Not reviewed, for internal circulation only

Draft version 0.2

ATLAS NOTE

September 29, 2017

Summary of and Recommendations for the "Better than Zero" Problem

The ATLAS Statistics Forum

Contributions from: K. Bjørke[a], W. Buttinger[b], D. Casadei[c], K. Cranmer[d], E. Gross[e], B. Malaescu[f], E. Nativ[g], A.L. Read[a], G. Redlinger[b], O. Vitells[g]

[a] *University of Oslo*
[b] *Brookhaven National Laboratory*
[c] *University of Birmingham*
[d] *New York University*
[e] *Weizmann Institute of Science*
[f] *Centre National de la Recherche Scientifique*
[g] *Formerly at Weizmann Institute of Science*

Abstract

In a Poisson counting experiment with a single source of events whose average event count (μ) is known exactly, the upper limit on μ when no events are observed is traditionally quoted to be 3 events at 95% CL. The "Better Than Zero Problem" is ATLAS jargon for the observation that the upper limit is seen to go below 3 with some methods when zero events are observed and systematic uncertainties are introduced. This note collects previous studies of the problem, and summarizes new work on some of the open issues. In addition, studies are presented on frequentist coverage, comparisons with asymptotic formulae, and a Bayesian perspective. The note concludes with a recommendation from the Statistics Forum: **the minimum value of the 95% CL upper limit on the number of events that should be reported is 3 if one uses the CL_s procedure in a counting experiment with non-zero background**.

Not reviewed, for internal circulation only

1 Introduction

In a Poisson counting experiment with a single source of events whose average event count (μ) is known exactly, the upper limit on μ when no events are observed is traditionally quoted as 3 events at 95% CL [1]. The "Better Than Zero Problem" is ATLAS jargon for the observation that in some situations with zero observed events, if systematic uncertainties are introduced the upper limit is seen to go below 3. The topic received much attention in March 2011, starting with a posting entitled "Better Than Zero"[1] to the Statistics Forum mailing list [2]. The various threads can be confusing when read today given that several other issues were being inter-mingled at the time such as the use of CL_s versus power-constrained limits (PCL) versus CL_{s+b}, frequentist vs Bayesian methods, and so on. The first document with a recommendation for a new "thumb rule" was Ref. [3]. This was quickly followed by two other documented studies in Refs. [4, 5]. A few issues remained unresolved after these earlier studies. Among them were:

- There appears to be a discontinuity in the upper limit when one goes from a situation with no systematic uncertainty to one in which there is an infinitesimally small uncertainty (though this is partially addressed in Refs. [4, 5]).
- The (more common) situation with a small background (with uncertainty) and an upper limit computed using the CL_s procedure was not explicitly studied.
- It has been observed over the years in several ATLAS analyses with zero observed events that the upper limit computed with CL_s goes below 3 if asymptotic formulae [6] are used.

This note attempts to clarify these issues. An understanding of different upper limits requires a discussion of frequentist coverage, which is explored for several different approaches, including objective Bayesian upper limits. Comparisons of full frequentist (and Bayesian) results with those obtained from asymptotic formulae are made. The note concludes by raising to the level of a Statistics Forum recommendation: **the minimum value of the 95% CL upper limit on the number of events that should be reported is 3 if one uses the CL_s procedure in a counting experiment with non-zero background**.

2 Summary of prior studies

We briefly summarize the afore-mentioned three documented studies on this subject.

The model under study [3] is a Poisson counting experiment with no background, and one source of signal with a systematic uncertainty on the signal efficiency. We write down the likelihood as

$$\mathcal{L}(\mu, \theta) = Pois(n|\theta\mu) \cdot Gaus(\tilde{\theta}|\theta, \sigma) \tag{1}$$

where n is the number of observed events, μ is the number of expected events, θ represents the signal efficiency, and the Gaussian term gives the distribution of the auxiliary measurement of the signal efficiency, given a nuisance parameter for the efficiency (θ) with a spread given by σ. The number of observed events is taken to be $n = 0$ and the auxiliary measurements of the efficiency are assumed to be centered at $\tilde{\theta} = 1$.

The one-sided profile-likelihood test statistic q_μ from Eq. 14 of Ref. [6] is used for limit setting:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \ , \\ 0 & \hat{\mu} > \mu \ , \end{cases} \tag{2}$$

[1] Tom LeCompte appears to be the originator of this clever aphorism.

Not reviewed, for internal circulation only

with

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} . \tag{3}$$

where $\hat{\hat{\theta}}$ in the numerator denotes the value of θ that maximizes $\mathcal{L}$ for the specified μ, while $\hat{\mu}$ and $\hat{\theta}$ in the denominator are the unconditional maximum likelihood estimators for μ and θ.

Figure 1 shows the normalized distribution of the test statistic for $\mu = 3$ with no systematic uncertainty (left) and with a systematic uncertainty on the signal efficiency of $\sigma = 0.05$ (right); the distribution is obtained by running pseudo-experiments on the model with RooStats.

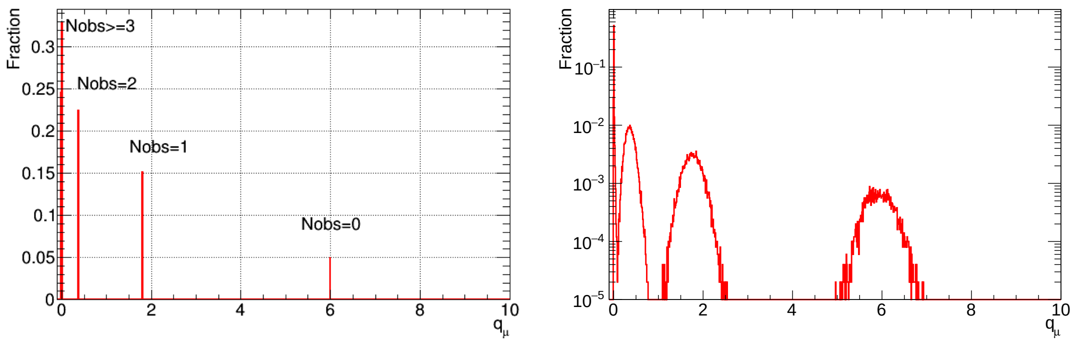


Figure 1: Normalized distribution (from pseudo-experiments obtained with RooStats) of the one-sided test statistic q_μ for $\mu = 3$ with no systematic uncertainty (left) and with a systematic uncertainty on the signal efficiency of $\sigma = 0.05$ (right).

From Fig. 1(left) it is apparent that when $\mu = 3$, 95% of the integral of the test statistic distribution is to the left of $q_\mu = 6$ (corresponding to $n_{obs} = 0$), leading to the classical 95% upper limit of $\mu = 3$. The right plot shows the distribution with a signal uncertainty of $\sigma = 0.05$, with the characteristic smearing of the test statistic distribution such that only half of the delta function at $q_\mu = 6$ is smeared to the left; since this factor of two from the smearing is introduced as soon as there is even an infinitesimal systematic uncertainty, leading to a seeming discontinuity in the behavior of the upper limit as the systematic uncertainty approaches 0.

Figure 2 shows a scatterplot of μ vs q_μ for the case with a signal efficiency systematic of $\sigma = 0.05$. The black line indicates the boundary where 95% of the integral of q_μ is to the left of the line. Following the usual rules of the Neyman construction, one then obtains the 95% CL upper limit on μ by the intersection of the line corresponding to the observation $n_{obs} = 0$ (i.e. the center of the lowest band in the figure) with the 95% boundary; this intersection occurs at $\mu \approx 2.3$ leading to an apparent reduction in the upper limit upon the introduction of a systematic uncertainty on the signal efficiency.

The behavior of upper limits under different conditions or different methods is intimately tied to the notion of coverage. For a given true (but unknown) value of μ_{true}, if one were to repeat the counting experiment N times (i.e. generating N different values of n_{obs}), computing an upper limit each time (i.e. generating a plot like Fig. 2 with toys for each experiment), coverage refers to the fraction of experiments in which the value of μ_{true} lies inside the confidence interval obtained in each experiment.[2]

[2] Such studies require a very large number of toy experiments if one were to do this by brute force. To be explicit, one scans over μ_{true}, generating N different values of n_{obs} for each value of μ_{true}. Then for each outcome of n_{obs} one runs a scan via toys of the test statistic distribution versus μ a la Fig. 2. Based on the scatterplot, one computes an upper limit on μ and then compares it to the value of μ_{true} to see if that particular experiment covered the true value or not.

Not reviewed, for internal circulation only

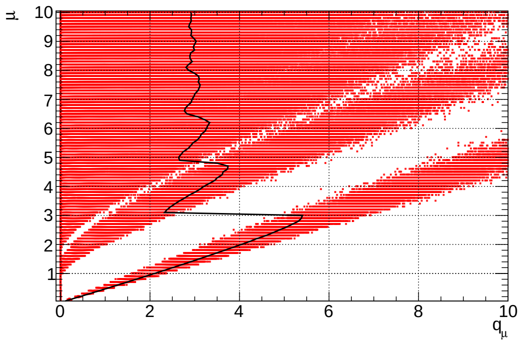


Figure 2: Scatterplot from toy experiments of μ vs q_μ for the case with a signal efficiency systematic of $\sigma = 0.05$. The black line indicates the boundary where 95% of the integral of q_μ for a given value of μ is to the left of the line.

If one is computing 95% CL upper limits, perfect coverage would imply that the value of μ_{true} lies inside the stated interval in 95% of the experiments.

It is well-known that the upper limit derived in classical Poisson statistics over-covers, i.e. the value of μ_{true} falls inside the confidence interval more often than it should. This is a consequence of the discrete nature of the test statistic distribution as illustrated in Fig. 1(left); it simply is not possible to define an exact 95% acceptance region due to the discrete distribution of the test statistic. Scanning over μ_{true}, the "dinosaur plot" of the coverage in the pure Poisson case is shown in Figure 3 where the overcoverage is apparent.

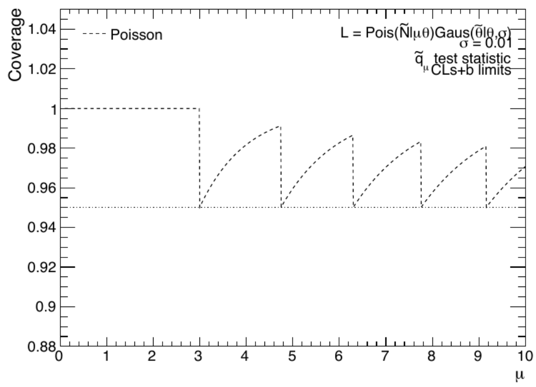


Figure 3: Coverage as a function of μ_{true} for a pure Poisson counting experiment with no uncertainties.

The introduction of a systematic uncertainty breaks the discreteness of the distribution of the test statistic and allows an acceptance region closer to the targeted value to be defined. The upper limit is lowered in the presence of systematic uncertainties, but the coverage becomes closer to the targeted value.[3]

Ref. [4] exploited the relationship with coverage further and derived a lower bound on the upper

[3]It is perhaps worth noting that introducing systematic uncertainties is not the only way to break the Poisson discreteness. An example (from Bill Murray) is a two-bin experiment with 0 events observed in the first bin (0 events expected for background and $3 - \epsilon$ events expected for signal, both without systematic uncertainties) and a second bin with large background and ϵ events of signal expected. The second bin "blurs" the expectation, leading to a similar breaking of the discreteness.

Not reviewed, for internal circulation only

limit if one was to guarantee a certain level of coverage, independent of the method used to compute the upper limit. Figure 4, taken from [4] shows the conclusion, with the reassuring behavior that the lower bound on the upper limit approaches the pure Poisson limit of 3 as the systematic uncertainty approaches 0.

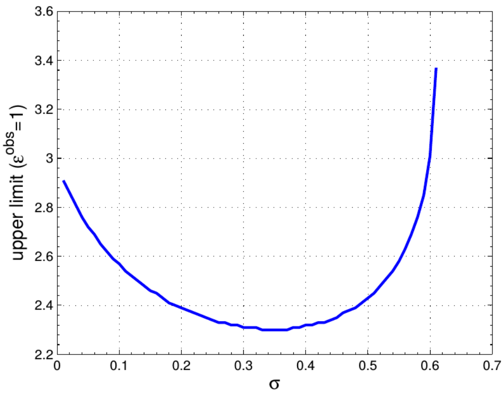


Figure 4: The lower bound on the upper limit as a function of the uncertainty on the signal efficiency. From [4].

Ref. [5] returned to the profile-likelihood test statistic and explored the change in the upper limit if the true central value of the efficiency differed from the one obtained from the auxiliary measurements. The conclusion was that if one wanted to cover deviations by up to 2σ from the true central value of the efficiency, the 95% CL upper limit would rise to a value close to the pure Poisson limit of 3 for small values of σ.

3 New Studies

3.1 mid-P values

The counter-intuitive behavior of the upper limit when going from a pure Poisson counting experiment to one with a systematic uncertainty on the signal efficiency has been explained in Sec. 2 as a symptom of the over-coverage of the standard p-value calculation in the pure Poisson case. Had a different recipe with better coverage properties been used in the pure Poisson case, the change in limits upon the introduction of systematic uncertainties would be more intuitive.

The mid-P value is often advocated in the statistics literature [7, 8] to avoid the bias coming from overcoverage, as noted for example in [9]. The mid-p-value is defined as the standard p-value for a test minus half the difference between it and the nearest lower possible value. This corresponds to splitting the delta functions in Fig. 1(left) down the middle when computing the tail fraction.

Not reviewed, for internal circulation only

3.2 Analytical studies of the profile construction

We consider two ways of constructing a confidence belt in the presence of nuisance parameters:

- In the profile construction, the test statistic distribution for each tested value of μ (the parameter of interest) is computed assuming the conditional maximum likelihood estimates (CMLE) of the nuisance parameters. The edge of the belt is defined by the values of the (μ-dependent) test statistic corresponding to 5% p-values (assuming 95% confidence limits). This is the standard approach at the LHC and has a number of good statistical properties plus the advantage that it is relatively inexpensive computationally to maximize the likelihood with respect to the nuisance parameters.

- In an alternative construction (hereafter called the "supremum construction") we sweep over all possible values of the nuisance parameters, computing the distribution of the test statistic for each set of nuisance parameters. The edge of the confidence belt is then defined by the value of the test statistic that ensures the p-value under all possible values of the nuisance parameter is less than or equal to 5%. In practice, this approach is computationally untenable in situations with a significant number of nuisance parameters. This approach also tends to over-cover [10].

The two questions we would like to address are:

1. What is the behaviour of these constructions in the small-uncertainty limit?

2. What are the coverage properties of these constructions? To what extend do these constructions overcover or undercover different values of the parameter of interest.

The determination of coverage can be computationally challenging, even in the case of the profile construction: a numerical approach would throw toys under a given hypothesis of the parameter of interest and nuisance parameters, and for each toy determine if the parameter of interest is inside the confidence interval corresponding to that toy. Determining the confidence interval for a toy will involve throwing further toys (as described for the profile construction in step 4 of [11]), therefore a fully numerical approach will involve generating many toys overall.

We therefore attempt an analytical treatment of the simple $s + b$ model, with the possibility of uncertainties on both terms:

$$\mathcal{L}(\mu;\theta) = Pois(\tilde{N} \mid \theta_\mu \mu + \theta_b b) \cdot Gaus(\tilde{\theta}_\mu \mid \theta_\mu, \sigma_\mu) \cdot Gaus(\tilde{\theta}_b \mid \theta_b, \sigma_b) \tag{4}$$

Figure 1 shows how the Gaussian constraint terms alter the (discrete) form of the q_μ test statistic distribution of a simple poisson model: each of the peaks that correspond to a different value of $\tilde{N}$ develop a gaussian shape depending on the magnitude of the uncertainty. The p-value for any given observed dataset ($\tilde{N}_{obs}, \tilde{\theta}_\mu^{obs}, \tilde{\theta}_b^{obs}$) can be estimated by assuming that the peaks corresponding to $\tilde{N} < \tilde{N}_{obs}$ will lie fully to the right of the observed q_μ, and that the contributing fraction of the peak that the observed q_μ lies in will be given by the outcomes where $\mu\tilde{\theta}_\mu + b\tilde{\theta}_b > \mu\tilde{\theta}_\mu^{obs} + b\tilde{\theta}_b^{obs}$. This is because increasing value of q_μ correspond to increasingly incompatibility with the μ hypothesis, which for a given value of $\tilde{N}$ will occur when the mean of the poisson term is greater than the mean obtained from the best fit to the observed data. The p-value of a given observed dataset if therefore approximately given by:

$$p_{\mu,\theta} = Pois(\tilde{N} < \tilde{N}_{obs} \mid \mu\theta_\mu + b\theta_b) + Pois(\tilde{N} = \tilde{N}_{obs} \mid \mu\theta_\mu) \cdot Prob(\mu\tilde{\theta}_\mu + b\tilde{\theta}_b > \mu\tilde{\theta}_\mu^{obs} + b\tilde{\theta}_b^{obs} \mid \theta_\mu, \sigma_\mu, \theta_b, \sigma_b). \tag{5}$$

In the case of the profile construction, we follow [11] and take θ_μ and θ_b to be given by the conditional maximum likelihood estimates (CMLE) $\hat{\hat{\theta}}_\mu$ and $\hat{\hat{\theta}}_b$. For the supremum construction these parameters are taken to be the ones that maximize $p_{\mu,\theta}$.

Not reviewed, for internal circulation only

The assumptions outlined above will only hold for small values of the uncertainties. The studies presented in the rest of this section will not be valid in cases where the uncertainties become large enough that the peaks start to overlap with one another to non-negligible degrees. But for the purposes of studying the small-uncertainty limiting cases these approximations for calculating the p-value are assumed to be valid.

3.2.1 The case of zero background and zero observed events

In the case where one observes 0 events ($\tilde{N}_{obs} = 0$) and expects 0 background ($b = 0$), the p-value from the tail integral of the test statistic (q_μ) distribution is given by:

$$p_{\mu,\theta} = Pois(\tilde{N} = 0 \mid \mu\theta_\mu) \cdot Gaus(\tilde{\theta}_\mu > \tilde{\theta}^{obs}_\mu \mid \theta, \sigma_\mu) \tag{6}$$

$$= e^{-\mu\theta_\mu} \cdot \Phi\left(\frac{\theta - \tilde{\theta}^{obs}_\mu}{\sigma_\mu}\right) \tag{7}$$

where Φ is the cumulative distribution of a Gaussian. The conditional maximum likelihood estimate (CMLE) of the nuisance parameter θ_μ is given by

$$\hat{\hat{\theta}}_\mu(\tilde{N}_{obs} = 0, \mu) = \tilde{\theta}^{obs}_\mu - \mu\sigma^2_\mu \tag{8}$$

so that

$$p_{\mu,\hat{\hat{\theta}}} = e^{-\mu\tilde{\theta}^{obs}_\mu} \cdot e^{\mu^2\sigma^2_\mu} \cdot \Phi(-\mu\sigma_\mu) \tag{9}$$

For small values of σ_μ the expression for $p_{\mu,\hat{\hat{\theta}}}$ in Eq. 9 approaches $0.5 \cdot e^{-\mu\tilde{\theta}^{obs}_\mu}$ which is half the value of the pure Poisson case. This is the same result as seen with toys in Sec. 2. For the supremum construction, we find the maximum p-value by differentiating Eq. 6 with respect to θ_μ, yielding the condition:

$$\mu\Phi\left(\frac{\theta - \tilde{\theta}^{obs}_\mu}{\sigma_\mu}\right) = \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{\frac{1}{2}\left(\frac{\theta - \tilde{\theta}^{obs}_\mu}{\sigma_\mu}\right)^2} \tag{10}$$

Figure 5 shows the numerical solution to Eq. 10 for two values of σ_μ (in the case where $\tilde{\theta}^{obs}_\mu = 1$) where it can be seen that the supremum occurs at $\theta_\mu \approx 1.003(1.0004)$ for $\sigma_\mu = 0.001(0.0001)$, differing slightly from the CMLE value of $\hat{\hat{\theta}}_\mu = 1.0$. Therefore the behaviour of the supremum construction is not the same as that of the profile construction: different limits will be obtained. Furthermore, these numerical solutions suggest that as $\sigma_\mu \to 0$, the value of θ_μ that maximises $p_{\mu,\hat{\hat{\theta}}}$ will tend to $\theta^{obs}_\mu + \infty\sigma_\mu$. In the limiting case of zero uncertainty ($\sigma_\mu = 0$) these results suggest that the suppremum p-value is

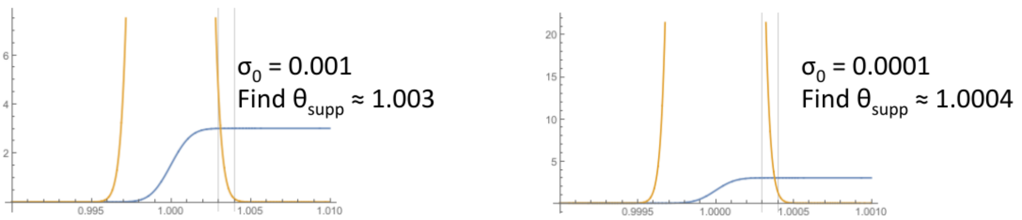


Figure 5: Numerical solutions to Eq. 10 for the value of θ_μ for two different values of σ_0 ($= \sigma_\mu$) in the case $\tilde{\theta}^{obs}_\mu = 1$.

Not reviewed, for internal circulation only

given by:

$$p_{\mu,\theta_{supp}} \to e^{-\mu\tilde{\theta}_\mu^{obs}} \cdot \Phi(\infty) = e^{-\mu\tilde{\theta}_\mu^{obs}} \tag{11}$$

which is identical to the p-value for a simple poisson case. This indicates that the supremum construction approaches the pure Poisson case smoothly as the systematic uncertainty reduces to 0. It can therefore be concluded that the apparent discontinuity in the upper limit of μ at $\sigma_\mu = 0$ arises from the properties of the profile construction itself. The supremum construction (which is computationally impractical except for the simplest of problems) shows no discontinuity.

3.2.2 Solutions to the general case

In order to calculate profile construction p-values in the general case of equation 5 we require $\hat{\hat{\theta}}_\mu$ and $\hat{\hat{\theta}}_b$, the CMLE values of the nuisance parameters. These are given by differentiating the likelihood (equation 4) by θ_μ and θ_b respectively to give two equations that would need to be solved simultaneously:

$$\hat{\hat{\theta}}_\mu = \frac{1}{2}\left(\tilde{\theta}_\mu^{obs} - \mu\sigma_\mu^2 - \frac{b\hat{\hat{\theta}}_b}{\mu} + \sqrt{\left(\tilde{\theta}_\mu^{obs} - \mu\sigma_\mu^2 - \frac{b}{\mu}\hat{\hat{\theta}}_b\right)^2 + 4\left(\tilde{N}_{obs}\sigma_\mu^2 - b\hat{\hat{\theta}}_b\sigma_\mu^2 + \frac{b}{\mu}\hat{\hat{\theta}}_b\tilde{\theta}_\mu^{obs}\right)}\right) \tag{12}$$

$$\hat{\hat{\theta}}_b = \frac{1}{2}\left(\tilde{\theta}_b^{obs} - b\sigma_b^2 - \frac{\mu\hat{\hat{\theta}}_\mu}{b} + \sqrt{\left(\tilde{\theta}_b^{obs} - b\sigma_b^2 - \frac{\mu}{b}\hat{\hat{\theta}}_\mu\right)^2 + 4\left(\tilde{N}_{obs}\sigma_b^2 - \mu\hat{\hat{\theta}}_\mu\sigma_b^2 + \frac{\mu}{b}\hat{\hat{\theta}}_\mu\tilde{\theta}_b^{obs}\right)}\right) \tag{13}$$

Solving these simultaneous equations is difficult, so instead we consider two simple cases where the solution to one of the equations if trivial: a zero background ($b = 0$) scenario with only uncertainty on the signal expectation ($\sigma_b = 0$), and a $b = 1$ scenario with only uncertainty on the background ($\sigma_\mu = 0$).

3.2.3 Coverage of profile construction in zero background case

In this case, $b = 0$ with no uncertainty, therefore equation 13 can be neglected and equation 12 simplifies to:

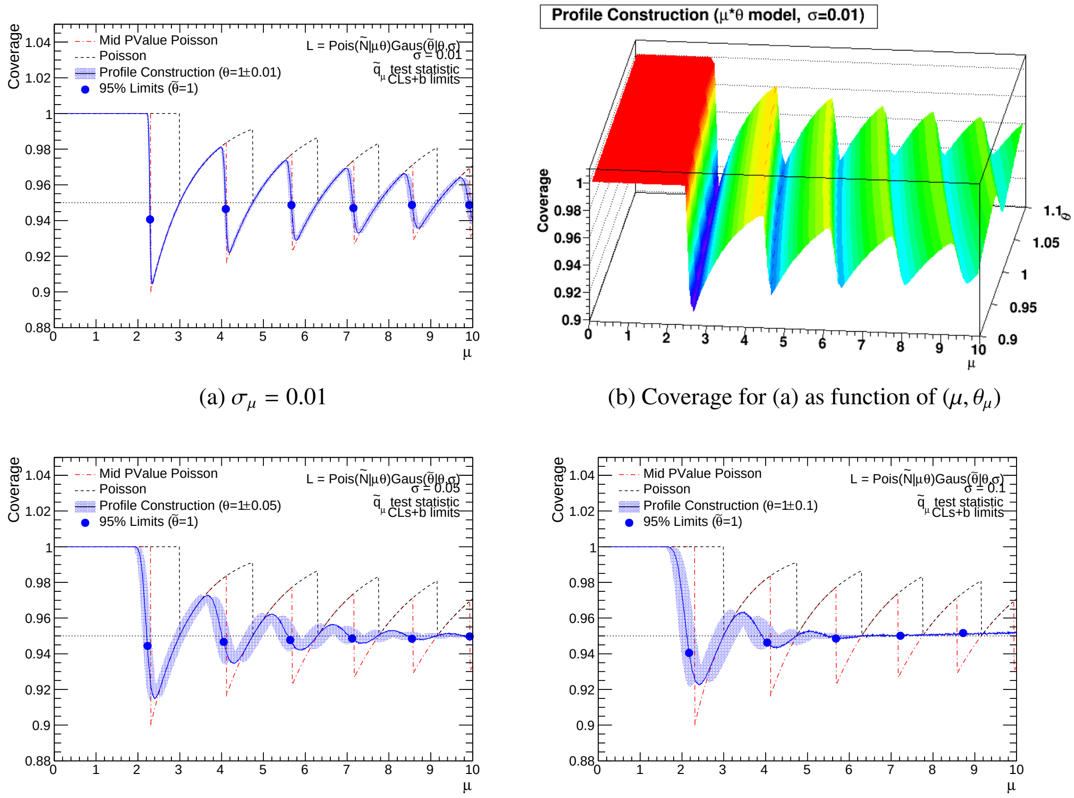
$$\hat{\hat{\theta}}_\mu = \frac{1}{2}\left(\tilde{\theta}_\mu^{obs} - \mu\sigma_\mu^2 + \sqrt{\left(\tilde{\theta}_\mu^{obs} - \mu\sigma_\mu^2\right)^2 + 4\tilde{N}_{obs}\sigma_\mu^2}\right) \tag{14}$$

This expression is combined with equation 5 and taking $\theta_\mu = \hat{\hat{\theta}}_\mu$ to estimate the p-value for toys generated under the $(\mu, \theta_\mu = 1))$ hypothesis. Figure 6 shows the result of this calculation for several different values of σ_μ. The "uncertainty" band of the profile construction coverage is calculated from repeating the coverage calculations at hypotheses $(\mu, \theta_\mu = 1 \pm \sigma_\mu)$, i.e. the band shows the spread of coverages in between two slices through the coverage asa function of (μ, θ_μ) (an example of which is shown in figure 6).

The coverage results (blue band) should be compared to the traditional poisson coverage (dashed black line) and mid-P value coverage (red dot-dashed line). These results confirm that the profile construction has coverage properties similar to and converging on the mid-P value construction.

The profile construction can be seen to undercover by as much as 5% for $\sigma_0 = 0.01$, though only in a restricted range of μ values. The undercoverage of the profile construction is seen to decrease as the systematic uncertainty on the signal efficiency increases. A brute force calculation with toy experiments of the coverage using RooStats confirms the results shown here.

Not reviewed, for internal circulation only



(a) $\sigma_\mu = 0.01$

(b) Coverage for (a) as function of (μ, θ_μ)

(c) $\sigma_\mu = 0.05$

(d) $\sigma_\mu = 0.1$

Figure 6: Coverage as a function of μ_{true} (and θ_μ in the case of (b)) of the 95% CL upper limit in a simple signal-only model (no background) for three different methods: traditional Poisson limit (black dashed), mid-P value (red dot-dashed) and profile likelihood construction (blue). The solid blue dots mark the upper limit obtained from an observation of $\tilde{\theta}_\mu^{obs} = 1, \tilde{N}_{obs} = 0, 1, 2, 3...$ See text for details.

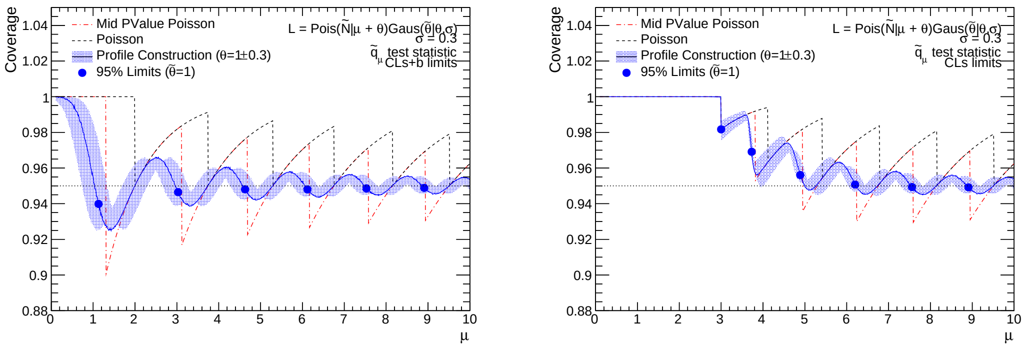
Not reviewed, for internal circulation only

3.2.4 Coverage of profile construction in zero signal uncertainty case

A similar exercise to that performed in the last section is performed with the model where $\theta_\mu = 1$ with no uncertainty, and $b = 1$. Equation 12 can be neglected ($\hat{\hat{\theta}}_\mu = 1$) and equation 13 simplifies to:

$$\hat{\hat{\theta}}_b = \frac{1}{2}\left(\tilde{\theta}_b^{obs} - \sigma_b^2 - \mu + \sqrt{\left(\tilde{\theta}_b^{obs} - \sigma_b^2 - \mu\right)^2 + 4\left(\tilde{N}_{obs}\sigma_b^2 - \mu\sigma_b^2 + \mu\tilde{\theta}_b^{obs}\right)}\right). \tag{15}$$

The results of this calculation are shown in Figure 7a. The damping effect on the coverage that is induced by the uncertainty is seen to be less significant for uncertainties on the background term compared to an equivalent relative uncertainty on the signal term. Additionally, the calculation here was repeated for the CL_s limit method, and is also shown in Figure 7b. The upper limit calculated by CL_s also approximately follows the coverage of the mid-P value Poisson case, overcovering at low values of μ.



(a) CL_{s+b} coverage

(b) CL_s coverage

Figure 7: Coverage of the profile construction (blue) as a function of μ of the 95% CL_{s+b} and CL_s upper limits for the simplified model described in section 3.2.4 with $\sigma_b = 0.3$. The coverage can be compared to: traditional Poisson limit (black dashed), mid-P value (red dot-dashed). The solid blue dots mark the upper limit obtained from an observation of $\tilde{\theta}_\mu^{obs} = 1, \tilde{N}_{obs} = 0, 1, 2, 3...$ See text for details.

3.2.5 Coverage properties of the supremum construction

The coverage properties of the supremum construction can be seen in Figure 8, calculated with the simple model described in section 3.2.3 for a systematic uncertainty on the signal efficiency of $\sigma_0 = 0.01$. The over-coverage of the supremum construction, similar to the traditional Poisson upper limit, is apparent.

It can be concluded that the introduction of a systematic uncertainty does indeed lower the upper limit from the profile construction relative to the limit from the pure Poisson case. This comes at the expense of a slight undercoverage of the profile construction for small values of the systematic uncertainty. Nevertheless, it must be emphasized that the overall coverage of the profile construction more accurately corresponds to the targeted coverage.

Not reviewed, for internal circulation only

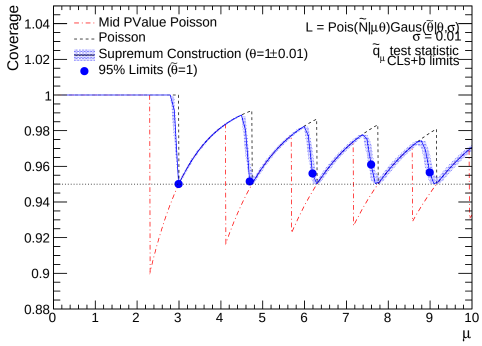


Figure 8: Coverage as a function of μ_{true} of the 95% CL upper limit for three different methods: traditional Poisson limit (black dashed), mid-P value (red dot-dashed) and supremum construction (blue). The solid blue dots mark the upper limit obtained from an observation of $\tilde{\theta}^{obs}_{\mu} = 1, \tilde{N}_{obs} = 0, 1, 2, 3...$ See text for details.

Not reviewed, for internal circulation only

3.3 Studies of a model with background

A counting model consisting of background and signal sources with an uncertainty on the background has also been studied.

$$\mathcal{L}(s, b) = Pois(n|s + b) \cdot Gaus(\tilde{b}|b, \sigma) \tag{16}$$

3.3.1 One-sided discovery

The one-sided profile-likelihood discovery test statistic (q_0) from Eq. 12 of Ref. [6] reduces to

$$q_0 = 2(n \ln \frac{n}{\hat{\hat{b}}} + \hat{\hat{b}} - n) + \left(\frac{\hat{\hat{b}} - b_0}{\sigma_b} \right)^2 \tag{17}$$

for $n > b_0$ ($q_0 = 0$ otherwise) where

$$\hat{s} = n - b_0 \tag{18}$$

$$\hat{b} = b_0 \tag{19}$$

$$\hat{\hat{b}} = \frac{(b_0 - \sigma_b)^2 + \sqrt{(b_0 - \sigma_b)^2 + 4n\sigma_b}}{2} \tag{20}$$

Figure 9 shows the distribution of q_0 for the case with $n_{obs} = 3, b = 0.78 \pm 0.18$ together with a few variations. The comparison with the q_0 distribution in the asymptotic approximation (i.e. the half-χ^2 distribution), shown as the red curve, illustrates how the asymptotic approximation breaks down for small background, but perhaps more interestingly how the asymptotic approximation can be valid for discovery even for small background if the background uncertainty starts to becomes significant.

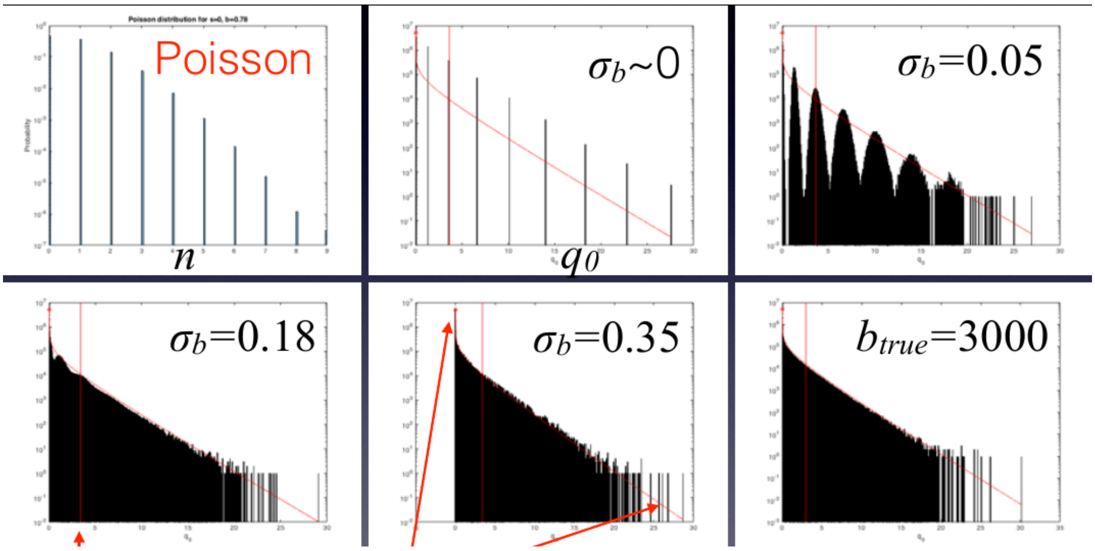


Figure 9: Distribution of the discovery test statistic q_0 for $n_{obs} = 3, b = 0.78 \pm 0.18$ together with a few variations. The red curve shows the q_0 distribution in the asymptotic approximation.

Table 1 examines a few different scenarios of n_{obs}, b and σ_b and compares the p-value calculated in a number of different ways: nominal background (i.e. no background uncertainty), profile-likelihood

Not reviewed, for internal circulation only

n_{obs}	$b \pm \sigma_b$	p_0 (nom b)	p_0 (asym)	p_0 ($\hat{\hat{b}}$)	p_0 (supp)
6	1.0 ± 0.4	$(1.03 \pm 0.01) \times 10^{-3}$	1.21×10^{-3}	$(1.03 \pm 0.01) \times 10^{-3}$	$(1.18 \pm 0.01) \times 10^{-3}$
8	1.0 ± 0.4	$(4.14 \pm 0.11) \times 10^{-5}$	4.85×10^{-5}	$(3.96 \pm 0.12) \times 10^{-5}$	$(4.96 \pm 0.22) \times 10^{-5}$
8	4.0 ± 0.1	$(3.61 \pm 0.01) \times 10^{-2}$	3.96×10^{-2}	$(3.55 \pm 0.01) \times 10^{-2}$	$(3.91 \pm 0.01) \times 10^{-2}$
7	0.78 ± 0.18	$(1.68 \pm 0.07) \times 10^{-5}$	2.17×10^{-5}	$(1.64 \pm 0.07) \times 10^{-5}$	$(1.95 \pm 0.08) \times 10^{-5}$
15	4.0 ± 0.1	$(1.19 \pm 0.06) \times 10^{-5}$	1.38×10^{-5}	$(1.17 \pm 0.06) \times 10^{-5}$	$(1.32 \pm 0.07) \times 10^{-5}$
15	4.0 ± 1.5	$(4.70 \pm 0.04) \times 10^{-4}$	5.03×10^{-4}	$(4.72 \pm 0.04) \times 10^{-4}$	$(4.96 \pm 0.04) \times 10^{-4}$
40	10.0 ± 3.0	$(1.0 \pm 0.3) \times 10^{-7}$	1.31×10^{-7}	$(1.8 \pm 0.4) \times 10^{-7}$	$(1.6 \pm 0.4) \times 10^{-7}$
150	100 ± 10	4.9×10^{-4}	5.0×10^{-4}	4.9×10^{-4}	5.0×10^{-4}

Table 1: Discovery p-value calculated in a number of different ways: nominal background (i.e. no background uncertainty), profile-likelihood with asymptotic approximation, profile likelihood with 10^6 toys, and supremum construction. The values are shown for a few representative cases of n_{obs}, b and σ_b.

with asymptotic approximation, profile likelihood, and supremum construction. Two trends (apparently independent of n_{obs}, b and σ_b) seem to be visible in this table: 1. the profile construction slightly under-estimates the discovery p-value compared to the supremum construction (it is therefore slightly aggressive), and 2. the asymptotic approximation is rather close to the result from the supremum construction.

3.3.2 Exclusion based on CL_s

The intuitive expectation for the behavior of CL_s limits with and without systematics when $n_{obs} = 0$ is that both the numerator and denominator of CL_s are affected in the same way by the smearing of the delta functions due to the systematic uncertainty. The ratio is therefore expected to give an upper limit very similar to the pure Poisson case. Figure 10 shows the distribution of the test statistic q_μ for the signal+background and background-only hypotheses for a background of $b = 1.2$. The red curves in Fig. 10 show the q_μ distribution from the asymptotic approximation. The deviation from the asymptotic approximation is different for the $s + b$ and *b only* hypotheses and thus the "cancellation" in CL_s is not expected to work as well. Thus, while it is possible (see Sec 3.4) for the upper limit from CL_s to go below 3 when $n_{obs} = 0$ when the asymptotic approximation is used, such results should not be quoted in ATLAS publications because the asymptotic approximation is clearly not valid.

Not reviewed, for internal circulation only

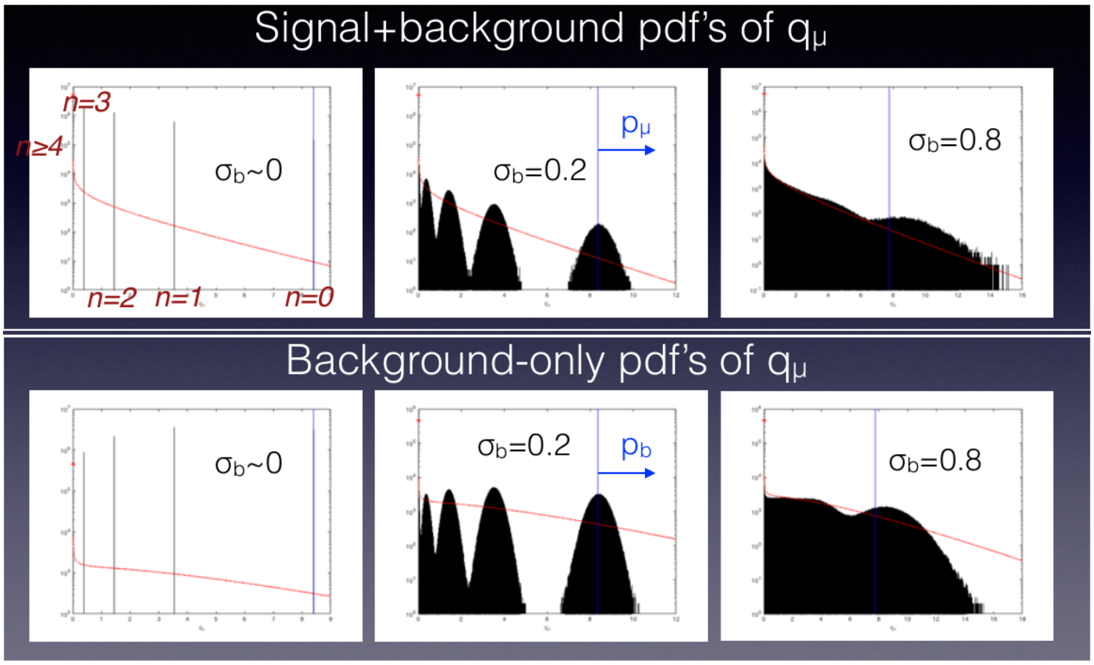


Figure 10: Distribution of the test statistic q_μ for a background $b = 1.2$ and different assumptions on σ_b. The top (bottom) row shows the test statistic for the signal+background (background-only) hypothesis. The red curves show the q_μ distribution based on the asymptotic approximation.

Not reviewed, for internal circulation only

3.4 Numerical comparisons of asymptotics versus toys

This section summarizes numerical comparisons of upper limits derived with the asymptotic approximation versus toys. The model is the same as Eq. 16. Tables 2 and 3 show the 95% CL_s upper limits for the case where $n_{obs} = 0$ with varying levels of expected background and background uncertainty, comparing the asymptotic approximation with toys, respectively. In the case of toys, 10000 (5000) pseudo-experiments are thrown for the Null (Alt) hypotheses, where the Null (Alt) hypothesis is S+B (B only). The upper limit from asymptotics can be as low as 1.9 events, to be compared to the of 3 events from toys.

	σ_b/b							
b	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
3.0	2.037	2.530	2.566	2.568	2.568	2.568	2.568	2.568
1.0	2.048	2.326	2.336	2.337	2.337	2.337	2.337	-
0.3	2.061	2.134	2.137	2.137	2.137	2.137	-	-
0.1	2.014	2.025	2.025	2.026	2.026	-	-	-
0.03	1.964	1.965	1.965	1.965	-	-	-	-
0.01	1.940	1.940	1.940	-	-	-	-	-
0.003	1.928	1.928	-	-	-	-	-	-
0.001	1.924	-	-	-	-	-	-	-

Table 2: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} = 0$ for different values of expected background and uncertainty. The results are obtained with the asymptotic approximation and can be compared to the results from toys in Table 3.

	σ_b/b							
b	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
3.0	3.06 ± 0.05	3.06 ± 0.15	3.09 ± 0.19	2.93 ± 0.19	3.07 ± 0.24	3.06 ± 0.23	3.06 ± 0.23	3.06 ± 0.23
1.0	3.05 ± 0.05	2.98 ± 0.08	2.83 ± 0.07	2.83 ± 0.07	2.83 ± 0.07	2.83 ± 0.07	2.83 ± 0.07	-
0.3	3.04 ± 0.05	3.05 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	-	-
0.1	2.99 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	3.04 ± 0.05	-	-	-
0.03	2.97 ± 0.04	2.98 ± 0.04	2.97 ± 0.04	2.97 ± 0.04	-	-	-	-
0.01	2.99 ± 0.05	2.98 ± 0.05	2.98 ± 0.05	-	-	-	-	-
0.003	3.00 ± 0.05	3.00 ± 0.05	-	-	-	-	-	-
0.001	3.01 ± 0.05	-	-	-	-	-	-	-

Table 3: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} = 0$ for different values of expected background and uncertainty. The results are obtained with 10000 (5000) toys for the Null (Alt) hypotheses, where the Null (Alt) hypothesis is S+B (B only); the same starting seed is used in each cell of the table.

On the other hand, in the case where $n_{obs} = b$, the asymptotic approximation is seen to be quite robust with respect to toys. Tables 4 and 5 compare the 95% CL_s upper limits for the model of Eq.16 in the case where $n_{obs} = b$ with varying levels of uncertainty on b. A comparison of the tables suggests that the asymptotic approximation is good to within 10% of the result from toys even for $n_{obs} = b$ as low as 1 event.

Finally, Appendix A shows a comparison of the asymptotic approximation versus toys for the cases where $n_{obs} < b$ for various values of n_{obs} and b. Though the statistical uncertainty from some of the

Not reviewed, for internal circulation only

	σ_b/b							
b	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
1	3.986	3.450	3.408	3.404	3.404	3.404	3.404	3.404
2	5.610	4.340	4.190	4.178	4.177	4.177	4.177	4.177
3	7.106	5.111	4.808	4.783	4.780	4.780	4.780	4.780
4	8.528	5.824	5.339	5.298	5.294	5.293	5.293	5.293
5	9.916	6.507	5.834	5.772	5.765	5.764	5.764	5.764
6	11.25	7.161	6.268	6.180	6.171	6.170	6.170	6.170
7	12.56	7.800	6.669	6.553	6.542	6.541	6.541	6.541
8	13.86	8.419	7.046	6.922	6.908	6.906	6.906	6.906
9	15.12	9.023	7.422	7.258	7.240	7.238	7.238	7.238
10	16.38	9.620	7.773	7.564	7.545	7.543	7.543	7.543
11	17.63	10.22	8.098	7.881	7.856	7.854	7.854	7.854
12	18.87	10.82	8.434	8.165	8.134	8.131	8.131	8.131
13	20.08	11.40	8.751	8.448	8.417	8.414	8.414	8.414
14	21.29	11.98	9.051	8.717	8.677	8.674	8.673	8.673
15	22.50	12.56	9.360	8.977	8.937	8.933	8.933	8.933

Table 4: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} = b$ for different values of expected background and uncertainty. The results are obtained with the asymptotic approximation and can be compared with the results from toys in Table 5.

	σ_b/b							
b	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
1	4.13 ± 0.06	3.71 ± 0.07	3.76 ± 0.07	3.79 ± 0.07	3.78 ± 0.07	3.77 ± 0.07	3.49 ± 0.06	3.13 ± 0.05
2	5.99 ± 0.06	4.59 ± 0.08	4.65 ± 0.09	4.52 ± 0.08	4.53 ± 0.08	4.53 ± 0.08	4.44 ± 0.08	4.65 ± 0.09
3	7.74 ± 0.08	5.42 ± 0.09	4.85 ± 0.08	5.09 ± 0.09	5.10 ± 0.09	5.10 ± 0.09	4.75 ± 0.07	4.62 ± 0.07
4	9.31 ± 0.09	6.09 ± 0.09	5.46 ± 0.09	5.47 ± 0.09	5.48 ± 0.09	5.49 ± 0.09	5.46 ± 0.09	5.46 ± 0.09
5	10.58 ± 0.09	6.80 ± 0.10	5.95 ± 0.10	5.89 ± 0.10	5.91 ± 0.10	5.91 ± 0.09	5.89 ± 0.09	5.76 ± 0.08
6	12.06 ± 0.10	7.39 ± 0.11	6.36 ± 0.10	6.42 ± 0.11	6.45 ± 0.11	6.44 ± 0.11	6.43 ± 0.11	6.28 ± 0.09
7	13.61 ± 0.12	8.10 ± 0.12	6.80 ± 0.11	6.78 ± 0.11	6.78 ± 0.11	6.79 ± 0.11	6.77 ± 0.11	6.72 ± 0.10
8	14.85 ± 0.12	8.48 ± 0.11	7.11 ± 0.11	7.14 ± 0.11	7.14 ± 0.11	7.13 ± 0.11	7.15 ± 0.11	7.03 ± 0.10
9	16.07 ± 0.13	9.19 ± 0.12	7.71 ± 0.12	7.59 ± 0.13	7.49 ± 0.12	7.46 ± 0.12	7.48 ± 0.12	7.43 ± 0.11
10	17.42 ± 0.14	9.85 ± 0.13	7.98 ± 0.13	7.60 ± 0.12	7.59 ± 0.12	7.61 ± 0.12	7.62 ± 0.11	7.60 ± 0.11
11	18.89 ± 0.15	10.75 ± 0.15	8.24 ± 0.12	8.00 ± 0.12	8.03 ± 0.13	8.02 ± 0.12	8.01 ± 0.12	8.01 ± 0.12
12	20.12 ± 0.16	11.19 ± 0.15	8.50 ± 0.13	8.35 ± 0.14	8.31 ± 0.12	8.34 ± 0.13	8.33 ± 0.12	8.32 ± 0.11
13	21.31 ± 0.16	11.79 ± 0.15	8.96 ± 0.14	8.64 ± 0.13	8.77 ± 0.14	8.80 ± 0.14	8.80 ± 0.14	8.77 ± 0.13
14	22.44 ± 0.16	12.53 ± 0.17	9.29 ± 0.14	9.01 ± 0.15	8.90 ± 0.14	8.91 ± 0.14	8.94 ± 0.14	8.98 ± 0.13
15	23.71 ± 0.18	12.91 ± 0.17	9.37 ± 0.14	9.23 ± 0.15	9.02 ± 0.14	9.06 ± 0.14	9.08 ± 0.14	9.067 ± 0.14

Table 5: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} = b$ for different values of expected background and uncertainty. The results are obtained with 10000 (5000) toys for the Null (Alt) hypotheses, where the Null (Alt) hypothesis is S+B (B only); the same starting seed is used in each cell of the table.

toys is significant, it appears that again, the asymptotic approximation reproduces the results from toys to within 10% for $b < 6$ and $1 < n_{obs} < b$.

Not reviewed, for internal circulation only

Not reviewed, for internal circulation only

3.5 Bayesian perspective

The "better than zero" problem has also been studied from the Bayesian perspective [12] in the framework of what is known as "reference analysis" [13], which obtains an objective posterior by making use of a reference prior. The latter is formally defined with an information-theoretic approach as the "least informative prior", in the sense that it maximizes the amount of prior *missing* information. The model addressed by [12] is a Poisson counting experiment with two parameters s and b describing the expected numbers of events coming from signal and background processes. An informative prior is used for the nuisance parameter b and a reference prior is used for the parameter of interest s.

The background prior is written in the form of a Gamma density

$$p(b) = \mathrm{Ga}(b \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} b^{\alpha-1} e^{-\beta b} \tag{21}$$

Writing the expected background as $b = b_0 \pm \sigma_b$ the Gamma parameters are fixed via $b_0 = E[b] = \alpha/\beta$ and $\sigma_b^2 = V[b] = \alpha/\beta^2$, which give $\alpha = (b_0/\sigma_b)^2$ and $\beta = b_0/\sigma_b{}^2$. The reference prior for the signal $\pi(s)$ has a complicated form [12], but is implemented in BAT and is very often well approximated by $\pi_0(s) = \sqrt{b_0/(s+b_0)}$ [14], which tends to the flat prior for large values of b_0.

The joint prior is

$$p(s, b) = \pi(s)\, p(b) \tag{22}$$

and the joint posterior density for s and b is obtained via Bayes' theorem as

$$p(s, b|n) \propto \mathrm{Poi}(n|s+b)\, \pi(s)\, p(b) \tag{23}$$

while the marginal posterior density for s (independent of the background) is obtained by integrating over the background:

$$p(s|n) = \int p(s, b|n)\, \mathrm{d}b \tag{24}$$

The upper bounds on s at 95% credibility level when $n = 0$ in a model with signal and non-zero background and uncertainty are summarized in Table 6. In the limit $b_0, \sigma_b \to 0$ the upper bound changes smoothly and approaches the value of 2.41 (no jumps when $b_0 \to 0$ nor when $\sigma_b \to 0$).

	σ_b/b_0			
b_0	10%	20%	50%	100%
0.5	2.55	2.55	2.56	2.57
1.0	2.65	2.66	2.68	2.66
2.0	2.77	2.78	2.80	2.74
4.0	2.88	2.89	2.91	2.78
8.0	2.96	2.98	2.99	2.81

Table 6: Upper bounds at 95% credibility level with $n_{obs} = 0$ for different values of expected background and uncertainty. The results are obtained following the so-called reference analysis for a Poisson counting experiment with two parameters s and b describing the expected numbers of events coming from signal and background processes.

The difference in the upper bound compared to the upper limits from the CL_s procedure can be attributed to a difference in the coverage properties of the Bayesian results. Reference priors are the "probability matching" priors whose q-credible intervals achieve the coverage q most quickly for increasing sample sizes [15]. When averaging over all possible true values of s and b in $\mathrm{Poi}(s+b)$ one gets a coverage which equals the posterior probability, but scanning the (s, b) parameter space one gets under- and over-coverage because of the discrete nature of the model. The coverage tends to be smaller than the posterior probability for very small values of the true signal ($s < 0.2 - 0.5$) and to be greater than the posterior probability with $s > 1$, with sizable oscillations. Coverage studies of the objective Bayesian results are reported in the appendix of [12], and show that for fixed s there is a tendency to overcover whenever the true background is smaller than b_0, whereas a true background larger than expected causes undercoverage.

Not reviewed, for internal circulation only

Not reviewed, for internal circulation only

4 Discussion

The profile construction has been studied for the case where the number of observed events $n_{obs} = 0$, with two counting models: one in which there is no background but a systematic uncertainty on the signal efficiency, and the second in which there is no signal uncertainty but there is background with an associated uncertainty. The upper limit from the profile construction in the presence of systematic uncertainties has been compared with the case of no uncertainty, calculated in the traditional Poisson way and via the mid-P value. A comparison with the supremum construction has also been performed.

In a simple Poisson event counting model with no uncertainty, the 95% CL_{s+b} limit for a zero-event observation ($n_{obs} = 0$) corresponds to $\approx$ 3 events when calculated with the traditional definition of p-values, or $\approx$ 2.3 events if using the mid-P value definition of p-values (see section 3.1 for further detail of the mid-P value). The CL_s limit in both cases, however, corresponds to $\approx$ 3 events. The is summarised by table 7.

	p_μ	p_0	$p_{CL_s} = \frac{p_\mu}{p_0}$	CL_{s+b}	CL_s
Traditional	$e^{-\mu}$	$e^{-0} = 1.0$	$e^{-\mu}$	$p_\mu = 0.05$ when $\mu \approx 3$	$p_{CL_s} = 0.05$ when $\mu \approx 3$
mid-P value	$\frac{1}{2}e^{-\mu}$	$\frac{1}{2}e^{-0} = 0.5$	$e^{-\mu}$	$p_\mu = 0.05$ when $\mu \approx 2.3$	$p_{CL_s} = 0.05$ when $\mu \approx 3$

Table 7: P values for the $n_{obs} = 0$ observation in a simple Poisson counting model with no uncertainty.

When nuisance parameters are introduced to the model (to describe systematic uncertainties, for example), the standard recommendation is to use the profile construction based on the q_μ test statistic for confidence intervals. The studies in section 3.2.3 demonstrate that the typical profile construction of CL_{s+b} limits follows the mid-P value behaviour for limits and coverage. As the uncertainties approach 0, the limit obtained with the profile construction will tend towards the mid-P value limit. One must use supremum construction to have a CL_{s+b} limit that tends towards the traditional p value Poisson limit in the 0-uncertainty limit. However, the supremum construction is computationally untenable, except in the simplest problems.

The coverage of the CL_{s+b} limits from profile construction oscillate about the target coverage of 95%, with undercoverage of up to 5% in some regions of the parameter space. However, with any moderate degree of uncertainty in the model (on signal or background) the coverage is within 1% of the target. The supremum construction guarantees coverage of at least 95%.

The coverage of the CL_s limits from the profile construction exhibit less undercoverage than the CL_{s+b} limits, with more overcoverage in the parameter space $\mu < 2$ (see figure 7).

Asymptotic formulae for the distribution of the q_μ test statistic must be used with care. For the computation of discovery p-values, the asymptotic formulae seems to provide a good approximation of the p-value calculated from the supremum construction, which is slightly more conservative compared to the p-value computed from the profile construction (see table 1). The asymptotic formulae also serve as a good approximation to the p-value when the background uncertainty is relatively large, even if the background is small.

When calculating exclusion limits with CL_s for $n_{obs} = 0$, however, asymptotic formulae must not be used; it is possible for the upper limit from CL_s to go below a signal strength corresponding to 3 signal events with asymptotic formulae when $n_{obs} = 0$ (see table 2), but the limit obtained from toys does not go below this value (see table 3). On the other hand, if $n_{obs} = b$, the asymptotic formulae are correct to within 10% (or better) of the result from toys, even for values of n_{obs} as low as 1 event (see tables 4 and 5). In these cases the CL_s limit obtained with asymptotic formulae correspond to signal yields greater than three events. **It is therefore a good indicator that a CL_s limit obtained from asymptotic formulae is invalid if it corresponds to a signal yield of fewer than three events.**

Bayesian upper bounds for $n_{obs} = 0$ for the model with a perfectly known signal plus a background

Not reviewed, for internal circulation only

with uncertainty have been studied in [12], based on so-called "reference analysis". An attractive feature of this analysis is that there is no discontinuity in the upper bound as the background uncertainty approaches 0. The upper bounds from the Bayesian analysis can differ significantly from those obtained by the CL_s method. Again, this is simply a reflection of the different coverage properties of the two approaches.

One can argue about which method has the "best" coverage properties. Because of the discrete nature of Poisson counting experiments (even in the presence of systematic uncertainties), no method can have exactly the desired coverage everywhere and there is necessarily a trade-off between average coverage versus under/over coverage in a restricted range of μ. Historically, ATLAS agreed with CMS to use the CL_s method, which is conservative in its coverage properties, i.e. it tends to over-cover.

5 Conclusion

We conclude by reinforcing the recommendation to continue our practice of using CL_s for setting exclusion limits in counting experiments with non-zero background due to the fact that its coverage is conservative for low values of n_{obs}. We extend this recommendation such that the upper limit obtained using CL_s should never be reported to be smaller than 3 events at 95% CL, and that asymptotic formulae should not be used to derive upper limits if the obtained limit corresponds to a signal yield of fewer than 3 events.

In the (rare) case of a counting experiment with 0 background and $n_{obs} = 0$, it would not be incorrect to report a 95% CL_{s+b} upper limit less than 3, but additional remarks about the possibility of a slight undercoverage of the method might be useful in a publication.

A Further comparisons of asymptotics versus toys

CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} < b$ for different values of n_{obs} are shown in Tables 8 and 9, based on asymptotic formulae and toys, respectively.

Not reviewed, for internal circulation only

Not reviewed, for internal circulation only

		σ_b/b							
n_{obs}	**b**	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
1	2	4.105	3.306	3.215	3.208	3.207	3.207	3.207	3.207
1	3	4.137	3.249	3.140	3.130	3.129	3.129	3.129	3.129
1	4	4.151	3.217	3.103	3.092	3.091	3.091	3.091	3.091
1	5	4.178	3.228	3.092	3.076	3.074	3.074	3.074	3.074
1	6	4.181	3.217	3.075	3.057	3.055	3.055	3.055	3.055
2	3	5.713	4.120	3.881	3.861	3.859	3.859	3.859	3.859
2	4	5.755	3.980	3.702	3.677	3.675	3.674	3.674	3.674
2	5	5.801	3.900	3.595	3.561	3.557	3.557	3.557	3.557
2	6	5.814	3.834	3.510	3.483	3.480	3.480	3.480	3.480
3	4	7.195	4.858	4.441	4.407	4.403	4.403	4.403	4.403
3	5	7.258	4.686	4.220	4.173	4.168	4.168	4.168	4.168
3	6	7.286	4.526	4.030	3.989	3.984	3.984	3.984	3.984
4	5	8.611	5.552	4.952	4.902	4.897	4.896	4.896	4.896
4	6	8.660	5.353	4.668	4.595	4.587	4.587	4.587	4.587
5	6	9.987	6.232	5.408	5.334	5.326	5.325	5.325	5.325

Table 8: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} < b$ for different values of n_{obs}, expected background and uncertainty. The results are obtained with the asymptotic approximation and can be compared with the results from toys in Table 9.

		σ_b/b							
n_{obs}	**b**	100%	33%	10%	3.3%	1%	0.33%	0.1%	0.033%
1	2	4.39 ± 0.06	3.50 ± 0.09	3.68 ± 0.11	3.69 ± 0.11	3.69 ± 0.11	3.69 ± 0.11	3.71 ± 0.11	3.76 ± 0.10
1	3	4.24 ± 0.06	3.33 ± 0.11	3.39 ± 0.15	3.62 ± 0.17	3.64 ± 0.17	3.68 ± 0.17	3.69 ± 0.17	3.69 ± 0.17
1	4	4.24 ± 0.06	3.32 ± 0.16	3.30 ± 0.22	3.42 ± 0.21	3.39 ± 0.20	3.38 ± 0.20	3.35 ± 0.19	3.48 ± 0.20
1	5	4.44 ± 0.08	3.36 ± 0.22	3.21 ± 0.25	3.54 ± 0.58	3.54 ± 0.59	3.54 ± 0.57	3.54 ± 0.57	3.54 ± 0.57
1	6	4.44 ± 0.08	3.43 ± 0.33	2.52 ± 0.27	3.0 ± 1.4	3.5 ± 2.6	3.5 ± 3.8	3.6 ± 3.4	3.6 ± 2.8
2	3	6.15 ± 0.07	4.42 ± 0.10	4.24 ± 0.11	4.14 ± 0.10	4.13 ± 0.10	4.14 ± 0.10	4.14 ± 0.10	4.14 ± 0.10
2	4	6.12 ± 0.07	4.19 ± 0.12	3.86 ± 0.13	3.99 ± 0.13	3.40 ± 0.13	4.00 ± 0.15	3.99 ± 0.15	3.99 ± 0.12
2	5	6.07 ± 0.07	4.22 ± 0.16	3.85 ± 0.20	3.83 ± 0.21	3.82 ± 0.20	3.85 ± 0.20	3.84 ± 0.20	3.87 ± 0.20
2	6	6.05 ± 0.07	4.05 ± 0.18	3.43 ± 0.24	3.03 ± 0.19	3.54 ± 0.28	4.09 ± 0.46	3.54 ± 0.30	3.54 ± 0.30
3	4	7.78 ± 0.08	5.26 ± 0.11	4.71 ± 0.11	4.70 ± 0.11	4.73 ± 0.11	4.74 ± 0.11	4.75 ± 0.11	4.72 ± 0.11
3	5	7.73 ± 0.07	4.81 ± 0.12	4.50 ± 0.15	4.48 ± 0.15	4.49 ± 0.15	4.51 ± 0.15	4.49 ± 0.15	4.53 ± 0.16
3	6	7.66 ± 0.08	4.71 ± 0.14	4.22 ± 0.19	4.27 ± 0.19	4.23 ± 0.18	4.24 ± 0.18	4.29 ± 0.18	4.29 ± 0.18
4	5	9.07 ± 0.08	5.95 ± 0.12	5.22 ± 0.12	5.21 ± 0.12	5.27 ± 0.12	5.28 ± 0.12	5.30 ± 0.12	5.29 ± 0.12
4	6	9.04 ± 0.08	5.73 ± 0.12	4.87 ± 0.16	4.74 ± 0.13	4.79 ± 0.13	4.80 ± 0.15	4.85 ± 0.15	4.94 ± 0.15
5	6	10.62 ± 0.09	6.53 ± 0.12	5.42 ± 0.11	5.44 ± 0.12	5.44 ± 0.12	5.48 ± 0.12	5.51 ± 0.11	5.62 ± 0.11

Table 9: CL_s 95% upper limits for the model of Eq. 16 with $n_{obs} < b$ for different values of n_{obs}, expected background and uncertainty. The results are obtained with 10000 (5000) toys for the Null (Alt) hypotheses, where the Null (Alt) hypothesis is S+B (B only); the same starting seed is used in each cell of the table.

Not reviewed, for internal circulation only

References

[1] G. Cowan, *Statistics section in PDG review*, http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf.

[2] The archives are available at: https://groups.cern.ch/group/hn-atlas-physics-Statistics/default.aspx.

[3] K. Cranmer, *Rules of thumb for limits with* $b << 1$ *and the unconditional ensemble*, https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/BetterThanZero.Cranmer.pdf.

[4] O. Vitells, *Upper limit for a Poisson measurement with systematic uncertainty*, https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/BetterThanZero.Vitells.pdf.

[5] E. Gross and E. Nativ, *Upper limit for a Poisson measurement with zero observed events and a systematic uncertainty*, https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/BetterThanZero.Gross_Nativ.pdf.

[6] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554.

[7] R. Routledge, *Practicing safe statistics with the mid-p*, The Canadian Journal of Statistics **22** (1994) 103–110.

[8] G. Berry and P. Armitage, *Mid-P confidence intervals: a brief review*, The Statistican **44** (1995) 417–423.

[9] L. Demortier, *Constructing Ensembles of Pseudo-Experiments*, eConf **C030908** (2003) WEMT003, `arXiv:physics/0312100 [physics.data-an]`. [,256(2003)].

[10] G. Punzi, *Ordering algorithms and confidence intervals in the presence of nuisance parameters*, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, September 12-15, 2005*, pp. 88–92. 2005. `arXiv:physics/0511202 [physics.data-an]`.

[11] The Statistics Forum, *Frequentist Limit Recommendations*, https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/Frequentist_Limit_Recommendation.pdf.

[12] D. Casadei, *Reference analysis of the signal + background model in counting experiments*, JINST **7** (2012) P01012, `arXiv:1108.4270 [physics.data-an]`.

[13] J. O. Berger, J. M. Bernardo, and D. Sun, *The formal definition of reference priors*, Annals of Statistics **37** (2009) 905–038, `arXiv:0904.0156 [math.ST]`.

[14] D. Casadei, *Reference analysis of the signal + background model in counting experiments II. Approximate reference prior*, JINST **9** (2014) no. 10, T10006, `arXiv:1407.5893 [physics.data-an]`.

[15] J. M. Bernardo, *Objective Bayesian point and region estimation in location-scale models*, Sort **31** (2007) 3–44.