PRIOR PROBABILITIES REVISITED

Norman Dalkey

April 1985
CSD-850007

PRIOR PROBABILITIES REVISITED

N.C. Dalkey
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles

Presented at the:
Fourth Workshop on Maximum Entropy and Bayesian
Methods in Applied Statistics
University of Calgary, August 5-8, 1984

This work was supported in part by National Science Foundation Grant IST 84-05161

ABSTRACT

Unknown prior probabilities can be treated as intervening variables in the determination of a posterior distribution. In essence this involves determining the minimally informative information system with a given likelihood matrix.

Some of the consequences of this approach are non-intuitive. In particular, the computed prior is not invariant for different sample sizes in random sampling with unknown prior.

PRIOR PROBABILITIES REVISITED

N. C. Dalkey

1. *Generalities*

The role of prior probabilities in inductive inference has been a lively issue since the posthumous publication of the works of Thomas Bayes at the close of the 18th century. Attitudes on the topic have ranged all the way from complete rejection of the notion of prior probabilities (Fisher, 49) to an insistence by contemporary Bayesians that they are essential (de Finetti, 75). A careful examination of some of the basics is contained in a seminal paper by E.T. Jaynes, the title of which in part suggested the title of the present essay (Jaynes, 68).

The theorem of Bayes, around which the controversy swirls, is itself non-controversial. It is, in fact, hardly more than a statement of the law of the product for probabilities, plus the commutativity of the logical product. Equally straightforward is the fact that situations can be found for which representation by Bayes theorem is unassailable. The classic classroom two-urn experiment is neatly tailored for this purpose. Thus, the issue is not so much a conceptual one, involving the "epistemological status" of prior probabilities, as it is a practical one. In practice, the required prior probabilities are often unknown, or poorly known.

The present paper presents an approach to the estimation of prior probabilities when these are unknown. The approach is a generalization of maximum entropy methods. It was derived with a quite different rationale, and thus represents a convergence of two different streams of thought.

2. *Figures of Merit*

As a foundation for a theory of estimation, it is necessary to introduce a figure of merit, a measure of the excellence of an estimate.

Figures of merit are commonly some form of discrepancy measure; e.g., if I am asked to guess the height of a distant tree, the excellence of my guess is determined by comparing it with the actual height. In the measurement literature a wide variety of scores can be found--absolute difference, squared difference, percentage difference, and the like.

Estimates of probabilities have the difficulty that the true or actual probability is rarely available for comparison. An ingenious way to sidestep this difficulty has been found in the theory of proper scores (Savage, 71). Let E be a partition on an event space, and e an unspecified member of E. Let R be an estimate of the probability distribution on E. Finally, let $S(R, e)$ be a function which assigns the score (rating, reward, payoff, etc.) if R is the estimated probability distribution on E and the event e occurs. If $P(E)$ is the actual probability distribution on E, the expected score for the estimate R is $\sum_E P(e)S(R, e)$. Notice that the score $S(R, e)$ can be assigned knowing only the estimate R and the event e that actually occurs, without knowing the actual probability P.

A score rule is called proper (reproducing, honesty-promoting, admissible, etc.) if it fulfills the condition:

$$\sum_E P(e)S(R, e) \leq \sum_E P(e)S(P, e) \qquad (1)$$

i.e., a score is called proper if the expected score is a maximum when the estimate is the same as the distribution which determines the expectation. (1) is analogous to the requirement for a discrepancy score that the "error" be a minimum when the estimate is precisely the same as the actual quantity.

It is convenient to introduce some definitions:

$$G(P, R) = \sum_E P(e)S(R, e)$$

$$H(P) = G(P, P) = \sum_E P(e)S(P, e)$$

$$N(P, R) = H(P) - G(P, R)$$

$G(P, R)$ is the expected (discrepancy) score if R is the estimate and P is the actual distribution. $H(P)$ plays a special role for probabilistic scores. For error measures that are analogous to a distance, e.g., the absolute difference, $H(x) = |x-x| = 0$ for all x. However, for proper scores, $H(P)$ represents a measure of the excellence of a distribution P on its own, so to speak. $N(P, R)$ is the net score if R is estimated and P obtains. Note that from (1) $N(P, R)$ is always non-negative.

Since $G(P, R)$ is an expectation, it is linear in P. An important property of $H(P)$ is that it is convex (Dalkey, 82).

There is a very large family of scores that fulfill (1). They range from scores derived from decisional payoff matrices to scores appropriate primarily for scientific contexts (Dalkey, 80). The most widely used of the latter is the logarithmic score, $S(R, e) = \log R(e)$. Note that $-H(P)$ for the logarithmic score is precisely the Shannon entropy for the distribution P.

Proper scores can play the same role in inductive logic that truth-value plays in traditional logic. In fact, the truth-value is a form of proper score. If an individual believes a given statement is true, but asserts the negation, his expected "score" is *false*, clearly less excellent than if he had asserted what he believed.

Proper scores enable the verification of statements of the probability of a single case, a possibility usually denied in the literature of probability theory. If an estimator asserts $P(e) = p$, where e is a specific event such as "rain tomorrow", one need only wait until tomorrow and (for the logarithmic score) award the prediction with the score $\log p$ if it rains, or $\log (1-p)$ if it doesn't. The dependence upon the occurrence of a specific event gives the requisite tie to reality needed for a verification procedure, and the dependence on the asserted probability furnishes the requisite dependence on the content of the assertion.

3. *Min-score Induction*

Given an appropriate figure of merit, it is feasible to formulate an inductive logic that is an extension of classical logic. A general structure for a logic is a collection of rules which transform a set of premises into a conclusion. In the classical case, if the premises are true and the inference is valid (i.e., follows the rules), then the conclusion must be true. That simple guarantee is, of course, precisely what makes classical logic useful in inquiries. As might be expected, the nature of the guarantee is somewhat more complex in inductive logic.

In the most elementary case, consider a partition E on an event space, where E represents the events of interest, i.e., E specifies the events for which a probability distribution is desired. We assume that there is a probability distribution on the event space, and thus, in particular, there is a distribution $P(E)$ on the partition E. Suppose the partial information consists of knowing that $P(E)$ is in some class K of distributions on E. In the extreme case of no information, K is the set of all possible distributions on n events, where n is the number of events in E; i.e, K is just the simplex Z_n of all probability distributions on n events. If K is a unit class, then $P(E)$ is completely known. In intermediate cases, K is some subset of Z_n.

We can take the specification of K as the premises of an inference. What is desired as a conclusion is some estimate $R(E)$ of the distribution on E. Since by assumption the actual distribution P is in K, it might be supposed that R must be selected from K. However, there is no formal constraint that R be in K; it could be any distribution in Z. Assuming that a score rule S has been adopted, the actual expectation is $G(P, R)$. The inductive rule to be employed in this paper is derived from two postulates:

P1. The selection rule should guarantee at least the expected score of R--i.e., it should guarantee $H(R)$. Formally, this requires $G(P, R) \geq H(R)$, for any P in K.

P2. The selection rule should assure the positive value of information--i.e., if additional information is obtained, then the expected score should not decrease. Formally, if $K' \subset K$, then $H(R') \geq H(R)$.

These two postulates lead to a specific selection rule which could be called the min-score rule: Select the R in the closure of the convex hull of K that minimizes $H(R)$ (Dalkey, 82). If K is convex and closed, then R will be in K; if K is not convex and closed, then R may not be in K, but will be in the closure of the convex hull of K (Dalkey, 85).

P1 appears to be essential for any kind of inference. The user of the conclusion must be confident that he will achieve at least as high an expected score as the conclusion promises. P2 is more germane to induction. In the case of complete information, the positive value of information is a theorem (Lavalle, 78). It appears *a-fortiori* plausible that additional information should be constructive in the case of incomplete information.

If the score rule adopted is the logarithmic score, then for the elementary case under consideration, the min-score rule is precisely the maximum entropy procedure. As noted above, the expected log score is just the negative of the entropy. For more highly structured problems, the min-score approach may lead to a different analysis than current practice with maximum entropy methods. This divergence will show up in the analysis of unknown prior probabilities.

4. *Prior Probabilities*

The elementary min-score rule does not involve the distinction between prior and posterior probabilities. The information class K does represent "prior information", but is not expressed as a probability.

Historically, the notion of prior probability has been employed in the context of "updating". A probability distribution is known for an event set E. New evidence I, either planned as in an experiment, or fortuitous as in casual observation, comes to attention, and the problem arises of revising the probability distribution on E to reflect the new evidence. In this case, the old distribution $P(E)$ is the prior and the new distribution $P(E|I)$ is the "posterior". Of course, $P(E|I)$ can operate as a new prior if further evidence arises. The distinction is significant only for a given instance of updating. Another way of putting the same point is that the distinction between primary events and evidence is not a formal aspect of the calculus of probabilities.

As long as the updating is conducted with complete information (all the relevant probabilities known), there is no conceptual difficulty. A variety of updating procedures is available, depending on what is known concerning the relationships between the evidence and the primary events. The one most frequently employed is the theorem of Bayes, $P(E|I) = P(E)P(I|E) / P(I)$.

Difficulties do arise, of course, if the relevant probabilities are not completely known. Essentially, what the analyst needs to know for the updating step is the joint distribution $P(E.I)$. A frequent situation is that in which the likelihoods $P(I|E)$ are known, but not the joint distribution.

In the context of min-score inference, the class K can be taken to be a set of joint distributions, constrained by the requirement that they generate the known likelihoods, i.e., $P(E.I)$ is in K if $P(E.I) / \sum_{I} P(E.I) = P(I|E)$.

In the given instance, the class K can be characterized equivalently by the set of joint distributions $P(E.I) = P(E)P(I|E)$ where $P(E)$ can be any distribution on n events (since $P(E)$ is totally unknown). However, it is clearly incorrect to select the min-score distribution in Z_n for $P(E)$ since this ignores the role of the score rule. The score for the updating problem is related to the posterior probability $P(E|I)$, not to the prior. In colloquial terms, the analyst is not being paid to estimate the prior, or, from the standpoint of the decision maker, his payoff will be determined by implementing the posterior, not the prior.

A further complication arises in imposing the score rule for the case of incomplete information. With complete information, it is legitimate to ignore all potential evidence except the specific item that actually obtains. Considering I as a set of possible items of evidence (observations, data, signals, etc.), and i as a member of I, then in practice what is wanted is $P(E|i)$ when i is known. This feature has been elevated to the status of a principle by some writers--the posterior determined by an item of evidence i should be a function solely of i and not of any other potential evidence that might have been observed.

That principle cannot be maintained in the case of incomplete information. For the illustrative case where the likelihoods, but not the prior, are known, the information in the likelihood matrix concerning potential (but not observed) evidence is relevant to the assessment of the observed evidence. As a simple example, consider the case of two events e and $\bar{e}$ (the bar indicating negation). Suppose there are two possible pieces of evidence, i and $\bar{i}$. Let $P(i|e) = q$ and $P(i|\bar{e}) = r$. Without loss of generality, we can let

$q > r$. (If $q = r$, the evidence is trivial.) Set $P(e) = p$. With $p = r/(q+r)$, $P(e|i) = 1/2$. Thus, whatever q and r, a prior probability can be assigned that makes the evidence completely uninformative (at least for any symmetrical score rule).

The example clearly generalizes to several events and several potential items of evidence. Thus, for the assessment of evidence in the case of incomplete information, it is necessary to treat the evidence and the events of interest as an information system, and the selection of a prior probability as the design of a min-score information system. For the logarithmic score, this requirement can be restated as designing a minimally informative information system (Dalkey, 80).

Summarizing: For the updating problem, the probabilities of interest are the posterior conditional probabilities $P(E|I)$; it is the expected score of these probabilities which determines the value of the new evidence. However, there is a separate posterior for each potential item of evidence; thus, the complete assessment consists of the average of these expectations over the potential items of evidence. Denoting the average expected score by $H(E|I)$, we have

$$H(E|I) = \sum_{I} P(i) \sum_{E} P(e|i) S(i, e) \qquad (2)$$

where $S(i, e)$ is shorthand for "the score given that $P(E|i)$ is the estimate, and e occurs".

For the logarithmic score, (2) can be unpacked in the form of a well-known formula in information theory:

$$H(E|I) = H(E) + H(I|E) - H(I) \tag{3}$$

That is, the average information furnished by an information system (E, I) is the information contained in the prior distribution $P(E)$, plus the average information in the likelihood matrix $P(I|E)$ minus the information in the initial distribution on the evidence $P(I)$. Notice that there is a simple duality between events and evidence. From (3), $H(I|E) = H(I) + H(E|I) - H(E)$.

If the prior probabilities $P(E)$ are not known, the min-score inference rule prescribes minimizing $H(E|I)$ as a function of the distribution $P(E)$ over the class K of joint distributions $P(E.I)$ constrained by the likelihood matrix $P(I|E)$. The maximum entropy rule (for the log score) is now extended to a maximum expected entropy rule.

For the illustrative case of two events described above,

$H(E) = p\log p + (1-p)\log(1-p)$,

$H(E|I) = p(q\log q + (1-q)\log(1-q)) + (1-p)(r\log r + (1-r)\log(1-r))$,

$H(I) = (pq + (1-p)r)\log(pq + (1-p)r) + (p(1-q) + (1-p)(1-r))\log(p(1-q) + (1-p)(1-r))$.

(I've expanded this elementary case in somewhat tedious detail because the role of the prior probability p is different from the usual form of max entropy analysis.) $H(E|I)$ can be minimized as a function of p by elementary differentiation and setting the result equal to 0. The solution is obtained by solving for p the implicit equation

$$\frac{p}{1-p} e^{H(q)-H(r)} = \left(\frac{pq + (1-p)r}{p(1-q) + (1-p)(1-r)} \right)^{q-r} \tag{4}$$

The solution is not particularly intuitive. If q and r are symmetric, i.e., $r = 1-q$, the min-score p is the classic uniform distribution, $p = 1/2$. However, if q

and r are not symmetric, and each is rather far from 1/2, the min-score prior is not uniform. For example, if $q = .9$ and $r = .025$, the min-score prior is about .63. Roughly speaking, the min-score solution puts greater weight on the "less informative" prior event.

An even less intuitive result is obtained if the observation is iterated, e.g., if two independent observations are made. The min-score prior computed from the extension of (4) to two observations is not the same as the prior computed form one observation; e.g., the min-score prior for $q = .8$, $r = 0$, is .625 for one observation and is .69 for two independent observations. The "discounting" of the more informative event is more drastic for the two-observation case; the difference between q and r has a more pronounced effect on the likelihoods for two observations.

In the classic calculus of probabilities, the effect of an additional observation can be computed by "updating", i.e., by using the posterior probability for one observation as the new prior for an additional observation. This procedure is not valid for the case of an unknown prior. One way of expressing what is going on is to note that in the min-score analysis, the solution is sensitive not only to the inputs , but also the precise question being asked. As remarked above, the question being asked in the case of additional evidence is the posterior probability given the evidence. If the evidence changes, then a new prior must be computed. Another way of saying the same thing is that the relevant K for the case of one observation is a set of joint distributions of the form $P(E.I)$; for two observations the relevant K is a subset of distributions of the form $P(E.I_1.I_2)$. This characteristic of the min-score rule has serious implications for general

purpose inference mechanism, e.g., expert systems. In a medical expert system, for example, there is a basic difference between the diagnostic and the prognostic use of data from the min-score point of view. A system could not use the same set of "best-guess priors" for both types of estimate.

Some readers may find this dependence on the specific question being asked a serious drawback to min-score procedures. There is no question but that it is a serious practical complication. A single prior distribution cannot be computed and then plugged into each new problem. However, the "difficulty" serves to emphasize the basic difference between complete and partial information. In the case of partial information and updating on new evidence, the prior probabilities are "intervening variables", serving to complete the analysis, not to advance knowledge. The new knowledge is contained in the posterior estimates.

5. *Random Sampling with Unknown Prior*

The classroom example of the previous section has a highly structured frame of reference. In practice most problems are not so neatly packaged. A case in point is random sampling with unknown prior. An elementary example is an exotic coin with unknown probability of heads. Another example is the case of the possible loaded die treated by Jaynes (Jaynes, 82). In the classroom example, there are two well-defined "states of nature" and a fairly clear interpretation of the prior probability--someone presumably selected one of the two states, e.g., one of two urns, according to a specific probability distribution. In the case of the exotic coin, the states of nature are not "given" and a mechanism to incarnate a prior probability is even less apparent.

A frame of reference for such problems was devised by Laplace. For the exotic coin, each potential probability of heads is considered to be a separate state. For the loaded die, each potential probability distribution on the six faces is a state. The prior probability, then, is a distribution on these states. If it is assumed for the coin that any probability of heads from 0 to 1 is possible, then a prior probability would be density on the interval 0--1. For the die, with similar freedom, a prior probability is a density on the simplex of distributions for six events.

The model is illustrated in Fig. 1 for the coin. There is a continuum of states, labeled by the probability p of heads. The prior is a density $D(p)$ whose integral is 1; and the likelihoods are just the probabilities p for a single flip of the coin. For multiple flips, assuming independence, the likelihoods are the Bernoulli probabilities $P(p, n, m) = \binom{n}{m} p^m (1-p)^{n-m}$ for the case of m heads in n flips of the coin.

For the logarithmic score, the continuous version of (3) holds, where

$$H(E) = \int_0^1 D(p) \log D(p) dp$$

$$H(I|E) = \int_0^1 D(p) [\sum_{m=0}^{n} P(p, n, m) \log P(p, n, m)] dp$$

$$H(I) = \sum_{m=0}^{n} \int_0^1 D(p) P(p, n, m) dp \log \int_0^1 D(p) P(p, n, m) dp$$

The min-score problem, then, is to find $D^o(p)$ which minimizes $H(E|I)$.

Because of the symmetry of the state space--for every state p there is an antisymmetric state $1-p$--we can expect the min-score $D^o(p)$ to be symmetrical in p. For one observation, symmetry implies that $P(I)$ is uniform, i.e., $P(\text{heads}) = 1/2$. Thus, the term $H(I)$ is invariant under changes in $D(p)$, and we have

$$D^o(P) \propto e^{-H(p)} \tag{5}$$

and since $-H(p) = \text{Entropy}(p)$, equivalently

$$D^o(p) \propto e^{\text{Ent}(p)} \tag{5}$$

The posterior density of p, given the observation i, is then

$$D^o(p|i) = \frac{pe^{-H(p)}}{\int_0^1 pe^{-H(p)}dp} \tag{6}$$

In Fig. 2, the prior and posterior densities are drawn for the case of a single observation. Also shown in dashed lines are the uniform prior and the posterior density for the uniform prior. If we take the average of p for the posterior distribution as the "best guess" p, then, for the uniform prior it is 2/3, and for the min-score prior it is .64. The difference is not large for a single observation.

For multiple observations, it is no longer the case that $H(I)$ is invariant under changes in $D(P)$. It is instructive, however, to glance at the result if $H(I)$ is assumed to be invariant. In that case we would have

$$D^o(p) \propto e^{-nH(p)} \tag{7}$$

i.e., the prior density becomes increasingly concentrated around 1/2 with increasing n, where n is the number of observations. Fig. 3 shows this "first approximation" $D^o(p)$

for several n. It is clear that for this approximation, $D^o(p)$ converges to a distribution concentrated at $p = 1/2$ as $n \to \infty$.

I do not have an exact solution for the case of $n>1$. If we consider two extreme distributions, the uniform distribution $D_u(p) = 1$, and the distribution $D_{1/2}(p)$ concentrated at $p = 1/2$, we can say that $H(I|E) = H(I)$ for $D_{1/2}(p)$, i.e., $H(I|E) - H(I) = 0$, and $H(E) = \infty$. At the other extreme, $D_u(p)$, $H(E) = 0$, $H(I|E) = \frac{-n}{2}$ and $H(I) = \log \frac{1}{n+1} - \frac{1}{n+1} \sum_{m=0}^{n} \log \binom{n}{m}$. Since H is convex, and $H(I|E)$ is an average while $H(I)$ is an H of averages, $H(I|E)$ is always greater than $H(I)$, but the difference is concave. Thus, D^o is intermediate between D_u and $D_{1/2}$. $H_u(I|E) - H_u(I) \to \infty$ as $n \to \infty$, thus $D^o \to D_{1/2}$ as $n \to \infty$.

Even without an exact solution, then, we can conclude that as $n \to \infty$, the asymptotic D^o is massed at 1/2. Qualitatively, this implies that the amount of information in large samples with unknown prior is less than classical theory would imply. For binary events, this implies that the best guess is closer to 1/2 than the classic $\frac{m}{n}$ guess, where m is the observed frequency of an event in n trials. This result is compatible, e.g., with the observation that opinion polls are, on the average, too extreme, i.e., they tend to predict a larger margin of winning than is actually observed (Dembert, 84). Of course, political polls involve potential forms of error other than the statistical analysis, and thus the compatibility with our present result is only suggestive.

The present treatment of random sampling with unknown prior deals with fixed-sample experiments. The number n of samples is fixed before hand, and a posterior distribution, $P^o(E|i)$, is computed for each potential sampling pattern i. Furthermore, the computation is conducted under the supposition that the score will be determined by the posterior distribution $P^o(E|i)$. In effect, this requires that the actual probability p be observed before the score can be awarded. For the textbook two-urn case there is no difficulty; determining which urn was selected is straight-forward. However, in more realistic contexts, this requirement may not be implementable. For the possibly biased coin case, there is no way to directly observe the actual probability.

What can be observed (in theory, at least) is further samples. It might be supposed that an operational scoring mechanism could be devised by using the computed posterior to predict the probability of further observations, and base the score on the occurrence or non-occurrence of these observations. However, as we have seen, to introduce further observations requires expanding the frame of reference to $(E.I_1.I_2)$, where I_1 is the initially observed sample, and I_2 is the predicted sample. This analysis can, of course, be carried out, and is a legitimate application of the min-score formalism; but, as seen earlier, this requires, in essence, computing a new best guess prior. In the former case, the figure of merit is $H(E|I_1)$; in the latter case, the figure of merit is $H(I_2|I_1)$, and these two may give divergent results.

6. *Constraints and Statistics*

The justification of the min-score rule assumes that the unknown probability function P is contained in the knowledge set K. In other words, it is assumed that whatever constraints delimit K are categorical. In contrast, it is common in applications of maxent methods to introduce constraints that do not have this property. A frequently used form of constraint is one derived from an observed statistic; i.e., given a statistic S, with observed value s, it si assume that the class K consists of those probability distributions whose expectation S is the observed value s. As an example, in the case of the biased die analyzed by Jaynes, he assumes that the probability distribution under investigation has a theoretical average equal to the observed sample average.

It is clear that such statistical constraints are not categorical. For the die example, any distribution within the interior of the simplex of all distributions on six events could give rise to the observed statistic. Many of these would have very small likelihoods of generating a sample with the observed average, but that is a fact to be exploited by the analysis, not ignored. The justification of the step from sample data to theoretical expectations is somewhat obscure in the literature. Jaynes uses terms such as "compatible with the given information" (Jaynes, 68) or "allowed by the data" (Jaynes, 82), but in light of the compatibility of the observed statistic with any underlying distribution, it is not clear how these terms are being used.

If the complete frequency table arising from a random experiment is available to the analyst then the maxent procedure becomes irrelevant. The constraint convention--expectation = observed value--leads to $P_i = f_i/N$ where f_i is the observed frequency for event i, and N is the total number of sample points. Some obscurity arises in this regard

concerning the question whether the justification of the procedure is intended to be asymptotic (infinite sample) or not. But, in practice, it seems clear that the procedure is intended to apply to finite samples.

In the min-score approach, observations are not considered as constraints on the knowledge class K, but rather are elements of an information system. In the case of the loaded die, what is sought is the best-guess posterior density on probability distributions on the faces of the die given the observed average. The class K is all joint densities $D(P,A)$, where P is a probability distribution on six events, and A is a potential observed average over N tosses of the die. Analysis consists of finding the minimally informative information system with this structure. As in the case of the binary event sampling analyzed in the previous section, there is no constraint on the possible probability distributions P.

If it is presumed that uncertainty arises in a given experiment, not from "error" but from the fact that the expectation of a statistic need not be the same as the observed value, then the min-score procedure is a way to deal with uncertainty without the addition of an error term.

7. *Comments*

The analysis of unknown prior probabilities presented above leaves a great deal to be desired as far as mathematical implementation is concerned; but there does not appear to be any deep mathematical issues involved.

The same cannot be said for logical issues. One that appears particularly critical is the fact that a min-score estimate is simply a best guess that depends on the score rule and on the specific question being asked. This characteristic seems to deny the possibility that min-score inference can be used to add to the store of knowledge. In a sense, this result is inherent in the formalism. By definition, all that is *known* is the class K and observations I.

This issue will be explored at somewhat greater depth in an upcoming paper (Dalkey, 85).

References

Dalkey, N. C. (1980) "The Aggregation of Probability Estimates". UCLA-ENG-CSL-8025.

Dalkey, N. C. (1982) "Inductive Inference and the Maximum Entropy Principle", UCLA-ENG-CSL-8270, presented at the 2nd Workshop on Maximum Entropy, U. of Wyoming, Aug., 1982.

Dalkey, N. C. (1985) "The Representation of Uncertainty" (In preparation).

de Finetti, B. (1975) *Theory of Probability*, Vol. II, John Wiley and Sons, New York.

Dembert, L. (1984) "Roper Cites Flaws in Political Polls", *Los Angeles Times*, May 26, p. 16.

Fisher, R. A. (1949) *The Design of Experiments*, Oliver and Boyd, London.

Jaynes, E. T. (1968) "Prior Probabilities", *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, pp. 227-241.

Jaynes, E. T. (1982) "On the Rationale of Maximum-Entropy Methods", *Proceedings of the IEEE*, 70, pp. 939-952.

Lavalle, I. H. (1978) *Fundamentals of Decision Analysis*, Holt, Rinehart & Winston, New York.

Savage, L. J. (1971) "Elicitation of Personal Probabilities and Expectations", *Jour. Amer. Stat. Assoc.* 66, pp. 783-801.

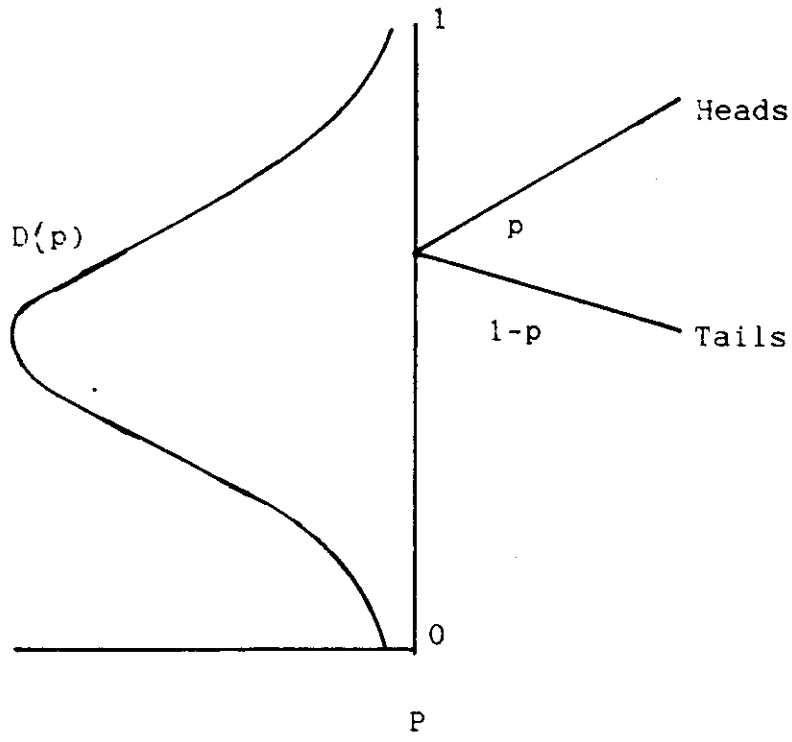

Figure 1. Laplace model for binary-event random sampling with unknown prior.

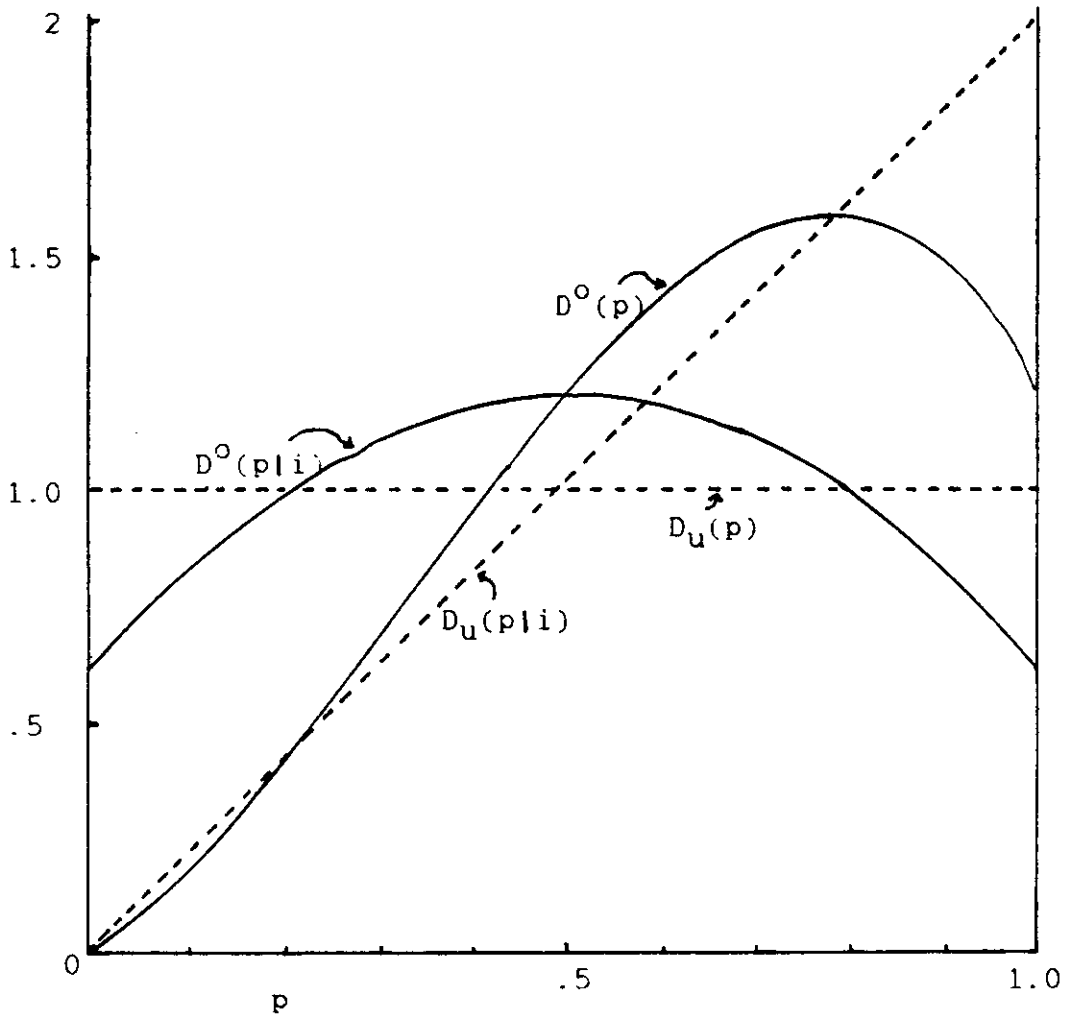

Figure 2. Min-score prior density $D^o(p)$, and min-score posterior density $D^o(p|i)$ for single observation (solid lines), with uniform prior $D_u(p)$ and corresponding posterior density $D_u(p|i)$ (dashed lines).

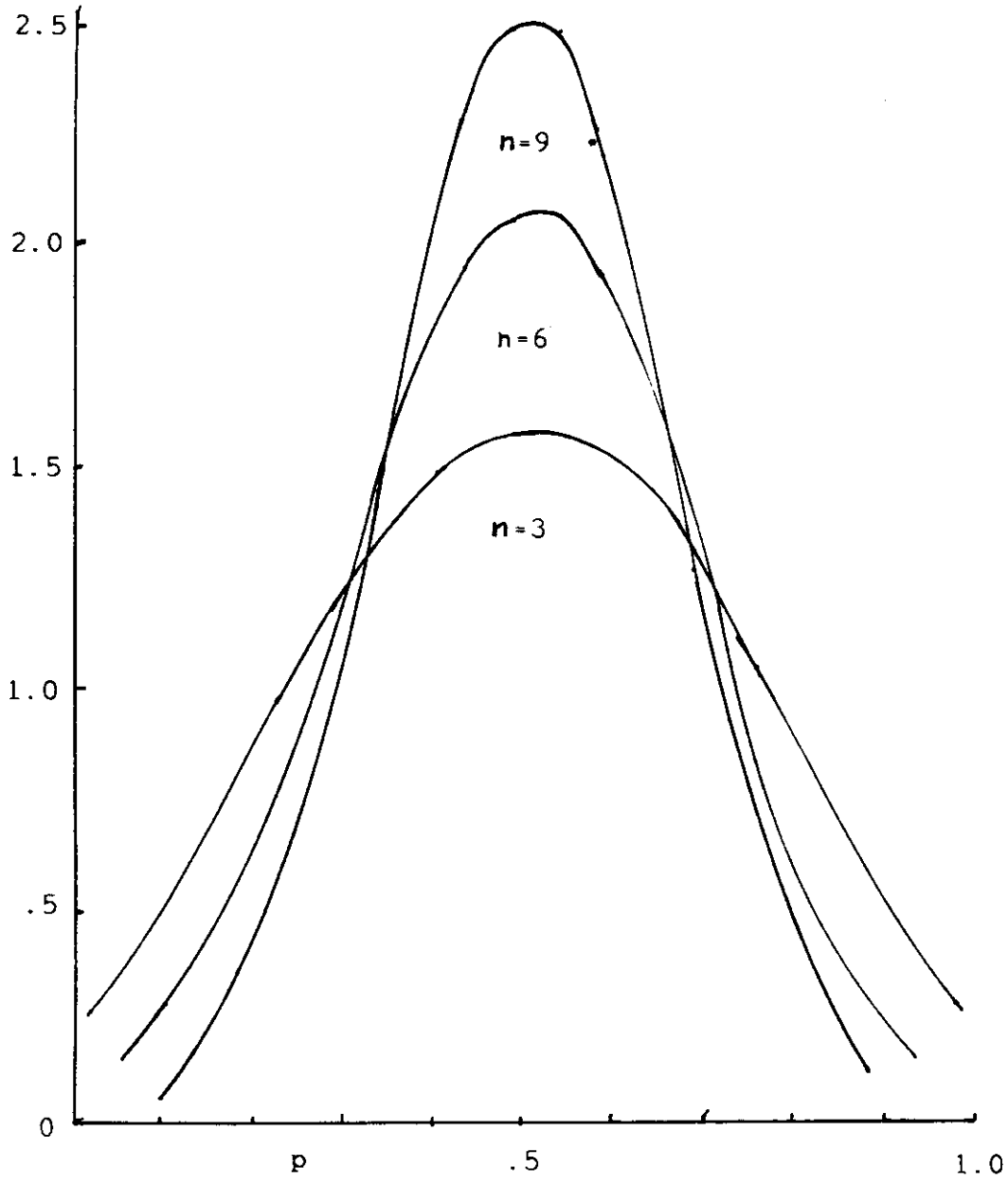

Figure 3. Approximate min-score priors, $ke^{-nH(p)}$, for various sample sizes n.