Combined searches for the Higgs boson with ATLAS and CMS

Kyle Cranmer, on behalf of the LHC Higgs Combination Group
New York University

Abstract
This document outlines the conceptual challenges involved in forming the combined statistical model of several ATLAS and CMS searches for the Higgs boson, as well as the technology developed to form such a complicated model and the statistical tests the group is considering for the initial result.

1 Introduction

Perhaps the most pressing open question in fundamental particle physics is why the W and Z bosons, the particles associated with the weak interaction, are massive. The fact that these bosons have mass is associated to a phenomena called electroweak symmetry breaking; however, the mechanism for the electroweak symmetry breaking has not been established. Within the standard model (SM) of particle physics, a specific manifestation of the so-called "Higgs mechanism" gives mass to the W and Z bosons and also predicts a new particle, H, called the Higgs boson, which has not been observed. The search for the Higgs boson is one of the primary goals of the Large Hadron Collider (LHC).

As will be detailed below, the Higgs boson can be produced and decay in many different ways. Dedicated searches are designed for the most promising of these possibilities, or *channels*, and each can be related to the same underlying physics theory. Thus, it is common for a collaboration to bring together the different searches and form a more powerful combined analysis. These combined analyses are not based on the results of the individual searches, as there is no satisfactory way to combine p-values; instead a joint statistical model is formed and tested. Forming this joint statistical model requires coordination as many systematic effects are common to the individual searches. Continuing on this logical path, different experiments at the same accelerator complex have formed combined analyses for the Higgs boson. This was first done by the four LEP experiments at CERN [1] and has also been done at the two Tevatron experiments [2]. The ATLAS and CMS collaborations have now formed an LHC Higgs Combination Group (LHC-HCG) with the goal of providing a first combined result in the summer of 2011.

This document outlines the conceptual issues involved in forming the combined statistical model across experiments, the technology developed to form such a complicated model, and the statistical tests the group is considering for the initial result. The focus will be on lessons learned from a toy combination completed in the summer of 2010 [3] and open questions the LHC-HCG is currently deliberating.

2 The probability model of a typical search channel

The SM is an impressively predictive theory formulated in a more general formalism called quantum field theory (QFT). Within QFT the fundamental mathematical object is called the Lagrangian, which encodes the particle content and the interactions in the theory with some free parameters that must be determined experimentally. All of the parameters of the SM Lagrangian have been measured, except for one: M_H, the mass of the Higgs boson itself. Once one specifies M_H – either by hypothesizing a particular value or measuring it – the SM is a fully specified theory that makes numerous predictions. In particular, the SM predicts the production rates via each of the possible Higgs production modes, branching ratios for different decay modes, and distributions of kinematic properties of the final-state particles.

The LHC is a proton-proton collider, currently with a center-of-mass energy of 7 TeV, giving it the capability to produce Higgs bosons. There are a few different types of interactions $pp \to H + X$ that can produce a Higgs boson (perhaps in association with other particles, generically referred to as X) in an individual collision. Each of these production modes, indexed by I, has an associated production

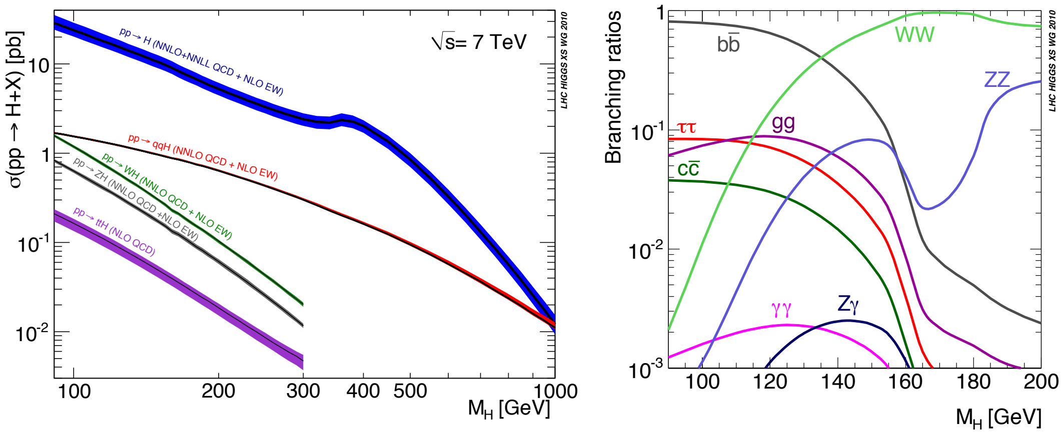

Fig. 1: Production cross-sections and branching ratios of the standard model Higgs boson taken from Ref. [4]

cross-section σ_I (in pb = picobarns = 10^{-36}cm^{-2} in the SM. Figure 1 shows the predicted cross-section in the SM for various production modes as a function of the unknown Higgs mass parameter M_H [4]. The number of collisions predicted to undergo $pp \to H + X$ is given by the product of this production cross-section and the time-integrated luminosity of the proton beams, L, (measured in 1/cm^2) which is controlled by the collider itself.

The Higgs boson is also predicted to be unstable and decay very quickly, well before coming in contact with any component of our particle detectors. The Higgs may decay in several different ways, and the relative proportions for these decays, indexed by f, are called branching ratios, denoted B_f. Figure 1 shows the dependence of the branching ratios for the various decay modes on the unknown Higgs mass parameter M_H [4]. In some cases, such as the decay $H \to ZZ$, there is a further cascade of decays before reaching the final-state particles that interact with the detector. For example, Fig. 2 shows a real collision observed by the CMS detector, that is compatible with the hypothesis of a $H \to ZZ \to \mu^+\mu^-\mu^+\mu^-$ decay. In addition to the overall rate of events like this, the SM also predicts the joint distribution of particles in *phase space*: the space that describes the energies and directions of the final state particles. The interaction of these final state particles with the detector is modeled using detailed computer simulations and reconstruction algorithms are developed to estimate the angles and energies of the particles based on the signatures left in the detector components.

Other types of interactions that may lead to the same final state particles or mimic them in the detector are referred to as backgrounds. Because the production of Higgs bosons is quite rare in comparison to backgrounds, event selection criteria are developed to reject the bulk of these background events, thus defining a *signal region* that is relatively rich in the Higgs signal events. The fraction of signal events from production mode I and decay mode f satisfying the criteria for search channel c is called the *efficiency* of the event selection, ϵ^c_{If}. Thus, the total number of Higgs boson events expected to satisfy the event selection criteria for a given channel is given by

$$s_c = \sum_{I \in \text{production}} \sum_{f \in \text{decay}} \epsilon^c_{If} L \sigma_{I,\text{sm}} B_{f,\text{sm}}. \tag{1}$$

Note that s_c is implicitly a function of the Higgs mass parameter M_H as the production cross-section, branching ratio, and efficiency all depend on M_H.

While the cross-section and branching ratio are predicted to have a specific value in the SM once M_H is specified, it is common to generalize the situation by considering events of a similar efficiency but with a modified rate. This is accomplished by introducing a new parameter $\mu = \sigma_I B_f / \sigma_{I,\text{sm}} B_{f,\text{sm}}$, where $\mu = 1$ corresponds to the presence of a SM Higgs boson, $\mu = 0$ corresponds to the background-only hypothesis, and other values would indicate a non-standard Higgs boson or possibly some other form of new physics beyond the SM. It is worth noting that in these non-SM situations there is no particular reason to assume a common μ for each production mode I and decay mode f. This common structure across production and decay modes can also be broken by theoretical uncertainties, which will be addressed below.

Of course, we also need an estimate of the background in this selection region. Some backgrounds are estimated using the same first-principles procedure based on theoretical predictions and detector simulation as is used for the signal. However, for some backgrounds this simulation-based approach is highly sensitive to the details of the interaction in the detector or rely heavily on aspects of the theory that require approximate solutions. In these cases, experimentalists prefer to use *data-driven* techniques, where one utilizes known or assumed relationships between the observations in *control regions* and the signal region. Let us refer to the estimates on the number of background events using the simulation-based and data-driven approaches as $b_{\text{sim.}}$ and $b_{\text{d.d.}}$, respectively, and assume that the contributions from those background processes are disjoint, exhaustive, and sum to a total background rate b.

At this point we can write the probability model, neglecting uncertainties, for observing N_c events in the event-selection for the c^{th}-channel based on the background estimates, the total signal expectation for a given M_H, and the parameter of interest μ as $\text{Pois}(N_c|\mu s_c + b)$.

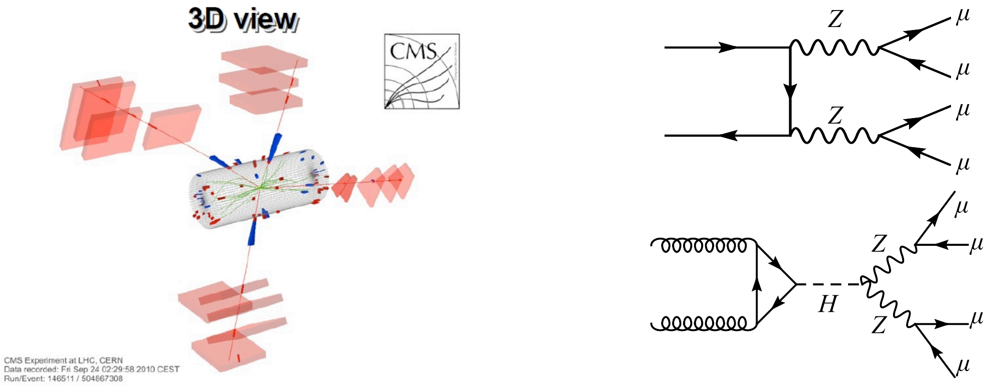

Fig. 2: An event display of a $H \to ZZ \to \mu^+\mu^-\mu^+\mu^-$ candidate event along with Feynman diagrams of compatible background and signal processes

The distribution of final state particles in phase-space can be quite complex and often have pronounced structures. For instance, a certain combination of the angles and energies of the four muons in the $H \to ZZ \to \mu^+\mu^-\mu^+\mu^-$ decay called the invariant mass is a direct estimator of M_H. In some cases, multivariate algorithms such as neural networks and boosted decision trees are used to form a discriminating variable, which may also include information on particle identification. Let us generically denote these discriminating variables, or *marks*, as x and the probability density function describing the distribution of x for a signal and background processes as $f_s(x)$ and $f_b(x)$, respectively. We can extend the simple Poisson model to include this shape information by building what statisticians refer to as a *marked Poisson* process with $\boldsymbol{x}_c = \{x_1, ..., x_{N_c}\}$:

$$P(\boldsymbol{x}_c|\mu) = \text{Pois}(N_c|\mu s_c + b) \prod_j^{N_c} \frac{\mu s_c f_s(x_j) + b f_b(x_j)}{\mu s_c + b} \, . \tag{2}$$

This is the same type of expression physicists use when performing an unbinned extended maximum likelihood fit, where the rate and shape information are both related to the parameter of interest μ.

3 Incorporating uncertainties into the model

Arguably, the most involved and delicate aspect to Higgs searches is controlling and understanding the many sources of uncertainty that modify the expected rate and distributions of signal and background processes. As a result, the model in Eq. 2 is extended to a family of models parametrized with several nuisance parameters $\boldsymbol{\nu}$: $P(\boldsymbol{x}_c|\mu) \to P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu})$. Modeling of these effects must be coordinated among the different searches as a single source of uncertainty may be common to many searches and the correlations must be taken into account. When combining searches from different experiments, the most acute correlated effects arise from the use of common theoretical tools and the luminosity of the beam.

There are systematic uncertainties associated with detector simulation and performance; statistical uncertainties associated with the auxiliary measurements used in data-driven background estimates; the residual statistical and systematic uncertainty of theoretical predictions associated with the measurement of the strong coupling constant and the parton density functions; and theoretical uncertainties that are not statistical in nature, but result from neglecting higher-order terms in the perturbative expansion of the theory. These uncertainties are quite different in nature, which is reflected in their statistical modeling.

It is helpful to think of the modeling in two steps. The first step is to parametrize the effect of the uncertainty on the primary measurement $\boldsymbol{x}_c$: $P(\boldsymbol{x}_c|\mu) \to P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu})$. The second is to incorporate additional *constraint terms* $P(\boldsymbol{y}_i|\boldsymbol{\nu})$ that describe how auxiliary measurements $\boldsymbol{y}_i = \{y_1, ..., y_{M_i}\}$ depend on the nuisance parameters. Thus, the probability model expands $P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu}) \to P(\boldsymbol{x}_c, \boldsymbol{y}|\mu, \boldsymbol{\nu})$.

Parametrization of the effect of uncertainty on the primary measurement of the marks, $P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu})$, is typically dealt with in one of two ways. In the first approach, the model is *explicitly parametrized* in terms of elemental sources of uncertainty, such as electron identification efficiency, jet energy scale uncertainty, parton density functions, etc. A single nuisance parameter ν_i parametrizes the effect of changing a single source of uncertainty, which can be estimated by running the simulation with modified settings or by correcting the simulation in some way. This approach typically requires some form of interpolation as the variation in the selection efficiency $\epsilon(\nu_i)$ and distributions $f(m; \nu_i)$ can only be estimated for discrete values of ν_i. The advantage of this approach is that it is straightforward to understand the correlated effect on the individual searches, by simply identifying the ν_i that correspond to the same source of uncertainty. This approach was used by the CMS inputs to the toy combination [5] and recent ATLAS analyses [6,7]. The second common approach is to use some parametric function that is believed to be flexible enough to capture potential variations due to the underlying sources of uncertainty as an effective model for $P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu})$. These effective models may be well-motivated by understanding of the physical processes, such as an exponentially falling distribution [8], or ad hoc choices, such as polynomials. In this approach, the effect of individual sources of uncertainty are *implicitly parametrized* by the effective model. The advantage of this approach is that the interpolation in $\boldsymbol{\nu}$ is built into the effective model; however, the disadvantage of this approach is that it is difficult to introduce the correlated effect of a specific source of uncertainty across individual searches – an issue the LHC-HCG currently faces.

Ideally, the constraint terms describe other auxiliary measurements $\boldsymbol{y}_i$ in such a way that the uncertainty in $\boldsymbol{\nu}$ can be dealt with in a clear statistical sense. A simple example is the well-studied "on/off" problem [9] in which the unknown background rate in the signal region is related to an auxiliary counting experiment via a known constant τ: $P(\boldsymbol{x}_c, \boldsymbol{y}_i|\mu, b) = \text{Pois}(N_c|\mu s_0 + b) \cdot \text{Pois}(M_i|\tau b)$, where N_c and M_i are the number of events in the main and auxiliary counting experiment and the distributions of x and y are not taken into account. An extension of the "on/off" problem in the marked Poisson model comes from using a control region in the data that is devoid of signal and has a similar distribution for the discriminating variable: $f_b(x; \boldsymbol{\nu}) = f_b(y; \boldsymbol{\nu})$. In most realistic situations, the extrapolation coefficient τ is also uncertain and the distributions in x and y are not identical. In particular, τ is often estimated from simulations and is subject to both experimental and theoretical uncertainties.

It is common that the uncertainty on a nuisance parameter can be estimated from an auxiliary measurement or from experience, but an explicit probability model relating $\boldsymbol{y}_i$ and ν_i is not available for practical reasons. In these cases, it is common to idealize the situation and choose an ad hoc constraint term that summarizes the auxiliary measurement or captures intuition about the uncertainty in the nuisance parameter. For example, Gaussian constraint terms are very common idealizations of auxiliary measurements. Here Bayesian reasoning is deceptively natural as one often refers to the prior $\pi(\nu_i)$ in informal conversation without recognizing a Bayesian probability inversion. Similarly, one often refers to a "Gamma" prior on a nuisance parameter, which can be interpreted as the posterior resulting from an idealized auxiliary counting experiment with a uniform prior via Bayes theorem: $\pi(\nu_i) \propto \mathrm{Pois}(M|\tau\nu_i) \cdot \mathrm{Uniform}(\nu_i)$. In order use a consistent probability model in both frequentist and Bayesian statistical formalisms, it is important to incorporate the Poisson term into the probability model and separate the original uniform prior $\eta(\nu)$ for Bayesian techniques. Another popular form for an ad hoc constraint term is the log-normal distribution, particularly for non-negative nuisance parameters with large relative uncertainty (>20%). In this case, one must be more careful about what is assumed to be log-normally distributed. If one assumes the observable y in the auxiliary measurement is log-normally distributed (as is implied when invoking multiplicative measurement errors) and uses a uniform prior on ν_i (as in the more familiar Gaussian and Gamma case), then the posterior $\pi(\nu_i)$ does not have a log-normal form. On the other hand, if one means that the posterior is log-normally distributed, then the likelihood function and prior must be specified to provide a consistent frequentist treatment of the problem. While a $\eta(\nu_i) \propto 1/\nu_i$ prior allows both the PDF and the posterior to have a log-normal form, the likelihood function and the posterior are no longer proportional (as they were in the Gaussian and Gamma case). The lesson here is that one cannot simply appeal to idealized measurements and hope for an unambiguous interpretation when there are large uncertainties involved.

3.1 Forming the Combined Model and Technical Implementation with RooFit/RooStats

Once one has prepared the model for each individual search channel together with the associated auxiliary measurements and constraint terms $P(\boldsymbol{x}_c, \boldsymbol{y}|\mu, \boldsymbol{\nu})$, the combined model can be formed by multiplying the individual terms, identifying the common parameters, and perhaps introducing additional terms that would impose non-trivial correlations or functional relationships among the parameters. This requires that the selection regions are disjoint and that the common parameters not only parametrize the same effect, but also have the same conventions. We refer to the combined model by $P(\boldsymbol{x}, \boldsymbol{y}|\mu, \boldsymbol{\nu})$, dropping the c and i subscripts for the individual terms: $\boldsymbol{x} = \{\boldsymbol{x}_c\}$ and $\boldsymbol{y} = \{\boldsymbol{y}_i\}$.

$$P(\boldsymbol{x}, \boldsymbol{y}|\mu, \boldsymbol{\nu}) = \prod_{c \in \mathrm{channel}} P(\boldsymbol{x}_c|\mu, \boldsymbol{\nu}) \prod_{i \in \mathrm{aux.meas.}} P(\boldsymbol{y}_i|\boldsymbol{\nu}) \tag{3}$$

Since the PhyStat conference in 2007, there has been a dedicated effort to develop technologies capable of creating and communicating, and testing complex probability models within the context of the related ROOT projects RooFit and RooStats [10–12]. The `RooWorkspace` class utilizes the ROOT I/O technology to save these complicated models into a persistent file, which can be shared easily. Much of the effort has been to cleanly separate and organize the information needed for the various statistical tests: in particular the `ModelConfig` class keeps track of the PDF $P(\boldsymbol{x}, \boldsymbol{y}|\mu, \boldsymbol{\nu})$, the priors $\eta(\mu)$ and $\eta(\boldsymbol{\nu})$, the parameter of interest μ, the nuisance parameters $\boldsymbol{\nu}$, the observables $\boldsymbol{x}$, and the auxiliary observables $\boldsymbol{y}$.

In the summer of 2010 the ATLAS and CMS collaborations embarked on a toy combination exercise with unofficial, though realistic mock-ups of the searches for $H \to W^+W^- \to l^+l^-\nu\bar{\nu} + 0$ jets with $L = 1$ fb^{-1}, where $l = e, \mu$ [3]. The ATLAS model was based on counting events in the signal region and three control-regions for each of the three $ee, e\mu$, and $\mu\mu$ final states [13]. This setup is very similar to the "on/off" problem where the background rate is a nuisance parameter and auxiliary counting measurements are made explicit; however, here there were three τ coefficients for the different control regions as well as coefficients due to cross-contamination of the different background processes in the individual control regions. The extrapolation coefficients also had large uncertainty that was represented by a Gaussian constraint term (truncated for $\tau < 0$). Unfortunately, the variance of this Gaussian was given by the sum in quadrature of the variation in τ due to the individual sources of uncertainty, making it impossible to identify the effect of theoretical of uncertainties common to both ATLAS and CMS. In contrast, the CMS model did not incorporate actual auxiliary measurements into the model, but used log-normal constraint terms for 37 nuisance parameters that explicitly parametrized the effect of 37 sources of uncertainty [5]. A visualization of the combined model in terms of a directed acyclic graph is shown in Fig. 3. The models that the LHC-HCG group are considering now are drastically more complex.

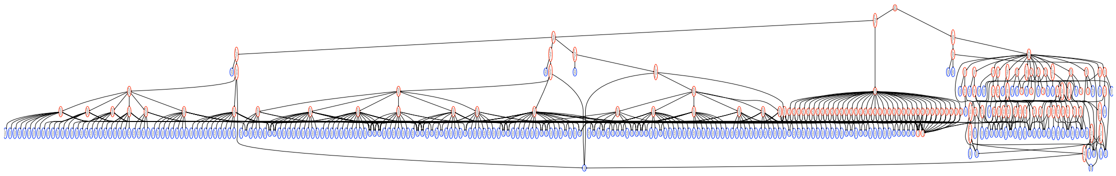

Fig. 3: Visualization of the combined model for the toy ATLAS and CMS Higgs combination performed in the summer of 2010. The top node represents the the likelihood, the left portion of the graph represents the CMS model, the right portion represents the ATLAS model, and the lowest node in the middle of the graph represents μ.

3.2 Statistical Tests

The emphasis on cleanly separating the objective PDF from the Bayesian prior is largely motivated by the desire to retain flexibility in the type of statistical tests that can be used. In particular, the strategy has been to put effort into a single probability model and then consider different statistical procedures. The RooStats framework has implementations of most of the commonly used statistical procedures, including Bayesian methods based on Markov-Chain Monte Carlo, fully frequentist methods based on the Neyman-Construction and hybrid-resampling (also referred to as the 'profile construction') [10, 14–16], likelihood-based methods that utilize the asymptotic results of Wilks and Wald together with numerical procedures for estimating the non-centrality parameter [17–19], as well as mixed Bayesian-Frequentist procedures [20, 21] and CL_s [22, 23].

In the Bayesian realm, the hope is that the auxiliary measurements or idealized constraint terms represented by $P(\boldsymbol{y}|\boldsymbol{\nu})$ are sufficiently informative to dominate the priors on the nuisance parameter $\eta(\boldsymbol{\nu})$, though this has not been studied in much detail. Instead, physicists are more keenly aware of the sensitivity to the prior on the parameter of interest $\eta(\mu)$. While uniform priors on μ reign supreme, there is interest in the use of Jeffreys prior and reference priors [24–26]. Recently, progress has been made in estimating these priors directly [27] and the efficient numerical techniques for calculating the Fisher information matrix, which is a necessary ingredient [19].

On the frequentist side, the emphasis has been on the choice of the test statistic and the details of the ensemble used to compute p-values. At LEP, systematic uncertainties were small and the test statistic was the simple likelihood ratio $Q_{LEP}(\boldsymbol{x}) = P(\boldsymbol{x}|\mu\!=\!1)/P(\boldsymbol{x}|\mu\!=\!0)$. At the Tevatron uncertainties are larger and a profiled generalization of the LEP test statistic has been used

$$Q_{\text{Tev}}(\boldsymbol{x},\boldsymbol{y}) = \frac{P(\boldsymbol{x},\boldsymbol{y}\,|\,\mu\!=\!0,\hat{\hat{\boldsymbol{\nu}}}(\mu\!=\!0;\boldsymbol{x},\boldsymbol{y}))}{P(\boldsymbol{x},\boldsymbol{y}\,|\,\mu\!=\!1,\hat{\hat{\boldsymbol{\nu}}}(\mu\!=\!1;\boldsymbol{x},\boldsymbol{y}))} = \frac{\lambda(\mu\!=\!0;\boldsymbol{x},\boldsymbol{y})}{\lambda(\mu\!=\!1;\boldsymbol{x},\boldsymbol{y})}\,, \tag{4}$$

where $\lambda(\mu;\boldsymbol{x},\boldsymbol{y})$ is the profile-likelihood ratio

$$\lambda(\mu;\boldsymbol{x},\boldsymbol{y}) = \frac{P(\boldsymbol{x},\boldsymbol{y}\,|\,\mu,\hat{\hat{\boldsymbol{\nu}}}(\mu;\boldsymbol{x},\boldsymbol{y}))}{P(\boldsymbol{x},\boldsymbol{y}\,|\,\hat{\mu}(\boldsymbol{x},\boldsymbol{y}),\hat{\boldsymbol{\nu}}(\boldsymbol{x},\boldsymbol{y}))}\,. \tag{5}$$

Our field has known for many years that the distribution $f(-2\log\lambda(\mu)|\mu,\boldsymbol{\nu})$ is asymptotically related to a chi-square distribution and independent of $\boldsymbol{\nu}$. However, we have only recently appreciated that

$f(-2\log\lambda(\mu)|\mu',\boldsymbol{\nu})$ is asymptotically a non-central chi-square distribution with a non-centrality parameter that depends on μ' and $\boldsymbol{\nu}$. As a result, $f(Q_{\rm TeV}|\mu=1$ or $\mu=0,\boldsymbol{\nu})$ necessarily depend on the true values of the nuisance parameters [19]. For this reason, there is growing support in the LHC-HCG to move to $\lambda(\mu)$ for the test statistic, perhaps restricted to a one-sided alternative. There has also been some progress in understanding how the the look-elsewhere effect modifies the distribution of the test statistic as M_H is a free parameter that cannot be identified in the $\mu=0$ hypothesis [28].

Another area of development that can be compared to the Tevatron procedure is the precise way in which the ensembles are generated. The Tevatron Higgs combination group primarily uses a mixed Bayesian-Frequentist procedure in which the nuisance parameters are marginalized with respect to $\pi(\boldsymbol{\nu})$ in the process of generating pseduo-experiments [20, 21]. The fully frequentist procedure emerging at the LHC is based on the hybrid-resampling procedure [10, 14–16], in which the distribution $f(-2\log\lambda(\mu)|\mu',\hat{\hat{\boldsymbol{\nu}}}(\mu';\boldsymbol{x},\boldsymbol{y}))$ is constructed at a particular value of the nuisance parameter expected to be most relevant given the data. This approach has been used by ATLAS in its first Higgs results [6,7,29] and Fig. 4 is a demonstration in the context of discovery with $\mathcal{O}(10^7)$ pseudo-experiments and a model of realistic complexity. As a result of going to this fully frequentist approach, the ensemble includes both variations in $\boldsymbol{x}$ as well as $\boldsymbol{y}$. The presence of variations in $\boldsymbol{y}$ breaks the discreteness in the test statistic and modifies familiar rules of thumb when the observed count N is zero and expected background rate b is small. This is relevant in cases such as $H \to ZZ \to 4l$ [7] and gives results similar to the Lancaster's mid-P [30].

A new element to the discussion is the potential for conditioning. In the simplest situation represented by the "on/off" problem Pois$(N|\mu s+b)$ Pois$(M|\tau b)$, the hypothesis test of $\mu = 0$ can be reformulated in terms of the ratio of Poisson means $\beta = (\mu s + b)/\tau b$ where it is clear that the total count $N+M$ has no information on the ratio [31–34]. However, in the context of confidence intervals on μ this conditioning is not appropriate as the total count does carry information on the magnitude of μ. Thus, it is not yet clear to the LHC-HCG if there is an appropriate conditioning procedure in this context.

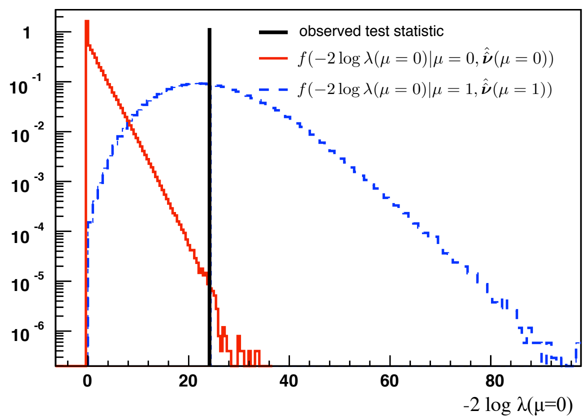

Fig. 4: An example distribution of $-2\log\lambda(\mu = 0)$ evaluated $\sim 10^7$ pseudo-experiments for background-only and signal plus background hypotheses. Evaluating $-2\log\lambda(\mu = 0)$ requires two fits to the full model, which typically has $\mathcal{O}(50)$ nuisance parameters. This requires batch or PROOF enabled computing clusters.

4 Conclusions

The LHC is performing exceptionally well breaking records in both energy and intensity at a hadron collider. Sensitivity studies suggest that within the next year or two ATLAS and CMS will be in the position to make very strong statements about the existence or non-existence of a SM Higgs boson. The LHC-HCG is aiming to show the first combined results from ATLAS and CMS Higgs searches in the summer of 2011 – roughly two months from the writing of this document.

Acknowledgements

The author would like to thank his colleagues in the LHC-HCG and the statistics forums and Higgs groups of ATLAS and CMS. This work was supported by NSF grants PHY-0854724 and PHY-0955626.

References

[1] R. Barate et al. *Phys.Lett.*, B565:61–75, 2003.

[2] The CDF and D0 Collaborations. 2011. arxiv:1103.3233.

[3] The ATLAS and CMS Collaborations. Jul. 2010. `http://indico.cern.ch/conferenceDisplay.py?confId=100458`.

[4] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). *CERN-2011-002*, 2011. [arxiv:1101.0593].

[5] W. Quayle. Jul. 2010. `http://indico.cern.ch/conferenceDisplay.py?confId=100458`.

[6] The ATLAS Collaboration. Mar 2011. ATLAS-CONF-2011-026.

[7] The ATLAS Collaboration. Mar 2011. ATLAS-CONF-2011-048.

[8] (ATLAS-CONF-2011-004), Feb 2011.

[9] R. D. Cousins, J. T. Linnemann, and J. Tucker. *Nucl.Instrum.Meth.*, A595:480–501, 2008.

[10] K. Cranmer. *PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics*, 2008. oai:cds.cern.ch:1021125. `http://cdsweb.cern.ch/record/1099969`.

[11] W. Verkerke. *PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics*, 2008. oai:cds.cern.ch:1021125. `http://cdsweb.cern.ch/record/1099988`.

[12] L. Moneta, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, et al. *PoS*, ACAT2010:057, 2010. [arxiv:1009.1003].

[13] The ATLAS Collaboration. Jul 2010. ATL-PHYS-PUB-2010-009.

[14] G. Feldman. 2000. Talk at the FermiLab Workshop on Confidence Limits.

[15] C. Chuang and T. L. Lai. *Statist. Sinica*, 10:1–50, 2000. `http://www3.stat.sinica.edu.tw/statistica/oldpdf/A10n11.pdf`.

[16] M. Walker B. Sen and M. Woodroofe. *Statist. Sinica*, 19:301–314., 2009. `http://www3.stat.sinica.edu.tw/statistica/oldpdf/A19n116.pdf`.

[17] S.S. Wilks. *Ann. Math. Statist.*, 9:60–2, 1938.

[18] A. Wald. *Transactions of the American Mathematical Society*, 54, No. 3, 1943.

[19] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. *Eur.Phys.J.*, C71:1554, 2011.

[20] R. D. Cousins and V. L. Highland. *Nucl.Instrum.Meth.*, A320:331–335, 1992. Revised version.

[21] T. Junk. *Nucl.Instrum.Meth.*, A434:435–443, 1999.

[22] A. L. Read. *J. Phys. G: Nucl. Part. Phys.*, 28, 2002.

[23] A. L. Read. *in Proceedings of the First Workshop on Confidence Limits, CERN*, 2000.

[24] J. M. Bernardo. *J. R. Statist. Soc. B*, 41:113, 1979.

[25] J. O. Berger and J. M. Bernardo. *J. Amer. Statist. Assoc.*, 84:200, 1989.

[26] J. O. Berger and J. M. Bernardo. *Biometrika*, 79:25, 1992.

[27] L Demortier, S. Jain, and H. B. Prosper. *Phys.Rev.*, D82:034002, 2010.

[28] E. Gross and O. Vitells. *Eur. Phys. Jour. C*, 70:525–530.

[29] The ATLAS Collaboration. Feb 2011. ATLAS-CONF-2011-005.

[30] H.O. Lancaster. *Biometrika*, 39:419–422, 1949.

[31] J. Przyborowski and H. Wilenski. *Biometrika*, 31:313, 1940.

[32] F. James and M. Roos. *Nuclear Physics B*, 172:475, 1980.

[33] N. Reid. *Stat. Sci.*, 10:138, 1995.

[34] R. D. Cousins, K. E. Hymes, and J. Tucker. *Nucl.Instrum.Meth.*, A612:388–398, 2010.