Nuclear Instruments and Methods in Physics Research A320 (1992) 331–335
North-Holland

NUCLEAR
INSTRUMENTS
& METHODS
IN PHYSICS
RESEARCH
Section A

Incorporating systematic uncertainties into an upper limit

Robert D. Cousins
Physics Department, University of California, Los Angeles, CA 90024, USA

Virgil L. Highland
Physics Department, Temple University, Philadelphia, PA 19122, USA

Received 27 March 1991 and in revised form 19 February 1992

We discuss the problem of incorporating the uncertainty in the experimental sensitivity into the calculation of an upper confidence limit on a branching ratio or similar quantity. If the number of events is small or zero but without background, the correction to the usual result is given by a simple, easily applied formula. The case of an accurately known background also has a simple solution.

1. Introduction

In this note we consider a problem that arises in rare decay experiments that observe no, or few, candidate events: How can one include the systematic uncertainty of the experimental sensitivity in the upper limit that is reported for the branching ratio? Similar problems arise in measuring limits on the production cross sections of rare particles and elsewhere. Published experimental papers have until recently usually ignored the problem, to the extent that it has been rare for the systematic uncertainties even to have been quoted. This general policy is no doubt based on an intuition that the systematic uncertainty in these cases is relatively unimportant compared with the Poisson fluctuations. It is, however, equally clear intuitively that an experiment with a small systematic uncertainty is to be preferred to an equivalent one with a larger uncertainty. The formulas derived in this paper allow that preference to be made quantitative with a minimum of calculation.

In section 2 we begin with a discussion of the basic problem and consider in detail the common case of no observed events ($n = 0$) and 90% confidence level (CL). We use this case to illustrate several practical methods of computing the upper limit and to clarify the considerations common to them. For a wide class of situations, the calculation can be reduced to a simple, easily applied formula. In section 3 we treat the more general case $n \geq 0$ and confidence levels other than 90%, but with no expected background. An equally simple formula results. Finally, section 4 considers this problem for the case of a known background rate.

Our statistical approach includes both classical and Bayesian elements [1]. Our treatment of the Poisson parameter is classical, the type of statistics we generally prefer. Because we average over a probability distribution for the experimental sensitivity, our treatment of that quantity is necessarily Bayesian. For the problem at hand, a consistently classicial approach yielded results that were physically unacceptable in the sense that a larger systematic error produced a smaller upper limit [2].

2. The case $n = 0$

Consider an experiment to determine the value of a branching ratio R by observing the number of counts n in a detector. Let R_t be the true value of the branching ratio and S_t be the true value of the sensitivity factor, a combination of beam flux, detector acceptance, etc., defined such that the mean number of events expected is $\mu_t = R_t S_t$. The number of events actually observed in an experiment will be a sample drawn from a Poisson distribution $P(n; \mu_t)$. To draw an inference about the value of R_t requires information about the value of S_t. This information is obtained from subsidiary measurements that give an estimate $\hat{S}$ for S_t, along with an estimate of the uncertainty σ. We follow the useage common in this branch of physics by referring to the

expected fluctuation in the number of events as the "statistical" uncertainty and the uncertainty in the sensitivity factor as the "systematic" uncertainty.

In considering how to specify an upper limit $\bar{R}_0$ on R_t in the case that the experiment observes no events, we wish to make contact with two well-understood situations:

i) *A significant number of events is observed ($n \gg 1$).* The point estimate of R_t is then $\hat{R} = n/\hat{S}$. Using a standard approximation, an experimentalist will normally assign an uncertainty $\hat{R}\sqrt{1/n + \sigma_r^2}$ to this estimate as a measure of the central confidence interval, where $\sigma_r = \sigma/\hat{S}$ is the relative uncertainty in $\hat{S}$.

ii) *No events are observed ($n = 0$), but there is no uncertainty in the value of S_t ($\sigma = 0$).* It is well-known that, from either a classical or a Bayesian viewpoint [1], the 90% confidence-level upper limit on the parameter μ_t is $\ln 10 \approx 2.30$, and hence the upper limit on R_t is $\bar{R}_{00} = 2.30/\hat{S}$. (The first zero in the subscript identifies the $n = 0$ case and the second distinguishes the zero variance upper limit from the more general case discussed below.) The classical upper limit is given by the value of μ that makes $P(0; \mu) = 0.1$, i.e. the value for which an observation as small as or smaller than the actual one is no more than 10% probable. If there were not an analytic solution to this equation, one could explore the function $P(0; \mu)$ numerically or by a Monte Carlo technique to find the apppropriate value of μ.

The problem we are considering is like case (II), but with non-zero variance for S_t. The probability of observing no events $p_0(R_t, S_t) = P(0; R_t, S_t)$ is a function of both R_t and S_t. The value of R_t is unknown and indeed is what we are trying to determine, or at least to bound. The value of S_t is known in a probabilistic sense from the subsidiary measurements. Assuming $\hat{S}$ is an unbiased estimator and that the Bayesian prior probability distribution of S_t is a constant, the posterior probability of S_t being in any neighborhood of $\hat{S}$ is distributed about $\hat{S}$ with variance σ^2. We designate the probability density function (pdf) of this distribution by $W(S; \hat{S}, \sigma)$. (Here and subsequently we use simply S instead of S_t to designate the true value of the sensitivity.) Note that the upper limit on the mean number of events that would be observed in many repetitions of the experiment is 2.30 and is unaffected by our ignorance of the exact value of the sensitivity, which is required only for deducing R_t.

A reasonable generalization of the definition of a 90% confidence-level upper limit in this case is to explore the function $p_0(R, S)$ and find the value $\bar{R}_0$ that makes $\langle p_0 \rangle = 0.1$, where $\langle\ \rangle$ indicates an average over $W(S; \hat{S}, \sigma)$ throughout this paper. This evaluation can be carried out with a straightforward Monte Carlo calculation, which we have found quite useful. However, additional insight is obtained by considering the problem from two analytical approaches.

2.1. Integral method

Beginning with the case of no observed events ($n = 0$), the first analytic approach expresses the above average as an explicit integral [3–5] over the Poisson probabilities of having zero events, weighted by the probability density $W(S; \hat{S}, \sigma)$:

$$p_0(R) \equiv \langle p_0(R, S) \rangle = \int_0^\infty e^{-RS} W(S; \hat{S}, \sigma)\, dS. \quad (1)$$

In the most complex cases this integral must be performed numerically, with a Monte Carlo approach being one option, but simplification may be available in some cases. Very commonly W can be considered Gaussian, so that

$$p_0(R) = \int_0^\infty e^{-RS} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(S-\hat{S})^2/2\sigma^2}\, dS. \quad (2)$$

This form for W requires σ to be small compared to $\hat{S}$; we can them extend the lower limit of the integral to $-\infty$ with negligible change. (If that is not the case, the suitability of a Gaussian form for W is doubtful; a log-normal distribution [6] may provide a better model.) Upon completing the square in the exponent and carrying out the integration, we have

$$p_0(R) = \exp\left(-R\hat{S} + R^2\sigma^2/2\right)$$
$$= \exp\left(-R\hat{S} + \left(R\hat{S}\sigma_r\right)^2/2\right). \quad (3)$$

We find the 90% confidence-level upper limit $\bar{R}_0$ by solving the equation $p_0(\bar{R}_0) = 0.1$, yielding

$$\bar{R}_0 = \left(1 - \sqrt{1 - 2\sigma_r^2 \ln 10}\right)/\left(\hat{S}\sigma_r^2\right). \quad (4)$$

Expanding the square root for small σ_r we have

$$\bar{R}_0 = 2.30\left(1 + 2.30\sigma_r^2/2\right)/\hat{S}. \quad (5)$$

As shown below in a second approach, this equation is quite general and not restricted to the Gaussian distribution. We defer discussion of this result to section 2.2.

If there are events ($n > 0$), the principle is the same, though the amount of computation increases in this approach. For Gaussian W, the probability of observing one event is

$$p_1(R) = R(\hat{S} - R\sigma^2) \exp\left(-R\hat{S} + R^2\sigma^2/2\right)$$
$$= \left(R\hat{S} - \left(R\hat{S}\sigma_r\right)^2\right) p_0(R). \quad (6)$$

Given one event, the upper limit is found by solving numerically for $\bar{R}_1$ in the equation $p_0(\bar{R}_1) + p_1(\bar{R}_1) = 0.1$. One could again expand this equation in σ_r, or perhaps make a series expansion in the integrand.

Instead, we proceed to the second approach, which naturally incorporates the series expansion from the beginning.

2.2. Expansion method

In the second analytic approach we work directly with the moments of $W(S; \hat{S}, \sigma)$ to derive upper limits without explicit integration. This approach has the advantage of being conveniently generalized, as shown in sections 3 and 4. We start with the observation that, because the probability of observing zero events, $p_0(R, S) = e^{-RS}$, is a concave upward function, the value averaged over an interval in S is greater than the value at the center of the interval. Hence R must be slightly larger to give the same value of $\langle p_0 \rangle$ as obtained without the averaging.

We proceed by making a Taylor expansion of $p_0(R, S)$ about the point one would have for the upper limit if there were no uncertainty in S. This is the point $R = \bar{R}_{00}$, $S = \hat{S}$, as in case II above. Carrying out the expansion in two variables through second-order terms we have

$$\Delta p_0 = \frac{\partial p}{\partial S}\Delta S + \frac{\partial p}{\partial R}\Delta R + \left(\frac{\partial^2 p}{\partial S^2}(\Delta S)^2 + 2\frac{\partial^2 p}{\partial S \partial R}\Delta S \Delta R + \frac{\partial^2 p}{\partial R^2}(\Delta R)^2\right)/2. \tag{7}$$

Inserting the derivatives of $p_0(R, S)$ we obtain

$$\Delta p_0 = \left\{-\bar{R}_{00}\Delta S - \hat{S}\Delta R + \bar{R}_{00}^2(\Delta S)^2/2 + \left(\bar{R}_{00}\hat{S} - 1\right)\Delta S \Delta R + \hat{S}^2(\Delta R)^2/2\right\} e^{-\bar{R}_{00}\hat{S}}. \tag{8}$$

Averaging eq. (8) over S yields

$$\langle \Delta p_0 \rangle = \left(-\hat{S}\Delta R + \bar{R}_{00}^2\sigma^2/2\right) e^{-\bar{R}_{00}\hat{S}}, \tag{9}$$

where we have used the definition $\sigma^2 \equiv \langle \Delta S^2 \rangle$ and the assumption that $\hat{S}$ is unbiased, from which it follows that $\langle \Delta S \rangle = 0$. We have also neglected the second-order term in ΔR as compared to the first-order term, on the assumption that $\hat{S}\Delta R \ll 2$. This term is retained in the more general treatment of section 3.

The average probability at the shifted point is

$$\langle p_0(\bar{R}_{00} + \Delta R, S)\rangle = p_0(\bar{R}_{00}, \hat{S}) + \langle \Delta p_0 \rangle. \tag{10}$$

Both evaluations of p_0 in this equation equal 0.1 by definition, because the first is the average probability of zero events at the redefined upper limit $\bar{R}_0 = \bar{R}_{00} + \Delta R$ and the second is the probability at the expansion point, the zero variance case II. Therefore $\langle \Delta p_0 \rangle = 0$ and it follows from eq. (9) that

$$\Delta R = \bar{R}_{00}^2\sigma^2/(2\hat{S}). \tag{11}$$

Finally, the new upper limit for the case $n = 0$ is

$$\bar{R}_0 = 2.30\left(1 + 2.30\sigma_r^2/2\right)/\hat{S}. \tag{12}$$

Eq. (12) (or eq. (5)) is a useful answer to the problem with which we began, having several desirable properties. Most importantly it goes over into the standard result of case II as $\sigma_r \to 0$ and it depends on σ_r quadratically, which agrees with one's expectations based on case I. It justifies to a large degree the long standing custom of ignoring the contribution of σ_r, since the effect of this term turns out to be fairly small (see the fifth column of table 1). But eq. (12) appropriately assigns a slightly smaller limit, other things being equal, to the experiment which has the smaller normalization uncertainty, while making clear that the most important consideration is to have S as large as possible. According to eq. (12) it is unnecessarily conservative to incorporate the systematic uncertainty by relaxing the bound for $\sigma = 0$ by a factor of $(1 + \sigma_r)$, as has sometimes been the practice. The formula gives some guidance on how much effort it is worth devoting to measuring S as opposed to pursuing the main measurement.

3. General case

In the general case, the probability of observing n or fewer events is

$$p_n(R, S) = \sum_{j=0}^{n} (RS)^j e^{-RS}/j!. \tag{13}$$

Again we wish to expand about the upper limit point $(\bar{R}_{n0}, \hat{S})$ that one would adopt in the zero variance case. In taking the first derivative of p_n we note that all terms in the sum but one cancel, so

$$\partial p_n/\partial R = -R^n S^{n+1} e^{-RS}/n!, \tag{14}$$

with a symmetrical expression for $\partial p_n/\partial S$. The remaining derivatives in eq. (7) are readily evaluated, so that after averaging over S we have a result similar to eq. (9):

$$\langle \Delta p_n \rangle = \left\{-\left(\bar{R}_{n0}\hat{S}\right)^n \hat{S}\Delta R + \left(\bar{R}_{n0}\hat{S}\right)^{n-1}\left(\bar{R}_{n0}\hat{S} - n\right)(\hat{S}\Delta R)^2/2 + \left(\bar{R}_{n0}\hat{S}\right)^{n+1}\left(\bar{R}_{n0}\hat{S} - n\right)\sigma_r^2/2\right\} e^{-\bar{R}_{n0}\hat{S}}/n!. \tag{15}$$

This time the second-order terms in ΔR have been retained. The value of eq. (15) is zero by the same

reasoning as before and one can readily solve the resulting quadratic equation for ΔR.

Before doing so, we simplify the notation by defining $U_n \equiv \bar{R}_n \hat{S}$ and similarly for U_{n0}. Solving for $\Delta U_n \equiv \Delta R \hat{S}$ we have

$$\Delta U_n = U_{n0}\left\{1 - \left(1 - \sigma_r^2 E_n^2\right)^{1/2}\right\}/E_n, \qquad (16)$$

where we have defined $E_n \equiv U_{n0} - n$ for compactness. E_n represents the excess of the upper limit of a Poisson parameter over the number n of observed events, for a specified confidence level. The value of U_n itself is

$$U_n = U_{n0}\left[1 + \left\{1 - \left(1 - \sigma_r^2 E_n^2\right)^{1/2}\right\}/E_n\right]. \qquad (17a)$$

As in section 2.1, we can simplify this equation by expanding the square root and retaining terms either through second order in σ_r^2:

$$U_n = U_{n0}\left[1 + E_n(\sigma_r^2/2)\left\{1 + (E_n\sigma_r/2)^2\right\}\right], \qquad (17b)$$

or only through first order:

$$U_n = U_{n0}\left(1 + E_n\sigma_r^2/2\right). \qquad (17c)$$

Eq. (17c) is the generalization of eq. (12) for arbitrary n and reduces to it for $n = 0$. It has all the desirable properties enumerated for eq. (12) and applies to any confidence level. For 90% CL the values of U_{n0} are the familiar 2.30, 3.89, ... for $n = 0, 1, \ldots$, and $E_n = 2.30$, 2.89, etc. U_n is the modified value, taking into account the systematic uncertainty. From the definition of U_R, the upper limit on the branching ratio is $\bar{R}_n = U_n/\hat{S}$. Eq. (17a) should in principle be more accurate than eq. (17c), at the cost of some complexity. The square root in eq. (17a) shows that the result cannot be correct for values of σ_r which are too large, and that the limiting value of σ_r must become smaller as n increases since it requires

$$\sigma_r \leq 1/E_n. \qquad (18)$$

Clearly this behavior is a result of the Taylor approximation and indicates where it breaks down.

To test eqs. (17a), (17b) and (17c), a numerical evaluation of U_n was made via a Monte Carlo calculation, assuming Gaussian systematic uncertainty. The results are displayed in table 1 for $n = 0$ to 3. One sees that out to $\sigma_r = 0.3$ and $n = 3$ the simplest equation (17c) is perfectly adequate. The better approximation (17b) does slightly better at the largest values, but the "exact" eq. (17a) is actually somewhat worse, especially considering that it is undefined for $n = 3$ and $\sigma_r > 0.27$. The approximation of expanding the square root seems to compensate the Taylor approximation to some degree.

The simple and intuitive result of eq. (17c) is adequate for most practical purposes. The unusual case with $\sigma_r > 0.3$ would probably be dealt with by numerical integration, and if an observation gave $n > 3$ events one would normally give a point estimate rather than a limit. But as long as the value of σ_r does not violate condition (18) by very much, it appears that the formula will still work for even larger n. Otherwise one can use eq. (17b), and in extreme cases go back to the explicit integral.

Table 1
For various n, the 90% confidence-level upper limit is given as a function of σ_r according to the three different approximations in the text and compared with a Monte Carlo result (M.C.)

n	σ_r	Eq. (17a)	Eq. (17b)	Eq. (17c)	M.C.
0	0.00	2.30	2.30	2.30	2.30
	0.10	2.33	2.33	2.33	2.33
	0.20	2.41	2.41	2.41	2.42
	0.30	2.58	2.57	2.54	2.60
1	0.00	3.89	3.89	3.89	3.89
	0.10	3.95	3.95	3.95	3.94
	0.20	4.14	4.13	4.11	4.13
	0.30	4.57	4.49	4.40	4.51
2	0.00	5.32	5.32	5.32	5.32
	0.10	5.41	5.41	5.41	5.41
	0.20	5.72	5.71	5.67	5.71
	0.30	6.78	6.31	6.11	6.30
3	0.00	6.68	6.68	6.68	6.68
	0.10	6.81	6.81	6.80	6.80
	0.20	7.27	7.24	7.17	7.21
	0.30	...	8.12	7.79	8.05

4. Background subtraction

If there is a background that must be subtracted, the determination – even the definition – of an upper limit becomes more difficult [5,7–9]. In the case that the mean rate of the background, B, is known with negligible uncertainty, the Particle Data Group has recommended [10] the adoption of the Bayesian procedure suggested by Helene [5]. In our notation their eq. (II.22) for the probability of observing n or fewer events becomes

$$p_n = K\, e^{-RS} \sum_{j=0}^{n} (B + RS)^j/j!, \qquad (19a)$$

where the normalization K depends on B and n through

$$K^{-1} = \sum_{j=0}^{n} B^j/j!. \qquad (19b)$$

To include the systematic uncertainty, a treatment like that given in section 3, but for simplicity going only to

linear terms in ΔR, gives [11]

$$\Delta U_n = \{(U_{n0} + B - n)/(U_{n0} + B)\} U_{n0}^2 \sigma_r^2/2. \quad (20)$$

In this instance U_{n0} represents the upper limit that would be read off the graphical solutions provided for this analysis [5,10]. For $B = 0$ this result agrees with eq. (17c) for all n, as it should. For $n = 0$ it agrees with eq. (12) for all B – a consequence of the Bayesian treatment of the background subtraction. It would also be of interest to be able to include an uncertainty in the value of B into this analysis, but the complex dependence of eqs. (19) on B seams to preclude anything other than the numerical approach suggested in section 2.1. A similar remark would apply to the various algorithms for background subtraction suggested by Prosper [9].

Swartz [12] has recently extended the analysis leading to eqs. (19) to include normalization to a sample of subsidiary events. There are then normalization uncertainties due to Poisson (actually trinomial) fluctuations in the number of those events. Within the statistical model adopted, the result (eq. (5) in his paper) is exact but quite complex; some numerical results are given in his tables I and II. Since Swartz's result and our eq. (20) are based on the same formalism, it is to be expected that our equation should give results corresponding to his tables. Using $\sigma_r = 1/\sqrt{D}$ in eq. (20), where D is the number of normalizing events, we find agreement [13] within $\sim 1\%$ for the $B = 0$ cases. For the $B \neq 0$ cases the agreement is within $\sim 2\%$ for $D = 50$ or 100. Our approximations lead to differences of $\sim 10\%$ when $D = 10$, since σ_r is then quite large ($> 30\%$). As in section 3, we conclude that, for the smaller values of σ_r commonly encountered, our simple formula suffices also for this case.

5. Conclusion

We have found a convenient solution to the problem of introducing the systematic uncertainty into the upper limit $\bar{R}_n$ quoted for a branching ratio experiment that has n candidate events and a best estimate $\hat{S}$ of the sensitivity with relative uncertainty σ_r: $\bar{R}_n = U_n/\hat{S}$, where

$$U_n = U_{n0}\left(1 + (U_{n0} - n)\sigma_r^2/2\right). \quad (21)$$

U_{n0} is the usual upper limit quoted for a Poisson parameter when n events are observed. For the most common case of $n = 0$ and 90% CL, eq. (21) reduces to

$$U_0 = 2.30\left(1 + 2.30\sigma_r^2/2\right). \quad (22)$$

As expected, the Poisson upper limit is relaxed when one incorporates the systematic uncertainty, but the effect is quite small in common cases (e.g., $\sigma_r < 10\%$). The formula allows one to combine quantitatively the statistical and systematic effects, in the same spirit (and with the same ease) that one normally combines them in the total uncertainty on a point estimate. The same general approach can be used when there is an accurately known background. More complex cases can require a numerical treatment.

Acknowledgements

We are grateful to Frederick James for a discussion of Bayesian and classical statistics in this context. One of us (R.D.C.) would like to thank Harvard University for its hospitality and partial support while part of this work was being done. This work was supported in part by the U.S. Department of Energy.

References

[1] See W.T. Eadie, D. Drijard, F.E. James, M. Roos and B. Sadoulet, Statistical Methods in Experimental Physics (American Elsevier, New York, 1971) pp. 190–202 and p. 213 for a general discussion of classical and Bayesian confidence limits. Ref. [10] explains the Poisson upper limit needed here.

[2] This problem can also occur in the case of a background subtraction. See ref. [10], p. 35 and F. James and M. Roos, Phys. Rev. D44 (1991) 299.

[3] C. Newman-Holmes, CDF internal memo, as cited in M. Contreras, Ph. D. thesis, Brandeis University (1990) and R.J. Hollebeek, P.K. Sinervo and H.H. Williams, The evaluation of upper limits for top quark production using combined measurements, CDF Note 1109 (January, 1990), both unpublished.

[4] T.M. Huber, A.R. Kunselman, A.C. Jannisen, G.A. Beer, G.R. Mason, A. Olin, T. Bowen, P.G. Halverson, C.A. Fry, K.R. Kendall, G.M. Marshall, B. Heinrich, K. Myrtle and J.B. Warren, Phys. Rev. D41 (1990) 2709.

[5] O. Helene, Nucl. Instr. and Meth. A212 (1983) 319. See also G. Zech, ibid., A277 (1989) 608, who claims to put Helene's Bayesian results for background subtraction on a classical statistical footing.

[6] J. Aitchison and J.A.C. Brown, The Lognormal Distribution (Cambridge University Press, 1957); ref. [1], pp. 79–80.

[7] O. Helene, Nucl. Instr. and Meth. A228 (1984) 120.

[8] F. James, Nucl. Instr. and Meth. A240 (1985) 203.

[9] H.B. Prosper, Nucl. Instr. and Meth. A238 (1985) 500; ibid. A241 (1985) 236; Phys. Rev. D37 (1988) 1153.

[10] Particle Data Group, Phys. Lett. B239 (1990) 1, p. III.36. Note the caveat against taking the result seriously if the observed n is statistically inconsistent with B.

[11] If the background were calculated rather than measured, it would likely include the sensitivity as a factor: $B = bS$, where b is the calculated portion. In that case the analysis of eqs. (19) becomes intractable by this means.

[12] M. Swartz, Nucl. Instr. and Meth. A294 (1990) 278. See also F. James and M. Roos, Nucl. Phys. B172 (1980) 475.

[13] The agreement is somewhat better for the less extreme values of the parameters, somewhat worse for the more extreme. The reader can readily compare values by evaluating eq. (20) and noting that the last column in Swartz's tables gives the values of U_{n0} needed.