Prepared for submission to JHEP

Improved Asymptotic Formulae for Statistical Interpretation Based on Likelihood Ratio Tests

Li-Gang Xia

School of Physics, Nanjing University
No. 22 Hankou Road, Nanjing, China

E-mail: ligang.xia@cern.ch

Abstract: The asymptotic formulae to describe the probability distribution of a test statistic in G. Cowan *et al.*'s paper [1] are deeply based on Wald's approximation [2]. Wald's approximation is valid if the sample size is big enough. It works well in most cases of searching for new physics. In this work, the asymptotic formulae are improved with considering the sub-leading contributions due to limited sample size and non-negligible signal-to-background ratio. The probability distribution of the test statistics from the new asymptotic formulae is closer to that from MC simulations. Let $\Delta\text{UL}(old/new)$ denote the difference between the upper limit from the old/new formulae and that from the toy MC method. Based on the test statistics $\tilde{q}_\mu$ and the examples presented in this work, we find $\Delta\text{UL}(new)$ is 57-81% of $\Delta\text{UL}(old)$. In addition, a conjecture about the standard deviation, proposed in G. Cowan *et al.*'s paper, is also clarified.

ArXiv ePrint: 2101.06944

Contents

1 Introduction

Searching for new physics is always the goal for most experimenters in particle physics, especially after the discovery of the Higgs boson [3, 4]. Once a measurement is done, it is important to report the results in a precise and well-accepted way. One often reports two things if no significant signal is observed. One is the probability that the observation is due to the fluctuation of known backgrounds. This is used to represent the statistical significance of a signal and to establish its discovery. The other is the parameter space about the new signal that the measurement can exclude for a given confidence level (C.L.). To interpret the results, we usually build a test statistic based on the likelihood ratio, which is the most powerful discriminant. To find the statistical significance and the exclusion limits, we need to know the probability distribution of the statistical test with many different signal strengths (or other parameter of interest). We can resort to toy Monte Carlo (MC) simulation. But it is usually computationally expensive.

Fortunately, G. Cowan *et al.* have found asymptotic formulae [1] to describe the distribution of the likelihood ratio tests if the sample size is big enough. Therefore, one can easily obtain the expected statistical significance and exclusion limits for a new signal based on the idea of "asimov" dataset [1]. The validity of the asymptotic formulae is due to a theorem by

Table 1. Summary of the test statistics based on the likelihood ratio

Test statistic	Purpose
t_0	to establish the discovery of a signal
t_μ	to set a confidence interval of a signal at a given level
q_μ	to set an upper limit of a signal at a given level
q_0	to establish the discovery of a positive signal
$\tilde{t}_\mu$	to set a confidence interval of a positive signal at a given level
$\tilde{q}_\mu$	to set an upper limit of a positive signal at a given level

Wald [2] and the condition is that the sample size is sufficiently big. The asymptotic formulae are believed to work if the number of event is over 5. Recently, the author has finished a study of the feasibility to search for leptoquarks in Pb-Pb ultra-peripheral collisions [5] and the background level in that case is very low (the expected number of background events is much less than 1). It is the direct motivation of the current work to explore the limits where the asymptotic formulae work. During the exploration, new asymptotic formulae are found from a different perspective and show better agreement with the toy MC simulation results than the old ones in Ref. [1].

In Sec. 2, we will have a brief review about the test statistic and the old asymptotic formulae. In Sec. 3, we will elaborate two improvements and present the new formulae. The two sets of asymptotic formulae are compared using two examples in Sec. 4. Some discussions about the standard deviation are presented in Sec. 5. Sec. 6 is a short summary.

2 Review of the test statistic and the asymptotic formulae

We will review the test statistics and the asymptotic formulae according to Ref. [1]. To test a hypothesis with the signal strength μ, we consider the likelihood ratio

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \,, \tag{2.1}$$

where $\boldsymbol{\theta}$ denotes a set of nuisance parameters; $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the optimal values to maximize the likelihood function; $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ are the optimal values with μ fixed and can be see functions of μ. Based on this ratio, six test statistics ($t_0, q_0, t_\mu, \tilde{t}_\mu, q_\mu$ and $\tilde{q}_\mu$) are defined for different purposes. They are summarized in Tab. 1.

For example, to set an upper limit and consider the constraint $\mu > 0$ (assuming that the signal contribution is positive to the observed number of events), the recommendation is $\tilde{q}_\mu$.

$$\tilde{q}_\mu = \begin{cases} 0 & \hat{\mu} > \mu \,, \\ -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \mu \geq \hat{\mu} \geq 0 \,, \\ -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \,. \end{cases} \tag{2.2}$$

To reject the background-only hypothesis (namely, $\mu = 0$), we use the test statistic q_0.

$$q_0 = \begin{cases} -2\ln\frac{\mathcal{L}(0,\hat{\hat{\boldsymbol{\theta}}}(0))}{\mathcal{L}(\hat{\mu},\hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,. \end{cases} \tag{2.3}$$

The asymptotic formulae in Ref. [1] to describe the probability distribution of these test statistics are based on Wald's theorem [2]. It says that the logarithmic likelihood ratio, seen as a random variable, satisfies the following relation

$$t_\mu \equiv -2\ln\lambda(\mu) = \frac{(\hat{\mu}-\mu)^2}{\sigma^2} + \mathcal{O}(\frac{1}{\sqrt{N}})\,, \tag{2.4}$$

where $\hat{\mu}$ abides by a Gaussian distribution with a mean μ_H and standard deviation σ; and N represents the data sample size. Here it is worth mentioning that N is basically the background sample size if we are searching for new physics signals. The standard deviation σ can be obtained from either the Fisher information matrix (second-order derivatives of the logarithmic likelihood function) [8, 9] (denoted by σ(d2L)) or from Wald's theorem (Eq. 3.7) based on an Asimov dataset (denoted by σ(Wald)). In the large sample limit, we can ignore the term $\mathcal{O}(\frac{1}{\sqrt{N}})$ in Eq. 3.7 (we call it "Wald's approximation" throughout this paper). The asymptotic relations between $\tilde{q}_\mu$ (q_0) and $\hat{\mu}$ are

$$\tilde{q}_\mu = \begin{cases} 0 & \hat{\mu} > \mu\,, \\ \frac{(\hat{\mu}-\mu)^2}{\sigma^2} & \mu \geq \hat{\mu} \geq 0\,, \\ \frac{\mu^2-2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0\,. \end{cases}, \tag{2.5}$$

and

$$q_0 = \begin{cases} \frac{\hat{\mu}^2}{\sigma^2} & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,. \end{cases}, \tag{2.6}$$

from which the probability distribution function (PDF) of $\tilde{q}_\mu$ (q_0) can be easily obtained via $\hat{\mu}$.

3 New asymptotic formulae

3.1 Asymptotic form of t_μ as a function of $\hat{\mu}$

Throughout this paper, we use μ', $\hat{\mu}$, μ and μ_H to denote the true value of the signal strength, the best-fit value from the maximum likelihood estimation, the test value we want to see if it is compatible with data, and the hypothesized value in an asimov dataset, respectively. First of all, we derive Wald's approximation following Wald's idea roughly in the case of no other nuisance parameters. Then we will explain the extension.

Although μ' is unique and will not change by any means, different μ's mean different data. So we can see $-2\ln\mathcal{L}$ as a function of μ'. Writing $-2\ln\mathcal{L}(\mu')$ as $f(\mu')$, we perform the Taylor expansion around $\hat{\mu}$ and μ.

$$f(\mu') = f(\hat{\mu}) + f^{(1)}(\hat{\mu})(\mu'-\hat{\mu}) + \frac{1}{2}f^{(2)}(\xi_{[\hat{\mu},\mu']})(\mu'-\hat{\mu})^2\,, \tag{3.1}$$

$$f(\mu') = f(\mu) + f^{(1)}(\mu)(\mu'-\mu) + \frac{1}{2}f^{(2)}(\xi_{[\mu,\mu']})(\mu'-\mu)^2\,, \tag{3.2}$$

$$f^{(1)}(\mu) = f^{(1)}(\hat{\mu}) + f^{(2)}(\xi_{[\hat{\mu},\mu]})(\mu-\hat{\mu})\,, \tag{3.3}$$

where $f^{(1)}(\xi)$ and $f^{(2)}(\xi)$ denote the first and second derivative at ξ respectively; $\xi[a,b]$ denotes a value between a and b used in the Taylor expansion remainders. Assuming $f^{(2)}(\xi)$ is always positive and denoting $\frac{1}{2}f^{(2)}(\xi)$ by $\frac{1}{\sigma^2(\xi)}$, we have

$$\begin{aligned} t_\mu = & \quad f(\mu) - f(\hat{\mu}) & (3.4)\\ = & \quad \frac{(\hat{\mu}-\mu')^2}{\sigma^2(\xi_{[\hat{\mu},\mu']})} - \frac{(\mu-\mu')^2}{\sigma^2(\xi_{[\mu,\mu']})} + 2\frac{(\mu-\mu')(\mu-\hat{\mu})}{\sigma^2(\xi_{[\hat{\mu},\mu]})} & (3.5)\\ = & \quad \frac{(\hat{\mu}-\mu')^2}{\sigma^2(\xi_{[\hat{\mu},\mu']})} - \frac{(\mu-\mu')^2}{\sigma^2(\xi_{[\mu,\mu']})} + 2\frac{(\mu-\mu')(\mu-\mu'+\mu'-\hat{\mu})}{\sigma^2(\xi_{[\hat{\mu},\mu]})} & (3.6)\\ = & \quad \frac{(\hat{\mu}-\mu')^2}{\sigma^2(\xi_{[\hat{\mu},\mu']})} + (2\frac{\sigma^2(\xi_{[\mu,\mu']})}{\sigma^2(\xi_{[\hat{\mu},\mu]})} - 1)\frac{(\mu-\mu')^2}{\sigma^2(\xi_{[\mu,\mu']})} - 2\frac{(\mu-\mu')(\hat{\mu}-\mu')}{\sigma^2(\xi_{[\hat{\mu},\mu]})} \,. & (3.7) \end{aligned}$$

In the limit of large number of observations, Wald's theorem shows that all three σs are the same (denoted by σ). We come to Wald's approximation

$$t_\mu = -2\ln\lambda(\mu) = \left(\frac{\hat{\mu}-\mu'}{\sigma} + \frac{\mu'-\mu}{\sigma}\right)^2 = \frac{(\hat{\mu}-\mu)^2}{\sigma^2} \,, \tag{3.8}$$

where $\hat{\mu}$ abides a Gaussian distribution with the mean μ' and the standard derivation σ. In practice, it is better to use $\sigma = \sigma(\text{Wald})$ as observed in Ref. [1].

The asymptotic form of t_μ as a function of $\hat{\mu}$ from Wald's theorem satisfies two conditions: 1) $t_\mu \geq 0$ for any $\hat{\mu}$; 2) t_μ reaches 0 at $\hat{\mu} = \mu$. Although they are fairly reasonable, we loosen the latter condition because $\hat{\mu}$ must be close to μ' from an unbiased estimation with a large sample and $\hat{\mu}$ would have a small probability to be around μ if $\mu \neq \mu'$. The latter condition is then loosened to be that t_μ reaches 0 at $\hat{\mu} = \mu$ only if $\mu = \mu'$. Requiring that the minimum value of t_μ is 0 in Eq. 3.7, the three σs should satisfies

$$\frac{1}{\sigma^2(\xi_{[\mu,\mu']})} = \frac{2\sigma^2(\xi_{[\hat{\mu},\mu]}) - \sigma^2(\xi_{[\hat{\mu},\mu']})}{\sigma^4(\xi_{[\hat{\mu},\mu]})} \,. \tag{3.9}$$

Using this equation, the proposed extension of Wald's approximation in this work is

$$t_\mu = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 \,, \tag{3.10}$$

where the true value μ' is replaced by μ_H which is the hypothesized signal strength in an asimov dataset because the real value is unknown; $\sigma_0 \equiv \sigma(\xi_{[\hat{\mu},\mu_H]})$ and $\sigma_1 \equiv \sigma^2(\xi_{[\hat{\mu},\mu]})/\sigma(\xi_{[\hat{\mu},\mu_H]})$. $\hat{\mu}$ abides by a Gaussian distribution with mean μ_H and standard deviation σ. We recommend use $\sigma = \sigma(\text{d2L})$, which is obtained from a likelihood fit for the new asymptotic formulae as we will see in Sec. 5. This choice is different from that in Ref. [1], but seems more natural.

From a different perspective, we present another motivation for this extension. For a binned dataset, the likelihood function without any nuisance parameter is

$$\mathcal{L}(\mu') = \Pi_{i=1}^{N_{\text{bins}}}\mathcal{P}(n_i|b_i+\mu' s_i) = \Pi_{i=1}^{N_{\text{bins}}}\frac{(b_i+\mu' s_i)^{n_i}}{n_i!}e^{-(b_i+\mu' s_i)} \,, \tag{3.11}$$

where $N_{\rm bins}$ is the number of bins; b_i, s_i and n_i are the number of background events, signal events and data events in the i-th bin, respectively. The logarithmic likelihood function is then

$$\ln \mathcal{L}(\mu') = \sum_{i=1}^{N_{\rm bins}} n_i \ln(b_i + \mu' s_i) - (b_i + \mu' s_i) = -(b + \mu' s) + \sum_{i=1}^{N_{\rm bins}} n_i \ln(b_i + \mu' s_i) \,, \quad (3.12)$$

where $b \equiv \sum_{i=1}^{N_{\rm bins}} b_i$ and $s \equiv \sum_{i=1}^{N_{\rm bins}} s_i$ are the total number of background and signal events; and the constant terms are omitted. The likelihood ratio is

$$t_\mu = -2 \ln \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})} = 2(\mu - \hat{\mu}) s - 2 \sum_{i=1}^{N_{\rm bins}} n_i \ln \frac{b_i + \mu s_i}{b_i + \hat{\mu} s_i} \,. \quad (3.13)$$

Under the hypothesis with a signal strength μ_H, n_i can be approximated well by $b_i + \hat{\mu} s_i$ if the sample size is large enough and $\hat{\mu}$ is close to μ_H. We can expand $\lambda(\mu)$ around $\hat{\mu} = \mu_H$.

$$\begin{aligned} t_\mu \approx\ & 2(\mu - \hat{\mu}) s - 2 \sum_{i=1}^{N_{\rm bins}} (b_i + \hat{\mu} s_i) \ln \frac{b_i + \mu s_i}{b_i + \hat{\mu} s_i} & (3.14) \\ \approx\ & C_0 + 2C_1(\hat{\mu} - \mu_H) + C_2(\hat{\mu} - \mu_H)^2 \,, & (3.15) \end{aligned}$$

where C_0, C_1 and C_2 are functions of μ_H or (and) μ and defined below.

$$\begin{aligned} C_0 &= 2(\mu - \mu_H) s - 2 \sum_{i=1}^{N_{\rm bins}} (b_i + \mu_H s_i) \ln \frac{b_i + \mu s_i}{b_i + \mu_H s_i} \,, \\ C_1 &= \sum_{i=1}^{N_{\rm bins}} s_i \ln \frac{b_i + \mu_H s_i}{b_i + \mu s_i} \,, \\ C_2 &= \sum_{i=1}^{N_{\rm bins}} \frac{s_i^2}{b_i + \mu_H s_i} \,, \end{aligned} \quad (3.16)$$

The parameterization of $\tilde{q}_\mu$ as a function of $\hat{\mu}$ (Eq. 3.15) can be seen as an extension of Wald's approximation (Eq. 3.7). It turns to Wald's approximation by neglecting the terms of the order of $(s_i/b_i)^2$ or higher, namely,

$$C_0 = \frac{(\mu - \mu_H)^2}{\sigma_*^2} \,, \quad C_1 = \frac{\mu_H - \mu}{\sigma_*^2} \,, \quad C_2 = \frac{1}{\sigma_*^2} \,. \quad (3.17)$$

where $\sigma_* \equiv 1/\sqrt{C_2}$ is just σ(d2L) for an asimov dataset with signal strength μ_H in this simplified model.

Despite its more generality, the new parameterization does not guarantees the non-negativeness ($t_\mu \geq 0$ by definition). We impose a constraint to solve the drawback. The smallest value of t_μ is $C_0 - C_1^2/C_2$. We require it to be 0 and replace C_2 by C_1^2/C_0 in Eq. 3.15 because the C_2 term should be affected by high order contribution more than the C_0 and C_1 terms. Hence we obtain

$$t_\mu = C_0 + 2C_1(\hat{\mu} - \mu_H) + \frac{C_1^2}{C_0}(\hat{\mu} - \mu_H)^2 \,, \quad (3.18)$$

which is equivalent to the extension of Wald's approximation in Eq. 3.10 with the following relations

$$\sigma_0 = \sqrt{\frac{C_0}{C_1^2}}\,, \quad \sigma_1 = \frac{|\mu_H - \mu|}{\sqrt{C_0}}\,. \tag{3.19}$$

According to the analysis above, Wald's approximation in Eq. 3.7 seems to hold well with two conditions: 1) the number of total events is large so that n_i can be approximated well by $b_i + \hat{\mu} s_i$; 2) the signal-to-background ratio $\mu s_i/b_i$ is small so that the difference between σ_0 and σ_1 can be neglected (via C_0 and C_1 in Eq. 3.16 and 3.19). This argument is probably not solid mathematically, but it seems true in real examples as we will see in Sec. 5. In the original paper [2] of Wald, it seems that only the first condition is enough.

As C_0 and C_1 can be calculated explicitly, Eq. 3.16 and 3.19 provide a way to calculate σ_0 and σ_1. However, we recommend a different way to obtain σ_0 and σ_1 using asimov datasets. Let $t_\mu^A(\mu_A)$ denote the value of t_μ obtained from an asimov dataset with the signal strength μ_A. According to Eq. 3.10, σ_0 and σ_1 are then

$$\sigma_0 = \lim_{\epsilon\to 0} \frac{|\epsilon|}{|\sqrt{t_\mu^A(\mu_H+\epsilon)} - \sqrt{t_\mu^A(\mu_H)}|}\,, \quad \sigma_1 = \lim_{\epsilon\to 0} \frac{|\mu_H + \epsilon - \mu|}{\sqrt{t_\mu^A(\mu_H+\epsilon)}}\,, \tag{3.20}$$

which is similar to the way to obtain $\sigma(\text{Wald})$ from Wald's approximation in Ref. [1]

$$\sigma(\text{Wald}) = \lim_{\epsilon\to 0} \frac{|\mu_H + \epsilon - \mu|}{\sqrt{t_\mu^A(\mu_H+\epsilon)}}\,, \tag{3.21}$$

where the limitation is used so that it is applicable if $\mu_H = \mu$. In reality, a choice like $\epsilon = 0.01\sigma(\text{d2L})$ is reasonable. Comparing Eq. 3.20 and Eq. 3.21, we have $\sigma_1 = \sigma(\text{Wald})$ in all cases and $\sigma_0 = \sigma(\text{Wald})$ only if $\mu_H = \mu$.

3.2 Asymptotic form for $\hat{\mu} < 0$

For the test statistics, $\tilde{t}_\mu$ and $\tilde{q}_\mu$, we have to pay attention to the case of $\hat{\mu} < 0$

$$-2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} = -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})} - (-2\ln\frac{\mathcal{L}(0,\hat{\hat{\theta}}(0))}{\mathcal{L}(\hat{\mu},\hat{\theta})}) = t_\mu - t_0\,. \tag{3.22}$$

According to the proposed extension in Eq. 3.10, it is

$$t_\mu - t_0 = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1})^2 - (\frac{\hat{\mu}-\mu_H}{\sigma_0'} + \frac{\mu_H}{\sigma_1'})^2\,. \tag{3.23}$$

For better illustration, $\tilde{q}_\mu$ as a function of $\hat{\mu}$ is plotted in Fig. 1 as well as the predicted form from Wald's approximation. The blue curve in Fig. 1 clearly shows two features. This produces two issues we have to fix reasonably.

Firstly, the asymptotic form above indicates that $t_\mu - t_0$ may be negative for very negative $\hat{\mu}$ if $\sigma_0 > \sigma_0'$. This is unlikely for $\hat{\mu} < 0$ if the likelihood function is approximately Gaussian-like around $\hat{\mu}$. On the other hand, a signal strength μ would not be too negative, otherwise the expected number of events $b + \mu s$ is negative and meaningless. The latter is

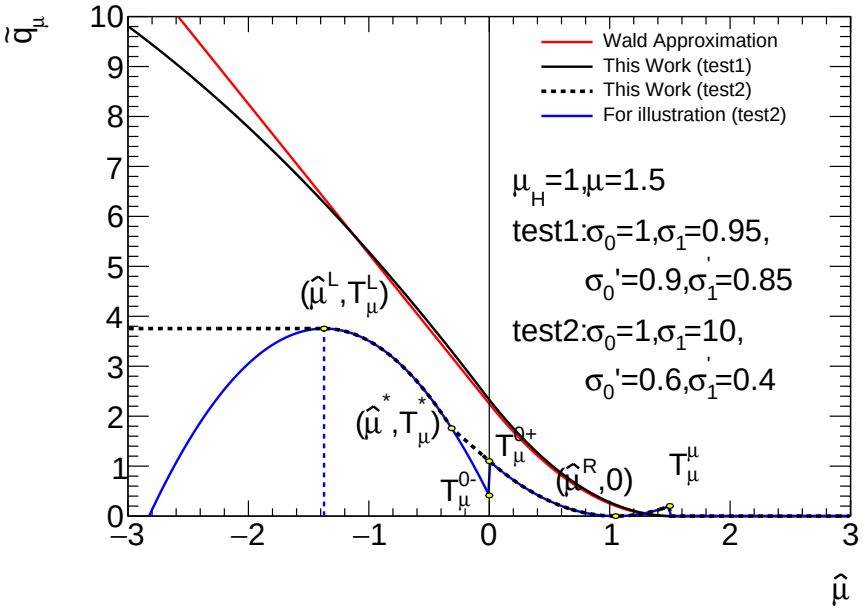

Figure 1. The asymptotic form of $\tilde{q}_\mu$ as a function of $\hat{\mu}$. The red curve represents the Wald's approximation. The black and blue curves represent the asymptotic form proposed in this work with different values for the σs. If all σ are equal, Wald's approximation is restored.

self-evident in the distributions of $\hat{\mu}$ as we will see in Sec. 4. In appendix C, we provide a finer description of $\hat{\mu}$ by including a sub-leading correction. As $t_\mu - t_0$ is a second-order polynomial about $\hat{\mu}$, the extreme value is reached at $\frac{\partial(t_\mu - t_0)}{\partial \hat{\mu}} = 0$, namely,

$$\hat{\mu}^L = \mu_H + \frac{\sigma_0\sigma_0'}{\sigma_1\sigma_1'}\frac{\mu_H\sigma_0\sigma_1 - (\mu_H - \mu)\sigma_0'\sigma_1'}{\sigma_0'^2 - \sigma_0^2} \,. \tag{3.24}$$

To make the asymptotic form consistent with the physics fact, we impose the following constraint for $\hat{\mu} < \hat{\mu}^L$

$$t_\mu - t_0 = T_\mu^L \equiv (t_\mu - t_0)|_{\hat{\mu}=\hat{\mu}^L} \,, \quad \text{for } \sigma_0 > \sigma_0', \hat{\mu} < \hat{\mu}^L \,, \tag{3.25}$$

This means that $t_\mu - t_0$ is fixed at $(t_\mu - t_0)|_{\hat{\mu}=\hat{\mu}^L}$ for $\hat{\mu} < \hat{\mu}^L$ in the case of $\sigma_0 > \sigma_0'$.

The second issue is that $\tilde{q}_\mu$ as a function of $\hat{\mu}$ is not continuous around $\hat{\mu} = 0$ generally. This is also not consistent with the definition of $\tilde{q}_\mu$. Noticing that the asymptotic function based on Wald's approximation is continuous and smooth around $\hat{\mu} = 0$. Especially, $\tilde{q}_\mu$ is a linear function of $\hat{\mu}$ for $\hat{\mu} < 0$. We can apply the following remedy.

$$(t_\mu - t_0)|_{\hat{\mu}<0} = \begin{cases} (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\hat{\mu}}{\sigma_0})^2 \,, & \hat{\mu}^* \le \hat{\mu} < 0 \\ (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\hat{\mu}-\mu_H}{\sigma_0'} + \frac{\mu_H}{\sigma_1'})^2 \,, & \hat{\mu} < \hat{\mu}^* \end{cases} \tag{3.26}$$

Here $t_\mu - t_0$ is split into two pieces for $\hat{\mu} < 0$. The piece with $\hat{\mu}^* \le \hat{\mu} < 0$ is a straight line. It intersects the other piece at $\hat{\mu}^*$. $\hat{\mu}^*$ is calculated to be

$$\hat{\mu}^* = -\frac{\frac{-1}{\sigma_0'} + \frac{1}{\sigma_1'}}{\frac{1}{\sigma_0'} + \frac{1}{\sigma_0}}\mu_H \,, \tag{3.27}$$

and the corresponding value of $t_\mu - t_0$ is (denoted by T^*_μ)

$$T^*_\mu = (\frac{2\hat{\mu}^* - \mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1})(\frac{-\mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1}) \,. \tag{3.28}$$

In this way, $\tilde{q}_\mu$ is continuous everywhere and the monotonic relation between $\tilde{q}_\mu$ and $\hat{\mu}$ is preserved. In Fig. 1, the blue curve clearly indicates the two features; the black dashed curve is the final asymptotic form proposed in this work; and the black solid curve shows a great similarity to the prediction from Wald's approximation if the difference among σs is small.:w

3.3 Asymptotic formulae for test statistics $\tilde{q}_\mu$

In this section, we present the cumulative probability distribution (CDF) of the test statistic $\tilde{q}_\mu$. For the other test statistics, t_0, q_0, t_μ, $\tilde{t}_\mu$ and q_μ, their CDFs are presented in Appendix A.

$$\tilde{q}_\mu = \begin{cases} 0 \,, & \hat{\mu} > \mu \,, \\ (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 \,, & 0 \le \hat{\mu} \le \mu \,, \\ (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\hat{\mu}-\mu_H}{\sigma_0})^2 \,, & \hat{\mu}^* \le \hat{\mu} < 0 \,, \\ (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\hat{\mu}-\mu_H}{\sigma'_0} + \frac{\mu_H}{\sigma'_1})^2 \,, & \hat{\mu} \le \hat{\mu}^* \text{ and } \hat{\mu} \ge \hat{\mu}^L \text{ if } \sigma_0 > \sigma'_0 \,, \\ T^L_\mu \,, & \hat{\mu} < \hat{\mu}^L \text{ if } \sigma_0 > \sigma'_0 \,. \end{cases} \tag{3.29}$$

We introduce the following quantities to better describe the features of this asymptotic form. They are shown in Fig. 1 for better understanding.

1. $\hat{\mu}^L$: it is defined in Eq. 3.24 and only introduced if $\sigma_0 > \sigma'_0$. It indicates that $\hat{\mu}$ cannot be too negative as explained in last section. $\tilde{q}^L_\mu$ is the value of $\tilde{q}_\mu$ at $\hat{\mu} = \hat{\mu}^L$, namely,

$$T^L_\mu \equiv (\frac{\hat{\mu}^L - \mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1})^2 - (\frac{\hat{\mu}^L - \mu_H}{\sigma'_0} + \frac{\mu_H}{\sigma'_1})^2 \,, \quad \text{for } \sigma_0 > \sigma'_0 \,. \tag{3.30}$$

2. $T^{0\pm}_\mu$: the value of $\tilde{q}_\mu$ at $\hat{\mu} = 0^\pm$, namely,

$$T^{0-}_\mu \equiv (\frac{-\mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1})^2 - (\frac{-\mu_H}{\sigma'_0} + \frac{\mu_H}{\sigma'_1})^2 \,, \tag{3.31}$$

$$T^{0+}_\mu \equiv (\frac{-\mu_H}{\sigma_0} + \frac{\mu_H - \mu}{\sigma_1})^2 \,. \tag{3.32}$$

3. T^*_μ: the value of $\tilde{q}_\mu$ at $\hat{\mu} = \hat{\mu}^*$. $\hat{\mu}^*$ and T^*_μ are defined in Eq. 3.27 and Eq. 3.28.

4. $\hat{\mu}^R$: the value of $\hat{\mu}$ such that $\tilde{q}_\mu = 0$, namely,

$$\hat{\mu}^R \equiv \frac{\sigma_0}{\sigma_1}\mu + (1 - \frac{\sigma_0}{\sigma_1})\mu_H \,. \tag{3.33}$$

5. T^μ_μ: the value of $\tilde{q}_\mu$ at $\hat{\mu} = \mu$, namely,

$$T^\mu_\mu \equiv (\frac{1}{\sigma_0} - \frac{1}{\sigma_1})^2(\mu - \mu_H)^2 \,. \tag{3.34}$$

The CDF of $\tilde{q}_\mu$ is

$$F(\tilde{q}_\mu) = \begin{cases} \Phi(-\frac{\mu-\mu_H}{\sigma}) + \theta(\mu-\hat{\mu}^R)[\Phi(\frac{\sigma_0}{\sigma}(\sqrt{\tilde{q}_\mu}+\frac{\mu_H-\mu}{\sigma_1})) \\ \quad -\Phi(\frac{\sigma_0}{\sigma}(-\sqrt{\tilde{q}_\mu}+\frac{\mu_H-\mu}{\sigma_1}))] \,, & \tilde{q}_\mu \le T_\mu^\mu \\ \Phi(\frac{\sigma_0}{\sigma}(\sqrt{\tilde{q}_\mu}+\frac{\mu_H-\mu}{\sigma_1})) \,, & T_\mu^\mu < \tilde{q}_\mu \le T_\mu^{0+} \\ \Phi(\frac{\mu_H}{\sigma} - \frac{\sigma_0(\tilde{q}_\mu - T_\mu^{0+})}{2\sigma(\frac{-\mu_H}{\sigma_0}+\frac{\mu_H-\mu}{\sigma_1})}) \,, & T_\mu^{0+} < \tilde{q}_\mu \le T_\mu^* \\ \Phi(-\frac{\hat{\mu}^-(\tilde{q}_\mu)-\mu_H}{\sigma}) \,, & T_\mu^* < \tilde{q}_\mu \text{ and } \tilde{q}_\mu \le T_\mu^L \text{ if } \sigma_0 > \sigma_0' \\ 1 \,, & \tilde{q}_\mu > T_\mu^L \text{ if } \sigma_0 > \sigma_0' \end{cases} \tag{3.35}$$

where $\theta(x)$ is the step function; and $\hat{\mu}^-(x)$ is defined as

$$\hat{\mu}^-(x) \equiv \mu_H - \frac{(\frac{\mu_H-\mu}{\sigma_0\sigma_1} - \frac{\mu_H}{\sigma_0'\sigma_1'}) + \sqrt{(\frac{\mu_H-\mu}{\sigma_0\sigma_1} - \frac{\mu_H}{\sigma_0'\sigma_1'})^2 - (\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2})[(\frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\mu_H}{\sigma_1'})^2 - x]}}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2}} \,. \tag{3.36}$$

For the special case of $\mu_H = \mu$, we have $\sigma_0 = \sigma$. The CDF is

$$F(\tilde{q}_\mu|\mu_H=\mu) = \begin{cases} \Phi(\sqrt{\tilde{q}_\mu}) \,, & \tilde{q}_\mu \le T_\mu^{0+} \\ \Phi(\frac{\sigma}{2\mu_H}\tilde{q}_\mu + \frac{\mu_H}{2\sigma}) \,, & T_\mu^{0+} < \tilde{q}_\mu \le T_\mu^* \\ \Phi(-\frac{\hat{\mu}^-(\tilde{q}_\mu)-\mu_H}{\sigma}) \,, & T_\mu^* < \tilde{q}_\mu \text{ and } \tilde{q}_\mu \le T_\mu^L \text{ if } \sigma_0 > \sigma_0' \\ 1 \,, & \tilde{q}_\mu > T_\mu^L \text{ if } \sigma_0 > \sigma_0' \end{cases} \tag{3.37}$$

where $\hat{\mu}^-(x)$ is simplified to be

$$\hat{\mu}^-(x) = \mu_H - \frac{-\frac{\mu_H}{\sigma_0'\sigma_1'} + \sqrt{(\frac{\mu_H}{\sigma_0'\sigma_1'})^2 - (\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2})[-(\frac{\mu_H}{\sigma_1'})^2 - x]}}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2}} \,. \tag{3.38}$$

For the special case of $\mu_H = 0 < \mu$, we have $\sigma_0' = \sigma$. The CDF is

$$F(\tilde{q}_\mu|\mu_H=0) = \begin{cases} \Phi(-\frac{\mu}{\sigma}) + \theta(\mu-\hat{\mu}^R)[\Phi(\frac{\sigma_0}{\sigma}(\sqrt{\tilde{q}_\mu}+\frac{-\mu}{\sigma_1})) \\ \quad -\Phi(\frac{\sigma_0}{\sigma}(-\sqrt{\tilde{q}_\mu}+\frac{-\mu}{\sigma_1}))] \,, & \tilde{q}_\mu \le T_\mu^\mu \\ \Phi(\frac{\sigma_0}{\sigma}(\sqrt{\tilde{q}_\mu}+\frac{-\mu}{\sigma_1})) \,, & T_\mu^\mu < \tilde{q}_\mu \le \frac{\mu^2}{\sigma_1^2} \\ \Phi(-\frac{\hat{\mu}^-(\tilde{q}_\mu)}{\sigma}) \,, & \frac{\mu^2}{\sigma_1^2} < \tilde{q}_\mu \text{ and } \tilde{q}_\mu \le T_\mu^L \text{ if } \sigma_0 > \sigma_0' \\ 1 \,, & \tilde{q}_\mu > \tilde{q}_\mu^L \text{ if } \sigma_0 > \sigma_0' \end{cases} \tag{3.39}$$

where $\hat{\mu}^-(x)$ is simplified to be

$$\hat{\mu}^-(x) = -\frac{\frac{-\mu}{\sigma_0\sigma_1} + \sqrt{(\frac{\mu}{\sigma_0\sigma_1})^2 - (\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2})[(\frac{\mu}{\sigma_1})^2 - x]}}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0'^2}} \,. \tag{3.40}$$

4 Performance comparison

In this section, we use three examples with different sample sizes to investigate the performance of the new formulae for $\tilde{q}_\mu$. All test statistics are compared in Appendix B. These

Table 2. Summary of the yields expected in the mass region $123 < m(\gamma\gamma) < 127$ GeV in three examples.

Yield	signal	background
Ex. 1	0.91	0.64
Ex. 2	0.91	2.79
Ex. 3	0.91	9.40

Table 3. Summary of systematic uncertainties in the examples.

Uncertainty	Luminosity	Higgs mass	Spurious signal
Ex. 1	$\pm 2\%$	± 0.2 GeV	± 0.5
Ex. 2	$\pm 2\%$	± 0.2 GeV	± 1.0
Ex. 3	$\pm 2\%$	± 0.2 GeV	± 3.0

examples originate from searching for Higgs boson using the $\gamma\gamma$ final state. They are denoted by Ex. 1, Ex. 2 and Ex. 3 with increasing sample size, respectively. Table 2 summarizes the expected signal and background yields in the mass region $123 < m(\gamma\gamma) < 127$ GeV. The signal shape is simulated by a Gaussian distribution while the background shape is simulated by an exponential distribution. They are shown in Fig. 2.

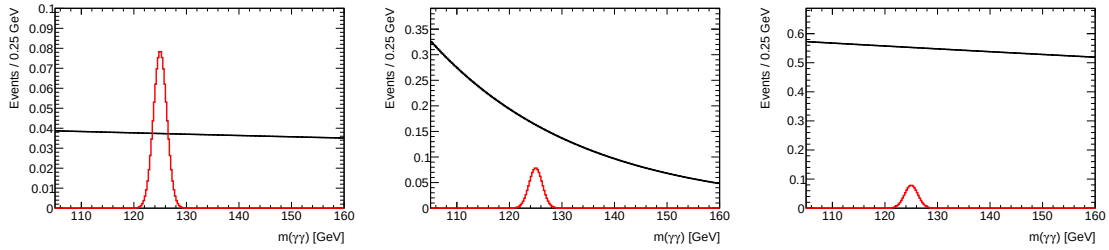

Figure 2. The distribution of $m(\gamma\gamma)$ in the three examples. From left to right, it is Ex. 1, Ex. 2 and Ex. 3.

We further consider three systematic uncertainties. They are due to the luminosity measurement, our knowledge on Higgs mass and the spurious signal, the last of which directly affects the expected signal yield and is usually dominant in real analyses. The uncertainty sizes are summarized in Table 3.

The CLs [6, 7] curves obtained from the asymptotic formulae in Ref. [1], this work and the toy MC method are shown in Fig. 3 without including any systematic uncertainty and in Fig. 4 with all three systematic uncertainties considered. The upper limits are summarized in Tab. 4. We can see that: 1) the predictions from the asymptotic formulae are better with larger sample size; 2) the new asymptotic formulae predict upper limits closer to those from toy MC method in all cases; 3) the distributions of $\tilde{q}_\mu$ predicted from the new formulae are also closer to those from the toy MC method in most of the cases. Some cases are shown in Fig. 5 for Ex. 1, Fig. 6 for Ex. 2 and Fig. 7 for Ex. 3. In these plots, the distribution of $\hat{\mu}$ and a finer description of its PDF are also shown. The PDF of $\hat{\mu}$ is elaborated in

Table 4. Summary of the upper limits at 95 % confidence level from the asymptotic formulae and the toy MC method. The percentages in the brackets indicate the difference with respective to the toy MC result.

Sys. Unc.?		Toy MC	Ref. [1]	This work	ΔUL(new)/ΔUL(old)
No	Ex. 1	3.39	3.16 (-6.9%)	3.25 (-4.1%)	0.59
	Ex. 2	5.16	4.95 (-3.9%)	5.00 (-2.9%)	0.74
	Ex. 3	8.13	7.91(-2.7%)	7.96 (-2.2%)	0.81
Yes	Ex. 1	3.53	3.26 (-7.7%)	3.36 (-4.8%)	0.62
	Ex. 2	5.42	5.23 (-3.5%)	5.32 (-2.0%)	0.57
	Ex. 3	9.86	9.64 (-2.2%)	9.72 (-1.4%)	0.64

Appendix C.

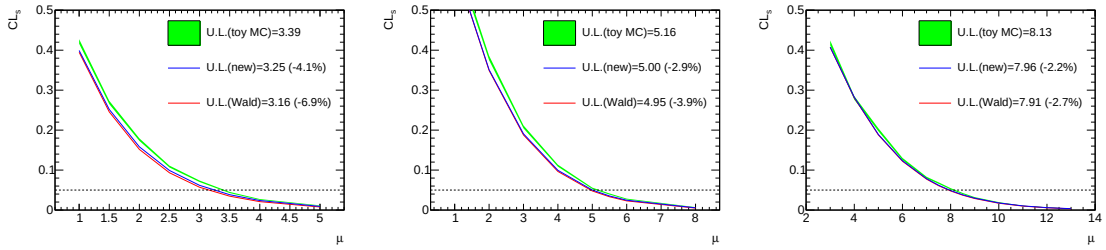

Figure 3. The CLs curves from the asymptotic formulae in Ref. [1](red), the new asymptotic formulae in this work (blue), and the toy MC method (green band) in the three examples. From left to right, it is Ex. 1, Ex. 2 and Ex. 3. No systematic uncertainties are considered. The upper limits at 95 % confidence level of the signal strength are shown in the legend. The percentages in the brackets indicate the difference with respective to the toy MC results.

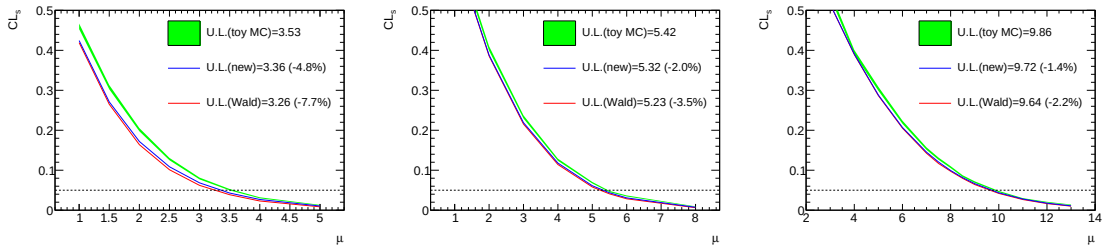

Figure 4. The CLs curves from the asymptotic formulae in Ref. [1](red), the new asymptotic formulae in this work (blue), and the toy MC method (green band) in the three examples. From left to right, it is Ex. 1, Ex. 2 and Ex. 3. All three systematic uncertainties are considered. The upper limits at 95 % confidence level of the signal strength are shown in the legend. The percentages in the brackets indicate the difference with respective to the toy MC results.

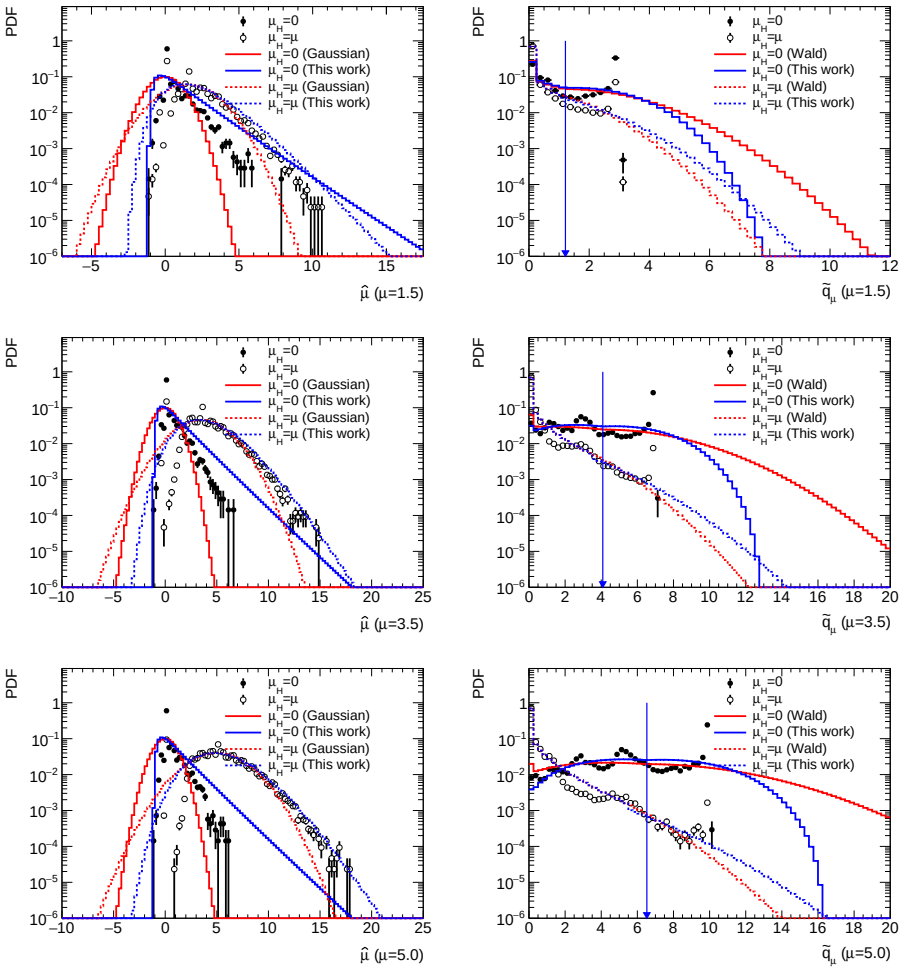

Figure 5. The probability distribution of $\hat{\mu}$ (L) and $\tilde{q}_\mu$ (R) in Ex 1. From top to bottom, the test signal strength is 1.5, 3.5 and 5.0. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed $\tilde{q}_\mu$. This is the case with all systematic uncertainties considered.

5 A clarification on the standard deviation of $\hat{\mu}$

We have known two methods to obtain the standard deviation of $\hat{\mu}$ in the PDF of the test statistics. The standard way is to use the Fisher information matrix obtained from the second derivatives of the logarithmic likelihood function. The alternative way is to use Wald's approximation based on asimov datasets. In Ref. [1], it has seen that both methods give similar results and the latter is the recommendation as it gives a distribution closer to the true sampling distribution in several cases. As stated in Sec. 3, we recommend to use σ(d2L) as the standard deviation in $\hat{\mu}$'s distribution. If we do not use the recommendations, Fig. 8 shows an example where we use σ(d2L) in the old formulae in the left plot while σ(Wald) in the new formulae in the right plot. The disagreement for $\mu_H \neq \mu$ compared to

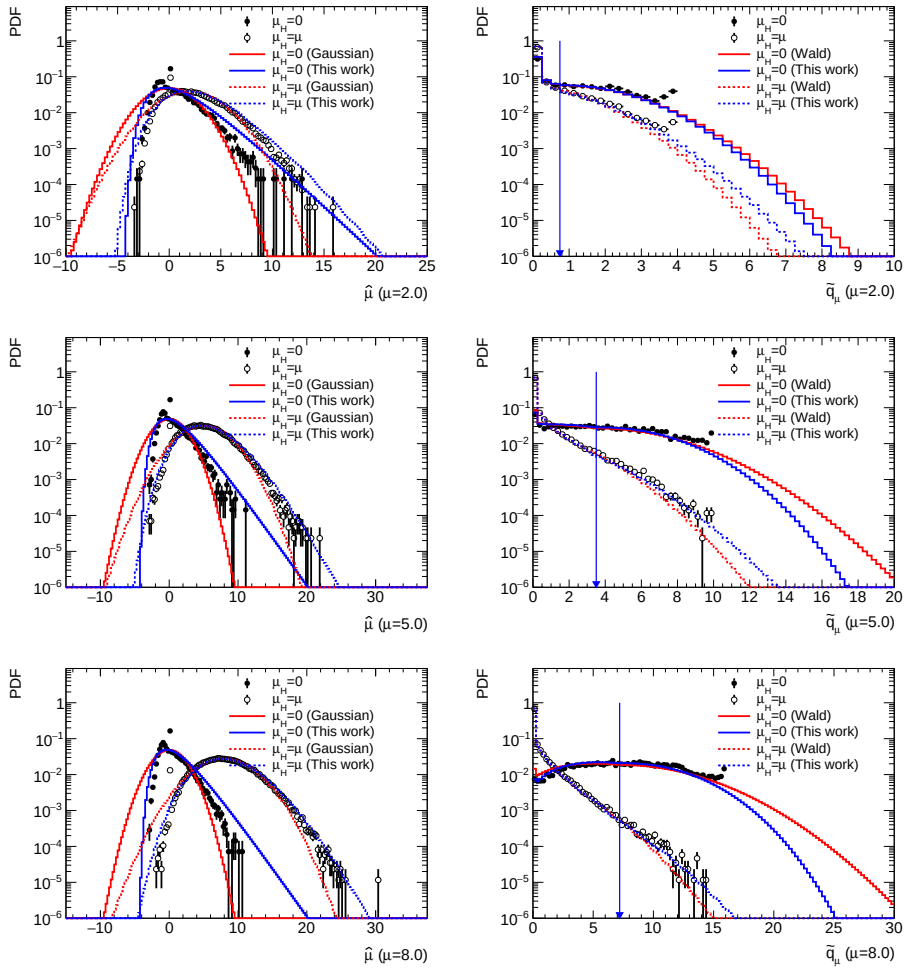

Figure 6. The probability distribution of $\hat{\mu}$ (L) and $\tilde{q}_\mu$ (R) in Ex 2. From top to bottom, the test signal strength is 2, 5 and 8. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed $\tilde{q}_\mu$. This is the case with all systematic uncertainties considered.

the toy MC result is significant.

The two methods are connected in Eq. 3.1 and Eq. 3.7. If the test value of signal strength happens to be the true value, $\mu = \mu'$, an unbiased estimation for a large sample gives $\hat{\mu} = \mu'$ and hence we have

$$\sigma(\text{d2L}) \equiv \sigma(\xi_{[\hat{\mu},\mu']}) = \sigma(\xi_{[\hat{\mu},\mu]}) = \sigma(\xi_{[\hat{\mu},\mu]}) \,. \tag{5.1}$$

The alternative method based on Wald's theorem leads to

$$\sigma(\text{Wald}) = \sigma(\text{d2L}) \,. \tag{5.2}$$

If $\mu \neq \mu'$, the equality above is only good if the sample size is large enough according to Wald's theorem. Therefore, we would expect that their difference will decrease if the sample

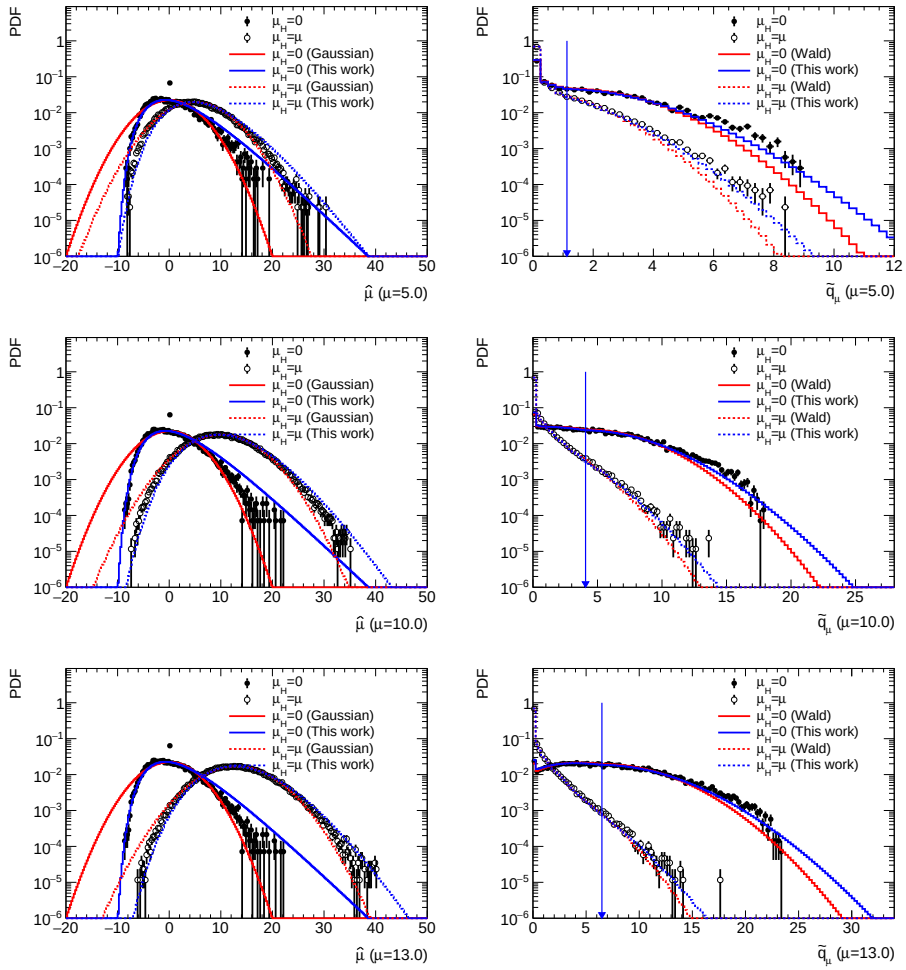

Figure 7. The probability distribution of $\hat{\mu}$ (L) and $\tilde{q}_\mu$ (R) in Ex 3. From top to bottom, the test signal strength is 5, 10 and 13. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed $\tilde{q}_\mu$. This is the case with all systematic uncertainties considered.

size increases. Table 5 summarizes the σs in some cases. Here we define δ_σ as the relative difference, $\delta_\sigma \equiv |\sigma(\text{Wald}) - \sigma(\text{d2L})|/\sigma(\text{d2L})$. These examples confirm the conclusion above. Furthermore, comparing the new CDF of $\tilde{q}_\mu$ in Eq. 3.35 to that in Ref. [1] and keeping in mind that $\sigma_0 \approx \sigma(\text{d2L})$ and $\sigma_1 \equiv \sigma(\text{Wald})$, we can understand why it is recommended to use $\sigma(\text{Wald})$ as the standard deviation of $\hat{\mu}$ in Ref. [1]. The authors in Ref. [1] conjecture that $\sigma(\text{Wald})$ absorbs some subleading effects compared to $\sigma(\text{d2L})$. So our proposed extension of Wald's approximation explains this conjecture.

On the other hand, we probably only care about an upper limit at 95 % C.L. or a similar level in reality. But if we are too conservative and want to determine an upper limit at 99.99 % C.L., we have to find μ corresponding to CLs equal to 1×10^{-4}. Taking $\mu = 12$ Ex. 2 as example (see Fig. 9), the toy MC method, new formulae and old formulae

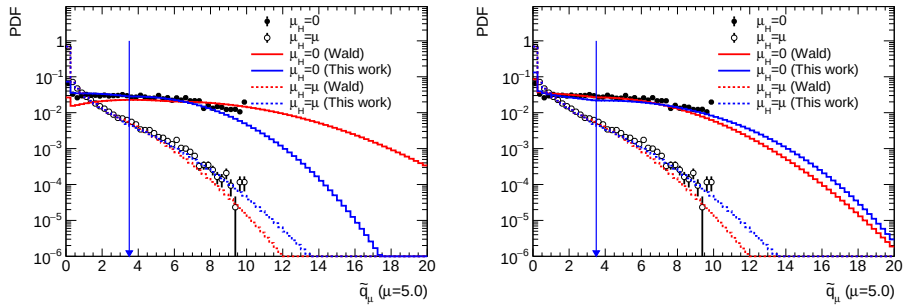

Figure 8. The probability distribution of $\tilde{q}_\mu$ in Ex 2 for $\mu_H = 5$. We use σ(d2L) in the old formulae in the left plot while σ(Wald) in the new formulae in the right plot. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed $\tilde{q}_\mu$. This is the case with all systematic uncertainties considered.

Table 5. Summary of different σs in the cases of including systematic uncertainties.

Ex.	Case	σ(d2L)	σ(Wald)	δ_σ	σ_0	σ_1	σ'_0	σ'_1
1	$\mu = 3.5, \mu_H = 0$	0.995584	1.73129	73.8%	1.20817	σ(Wald)	σ(d2L)	-
	$\mu = 3.5 = \mu_H$	2.17809	2.14136	1.7%	σ(d2L)	-	2.10577	1.58559
2	$\mu = 5, \mu_H = 0$	2.05339	2.67143	30.1%	2.17675	σ(Wald)	σ(d2L)	-
	$\mu = 5 = \mu_H$	3.1291	3.13158	0.08%	σ(d2L)	-	3.1086	2.60244
3	$\mu = 10, \mu_H = 0$	4.46857	4.95359	10.9%	4.26207	σ(Wald)	σ(d2L)	-
	$\mu = 10 = \mu_H$	5.60164	5.60216	0.01%	σ(d2L)	-	5.81383	5.17148

give CLs= $(4.4 \pm 0.5) \times 10^{-4}$, 3.3×10^{-4} and 2.4×10^{-4}, respectively. In this case, the big signal-to-background ratio, $\mu s/b >> 1$, makes the prediction from Wald's approximation break down (the upper limit difference compared to the toy MC result is about 45 %). In Eq. 3.16 and Eq. 3.17, we have seen that $C_{0,1}$ (or $\sigma^{(\prime)}_{0,1}$ via Eq. 3.19) reduce to the simple form in Wald's approximation if $(s/b)^2$ or higher-order terms are omitted. Therefore, based on the analysis above, the difference among σ_0, σ_1 and σ(d2L) in the new formulae (or the difference between σ(Wald) and σ(d2L) in the old formulae) can be attributed to limited sample size and non-negligible signal-to-background ratio. But these effects are partially considered in the new formulae. This is why the new formulae work better than the old ones.

6 Summary

In summary, we have provided a set of new asymptotic formulae to describe the probability distributions of the likelihood-ratio test statistics from a different perspective. They reduce to the old formulae [1] in the limit of large sample. A few examples are presented with different sample sizes. The new formulae are found to agree better with the toy MC simulations. Let ΔUL(old/new) denote the difference between the upper limit from the new/old

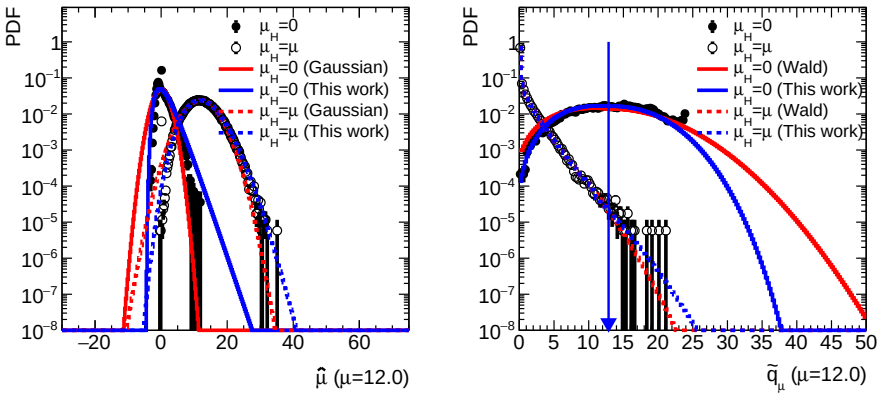

Figure 9. The probability distribution of $\hat{\mu}$ (L) and $\tilde{q}_\mu$ (R) in Ex 2 with $\mu = 12$ and all systematic uncertainties considered. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed $\tilde{q}_\mu$.

formulae and that from the toy MC method. We find $\Delta\text{UL}(new)$ is 57-81% of $\Delta\text{UL}(old)$ based on the examples. Besides, they have been used to explain a conjecture proposed in Ref. [1].

Acknowledgments

I would like to thank Fang Dai for her encouragement and partial financial support. This work is supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 12105140).

A Asymptotic formulae for the CDF of t_μ, $\tilde{t}_\mu$, q_μ, t_0 and q_0

In this appendix, we present the asymptotic formulae for the other 5 test statistics, t_μ, $\tilde{t}_\mu$, q_μ, t_0 and q_0. According to the proposed asymptotic form of t_μ in Eq. 3.10, the CDF of t_μ is

$$F(t_\mu) = \Phi(\frac{\sigma_0}{\sigma}(\frac{\mu_H - \mu}{\sigma_1} + \sqrt{t_\mu})) - \Phi(\frac{\sigma_0}{\sigma}(\frac{\mu_H - \mu}{\sigma_1} - \sqrt{t_\mu}))\,. \tag{A.1}$$

For the special case of $\mu_H = \mu$, the CDF is $F(t_\mu|\mu_H = \mu) = 2\Phi(\sqrt{t_\mu}) - 1$, which is the same as that in Ref. [1].

For $\tilde{t}_\mu$, the asymptotic form is

$$\tilde{t}_\mu = \begin{cases} t_\mu = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2\,, & \hat{\mu} \geq 0 \\ t_\mu - t_0 = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2 - (\frac{\hat{\mu}-\mu_H}{\sigma_0'} + \frac{\mu_H}{\sigma_1'})^2\,, & \hat{\mu} < 0 \end{cases} \tag{A.2}$$

The CDF of $\tilde{t}_\mu$ is

$$F(\tilde{t}_\mu) = \begin{cases} \Phi(\frac{\sigma_0}{\sigma}(-\frac{\mu_H-\mu}{\sigma_1}+\sqrt{\tilde{t}_\mu})) - \Phi(\frac{\sigma_0}{\sigma}(-\frac{\mu_H-\mu}{\sigma_1}-\sqrt{\tilde{t}_\mu}))\,, & \tilde{t}_\mu \le T_\mu^{0+} \\ \Phi(\frac{\sigma_0}{\sigma}(-\frac{\mu_H-\mu}{\sigma_1}+\sqrt{\tilde{t}_\mu})) - \Phi(-\frac{\mu_H}{\sigma}+\frac{\sigma_0(\tilde{t}_\mu-T_\mu^{0+})}{2\sigma(\frac{-\mu_H}{\sigma_0}+\frac{\mu_H-\mu}{\sigma_1})})\,, & T_\mu^{0+} < \tilde{t}_\mu \le T_\mu^* \\ \Phi(\frac{\sigma_0}{\sigma}(-\frac{\mu_H-\mu}{\sigma_1}+\sqrt{\tilde{t}_\mu})) - \Phi(\frac{\hat{\mu}^-(\tilde{t}_\mu)-\mu_H}{\sigma})\,, & T_\mu^* < \tilde{t}_\mu \text{ and } \tilde{t}_\mu \le T_\mu^L \text{ if } \sigma_0 > \sigma_0' \\ \Phi(\frac{\sigma_0}{\sigma}(-\frac{\mu_H-\mu}{\sigma_1}+\sqrt{\tilde{t}_\mu}))\,, & \tilde{t}_\mu > T_\mu^L \text{ if } \sigma_0 > \sigma_0' \end{cases}. \tag{A.3}$$

For the special case of $\mu_H = \mu$, it becomes

$$F(\tilde{t}_\mu|\mu_H=\mu) = \begin{cases} 2\Phi(\sqrt{\tilde{t}_\mu}) - 1\,, & \tilde{t}_\mu \le T_\mu^{0+} \\ \Phi(\sqrt{\tilde{t}_\mu}) - \Phi(-\frac{\sigma}{2\mu_H}\tilde{t}_\mu - \frac{\mu_H}{2\sigma})\,, & T_\mu^{0+} < \tilde{t}_\mu \le T_\mu^* \\ \Phi(\sqrt{\tilde{t}_\mu}) - \Phi(\frac{\hat{\mu}^-(\tilde{t}_\mu)-\mu_H}{\sigma})\,, & T_\mu^* < \tilde{t}_\mu \text{ and } \tilde{t}_\mu \le T_\mu^L \text{ if } \sigma_0 > \sigma_0' \\ \Phi(\sqrt{\tilde{t}_\mu})\,, & \tilde{t}_\mu > T_\mu^L \text{ if } \sigma_0 > \sigma_0' \end{cases}. \tag{A.4}$$

For q_μ, the asymptotic form is

$$q_\mu = \begin{cases} 0\,, & \hat{\mu} > \mu \\ t_\mu = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H-\mu}{\sigma_1})^2\,, & \hat{\mu} \ge \mu \end{cases} \tag{A.5}$$

The CDF of q_μ is

$$F(q_\mu) = \begin{cases} \Phi(-\frac{\mu-\mu_H}{\sigma}) + \theta(\mu-\hat{\mu}^R)[\Phi(\frac{\sigma_0}{\sigma}(\sqrt{q_\mu}+\frac{\mu_H-\mu}{\sigma_1})) \\ \quad -\Phi(\frac{\sigma_0}{\sigma}(-\sqrt{q_\mu}+\frac{\mu_H-\mu}{\sigma_1}))]\,, & q_\mu \le T_\mu^\mu \\ \Phi(\frac{\sigma_0}{\sigma}(\sqrt{q_\mu}+\frac{\mu_H-\mu}{\sigma_1}))\,, & q_\mu > T_\mu^\mu \end{cases}. \tag{A.6}$$

For the special case of $\mu_H = \mu$, it is $F(q_\mu|\mu_H=\mu) = \Phi(\sqrt{q_\mu})$, which is the same as that in Ref. [1].

The asymptotic form of t_0 as a function of $\hat{\mu}$ is just t_μ with $\mu = 0$. Its CDF is given in Eq. A.1 with $\mu = 0$.

The asymptotic form of q_0 as a function of $\hat{\mu}$ is

$$q_0 = \begin{cases} t_0 = (\frac{\hat{\mu}-\mu_H}{\sigma_0} + \frac{\mu_H}{\sigma_1})^2\,, & \hat{\mu} \ge 0 \\ 0\,, & \hat{\mu} < 0 \end{cases}. \tag{A.7}$$

The CDF of q_0 is

$$F(q_0) = \begin{cases} \Phi(\frac{-\mu_H}{\sigma}) + \theta(\sigma_1-\sigma_0)[\Phi(\frac{\sigma_0}{\sigma}(\sqrt{q_0}-\frac{\mu_H}{\sigma_1})) - \Phi(\frac{\sigma_0}{\sigma}(-\sqrt{q_0}-\frac{\mu_H}{\sigma_1}))]\,, & q_0 \le (\frac{\sigma_0-\sigma_1}{\sigma_0\sigma_1})^2\mu_H^2 \\ \Phi(\frac{\sigma_0}{\sigma}(\sqrt{q_0}-\frac{\mu_H}{\sigma_1}))\,, & q_0 > (\frac{\sigma_0-\sigma_1}{\sigma_0\sigma_1})^2\mu_H^2 \end{cases}. \tag{A.8}$$

For the special case of $\mu_H = 0$, the CDF is

$$F(q_0|\mu_H=0) = \Phi(\sqrt{q_0})\,, \tag{A.9}$$

which is exactly the same as that based on Wald's approximation.

Table 6. Summary of the upper limits from the asymptotic formulae and the toy MC method.

		t_μ	$\tilde{t}_\mu$	q_μ	$\tilde{q}_\mu$
Ex. I	Toy MC	4.30	4.21	3.46	3.53
	This Work	4.13	4.10	3.37	3.36
	Ref. [1]	4.13	4.00	3.37	3.25
Ex. II	Toy MC	6.32	6.46	5.52	5.42
	This Work	6.37	6.37	5.37	5.32
	Ref. [1]	6.37	6.30	5.37	5.23
Ex. III	Toy MC	11.12	11.32	10.03	9.86
	This Work	11.28	11.28	9.72	9.72
	Ref. [1]	11.28	11.22	9.72	9.64

B Comparison of different test statistics

In the section, the test statistics (t_μ, q_μ, $\tilde{t}_\mu$ and $\tilde{q}_\mu$) are compared. Figure 10, 11 and 12 show the distributions of the test statistics in the three examples respectively. Table 6 summarizes upper limits in the three examples for the four test statistics proposed in Ref. [1] as well as the predictions from the old and new asymptotic formulae. We have the following observations.

- The old and new asymptotic formulae have almost the same predictions for t_μ and q_μ. This is not surprising because they give exactly the same probability distribution for the special case of $\mu_H = \mu$.

- The difference between the old and new formulae is only significant for $\tilde{t}_\mu$ and $\tilde{q}_\mu$; and the new formulae give closer results to those from the toy MC method. This is mainly because the new formulae treats the part with $\hat{\mu} < 0$ very differently.

- Let $UL(t)$ denote the upper limit based on the test statistic t. We find that $UL(q_\mu) \approx UL(\tilde{q}_\mu) < UL(t_\mu) \approx UL(\tilde{t}_\mu)$. This is not surprising because q_μ and $\tilde{q}_\mu$ do not consider the data with $\hat{\mu} > \mu$ as incompatible with a hypothesized value μ by definition. t_μ and $\tilde{t}_\mu$ are suitable for setting a confidence interval as summarized in Tab. 1.

C A finer description of the PDF of $\hat{\mu}$

According to Wald's theorem, $\hat{\mu}$ abides by a Gaussian distribution approximately. Here we try to provide a finer description. The optimal value $\hat{\mu}$ is determined by $\partial \ln \mathcal{L}/\partial\mu = 0$, which leads to

$$\sum_{i=1}^{N_{\rm bins}} \frac{n_i s_i}{b_i + \hat{\mu} s_i} - s = 0\,. \tag{C.1}$$

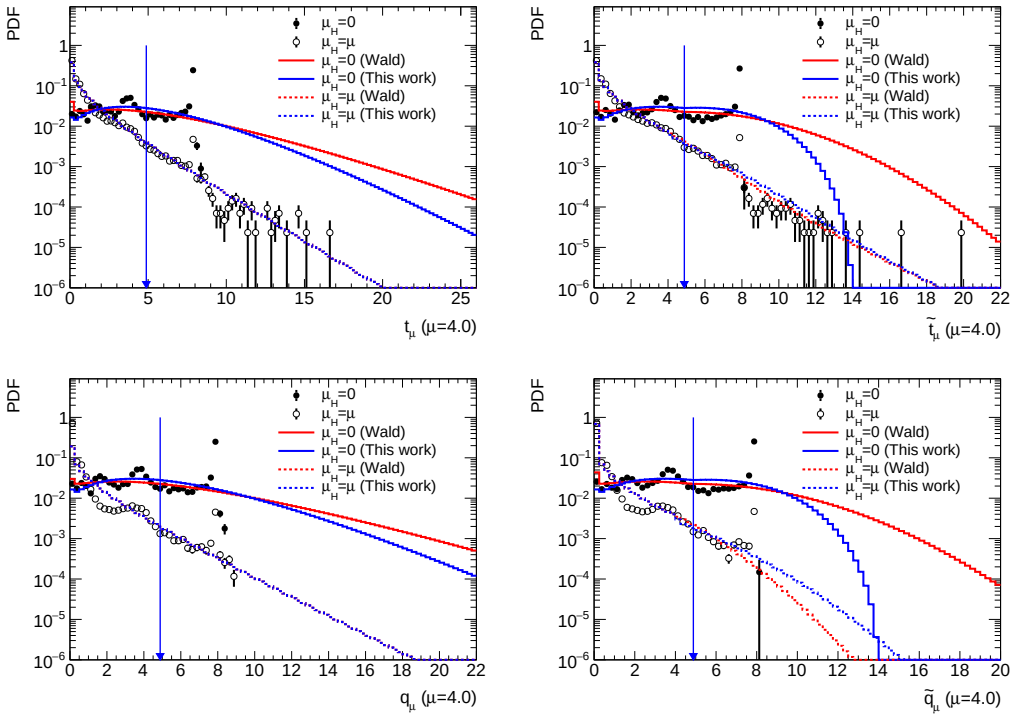

Figure 10. The probability distribution of t_μ (top left), $\tilde{t}_\mu$ (top right), q_μ (bottom left) and $\tilde{q}_\mu$ (bottom right) in Ex 1 for $\mu_H = 4$. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed value of the test statistic. This is the case with all systematic uncertainties considered.

If the data distribution n_i is consistent with the hypothesis with a signal strength μ_H, then $\hat{\mu}$ is μ_H. We can expand the equation around $\hat{\mu} = \mu_H$.

$$\sum_{i=1}^{N_{\rm bins}} \frac{n_i s_i}{b_i + \mu_H s_i}\left(1 - \frac{(\hat{\mu} - \mu_H)s_i}{b_i + \mu_H s_i}\right) - s \approx 0 \tag{C.2}$$

$$\rightarrow \sum_{i=1}^{N_{\rm bins}} \frac{n_i s_i}{b_i + \mu_H s_i} - \sum_{i=1}^{N_{\rm bins}} \frac{s_i^2}{b_i + \mu_H s_i}(\hat{\mu} - \mu_H) - s \approx 0 \tag{C.3}$$

$$\rightarrow \frac{\hat{\mu} - \mu_H}{\sigma^2} + s \approx \sum_{i=1}^{N_{\rm bins}} \frac{n_i s_i}{b_i + \mu_H s_i}\,, \tag{C.4}$$

where $n_i \approx b_i + \mu_H s_i$ is used in the second term in the left-hand side (LHS) of Eq. C.3 and σ is the same as defined in last section. Defining $\mu' \equiv \frac{\hat{\mu} - \mu_H}{\sigma^2} + s$ and $n'_i \equiv \frac{n_i s_i}{b_i + \mu_H s_i}$, we have

$$\mu' \approx \sum_{i=1}^{N_{\rm bins}} n'_i\,. \tag{C.5}$$

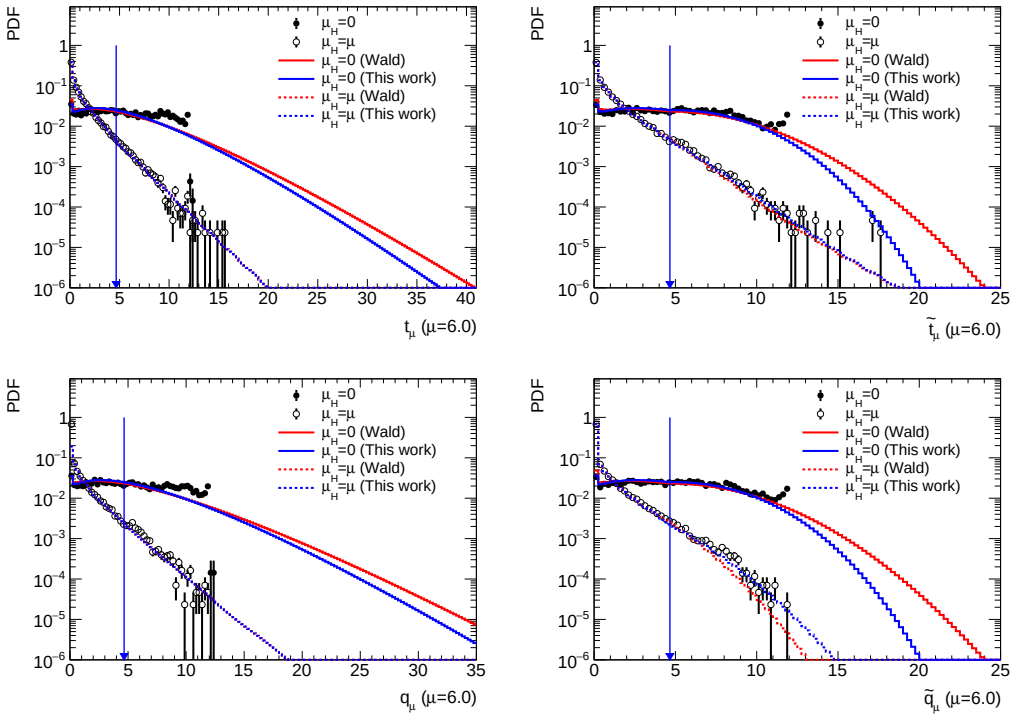

Figure 11. The probability distribution of t_μ (top left), $\tilde{t}_\mu$ (top right), q_μ (bottom left) and $\tilde{q}_\mu$ (bottom right) in Ex 2 for $\mu_H = 6$. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed value of the test statistic. This is the case with all systematic uncertainties considered.

To obtain the PDF of μ' (hence the PDF of $\hat{\mu}$) from the PDF of n_i's, we resort to the method of characteristic function [9]. Letting $\phi_X(k)$ denote the characteristic function for the random variable X, we have

$$\phi_{n_i'}(k) = \sum_{n=0}^{\infty} \frac{(b_i + \mu_H s_i)^{n_i}}{n_i!} e^{-(b_i+\mu_H s_i)} e^{ikn_i'} \tag{C.6}$$

$$= \sum_{n=0}^{\infty} \frac{(b_i + \mu_H s_i)^{n_i}}{n_i!} e^{-(b_i+\mu_H s_i)} e^{ik\frac{n_i s_i}{b_i+\mu_H s_i}} \tag{C.7}$$

$$= e^{(b_i+\mu_H s_i)(e^{i\frac{ks_i}{b_i+\mu_H s_i}}-1)} , \tag{C.8}$$

and

$$\phi_{\mu'}(k) = \Pi_{i=1}^{N_{\rm bins}} \phi_{n_i'}(k) \tag{C.9}$$

$$= \Pi_{i=1}^{N_{\rm bins}} e^{(b_i+\mu_H s_i)(e^{i\frac{ks_i}{b_i+\mu_H s_i}}-1)} . \tag{C.10}$$

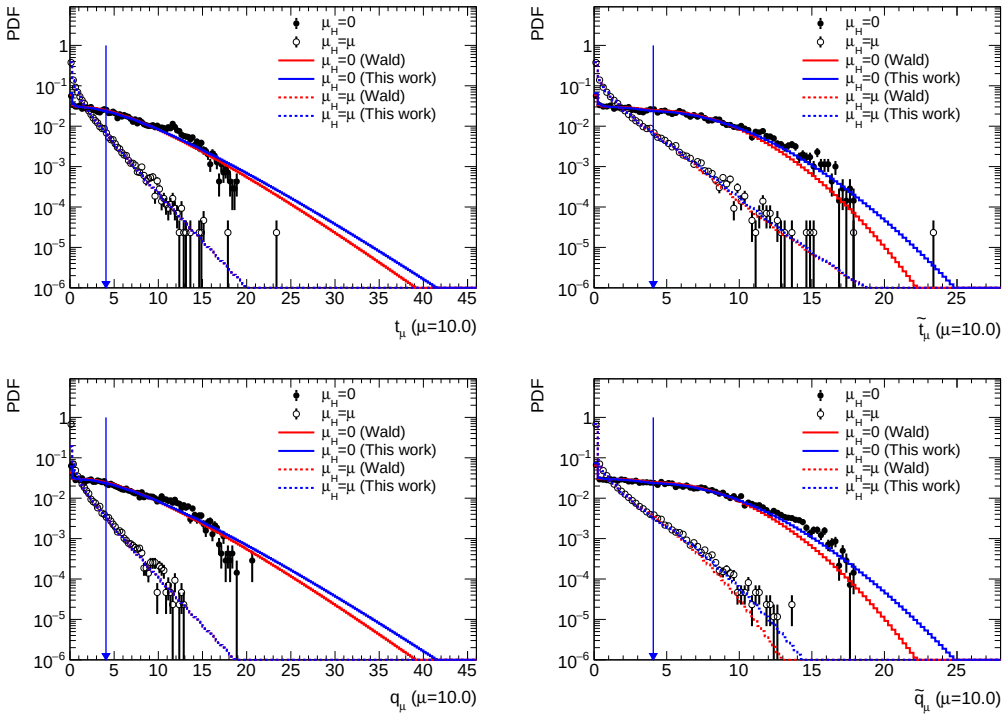

Figure 12. The probability distribution of t_μ (top left), $\tilde{t}_\mu$ (top right), q_μ (bottom left) and $\tilde{q}_\mu$ (bottom right) in Ex 3 for $\mu_H = 10$. The black dots and open circles represent the toy MC results. The blue solid/dashed histograms represent the new asymptotic formulae in this work while the red solid/dashed histograms represent the old asymptotic formulae from Wald's approximation. The blue arrow represents the observed value of the test statistic. This is the case with all systematic uncertainties considered.

Let us apply some approximations to $\phi_{\mu'}(k)$ to simplify the derivation.

$$\ln \phi_{\mu'}(k) = \sum_{i=1}^{N_{\text{bins}}} (b_i + \mu_H s_i)(e^{i\frac{ks_i}{b_i+\mu_H s_i}} - 1) \tag{C.11}$$

$$\approx \sum_{i=1}^{N_{\text{bins}}} (b_i + \mu_H s_i)(i\frac{ks_i}{b_i+\mu_H s_i} - \frac{1}{2}\frac{k^2 s_i^2}{(b_i+\mu_H s_i)^2} - \frac{i}{6}\frac{k^3 s_i^3}{(b_i+\mu_H s_i)^3}) \tag{C.12}$$

$$= iks - \frac{1}{2}\frac{k^2}{\sigma^2} - \frac{i}{6}c_3 k^3 \,. \tag{C.13}$$

where c_3 is defined as

$$c_3 = \sum_{i=1}^{N_{\text{bins}}} \frac{s_i^3}{(b_i+\mu_H s_i)^2} \,. \tag{C.14}$$

The PDF of μ' is then

$$g(\mu') = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_{\mu'}(k)e^{-ik\mu'}dk \tag{C.15}$$

$$= \frac{1}{2\pi}e^{-\frac{1}{2}\sigma^2(s-\mu')^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}(\frac{k}{\sigma}-i\sigma(s-\mu'))^2-\frac{i}{6}c_3k^3}dk\,. \tag{C.16}$$

Here we can see that this is a Gaussian distribution if c_3 and all higher-order terms are ignored. Since k centers around $i\sigma^2(s-\mu')$ mostly, we have the following approximation

$$\ln\phi_{\mu'}(k) \approx iks - \frac{1}{2}\frac{k^2}{\sigma^2} - \frac{i}{6}c_3k^2(i\sigma^2(s-\mu'))\,. \tag{C.17}$$

Clearly, the last term lead to the correction to the standard deviation. Hence the PDF of $\hat{\mu}$ is

$$f(\hat{\mu}) = \frac{1}{\sigma^2}g(\mu') \tag{C.18}$$

$$= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty}\phi_{\mu'}(k)e^{-ik\mu'}dk \tag{C.19}$$

$$\approx \frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty}e^{ik(s-\mu')-\frac{1}{2}\frac{k^2}{\sigma_0^2}}dk \tag{C.20}$$

$$= \frac{A}{\sqrt{2\pi}\sigma^*}e^{-\frac{1}{2}(\frac{\hat{\mu}-\mu_H}{\sigma^*})^2}\,, \tag{C.21}$$

where A is a constant normalization factor; σ^* is a function of $\hat{\mu}$.

$$\sigma^*(\hat{\mu}) = \sqrt{1+\frac{c_3}{3}\sigma^2(\hat{\mu}-\mu_H)}\sigma\,. \tag{C.22}$$

We can see the PDF of $\hat{\mu}$, $f(\hat{\mu})$, is approximately a Gaussian distribution with a $\hat{\mu}$-dependent standard deviation. Basically, the probability distribution of $\hat{\mu}$ is narrower if $\hat{\mu} < \mu_H$ and fatter if $\hat{\mu} > \mu_H$. There is a truncation at which $1+\frac{c_3}{3}\sigma^2(\hat{\mu}-\mu_H) = 0$, namely, $\hat{\mu} = \hat{\mu}^* \equiv \mu_H - \frac{3}{c_3\sigma^2}$. This is consistent with the intuitive picture that $\hat{\mu}$ cannot be too negative otherwise the expected number of events, $b_i+\mu s_i$, is negative and this is not allowed in reality. The correction to the standard deviation is only reasonable if $\frac{c_3}{6}\sigma^2(\hat{\mu}-\mu_H) < 1$ and definitely not valid if $\hat{\mu} \to +\infty$. This point is obvious in Fig. 5, Fig. 6 and Fig. 7.

References

[1] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C 71 (2011) 1554, Eur. Phys. J. C 73 (2013) 2501 (Erratum), arXiv:1007.1727

[2] A. Wald, *Tests of Statistical Hypothesis Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, Vol. **54**, No. 3, pp. 426-482.

[3] ATLAS Collaboration, Phys. Lett. B 716 (2012) 1, arXiv:1207.7214.

[4] CMS Collaboration, Phys. Lett. B 716 (2012) 30, arXiv:1207.7235.

[5] L.-G. Xia, JHEP **08** (2021) 071, arXiv:2012.15618, version 1.

[6] G. Zech, Nucl. Instrum. Meth. **A 277** (1989) 608.

[7] A. L. Read, J. Phys. **G 28** (2002) 2693.

[8] L.-G. Xia, J. Phys. **G 46** (2019) 085004, arXiv:1805.03961.

[9] G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford, 1998.