

Estimation of asymmetry in physics

Simon P. Wilson¹ and Kevin J. Coakley²

¹*Department of Statistics, University of Dublin, Trinity College, Ireland*

²*National Institute of Standards and Technology, Boulder, Colorado 80303*

(Received 1 May 1995; revised manuscript received 28 September 1995)

The notion of asymmetry arises in many physical experiments, being a natural expression of the relative difference between two quantities. In this paper, we discuss the estimation of asymmetries between the rates of two event processes subject to a common background. Such asymmetries have many applications in solid-state physics—a measurement of the difference in arrival rates of electrons with one of two possible spin orientations to a detector, for example. We describe three techniques for estimating asymmetries from observations subject to uncertainty: a naive method based on an unbiased estimate and a Gaussian approximation, a bootstrap approach, and a Bayesian approach. Differences in the interpretation of the estimates are discussed and their performances are compared by means of a simulation study. The aim is both to review the methodology on estimation of asymmetries and to contribute to the more general discussion of the relative merits of frequentist and Bayesian inference.

PACS number(s): 02.50.-r, 02.70.Lq, 34.90.+q

I. INTRODUCTION

Physicists are becoming more conscious of the need for sound statistical analysis of data that is obtained under uncertainty. Along with this awareness has come the realization that the outcome of a statistical analysis often depends on which of the two statistical paradigms—classical or Bayesian—is adopted. This is particularly the case in particle physics, a point that is admirably discussed in a recent paper by Cousins [1].

In this paper we investigate the estimation of asymmetries. The asymmetry between two comparable, non-negative numerical quantities X and Y , denoted by R , is given by

$$R = \frac{X - Y}{X + Y}. \quad (1)$$

The purpose of R is to give a relative measure of the difference between X and Y ; R takes values between -1 and $+1$, is dimensionless, and is also scale invariant. It is used in optics, for example, where it is a measure of light polarization, and can be applied to the analysis of thin films. Because an asymmetry is a relative scale, it allows one to ignore any scaling effects on X and Y that are present in an experiment—such as detector sensitivity—that otherwise would have to be carefully measured and accounted for.

The motivating example for our work comes from scanning electron microscopy with polarization analysis (SEMPA)—we refer to Pierce [2] and Pierce *et al.* [3] for details—but the ideas presented easily extend to other situations of interest, in particular where one is interested in the difference between two event rates in the presence of background noise. There are many more examples of asymmetries to be found in the physics literature; see, for example, [4–6].

The paper begins with a description of three different estimation techniques: a classical or “naive” technique, a

bootstrap, and a Bayesian approach. Then their properties are investigated by means of simulation experiments. The question of greatest interest is whether any method is clearly superior to the others; not surprisingly, there is no straight answer to this question, although we do conclude that the most commonly used classical estimation technique does possess undesirable properties that are not present in the other two. We demonstrate that the methods produce almost identical analyses for many situations, but that they differ in certain cases; namely, when there is a large component of noise in the data, when the asymmetry is close to a boundary value, or when the number of counts is small.

The general problem is as follows. We conduct an experiment where we observe the occurrence of two types of event (in the SEMPA example, this would be the detection of electrons of either spin, relative to some predetermined axis). We observe n_1 occurrences of the first event and n_2 occurrences of the second. The number of observed events of the two types are Poisson random variables with unknown means λ_1 and λ_2 ; such an assumption is motivated by a maximum entropy argument (see [7]). From the observed counts, we desire to estimate the asymmetry in means, $R = (\lambda_1 - \lambda_2) / (\lambda_1 + \lambda_2)$. The problem can be viewed as an extension of the situation where a single signal is recorded in the presence of a background, where various models, using both the frequentist and Bayesian methodology, have been proposed (see [7–10]).

Unfortunately, within n_1 and n_2 there are background events that have been counted and cannot be distinguished from the true events of interest (in the SEMPA example, these will be background electrons unconnected with the microscopy). Here we assume that both our counts are confounded by a background count that is a Poisson random variable with unknown mean λ , although we note that what we have to say can be extended to the case where both signals have a background of

different magnitude. We also note that this does not mean that the number of background counts within n_1 and n_2 are the same; they just have the same distribution. We have made an implicit assumption here that all counts are taken over the same interval of time, though we note that all the methods we discuss are easily modified to account for differing measurement times.

In order to gauge the unknown size of the background count, a separate measurement of the background alone is taken either before or after the experiment. We denote the size of this background count as n . We are assuming that this background is directly measured, and not inferred from theoretical considerations or by a Monte Carlo simulation.

The first aim is to produce a point estimate for the value of R from the data (n_1, n_2, n) . Having obtained a point estimate for R , it is then important to calculate the magnitude of the possible error in the estimate. Such an estimate takes the form of an interval over which, in light of the data, we believe R is likely to lie. All the methods we discuss will calculate an interval estimate to go with the point estimate.

The next three sections describe the three estimation techniques in detail. Section V compares the techniques, and we close with a discussion of the relative merits of the three techniques.

II. CLASSICAL ESTIMATION

This is the most commonly used estimation technique, which we will also call *naive* estimation. Suppose n_1 , n_2 , and n are the observed counts from the experiment and the observed background count, respectively. The naive estimator is motivated by noting that, since they are means of Poisson counts, unbiased estimates for λ_1 and λ_2 are $n_1 - n$ and $n_2 - n$, respectively. Substitution into the definition of asymmetry yields the *naive estimator*

$$\hat{R} = \frac{n_1 - n_2}{n_1 + n_2 - 2n}. \quad (2)$$

If either $n_1 < n$ or $n_2 < n$ then \hat{R} lies outside the admissible values of R (that is, $[-1, 1]$). When this happens, it is interpreted as evidence that R lies at a boundary value.

A confidence interval for R is found by the propagation of errors (POE) technique; see [11]. The variance of \hat{R} is approximated by a first order Taylor expansion of \hat{R} as a function of n_1 , n_2 , and n and from this, an expression for $\sigma_{\hat{R}}^2$ is obtained as

$$\sigma_{\hat{R}}^2 \approx \frac{4}{(n_1 + n_2 - 2n)^4} [n_1(n_2 - n)^2 + n_2(n - n_1)^2 + n(n_1 - n_2)^2]. \quad (3)$$

Now \hat{R} is assumed to be approximately normally distributed. Under this additional assumption, an approximate 95% confidence interval for R is $(\hat{R} - 1.96\sigma_{\hat{R}}, \hat{R} + 1.96\sigma_{\hat{R}})$.

Unfortunately, this naive estimation technique does possess some undesirable properties in the presence of high background counts. If $n_1 + n_2 < 2n$ then the denom-

inator of \hat{R} is negative and the estimate takes on fatuous values; for example, if $n_1 = 10$, $n_2 = 3$, and $n = 20$, then $\hat{R} = -0.26$, even though the data suggests that R is positive. If $n_1 + n_2 = 2n$, then \hat{R} and $\sigma_{\hat{R}}$ are infinite in value and, since this event occurs with a positive probability, \hat{R} has infinite mean and variance. So, \hat{R} is not an unbiased estimator for R and the confidence interval is potentially of infinite length.

III. BOOTSTRAP ESTIMATION

Because of the difficulties with the naive estimator, a bootstrap method is proposed by Coakley *et al.* [11], where a confidence interval estimate of R is obtained, and by Coakley [12], where a bootstrap point estimator for R is described.

In this method, n_1 , n_2 , and n are first observed. Then, independent samples are made from three independent Poisson random variables with means n_1 , n_2 , and n . Let the i th sample from the Poisson distributions be denoted by (n_1^i, n_2^i, n^i) . This set of values is checked to see if it satisfies the following three constraints:

$$\begin{aligned} n^i &\leq n_1^i, \\ n^i &\leq n_2^i, \\ 2n^i &< n_1^i + n_2^i. \end{aligned} \quad (4)$$

Those samples that do not satisfy the three constraints are discarded. From the remaining replications that do satisfy the constraints, the naive estimator \hat{R} [see Eq. (2)] is calculated. Satisfaction of the constraints ensures that the resulting collection of \hat{R} 's lie in $[-1, 1]$, so that they are all admissible, and also ensures that the denominator term of (2) is positive, so that the sign of the \hat{R} 's is not switched.

Once a large number of bootstrap samples have been generated, their mean is taken as a point estimate of R . A histogram of asymmetry values is formed and from this a confidence interval for R is constructed. If $n_1 < n$ or $n_2 < n$ then an interval with an endpoint at either -1 or 1 is given; otherwise the authors use a modified version of the bias correction method of Efron [13] to produce their interval estimate. In both cases, the confidence interval for R always lies in $[-1, 1]$.

IV. BAYESIAN ESTIMATION

In this section we propose an alternative approach under the Bayesian paradigm. The principal obstacle to the implementation of the Bayesian approach is, of course, the specification of a prior distribution on the three rate parameters λ_1 , λ_2 , and λ , and a large part of this section is devoted to how one might proceed with the specification.

We observe two counts from the experiment— n_1 and n_2 —and we also measure the background count n . All the counts are Poissonian and so the likelihood of the three observations is

$$P(n_1, n_2, n | \lambda_1, \lambda_2, \lambda) = \frac{(\lambda_1 + \lambda)^{n_1} (\lambda_2 + \lambda)^{n_2} \lambda^n}{n_1! n_2! n!} \times e^{-(\lambda_1 + \lambda_2 + 3\lambda)} \tag{5}$$

A. Specification of the prior distribution

Our likelihood is a function of λ_1 , λ_2 , and λ , so if we can specify a prior distribution on these three parameters then a simple application of Bayes' law will yield a posterior distribution. Since R is a function of λ_1 and λ_2 , both a prior and the associated posterior on R are implied. By using the simple laws of transformation of random variables, we can say that if $f(\lambda_1, \lambda_2)$ is a probability density on the two rate parameters, then the density of $R = (\lambda_1 - \lambda_2) / (\lambda_1 + \lambda_2)$ must be

$$\pi_R(r) = \frac{2}{(1+r)^2} \int_{\forall \lambda_1} \lambda_1 f \left[\lambda_1, \frac{1-r}{1+r} \lambda_1 \right] d\lambda_1 \tag{6}$$

We offer three forms for the prior density that might be used. The first is an informative prior that attempts to incorporate valid *a priori* information into the analysis. The second two priors are attempts to convey a state of ignorance or indifference as the state of prior knowledge about the parameters λ_1 , λ_2 , and λ .

1. Using an informative prior

We wish to incorporate some information about the likely values of the parameters λ_1 , λ_2 , and λ into our analysis. Unfortunately, it seems likely that much of the prior knowledge would concern R directly, and not these rate parameters. In this section we propose a certain form for the distribution of the rate parameters that enables us to incorporate information about R .

First of all, the background count rate λ is assumed stochastically independent of λ_1 and λ_2 . We declare λ to be Γ distributed with scale parameter α and shape parameter β ; this is the conjugate prior for the rate parameter of a Poisson process:

$$\pi(\lambda | \alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\alpha\lambda} \tag{7}$$

For the event rate parameters of interest, λ_1 and λ_2 , an assumption of stochastic independence is not reasonable because they are properties of the same experiment and are subject to many common scaling factors in the experimental setup and environment (in SEMPA, two such factors are electron beam intensity and detector efficiency). We assume that both λ_1 and λ_2 are also distributed as Γ random variables with shape parameters β_1 and β_2 , respectively, but with a common scale parameter ξ . Conditional on ξ , λ_1 , and λ_2 are stochastically independent,

$$\pi(\lambda_1, \lambda_2 | \beta_1, \beta_2, \xi) = \prod_{i=1}^2 \frac{\xi^{\beta_i}}{\Gamma(\beta_i)} \lambda_i^{\beta_i-1} e^{-\xi\lambda_i} \tag{8}$$

Dependence is introduced when ξ is itself described by a Γ distribution with a scale parameter a and shape parameter b . By averaging out ξ we obtain a joint prior density for λ_1 and λ_2 ,

$$\begin{aligned} \pi(\lambda_1, \lambda_2 | a, b, \beta_1, \beta_2) &= \int \pi(\lambda_1, \lambda_2 | \beta_1, \beta_2, \xi) \pi(\xi | a, b) d\xi \\ &= \frac{a^b \Gamma(b + \beta_1 + \beta_2)}{\Gamma(b) \Gamma(\beta_1) \Gamma(\beta_2)} \\ &\quad \times \frac{\lambda_1^{\beta_1-1} \lambda_2^{\beta_2-1}}{(a + \lambda_1 + \lambda_2)^{b + \beta_1 + \beta_2}} \end{aligned} \tag{9}$$

Why choose this form for the prior of λ_1 and λ_2 ? If one substitutes Eq. (9) into Eq. (6), then one sees that the implied prior on R is

$$\begin{aligned} \pi_R(r | \beta_1, \beta_2) &= \frac{\Gamma(\beta_1 + \beta_2)}{2^{\beta_1 + \beta_2 - 1} \Gamma(\beta_1) \Gamma(\beta_2)} \\ &\quad \times (1+r)^{\beta_1-1} (1-r)^{\beta_2-1} \end{aligned} \tag{10}$$

for $-1 \leq r \leq 1$, which is a β distribution in $[-1, 1]$ with two parameters β_1 and β_2 (a and b having been eliminated). The β is a very flexible density that can take many shapes and, because this β distribution has support $[-1, 1]$, we immediately remove the possibility of obtaining the inadmissible estimates and confidence intervals that plague the naive estimator.

Information on the likely values of R , quantified as a β distribution, specifies β_1 and β_2 and so gives information on the prior distribution for λ_1 and λ_2 . This assessment is helped by noting that

$$\begin{aligned} E(R | \beta_1, \beta_2) &= \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}, \\ V(R | \beta_1, \beta_2) &= \frac{4\beta_1\beta_2}{(\beta_1 + \beta_2)^2 (1 + \beta_1 + \beta_2)}, \end{aligned} \tag{11}$$

and

$$\text{mode}(R | \beta_1, \beta_2) = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2 - 2},$$

where $E(\dots)$ denotes expectation and $V(\dots)$ denotes variance. To complete the specification of Eq. (9), values must be assigned to the parameters a and b , relating to ξ .

So, the prior distribution on λ_1 , λ_2 , and λ is specified with six hyperparameters: a , b , β_1 , β_2 , α , and β . The specification procedure would involve first assessing β_1 and β_2 , using information about R , and then specifying the remaining parameters using knowledge about the rate parameters directly.

Standard calculations, from Bayes' law and the simple transformation of variables (6), gives the posterior density of R , following observation of data (n_1, n_2, n) , to be of the following form:

$$\pi_R(r|n_1, n_2, n) \propto (1-r)^{\beta_2-1} (1+r)^{b+\beta_1-1} \times \int_0^\infty \int_0^\infty \frac{\lambda_1^{\beta_1+\beta_2-1} \lambda^{n+\beta-1} (\lambda_1+\lambda)^{n_1} \left[\frac{1-r}{1+r} \lambda_1+\lambda \right]^{n_2}}{[2\lambda_1+a(1+r)]^{b+\beta_1+\beta_2}} e^{-[2/(1+r)]\lambda_1-(\alpha+3)\lambda} d\lambda_1 d\lambda. \quad (12)$$

Evaluation of this integral must be done numerically; shortly, we discuss this in more detail.

2. Using a uniform prior

Commonly, one would like inference on R to be independent of any *a priori* knowledge of its likely values, other than that it must lie in $[-1, 1]$. The naive and bootstrap estimates are both a function of the data only and so possess the necessary independence. How can the Bayesian approach be employed in this situation? One way is by specifying prior distributions that convey as little information as possible, so that the posterior distribution of R is as dependent upon the data alone as it can be.

One definition of such independence is that we are indifferent between the possible values of R *a priori*, that is, our prior distribution for R is uniform on $[-1, 1]$. Here, we can employ the prior of Sec. IV A 1 since the uniform distribution is a special case of the beta distribution, with $\beta_1 = \beta_2 = 1$. So, we start by adopting the informative prior of Sec. IV A 1 with these values for β_1 and β_2 . We still have the problem of specifying ξ , which would again imply the use of prior knowledge about λ_1 and λ_2 . By assuming that ξ is a fixed number and not

random, and letting ξ tend to zero, the density of λ_1 and λ_2 is given by (8) and becomes increasingly flat. In the limit, as ξ becomes zero, it fails to be a density function, but would essentially convey no information about the values of λ_1 and λ_2 . If $\xi = 0$ then the prior distribution on λ_1 and λ_2 is called *improper* to reflect the fact that it is not a proper density. However, we may still proceed to obtain a proper posterior distribution on the rate parameters via Bayes' theorem. We also note that by letting ξ take a value, as opposed to being a random variable, λ_1 and λ_2 are independent. Given that we have assigned a scale parameter of 0 and a shape parameter of 1 to the Γ distributions of both λ_1 and λ_2 , we do the same with λ ; thus, $\alpha = 0$ and $\beta = 1$.

So if we wish to express indifference on R , we can opt for an improper prior and use the following values for the parameters: $\alpha = \xi = 0$, $\beta_1 = \beta_2 = \beta = 1$. We have argued that this choice of parameters conveys very little information and implies indifference as to the values of R *a priori*.

Routine calculations [Bayes' law and the application of (6)] yield a posterior distribution of R , given the data n_1 , n_2 , and n , of the form

$$\pi_R(r|n_1, n_2, n) \propto \frac{1}{(1+r)^2} \int_0^\infty \int_0^\infty \lambda_1 \lambda^n (\lambda_1 + \lambda)^{n_1} \left[\frac{1-r}{1+r} \lambda_1 + \lambda \right]^{n_2} e^{-\{[2/(1+r)]\lambda_1 + 3\lambda\}} d\lambda_1 d\lambda. \quad (13)$$

Again, we address the evaluation of this integral shortly, but for now we note that at $r = -1$ there is a limit which is awkward numerically, but that this can be addressed by noting

$$\pi_R(r|n_1, n_2, n) = \pi_R(-r|n_2, n_1, n). \quad (14)$$

So, we can calculate the posterior density near $r = -1$ as that of another posterior density near $+1$, where there is no such numerical problem.

3. Using a reference prior

The notion of indifference is not the only one that can be used to form a noninformative prior distribution. Another commonly invoked invariance concept is due to Jeffreys [14]. Jeffreys argues that, in the absence of any knowledge concerning the parameters apart from their domain, a reasonable form for the prior density is one which is invariant to symmetry transformations of the

likelihood function. In other words, our absence of knowledge on model parameters should not depend on such transformations of the probability model.

Prosper [7] calculates the reference prior for the case of a single Poisson count (with mean λ_1) with additive background (of mean λ) and shows that it has the form

$$\pi(\lambda_1, \lambda) \propto \frac{1}{(\lambda_1 + \lambda)\lambda}. \quad (15)$$

One can extend Prosper's reasoning to the case of two Poisson counts with background; not unexpectedly, the reference prior for our case is of the form

$$\pi(\lambda_1, \lambda_2, \lambda) \propto \frac{1}{(\lambda_1 + \lambda)(\lambda_2 + \lambda)\lambda}. \quad (16)$$

Although this is an improper prior density, one can still proceed (as we had with the previous improper prior) and calculate a proper posterior density, only this time there is a restriction in that all three counts must be nonzero. The posterior for R is given by

TABLE I. Analysis of low background count data.

Data		n	Bayes' (Uniform)		Bayes' (Ref.)		Bootstrap		Naive/POE	
n_1	n_2		mean	P.I.	mean	P.I.	mean	C.I.	\hat{R}	C.I.
100	50	0	0.333	(0.18,0.48)	NA	NA	0.332	(0.18,0.48)	0.333	(0.18,0.48)
100	50	15	0.419	(0.22,0.62)	0.419	(0.22,0.62)	0.419	(0.23,0.62)	0.417	(0.22,0.61)
200	100	30	0.418	(0.275,0.556)	0.418	(0.275,0.555)	0.419	(0.28,0.56)	0.417	(0.28,0.56)

$$\pi_R(r|n_1, n_2, n) \propto \frac{1}{(1+r)^2} \int_0^\infty \int_0^\infty \lambda_1 \lambda^{n-1} (\lambda_1 + \lambda)^{n_1-1} \left[\frac{1-r}{1+r} \lambda_1 + \lambda \right]^{n_2-1} e^{-\{[2/(1+r)]\lambda_1 + 3\lambda\}} d\lambda_1 d\lambda, \quad (17)$$

for $n_1, n_2, n > 0$. If any of the counts are zero, the posterior is not well defined.

Note the similarity between this posterior density and that under the uniform prior assumption; in fact, under the reference prior, the posterior is that obtained under the uniform prior assumption but with one subtracted from all the observed counts. Thus, for large counts, there will be little difference between the posterior densities under the two improper priors. For small counts, some investigation is needed to ascertain the size of the difference.

B. Computational issues

The evaluation of the posterior densities of Eqs. (12), (13), and (17) must be done numerically. In each case, this requires that a double integral over the positive quadrant be approximated accurately. The integrand is the posterior density of λ_1, λ_2 , and λ and in all the cases is well behaved, so the principle obstacle to the integral's evaluation is finding the appropriate finite region of the quadrant where the mass of the integrand is enclosed.

For the two cases where the prior is improper, the posterior is essentially the likelihood. We know where the mass of the likelihood function is located; the maximum likelihood estimates for λ_1, λ_2 , and λ are $\max(0, n_1 - n)$, $\max(0, n_2 - n)$, and n , respectively, and they give both the modal values of the likelihood and estimates for the variance of the three parameters. From this information, we can form a finite region of the quadrant over which we are confident the mass of the integrand lies, and then proceed with the numerical approximation using any one of the common techniques.

The case of the proper prior is a little more complex, since here the posterior is a product of a prior and the likelihood. However, we know where the mass of the prior is located in the quadrant (since we completely specified it) and we have argued that we know where the likelihood is located, so the location of their product can be calculated easily. From this, a suitable finite region can be formed and the numerical integration performed.

V. COMPARATIVE STUDY

In this section, we will compare the naive, bootstrap, and Bayesian techniques for estimating asymmetry, with the aim of highlighting any disagreements between them.

For the frequentist models, we will compute the point and 95% confidence interval estimates that have been described. For the Bayesian method, we will use one of the improper priors, since these are the ones most likely to be used in an analysis of experimental data. As a point estimate in the Bayesian approach, we take the posterior mean of R , and for an interval estimate we take the shortest 95% posterior probability interval, that is, the interval of shortest length over which 95% of the posterior probability lies (since the posterior is always unimodal, this interval is unique).

It should be emphasized that the two types of interval estimates—confidence interval and posterior probability interval—have different interpretations. The classical and bootstrap approaches come under the frequentist paradigm and so the confidence interval must be interpreted in the following way:

TABLE II. Analysis of high background count data.

Data		n	Bayes'		Bootstrap		\hat{R}	POE	
n_1	n_2		mean	P.I.	mean	C.I.		C.I.	
220	59	65	0.917	(0.796,1)	0.924	(0.803,1)	1.081	(0.769,1.39)	
243	63	64	0.911	(0.791,1)	0.916	(0.796,1)	1.011	(0.760,1.26)	
227	50	53	0.924	(0.816,1)	0.928	(0.817,1)	1.035	(0.796,1.27)	
224	50	56	0.929	(0.824,1)	0.935	(0.829,1)	1.074	(0.810,1.34)	

TABLE III. Summary of the six simulation studies.

Study no.	λ_1	λ_2	λ	R
1	19	1	5	0.9
2	190	10	50	0.9
3	1900	100	500	0.9
4	4	4	20	0
5	40	40	200	0
6	400	400	2000	0

The 95% confidence interval for R is the interval such that, were the experiment to be repeated many times and this interval constructed, the true value of R would lie within it 95% of the time.

On the other hand, the Bayesian shortest 95% posterior probability interval is interpreted differently:

In light of the *a priori* assumptions on R and the observed data, there is a 95% probability that R lies in this interval. The interval is the shortest in length where this is true.

Thus, a confidence interval is a property of the experiment; in particular, it is *not* a probability statement and should not be interpreted in the same way as the Bayesian posterior probability interval.

To begin the comparison, we analyze three typical "nice" data sets, where the overall size of count is large and the background count is relatively small. Table I gives estimates of the asymmetry from four sources: the posterior distributions of R from both the uniform and reference priors, the bootstrap estimate, and the naive estimator with POE. The table displays the mean and shortest 95% probability interval of the posterior distributions, the mean and 95% confidence interval from the bootstrap, and the naive point and 95% confidence interval estimates.

The four approaches give very similar point and interval estimates for these well-behaved sets of data. In the first case, the posterior distribution obtained from the reference prior is still improper because one count was zero, and so cannot be used. This preliminary set of examples also shows that for moderately high counts the two uninformative priors give almost identical results. So, in what follows, where we analyze counts of this or greater magnitude, we use the uniform prior in the Bayesian analysis with the assumption that the reference prior would produce a very similar analysis.

Of most interest are cases where estimates from the

TABLE V. Summary of simulation study 2 ($\lambda_1 = 190$, $\lambda_2 = 10$, and $\lambda = 50$).

	Av. point estimate	Mean sq. error	Av. interval estimate	Coverage proport.
Naive	0.905	0.009 84	(0.709,1.101)	0.958
Bootstrap	0.865	0.004 47	(0.741,0.986)	0.974
Bayes'	0.862	0.004 52	(0.734,0.983)	0.970

three methods are likely to differ. One such case is where the background count is high and dominates one of the signals. Table II displays the analysis of several sets of such data and we observe that the Bayesian and bootstrap interval estimates are in broad agreement. In the table, the values of \hat{R} are according to Eq. (2) and are all inadmissible. The interval estimates provided by the POE also lie mostly outside admissible values of R . The Bayesian and bootstrap interval estimates are similar, although the Bayesian probability interval is always a little longer.

A. Simulation studies

As a final comparison between the models, simulation studies were run using the type of data that would most likely distinguish between the methods. The parameters λ_1 , λ_2 , and λ were fixed and then 4000 Poisson samples of n_1 , n_2 , and n were generated. Point and interval estimates were calculated for each sample, using all three methods, and these were summarized for comparison.

Six simulation studies were run. All of them simulated data that would emphasize differences between the methods, since they possessed a relatively high background count, an asymmetry near a boundary value, low counts overall, or a combination of these. The values of λ_1 , λ_2 , and λ for the three studies are given in Table III, and we can see that the studies are in two groups: the first three studies have an asymmetry of 0.9 and the last three have an asymmetry of 0. Within each group the magnitude of the counts ranges over two orders of magnitude. The studies with larger counts will have a larger signal to noise ratio; thus studies 1 and 4 have a low signal and studies 3 and 6 have a much higher signal.

Before moving to the studies, the question of how to compare the different techniques must be addressed. There are some difficulties to conducting any comparison.

(1) There are certain situations where one of the esti-

TABLE IV. Summary of simulation study 1 ($\lambda_1 = 19$, $\lambda_2 = 1$, and $\lambda = 5$).

	Av. point estimate	Mean sq. error	Av. interval estimate	Coverage proport.
Naive	0.981	0.657	(-0.476,2.438)	0.955
Bootstrap	0.734	0.0467	(0.401,0.991)	0.964
Bayes' (Uniform)	0.697	0.0578	(0.358,0.982)	0.963
Bayes' (Ref.)	0.714	0.0508	(0.394,0.980)	0.942

TABLE VI. Summary of simulation study 3 ($\lambda_1=1900$, $\lambda_2=100$, and $\lambda=500$).

	Av. point estimate	Mean sq. error	Av. interval estimates	Coverage proport.
Naive	0.901	0.000 979	(0.840,0.961)	0.947
Bootstrap	0.900	0.000 873	(0.843,0.960)	0.942
Bayes'	0.900	0.000 858	(0.843,0.959)	0.927

mates is not defined. If $n_1 + n_2 - 2n = 0$ then the naive estimate is not defined, whereas if any count is zero the posterior density under the reference prior is not proper. For the purposes of comparison, we ignore any samples where either of these events have occurred and merely note that they have arisen. In study 1, there are three cases where $n_1 + n_2 - 2n = 0$ and 31 cases where there is a zero count, so we ignore 34 of the 4000 simulations. In study 4, we ignore 133 cases where $n_1 + n_2 - 2n = 0$ and in study 5, we ignore two cases where $n_1 + n_2 - 2n = 0$ (there are no zero counts in these two studies).

(2) The aims of the estimates are different under the two paradigms. For the frequentist, the estimates want to be as close as possible to the "true" value of asymmetry. For the Bayesian, there is no true value of asymmetry; R is an unknown quantity and the posterior distribution reflects the current state of belief as to its likely values. For the simulation studies, however, we have fixed R beforehand and so it becomes of general interest to see how close our estimates are to this value, on the average. So, our comparisons will make use of (frequentist) concepts of closeness like biasedness, mean square error, and coverage.

For the point estimates, we look at two measures: biasedness and mean square error. We take the average of each point estimate over the 4000 simulations as the expected value of the point estimate and can then compare it with the true asymmetry value. Biasedness is a measure of the location of the point estimate, but gives no measure of the likely error; for this we take mean square error (MSE), which is the average of the squared difference between the point estimates and the true asymmetry value. The naive estimator has theoretical infinite mean and variance, and so infinite bias and mean square error, but by eliminating from consideration the simulations where $n_1 + n_2 - 2n = 0$, our simulated bias and mean square error will be finite.

For the interval estimate, the important measure (under the frequentist paradigm) is coverage, or the propor-

TABLE VIII. Summary of simulation study 5 ($\lambda_1=40$, $\lambda_2=40$, and $\lambda=200$).

	Av. point estimate	Mean sq. error	Av. interval estimates	Coverage proport.
Naive	0.0168	2.038	(-4.44,4.47)	0.993
Bootstrap	0.0008	0.0603	(-0.578,0.578)	0.918
Bayes'	0.0007	0.0571	(-0.579,0.580)	0.942

tion of confidence intervals that contain the true asymmetry value. From the definition of a confidence interval, we would expect 95% of the 4000 naive and bootstrap confidence intervals to contain the true asymmetry value, that is, a coverage of 0.95. For the Bayesian interval, coverage does not strictly have any meaning (since there is no true value of R) but it is still of interest to see how often the interval captures the true value in the simulation study. Indeed, Cousins [1] argues that a Bayesian method must exhibit good coverage if it is to be acceptable to physicists.

Tables IV–IX display summary information from each study. The studies that show greatest divergence between the methods are 1, 4, and 5. To highlight these, we display scatter plots of the different point estimates in Figs. 1–3.

VI. CONCLUSION

In this paper, we have described the current statistical approach to estimating asymmetries and highlighted some deficiencies. Two alternative approaches were then proposed.

The differences in the three methods were highlighted in the simulation studies. As regards the point estimates, the bootstrap and Bayesian approaches were in greater agreement with greatly smaller mean square error, for the most part, than the naive estimate. For the last three studies, where $R = 0$, a simple symmetry argument implies that all three point estimates are unbiased, but for studies 1 and 2, where the true asymmetry is 0.9 and the noise is high, the bias of the two approaches introduced here was seen to be significantly larger than that of the naive estimate. For studies 3 and 6, where the noise becomes relatively small, all three methods were in general agreement.

So, in terms of minimizing MSE, the bootstrap and Bayes' estimate of asymmetry are generally better than the naive estimate. However, according to the criteria of minimizing bias, the naive estimate performed generally

TABLE VII. Summary of simulation study 4 ($\lambda_1=4$, $\lambda_2=4$, and $\lambda=20$).

	Av. point estimate	Mean sq. error	Av. interval estimate	Coverage proport.
Naive	0.0255	4.43	(-1.214,1.219)	0.999
Bootstrap	-0.0031	0.0970	(-0.770,0.767)	0.938
Bayes' (Uniform)	-0.0035	0.0724	(-0.743,0.736)	0.964
Bayes' (Ref.)	-0.0036	0.0750	(-0.738,0.731)	0.960

TABLE IX. Summary of simulation study 6 ($\lambda_1=400$, $\lambda_2=400$, and $\lambda=2000$).

	Av. point estimate	Mean sq. error	Av. interval estimates	Coverage proport.
Naive	-0.001 20	0.007 85	(-0.177,0.174)	0.963
Bootstrap	-0.000 95	0.008 22	(-0.185,0.182)	0.954
Bayes'	-0.001 25	0.008 06	(-0.185,0.182)	0.964

better than the other two, with the proviso that we have ignored cases where \hat{R} is infinite.

Now we consider the interval estimates. We can use the coverage proportion to detect whether the intervals are capturing the true value of R about 95% of the time; the sampling error from 4000 simulations is 0.0034, so that any coverage proportions lying above 0.96 or below 0.94 provides strong evidence against proper coverage.

Aside from study 4, the POE interval had coverage higher than the desired level of 0.95. For study 3, the difference between the observed coverage 0.947 and 0.95 was within the sampling error; otherwise the coverage was significantly different from that expected if the coverage were 0.95. For the low signal to noise situations (studies 1, 4, and 5), the propagation of errors method yielded dramatically wider intervals than did the bootstrap confidence intervals and the Bayesian probability intervals. In studies 4 and 5 this led to a very high coverage above 0.99, much larger than with the alternative methods.

The coverage of the Bayesian and bootstrap intervals was sometimes significantly more and sometimes significantly less than 0.95. For study 4, which has the lowest signal to noise ratio, the Bayesian coverage was more than 0.95 but the bootstrap coverage was less than 0.95. For study 5, where there is a more moderate signal to noise ratio, the Bayesian and bootstrap coverages are a bit less than 0.95. However, for study 3, which had one of the highest signal to noise ratios, the coverage of the Bayesian method seemed to drop quite severely.

Our conclusions are that the naive approach to estimating asymmetries is easy to implement and works

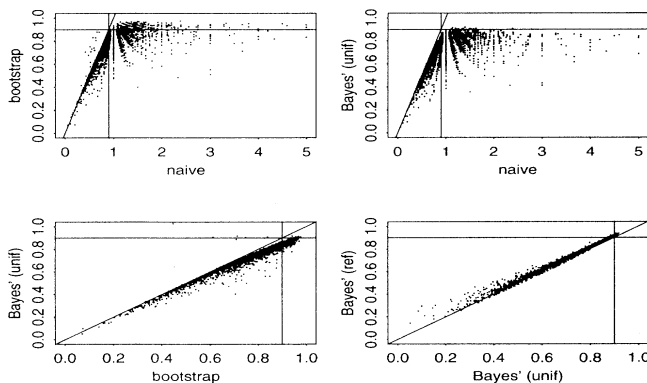


FIG. 1. Point estimates of R from simulation study 1 ($R=0.9$).

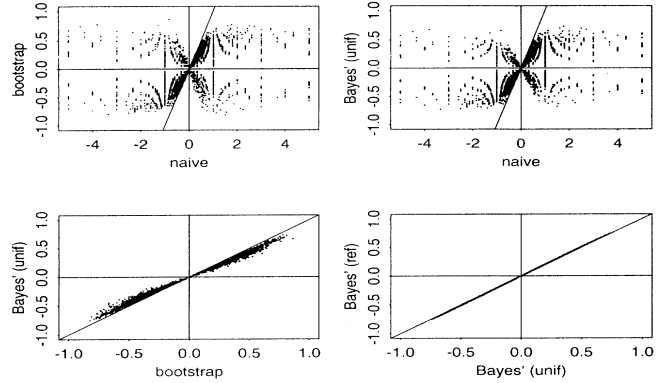


FIG. 2. Point estimates of R from simulation study 4 ($R=0$).

well in many situations, such as those of Table I in Sec. V, but the assumptions that underlie the method begin to fail under certain conditions, such as high background noise and asymmetries close to ± 1 . In these situations, naive estimates of asymmetry can lie outside physically possible values and, since the denominator can be zero with a sizeable probability, might well be infinite. It would seem sensible to search for alternative approaches that maintain admissibility under these difficult circumstances, and there are two natural alternatives that we have proposed: one is a bootstrap and the other is a Bayesian method. Both guarantee admissibility, although evidence from our studies seems to suggest that this is at the cost of some bias in the point estimate when the asymmetry is close to a boundary value. Given this, in cases of high asymmetry, the choice of estimate may depend upon the goal of the analysis; if minimizing bias is important, say, where there are many measurements to be combined into one estimate, then one may still want to use the naive estimate, while if the goal is to get an estimate with minimum error then Bayes' or bootstrap is superior. However, when the asymmetry is near 0, the bias of the Bayes' or bootstrap methods become as good as that of the naive, while maintaining their superiority in

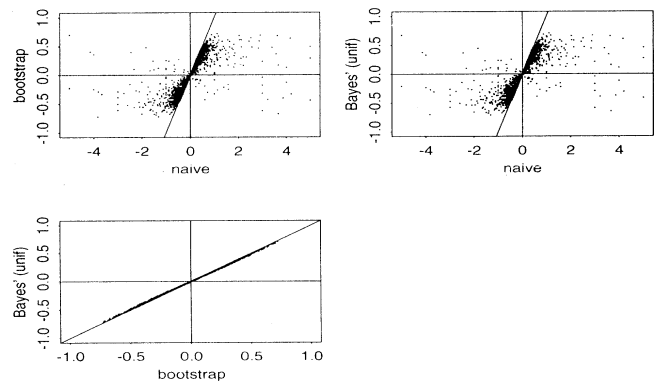


FIG. 3. Point estimates of R from simulation study 5 ($R=0$).

mean square error, so are to be recommended.

As regards an interval estimate, no method provides particularly accurate coverage, although there is evidence of gross overcoverage with propagation of errors under a low signal. This is consistent with the fact that the mean length of the POE interval estimate is longer than that provided by other methods. Also, there is a dropoff in coverage with the Bayesian approach in study 3 and with the bootstrap approach in studies 4 and 5. However, it is difficult to make any firm conclusions about coverage performance, other than to say that no method is ideal.

We finish by reiterating that the proponent of Bayesian methodology would address concerns about bias and coverage by saying that they are meaningless concepts under the Bayesian paradigm, since R has no true value and is a random quantity. Even if we disregard this retort, we believe that the current naive technique has sufficient problems, under difficult experimental conditions, to make

one consider either of the alternative approaches in the presence of high noise or where the asymmetry is expected to be close to ± 1 . As to which of these two should be used, we will confine ourselves to saying that this is a choice to be made by the experimenter, according to his or her preference.

ACKNOWLEDGMENTS

This paper was largely written while the first author was an ASA/NSF/NIST Research Associate at the National Institute of Standards and Technology, Gaithersburg, MD. Useful conversations were held with Jabez McClelland and John Unguris, both of the Electron Physics Group at the National Institute of Standards and Technology. We would like to thank Nozer Singpurwalla of the George Washington University for his contributions to the paper.

-
- [1] R. D. Cousins, *Am. J. Phys.* **63**, 398 (1995).
 - [2] D. T. Pierce, *Phys. Scr.* **38**, 291 (1988).
 - [3] D. T. Pierce, M. R. Scheinfein, J. Unguris, and R. J. Celotta, in *Growth, Characterization and Properties of Ultrathin Magnetic Films and Multilayers*, edited by B. T. Jonker, J. P. Heremans, and E. E. Marinero, MRS Symposia Proceedings Vol. 151 (Materials Research Society, Pittsburgh, 1989), pp. 49–58.
 - [4] D. T. Campbell and P. S. Farago, *J. Phys. B* **20**, 5133 (1987).
 - [5] A. Crowe and M. R. H. Rudge, in *Proceedings of the International Symposium on Correlation and Polarization in Electronic and Atomic Collisions* (World Scientific, Singapore, 1988).
 - [6] J. J. McClelland, M. H. Kelley, and R. J. Celotta, *Phys. Rev. A* **40**, 2321 (1989).
 - [7] H. B. Prosper, *Phys. Rev. D* **37**, 1153 (1988).
 - [8] V. Innocente and L. Lista, *Nucl. Instrum. Methods Phys. Res. Sect. A* **340**, 396 (1994).
 - [9] F. James and M. Roos, *Nucl. Phys. B* **172**, 475 (1980).
 - [10] F. James and M. Roos, *Phys. Rev. D* **44**, 299 (1991).
 - [11] K. J. Coakley, J. J. McClelland, M. H. Kelley, and R. J. Celotta, *Rev. Sci. Instrum.* **64**, 1888 (1993).
 - [12] K. J. Coakley, *Rev. Sci. Instrum.* **64**, 1895 (1993).
 - [13] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
 - [14] H. Jeffreys, *Theory of Probability* (Clarendon, Oxford, 1948).