Received: *November 17, 2000*, Accepted: *November 27, 2000*

Coverage of confidence intervals based on conditional probability

Mark Mandelkern and Jonas Schultz

Department of Physics and Astronomy, University of California
Irvine, California 92697, USA
E-mail: markm@uci.edu, jschultz@uci.edu

Abstract: We consider the Roe-Woodroofe construction of confidence intervals for the case of a Poisson distributed variate, where the mean is the sum of a known background and an unknown non-negative signal. The RW construction is based on a conditional pdf and posseses *conditional coverage* such that the intervals contain the true value of the signal with prescribed probability only for subsets of the full set of repeated experiments. The intervals do not have coverage in the usual sense but can be made to have such with a modification that does not affect the believability and other desirable features of this attractive construction. A similar modification can be used to provide coverage to a recently discussed application of this method to the gaussian-with-boundary problem. In both cases the result is equivalent to a construction in which the defining contour is the bayesian upper limit (uniform prior) for observations near the physical bound and the bayesian (or frequentist) central confidence interval asymptotically.

Keywords: Statistical Methods, Neutrino Physics.

JHEP11(2000)036

Contents

JHEP11(2000)036

1. Introduction

A problem of long-standing interest is that of setting confidence intervals when a measurement is *a priori* improbable. There are two cases that appear frequently in experimental physics. One of these is where a continuous variate, known to be normally distributed about a non-negative mean, is observed. A negative measurement leads to an unphysical estimate for the mean. The second and closely related case is where a non-negative discrete variate, known to be Poisson distributed such that the mean is the sum of a known non-negative background and an unknown non-negative signal, is observed. A measurement smaller than the mean background again leads to a negative (unphysical) estimate of the signal.

This problem has recently become particularly interesting because of the current topics of interest in high energy physics. There are ambitious experimental programs to measure the electron neutrino mass in beta decay. A major goal of accelerator programs is the discovery of the Higgs boson and of possible sypersymmetric states. Neutrino oscillations are actively sought in accelerator, reactor and cosmic ray experiments. The intermediate energy programs are dominated by searches for undiscovered or unconfirmed states such as the radial excitations of heavy quarkonium states, hybrids and exotics. Each of these experiments consists of measurements of parameters that are close to null or searches for small signals in the presence of significant backgrounds.

The continuous variate problem is nicely illustrated by the results from the tritium beta decay experiment at the Institute for Nuclear Research in Moscow, where three of the four measurements reported in the past 6 years, and the combined result, $m_\nu^2 = -1.9 \pm 3.4_{\mathrm{stat}} \pm 2.2_{\mathrm{syst}}\,\mathrm{eV}^2/\mathrm{c}^4$, are negative for the intrinsically positive parameter m_ν^2 [1].

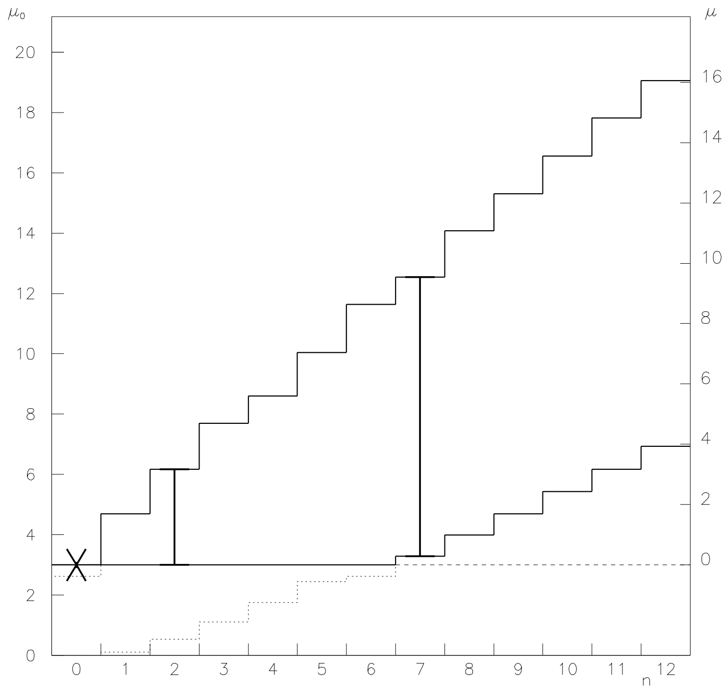

Figure 1: 90% classical Poisson confidence belt for unknown non-negative signal μ in the presence of a background with known mean b taken to be 3.0, where n is the result of a single observation. Here $\mu_0 = \mu + b$ is the parameter representing the mean of signal plus background. This construction approach leads to a null confidence interval for sufficiently small $n < b$, in this case $n = 0$. Representative confidence intervals are shown.

The discrete variate (Poisson with background) problem is likely to occur when small numbers of events are observed, which is nearly inevitable in new particle searches and neutrino oscillations appearance experiments. The difficulty here is illustrated by the initial KARMEN results [2], where 0 events were observed for a mean background of 2.88.

There have been a number of recent attempts to address these problems [3]–[6]. We address that of Roe and Woodroofe [5] who discuss the setting of confidence intervals for an unknown non-negative signal μ in the presence of a known mean background b when the measurement n is Poisson distributed as $p(n; \mu + b)$. When $n < b$, the usual estimate for μ, i.e. $n - b$, is negative, leading in most constructions to small upper limits that imply unrealistically high confidence in small values of μ. In figure 1 we show the classical construction [7] of frequentist confidence intervals for the case $b = 3$, where the observation $n = 0$ leads to the null interval. In a recent paper, Roe and Woodroofe (RW) [5] propose a construction that produces more believable intervals and contains the unifying feature that one need not decide beforehand whether to set a confidence interval or an upper confidence bound. However, since the RW confidence belt (of confidence level α) is not constructed from an unconditional probability density and does not have coverage in the usual sense (i.e. unconditional coverage), one cannot state that the probability of the interval

JHEP11(2000)036

enclosing the true value is at least α. A straightforward modification of the RW confidence belt gives it coverage, making the construction effectively an *ordering principle* applied to the Poisson pdf, albeit reached by circuitous means.

2. Roe-Woodroofe confidence intervals

Roe and Woodroofe are motivated by the observation [5] that the measurement $n = 0$ implies that zero signal (as well as zero background) is seen; thus, the resulting estimate for μ is zero, independent of b.

They argue, therefore, that the confidence interval for μ for n=0 must be independent of b. Extending the argument, they note that for any observation n, one has observed a result n from the Poisson pdf $p(n; \mu + b)$ *and* a background of at most n. They formulate a method of obtaining confidence intervals based on the conditional probability to observe n given a background $\leq n$ and obtain the desired result for $n = 0$ and approximately the classical confidence intervals for $n > b$. While they identify their method as an ordering principle, it is not one in the same sense as refs. [3] and [4] which, applying the Neyman method, explicitly choose a confidence belt of probability α from the Poisson pdf $p(n; \mu + b)$ ordered by likelihood ratio, and invert it to find confidence intervals. The latter methods do not obtain intervals that are independent of b for $n = 0$ and yield confidence intervals so small for $n < b$ that, when interpreted as measures of uncertainty, convey the misleading impression that μ is very small and well measured.

Roe and Woodroofe consider the pdf $P(n'|\mu, b, n_b \leq n)$, the probability for observing $n' = n_s + n_b$ events conditioned on $n_b \leq n$. For each n a confidence belt of probability α is constructed using this conditional probability. The belt is ordered by likelihood ratio, such that, for each μ, the n' included are those with the largest $R = [P(n'|\mu, b, n_b \leq n)]/P(n'|\mu_{\text{best}}, b, n_b \leq n)$ subject to $\sum_{n'} P(n'|\mu, b, n_b \leq n) \geq \alpha$. Here μ_{best} is the allowed value of μ which maximizes the conditional probability for an observation n'. Inverting, we have for each n' an interval. One can construct such a *conditional confidence belt* for any n. If one considers the subset of experiments such that $n_b \leq n$ then for those experiments the construction has coverage, meaning that the probability of finding the true value μ within the *conditional confidence interval* is α.

R-W then use the information that for an observation n, one knows that $n_b \leq n$, by constructing an overall confidence belt, given in figure 2 for $b = 3$ and $\alpha = 0.90$, so that the interval for $n' = n$ is taken from that conditional confidence belt constructed for $n_b \leq n$. We easily see that the requirement initially motivating RW, that the overall confidence interval for $n = 0$ be independent of b, is obtained. The conditional probability $P(n'|\mu, b, n_b = 0) = p(n'; \mu)$ and resulting confidence belt for n' vs. μ is independent of b giving, for $n = 0$, the upper limit 2.44 for μ. We also see that the R-W intervals approach the usual intervals for $n \gg b$ since in that regime $P(n|\mu, b, n_b \leq n)$ approaches $p(n; \mu + b)$.

JHEP11(2000)036

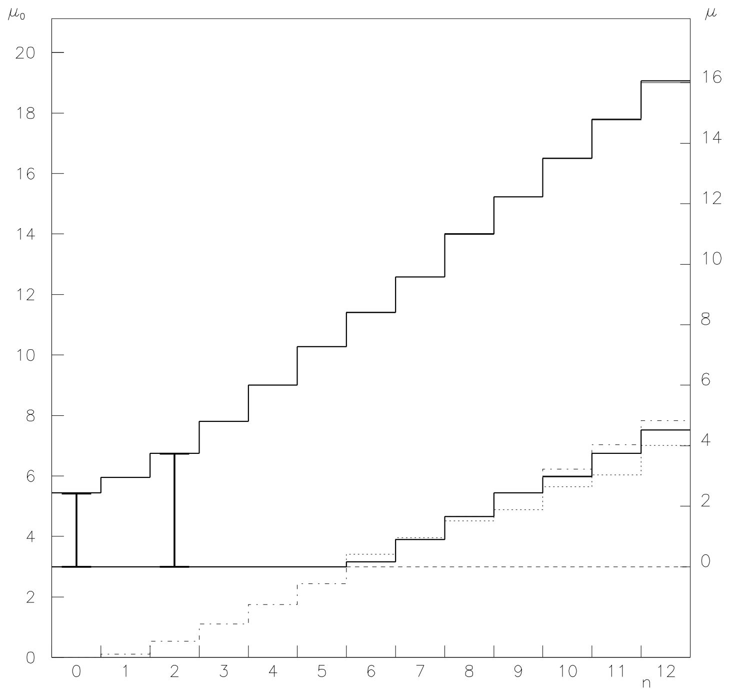

Figure 2: 90% Poisson confidence belts in the Roe-Woodroofe (solid-dotted) and modified Roe-Woodroofe (solid-solid) constructions, where $b = 3.0$. As in figure 1, $\mu_0 = \mu + b$. The dotted-dashed lines are the one-sided boundary such that $\geq 90\%$ of the probability for the pdf $p(n; \mu_0)$ lies to its left. The RW construction undercovers for those μ, where the dotted lines are to the left of the solid lines and overcovers for those μ, where they lie to the right. Representative confidence intervals are shown.

The overall confidence belt does not have unconditional coverage; it is not in general true that the probability of finding the true value μ within the confidence interval is $\geq \alpha$. While the construction has conditional coverage as described above for any of the intersecting subsets of measurements identified by $n_b \leq n$, only that for $n_b \leq \infty$ corresponds to the full set of measurements.

It is our view that coverage in the usual sense, i.e. unconditional coverage, is a crucial ingredient for a frequentist confidence interval. The latter must have the property that the true value μ should be contained within the confidence interval with the prescribed frequency in a large set of repeated measurements that are not subject to extra conditions. Although the RW construction does not have unconditional coverage, it can be easily modified to have such, at the expense of conditional coverage, by retaining the left-hand boundary of the confidence belt and adjusting the right-hand boundary so that for all μ the horizontal intervals contain probability $\geq \alpha$. In figure 2 we show the RW 90% intervals for $b = 3$ along with the one-sided confidence boundary for the Poisson distribution without background. We note that the RW horizontal intervals do not coincide with the one-sided intervals shown for $\mu < 2.44$. Therefore for some values of μ in this range, the confidence belt does not satisfy the coverage requirement that $\geq$90% of the proba-

JHEP11(2000)036

	Roe-Woodroofe		Modified	
n(observed)	Lower	Upper	Lower	Upper
0	0.0	2.44	0.0	2.44
1	0.0	2.95	0.0	2.95
2	0.0	3.75	0.0	3.75
3	0.0	4.80	0.0	4.80
4	0.0	6.01	0.0	6.01
5	0.0	7.28	0.0	7.28
6	0.41	8.42	0.16	8.42
7	0.96	9.58	0.90	9.58
8	1.51	11.02	1.66	11.02
9	1.88	12.23	2.44	12.23
10	2.64	13.51	2.98	13.51
11	3.04	14.77	3.75	14.77
12	4.01	16.01	4.52	16.01

Table 1: Comparison of 90% Poisson confidence intervals for the Roe-Woodroofe and modified Roe-Woodroofe constructions for $b = 3$.

bility is contained. Because coverage cannot be exact when the variable is discrete, the *error* for the example given here is not of great numerical significance. The minimum coverage of ~ 0.87 is obtained at $\mu \sim 0.4$. We have chosen the example here, $b = 3$, only because it is the example given by RW. Undercoverage is more severe for greater b; for $b = 10$, the minimum coverage is ~ 0.78. In any case it is desirable to have coverage, which we obtain as shown in figure 2, where we change the right side of the confidence belt (solid lines) so that the horizontal intervals contain probability $\geq$90%. We note that the confidence intervals for small n, i.e. $n < b$, are unchanged. Intervals for both constructions are given in table 1.

It would be nice to devise an ordering principle that can be directly applied to the Poisson pdf $p(n; \mu + b)$ to obtain the confidence belt shown in figure 2, if only because the construction we have used here is aesthetically unpleasing. This method, which consists of first determining vertical intervals per ref. [5], and then fixing them, leaves something to be desired. However, in the end the method of construction does not really matter. What results here is an ordering procedure that yields a confidence belt with coverage and produces sensible intervals.

Our modification is equally applicable to another construction, in which the Roe-Woodroofe method of conditioning is applied to the gaussian-with-boundary [8] problem. Here, for example, an interval of confidence level α is sought for an unknown non-negative signal μ, where the measurements x are normally distributed as $n(x; \mu)$. Just as in the Poisson-with-background problem, the classical construction [7] yields unrealistically small confidence intervals for small measurements x.

JHEP11(2000)036

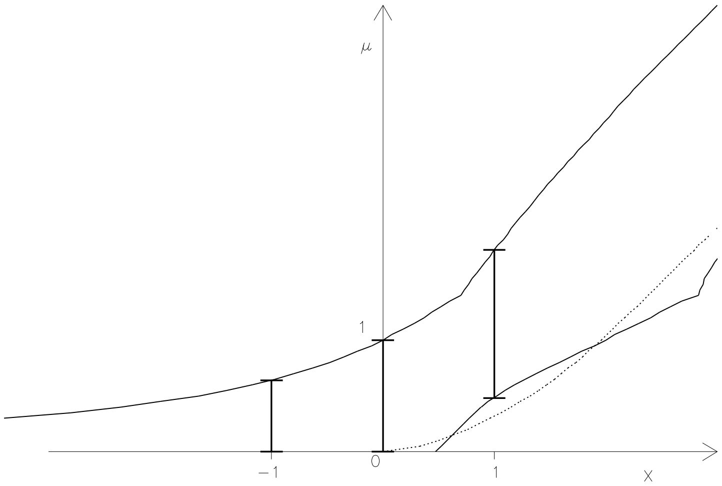

Figure 3: 68.27% gaussian confidence belts in the Roe-Woodroofe (solid-dotted) and modified Roe-Woodroofe (solid-solid) constructions. The former undercovers for those μ, where the dotted lines are to the left of the solid lines and overcovers for those μ where they lie to the right. Representative confidence intervals are shown.

Given a measured x, we seek a probability density for x', conditioned on the constraint that $\mu \geq 0$. In analogy with the Poisson case above, where $n' = n_s + n_b$ and $n_s \geq 0$ implies for observation n that $n_b \leq n$, leading to the conditional pdf $P(n'|\mu, b, n_b \leq n)$, here we have $x' = \mu + \epsilon$ and $\mu \geq 0$ implies for measurement x that $\epsilon \leq x$, i.e. $x' \leq x + \mu$, leading to [8]:

$$
\begin{aligned}
P(x'|\mu, x' \leq x + \mu) &= \frac{\frac{2}{\sqrt{2\pi}} \exp[-(x'-\mu)^2/2]}{\left(\mathrm{erf}(x/\sqrt{2}) + 1\right)}\,, && \text{if } x' \leq \mu + x \\
P(x'|\mu, x' \leq x + \mu) &= 0\,, && \text{if } x' > \mu + x\,.
\end{aligned}
\tag{2.1}
$$

One can again, for all x, construct a confidence belt of probability α for x' vs. μ. The relevant confidence interval for observation x is the interval obtained for $x' = x$, which is then *transferred* to an overall confidence belt, shown in figure 3 for α=0.6827. As for the Roe-Woodroofe construction referred to above, this construction produces sensible confidence intervals for all x including $x < 0$. However it significantly undercovers for $\mu < 0.15$ and $\mu > 0.9$ and significantly overcovers in between. In order to produce exact coverage, we retain the left hand (upper) curve of the confidence belt $x_l(\mu)$ and recalculate the right hand (lower) curve $x_r(\mu)$ so that the horizontal intervals contain probability α using:

$$
2\alpha = \mathrm{erf}\left(\frac{\mu - x_l}{\sqrt{2}}\right) + \mathrm{erf}\left(\frac{x_r - \mu}{\sqrt{2}}\right). \tag{2.2}
$$

The resulting modified confidence belt is given in figure 3 and the confidence intervals tabulated in table 2.

JHEP11(2000)036

We observe that the upper boundary of the RW confidence belt for the Poisson case is nearly identical with the bayesian upper limit contour for small n and identical to the upper contour of the bayesian (or frequentist) central confidence belt asymptotically, all for the same confidence level. Similarly, the closed form expression for the RW upper contour in the Gaussian case is identical to that for the Bayesian upper limit contour for $x < 0$ and asymptotically equal to the bayesian (or frequentist) upper contour. It is however not clear to us how the RW constuction interpolates between the relevant bayesian constructions.

	Roe-Woodroofe		Modified	
x(observed)	Lower	Upper	Lower	Upper
-3.0	0.0	0.33	0.0	0.33
-2.0	0.0	0.45	0.0	0.45
-1.0	0.0	0.64	0.0	0.64
0.0	0.0	1.00	0.0	1.00
1.0	0.32	1.81	0.48	1.81
2.0	1.04	2.97	1.00	2.97
3.0	2.01	4.00	1.75	4.00

Table 2: Comparison of 68.27% gaussian confidence intervals for the RW and modified RW constructions. Only the lower limits are affected by the modification used to provide coverage.

3. Conclusion

For the case of Poisson distributed measurements n with a non-negative signal mean μ and known mean background b, the Roe-Woodroofe construction produces well-behaved confidence intervals, particularly for $n < b$, where some other constructions yield unrealistically small intervals. Since the construction is not based on integrating probabilities that arise from an unconditional pdf, it does not produce a confidence belt with coverage in the usual frequentist sense. We suggest a modification that provides coverage while preserving the believability of the intervals and other desirable features of the construction. While the changes introduced by this modification are relatively small for the example given here (that provided by Roe and Woodroofe), they are substantially larger for greater b.

A similar modification provides coverage for an application of the Roe-Woodroofe method recently discussed elsewhere for the gaussian-with-boundary problem, where the original confidence belt substantially undercovers for some values of the true mean and substantially overcovers for others.

The resulting confidence intervals for both cases are believable and possess frequentist coverage. They are suitable for the quotation of experimental errors, especially when the measured quantity is improbable because of an explicit constraint on the parameter of interest.

These constructions produce confidence belts in which the upper defining contours are equal to the bayesian upper limit contours for small observations near

JHEP11(2000)036

the physical boundary and produce confidence belts equal to the usual bayesian or frequentist central confidence belts asymptotically.

References

[1] V. M. Lobashev et al., *Direct search for mass of neutrino and anomaly in the tritium β spectrum*, *Phys. Lett.* **B 460** (1999) 227.

[2] KARMEN collaboration, *The search for neutrino oscillations $\bar{\nu}_\mu \to \bar{\nu}_e$ with* KARMEN, *Nucl. Phys.* **77** *(Proc. Suppl.)* (1999) 212 [hep-ex/9809007].

[3] G.J. Feldman and R.D. Cousins, *A unified approach to the classical statistical analysis of small signals*, *Phys. Rev.* **D 57** (1998) 3873 [physics/9711021].

[4] C. Giunti, *A new ordering principle for the classical statistical analysis of poisson processes with background*, *Phys. Rev.* **D 59** (1999) 053001 [hep-ph/9808240].

[5] B.P. Roe and M.B. Woodroofe, *Improved probability method for estimating signal in the presence of background*, *Phys. Rev.* **D 60** (1999) 053009 [physics/9812036].

[6] M. Mandelkern and J. Schultz, *The statistical analysis of gaussian and poisson signals near physical boundaries*, *J. Math. Phys.* **41** (2000) 5701 [hep-ex/9910041].

[7] D.R. Cox and D.V. Hinkley, *Theoretical statistics*, Chapman and Hall 1974.

[8] R.D. Cousins, *Application of conditioning to the gaussian-with-boundary problem in the unified approach to confidence intervals*, physics/0001031.

JHEP11(2000)036