Richard C. Carrier, Ph.D.

"Bayes' Theorem for Beginners: Formal Logic and Its Relevance to Historical Method — Adjunct Materials and Tutorial"

The Jesus Project Inaugural Conference
"Sources of the Jesus Tradition: An Inquiry"
5-7 December 2008 (Amherst, NY)

Table of Contents for Enclosed Document

NOTE: A chapter of the same title was published in 2010 by Prometheus Press (in *Sources of the Jesus Tradition: Separating History from Myth*, ed. R. Joseph Hoffmann, 2010) discussing or referring to the contents of this online document. That primary document (to which this document is adjunct) has also been published in advance as **"Bayes' Theorem for Beginners: Formal Logic and Its Relevance to Historical Method"** in ***Caesar: A Journal for the Critical Study of Religion and Human Values*** 3.1 (2009): 26-35.

Richard C. Carrier, Ph.D.

"Bayes Theorem for Beginners: Formal Logic and Its Relevance to Historical Method"

December 2008 (Amherst, NY)

Notes and Bibliography

1. Essential Reading on "Historicity Criteria"

Stanley Porter, *The Criteria for Authenticity in Historical-Jesus Research: Previous Discussion and New Proposals* (Sheffield Academic Press: 2000).

Christopher Tuckett, "Sources and Methods," *The Cambridge Companion to Jesus*, edited by Markus Bockmuehl (Cambridge University Press: 2001): pp. 121-37.

Gerd Theissen and Dagmar Winter, *The Quest for the Plausible Jesus: The Question of Criteria* (John Knox Press: 2002).

2. Example List of Popular Historicity Criteria

Incomplete List (names often differ, criteria often overlap – here are 17; there are two or three dozen):

Dissimilarity	- dissimilar to independent Jewish or Christian precedent
Embarrassment	- if it was embarrassing, it must be true
Coherence	- coheres with other confirmed data
Multiple Attestation	- attested in more than one independent source
Contextual Plausibility	- plausible in a Jewish or Greco-Roman cultural context
Historical Plausibility	- coheres with a plausible historical reconstruction
Natural Probability	- coheres with natural science (etc.)
Explanatory Credibility	- historicity better explains later traditions
Oral Preservability	- capable of surviving oral transmission
Fabricatory Trend	- isn't part of known trends in fabrication or embellishment
Least Distinctiveness	- the simpler version is the more historical
Vividness of Narration	- the more vivid, the more historical
Crucifixion	- explains why Jesus was crucified
Greek Context	- if whole context suggests parties speaking Greek
Aramaic Context	- if whole context suggests parties speaking Aramaic
Textual Variance	- the more invariable a tradition, the more historical
Discourse Features	- if J's speeches cohere in style but differ fr. surrounding text

3. Formal Logical Analysis – discussion & examples available by online document:

www.richardcarrier.info/CarrierDec08.pdf

4. Recommended Reading on Bayes' Theorem

Eliezer Yudkowsky, "An Intuitive Explanation of Bayesian Reasoning (Bayes' Theorem for the Curious and Bewildered: An Excruciatingly Gentle Introduction)" at yudkowsky.net/bayes/bayes.html

Douglas Hunter, *Political [and] Military Applications of Bayesian Analysis: Methodological Issues* (Westview Press: 1984).

Luc Bovens and Stephan Hartmann, *Bayesian Epistemology* (Oxford: 2003).

Timothy McGrew, "Bayesian Reasoning: An Annotated Bibliography" at homepages.wmich.edu/~mcgrew/bayes.htm

Wikipedia on "Bayes' Theorem" (for English: en.wikipedia.org/wiki/Bayes'_theorem)

5. Bayes' Theorem (Complete)

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{[\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]}$$

6. Bayes' Theorem (Abbreviated)

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{P(e|b)}$$

Note that $P(e|b) = [\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]$ so I recommend the longer form instead, as a check against error (especially that of assuming $P(e|b) = P(e|\sim h.b)$).

7. Explanation of the Terms in Bayes' Theorem

P = Probability (epistemic probability = the probability that something stated is true)

h = hypothesis being tested

~h = all other hypotheses that could explain the same evidence (if *h* is false)

e = all the evidence directly relevant to the truth of *h* (*e* includes both what *is* observed *and* what is *not* observed)

b = total background knowledge (all available personal and human knowledge about anything and everything, from physics to history)

P(h|e.b) = the probability that a hypothesis (*h*) is true given all the available evidence (*e*) and all our background knowledge (*b*)

P(h|b) = **the prior probability that *h* is true** = the probability that our hypothesis would be true given only our background knowledge (i.e. if we knew nothing about *e*)

P(e|h.b) = **the consequent probability of the evidence (given *h* and *b*)** = the probability that all the evidence we have would exist (or something comparable to it would exist) if the hypothesis (and background knowledge) is true.

P(~h|b) = 1 – P(h|b) = **the prior probability that *h* is false** = the sum of the prior probabilities of all alternative explanations of the same evidence (e.g. if there is only one viable alternative, this means the prior probability of all other theories is vanishingly small, i.e. substantially less than 1%, so that P(~h|b) is the prior probability of the one viable competing hypothesis; if there are many viable competing hypotheses, they can be subsumed under one group category (~*h*), or treated independently by expanding the equation, e.g. for three competing hypotheses [P(h|b) x P(e|h.b)] + [P(~h|b) x P(e|~h.b)] becomes [P(h_1|b) x P(e|h_1.b)] + [P(h_2|b) x P(e|h_2.b) + [P(h_3|b) x P(e|h_3.b)])

P(e|~h.b) = **the consequent probability of the evidence if *b* is true but *h* is false** = the probability that all the evidence we have would exist (or something comparable to it would exist) if the hypothesis we are testing *is false*, but all our background knowledge is still true. This also equals the posterior probability of the evidence if some hypothesis *other* than *h* is true—and if there is more than one viable contender, you can include each competing hypothesis independently (per above) or subsume them all under one group category (~*h*).

8. The Advantages of Bayes' Theorem

1. Helps to tell if your theory is *probably* true rather than merely *possibly* true
2. Inspires closer examination of your backgr. knowl. and assumptions of likelihood
3. Forces examination of the likelihood of the evidence on competing theories
4. Eliminates the Fallacy of Diminishing Probabilities
5. Bayes' Theorem has been proven to be formally valid
6. Bayesian reasoning with or without math exposes assumptions to criticism & consequent revision and therefore promotes progress

9. Avoiding Common Errors with Bayes' Theorem

1. The Fallacy of False Precision

   (SOLUTION: include reasonable margins of error)

2. The Fallacy of Confusing Evidence with Theories

   (SOLUTION: try to limit contents of *e* to tangible physical facts, i.e. actual surviving artifacts, documents, etc., and straightforward generalizations therefrom)

3. The Fallacy of Confusing Assumptions with Knowledge

   (SOLUTION: try to limit contents of *b* to universally agreed expert knowledge)

NOTE: Further discussion of all these topics (and in considerable detail) will appear in my forthcoming book: Richard Carrier, *On the Historicity of Jesus Christ* (there especially in chapter two and a mathematical appendix). For contact information and more about my work in other aspects of Jesus studies and beyond, see **www.richardcarrier.info**.

To employ Bayesian reasoning without doing any math, employ relative (but non-numerical) estimates of likelihood on the following flowchart:

Bayesian Reasoning Flowchart

```
                                        - P(e|~h.b) is same or higher = low P(h|e.b) = h is probably false
                    -- low P(e|h.b) ----|
                    |                   - P(e|~h.b) is lower = low P(h|e.b) unless prior is high enough
                    |
low P(h|b) ---------|
                    |
                    |                   - P(e|~h.b) is lower = low P(h|e.b) unless prior is high enough
                    -- high P(e|h.b) ---|
                                        - P (e|~h.b) is same or higher = low P(h|e.b) = h is probably false


                                        - P(e|~h.b) is same or lower = high P(h|e.b) = h is probably true
                    -- low P(e|h.b) ----|
                    |                   - P(e|~h.b) is higher = low P(h|e.b) unless prior is high enough
                    |
high P(h|b) --------|
                    |
                    |                   - P (e|~h.b) is higher = low P(h|e.b) unless prior is high enough
                    -- high P(e|h.b) ---|
                                        - P(e|~h.b) is same or lower = high P(h|e.b) = h is probably true
```

"Low" and "High" = lower than 0.5 (50%) and higher than 0.5 (50%), respectively ; when P(h|b) = 0.5, so does P(~h|b): then the hypothesis with the higher P(e|b) is probably true.

"Prior is high enough" = when P(h|b) is higher than the Bayesian ratio between P(e|~h.b) and P(e|h.b) by enough to overcome their difference and thus produce the opposite result.

Richard C. Carrier, Ph.D.

"Bayes Theorem for Beginners: Formal Logic and Its Relevance to Historical Method"

December 2008 (Amherst, NY)

At our Jesus Project conference I will refer to items in this document as I go, in lieu of slides. But I won't discuss every item in it.

1. Essential Reading on "Historicity Criteria"

Stanley Porter, *The Criteria for Authenticity in Historical-Jesus Research: Previous Discussion and New Proposals* (Sheffield Academic Press: 2000).

Christopher Tuckett, "Sources and Methods," *The Cambridge Companion to Jesus*, edited by Markus Bockmuehl (Cambridge University Press: 2001): pp. 121-37.

Gerd Theissen and Dagmar Winter, *The Quest for the Plausible Jesus: The Question of Criteria* (John Knox Press: 2002).

2. Examining Historicity Criteria

Typical Problems:

1. The criterion is invalidly applied (e.g. the text actually fails to fulfill the criterion, contrary to a scholar's assertion or misapprehension)

2. The criterion itself is invalid (e.g. the criterion depends upon a rule of inference that is inherently fallacious, contrary to a scholar's intuition)

Required Solutions:

1. Scholars are obligated to establish with clear certainty that a particular item actually fulfills any stated criterion (and what exactly that criterion is).

2. Scholars are obligated to establish the formal logical validity of any stated criterion (especially if establishing its validity requires adopting for it a set of qualifications or conditions previously overlooked).

<u>Incomplete List</u> (names often differ, criteria often overlap – here are 17; there are two or three dozen):

Dissimilarity	- dissimilar to independent Jewish or Christian precedent
Embarrassment	- if it was embarrassing, it must be true
Coherence	- coheres with other confirmed data
Multiple Attestation	- attested in more than one independent source
Contextual Plausibility	- plausible in a Jewish or Greco-Roman cultural context
Historical Plausibility	- coheres with a plausible historical reconstruction
Natural Probability	- coheres with natural science (etc.)
Explanatory Credibility	- historicity better explains later traditions
Oral Preservability	- capable of surviving oral transmission
Fabricatory Trend	- isn't part of known trends in fabrication or embellishment
Least Distinctiveness	- the simpler version is the more historical
Vividness of Narration	- the more vivid, the more historical
Crucifixion	- explains why Jesus was crucified
Greek Context	- if whole context suggests parties speaking Greek
Aramaic Context	- if whole context suggests parties speaking Aramaic
Textual Variance	- the more invariable a tradition, the more historical
Discourse Features	- if J's speeches cohere in style but differ fr. surrounding text

3. Formal Logic: The Basic Syllogism

Major Premise: [Some general rule]
Minor Premise: [Some specific fact satisfying the general rule]
Conclusion: [That which follows necessarily from the major and minor premise]

<u>Strongest Argument</u>:

Major Premise: All working wagons had wheels.
Minor Premise: Jacob owned a working wagon.
Conclusion: Therefore, Jacob owned a wheel.

[This is true by virtue of the definition of 'working wagon']

<u>Stronger Argument</u>:

Major Premise: All major cities in antiquity had sewers.
Minor Premise: Jerusalem was a major city in antiquity.
Conclusion: Therefore, Jerusalem had sewers.

[This is true to a very high degree of probability by virtue of abundant background evidence, including archaeological and textual, that is uncontestable for its scope and degree]

= *Major Premise*: [Very probably] all major cities in antiquity had sewers.
Minor Premise: [Very probably] Jerusalem was a major city in antiquity.
Conclusion: Therefore, [very probably] Jerusalem had sewers.

Weaker Argument:

Major Premise: All major cities in antiquity had public libraries.
Minor Premise: Jerusalem was a major city in antiquity.
Conclusion: Therefore, Jerusalem had a public library.

[This is true but not to a high degree of probability by virtue of archaeological and textual evidence that is contestable to some degree]

= *Major Premise*: [Most] major cities in antiquity had public libraries.
= [Somewhat probably] any major city in antiquity had a public library.
Minor Premise: [Very probably] Jerusalem was a major city in antiquity.
Conclusion: Therefore, [somewhat probably] Jerusalem had a public library.

[Here basic theory would produce a probability for the conclusion, P(Jerusalem had a public library) = P(JL), equal to the *product* of the probabilities of the major and minor premise, due to the Law of Conditional Probability: so although one might say the conclusion is "[somewhat probably] Jerusalem had a public library" this "somewhat probably" will be slightly less than the "somewhat probably" implied in the major premise. For example, if the major premise has a probability of 60% and the minor premise a probability of 90%, then P(JL) = 0.6 x 0.9 = 0.54 = 54%, not 60%. This can lead to a fallacy of diminishing probabilities, where the more evidence you have, the lower the probability of the conclusion (see next), which clearly cannot be correct. The solution is Bayes' Theorem, which eliminates the fallacy of diminishing probabilities. Hence the importance of Bayes' Theorem.]

Complex Argument:

Major Premise 1: [Probably] all major cities in antiquity had public libraries.
Major Premise 2: [Probably] when the surviving text of something written by an ancient author mentions consulting books in a city's public library, then that city had a public library.
Minor Premise 1: [Very probably] Jerusalem was a major city in antiquity.
Minor Premise 2: [Very probably] we have a 3rd century papyrus fragment of an encyclopedia written by Hippolytus a few decades before, which mentions his consulting books in Jerusalem's public library.
Conclusion: Therefore, [very probably] Jerusalem had a public library.

[Here the probability that Jerusalem had a public library should be increased by the fact of having two kinds of evidence mutually supporting the same conclusion, including both general evidence, and specific evidence. More evidence of either kind could be added to raise the probability of the conclusion even more. But observe, if P(Major 1) = 60% and P(Major 2) = 60% and P(Minor 1) = 90% and P(Minor 2) = 90%, we get P(Conclusion) = P(Jerusalem had a public library) = P(JL) = 0.6 x 0.6 x 0.9 x 0.9 = 0.29 = 29%. Thus the conclusion appears to be less probable when we get more evidence, which clearly cannot be correct. Therefore we need Bayes' Theorem, which avoids this fallacious application of probabilities.]

4. Nesting as a Method of Verification

It's easy to determine if a syllogism is valid (just observe if the conclusion follows from the premises), but harder to determine if a syllogism is sound (since that requires all the premises to be true—as a conclusion can only be as true as its weakest premise). A good method to check for errors in the way you derive confidence in your premises (and thus in how you determine a premise is true), is to build out the required syllogisms supporting each premise, nesting one set of syllogisms within the other.

For example:

Major Premise 1: All major cities in antiquity had public libraries.
Minor Premise 1: Jerusalem was a major city in antiquity.
Conclusion: Therefore, Jerusalem had a public library.

Nesting Test for Major Premise 1:

Major Premise 2: If archaeologists and textual historians together find that a large number of major cities in antiquity had public libraries, and have insufficient data to confirm there was any major city that lacked a public library, then all major cities in antiquity had public libraries.
Minor Premise 2: Archaeologists and textual historians together have found that a large number of major cities in antiquity had public libraries, and have insufficient data to confirm any major city lacked a public library.
Conclusion (= MjP 1): Therefore, all major cities in antiquity had public libraries.

Nesting Test for Major Premise 2:

Minor Premise 3: If a large number of representatives of a class have property *x*, and there is insufficient data to confirm any members of that class lack *x*, then all members of that class have *x*.
Major Premise 3: If it is the case that {MnP 3: if a large number of representatives of a class have property *x*, and there is insufficient data to confirm any members of that class lack *x*, then all members of that class have *x*}, then it is the case that {MjP 2: if archaeologists and textual historians together find that a large number of major cities in antiquity had public libraries, and have insufficient data to confirm there was any major city that lacked a public library, then all major cities in antiquity had public libraries}.
Conclusion (= MjP 2): Therefore, if archaeologists and textual historians together find that a large number of major cities in antiquity had public libraries, and have insufficient data to confirm there was any major city that lacked a public library, then all major cities in antiquity had public libraries.

Nesting Test for Minor Premise 3:

Major Premise 4: If there can be no exceptions to a rule {if A, then B} then it is always the case that {if A, then B}.
Minor Premise 4: There can be no exceptions to the rule {if a large number of representatives of a class have property *x*, and there is insufficient data to confirm any members of that class lack *x*, then all members of that class have *x*}.
Conclusion (= MnP 3): Therefore, if a large number of representatives of a class have property *x*, and there is insufficient data to confirm any members of that class lack *x*, then all members of that class have *x*.

Defeat of Minor Premise 4 (hence of Major Premise 1):

There can be exceptions to the rule {if a large number of representatives of a class have property *x*, and there is insufficient data to confirm any members of that class lack *x*, then all members of that class have *x*}.

Correction of MnP 4 (hence of MjP 1):

There can be no exceptions to the rule {if a large number of representatives of a class have property *x*, and those members were effectively selected at random from among all members of that class, and there is insufficient data to confirm any members of that class lack *x*, even though it is somewhat probable we would have such data in at least one instance if many members of that class lacked *x*, then it is at least somewhat probable that any given member of that class has *x*}

RESULT: Through nesting we locate the underlying rule, discover its flaw, and when we correct that flaw, we discover the necessary qualifications and analyses we may have overlooked before. For instance, in this case: (a) we now know we should qualify our premises (and thus our conclusion) as a matter of probability, and a probability less than what we would consider a historical certainty; (b) we now know to ask whether we should even *expect* evidence of major cities lacking public libraries, if any did in fact lack them; and (c) we now know to ask: is the sample of major cities for which we have confirmed public libraries effectively a random sample of all major cities—and if not, will the known bias in sampling affect our conclusion?

As to (b), we might be able to say that if there were *many* such deprived cities, we should have evidence of at least *one* case, whereas if there were only a few, we might not have evidence of that, and as to (c) we might observe that the bias now is in fact against having evidence for the largest of cities (since modern cities often stand on top of the most successful ancient cities, making archaeological surveys spotty at best), and since it is highly improbable that numerous lesser cities would have public libraries while yet greater cities lacked them, the bias is actually *in favor* of our conclusion that all major cities had public libraries.

Therefore, logical analysis can be a useful tool in history: to identify or check against possible errors, by identifying underlying assumptions regarding rules of inference and trends and generalizations in the evidence, and to discover new avenues of analysis, qualification, and inquiry that could improve our methods or results.

5. Syllogistic Representation of Common Historicity Criteria

<u>EXAMPLE 1</u>: The Criterion of Dissimilarity : "If a saying attributed to Jesus is dissimilar to the views of Judaism and to the views of the early church, then it can confidently be ascribed to the historical Jesus"

Major Premise: If any saying *x* attributed to Jesus is dissimilar to the views of Judaism and to the views of the early church, then Jesus said *x*.
Minor Premise: Saying *x* [= Jesus addressing God as his Father] is dissimilar to the views of Judaism and to the views of the early church.
Conclusion: Therefore, Jesus said *x* [= Jesus addressed God as his Father].

Nesting Test for the Minor Premise:

Major Premise 2: If we have no evidence of saying *x* [addressing God as one's Father] from Jews (prior to or contemporary with Jesus) or the early Church (without attribution to Jesus), then saying *x* [Jesus addressing God as his Father] is dissimilar to the views of Judaism and to the views of the early church.
Minor Premise 2: We have no evidence of saying *x* [addressing God as one's Father] from Jews (prior to or contemporary with Jesus) or the early Church (without attribution to Jesus)
Conclusion (MnP): Therefore, Saying *x* [Jesus addressing God as his Father] is dissimilar to the views of Judaism and to the views of the early church.

Etc.

RESULT: If we continue this nesting analysis we will find both MjP 2 and MnP 2 insupportable: because we *do* have evidence of Jews addressing God as one's Father, and it is *not* the case that if we have no evidence of a practice that it did not exist.[1] The Criterion of Dissimilarity thus reduces to an *Argumentum ad Ignorantiam*, a textbook fallacy. Applying the criterion to produce a conclusion of any confidence requires just as much confidence that the practice didn't exist, which is very difficult to establish (one

[1] Cf. Mary Rose D'Angelo, "*Abba* and Father: Imperial Theology in the Contexts of Jesus and the Gospels," *The Historical Jesus in Context*, edited by Amy-Jill Levine, Dale C. Allison, Jr., and John Dominic Crossan (Princeton University Press: 2006): pp. 64-78.

must thoroughly survey all relevant evidence and scholarship, no simple task) and often *impossible* to establish (since we know for a fact there was a great deal more diversity in Jewish beliefs and practice than we presently know any specifics of, and since the survival of sources is so spotty, no valid conclusion can be reached about what *no* Jews ever thought, said, or did).

That invalidates the Minor Premise ("Saying *x* [Jesus addressing God as his Father] is dissimilar to the views of Judaism and to the views of the early church."). But even the Major Premise here will be found indefensible on a complete nesting analysis. "If any saying *x* attributed to Jesus is dissimilar to the views of Judaism and to the views of the early church, then Jesus said *x*" assumes an invalid rule of inference: that only Jesus could innovate. But if Jesus could innovate a saying, then so could anyone, including an actual Gospel author (or other intermediary source)—note how Paul innovated a law-free Gentile mission, and if Paul could do that, so could anyone innovate anything, and we know too little about the many Christians and Jews who lived prior to the Gospels to rule any of them out, yet we would have to rule them all out in order to isolate Jesus as the only available innovator we can credit for the innovation.

For instance, any saying (*x*) we think we can isolate as being unique to Jesus, may in fact be unique to Peter instead, who uniquely imagined or hallucinated or dreamed or invented Jesus saying (*x*). There is no more reason to assume the innovation was of Jesus' own making than of Peter's (whether consciously, for a specific innovative purpose, or unconsciously, as a construct of what Peter took to be visions or revelations, but were actually the product of his subconscious creatively responding to the problems and ambitions of his time). So how are we to tell the difference?

<u>EXAMPLE 2</u>: Multiple Attestation : "If a tradition is independently attested in more than one source, then it is more likely to be authentic than a tradition attested only once."

Major Premise: If a tradition is independently attested in more than one source, then it is more likely to be authentic than a tradition attested only once.
Minor Premise: Jesus healing the sick is independently attested in more than one source.
Conclusion: Therefore, Jesus healed the sick.

RESULT: A nesting analysis will find flaws in both the major and minor premise here.

As to the Minor Premise: Jesus healing the sick appears only in the Gospels and later documents influenced by the Gospels, and not in any of the prior Epistles. The Synoptic and Apocryphal Gospels all show mutual dependence and therefore are not independent sources, and modern scholarship has established that the Gospel of John was probably influenced by that of Mark and Luke, and therefore there are no independent attestations of Jesus healing the sick: that concept appears only in documents ultimately derivative of the Gospel of Mark, which is the earliest mention of the tradition. Therefore the criterion of multiple attestation might not apply.

As to the Major Premise: it is well-documented that even a false claim can be multiply attested in independent sources (e.g. multiple independent sources attest to the labors of Hercules), and folklorists have documented that this can occur very rapidly (there is no actual limit to how rapidly multiple sources can transform and transmit the same story). Therefore the rule of inference this criterion depends on is invalid. A more rigorous rule must be developed that can distinguish between multiple attestation that is caused by the authenticity of what is reported and multiple attestation that is caused by spreading stories of false origin. Such a rule might not be possible, or will be very limited in application because of the extent of the qualifications and conditions that must be met.

A third flaw is that "more likely to be authentic" is vague as to whether (or when) "more likely" means "likely enough to warrant believing it's true." When is the evidence enough to warrant belief? None of the historicity criteria developed provide any help in answering this question. But Bayes' Theorem does.

EXAMPLE 3: The Criterion of Embarrassment : "Since Christian authors would not invent anything that would embarrass them, anything embarrassing in the tradition must be true."

Major Premise 1: Christians would not invent anything that would embarrass them.
Minor Premise 1: The crucifixion of Jesus would embarrass Christians.
Conclusion 1: Therefore, Christians did not invent the crucifixion of Jesus.

Major Premise 2: A report is either invented or it is true.
Minor Premise 2 (= Conclusion 1): The crucifixion of Jesus was not invented.
Conclusion 2: Therefore, the crucifixion of Jesus is true.

Another way to test rules of inference is to try them out on contrary cases. For example:

Major Premise 1: Cybeleans would not invent anything that would embarrass them.
Minor Premise 1: The castration of Attis would embarrass Cybeleans.
Conclusion 1: Therefore, Cybeleans did not invent the castration of Attis.

Major Premise 2: A report is either invented or it is true.
Minor Premise 2 (= Conclusion 1): The castration of Attis was not invented.
Conclusion 2: Therefore, the castration of Attis is true.

RESULT: This is obviously not a credible conclusion. We have no good reason to believe there was ever an actual Attis who was castrated and it is commonly assumed the story was invented for some particular symbolic reason. The same, then, could be true of the crucifixion of Jesus. Tacitus reports that the castration of Attis was indeed embarrassing (it is grounds for his disgust at the religion), yet the castration of Attis is not a credible story, therefore the criterion of embarrassment is in some manner fallacious.

An example within the Christian tradition is the astonishing stupidity of the Disciples, especially in the earliest Gospel of Mark. Their depiction is in fact so unrealistic it isn't credible (real people don't act like that), which means Mark (or his sources) invented that detail *despite* its potential embarrassment. Hence the flaw in the criterion of embarrassment is in assuming that historical truth is the only factor that can overcome the potential embarrassment of some reported detail, when in fact moral or doctrinal or symbolic truth can *also* override such concerns.

For example, Dennis MacDonald argues this attribute emulates the equally-unrealistic stupidity of the crew of Odysseus and thus stands as a marker of the same things that *their* stupidity represented. That may be true. But I also argue it furthers a literary theme found throughout Mark of the Reversal of Expectation.[2] Thus everything that seems embarrassing in Mark might be an intentional fabrication meant to convey a lesson. Mark echoes the gospel theme that “the least shall be first” in his construction of all his stories: although Jesus tells Simon Peter he must take up the cross and follow him, Simon the Cyrenean does this instead; although the pillars James and John debate who will sit at Jesus’ right and left at the end, instead two nameless thieves sit at his right and left at the end; although the lofty male Disciples flee and abandon Jesus, the lowly female followers remain faithful, and as a result the *least* are the *first* to discover that Christ is risen; and while Mark begins his Gospel with the “good news” of the “voice crying out” of the lone man who boldly came forward as a “messenger who will prepare our way,” he ends his Gospel with *several women*, *fleeing* in *fear* and *silence*, and *not* delivering the good news, exactly the opposite of how his book began. So since details that seem embarrassing in Mark might serve his literary intentions, we can’t be certain they’re true.

This final example exposes the importance of testing criteria by comparing them with alternative theories of the evidence. You must ask yourself, *what if I’m wrong?* What *other* reasons might Christians have for inventing potentially embarrassing stories? And how do those reasons compare with the theory that they reported embarrassing stories because they were true? Bayes’ Theorem suits exactly such an analysis.

6. Recommended Reading on Bayes’ Theorem

Eliezer Yudkowsky, “An Intuitive Explanation of Bayesian Reasoning (Bayes’ Theorem for the Curious and Bewildered: An Excruciatingly Gentle Introduction)” at yudkowsky.net/bayes/bayes.html

Douglas Hunter, *Political [and] Military Applications of Bayesian Analysis: Methodological Issues* (Westview Press: 1984).

Luc Bovens and Stephan Hartmann, *Bayesian Epistemology* (Oxford University Press: 2003).

Timothy McGrew, “Bayesian Reasoning: An Annotated Bibliography” at homepages.wmich.edu/~mcgrew/bayes.htm

Wikipedia on “Bayes’ Theorem” (for English: en.wikipedia.org/wiki/Bayes'_theorem)

[2] See my contributions to *The Empty Tomb: Jesus beyond the Grave*, edited by Jeff Lowder and Robert Price (Prometheus: 2005) and my online book *Was Christianity Too Improbable to be False?* (The Secular Web: 2006) at www.infidels.org/library/modern/richard_carrier/improbable.

7. Bayes' Theorem (Complete)

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{[\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]}$$

8. Explanation of the Terms in Bayes' Theorem

See Handout

9. Bayes' Theorem (Abbreviated)

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{P(e|b)}$$

Note that P(e|b) = [P(h|b) x P(e|h.b)] + [P(~h|b) x P(e|~h.b)] so I recommend the longer form instead, as a check against error (especially that of assuming P(e|b) = P(e|~h.b)).

10. Employing Bayes' Theorem

If we use Bayes' Theorem to determine the likelihood that Jerusalem had a public library, and if the following data is the same (note that this model of the problem and these percentages are deliberately unrealistic—realistically they should all be much higher, and the evidence is more complicated):

60% chance that any major city in antiquity had a public library
90% chance that Jerusalem was a major city
60% chance that there was a library if there is a document attesting to a library
90% chance that we have a document attesting to a public library in Jerusalem

P(h|b) = (0.6 x 0.9) + *x* = 0.54 + *x* = The prior probability that Jerusalem was a major city and (as such) would have a public library (based on our background knowledge alone, i.e. what we know of other cities and of the archaeological and textual evidence of the size and success of Jerusalem as a city). The added factor *x* is the prior probability that Jerusalem had a public library even if it *wasn't* a major city. Here for simplicity this is arbitrarily assumed to be zero, but realistically it needn't be. And if it wasn't, *x* would equal 0.1 x *y*, where 0.1 is the probability that Jerusalem *wasn't* a major city (i.e. the converse of the probability that it *was*, i.e. 0.9) and *y* is the probability of a non-major city having a public library (whatever we deemed that to be from available evidence).

P(~h|b) = 1 – 0.54 = 0.46 = The prior probability that Jerusalem did *not* have a public library = the converse of the other prior (i.e. all prior probabilities that appear in a Bayesian equation must sum to exactly 1, no more nor less, because together they must encompass all possible explanations of the evidence for Bayes' Theorem to be valid).

P(e|h.b) = 1.0 = The consequent probability that we would have either some specific evidence of a public library at Jerusalem *and / or* no evidence against there being one. For example, if we assume there is an 80% chance that no evidence would survive even if there was a library there, then there is a 20% chance that some evidence would survive, but this doesn't mean the survival of such evidence drops P(e|h.b) to 20% (as if having no evidence made the library more likely than if we had evidence). For on *h* we can expect *either* no evidence *or* some evidence, so the total probability of having what we have (either none or some) is 100% (since having some constitutes having 'either none or some'). Notably this renders irrelevant the question of whether we actually do have a document attesting to a library, so the probability that we do makes no difference to the likelihood of *h*. The only thing we *don't* expect on *h* is evidence *against* there being a public library at Jerusalem, any of which (no matter how indirect or indefinite) would reduce P(e|h.b) below 100% (to whatever degree was appropriate, i.e. in accordance with how strongly or surely or clearly that evidence argues against a public library there).

P(e|~h.b) = 0.4 = The consequent probability that there would be a document attesting to a public library in Jerusalem even when there wasn't a public library there (i.e. the converse of the 60% chance that such a library existed if we have such a document).

The probability that we don't have such a document can affect P(e|~h.b), by upping it slightly. But that would require a longer digression.

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{[\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]}$$

$$P(h|e.b) = \frac{0.54 \times 1.00}{[\ 0.54 \times 1.00\ ] + [\ 0.46 \times 0.40\ ]} = \frac{0.54}{[\ 0.54\ ] + [\ 0.184\ ]}$$

$$= \frac{0.54}{[\ 0.724\ ]} = 0.746 = 75\% = P(h|e.b)$$

RESULT: Instead of the implausible 29% chance that Jerusalem had a public library, we get the plausible result of a 75% such chance. And this was found with unrealistic percentages that biased the result *against* there being a library there, which means *a fortiori* we can be highly certain Jerusalem had a public library. This illustrates the advantage of using unrealistically hostile estimates against a hypothesis, since if the conclusion follows even then, we can have a high confidence in that conclusion.

11. The Advantages of Bayes' Theorem

1. Bayes' Theorem will help you determine how to tell if your theory is *probably* true rather than merely *possibly* true.

It achieves this (a) by forcing you to compare the relative likelihood of different theories of the same evidence (so you *must* think of *other* reasons the evidence we have might exist, *besides* the reason you intuitively think is most likely), and (b) by forcing you to examine what exactly you mean when you say something is 'likely' or 'unlikely' or is 'more likely' or 'less likely'. With Bayes' Theorem you have to think in terms of relative probabilities, as in fact must be done in all sound historical reasoning, which ultimately requires matching numbers (or ranges of numbers) to your vocabulary of likelihood.

2. Bayes' Theorem will inspire a closer examination of your background knowledge, and of the corresponding objectivity of your estimates of prior probability.

Whether you are aware of it or not, all your thinking relies on estimations of prior probability. Making these estimations explicit will expose them to closer examination and test. Whenever you say some claim is implausible or unlikely because 'that's not how things were done then', or 'that's not how people would likely behave', or 'other things happened more often instead', you are making estimates of the prior probability of what is being claimed. And when you make this reasoning explicit, unexpected discoveries can be made.

For example, as Porter and Thiessen have both observed, it's inherently *unlikely* that any Christian author would include *anything* embarrassing in a written account of his beliefs, since he could choose to include or omit whatever he wanted. In contrast, it's inherently *likely* that anything a Christian author included in his account, he did so for a deliberate reason, to accomplish something he *wanted* to accomplish, since that's how all authors behave, especially those with a specific aim of persuasion or communication of approved views. Therefore, already the prior probability that a seemingly embarrassing detail in a Christian text is in there because it is true *is low*, whereas the prior probability that it is in there for a specific reason *regardless* of its truth *is high*.

3. Bayes' Theorem will force you to examine the likelihood of the evidence on competing theories, rather than only one—in other words, forcing you to consider what the evidence should look like if your theory happens to be false (What evidence can you then expect there to be? How would the evidence in fact be different?). Many common logical errors are thus avoided. You may realize the evidence is just as likely on some alternative theory, or that the likelihood in either case is not sufficiently different to justify a secure conclusion.

For example, Paul refers to James the Pillar as the Brother of the Lord, and to the Brothers of the Lord as a general category of authority besides the Apostles. It is assumed this confirms the historicity of Jesus. But which is more likely, that a historical (hence biological) brother of Jesus would be called the Brother of the Lord, or that he would be called the Brother of *Jesus*? In contrast, if we theorize that 'Brother of the Lord' is a rank in the Church, not a biological status, then the probability that we would hear of authorities being called by that title is just as high, and therefore that Paul mentions this title is not by itself sufficient evidence to decide between the two competing theories of how that title came about.

Estimates of prior probability might then decide the matter, but one then must undertake a total theory of the evidence (extending beyond just this one issue), since there is no direct evidence here as to what was normal (since there is no precedent for calling anyone "Brother of the Lord" as a biological category, and only slim or inexact precedent for constructing such a title as a rank within a religious order).

4. Bayes' Theorem eliminates the Fallacy of Diminishing Probabilities.

Bayes' Theorem requires a total theory of the evidence, as all historical reasoning should, rather than focusing on isolated pieces of information without regard for how they all fit together. But it does this by balancing every term in the numerator with a term in the denominator.

For example, in the public libraries argument we saw that adding two pieces of evidence together reduced the probability of the conclusion, contrary to common sense. It would have been even worse if we had *ten* items of evidence. But in Bayes' Theorem, for every element of P(e|h.b) there is a corresponding element of P(e|~h.b), producing the mathematical result that the more evidence you add, the *higher* the probability of the hypothesis, exactly as we should expect.

For instance, if we had four items of evidence, each 60% likely if *h* is true, on a straight calculation the probability of having all four items of evidence on *h* is 0.6 x 0.6 x 0.6 x 0.6 = 0.6^4 = 0.1296 = 13%, which means the more evidence we have, the less likely *h* is, contrary to reason. But on Bayes' Theorem we ask how likely that same evidence is if *h* is false (and any other hypothesis is true instead). If each item of evidence in our hypothetical case was only 40% likely on ~*h*, then Bayes' Theorem would give us (for simplicity's sake assuming the prior probabilities are equal):

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{[\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]}$$

$$P(h|e.b) = \frac{0.5 \times [0.6 \times 0.6 \times 0.6 \times 0.6]}{[\ 0.5 \times [0.6 \times 0.6 \times 0.6 \times 0.6]\ ] + [\ 0.5 \times [0.4 \times 0.4 \times 0.4 \times 0.4]\ ]}$$

$$P(h|e.b) = \frac{0.5 \times 0.1296}{[0.5 \times 0.1296] + [0.5 \times 0.0256]} = \frac{0.0648}{0.0648 + 0.0128} = \frac{0.0648}{0.0776}$$

$$P(h|e.b) = 0.835 = 84\%$$

RESULT: The more evidence you have, the higher the probability of the hypothesis, exactly as common sense would expect.

5. Bayes' Theorem has been proven formally valid. Any argument that violates a valid form of argument is itself invalid. Therefore any argument that violates Bayes' Theorem is invalid. All valid historical arguments are described by Bayes' Theorem. Therefore any historical argument that cannot be described by a correct application of Bayes' Theorem is invalid. Therefore Bayes' Theorem is a good method of testing any historical argument for validity.

6. You can use Bayesian reasoning without attempting any math (see below), but the math keeps you honest, and it forces you to ask the right questions, to test your assumptions and intuitions, and to actually give relative weights to hypotheses and evidence that are not all equally likely. *But above all, using Bayes' Theorem exposes all our assumptions to examination and criticism, and thus allows progress to be made, as we continually revise our arguments in light of the flaws detected in our reasoning*.

12. Avoiding Common Errors with Bayes' Theorem

1. The Fallacy of False Precision.

One common objection to using Bayes' Theorem in history is that Bayes' is a model of mathematical precision in a field that has nothing of the kind. This precision of the math can create the illusion of precision in the estimates and results. But as long as you do not make this mistake, it will not affect your results.

The correct procedure is to choose values for the terms in the equation that are at the limit of what you can reasonably believe them to be, to reflect a wide margin of error, thus ensuring a high confidence level (producing an argument *a fortiori*), regardless of any inexactness in your estimations.

For example, surely more than 60% of major cities in antiquity had public libraries (the evidence is compellingly in favor of a much higher percentage, provided 'major city' is reasonably defined). But since we don't have exact statistics, we can say that the percentage of such cities must fall between 60% and 100% (= 80% with a margin of error +/-20%). With such a wide margin of error, our confidence level remains high (see appendix). We are in effect saying that we might not be sure it was 100% (or 90% or even 80%), even though we may believe it is, but we *can* be sure it was *no less* than 60%. Since that is the limit of what we deem reasonable, so will our conclusion be (the conclusion is only as strong as an argument's weakest premise, and each probability assigned in a Bayesian equation is the formal equivalent of a premise).

2. The Fallacy of Confusing Evidence with Theories.

A single example will suffice: William Lane Craig frequently argues that historians need to explain the evidence of the empty tomb. But in a Bayesian equation, the evidence is not the discovery of an empty tomb, but the production of a *story* about the discovery of an empty tomb. That there was an actual empty tomb is only a theory (a hypothesis, i.e. h) to explain the production of the story (which is an element of e). But this theory must be compared with *other* possible explanations of why that story came to exist ($= \sim h$, or $= h_2$, h_3, etc.), and these must be compared on a total examination of the evidence (*all* elements of e, in conjunction with b and the resulting prior probabilities).

Hence a common mistake is to confuse actual hypotheses about the evidence, with the actual evidence itself (which should be tangible physical facts, i.e. actual surviving artifacts, documents, etc., and straightforward generalizations therefrom). Though hypotheses can in principle be treated as evidence, this is often only mathematically practical (or non-fallacious) when such hypotheses are so well confirmed as to be nearly as certain as the existence of the evidence that confirms them.

3. The Fallacy of Confusing Assumptions with Knowledge.

Assumptions in, assumptions out. Therefore, assumptions usually should have no place in Bayesian argument, as its conclusions will only be as strong as their weakest premise, and an assumption is a very weak premise indeed.

The term b in Bayes' Theorem establishes that all the probabilities in the equation are conditional probabilities, i.e. probabilities conditional on the truth of our background knowledge. Therefore, only background *knowledge* should be included in b and thus considered in assigning probabilities, not background assumptions or dogmatic beliefs. The difference between professional and unprofessional history is the acceptance in b of only what has been more or less accepted by peer review as an established fact (although the possibility of something can be accepted even when the certainty of that something is not, but in any case the only position that counts as professional when determining background knowledge is the position all experts can agree is acceptable).

This error leads to a common misapprehension that, for example, prior probabilities in Bayes' Theorem are worldview-dependent. They are not. For example, it doesn't matter whether you are a naturalist and believe no miracles exist, or a Christian and believe they do. Either way, if you are acting professionally, you both must agree that so far as is *objectively known*, most miracle claims in history have turned out to be bogus and none have been confirmed as genuine, therefore the prior probability that a miracle claim is genuine must reflect the *fact* that most miracle claims are not, and that is a fact even if genuine miracles exist.

In other words, the naturalist must allow that he could be wrong (so he must grant *some* probability that there might still be a genuine miracle somewhere, whatever that probability must be) and the Christian must allow that most miracle claims are false (not only because investigated claims trend that way, but also because the Christian already grants that most miracle claims validate other religious traditions, and therefore must be false if Christianity is true, and many even within the Christian tradition strain even a Christian's credulity, and therefore are probably false even if Christianity is true). If most miracle claims are false, then the prior probability that any miracle claim is false must be high, regardless of whether miracles exist or whether Christianity is true.

So although worldview considerations *can* be brought into *b*, Bayes' Theorem does not require this. And when such considerations *are* brought into *b*, that only produces conditional probabilities that follow only when the adopted worldview is true. But if a certain worldview is already assumed to be true, most arguments don't even have to be made, as the conclusion is already foregone. Therefore, *b* should include only objectively agreed knowledge (and probabilities assessed accordingly), unless arguing solely to audiences *within* a single worldview community.

NOTE: Further discussion of all these topics (and in considerable detail) will appear in my forthcoming book: Richard Carrier, *On the Historicity of Jesus Christ* (there especially in chapter two and a mathematical appendix). For contact information and more about my work in other aspects of Jesus studies and beyond, see **www.richardcarrier.info**.

Additional Notes for Further Consideration:

A. Bayesian Reasoning without Mathematics

1. You don't have to use scary math to think like a Bayesian, unless the problem is highly complex, or you want clearer ideas of relative likelihood. For instance...

2. Easy Case of Rejection: If you estimate that the prior probability of *h* must be at least somewhat low (any degree of "*h* is unlikely, given just *b*"), and you estimate the evidence is no more likely on *h* than on ~*h*, then *h* is probably false ("*h* is unlikely even given *e*").

EXAMPLE: Talking donkeys are unlikely, given everything we know about the world. That there would be a story about a talking donkey is just as likely if there were a real talking donkey than if someone just made up a story about a talking donkey. Therefore, *e* is no more likely on *h* (Balaam's donkey actually spoke) than on ~*h* (Someone made up a story about Balaam's donkey speaking), and *h* (Balaam's donkey actually spoke) is already initially unlikely, whereas ~h (Someone made up a story about Balaam's donkey speaking) is initially quite likely (since on *b* we know people make up stories all the time, but we don't know of any talking donkeys). Therefore, we can be reasonably certain that Balaam's donkey didn't talk. Note how this conclusion is worldview-independent. It follows from plain facts everyone can agree on.

3. Unexpected Case of Acceptance: If you estimate that the prior probability of *h* must be at least somewhat high (any degree of "*h* is likely given just *b*"), even if you estimate the evidence is no more likely on *h* than on ~*h*, then *h* is still probably true ("*h* is likely even given *e*").

EXAMPLE: That Julius Caesar regularly shaved (at least once a week) is likely, given everything we know about Julius Caesar and Roman aristocratic customs of the time and human male physiology. That we would have no report of his shaving anytime during the week before he was assassinated is as likely on *h* (Caesar shaved during the week before he was assassinated) as on ~*h* (Caesar didn't shave any time during the week before he was assassinated), since, either way, we have no particular reason to expect any report about this to survive. Nevertheless, *h* is probably true: Caesar shaved sometime during the week before he was assassinated.

4. In the above examples, exact numbers and equations were unneeded, just the innate logic of Bayesian reasoning sufficed. This is the case for most historical problems. Only when the problem is complex does math become a necessary tool.

For example, if you want some idea of *how likely* a hypothesis is, then you may need to do some math, unless the relative degrees of likelihood are so clear you can reason out the result without any equation. For instance, it is *very* unlikely that any donkey spoke, therefore it is *very* unlikely that Balaam's story is true. But things are not always this clear.

Or, for example, when the prior probability of a hypothesis is low, but the evidence is still much more likely on that hypothesis than any other (or vice versa), then you need to evaluate how low and how high these probabilities are, in order to determine if they balance each other out.

A flow chart results, wherein the mathematical terms (as given below) can be replaced with ordinary sentences in your native language:

Bayesian Reasoning Flowchart

low P(h|b) --------

-- low P(e|h.b) ---

- P(e|~h.b) is same or higher = low P(h|e.b) = *h* is probably false

- P(e|~h.b) is lower = low P(h|e.b) unless prior is high enough — i.e. For a high P(h|e.b), P(e|~h.b) must be a lot lower than P(e|h.b), in fact the lower P(h.b) is, the lower P(e|~h.b) must then be.

-- high P(e|h.b) ---

- P(e|~h.b) is lower = low P(h|e.b) unless prior is high enough — (see box above)

- P (e|~h.b) is same or higher = low P(h|e.b) = *h* is probably false

high P(h|b) -------

-- low P(e|h.b) ---

- P(e|~h.b) is same or lower = high P(h|e.b) = *h* is probably true

- P(e|~h.b) is higher = low P(h|e.b) unless prior is high enough — i.e. For a high P(h|e.b), P(e|~h.b) must not be higher than P(e|h.b) by too much, but the higher P(h.b) is, the higher P(e|~h.b) can be and still not be too much.

-- high P(e|h.b) ---

- P (e|~h.b) is higher = low P(h|e.b) unless prior is high enough — (see box above)

- P(e|~h.b) is same or lower = high P(h|e.b) = *h* is probably true

"Low" and "High" = lower than 0.5 (50%) and higher than 0.5 (50%), respectively; when P(h|b) = 0.5, so does P(~h|b): then the hypothesis with the higher P(e|b) is probably true.

"Prior is high enough" = when P(h|b) is higher than the Bayesian ratio between either P(e|~h.b) and P(e|h.b) or vice versa, enough to overcome the gap and thus produce the opposite result.

B. The Application of Probability Theory to Bayesian Reasoning

1. All the rules of probability theory and statistical analysis apply to Bayesian calculations, so the more you know about the former, the more you can do with the latter.

   For example:

2. Probability estimates in history are usually intuitive and thus inexact, therefore to ensure a high confidence level (i.e. a strong enough certainty to warrant designating your conclusion an objective result of historical inquiry), you must adopt a wide margin of error.

   Since in probability theory (e.g. in scientific polls) the wider the margin of error, the higher the confidence level, if you widen your margin of error as far as you can reasonably believe it possible to be, given the evidence available to all expert observers, then your confidence level will conversely be such that you cannot reasonably believe the conclusion is false given the evidence currently available to all expert observers. That is the highest state of objective certainty possible in historical inquiry (which is not absolute certainty, but reasonable certainty, which can change with new data).

For instance, you may see a poll result that says 20% of teens smoke, but in a footnote you see 97% at +/-3%. This means the data entail there is a 97% chance that the percentage of teens who smoke falls between 17% and 23%. The first number (97%) is the confidence level, the second (+/-3%) is the margin of error. Given any set of data, raising the confidence level widens the margin of error (and vice versa) according to a strict mathematical formula. So if you lack scientifically precise data, you can compensate by setting your margins of error as wide as you can reasonably believe them to be (e.g. you may think it unreasonable from your own personal knowledge and consultation with others that the percentage of teen smokers could be anything above 33%, which is in effect saying you are highly confident, perhaps 99% certain, that the percentage can be no more than that, but it could easily be much less than that).

3. If a sample of a population is genuinely random, the statistical results are deductively certain (not inductively inferred, but literally undeniable, insofar as the data are accepted as given), but even if the sample is not random, formal techniques exist to weight the data to account for any known or presumed bias in the sample.

You may have heard the example in inductive logic of drawing beads from a barrel. If you draw a hundred red beads in a row, you may conclude the barrel is full of only read beads. But if the bottom half is full of blue beads, your sampling procedure will have misled you. However, if you sample beads from the whole barrel *at random*, then you are pulling beads randomly from every part of the barrel. As long as you are sampling at random, then if you draw a hundred red beads in a row, it is necessarily the case that there are very probably no other beads in the barrel (a precise confidence level and margin of error can be calculated), or extremely few such beads (the margin of error indicates how many beads could still be some other color). Random sampling is thus far more powerful than arbitrary sampling. But if you can only sample from the top of the barrel and yet you know there was a color bias in the way the barrel was loaded, then you can adjust the math to account for this. For example, if you know the barrel was loaded in such a way that would have heavily biased the top half with red beads, then finding ten red beads in a row from sampling the top will produce a mathematically low or even negligible confidence level in the conclusion that the barrel is full of red beads. But if the bias was weaker, if for example there could not have been any more than a 2:1 bias (twice as many red beads ended up near the top than elsewhere in the barrel, if there were any other beads at all), then even after taking this into account, drawing 100 red beads in a row from the top would be highly improbable unless there were only red beads throughout the barrel. Thus, random samples are maximally powerful, but even biased samples can give us highly confident results, if the bias isn't extreme and we have some idea of what it could be.

4. There are numerous statistical fallacies and statistical illusions (where the correct result violates common sense intuitions). The more of these you are aware of, the better your reasoning will be.

An example of a pertinent statistical fallacy arises from the Talpiot tomb discovery (the so-called "Jesus Tomb"), where it was claimed that a particular conjunction of five names was highly improbable, but the fact that there were ten burials in that same tomb was ignored, a serious error. The probability of getting five specific names in a sample of five is indeed low, but the probability of getting those same five names in a sample of *ten* is much greater, *especially when you don't know what the other names are*. For example, if 1 in 4 people were named Mark, and you picked three people at random, the odds that they would all be named Mark would be $0.25^3 = 0.016 = 1.6\%$, in other words very unlikely. But if you picked ten people at random, the odds that at least three of those ten were named Mark would be the converse of the probability of there being less than three Marks in those ten (i.e. the probability of finding only 0 Marks, plus the probability of finding only 1 Mark, plus the probability of finding only 2 Marks), or 1-[(10!/2!8!)(0.25^2)(0.75^8)+(10!/1!9!)(0.25^1)(0.75^9)+(10!/0!10!)(0.25^0)(0.75^10)] = 1-[(45)(0.0625)(0.1001)+(10)(0.25)(0.0751)+(1)(1)(0.0563)] = 1-[0.2816+0.1877+0.0563] = 1 - 0.5256 = 0.4744. In other words almost 50/50, *which is very likely*. If there is almost a 50/50 chance of at least three Marks in a tomb of ten, finding three named Mark there is not improbable at all.[1] This is the kind of statistical fallacy you need to be aware of if you decide to employ statistical logic in your historical method.

5. The application of all these facts to historical reasoning is limitless.

For example, estimating how many early Christians would likely have been literate, using abundant but spotty data about ancient literacy. The sampling isn't random but isn't inordinately biased, and the bias can be accounted for, while widening margins of error can increase confidence level in your final estimate.

1 The probability of three names chosen at random from those ten then being these Marks will be less than this (and differently so for observing three Marks if five names are selected). But in the Talpiot case, the accident of which names remain unmarked is irrelevant to the hypothesis being tested (and in any case requires an even more complex equation to account for it).

Simple Tutorial in Bayesian Reasoning

by Richard Carrier

Preparation

1. To practice the principles of Bayes' Theorem it can be helpful to start with a silly example. For example, imagine you are a teacher and a student comes to you and begs for an extension on their homework because their homework was eaten by their dog. Setting aside all questions of actual policy, assume you genuinely care whether the student is telling the truth. Should you believe the student? What would it take to convince you?

2. Write down a brief list of the kinds of things (alone or in conjunction—specify which) that would indeed convince you your student was telling the truth. Then write down a brief list of the kinds of things that would make you suspicious or convince you your student was lying.

3. Probabilities are usually expressed as percentages, but they can also be expressed as numbers, by dividing the corresponding percentage by 100. So 100% = 1.0, 20% = 0.2, 1% = 0.01, etc. From here on out it is assumed the reader knows all the other rudiments of mathematical procedures and notations.

4. Bayes' equation (in full) reads:

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{[\ P(h|b) \times P(e|h.b)\ ] + [\ P(\sim h|b) \times P(e|\sim h.b)\ ]}$$

5. Note the form of this is:

$$P = \frac{A}{A + B}$$

i.e. the term above (i.e. A) is repeated identically at the bottom, where it is added to another term (i.e. B), so when you input the numbers for A, they are input the same, both above and below.

Defining the Hypothesis

1. First you must define the hypothesis to be tested. In this case, the hypothesis being tested is the claim "The dog ate my homework" (in that specific context, from that specific student). Therefore:

h = "the dog ate the student's homework"

(= the student's claim is true)

2. Then define the negation of this hypothesis. This is what is the case if the hypothesis is false, which must be what would explain the same evidence if that hypothesis were false. There may of course be many alternative explanations. There are different ways to handle that circumstance. We'll bypass this by sticking with the artificial assumption that there is only one alternative:

$\sim h$ = "the student lied about the dog eating her homework"

(= the student's claim is false)

3. Note that $\sim h$ could be generalized (to include all alternative explanations) as simply (and obviously) "The dog did not eat the student's homework," but as this includes no explicit explanation for any of the evidence (e.g. why the homework is missing, or how any other evidence came about that the student may have presented in support of his claim), it has limited explanatory value. This risks becoming a straw man. Specific theories of the evidence are more robust as explanations and thus pose a greater challenge to any hypothesis being tested. It is therefore always a good idea to think of what the most plausible alternative explanation of the evidence may be (or several, if there are several, and, say, rank them according to what you deem to be their relative plausibility).

Determining Prior Probabilities

1. The prior probability is a *relative* measure of what is *typical*. It is a measure of what is *typical* in that it draws on what you know about the world, the context, etc. (and students in general, this student in particular, etc.). It is a *relative* measure because its value is determined relative to all other explanations, and not in isolation from them. Thus, all possible explanations must be examined to determine the prior probability of any.

2. Hence the prior probability of *h* ("the dog ate the student's homework") is *not* the frequency with which dogs eat homework (which is certainly a very, very low frequency and thus would entail a very, very low prior probability) nor even the frequency with which missing homework went missing because dogs ate it (which is a higher frequency than the former, but still a very low one—since the category of missing homework is narrower and includes all cases of dogs eating it, but even still this is not a typical cause of homework going missing).

3. Rather, the prior probability of *h* ("the dog ate the student's homework") is the frequency with which *claims* of dogs eating homework are actually *true* (since we are dealing in this case with a student making a claim, rather than, e.g., a prediction whether a dog will eat that student's homework). Although the frequency of dogs eating homework does affect the frequency of such claims being true, these two frequencies are not necessarily identical (since one must also factor in the frequency with which students tell lies to get extensions on homework they haven't finished, etc.).

4. This is called assigning a 'reference class'. The reference class in this case is "all claims of homework being eaten by dogs" (more strictly, "all missing homework claimed to have been eaten by dogs"). These claims can be generated by various causes (such as by lying, by being mistaken or misled, or by the claim actually being true), each one would correspond to its own exclusive hypothesis.

5. We are testing two hypotheses ("the dog ate the student's homework" against "the student lied about the dog eating her homework") only because we are assuming all other possible explanations can be ruled out (i.e. assigned a prior probability so low it wouldn't even show in our math and thus can be safely ignored—of course, *logically impossible* explanations by definition have a prior probability of zero and thus can *always* be ignored).

6. A Bayesian equation must include all possible explanations of the evidence (in some fashion, somewhere). The prior probabilities of all those explanations must sum to 1 (i.e. 100%)—no more, no less—and this must show in the equation. Strictly speaking, in our silly example this means we would have three numbers that must add to 1: the probability that (dogs did eat) + the probability that (student is lying) + the probability that (every other explanation for the evidence—the evidence consisting at the very least of the lack of homework and the student's claim regarding what happened, but it could include other things, like things on those lists you drew up). In other words:

$$P(\text{ate}) + P(\text{lied}) + P(\text{other}) = 1$$

But we are assuming P(other) is too small to care about—let's say, the sum of *all* of them only comes to 1 in 100,000 (or 0.001% or 0.00001). In that case, P(other) is so close to 0 it won't make any difference, so we just exclude it and assume:

$P(ate) + P(lied) = 1$

7. An easy way to think about this is to draw pie charts. The pie (the circle) represents the total probability space (all 100%, i.e. there is a 100% chance that the evidence we have is explained by *something* occupying space in that pie). So we carve it up to represent the relative likelihood of different explanations, based on what we know to be typical.

8. For example, suppose we think that when students claim dogs at their homework, *typically* they are lying, but you would grant that's not *always* the case. Maybe occasionally such a story really is true. We could model this judgment by saying the number of claims of dogs eating homework that are actually *true* is only 1 in 10 (don't worry for now why we should settle on that number—we're just ballparking it this time, we'll get to the problem of determining numbers later). So only a tenth of the pie is devoted to that, while the rest, or 90%, is then occupied by students lying.

9. That makes for a prior probability of:

$P(h|b) = 0.10$
and
$P(\sim h|b) = 0.90$

Since $P(h|b) + P(\sim h|b)$ must = 1, and $0.10 + 0.90 = 1.00$, we can see that with these numbers we'll have a valid Bayesian equation.

10. If your background knowledge was such that you had more trust in your students, or there was a genuinely well-documented plague of dogs suddenly eating homework, or this particular student you know to be so honest you would actually be *surprised* if she were lying, even about dogs eating her homework, then you would divide the pie differently. If you had more trust in general student honestly, but were still skeptical of this sort of claim, you might say 1 in 4 claims of dogs eating homework are true, in which case $P(h|b) = 0.25$ and $P(\sim h|b) = 0.75$. In the last case, if you would be genuinely surprised, then the odds flip all the way over, e.g. $P(h|b) = 0.80$ and $P(\sim h|b) = 0.20$ (i.e. the claim being *true* is then initially a lot more likely, instead of the other way around).

11. Ideally, you would work from perfect knowledge (you would have a completely reliable database of all dog-ate-my-homework claims and whether they were true or false), from which you could confidently derive precise numbers (the exact ratio of claims in that database being true and false). But such ideal conditions (just as in physics) never exist in real life. So what you are doing is trying to make as accurate a guess as possible what the true frequency would be in such a database. The more information you have, the better your guess will be. The remainder can be accommodated by using margins of error, i.e. the highest and lowest you can believe the frequency to be (if you could get a peek at the hypothetical database).

> Even more ideally (and sometimes necessarily), the database you are approximating is not that of all actual occasions in history (of claims that dogs ate homework), but all hypothetical occasions as well. For example, if there was only ever one claim in history of a dog eating homework, and it was true, it would not follow that the prior probability of further such claims being true is 100%. If we imagine all possible occasions in which homework would go missing and this claim then made, we would surely find in that imagined collection numerous cases of lying, in fact possibly far more. Thus, we must often use our background knowledge to *model human and other behavior* in trying to guess at the true frequencies of various phenomena. But again, the more we know, the better at this we will be.

12. If you run a full Bayesian equation with the highs and the lows (in the right way), you will get a result that is also a high and a low (e.g. "There is a 10% to 40% chance the student is telling the truth"). In that case, you won't be able to say where within that range the likelihood really is, but you *will* be able to say that you cannot reasonably believe it to be any less than 10% or any more than 40%. As long as your margins of error represent the limits of what you can reasonably believe (given what you know), so will your conclusion represent the limits of what you could reasonably believe (given what you know).

14. It is not always necessary to run the equation multiple times to attend to both ends of all margins of error. If your aim is to assess the likelihood your student is telling the truth, for example, you need only use the values at the furthest extremes favoring your student's lying. For if you get a result that the student is probably telling the truth *even with the most unfavorable estimates*, then you needn't run the numbers on the other end of the margins, since that will only get you a result *even more strongly* supporting the student (hence the first calculation produces the same conclusion *a fortiori*). But if you *do* get a result of the student lying when using the most unfavorable estimates, you will need to run the other numbers to see where the other margin falls in your conclusion—if the result is then still on the side of lying even with the *most* favorable estimates, you have an argument *a fortiori* for lying, but if the new result is different, you'll be left with an inconclusive result.

For instance, you may end up with something like "the probability that the dog really ate the homework is somewhere between 30% and 60%," which means you don't really know—maybe you will be leaning towards it not being true (this result can be expressed as 45% true +/-15%), but you can't be sure, and thus are left a 'dog-ate-her-homework' agnostic. On the other hand, if you run with just the most *un*-favorable numbers and get a result of "the probability that the dog really ate the homework is between 60% and (some higher value not yet calculated)" you don't have to calculate that other value (unless you still have reason to know what the upper margin is), since you know it won't be any lower than 60%.

15. When you genuinely don't know what is typical, or have no idea what would be typical (i.e. if you had no idea whether it was common for students to lie or dogs to eat homework), then the prior probabilities are equal, i.e. P(h|b) = 0.50 and P(~h|b) = 0.50, if you would agree there is as much chance of one hypothesis being true as the other (margins of error can still be assigned depending on how uncertain you are of even *that*).

16. All that aside, we will assume for our example that P(h|b) = 0.10 and P(~h|b) = 0.90 (our first estimates, i.e. we estimated that typically 90% of all 'dog ate my homework' stories are false, and only 10% are true). Therefore:

$$P(h|e.b) = \frac{0.10 \times P(e|h.b)}{[\ 0.10 \times P(e|h.b)\ ] + [\ 0.90 \times P(e|{\sim}h.b)\ ]}$$

Determining Consequent Probabilities

1. There are two prior probabilities in the Bayesian equation (which together sum to 1, covering a complete probability space) and two consequent probabilities (which do not have to sum to 1 or any other value, as they are completely independent of each other). The first consequent probability is that of having the evidence we do if *h* is true, the second consequent probability is that of having the evidence we do if *h* is false. Those four values complete the Bayesian equation. They are in effect four premises in a formal logical argument, so if you can defend these premises, you can defend the conclusion.

2. If there is no reason to expect any evidence of the incident other than the missing homework (and the student's story), then the probability of having the evidence we do if *h* ("the dog ate the student's homework") is true is 100% (because we have exactly what we expect and nothing more nor less), as high as can be, and therefore:

$$P(e|h.b) = 1.00 \qquad \text{(common case)}$$

3. What would it take to lower this value? If any of the evidence we have were contrary to expectation. For example, if you think there was a small chance (say, 20%) that your student, given her story, could at least produce scraps of dog-mangled homework as evidence, then your student's inability to produce that evidence is somewhat contrary to expectation, and therefore the absence of this expected evidence drops the probability of all the evidence we have by as much as the missing evidence is expected, in other words by 20% (or 0.20). Therefore:

$$P(e|h.b) = 1.00 - 0.20 = 0.80 \qquad \text{(case with small incongruity)}$$

This represents the fact that the evidence we have is at least slightly less probable than we would expect if the student's claim were true, but not greatly so.

4. If this expectation is stronger, e.g. a dog may have completely consumed paper, but if the student claims the dog mangled her computer, there definitely would be a mangled computer for evidence, and the student's inability to present it for inspection even when asked may be very unlikely indeed—or if, for example, the student claims it was thrown in the garbage and the garbage was already picked up, but you check and confirm that garbage on that student's street is not scheduled to be picked up for several days yet (which would be evidence the student is lying to cover up the lack of evidence for the first claim)—then you may deem the evidence we have (in this case, a particular item of missing evidence) to be very unlikely. You might say there is a 90% chance (at least) that this student could produce the mangled computer (for example), in which case the failure to do this lowers the consequent probability thus:

$$P(e|h.b) = 1.00 - 0.90 = 0.10 \qquad \text{(case with strong incongruity)}$$

5. Look over the list you drew up of evidence that would convince you the student was lying, and think how any of it would lower this probability (and then think how all of it together would do so), by thinking about how unexpected that evidence would be if the student were actually telling the truth.

6. If the student is lying, typically there would be no reason to expect any evidence of this besides the missing homework (and the student's story), in other words *exactly the same evidence* we would often expect if the story were true (so prior probabilities will decide the matter, because the evidence can't—unless we get more evidence, hence the role of investigations and inquiries, research and interrogations, etc.). So again, the probability of having the evidence we do if *h* ("the dog ate the student's homework") is *false* is also 100% (because we still have exactly what we expect and nothing more nor less), which again is as high as can be. So:

$$P(e|\sim h.b) = 1.00 \qquad \text{(common case)}$$

7. What would it take to lower this value? Again, if any of the evidence we have were contrary to expectation. But now this means if any evidence were contrary to expectation if the hypothesis ("the dog ate the student's homework") is *false*, i.e. (in our simplified case) if any evidence were contrary to expectation if the student lied. For example, if the student's parent attests to having witnessed the incident, that would be contrary to expectation if the student was lying, although the parent could still be covering for their child (they aren't necessarily a disinterested party). So you may deem the parent's testimony to have some but only small weight. For example, if you think there is only a

33% chance the parent is telling the truth, then the evidence is only contrary to expectation by that much:

$$P(e|\sim h.b) = 1.00 - 0.33 = 0.67$$ (case with some support)

8. Stronger supporting evidence would generate a stronger factor. For example, if a trusted colleague of yours, who has no connection to the student (and thus no interest in the matter), attests to witnessing the incident, or there is a video recording of it, etc., you may think there is only a 5% chance he would lie (or the video was forged or misleading you, etc.), in which case there is only a 5% chance you would have the evidence you do (missing homework, student's story, *and* trusted colleague's eyewitness report), which means quite simply:

$$P(e|\sim h.b) = 0.05$$ (case with strong support)

9. So assuming the common case (no evidence or expectations of evidence beyond the missing homework and the story, and using the prior probabilities concluded with above):

$$P(h|e.b) = \frac{0.10 \times 1.00}{[\,0.10 \times 1.00\,] + [\,0.90 \times 1.00\,]} = 0.10$$

= There is only a 10% chance the student is telling the truth (in this case relying solely on your background evidence regarding both student honesty and frequencies of dogs eating homework, i.e. how often stories of dogs eating homework are typically true)

10. Assuming the case with small incongruity (some expectation that the student should be able to present more evidence but can't) but some support (parent's testimony):

$$P(h|e.b) = \frac{0.10 \times 0.80}{[\,0.10 \times 0.80\,] + [\,0.90 \times 0.67\,]} = 0.117 = 0.12$$

= There is still only a 12% chance the student is telling the truth.

11. Assuming the case with small incongruity (some expectation that the student should be able to present more evidence but can't) but strong support (trusted colleague's testimony):

$$P(h|e.b) = \frac{0.10 \times 0.80}{[\ 0.10 \times 0.80\ ] + [\ 0.90 \times 0.05\ ]} = 0.64$$

= There is now at least a 64% chance the student is telling the truth (demonstrating the power of evidence you trust to overcome even a strong contrary expectation).

12. You can run the numbers for other combinations of possibilities as above. Or go straight to the list you compiled of evidence that would convince you the student was telling the truth, and generate numbers and run the math for any of those items, then for all of them, by thinking about how unexpected that evidence would be if the student were lying.

Lessons Regarding Consequents

Lesson 1: If evidence would support one side or the other but isn't expected, then its absence does not affect either consequent probability, and its presence only affects one (not both) consequent probabilities. For example, if there are only three items of evidence (missing homework and student's story as one, and note from parent and presentation of a mangled computer as two more, and nothing else, either pro or con), then the consequent probability of *all* this evidence remains 100% (it can't go any higher), which is the same as if you had only one of these (still 100%), so long as you *expected* nothing else. The addition of any more evidence thus only lowers the consequent probability of the other value (the consequent probability on ~*h*). Only when it is in fact *unlikely* that you would have more evidence, does the probability of having it go down, e.g. having *all three* items of evidence (listed above) is less probable on ~*h* than having only one or two of them.

Thus consequent probabilities are about how *expected* the evidence is (including how expected the *absence* of certain evidence is). Unexpected evidence that turns up thus only affects the hypothesis on which that additional evidence is actually unlikely. You may not expect a student to produce your trusted colleague's testimony (as there is rarely any expectation that he would just happen to have been present at just the right time, etc.), so producing it is unexpected on *h*, and is unlikely on *b* (in the sense that such a coincidence is unlikely, given all you know about the world, etc.), but it is not unlikely on *h* (i.e. there is nothing about the student's claim that entails such a colleague's testimony wouldn't be available), so it has no effect on P(e|h.b). Technically it would, but it would have *exactly the same* effect on P(e|~h.b), which cancels out, so the natural improbability of such evidence (i.e. the unlikelihood of a trusted colleague just happening to be in a position to witness the event) can be ignored. Of course, that's not so if other evidence makes such a coincidence suspicious (i.e. you have evidence your colleague was set up), but that's a much more complex example.

Lesson 2: When there is good evidence for a hypothesis, it is 'good' simply because it makes other hypotheses less likely, because those hypotheses have a harder time explaining how that evidence came about. Therefore, evidence supporting *h* lowers the probability of ~*h* (more specifically, it lowers the probability of *e* on ~*h*). When there is evidence *against* a hypothesis, it is 'against' it simply because it makes other hypotheses more likely, because the tested hypothesis has a harder time explaining that evidence than they do. Therefore, evidence against *h* lowers the probability of *h* (more specifically, it lowers the probability of *e* on *h*). Which also means evidence *for* an alternative hypothesis is evidence against *h*, but *only* if that evidence is relevantly unexpected on *h*. If, instead, it is just as predicted by *h* as by the alternative, then it is no longer evidence for the alternative—it is then *inconclusive* evidence. Likewise, evidence believed to support *h* is also, in fact, inconclusive (and thus doesn't actually support *h* in any notable way) if that same evidence is just as likely on an alternative hypothesis (~*h*).

§

Fully understanding both lessons, and why they are true, is one of the keys to understanding why Bayes' Theorem works and how to think like a Bayesian even without the math. Equally key is a full understanding of why prior probability is not just the raw frequency of such things happening (i.e. such things happening as are proposed by *h*), but the relative frequency of different explanations of the same evidence.

Advancing to Increasingly Objective Estimates

This has been a simple example of the mechanics of Bayes' Theorem. In reality you will want a more informed and carefully considered estimate for each of the four probability values that must go into any Bayesian analysis. In other words, the question must be asked, "How do you get those values? Do you just pull them out of your ass?" In a sense, yes, but in a more important sense, no.

You aren't just blindly making up numbers. You have some reason for preferring a low number to a high one, for example (or a very low one to one that's merely somewhat low, and so on), and the strength of that reason will be the strength of any conclusion derived from it—by the weakest link principle, i.e. any conclusion from Bayes' Theorem will only be as strong as the weakest premise in it (i.e. the least defensible probability estimate you employ). So to make the conclusions of a Bayesian analysis stronger, you must make history more scientific, and the degree to which you fall short of that is the degree to which your conclusions (as a historian) carry less weight and certainty than the conclusions of scientists. We already concede that the conclusions of historical analysis carry less weight and certainty than the conclusions of science (often very much less), so we should expect there to be a reason this is so, and that reason happens to be this very one: that we can't get the kinds of data to produce the kinds of probability estimates that scientists can. That is no more an argument against using Bayes' Theorem in history than it is an argument for dismissing all history as unknowable. In other words, this is no more a defect of Bayesian method than of all historical methods whatever.

But we don't have to be lax. The fact that we can improve the certainty of our conclusions by improving the certainty of our estimates of likelihood and unlikelihood is precisely what historians need to learn from Bayes' Theorem. They cannot avoid the consequences of not learning this lesson by avoiding Bayes' Theorem. Their conclusions will still have just as feeble a foundation—unless they do something about it, the very thing Bayesian reasoning teaches them needs to be done.

There are four basic stages of increasing 'scientific' precision in historical analysis, i.e. four steps in that direction. Completing each step improves the reliability of any numbers you enter into the equation, and thus narrows their margins of error and raises your confidence level (such progress also exposes and corrects errors). Historians should endeavor to move their estimates up these levels as far as they can, and exactly what specific research is needed to do this is one of the very things that modeling our arguments in Bayesian terms will teach us.

In addition to this, of course, is the process of peer review and debate, in which any actual errors or mistakes we make, or evidence or knowledge we overlook, is pointed out to us, leading us to revise our numbers and math, and thus continually hone our conclusion toward what is least assailable, most defensible, and ultimately capable of persuading a consensus of fellow experts (which should always be our final aim).

Stage 1: Initial Best Guess

This is the first step, similar to what we did in the simple example above (except, using your own estimates instead of the made-up ones I employed). We use our informed expert intuitions to make a best guess given all we know.

Stage 2: Refined Guess

After making our initial best guess, we can spend time thinking about it, and the evidence, and hypotheses, and background knowledge, and test out why we would reject different estimates than ours—or discovering that we would not. The first discovery (why we would reject other particular estimates) shores up our case for maintaining the estimates we do by exposing and delineating the evidence that supports our estimates, thereby making our 'premises' stronger (more defensible, less contestable). The second discovery (that we wouldn't in fact reject certain other estimates) will start to uncover our confidence interval (i.e. the margins of error we should be working with), thus allowing us to start building an *a fortiori* argument for our conclusion, by using only those margins that work most against our conclusion (or at least revealing to us what our conclusion must then be given those margins).

Stage 3: Initial Best Data Survey

The ideal would be to know all possible cases (actual and hypothetical, but considering for the moment only actual), in order to know what was in fact typical and what the relative frequencies really are. For example, if we actually knew how many 'dog ate my homework' stories were really true and how many really false, throughout all history, then we would have a concrete, extremely reliable estimate of prior probability. But such ideal knowledge is rare.

The first step toward it, however, especially for historians, is to make your best effort to gather as many cases as you can. All available cases may be too numerous or difficult to find, but you can gather a large number, as many as are easily culled. Then you can generate numbers based on a formal analysis of this sample, treating it as a random sample of all cases in history (a random sample with, if necessary, certain selection biases, whatever biases would be produced by how you 'found' or 'chose' your data, as well as any produced by how that data was selected for survival, and weighting your estimates according to how strong these biases are). Basic statistics can then be run on the data, complete with precise margins of error and calculable confidence levels (any decent introductory college textbook on statistics will teach you everything you need to know about how to do that). This will produce a more reliable result than even your best refined guess (and thus serves as a good check on the reliability of your expert intuitions and the completeness of your evidence and background knowledge).

Stage 4: Refined Data Survey

If you can complete your data collection (so that you have, or can reasonably expect you have, every available known case that is relevant), you can produce an even stronger result than with only a best effort survey. For example, estimating the frequency with which miracle claims are false could benefit from having a complete database of all miracle claims that have been investigated to a reliable conclusion (vs. miracle claims that could not be accessed by any reliable investigation or were not so accessed or were accessed but the needed records of the inquiry were not preserved, etc.). There may be even more refined ways to secure a reliable statistical estimate of the whole population (i.e. all comparable instances in history, known and unknown, examined and unexamined) using the available sample (i.e. all known and examined instances) as a random sample (or a random sample with known and accountable biases). And of course, scientists can control and expand access to their data in ways historians rarely or never can, so even stage 4 does not rise entirely to the level of scientific certainty. But it reaches as high a reliability as any historical argument ever will.

For Future Learning

1. Read about Bayes’ Theorem and find and study examples of actual applications of it.

2. Any argument you make, read, or hear can be modeled with Bayes’ Theorem (if the argument is valid and sound, although if it isn’t, Bayesian modeling will also expose that fact, too). So to practice and thereby develop more understanding of Bayes’ Theorem, a good plan is to develop a Bayesian analysis of any empirical argument you find convincing (or unconvincing) in order to see how that everyday argument would be represented correctly in Bayesian terms. With enough practice at reverse-engineering standard arguments into Bayesian form, you will be better equipped to directly engineer Bayesian arguments from the ground up.

3. You don’t always have to use Bayes’ Theorem in publications. You can use Bayes’ Theorem to model and thus test your own (more colloquial) arguments ‘behind the scenes’ as it were, in anything you plan for publication, just to see if this produces insights that you can convert for use in your planned publication (such as discovering overlooked consequences of certain assumptions, or additional arguments you can add to and thus improve your case, or flaws in your reasoning you can correct, or the need to be more moderate or less confident in your conclusions, etc.).

For more help and advice see: http://www.richardcarrier.info/bayescalculator.html