthropol., Univ. N.M., Albuquerque

Athens JS, Rieth TM, Dye TS. 2014. A paleoenvironmental and archaeological model-based age estimate for the colonization of Hawai'i. *Am. Antiq.* 79:144–55

Bayes T. 1763. An essay towards solving a problem in the doctrine of chances. *Philos. Trans.* 53:370–418

Bayliss A. 2009. Rolling out revolution: using radiocarbon dating in archaeology. *Radiocarbon* 51:123–47

Bayliss A. 2015. Quality in Bayesian chronological models in archaeology. *World Archaeol*. 47:677–700

Bayliss A, Brock F, Farid S, Hodder I, Southon J, Taylor RE. 2015. Getting to the bottom of it all: a Bayesian approach to dating the start of Çatalhöyük. *J. World Prehist.* 28:1–26

Bayliss A, Ramsey CB, McCormac FG. 1997. Dating Stonehenge. *Proc. Br. Acad.* 92:39–59

Béland SL, McLeod BA, Martin J, Martin GM, Darling JD, Frasier TR. 2018. Species composition of First Nation whaling hunts in the Clayoquot Sound region of Vancouver Island as estimated through genetic analyses. *J. Archaeol. Sci. Rep.* 17:235–41

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, et al. 2018. Redefine statistical significance. *Nat. Hum. Behav.* 2:6–10

Beramendi-Orosco LE, Gonzalez-Hernandez G, Urrutia-Fucugauchi J, Manzanilla LR, Soler-Arechalde AM, et al. 2009. High-resolution chronology for the Mesoamerican urban center of Teotihuacan derived from Bayesian statistics of radiocarbon and archaeological data. *Quat. Res.* 71:99–107

Binford LR. 1964. A consideration of archaeological research design. *Am. Antiq.* 29:425–41

Bonneau A, Brock F, Higham T, Pearce DG, Pollard AM. 2011. An improved pretreatment protocol for radiocarbon dating black pigments in San rock art. *Radiocarbon* 53(3):419–28

Borradaile GJ. 2003. Viscous magnetization, archaeology and Bayesian statistics of small samples from Israel and England. *Geophys. Res. Lett.* 30:35

Bronk Ramsey C. 1994. Analysis of chronological information and radiocarbon calibration: the program OxCal. *Archaeol. Comput. Newsl.* 41:e16

Bronk Ramsey C. 2008. Radiocarbon dating: revolutions in understanding. *Archaeometry* 50:249–75

Bronk Ramsey C, Dee MW, Rowland JM, Higham TF, Harris SA, et al. 2010. Radiocarbon-based chronology for dynastic Egypt. *Science* 328:1554–57

Bronk Ramsey C, Schulting R, Goriunova OI, Bazaliiskii VI, Weber AW. 2014. Analyzing radiocarbon reservoir offsets through stable nitrogen isotopes and Bayesian modeling: a case study using paired human and faunal remains from the Cis-Baikal region, Siberia. *Radiocarbon* 56:789–99

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* 4:1764

Buck C. 1993. The provenancing of archaeological ceramics: a Bayesian approach. In *Computing the Past: Computer Applications and Quantitative Methods in Archaeology*, ed. J Andersen, T Madsen, I Scollar, pp. 293–301. Aarhus, Den.: Aarhus Univ. Press

Buck CE, Bard E. 2007. A calendar chronology for Pleistocene mammoth and horse extinction in North America based on Bayesian radiocarbon calibration. *Quat. Sci. Rev.* 26:2031–35

Buck CE, Cavanagh WG, Litton CD. 1996. *Bayesian Approach to Interpreting Archaeological Data*. West Sussex, UK: Wiley

Buck CE, Christen JA, James GN. 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeol*. 7. **https://doi.org/10.11141/ia.7.1**

Buck CE, Kenworthy JB, Litton CD, Smith AFM. 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65:808–21

Buck CE, Litton CD, Smith AF. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *J. Archaeol. Sci.* 19:497–512

Burley DV, Edinborough K. 2014. Discontinuity in the Fijian archaeological record supported by a Bayesian radiocarbon model. *Radiocarbon* 56:295–303

Cadwallader L, Torres SA, O'Connell TC, Pullen AG, Beresford-Jones DG. 2015. Dating the dead: new radiocarbon dates from the Lower Ica Valley, south coast Peru. *Radiocarbon* 57:765–73

Canning S. 2003. *Site unseen: archaeology, cultural resource management, planning and predictive modelling in the Melbourne metropolitan area*. PhD Thesis, Sch. Eur. Hist. Stud., La Trobe Univ., Bundoora

Capuzzo G, Barceló JA. 2015. Cultural changes in the second millennium BC: a Bayesian examination of radiocarbon evidence from Switzerland and Catalonia. *World Archaeol*. 47:622–41

Carlson DL. 2017. *Quantitative Methods in Archaeology Using R*. Cambridge, UK/New York: Cambridge Univ. Press

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, et al. 2017. Stan: a probabilistic programming language. *J. Stat. Softw.* 76(1). **https://doi.org/10.18637/jss.v076.i01**

Cherkinsky A, Di Lernia S. 2013. Bayesian approach to 14C dates for estimation of long-term archaeological sequences in arid environments: the Holocene site of Takarkori Rockshelter, Southwest Libya. *Radiocarbon* 55:771–82

Chirikure S, Pollard M, Manyanga M, Bandama F. 2013. A Bayesian chronology for Great Zimbabwe: re-threading the sequence of a vandalised monument. *Antiquity* 87:854–72

Clarke DL. 1968. *Analytical Archaeology*. London: Methuen

Coltrain JB, Janetski JC. 2013. The stable and radio-isotope chemistry of southeastern Utah Basketmaker II burials: dietary analysis using the linear mixing model SISUS, age and sex patterning, geolocation and temporal patterning. *J. Archaeol. Sci.* 40:4711–30

Cooper DB, Willis A, Andrews S, Baker J, Cao Y, et al. 2001. Assembling virtual pots from 3D measurements of their fragments. In *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage*, ed. D Arnold, A Chalmers, D Fellner, pp. 241–54. New York: ACM

Coqueugniot H, Weaver TD, Houët F. 2010. Brief communication: a probabilistic approach to age estimation from infracranial sequences of maturation. *Am. J. Phys. Anthropol.* 142:655–64

Cowgill GL. 1989. Formal approaches in archaeology. In *Archaeological Thought in America*, ed. CC Lamberg-Karlovsky, pp. 74–88. Cambridge, UK: Cambridge Univ. Press

Cowgill GL. 1993. Distinguished lecture in archeology: beyond criticizing New Archeology. *Am. Anthropol.* 95:551–73

Cowgill GL. 2001. Past, present, and future of quantitative methods in United States archaeology. In *Computing Archaeology for Understanding the Past. CAA 2000. Computer Applications and Quantitative Methods in Archaeology*, ed. Z Stančič, T Veljanovski, pp. 35–40. Oxford, UK: Archaeopress

Cowgill GL. 2015. Some things I hope you will find useful even if statistics isn't your thing. *Annu. Rev. Anthropol.* 44:1–14

Culleton BJ, Prufer KM, Kennett DJ. 2012. A Bayesian AMS ^{14}C chronology of the classic Maya center of Uxbenká, Belize. *J. Archaeol. Sci.* 39:1572–86

Dee M, Wengrow D, Shortland A, Stevenson A, Brock F, et al. 2013. An absolute chronology for early Egypt using radiocarbon dating and Bayesian statistical modelling. *Proc. R. Soc. A* 469:20131395

Diez DM, Barr CD, Çetinkaya-Rundel M. 2017. *OpenIntro Statistics*. North Charleston, SC: CreateSpace

Dodge DR. 2011. *A molecular approach to the Neanderthal extinction: the applicability of a Bayesian programme to the analysis of Neanderthals*. PhD Thesis, Dep. Archaeol., Univ. Sheffield

Edinborough K. 2005a. Weapons of maths instruction: a thousand years of technological stasis in arrowheads from the South Scandinavian Middle Mesolithic. *Pap. Inst. Archaeol.* 16:50–58

Edinborough KSA. 2005b. *Evolution of bow-arrow technology*. PhD Thesis, Univ. Coll. London

Faith JT, Tryon CA, Peppe DJ, Fox DL. 2013. The fossil history of Grévy's zebra (*Equus grevyi*) in equatorial East Africa. *J. Biogeogr.* 40:359–69

Fernandes R. 2016. A simple(R) model to predict the source of dietary carbon in individual consumers. *Archaeometry* 58:500–12

Fernandes R, Grootes P, Nadeau M-J, Nehlich O. 2015. Quantitative diet reconstruction of a Neolithic population using a Bayesian mixing model (FRUITS): the case study of Ostorf (Germany). *Am. J. Phys. Anthropol.* 158:325–40

Fernández-López de Pablo J, Barton CM. 2015. Bayesian estimation dating of lithic surface collections. *J. Archaeol. Method Theory* 22(2):559–83

Finkelstein I, Piasetzky E. 2010. Radiocarbon dating the Iron Age in the Levant: a Bayesian model for six ceramic phases and six transitions. *Antiquity* 84:374–85

Finkelstein I, Piasetzky E. 2015. Radiocarbon dating Khirbet Qeiyafa and the Iron I–IIA phases in the Shephelah: methodological comments and a Bayesian model. *Radiocarbon* 57:891–907

Fisher DC. 1987. Mastodont procurement by Paleoindians of the Great Lakes region: hunting or scavenging. *Evol. Hum. Hunting*: 309–421

Fisher RA. 1925. *Statistical Methods for Research Workers*. Edinburgh/London: Oliver and Boyd

Fisher RA. 1929. The statistical method in psychical research. *Proc. Soc. Psychical Res.* 39:189–92

Fisher RA. 1935. Letter to the Editor: statistical tests. *Nature* 136:474

Fletcher M, Lock GR. 2005. *Digging Numbers: Elementary Statistics for Archaeologists*. Oxford, UK: Oxford Press

Ford A, Clarke KC, Raines G. 2009. Modeling settlement patterns of the late classic Maya civilization with Bayesian methods and geographic information systems. *Ann. Assoc. Am. Geogr.* 99:496–520

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, et al. 2013a. DNA analysis of an early modern human from Tianyuan Cave, China. *PNAS* 110:2223–27

Fu Q, Mittnik A, Johnson PL, Bos K, Lari M, et al. 2013b. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23:553–59

Gearey BR, Marshall P, Hamilton D. 2009. Correlating archaeological and palaeoenvironmental records using a Bayesian approach: a case study from Sutton Common, South Yorkshire, England. *J. Archaeol. Sci.* 36:1477–87

Gelman A. 2017. The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Pers. Soc. Psychol. Bull.* 44:16–23

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014. *Bayesian Data Analysis*. Boca Raton: CRC Press

Gill J. 1999. The insignificance of null hypothesis significance testing. *Political Res. Q.* 52:647–74

Goring S, Williams J, Blois JL, Jackson ST, Paciorek CJ, et al. 2012. Deposition times in the northeastern United States during the Holocene: establishing valid priors for Bayesian age models. *Quat. Sci. Rev.* 48:54–60

Gowland RL, Chamberlain AT. 2002. A Bayesian approach to ageing perinatal skeletal material from archaeological sites: implications for the evidence for infanticide in Roman-Britain. *J. Archaeol. Sci.* 29:677–85

Greco C, Palamarczuk V. 2014. Strategy for radiocarbon chronological assessment of ceramic styles: an example from Prehispanic northwestern Argentina. *Radiocarbon* 56:1093–106

Griffiths S. 2014. A Bayesian radiocarbon chronology of the early Neolithic of Yorkshire and Humberside. *Archaeol. J.* 171:2–29

Hamilton WD, Haselgrove C, Gosden C. 2015. The impact of Bayesian chronologies on the British Iron Age. *World Archaeol*. 47:642–60

Harris JA, Marean CW, Ogle K, Thompson J. 2017. The trajectory of bone surface modification studies in paleoanthropology and a new Bayesian solution to the identification controversy. *J. Hum. Evol.* 110:69–81

Heuzé Y, Cardoso HF. 2008. Testing the quality of nonadult Bayesian dental age assessment methods to juvenile skeletal remains: the Lisbon collection children and secular trend effects. *Am. J. Phys. Anthropol.* 135:275–83

Hey G, Bayliss A, Boyle A. 1999. Iron Age inhumation burials at Yarnton, Oxfordshire. *Antiquity* 73:551–62

Higham C, Higham T. 2009. A new chronological framework for prehistoric Southeast Asia, based on a Bayesian model from Ban Non Wat. *Antiquity* 83:125–44

Higham T, van der Plicht J, Bronk Ramsey C, Bruins HJ, Robinson M, Levy TE. 2005. Radiocarbon dating of the Khirbat-en Nahas site (Jordan) and Bayesian modeling of the results. In *The Bible and Radiocarbon Dating: Archaeology, Text and Science*, ed. T Levy, T Higham, pp. 164–78. New York: Routledge

Hill JB, Devitt M, Sergeyeva M. 2006. *Predictive modeling and cultural resource preservation in Santa Cruz county, Arizona*. Work. Pap., Cent. Desert Archaeol., Tucson. **https://www.archaeologysouthwest.org/pdf/Hilletal_scnha_pred_mod.pdf**

Holdaway S, Shiner J, Fanning P. 2004. Hunter-gatherers and the archaeology of discard behavior: an analysis of surface stone artifacts from Sturt National Park, western New South Wales, Australia. *Asian Perspect.* 43:34–72

Hoppa RD, Vaupel JW. 2008. The Rostock Manifesto for paleodemography: the way from stage to age. In *Paleodemography: Age Distributions from Skeletal Samples*, ed. RD Hoppa, JW Vaupel, pp. 1–8. Cambridge, UK: Cambridge Univ. Press

Horsburgh KA, Prost S, Gosling A, Stanton J-A, Rand C, Matisoo-Smith EA. 2013. The genetic diversity of the Nguni breed of African Cattle (*Bos* spp.): complete mitochondrial genomes of haplogroup T1. *PLOS ONE* 8:e71956

Kadane JB, Hastorf CA. 1987. *Bayesian paleoethnobotany*. Pap., Dep. Stat., Carnegie-Mellon Univ.

Kammonen J, Sundell T, Moltchanova E, Pesonen P, Oinonen M, et al. 2012. Bayesian spatial analysis of archaeological finds and radiocarbon dates: an example from Finland 4000–3500 cal BC. In *Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 39th International Conference, Beijing, April 12–16*, ed. M Zhou, I Romanowska, Z Wu, P Xu, P Verhagen, pp. 318–25. Amsterdam: Pallas

Kirkinen T. 1999. GIS-assisted data analysis-finding meanings in complex spatial data sets. In *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*, ed. JA Barceló, I Briz, A Vila, pp. 255–58. Oxford, UK: Archaeopress

Konigsberg LW, Frankenberg SR. 1992. Estimation of age structure in anthropological demography. *Am. J. Phys. Anthropol.* 89:235–56

Konigsberg LW, Frankenberg SR. 1994. Paleodemography: "Not quite dead." *Evol. Anthropol.* 3:92–105

Konigsberg LW, Frankenberg SR. 2002. Deconstructing death in paleodemography. *Am. J. Phys. Anthropol.* 117:297–309

Konigsberg LW, Hens SM, Jantz LM, Jungers WL. 1998. Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. *Am. J. Phys. Anthropol.* 107:65–92

Kopperl RE, Taylor AK, Miss CJ, Ames KM, Hodges CM. 2015. The Bear Creek Site (45KI839), a Late Pleistocene–Holocene transition occupation in the Puget Sound Lowland, King County, Washington. *PaleoAmerica* 1:116–20

Kraitsek S, Baskerville B, Ho SY, Gongora J. 2013. Australian and Pacific contributions to the genetic diversity of Norfolk Island feral chickens. *BMC Genet*. 14:91

Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, et al. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *PNAS* 104:15276–81

Lee S, Bronk Ramsey C, Mazar A. 2013. Iron Age chronology in Israel: results from modeling with a trapezoidal Bayesian framework. *Radiocarbon* 55:731–40

Lehmann EL. 1959. *Testing Statistical Hypotheses*. New York: Wiley

Levy TE, Higham T, Najjar M. 2006. Response to van der Steen & Bienkowski. *Antiquity* 80:3–5

Litton C, Leese M. 1990. Some statistical problems arising in radiocarbon calibration. *Comput. Appl. Quant. Methods Archaeol.* 565:101–9

Long T, Taylor D. 2015. A revised chronology for the archaeology of the lower Yangtze, China, based on Bayesian statistical modelling. *J. Archaeol. Sci.* 63:115–21

Lull V, Micó R, Rihuete-Herrada C, Risch R. 2013. Funerary practices and kinship in an Early Bronze Age society: a Bayesian approach applied to the radiocarbon dating of Argaric double tombs. *J. Archaeol. Sci.* 40:4626–34

Mannino MA, Thomas KD. 2001. Intensive Mesolithic exploitation of coastal resources? Evidence from a shell deposit on the Isle of Portland (Southern England) for the impact of human foraging on populations of intertidal rocky shore molluscs. *J. Archaeol. Sci.* 28:1101–14

Mays S. 2012. An investigation of age-related changes at the acetabulum in 18th–19th century AD adult skeletons from Christ Church Spitalfields, London. *Am. J. Phys. Anthropol.* 149:485–92

McElreath R. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: CRC Press/Taylor & Francis

McShane BB, Gal D. 2017. Statistical significance and the dichotomization of evidence. *J. Am. Stat. Assoc.* 112:885–95

Meiri M, Huchon D, Bar-Oz G, Boaretto E, Horwitz LK, et al. 2013. Ancient DNA and population turnover in southern Levantine pigs—signature of the sea peoples migration? *Sci. Rep.* 3:3035

Meiri M, Lister AM, Collins MJ, Tuross N, Goebel T, et al. 2014. Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proc. R. Soc. B* 281:20132167

Mellars P, Gori KC, Carr M, Soares PA, Richards MB. 2013. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *PNAS* 110:10699–704

Millard AR. 2006. Bayesian analysis of ESR dates, with application to Border Cave. *Quat. Geochronol.* 1(2):159–66

Millard A, Gowland R. 2002. A Bayesian approach to the estimation of the age of humans from tooth development and wear. *Archeol. Calcolatori* 13:197–210

Myers OH. 1950. *Some Applications of Statistics to Archaeology*. Cairo: Serv. Antiq. Egypte

Nagaoka T, Abe M, Shimatani K. 2012. Estimation of mortality profiles from non-adult human skeletons in Edo-period Japan. *Anthropol. Sci.* 120:115–28

Nagaoka T, Sawada J, Hirata K. 2008. Did the Jomon people have a short lifespan? Evidence from the adult age-at-death estimation based on the auricular surface of the ilium. *Anthropol. Sci.* 116:161–69

Nagarajan M, Nimisha K, Kumar S. 2015. Mitochondrial DNA variability of domestic river buffalo (*Bubalus bubalis*) populations: genetic evidence for domestication of river buffalo in Indian subcontinent. *Genome Biol. Evol.* 7:1252–59

Naylor JC, Smith AFM. 1988. An archaeological inference problem. *J. Am. Stat. Assoc.* 83:588–95

Neyman J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. R. Stat. Soc.* 97:558–625

Neyman J. 1937. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. A* 236:333–80

Neyman J, Pearson ES. 1933. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. A* 231:289–337

Nomokonova T, Losey RJ, Ol'ga IG, Novikov AG, Weber AW. 2015. A 9,000 year history of seal hunting on Lake Baikal, Siberia: the zooarchaeology of Sagan-Zaba II. *PLOS ONE* 10:e0128314

Nunn PD, Petchey F. 2013. Bayesian re-evaluation of Lapita settlement in Fiji: radiocarbon analysis of the Lapita occupation at Bourewa and nearby sites on the Rove Peninsula, Viti Levu Island. *J. Pac. Archaeol.* 4:21–34

Ortman SG, Varien MD, Gripp TL. 2007. Empirical Bayesian methods for archaeological survey data: an application from the Mesa Verde region. *Am. Antiq.* 72(2):241–72

Otárola-Castillo E, Torquato MG, Hawkins HC, James E, Harris JA, et al. 2018. Differentiating between cutting actions on bone using 3D geometric morphometrics and Bayesian analyses with implications to human evolution. *J. Archaeol. Sci.* 89:56–67

Papageorgiou I, Liritzis I. 2007. Multivariate mixture of normals with unknown number of components: an application to cluster Neolithic ceramics from Aegean and Asia Minor using portable XRF. *Archaeometry* 49:795–813

Pearson K. 1922. On the χ^2 test of goodness of fit. *Biometrika* 14:186–91

Pesonen P, Oinonen M, Carpelan C, Onkamo P. 2012. Early Subneolithic ceramic sequences in eastern Fennoscandia—a Bayesian approach. *Radiocarbon* 54:661–76

Petrie CA, Torrence R. 2008. Assessing the effects of volcanic disasters on human settlement in the Willaumez Peninsula, Papua New Guinea: a Bayesian approach to radiocarbon calibration. *Holocene* 18:729–44

Pluckhahn TJ, Thompson VD, Cherkinsky A. 2015. The temporality of shell-bearing landscapes at Crystal River, Florida. *J. Anthropol. Archaeol.* 37:19–36

Quiles A, Aubourg E, Berthier B, Delqué-Količ E, Pierrat-Bonnefois G, et al. 2013. Bayesian modelling of an absolute chronology for Egypt's 18th Dynasty by astrophysical and radiocarbon methods. *J. Archaeol. Sci.* 40:423–32

R Core Team. 2018. R: a language and environment for statistical computing. *R Found. Stat. Comput.* **https://www.R-project.org/**

Rasheed NA, Nordin MJ. 2015. A survey of computer methods in reconstruction of 3D archaeological pottery objects. *Int. J. Adv. Res.* 3:712–14

Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, Excoffier L. 2009. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol. Biol. Evol.* 27:337–45

Rhodes E, Bronk Ramsey C, Outram Z, Batt C, Willis L, et al. 2003. Bayesian methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old Scatness Broch excavations, Shetland Isles. *Quat. Sci. Rev.* 22:1231–44

Riede F, Edinborough K. 2012. Bayesian radiocarbon models for the cultural transition during the Allerød in southern Scandinavia. *J. Archaeol. Sci.* 39:744–56

Rieth TM, Mills PR, Lundblad SP, Morrison AE, Johnson A. 2013. Variation in lithic sources utilized by late pre-contact elites in Kona, Hawai'i Island. *Hawaiian Archaeol.* 13:103–30

Rissech C, López-Costas O, Turbón D. 2013. Humeral development from neonatal period to skeletal maturity—application in age and sex assessment. *Int. J. Legal Med.* 127:201–12

Robert C, Casella G. 2011. A short history of Markov Chain Monte Carlo: subjective recollections from incomplete data. 26:102–15

Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O. 2013. Y chromosome analysis of dingoes and Southeast Asian village dogs suggests a Neolithic continental expansion from Southeast Asia followed by multiple Austronesian dispersals. *Mol. Biol. Evol.* 30:1103–18

Salsburg D. 2001. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. New York: Macmillan

Savage SH. 1998. AMS radiocarbon dates from the Predynastic Egyptian cemetery, N7000, at Naga-ed-Dêr. *J. Archaeol. Sci.* 25:235–49

Séguy I, Caussinus H, Courgeau D, Buchet L. 2013. Estimating the age structure of a buried adult population: a new statistical approach applied to archaeological digs in France. *Am. J. Phys. Anthropol.* 150:170–83

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, et al. 2004. Rise and fall of the Beringian steppe bison. *Science* 306:1561–65

Shapiro B, Hofreiter M, ed. 2012. *Ancient DNA Methods and Protocols*. New York: Humana

Smith NG, Levy TE. 2008. The Iron Age pottery from Khirbat en-Nahas, Jordan: a preliminary study. *Bull. Am. Schools Orient. Res.* 2008(352):41–91

Spaulding AC. 1953. Statistical techniques for the discovery of artifact types. *Am. Antiq.* 28:305–313

Speller CF, Burley DV, Woodward RP, Yang DY. 2013. Ancient mtDNA analysis of early 16th century Caribbean cattle provides insight into founding populations of New World creole cattle breeds. *PLOS ONE* 8:e69584

Stan Dev. Team. 2018. RStan: the R interface to Stan. *R package version 2.17.3*. **http://mc-stan.org/**

Tocheri MW, Dupras TL, Sheldrick P, Molto JE. 2005. Roman period fetal skeletons from the east cemetery (Kellis 2) of Kellis, Egypt. *Int. J. Osteoarchaeol.* 15:326–41

Tsutaya T, Naito YI, Ishida H, Yoneda M. 2014. Carbon and nitrogen isotope analyses of human and dog diet in the Okhotsk culture: perspectives from the Moyoro site, Japan. *Anthropol. Sci.* 122:89–99

Unkel I, Kromer B, Reindel M, Wacker L, Wagner G. 2007. A chronology of the pre-Columbian Paracas and Nasca cultures in South Peru Based on AMS 14C dating. *Radiocarbon* 49:551–64

van Holst Pellekaan S. 2013. Genetic evidence for the colonization of Australia. *Quat. Int.* 285:44–56

Warmuth V, Manica A, Eriksson A, Barker G, Bower M. 2013. Autosomal genetic diversity in non-breed horses from eastern Eurasia provides insights into historical population movements. *Anim. Genet.* 44:53–61

Weber A. 2012. Patterns of cemetery use at Kurma XI: Bayesian approach to the examination of radiocarbon dates. In *Kurma XI, A Middle Holocene Hunter-Gatherer Cemetery on Lake Baikal, Siberia: Archaeological and Osteological Materials*, Vol. 6, ed. MA Weber, HG McKenzie, AR Lieverse, OI Goriunova, pp. 141–72. Edmonton: Univ. Alberta, Can. Circumpolar Inst. Press

Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, et al. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *PNAS* 111:4832–37

Willis AR, Cooper DB. 2004. Bayesian assembly of 3d axially symmetric shapes from fragments. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. New York: IEEE

Wohlfarth B, Blaauw M, Davies SM, Andersson M, Wastegård S, et al. 2006. Constraining the age of Lateglacial and early Holocene pollen zones and tephra horizons in southern Sweden with Bayesian probability methods. *J. Quat. Sci.* 21:321–34

Wolfhagen J, Price MD. 2017. A probabilistic model for distinguishing between sheep and goat postcranial remains. *J. Archaeol. Sci. Rep.* 12:625–31

Xie G, Feng S, Feng X, Zhu J, Yan L, Li L. 2009. Dating ancient Chinese celadon porcelain by neutron activation analysis and Bayesian classification. *Atomic Energy Sci. Technol.* 43:561–65

Yue X-P, Li R, Xie W-M, Xu P, Chang T-C, et al. 2013. Phylogeography and domestication of Chinese swamp buffalo. *PLOS ONE* 8:e56552

Zeidler JA, Buck CE, Litton CD. 1998. Integration of archaeological phase information and radiocarbon results from the Jama River Valley, Ecuador: a Bayesian approach. *Latin Am. Antiq.* 9:160–79

Annual Review of Anthropology

Volume 47, 2018

Contents

Anthropology of Language and Communicative Practices

Sociocultural Anthropology

Theme I: Ethics

Theme II: Food

Indexes

Errata

An online log of corrections to *Annual Review of Anthropology* articles may be found at http://www.annualreviews.org/errata/anthro