PHYSICAL REVIEW D VOLUME 37, NUMBER 5 1 MARCH 1988

Small-signal analysis in high-energy physics: A Bayesian approach

Harrison Bertrand Prosper*

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 26 February 1987)

The statistics of small signals masked by a background of imprecisely known magnitude is addressed from a Bayesian viewpoint using a simple statistical model which may be derived from the principle of maximum entropy. The issue of the correct assignment of prior probabilities is resolved by invoking an invariance principle proposed by Jaynes. We calculate the posterior probability and use it to calculate point estimates and upper limits for the magnitude of the signal. The results are applicable to high-energy physics experiments searching for new phenomena. We illustrate this by reanalyzing some published data from a few experiments.

I. INTRODUCTION

Searches for new phenomena in high-energy physics naturally yield very few events; often the background contribution to these events is poorly known and sometimes the estimate of background level is greater than the number of events found. This immediately leads to difficulties and invariably to heated debate. The disagreements stem from the lack of a standard statistical method for dealing with small-signal problems. The problem is of sufficient practical importance to merit a close examination. That is what we aim to do in this paper.

We have decided to address this problem from a Bayesian viewpoint. In so doing we are merely acknowledging the fact that a coherent solution to the small-signal problem is more easily achieved within a Bayesian framework than one which uses the methods of "classical" (i.e., orthodox) statistics.[1] This requires of course that we accept the notion that probability is to be interpreted subjectively; the relative-frequency interpretation is considered less fundamental. Therefore, statements about the probability of a parameter assuming certain values are deemed meaningful.

Bayesian methods, however, should not be considered entirely satisfactory until the old problem of how to quantify prior information in terms of prior probabilities has been resolved; in particular, how to represent a state of "complete initial ignorance." For large signals, the problem of prior probabilities is largely irrelevant; however, for the very small signals discussed here the problem may not be brushed aside.

In fact, a considerable step towards resolving this issue has been taken by Jaynes[2] in his statement of the following principle: "in two problems where we have the same prior information, we should assign the same prior probabilities." While some would argue otherwise, we regard the foregoing principle as being both reasonable and constructive. Jaynes has shown that in several simple but important examples it leads to unique assignments of prior probabilities. If, nonetheless, one is inclined to reject this idea then presumably prior probabilities are to be considered subjective in a more thoroughgoing sense: namely, that they are personal assessments of prior information.[3] For example, if prior to performing an experiment our only knowledge concerning the value of a parameter is that it lies within a specified interval it would be reasonable, from the "personal" viewpoint, to assign to this information a uniform prior distribution. Indeed, since the time of Bayes a uniform prior distribution has been the form most commonly adopted.

In this paper, however, we specifically reject the "personal" approach as inappropriate for the statistical analysis of experiments in physics; our point of view is that of Jeffreys:[4] namely, that the prior probability should be as "impersonal as possible." This point of view, we believe, is in keeping with the established tradition in the mathematical sciences; indeed, it seems to us that if we want the assignment of prior probabilities to be less a matter of individual judgment and more a matter of "impersonal" mathematics we should follow that tradition: that is, make a statement of basic principles and follow through to some ultimate conclusions. This is what we attempt to do here.

The paper is organized as follows. In Sec. II we discuss the statistical model we have used for our analysis. In Sec. III we obtain a unique prior distribution for this model. Section IV deals with the calculation of the posterior distribution of the signal and its use in making statistical inferences. In Sec. V we illustrate the use of our results by reanalyzing some published data. We consider data from the LENA Collaboration[5] on hadronic decays of the $\Upsilon(1S)$; we look at data on flavor-changing neutral currents (FCNC's) from JADE (Ref. 6) and finally we examine data from the CRISP Collaboration[7] on $n\bar{n}$ oscillations in free neutrons. Some concluding remarks are made in Sec. VI.

II. THE STATISTICAL MODEL

The results of a high-energy physics experiment searching for new phenomena are typically the number N of signal plus background events, drawn from a distribution with unknown rate λ, and an estimate $(\hat{\mu},\hat{\sigma})$ of the background rate μ. Ordinarily, the background estimate is obtained either from a real control experiment or from a Monte Carlo simulation. In the first case we obtain an estimate B of the background rate μ_B of the control experiment which, together with the relation $\mu_B=\mu_B(\mu)$, leads to an estimate of μ. For background rate estimates derived from a Monte Carlo simulation the situation is a bit more complicated. Ideally, the error $\hat{\sigma}$ quantifies the magnitude of the *systematic* uncertainties in the value of μ; in principle, the statistical error can be rendered negligible by running the Monte Carlo code long enough. In this case the distribution of possible background means would be incorporated into the overall prior distribution for the problem. Sometimes, however, data are published (see, for example, Sec. V B) in which the statistical error in the Monte Carlo background rate estimate may not be neglected. This statistical error can be dealt with in precisely the same manner as for a real control experiment. We note, however, that this does not preclude the incorporation of knowledge of systematic errors into the prior probability. For example, if it is known that the background rate lies within a certain interval this information can be trivially incorporated into the prior probability.

For many real experiments

$$\mu_B(\mu)=\eta\mu \ , \tag{1}$$

where η is a known constant. The background rate estimate is usually taken to be

$$(\hat{\mu},\hat{\sigma})=(B,\sqrt{B}\,)/(\eta t) \ , \tag{2}$$

where t is the observation time for the experiment.

If $P(N\mid\lambda,t)$ is the probability of the result N and $P(B\mid\mu_B,t_B)$ the probability to obtain the result B, then given that the two experiments are independent, the joint probability of the results, $P(N,B\mid s,t;\mu,t_B)$, is given by

$$P(N,B\mid s,t;\mu,t_B)=P(N\mid\lambda,t)P(B\mid\mu_B,t_B) \ , \tag{3}$$

where $s=\lambda-\mu$ is the unknown signal rate and t_B is the observation time for the background experiment. It would be reasonable to assume that the function $P(n\mid\alpha,t)$ is the Poisson distribution and we shall in fact assume this. However, from a strict Bayesian viewpoint the probability $P(n\mid\alpha,t)$ need have nothing whatsoever to do with the relative frequency of events. In fact, as we now demonstrate, the Poisson distribution can be derived from first principles without appealing to this interpretation of probability. Our derivation makes use of the principle of *maximum entropy.*

Consider the time interval $[0,t]$. Let it be partitioned into K intervals of duration $\delta t=t/K$. We assume that events occur one at a time and that K is so large that no more than one event can occur per interval δt. Eventually, we shall let $K\to\infty$. We shall call each sequence of events in $[0,t]$ an event history. The set of all event histories contains 2^K elements. To the ith element of this set we assign the probability q_i. Let n_i be the total number of events within the ith history. The entropy associated with the distribution $\{q_i\}$ is defined by

$$S=-\sum_{i=1}^{2^K} q_i\ln q_i \ , \tag{4}$$

which can be resummed as

$$S=-\sum_{n=0}^{K}\ \sum_{i\in\{i:n_i=n\}} q_i\ln q_i \ . \tag{5}$$

If we now assume that within each class of event histories defined by the set $\{i:n_i=n\}$ the associated probabilities are identical and that

$$q_i=Q_n \ , \tag{6}$$

say, we can write the entropy as

$$S=-\sum_{n=0}^{K}\begin{bmatrix}K\\n\end{bmatrix}Q_n\ln Q_n \ , \tag{7}$$

which, with the definition

$$P_n\equiv\begin{bmatrix}K\\n\end{bmatrix}Q_n \ ,$$

that is, the probability associated with the class of event histories in which n events occur, becomes

$$S=-\sum_{n=0}^{K}P_n\ln(n!P_n)+\sum_{n=0}^{K}P_n\ln[K!/(K-n)!] \ . \tag{8}$$

Our problem then is to maximize S subject to the constraints

$$\sum_{n=0}^{K}P_n=1 \tag{9}$$

and

$$\sum_{n=0}^{K}nP_n=\alpha t \ . \tag{10}$$

After some simple algebra and making use of Eq. (10) we can write the second term in Eq. (8) in two parts:

$$\sum_{n=0}^{K}P_n\ln[K!/(K-n)!]=\alpha t\ln K+(1/K)\sum_{n=0}^{K}P_n\sum_{i=0}^{n-1}\ln[(1-i/K)^K] \ . \tag{11}$$

The second term in Eq. (11) is of $O(1/K)$ and therefore vanishes in the limit $K\to\infty$. Applying the maximum-entropy principle, $dS=0$, and then taking the limit $K\to\infty$ leads to the following set of equations whose simultaneous solution is required:

$$\sum_{n=0}^{\infty}[\ln(n!P_n)+1]dP_n=0 \ , \tag{12}$$

$$\sum_{n=0}^{\infty}dP_n=0 \ , \tag{13}$$

and

$$\sum_{n=0}^{\infty}n\,dP_n=0 \ . \tag{14}$$

The solution is

$$P_n = \frac{1}{Z(\omega)} \frac{e^{-\omega n}}{n!} , \quad (15)$$

where

$$Z(\omega) = \sum_{n=0}^{\infty} \frac{e^{-\omega n}}{n!}$$

$$= e^{e^{-\omega}} , \quad (16)$$

and ω is a Lagrange multiplier determined from the constraint equation

$$\frac{-\partial \ln Z}{\partial \omega} = \alpha t . \quad (17)$$

Imposition of the constraint leads finally to the result

$$P_n \equiv P(n \mid \alpha, t) = \frac{(\alpha t)^n e^{-\alpha t}}{n!} . \quad (18)$$

Using Eqs. (18) and (1) in the general formula, Eq. (3), leads to the probability

$$P(N,B \mid s,t;\mu,t_B) = \frac{1}{N!B!}[(s+\mu)t]^N \times (\eta\mu t_B)^B e^{-(s+\mu)t - \eta\mu t_B} , \quad (19)$$

which defines our statistical model. The posterior probability of the signal $P(s \mid N,B)$, that is, the probability distribution of the signal after the experiment has been performed, is given by Bayes theorem:[4]

$$P(s \mid N,B) = \frac{\int_0^\infty P(N,B \mid s,t;\mu,t_B) P(s,\mu)}{\int_0^\infty \int_0^\infty P(N,B \mid s,t;\mu,t_B) P(s,\mu)} , \quad (20)$$

where the integral in the numerator is over the background mean μ and $P(s,\mu) = H(s,\mu) ds\, d\mu$ is the prior probability.

III. PRIOR PROBABILITY FROM PRIOR IGNORANCE

In the previous section we were able to derive the direct probability, or likelihood function, $P(N,B \mid s, t;\mu,t_B)$ from a well-established principle. Unfortunately, there are no well-established principles for deriving the prior probability distribution which, as Jeffreys has it, "will enable the theory to begin." In particular, there are no agreed upon principles whereby the form of our prior probability $P(s,\mu)$ can be derived given the following minimal prior information: the point (s,μ) is known to lie in the space $[0,\infty) \otimes [0,\infty)$.

However, while not firmly established, the proposal by Jaynes alluded to earlier is a serious attempt to address this well-known difficulty of the non-"personal" Bayesian statistics.

Here we apply invariance arguments in the spirit of Jaynes proposal to deduce the form of the prior probability. These arguments are based on the observation that our vague knowledge concerning the parameters is supplemented by our knowledge of the functional form of the likelihood function for our experiment. Therefore, we know, in principle, the transformations which leave the likelihood function invariant. Now let us suppose that two experimenters have *exactly* the same prior information regarding the rate parameters of the likelihood function; in this case, they know that s and μ are non-negative numbers. Given the vagueness of their prior knowledge they have no choice but to consider *all* possible hypotheses regarding the magnitude of s and μ, and to assign to each hypothesis a prior probability. Following Jaynes we assume that these experimenters, being rational, must agree on the assignment of prior probabilities.[8] More precisely: if they have the same prior information and if they make equivalent hypotheses about the signal and background rates they must assign the same prior probability to the particular hypothesis under consideration. Furthermore, since for equivalent hypotheses our experimenters must compute the same likelihood for a given experimental outcome the parameters used by one will be related, by the symmetry transformations, to those used by the other.

The above considerations suggest the following basic premise: *If, prior to performing an experiment, we know nothing more than the form of the likelihood function and the domain in which it is defined then the prior probability pertaining to this knowledge is invariant with respect to the symmetry transformations of the likelihood function.* We offer the foregoing as a reasonable point of departure; if subsequently it is shown to be unreasonable then of course it should be abandoned. In the meantime, however, we shall accept it as a kind of postulate; that is, it is held to be true without proof.

Evidently, our first task is to establish the symmetries, if any, of the likelihood function $P(N,B \mid s,t;\mu,t_B)$. The precise form of the symmetry transformations will depend, of course, on the particular pair of parameters we choose to use. It is convenient to choose those parameters in terms of which the symmetries of the likelihood function are most clearly exhibited. For our likelihood function the symmetry transformations take their simplest form when expressed in terms of the rates λ and μ_B. The transformations could also be given in terms of s and μ; however, their form would be more complicated.

We first introduce the rate parameters $Q = N/t$ and $Q_B = B/t_B$. It is evident that the likelihood function is invariant with respect to the transformations

$$\begin{aligned} &\lambda \to \lambda'/p, \quad t^{-1} \to t'^{-1}/p, \quad Q \to Q'/p , \\ &\mu_B \to \mu'_B/q, \quad t_B^{-1} \to t_B'^{-1}/q, \quad Q_B \to Q'_B/q , \end{aligned} \quad (21)$$

and

$$\begin{aligned} &\lambda \to \mu''_B/q', \quad t^{-1} \to t_B''^{-1}/q', \quad Q \to Q''_B/q' , \\ &\mu_B \to \lambda''/p', \quad t_B^{-1} \to t''^{-1}/p', \quad Q_B \to Q''/p' , \end{aligned} \quad (22)$$

where p, q, p', and q' can be any set of nonzero numbers. The transformations given in Eqs. (21) and (22) together form a three-parameter (non-Abelian) group containing a two-parameter subgroup defined by Eq. (21) (for details refer to the Appendix). Therefore, we expect the form of the prior probability to be uniquely determined.

In accordance with our basic premise the prior probability must satisfy the constraints

$$P(\lambda',\mu_B')=P(\lambda,\mu_B) \tag{23}$$

and

$$P(\lambda'',\mu_B'')=P(\lambda,\mu_B) . \tag{24}$$

If we define the density $f(\lambda,\mu_B)$ by

$$P(\lambda,\mu_B)=f(\lambda,\mu_B)d\lambda\, d\mu_B$$

we obtain from Eqs. (23) and (24), respectively, the functional equations

$$pqf(p\lambda,q\mu_B)=f(\lambda,\mu_B) \tag{25}$$

and

$$p'q'f(p'\mu_B,q'\lambda)=f(\lambda,\mu_B) . \tag{26}$$

The solution of Eqs. (25) and (26) is

$$f(\lambda,\mu_B)\propto\frac{1}{\lambda\mu_B} . \tag{27}$$

Without loss in generality, we can set the constant of proportionality to unity whereupon the prior probability takes the form

$$P(\lambda,\mu_B)=\frac{d\lambda\, d\mu_B}{\lambda\mu_B} , \tag{28}$$

which, when expressed in terms of the parameters s and μ, recalling that $\lambda=s+\mu$ and that $\mu_B=\eta\mu$, becomes

$$P(s,\mu)=\frac{ds\, d\mu}{(s+\mu)\mu} . \tag{29}$$

Equation (29) is the basic result of this paper. It is important to note that had we used, from the start, the parameters s and μ or any other pair of parameters for that matter we would still have arrived at Eq. (29), albeit via a more involved derivation. Indeed, if this were not the case the result would be mathematically inconsistent. It is a direct consequence of our basic premise. One might balk at this result; some will consider it to be decidedly nonintuitive. If so, they must reject our basic premise and replace it with something else. One alternative premise is provided by the "personal" interpretation: the prior probability should reflect an individual's personal assessment of prior information. We contend, however, that if we accept this premise then we forego all possibility of agreement on the form of the prior probability distribution; for, being a matter of individual judgment, every prior probability would be valid *a priori.* Yet it is precisely this lack of some basic principle for assigning prior probabilities, this apparent invitation to arbitrariness, rather than any philosophical disagreements about the meaning of probability that has engendered so much controversy. Not surprisingly, this has hampered the introduction of Bayesian methods in physics and other fields which presume to deal with "objective" knowledge and which therefore eschew the use of overly subjective methodologies.

What is the content of Eq. (29)? The latter shows that our prior hypotheses regarding s and μ can be divided into two classes: hypotheses which are "nonfactorizable" in the sense that neither s nor μ can be assigned prior probabilities independently of one another, and hypotheses which allow the specification of a prior probability for the signal independently of the assumed background rate; that is, hypotheses which are "factorizable." When the rates are assumed to be roughly comparable we can make only "nonfactorizable" hypotheses regarding s and μ; when $\mu \gg s$ we should assign equal probabilities to the "nonfactorizable" hypotheses which differ only in the assumed signal rate. Roughly speaking, if we hypothesize that the background rate is very large compared with the signal rate then every prior assumption regarding the magnitude of the latter should be regarded as equally plausible. If, on the other hand, we assume that $\mu \ll s$ the prior probability factorizes in terms of these parameters and we can then assign prior probabilities for the signal rate independently of the background rate, and vice versa.

We do not presume to justify Eq. (29) with the above comments; our purpose is merely to suggest that an intuitive understanding of it is possible.

IV. MAKING INFERENCES ABOUT THE SIGNAL

For the proponents of Bayesian statistics its great merit is the availability of a probability distribution which quantifies the degree of belief to be associated with different values of an unknown parameter and the fact that this distribution can be systematically modified to take account of relevant new information. The following criticism is sometimes leveled against the use of a probability distribution over a parameter space: that on physical grounds it is absurd to contemplate a distribution of values for certain quantities, for example, the speed of light in vacua, or the rest mass of an electron. However, we have already alluded to the fact that probability, according to Bayesians, need have no direct connection with relative frequencies; therefore, the fact that we may choose to represent our state of knowledge by a finite-width probability distribution makes no statement about whether or not the quantity itself assumes a distribution of values. A quantity can, of course, have a unique value; the point is that because our knowledge of this value is imprecise the probability distribution we use to describe our knowledge will have a finite width.

In this section we would like to infer certain statements about the magnitude of the signal in our model experiment. First we must calculate the posterior probability $P(s\mid N,B)$. Following the method outlined in Ref. 9 we obtain

$$P(s\mid N,B)=d(st)e^{-st}(st)^{N-1}\sum_{i=0}^{N-1}x_i/[\Gamma(N-i)(st)^i], \tag{30}$$

where

$$x_i=\frac{\Gamma(B+i)/[(1+\eta)^i i!]}{\sum_{j=0}^{N-1}\Gamma(B+j)/[(1+\eta)^j j!]} . \tag{31}$$

As our knowledge of the background rate μ improves, that is when B and $\eta \rightarrow \infty$ in such a way that the ratio $B/\eta \rightarrow \mu_0 t$ where μ_0 is the true background rate, x_i will approach the limit

$$x_i = \frac{(\mu_0 t)^i / i!}{\sum_{j=0}^{N-1} (\mu_0 t)^j / j!} . \tag{32}$$

We shall make two kinds of inference using $P(s \mid N,B)$: a point estimate $\hat{s}$ of the signal and an interval estimate $[0,\bar{s}]$ containing the true value of the signal rate with probability β. The upper bound $\bar{s}$ is the $100\beta\%$-confidence-level (C.L.) upper limit on the signal rate s. The quantities $\hat{s}$ and $\bar{s}$ are defined, respectively, by

$$\hat{s} = \int_0^\infty sP(s \mid N,B) \tag{33}$$

and the implicit formula

$$\beta = \int_0^{\bar{s}} P(s \mid N,B) , \tag{34}$$

and are given, respectively, by

$$\hat{s} = \left[N - \sum_{i=0}^{N-1} i x_i \right] \Big/ t \tag{35}$$

and

$$\beta = 1 - e^{-\bar{s}t} \sum_{i=0}^{N-1} \frac{(\bar{s}t)^i}{i!} \sum_{j=0}^{N-1-i} x_j . \tag{36}$$

The above formulas present a coherent solution within the non-"personal" Bayesian framework to the problem of estimating the magnitude of a small signal masked by background. The utility of these expressions lies in the fact that the idealized experiment to which these expressions refer is a reasonably good model of some real-life experiments. As examples we cite experiments searching for neutron-antineutron oscillations[7] or proton decay.[10]

In some cases the background rate estimate is not derived from an event count B. Nonetheless, we may still use Eqs. (35) and (36) approximately by deriving an effective scale factor and an effective event count in the following obvious manner:

$$\begin{aligned} \eta &\approx \hat{\mu}/(\hat{\sigma}^2 t) , \\ B &\approx \eta \hat{\mu} t . \end{aligned} \tag{37}$$

The derivation of our formulas for $\hat{s}$ and $\bar{s}$ implicitly assumes that both N and B are >0. However, we can continue the functions to $B=0$ by taking the limit of x_i as $B \rightarrow 0$. We obtain

$$x_i = \begin{cases} 1 & \text{if } i=0 , \\ 0 & \text{otherwise} . \end{cases} \tag{38}$$

Therefore, for the case $B=0$ the formulas simplify to

$$\hat{s} = N/t \tag{39}$$

and

$$\beta = 1 - e^{-\bar{s}t} \sum_{i=0}^{N-1} \frac{(\bar{s}t)^i}{i!} , \tag{40}$$

which we note are independent of η. What of the case $N=0$? Well, in this case our prior distribution, Eq. (29), gives rise to a posterior probability distribution which cannot be normalized. As a consequence we cannot make any statements which involve absolute probabilities; only statements involving the relative probabilities of different hypotheses are possible. Therefore, for $N=0$ there is no formal solution for the upper limit. This asymmetry between the cases $N=0$ and $N>0$ is a well-known result of the Jeffreys-Jaynes prior probability, to which Eq. (29) reduces in the absence of background, and is regarded by some as a difficulty. However, there is an intrinsic asymmetry between the two sets of outcomes: if $N>0$ we know with probability one that the rate $\lambda \equiv \langle N \rangle / t$ is nonzero; if we obtain $N=0$ we cannot make such an assertion. Perhaps, the lack of a formal upper limit for $N=0$ is a reflection of this basic asymmetry. In any case, we can always *assign* to the case $N=0$ the upper limit for $N=1$. That this is always possible, and also consistent, is clear from the following observation: $\bar{s}(N,B) \leq \bar{s}(N+1,B)$; therefore, if we know with 90% confidence that $s<2.3$, say, for $N=1$, then we certainly know with at least that degree of confidence that $s<2.3$ if instead of $N=1$ we had obtained the result $N=0$.

To summarize, given the data $(N,\hat{\mu},\hat{\sigma}) \equiv (N,B,\eta)$ we can, using Eqs. (35) and (36), make some plausible inferences about the magnitude of the signal rate. Moreover, if the data are binned these formulas may be applied to each bin.

V. APPLICATIONS

As an illustration of the use of the above equations we shall consider the published data from three experiments: the search for hadronic decays of the $\Upsilon(1S)$ by the LENA group; the search for flavor-changing neutral currents performed by the JADE Collaboration at DESY and the $n\bar{n}$ experiment carried out by the CRISP Collaboration at the ILL reactor in Grenoble.

A. LENA: hadronic decays of the $\Upsilon(1S)$

Our first example is taken from the LENA Collaboration which has searched for the decay $\Upsilon(1S) \rightarrow \rho^0 \pi^0$ at the DESY storage ring DORIS.[5] This experiment serves as an example of one in which the background was determined by measurement rather than by calculation. From a data sample, corresponding to an integrated luminosity of 701 nb^{-1}, obtained at the Υ resonance 2 candidate events were found with opening angles greater than 11.5°. In the continuum about the resonance 5 events were found within a data sample of 1199 nb^{-1}. The background level was estimated by normalizing the integrated ed luminosity of the continuum data to that of the data collected at the Υ. Evidently, this experiment is well described by our statistical model with $N=2$, $B=5$, and $\eta = 1199/701$, that is, 1.71. The associated upper limit is obtained from Eq. (36). The result is 3.0 events (90% C.L.) and leads to a branching ratio of

$$\begin{aligned} B(\Upsilon \rightarrow \rho^0 \pi^0) &= \frac{st}{37.7} \% \\ &= 8 \times 10^{-2} \% \quad (90\% \text{ C.L.}) . \end{aligned} \tag{41}$$

TABLE I. Point estimates of the signal rate $s=\lambda-\mu$, where $\lambda t=\langle N\rangle$ and $\mu t=\langle B\rangle$.

N \ B	0	1	2	3	4	5	6	7
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	2.00	1.67	1.50	1.40	1.33	1.29	1.25	1.22
3	3.00	2.43	2.09	1.88	1.73	1.62	1.54	1.48
4	4.00	3.27	2.77	2.43	2.19	2.01	1.88	1.77
5	5.00	4.16	3.53	3.06	2.72	2.46	2.26	2.11
6	6.00	5.10	4.35	3.77	3.32	2.98	2.71	2.50
7	7.00	6.06	5.23	4.54	3.99	3.56	3.21	2.93

The above result differs little from that quoted in Ref. 5; however, here the result takes account of the Poisson nature of the data while the LENA group makes use of standard Gaussian statistics, a procedure which for so few events is inappropriate.

B. JADE: Flavor-changing neutral currents

Here we discuss an experiment in which the background level is estimated from a Monte Carlo simulation. The JADE Collaboration has searched for flavor-changing neutral currents in the decay of the b quark.[6] The analysis was based on a sample of 21 494 hadronic events at a mean center-of-mass energy of 34.6 GeV. A total of 138 243 hadronic events were simulated which were then passed through the same trigger and analysis chain as used for the real data, leaving 106 926 events. Therefore, for this experiment $\eta=106\,926/21\,494$, that is, 4.97. Cuts were applied to select dimuon events from the decays $b(\bar{b})\to s(\bar{s})\mu^+\mu^-$ where here s stands for the strange quark. In particular, a cut, $20°<\delta<80°$, on the dimuon opening angle δ was imposed to enhance the sensitivity to an FCNC signal. Three opposite-sign, same-jet dimuon events were found with an overall efficiency of 0.167. The Monte Carlo simulation yielded 30 background events, that is, $(\hat{\mu},\hat{\sigma})=(6.0,1.1)/t$ events when scaled down by η.

Again, this experiment may be analyzed by the method explained above. The results $(N,B,\eta)=(3,30,4.97)$ imply an upper limit on a possible FCNC signal in the JADE experiment of 3.1 events (90% C.L.). For this experiment the branching ratio may be expressed as

$$B(b(\bar{b})\to s(\bar{s})\mu^+\mu^-)=\frac{st}{6.54}\% \ , \tag{42}$$

which leads to the upper limit 0.5% (90% C.L.).

C. CRISP: $n\bar{n}$ oscillations

In the CRISP experiment[7] the background level was known with about the same degree of precision as the combined level of signal and background. We may characterize this situation by taking $\eta=1$. (In fact, $\eta=0.97$ for this experiment. It is the ratio of the observation time for the real experiment to that of the control experiment.) Owing to the magnetic properties of the neutron it is possible to switch off the oscillation effect by the application of a suitable magnetic field. By this means a direct measurement of the background level can be made. From control experiments 7 background events were found while the actual experiment yielded 3 events. This result, of course, is statistically perfectly reasonable; however, within the framework of "classical" statistics it does present a problem.[11] The Bayesian solution on the other hand is straightforward and is again given by Eq.(36).

Table I shows some results for $\hat{s}t$ when $\eta=1$. The corresponding 90%-C.L. upper limits are listed in Table II. For this experiment the $n\bar{n}$ mixing time $\tau_{n\bar{n}}$ may be written as

$$\tau_{n\bar{n}}=\frac{1.48}{\sqrt{st}}10^6 \text{ sec} \ , \tag{43}$$

which, with the upper limit of 3.33 antineutron events, corresponding to the results $(N,B,\eta)=(3,7,1)$ gives the

TABLE II. Upper limits for the signal rate $s=\lambda-\mu$ at 90% C.L., where $\lambda t=\langle N\rangle$ and $\mu t=\langle B\rangle$. For the case $N=0$, refer to Sec. IV of the text.

N \ B	0	1	2	3	4	5	6	7
0								
1	2.30	2.30	2.30	2.30	2.30	2.30	2.30	2.30
2	3.89	3.51	3.27	3.11	3.00	2.91	2.84	2.78
3	5.32	4.75	4.32	4.01	3.77	3.59	3.44	3.33
4	6.68	5.99	5.43	4.99	4.63	4.35	4.12	3.93
5	7.99	7.24	6.59	6.03	5.57	5.18	4.87	4.60
6	9.28	8.49	7.76	7.12	6.56	6.08	5.68	5.34
7	10.53	9.72	8.95	8.24	7.60	7.04	6.55	6.14

lower limit 0.8×10^{6} sec for the neutron-antineutron mixing time.

We note the absence, in both tables, of negative values. This, of course, is to be expected; the parameter space has been restricted to the positive quadrant. Therefore, negative values cannot occur. This is in striking contrast to the "classical" calculation[11] which, for values of B sufficiently greater than N, will eventually lead to negative upper limits. In the "classical" case this reflects the fact that the *unbiased* estimate of the signal rate is given by $\hat{s}=(N-B)/t$, which can of course be negative. However, for values of $N \gg B$ the Bayesian and classical calculations give essentially the same numerical results.

VI. CONCLUSIONS

Many of the commonly used statistical techniques are valid strictly in the limit of large signals and samples. As a consequence the analysis of small signals using the standard methods is rendered difficult. Indeed, it is easy to think of statistical problems in high-energy physics for which there are no satisfactory solutions using standard methods. Bayesian methods offer a simple solution to this and related problems. For a recent interesting application of Bayesian statistics in high-energy physics we cite the methods discussed in Ref. 12. However, granted the acceptance of the subjective interpretation of probability one point of contention still remains: the manner in which prior probabilities are to be assigned. We believe that by accepting the premise that the prior probability should be invariant with respect to the symmetry transformations of the likelihood function some progress can be made towards eliminating what critics perceive as the arbitrariness of Bayesian statistics with regard to the assignment of prior probabilities. This premise pertains to the case in which the form of the likelihood function and the domain in which its parameters are defined is the only prior information we have.

We have tried to present a coherent Bayesian analysis of small signals buried in background noise. In particular, granted our premise, we have obtained a unique prior distribution for the problem addressed. This prior distribution differs from the divergent form tentatively suggested in Ref. 9. We regard the form given in that paper as untenable because it leads to an unnormalizable posterior probability for *all* values of N and it is incompatible with our basic premise. On the other hand it is clear from the results presented in that paper that use of a uniform prior probability distribution $ds\,d\mu$, will avoid the first problem; however, for the model used here such a prior probability is again inconsistent with the principle of prior probability invariance under the symmetry group of the likelihood function.

The results obtained in this paper are directly applicable to real experiments; a few examples from the literature were discussed. A particular advantage of our method is that the uncertainty in the background estimate can be accounted for in a manner which is both simple and systematic. As a consequence the results from different experiments can be compared without difficulty. From a practical viewpoint the primary difference between our Bayesian results and those from a more orthodox approach is the explicit elimination in the former case of inferences that might be construed as unphysical, for example, negative values for the signal. However, it should be recognized that inferences based on Bayesian statistics[4] are interpreted differently from those based on more orthodox statistical approaches.[1]

ACKNOWLEDGMENTS

I would like to thank Fred James and Maria Fidecaro for valuable discussions on this subject. This work was supported in part by the National Science Foundation.

APPENDIX

If we construct a column vector from the quantities λ, t^{-1}, Q, μ_B, t_B^{-1}, and Q_B, Eqs. (21) and (22) define the transformation matrix

$$A(p,q,r)\equiv\begin{pmatrix} rp & 0 & 0 & p-rp & 0 & 0\\ 0 & rp & 0 & 0 & p-rp & 0\\ 0 & 0 & rp & 0 & 0 & p-rp\\ q-rq & 0 & 0 & rq & 0 & 0\\ 0 & q-rq & 0 & 0 & rq & 0\\ 0 & 0 & q-rq & 0 & 0 & rq \end{pmatrix}, \tag{44}$$

where $|p|$ and $|q|>0$ and r is a discrete parameter with values 0 and 1. Given this matrix representation it is straightforward to verify that

$$\begin{gathered} A(p,q,1)A(p',q',1)=A(pp',qq',1),\quad A(p,q,1)A(p',q',0)=A(pp',qq',0)\,,\\ A(p,q,0)A(p',q',0)=A(pq',qp',1)\,, \end{gathered} \tag{45}$$

thereby demonstrating that the infinite set of elements $\{A(p,q,r)\}$ form a three-parameter non-Abelian group with $A(1,1,1)$ as the identity element.

*Now at the Fermi National Accelerator Laboratory, Batavia, IL 60510.

[1]M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Griffin, London, 1973).

[2]E. T. Jaynes, IEEE Trans. Syst. Man Cybern SSC-4, 227 (1968).

[3]L. J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954).

[4]H. Jeffreys, *Theory of Probability* (Clarendon, Oxford, 1948).

[5]LENA Collaboration, B. Niozyporuk *et al.*, Z. Phys. C **17**, 197 (1983).

[6]JADE Collaboration, W. Bartel *et al.*, Phys. Lett. **132B**, 241 (1983). See also C. K. Bowdery, University of Lancaster report, 1986 (unpublished).

[7]CRISP (CERN, RAL, ILL, Sussex University, Padova University) Collaboration, G. Fidecaro *et al.*, Phys. Lett. **156B**, 122 (1985).

[8]We are all familiar with the observation that apparently rational individuals often disagree; this is so because no two individuals ever have *precisely* the same mental history, that is, prior information. However, if two individuals can at least agree to exclude all prior information other than that which is relevant to the problem at hand then if they truly think according to the laws of reason they must surely agree. The difficulty, of course, is arriving at the first agreement.

[9]H. B. Prosper, Nucl. Instrum. Methods **A240**, 236 (1985).

[10]IMB Collaboration, T. W. Jones *et al.*, Phys. Rev. Lett. **52**, 720 (1984).

[11]H. B. Prosper, Nucl. Instrum. Methods **A238**, 500 (1985).

[12]J. F. Gunion and D. E. Soper, Phys. Rev. D **35**, 179 (1987).