CMS Analysis Note

The content of this note is intended for CMS internal use and distribution only

LandS: software package for calculations of exclsion limits and signifcance

Mingshui Chen and Andrey Korytov

University of Florida, Gainesville, FL, USA

CMS collaboration

Abstract

Software package LandS (Limits-and-Significance) allows for calculations of exclusion limits and signifcances in the context of combining binned distributions. It allows one to account for multiple background contributions per bin and multiple sources of uncertainties, including a complete matrix of systematic error correlations between signal and different backgrounds within and across bins. A number of pre-defined *pdf* types for systematic errors are readily available. The package is optimized for a high speed of calculations. At the end of the note, the available features and the overall performance of the LandS package are compared to a number of other tools such as RooStats, mclimit, genlimit, TLimit, etc.

4 May 2010, Version 1.0

1 Introduction

There are a number of a very comrehensive statistical analysis tools already available on the "HEP market". To name a few: mclimit [1] (now including genlimit [2]) has been extensively used at Tevatron; RooStats [3] is being under development by the LHC experiments. The LandS package has been put together targetiing only a few specific tasks to be discussed below. Being an "in-house" product, it allows for a mximum flexibility to quickly adopt for ever changing needs and provides a good playground for developing improvements. For example, at the time of the first release, it was more than 10,000 times faster than the then-available RooStats versions (see last section) [1]. In comparison to mclimit/genlimit, LandS has more pre-defined options for systematic error *pdf*'s (e.g, lognormal, gamma function, which, as we will argue in the note, are very important options). Also, being an independent technical implementation, it gives means for crosschecking results and performance of other packages.

Current version of LandS (1.0) has the following features:

- Statatistical analyses are performed starting from binned distributions. This is well suited for combyning different physics channels, each of which is treated as a simple counting experiment. One can also analyze single-channel or multi-channel binned distributions (e.g. MVA output). An analysis for a single-channel counting experiment (single bin) is of course possible as well.
- Each bin is allowed to have multiple background contributions.
- LandS is designed to handle multiple sources of uncertainties, including a complete matrix of systematic error correlations between signal and background event rates within a single bin as well as across different bins. Correlations are assumed to be either $\pm 100\%$ or 0%, which are the natural choices once all systematic errors are broken down to a complete table of independent sources of uncertainties. Note that 100%-correlated errors do not necessarily have to have the same absolute scale.
- LandS provides a predefined set of typical systematic error *pdf*'s. Currently implemented functions are log-normal and trancated Gaussian. Coming up in the next release are Gamma function, Gamma function convoluted with lognormal/Gaussian, ABCD-method *pdf*, and more.
- Taking all the input information listed above, LandS allows one to calculate
  - Exclusion Limits using the Modified Frequentist (CLs) method for a given experimental outcome as well as projections (mean, median, 68% and 95% bands)
  - Exclusion Limits using the Bayesian method for a given experimental outcome and projections
  - Significance (Bayesian-Frequentist Hybrid) for a given experimental outcome and projections

Additional features will become available in the next version.

The LandS code is inteded for calculations of the final results. Although being relatively fast, it should not to be used with sophysticated tools for multi-dimentional cut optimization and/or MVA training—it simply may not be fast enough. For purposes of a cut optimization and MVA training, one would be better off by making use of various fast approximations for evaluating a significance, an exclusion limit, or any other number of merit one may want to consider.

[1] Inspired by such dramatic comparisons, RooStats's speed has been recently improved by almost a factor of 100.

2 Types of systematic error *pdf*'s

Throughout this section, we will use $\tilde{b}$ to denote a best estimate for event rate due to some background b. An uncertainty on b, whether theoretical or experimental, can be in general described in a form of some probability density function $pdf(b)$. This section describes commonly used *pdf*'s that are readily pre-defined in LandS. It remains up to user's judgment to decide which of the provided *pdf*'s would be the most appropriate for a particular uncertainty.

Of coursbe, the whole discussion of this section is equally applicable to the expected signal rates s and their uncertainties described by $pdf(s)$.

Instead of intgrating over *pdf*'s of neusance parameters, LandS utilizes a method of averaging over a large anssemble of random numbers generated according to the chosen *pdf*. Therefore, we also describe how random fluctuations are generated in LandS so that they have distributions with the desired *pdf*'s

2.1 Truncated Gaussian *pdf*

The truncated Gaussian distribution is the Gaussian distribution with the unphysical tail below $b = 0$ simply chopped off by hand:

$$\text{for } b > 0, \quad \frac{dp}{db} = \frac{C}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(b-\tilde{b})^2}{2\sigma^2}\right), \tag{1}$$

$$\text{for } b \leq 0, \quad \frac{dp}{db} = 0, \tag{2}$$

$$C = \frac{1}{0.5 + 0.5\,\text{erf}\left(1/(\sqrt{2}\,\epsilon)\right)} \tag{3}$$

where $\tilde{b}$ is our best estimate of the event rate b, $\sigma = \epsilon\tilde{b}$ is the dispersion (ϵ is the relative dispersion with respect to $\tilde{b}$), and C is the new normalization constant needed to preserve the unit integral under the *pdf* after the Gaussian tail for $b < 0$ has been chopped off.

Note that after truncation of the tail, the mean value and the dispersion of b change.

To generate b's with the truncated *pdf*, one can use the normal random number generator to get x's and make

$$b = \tilde{b} \cdot (1 + \epsilon \cdot x). \tag{4}$$

Should one find that $(1 + \epsilon \cdot x) \leq 0$, the random number trial must be dismissed and re-tried again.

If there are a number of backgrounds b_i's assumed to be 100% correlated, one simply modifies all these 100% correlated backgrounds in sync by using only one random number x for all of them, namely:

$$b_i = \tilde{b}_i \cdot (1 + \epsilon_i \cdot x)\,. \tag{5}$$

All x-trials giving $(1 + \epsilon_{max} \cdot x) \leq -1$ are ignored and re-taken again. This gives the desired truncation so that none of backgrounds b_i take a negative value.

2.2 Log-normal *pdf*

The log-normal *pdf* is the normal distribution in the log-scale for event rates. E.g., if $\tilde{b}$ is our best estimate for a background b, then the log-normal distribution around this cenral value is described by the following equation:

$$\frac{dp}{db} = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(b/\tilde{b}))^2}{2(\ln\kappa)^2}\right)\frac{1}{b}, \tag{6}$$

where κ describes the spread of the distribution.

One can see that $dp/d\ln b$ is indeed a Gaussian with a mean at $\ln\tilde{b}$ and sigma being $\ln\kappa$. From this, it is clear tha the log-normal distribution assigns equal probabilities for background to be a factor of κ^n larger or smaller than the best estimate $\tilde{b}$, i.e.

$$p(b > \tilde{b} \cdot \kappa^n) = p(b < \tilde{b}/\kappa^n) \tag{7}$$

For $n = 1$, these probabilities are 16% each, for $n = 2$, they are 2.5%, and so on (following the usual Gaussian tail probabilities).

For small errors, κ is close to 1 and one can write $\kappa = 1 + \sigma$, where σ is a small number (much smaller than one). In this case, by expanding logarithms, one can see that the log-normal distribution (Eq. 6) becomes a familiar Gaussian. The most important differences with respect to the Gaussian distribution are as follows:

- The log-normal *pdf* has a larger tail than a Gaussian. Consequently, a significance of an event excess will more conservative when one uses a log-normal *pdf* for the background uncertainties.
- The log-normal *pdf* has no tail toward negative values of b and the probability density at $b = 0$ is always zero.
- When uncertainties become very large (e.g., when we say "a factor of two uncertainty"), they map very naturally onto the log-normal *pdf*, while the Gaussian distribution obviously becomes completely inappropriate.
- Moreover, when the overall uncertainty in background b arises from uncertainties in multiple *multiplicative* factors (like various efficiencies), the central limit theorem implies that it is the $\ln b$ distribution that would tend to become Gaussian.

To generate b's with a log-normal *pdf*, one can use the normal random number generator to get x's and make

$$b = \tilde{b} \cdot \kappa^x, \tag{8}$$

which will have the desired log-normal distribution. If there are a number of backgrounds b_i's assumed to be 100% correlated, then one simply modifies them in sync by using only one random number x for all of them, namely:

$$b_i = \tilde{b}_i \cdot \kappa_i^x. \tag{9}$$

2.3 Gamma function *pdf*

[THIS FUNCTION IS NOT YET IMPLEMENTED]

Another distinct *pdf* arises when one evaluates backgrounds via counting events in a control sample:

$$b = \rho \cdot B, \tag{10}$$

where B is a number of events in some control region and ρ is a factor relating the event count in the control region B and the number of background events in the signal region b.

In this case, one needs to worry about statistical uncertainty related to the number of observed events in the control region and systematic errors associated with the factor ρ. Similar situations arise when one relies on Monte Carlo and ends up with only a few Monte Carlo events after all the cuts (in this case, the MC event weight plays the role of ρ and number of MC events surving cuts is B).

If one can neglect systematic erros on ρ, then *pdf* for the estimated background becomes the Gamma function (derived using the Bayesian methodology with a flat prior on what b might be):

$$\frac{dp}{db} = \frac{1}{\rho} \frac{(b/\rho)^B \exp(-b/\rho)}{\Gamma(B+1)}. \tag{11}$$

Mean value of this distribution is $\rho(B+1)$ and its dispersion is $\rho\sqrt{B^2+1}$. Note that even if one observes $B = 0$ in the control region in data or in a MC sample, in fact, it does NOT imply that the estimated event rate b is zero.

Should ρ have its own uncertainties, this equation must be further convoluted with uncertainties on the parameter ρ.

2.4 Gamma function convoluted with lognormal/gaussian *pdf*

[THIS FUNCTION IS NOT YET IMPLEMENTED]

...

2.5 ABCD method *pdf*

[THIS FUNCTION IS NOT YET IMPLEMENTED]

...

2.6 Flat *pdf*

The flat pdf is given by the following equation:

$$\text{for } |b-\tilde{b}| < \Delta/2, \quad \frac{dp}{db} = \frac{1}{\Delta}, \tag{12}$$

$$\text{otherwise,} \quad \frac{dp}{db} = 0, \tag{13}$$

where $\tilde{b}$ is our best estimate of the event rate b and Δ is the allowed range. The mean value of b is $\tilde{b}$ and the relative dispersion is $\Delta/\sqrt{12}/\tilde{b}$. In order to keep $b > 0$, $\epsilon = (\Delta/2)/\tilde{b}$ must be smaller than one.

To generate b's with the flat pdf, one can use the flat random number generator to get x's beween 0 and 1 and then make

$$b = \tilde{b} + \Delta \cdot (x - 0.5)). \tag{14}$$

If there are a number of backgrounds b_i's assumed to be 100% correlated, one simply modifies all these 100%-correlated backgrounds in sync by using only one random number x for all of them, namely:

$$b_i = \tilde{b}_i + \Delta_i \cdot (x - 0.5)). \tag{15}$$

3 Binned distributions: organization of the input information

The input information is expected to be organized in a form of the following table Table 1

For each channel (bin), a user must provide number of observed events (this will be ignored when one runs a code to generate projections), number of expected signal and all backgrounds that are deemed to be non-negligible. These entries define columns of the table.

Rows must give all independent sources of uncertainties that are belived to be important. A user should ascribe a type of pdf to be associated with each source of uncertainties.

Finally, the table cells must show how much each particualr background or signal in each bin are expected to be affected by the listed uncertainties, namely, the table cells contain the following parameters:

- $\sigma/\tilde{b}$ for a truncated Gaussian pdf's,
- κ for the log-normal pdf
- $(\Delta/2)/\tilde{b}$ for the flat pdf

All systematic errors along any single line are assumed to be 100% correlated, while errors in different lines are assumed to have no correlations. It is important to stress that correlated errors may not necessarily have the same absolute scale. E.g., uncertainties in electron reconstruction efficiency will result in systematic errors in the $H \to WW \to 2e2\nu$ and $H \to WW \to e\mu2\nu$ event rates. The two channels will have 100% corrleated errors that are two times different in magnitude.

Note that in many analyses, the signal hypothesis may have parameters over which one has to scan in a search (e.g., Higgs boson mass). The whole analysis may be even tailored to be signal hypothesis dependent (e.g., cuts may have to be optimized differently for a differnet Higgs boson masses). If the reasonable step in probing such signal parameters is small in comparison to the range of scanning, producing of such tables for all possible steps may become not practical. In these cases, cells may represnt pointers to the functions describing how event counts and systematic errors change as one scans over the available phase space of the signal hypothesis parameters.

Channels		**bin 1**					**bin 2**					**bin i**				
Observed events		n_1					n_2					n_i				
Processes		signal	bkgd 1	bkgd 2	...	bkgd j	signal	bkgd 1	bkgd 2	...	bkgd j	signal	bkgd 1	bkgd 2	...	bkgd j
Expected Rate		s_{10}	b_{11}	b_{12}	...	b_{1j}	s_{20}	b_{21}	b_{22}	...	b_{2j}	s_{i0}	b_{i1}	b_{i2}	...	b_{ij}
Sources of uncertainties	pdf type															
Uncertainty 1	lognormal	κ_{110}	-	-	-	-	κ_{120}	-	-	-	-	κ_{1i0}	-	-	-	-
Uncertainty 2	tr. gauss	κ_{210}	ϵ_{211}	ϵ_{212}	...	ϵ_{21j}	κ_{220}	ϵ_{221}	ϵ_{222}	...	ϵ_{22j}	κ_{2i0}	ϵ_{2i1}	ϵ_{2i2}	...	ϵ_{2ij}
Uncertainty ...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Uncertainty k	gamma pdf	-	-	-	-	-	-	-	-	-	-	-	-	-	-	B_{kij}
Uncertainty ...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Table 1: Input table witha complete breakdown of all individual processes, all *independent* sources of uncertainties, and their impact on event rates associated with individual processes expressed in terms of parameters of pdf's. All systematic errors along any single line are assumed to be 100% correlated. Errors in different lines are assumed to have no correlations.

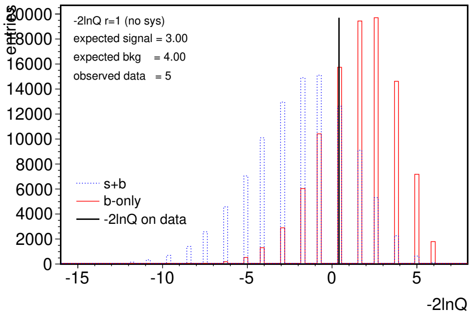

Figure 1: An example of $-2\ln Q$ distributions of *background-only* (red, solid) and *background+signal* (blue, dashed) hypotheses.

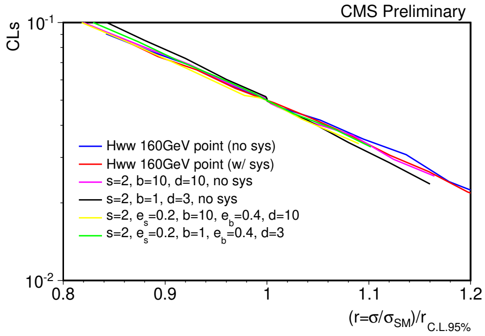

Figure 2: Dependence of CL_s vs $r/r_{95\%}$ Different curves represent different assumptions on signal, background, systematic errors. The HWW curves correspond to a three-channel combination ($H \to WW \to 2\ell 2\nu$ at 7 TeV, 1 fb^{-1}).

4 Statistical analysis options

All the methods presented in this section have been routinely used in the high energy physics for the last two decades; in particular, by LEP and Tevatron experiments in Higgs searches. As mentioned earlier, one can view individual bins as separate single-channel counting experiments we try to combine or as a single-channel shape based analysis or a combination of both. From the point of view of a statistical analysis—once Table 1 is fully defined—there is not any distinction between all these cases.

For each of the bins, we define:

- n_i is a number of observed events in a bin i;
- s_i^* is our best estimate of the expected signal events in bin i;
- b_{ij}^* is our best estimate of the expected background events in bin i for each j-th background considered in the analysis (and $b_i^* = \sum_j b_{ij}^*$);
- r is the "signal strength modifier" (common for all bins)

4.0.1 Exclusion limits: the Modified Frequentist method

For a given observation (set of n_i's), the Modified Frequentist approach [**?**] (often referred to as $\mathrm{CL_s}$) starts out from defining a likelihood ratio Q:

$$Q = \frac{p(observation|signal + bkgd)}{p(observation|bkgd)} = \frac{\prod_i \frac{(\tilde{b}_i + r\cdot\tilde{s}_i)^{n_i}}{n_i!} e^{-(\tilde{b}_i + r\cdot\tilde{s}_i)}}{\prod_i \frac{(\tilde{b}_i)^{n_i}}{n_i!} e^{-\tilde{b}_i}} = \prod_i \left(1 + r\frac{\tilde{s}_i}{\tilde{b}_i}\right)^{n_i} \cdot e^{-r\cdot\tilde{s}_i}, \qquad (16)$$

which is then converted into $-2lnQ$:

$$-2lnQ = 2r\sum_i \tilde{s}_i - 2\sum_i n_i \cdot ln\left(1 + r\frac{\tilde{s}_i}{\tilde{b}_i}\right). \qquad (17)$$

Next, for a given trial value of r, we run 10^5 toy experiments to build a $-2lnQ$ distribution for the *background-only* hypothesis and, then, we run another 10^5 toy experiments to build a $-2lnQ$ distribution for the *background+signal* hypothesis. An example of two such distributions is shwon in Fig. 1. In each of these toy experiments, b_{ij}'s and s_i's are modified according to the table of systematic errors, including bin-to-bin correlatations. Note that the values of $-2lnQ$ are always calculated using our best estimates $\tilde{b}_{ij}$ and $\tilde{s}_i$, regardless of the actual variations that we used in a particular toy experiment.

From the two $-2lnQ$ distributions obtained this way, we evaluate CL_b and CL_{sb} probabilities and their ratio CL_s as follows:

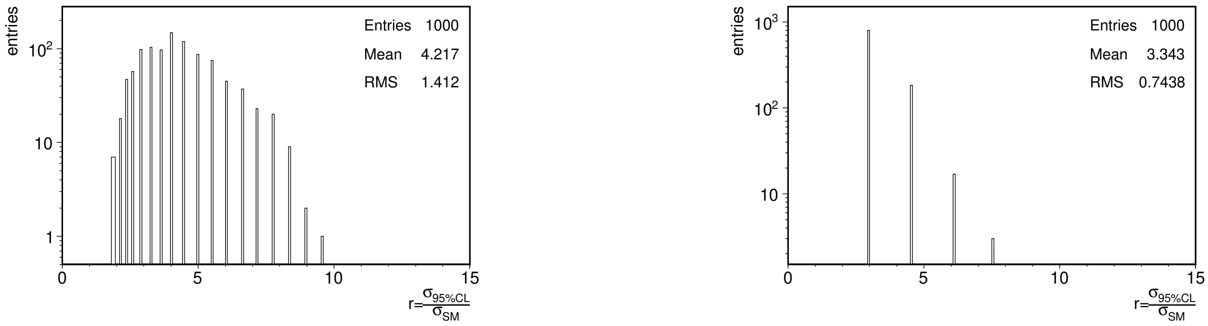

Figure 3: Differential distributions of possible experimental outcomes in terms of r for the case of high background rate (left; $\tilde{b} = 10$ and $\tilde{s} = 2$) and for the case of when only a very few background events are expected (right; $\tilde{b} = 0.2$ and $\tilde{s} = 1$).

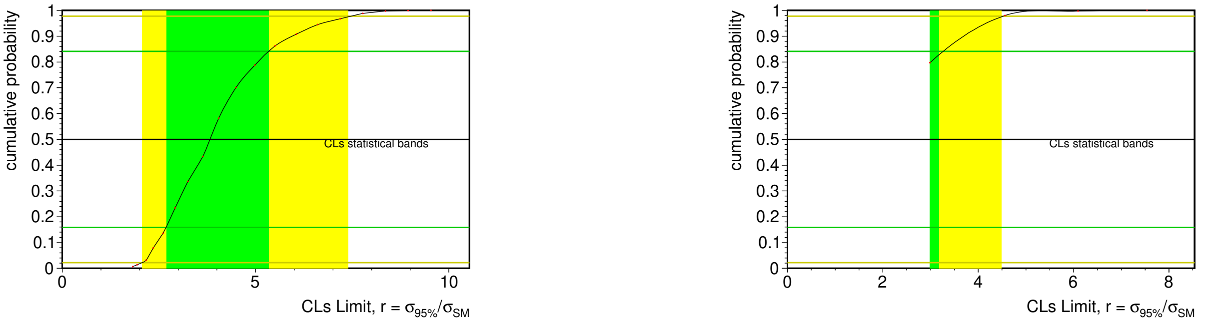

Figure 4: Cumulative probabilities for possible experimental outcomes in terms of r for the case of high background rate (left; $\tilde{b} = 10$ and $\tilde{s} = 2$) and for the case of when only a very few background events are expected (right; $\tilde{b} = 0.2$ and $\tilde{s} = 1$).

- $CL_b = P(-2lnQ \geq -2lnQ_{obs})$ from the background-only distribution,
- $CL_{sb} = P(-2lnQ \geq -2lnQ_{obs})$ from the signal+background distribution,
- and their ratio $CL_s = CL_{sb}/CL_b$

For an exclusion at 95% Confidence Level, we then adjust r (each adjustment requires $2 \cdot 10^5$ toy experiments) untill we get CL_s within the ± 0.001 tolerance band about the desired value of 0.05. With 10^5 toy-experiments we use to generate $-2lnQ$ distributions and the ± 0.001 tolerance on matching the desired value of CLs, the precision on the evaluated values of r is about 1%.

To build projections, the LandS package goes along the following prcedure.

We run 1,000 pseudo-experiments in assumption of the *background-only* hypothesis. In each of these 1,000 pseudo-experiments, n_i's are generated according to Poisson probabilities for a mean event rate $b_i = \sum_j b_{ij}$, where b_{ij}'s are modified each time according to systematic error *pdf*'s. This gives us a bulk of possible experimental outcomes n_i's, for each of which we derive an exclusion limit r as described above. At the end, we have a set of possible experimental outcomes, their probabilities, and associated exclusion limits r (Fig. 3. This allows us to report mean $< r >$, which is an average (or median) r over the 1000 pseudo-experiments. In addition, we can show bands that contain 68% (95%) of the 1000 pseudo-experimental outcomes. The way median limits and the 68% (95%)bands (so-called green/yellow bands) are defined is illustrated in Fig. 4.

4.0.2 Exclusion limits: the Bayesian method

For a given observation (set of n_i's), the Bayesian method starts out from building a likelihood function:

$$L(r) = \frac{p(\,observation\,|\,b + r \cdot s\,) \cdot \pi(r)}{\int_0^\infty p(\,observation\,|\,b + r \cdot s\,) \cdot \pi(r)\,dr}, \tag{18}$$

where $\pi(r)$ is the prior on the value of r, which is customary taken to be "flat". Another option is a so-called "correlated" prior. More details on "flat" and "correlated" priors can be found elsewhere [**?**, 2]. Both options are available in LandS.

The 95% Confidence Level exclusion limit on signal strength r is derived from the following equation:

$$\int_r^\infty L(r)\,dr = 0.05. \tag{19}$$

Without systematic errors, $p(\,observation \,|\, b + r \cdot s\,)$ is calculated according to the following equation:

$$p(\,observation \,|\, b + r \cdot s\,) = \prod_i \frac{(\tilde{b}_i + r \cdot \tilde{s}_i)^{n_i}}{n_i!} e^{-(\tilde{b}_i + r \cdot \tilde{s}_i)}. \tag{20}$$

To include systematic errors, $p(\,observation \,|\, b + r \cdot s\,)$ is calculated by taking into account uncertainties on backgrounds, signals, and their correlations.

In the LandS code, this is achieved by throwing 20K toy-experiment conditions (sets of b_{ij} and s_i) and averaging $p(\,observation \,|\, b + r \cdot s\,)$ over them. In fact, to save on CPU, one can perform integration over r in the equations 18, 19 before performing the averaging over uncertainties (for details, see, for example, Ref [2]). The very same methodology is implemented in the LandS package.

As with the CL_s implementation, the projected Bayesian exclusion limits r's (mean, median, and the 68%/95% bands) are obtained for 1,000 pseudo-experiments played out according to the *background-only* hypothesis and varied event rates s_i and b_{ij} following the table of systematic errors (Table 1).

4.0.3 Significance: the Hybrid Frequentist-Bayesian method

For a given observation (set of n_i's), the LandS package allows one to evaluate significance using the Hybrid Bayesian-Frequentist method. The method calls for an evaluation of the p-value, i.e. the probability to observe $-2\ln Q \leq -2\ln Q_{obs}$, by a brute-force of generating an adequate number of pseudo-experiments (as always, b_{ij} are varied from one toy-experiment to another following the table of systematic errors) and building a *background-only* $-2lnQ$ distribution, where $-2lnQ$ is calculated according to Eq. 17. One of the bins in this distribution corresponds to $-2\ln Q_{obs}$.

The p-value is then converted into significance Z following the *one-sided* Gaussian tail convention:

$$p - \text{value: } p = P(-2\ln Q \leq -2\ln Q_{obs} \,|\, b); \tag{21}$$

$$if(p < 0.5),\;\; p = \frac{1}{\sqrt{2\pi}} \int_Z^\infty e^{-x^2/2}\,dx; \quad \text{otherwise, } Z = 0. \tag{22}$$

This brute-force approach gives as accurate a measure of significance as accurate our assumptions on systematic errors are. It has no internal approximations or shortcuts. The only shortcoming of this method is that we hit the CPU limitations as one starts probing significances $Z > 5$ ($Z = 5$ implies p-value $p = 3 \cdot 10^{-7}$, which calls for a minimum of 10^8 pseudo-experiments). However, this does not concern the current note since at 10 TeV and luminosities ≤ 1 fb^{-1}, we are not any close to so high signfcances, as will be shown below. Also, in any real search, a discussion of a significance beyond $Z \sim 5$ is hardly ever needed. One of the reasons is that it is not realistic to really control tails of systematic errors with such a level of precision.

To make a projection for what one may expect in the real experiment, should the signal be present, 1,000 pseudo-experiments (thousand sets of possible outcomes n_i) are generated according to the *signal+background* hypothesis with varying background and signal rates b_{ij} and s_i following the table of systematic errors (Table 1). Each such set of pseudo-measurements, gives a particular $-2\ln Q$ value and its corresponding significance Z. Using this information, we then evaluate the mean expected significance $< Z >$, the median value Z_{median}, as well as the 68% and 95% bands.

data	1	2	3	4	5	6
LandS	0.70868	0.889756	1.09499	1.33386	1.57428	1.57428
RooStats	0.701835	0.882929	1.10647	1.32323	1.57232	1.80109

Table 2: An example for comparing results by LandS and HybridCalculator of RooStats: expected signal events = 5.39, background = 2.067, C.L. 95% upper limits on the signal strenth r for different observed data samples.

data	RooStats	LandS
1	0.70647	0.702283
1	0.71417	0.699838
1	0.682842	0.701599
1	0.719481	0.70816
1	0.698538	0.701605
1	0.688738	0.707329
1	0.725877	0.702908
1	0.699042	0.705187
1	0.709633	0.70196
1	0.698447	0.70529
mean	0.704324	0.703616
standard deviation	0.013434	0.002731

Table 3: An example showing the stability of results from LandS and HybridCalculator of RooStats: expected signal events = 5.39, background = 2.067, C.L. 95% upper limits on the signal strenth r for observed data = 1.

5 Code performance and validation

5.1 LandS v.s. RooStats: 95% C.L. upper limits using Modified Frequentist approach without systematics

HybridCalculator in RooStats in latest ROOT v5-26-00b was tried and it's found that the time of throwing toys in RooStats was unacceptablly slow.

The input is taken to be with expected signal events = 5.39, and expected background events = 2.065. 95% C.L. upper limits are computed for the observed data = 1, 2, 3, 4, 5, 6 separately. As shown in Table 2, results from both tools are perfectly consistent with each other (within statistical precision).

With RooStats, it takes 4 minutes to throw 5,000 toys for building -2lnQ distribution to calculate the value of CLs. To get 95% C.L. upper limit on r (CLs=0.05), one needs 5 iterations in average, about 20 minutes. While in LandS, it takes 0.5 seconds to throw 100,000 toys and 2 seconds to get 95% C.L upper limit on r. In RooStats the tolerance of CLs=0.05 is 0.005, while in LandS it's 0.001.

To demostrate the stability of the results, 10 trials with data = 1 and different seeds were exercised. The results are shown in Table 3. As expected, with more toys (which is doable due to much faster generation of toys in LandS), the precision is improved.

5.2 LandS v.s. RooStats: 95% C.L. upper limits calculation using Bayesian technique without systematics

Recently in the note [2], RooStats and LandS were used for calculating C.L. 95% upper limit on the 3-channel combination. The results from both Bayesian techniques of RooStats and LandS are consistent with each other.

5.3 LandS v.s. TLimit: CLs values

TLimit which is available in ROOT is a package to do limit computation with CLs technique. Unfortunately, it's not able to calculate upper limit at desired confidence level. Hence we will only compare the CLs values between TLimit and LandS.

An 3-channels combination exercise was performed both for LandS and TLimit. The inputs are listed in Table 4, truncated Gaussain pdf are assumed on all uncertainties of signal and backgrounds (that's the only pdf implemented in the TLimit package).

Channels		channel 1		channel 2		channel 3	
Observed events		10		10		12	
Processes		signal	bkgd	signal	bkgd	signal	bkgd
Expected Rate		1.798	10.172	3.481	10.342	2.107	12.187
Sources of uncertainties	*pdf* type						
uncertainty	tr. gauss	0.11	0.28	0.3	0.28	0.5	0.28

Table 4: An example of 3-channel combination inputs.

category	A	B	C
TLimit	0.514502	0.457196	0.322654
LandS	0.515046	0.462152	0.322961

Table 5: CLs values calculated by TLimit and LandS for the inputs(Table 4). Categories A, B, C are detailed in context.

	genlimit			*LandS*		
data	0	1	2	0	1	2
baysian "flat" prior	3.3885	5.0191	6.7932	3.37894	5.00461	6.77293
baysian "correlated" prior	3.1987	4.7293	6.3881	3.19541	4.72454	6.38166
CLs technique	-	-	-	3.21101	4.69774	6.37697

Table 6: 95% C.L. upper limits on the cross section ratio from *genlimit* and *LandS*.

For the uncertainties correlation, it was further broken down to 3 categories:

- A. 100% correlated among all channels and among signal and backgrounds
- B. only 100% correlated between signal and background inside each channel
- C. uncorrelated at all

The CLs values of each category computed from TLimit and LandS are listed in Table 5. The results from TLimit and LandS are perfectly consistent.

5.4 LandS v.s. genlimit: 95% C.L. upper limits with Bayesian technique including systematics

genlimit is a package with Bayesian technique implemented, it can do limit calculation with/without systematics. Also it has choices to choose "*flat*" or "*correlated*" prior on signal cross section. However, there is no pre-implemented choices of uncertainties pdf in *genlimit*.

In this section we take expected rate of signal to be 1 with uncertainty 20%, and expected rate of background to be 0.46 with no uncertainty. The prior on the signal uncertainty is assumed to be truncated Gaussian pdf.

Since LandS borrowed the whole idea of doing integration via Gauss-Laguerre Quadrature technique and elimiting nuisance parameters through Montre Carlo sampling, it's expected to get the identical results for *genlimit* and *LandS* with the same prior on the cross section. The results are listed in Table 5.

In the Table 5, there are also results from LandS using CLs technique. One may already notice that with "correlated" prior on the cross section, Baysian technique gives the same results as that of CLs technique.

5.5 LandS v.s. RooStats: significance calculation

In the note [2], we also exercised the calculation of significance in LandS through the Hybrid-Frequentist-Bayesian technique. And the results are compared with that of Profile-Likelihood technique implemented in RooStats. As shown in the note, the results are comparable (without systematics).

References

[1] Mingshui Chen et al., *LandS software package* http://cern.ch/mschen/LandS.

[2] S. Baffioni et al., *Search strategy for the Higgs boson in the $ZZ^{(*)}$ decay channel at $\sqrt{s}$ = 10 TeV with the CMS experiment.*

6 Conclusions

...

References

[1] mclimit

[2] Joel Heinrich, *GENLIMIT software package* CDF note CDF/MEMO/STATISTICS/PUBLIC/7587 (April 15, 2005). Now genlimit has been absorbed into the mclimit package.

[3] RooStats

[4] Mingshui Chen et al., *LandS: Limits and Significance* The code and instructions can be found at https://mschen.web.cern.ch/mschen/LandS/

[5] CLs

[6] Bayes

[7] Bayesian-Frequentist Hybrid

[**?**] Luc Demortier, *Objective Bayesian Upper Limits for Poisson Processes*, CDF note CDF/MEMO/STATISTICS/PUBLIC/5928 (April 18, 2002; revised June 24, 2005).

TO BE FILLED OUT