On uncertainties associated with expected backgrounds in planned experiments

Sergey Bityukov*, Nikolai Krasnikov† and Vera Smirnova**

* *Institute for high energy physics, 142281 Protvino, Russia*
†*Institute for nuclear research RAS, Prospect 60-letiya Octyabrya, 7a, 117312 Moscow, Russia*
***Institute for high energy physics, 142281 Protvino, Russia*

Abstract. The expected numbers of events of several backgrounds in experiment are estimated from Monte Carlo experiments. In the analysis we take into account an integrated luminosity of Monte Carlo experiments. The expected number of events allows to construct the distribution of probabilities of number of events which in real experiment may be observed (in accordance with formulae in [1]). The formulae allow to take into account statistical uncertainty of corresponding Monte Carlo experiment. The influence of systematics is determined by additional Monte Carlo experiments with expected number of events.

Keywords: Uncertainty, Confidence distributions, Estimation
PACS: 02.50.Tt, 06.20.Dk, 07.05.Kf

INTRODUCTION

Often in experiments the signal and background events have the same signature. In this case the number of background events can be estimated from additional measurements and/or Monte Carlo calculations. It implies uncertainties associated with the estimation. The incorporation of uncertainties to calculations of expected backgrounds in experiment is an actual task.

We propose a method to estimate the probability that background fluctuates above the expected value of signal plus background. A short description of the method follows. The expected number of events of each background is estimated from Monte Carlo experiments. In the analysis we take into account the integrated luminosities of Monte Carlo experiments. The expected number of events allows construction of the distribution of probabilities of the number of events which in a real experiment may be observed (in accordance with formulae in papers [1]). These formulae take into account statistical uncertainty of the corresponding Monte Carlo experiment [2]. The influence of systematics is determined by additional Monte Carlo experiments with the expected number of events. The result is the distribution of probabilities to observe k background events $p(background\ in\ experiment = k),\ k = 0, 1, \ldots$. To estimate the significance of the excess of the expected sum of signal and background events above the number of pure background events, we use the significance S_{cP} [3, 4] (Z_N in astrophysics notation [5]).

In the remainder of the paper, the process of treatment of statistical and systematic [1]

[1] The notion of systematic uncertainties is used in high energy physics and astrophysics widely (see, for example, ref. [6]). The recommendation INC-1 by the Working Group on the Statement of Uncertainties

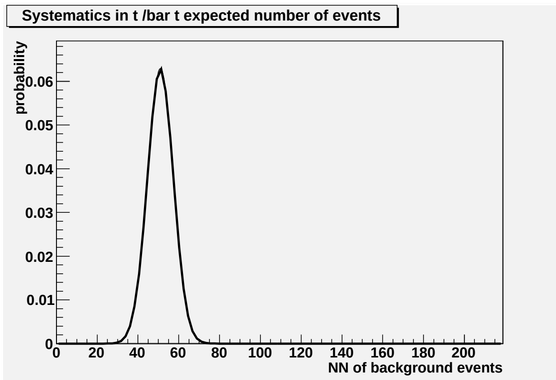

FIGURE 1. The probabilities of expected number of background events: $t\bar{t}$.

uncertainties in the estimation of backgrounds for one of the models of the experiment for searching of single top quark production [8] is presented.

THE METHOD

The procedure contains several steps.

1. For each background [2] $i,\ i = t\bar{t}, tW, W + jets, Wc, QCD, etc.$ we take into account the systematic uncertainties by performing Monte Carlo experiments. These uncertainties reflect the inaccuracy in the knowledge of N_{b_i} - the expected number of events of background i (see an approach of incorporating systematic uncertainties into an upper limit in paper [9]). The outcome of Monte Carlo experiments is the set of probabilities that N_{b_i} arise due to systematic uncertainty from one of numbers $\ldots, N_{b_i} - 1, N_{b_i}, N_{b_i} + 1, \ldots$ which we can consider as true value of expected number of background events, i.e. we change the constant value N_{b_i} to the distribution of possible values with different probabilities (see, Fig.1).
2. The second step is the inclusion of statistical uncertainty (which relates to the given integrated luminosity) and, correspondingly, the forecasting of the probabilities to see k background events in experiment. These probabilities are calculated for each value $\ldots, N_{b_i} - 1, N_{b_i}, N_{b_i} + 1, \ldots$ by the formula [1]:

$$p(background\ in\ experiment = k | M_b, m) = C^k_{M_b+k} \frac{m^{1+M_b}}{(m+1)^{1+M_b+k}}, \qquad (1)$$

(Bureau International des Poids et Mesures) [7] about category B (type B uncertainty) is internally inconsistent. The using of statistical formulae for nonstatistical evaluation, as seems, is not correct.

[2] We consider different uncorrelated sources of backgrounds which can imitate the production of single top quark: production of pair of top quarks $t\bar{t}$, associated production of W-boson with heavy quark, hadron jets or W/Z boson, QCD (Quantum ChromoDynamics) production of high-energy quarks or gluons and so on (see, [3]).

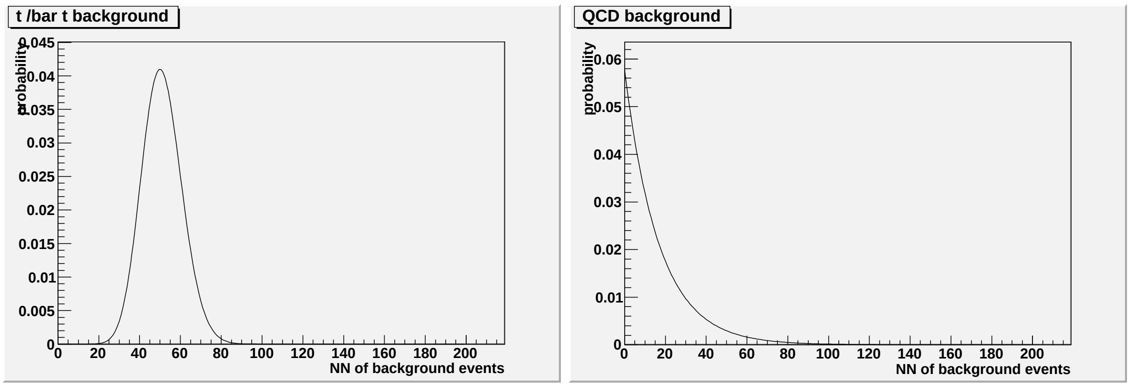

FIGURE 2. The expected probabilities of number of background events: $t\bar{t}$(left) and QCD (right).

where $k = 0, 1, \ldots$ is the possible number of background events in a real experiment, N_b is a possible expected number of background events in real experiment $(\ldots, N_{b_i} - 1, N_{b_i}, N_{b_i} + 1, \ldots)$, $M_b = m \cdot N_b$ is the expected number of background events from Monte Carlo experiment with integrated luminosity which is equal to m integrated luminosities of a real experiment and $C^k_{M_b+k}$ is $\dfrac{(M_b+k)!}{M_b!k!}$.
Then all distributions are summed with corresponding weights which are determined in the first step as probabilities. Note that this formula is applicable both for $m \geq 1$ and for $m < 1$. It allows Monte Carlo experiments to be taken into account, which give zero background under incomplete integrated luminosity. The second step is performed for each background. As an example, two backgrounds ($t\bar{t}$ and QCD) are presented in Fig.2. Distributions in Fig.2 show the summary influence of systematic uncertainties (step 1) and of statistical uncertainty (step 2) for background from $t\bar{t}$ production (left figure) and background from QCD processes (right figure).

3. The third step is the integration of all backgrounds $i,\ i = t\bar{t}, tW, W + jets$, etc. Each background is considered as independent from another background. For example, the integration of the pair of probability distributions of background 1 and background 2 can be written as

$$p(background\ in\ experiment = k | N_{b_1} + N_{b_2}) =$$

$$\sum_{j=0}^{k} p(background\ in\ experiment = j | N_{b_1}) \cdot p(background\ in\ experiment = k - j | N_{b_2}). \tag{2}$$

The combined backgrounds without QCD (left figure) and with QCD (right figure) are presented in Fig.3. These figures show distributions of probabilities to see k background events in the experiment from all background processes.

4. The final step is the determination of the significance of the excess of expected number of signal plus background events above pure background. We calculated the probability ε of the appearance of the number of background events above the

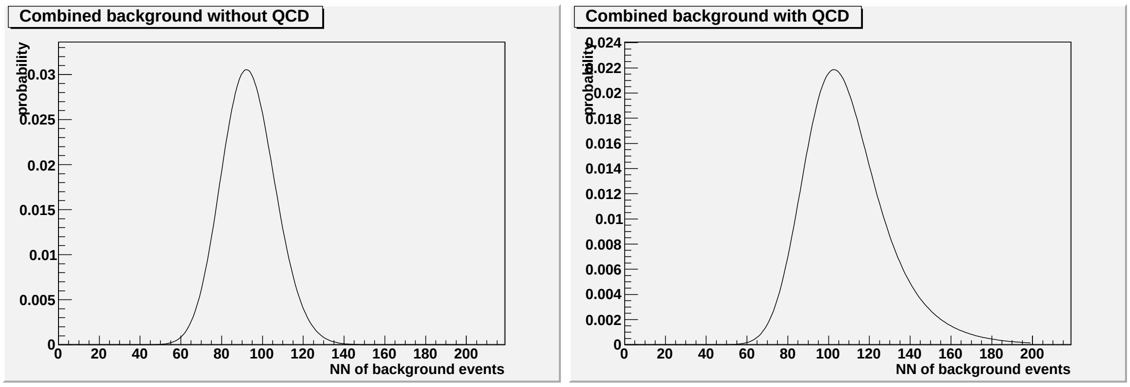

FIGURE 3. The combined background without QCD (left) and combined background with QCD (right).

expected number of signal plus background events in real experiment, i.e. the right tail above signal plus background in the distribution determined in step 3 (see, Fig.3). This probability was converted to the significance S_{cP}.

$$\int_{S_{cP}}^{\infty} \varphi(x)dx = \varepsilon \qquad (3)$$

where $\varepsilon = \sum_{k=expected\ signal+expected\ background}^{\infty} p(background\ in\ experiment = k)$, $\varphi(x)$ is the probability density of standard normal distribution and ε is the upper tail of probability distribution from the sum of expected signal and expected background.

NUMERICAL RESULTS

In the analysis we take into account the integrated luminosities of Monte Carlo experiments. For each background we determine the ratio m of integrated luminosities Monte Carlo experiment and real experiment by using the data from Table 1.

In calculations with equation (1) we use the number of Monte Carlo events from Table 2, namely, the value N_b in formula (1) is the value $N^{expected}$ in Table 2.

The influence of systematics (jets energy scale, missing transverse energy, etc.) is determined by additional Monte Carlo experiments using the expected number of events. We suppose that systematics have a normal distribution (Table 3). The reason of the using of this approach for the analysis is a relatively small influence of systematics on the value of significance in our case. In the table we present the impact of the considered systematic errors of instrumental origin on the final selection efficiencies. In principle, more complicated distributions for systematics are produced but in this analysis are not used due to the small effect on the final result (see Table 4).

TABLE 1. The ratio of integrated luminosities Monte Carlo experiment and real experiments (number of events before applying of cuts).

process	N_b^{exp}	N_b^{MC}	$m = N_b^{MC} / N_b^{exp}$
$t\bar{t}$	992	12324	12.42
VQQ	76	505	6.64
$W + jets$	229	260	1.14
tW	79	2589	32.77
Wc	60	432	7.20
WW	3	46	15.33
WZ	5	186	37.20
$Z + light\ partons$	6	9	1.5
QCD	587	36	0.06

TABLE 2. The number of expected events which survive a cut on Super Neutral Network output greater than 0.75 for 200 pb^{-1}.

process	$N^{expected}$ (N^{MC})
$t\bar{t}$	51 (635)
$Wb\bar{b}j$	6 (39)
W+jets	19 (21)
tW	6 (209)
Wc	10 (72)
WW	0 ± 0.13 (1)
WZ	0 ± 0.04 (9)
$Z + light\ partons$	0 ± 1.1 (0)
QCD	0 ± 29 (0)
Bkgd. total	92 (986)
t-channel Signal	91 (3081)

TABLE 3. Systematics which are taken into account for calculation of significance.

Process	total systematics
$t\bar{t}$	12.43%
$Wb\bar{b}j$	34.15%
$W + jets$	14.11%
tW	8.42%
Wc	10.05%

The expected number of signal events for the given integrated luminosity (200 pb^{-1}) equals 91. The values of significance S_{cP} are presented in Table 4.

As it follows from Table 4, the precise knowledge of systematics has a relatively small

TABLE 4. The dependence of the significance S_{cP} on systematics and the integrated luminosity of the Monte Carlo experiment. The third column shows the expected significance if we multiply the total systematic uncertainty by a factor of 1.5.

Backgrounds	systematics	systematics*1.5
All, except QCD	5.72	5.00
All, QCD m=0.06	2.46	2.41
All, QCD m=0.27	5.0	4.56

influence (in considered example) on the significance of the planned experiment using Neural Networks selection. The adequate statistics of the Monte Carlo experiment has a crucial influence on the expected significance of the real experiment.

CONCLUSION

In many searches for small signals, a significant limiting factor is the relative size and nature of systematic uncertainties on the measurement of background processes. The finite statistics of simulated samples which are used to predict the rates of different classes of events also represents a particular challenge in searches for small signals. It is therefore important to incorporate the statistical and systematic uncertainties into calculation of backgrounds.

We proposed a method for combined estimation of expected backgrounds with statistical and systematic uncertainties in planned experiment. The method has a clear probabilistic interpretation. We also show the applicability of the method for the planning of experiment for searching of single top quark production.

ACKNOWLEDGMENTS

We are grateful to Vassili Kachanov, Victor Matveev and Nikolai Tyurin for interest and support of this work. We thank Lev Dudko and Dmitry Konstantinov for helpful and constructive comments. We also thank Sergei Gleyzer, Yuri Gouz, Valery Zhukov, Alexandre Nikitenko and Claudia Wulz for very useful discussions. This work was supported by the Grants RFBR 08-02-91007-CERN and 10-02-00468. The participation of S.B. in the Workshop has been supported by the Organized Committee of MaxEnt'2010.

REFERENCES

1. S.I.Bityukov, *JHEP* **09** (2002) 060 [hep-ph/0207130];
   S.I. Bityukov, N.V. Krasnikov, *Nucl.Inst.&Meth.* **A502**, (2003) 795.
2. S.I. Bityukov et al., "Confidence distributions in statistical inference," in these Proceedings, *MaxEnt'2010*.

3. CMS Technical Design Report, vol.II: Physics Performance. The CMS Collaboration. *Journal of Physics G, Nuclear and Particle Physics, 34,* 995-1579 (2007).
4. S.I. Bityukov, S.E. Erofeeva, N.V. Krasnikov, A.N. Nikitenko, "Program for evaluation of the significance, confidence intervals and limits by direct calculation of probabilities", in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, edited by Louis Lyons and Muge Karagoz Unel, Conference Proceedings, Imperial College Press, 2006, pp. 106-107.
5. J.T. Linnemann, "Measures of significance in HEP and astrophysics", in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, edited by L. Lyons, R. Mount, R. Reitmeyer, Conference Proceedings, SLAC, Menlo Park, California, 2003, pp 35-40.
6. P.K. Sinervo, "Definition and Treatment of Systematic uncertainties in High Energy Physics and Astrophysics", in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, edited by L. Lyons, R. Mount, R. Reitmeyer, Conference Proceedings, SLAC, Menlo Park, California, 2003, pp 122-129.
7. http://physics.nist.gov/cuu/Uncertainty/international1.html
8. E. Boos et al., CMS AN-2009/180 (2009).
9. R.D. Cousins and V.L. Highland, *Nucl.Instr.*&*Meth.* **A320** 331-335 (1992).