

## Setting confidence belts

Byron P. Roe and Michael B. Woodroffe

*Department of Physics (B.P.R.) and Department of Statistics (M.B.W.), University of Michigan, Ann Arbor, Michigan 48109*

(Received 14 July 2000; published 8 December 2000)

We propose using a Bayes procedure with uniform, improper, prior to determine credible belts for the mean of a Poisson distribution in the presence of a background and for the continuous problem of measuring a non-negative quantity  $\theta$  with a normally distributed measurement error. Within the Bayesian framework, these belts are optimal. The credible limits are then examined from a frequentist point of view and found to have good frequentist and conditional frequentist properties.

DOI: 10.1103/PhysRevD.63.013009

PACS number(s): 06.20.Dk, 14.60.Pq

### I. INTRODUCTION

We consider two simple problems for setting confidence belts. The first problem is that of the Poisson distribution in the presence of background. The second problem is the measurement of a non-negative parameter  $\theta$ , with a normally distributed error. Although simple problems, there has been considerable work and discussion on them in the last few years. Most of the discussion has centered on the Poisson case, when the number of observed events is small. This is an important current topic because there are many search experiments involving small numbers of events. These include Higgs particle searches, supersymmetry searches, and neutrino oscillation searches such as at the Liquid Scintillation Neutrino Detector (LSND) and KARMEN.

We will briefly describe some recent attempts to address these problems.

### II. THE FELDMAN-COUSINS UNIFIED APPROACH

The standard method for setting 90% C.L. bounds has been to select a region with a 5% probability above and a 5% probability below the bounds. About two years ago, Feldman and Cousins FC [1] suggested a method, new to physics analyses, called the unified procedure. For the Poisson case, for each possible value of the parameter  $\theta$ , they looked at the ratio of the probability of getting the observed number of events  $n$  for that  $\theta$  compared to the maximum probability for all physically allowed  $\theta$ 's, and picked  $n$ 's with the highest ratio to build a 90% confidence region. This solved two problems with the old procedure. It automatically transitioned from an upper limit to a confidence belt and it always produced confidence sets in the physical region.

The unified procedure works well for many problems and is a significant improvement over the symmetric tails procedure. However it has a serious problem if few or no events are observed. If 0 events are seen, there are 0 signal and 0 background events. That there are 0 background events is interesting, but irrelevant to the question of whether signal events are seen. A 90% C.L. limit on  $\theta$  should come from  $p_\theta(0) \leq 10\%$ , which sets a limit at  $\theta = 2.3$ . (Throughout this article probabilities for a fixed parameter  $\lambda$  are denoted by a subscript  $p_\lambda$ . Probability densities and probability masses are denoted with lower case letters and distribution functions by upper case letters.) Such a case did occur in the Summer of 1998 from initial KARMEN results [2]. They had 0 events

with a background of 2.88, and using the unified method obtained an upper limit of  $\theta = 1.08$ . It is desirable to have a method set a confidence limit near 2.3, independent of background, for a 90% C.L., when no events are seen.

### III. THE ROE-WOODROOFE PROCEDURE

We [3] presented a variant of the Feldman-Cousins unified procedure which corrected this problem and introduced the concept of conditional coverage. For any observation we know that the background is less or equal to the observed number of events,  $n_{observed}$ . We suggested the use of a sample space not of all experimental outcomes, but of outcomes with background  $\leq n_{observed}$ . Thus, if  $n_{observed} = 4$ , we considered only measurements in which there are  $\leq 4$  background events, even for 20 events observed in the sample space. This probability was used both to set the ratio used in the unified method and to calculate the coverage probability. This method had the advantage of giving an upper limit of about 2.42 for 0 observed events independently of the background mean  $b$ . See Fig. 1.

Cousins [4] found that this procedure had a problem with its lower limit when applied to a continuous observable. Suppose one measures a parameter  $\theta \geq 0$  with a measurement error  $\Delta$ ,  $x = \theta + \Delta$ , where  $\Delta$  is normal (0,1). The Roe-Woodroffe (RW) procedure eliminated  $\theta = 0$  for all  $x > 0$ . However, if  $\theta = 0$ , the probability of measuring  $x > 1.39$  is 10%. See Fig. 2. It appeared that the RW upper limit and the FC lower limit were needed.

Since the conditional coverage concept is, as yet, unfamiliar, it is desirable that a method should have reasonable conventional coverage as well as conditional coverage.

### IV. BAYESIAN CREDIBLE INTERVALS

In this section, we derive Bayesian credible intervals when the expected signal is given a uniform prior distribution. These intervals have several desirable properties. Like the unified intervals, Bayesian intervals lie in the physical region and automatically change from confidence bounds to two-sided intervals. They avoid the known pitfalls mentioned above, the problems with low counts in unified method and the large lower boundary of the conditional confidence intervals in the continuous case. Finally, and importantly, the Bayesian intervals are optimal on their own terms. The intervals derived minimize length among all credible intervals.

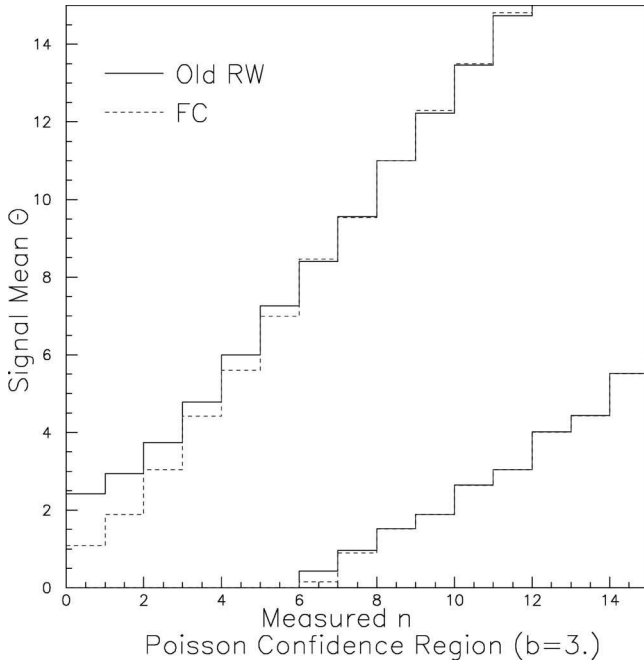


FIG. 1. The 90% C.L. belt for a Poisson probability with  $b = 3$ , using the old RW procedure (solid line) and the Feldman-Cousins unified procedure (dashed line).

The price paid for these desirable properties is dependence on the prior and metric. We address the first of these concerns by showing that the frequentist coverage probability of the Bayesian intervals is quite close to the Bayesian posterior credible level. That is, while the two probabilities are conceptually quite different, they are close numerically. In addition,

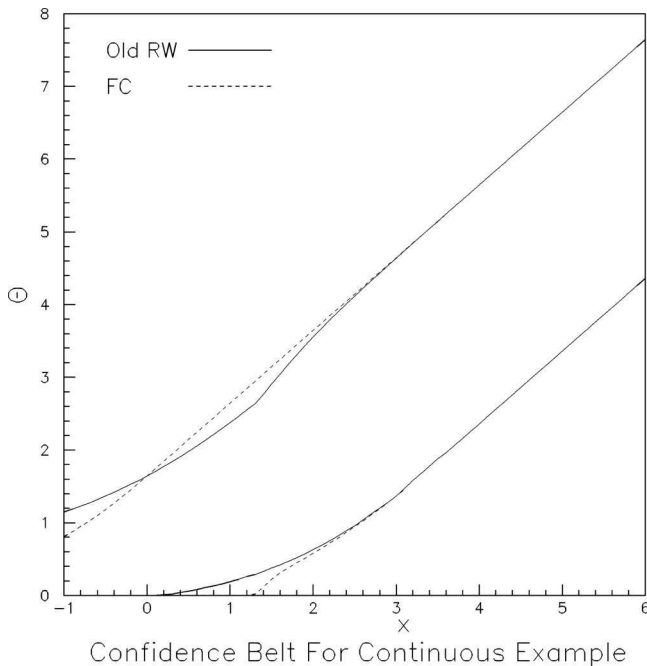


FIG. 2. The 90% C.L. belt for the continuous probability example, using the old RW procedure (solid line) and the Feldman-Cousins unified procedure (dashed line).

tion, we show that the Bayesian credible intervals have exact conditional (frequentist) coverage probability, except for discreteness in the Poisson case. With regard to the metric, it is primarily the optimality of our procedures that depends on the metric. See Sec. VII.

We will illustrate with two examples, a Poisson case in which the mean is composed of an unknown signal mean  $\theta \geq 0$  and a known background mean  $b$  and the measurement  $x$  of a parameter  $\theta \geq 0$ , with a measurement error  $\Delta$  which is normal  $(0,1)$ . We use a uniform (improper) prior of 1 for  $0 \leq \theta < \infty$  in both problems.

### A. The continuous example

This is a simple problem in which the statistical issues are clear and which approximates the Poisson problem for large  $b + \theta$ . We measure  $x = \theta + \Delta$ , where  $\theta \geq 0$ , and the density function for  $x$  is

$$f(x|\theta) \equiv \phi(x - \theta) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x - \theta)^2}. \quad (1)$$

For the uniform prior probability  $pr(\theta) = 1$ , the marginal density of  $x$  becomes

$$f(x) = \int_0^\infty \phi(x - \theta) pr(\theta) d\theta = \Phi(x),$$

where

$$\Phi(x) \equiv \int_{-\infty}^x \phi(y) dy,$$

the standard normal distribution function. We now use Bayes theorem  $f(x|\theta) \times pr(\theta) = f(\theta|x) \times f(x)$ . The conditional density of  $\theta$  given  $x$  is

$$f(\theta|x) = \frac{\phi(x - \theta)}{\Phi(x)}. \quad (2)$$

$f(\theta|x)$  is proper; the improper (infinite) prior has cancelled out. We wish to find an upper limit  $u$  and a lower limit  $l$ , dependent on  $x$ , for which

$$\text{Prob}\{l \leq \theta \leq u | x\} = 1 - \epsilon. \quad (3)$$

In Bayes theory, such intervals are called *credible intervals*. It is desirable to minimize the interval  $[l, u]$ , subject to Eq. (3). We do that by picking  $\theta$ 's with the largest probability density to be within the interval,

$$[l, u] = \{\theta: f(\theta|x) \geq c\},$$

where  $0 \leq c \leq 1/\sqrt{2\pi}$  is chosen to satisfy Eq. (3). Now,  $f(\theta|x) \geq c$  if and only if  $|\theta - x| \leq d$ , where

$$d = \sqrt{-2 \ln c - \ln(2\pi) - 2 \ln[\Phi(x)]}.$$

There are two cases to be considered. If  $d \leq x$ , then the condition (3) becomes

$$1 - \epsilon = \int_{x-d}^{x+d} f(\theta|x) d\theta = \int_{-d}^d \frac{\phi(\omega)}{\Phi(x)} d\omega = \frac{2\Phi(d) - 1}{\Phi(x)};$$

$$d = \Phi^{-1} \left[ \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x) \right], \quad (4)$$

where  $\Phi^{-1}$  denotes the inverse function to  $\Phi$ . If  $x < d$ , Eq. (3) becomes

$$\epsilon = \int_{x+d}^{\infty} f(\theta|x) d\theta = \int_d^{\infty} \frac{\phi(\omega)}{\Phi(x)} d\omega = \frac{1 - \Phi(d)}{\Phi(x)};$$

$$d = \Phi^{-1} [1 - \epsilon\Phi(x)]. \quad (5)$$

If  $x_0$  is the point where these two curves meet, then,

$$x_0 = \Phi^{-1} \left( \frac{1}{1 + \epsilon} \right). \quad (6)$$

Hence, the desired credible interval is  $[l, u] = [\max(x-d, 0), x+d]$ , where  $d = \Phi^{-1} [1 - \epsilon\Phi(x)]$ , if  $-\infty < x \leq x_0$ , and  $d = \Phi^{-1} [\frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x)]$  if  $x_0 < x < \infty$ .

**B. Poisson example**

The probability mass function and distribution function for the Poisson distribution with mean  $\lambda$  are

$$p_\lambda(n) = \frac{e^{-\lambda} \lambda^n}{n!} \quad (7)$$

and  $P_\lambda(n) = p_\lambda(0) + \dots + p_\lambda(n)$ . For our present problem  $\lambda = \theta + b$ , where  $b \geq 0$  is the fixed (known) ‘‘background’’ mean. If we let  $\theta$  have a prior uniform distribution  $pr(\theta) = 1$  for  $\theta \geq 0$ , we then have

$$p(n) = \int_0^\infty \frac{1}{n!} (\theta + b)^n e^{-(\theta + b)} pr(\theta) d\theta$$

$$= \int_b^\infty \frac{1}{n!} \omega^n e^{-\omega} d\omega = P_b(n) \quad (8)$$

as derived in our previous paper [3]. The derivation proceeded by expanding  $(\theta + b)^n$  and noting that  $\int_0^\infty \theta^k e^{-\theta} d\theta = k!$ . We again use Bayes’ Theorem  $p_{\theta+b}(n) \times pr(\theta) = p_b(n, \theta) = p_b(\theta|n) \times p(n)$ . Then, recalling that  $pr(\theta) = 1$  for  $\theta \geq 0$  and  $p(n) = P_b(n)$ , we have

$$p_b(\theta|n) = \frac{P_{\theta+b}(n)}{P_b(n)}. \quad (9)$$

To set upper and lower confidence bounds for  $\theta$ , we need to solve the equation

$$1 - \epsilon = \int_l^u p_b(\theta|n) d\theta = \frac{P_{b+l}(n) - P_{b+u}(n)}{P_b(n)} \quad (10)$$

for  $l$  and  $u$ . A 90% C.L. corresponds to setting  $\epsilon = 0.1$ . The second equality here is obtained by writing  $\int_l^u = \int_l^\infty - \int_u^\infty$  and

using the reasoning described after Equation (8). To get the shortest interval  $u-l$  we need to include values of  $\theta$  with highest density, i.e.,

$$[l, u] = [\theta: p_b(\theta|n) \geq c] \quad (11)$$

for some  $c$ . The conditions (10) and (11) may be solved numerically by an iterative procedure. This procedure is described in Sec. VI in a more general context.

**V. FREQUENTIST PROPERTIES**

The Bayesian credible intervals just derived have exact posterior coverage probability  $1 - \epsilon$ , by construction. In this section, we show that they also have high frequentist coverage probability.

**A. The continuous example**

In the continuous example,  $x = \theta + \Delta$ ,  $\theta \geq 0$ , and  $\Delta \sim \text{normal}(0, 1)$ . Fix a  $\theta$  and find the points at this fixed  $\theta$  where  $x$  meets the upper and lower Bayesian limits. Let  $x_l(\theta)$  be that  $x$  which meets the upper limit,  $\theta = u(x_l) = x_l + d(x_l)$  and  $x_u(\theta)$  be that  $x$  which meets the lower limit,  $\theta = l(x_u) = x_u - \min[d(x_u), x_u]$ . Then  $[l(x), u(x)]$  covers  $\theta$  iff  $x_l \leq x \leq x_u$ , and the conventional unconditional coverage is  $\Phi(x_u - \theta) - \Phi(x_l - \theta)$ . This is not exactly  $1 - \epsilon$ . A lower limit on the coverage is shown in the Appendix to be

$$\Phi(x_u - \theta) - \Phi(x_l - \theta) \geq \frac{1 - \epsilon}{1 + \epsilon}, \quad (12)$$

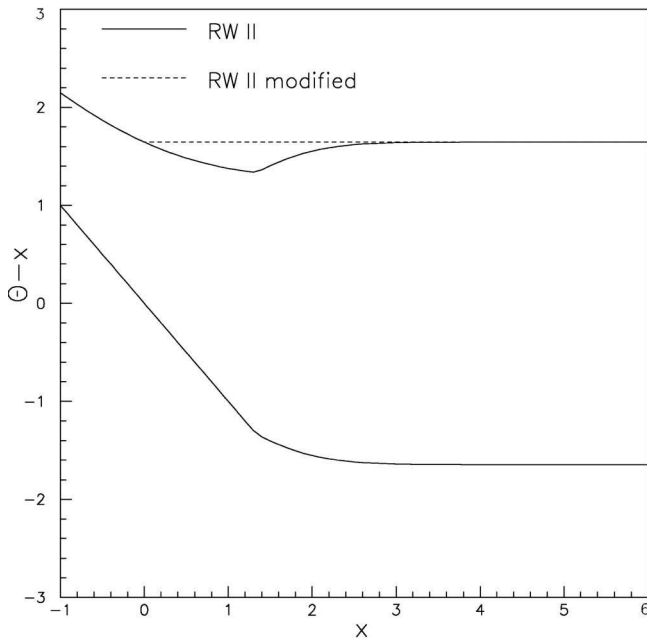
but this is a very conservative limit. For  $\epsilon = 0.1$ , this lower limit is 0.8182, while a numerical calculation finds a minimum coverage of about 0.86.

The conventional coverage probability can be improved for a very small increase in limits. We will consider a conservative *ad hoc* modification of the upper limit on  $\theta$  using the one-sided limit for a higher confidence level,  $\epsilon' = \epsilon/2$ . Let

$$u'(x) = \max \left[ u(x), x + \Phi^{-1} \left( 1 - \frac{1}{2} \epsilon \right) \right].$$

Let  $x'_l$  be the  $x$  corresponding to  $u'(x) = \theta$ . The formula for the conventional coverage is  $\Phi(x_u - \theta) - \Phi(x'_l - \theta)$ , as above. The undercoverage, derived in the Appendix, is very small for this conservative modification. For a 90% C.L., the conventional coverage is at least .900 everywhere to three significant figures. Figure 3 shows a plot of the coverage bands, plotted as  $\theta - x$  vs  $x$ . The confidence bands are shown in Fig. 4 and compared with the old RW upper bound and the Feldman-Cousins unified procedure lower bound. The conventional frequentist coverage for the Bayesian model and for this conservative modification are shown in Fig. 5.

The Bayesian credible intervals also have an exact conditional frequentist property, in terms of the error  $\Delta = x - \theta$ . Note that if  $x$  is observed, then, necessarily,  $\Delta \leq x$ , since  $\theta \geq 0$ .  $\Phi(x)$  is just the probability that  $\Delta \leq x$ . Let  $a(x) = \min[d(x), x]$ . Then the interval  $[l(x), u(x)]$  covers  $\theta$  iff



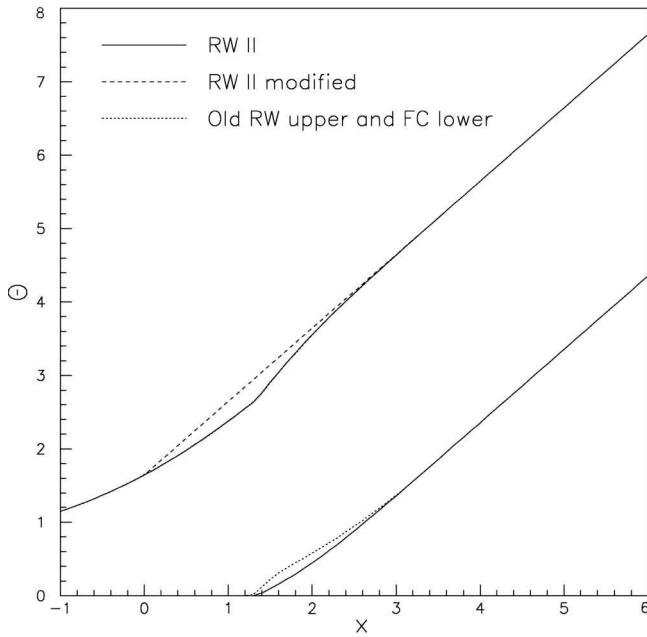
Confidence Belt For Continuous Example

FIG. 3. The 90% C.L. belt plotted as  $\theta - x$  vs  $x$  using the Bayesian procedure and a conservative modification (dashed) for the continuous example.

$l(x) \leq \theta \leq u(x)$  or, equivalently,  $-d(x) \leq \Delta \leq a(x)$ . In the Appendix, it is shown that if  $x' = \theta + \Delta'$  is an independent copy of  $x$ , then

$$\text{Prob}_{\theta}[-d(x) \leq \Delta' \leq a(x) | \Delta' \leq x] = 1 - \epsilon$$

and  $[l(x), u(x)]$  is the shortest interval with these properties.



Confidence Belt For Continuous Example

FIG. 4. The 90% C.L. belt using the Bayesian procedure (solid line) and the conservative modification (dashed line) for the continuous example. The old RW upper limit and the F-C unified method lower limit are shown as dotted lines.

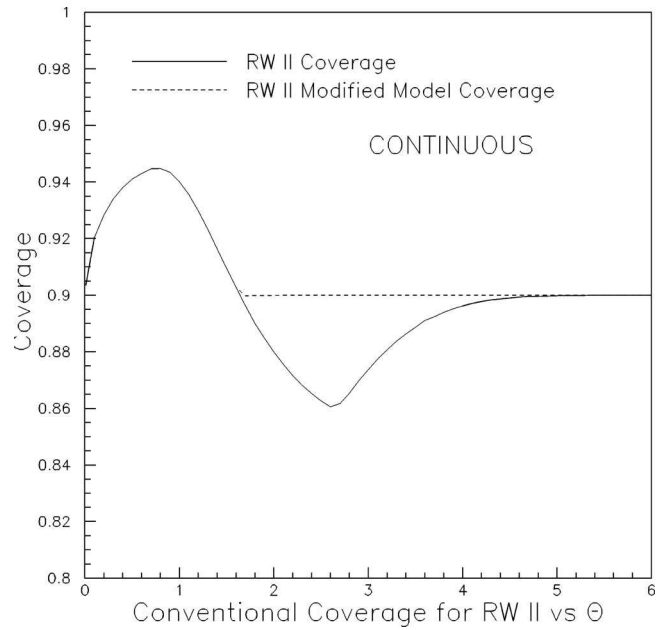


FIG. 5. The conventional frequentist coverage for a Bayesian 90% C.L. belt for the Bayesian continuous model (solid line) and the conservative modification (dashed line).

**B. The Poisson example**

As in the continuous case, the conventional coverage is not exact in the Poisson case. For  $b=3$  and  $1 - \epsilon = 0.9$ , for example, the conventional coverage varies from about 86% to 96.6%. Figure 6 shows the resulting confidence belt for  $b=3, \epsilon=0.1$ .

The conditional coverage is shown in the Appendix to be  $\approx 1 - \epsilon$ , except for discreteness.

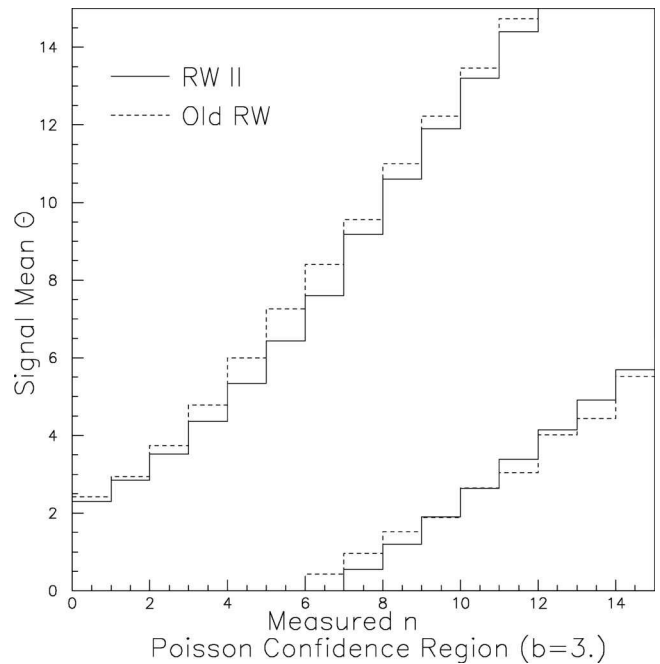


FIG. 6. The 90% C.L. belt for the Poisson distribution with  $b=3$  using the Bayesian procedure and the old RW procedure (dashed line).

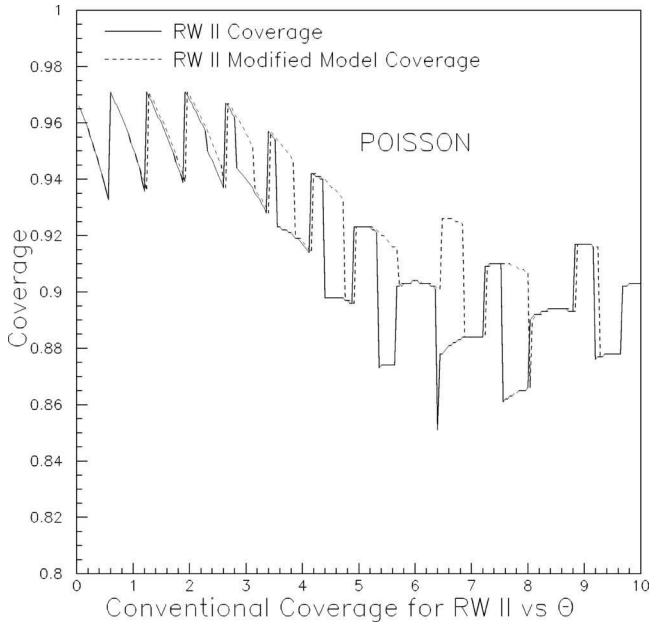


FIG. 7. The conventional frequentist coverage obtained by this Bayesian procedure (solid line) and by the conservative modification (dashed line) for  $b=3$ .

The conventional coverage can be improved by a small *ad hoc* modification similar to that used for the continuous case. Consider an alternate Bayesian upper limit  $u'$  for  $\theta$  defined as the one-sided limit for a credible level  $= 1 - \epsilon'$ . Take the modified upper limit as the maximum of  $u$  and  $u'$ . In the continuous example we chose  $\epsilon' = \epsilon/2$ . Here, we make a different *ad hoc* choice. The effect of the modification for  $b=3, \epsilon=0.1$  and  $\epsilon'=0.08$  is shown in Fig. 7. The value of the limit for  $n=0$  then increases from 2.3 to 2.53. See Fig. 8. For a fixed  $\epsilon'$  the limit at  $n=0$  is independent of  $b$ , but the appropriate choice of  $\epsilon'$  will depend weakly on  $b$ .

## VI. PARTIAL BACKGROUND-SIGNAL SEPARATION

In this section, we extend the method to processes in which we have some information concerning whether a given event is a signal or a background event. Suppose on each event one measures a statistic  $x$  (or a vector of statistics) for which the density function is  $g(x)$  for signal events and  $h(x)$  for background events. Suppose, further, that the number of observed events  $n$  has a Poisson distribution with mean  $b + \theta$ , where  $\theta$  is the expected number of signal events. We assume that  $b$ , the expected number of background events, is known. Then the joint probability mass function/density for observing  $n$  events and parameters  $x_1, \dots, x_n$  is

$$\begin{aligned} f_{\theta}(n, x) &= \frac{(b + \theta)^n}{n!} e^{-(b + \theta)} \prod_{k=1}^n \left[ \frac{\theta g(x_k) + b h(x_k)}{b + \theta} \right] \\ &= \frac{1}{n!} e^{-(b + \theta)} \prod_{k=1}^n [b + \theta r(x_k)] \prod_{k=1}^n h(x_k), \end{aligned} \quad (13)$$

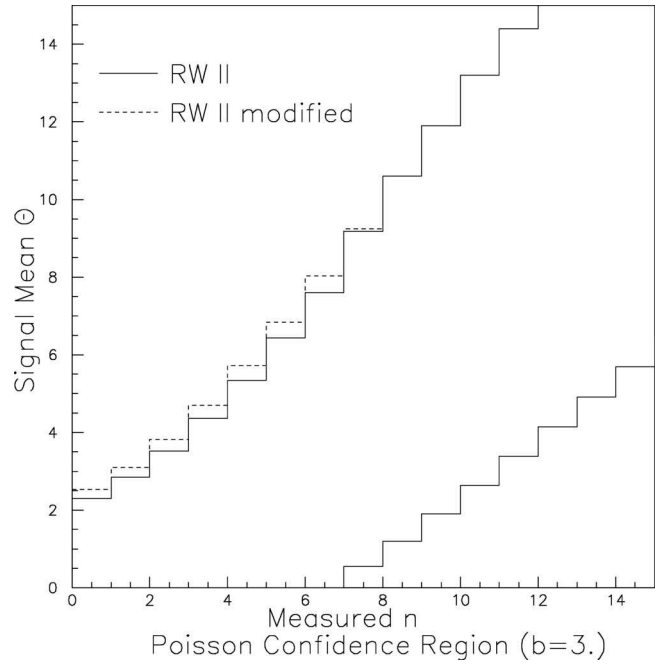


FIG. 8. The 90% C.L. belt for the Poisson distribution with  $b=3$  using the Bayesian procedure (solid line) and the conservative modification (dashed line).

where  $r(x_k) = g(x_k)/h(x_k)$ . For a uniform prior,  $pr(\theta) = 1$  for  $0 \leq \theta < \infty$ , the marginal and posterior probability mass function/densities of  $n$  and  $x_1, \dots, x_n$ , and of  $\theta$  are then

$$\begin{aligned} f(n, x) &= \int_0^{\infty} f_{\theta}(n, x) d\theta \\ &= \frac{1}{n!} \prod_{k=1}^n h(x_k) \int_0^{\infty} e^{-(b + \theta)} \\ &\quad \times \prod_{k=1}^n [b + \theta r(x_k)] e^{-(b + \theta)} d\theta \\ q(\theta | n, x) &= \frac{f_{\theta}(n, x)}{f(n, x)}. \end{aligned} \quad (14)$$

Observe that

$$\int_0^{\infty} e^{-(b + \theta)} \prod_{k=1}^n [b + \theta r(x_k)] d\theta = n! \sum_{m=0}^n C_{n,m} \frac{b^{n-m}}{(n-m)!} e^{-b},$$

where

$$C_{n,m} = \left[ 1 / \binom{n}{m} \right]_{j_1 + \dots + j_n = m} \prod_{k=1}^n r(x_k)^{j_k}, \quad (15)$$

and where  $j_1 \dots j_n$  are restricted to the values 0 or 1. Cancelling common factors,

$$q(\theta|n,x) = \frac{e^{-\theta} \prod_{k=1}^n [b + \theta r(x_k)]}{n! \sum_{m=0}^n C_{n,m} \frac{b^{n-m}}{(n-m)!}}. \quad (16)$$

Observe also that if  $g=h$ , then  $C_{n,m}=1$  and  $q(\theta|n,x)$  is the same as the posterior density of  $\theta$  given  $n$  in the Poisson case without the extra variables  $x$ .

We want to find upper and lower limits  $u, l$  for  $\theta$  such that  $\text{Prob}\{l \leq \theta \leq u | n, x\} = 1 - \epsilon$  and to minimize that interval. This means we want  $[l, u] = \{\theta : q(\theta|n, x) \geq c\}$ . We first find the value  $\theta_{max}$  at which  $q(\theta|n, x)$  is maximum, or, equivalently,  $\ln[q(\theta|n, x)]$  is maximum. Since the denominator in Eq. (16)

$$D = n! \sum_{m=0}^n \frac{b^{n-m}}{(n-m)!} C_{n,m} \quad (17)$$

does not depend on  $\theta$ ,

$$\frac{d}{d\theta} \ln[q(\theta|n, x)] = \sum_{k=1}^n \frac{r(x_k)}{b + \theta r(x_k)} - 1.$$

If  $r(x_1) \cdots r(x_n) < b$ , then  $\theta_{max} = 0$ . Otherwise, setting the derivative equal to 0 then leads to the equation

$$\theta_{max} = \sum_{k=1}^n \frac{\theta_{max} r(x_k)}{b + \theta_{max} r(x_k)}. \quad (18)$$

If  $g=h$ , then it is easy to see that  $\theta_{max} = \max[0, n-b]$ . It is clear from Eq. (18) that  $\theta_{max} \leq n$ , and it can be shown that the obvious iteration starting at  $\theta=n$  converges to  $\theta_{max}$ . This iteration is a special case of the EM algorithm [5].

Next integrate  $q(\theta|n, x)$ . Let

$$Q(a|n, x) \equiv \int_0^a q(\theta|n, x) d\theta.$$

Then

$$\begin{aligned} 1 - Q(a|n, x) &= \frac{1}{D} \int_a^\infty e^{-\theta} \prod_{i=1}^n [b + \theta r(x_i)] d\theta \\ &= \frac{1}{D} \sum_{m=0}^n b^{n-m} C_{n,m} \binom{n}{m} \int_a^\infty \theta^m e^{-\theta} d\theta. \end{aligned} \quad (19)$$

We use the reasoning described after Eq. (8) to evaluate this integral:

$$\begin{aligned} 1 - Q(a|n, x) &= \frac{1}{D} \sum_{m=0}^n b^{n-m} C_{n,m} \frac{n!}{m!(n-m)!} m! e^{-a} \sum_{l=0}^m \frac{a^l}{l!} \\ &= \frac{e^{-a} \sum_{i=0}^n \frac{b^{n-i} C_{n,i}}{(n-i)!} \sum_{l=0}^i \frac{a^l}{l!}}{\sum_{m=0}^n \frac{b^{n-m}}{(n-m)!} C_{n,m}}. \end{aligned} \quad (20)$$

We can either use Eq. (19), recognizing that the integral is an incomplete gamma function, or use Eq. (20) to find  $Q(a|n, x)$ . The limits can then be found by iterations as follows.

(1) First solve the equation  $Q(z|n, x) = 1 - \epsilon$  for  $z$ . This is straightforward, since  $Q(z|n, x)$  is increasing in  $z$ .

(2) If  $q(z|n, x) \leq q(0|n, x)$ , then  $l=0$  and  $u=z$ .

(3) If  $q(0|n, x) < q(z|n, x)$ , solve the equations

$$q(y|n, x) = c,$$

$$q(z|n, x) = c,$$

$$Q(z|n, x) - Q(y|n, x) = 1 - \epsilon$$

for  $c, y, z$  by a double iteration. In solving these equations, start with  $c = (c_{min} + c_{max})/2$ , where  $c_{min} = q(0|n, x)$  and  $c_{max}$  is the maximum value of  $q(\theta|n, x)$ , corresponding to  $\theta = \theta_{max}$ . Solve the first two equations for  $y < \theta_{max} < z$ . This is straightforward, since  $q(\theta|n, x)$  is increasing for  $\theta \leq \theta_{max}$  and decreasing for  $\theta > \theta_{max}$ . Then replace  $c_{max}$  (respectively,  $c_{min}$ ) by  $c$ , accordingly as  $Q(z|n, x) - Q(y|n, x) < 1 - \epsilon$  (respectively,  $> 1 - \epsilon$ ).

Note that the iteration procedure for the Poisson example without additional parameters measured is just a special case of this iteration procedure.

## VII. CONCLUDING REMARKS

We have proposed methods for setting credible or confidence intervals for the *means*  $\theta$  of a Poisson variable, observed in the presence of background, and of a normal distribution, when  $\theta$  is known to be non-negative. In the process, we have made specific choices for the prior distributions and loss structure. The (uniform) prior distributions were chosen for mathematical tractability and agreement between the Bayesian credible level and conventional confidence level; and one of the main findings is that such agreement is possible. The intervals have an optimality property: they minimize the length of credible intervals in the  $\theta$  scale for the uniform prior. For both problems, the mean provides a physically meaningful and mathematically tractable metric. Our procedures depend on the choice of metric, prior, and loss structure as follows: if the interval for  $\theta$  is  $l \leq \theta \leq u$  and the metric were changed, to  $\tau = 1/\theta$  say, then a valid interval for  $\tau$  is  $1/u \leq \tau \leq 1/l$ . This has the same frequentist coverage as the original interval and is an exact credible interval for the induced prior,  $d\tau/\tau^2$ . The optimality is lost, however. The transformed interval does not minimize length in the  $\tau$ -scale. The situation for our method is similar to that in the Cramer-Rao limit, Eq. 13.10 of Roe [6]. This very useful bound also depends on the metric.

The derivations of the *ad hoc* modifications in Sec. V mix

Bayesian and frequentist reasoning, and this may seem inelegant. We believe that any inelegance is mitigated by the minor nature of the changes and the resulting good properties of the procedure from both the Bayesian and frequentist viewpoints.

In the case of a Poisson distribution without auxiliary variables, there is a close connection between our method and  $CL_s$ : the  $CL_s$  procedure is mathematically equivalent to an upper credible bound with a uniform prior distribution. Read [7] calls this a coincidence. We think that there is a deeper connection [8]. The optimality claimed for  $CL_{b+s}$  in Read's work is optimality for testing background only versus background plus a specified signal. Our procedures are designed to produce intervals of shortest length, not most powerful tests. That these two criteria can lead to different procedures was noted in the statistical literature by Pratt [9].

**ACKNOWLEDGMENTS**

This research is supported by the National Security Agency under MDA904-99-1-004 and the National Science Foundation under PHY-9725921.

**APPENDIX: PROOFS OF COVERAGE AND OPTIMALITY THEOREMS**

**1. Continuous example**

Recall that if  $x = \theta + \Delta$  is observed, then necessarily  $\Delta \leq x$  and that  $\theta \in [l(x), u(x)]$  iff  $-d(x) \leq \Delta \leq a(x)$ , where  $d(x) = u(x) - x$  and  $a(x) = x - l(x) = \min[d(x), x]$ . Let  $x' = \theta + \Delta'$  be an independent copy of  $x$ .

**Theorem:** For every  $x$ ,  $\text{Prob}_\theta[-d(x) \leq \Delta' \leq a(x) | \Delta' \leq x] = 1 - \epsilon$ , and  $[l(x), u(x)]$  is the shortest interval with this property. The first statement says that the conditional coverage of  $x - \theta$  is exact.

**Proof:** The result follows from

$$\text{Prob}_\theta[\Delta' \leq z | \Delta' \leq x] = \frac{\Phi[\min(x, z)]}{\Phi(x)} = \text{Prob}[x - \theta \leq z | x]. \tag{A1}$$

In other words, the conditional frequentist distribution of  $x' - \theta$  is the same as the posterior distribution of  $x - \theta$ . It follows that

$$\begin{aligned} \text{Prob}_\theta[-d(x) \leq \Delta' \leq a(x) | \Delta' \leq x] &= \text{Prob}[x - u(x) \leq x - \theta \leq \min(x, d(x)) | x] \\ &= \text{Prob}[l(x) \leq \theta \leq u(x) | x] = 1 - \epsilon, \end{aligned}$$

where the last equality follows from the definitions of  $l(x)$  and  $u(x)$ . Next, if  $[L(x), U(x)]$  is any interval for which  $\text{Prob}_\theta[-D(x) \leq \Delta' \leq A(x) | \Delta' \leq x] = 1 - \epsilon$ , where  $D(x) = U(x) - x$  and  $A(x) = x - L(x)$ , then

$$\begin{aligned} \text{Prob}[L(x) \leq \theta \leq U(x) | x] &= \text{Prob}[-D(x) \leq x - \theta \leq A(x) | x] = 1 - \epsilon \end{aligned}$$

and, therefore,  $U(x) - L(x) \geq u(x) - l(x)$ , since the Bayesian limits were chosen to minimize the length of the interval.  $\diamond$

To put this in more conventional terms, let the interval  $C(x, x') = [x' - a(x), x' + d(x)]$ . Then

$$\text{Prob}_\theta[C(x, x') \ni \theta | \Delta' \leq x] = 1 - \epsilon, \tag{A2}$$

where  $x'$  is the random variable.

**Theorem:** A lower limit for the conventional coverage is  $(1 - \epsilon)/(1 + \epsilon)$ .

**Proof:** Recall that  $x_l$  and  $x_u$  were the values of  $x$  for a given  $\theta$  meeting the upper and lower Bayesian limits, so  $\theta = u(x_l) = x_l + d(x_l)$  and  $\theta = l(x_u) = x_u - \min[d(x_u), x_u]$ . The unconditional frequentist coverage probability is

$$\begin{aligned} \text{Prob}_\theta[l(x) \leq \theta \leq u(x)] &= \text{Prob}_\theta[x_l \leq x \leq x_u] \\ &= \Phi(x_u - \theta) - \Phi(x_l - \theta). \end{aligned}$$

Recall the definition of  $x_0$  and observe that  $x_u \geq x_0$  for all  $\theta > 0$ , since  $\theta = l(x_u)$  and  $l(x) = 0$  for  $x \leq x_0$ . First suppose that  $x_l \leq x_0$ . Then  $\Phi(x_l) \leq \Phi(x_0) \leq \Phi(x_u)$  and  $\Phi(x_0) = 1/(1 + \epsilon)$  by Eq. (6). Also, by Eqs. (4) and (5),  $\Phi(x_l - \theta) = \Phi[-d(x_l)] = 1 - \Phi[d(x_l)] = \epsilon\Phi(x_l)$ ,  $\Phi(x_u - \theta) = \Phi[d(x_u)] = \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_u)$ , and

$$\begin{aligned} \Phi(x_u - \theta) - \Phi(x_l - \theta) &= \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_u) - \epsilon\Phi(x_l) \\ &\geq \frac{1}{2} + \frac{1}{2} \frac{1 - \epsilon}{1 + \epsilon} - \frac{\epsilon}{1 + \epsilon} \\ &\geq \frac{1 - \epsilon}{1 + \epsilon}. \end{aligned}$$

Next suppose  $x_l > x_0$ . Then  $\Phi(x_u) \geq \Phi(x_l) \geq \Phi(x_0)$ ,  $\Phi(x_l - \theta) = \Phi[-d(x_l)] = 1 - [\frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_l)]$ , and

$$\begin{aligned} \Phi(x_u - \theta) - \Phi(x_l - \theta) &= \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_u) - 1 \\ &\quad + \left[ \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_l) \right] \\ &= \frac{1}{2}(1 - \epsilon)[\Phi(x_l) + \Phi(x_u)] \\ &\geq \frac{1 - \epsilon}{1 + \epsilon}. \end{aligned} \tag{A3}$$

**Theorem:** For the modified procedure a lower limit to the conventional coverage is given by  $\frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_u) - \min(\frac{1}{2}, \Phi(x_l))\epsilon$ .

**Proof:** For the modified procedure, the upper limit is defined as

$$u'(x) = \max \left[ u(x), x + \Phi^{-1} \left( 1 - \frac{1}{2} \epsilon \right) \right].$$

We set  $x'_l$  to be the  $x$  corresponding to  $u'(x) = \theta$ . Then  $x'_l = \min[x_l, \theta - \Phi^{-1}(1 - \frac{1}{2}\epsilon)]$ , and  $x'_l - \theta \leq -\Phi^{-1}(1 - \epsilon/2)$ . Observe that  $u'(x) = u(x)$  for  $x \leq 0$  and  $u'(x) = \Phi^{-1}(1 - \frac{1}{2}\epsilon)$  for  $x > 0$ . For if  $x \leq 0$ , then  $u(x) - x = d(x) = \Phi^{-1}[1 - \epsilon\Phi(x)] > \Phi^{-1}(1 - \frac{1}{2}\epsilon)$  using Eq. (5). Similarly, if  $0 < x < x_0$ , then  $u(x) - x = d(x) = \Phi^{-1}[1 - \epsilon\Phi(x)] < \Phi^{-1}(1 - \frac{1}{2}\epsilon)$ ; and if  $x_0 < x < \infty$ , then  $u(x) - x = d(x) = \Phi^{-1}[\frac{1}{2} - \frac{1}{2}(1 - \epsilon)\Phi(x)] < \Phi^{-1}(1 - \frac{1}{2}\epsilon)$ . It follows that  $x'_l = x_l$  for  $x_l < 0$  and  $x'_l = \theta - \Phi^{-1}(1 - \frac{1}{2}\epsilon)$  if  $x_l > 0$ .

So, if  $x_l > 0$ , then  $x'_l - \theta = -\Phi^{-1}(1 - \frac{1}{2}\epsilon) = \Phi^{-1}(\frac{1}{2}\epsilon)$  and, therefore,

$$\Phi(x'_l - \theta) = \Phi \left[ \Phi^{-1} \left( \frac{1}{2} \epsilon \right) \right] = \frac{1}{2} \epsilon.$$

If  $x_l < 0 (< x_0)$ , then

$$\Phi(x'_l - \theta) \equiv \Phi(x_l - \theta) = \epsilon \Phi(x_l).$$

Hence

$$\Phi(x'_l - \theta) = \min \left[ \frac{1}{2}, \Phi(x_l) \right] \epsilon, \quad (\text{A3})$$

and the theorem follows from  $\Phi(x_u - \theta) - \Phi(x'_l - \theta) = \frac{1}{2} + \frac{1}{2}(1 - \epsilon)\Phi(x_u) - \Phi(x'_l - \theta)$ .  $\diamond$

## 2. Discrete (Poisson) example

*The Bayesian credible intervals are (nearly) exact conditional confidence intervals (except for discreteness) if we convert to a scale that is natural for conditional confidence.*

To see this, let

$$H_a(y) = \int_0^y \frac{x^a}{\Gamma(a+1)} e^{-x} dx,$$

where  $\Gamma$  is the Gamma function. Then, using reasoning similar to that following Eq. (8),  $H_n(\mu) = 1 - P_\mu(n)$ . From Eq. (9), the posterior distribution function of  $\theta$  is

$$\begin{aligned} \text{Prob}(\theta \leq \theta_0 | n) &\equiv Q(\theta_0 | n) = \int_0^{\theta_0} p_b(\theta | n) d\theta \\ &= \frac{H_n(b + \theta) - H_n(b)}{P_b(n)}. \end{aligned}$$

Fix  $b$  and let  $K_{\theta,n}(m)$  be the conditional probability of at most  $m$  events from the conditional sample space, given at most  $n$  background events in the experiment, i.e.,  $K_{\theta,n}(m) = \text{Prob}_\theta(\leq m \text{ events obtained} | n_{\text{background}} \leq n)$ . Then

$$K_{\theta,n}(m) = \frac{P_{\theta+b}(m)}{P_b(n)}$$

for  $m \leq n$ , and

$$K_{\theta,n}(m) = \frac{1}{P_b(n)} \sum_{k=0}^n \frac{e^{-b}}{k!} b^k P_\theta(m-k)$$

for  $m > n$ . Letting  $M$  be the random count obtained in a particular experiment

$$\text{Prob}_\theta[K_{\theta,n}(M) \leq y | n_{\text{background}} \leq n] \approx y \quad (\text{A4})$$

for  $0 \leq y \leq 1$ , where the approximation arises because  $m$  is discrete. Specifically, Eq. (A4) is exact for  $y$  of the form  $y = K_{\theta,n}(m)$ , since  $\text{Prob}_\theta[K_{\theta,n}(M) \leq K_{\theta,n}(m) | n_{\text{background}} \leq n] = \text{Prob}_\theta[M \leq m | n_{\text{background}} \leq n] = K_{\theta,n}(m)$  for all  $m = 0, 1, 2, \dots$ .

Clearly,  $l_n \leq \theta \leq u_n$  iff  $b + l_n \leq b + \theta \leq b + u_n$  iff  $H_n(b + l_n) \leq H_n(b + \theta) \leq H_n(b + u_n)$ . Thus,  $l_n \leq \theta \leq u_n$  iff  $1 - H_n(b + u_n) \leq P_{\theta+b}(n) \leq 1 - H_n(b + l_n)$ , or equivalently,

$$\frac{1 - H_n(b + u_n)}{P_b(n)} \leq K_{\theta,n}(n) \leq \frac{1 - H_n(b + l_n)}{P_b(n)}.$$

Using Eq. (A4), the conditional probability of the latter event given  $n_{\text{background}} \leq n$  is approximately

$$\frac{H_n(b + u_n) - H_n(b + l_n)}{P_b(n)} = 1 - \epsilon, \quad (\text{A5})$$

where the last equality follows from the definitions of  $l_n$  and  $u_n$ .

To appreciate Eq. (A5), it is instructive to use a different approach. Suppose that we knew apriori that  $n_{\text{background}} \leq n$ . Then the mass function of  $M$  is then

$$\begin{aligned} \text{Prob}_\theta(M = m | n_{\text{background}} \leq n) &\equiv k_{\theta,n}(m) \\ &= K_{\theta,n}(m) - K_{\theta,n}(m-1). \end{aligned}$$

To form confidence intervals in this model, we might construct tests that the parameter has a given value and then invert this family of tests. This requires finding  $n_l$  and  $n_u$  for which  $\text{Prob}_\theta(n_l \leq M \leq n_u | n_{\text{background}} \leq n) \geq 1 - \epsilon$  for each  $\theta$ , where  $n_l$  and  $n_u$  are functions of  $\theta$  and  $n$ . This implies and is nearly equivalent to

$$\begin{aligned} \text{Prob}_\theta[K_{\theta,n}(n_l) \leq K_{\theta,n}(M) \leq K_{\theta,n}(n_u) | n_{\text{background}} \leq n] \\ \geq 1 - \epsilon. \end{aligned} \quad (\text{A6})$$

Except for discreteness, the left side of Eq. (A6) is  $K_{\theta,n}(n_u) - K_{\theta,n}(n_l)$ . The Bayesian credible intervals implicitly determine values of  $n_l$  and  $n_u$ , namely

$$\begin{aligned} K_{\theta,n}(n_l) &= \frac{1 - H_n(b + u_n)}{P_b(n)} \\ K_{\theta,n}(n_u) &= \frac{1 - H_n(b + l_n)}{P_b(n)}. \end{aligned}$$

The relation (A5) shows that these values nearly satisfy Eq. (A6), except for discreteness.

- [1] G.J. Feldman and R.D. Cousins, *Phys. Rev. D* **57**, 3873 (1998).
- [2] K. Eitel and B. Zeitnitz, in *Neutrino Physics and Astrophysics*, Proceedings of the XVIII International Conference on Neutrino Physics and Astrophysics, Takayama, Japan, edited by Y. Suzuki and Y. Totsuka (North-Holland, Amsterdam, 1999), p. 212.
- [3] B.P. Roe and M.B. Woodroffe, *Phys. Rev. D* **60**, 053009 (1999).
- [4] Robert D. Cousins, *Phys. Rev. D* **62**, 098301 (2000).
- [5] A.P. Dempster, N.M. Laird, and D.B. Rubin, *J. R. Stat. Soc. Ser. B. Methodol.* **39**, 1 (1977).
- [6] B.P. Roe, *Probability and Statistics in Experimental Physics*, 2nd ed. (Springer-Verlag, New York, in press).
- [7] A.L. Read, Proceedings of the Workshop on Confidence Limits, CERN, Geneva, Switzerland, edited by F. James, L. Lyons, and Y. Perrin, CERN 2000-005, 80 (2000).
- [8] H. Wang and M. Woodroffe, *Ann. Stat.* (to be published).
- [9] J. Pratt, *J. Am. Stat. Assoc.* **56**, 549 (1961).