VALIDATION WARRANTS ARE INVERSE PROOFS

Abstract. Verification has a long history of development in science, engineering, and mathematics. Validation has garnered less overt attention except among philosophers. We show that validation and verifications are related by developing validation methods from the proof associated with a model.

“A habit of basing convictions upon evidence, and giving to them only that degree or certainty which the evidence warrants, ... [would] cure most of the problems in the world.”
Bertrand Russell.

1. Introduction

Our goal is to satisfy Russell’s dictum by developing computational relationships between logical proof and validation. These would be extension of concepts described by Pólya [7] and Jaynes [3].

1.1. Why a New Term? The terms verification and validation are well-defined in the literature. Technically, verification (testing) is an act of justification. Justification includes logical proof among its methods. Justification methods can be described as impersonal, universal, self-sufficient, and definitive [7] and definitely finite.

Validation is an infinite process. However, every time a successful experiment is conducted, we are more confident that the process as described is correct. Validation shares with plausible or credible reasoning the traits of impersonal in method but personal in strength; universal but having weight of evidence; self-sufficient but not durable. Validation is not definitive, but provisional and ephemeral [7].

There is no commonly accepted term to describe one act of validation, like proof is for justification.

Note. The term accreditation is loosely understood within the U.S. military establishment where it is taken to mean an administrative action certifying that a model or simulation meets a particular need. One would hope that accreditation occurs only after suitable warrants are found.

2. The Background

Current technical literature such as Oberkampf and Trucano [6] and Roache [8] focus on standard statistical processes. Science and engineering models often have well developed historical attachments to particular statistical models; indeed, many statistical methods were derived out specific needs in science. Unfortunately, such highly evolved physical models bear little or no resemblances to such things as “people-in-the-loop” simulations. This makes the application of statistical methods difficult and open to severe criticisms.

In order to describe this dicotomy between engineering and non-engineering uses, we classify systems in two ways. Natural systems are systems over which we have no control, such as physical processes. Synthetic systems are systems over which we have some or total control such as command, control, and communications systems or shop floor simulations. There is obviously a classification of hybrid systems that incorporate both natural and synthetic systems.

We focus on synthetic systems in this paper. The major difference between the two is the role of logic. While logic can be made to play a strong role in scientific systems [9], its role is pervasive in defining synthetic systems. While we do not intend to “prove” properties of synthetic systems in practice, the role of logic makes for a great theoretical tool.

2.1. Laplace Said It All. Not only to understand our approach, but to also understand that the problem has been around for 176 years, we turn to Laplace’s (1749–1827) seminal work on probability, Essai Philosophique sur les Probabilités [5].

To the non-practioners of probability and statistics, the subjects can appear arcane, almost mystical. As pointed out above, probabilistic reasoning is quite different form the deterministic methods of classical
science. The subject has a long history: Jakob (James) Bernoulli (1654–1705) is credited with the first treatise in 1713; Thomas Bayes (1702–1761) and his (in)famous “Bayes Law” were first published in 1764. Laplace accepted Bayes principles in 1781. It was Boole (1815–1864) who began the controversy over Bayes’ principles in his treatise *Laws of Thought*.²

Laplace gives two principles.

1. **Principle of Sufficient Reason.** “The connexion between present and preceding events is based on the evident principle that a thing cannot come into existence without there being a cause to produce it”

2. **Principle of Indifference:**

   The theory of chances consists in reducing all events of the same kind to a certain number of equally possible cases, that is to say, to cases whose existence we are equally uncertain of, and in determining of the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all possible cases is the measure of this probability, which is thus only a fraction whose numerator is the number of favourable cases, and whose denominator is the number of all possible cases.”

History has shown that those who stray from these principles get in trouble [3].

Laplace also understood the difficulties in using probabilities to understand Nature. (See Appendix A.) There have been great debates concerning the correct interpretation of these ideas which tends to cloud the true issues. Some of these positions are

- There are two views of probability — both described by Laplace. One is the frequentist account that focuses on distributions and the degree of belief account that calls for subjective probabilities. We will not discuss these issues.
- Karl Pearson proposed that probabilities and statistics cannot deal with causality, even though Laplace demands it.
- Axiomatic probability theory dominates probability and statistical practice. Measure theory dominates the textbooks making the subject more intractable to the practitioner.
- There are the Bayesians and the Classical accounts of probability. The Bayesian account is more modern, even though Thomas Bayes died before Laplace was twenty.

Our account relies on some names not really associated with the standard theories: Sir Harold Jeffreys [4], George Polya [7], and Edwin T. Jaynes. While Jeffreys (1891–1989) and Polya (1887–1985) are well-known, a word of introduction about E. T. Jaynes (1922-1998). Professor Jaynes was an accomplished physicist, publishing several books and over 70 papers. His lifelong interest was in probability. His online text in probability is must reading [3]. The current work builds on the work of Polya and Jaynes.

By their account, probability is (1) Bayesian and (2) an extension to classical logic. We will bypass the Bayesian-classical controversy entirely.

## 3. Pólya’s Concepts of Credibility

Pólya’s idea is that credibility follows the laws of probability but we are interested in the qualitative, not quantitative, relationships.

### 3.1. Fundamental Method

Let us assume that $A \rightarrow B$ and we are able to observe $B$. We wish to investigate the relationship $A|B$; that is, what plausibility is gained in $A$ by finding $B$? By elementary probability methods:

$$P(B) = P(AB) + P(\bar{A}B),$$

$$= P(B|A)P(A) + P(B|\bar{A})P(\bar{A}),$$

$$= P(A) + (1 - P(A))P(B|\bar{A}).$$

In the last line, we used the fact that $P(B|A) \equiv 1$ since $A \rightarrow B$. From Bayes’ Theorem $P(B|A)P(A) = P(A|B)P(B)$ we can use solve for $P(A|B)$ and substituting Eq (1) we get

$$P(A|B) = \frac{P(A)}{P(B)} = \frac{P(A)}{P(A) + (1 - P(A))P(B|\bar{A})}.$$  

Given this equation and taking $P(A)$ as a constant what qualitative result can we present about $P(A|B)$?

¹ *An investigation into the Laws of Thought, on Which are founded the Mathematical Theories of Logic and Probabilities.*
As $P(B|\bar{A})$ goes from 1 to zero,
$P(A|B)$ goes from $P(A)$ to 1.

In other words, as we move from $\bar{A} \rightarrow B$ towards $A \rightarrow B$, our credence of (the plausibility of) $A$ increases and the closer to impossible the first relationship becomes the faster our credence in $A$ increases.

3.2. Other Relationships. Pólya [7] investigates a great number of these relationships. It is not our purpose here to rederive his results; rather to bring these results to the V&V community. Pólya’s results seem to be lost to the education of scientists, engineers, and even mathematicians.

In Example 5 of [7, p. 132], Pólya discusses caveats to this method (which equally well apply to probability and statistics). This example derives credibility in terms of likelihoods (ratios of probabilities) and admonishes about reading too much into these methods. In particular, he admonishes the non-statistician into reading too much into these probabilistic ideas.

4. Good’s Formulation

I. J. Good popularized the concept of subjective probabilities with his 1950 text *Probability and the Weighing of Evidence*. Papers germane to his views were published in 1983 as an anthology [1]. He invented the notation $(H : E|G)$ to mean the evidence given by the data $E$ for the hypothesis $H$ given the knowledge base $G$. Note that $G \neq \emptyset$.

Good proposes measures for (1) information as a measure of relevance; (2) evidence which can be for or against; (3) corroboracion; (4) causality; and (5) the weight of evidence. The two most interesting to this discussion are information which is defined as

$$\log \frac{P(E|FG)}{P(E|G)}$$

and weight of evidence given as

$$\log \frac{O(E|FG)}{O(E|G)} = I(H : E|G) - I(H : E|G).$$

Briefly, he proposed several measures:

**Information, Evidence Definitions here from Good’s paper**

5. Probability Theory *qua* Logic

Jaynes published his first paper outlining his concepts of probability theory as logic in [2] and continued to work on these ideas until his death in 1998. The physics department at Washington University continues to work on these ideas [3]. Jaynes makes two substantial contributions:

1. The Mind Projection Fallacy. This fallacy considers the tenancy of humans to insist that Nature is as they want it, rather than as it truly is. The fallacy takes on two forms:

   (A) My own imagination implies a real property of Nature
   (B) My own ignorance implies that Nature is indeterminate

Much of the paper shows how this plays out in probability and statistical theory. Often the fallacy is used to confuse *physical* independence with *logical* independence.

2. Probability Theory as Logic. By this, he means that we should only accept probabilistic models built up from the standard sum and product rules plus Bayes’ rule. [2] considers several examples; [3] is far more complete but unfinished. The difference of this view from the traditional one is explained by

“...[conventional probability theory] allows only sampling distributions, interprets them as physically real frequencies of ‘random variables’, and rejects the notion of probability of an hypothesis as being meaningless. We take the opposite position: that the probability of an hypothesis is the fundamental, necessary ingredient in all inference, and the notion of ‘randomness’ is a red herring, at best irrelevant.” [2].

Jaynes viewpoint is decidedly non-platonic, closer to constructive mathematics than to conventionally taught Plutonic mathematics. Here is the difference between orthodox probability/statistics and Jaynes’ position:
"The philosophical difference between conventional probability theory and probability theory as logic is that the former allows only sampling distributions, interprets them physically as real frequencies, and rejects the notion of probability of an hypothesis as being meaningless. We take the opposite position: that the probability of an hypothesis is the fundamental, necessary ingredient in all inference, and the notion of "randomness" is a red herring, at best irrelevant." [2]

The central issues are (1) slavish adherence to the view of Nature’s physical behavior and (2) inclusion of only relevant relationships and (3) evidence is accumulated over time.

Our view is that validation is the inverse process to mathematical justification (proof). The basic tenets are

1. The only real models are those that predict states.
2. Given state descriptions, we can derive logical statements about models either scientific [9] or synthetic.
3. From the proofs, we have proposed causality relationships.
4. The variables and processes in the conclusion are thereby linked to variables and processes in the hypotheses.
5. We define evidentiary processes using Bayesian concepts whereby we can compute the weight of evidence and compare it to a goal. These processes are developed by “inverting” the proof in a way made clear below.

### 6. Validation as Inverse Proof

In order to obey Russell’s dictum, we move to the following “maxim”:

“To form a judgment about the likely truth or falsity of any proposition \( A \), the correct procedure is to calculate the probability that \( A \) is true:

\[
P(A | E_1 E_2 \ldots)
\]

conditioned on all the evidence at hand.” [3].

For this discussion let

\[
X = \text{Prior information, which is really context} \\
H = \text{Some hypothesis to be tested} \\
D = \text{Observed data}
\]

then we seek formulas of the form

\[
P(H | D \land X)
\]

To make the notation a bit cleaner, we use Good’s notation [1]

\[
P(H : D | X)
\]

which is read “the probability of \( D \) supporting \( H \) given information \( X \).”

The major criticism of Jaynes’ approach is that it requires can require a separate model. That is, we are forced to posit a probabilistic model ab initio, which means we have to validate the connection of the original to the new one. This is obviously an infinite regress problem.

Instead, we suggest that we can develop the probabilistic model from the original model. This is particularly clear in synthetic models, since they tend to be logical models, too. In conventional probability, the models are proposed as encodings of boolean operators of sets, namely sum, product, and conditional operations for boolean inclusive or, and, and implication. Our approach is to concentrate on a different set of operations: the rules of inference used to derive the proof. Using the rules of inference establishes the direction of causality and the inverse of these rules establishes the probability. The central operation in proofs is that of modus ponens or the law of detachment given below in modern dress:

\[
A \rightarrow B \quad A \quad MP
\]

What should inverse be? Bayes theorem suggests that

\[
P(H : D | X) = P(D | HX) P(H | X) = P(H | DX) P(D | X).
\]
From this, we can derive Russell’s dictum:

\[ P(H : D|X) = P(H|X) \frac{P(D|HX)}{P(D|X)} \]

The last factor, \( P(D|HX)/P(D|X) \), is the likelihood function \( L(H) \). Even though the above derivation is elementary, the likelihood factor is a crucial part: likelihoods are not probabilities but rather scaling factors. For example, let \( y(D) \) be a non-negative function of \( D \). Then \( L(H) = y(D)P(D|H_1X) \) is equally deserving of the name likelihood. Therefore, problem dependence enters through likelihoods.

The inverse should give the probability of \( A \) in terms of the probability of \( B \) and the implication. For this paper, we use the following idea. Let \( \{A \rightarrow B, A_1 \rightarrow B, \ldots, A_n \rightarrow B\} \) be formulas, with the \( A_i \)'s disjoint. Then

\[ P(A : B|X) = P(B|X)\rho(A \rightarrow B) \]

where \( \rho(A \rightarrow B) \) is the percentage of the time we would expect \( A \) out of all the cases of \( \{A, A_1, \ldots, A_n\} \).

7. Moving on to Odds and Evidence

Consider now the simplest possible decision criteria for a validation: we will accept a hypothesis as true if we find that the probability for the hypothesis is greater than the against the hypothesis. In other words,

\[ \frac{P(H|DX)}{P(\overline{H}|DX)} \]

Another way to look at this is to consider the odds of the hypothesis as being true:

\[ O(H|DX) = \frac{P(H|DX)}{P(\overline{H}|DX)} \]

Therefore, we have from (Eq 7)

\[ O(H|DX) = \frac{P(H|DX)}{P(\overline{H}|DX)} = \frac{P(H|X) \frac{P(H|DX)}{P(H|X)}}{P(\overline{H}|X) \frac{P(\overline{H}|DX)}{P(\overline{H}|X)}} = O(H|X) \frac{P(H|DX)}{P(\overline{H}|DX)} \]

We need two things: how to measure evidence and who to compute a value for this function. The two are closely related.

7.1. Definition of Evidence. Following Jaynes, define evidence by

\[ e(H|DX) = b \log_b O(H|DX) \]

where \( b \) is conventionally taken as 10 in engineering applications (decibels). For computer applications we would probably prefer to measure evidence in bits so the above becomes

\[ \varepsilon(H|DX) = \log_2 O(H|DX). \]

7.2. Computing Evidence. The computational rule now becomes the obvious. From (Eq 5) we get

\[ e(H|DX) = e(H|X) + \varepsilon(H|DX) \]

and the initial value would be zero since \( \log_b 1 = 0 \). That is, since the odds for a hypothesis before any data is taken as equally likely.
For illustrative purposes suppose we have a logical derivation of a fact. Our purpose is to show the relationship between a logical model and the validation method.

If Company R has entered into a contract and that contract is legal (L) and if he has performed (P) the contract then Company J will win the lawsuit. If Company R had not accepted (~A), then the contract is not legal (~L). Company R had not accepted. Therefore, Company J will not win (~J).

This has a simple logical model:

\[ R \land L \land P \rightarrow J \]

\[ \neg A \rightarrow \neg L \]

\[ \neg A \]

\[ \neg J \] (8)

Suppose now we want to validate this logical model. We would scour the newspapers for instances in which the hypotheses are true and then consider the conclusion.

**Note 1.** There are two independent facts in this story: ~J and ~A. For simplicity sake, let us assume that ~A is not in dispute and we observe both J and ~J. What does this say about the legality of Company R’s into the contract (the proposition L)? There are five propositional variables R, L, P, A, J. By truth table methods, then, there are 2^5 = 32 possible observations. The ratio of the counts of values for ~J and those for J is the likelihood L(J/~J). This likelihood can now be used to conduct a validation experiment.

**Note 2.** The proof is no proof at all! This proof is the trap of denying the antecedent (hypothesis).

Using our formulas, there are four possible ways to consider the evidence as shown in Figure 1; these factors are not independent. We choose to validate (L|JX).

We can now develop a plan to validate our model. Every article that fits the hypothesis can be used as an element of data. We can continue the process until the evidence “converges”. Just as with convergence of infinite series, the investigator must choose the level of accuracy; See Fig 2. When the investigation converges, we make a decision.
For each warrant, we have the following

\[ \varepsilon(L|JX) = 10 \log_{10} \frac{P(J|LX)}{P(J|L)} = 10 \log_{10} \frac{2/3}{14/32} = -8.45 \]

\[ \varepsilon(L|\bar{J}) = 8.45 \]

Evidently, if the two situations \((L|JX)\) and \((L|\bar{J}X)\) are equiprobable, then the sequence converges to zero. This is exactly what should happen.

Suppose we find the following situation. In a particular jurisdiction there are eleven courts. We find that one court (\(\alpha\)) acquires one case in three when the bad condition \(L \land \beta\); the other ten courts (\(\beta\)) have a rate of one in six. The intent of the original model was that observing conditions \(L\) (and the two others) should result in a judgement for \(J\). The complement of bad will be taken as good = \(L \land \bar{J}\). Can we identify the system out of compliance?

The probabilities are

\[
P(bad|\alpha X) = 1/3 \quad P(good|\alpha X) = 2/3
\]

\[
P(bad|\beta X) = 1/6 \quad P(good|\beta X) = 2/3
\]

Our factor matrix this time is

\[
\varepsilon(\alpha|badX) = 10 \log_{10}(1/3)/(1/6) = 3.103 \approx 3
\]

\[
\varepsilon(\alpha|goodX) = 10 \log_{10}(2/3)/(5/6) = -0.9691 \approx -1
\]

It is clear in this case that the evidence for the batch coming from \(\alpha\) is increased by 3 for every bad decision and decremented by one for each good one. Suppose we wanted to be 99 percent assured that a batch came from \(\alpha\) before saying anything. If there are \(N\) total cases for which \(n_g\) are good and \(n_b\) are bad judgements, then we see that we need

\[ 3n_b - n_g \geq 20 \]

to attain this assurance.

9. Conclusions

We have explored the concept of validation as the inverse process to verification. This was based on the interpretation of probability given by Laplace, Jeffreys, Pólya, and Jaynes; probability is an extension to normal logic. Our organizational principle is that of the system theory.

References


Appendix A. Laplace on Plausible Reasoning in Science

“In matters that are only probable, the difference in the data that each man has on them is one of the principal causes of the diversity of opinions found to hold on such matters. ....

“It is thus the same matter recounted before a large crowd of people, finds various degrees of belief according to the extent of the listeners’ knowledge. If the man who reports it is deepley convinced of it [i.e., the truth], and if by his calling and character he inspires great confidence, his accoun, however extraordinary it may be, will have the same degreee of likelihoood {or plausibility} for ignorant listeners as an ordinary matter repored by the same man, anhd they will believe it impicitely.... and the matter will be judged false by well-infomred listenrs who deem it inconsistent, wither with well-authenticated matters or with the immutable laws of nature.

....

It is to the influence of the viewpoint of those whom popular opinion judges the best informed, and to whom has been accustomed to give its trust in the most important matters of life, that the propogation of those errors is due that in times of ingnorance, have covered the face of the earth.”