Coverage Studies for $H \to \gamma\gamma$

Kyle Cranmer
Brookhaven National Laboratory

Stathes Paganis
University of Wisconsin - Madison

March 4, 2005

Abstract

Several methods have been proposed to characterize the statistical significance of an observation in the $H \to \gamma\gamma$ channel. In order to compare these methods, we introduce the notion of a coverage study to assess the performance of an arbitrary discovery criterion. We have performed coverage studies for several criteria in the context of $H \to \gamma\gamma$. We have found that the residual uncertainty on the background after a sideband fit is not negligible and must be incorporated into the significance calculation. To that end, we have derived an equation for the residual background uncertainty, incorporated it into both Cousins-Highland and fully frequentist calculations, and found that those methods give the correct coverage.

Contents

1 Introduction

The inclusive $H \to \gamma\gamma$ analysis has a huge continuum background with a simple shape. The strategy for this channel has been to use the sidebands to extract the number of expected background events in the signal-like region. The fit takes into account both the electromagnetic energy-scale uncertainty and the cross-section uncertainty. Given this channel's low s/b, the uncertainty on the background must be less than about 0.2% for the expected significance to be greater than 5σ. By using the sideband technique, the uncertainty on the expected background was expected to be negligible. As we will show, the uncertainty is not negligible, but the method is robust against uniform energy scale uncertainties.

The use of a fitted background as a substitute for a true prediction could be substantiated if the fits to both Monte Carlo background events and data provide an acceptable χ^2. If the χ^2 is not acceptable, then the parametric form either needs to be rejected or extended. Extending the parametric form of the continuum background will most likely result in a higher uncertainty in the extrapolation of background events into the signal-like region. Given the low tolerance of the $H \to \gamma\gamma$ analysis to background uncertainty, it needs to be confirmed that an acceptable parametric form and small background uncertainty can be achieved simultaneously.

As a first step in this direction, a coverage study for $H \to \gamma\gamma$ has been performed. In Section 2, we define what is meant by a *coverage study* and explain why it is a useful concept. In Section 3, we give the details for our Toy Monte Carlo that was used to perform the coverage studies in this note. Section 4 describes a few methods for calculating the statistical significance of an observation in the $H \to \gamma\gamma$ channel, two of which incorporate the residual background uncertainty after a fit to the sidebands. In Section 5 we present the results of the coverage studies for these different discovery criteria. Lastly, in Section 6 we discuss future extensions to this work and draw conclusions in Section 7.

2 Definition of a Coverage Study

When we perform a search for a new particle, we are performing a hypothesis test between the null (background-only) hypothesis, H_0, and the alternate (signal-plus-background) hypothesis, H_1. When we say that we have a signal that is 3σ, we are referring to the probability that the background could have fluctuated to produce the observation. Similarly, when we set 5σ to be our discovery threshold, it refers to a probability that the background could produce such an observation. The chance that we claim a discovery in the presence of background only is called the rate of Type I error, α, and the relationship to $N\sigma$ is given by

$$\alpha = \frac{1 - \mathrm{erf}(N/\sqrt{2})}{2}. \qquad (1)$$

The idea of coverage is simple: a method *should* commit a Type I error with probability α [1]. If the method claims discovery too often (is "optimistic"), it is said to

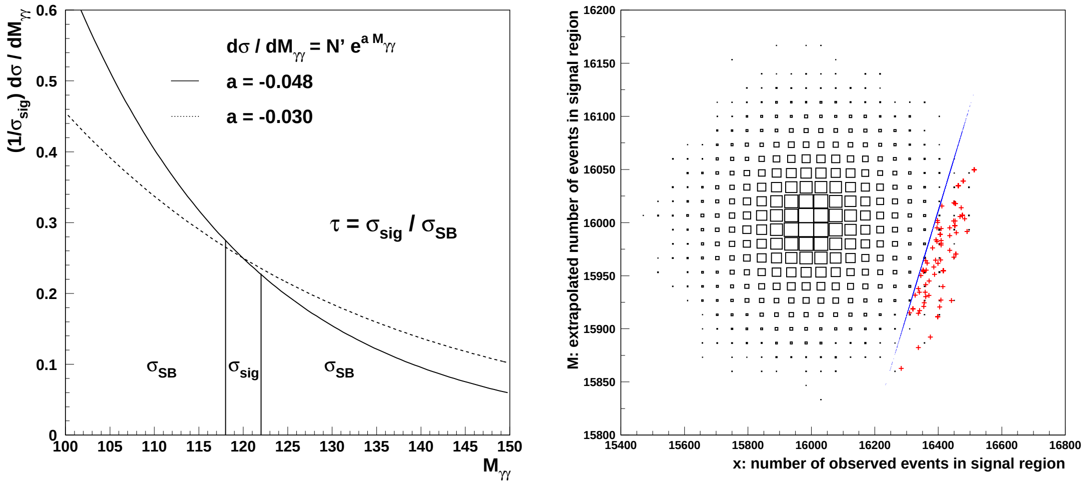

Figure 1: Left: exponential form used for Toy Monte Carlo. Right: number of events from sideband fit extrapolated into the signal-like region vs. observed number of events in the signal-like region. The red points represent experiments considered as 3σ discoveries.

undercover (for H_0). Conversely, if the method claims discovery too infrequently (is "conservative"), it is said to overcover (for H_0). The nice thing about the coverage concept is that you can apply it to *any* well-defined method. To perform these coverage studies we have used a simple Toy Monte Carlo to generate toy experiments.

3 Details of the Toy Monte Carlo

A toy Monte Carlo was used to generate a number of experiments with a $M_{\gamma\gamma}$ spectrum given by a simple exponential form (see Figure 1 and Equation 2). The exponential was normalized such that $\mathcal{N}$, the average number of events generated in the mass window $118 < M_{\gamma\gamma} < 122$ GeV, was 16000[1]. For each Toy Monte Carlo experiment, the spectrum was sampled in the range 100-150 GeV (about 200,000 events). We then varied the exponent in the range $[-0.048, -0.030]$ GeV^{-1} in steps of 0.002 GeV^{-1} producing about 80,000 toy Monte Carlo experiments per exponent tested.

$$\frac{d\sigma}{dM_{\gamma\gamma}} = \underbrace{\left(\frac{a\mathcal{N}}{e^{am^+} - e^{am^-}} \right)}_{N'} e^{aM_{\gamma\gamma}} \tag{2}$$

[1]Given current Monte Carlo estimates, this corresponds to about 40 fb^{-1} of data.

4 Various Discovery Criteria

For each Toy Monte Carlo experiment, we evaluated various discovery criteria. Each experiment was treated as if it were real data, with no access to the parameters of the Toy Monte Carlo. The original goal of this study was to address the residual background uncertainty after a sideband fit. Thus, we fit the background to a simple exponential shape in the range 100-150 GeV, excluding the pre-defined mass window $118 < M_{\gamma\gamma} < 122$. For each experiment, the result of the sideband fit was used to extrapolate the expected background in the signal-like region, denoted M_0. In total, nearly a million MINUIT fits were performed.

For clarity, let us distinguish between the variables b and M. The variable b *is not* an experimental observable, but a "truth" quantity defined as the average number of events in the mass window. Formally, $b = \int_{m_-}^{m_+} dM_{\gamma\gamma}(\mathcal{L}\, d\sigma/dM_{\gamma\gamma})$. The variable M *is* an experimental observable, it is the number of events in the signal-like region extrapolated from the fit to the sideband. Even though b is fixed, the measured value of M will vary from experiment to experiment due to fluctuations in the sideband.

Observations on the Fit Procedure

We arrived at several interesting results. First, we found that a *modified* least squares fit to a binned $M_{\gamma\gamma}$ spectrum leads to a background extrapolation that is biased by about 10 events (which is about a 0.08σ effect)[2]. Second, we found that the variation in the number of extrapolated background events, δM, from the mean value was about 38 events across the range of exponents tested (see Figure 1 right). Because the same exponential form was used to generate the toy Monte Carlo, it is not surprising that the background uncertainty is exactly what one would expect from the number of events in the side band region. For convenience let us use τ to denote the ratio of the cross section in the signal like region, σ_{sig}, to the cross section in the sidebands, σ_{SB}. Thus the background uncertainty due to the statistical fluctuations in the sideband region is given by

$$\delta M \approx \tau \sqrt{N_{SB}}. \tag{3}$$

The $s/\sqrt{b}$ Method

The typical significance calculation for this channel is $s/\sqrt{b}$. First, one should realize that this is a calculation of the *expected* significance and can not be directly applied to data. Second, one should realize that this method does not take into account any uncertainty on the mean background. In the same spirit as $s/\sqrt{b}$, we tested the discovery criterion in which one claims an $N\sigma$ discovery if the number of observed events in the signal like region, x, was greater the extrapolated background M plus N times the statistical fluctuation on M events. Formally, a discovery was claimed if

$$x_0 > M_0 + N\sqrt{M_0}. \tag{4}$$

[2]This is due to the fact that when the bin fluctuates down, the error used in the *modified* least squares fit also goes down. In the unmodified least squares fit, the error is defined by the model not the observation.

The Cousins Highland Method

The Cousins-Highland formalism [2] has been used in ATLAS to incorporate background uncertainty (systematic error on the background) in the TDR [3] and more recent notes [4, 5]. The basic idea of this method is to perform a convolution of the statistical and systematic errors. The Cousins-Highland method has some formal inconsistencies and implicitly depends on prior assumptions about the background; however, it performs well in most practical situations. Furthermore, in the context of $H \to \gamma\gamma$, the number of events is so large that it dominates the effect of any *a priori* assumptions.

The formula that was used in the TDR to calculate the expected significance was

$$\sigma^{CH} = \frac{s}{\sqrt{b(1+\alpha^2 b)}}, \tag{5}$$

where α is a relative background uncertainty. As in the case of $s/\sqrt{b}$, this formula can not be directly applied to data. Furthermore, we did not previously know the value of α. One of the major results of this note was that $\delta M \approx \tau\sqrt{N_{SB}}$ and $M = \tau N_{SB}$, thus one arrives at the equivalent equation that describes the $N\sigma$ discovery criterion:

$$x_0 > M_0 + N\sqrt{M_0(1+\tau)} \tag{6}$$

A Purely Frequentist Method

In this section we derive a fully frequentist result based in part on the work presented in Ref. [6]. The basic idea of the frequentist technique is that it is explicitly designed to have the correct coverage[3].

Equations 4 and 6 define two acceptance regions for the alternate hypothesis (the complement of which is an acceptance region for the null hypothesis). These acceptance regions, as presented, are *ad hoc*. Now we shall derive a form for the acceptance region that has the correct coverage (implicitly forming the Neyman Construction).

For each value of the parameters of H_0 there is a distribution $L(x, M|H_0, \mathcal{N}, a)$. If we can find a region W with the property

$$\int_W L(x, M|H_0, \mathcal{N}, a)dxdM = \alpha, \tag{7}$$

for every value of the nuisance parameters $\mathcal{N}$ and a, then we have a similar test which should provide the correct coverage. For W of the form $W = \{x, M|x > M + \eta\sqrt{M}\}$, the challenge is to find the η which satisfies Equation 7. If we write the boundary as a function $x(M) = M + \eta\sqrt{M}$, and expand it about M_0, then the linear form of the boundary is

$$x(M) \approx M + \eta\sqrt{M_0} + (M - M_0)\underbrace{\left(1 + \frac{\eta}{2\sqrt{M_0}}\right)}_{m^{-1}} \tag{8}$$

[3]This is not possible in general if the null hypothesis has a continuously varying nuisance parameter, but as we shall see, for this problem, we achieve approximately similar tests.

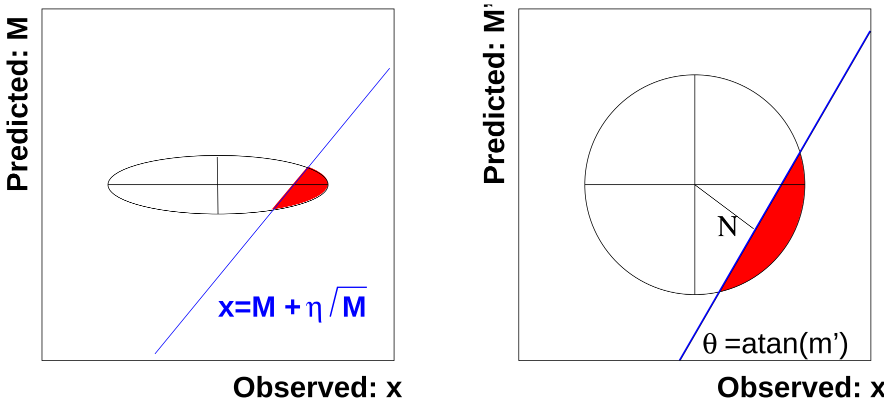

Figure 2: Determination of η via a change of variables.

Considering contours of $L(x, M|H_0, \mathcal{N}, a)$ as ellipses with eccentricity

$$\epsilon = \frac{\Delta M}{\Delta x} = \frac{\tau\sqrt{N_{SB}}}{\sqrt{M}} = \sqrt{\tau} \tag{9}$$

and the critical boundary as lines with slope $m = (1 + \frac{\eta}{2\sqrt{M_0}})^{-1}$, the goal is to find η that satisfies Equation 7. By a change of variables $M' = M/\epsilon$, the contours of $L(x, M|H_0, \mathcal{N}, a)$ become circles and the critical boundary has slope $m' = m/\epsilon$. In this new space, the coverage requirement is satisfied if the perpendicular bisector, has a length N (in number of Gaussian σ) that corresponds to α according to Equation 1. Here we have $\theta = \tan^{-1} m/\epsilon$ and $\eta = N/\sin\theta$. Note, the x-direction was not modified by the change of coordinates. We can re-write $\eta = N/\sin(\tan^{-1}(m/\epsilon)) = N\sqrt{1+\epsilon^2/m^2}$.

Thus the expected significance can be calculated with the equation

$$\sigma^F = \underbrace{\frac{s}{\sqrt{b(1+\tau/m^2)}}}_{\text{new result}} \tag{10}$$

and the discovery criterion is given by

$$x_0 > M_0 + N\sqrt{M(1+\tau/m^2)}, \tag{11}$$

where

$$m = \left(1 + \frac{N}{2\sqrt{M_0}}\right)^{-1} \tag{12}$$

and N is the desired significance of the test.

The quantity m, which is less than unity, can be seen as a correction to the Cousins-Highland result. In the case of $H \to \gamma\gamma$, the correction is minuscule, and the Cousins-Highland result is an excellent approximation.

Exponent	BINNED	UNBINNED	COUSINS-HIGHLAND	FREQUENTIST	N_{exper}
-0.030	0.18%	0.12%	0.08%	0.08%	83726
-0.032	0.19%	0.15%	0.10%	0.10%	46835
-0.034	0.38%	0.33%	0.27%	0.27%	84548
-0.036	0.19%	0.18%	0.16%	0.16%	84628
-0.038	0.21%	0.17%	0.11%	0.11%	84294
-0.040	0.27%	0.23%	0.16%	0.16%	90020
-0.042	0.22%	0.18%	0.10%	0.10%	90020
-0.044	0.21%	0.16%	0.10%	0.10%	90020
-0.046	0.32%	0.27%	0.15%	0.15%	85630
-0.048	0.23%	0.19%	0.12%	0.12%	78514
all	0.25%	0.21%	0.14%	0.14%	818235

Table 1: Results of the $H \rightarrow \gamma\gamma$ coverage study (see text).

5 Results: The Impact of Background Uncertainty

Let us examine the impact of background uncertainty on the $H \rightarrow \gamma\gamma$ significance. If we assume that the background uncertainty is negligible, then for $M_H = 120$ GeV $\sigma = s/\sqrt{b} = 3.2\sigma$ (using TDR results). In our studies, the sideband region ranged from 100-150 GeV (which is quite large), thus $\tau = 8\%$ and $\sigma^{CH} = 3.1\sigma$. However, if we use the sideband region shown in the TDR which ranged from 105-135 GeV, $\tau = 25\%$ and $\sigma^{CH} = 2.9\sigma$.

In Table 1, we show the probability to claim a 3σ discovery given background-only experiments for several methods. The chance for such a discovery should be 0.135% via Equation 1. The label BINNED refers to the scenario in which the background is estimated from a χ^2 fit to a binned $M_{\gamma\gamma}$ spectrum (which causes a bias in the expected number of events). The label UNBINNED refers to the unbinned extended likelihood fit procedure, for which no bias was observed. In both the BINNED and UNBINNED cases, an experiment was classified as a discovery according to Equation 4. The labels COUSINS-HIGHLAND and FREQUENTIST refer to the unbinned case when discovery was claimed with Equation 6 and Equation 11, respectively. The statistical error on the entries is approximately 10% for each exponent, and 3% for the sum over all exponents.

6 Future Directions

The obvious extension of this work is to use different parametric forms for the generation of the Toy Monte Carlo and the fit to the sideband. We now have NLO predictions of the reducible and irreducible background components [7], which can be used as templates either to produce Toy Monte Carlo or to fit the background.

7 Conclusions

In conclusion, coverage studies like those presented in this note are powerful tools to assess a proposed statistical method or discovery criterion. We have focused on the performance of a few discovery criteria and two sideband fitting methods. The sideband fit technique is a powerful method, but has some subtleties. First, one must use an extended unbinned likelihood fit (or some other unbiased fit method) to avoid undercoverage. Secondly, even after the sideband fit, there is a residual background uncertainty that is not negligible. The Cousins-Highland and frequentist approaches take into account the background uncertainty and produce the correct coverage. Lastly, one must understand the trade-off between a larger sideband with larger shape systematics and a smaller sideband with larger statistical fluctuations.

References

[1] A. Stuart, J.K. Ord, and S. Arnold, Kendall's Advanced Theory of Statistics, Vol 2A (6th Ed.) (Oxford University Press, New York, 1994).

[2] R.D. Cousins and V.L. Highland, Nucl. Instrum. Meth. **A320** (1992) 331–335.

[3] ATLAS Collaboration, Detector and Physics Performance Technical Design Report (Volume II), CERN-LHCC/99-15 (1999).

[4] K. Cranmer, P. McNamara, B. Mellado, W. Quayle, Sau Lan Wu, Confidence Level Calculations for $H \to W^+W^- \to l^+l^-\not{p}_T$ for $115 < M_H < 130\,\mathrm{GeV}$ Using Vector Boson Fusion, ATLAS Note ATL-PHYS-2003-008 (2003).

[5] K. Cranmer, B. Mellado, W. Quayle, Sau Lan Wu, Statistical Methods to Assess the Combined Sensitivity of the ATLAS Detector to the Higgs Boson in the Standard Model, ATLAS Note ATL-PHYS-2004-034 (2004).

[6] K. Cranmer, Frequentist Hypothesis Testing with Background Uncertainty, PhyStat2003 physics/0310108 (2003).

[7] M. Escalier, Higgs to $\gamma\gamma$ at NLO with full simulation of signal, http://agenda.cern.ch/fullAgenda.php?ida=a051030.

Symbol	**Meaning**
H_0	The null (background-only) hypothesis.
H_1	The alternate (signal-plus-background) hypothesis.
$\alpha = CL_b$	The probability of Type I error, the confidence in the background-only hypothesis.
$\mathrm{erf}(x)$	The error function.
a	The exponent in the Toy MC with units GeV^{-1}.
$\mathcal{N}$	The normalization of the exponential in the Toy MC. Defined as the number of events expected in the signal like region so as to remove dependence on a.
σ_{SB}	The expected cross-section in the sideband region (excluding the signal-like region).
N_{SB}	The integrated luminosity times σ_{SB}, *i.e.* the expected number of events in the sideband region.
σ_{sig}	The expected cross-section in the signal-like region.
τ	The ratio of σ_{sig} to σ_{SB}.
x	The number of events in the signal-like region.
M	The number of events expected/predicted in the signal-like region from fitting the sideband.
$L(x, M \vert H_0, a, \mathcal{N})$	The likelihood to observe x and M given the background-only hypothesis with parameters a and $\mathcal{N}$.
x_0	The measured value of x for a particular experiment.
M_0	The measured value of M for a particular experiment.
ϵ	The ratio of ΔM to Δx, *i.e.* the eccentricity of a contour of $L(x, M \vert H_0, a, \mathcal{N})$.
W	The acceptance region for H_1 in the $x - M$ plane.
η	The parameter that defines the a acceptance region in the formula $W = \{x, M \vert x < M + \eta\sqrt{M}\}$.
m	The slope (dM/dx) of the boundary of W (at $M = M_0$).
M'	M after a change of variables, $M' = M/\epsilon$.
s	The number of expected signal events in the signal-like region. Used for estimating the expected statistical significance, not in an actual experiment.
b	The number of expected background events in the signal-like region. Used for estimating the expected statistical significance, not in an actual experiment.
N	The desired significance of the test, or the threshold for discovery in Gaussian σ.
σ^F	The expected significance from the frequentist method.
σ^{CH}	The expected significance from the Cousins-Highland method.