Scan Statistics in High Energy Physics

F. Terranova
I.N.F.N., Laboratori Nazionali di Frascati, Frascati (Rome), Italy

Scan Statistics are useful tools to signal a departure from the underlying probability model that describes the experimental data. They are commonly used in many research areas such as bioinformatics, control theory and medicine [1]; applications to astrophysics have also been suggested [2] . We consider, here, possible applications to high energy physics (HEP). It is shown that local perturbations ("bumps" of events or unexpected narrow resonances) 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. We also focus on the extensions needed to meet the challenges of particle physics problems and detail the differences between HEP and non-physics applications.

1. INTRODUCTION

Local perturbations of the expected distribution in a given kinematic variable can signal a departure from the underlying model used to describe the experimental data ("null hypothesis", NH). In high energy physics this phenomenon is related, for instance, to the appearance of an unexpected resonance and the determination of the statistical significance of the excess is a delicate task especially if aimed at claiming a discovery or planning a confirmatory data taking. Global distortions are commonly dealt with by the (unbinned) Kolmogorov-Smirnov test statistics or their extensions. However, the power[1] of these tests is drastically reduced for local perturbations. Conversely, the Pearson χ^2 test performs a binning of the range and compares the content of each bin (k_i) with its expectation under NH (b_i). If the parameters defining NH are known, the Pearson test statistic T

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

behaves as $\chi^2(N_{\rm bin})$ in the asymptotic limit. This test is better suited for local perturbations. Note, however, that the test is unbiased only if the binning grid is fixed *a priori* and the power depends strongly on the peak position. In general, local perturbations shared among different bins are tagged less effectively than peaks where the data cluster around one bin. In this respect, scanning the distribution with a running window of fixed length would be more appropriate and would avoid *a priori* fixing of the binning grid. This technique is sometimes employed in particle physics [3] but no quantitative estimates of the p-value for the null hypothesis are provided due to the strong correlation of the contents of nearby bins. In fact, the problem of the correlations can be solved analytically in the framework of Scan Statistics. Given N events distributed along the $[A, B]$ range, we call $S(w)$ ("scan statistic") the largest number of events in a window of fixed length w. If the distribution of $S(w)$ is known, it is possible to compute the probability $P(S(w) \geq k)$ for the null hypothesis to produce a cluster $S(w)$ greater or 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 statistic $S(w)$ is not affected by the drawbacks of the Kolmogorov-Smirnov (KS) tests.

[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. On the other hand, the significance level of a hypothesis test is a fixed probability of wrongly rejecting the null hypothesis, if it is in fact true.

2. THE PROPERTIES OF THE SCAN STATISTIC

For very simple cases, the computation of $P(S(w) \geq k)$ can be carried out through direct integration of the probability density function (p.d.f.). This is the case, for instance, of $P(S(w) \geq 2)$ when the events are $X_1 \ldots X_N$ independent and identically distributed random variables with common density $f(x) = 1$ for $x \in (0, 1)$ and zero elsewhere [1]. Here, the problem can be solved noting that $P(S(w) \geq 2) = P(W_2 \leq w)$, W_2 being the size of the smallest interval that contains 2 events. In fact

$$W_k = \min_{1 \leq i \leq N-1} \{X_{(i+1)} - X_{(i)}\} \tag{2}$$

where $X_{(i)}$ denote the *ordered* value of the X's and

$$P(W_2 > w) = P\left\{ \bigcap_{i=1}^{N-1} \left[X_{(i+1)} - X_{(i)} > w\right] \right\} \tag{3}$$

Thus, we are interested on the simultaneous occurrence of the conditions $X_{(i+1)} - X_{(i)} > w$ for all $i = 1, \ldots N-1$. $P(W_2 > w)$ results from the integration of the p.d.f. of the ordered distribution once the boundary of the integration is chosen in a way to fulfill $X_{(i+1)} - X_{(i)} > w$. If $w < 1/(N-1)$ we end up with the following integral:

$$P(W_2 > w) = N! \int_{(N-1)w}^{1} dx_N \int_{(N-2)w}^{x_N - w} dx_{N-1} \ldots \int_{w}^{x_3 - w} dx_2 \int_{0}^{x_2 - w} dx_1 = (1 - (N-1)w)^N \tag{4}$$

otherwise $P(W_2 > w) = 0$. Note that the joint p.d.f. of $X_{(1)} \ldots X_{(N)}$ is $N!$ times the original joint p.d.f. for $X_1 \ldots X_N$. This follows from the fact that the ordering function maps $X_1 \ldots X_N$ into $X_{(1)} \ldots X_{(N)}$, i.e. it represents a $N!$ to 1 transformation. The p.d.f. of the transformed variable is the sum of the contributions from each of these component mappings, hence it is $N!$ denser. Unfortunately, for values of k greater than 2 or 3 and smaller than N or $N-1$ this direct integration approach becomes overly complicated and a combinatorial approach based on the Karlin-McGregor theorem turns out to be more appropriate.

In fact, in the vast majority of HEP applications, events are produced randomly through a Poisson process and are characterized by a kinematic variable that is randomly distributed over an interval. Let us consider an interval $[A, B]$ 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 $P(Y_x(w) = k) = e^{-\lambda w}(\lambda w)^k/k!$ The number of events in any disjoint non-overlapping intervals are independently distributed. Again, the scan statistic is the largest number of events to be found in any subinterval of $[A, B]$ of length w. The formulas for fixed N can be extended to the Poisson case. Moreover in most of particle physics analysis a simple approximation by Naus [4] can be implemented. In this case $P(S(w) < k) \simeq Q_2 \left[Q_3/Q_2\right]^{\frac{\Delta}{w}-2}$ where $\Delta \equiv B - A$; Q_2 and Q_3 are functions of the Poisson probability $P(k, \lambda w)$ and their cumulative. Full analytic formulas for $P(S(w) < k)$, Q_2 and Q_3 can be found in [1, 5].

3. SEARCH FOR NARROW RESONANCES

From the discussion of Sec. 1, it is clear that the ideal situation to employ a goodness-of-fit test based on Scan Statistics (SS) is the search for narrow resonances along a 1-dim distribution of kinematic variables [2]. In this case the scanning width is fixed *a priori* and it corresponds to the expected width due to the finite detector resolution. To determine the power of the SS-based test we considered as alternative hypotheses local perturbations of the uniform distribution which leads to the appearance of a "excess" of events. The alternative functions are Poisson processes of mean S. The signal events are spread along $[A, B]$ according to a normal distribution of mean x_S and sigma σ_S. The significance of the test is at 95%. Fig. 1 shows the power of the KS, SS and χ^2 tests as a function of the signal position x_S. Here, $[A, B] = [0, 1]$, $\sigma_S = 0.05$, $w = 4\sigma_S$, $\lambda\Delta = 100$ and $S = 20$. The optimal bin size for the χ^2 test has been computed following the prescription [6] $N_{\rm bin} = 2(\lambda\Delta)^{2/5}$, where $\lambda\Delta$ is the expected sample size in case of null hypothesis. 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. Signal events generated beyond the interval $[0, 1]$ are ignored (out of the sensitivity region $[A, B]$).

As anticipated in Sec. 1 the KS 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 with respect to SS since the correlations among the bins are ignored. 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 actual width σ_S is not the same as the expected instrumental resolution sigma, because the scanning window is no more optimized. However, it can be shown [5] that within a large range of mismatch (up to a factor three) between

[2] E.g. the case of [3] where the resonance was sought for through the distribution of the dijet invariant mass sum.

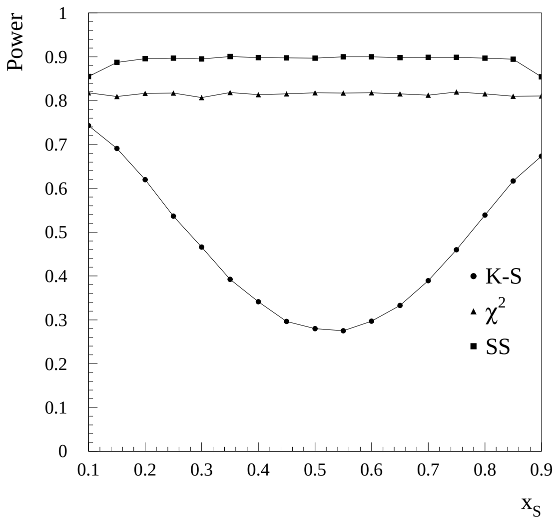

Figure 1: The power of the test statistic versus the peak positions for $S = 20$, $B = 100$, $\sigma_S = 0.05$ and $w = 4\sigma_S$.

the resonance width and the scanning window w the SS-based test still has higher power than Pearson χ^2. Moreover, the SS p-value tables turn out to be correct [5] even when very few events are observed. This is due to the fact that $P(S(w) < k)$ is derived invoking neither the Central Limit Theorem nor any normality approximation. Conversely, the corresponding tables for χ^2 hold in the asymptotic limit only and need to be recomputed by Monte Carlo experimentation if the number of events per bin is not sufficiently high. A more detailed comparison among the various tests can be found in [5].

4. NON UNIFORM BACKGROUND AND PARAMETER ESTIMATION

The assumption of uniform background in $[A, B]$ is rarely fulfilled in HEP. In most of the cases the expected events per unit interval is a function of the position along $[A, B]$: $\lambda = \lambda(x)$. The most straightforward extension of SS is due to Weinstock [7] and allows for stretching the window during the scan: the scan statistic is computed with a x-dependent window of variable width $w'(x)$ that always contains $w/(B - A)$ percent of the expected events under the null hypothesis. The advantage of this approach is that we expect the p-value tables for NH to be identical to the ones of the standard SS with width w and uniform background. On the other hand, the optimal pulse alternative for this test is no more the optimal pulse of $S(w)$ but the corresponding pulse after stretching of its domain. In general, this causes a non negligible loss of power only for strongly varying background (e.g. exponential decays). In those cases, more powerful generalizations of $S(w)$ exist [8] but the p-value tables have to be computed numerically.

The null hypothesis is often specified up to a set of free parameters which have to be extracted from the data. The concept of degree-of-freedom allows to correct the p-value of the Pearson χ^2 for NH keeping into account the fact that NH is defined using part of the experimental information. If b_i of Eq. 1 are functions of a set of M unknown parameters $\underline{\theta}$, the test statistic $T(k_i, b_i(\hat{\underline{\theta}}))$ behaves asymptotically as $\chi^2(N_{bin} - M)$, or $\chi^2(N_{bin} - M - 1)$ if the normalization is fixed. $\hat{\underline{\theta}}$ are the estimated parameters from the data through χ^2 minimization. Unfortunately, this concept cannot be implemented in a straightforward manner to SS. Here, SS should be extended to devise the optimal estimate of the underlying background density 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 for a generic function. Unbiased estimators have been obtained for simple functional dependencies as in the case of the linear regression [1]. In general, this implies an optimal splitting of the range $[A, B]$ between a "minimization region" used to draw $\hat{\underline{\theta}}$ and a complementary "scanning region" where the resonance is searched for. This technique allows to retain high power at least in the scanning region. Clearly, it is always possible to estimate $\hat{\underline{\theta}}$ using the whole range, determine naively $S(w', \hat{\underline{\theta}})$ instead of $S(w', \underline{\theta}_{true})$ and re-compute numerically the p-value tables (e.g. by MC). However, in general this approach results in a significant power loss.

5. CONCLUSIONS

Tests based on Scan Statistics are currently applied in several research areas to investigate local anomalies in time series of events. High energy physics offer many case studies where SS tests are expected to be powerful and easy to implement. The search for narrow resonances and the determination of the statistical relevance of local distortions in the distribution of the experimental data are two of them.

The Scan Statistics can be adapted to cope with common situations in particle physics such as the occurrence of non-uniform background (see Sec. 4) and scan over periodical angular variables [5]. On the other hand, their use become less straightforward when the scanning width is unknown *a priori* or the parameters of the null hypothesis must be estimated from the data. These cases were discussed in Sec. 3,4 and even in these conditions SS provide significant improvements with respect to traditional goodness-of-fit tests.

References

[1] J. Glaz, J. Naus and S. Wallenstein, “Scan Statistics”, Springer, New York, 2001.
[2] K.J. Orford, J. Phys. G26 (2000) R1.
[3] K. Ackerstaff et al. [OPAL Collaboration], Phys. Lett. B429 (1998) 399.
[4] J.I. Naus, J. Amer. Stat. Ass. 77 (1982) 177.
[5] F. Terranova, “Peak finding through Scan Statistics”, to appear in Nucl. Instr. Meth., arXiv:physics/0311020.
[6] D. Moore, in Goodness of fit techniques, R.B. D’Agostino and M.A. Stephens eds., Dekker, New York, 1986.
[7] M. Weinstock, Int. J. Epidem. 10 (1981) 289.
[8] M. Kulldorff et al., Stat. Med. 14 (1995) 799.