CMS CR 2000/004

CMS Conference Report

23 January 2000

On observability of signal over background

S.I. Bityukov

Institute for High Energy Physics, Protvino Moscow Region, Russia

N.V. Krasnikov

Institute for Nuclear Research RAS, Moscow, Russia

presented by S. Bityukov

Abstract

Several criteria used by physicists to quantify the ratio of signal to background in planned experiments are compared. An equal probabilities test is proposed for the evaluation of the uncertainty in planned search experiments. We also consider a probability of discovery as a quantity for comparison of proposals for future search experiments.

Presented at *Workshop on "Confidence Limits"*, CERN, Geneva, Switzerland, January 17-18, 2000

1 Introduction

The aim of a search experiment is to detect predicted new phenomena. Usually, the theoretical estimations of expected mean number of signal events of a new phenomenon N_s and that of background events N_b are known, and we can define some value of "significance" as a characteristic of the observability of the phenomenon. Some function of the observed number of events x (a statistic) is used to draw a conclusion on observation or non-observation of the phenomenon. The value of this statistic allows one to find the degree of confidence of the conclusion. There exist two types of mistake: to state that a phenomenon does not exist while it actually exists (Type I error), or to state that a phenomenon exists while it does not (Type II error).

In this paper we compare the "signal significances" to estimate the discovery potential of a future experiment:

– "significance" $S_1 = \dfrac{N_s}{\sqrt{N_b}}$ [1],

– "significance" $S_2 = \dfrac{N_s}{\sqrt{N_s + N_b}}$ [2, 3],

– "significance" $S_{12} = \sqrt{N_s + N_b} - \sqrt{N_b}$ [4].

For this purpose we apply an equal-tailed test to study the behaviour of Type I and Type II errors as a function of N_s and N_b in planned search experiments with specified values of the "significances" S_1, S_2 and S_{12}. An equal probabilities test is proposed to estimate the uncertainty in separation of two hypotheses on observability of predicted phenomenon in these experiments. The hypotheses testing results obtained by Monte-Carlo calculations are compared with the result obtained by the direct calculation of probability distributions.

2 Notations

Let us assume that the average number of signal events coming from a new phenomenon (N_s) and the average number of background events (N_b) in the experiment are given. We suppose that the events have a Poisson distribution with parameters N_s and N_b, i.e. the random variable $\xi \sim Pois(N_s)$ describes the signal events and the random variable $\eta \sim Pois(N_b)$ describes the background events. Assume that we observed x events – the realization of the process $X = \xi + \eta$ (x is the sum of signal and background events in the experiment). Here N_s, N_b are non-negative real numbers and x is an integer. The classical frequentist methods of testing a precise hypothesis allow one to construct a rejection region and determine associated error probabilities for the following "simple" hypotheses:

$H_0 : X \sim Pois(N_s + N_b)$ versus $H_1 : X \sim Pois(N_b)$, where $Pois(N_s + N_b)$ and $Pois(N_b)$ have the probability distributions
$f_0(x) = \dfrac{(N_s + N_b)^x}{x!} e^{-(N_s+N_b)}$ for the case of presence, and $f_1(x) = \dfrac{(N_b)^x}{x!} e^{-(N_b)}$
for the case of absence of signal events in the whole population.

The probability distributions $f_0(x)$ (a) and $f_1(x)$ (b) for the case of $N_s + N_b = 104$ and $N_b = 53$ ([3], Table.13, cut 6) are shown in Fig.1. As we see, the intersection of these distributions takes place. Let us denote the threshold (critical value) that divides the abscissa in Fig.1 into the rejection region and the area of accepted hypothesis H_0 via N_{ev}. The incorrect rejection of the null hypothesis H_0, the Type I error (a phenomenon is taken to be absent, while it exists), has the probability $\alpha = \sum_{x=0}^{N_{ev}} f_0(x)$, and the incorrect acceptance of H_0, the Type II error (a phenomenon is taken to be present, while it is absent), has the probability $\beta = \sum_{x=N_{ev}+1}^{\infty} f_1(x)$. The α and β dependences on the value of N_{ev} for the above example are presented in Fig.2.

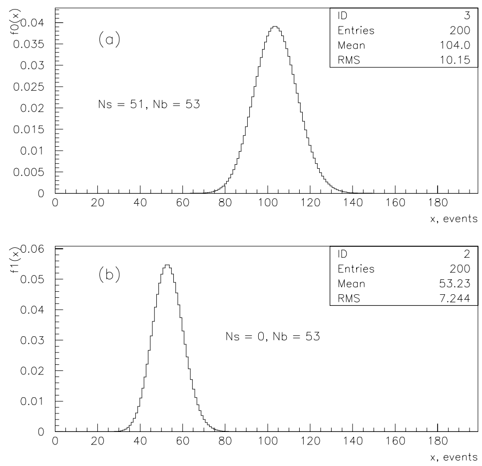

Figure 1: The probability distributions $f_0(x)$ (a) and $f_1(x)$ (b) for the case of 51 signal events and 53 background events obtained by direct calculations of the probabilities.

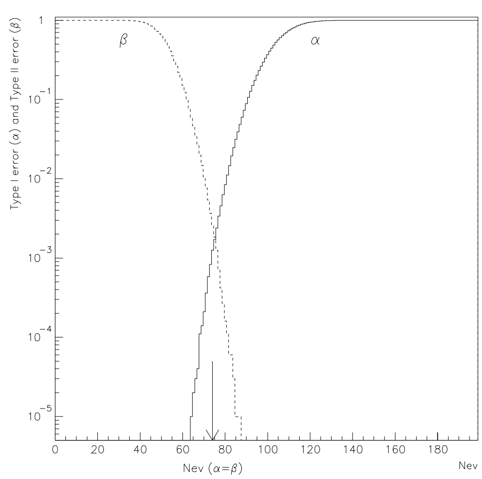

Figure 2: The dependence of Type I α and Type II β errors on critical value N_{ev} for the case of 51 signal events and 53 background events.

3 Hypotheses testing

In this Section the construction of a rejection region for the statistic x, the number of observed events, is described. The decision to either reject or accept H_0 will depend on the observed value of x, where small values of x correspond to the rejection of H_0, i.e.

if $x \leq N_{ev}$, reject H_0,

if $x > N_{ev}$, accept H_0.

In compliance with this test, the frequentist reports the Type I and Type II error probabilities as $\alpha = P_0(X \leq N_{ev}) \equiv F_0(N_{ev})$ and $\beta = P_1(X > N_{ev}) \equiv 1 - F_1(N_{ev})$, where F_0 and F_1 are cumulative distribution functions of X under H_0 and H_1, respectively.

The Type I error α is also called a significance level of the test. The value of β is meaningful only when it is related to the alternative hypothesis H_1. The dependence $1 - \beta$ is referred to as a power function that allows one to choose a favoured statistic for the hypothesis testing. It means that for the specified significance level we can determine the critical value N_{ev} and find the power $1 - \beta$ of this criterion. The larger the value of $1 - \beta$, the better the statistic separates hypotheses for a specified value of α.

For a conventional equal-tailed test [1)] with $\alpha = \beta$, the critical value N_{ev} satisfies the relation $F_0(N_{ev}) \equiv 1 - F_1(N_{ev})$.

In a similar way we can construct the rejection region, finding the critical values c_1, c_2 and c_{12}, for the statistics $s_1 = \dfrac{x - N_b}{\sqrt{N_b}}$ ("significance" S_1), $s_2 = \dfrac{x - N_b}{\sqrt{x}}$ ("significance" S_2) and $s_{12} = \sqrt{x} - \sqrt{N_b}$ ("significance" S_{12}). The probability distributions of statistics under consideration can be obtained in analytical form or by a Monte-Carlo simulation of the results of a large number of experiments (see as an example [6]) for the given values N_s and N_b. The both approaches were used in our study. The probability distributions for the case of $N_s + N_b = 104$ and $N_b = 53$ events obtained as a result of 10^5 simulations with random variables ξ and η are shown in Fig.3. There is no significant difference between these distributions compared with the distributions resulting from direct calculations of the probabilities (Fig.1).

The probability distributions of statistic s_2 for the case of $N_s = 51$, $N_b = 53$ (a) and the case of $N_s = 0$, $N_b = 53$ (b) are shown in Fig.4. The behaviour of probabilities α and β as a function of the critical value c_2 for the statistic s_2 is also presented in Fig.4 (c).

We stress that the second approach allows to construct the probability distributions and, correspondingly, the acceptance and the rejection regions for complicated statistics, taking into account the systematic errors and the uncertainties in the estimations of N_b and N_s.

---

1) See for example ref. [5].

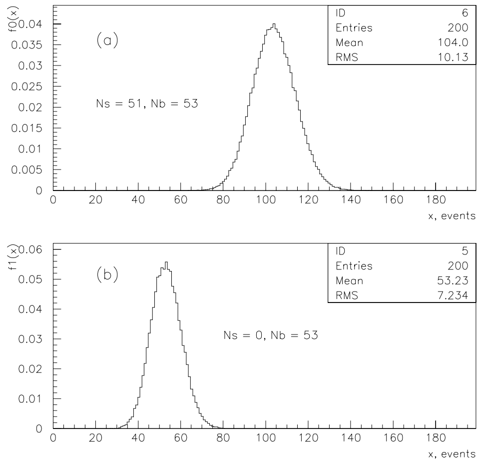

Figure 3: The probability distributions $f_0(x)$ (a) and $f_1(x)$ (b) for the case of 51 signal events and 53 background events obtained by Monte Carlo simulation (10^5 Monte-Carlo trials).

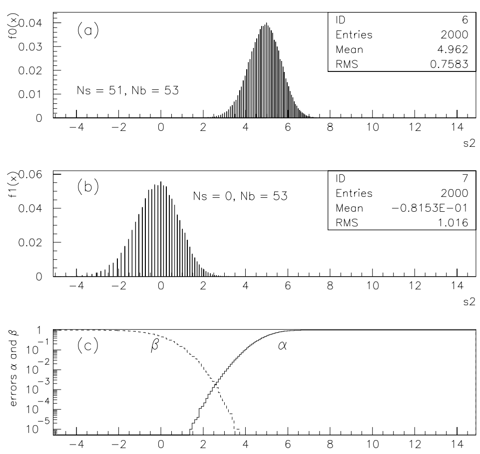

Figure 4: The probability distributions $f_0(x)$ (a) and $f_1(x)$ (b) for statistic s_2. The dependence of Type I and Type II errors on the critical value c_2 (c) for the case of 51 signal events and 53 background events.

4 Equal-tailed test

What is the exact meaning of the statement that

$$S_1 = \frac{N_s}{\sqrt{N_b}} = 5 \text{ or } S_2 = \frac{N_s}{\sqrt{N_s + N_b}} = 5 \text{ ?}$$

Tables 1 and 2 give the answer to this question. Here the values α and β have been determined by applying equal-tailed test (in this study we use the conditions $min(\beta - \alpha)$ and $\alpha \leq \beta$). One can see the dependence of α (or β) on the value of N_s and N_b. The case of $N_s = 5$ and $N_b = 1$ for S_1 (Fig.5) is perhaps the most dramatic example. Having 5σ deviation and rejecting the hypothesis H_0, we are mistaken in 6.2% of the cases; if we accept the hypothesis H_0, we are mistaken in 8.0% of the cases.

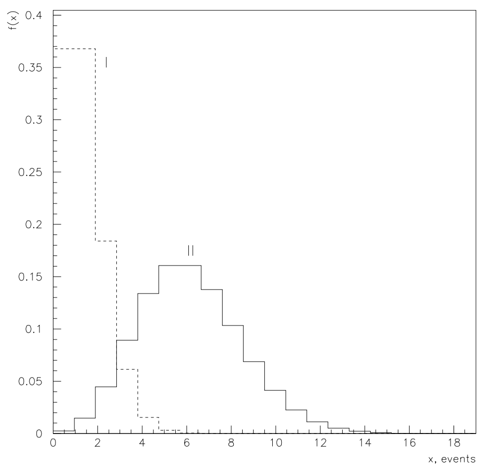

Figure 5: The probability distributions $f_0(x)$ (II) and $f_1(x)$ (I) of statistic s_1 for the case of 5 signal events and 1 background events.

One can point out that the values of α and β for S_1 and S_2 converge with increasing of a number of events. It means that, for a sufficiently large value of N_b, the values of α and β obtained by equal-tailed tests have a constant value close to 0.0062 for both S_1 and S_2. The standard deviation tends to be unity both for the distribution of s_1 (Fig.6) and for the distribution of s_2, i.e. these distributions in case of large N_b and N_s can be approximated by a standard Gaussian function $\mathcal{N}(0, 1)$ [2)] for a pure background and by a Gaussian function $\mathcal{N}(5, 1)$ for a signal mixed with a background. Therefore, the equal-tailed test for the normal distributions gives the critical value $c_1 = 2.5$ and $\alpha = \beta = 0.0062$. These are the limiting values of α and β for the requirement $S_1 = 5$, or $S_2 = 5$, or $S_{12} = 2.5$.

In a similar way we can determine the Type I and Type II errors for small values N_s and N_b and predict the limiting values of α and β for a large number of events for other statements about "significance" S_1 (Table 3) or any other estimator.

[2)] $\mathcal{N}$(mean,variance) is a traditional notation for normal distribution.

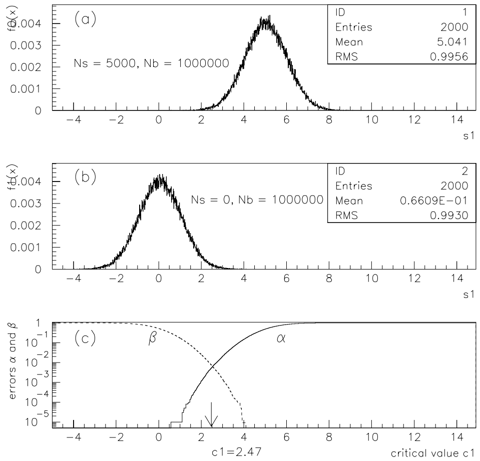

Figure 6: The probability distributions $f_0(x)$ (a) and $f_1(x)$ (b) of statistic s_1. The dependence of Type I and Type II errors on the critical value c_1 (c) for the case of 5000 signal events and 10^6 background events

5 Equal probabilities test

Last columns in Tables 1, 2 and 3 contain the value of probability κ [4] which is a characteristic of the observability of a phenomenon in future experiments with given N_s and N_b. In particular, it is the fraction of probability distribution $f_0(x)$ for a statistic x that can be described by the fluctuation of the background. The value of κ equals to the area of overlapping of probability distributions $f_0(x)$ and $f_1(x)$ (Fig.1). If we superimpose the distributions $f_0(x)$ and $f_1(x)$ and choose the intersection point ($N_{ev} = [\dfrac{N_s}{ln(1+\frac{N_s}{N_b})}]$) as a critical value for the hypotheses testing, we obtain $\kappa \equiv \alpha + \beta$. In this point $f_0(N_{ev}) = f_1(N_{ev})$ (in our case conditions $min(f_0(N_{ev}) - f_1(N_{ev}))$ and $f_1(N_{ev}) \leq f_0(N_{ev})$ are used). It means that this kind of check can be named the equal probabilities test. The dependences of κ on the number of signal events for the criteria $S_1 = 5,\ S_2 = 5$ and $S_{12} = 2.5$ are shown in Fig.7. Correspondingly, the dependences of N_b versus N_s for these criteria are presented in Fig.8. Note that the equal probabilities test can be applied for probability distributions with several points of intersection (Fig.9). The relative uncertainty of the observability of a new phenomenon in future experiment $\tilde{\kappa}$ is equal to $\dfrac{\kappa}{2-\kappa}$.

As is seen from Tables 1, 2 and 3, the value of κ is also close to the sum of $\alpha + \beta$ determined by using the equal-tailed test. Clearly, the accuracy of the determination of κ by Monte-Carlo calculations depends on the number of Monte-Carlo trials. Fig.10 shows the distribution of 40 estimations of the $\alpha + \beta$ for the case $N_s = 100$, $N_b = 500$ and for the 10^5 Monte-Carlo trials in each estimation. The result obtained by the direct calculation of the probability distributions is also given in the Fig.10.

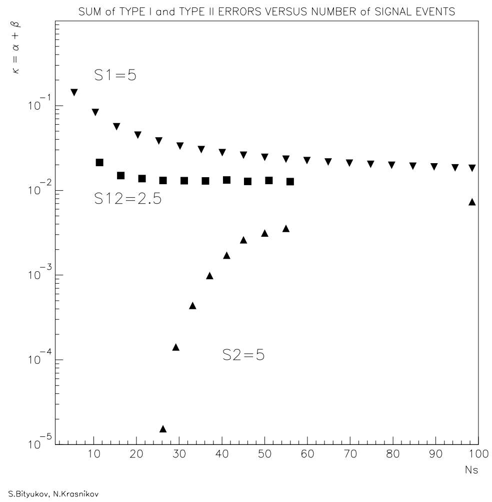

Figure 7: The dependences of κ on the number of signal events for “significances” $S_1 = 5,\ S_2 = 5$ and $S_{12} = 2.5$.

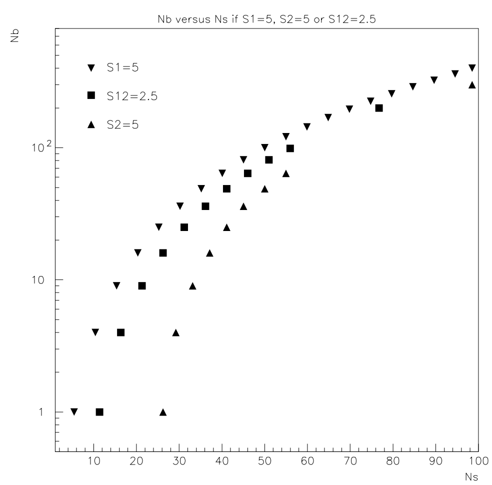

Figure 8: The dependences of the number of background events on the number of signal events for “significances” $S_1 = 5,\ S_2 = 5$ and $S_{12} = 2.5$.

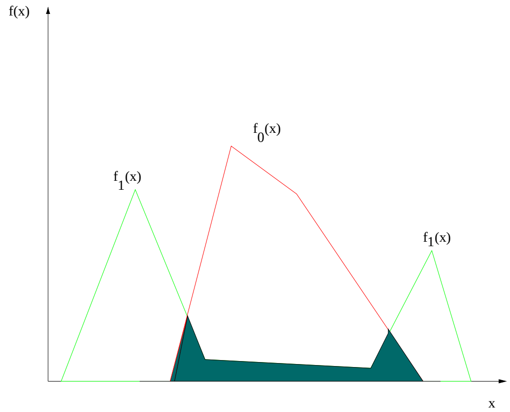

Figure 9: The estimation of uncertainty in hypotheses testing for arbitrary distributions by using of equal probabilities test.

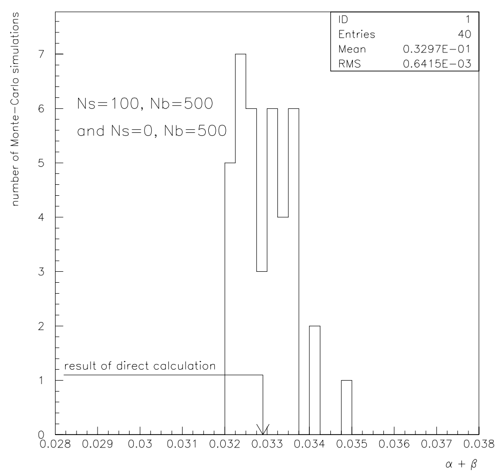

Figure 10: The variation of $\alpha + \beta$ in the equal-tailed hypotheses testing ($N_s = 100$, $N_b = 500$ versus $N_s = 0$, $N_b = 500$ in 40 Monte Carlo simulations of probability distributions).

6 The probability of new physics discovery

It is also very important to determine the probability of new physics discovery in future experiment. According to common definition (for example, [7, 8]) the new physics discovery corresponds to the case when the probability that background can imitate signal is less than 5σ or in terms of the probability less than $5.7 \cdot 10^{-7}$ (here of course we neglect any possible systematic errors).

So we require that the probability $\beta(\Delta)$ of the background fluctuations for $n > n_0(\Delta)$ is less than Δ, namely

$$\beta(\Delta) = \sum_{n=n_0(\Delta)+1}^{\infty} P(N_b, n) \leq \Delta \qquad (1)$$

The probability $1 - \alpha(\Delta)$ that the number of signal events will be bigger than $n_0(\Delta)$ is equal to

$$1 - \alpha(\Delta) = \sum_{n=n_0(\Delta)+1}^{\infty} P(N_b + N_s, n) \qquad (2)$$

It should be stressed that Δ is a given number and $\alpha(\Delta)$ is a function of Δ. Usually physicists claim the discovery of phenomenon [7, 8] if the probability of the background fluctuation is less than 5σ that corresponds to $\Delta_{dis} = 5.7 \cdot 10^{-7}$ [3]. So from the equation (4) we find $n_0(\Delta)$ and estimate the probability $1 - \alpha(\Delta)$ that an experiment will satisfy the discovery criterion.

As an example consider the search for standard Higgs boson with a mass $m_h = 110\ GeV$ at the CMS detector. For total luminosity $L = 3 \cdot 10^4 pb^{-1} (2 \cdot 10^4 pb^{-1})$ one can find [8] that $N_b = 2893(1929), N_s = 357(238)$, $S_1 = \dfrac{N_s}{\sqrt{N_b}} = 6.6(5.4)$. Using the formulae (4, 5) for $\Delta_{dis} = 5.7 \cdot 10^{-7}$ (5σ discovery criterion) we find that $1 - \alpha(\Delta_{dis}) = 0.96(0.73)$. It means that for total luminosity $L = 3 \cdot 10^4 pb^{-1} (2 \cdot 10^4 pb^{-1})$ the CMS experiment will discover at $\geq 5\sigma$ level standard Higgs boson with a mass $m_h = 110\ GeV$ with a probability 96(73) percent.

7 Conclusions

In this paper the discussion on the observation of new phenomena is restricted to the testing of simple hypotheses in the case of the predicted values N_s and N_b and the observable value x. As is stressed in [5], the precise hypothesis testing should not be done by forming a traditional confidence interval and simply checking whether or not the precise hypothesis is compatible with the confidence interval. A confidence interval is usually of considerable importance in determining where the unknown parameter is likely to be, given that the alternative hypothesis is true, but it is not useful in determining whether or not a precise null hypothesis is true.

To compare several criteria used for the hypotheses testing, we employ as a method that allows one to construct the rejection regions via the determination the probability distributions of these statistics by Monte-Carlo calculations,

[3] The approximation of Poisson distribution by Gaussian for tails with area close to or less than Δ_{dis} for values of N_s and N_b under consideration gives strong distinction in determination of $1 - \alpha$.

so direct calculations of probabilities distributions. The equal-tailed test was used to compare the criteria. An equal probabilities test is proposed to estimate the uncertainty in separation of two hypotheses about observability of predicted phenomenon in planned experiment. We also consider a probability of discovery as a quantity for comparison of proposals for future search experiments. The methods were used to draw a conclusion on the observability of some predicted phenomena [4, 9].

Acknowledgements

We would like to thank the Workshop co-convenors Fred James and Louis Lyons and local organizer Yves Perrin. We are greatly indebted to M.Dittmar for useful discussions which were one of the motivations to perform this study. We are grateful to V.Genchev, V.A.Matveev, V.F.Obraztsov, V.L.Solovianov and Yu.P.Gouz for the interest and valuable comments. This work has been supported by RFFI grants 99-02-16956 and 99-01-00091.

References

[1] See as an example,

V.Tisserand, *The Higgs to Two Photon Decay in the ATLAS Detector,* Talk given at the VI International Conference on Calorimetry in High Energy Physics, Frascati (Italy), June 8-14, 1996.

S.I.Bityukov and N.V.Krasnikov, *The Search for New Physics by the Measurement of the Four-jet Cross Section at LHC and TEVATRON,* Modern Physics Letters **A12**(1997)2011.

M.Dittmar and H.Dreiner, *LHC Higgs Search with* $l^+\nu l^-\bar{\nu}$ *final states*, CMS Note 97/083, October 1997.

[2] See as an example,

D.Denegri, L.Rurua and N.Stepanov, *Detection of Sleptons in CMS, Mass Reach,* CMS Note CMS TN/96-059, October 1996.

F.Charles, *Inclusive Search for Light Gravitino with the CMS Detector,* CMS Note 97/079, September 1997.

S.Abdullin, *Search for SUSY at LHC: Discovery and Inclusive Studies,* Presented at International Europhysics Conference on High Energy Physics, Jerusalem, Israel, August 19-26, 1997, CMS Conference Report 97/019, November 1997.

[3] S.I.Bityukov and N.V.Krasnikov, *The Search for Sleptons and Flavour Lepton Number Violation at LHC (CMS),* Physics of Atomic Nuclei, **62**(1999)1213; CMS Note 98/040, July, 1998.

[4] S.I.Bityukov and N.V.Krasnikov, *New Physics Discovery Potential in Future Experiments,* Modern Physics Letter **A13**(1998)3235, also physics/9811025, also S.I.Bityukov and N.V.Krasnikov, *Observability and Probability of Discovery in Future Experiments,* CMS IN 1999/027, also hep-ph/9908402.

[5] J.O.Berger, B.Boukai and Y.Wang, *Unified Frequentist and Bayesian Testing of a Precise Hypothesis,* Statistical Science **12**(1997)133.

[6] M.A.Stephens, EDF *statistics for goodness-of-fit and some comparisons,* J.Amer.Statist.Assoc., 1974, **69, N 347**, p.730.

T.E.Dielman and E.L.Rose, *A bootstrap approach to hypothesis testing in least absolute value regression,* Computational Statistics and Data Analysis, **20**(1995)119.

[7] J.J.Hernandez, S.Navas and P.Rebecchi, *Discovery limits in prospective studies,* Nucl.Instr.&Meth. A **372**, 1996, p.293.

[8] *The Compact Muon Solenoid.* Technical Proposal, CERN/LHCC 94-38, 1994.

[9] S.I.Bityukov and N.V.Krasnikov, *Search SUSY at LHC in jets +* E_T^{miss} *final states for the case of nonuniversal gaugino masses,* Phys.Lett., **B469**(1999)149.

Table 1: The dependence of α and β determined by using the equal-tailed test on N_s and N_b for $S_1 = 5$; κ is the area of intersection of probability distributions $f_0(x)$ and $f_1(x)$.

N_s	N_b	α	β	κ
5	1	0.0620	0.0803	0.1423
10	4	0.0316	0.0511	0.0828
15	9	0.0198	0.0415	0.0564
20	16	0.0141	0.0367	0.0448
25	25	0.0162	0.0225	0.0383
30	36	0.0125	0.0225	0.0333
35	49	0.0139	0.0164	0.0303
40	64	0.0114	0.0171	0.0278
45	81	0.0124	0.0136	0.0260
50	100	0.0106	0.0143	0.0245
55	121	0.0114	0.0120	0.0234
60	144	0.0100	0.0126	0.0224
65	169	0.0106	0.0109	0.0216
70	196	0.0095	0.0115	0.0209
75	225	0.0101	0.0102	0.0203
80	256	0.0091	0.0107	0.0198
85	289	0.0096	0.0097	0.0193
90	324	0.0088	0.0101	0.0189
95	361	0.0081	0.0106	0.0185
100	400	0.0086	0.0097	0.0182
150	900	0.0078	0.0084	0.0162
500	10^4	0.0068	0.0068	0.0136
5000	10^6	0.0062	0.0065	0.0125

Table 2: The dependence of α and β determined by using the equal-tailed test on N_s and N_b for $S_2 \approx 5$. Here κ is the area of intersection of probability distributions $f_0(x)$ and $f_1(x)$.

N_s	N_b	α	β	κ
26	1	$0.519 \cdot 10^{-5}$	$0.102 \cdot 10^{-4}$	$0.154 \cdot 10^{-4}$
29	4	$0.661 \cdot 10^{-4}$	$0.764 \cdot 10^{-4}$	$0.142 \cdot 10^{-3}$
33	9	$0.127 \cdot 10^{-3}$	$0.439 \cdot 10^{-3}$	$0.440 \cdot 10^{-3}$
37	16	$0.426 \cdot 10^{-3}$	$0.567 \cdot 10^{-3}$	$0.993 \cdot 10^{-3}$
41	25	$0.648 \cdot 10^{-3}$	$0.118 \cdot 10^{-2}$	$0.172 \cdot 10^{-2}$
45	36	$0.929 \cdot 10^{-3}$	$0.193 \cdot 10^{-2}$	$0.262 \cdot 10^{-2}$
50	49	$0.133 \cdot 10^{-2}$	$0.185 \cdot 10^{-2}$	$0.314 \cdot 10^{-2}$
55	64	$0.178 \cdot 10^{-2}$	$0.179 \cdot 10^{-2}$	$0.357 \cdot 10^{-2}$
100	300	$0.317 \cdot 10^{-2}$	$0.428 \cdot 10^{-2}$	$0.735 \cdot 10^{-2}$
150	750	$0.445 \cdot 10^{-2}$	$0.450 \cdot 10^{-2}$	$0.894 \cdot 10^{-2}$

Table 3: The dependence of α and β determined by using equal-tailed test on N_s and N_b for $S_1 = 2$, $S_1 = 3$, $S_1 = 4$, $S_1 = 6$ and $S_1 = 8$. Here κ is the area of intersection of probability distributions $f_0(x)$ and $f_1(x)$

S_1	N_s	N_b	α	β	κ
2	2	1	0.199	0.265	0.4634
	4	4	0.192	0.216	0.4061
	6	9	0.184	0.199	0.3817
	8	16	0.179	0.188	0.3680
	∞	∞	0.1587	0.1587	0.3174
3	3	1	0.0906	0.263	0.3184
	6	4	0.0687	0.216	0.2408
	9	9	0.0917	0.123	0.2159
	12	16	0.0722	0.131	0.1952
	∞	∞	0.0668	0.0668	0.1336
4	4	1	0.0400	0.263	0.2050
	8	4	0.0459	0.110	0.1406
	12	9	0.0424	0.0735	0.1130
	16	16	0.0407	0.0572	0.0977
	∞	∞	0.0228	0.0228	0.0456
6	6	1	0.0301	0.0806	0.1008
	12	4	0.0217	0.0217	0.0434
	18	9	0.0089	0.0224	0.0271
	24	16	0.00751	0.0132	0.0198
	∞	∞	0.00135	0.00135	0.0027
8	8	1	0.0061	0.0822	0.0402
	16	4	0.0049	0.0081	0.0131
	24	9	0.0016	0.0052	0.00567
	32	16	0.00128	0.00237	0.00331
	∞	∞	0.000032	0.000032	0.000064

Discussion

Louis Lyons: Would you mind explaining how you derived the formula for "significance" S_{12} ?

Sergei Bityukov: Let us draw two Poisson distributions with parameters $\mu_1 = N_b$ and $\mu_2 = N_s + N_b$. Let N_b be large enough to approximate these distributions by normal distributions $\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_2, \sigma_2)$, where $\sigma_1 = \sqrt{\mu_1}$ and $\sigma_2 = \sqrt{\mu_2}$. The transformation of the distributions to standard normal distribution (see Fig.11) and exploitation of the equalities

$$x_t = \frac{x_0 - N_b}{\sqrt{N_b}} = -\frac{x_0 - (N_s + N_b)}{\sqrt{N_s + N_b}}$$

allows one to find the points $x_0 = \sqrt{N_s + N_b}\sqrt{N_b}$ and, correspondingly, $x_t = \sqrt{N_s + N_b} - \sqrt{N_b}$. It allows us to use both the language of probability and the language of standard deviations. Note that in this approximation an equal-tailed test coincides with equal probabilities test.

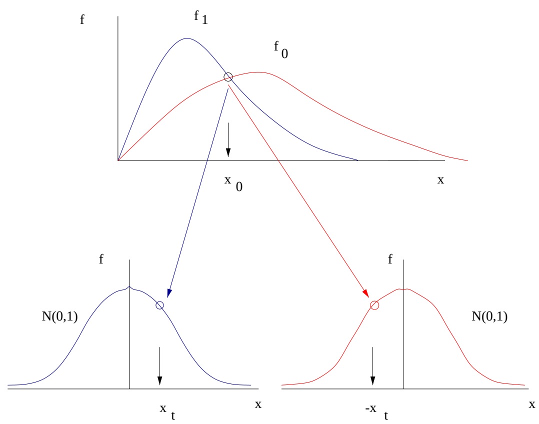

Figure 11: A sketch of transformation of Poisson probability distributions to standard normal probability density function.