PROCEEDINGS OF SCIENCE

Statistical Issues for Higgs Physics

Eilam Gross[1]
Weizmann Institute of Science
Rehovot, 76100, ISRAEL
E-mail: `eilam.gross@weizmann.ac.il`

Ofer Vitells
Weizmann Institute of Science
Rehovot, 76100, ISRAEL
E-mail: `ofer.vitells@weizmann.ac.il`

How should we interpret discovery sensitivity plots and exclusion plots? What are the expected limits vs the actual ones? Was the look elsewhere effect taken into account? Why do Bayesian and frequentist approaches sometimes give very different answers? In a short review we will try to answer these and other questions that some of us wanted to ask but wouldn't dare.

The 2009 Europhysics Conference on High Energy Physics,
Krakow , Poland
July 16-22, 2009

---

[1] Speaker

1. Introduction

Higgs statistics is about testing one hypothesis against another hypothesis. For example, one hypothesis is the Standard Model (**SM**) with no Higgs Boson (let us denote this hypothesis by H_0). The other hypothesis is the SM with a Higgs Boson with a specific mass m_H (H_1). The other hypothesis could also be a SM Higgs boson at some unspecified mass.

Rejecting the No-Higgs (H_0) hypothesis is considered to be a discovery, while rejecting the Higgs hypothesis (H_1) is considered exclusion.

In order to explain the different methods for testing a hypothesis we have generated toy Monte Carlo (**MC**) simulations of data with and without the Higgs signal.

2. The Discovery Case

2.1 The Toy Model

We assume a Gaussian higgs signal (**s**) on top of a Rayleigh shaped background (**b**) Let μ denote the signal strength with the expectation of the data being $\langle n \rangle = \mu s + b$[1]. μ =1 for the SM higgs (H_1), and μ =0 for background only (H_0). The background is either known from MC with some systematic uncertainty or could be measured with a control data sample. Since any uncertainty on the predicted background can be taken into account by using it as an artificial measurement, we assume two hypothetical measurements; one for the data which might include the Higgs boson, i.e. $n \sim \mu s + b$ depending on nature, and another one using a data control sample which contains no signal, $m \sim \tau b$ with τ being the scaling factor between the prospective signal and the background-only regions (see Figure 1).

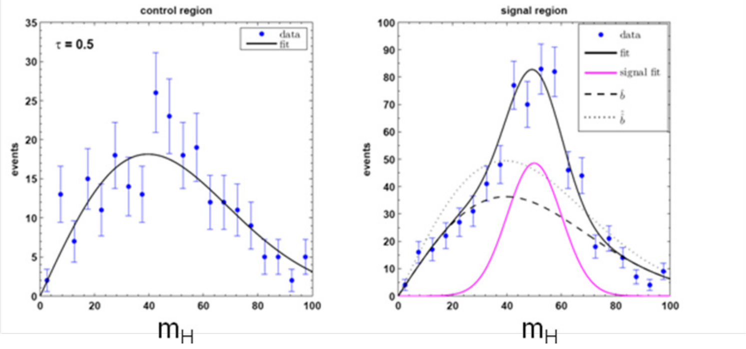

Figure 1: The Rayleigh shaped background control sample (left) and the simulated data containing a Gaussian signal (right). $\hat{b}$ *and* $\hat{\hat{b}}$ *are MLEs explained in the text. In this example the assumed "Higgs mass" is 50.*

In this non-trivial toy model the data sets are binned in histograms. The background is described by one parameter, θ, i.e. $b(\theta)$. The parameter θ determines the Rayleigh

---

[1] Note that we use a simplified notation for the data, n, and the signal and background, s and b. Each one is actually a binned distribution.

distribution. Although, the histograms are binned, we use a symbolic notation. The Likelihood for $s+b(\theta)$ is given by

$$L\left(n,m\,|\,s+b(\theta)\right)=\prod_{i=1}^{nbins}Poisson\left(n_i;s_i+b_i(\theta)\right)\cdot Poisson\left(m_i;\tau b_i(\theta)\right) \qquad (1.1)$$

The Likelihood can be maximized with respect to θ. The Maximum Likelihood Estimator (**MLE**) for the background is denoted by $\hat{b}\equiv b(\hat{\theta})$, while the conditional MLE, under the assumption of No-Signal (μ=0) is denoted by $\hat{\hat{b}}\equiv b(\hat{\hat{\theta}})$.

2.1.1 Systematics or Nuisance Parameters

The background, *b(θ)*, has an uncertainty which has to be taken into account via a real or virtual measurement. θ is called a nuisance parameter (which we associate with background systematics). The simplest way to include the systematics is profiling via MLEs (see 2.2.3). A popular way is using marginalization known as the Cousins-Highland hybrid way[1]. The Likelihood (1.1) is integrated in a Bayesian manner using a flat prior in θ, $\pi(\theta)$.

$$L(s+b)=\int L\left(n,m\,|\,s+b(\theta)\right)\pi(\theta)d\theta;\quad L(b)=\int L\left(n,m\,|\,b(\theta)\right)\pi(\theta)\,d\theta \qquad (1.2)$$

2.2 The Likelihood Ratio (LR) CL$_b$ method.

2.2.1 The Neyman-Pearson Lemma

The basis for the Likelihood Ratio method is the Neyman-Pearson lemma[2]. It states (in a layman's language) that when performing a hypothesis test between two simple hypotheses, H$_0$ and H$_1$, the LR test, which rejects H$_0$ in favor of H$_1$ is the most powerful test. The test statistic is therefore defined as $\Lambda=\frac{L(H_1)}{L(H_0)}\equiv\frac{L(H_1\,|\,x)}{L(H_0\,|\,x)}$ where x symbolizes the data. Note that Λ is data dependent even though it is usually not specified.

2.2.2 The Frequentist LR (CL$_b$) Method

Following the above discussion, a test statistic is defined as

$$\Lambda=\frac{L(H_1)}{L(H_0)}=\frac{L\left(s+b\right)}{L\left(b\right)} \qquad (1.3)$$

One now generates the pdf of Λ under H$_0$ (b-only) and H$_1$ (s+b). If Λ_{obs} is the result of one experiment (by experiment we mean e.g. the LHC or the TEVATRON), one calculates the probability to get an observation which is less b-like than the observed one. This probability is called the p-value. Figure 2 illustrates the pdf distributions and the p-value.

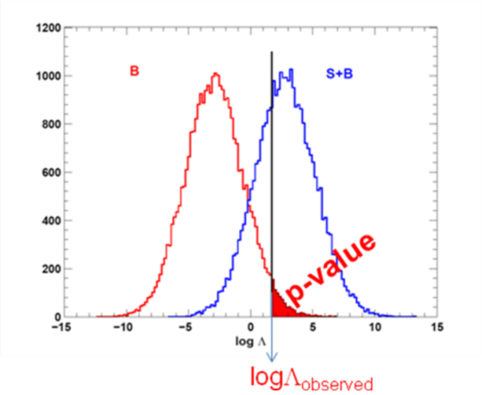

Figure 2: The pdf of the LR distribution under H_0 (red) and H_1 (blue) with the LR test statistic, Λ. The red area is the p-value corresponding to Λ_{obs}.

If the result of the experiment (LHC) yields a p-value< $2.8\cdot 10^{-7}$ a 5σ discovery is claimed (this is a convention motivated by Gaussian shaped pdf)[2].

Note that the p-value can be interpreted as a frequency; hence, this is a frequentist approach.

2.2.3 The Profiled LR CL$_b$ way

The background systematics can be taken into account by either marginalization of the nuisance parameter θ (see Eq. (1.2)) or by profiling.

Let $\hat{\hat{\theta}}_{s+b}$ be the MLE of θ under H$_1$. Let $\hat{\theta}_b$ be the MLE of θ under H$_0$.
The test statistic is defined as

$$\Lambda_{PL} = \frac{L(H_1)}{L(H_0)} = \frac{L\left(n,m \mid s+b(\hat{\hat{\theta}}_{s+b})\right)}{L\left(n,m \mid b(\hat{\theta}_b)\right)} \tag{1.4}$$

We can now estimate the expected sensitivity of an experiment by generating MC experiments and calculate the p-value equivalent significance for each one. The expected sensitivity is defined as the median of the significance distribution (see Figure 3). This sensitivity can be approximated by using one representative data set, called the **Asimov** data set[3][3] which is the expected s+b (for the data measurement, set $n = s+b$) and the expected control sample measurement ($m = \tau b$) with no statistical fluctuations.

---

[2] In LEP days this p-value was wrongly expressed in terms of the background confidence level (**CL**), i.e. p=p$_0$=1-CL$_b$.

[3] The name of the Asimov data set is inspired by the short story Franchise, by Isaac Asimov. In it, elections are held by selecting a single voter to represent the entire electorate.

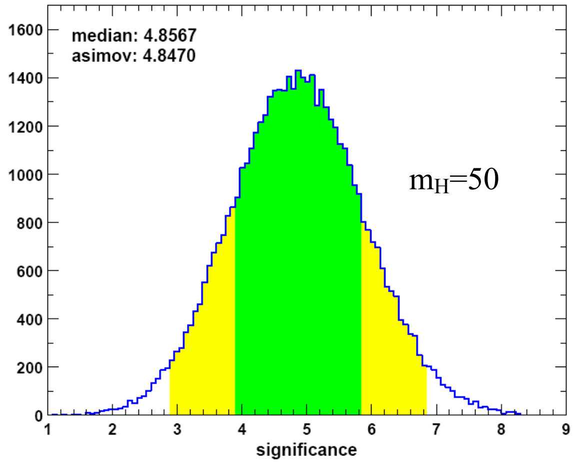

Figure 3: The distribution of the significances obtained with s+b MC experiments in the profile LR CL_b method. The green and yellow bands indicate the 1σ and 2σ significance bands. The median and Asimov significances are indicated. Note, the toy signal is assumed to be with a mass of 50 GeV (m_H=50).

The significance can be calculated for each assumed fixed Higgs mass. Figure 4 shows the sensitivity of the experiment as a function of the Higgs mass. One can see that in this experiment the Higgs is expected to be discovered (significance>5σ) if its mass is m_H<32 or m_H>52. To clarify the meaning of this, the correct statement is that if the Higgs is in the above quoted mass ranges, it will be discovered in 50% of hypothetical experiments (i.e. there is a 50% chance of claiming a 5σ discovery claim).

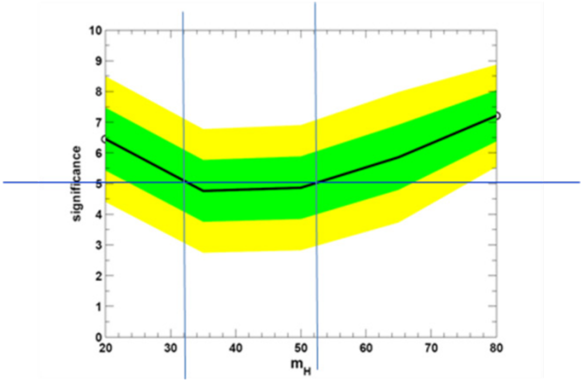

Figure 4: The expected significance as a function of the Higgs mass. The green and yellow bands are the 1σ and 2σ bands.

2.2.4 The Profile Likelihood Ratio method (PLR)

While likelihood ratio defined in Equation (1.4) tested the s+b hypothesis with respect to the b-hypothesis, the PL ratio method[4] tests the H_μ hypothesis with respect to the best hypothesis, preferred by the data, i.e. the one with $\mu = \hat{\mu}$. So we test H_1 (or H_0) with respect to $H_{\hat{\mu}}$, and let the data determines the best alternate hypothesis.

Therefore we test the H_μ hypothesis using the following test statistic

$$q_\mu = -2\ln\lambda(\mu); \quad \lambda(\mu) = \frac{L\left(\mu\cdot s + b(\hat{\hat{\theta}}(\mu))\right)}{L\left(\hat{\mu}\cdot s + b(\hat{\theta})\right)} \tag{1.5}$$

$\hat{\hat{\theta}}(\mu)$ and $\hat{\theta}$ are the MLEs defined in section 2.2.3. The presence of the nuisance parameter (θ) broadens the profile likelihood ratio as a function of μ relative to what one would have if their values were fixed. This reflects the loss of information about μ due to the systematic uncertainties.

In order to establish a discovery we try to reject the H_0 hypothesis. The test statistic is therefore $q_0 = -2\ln\lambda(0)$.

We can follow the same procedure of toy MCs and derive the p-values. The resulting discovery sensitivity is shown in Figure 5 (dash line).

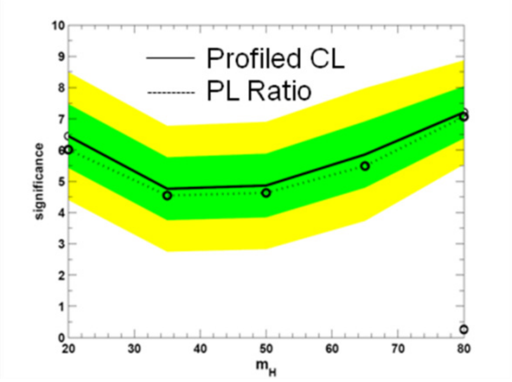

Figure 5: The discovery sensitivity of the PL ratio (dash line) compared with that of the LR CL_b method (full line).

It is compared with the Profiled CL_b sensitivity (full line), and shows slightly less sensitivity. We believe that even though the Neyman Pearson Lemma does not fully apply here (since we are not dealing with simple hypotheses) it suggests that the LR Profiled CL_b (Eq.(1.4)) may still be close to optimal.

So why use a method which is a bit less sensitive? The reason is its unique characteristics expressed by Wilks theorem[5]: Under a set of regularity conditions and for a sufficiently large data sample, the pdf of the statistic $q_0 = -2\ln\lambda(\mu)$ approaches the chi-square pdf for one degree of freedom for a hypothesized value of μ. In particular the pdf of q_0 approaches a chi-square distribution with one d.o.f. for b-only experiments (H_0). So in order to calculate a p-value of some q_{obs} one does not need to perform over 10^9 toy MC experiments, but simply calculate the significance Z given by the following relationship

$$Z_{obs} = \sqrt{q_{obs}} = \sqrt{-2\ln\lambda(\mu = 0 \mid x_{obs})} \qquad (1.6)$$

Moreover, to calculate the median significance one generates the Asimov data set $x_A = s + b$ and calculate Z_A in a straightforward manner. This can save hundreds of hours of computing time.

Figure 6 shows the pdf of the PL discovery test statistic for b-only experiments (blue), which indeed approaches a chi squared, for s+b experiments (red), and compares the median with the Asimov expectation. The corresponding expected significance in this toy example is $Z_{med} = Z_A = \sqrt{q_A} = \sqrt{-2\ln\lambda(0 \mid x_A)} = \sqrt{27} = 5.2\sigma$.

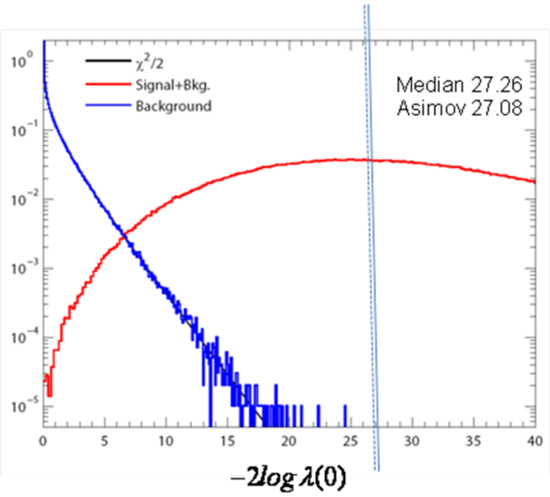

Figure 6: the pdf of the PL discovery test statistic for b-only experiments (blue), for s+b experiments (red). A chi squared line is also drawn and the median and Asimov sensitivities are indicated.

2.2.5 The Bayes Factor and the Frequentist significance

In a Bayesian approach one tests a hypothesis based on the one experimental observation using priors. Some consider priors to be the weakness of Bayesian statistics.

The posterior for the hypothesis H_i is given by

$$P(H_i \mid x) = \frac{L(x \mid H_i)\pi(H_i)}{P(x)} \tag{1.7}$$

where x is the data, and $\pi(H_i)$ is the prior for the hypothesis H_i. To claim an evidence of H_1 over H_0 (a discovery) one defines the Bayes factor B_{10} [6] as the ratio of the posteriors to priors odds, i.e.

$$B_{10} = \frac{P(H_1 \mid x)/P(H_0 \mid x)}{P(H_1)/P(H_0)} \, . \tag{1.8}$$

Integrating over the nuisance parameters with priors $\pi(\theta)$ we find

$$B_{10}(\mu) = \frac{\int L\left(n,m \mid \mu\cdot s + b(\theta)\right)\pi(\theta)d\theta}{\int L\left(n,m \mid b(\theta)\right)\pi(\theta)\,d\theta} \tag{1.9}$$

and integrating over the signal strength μ with a prior $\pi(\mu)$ we find that the Bayes factor is given by

$$B_{10} = \frac{\iint L\left(n,m \mid \mu\cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)d\theta d\mu}{\int L\left(n,m \mid b(\theta)\right)\pi(\theta)\,d\theta} \tag{1.10}$$

We therefore find that **for flat priors**, $\pi(\theta)$, which are very common (not necessarily justified) in HEP,

$$B_{10} = \frac{\int d\mu\,\pi(\mu)\int L\left(n,m \mid \mu\cdot s + b(\theta)\right)d\theta}{\int L\left(n,m \mid b(\theta)\right)d\theta} \tag{1.11}$$

One can show, using the saddle point approximation, that for a flat prior in θ

$$B_{10}(\mu)=\frac{\int L\big(\mu\cdot s+b(\theta)\big)d\theta}{\int L\big((\mu=0)\cdot s+b(\theta)\big)d\theta}=\frac{\int e^{\log L(\mu\cdot s+b(\theta))}d\theta}{\int e^{\log L((\mu=0)\cdot s+b(\theta))}d\theta}\approx\frac{e^{\log L\left(\mu\cdot s+b\left(\hat{\hat{\theta}}(\mu)\right)\right)}}{e^{\log L\left((\mu=0)\cdot s+b\left(\hat{\hat{\theta}}(\mu=0)\right)\right)}}=\frac{\lambda(\mu)}{\lambda(0)} \tag{1.12}$$

This relationship between the Bayes factor and the frequentist PL ratio, though disturbing in the first place, is not surprising when you come to think about it. Wilks theorem ensures that using the PL ratio you do not need to perform any toy MC experiments to tell a significance of an observation based on the one observed data set. This is also the characteristics of a Bayesian hypothesis test. Actually we have tested Eq. (1.12) on our assumed Higgs sample and found out an extraordinary agreement (Figure 7)

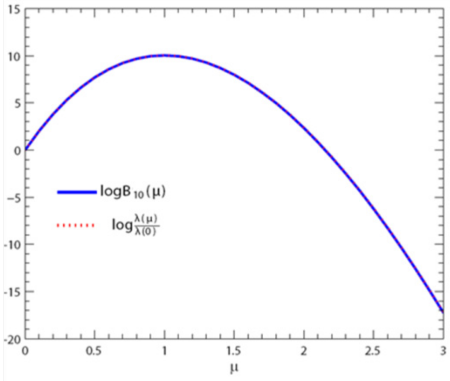

Figure 7: The Bayes factor $B_{10}(\mu)$ (blue) and the PL ratio (red). The agreement is perfect.

It is now straightforward to show that with a flat prior in θ

$$\begin{aligned}B_{10}&=\int B_{10}(\mu)\pi(\mu)d\mu\approx\frac{1}{\lambda(0)}\int\lambda(\mu)\pi(\mu)d\mu\\ \log B_{10}&\approx-\log\lambda(0)+\log\int\lambda(\mu)\pi(\mu)d\mu\approx-\log\lambda(0)+C\end{aligned} \tag{1.13}$$

To evaluate C we can plug in $\lambda(\mu)\approx e^{-\frac{(\mu-\hat{\mu})^2}{2\sigma_\mu{}^2}}$. The median significance is obtained with the Asimov s+b data, i.e. $\hat{\mu}=1$. For this data set we find

$$\lambda(0\,|\,\hat{\mu}=1)\approx e^{-\frac{1}{2\sigma_\mu{}^2}}\Rightarrow\sigma_\mu\approx\frac{1}{\sqrt{-2\ln\lambda(0\,|\,\hat{\mu}=1)}}=\frac{1}{Z_A} \tag{1.14}$$

If we take an improper flat prior for μ, $\pi(\mu)=1$ we can approximate the integral (for Z>1) to be

$$C=\ln\int_0^\infty\lambda(\mu)\pi(\mu)d\mu\approx\ln\sqrt{2\pi}\sigma_\mu=\ln\frac{\sqrt{2\pi}}{Z_A} \tag{1.15}$$

and therefore for the Asimov s+b deta we get that the expected Bayes factor is related to the frequentist significance via

$$\ln B_{10} \approx \frac{Z_A{}^2}{2} + \ln\frac{\sqrt{2\pi}}{Z_A} \tag{1.16}$$

Eq. (1.16) can be used to relate the Bayes factor to the frequentist significance. For example, the 3σ observation and 5σ discovery are equivalent to a Bayes factor of $B_{10} \sim 100$ and $B_{10} \sim 30000$ which are traditionally considered by Bayesians as an evidence and a very strong evidence (discovery).

2.2.6 The Look Elsewhere Effect or The Floating Higgs Mass Case

In all the examples shown so far, the alternate hypothesis was a Higgs Boson with a fixed mass m_H. One could also pose a different alternate hypothesis, i.e. a Higgs boson with some mass m_H in a given mass range.

The sensitivity will be reduced because the probability for the background to fluctuate and mimic a signal anywhere in the mass range (“elsewhere”) is much bigger than its probability for such a fluctuation at a given mass.

To quantify the look elsewhere effect one defines the trial factor which is the ratio of the p-values of the floating mass hypothesis and the fixed Higgs mass hypothesis, i.e.

$$trial\,\# = \frac{p_{float}}{p_{fix}} \tag{1.17}$$

It is tempting to estimate the trial factor as the ratio between the search range and the mass resolution (Γ / σ_{m_H}), but one has to be careful not to do it. It is wrong.

To calculate the p-value under the floating mass hypothesis one can use the PL ratio with two parameters of interest, the signal strength μ, and the Higgs mass m_H,

$$\lambda(\mu, m_H) = \frac{L\left(\mu \cdot s(m_H) + b(\hat{\hat{\theta}}_{(\mu=0)})\right)}{L\left(\hat{\mu} \cdot s(\hat{m}_H) + b(\hat{\theta})\right)}; \quad \lambda(\mu = 0) = \frac{L\left(b(\hat{\hat{\theta}}_{(\mu=0)})\right)}{L\left(\hat{\mu} \cdot s(\hat{m}_H) + b(\hat{\theta})\right)} \tag{1.18}$$

Strictly speaking, Wilks theorem does not apply here. However we have seen that the test statistic pdf under b-only experiments (H_0), distributes approximately as a chi squared with two degrees of freedom. This can be seen in Figure 8 below.

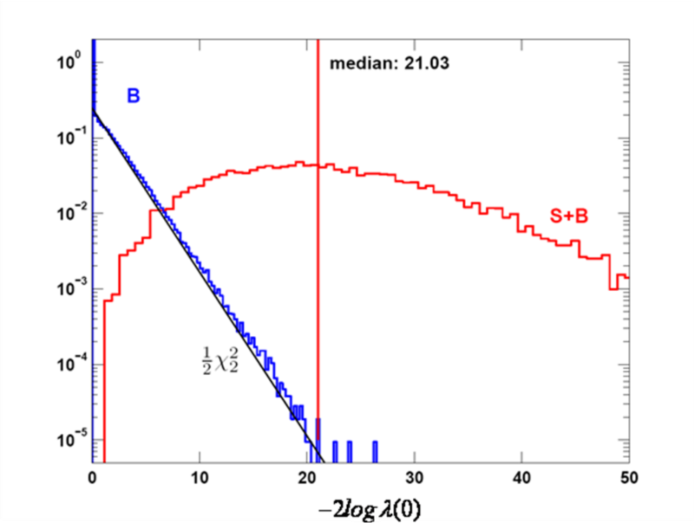

Figure 8: The pdf of the discovery test statistic under b-ony experiments (blue) and under s+b experiments (red). Also shown is a Chi squared distribution with two degrees of freedom and the median sensitivity.

A comparison between the fixed and floating mass sensitivities and the trial factor as a function of the Higgs mass is shown in Figure 9. Note the reduced sensitivity of the floating mass sensitivity due to the look elsewhere effect. Note also that the trial factor cannot be estimated by the above mentioned rule of thumb ($trial\,\# \approx \Gamma / \sigma_{m_H}$).

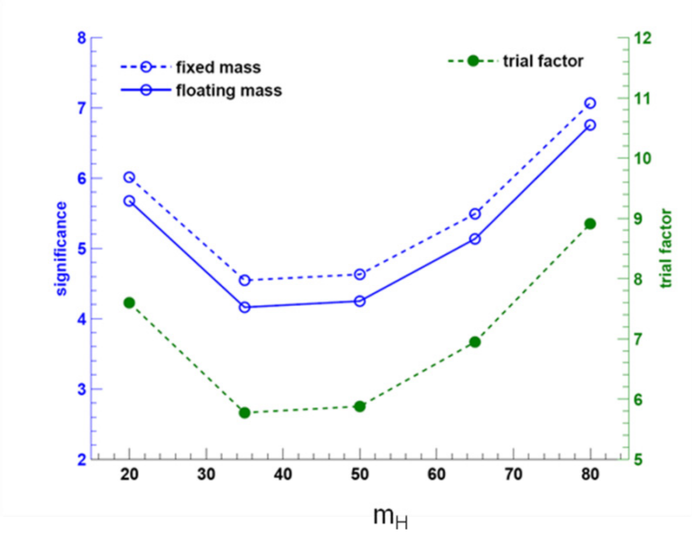

Figure 9: The Profiled LR significance for the fixed Higgs mass (dash blue), floating mass (blue) and the trial factor (dash green).

3. The Exclusion Case

3.1 The Toy Model

In order to announce a discovery one requires to reject the H_0 hypothesis at the level of 5σ, which means that the corresponding p-value should be $< 2.8\cdot 10^{-7}$. The requirements for exclusion are less stringent. It is custom to announce an exclusion of the H_1 hypothesis if one manages to reject it at the 95% confidence level. This corresponds to a (one sided) p-value of 5%, or a corresponding Gaussian significance of ~1.64σ (which is equivalent to a 2-sided p-value of 10%). To demonstrate the exclusion statistical methods we used a similar toy model to the one described in section 2.1 except that the signal strength was reduced in order to make the exclusion procedure more challenging and realistic. An example of some exclusion toy models is shown in Figure 10.

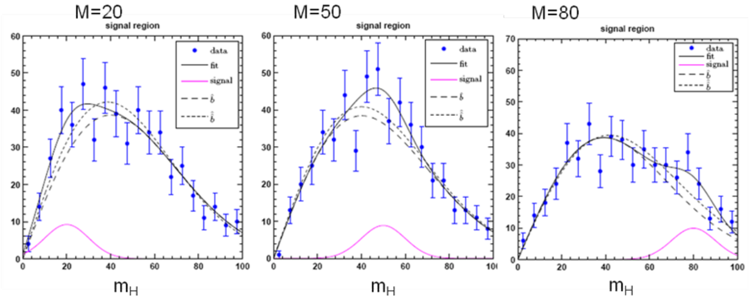

Figure 10: Exclusion toy MCs: The Gaussian shaped signal on top of the Rayleigh shaped background for m=20,50,80. $\hat{b}$ *and* $\hat{\hat{b}}$ *are MLEs.*

3.2 Exclusion using the Profile Likelihood ratio

The test statistic is given by Eq. (1.5) with μ=1, i.e. q_1. We test the $s(m_H)$+b hypothesis. Following Wilks theorem, q_1 distributes as a χ^2 with one d.o.f. under $s(m_H)$+b experiments (H_1). The exclusion significance can be easily calculated via

$$Z_{obs} = \sqrt{q_1} = \sqrt{-2\ln\lambda(\mu=1\,|\,x_{obs})}\,. \tag{1.19}$$

It is a custom to express it in terms of exclusion Confidence Level of the s+b ($CL \equiv 1 - CL_{s+b}$) which is related to the observed p-value via

$$p_1 = p_{s+b} = 1 - CL\,. \tag{1.20}$$

Note that a significance of 1.96σ corresponds to exclusion at the 97.5% CL, while a significance of 1.64σ corresponds to exclusion at the 95% CL.

The exclusion sensitivity is the median CL, which can be obtained using toy MCs or via the Asimov data set which is the expected background in this case (n=b).

The exclusion sensitivity as a function of the assumed Higgs mass is shown as the full blue line in Figure 11. Since 1-CL is below the 5% in all of the search range, the Higgs is excluded in all of the mass range with >95% CL.

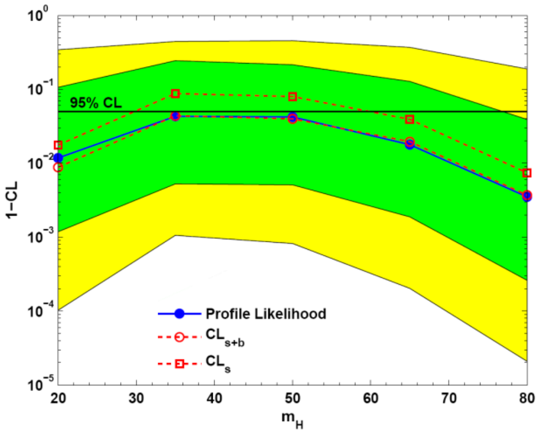

Figure 11: The expected exclusion sensitivity for the PL ratio (blue), CLs (top red dash) and CLs+b (bottom dash red) methods as a function of the Higgs mass. The green and yellow bands are the 1σ and 2σ bands for the PL ratio method. The region below the 95% CL line is excluded at >95% CL.

3.3 A Bayesian Exclusion

The signal strength μ can be interpreted as the ratio between the observed Higgs production cross section and the expected (SM) one, i.e. $\mu = \frac{\sigma}{\sigma_{SM}}$. Bayesians will set an upper limit on the signal strength by calculating the credibility interval $[0,\mu_{95}]$ where μ_{95} is calculated by integrating the posterior probability for μ, i.e.

$$0.95 = \int_0^{\mu_{95}} \Pr ob(\mu \mid n,m) d\mu \tag{1.21}$$

The posterior probability is calculated by integrating the likelihood using priors for the signal strength, $\pi(\mu)$, and the background $b(\theta)$, $\pi(\theta)$, i.e.

$$prob(\mu \mid n,m) = \frac{\int L\left(n,m \mid \mu\cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)d\theta}{\iint L\left(n,m \mid \mu\cdot s + b(\theta)\right)\pi(\mu)\pi(\theta)\,d\theta d\mu} \tag{1.22}$$

Note that the pdf of the posterior is based on the one observed experimental result with the likelihood integrated over the nuisance parameters. The posterior pdf with flat priors, is shown in Figure 12 for the Asimov data set, $n=b$. The credibility interval, $[0,\mu_{95}]$, can be easily found from the posterior probability distribution by integration up to an area of 95%. In this example, flat priors were used and the SM Higgs is not excluded since $\mu_{95}=1$ is included in the credibility interval. One will interpret the credibility interval by stating (though it's a wrong jargon) that $\mu<1.14$ at the 95% CL.

Even though there was no need for toy MC experiments to determine the credibility interval for a given experimental outcome, one needs to generate toy b-only MC experiments (each representing e.g. the LHC) to determine the expected exclusion sensitivity. The distribution of μ_{95} in toy MC experiments for some assumed Higgs mass, is shown in Figure 12.

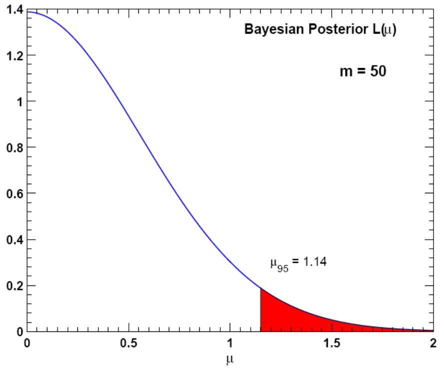

Figure 12: The posterior distribution as a function of the signal strength μ for the Asimov data set (n=b). The credibility interval in this example is [0,1.14].

The median and Asimov expected sensitivities coincide ($\mu_A=\mu_{median}=1.14$ in this example) and the 1σ and 2σ bands are indicated by the green and yellow bands.

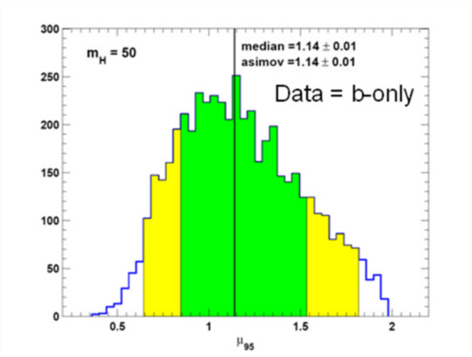

Figure 13: The μ_{95} distribution in toy b-only experiment. The 1σ and 2σ bands are indicated by the green and yellow bands. The expected median coincides with the Asimov sensitivity, $\mu_{95}=1.14$.

The expected Bayesian μ_{95} for various assumed Higgs masses is shown in Figure 14 (solid blue line). We find that the credibility interval $[0,\mu_{95}]$ does not contain $\mu_{95}=1$ (SM Higgs) for $m_H<28$ or $m_H>61$. This is sometimes wrongly expressed as an exclusion of the Higgs boson with the above quoted masses at the 95% CL. Also shown in Figure 14 is the 95% CL PL ratio exclusion sensitivity (dash red triangles) and the 97.5% CL PL ratio exclusion sensitivity (dash red circles). It is noticeable that the 95% Bayesian exclusion and the 97.5% CL PL exclusion coincide. We try to explain it in the next subsection.

3.4 Bayesian Exclusion and the PL Ratio

The dedicated Bayesians or Frequentists will say it's a blasphemy to try and find a connection between the two approaches as we attempted to do in section 2.2.5. However, though, comparing oranges to apples

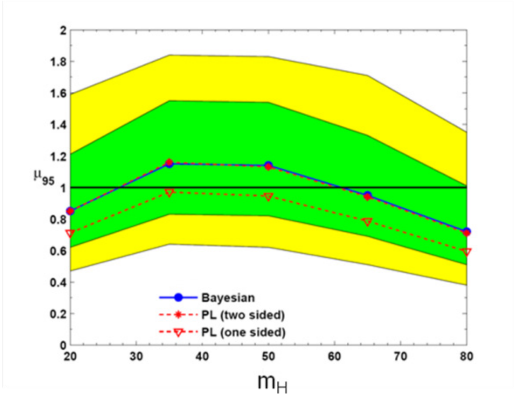

Figure 14: The expected μ_{95} sensitivity as a function of the Higgs mass in the Bayesian way (solid blue), the PL (at the 95% CL, bottom dashed red line)and the PL (at the 97.5% CL, top dashed red line).

make no sense, the coincident between the expected PL Ratio and Bayesian exclusions shown in Figure 14 led us to further investigation. If we plug in flat priors in the posterior, Eq. (1.22) we find using the saddle point approximation

$$prob(\mu\,|\,n,m)=\frac{\int L\big(n,m\,|\,\mu\cdot s+b(\theta)\big)\,d\theta}{\iint L\big(n,m\,|\,\mu\cdot s+b(\theta)\big)\,d\theta\, d\mu}=\frac{e^{\ln L\left(\mu\cdot s+b(\hat{\hat{\theta}})\right)}}{e^{\ln L\left(\hat{\mu}\cdot s+b(\hat{\theta})\right)}}=\lambda(\mu\,|\,n,m) \qquad (1.23)$$

Plugging in the b-only Asimov data sets, n=m=b, we find $\hat{\mu}=0$ and therefore

$$-2\log\lambda(\mu)\sim\frac{(\mu-\hat{\mu})^2}{{\sigma_\mu}^2};\quad prob(\mu\,|\,n=m=b)\sim\lambda(\mu)\sim e^{-\frac{\mu^2}{2{\sigma_\mu}^2}} \qquad (1.24)$$

For the Asimov b-only data we find $\sigma_\mu\sim\dfrac{1}{\sqrt{-2\log\lambda(1\,|\,n=m=b)}}$

This is exactly the normalized shape seen in Figure 12. If we define μ_{95} as in Eq. (1.21) we find the following equivalence

$$0.95=\int_0^{\mu_{95}}\Pr ob(\mu\,|\,n,m)\,d\mu \;\Leftrightarrow\; \sqrt{-2\log\lambda(\mu_{97.5})}=1.96 \qquad (1.25)$$

Meaning, that one has to identify the Bayesian μ_{95} with the frequentist PL CL of 97.5%, i.e. $\mu_{97.5}$. This is a result of thinking 5% one-sided or 5% two-sided…. This is exactly the equivalence we observe in Figure 14.

3.5 The CL$_{s+b}$ Method

Following the discussion in section 2.2 we use the LR as a test statistic (Eq. (1.3)). To take systematics into account we can either integrate it out (Eq. (1.2)) or profile it (Eq. (1.4)) . However, in this case the tested hypothesis is the s+b hypothesis (H_1), so in order to reject this hypothesis we calculate the observed s+b Confidence Level (CL$_{s+b}$) which is equivalent to the observed p_{s+b} and require that it is less than 5%. Figure 15 shows the test statistic pdf under b-only (red) and s+b (blue) hypotheses. The s+b p-value is indicated.

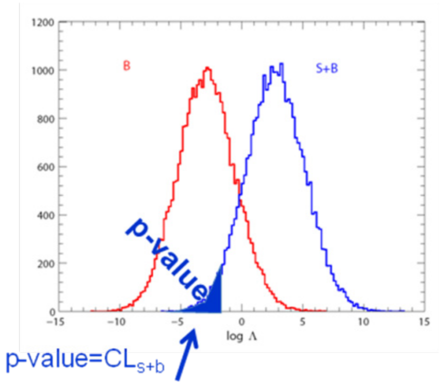

Figure 15: The LR test statistic pdf for b-only (red) and s+b (blue) experiments. The p-value is indicated.

3.6 The Modified Frequentist CL$_s$

The CL$_{s+b}$ method tests the s(m$_H$)+b hypothesis. If one excludes it, one excludes the s(m$_H$)+b hypothesis for a Higgs with a mass m$_H$. The subtlety here is that a downward fluctuation in the background might lead to an exclusion of a signal to which one is not sensitive (with a very low cross section). For a given observation, $n_{obs} < b$, the probability to observe even less events is given by $1 - p_b$. To protect against such fluctuations CL$_{s+b}$ was modified and CL$_s$ was defined [7]

$$CL_{s+b} \to CL_s \equiv \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b} \sim \frac{p(n_{obs} \le s+b)}{p(n_{obs} \le b)}. \qquad (1.26)$$

Statisticians do not like this p-value ratio because it does not provide a frequentist insight. However, some physicists insist on it since physics-wise it is conservative in a sense of coverage. Meaning, an observation might lead, using the CL$_{s+b}$ criteria, to an exclusion of the s+b hypothesis in 5% of toy s+b experiments, but it will be excluded in <5% of the times, if one uses the CL$_s$ criterion. This is clearly seen in Figure 16 where 1-CL is plotted vs. the assumed Higgs mass. The Higgs is excluded in all the measured mass region using the PL ratio or its almost equivalent CL$_{s+b}$ method, yet, using CL$_s$ the exclusion sensitivity is reduced and one can only exclude the Higgs if its mass is 30<m<60 . In that sense the CL$_s$ method is conservative, yet, quoting its result does not have any statistical clear frequentist interpretation.

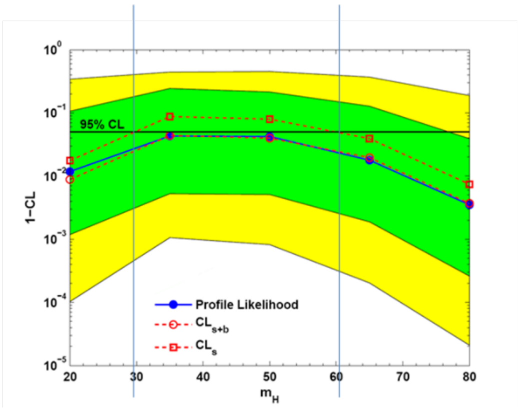

Figure 16: The expected 1-CL exclusion for the PL method (blue) with its 1σ (green) and 2σ (yellow) bands, the CL_{s+b} (dashe red circles) and CL_s (top dash red squares) as a function of the Higgs mass.

4. Conclusions

We have explored and compared all the methods to test hypotheses that are currently in use in the High Energy Physics market (PL ratio, CL_{s+b}, CL_s and Bayesian). We have shown a way to appreciate the Bayes factor by comparing it to the PLR. We have shown equivalence between the Bayesian exclusion with flat priors and the frequentist PL ratio. All methods tend to give similar sensitivities whether one integrates out the nuisance parameters or profile them.

Even though we have used non trivial typical case studies, we suggest that with real data all available methods be explored.

One of us (E.G.) is obliged to the Benoziyo center, to the the Israeli Science Foundation (ISF), the Minerva Gesellschaft and the German Israeli Foundation (GIF) for supporting this work.

Bibliography

[1] R. D. Cousins and V. L. Highland. "Incorporating systematic uncertainties into an upper limit." *Nucl. Instrum. Meth.*, 1992: A320:331.

[2] J. Neyman, E. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231*, 1933: 289-337.

[3] G. Cowan K. Cranmer, ,E. Gross and O. Vitells. "The Asimov Representative Data Set for Estimating Experimental Sensitivities with the Profile Likelihood." *In Preparation*, 2010.

[4] F. James, M. Ross. *Comput.Phys.Commun, 10.*, 1975: 343.

[5] Wilks, S.S. "The large-sample distribution of the likelihood ratio for testing composite hypotheses." *Ann. Math. Stat.*, 1938: 60.

[6] See for example Kass and Raftery. "Bayes Factors." *J. Am. Stat. Assoc. 90*, 1995: 773.

[7] Read, A. L. "Presentation of search results: the CLs technique." *J. Phys. G: Nucl. Part. Phys. 28*, 2002: 2693.