The Laplace Rule of Succession Under A General Prior

Kalyan Raman
University of Michigan in Flint
School of Management
Flint, MI 48502
May 2000

Kalyan Raman is an Associate Professor at the University of Michigan in Flint, School of Management, Flint, MI 48502. The author is grateful to Professor Malay Ghosh (Department of Statistics, University of Florida) for his useful comments on an earlier draft of this paper and suggestions for improving it.

The Laplace Rule of Succession Under A General Prior

1. Introduction

"Even in the mathematical sciences, our principal instruments to discover the truth are induction and analogy" (Laplace, 1814)

Scientific knowledge depends heavily on inductive reasoning. Indeed inductive reasoning is required not only in science but also in routine decision making. Much of our ability to learn from experience is based on inductive processes. The mathematician George Polya regards both induction and analogy as particular cases of plausible reasoning, and develops these rules using the basic axioms of probability theory (1954, 1968). Inferences made on the basis of plausible reasoning are inadmissible in traditional deductive logic--no number of previous instances of a rule can conclusively prove that the rule will hold in a new instance, because the formal possibility of an exception can never be eliminated.

Laplace (1814) appears to have been the first person to link induction with probability. His work led to the controversial Rule of Succession. This rule has been critiqued or modified by many authors (Keynes 1929, Pearson 1907, Jeffreys 1961, Polya 1968), and is described in many texts on probability, e.g. Ross (1985) is an excellent source. The Laplace rule of succession has been linked to the inductive process by many authors (Pearson 1907, Jeffreys 1961); in particular, Polya (1968, p.133-135) gives a succinct discussion of the issues. The major objections raised against making such a linkage include the assumptions of (A) constant probability of success p on each trial, with no updating of p, (B) a uniform prior for p, and (C) independent trials. I provide an approach here which (A) updates the probability after each outcome; hence the probability varies as a function of current information, (B) places no restrictions on the prior apart from continuity requirements, and (C) shows that the general induction probability is a linear combination of the Laplacian probabilities. This representation suggests that there are cases where the uniform prior assumption can work well, in the sense that the substantive implications of the Laplace rule for the inductive process remain invariant under some general priors. Hence, this generalization could potentially serve as a basis to analyze the implications of inductive inference for replications in scientific experimentation (Raman 1994).

2. Linking Induction With Probability

Laplace's attempt to link induction with probability is often described in terms of random drawings from urns. In a simple version of the problem, an urn contains green and red balls in unknown proportions. A drawing of n balls with replacement yields k green and n-k red balls. What is the probability that an additional drawing will result in a green ball? A particular case of this problem may be interpreted as the fundamental issue about inductive inference reduced to its simplest terms. Suppose that k = n, i.e. the first n drawings all yields green balls. What is the probability that the next drawing will also be green?

Laplace gave a solution by using a uniform prior for the proportion p of green balls. Under this assumption, the solution to the problem is straightforward. Let u_n denote the probability that if n drawings yielded only green balls, the next one is also green, given a **uniform** prior distribution for p. Then:

$$u_n = (n+1)/(n+2) \qquad (1)$$

This is the controversial Rule of Succession, proved in many elementary texts on probability.

2.1 Laplace's Rule of Succession

Laplace's Rule of Succession has been applied rather indiscriminately in situations where its underlying assumptions are unlikely to be met. A famous example due to Laplace himself involves computing the odds of the sun rising tomorrow, given that it always rose in the past. The problem with such applications is that the rule was derived on the assumption of independent drawings with the same probability of success on each draw (Ross 1985, p.118). Both these assumptions are often violated in situations like the one described above. Furthermore, when the probability of success p is unknown, and is therefore assumed to be a random variable, each outcome gives information about p. When this information is used to update the distribution of p, the probability varies across trials.

In Laplace's day, it was his uniform distributional assumption for the unknown probability p that provoked the most heated debates. Many scientists and mathematicians felt that this was an arbitrary assumption, arguing that not knowing the probability of an event did not justify regarding it to be as likely as other events whose probabilities were also unknown. The debate on this issue remains somewhat mixed today. Geisser (1984) presents a compelling case for the uniform prior used by Laplace. Similarly, Berger (1985, p.90) makes a strong case for non-informative priors and suggests that Laplace's approach has much to commend it. On the other hand, Zellner (1984) cautions that the form of the prior does matter, particularly in small-sample or extreme situations. Indeed, Geisser (1980) reviews Jeffreys' (1961) analysis of the Laplace model, in which the latter shows that the Laplace prior can be very unsatisfactory in some cases. These conflicting views signal the desirability of considering the Laplace rule under a general non-uniform prior. My analysis will show that some of the substantive implications of the Laplace rule for induction are inherited by rules under general priors.

2.2 A Generalized Rule of Succession

Assume that n experiments of a theory are to be performed. Before the first experiment, assume that our prior probability p that the theory is true follows a continuous distribution g(p), which we express by writing p ~ g(p). By the Weierstrass Approximation Theorem (Simmons 1963, p.154), we can always find a polynomial to approximate a continuous function to any desired degree of accuracy. Therefore, let g(p) be represented by $\sum_{i \geq 0} a_i p^i$, such that the approximation error is less than an arbitrarily small ε. Let g_n denote the probability that if n experiments were successful, the next one is too, given that the prior distribution for p is g(p). My purpose here is to derive g_n and establish a simple relationship between g_n and u_n, the corresponding probability under an uniform prior.

Define random variables X_i=1 if the i-th trial is a success and zero otherwise. Let p_i = P{X_i=1}. Then p_1 has the same distribution as p before the first experiment is performed. The result of the first experiment yields either X_1=1 (success) or X_1=0 (failure). After observing the outcome, the distribution of p_1 is updated. The probability p_2 that a repetition of the experiment will be a success is now drawn from this updated distribution and the outcome for X_2 is observed. After observing the outcome, the distribution of p_2 is updated. This procedure is repeated until the n experiments have been performed. We now ask: given that {X_i=x_i,i=1,..,n}, where each x_i is either 1 or 0, what is the probability that the (n+1)th experiment, if it is performed, will be a success? In other words, given any history of previous confirmations and refutations of our theory, what is the probability that the next experiment will be in its favor? The particular case in which all x_i, i=1,..,n, are 1 is the induction probability g_n; i.e. the probability that, given that n experiments confirmed a theory, the (n+1)-th experiment will produce a confirmation.

2.3 Derivation of g_n

We obtain $g_n = P[X_{n+1} = 1|\{X_i=1\}, 1 \quad i \quad n]$ as a special case of the more general probability $P[X_{n+1} = x_{n+1}|\{X_i = x_i\}, 1 \quad i \quad n]$.

$$p_1 = p \sim \sum_{i \geq 0} a_i p^i \text{, and } P[X_1 = x_1|p_1] = p_1^{x_1}(1-p_1)^{1-x_1}$$

Therefore the joint likelihood of X_1 and p_1 is proportional to

$\sum_{i \geq 0} a_i B(p_1|i+1+x_1,\ 2-x_1)$, where $B(.|\alpha,\ \beta)$ denotes a Beta density function with parameters α and β, and x_1 is 0 or 1; therefore the updated density of p_1 is a

$\sum_{i \geq 0} a_i B(p_1|i+1+x_1,\ 2-x_1)$ density. The probability that the second experiment is a success, i.e. p_2, is drawn from this density. Therefore, given that $X_1 = x_1$, the joint likelihood of X_2 and p_2 is proportional to $\sum_{i \geq 0} a_i B(p_2|i+1+x_1+x_2, 3-x_1-x_2)$, where each of x_1 and x_2 is 0 or 1, and therefore the updated density of p_2 is a $\sum_{i \geq 0} a_i B(p_2|i+1+x_1+x_2, 3-x_1-x_2)$ density. Continuing this updating process, it is clear that the conditional distribution of p_{n+1} is $H_n(p_{n+1}) = \sum_{i \geq 0} a_i B(p_{n+1}|\alpha_i, \beta_n)$, where $B(p_{n+1}|\alpha_i, \beta_n)$ denotes the Beta distribution for p_{n+1}, with parameters α_i and β_n, where $\alpha_i = i + 1 + \sum_{j=1}^{j=n} x_j$ and $\beta_n = n + 1 - \sum_{j=1}^{j=n} x_j$. A result in DeGroot (1975, p.265) assures us that this sequential updating of g(p) is equivalent to updating g(p) after observing all the events $\{X_i=x_i\}$, $1 \quad i \quad n$. Since $g_n = E(p_{n+1})$, and $p_{n+1} \sim H_n(p_{n+1})$, it follows that g_n is given by: $g_n = \sum_{i \geq 0} a_i(i+1+\sum_{j=1}^{j=n} x_j /(i+2+n))$.

Therefore, if k of the previous experiments were successful, it follows that $\sum_{i=1}^{i=n} x_i = k$, and $g_n = \sum_{i \geq 0} a_i(i+1+k)/(i+2+n)$. If the first n experiments were all successes, i.e. if k = n, then we obtain the following generalization of the classic rule of succession:

$$g_n = \sum_{i \geq 0} a_i\ (i+1+n)/(i+2+n) \tag{2}$$

The above expression gives the expected probability of success for the (n+1)th experiment, given that the first n were all successful and given our prior beliefs about the theory. Next, I show that g_n is a linear combination of the Laplacian probabilities u_n.

2.4 Relationship Between g_n and u_n

Recall that u_{n+i} is the probability of an additional success after n+i successes, i = 0,1,.., under the assumption of a uniform prior for the success probability, and therefore each u_{n+i} is given by Laplace's rule. Hence for each i=0,1,.., $u_{n+i} = (i+1+n)/(i+2+n)$; but this is precisely the i-th term in the above expression for g_n and so we may write g_n as:

$$g_n = \sum_{i \geq 0} a_i u_{n+i} \qquad (3)$$

Thus, the general probability, given n successes, is a weighted sum of the Laplacian probabilities of an additional success, given n+1, n+2,..successes. The weights a_i arise from the polynomial approximation of the prior density g(p) (e.g. through a Maclaurin series representation). Alternatively, the weights a_i may be subjectively chosen to represent the prior density. In any case, care should be taken to appropriately normalize g(p) and ensure that it is nonnegative for all 0 p 1. For the choice $a_0 = 1$, and $a_i = 0$ for all i 1, we extract Laplace's rule of succession $g_n = u_n = (n+1)/(n+2)$, which is as it should be since g(p) is a uniform distribution with this choice of a_i.

An interesting and useful special case arises through the choice of a_i as the coefficients in the polynomial representation of a $B(p|\alpha, \beta)$ density, i.e., take the prior g(p) to be a Beta density with parameters α and β. In this case, g_n has the following simpler form which I will call b_n:

$$b_n = (\alpha + n)/(\alpha + \beta + n) \qquad (4)$$

In this case, the choice $\alpha = \beta = 1$ gives a B(p|1,1) density for the prior, which is a uniform density, and we recover the Laplace rule. Given the flexibility of the Beta distribution, it is an attractive way of representing a variety of prior beliefs. We now analyze the general rules g_n and b_n to discover their implications for the inductive process.

3. Implications For Induction From The General Rules

The representation of the general induction probability g_n as a linear combination of Laplacian probabilities has interesting implications for the induction process. This analysis shows that the Laplace rule is a special case of both g_n and b_n corresponding to the assumption of a uniform prior. Next, let us examine the assumption of a uniform prior more closely. I will argue that, under some circumstances, this is not a very objectionable assumption either, in the sense that some of the **implications** for the inductive process remain invariant with respect to more general specifications of the prior. Intuitively, this happens because g_n is a linear combination of the u_{n+i}, and therefore inherits some of the latter's properties. I begin by analyzing the implications of the Laplace rule u_n and then show that implications that are fundamental to inductive reasoning do remain invariant with respect to the more general rules g_n and b_n.

The following interesting features of u_n are easily established through straightforward calculus. As n increases, the probability increases ($u_n/\ n>0$); thus the more the number of successful experiments in the past, the more confidently we expect the next experiment to be a success. Furthermore, the probability approaches 1 as n tends to infinity. Thus, we can expect to get closer and closer to certainty by collecting more and more verifications of our theory. Furthermore, although it is true that each new verification adds to our confidence, it adds less and less when it arrives after more and more previous verifications ($^2u_n/\ n^2<0$). These properties are discussed in more detail in Polya (1968). Note that the probability is always less than 1 as long as n remains finite. This corresponds to the generally held notion that a finite number of experiments, no matter how large, can never prove a theory beyond all shadow of doubt. There always remains the formal possibility of a violation to the theory. (To use Popper's analogy (1968), no matter how many white swans we see, the next swan could still turn out to be black). The rule b_n also has all the above properties. The general succession rule g_n offers greater flexibility than b_n in that it permits richer behavior of the limiting probability as n tends to infinity.

Notice that both u_n and b_n approach 1 as n tends to infinity, i.e. we can expect to get closer and closer to certainty by collecting more and more verifications of our theory. Thus, even a theory which is maximally uncertain *a priori*, can be made highly likely to be true by collecting a sufficiently large number of verifications. The rule g_n admits a larger number of possibilities. As n tends to infinity, g_n tends to $\sum_{i \geq 0} a_i$, which can be made less than 1 by suitable choice of the a_i. For example, suppose the prior density for p is $g(p) = 0.832776 + 1.00334p - 1.00334p^2$. It is easy to check that $g(p) = 0$ for $0 \quad p \quad 1$, and $\int_0^1 g(p)\,dp = 1$, so that g(p) is a proper probability density function. Then the modal prior probability of the theory is 0.5 and the limiting probability is 0.832776.

Note that the expected likelihood of the theory being true, before <u>any</u> experiments have been performed (i.e. n=0), is given by $\alpha/(\alpha + \beta)$ under b_n, and by 1/2 under u_n. One of the problems with u_n is that it gives a probability of 1/2 to the **untested** theory. The rule b_n gives sensible results in this case. The probability $\alpha/(\alpha + \beta)$ may be interpreted as the prior belief of the investigator regarding the truth of the theory, i.e. as an assessment of its face validity. For example, the choice of $\alpha = 1/2$, $\beta = 2$ gives a mean prior probability of 0.2 that the theory is true. The rule g_n gives a prior probability of $g_0 = \sum_{i \geq 0} a_i(i+1)/(i+2)$ to the untested theory, whose magnitude depends on the a_i, i.e. on the investigator's prior beliefs. Both the general rules b_n and g_n offer greater flexibility in incorporating prior beliefs.

Writing $g_n = a_0 u_n + \sum_{i \geq 1} a_i u_{n+i}$, and noting that $u_n = (n + 1)/(n + 2)$, we see that the terms under the summation sign and the amount by which a_0 differs from 1 may be regarded as a correction to the Laplace rule, for departures of the prior density from a uniform prior. When a_0 is close to 1 and the a_i, $i \quad 1$, are close to zero, the prior is not very different from the uniform distribution and in that case, the Laplace rule remains approximately valid.

4. Conclusions

I have shown how an inductive argument can be assessed probabilistically in Laplace’s framework without assuming a uniform prior. Banishing the uniform prior amounts to slicing off the hydra's head--it results in more of the Laplacian probabilities, since the rule of succession, under a general prior, is a linear combination of Laplacian rules. Each linear combination corresponds to a particular specification of prior beliefs. Many of the substantive implications of the Laplace rule for induction remain unchanged under some prior beliefs, and so in this sense, the uniform prior works reasonably well. The general rule offers greater flexibility and shows how the investigator's prior beliefs modify her confidence in a theory, given successive verifications of that theory.

REFERENCES

Berger, James. (1985) Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

DeGroot, Morris. (1975) Probability and Statistics, Addison-Wesley Publishing Company, Reading, Massachusetts.

Geisser, Seymour (1980) "Jeffreys' Contributions to Bayesian Inference", in Bayesian Analysis in Econometrics and Statistics:Essays in Honor of Harold Jeffreys, ed. A.Zellner, Amsterdam: North-Holland Publishing Co.,13-34.

Geisser, Seymour (1984) "On Prior Distributions for Binary Trials", The American Statistician, Volume 38, Number 4 (November), 244-247.

Jeffreys, Harold (1961) Theory of Probability (3rd edn.), Oxford University Press.

Keynes, J. M. (1929) A Treatise on Probability, Macmillan and Co., Limited, London.

Laplace (1814), Essai philosophique sur les probabilities, Paris.

Pearson, Karl. (1907) "On the Influence of Past Experience on Future Expectation", Philosophical Magazine (6), Volume 13, 365-378.

Polya, G. (1954) Mathematics and Plausible Reasoning:Volume I, Princeton University Press, Princeton, New Jersey.

Polya, G. (1968) Mathematics and Plausible Reasoning:Volume II, Princeton University Press, Princeton, New Jersey.

Popper, Karl (1968) The Logic of Scientific Discovery, Harper & Row, Publishers, New York.

Raman, Kalyan (1994) "Inductive Inference and Replications: A Bayesian Perspective," Journal of Consumer Research, Vol. 20, March, 633 – 643.

Ross, Sheldon (1985) Introduction to Probability Models, Academic Press Inc., New York.

Simmons, G.F. (1963), Introduction to Topology and Modern Analysis, McGraw-Hill Book Co., London.

Winkler, Robert L. (1972), An Introduction To Bayesian Inference and Decision, Holt, Rinehart and Winston, Inc., New York

Zellner, Arnold (1984) "Comment", The American Statistician, Volume 38, Number 4, (November), 249-250.