John Craig and the Probability of History: From the Death of Christ to the Birth of Laplace

Author(s): Stephen M. Stigler

Source: *Journal of the American Statistical Association*, Dec., 1986, Vol. 81, No. 396 (Dec., 1986), pp. 879-887

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2289056

John Craig and the Probability of History: From the Death of Christ to the Birth of Laplace

STEPHEN M. STIGLER*

In 1699 John Craig published an underappreciated book on the probability of historical events. A new interpretation of Craig's work is offered, and it is argued that his formula for the probability of testimony was tantamount to a logistic model for the posterior odds. A modern model for the transmission of evidence is given that is qualitatively similar to Craig's, and Craig's attempt to date the Second Coming of Christ is discussed. Finally, Craig's model is fit to more modern data (from 1882) on the birth date of Laplace.

KEY WORDS: History of statistics; Logistic regression; Evidence; Likelihood ratio; Posterior odds; Apocalyptic inference.

1. INTRODUCTION

I like to think of the constant presence in any sound Republic of two guardian angels: the Statistician and the Historian of Science. The former keeps his finger on the pulse of Humanity, and gives the necessary warning when things are not as they should be. . . . If the Statistician is like a physician, the Historian is like a priest,—the guardian of man's most precious heritage, of the one treasure which, whatever may happen, can never be taken away from him—for the past is irrevocable.

George Sarton (1935, p. 23)

We, like George Sarton, tend to regard the past as fixed. And yet we also know that historical accounts can be extremely uncertain, that in a very real sense our past *can* be taken from us by a series of unreliable historians. "Written on the sands of time" is a metaphor that might have been coined by a statistician, as statisticians and probabilists have long tried to bring their crafts to bear on the question of determining the probability of historical events. From the middle of the 1700s through the nineteenth century, the probability of miracles held a particular fascination for mathematical minds (Kruskal 1982). My concern, however, is principally with an earlier period, with an underappreciated work that John Craig published in 1699. Craig's approach has long been the target of scorn, but I hope to convince the reader that in many respects he was three centuries ahead of his time, that his work was a remarkable early effort at the mathematical modeling of social processes, and that from one perspective it may legitimately be viewed as an anticipation of modern logistic regression.

2. JOHN CRAIG AND THE PROBABILITY OF TESTIMONY

John Craig was a Scottish mathematician and contemporary of Newton; Craig (sometimes spelled Craige) is better known for the early date of his interest in and appreciation of Newton's calculus than for his contributions to it. The date of his birth is unknown, but he wrote on mathematical topics between 1685 and 1718 and he died in 1731. What reputation Craig has as an independent thinker rests entirely on one work, the short book *Theologiae Christianae Principia Mathematica* (*Mathematical Principles of Christian Theology*) that he wrote in 1696 and published in 1699 (Craig 1699, 1964). The major portion of the book is concerned with a mathematical investigation of the probability to be attached to historical events in the light of subsequent accounts of the events, with particular emphasis on the story of Christ.

Craig's use of mathematical analysis on this question has not been treated kindly by those few commentators who have troubled themselves to look at it. Augustus De Morgan (1837), writing in the *Penny Cyclopedia*, called it "a very silly attempt to apply numerical reasoning to historical evidence" (p. 136). To Lubbock and Drinkwater (1830) [quoted in Todhunter (1865, p. 54)], "it has the appearance of an insane parody of Newton's Principia." Charles Gouraud, in his history of probability, called it "quite bizarre" (Gouraud 1848, p. 35). Karl Pearson (1978, pp. 465–466) admired Craig's originality, but thought the formula "perfectly arbitrary." Today, these characterizations do not seem fair.

Craig began his development with a series of nonmathematical definitions, starting with "probability":

Probability is the appearance of agreement or of disagreement of two ideas through arguments whose conclusion is not fixed, or at least is not perceived to be so.

He categorized probability as being "natural" or "historical" according to whether it is based on our own experience or the testimony of others. Craig's main topic was how probability changes as different factors change, for example, how our assessment of the probability of an historical event changes with the number of primary witnesses, with the number of successive witnesses through whom the testimony is handed, and with the distance in time or space from the event in question. He called this a change in "suspicion," in these terms:

Suspicion of historical probability is the application of the mind to the contradictory sides of an historical event.

Velocity of suspicion is the faculty through which the mind is driven, as though through a particular space of time, to see the contradictory sides of an historical account.

Scholium. By space I here mean the degree of assent which the mind gives to opposed arguments of history. Of course I conceive the mind as a moving thing, and arguments as the motive forces driving it in one direction or the other.

At first glance these definitions seem more than vague; they suggest that we are reading the work of a crank. Phrases like "velocity of suspicion" and "the mind as a moving thing" would seem to be ample justification for the strictures of the nineteenth century commentators. Also

* Stephen M. Stigler is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. This work was supported in part by a grant from the National Science Foundation. The author is grateful to the participants in the University of Chicago statistics workshop for their critical comments.

Journal of the American Statistical Association
December 1986, Vol. 81, No. 396, Applications

at first glance, these dark judgments are not dispelled by a look at what emerges from Craig's mathematical analysis: for example, he found the probability of a history transmitted over a distance D in a time T by M successive witnesses to be

$$P = x + (M - 1)s + T^2k/t^2 + D^2q/d^2, \quad (1)$$

where x is the probability attached to the first (or primary) witness, s is the suspicion attached to each of the remaining $M - 1$ witnesses, k is the suspicion known to arise over a time t (all other things being held constant), and q is the suspicion known to arise over a distance d (all other things being held constant). Yet there is a simple way of viewing all that Craig has done that puts him in quite a different light, a simple interpretation that shows him to have been operating intuitively as a highly sophisticated twentieth century statistician.

The key, we must realize, is that Craig's "probability" was not our probability. He was writing at a time before any of the famous treatises of Montmort, Bernoulli, or De Moivre had appeared, before "probability" came to be nearly universally taken to be a measure of uncertainty on a scale of zero to one, where, in the case of independent events, combination was by multiplication. The mathematical reasons for our modern scale are sound, but they are removed from the intuitive level at which Craig was operating. Craig was effectively operating de novo, and there was no reason for him to work directly in terms of our later concepts, yet there is a simple relationship that captures what he was doing quite well. I propose that we should interpret Craig's P as the logarithm of the likelihood ratio in favor of the historical event.

3. CRAIG AND THE LOG-LIKELIHOOD RATIO

To be more precise, let us represent the total available evidence or data, the historical testimony *as received at the present time*, by the symbol E, and let H represent the historical hypothesis or event in question, say the resurrection of Christ. Then I wish to argue that Craig's P should be considered to be

$$P = \log\left(\frac{\Pr\{E \mid H\}}{\Pr\{E \mid \text{not } H\}}\right). \quad (2)$$

We may note that if $\Pr\{H\}$ is the a priori probability of H, independent of any testimony, the posterior odds in favor of H are just

$$\frac{\Pr\{H \mid E\}}{\Pr\{\text{not } H \mid E\}} = \frac{\Pr\{H\}}{\Pr\{\text{not } H\}} \cdot \frac{\Pr\{E \mid H\}}{\Pr\{E \mid \text{not } H\}},$$

so the logarithm of the posterior odds is

$$\log\left(\frac{\Pr\{H \mid E\}}{\Pr\{\text{not } H \mid E\}}\right) = \log\left(\frac{\Pr\{H\}}{\Pr\{\text{not } H\}}\right) + \log\left(\frac{\Pr\{E \mid H\}}{\Pr\{E \mid \text{not } H\}}\right).$$

Thus if my interpretation of Craig is accepted, his P is just the change in the log odds due to the available testimony of all of the witnesses and his Equation (1) is equivalent to a quadratic model for the posterior log odds. Taking t as the time unit and d as the distance unit, (1) becomes

$$P = x + (M - 1)s + T^2k + D^2q,$$

giving us posterior log odds

$$\log\left(\frac{\Pr\{H \mid E\}}{1 - \Pr\{H \mid E\}}\right) = \alpha + Ms + T^2k + D^2q, \quad (3)$$

where $\alpha = \log[(\Pr\{H\})/(1 - \Pr\{H\})] + x - s$; M, T, and D are explanatory variables; and α, s, k, and q are coefficients.

Now, the attribution of so sophisticated a modern idea as the log-likelihood ratio to a seventeenth century mathematician requires qualification and explanation. First, it should be emphasized that Craig most certainly did not show any indication of awareness of the complicated mathematical relationship we associate with the term "log-likelihood ratio"; indeed, he could not have been aware of it because he did not have our "probability" available to him as a concept to refer to. Rather, he was in effect starting from scratch, and he chose as his primitive notion not "fraction of certainty" but the amount of change in a vaguely understood likelihood due to the evidence or data. There are two principal reasons for identifying Craig's primitive notion (which he called probability) with our likelihood ratio. The first is that he was definitely after a concept referring to *changes* in the weight of evidence due to data (the mind being "driven" by "arguments as motive forces"), and the second is that it is clear from his sequence of propositions that he had a measure in mind that was additive, with zero being the baseline representing no net evidential value in the data.

Craig deduced his formula (1) (and others related to it) from the definitions quoted previously through a series of theorems and lemmas. This deduction was not rigorous to our eyes, though, for the proofs usually were circular: the way in which he chose to interpret the vaguely worded hypotheses led immediately to the conclusions. It would be more accurate to view them as a series of definitions, each giving a more precise and more directly mathematical expression to the initially vague idea of "velocity of suspicion." His Theorem II asserted that

> Historical probability increases in proportion to the number of primary witnesses.

Thus his P was to be linearly increasing with the number of primary witnesses. This occurs, for example, with the log-likelihood ratio. Let $\Pr\{E_i \mid H\}$ and $\Pr\{E_i \mid \text{not } H\}$ be the likelihoods of the testimony of the ith primary witness, testimony in favor of the hypothesis or event H. Craig explicitly assumed that all primary witnesses are equally reliable and, further, that "Any history (not contradictory) confirmed by the testimony of one first witness has a certain degree of probability" (his Theorem I), a statement we may interpret as meaning $\Pr\{E_i \mid H\} > \Pr\{E_i \mid \text{not } H\}$. But then if we take him as implicitly assuming independence, we have for n primary witnesses (with the available evidence being $E = \cap_{i=1}^{n} E_i$)

$$\log\left(\frac{\Pr\{E \mid H\}}{\Pr\{E \mid \text{not } H\}}\right) = \log\left(\prod_{i=1}^{n} \frac{\Pr\{E_i \mid H\}}{\Pr\{E_i \mid \text{not } H\}}\right)$$
$$= n \log\left(\frac{\Pr\{E_1 \mid H\}}{\Pr\{E_1 \mid \text{not } H\}}\right)$$
$$= nx, \text{ say.}$$

Since $\Pr\{E_1 \mid H\} > \Pr\{E_1 \mid \text{not } H\}$ by Theorem I,

$$x = \log(\Pr\{E_1 \mid H\}/\Pr\{E_1 \mid \text{not } H\}) > 0$$

and this measure of "historical probability" increases linearly with n, the number of primary witnesses.

We shall see later how Craig intended that a value of zero for his P corresponded to a "vanishing" of the historical probability; that is, zero corresponded to no net evidence being supplied by the available testimony. Both of these, the linearity in the number of primary witnesses and the use of zero as the value for no relevant evidence, fit precisely with my interpretation of P as a log-likelihood ratio. What about the remaining terms, those relating to successive witnesses, time, and distance? There we encounter some difficulties that show that Craig's notion of historical probability was incompletely formed, that in attempting to go directly to a measure of *change* in evidence, instead of starting with something like our probability, a measure of *amount* of evidence, he had been led to a somewhat ad hoc model. The model gave only an approximation to the process that Craig wished to represent, and there is some indication that he realized this.

4. THE TRANSMISSION OF EVIDENCE THROUGH SECONDARY WITNESSES

Craig's problem was this: Evidence would not only accumulate through the addition of primary witnesses, it would also decay, for example, as it passed through the hands of a not perfectly reliable chain of reporters. Algebraically, Craig's P would not only increase, but it would also decrease through suspicion attached to subsequent testimony by secondary witnesses. His Theorem III announced:

> Suspicions of historical probability transmitted through single successive witnesses (other things being equal) increase in proportion to the numbers of witnesses through whom the history is handed down.

In terms of his model, Craig would incorporate this accumulating suspicion by subtracting terms from the historical probability due to the primary witness. With $M - 1$ successive secondary witnesses, the result would be

$$P = x + (M - 1)s,$$

where $s < 0$ was the "suspicion" due to the passage of evidence through one link in this chain. All successive secondary witnesses were equally reliable, hence each should decrease the historical probability equally.

The difficulty with this approach is apparent to us. For if M is large, this makes P negative! In terms of my interpretation, this gives us posterior odds for the event or hypothesis H that are lower than the prior odds, despite the availability of favorable testimony. Even in Craig's eyes this seems to have been unsatisfactory, for he later added a theorem to the effect that his historical probability could not reach zero, Theorem VI:

> Historical probability transmitted by one historian, and through only one series of witnesses, although it continually decreases, nevertheless in no given time utterly vanishes.

His proof of this amounted to a demonstration of the absurdity of the converse: If we imagine a very large number a of independent series of witnesses, then (by Theorem II) they would imply a historical probability aP for the event. Clearly, if a is immense, aP cannot be zero. So P cannot be zero. Although this argument indeed showed the difficulty of permitting $P = 0$, it did not patch the formula. Craig evidently realized this, adding a remark to the effect that P should be capable of becoming imperceptibly small and we could interpret a zero value from the formula as meaning that. He had no interest in the more extreme case of negative P; he seems to have intended that his formula be interpreted as zero in such cases (once historical probability vanished, it is gone, it cannot become "more gone"!).

5. A MODERN PROBABILISTIC VERSION OF CRAIG'S MODEL

To understand better the nature of the difficulty, let us consider a model that plausibly captures what Craig had in mind. Consider a single primary witness, who reports that H has occurred with probability p_1 if H has occurred and with probability p_2 if H has not occurred. Otherwise, that witness reports "not H." We may suppose that $p_1 > p_2$. Further suppose that each subsequent witness reports reliably with probability r what the previous witness said, and with probability $1 - r$ he ad-libs (says "H" with probability q, says "not H" with probability $1 - q$).

At any given stage, the data or evidence available to us consists only of the testimony of the most recent witness. If E_k is the data available at stage k (a report of either H or not H), then the successive reports form a two-state Markov chain, with initial probabilities

$$\Pr\{E_1 = H \mid H\} = p_1,$$
$$\Pr\{E_1 = H \mid \text{not } H\} = p_2,$$

and transition probabilities

$$\Pr\{E_i = H \mid E_{i-1} = H\} = r + (1 - r)q,$$
$$\Pr\{E_i = H \mid E_{i-1} = \text{not } H\} = (1 - r)q.$$

Let

$$P_k(H) = \Pr\{E_k = H \mid H\}$$

and

$$P_k(\text{not } H) = \Pr\{E_k = H \mid \text{not } H\},$$

so the likelihood ratio at stage k is

$$P_k(H)/P_k(\text{not } H).$$

Now $P_1(H) = p_1$ and $P_k(H)$ satisfies

$$P_{k+1}(H) = P_k(H)r + (1 - r)q.$$

It is easy to solve this to find

$$P_k(H) = (p_1 - q)r^{k-1} + q.$$

Similarly,

$$P_k(\text{not } H) = (p_2 - q)r^{k-1} + q,$$

and the log-likelihood ratio is

$$l_k = \log \frac{P_k(H)}{P_k(\text{not } H)} = \log\left(\frac{(p_1 - q)r^{k-1} + q}{(p_2 - q)r^{k-1} + q}\right).$$

If $p_1 > p_2$ and $r < 1$, it is easy to see that l_k decreases toward zero as k increases. The evidential content of the first report diminishes with its retelling.

In Craig's terms, we might call $l_1 = \log(p_1/p_2) > 0$ the historical probability due to the single primary witness and $s_k = l_k - l_{k-1} < 0$ the suspicion that arises with the kth witness ($k > 1$). Since all secondary witnesses are equally reliable, we might (following Craig) be tempted to set $s_2 = s_3 = \cdots = s_k$. But we see that we cannot do that, for the successive suspicions s_k decrease with k as the testimony of the primary witness erodes. [For large k, $s_k = -r^{k-2}(1 - r)(p_1 - p_2)/q$, approximately.] At an intuitive level Craig understood a basic principle of reliability theory: systems in parallel (e.g., primary witnesses) are more reliable than the same systems in series. But by taking suspicion as accumulating at a constant rate he was led to an ad hoc model that is implausible in extreme cases. It is true that we can concoct probability models that fit Craig's supposition of constant suspicion, but only by making the witnesses behave in perverse ways, lying about what they have been told, or by varying their successive r's and q's in pathological ways. Still, for shorter series of successive witnesses, Craig's formula could be useful.

6. CRAIG'S ANALYSIS OF THE EFFECT OF TIME

We thus far have accounted for the first two terms of Craig's Equation (1) for P, the historical probability due to the primary witness and its erosion by suspicion arising from later witnesses. He next moved on to consider the effects of time and distance, as distinct from the simple effects of transmission of testimony from one witness to another. We might now describe this as the differential or marginal effect of time or as the residual effect of time after allowing for number of witnesses. Craig argued that these effects were quadratic in time and distance, and his argument is interesting even if it is not compelling. It was here that the element of Newtonian analogy was the strongest.

Craig accepted it as obvious that time could only decrease the historical probability P; that is, it could only increase suspicion. The question was, at what rate? He argued that the velocity (rate of change) in suspicion increased linearly in time. As he put it in Lemma I,

Velocities of suspicion produced in equal periods of time increase in arithmetical progression.

His reasoning was clearly modeled after that Newton applied to the acceleration of a mass under the influence of constant force of gravity. For Newton's mass, Craig substituted suspicion; for gravity, he substituted the cause of suspicion. He reasoned that over a given interval of time (say t, our notation), the cause of suspicion would act to increase the velocity of suspicion a certain amount (say g, our notation). He asked, how much would the velocity of suspicion have increased due to time alone by time $2t$? Even if the cause of suspicion were suspended (that is, the accelerating effects of time suspended) at time t, the velocity of suspicion at time $2t$ would have increased an amount g by time $2t$. Those effects are not suspended, but they continue to act to increase the velocity another g units to $2g$. Thus *velocity* or rate of increase went up linearly, and it was an easy next step to conclude that the *amount* of suspicion increased as the square of time. As Craig put it in Theorem IV,

Suspicions of historical probability transmitted through any period of time (other things being equal) increase in double proportion to the times taken from the beginning of the history.

As an immediate corollary he found that if an amount of distrust k arose in a unit of time t, the total suspicion arising in a time T was T^2k/t^2; thus he had the third term in Equation (1) for P. His argument regarding the effect of distance, summarized in his Theorem V, was exactly as that for time, and the final term in (1) was in place. He then went on to generalize this equation by allowing for a number b primary witnesses (replace x by bx) and the possibility of several (presumably independent) series of witnesses [he would simply multiply (1) by the number of such series]. And he showed how these formulas could be used to compare histories, to conclude that one history was twice as probable as another, or to find ways of expressing one history (as if canonically) in terms of its equivalent for a different set of witnesses and a different time scale.

Craig's argument for a quadratic rate of increase for suspicion in time and distance does not seem compelling, even within Craig's own framework. In fact it is tempting to guess that he had made a mistake even on his own terms. Would it not be more plausible to recognize that his suspicion is already a change (in historical probability), and a direct Newtonian analogy would have the constant force of historical doubt increasing *that* change linearly? Thus suspicion itself would increase linearly, rather than, as Craig had it, the velocity of suspicion increasing linearly. And was not Craig's treatment here inconsistent with his treatment of successive witnesses, where the transfer of information between witnesses was a constant cause that only increased suspicion linearly but did not increase the velocity of suspicion? Of course, whether or not this "correction" is adopted (and a term linear in T/t substituted for Craig's quadratic term), the same objections as before hold. The equation can be no more than an approximation to a plausible model, and then only for a limited range of T, for otherwise extrapolation would lead to negative probability—in our terms, to posterior odds lower than prior odds.

Even so, Craig's result can be accepted as qualitatively reasonable for a limited range of T. We might view it in these terms: As a single individual holds information, he is constantly transferring it to himself—he is his own successive witness. For Craig, this view should (as we have just noted) have led to the linear accumulation of suspicion with time; for a model of the type discussed in Section 5, it would lead to a term like $\beta(1 - \alpha^T)$ ($\beta < 0$, $0 < \alpha < 1$) instead of Craig's kT^2/t^2. But although my literal interpretation of Craig's P as precisely a log-likelihood ratio does not hold up for these latter terms, the formula can still be viewed as an intuitively reasonable approximation to a log-likelihood ratio, intended to be interpreted in the same way, as an additive measure of change in likelihood.

7. CRAIG'S ANALYSIS OF THE PROBABILITY OF THE STORY OF CHRIST

As the principal application of his formula, Craig investigated the reliability of historical testimony about the story of Christ. The point at issue was the timing of the Second Coming of Christ. The Gospel of Luke (18:8) had hinted that the Second Coming would coincide with the disappearance of faith: "Nevertheless when the son of man cometh, shall he find faith on the earth?" Apparently some in Craig's time thought that the second coming was imminent; if Craig could use his analysis to estimate the date at which the worth of the historical evidence about Christ had decreased to become imperceptible, he would have addressed that question.

Craig first refined his analysis with the testimony of the four Gospels in mind. His initial formula had been cast in terms of oral testimony, retransmitted orally. On the other hand, each of the Gospels was then considered the written record of an on-the-scene historian and the transmission was made by recopying rather than retelling. Presumably the evidence in such written series was more reliable than that in oral series. Neglecting the effect of distance, Craig gave the present probability of the story of Christ to be

$$P = cz + (n - 1)f + T^2k/t^2,$$

where $c = 4$ is the number of primary historians (Matthew, Mark, Luke, and John), n is the number of successive recopyings of their testimony made to date, f is the suspicion arising with each recopying, and T^2k/t^2 is the suspicion arising over time T years (k is the amount arising in t years, where he took $t = 50$ years as a unit).

To evaluate this formula Craig needed values for z (the historical probability arising from one primary historian), n, f, and k. He argued that n was roughly proportional to time; he took $n = T/4t$, corresponding to one recopying every 200 years. He related the other quantities to x, the probability to be attached to one contemporary eyewitness's oral testimony. First, he set $z = 10x$—a primary historian's considered written record was assumed to be equivalent to 10 independent eyewitness oral accounts. Next, $f = -x/100$—the suspicion arising in 1 recopying is such that 100 recopyings are necessary to erode the testimony of one eyewitness to imperceptibility. And finally, $k = -x/100$—the suspicion arising by delaying the telling of the story by 50 years is equivalent to that of one recopying.

These values give

$$P = 40x - \left(\frac{T}{4t} - 1\right)\frac{x}{100} - \frac{T^2}{t^2}\cdot\frac{x}{100},$$

and for the year $T = 1696$ Craig found $P = 28x$. That is, the account of the story of Christ had initially had a likelihood equivalent to that of 40 independent eyewitnesses; after 1,696 years it would decay to be worth that of 28 eyewitnesses. The historical probability of this written testimony would not vanish entirely, Craig calculated, until the year 3150 (a modern recalculation with his numbers gives 3156). Craig's conclusion was that those who thought the Second Coming was at hand were gravely in error, for the conditions postulated for Christ's reappearance would not occur for another 1,454 (= 3,150 − 1,696) years.

From our present vantage point we can find much to criticize in Craig's work. His "historical probability" may have corresponded (precisely, in some respects) to a log-likelihood ratio, but his implicit assumptions of independence, even if viewed as only conditional independence, could be argued with. Were Matthew, Mark, Luke, and John really four independent accounts of Christ? Modern scholarship would deny that they are, and indeed, would date the earliest at least a generation after the crucifixion. And Craig's formula was in some respects only an ad hoc approximation to what we might view as a plausible probability model for sequential transmission of information, an approximation that could be expected to break down in extreme situations and be particularly unsuitable for extrapolation. Yet the fundamental question he addressed required extrapolation. In addition, Craig's use of the formula required the specification of several coefficients. He did not have hard statistical data to determine these coefficients, and he had to fall back on subjectively plausible guesses. Yet he did not examine how sensitive his conclusion could be to variations in these guesses.

From another point of view, though, that of historical analysis, these complaints are mere quibbles about an ingenious and fundamentally novel attack on an extraordinarily difficult problem. Others, from Craig's time to the present, examined questions of the reliability of testimony. The earliest explicitly mathematical one of these was published anonymously in the *Philosophical Transactions* in 1699. [Brown Grier (1982) of Northern Illinois University has recently found that the work was by George Hooper and has located an earlier nonalgebraic version buried in a religious tract published in 1689.] But without exception those other early attempts treated only simplistic situations (not allowing for covariates such as time and distance or the differences between oral and written testimony), and they were far more sensitive to the implicit assumption of independence than Craig's approach was. The literal interpretation of Craig's coefficients required independence, but short of extrapolation, the rough validity of the model did not. Alternative models offered little more than products of probabilities and the all-too-obvious observation that if enough of them were multiplied together the product would be nearly zero. Craig's ingenuity in starting with what I have argued was effectively the log-likelihood ratio may have been necessary (he had no well-developed alternative in 1696!), but he was ingenious nonetheless. We even recognize in his formulation an adumbration of one of today's hottest "new" statistical approaches, logistic regression. The dichotomization of hypotheses about historical events such as the story of Christ into two possible states, true or false, may be highly questionable, but given that dichotomy the approach through the logarithm of the odds ratio remains appealing. And even if Craig lacked statistical data, he again showed ingenuity in dealing with his application. He recognized the close collinearity between number of successive witnesses and time and turned it to his advantage by using it to eliminate one coefficient: in place of (3) (where M = number of secondary witnesses),

$$\log\left(\frac{\Pr\{H \mid E\}}{1 - \Pr\{H \mid E\}}\right) = \alpha + Ms + T^2k + D^2q,$$

he would effectively have

$$\log\left(\frac{\Pr\{H \mid E\}}{1 - \Pr\{H \mid E\}}\right) = \alpha + Ts' + T^2k + D^2q. \quad (4)$$

Craig was also far ahead of his time in recognizing the value of relating all variables to a common scale, choosing as his unit x, the worth of a single eyewitness's testimony. Without this convention, the different terms of his formula are incommensurable and the interpretation of the coefficients is a vexed question. Craig's treatise is a remarkable early example of the application of mathematical statistics to a problem in social science, despite the lack of statistical data.

8. AN EMPIRICAL TEST: THE BIRTH OF LAPLACE

With a few qualifications, Craig's formula for the decay in the reliability of historical testimony seems to be intuitively reasonable and in at least qualitative agreement with another more detailed model for the transmission of information. But a basic question remains: Does it fit the facts? How well would it survive an empirical test? Craig's own application to the story of the life of Christ was limited by the lack of available statistical data and the consequent need to estimate coefficients subjectively. And in any event, it is hard to see how his inferences could be verified, short of waiting for the Second Coming. I propose to examine the empirical applicability of Craig's formula to data concerning a more recent event, the birth of the French scientist Pierre Simon Laplace.

There is a note of irony in the application of Craig's model to Laplace, for Laplace was among those who commented adversely about the model. In his famous *Essai philosophique sur les probabilités*, Laplace (1814) dismissed Craig's conclusion about the Second Coming (which Laplace described as the end of the world) in these words: "His analysis is as mistaken as his hypothesis about the duration of the world is bizarre" (p. 85; 1951 ed., p. 125). This uncharitable characterization aside, Laplace's role in the history of statistics is so important that he seems an apposite choice as an historical subject on which to test Craig's formula. In addition, thanks to the labors of the nineteenth century bibliographer of the history of mathematics, Prince Baldassare Boncompagni, there are statistical data available on Laplace that are much more suitable for our purposes than those on Christ. In 1882, Boncompagni compiled a list of 65 biographical accounts that gave either the birth date or death date of Laplace (Table 1). Through diligent archival work, Boncompagni was able to establish beyond any reasonable doubt that Laplace was born on March 23, 1749, and died on March 5, 1827. Indeed, Boncompagni's historical accounts showed a preference for these dates, though the margin of preference was narrow. Table 2 gives a cross-classification of the 65 accounts by birth date and death date.

Of the 26 + 24 = 50 accounts that gave a birth date, 26, or 52%, were correct. It is interesting to note that the inaccurate accounts were not the work of a few consistently careless historians: what association there is between inaccuracies in the two dates is negative. The cross-classification shows a slightly greater tendency for these writers to get one of the two dates correct than if they erred at random.

Table 3 shows the data on birth dates grouped by time

Table 1. Data on Birth Dates (B), Death Dates (D), and Language (L) From 65 Biographical Accounts of Laplace, From Boncompagni (1882)

No.	Year	B	D	L	No.	Year	B	D	L
1	1799	C	N	G	34	1842	C	C	F
2	1802	C	N	F	35	1843	N	6	F
3	1812	C	N	E	36	1845	C	7	F
4	1818	C	N	F	37	1847	22	C	F
5	1823	28	N	F	38	1847	C	C	I
6	1826	C	N	F	39	1847	C	C	I
7	1826	C	N	F	40	1848	N	6	I
8	1826	C	N	F	41	1848	C	C	F
9	1827	N	C	F	42	1848	22	C	F
10	1827	C	C	F	43	1849	C	6	F
11	1827	C	C	F	44	1851	28	5	G
12	1827	C	C	F	45	1852	28	C	F
13	1827	N	C	F	46	1853	N	C	E
14	1827	N	6	F	47	1854	28	C	F
15	1827	N	C	G	48	1855	28	C	F
16	1828	N	C	F	49	1855	22	7	F
17	1828	27	6	F	50	1856	N	5	E
18	1828	N	C	I	51	1857	C	C	F
19	1828	28	6	F	52	1859	C	C	F
20	1829	C	5	F	53	1860	C	C	F
21	1830	28	C	F	54	1863	28	C	G
22	1833	N	C	F	55	1866	N	C	I
23	1834	28	6	F	56	1867	28	C	F
24	1834	N	6	F	57	1872	C	C	F
25	1835	28	5	G	58	1872	28	C	G
26	1836	C	7	F	59	1873	28	C	F
27	1836	C	7	F	60	1876	C	C	F
28	1837	C	5	F	61	1877	28	C	G
29	1839	N	6	I	62	1877	28	C	G
30	1840	22	C	F	63	1881	28	C	G
31	1841	C	C	F	64	1881	28	C	G
32	1841	N	6	F	65	1881	28	C	G
33	1842	22	C	F					

NOTE: C = Correct, N = None, F = French, G = German, I = Italian, E = English. Numbers in columns B and D indicate dates in March, except that "5" indicates "May 5."

Table 2. A Cross-Classification of 65 Accounts of the Life of Laplace, From Boncompagni (1882)

	Death date			
Birth date	Right	Wrong	None given	Totals
Right	13	6	7	26
Wrong	17	6	1	24
None given	8	7	0	15
Totals	38	19	8	65

NOTE: For example, 17 of the 65 accounts gave the wrong birth date and the correct death date. Note that the zero in column 3 is a structural zero; only accounts that gave at least one of the two dates were included. The more common wrong birth dates were March 22 (5 times) and March 28 (18 times); the death dates included March 6 (10 times), March 7 (4 times), and May 5 (5 times).

Table 3. Boncompagni's Data on Laplace's Birth Date, Grouped by Date of Publication of Account into Three Approximately Equal Groups

Birth date	1799–1830	1833–1849	1851–1881
Right	11	10	5
Wrong	4	6	14
None given	6	6	3
Totals	21	22	22

period, and it gives the quick impression of at least qualitative agreement with Craig's assumptions. There is an evident decrease in the reliability of the more recent accounts. Interestingly, seven of the eight accounts published during Laplace's lifetime are correct, as if Laplace was looking over the biographer's shoulder. Table 4 summarizes the same data grouped by language. Crudely speaking, we may identify accounts written in French as being at distance $D = 0$ from the place of Laplace's birth. The remaining accounts are in the national languages of neighbors of France; we may classify them as being at linguistic distance $D = 1$. The pattern in Table 4 is supportive of decreasing reliability with increasing distance.

Craig's model, as I have interpreted it in Section 3, amounts to representing the logarithm of the likelihood ratio as a possibly quadratic function $f(T, D)$ of time and distance: in the notation of that section,

$$\log\left(\frac{\Pr\{E \mid H\}}{\Pr\{E \mid \text{not } H\}}\right) = f(T, D). \tag{5}$$

In the present context, H would stand for the hypothesis (now known to be true, thanks to Boncompagni) "Laplace born March 23, 1749" and E would represent the evidence of one biographical account (published T years since Laplace's birth, a distance D from France) *reporting* that Laplace was born March 23, 1749. Since all of our data are gathered under the single hypothesis H, it is clear that we cannot address Craig's model directly (we cannot estimate $\Pr\{E \mid \text{not } H\}$) without further assumptions. The principal assumption that will be adopted here is that $\Pr\{E \mid \text{not } H\} = K \Pr\{\text{not } E \mid H\}$, where the constant K of proportionality does not vary with T and D. With this assumption, Craig's model (5) becomes

$$\log\left(\frac{\Pr\{E \mid H\}}{\Pr\{\text{not } E \mid H\}}\right) = \log K + f(T, D), \tag{6}$$

and (subject to one further assumption, to be mentioned subsequently) we can fit this model by logistic regression to obtain estimates of the coefficients of T and D in $f(T, D)$.

The plausibility of the assumption $\Pr\{E \mid \text{not } H\} = K \Pr\{\text{not } E \mid H\}$ can be argued as follows. Since a priori $\Pr\{H\}$ and $\Pr\{\text{not } H\}$ cannot depend on T and D, it suffices to show that $\Pr\{E \cap \text{not } H\} \propto \Pr\{\text{not } E \cap H\}$. Let "not H" $= \cup H_i$ and "not E" $= \cup E_i$, where the H_i's and E_i's, respectively, represent the different days of the year, excluding March 23. Then $\Pr\{E \cap \text{not } H\} = \sum \Pr\{E \cap H_i\}$ and $\Pr\{\text{not } E \cap H\} = \sum \Pr\{E_i \cap H\}$. Now it seems reasonable to expect a certain amount of "time symmetry"—for example, that biographers are as likely to miswrite March 28 for March 23 as they would be to write March 23 for March 28. But that implies that $\Pr\{E \mid H_i\} = \Pr\{E_i \mid H\}$ for all i, and if all birth dates are a priori equally likely, $\Pr\{H_i\} = \Pr\{H\}$ and $\Pr\{E \cap \text{not } H\} = \Pr\{\text{not } E \cap H\}$. This would give us $\Pr\{E \mid \text{not } H\} = K \Pr\{\text{not } E \mid H\}$, with $K = \Pr\{H\}/\Pr\{\text{not } H\}$, the prior odds, and we would in fact have that (6) gives the logarithm of the posterior odds. (Actually, the data as summarized in the note to Table 2 suggest that $\Pr\{E_i \mid H\}$ is near zero for all but a few i.) Let us then accept the assumption and conclude that estimating (6) will give us the coefficients we need to test Craig's model.

Consider only the $n = 50$ accounts that gave a birthdate for Laplace. Let $Y_i = 1$ if the ith account is correct and 0 if the ith account is wrong, so "$Y_i = 1$" represents an instance of positive evidence, "E," of the preceding discussion. Let

$$E(Y_i) = \mu_i = \Pr\{Y_i = 1 \mid T_i, D_i, H\} \tag{7}$$

and consider this version of (6):

$$\log_e\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 D_i, \tag{8}$$

where H stands as before for the hypothesis "Laplace born March 23, 1749," $T_i =$ the number of years since the birth of Laplace, and $D_i =$ distance from France ($D = 0$ if the language is French, $D = 1$ otherwise). This model was fit as a simple logistic regression model by using GLIM. This amounts to assuming that the time and distance capture

Table 4. Boncompagni's Data on Laplace's Birth Date Grouped by Language

Birth date	French accounts	Other languages
Right	22	4
Wrong	15	9
None given	8	7
Totals	45	20

NOTE: "Other languages" includes German (11 accounts, 10 with birth dates), Italian (6, 2 dated), and English (3, 1 dated).

Table 5. A Summary of Logistic Regression Analyses (using GLIM) of the Birth Date Data

Model	Terms in model	Estimated coefficients	Standard error	Deviance	Degrees of freedom
A	Constant	.08	.28	69.2	49
B	Constant	.38	.33	66.0	48
	Distance	−1.19	.69		
C	Constant	2.19	.84	60.4	48
	Time	−.0463	.0172		
D	Constant	2.17	.84	59.3	47
	Time	−.0419	.0175		
	Distance	−.80	.78		
E	Constant	2.43	1.71	60.35	47
	Time	−.0581	.0749		
	Time squared	.00013	.00077		
F	Constant	3.06	1.80	58.9	46
	Time	−.0870	.0798		
	Time squared	.00050	.00085		
	Distance	−1.01	.87		
G	Constant	1.29	.96	56.3	46
	Time	−.0214	.0210		
	Distance	3.4	3.4		
	Time * Distance	−.087	.067		

NOTE: Time = years since 1800.

all dependence between accounts and that the Y_i's are conditionally independent given T_i and D_i. In particular, it amounts to making the one further assumption alluded to earlier, that the biographical accounts in the study are like observations of an ongoing process and that they do not themselves influence that process. Since most of the accounts are of a decidedly minor nature, this additional assumption seems unlikely to be often far from the mark; even if a few accounts could be considered "authorities," we may hope that in the aggregate the effect on interpretation of our model is not great. In any event, although it would be interesting to attempt a more sophisticated model for these data, my primary aim here is to force Craig's model on the data to attempt to answer these questions: Is there a discernible effect due to distance? Is there a significant time effect? Is the time effect quadratic, as Craig hypothesized? We may note that the model (8) is essentially Craig's model (1) as I have interpreted it, following Craig in eliminating collinearity by substituting time for the number of successive witnesses. In addition, the distance measure we have is insufficiently detailed to permit the estimation of a term in D^2.

The results of this analysis are summarized in Table 5. For these data, Craig's T^2 term seems imperceptible: a linear term in time is all that is required. The simplest adequate model is

$$\log\left[\frac{\Pr(E \mid H)}{\Pr(\text{not } E \mid H)}\right] = 2.19 - .0463 \text{ Time},$$

where Time = years since 1800. A slight improvement can be obtained by including distance. Craig had suggested that more distant accounts would be less reliable, and this is not contradicted by these data, though the effect of distance is evidently not simply additive as in Craig's model, but multiplicative as in Model G:

$$\log\left[\frac{\Pr(E \mid H)}{\Pr(\text{not } E \mid H)}\right] = 1.29 - .0214 \text{ Time for } D = 0$$

$$= 4.69 - .1084 \text{ Time for } D = 1,$$

where Time = years since 1800. Thus the estimated rate of decay of information—of the log odds that a reported birth date is given correctly—is about 5 times as fast outside of France as in French accounts. Under our assumptions the same holds true for the posterior odds that a report is correct, given the report.

Boncompagni's data on Laplace's death date behave rather differently from those on the birth date. Table 6 captures in broad outline the main pattern confirmed by the analyses reported in Table 7: the effect of time is not linear. In fact, the model of choice here (Model E) includes Craig's T^2, but in a way that reflects decay followed by *increasing* reliability in time. The model (with "Time" = years since 1800)

$$\log\left(\frac{\Pr\{E \mid H\}}{\Pr\{\text{not } E \mid H\}}\right) = 11.0 - .58 \text{ Time} + .0073 \text{ (Time)}^2$$

Table 6. Boncompagni's Data on Laplace's Death Date, Grouped by Year of Publication of Account into Three Nearly Equal Groups

Death date	1799–1830	1833–1849	1851–1881
Right	9	10	19
Wrong	4	12	3
None given	8	0	0
Totals	21	22	22

Table 7. A Summary of Logistic Regression Analyses (using GLIM) of the Death Date Data

Model	Terms in model	Estimated coefficients	Standard error	Deviance	Degrees of freedom
A	Constant	.69	.28	72.6	56
B	Constant	.58	.33	72.2	55
	Distance	.38	.62		
C	Constant	−1.64	.95	65.3	55
	Time	.052	.021		
D	Constant	−1.68	.96	65.1	54
	Time	.055	.023		
	Distance	−.29	.71		
E	Constant	11.0	5.7	57.4	54
	Time	−.58	.29		
	Time squared	.0073	.0035		
F	Constant	11.1	5.7	57.1	53
	Time	−.58	.29		
	Time squared	.0073	.0036		
	Distance	−.42	.76		
G′	Constant	11.8	5.7	55.4	52
	Time	−.64	.30		
	Time squared	.0085	.0037		
	Distance	3.9	3.8		
	Time * Distance	−.103	.086		

NOTE: Time = years since 1800.

reaches a minimum at about Time = 40, or in the year 1840. Testimony on Laplace's death date would seem to have reached a minimum credibility within two decades after his death, well before Boncompagni's study. The effect of distance (as measured by language) is not statistically significant; the addition of a linear interaction term to Model F (as in Model G′) improves the fit only slightly. What these results show is that simple models that postulate a monotone decay of historical evidence will be insufficiently rich for practical use. The Laplace data suggest a tendency for convergence, but historians will be relieved to see that it is not always toward error.

[*Received June 1985. Revised February 1986.*]

REFERENCES

Boncompagni, Baldassare (1882), "Intorno Algi Atti di Nascita di Morte di Pietro Simone Marchese di Laplace," *Bulletino di Bibliografia e di Storia Delle Scienze Matematiche e Fisiche,* 15, 446–464.

Craig, John (1699), *Theologiae Christianae Principia Mathematica,* London: Timothy Child. (Reprinted with commentary by Daniel Titius in Leipzig in 1755.)

——— (1964), "Craig's Rules of Historical Evidence," in *History and Theory: Studies in the Philosophy of History* (Beiheft 4), pp. 1–31. [A reprint of the Latin original of most of the material in Craig (1699) through the end of chap. 2, with facing English translation.]

De Morgan, Augustus (1837), "Craig, John," *Penny Cyclopedia,* 8, 135–136.

Gouraud, Charles (1848), *Histoire du Calcul des Probabilités Depuis ses Origines Jusqu'à nos Jours,* Paris: Durand.

Grier, Brown (1982), "George Hooper and the Early Theory of Testimony," unpublished manuscript.

Kruskal, William H. (1982), "Miracles and Statistics: The Casual Assumption of Independence," unpublished ASA Presidential Address.

Laplace, Pierre Simon (1814), *Essai Philosophique sur les Probabilités,* Paris: Courcier. (Sixth edition translated as *A Philosophical Essay on Probabilities,* reprinted 1951 by Dover, New York.)

Lubbock, J., and Drinkwater, J. (1830), *On Probability,* London: Society for the Diffusion of Useful Knowledge.

Pearson, Karl (1978), *The History of Statistics in the 17th and 18th Centuries, Against the Changing Background of Intellectual, Scientific and Religious Thought* (Lectures from 1921–1933, edited by E. S. Pearson), London: Charles W. Griffin.

Sarton, George (1935), "Preface to Volume XXIII of Isis (Quetelet)," *Isis,* 23: 6–24.

Todhunter, Isaac (1865), *A History of the Mathematical Theory of Probability,* London: Macmillan. (Reprinted in 1965 by Chelsea, New York.)