H.G.B. Alexander Research Foundation
Graduate School of Business
University of Chicago

Generalizing the Standard Product Rule of Probability Theory and Bayes's Theorem

By

Arnold Zellner

University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637

Paper 0405 **Revised January 2006**
July 2004

Generalizing the Standard Product Rule of Probability Theory and Bayes's Theorem*

by

Arnold Zellner[1]

Graduate School of Business, University of Chicago,

5807 So. Woodlawn, Chicago, IL 60637, USA

Abstract

In this paper the usual product rule of probability theory is generalized by relaxing the assumption that elements of sets are equally likely to be drawn. The need for such a generalization has been noted by Jeffreys (1998, pp. 24-25), among others, in his work on an axiom system for scientific learning from data utilizing Bayes's Theorem. It is shown that by allowing probabilities of elements to be drawn to be different, generalized forms of the product rule and Bayes's Theorem are obtained that reduce to the usual product rule and Bayes's Theorem under certain assumptions that may be satisfactory in many cases encountered in practice in which the principle of insufficient reason is inadequate. Also, in comparing alternative hypotheses, allowing the prior

[1] Tel.: 1 773 702 7145; fax: 1 773 702 0458.
E-mail address: arnold.zellner@chicagogsb.edu

odds to be random rather than fixed provides a useful generalization of the standard posterior odds.

JEL Classification: C1

Key words: Proofs of product rule, Bayes's Theorem, posterior odds, principle of insufficient reason, random prior odds.

I. Introduction

The Bayesian learning model, Bayes's Theorem is widely used in science, industry and government. For example, Alan Greenspan (2004), Chairman of the Board of Governors of the U.S. Federal Reserve System wrote, "…the conduct of monetary policy in the United States has come to involve, at its core, crucial elements of risk management…In essence, the risk-management approach to monetary policymaking is an application of Bayesian decision-making." (p. 37). See also Feldstein's (2004) comments on Greenspan's article in which he states, "Chairman Greenspan emphasized that dealing with uncertainty is the essence of making monetary policy…The key to what he called the risk-management approach to monetary policy is the Bayesian theory of decision-making." (p. 42). Thus Bayesian theory is being used to solve economic policy problems of the kind considered many years ago in path-breaking, innovative work in Rotterdam by Tinbergen, Theil and others using non- Bayesian methods since Bayesian methods did not exist then for solving their policy problems. See, from among many other sources, Berger (2000), Geweke (2005), Koop (2003), Lancaster (2004), Rossi, Allenby and McCulloch (2005) and Zellner (1997) for an overview of modern Bayesian analysis and uses of Bayesian methods to solve many important problems. Also, the websites of the International Society for Bayesian Analysis (http://www.bayesian.org) and the American Statistical Association's Section on Bayesian Statistical Science (http://www.amstat.org) contain much useful information about Bayesian history, meetings, texts, computer programs, and applications.

That Bayes's Theorem or learning model has been employed widely and successfully in important applications is indeed impressive. However, as with all theorems or laws, assumptions are made in their proofs that may not be satisfied in all circumstances. In particular, below attention is focused on an assumption that is employed in deriving the product rule of probability which is usually utilized to prove Bayes's Theorem.

The product rule of probability, a central result, is broadly applied and used in the standard proof of Bayes's Theorem. Thus it is important to appreciate the assumptions employed in the proof of the product rule and when they may not be satisfied, as emphasized by Jeffreys (1998, pp. 24-25) in the following words: "The proof has assumed that the alternatives considered are equally probable…It has not been found possible to prove the theorem without using this condition… But it is necessary to further developments of the theory that we shall have some way of relating probabilities on different data, and Theorem 9 suggests the simplest general rule that they can follow if there is one at all. We therefore take the more general form as an axiom…" (p. 25). From these remarks, it is the case that the product rule is a basic input to Jeffreys's and others's axiom systems for using Bayes's Theorem in learning from data in a systematic and logical fashion. Thus, having the product rule proved under broader assumptions is an important issue.

Of course, Jeffreys was not the first to comment on the assumptions underlying the proofs of the product rule and Bayes's Theorem. As Stigler (1986) points out, "The principle of insufficient reason (to use a later name for the assumption that causes not known to have different a priori probabilities should be assumed a priori equally likely) would therefore be less a metaphysical axiom than a simplifying and approximative assumption to permit calculation….he [Laplace] supposed that if causes under one specification were known not to be equally likely, we would respecify them in such a way that they were equally likely, say by subdividing the more likely causes."(p. 103). He goes on to explain, "In Bayes's work the relationship [the product rule] is derived explicitly and rigorously from first principles, including a rather explicit argument for the assumption of equally likely causes. In Laplace's work the relationship emerges full grown, without proof, and with only a tacit reliance on equally likely causes." (p.104) See also Keynes (1952, Ch.IV) for a lengthy discussion of "The Principle of Indifference," his name for what others have called the Principle of Insufficient Reason or the Principle of Non-Sufficient Reason and Levy (1977) for an interesting discussion of four types of ignorance. However, he does not consider the broadening of the usual "equally likely" representation that is presented below but does remark, "All of these notions [of ignorance] have their uses. It is rather more important to identify them and evaluate their significance for deliberation and inquiry than it is to engage in verbal disputes." (p. 756).

Further, Jeffreys (1998) remarks, "In terms of our fundamental notions of the nature of inductive inference, to say that the probabilities are equal is a precise way of saying that we have no ground for choosing between the alternatives….To take the prior probabilities different in the absence of observational reason for doing so would be an expression of sheer prejudice. The rule that we should take them equal…is merely the formal way of expressing ignorance. It is sometimes referred to as the Principle of Insufficient Reason (Laplace) or the equal distribution of ignorance." (Pp. 33-34). For example, in use of posterior odds to evaluate two alternative hypotheses, it is usual practice to assume that the prior odds on the hypotheses is equal to one when there "…is no good reason for choosing between alternatives," as Jeffreys mentioned above. The use of the prior odds =1 assumption is present not only in Jeffreys's work but also in many others, e.g. Geweke (2005), Koop (2003), Lancaster(2004), Press (2003), Rossi et al.(2005), and many other publications, including my own. In what follows, we shall assume the prior odds to be random and assume that the prior mean of the odds is equal to one, a broadening of the usual "ignorance" assumption that allows for a broader range of possibilities than does the usual prior odds equals one assumption. For example, with two mutually exclusive hypotheses, taking the prior odds equal to one implies that the probability on each hypothesis equals 1/2. Knowing this precise value is much more informative than assuming that the prior odds is random with a mean equal to one. Examples involving prior and posterior odds will be presented below.

Here we remark briefly that taking probabilities associated with different alternatives all equal is a rather extreme form of representing ignorance about alternatives since it does not accommodate possible unknown variation in the probabilities over the hypotheses. Below, we shall allow for unknown differing probabilities to be associated with alternatives and it will be shown that by restricting the means of these probabilities, the usual product rule and Bayes's Theorem can be derived without any approximations, a result that Laplace might regard as "respecifying the probabilities"…see quote from Stigler (1986, p. 103) given above.

In Section II, the standard proof of the product rule of probability and the role that it plays in proving Bayes's Theorem are reviewed. Then in Section III, the assumptions underlying the usual product rule are broadened and more general versions of the product rule and of Bayes's theorem are derived. Assumptions needed for the broadened versions of the product rule and Bayes's theorem to reduce to the standard versions are provided. In Section IV, several examples, including an important example involving cancer data, presented in Yudkowsky (2005) as a guide to medical doctors, are analyzed to illustrate the general principles. Some concluding remarks are presented in Section V.

II. Proofs of the Product Rule and Bayes's Theorem

In the usual or standard proofs of the product rule and Bayes's Theorem, see, e.g., Jeffreys (1998) or Press(2003), (For "non-standard" proofs, that we do not consider herein, see Jaynes (2003, p.24ff.) who presents and discusses Cox's (1961) derivation, Good's (1971) and Bernardo's (1979) maximum expected utility derivation and Zellner's (1988, 2002, 2003) information conservation derivation.), we have sets A, B and the intersection of A and B with n_A, n_B and n_{AB} elements, respectively with $n = n_A + n_B$ as shown in Figure 1. Elements of the sets are assumed independently drawn with probability $1/n$, the same for all n elements. Then the probability of drawing an element of set A is Pr (A) = $n_A x 1/n = n_A/n$ using the addition rule. Further, the probability of drawing an element of B given that we are drawing from set A is Pr (B given A) $= n_{AB}/n_A$. And thus Pr (A) x Pr (B given A) =

$$(n_A/n)x(n_{AB}/n_A) = n_{AB}/n.$$

Similarly, Pr (B) x Pr (A given B) =

$$(n_B / n) x (n_{AB} / n_B) = n_{AB} / n$$

and thus,

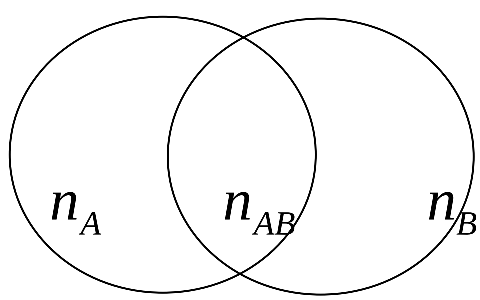
Pr (A) Pr (B given A) = Pr (B) Pr (A given B) = n_{AB} / n = Pr (A and B) (1)

which is the product rule of probability.

Figure 1

Sets and Numbers of Elements in Each*

Set A Intersection of Sets A and B Set B



* $n = n_A + n_B$

Further from (1), we have

Pr (B given A) = Pr (B) Pr (A given B)/Pr (A) (2)

which is a form of Bayes's theorem.

As mentioned in the introduction, the above results depend on the assumption that the unconditional probabilities of drawing elements of the various sets are all the same, namely 1/n. In the next section, this assumption will be relaxed and modified versions of (1) and (2) will be obtained and discussed.

III. Broadened Assumptions, the Product Rule and Bayes's Theorem

Herein we shall assume that the probability of drawing the i'th element of set A is not 1/n but is given by $\delta_{Ai}/n, i = 1, 2, ..., n_A$. Similarly, the probability of drawing an element from the set B is assumed to be $\delta_{Bi}/n, i = 1, 2, ..., n_B$ and from the intersection of A and B, $\delta_{ABi}/n, i = 1, 2, ..., n_{AB}$. With these probabilities, the following results are direct consequences of the addition rule given assumed independent draws:

Pr (A) =

$$\sum_1^{n_A} p_{A_i} = \sum_1^{n_A} \delta_{Ai}/n = (n_A/n)(\sum_1^{n_A} \delta_{Ai}/n_A) = (n_A/n)\bar{\delta}_A \tag{3}$$

Similarly,

Pr (B given A)

$$= \sum_{1}^{n_{AB}} p_{B/Ai} = (n_{AB}/n)\sum_{1}^{n_{AB}} \delta_{B/Ai}/n_{AB} = (n_{AB}/n)\overline{\delta}_{B/A} \tag{4}$$

Then,

Pr (AB) = Pr (A) x Pr (B given A)

$$= (n_A/n)\overline{\delta}_A (n_{AB}/n_A)\overline{\delta}_{B/A} = (n_{AB}/n)\overline{\delta}_A \overline{\delta}_{B/A} \tag{5}$$

and

Pr (AB) = Pr (B) Pr (A given B)

$$= (n_B/n)\overline{\delta}_B (n_{AB}/n_B)\overline{\delta}_{A/B} = (n_{AB}/n)\overline{\delta}_B \overline{\delta}_{A/B} \tag{6}$$

Further, on equating (5) and (6) and solving for Pr (B given A) = Pr (B) x Pr (A given B)/Pr (A), we have

Pr (B given A) = Pr (B) Pr (A given B)/Pr (A) (7)

$$= (n_B / n)\bar{\delta}_B (n_{AB} / n_B)\bar{\delta}_{A/B} / (n_A / n)\bar{\delta}_A$$

$$= [(n_B / n)(n_{AB} / n) / (n_A / n_B)]\bar{\delta}_B \bar{\delta}_{A/B} / \bar{\delta}_A.$$

In the last line of (7), the first factor in square brackets involving the n's is just the traditional result using the equally likely to be drawn hypothesis. The factor involving the $\bar{\delta}$'s is a correction to the usual expression for Pr (B given A) reflecting the assumed different probabilities of elements being drawn. To the extent that this factor departs from the value one, there will be errors involved in using the traditional result as is obvious by writing (7) as

Pr (B given A)/ Traditional Pr (B given A) = $\bar{\delta}_B \bar{\delta}_{A/B} / \bar{\delta}_A$. (8)

Of course, if the ratio on the right hand side of (8) equals one, there is no difference between the traditional results and results derived from the assumption that individual elements have differing probabilities of being drawn. One important case in which the ratio on the right side of (8) is equal to one is that in which each of the average deltas= 1, our "average" principle of insufficient reason, so-called because with this assumption the expressions in (3), (4) and (5) all become equal to their values given that each element has the same probability of being drawn.

It is possible to establish additional results for other assumptions regarding the underlying probabilities, as Laplace suggested; however, some of these alternative assumptions may not meet Jeffreys's goal of attaining a good representation of ignorance when it is needed. In many cases, the assumptions introduced above may be as good a representation of ignorance regarding the validity of alternatives as the equal probability assumption.

IV. Analyses of Applied Problems

A. Cancer Data Problem

In a very interesting and useful paper explaining Bayes's Theorem and Bayesian reasoning in connection with analyses of data relating to cancer incidence, Yudkowsky (2005, pp. 2-3) poses the following problem "…on which doctors fare best of all, with 46%-nearly half-arriving at the correct answer:

100 of 10,000 women at age forty who participate in routine screening have breast cancer. 80 of every 100 women with breast cancer will get a positive mammography. 950 out of 9,900 women without breast cancer will also get a positive mammography. If 10,000 women in this age group undergo a routine screening, about what fraction of women with positive mammographies will actually have breast cancer?"

Here we have set A comprising those 9900 women who do not have breast cancer and set B with 100 who have breast cancer. 950 of set A and 80 of set B, a total of 1030 women test positively. Yudkowsky (2005, p. 3) poses the following problem: "Given the above data [shown in Table 1], what answer should a doctor give to a woman patient with a positive mammography test who asks about the chance that she has breast cancer?"

He states that this chance or probability is given by 80/ (80 + 950) = .078, a result that flows from the "equal probability assumption," that is each individual in the group of 1030 women who tested positively has a probability, say p of having cancer given a positive mammography test result and then the result .078 follows by the addition rule. Or, with y(i) = p + u(i), where y(i) = 1 if cancer is present and y(i) = 0 if it is not, and u(i) a zero mean error term, an estimate of p is 80/1030=.078. Alternatively, one can assume that the probability is p(i) with

$\sum_{1}^{1030} p(i)/1030 = p,$ and then 80/1030 = .078 is an estimate of the mean of the p(i)'s assumed equal to p. The resulting number .078 is identical to that obtained above but has a quite different interpretation in that it allows for possible heterogeneity among women over and above the control for age used in generating these data.

Further, if the probability of having cancer given a positive test result, $P(C|+)$ satisfies $P(C|+) = P(C)P(+|C)/P(+)$, a form of Bayes's Theorem, then we have,

$P(C|+) = [100/10,000]x[80/100]/[1030/10,000] = 80/1030 = .078$ using the "equally likely" assumption. On the other hand, if the probabilities are not assumed equally likely, we have:

$$P(C|+) = [\sum_{i=1}^{100} \delta_{Ci} /10000][\sum_{i=1}^{80} \delta_{C+i} \ /100]/\sum_{i=1}^{1030} \delta_{+i} /10000$$

$$= [(100/10,000)\bar{\delta}_C][(80/100)\bar{\delta}_{+C}]/[(1030/10000)\bar{\delta}_+]$$

$$= 0.078 \text{ x } (\bar{\delta}_C \bar{\delta}_{+C} / \bar{\delta}_+)$$

which reduces to 0.078 if (1) all the $\bar{\delta}'s = 1$ or (2) the ratio involving the $\bar{\delta}'s = 1$.

Table 1

Cancer Incidence and Cancer Test Results for
10,000 Women Aged 40

		Without Cancer	With Cancer	Totals
Test	Positive	950	80	1030
Results	Negative	8950	20	8970
	Totals	9900	100	10000

B. Credit Ratings

We shall reinterpret the data in Table 1 to relate to results of credit checks (poor or favorable) and loan experience (no default or default). Of the 1,030 receiving poor credit ratings, 950 did not default on their loans and 80 did default. Of the 8,970 receiving favorable credit reports, 8,950 did not default on their loans while 20 did default. We can then ask, "Given a person with a poor credit rating, what is the probability of default?" The answer under the "equal probability" assumption is 80/1030=.078, as in the cancer example that can also be obtained without the "equal probability" assumption, as shown above. Similarly, the analysis relating to Bayes's Theorem carried out in the cancer example applies here as well.

C. Economists' Forecasts

Suppose we regard Table 1's data to relate to 10,000 upper cyclical turning point episodes for world economies involving economists' forecasts, downturn (DT) and no downturn (NDT) and to actual outcomes, i.e. DT and NDT outcomes for economies of the world. Assume that of 10,000 turning point episodes, 9,900 are NDT outcomes and 100 are DT outcomes. Also suppose that 1,030 NDT forecasts were followed by 950 NDTs and 80 DTs and that following the 8,970 NDT forecasts, 8,950 NDTs and 20 DTs were observed. As above, we can ask, given a NDT forecast, what is the probability of experiencing a DT? The answer, as considered above in the cancer example, is 80/1030= .078 (Would that this were true!) and other results relating to the cancer example can be carried over to apply to this example.

D. Posterior Odds

As mentioned earlier, another situation in which the "equally likely" assumption is widely employed is in the important problem of comparing and/or testing alternative hypotheses or models using prior odds and Bayes's factors. That is using Bayes's Theorem, see, e.g. Jeffreys (1998, Ch.Vff.), Press (2003, Ch. 9), Geweke (2005, p.16ff), Koop (2003, pp. 4-5), Lancaster (2004, p. 97ff), and Rossi, Allenby and McCulloch (2005, p. 160ff), we have for two hypotheses or models, H1 and H2, the posterior odds (K) equals the prior odds times the Bayes factor, i.e.

$$K = \text{Posterior odds} = \text{Prior Odds x Bayes Factor} \qquad (9)$$

$$= \Pr(H_1|D)/\Pr(H_2|D) = (P_1/P_2)xBF$$

where $\Pr(H_i|D)$ is the posterior probability associated with H_i given the data D, P_1/P_2 is the prior odds and BF is the Bayes Factor, namely , the ratio of the marginal densities for the data, y, $f_1(y|H_1)/f_2(y|H_2)$.

When we wish to be “fair” and/or don’t know much about the alternative hypotheses, it is customary to assume that the prior odds, P_1/P_2 = 1, a precise value which in a gambling situation is quite informative. For example, as mentioned above, if the hypotheses are exhaustive, the prior probabilities are P and 1-P and assuming P/(1-P) = 1 implies that P = 1/2, a rather informative value that leads to K = BF from (9). If, as an alternative to assuming the prior odds = 1, we consider the prior odds, denoted by $P_1/P_2=\varphi$, to be a subjectively random parameter, $0<\varphi<\infty$, with $E\varphi=1$, then K becomes subjectively random and

$$EK = E\varphi BF = BF \tag{10}$$

an estimate that is known to be optimal relative to a quadratic loss function, $L(K,\hat{K})=a(K-\hat{K})^2$. For other loss functions, as is well known, optimal point estimates are available, either analytically or numerically, e.g. the median vis a vis an absolute error loss function, etc.

Thus, (10) provides a result that allows for many possible alternative values of the prior odds subject to the condition that $E\varphi = 1$, a less informative assumption than the prior odds = 1 exactly. Also, with the prior odds random, K becomes subjectively random with moments, $EK^m = E\varphi^m x(BF)^m$.

If it is assumed that we know just that the prior density for the prior odds, φ, $0 < \varphi < \infty$, is proper and has mean = 1, a least informative probability density function (in terms of expected log height relative to uniform measure) satisfying these assumptions is the exponential density, namely, $f(\varphi) = \exp\{-\varphi\}$ which is proper and has mean 1. With this assumption the implied density for the posterior odds, K, given in (9) is:

$$g(K) = (1/b)\exp\{-K/b\} \qquad 0 < K < \infty \tag{11}$$

an exponential density, where b = BF, the Bayes Factor. The moments of K are readily available, namely $EK^m = b^m xm!$ for m =0,1,2,… Of course, if other prior densities, e.g. one that is invariant to a reciprocal transformation for φ (see Zellner (1998, p. 122) for derivation of such priors) are considered appropriate, they can be readily employed.

With the above approach, one can report the result that given the data relating to two alternative hypotheses or models, the posterior odds, K, has a mean = b, the Bayes Factor and the post data density given in (11). Also, it is possible to compute Pr ($a<K<c$ given b) = P*, for any given positive values of a and c using the exponential density for K given in (11) and the value of b = the Bayes Factor.

This example, that can be generalized in many different ways, e.g. to cases in which more than two hypotheses are considered, shows that a simple modification of the "equally likely" or "prior odds = 1" assumption, namely that the prior odds is a parameter that is subjectively random with a mean equal to one and a particular prior probability density provides a comparison of alternative hypotheses with EK = b, the Bayes Factor and a post data density for K, the posterior odds that can be used to make probability statements regarding the random posterior odds, K. Thus rather than just reporting, e.g., K= 3.2, it is now possible to report both a point estimate, e.g., EK = BF and probabilistic intervals for the random variable K, the posterior odds.

V. Summary and Conclusions

In this paper the usual derivations of the product rule of probability and Bayes's Theorem have been reviewed that depend on an "equally likely" assumption, also referred to as the principle of insufficient reason. Some regard the "equally likely" assumption to be a representation of our ignorance regarding alternatives and/or a fair and unprejudiced representation of views regarding alternatives. It was pointed out that having alternatives equally probable is indeed a strong assumption. A weaker assumption that was presented and analyzed involved giving alternatives differing unknown probabilities with means equal to an average probability that can be used to produce an extended form of the product rule of probability which, with a simple assumption about the means of the probabilities, reduces to the traditional forms of the product rule and Bayes's Theorem. Also, in comparing or testing alternative hypotheses, it was shown that broadening the assumption that the prior odds = 1 to the hypothesis that the prior odds is a random quantity with mean equal to one leads to the useful results that the posterior odds, K, becomes random, with EK = Bayes Factor, and that probability intervals are available for K. Of course, getting data relating to possible heterogeneity in the underlying probabilities can yield tests of the "equally likely" assumption and takes us beyond the assumed state of ignorance underlying this assumption, as has been recognized in the extensive literature dealing with heterogeneity in consumer, investor and other units' behavior. In the literature, procedures are readily available for computing posterior odds on fixed parameter versus random parameter models. See, e.g. Min and Zellner (1993) for derivations and applications of posterior odds for such problems, an operational approach that will help settle the issue empirically in an inductive, scientific manner, an issue that can not be resolved deductively.

*Research financed by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago. Thanks to Herman van Dijk and two referees for helpful comments on an initial draft of this paper.

References

Berger, J. O., 2000. Bayesian analysis: a look at today and thoughts of tomorrow.” Journal of the American Statistical Association 95, December, 1269-1276.

Bernardo, J.M., 1979. Expected information as expected utility. The Annals of Statistics 7, 686-690.

Cox, R.T., 1961. The algebra of probable inference. Johns Hopkins University Press, Baltimore, MD.

Feldstein, M., 2004. Discussion. American Economic Review 94, No. 2, May, 41-45.

Geweke, J., 2005. Contemporary Bayesian econometrics and statistics. John Wiley and Sons, Inc., Hoboken, New Jersey

Good, I.J., 1971. Discussion of "Measuring information and uncertainty" by R. J. Buehler in Godambe, V.P. and Sprott, D.A. (Eds.), Foundations of Statistical Inference. Holt, Rinehart & Winston, 337-339, Toronto

Greenspan, A., 2004. Risk and uncertainty in monetary policy. American Economic Review 94, No. 2, May, 33-40.

Jaynes, E.T., 2003. Probability Theory The Logic of Science. Bretthorst,G.L. (Ed.), Cambridge University Press, Cambridge.

Jeffreys, H., 1998. Theory of Probability, 3rd revised ed. (1967), reprinted in Oxford Classics Series Oxford University Press (1st ed., 1939), Oxford.

Keynes, J.M., 1952. A Treatise on Probability. Macmillan & Co. (1st ed. 1921), London.

Koop, G., 2003. Bayesian Econometrics. John Wiley and Sons, Inc., Hoboken, New Jersey.

Lancaster, T., 2004. An Introduction to Modern Bayesian Econometrics. Blackwell Publishing, Ltd., Oxford, UK.

Levi, I., 1977. Four types of ignorance. Social Research 44, No.4, 745-756

Min, C-k and Zellner, A., 1993. "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," Journal of Econometrics, Annals Issue, 56, 89-118, reprinted in Zellner, A. and Palm F.C., 2004. The Structural

Econometric Time Series Analysis Approach. Cambridge University Press, 559-589, Cambridge, UK.

Press, S.J., 2003. Subjective and Objective Bayesian Statistics: Principles, Models and Applications. John Wiley & Sons, Inc., Hoboken, New Jersey.

Rossi, P.E., Allenby, G.M. and McCulloch, R., 2005. Bayesian Statistics and Marketing. John Wiley & Sons, Ltd., Chichester, U.K.

Stigler, S.M., 1986. The History of Statistics. Harvard University Press, Cambridge, MA.

Yudkowsky, E., 2005. An Intuitive Explanation of Bayesian Reasoning, 38pp. at http://yudkowski.net/bayes/bayes.html

Zellner, A.,1988. Optimal information processing and Bayes's theorem. The American Statistician 42, No. 4, 278-294 with invited discussion and the author's reply.

Zellner, A., 1997. Bayesian Analysis in Econometrics and Statistics, invited contribution to M. Perlman and M. Blaugh, (Eds.), Economists of the Twentieth Century Series. Edward Elgar Publ. Ltd., Cheltenham, UK.

Zellner, A., 2000. Information processing and Bayesian analysis, presented at the American Statistical Association Meeting, August, 2001 and published in Golan, A., (Ed.), Annals Issue of the Journal of Econometrics 107 (2002), 41-50.

Zellner, A., 2003. Some aspects of the history of Bayesian information processing, 32 pp., paper presented at the ASA meeting in San Francisco, August, 2003 and at the ISBA 2004 World Meeting, Vina del Mar, Chile and to be published in Journal of Econometrics.