LABORATORI NAZIONALI DI FRASCATI
SIS-Pubblicazioni

PEAK FINDING THROUGH SCAN STATISTICS

Francesco Terranova
INFN, Laboratori Nazionali di Frascati, P.O. Box 13, I-00044 Frascati, Italy

Abstract

We discuss the conditions under which Scan Statistics can be fruitfully implemented to signal a departure from the underlying probability model that describes the experimental data. It is shown that local perturbations ("bumps" or "excesses" of events) are better dealt within this framework and, in general, tests based on these statistics provide a powerful and unbiased alternative to the traditional techniques related with the χ^2 and Kolmogorov distributions. Approximate formulas for the computation of Scan Statistics in the range of interest for high energy and nuclear physics applications are also presented.

1 Introduction

In the last two decades, the properties of "scan statistics" [1,2] have been extensively investigated and this subject still represents a rapidly developing branch of applied probability. Nowadays, scan statistics are used in several areas of science and engineering to analyze the occurrence of cluster of events and to assess their statistical significance. Applications range from control theory to molecular biology while, at present, their use in physics is mainly limited to the analysis of time series, especially in x and γ-ray astronomy [3]. In fact, a common problem in nuclear and particle physics is to determine whether an observed cluster of events has occurred by chance or if it signals a departure from the underlying probability model ("null hypothesis") for the data. Peaks or local excesses of events can appear during energy scans at the colliders, in the distributions of kinematical variables as invariant masses or four-momentum transfer, in the interpretation of Dalitz plots, etc. The traditional test statistics that are employed to challenge the null hypothesis after a data taking can be divided into two broad families. Binned goodness-of-fit tests are connected to the χ^2 distribution. If the region where the excess is expected is known *a priori* because it has already been observed by another experiment, i.e. a confirmatory experiment against a claim of discovery has been performed, the level of agreement between the null hypothesis and the data can be evaluated in a straightforward manner through a Pearson χ^2 test [4]. For one-dimensional distributions of the variable x ($x \in [\mathcal{A}, \mathcal{B}]$), let the number of observations in the candidate signal region $[a, b]$ be k_{ab}. The best estimate of the background in the region $[a, b]$ is given by

$$\hat{B}_{ab} = \int_a^b B(x, \hat{\underline{\theta}}) \underline{x} \tag{1}$$

where $\underline{\theta}$ is the set of parameters describing the background distribution and $\hat{\underline{\theta}}$ is its best estimate based on events outside the interval $[a, b]$, so that $\mathrm{cov}(\mathrm{k_{ab}}, \hat{\mathrm{B}}_{\mathrm{ab}}) = 0$. Hence, the test variable can be defined as:

$$T_{ab} \equiv \frac{(k_{ab} - \hat{B}_{ab})^2}{V(k_{ab} - \hat{B}_{ab})} \tag{2}$$

$V(k_{ab} - \hat{B}_{ab})$ being the variance of $k_{ab} - \hat{B}_{ab}$. Under the null hypothesis $V(k_{ab}) = \hat{B}_{ab}$ and $V(k_{ab} - \hat{B}_{ab}) = \hat{B}_{ab} + \hat{\sigma}_{ab}^2$, i.e. the variance is the quadratic sum of the estimated background rate in the signal region and its error. Finally, if the error on the background is negligible

$$T_{ab} \simeq \frac{(k_{ab} - \hat{B}_{ab})^2}{\hat{B}_{ab}} \tag{3}$$

and, in the asymptotic limit, the test variable behaves as $\chi^2(1)$. Clearly, if the bin number N_{bin} and the bin size $w \equiv (\mathcal{B} - \mathcal{A})/\mathcal{N}_{\mathrm{bin}}$ are specified in advance but no information on the position, size and width of the signal are available, the corresponding test is given by

$$T \equiv \sum_{i=1}^{N_{bin}} \frac{(k_i - b_i)^2}{b_i} \tag{4}$$

which behaves as $\chi^2(N)$ in the asymptotic limit. The power[1] of this test depends on the position and width of the signal compared to the binning since a cluster shared among several bins becomes harder to be detected while clusters appearing in too large bins are swamped by background. If the size and position of the bin is chosen *after* the inspection of the data, the power of the test is increased but the estimate of the p-value[2] of the null hypothesis inferred from $\chi^2(N)$ is unreliable; hence, the significance level of the hypothesis test is biased. The binning problem can be overcome employing a second family of test statistics connected to the Kolmogorov distribution. The most common test is the Kolmogorov-Smirnov test [4] which corresponds to the largest distance between the cumulative distribution of the data and the one of the null hypothesis. This distance has a characteristic distribution that can be computed analytically and, hence, provides the p-value of the null hypothesis for a given data taking. These tests are well-suited to detect global distortions of the x distribution but have limited power for strong local deviation from the null hypothesis (cluster of events). Moreover, the power depends significantly on the position of the signal peak.

The procedure used to seek for event clusters *after* the data taking suggests a possible alternative to those methods, but which does not suffer from the problems of choosing the region *a posteriori*. The search is performed scanning the $[\mathcal{A}, \mathcal{B}]$ interval to identify the region where an anomalous accumulation of events appears. Given N events distributed along the $[\mathcal{A}, \mathcal{B}]$ range, we call $S(w)$ the largest number of events in a window of fixed length w. If the distribution of $S(w)$ is known, it will be possible to compute the probability $Prob(S(w) \geq k)$ for the null hypothesis to produce a cluster $S(w)$ greater or

---

[1]The power of an hypothesis test against a specific alternative hypothesis is the chance that the test correctly rejects the null hypothesis when that alternative hypothesis is true; that is, the power is 100% minus the chance of a Type II error when that alternative hypothesis is true.

[2]For a null hypothesis H_0 and a test statistic T (e.g. the one of Eq. (4)) we define $g(T|H_0)$ as the p.d.f. of T in the occurrence of H_0. The p-value is defined as the probability to find T in the region of equal or less compatibility with H_0 than the level of compatibility observed with the actual data [5]. For example, if T is defined such that large values correspond to poor agreement with the hypothesis, then the p-value will be

$$\int_{T_{obs}}^{+\infty} g(T|H_0)\,dT$$

T_{obs} being the value of the test statistic obtained in the actual experiment.

equal than the one actually observed. Hence, the p-value of the null hypothesis can be assessed. In this context, an *a priori* binning similar to the one of the Pearson χ^2 test is no more needed. Moreover, the test statistics $S(w)$ ("scan statistics") is not affected by the drawbacks of the Kolmogorov-Smirnov (K-S) tests (see Sec. 3). Sometimes this approach is followed, at least qualitatively, in literature. For instance, the OPAL [6] data accumulated at LEP during the high energy run beyond the Z^0 resonance (LEP2) were used to falsify the ALEPH [7] claim of a peak in the dijet invariant mass M of the $e^+ e^- \rightarrow$ four jets final state at $M \simeq 105$ GeV. Clearly, this refutation was carried out using the test statistics of Eq. (3). Moreover, to test for a peak in the dijet mass sum distribution for arbitrary mass M and independent of histogram binning, the positions of the mass windows were scanned over the full range of M. However, no quantitative statement was drawn due to the strong correlations of the contents of nearby bins. In fact, accounting for this correlation is possible once the properties of the scan statistics $S(w)$ are determined. In Sec. 2 these properties are discussed and the formulas to compute $Prob(S(w) \geq k)$ are presented. The power and significance of the test statistics $S(w)$ is computed in Sec. 3 and compared with the Pearson χ^2 and K-S approach for one-dimensional distributions. Extensions of the tests based on $S(w)$ and further applications in particle physics data analyses are discussed in Sec. 4.

2 Scan statistics

Let us consider an interval $[\mathcal{A}, \mathcal{B}]$ of a continuous variable x and a Poisson process ("background") whose mean value per unit interval is denoted with λ. Hence, the probability of finding $Y_x(w)$ events in an interval $[x, x + w]$ is

$$Prob(Y_x(w) = k) \;=\; e^{-\lambda w} \frac{(\lambda w)^k}{k!} \;; \quad k = 0, 1, 2, \ldots \tag{5}$$

The number of events in any disjoint non-overlapping intervals are independently distributed. We call "scan statistic" (SS) the largest number of events to be found in any subinterval of $[\mathcal{A}, \mathcal{B}]$ of length w [3], i.e.

$$S(w) \equiv \max_{\mathcal{A} \leq x \leq \mathcal{B} - w} \{Y_x(w)\} \tag{6}$$

The probability that the number of events in a scanning window never reaches k will be denoted, following [1], as

$$Q^*(k, \lambda\Delta, w/\Delta) \equiv 1 - Prob(S(w) \geq k) \tag{7}$$

---

[3]The case of non-uniform background can be dealt with by allowing for a window of variable width $w(x)$ that always contains $w/(\mathcal{B} - \mathcal{A})$ percent of the expected events under the null hypothesis [8].

where $\Delta \equiv \mathcal{B} - \mathcal{A}$ and the suffix "*" indicates that unconditional probabilities are considered, i.e. that the overall number of events N in the interval $[\mathcal{A}, \mathcal{B}]$ is not fixed but it fluctuates according to Eq. (5) with $w = \Delta$. The exact form of Eq. (7) can be expressed in terms of the sum of products of two determinants [9]. The summation is over the set V of all the partitions of N into $2H + 1$ non-negative integers m_i satisfying $m_i + m_{i+1} < k$ for $i = 1, \ldots, 2H$, where H is the largest integer in Δ/w. The determinants are computed starting from the $(H + 1) \times (H + 1)$ matrix $\{h\}_{ij}$ and the $H \times H$ matrix $\{v\}_{ij}$ whose entries are:

$$
\begin{aligned}
h_{ij} &= \sum_{s=2j-1}^{2i-1} m_s - (i - j)k \quad 1 \leq j \leq i \leq H + 1 \\
&= -\sum_{s=2i}^{2j-2} m_s + (j - i)k \quad 1 \leq i < j \leq H + 1 \\
v_{ij} &= \sum_{s=2j}^{2i} m_s - (i - j)k \quad 1 \leq j \leq i \leq H \\
&= -\sum_{s=2i+1}^{2j-1} m_s + (j - i)k \quad 1 \leq i < j \leq H
\end{aligned}
$$

Using these definitions for V, h_{ij} and v_{ij}, we have for $k \geq 2$ and $w < \Delta$:

$$
Q^*(k, \lambda\Delta, w/\Delta) = \sum_V R^* \det|1/\mathrm{h}_{ij}!| \det|1/\mathrm{v}_{ij}!| \tag{8}
$$

In formula (8)

$$
R^* = N! \, d^M \, (\frac{w}{\Delta} - d)^{N-M} \, p(N, \lambda\Delta) \tag{9}
$$

$$
M = \sum_{j=0}^{H} m_{2j+1} \tag{10}
$$

being $d \equiv 1 - wH/\Delta$ and $p(N, \lambda\Delta)$ is the Poisson probability of having N events from an average rate $\lambda\Delta$.

A very useful approximation of Eq. (8) has been derived by Naus in 1982 [10], based on the exact values of the probabilities $Q_2 \equiv Q^*(k, 2\psi, 1/2)$ and $Q_3 \equiv Q^*(k, 3\psi, 1/3))$ [4]. It can be shown that

$$
Q^*(k; \psi L, 1/L) \simeq Q_2^* [Q_3^*/Q_2^*]^{L-2} \tag{11}
$$

where

---

[4]For later convenience we define $\psi \equiv \lambda w$ and $L = \Delta/w$.

$$Q_2^* = [F(k-1,\psi)]^2 - (k-1)\,p(k,\psi)p(k-2,\psi)$$
$$- (k-1-\psi)\,p(k,\psi)F(k-3,\psi) \tag{12}$$
$$Q_3^* = [F(k-1,\psi)]^3 - A_1 + A_2 + A_3 - A_4 \tag{13}$$

and

$$A_1 = 2\,p(k,\psi)F(k-1,\psi)\left\{(k-1)\,F(k-2,\psi) - \psi F(k-3,\psi)\right\}$$
$$A_2 = 0.5\,[p(k,\psi)]^2\left\{(k-1)\,(k-2)\,F(k-3,\psi)\right.$$
$$\left. -2\,(k-2)\,\psi F(k-4,\psi) + \psi^2 F(k-5,\psi)\right\}$$
$$A_3 = \sum_{r=1}^{k-1} p(2k-r,\psi)\,[F(r-1,\psi)]^2$$
$$A_4 = \sum_{r=2}^{k-1} p(2k-r,\psi)p(r,\psi)\left\{(r-1)\,F(r-2,\psi) - \psi F(r-3,\psi)\right\}$$

In the above formulas $F(k,\psi)$ denotes the cumulative distribution

$$F(k,\psi) = \sum_{i=0}^{k} p(i,\psi) \ ; \ \ p(i,\psi) = e^{-\psi}\frac{\psi^i}{i!} \tag{14}$$

and $F(k,\psi) = 0$ for $k < 0$. For large values of Δ/w an even simpler approximation due to Alm [11] can be implemented:

$$Q^*(k,\lambda\Delta,w/\Delta) \simeq$$
$$F(k-1,\lambda w)\exp\left\{-\frac{k-w\lambda}{k}\lambda(\Delta-w)\,p(k-1,\lambda w)\right\} \tag{15}$$

Eq. (15) is often used in astrophysics applications and in many time series problems but it is of limited use in the present case where the condition $w/\Delta \ll 1$ is rarely fulfilled. In the following, the test statistics based on SS will be studied relying on the approximation (11). For a systematic comparison of the various approximations of Eq. (8) we refer to [1].

3 Power and significance for one-dimensional distributions

Sec. 2 dealt with the distribution of the scan statistics under the null hypothesis. The class of alternative hypotheses considered hereafter describe a local perturbation of the uniform distribution which leads to the appearance of a "excess" of events. Long-range distortions

like anomalous angular distributions are better dealt with global K-S tests and will not be further considered here. The alternative functions are Poisson processes of mean S. The signal events are spread along $[\mathcal{A}, \mathcal{B}]$ according to a normal distribution of mean x_S and sigma σ_S [5]. In the following we reject the null hypothesis if its p-value is smaller than 5%. The actual significance of the test statistics, i.e. the number of experiments where the null hypothesis was rejected albeit true, has been computed by Monte Carlo experimentation. Similarly the rate of Type II errors was computed to estimate the power of the test. In general, some prior assumptions are made before the inspection of a distribution. The domain $[\mathcal{A}, \mathcal{B}]$ of the variable x accessible to the experiment depends on the particular apparatus and, in most of the cases, it is selected *a priori*; so it is not expected to be a source of biases. In fact, in many applications the final distribution of x is the result of sequential cuts on other kinematic variables which can severely bias the sample [12]. For narrow resonances, whose width does not exceed the (known) instrumental resolution of the detector, the scanning window w of $S(w)$ can be fixed *a priori*. In particular, for a gaussian perturbation of variance σ_S^2 a nearly optimal choice of w is $w \simeq 4\sigma_S$ [13].

Fig. 1 shows the power of the K-S, SS and χ^2 tests as a function of the signal position x_S. Here, $[\mathcal{A}, \mathcal{B}] = [0, 1]$, $\sigma_S = 0.05$, $B \equiv \lambda\Delta = 100$ and $S = 20$. The optimal bin size for the χ^2 test has been computed following the prescription [14] $N_{\text{bin}} = 2(\lambda\Delta)^{2/5}$, where $\lambda\Delta$ is the expected sample size in case of null hypothesis[6]. In Fig. 1 signal events generated beyond the interval $[0, 1]$ are ignored (out of the sensitivity region $[\mathcal{A}, \mathcal{B}]$). The power averaged over the peak positions is shown in Fig. 2 as a function of S. A few comments are in order. As anticipated in Sec. 1 the K-S test is not appropriate for local perturbations. The power is limited compared to other statistics and depends on the peak position, having the highest sensitivity at the border of the distribution. The Pearson χ^2 test has a much higher power but in general the peak detection efficiency is reduced when the peak is shared between two adjacent bins. On average the χ^2 test underperforms w.r.t. SS since the correlations among the bins are ignored[7]. However, the bin prescription for χ^2 is independent of the *a priori* knowledge of σ_S while SS makes use of this additional information. This is a drawback for SS if the cluster width is broader than the instrumental resolution because the scanning window is no more optimized. Fig. 3 shows the average power versus σ_S assuming $w = \Delta/N_{\text{bin}}$ and $S = 20$. The vertical line corresponds to

---

[5]This is the case, for instance, of a narrow resonance whose intrinsic width is smaller than the detector resolution. For resonances broader than the instrumental precision a relativistic Breit-Wigner or a Jacobian-peak would be more appropriate. However, for the present purposes the details of the alternative function are not critical.

[6]Other choices of the binning for the χ^2 test, based on the knowledge of σ_S, have been tested by Monte Carlo experimentation. The corresponding powers do not exceed the one shown in Fig. 1.

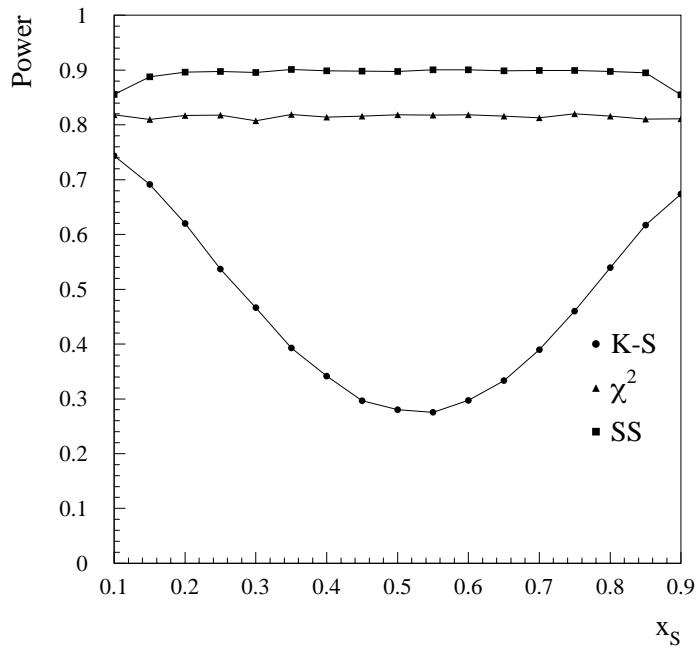[7]This is the reason why the χ^2 test and the Run Test [4] are complementary.

Figure 1: The power of the test statistics versus the peak positions for $S = 20$ and $B \equiv \lambda\Delta = 100$.
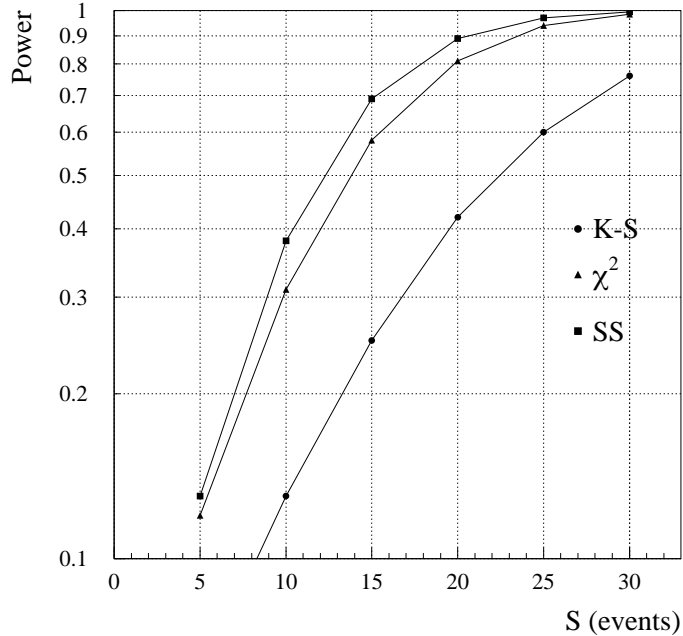
Figure 2: The power of the test statistics averaged over the peak positions versus the mean expected signal S.

$w = 4\sigma_S$. In fact, it is possible to compute the scan statistics for an *a posteriori* choice of w [15] but, clearly, this additional degree of freedom implies a deterioration of the power.

On the other hand, SS has a relevant feature which is not manifest in Figs. 1,2. The test statistics (4) behaves as $\chi^2(N)$ only in the asymptotic limit. This implies that the prescription $N_{\mathrm{bin}} = 2(\lambda\Delta)^{2/5}$ is appropriate only if the number of expected events per bin is such that the normal limit is justified. If this is not the case, the extraction of the p-value for the null hypothesis under the assumption $T \sim \chi^2(N)$ is biased and the proper behavior has to be restored increasing the bin size or computing the correct p-values by MC experimentation [16]. This fact is immaterial for SS, since the derivation of Eqs. (8) and (11) does not invoke the Central Limit theorem. The unbiaseness of the p-value for the null hypothesis even for few events expected in the scanning window has been checked by Monte Carlo experimentation. Fig. 4 shows the probability $Prob(S(w) \geq k) - Prob(S(w) \geq k + 1)$ of finding exactly k events after a scan, computed by Monte Carlo (crosses) and by Eq. (11). The upper plot shows the region with highest probability assuming $B = 10$, $S = 0$, $w = 0.2$; in this case the corresponding χ^2 test with optimal binning would have no more than 2 events per bin. The number of trials is 10^7 so the MC
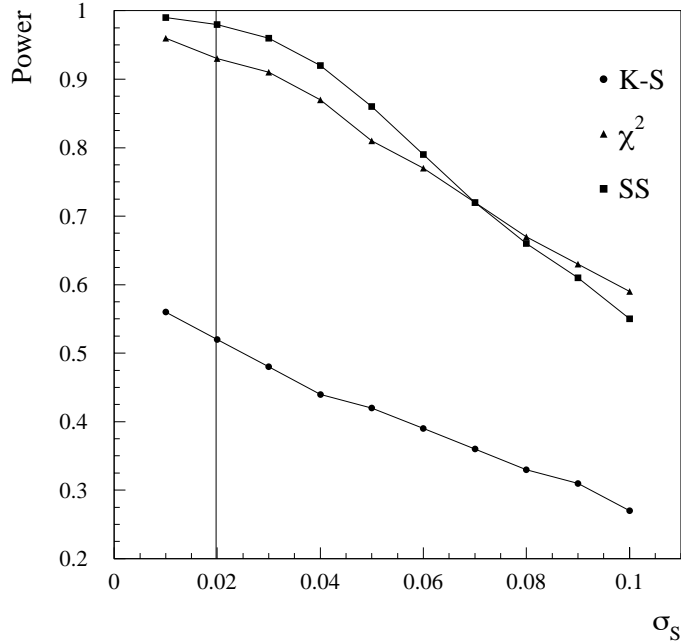
Figure 3: The power of the test statistics averaged over the peak positions versus the signal width σ_S for $B = 100$ and $S = 20$. For the Pearson χ^2 test, optimal binning N_{bin} is assumed; for SS $w = \Delta/N_{\mathrm{bin}} = 0.08$. The vertical line corresponds to $w = 4\sigma_S$.

error in the upper plot is negligible. The lower plot indicates the tail of the distribution. Note that the exact formula on which the Naus approximation is based holds for $k > 1$. Biases in the p-value will appear only when the approximation

$$Prob(S(w) = 0) \simeq 0 \implies Prob(S(w) = 1) \simeq 1 - Prob(S(w) \geq 2) \qquad (16)$$

does not hold, that is when the probability of having zero events after a full scan is non-negligible as in Fig. 5 where $\lambda\Delta = 2$ and the first empty dot indicates $Prob(S(w) = 0 \text{ or } S(w) = 1) = 1 - \mathrm{Prob}(S(w) \geq 2)$.

4   Extensions of the Scan Statistics

In Sec. 3 the alternative hypotheses (normal distributions centered at $x_S \in [\mathcal{A}, \mathcal{B}]$) were such that events generated outside the sensitivity region $[\mathcal{A}, \mathcal{B}]$ have been ignored. This implies a loss of power for the SS in the case of x_S lying near the border of the sensitivity region (see Fig. 1), i.e. when $x_S - \mathcal{A}$ or $\mathcal{B} - x_S$ is comparable or smaller than the
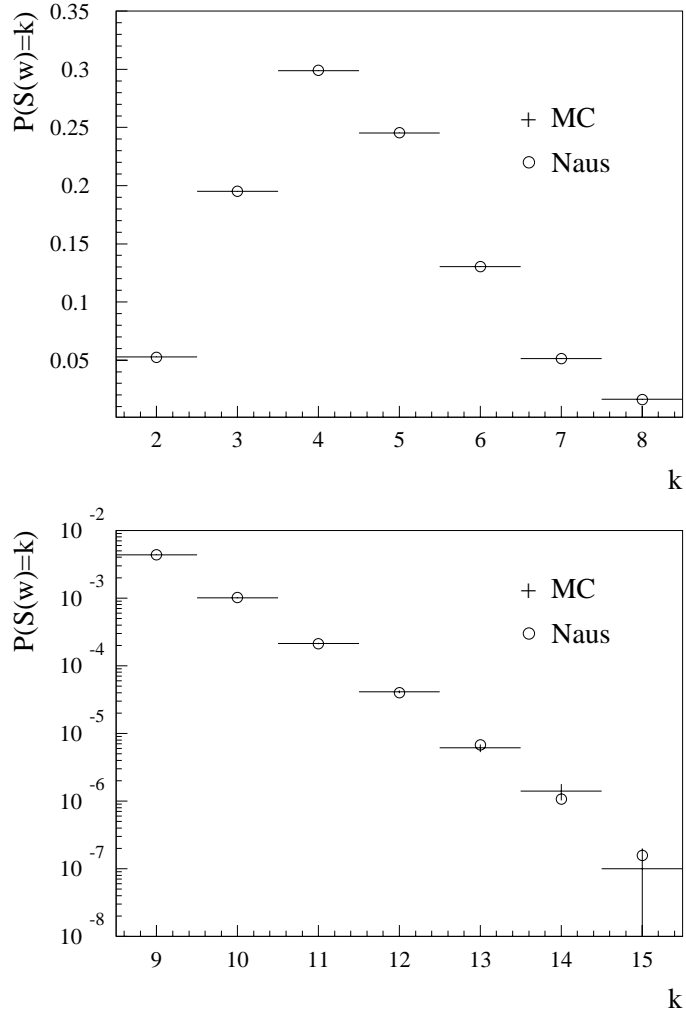
10

Figure 4: The probability of seeing k events after a scan with $B = 10$, $w = 0.2$ under the null hypothesis, computed by MC (crosses) and Eq. (11) (empty dots). The high-k tail of the distribution is shown in the lower plot.
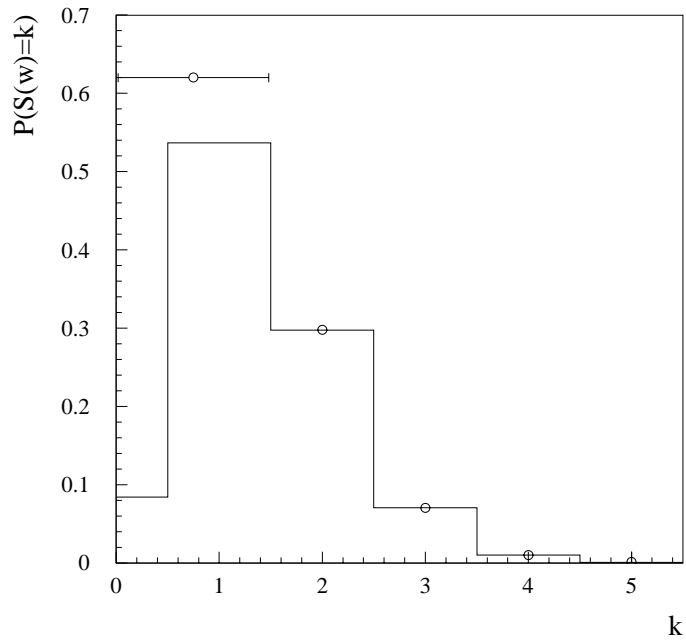
Figure 5: The probability of seeing k events after a scan with $B = 2$, $w = 0.2$ under the null hypothesis, computed by MC (crosses) and Eq. (11) (empty dots). The first empty dot indicates the probability of having $k < 2$ events after a full scan: $1 - Prob(S(w) = k \geq 2)$.

scanning window w. Clearly, these results can be extended in a straightforward manner to a bounded variable x on a range $[\mathcal{A}, \mathcal{B}]$ where the signal accumulates at the border. In this case the probability of finding a signal event between x and $x + dx$ is

$$Prob(\hat{x} \in [x, x+dx]) = \begin{cases} I^{(-\infty, \mathcal{A})} \, \delta(x - \mathcal{A}) & : & x = \mathcal{A} \\ G(x, x_S, \sigma_S) & : & x \in (\mathcal{A}, \mathcal{B}) \\ I^{(\mathcal{B}, \infty)} \, \delta(\mathcal{B} - x) & : & x = \mathcal{B} \end{cases} \tag{17}$$

and zero for x outside $[\mathcal{A}, \mathcal{B}]$; $G(x, x_S, \sigma_S)$ is the normal distribution with mean x_S and variance σ_S^2 and

$$I^{(a,b)} \equiv \int_a^b G(x, x_S, \sigma_S) \underset{\cdot}{x} \tag{18}$$

On the other hand, in many physics applications the alternative hypothesis (17) does not describe our signal expectation. For instance, if the bounded variable is connected with an angular distribution, a signal excess will manifest as a local perturbation of an uniform distribution of events along a unit circle. Event positions are described by the bounded variable $\theta \in [0, 1]$. In this case $S_c(w)$ is defined as the maximum number of points in any arc of length w and, following the notation of Sec. 2 with $\Delta = 1$, we have [10]:

$$Prob(S_c(w) \geq k) \simeq 1 - Q^*(k, 4\psi, 1/4) \frac{[Q^*(k, 3\psi, 1/3)]^{L-2}}{[Q^*(k, 2\psi, 1/2)]^{L-1}} \tag{19}$$

where $Q^*(k, 3\psi, 1/3)$ and $Q^*(k, 2\psi, 1/2)$ are given by Eqs. (12) and (13), and $Q^*(k, 4\psi, 1/4)$ can be derived from Eq. (8) with $L = 4$.

Note also that in Sec. 3 we considered the parameters describing the background known with high precision, so that it is possible to assume the null hypothesis to be fully specified. This is not the case if the parameters $\underline{\theta}$ of Eq. (1) have to be estimated *after* the data taking. In this case SS should be extended to devise the optimal estimate of the underlying background density $B(x, \underline{\theta})$ that is unbiased and consistent under both the null hypothesis and the occurrence of a local excess of width σ_S. This problem is still unsolved [1] for a generic function $B(x, \underline{\theta})$. Unbiased estimators have been obtained for simple functional dependences as in the case of the linear regression: for a discussion we refer to [1].

Finally, it is worth mentioning that, even if we focused on 1-dim distributions, SS has been extended to multivariate problems [17–19]. In particular, 2-dim applications are quite common e.g. in space analysis of arrival direction data of high energy cosmic rays, x and γ-ray bursts [3]. A description of multivariate unconditional scan statistics can be found in [1,18].

13

5 Conclusions

In this paper, we considered the conditions under which Scan Statistics can be implemented to signal a departure from the underlying probability model that describes the experimental data. In fact, local perturbations ("bumps" or "excesses" of events) are better dealt within this framework and, in general, tests based on $S(w)$ provide a powerful and unbiased alternative to the traditional techniques related with the χ^2 and Kolmogorov distributions. This holds in particular if the widths of the resonances are known *a priori*, e.g. when the event distribution is dominated by the instrumental resolution. Approximate formulas for the computation of SS in the range of interest for high energy and nuclear physics applications have been provided. Possible extensions to bounded variables and multivariate problems were also discussed.

6 Acknowledgements

References

[1] J. Glaz, J. Naus and S. Wallenstein, *Scan Statistics*, Springer, New York, 2001.

[2] J. Glaz and N. Balakrishnan, *Scan Statistics and applications*, Birkhauser, Boston, 1999.

[3] K.J. Orford, J. Phys. G26 (2000) R1.

[4] W.T. Eadie, D. Drijard, F. James, M. Roos, B. Sadoulet, *Statistical methods in experimental physics*, North-Holland, Amsterdam, 1971.

[5] K. Hagiwara et al., Phys. Rev. D66 (2002) 010001.

[6] K. Ackerstaff *et al.* [OPAL Collaboration], Phys. Lett. B429 (1998) 399.

[7] D. Buskulic *et al.* [ALEPH Collaboration], Z. Phys. C71 (1996) 179.

[8] M. Weinstock, Int. J. Epidem. 10 (1981) 289.

[9] R.J. Huntington and J. I. Naus, Ann. Prob. 3 (1975) 898.
S. Wallestein and J. I. Naus, Ann. Prob. 1 (1973) 188.

[10] J.I. Naus, J. Amer. Stat. Ass. 77 (1982) 177.

[11] S.E. Alm, *On the distribution of scan statistics in a Poisson process*, in Probability and mathematical statistics, A. Gut and L. Helst eds., Upsalla University Press, Upsalla, 1983.

[12] P.F. Harrison, J. Phys. G28 (2002) 2679.

[13] N. Cressie, J. App. Prob. 14 (1977) 272.

[14] D. Moore, *Test of Chi Squared type*, in Goodness of fit techniques, R.B. D'Agostino and M.A. Stephens eds., Dekker, New York, 1986.

[15] N. Nargawalla, Stat. Med. 15 (1996) 845.

[16] J. Heinrich, CDF note 6438 (2003).

[17] C. Loader, AT&T Bell Labs, Technical Memorandum 11214-901025-15TM.
C. Loader, Adv. Appl. Prob. 23 (1991) 751

[18] S.E. Alm, *Approximations of the distributions of scan statistics of Poisson processes*, in [2].

[19] P. Auer, K. Hornik and P. Revesz, Stat. Prob. Lett. 12 (1991) 91.